



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ
ΠΛΗΡΟΦΟΡΙΚΗΣ

Καινοτόμες Μεθοδολογίες Μηχανικής Μαθήσεως

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

του

Σωτηρίου Π. Χατζή

Αθήνα, Δεκέμβριος 2008



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ
ΠΛΗΡΟΦΟΡΙΚΗΣ

Καινοτόμες Μεθοδολογίες Μηχανικής Μαθήσεως

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

του

Σωτηρίου Π. Χατζή

Συμβουλευτική Επιτροπή:

Θ. Βαρβαρίγου, Ε. Ν. Πρωτονοτάριος, Μ. Θεολόγου

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την 1η Δεκεμβρίου 2008.

.....
Θ. Βαρβαρίγου
Καθ. Ε.Μ.Π.

.....
Γ. Στασινόπουλος
Καθ. Ε.Μ.Π.

.....
Μ. Θεολόγου
Καθ. Ε.Μ.Π.

.....
Β. Λούμος
Καθ. Ε.Μ.Π.

.....
Σ. Κόλλιας
Καθ. Ε.Μ.Π.

.....
Α. - Γ. Σταφυλοπάτης
Καθ. Ε.Μ.Π.

.....
Α. Δουλάμης
Επ. Καθ. Πολυτεχνείου Κρήτης

Αθήνα, Δεκέμβριος 2008

.....
Σωτήριος Π. Χατζής

Διδάκτωρ Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Σωτήριος Π. Χατζής, 2008.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

*Η διδακτορική μου διατριβή "Καινοτόμες Μεθοδολογίες Μηχανικής
Μαθήσεως" αφιερώνεται στους διαχρονικούς μέντορες των ποικίλων
εκφάνσεων της πνευματικής μου δημιουργικότητας ...*

Περίληψη

Στόχος της διδακτορικής αυτής διατριβής ήταν η ενδελεχής μελέτη των μεθοδολογιών μηχανικής μάθησης, και η χάραξη νέων δρόμων στον χώρο, με την εισαγωγή πρωτοτύπων μεθοδολογιών και καινοτόμων επαναστατικών θεωρήσεων αναγνώρισης προτύπων. Μεγάλη έμφαση εδώθη στις τεχνικές Variational Bayesian inference, που κατά την γνώμη του συγγραφέως αποτελούν το αύριο των μεθοδολογιών αναγνώρισης προτύπων των βασισμένων σε προσεγγίσεις statistical clustering, με συνεισφορά ενός πρωτότυπου μοντέλου εύρωστης αναγνώρισης προτύπων για πολυδιάστατα δεδομένα, καθώς και οι μεθοδολογίες fuzzy clustering. Σε αυτό τον τελευταίο χώρο εντοπίζεται και η μεγαλύτερη και σημαντικότερη συνεισφορά της παρούσης διατριβής, με την εισαγωγή μιας νέας θεώρησης του τί είναι fuzzy clustering, υπό την έννοια του τι εργασίες μηχανικής μάθησης μπορεί κανείς να περαιοώσει με χρήση fuzzy clustering, κατά την οποία ο αλγόριθμος FCM αναδεικνύεται σε μια πλεονεκτική εναλλακτική του EM αλγορίθμου (και λοιπών statistical clustering προσεγγίσεων) για την εκπαίδευση πολλών μορφών πιθανοτικών παραγωγικών μοντέλων. Πλέον αυτών, η εργασία αυτή παρήχε ακόμα ένα καινοτόμο αλγόριθμο hidden Markov models, προσφέρον εξαιρετικά πλεονεκτήματα σε ένα πολύ μεγάλο εύρος εφαρμογών σε σχέση με τις σημερινές τεχνικές, και τέλος, μια νέα μέθοδο ταυτοποίησης ομιλητή, στηριγμένη σε Gaussian process models.

Abstract

This thesis aimed at the extensive study of modern machine learning methodologies, and the provision of novel methodologies and groundbreaking regards towards pattern recognition techniques. A great deal of emphasis was put on Variational Bayesian inference techniques, which, to the author's mind, comprise the future of statistical clustering-based approaches, resulting in the provision of a novel model for robust recognition of high-dimensional patterns. Another key concept studied was fuzzy clustering. In this latter field can be found the most significant and seminal contribution of this thesis: we have introduced a novel regard towards fuzzy clustering, by showing that fuzzy clustering techniques, in the form of FCM algorithm variants, provide an interesting alternative to the EM algorithm (and other statistical clustering alternatives) for the treatment of probabilistic generative models. Furthermore, this work provided a novel hidden Markov model, offering significant benefits comparing to modern sequential data modeling techniques, as well as a new speaker verification methodology, based on Gaussian process classifiers.

Περιεχόμενα

1 Fuzzy clustering με χρήση factor analysis και κατανομών Student's- <i>t</i>	16
1.1 Προοίμιο	16
1.2 Εισαγωγή	18
1.3 Ανάλυση Παραγόντων κατά Student's- <i>t</i> (Factor Analysis on Student's- <i>t</i> distributions)	24
1.4 Το προτεινόμενο FMSFA model	30
1.4.1 Διατύπωση Μοντέλου	30
1.4.2 Εκτιμητική του FMSFA Model	39
1.4.3 Μεθοδολογίες Βέλτιστης Επιλογής Μοντέλου	45
1.5 Πειραματική Αξιολόγηση	46
1.5.1 Clustering Θορυβωδών Συνθετικών Δεδομένων	48
1.5.2 Αναγνώριση Χειρογράφων Ψηφίων	49
1.5.3 Σημασιολογική Αναγνώριση Αντικειμένων σε Κατα- τημένες Εικόνες	51
1.5.4 Εντοπισμός Spam E-Mail	52

1.5.5	Αξιολόγηση του Προσεγγιστικού Τύπου Υπολογισμού των Βαθμών Ελευθερίας	54
2	Μια Εύρωστη Μπεϋζιανή Τεχνική Μοντελοποίησης Πολυδιάστατων Σημάτων	65
2.1	Προοίμιο	65
2.2	Εισαγωγή	66
2.3	Διατύπωση Προβλήματος	69
2.3.1	Mixtures of Student's- t Factor Analyzers	69
2.3.2	To Variational Bayes Mixture of Student's- t Factor Analyzers	72
2.4	Variational Bayesian Inference	76
2.4.1	Variational Lower Bound	79
2.4.2	Variational Posteriors	80
2.4.3	Επιλογή Υπερπαραμέτρων	84
2.4.4	Επιλογή Μοντέλου και Χρήση της ARD Μεθόδου . .	86
2.4.5	Εκτίμηση της Κατανομής Πρόβλεψης	88
2.5	Εφαρμογές	89
2.5.1	Εύρωστη Επιλογή Μεγέθους Μοντέλου	90
2.5.2	Ταξινόμηση σημάτων ηλεκτροεγκεφαλογραφήματος	91
2.6	Συμπεράσματα	93
3	Μια Fuzzy Clustering Τύπου Προσέγγιση στα Hidden Markov Random Field Models	107
3.1	Προοίμιο	107
3.2	Εισαγωγή	108
3.3	Hidden Markov Random Field Models	112

3.4	Προτεινόμενη Προσέγγιση	117
3.4.1	Μια FCM-τύπου θεώρηση του HMRF model	117
3.4.2	Εκτιμητική των παραμέτρων του HMRF model με χρήση του HMRF-FCM αλγορίθμου	120
3.5	Πειραματική Μελέτη	124
3.5.1	Κατάτμηση θορυβωδών εικόνων	125
3.5.2	Κατάτμηση φυσικώς ληφθεισών εικόνων	128
3.5.3	Αξιολόγηση με το Berkeley image segmentation benchmark	128
3.5.4	Κατάτμηση πολυδιαστάτων εικόνων	135
4	Μια Καινοτόμος Μεθοδολογία Εύρωστης Μοντελοποίησης Ακολουθιακών Δεδομένων	145
4.1	Εισαγωγή	146
4.2	Διατύπωση Μοντέλου	149
4.2.1	Μοντέλα Πεπερασμένων Υπερθέσεων Κατανομών Student's- t (SMM)	149
4.2.2	Το Student's- t hidden Markov model (SHMM)	151
4.3	Επίλυση Μεγίστης Πιθανοφαιίας του SHMM Μοντέλου με Χρήση Πολλαπλών Ακολουθιών	153
4.4	Μια Τεχνική Μοντελοποίησης Συνδιακύμανσης για Χειραφέτηση του SHMM στην Περίπτωση Πολυδιάστατων Δεδομένων	163
4.4.1	Πεπερασμένες Υπερθέσεις Student's- t Αναλυτών Παραγόντων	165

4.4.2	Κρυφού Υποχώρου Μοντελοποίηση Συνδιακύμανσης για το SHMM Μοντέλο με Χρήση Ανάλυσης Παραγόντων	166
4.5	Εφαρμογές	173
4.5.1	Εύρωστη Ταξινόμηση Ακολουθιών με Χρήση του SHMM	175
4.5.2	Εύρωστη Κατάτμηση Ακολουθιών	177
4.5.3	Εύρωστη Ταξινόμηση Πολυδιάστατων Ακολουθιών με Χρήση του Factor Analyzed SHMM	182
4.6	Συμπεράσματα	186
5	Ταυτοποίηση Ομιλητή με Χρήση Gaussian Process Ταξινομητών	197
5.1	Προοίμιο	197
5.2	Εισαγωγή	199
5.3	Κυρίαρχες Τεχνικές Ταυτοποίησης Ομιλητή	202
5.3.1	Ανεξάρτητη Κειμένου Ταυτοποίηση με Χρήση GMM	202
5.3.2	Βάσει Κειμένου Ταυτοποίηση με Χρήση HMM	204
5.4	Δυναμική Ταξινόμηση με Χρήση Gaussian Processes	206
5.5	Ταυτοποίηση Ομιλητή με Χρήση Διακρινοσών Ακολουθίες Gaussian Processes	211
5.5.1	Ανεξάρτητη Κειμένου Ταυτοποίηση	215
5.5.2	Βάσει Κειμένου Ταυτοποίηση	217
5.6	Πειράματα	218
5.6.1	Ανεξάρτητη Κειμένου Ταυτοποίηση	218
5.6.2	Βάσει Κειμένου Ταυτοποίηση	221

Κατάλογος σχημάτων

1.1	Γραφική αναπαράσταση της μοντελοποίησης με factor analysis της εκατοστιαίας μεταβολής στις τιμές δέκα εταιρειών σε μια περίοδο 100 εβδομάδων.	25
1.2	Clustering Θορυβωδών Συνθετικών Δεδομένων: Περιγράμματα των λαμβανομένων clusters	50
3.1	(i) Αρχική 4-χρωμάτων εικόνα σκακιέρας. (ii) Διαβρωμένη δια θορύβου εικόνα. (iii) Κατάτμηση της εικόνας δια του αλγορίθμου FCM. (iv) Κατάτμηση της εικόνας δια του αλγορίθμου KL-FCM. (v) Κατάτμηση της εικόνας δια του αλγορίθμου ICM. (vi) Κατάτμηση της εικόνας δια του αλγορίθμου SF. (vii) Κατάτμηση της εικόνας δια του αλγορίθμου MF. (viii) Κατάτμηση της εικόνας δια του αλγορίθμου CA-SVFMM. (ix) Κατάτμηση της εικόνας δια του αλγορίθμου HMRF-FCM.	127

3.2	(i) Εικόνα σημαντήρα (buoy). (ii) Κατάτμηση της εικόνας δια του αλγορίθμου FCM. (iii) Κατάτμηση της εικόνας δια του αλγορίθμου KL-FCM. (iv) Κατάτμηση της εικόνας δια του αλγορίθμου ICM. (v) Κατάτμηση της εικόνας δια του αλγορίθμου SF. (vi) Κατάτμηση της εικόνας δια του αλγορίθμου MF. (vii) Κατάτμηση της εικόνας δια του αλγορίθμου CA-SVFMM. (viii) Κατάτμηση της εικόνας δια του αλγορίθμου HMRF-FCM.	129
3.3	(i) PET απεικόνιση του πνεύμονα ενός σκυλιού. (ii) Κατάτμηση της εικόνας δια του αλγορίθμου FCM. (iii) Κατάτμηση της εικόνας δια του αλγορίθμου KL-FCM. (iv) Κατάτμηση της εικόνας δια του αλγορίθμου ICM. (v) Κατάτμηση της εικόνας δια του αλγορίθμου SF. (vi) Κατάτμηση της εικόνας δια του αλγορίθμου MF. (vii) Κατάτμηση της εικόνας δια του αλγορίθμου CA-SVFMM. (viii) Κατάτμηση της εικόνας δια του αλγορίθμου HMRF-FCM.	130
3.4	Berkeley image segmentation data set and benchmark: Αρχικές Εικόνες	133
3.5	Berkeley image segmentation data set and benchmark: Κατατμήσεις Ληφθείσες από τον HMRF-FCM Αλγόριθμο.	134

3.6	Κατάτμηση RGB εικόνας: (i) Αρχική εικόνα. (ii) Κατάτμηση της εικόνας δια του αλγορίθμου FCM. (iii) Κατάτμηση της εικόνας δια του αλγορίθμου KL-FCM. (iv) Κατάτμηση της εικόνας δια του αλγορίθμου SF. (v) Κατάτμηση της εικόνας δια του αλγορίθμου MF. (vi) Κατάτμηση της εικόνας δια του αλγορίθμου CA-SVFMM. (vii) Κατάτμηση της εικόνας δια του αλγορίθμου HMRF-FCM.	137
3.7	Κατάτμηση θορυβώδους RGB εικόνας: (i) Αρχική εικόνα. (ii) Κατάτμηση της εικόνας δια του αλγορίθμου FCM. (iii) Κατάτμηση της εικόνας δια του αλγορίθμου KL-FCM. (iv) Κατάτμηση της εικόνας δια του αλγορίθμου SF. (v) Κατάτμηση της εικόνας δια του αλγορίθμου MF. (vi) Κατάτμηση της εικόνας δια του αλγορίθμου CA-SVFMM. (vii) Κατάτμηση της εικόνας δια του αλγορίθμου HMRF-FCM.	138
4.1	Γραφική αναπαράσταση του θεωρουμένου SHMM για T παρατηρήσεις, $\{\mathbf{y}_t\}_{t=1}^T$	155
4.2	Γραφική αναπαράσταση του factor analyzed SHMM για μια ακολουθία T παρατηρήσεων, $\{\mathbf{y}_t\}_{t=1}^T$	169
4.3	Βάσει Κειμένου Αναγνώριση Ομιλητή: Ρυθμοί σύγκλισης της μεθόδου GHMM για $J = 3$ και $J = 5$	178
4.4	Βάσει Κειμένου Αναγνώριση Ομιλητή: Ρυθμοί σύγκλισης της μεθόδου SHMM για $J = 3$ και $J = 5$	179

4.5	Αναγνώριση Νοητικής Κατάστασης με Χρήση Απεικόνισης Εγκεφάλου: Ρυθμοί σύγκλισης του αλγορίθμου εκπαίδευσης για τα μοντέλα (i)factor analyzed SHMM, (ii) διαγώνιο SHMM, και (iii) διαγώνιο GHMM.	184
5.1	Ανεξάρτητη Κειμένου Ταυτοποίηση: Ληφθήσες DET καμπύλες	220
5.2	Βάσει Κειμένου Ταυτοποίηση: Ληφθήσες DET καμπύλες . . .	223

Κατάλογος πινάκων

1.1	Clustering Θορυβωδών Συνθετικών Δεδομένων: Εκτιμήτριες των κέντρων των clusters	49
1.2	Αναγνώριση Χειρογράφων Ψηφίων: Ρυθμοί σφάλματος για βέλτιστη επιλογή των εκπαιδευομένων μοντέλων.	51
1.3	Σημσιολογική Αναγνώριση Αντικειμένων σε Κατατημένες Εικόνες: Ρυθμοί σφάλματος για βέλτιστη επιλογή των εκπαιδευομένων μοντέλων.	52
1.4	Εντοπισμός Spam E-Mail: Ρυθμοί σφάλματος για βέλτιστη επιλογή των εκπαιδευομένων μοντέλων.	53
2.1	Εύρωστη Επιλογή Μεγέθους Μοντέλου: Τιμές για διάφορα εύρη θορύβου	91
2.2	Ταξινόμηση HEΓ σημάτων: Ρυθμοί σφάλματος για βέλτιστη επιλογή μοντέλου	93
3.1	Berkeley Image Data Set: Ληφθέντες PR Δείκτες για τις Θεωρηθείσες Εικόνες	132
4.1	EM Αλγόριθμος για το SHMM	161
4.2	AECM αλγόριθμος για το factor analyzed SHMM	174

4.3	Αναγνώριση Χειρονομών: Ρυθμοί σφάλματος (%) για βέλτιστες τιμές J	176
4.4	Βάσει Κειμένου Αναγνώριση Ομιλητή: Ρυθμοί σφάλματος (%) των διαφόρων μεθόδων για βέλτιστη επιλογή του J	180
4.5	Κατάτμηση Ακολουθιών TIMIT: Ρυθμοί σφάλματος αναγνώρισης φθόγγων (%) για διάφορες τιμές του αριθμού συνιστωσών, J	180
4.6	Κατάτμηση Ακολουθιών TIMIT: Χρόνοι εκτέλεσης των αλγορίθμων εκπαίδευσης των αξιολογουμένων μοντέλων. . . .	180
4.7	Αναγνώριση Νοητικής Κατάστασης με Χρήση Απεικόνισης Εγκεφάλου: Μέσοι ρυθμοί σφάλματος και η τυπική τους απόκλιση (%) επί 30 εκτελέσεων των αλγορίθμων εκπαίδευσης. . . .	183
4.8	Αναγνώριση Νοητικής Κατάστασης με Χρήση Απεικόνισης Εγκεφάλου: Μέσοι χρόνοι εκτέλεσης των αλγορίθμων εκπαίδευσης των αξιολογουμένων μοντέλων.	183
5.1	Ανεξάρτητη Κειμένου Ταυτοποίηση	219
5.2	Βάσει Κειμένου Ταυτοποίηση	222

Κατάλογος Δημοσιεύσεων Συγγραφέως

- Sotirios Chatzis, Theodora Varvarigou, "Speaker Verification Using Sequence Discriminant Gaussian Process Classifiers," *IEEE Transactions on Audio, Speech, and Language Processing*, Accepted for Future Publication, 2008.
- Sotirios Chatzis, Theodora Varvarigou, "A Fuzzy Clustering Approach Towards Hidden Markov Random Field Models for Enhanced Spatially-Constrained Image Segmentation," *IEEE Transactions on Fuzzy Systems*, Accepted for Future Publication, 2008.
- Sotirios Chatzis, Dimitrios Kosmopoulos, Theodora Varvarigou, "Robust Sequential Data Modeling Using an Outlier Tolerant Hidden Markov Model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Accepted for Future Publication, 2008.
- Sotirios Chatzis, Theodora Varvarigou, "Factor Analysis Latent Subspace Modeling and Robust Fuzzy Clustering Using t Distributions," *IEEE Transactions on Fuzzy Systems*, Accepted for Future Publication, 2007.
- Sotirios Chatzis, Dimitrios Kosmopoulos, Theodora Varvarigou, "Signal Modeling and Classification Using a Robust Latent Space Model Based on t Distributions," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 949-963, March 2008.
- Sotirios Chatzis, Dimitrios Kosmopoulos, Theodora Varvarigou, "A Robust Approach Towards Sequential Data Modeling and its Application in Automatic Gesture Recognition," in proc. IEEE Int'l. Conference on Acoustics, Speech and Signal Processing 2008, pp. 1937-1940.

- Sotirios Chatzis, Theodora Varvarigou, “A Robust to Outliers Hidden Markov Model With Application in Text-Dependent Speaker Identification,” in proc. IEEE Int’l. Conference on Signal Processing and Communications, 2007.
- Sotirios Chatzis, Anastasios Doulamis, Theodora Varvarigou, “A Content-Based Image Retrieval Scheme Allowing for Robust Automatic Personalization,” in proc. ACM Int’l. Conference on Image and Video Retrieval 2007, pp. 1-8.

Κεφάλαιο 1

Fuzzy clustering με χρήση factor analysis και κατανομών Student's-*t*

1.1 Προοίμιο

Στην εργασία αυτή εισάγουμε ένα καινοτόμο υπόδειγμα ασαφούς συστοίχισης προτύπων (fuzzy clustering), υπό την θεώρηση ότι οι εξαγόμενες συστοιχίες ακολουθούν προκαθορισμένες κατανομές πιθανότητας. Ειδικότερα, εξετάζουμε την περίπτωση fuzzy clustering υπό την υπόθεση συστοιχιών της μορφής αναλυτών παραγόντων κατανεμημένων κατά Student's-*t* (Student's-*t* factor analyzers). Η θεμελίωση του πρωτοτύπου αυτού αλγορίθμου καθίσταται εφικτή δια της διατύπωσης στην παρούσα εργασία των αποτελεσματικών εκείνων εργαλείων που επιτρέπουν την εισαγωγή της υποθέσεως περί της κατανομής των μοντελοποιουμένων δεδομένων στην διαδικασία του fuzzy clustering. Σημειώνεται δε, πως παρότι στην παρούσα εργασία μελετάται ειδικά η περίπτωση του μοντέλου Student's-*t* factor analysis, η προτει-

νόμηση μεθοδολογία για την διεξαγωγή του fuzzy clustering υπό την a priori υπόθεση της πιθανοτικής κατανομής των εξαγομένων clusters είναι γενική, και ως τέτοια, μπορεί να εφαρμοστεί και για οιαδήποτε άλλη υπόθεση κατανομής.

Η ανάλυση παραγόντων είναι ένα μοντέλο κρυφού υποχώρου αρκετά δημοφιλές σε εφαρμογές τοπικού περιορισμού της διάστασης του χώρου παρατήρησης. Η κατανομή ενός factor analyzer ήθισται να λαμβάνεται ως κανονική (Gaussian). Ωστόσο, η χρήση της μακρύτερων άκρων κατανομής Student's-*t* έχει πρόσφατα αποδειχθεί ότι επιτρέπει την μεγαλύτερη ανθεκτικότητα των factor analyzers σε ατυπικά (θορυβώδη ή θορυβογενή) δεδομένα εν σχέσει με την Gaussian κατανομή. Έτσι, τα οφέλη από την χρήση του μοντέλου Student's-*t* factor analysis στα πλαίσια ενός υποδείγματος fuzzy clustering είναι τα ακόλουθα: Πρώτον, παρέχει μια καλά θεμελιωμένη μεθοδολογία μείωσης του χώρου παρατήρησης για αλγορίθμους fuzzy clustering, που επιτρέπει την ταυτόχρονη διενέργεια fuzzy clustering, και εντός κάθε εξαγομένου cluster, τοπικής μείωσης του χώρου παρατήρησης. Δεύτερον, εκμεταλλεύεται τα πλεονεκτήματα των Student's-*t* κατανομών αναφορικά με την ανθεκτικότητα τους σε ατυπικά δεδομένα ώστε να παρέχει ένα καλά θεμελιωμένο, μη ευρετικό υπόδειγμα για εύρωστο σε ατυπικά δεδομένα fuzzy clustering.

Η παρουσιαζόμενη κατωτέρω εργασία έχει δημοσιευθεί υπό το άρθρο μου:

Sotirios Chatzis, Theodora Varvarigou, "Factor Analysis Latent Subspace Modeling and Robust Fuzzy Clustering Using *t* Distributions," *IEEE Transactions on Fuzzy Systems*, Accepted for Future Publication, 2008.

1.2 Εισαγωγή

Η ανάλυση συστοιχιών (cluster analysis) είναι μια μεθοδολογία αναπτυχθείσα για την σύλληψη τοπικών δομών - προτύπων σε πολυμετάβλητα δεδομένα, δια της εφαρμογής ενός κριτηρίου συναφείας για την ομαδοποίηση των πολυμεταβλήτων σημείων - δεδομένων. Μια σημαντική έκφανση της ανάλυσης παραγόντων που βρίσκεται στο προσκήνιο, με αρκετή πολυπλοκότητα, είναι ο βαθμός της ακαμψίας (crispness) του επιθυμητού διαμερισμού των δεδομένων. Κάμπτοντας αυτό τον βαθμό ακαμψίας έδωσε στο παρελθόν ένα μεγάλο πεδίο έρευνας στο χώρο της cluster analysis. Πολλοί συγγραφείς έχουν προτείνει την ασαφή θεώρηση (fuzzy setting) ως την ενδεικνυόμενη λύση στο πρόβλημα. Η θεωρία ασαφών συνόλων (fuzzy set theory), προταθείσα από τον Zadeh [1] το 1965, παρέχει μια μορφή αβεβαιότητας του ανήκειν μιας παρατήρησης σε μια συγκεκριμένη συστοιχία δεδομένων (πρότυπο), περιγραφόμενης από μια συνάρτηση μετοχής (membership function). Πρώιμες εφαρμογές της θεωρίας ασαφών συνόλων στην cluster analysis περιλαμβάνουν τις εργασίες των Bellman et al. [2], και Ruspini [3]. Μια από τις πιο δημοφιλείς τεχνικές βάσει προτύπου (δηλ. εκ των προτέρων θεώρησης της μορφής των εξαγομένων clusters) fuzzy clustering είναι το fuzzy c -means (FCM) clustering. Ο FCM αλγόριθμος, αποτελών επέκταση του k -means algorithm, εισήχθη για πρώτη φορά από τον Dunn [4]. Ο Bezdek [5] παρείχε μια γενίκευση του FCM αλγορίθμου, εισάγοντας την έννοια του βαθμού ασαφείας (degree of fuzziness), λαμβάνοντας τιμές όχι μικρότερες από ένα, και παρέχοντας μια γενική μεθοδολογία εκτιμητικής με αποδεδειγμένη σύγκλιση. Οι Gustafson and Kessel [6] προέτειναν μια τροποποίηση του FCM αλγορίθμου με εφαρμογή μιας προσαρμοστικής μετρικής

απόστασης, για εντοπισμό clusters με διαφορετικά γεωμετρικά σχήματα. Οι Gath and Geva [7] όρισαν μια εκθετική μετρική απόστασης για τον FCM αλγόριθμο για να λάβουν μια νέα προσέγγιση υπό το ασαφές υπόδειγμα στην εκτιμητική μεγίστης πιθανοφανείας (maximum-likelihood estimation) των πεπερασμένων υπερθέσεων κανονικών κατανομών (Gaussian mixture models, GMMs) με χρήση του EM αλγορίθμου.

Πολλοί ερευνητές έχουν εκλάβει τον τρόπο με τον οποίο ο Bezdek εισήγαγε το degree of fuzziness στον k -means algorithm για να λάβει τον FCM algorithm, ως μια τεχνητή μεθόδευση, στερούμενη ισχυράς θεωρητικής θεμελίωσης. Ως μια λύση στο πρόβλημα αυτό, οι Li and Mukaidono [8] πρότειναν μια παραλλαγή του FCM algorithm εισάγοντας μια νέα προσέγγιση στο fuzzy clustering δια μέσου μιας τεχνικής συμπερασμού μεγίστης εντροπίας (maximum entropy inference, MEI). Οι Miyamoto and Mukaidono [9] αναδιαμόρφωσαν την MEI προσέγγιση θεωρώντας τον όρο εντροπίας ως ένα όρο συστηματοποίησης (regularization term). Υπό αυτή την θεώρηση, η ασάφεια αντιπροσωπεύεται από τον regularization term, υπό την μορφή ενός παράγοντα πολλαπλασιαζόμενου στην συνεισφορά του regularization function στο κριτήριο του clustering (αντικειμενική συνάρτηση). Οι Ichihashi et al. [10] προέτειναν μια περαιτέρω παραλλαγή του FCM με συστηματοποίηση μέσω της απόστασης Kullback-Leibler (KL) divergence αντί εντροπίας.

Οι αλγόριθμοι clustering τύπου fuzzy c -means (FCM) είναι στενά συνδεδεμένοι με τα GMMs στο αλγοριθμικό επίπεδο. Οι Gan et al. [11] έδειξαν ότι το GMM μπορεί να ερμηνευθεί ως ένα αθροιστικό ασαφές σύστημα. Οι Ichihashi et al. [10] έδειξαν ότι ο EM αλγόριθμος για το GMM μπορεί να εξαχθεί ως ειδική περίπτωση FCM-τύπου fuzzy clustering, δια της θεώρη-

σης της κανονικοποιημένης μέσω KL information ασαφούς αντικειμενικής συνάρτησης, και για μια κατάλληλη επιλογή της μετρικής απόστασης.

Παρά την μεγάλη τους δημοφιλία, οι FCM-τύπου clustering αλγόριθμοι είναι γνωστό ότι υποφέρουν από ένα σημαντικό μειονέκτημα, κοινό σε όλες τις GMM-τύπου ή GMM-σχετισμένες μεθοδολογίες clustering [12, 13]: η εκτιμητική των παραμέτρων των clusters δύναται να υποστεί σημαντική αρνητική επιρροή από την παρουσία ατυπικών δεδομένων στα δεδομένα εκπαίδευσης. Το πρόβλημα της προστασίας από ατυπικά δεδομένα σε προβλήματα με πολυδιάστατες παρατηρήσεις είναι γνωστό ως υψηλής δυσκολίας, αυξανόμενης με την διάσταση των δεδομένων [14]. Διάφοροι ερευνητές έχουν προσπαθήσει να προτείνουν λύσεις στον εν λόγω πρόβλημα. Μεταξύ αυτών, η noise clustering (NC) technique [15], που περιλαμβάνει την εισαγωγή ενός επιπλέον cluster για την αναπαράσταση των ατυπικών δεδομένων έχει καταστεί ένα από τα πιο δημοφιλή εργαλεία για την διαχείριση των ατυπικών δεδομένων (outliers) από τους fuzzy clustering αλγορίθμους (π.χ., [16--18]). Άλλες μέθοδοι πρόσφατα προταθείσες περιλαμβάνουν την εισαγωγή ενός είδους συνάρτησης απωλειών (loss function) στην ασαφή αντικειμενική συνάρτηση, ώστε να επιτρέψει τον εντοπισμό των outliers και την υποβάθμιση της συνεισφοράς τους στην εκτιμητική του μοντέλου (για μια εκτενή αναδρομή τέτοιων μεθόδων ανατρέξτε, π.χ., στο [13]). Ωστόσο, η ευρετική φύση των μεθόδων αυτών είναι ένα μεγάλο τους μειονέκτημα, καθώς η απόδοσή τους εξαρτάται από το ορθό προσδιορισμό των ευρετικών τους παραμέτρων. Αντιθέτως, αυτό που χρειάζεται είναι ένα μαθηματικό concept που να παρέχει μια γενικής εφαρμογής και αποδοτική τεχνική εντοπισμού και υποβάθμισης των outliers στα πλαίσια των αλγορίθμων

fuzzy clustering.

Οι πεπερασμένες υπερθέσεις κατανομών Student's- t (Student's- t mixtures models, SMMs) έχουν εσχάτως αναδειχθεί ως μια εναλλακτική στα GMMs παρέχουσα την αναγκαία ανθεκτικότητα σε outliers που λείπει από τα GMMs [19]. Η πολυμετάβλητη κατανομή Student's- t είναι μια εκθετική παραβολοειδής κατανομή με μακρύτερες ουρές και μια επιπλέον παράμετρο (τους αποκαλούμενους βαθμούς ελευθερίας) σε σχέση με την Gaussian, και τείνει στην Gaussian για μεγάλες τιμές των βαθμών ελευθερίας της. Τα SMMs έχει καταδειχθεί ότι μπορούν να προσεγγίζουν απείρως καλά άγνωστες τυχαίες κατανομές, περιλαμβανομένων και πολυκλάδων κατανομών, ενώ, παράλληλα, παρέχουν μια εύρωστη εναλλακτική στα GMMs, επιτρέπουν τον εντοπισμό και την υποβάθμιση των outliers κατά την εκπαίδευση του μοντέλου, μέσω μιας εγγενούς στο μοντέλο διαδικασίας, βασισμένης σε ένα μαθηματικά στέρεο στατιστικό υπόδειγμα [12].

Η ανάλυση παραγόντων (factor analysis) είναι ένα γραμμικό σχήμα κρυφών μεταβλητών που επίσης χρησιμοποιείται για τον εντοπισμό τοπικών δομών σε δεδομένα. Αντίστροφα όμως ως προς την cluster analysis, η factor analysis είναι τεχνική τοπικής μείωσης του χώρου παρατήρησης, μοντελοποιούσα συσχετίσεις μεταξύ των συνιστωσών πολυμεταβλήτων δεδομένων δια της εκφράσεως αυτών των συσχετίσεων σε ένα χαμηλών διαστάσεων υποχώρο, και άρα, μειώνουσα των χώρο παρατήρησης από ένα μεγαλύτερο αριθμό παρατηρήσιμων μεταβλητών σε ένα μικρότερο αριθμό από μη-παρατηρήσιμες μεταβλητές καλούμενες ως *παράγοντες*. Η factor analysis μοντελοποιεί κάθε παρατήρηση διαμερίζοντάς την σε ένα μη-παρατηρήσιμο τμήμα, λαμβανόμενο ως ο γραμμικός συνδυασμός των αντιστοίχων της πα-

ραγόντων, και ένα μη-παρατηρήσιμο τμήμα σφάλματος, του οποίου οι συνιστώσες λαμβάνονται ως ασυσχέτιστες. Μια περαιτέρω υπόθεση του μοντέλου factor analysis είναι ότι τα διανύσματα των παραγόντων και του σφάλματος ακολουθούν Gaussian κατανομές. Η factor analysis είναι στενά συνδεδεμένη με τη τεχνική principal component analysis (PCA) [20]. Ωστόσο, οι μέθοδοι αυτές έχουν μεταξύ τους πολλές θεωρητικές και αλγοριθμικές διαφορές, με την σπουδαιότερη να είναι ότι οι παράγοντες (factors) που δίνει η factor analysis δεν ορίζονται μονοσήμαντα, αλλά μόνο μέχρι τον τελεστή περιστροφής, εν αντιθέσει με την μοναδικότητα των principal vectors που παρέχει το PCA [21].

Μια πεπερασμένη υπέρθεση από factor analyzers (mixture of factor analyzers model, MFA) παρέχει μια γενική, μη-γραμμική προσέγγιση για την αναπαράσταση μιας πολυμετάβλητης παρατήρησης σε ένα ολιγοδιάστατο υποχώρο, και λαμβάνεται με την θεώρηση ενός αριθμού υπερτιθέμενων factor analysis μοντέλων (γραμμικά υπομοντέλα) για την αναπαράσταση της κατανομής του πλήρους διανύσματος παρατήρησης [22]. Το μοντέλο MFA είναι στην πραγματικότητα μια ειδική περίπτωση του GMM με περιορισμένο αριθμό παραμέτρων, που επιτρέπει την μείωση των βαθμών ελευθερίας των πινάκων συνδιακύμανσης των συνιστωσών το μοντέλο κατανομών, με ταυτόχρονη διατήρηση της επίδοσής του GMM σε εφαρμογές αναγνώρισης προτύπων. Έτσι, το μοντέλο MFA παρέχει την δυνατότητα επαρκούς μοντελοποίησης εξαιρετικά πολυδιαστάτων παρατηρήσεων, ενώ επιτρέπει παράλληλα την διεξαγωγή clustering και, εντός κάθε cluster, τοπικής μείωσης διάστασης (local dimensionality reduction), αποδίδοντας πολλά πλεονεκτήματα σε σχέση με ένα μοντέλο όπου clustering και local dimensionality reduc-

tion θα γίνονταν ανεξάρτητα η μια από την άλλη [22].

Μια πρώτη local dimensionality reduction τεχνική για fuzzy clustering αλγορίθμους ήταν ο fuzzy c -varieties (FCV) αλγόριθμος [23]. Ο FCV clustering αλγόριθμος μπορεί να θεωρηθεί ως μια τεχνική ταυτόχρονου fuzzy clustering και PCA, όπου τα πρότυπα είναι πολυμετάβλητες γραμμικές ποσότητες που αναπαρίστανται από μερικά διανύσματα local principal component. Στο [24], ο KLFCV αλγόριθμος έχει προταθεί ως μια τροποποίηση του FCV μέσω συστηματοποίησης με KL information, περιλαμβάνων τον FCV αλγόριθμο ως μια ειδική υποπερίπτωσή του.

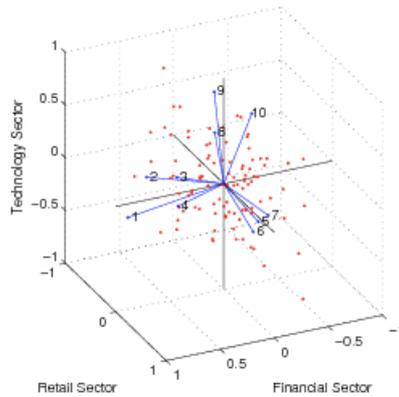
Στην παρούσα εργασία, προτείνουμε μια καινοτόμο προσέγγιση στο fuzzy clustering, που εν τέλει συνδυάζει τα πλεονεκτήματα του local dimensionality reduction με χρήση factor analysis και την ευρωστία σε outliers των κατανομών Student's- t . Η ανάλυσή μας ξεκινάει από την υπόθεση ότι τα πρότυπα που ακολουθούν τα δεδομένα υπό μελέτη μπορούν να αναπαρασταθούν με χρήση μιας πεπερασμένης υπέρθεσης factor analyzers, κατανεμημένων όχι κατά Gauss, ως ήθισται, αλλά κατά Student's- t . Αυτή είναι μια πρόσφατα προταθείσα τροποποίηση του μοντέλου MFA model, που επιτρέπει την εκμετάλλευση των προτερημάτων (αναφορικά με τα outliers) των κατανομών Student's- t [25,26]. Επί τη βάση της υποθέσεως αυτής, προτείνουμε μια γενική και θεωρητικά στέρεα μέθοδο για την εισαγωγή της θεωρηθείσης πληροφορίας σχετικά με τις κατανομές των προτύπων (clusters) των μοντελοποιουμένων δεδομένων στην διαδικασία του fuzzy clustering, την οποία και εφαρμόζουμε υπό το υπόδειγμα της παραλλαγής του FCM algorithm με την χρήση συστηματοποίησης μέσω KL information [10]. Έτσι, συνάγουμε έναν αλγόριθμο που ταυτόχρονα εκτελεί clustering and, και εντός κάθε cluster,

local dimensionality reduction με χρήση factor analysis. Επιπλέον, η αποτελεσματική εισαγωγή της πληροφορίας για την κατανομή που ακολουθούν τα εξακτέα clusters στην διαδικασία του fuzzy clustering αναμένεται να παρέχει σημαντική βελτίωση στην επίδοση του clustering, η οποία επαυξάνεται περαιτέρω λόγω της ευρωστίας της κατανομής Student's- t στα outliers, όπως περιεγράψαμε προηγουμένως.

Το υπόλοιπο του παρόντος κεφαλαίου έχει ως ακολούθως: Στην ενότητα 1.3 θα παρέχουμε μια σύντομη ανασκόπηση του μοντέλου factor analysis με χρήση Student's- t κατανομών. Η ενότητα 1.4 ξεκινάει με την εισαγωγή του προτεινόμενου μοντέλου fuzzy mixture of Student's- t factor analyzers (FMSFA), ως μιας υβριδικής προσέγγισης στην μοντελοποίηση υποχώρου με χρήση factor analysis υπό το πρίσμα της ασαφούς μοντελοποίησης, και με την θεώρηση δεδομένων κατανεμημένων κατά Student's- t . Περαιτέρω, εξάγομε ένα FCM-τύπου αλγόριθμο με συστηματοποίηση μέσω KL information για την εκτιμητική του FMSFA μοντέλου δοθέντος ενός συνόλου εκπαίδευσης. Τέλος, στην ενότητα 1.5, επιδεικνύουμε πειραματικά την απόδοση του μοντέλου μας.

1.3 Ανάλυση Παραγόντων κατά Student's- t (Factor Analysis on Student's- t distributions)

Η ανάλυση παραγόντων (factor analysis) είναι ένα γραμμικό μοντέλο υποχώρου για την σύλληψη τοπικών υποπροτύπων σε μοντελοποιούμενα δεδομένα. Έστω x_1, \dots, x_n ένα p -διάστατο τυχαίο δείγμα μεγέθους n . Η factor



Σχήμα 1.1: Γραφική αναπαράσταση της μοντελοποίησης με factor analysis της εκατοστιαίας μεταβολής στις τιμές δέκα εταιρειών σε μια περίοδο 100 εβδομάδων.

analysis μοντελοποιεί τις παρατηρήσιμες μεταβλητές, x_j , ως

$$x_j = \mu + \Lambda y_j + e_j \quad (1.1)$$

όπου y_j είναι ένα q -διάστατο ($q < p$) διάνυσμα από μη παρατηρήσιμες μεταβλητές, καλούμενες ως παράγοντες που αντιστοιχούν στην j οστη παρατήρηση, μ είναι ο μέσος των παρατηρήσεων x_j , Λ είναι ο $p \times q$ πίνακας των *factor loadings* (παράμετροι του μοντέλου), και e_j είναι το σφάλμα του μοντέλου που αντιστοιχεί στην j οστη παρατήρηση. Θεωρούμε ότι οι $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ είναι ανεξάρτητες και ισόνομες.

Στο Σχ. 1, παρέχουμε μια γραφική αναπαράσταση της μοντελοποίησης με factor analysis. Το θεωρούμενο σύνολο δεδομένων περιλαμβάνει την εκατοστιαία μεταβολή στις τιμές των μετοχών δέκα εταιρειών σε μια περίοδο 100 εβδομάδων. Έτσι, το σύνολο δεδομένων μας περιλαμβάνει 100 10-διάστατα δεδομένα, όπου κάθε διάσταση αντιπροσωπεύει μια εταιρεία. Από

τις 10 εταιρείες, οι 4 πρώτες ανήκουν στο χρηματοπιστωτικό τομέα, οι επόμενες 3 είναι εμπορικές, και οι τελευταίες 3 είναι υψηλής τεχνολογίας. Η υπόθεση λοιπόν ότι οι τιμές των μετοχών των εταιρειών του ίδιου τομέα τείνουν να μεταβάλλονται με τον ίδιο τρόπο καθώς οι οικονομικές συνθήκες εξελίσσονται, μπορεί να θεωρηθεί ως λογική. Η factor analysis μπορεί να δώσει ποσοτική ένδειξη ότι πράγματι αυτή η υπόθεση είναι σωστή. Προκειμένου λοιπόν να το ελέγξουμε αυτό, εκπαιδεύουμε ένα μοντέλο factor analysis με $q = 3$ factors, όσοι δηλαδή και οι τομείς όπου κατατάσσονται οι μετοχές, στα διαθέσιμα δεδομένα, με χρήση του EM αλγορίθμου για factor analyzers ως εξήχθη στο [22], και παρέχουμε μια γραφική απεικόνιση το αποτελέσματος, που μπορεί να βρείτε στο Σχ. 1. Μια εκάστη των 10 μετοχών παριστάνεται στο διάγραμμα αυτό με ένα διάνυσμα, η κατεύθυνση και το μήκος του οποίου δείχνει πώς η τιμή της μετοχής εξαρτάται από τους 3 factors. Επί παραδείγματι, οι 4 πρώτες μετοχές, που ανήκουν στο χρηματοπιστωτικό τομέα, έχουν θετικά factor loadings για τον πρώτο factor, που αντιστοιχεί στον financial sector, και ασήμαντα loadings για τους άλλους δυο factors. Αυτό έχει σαν αποτέλεσμα τα διανύσματα που αντιστοιχούν στις 4 πρώτες μετοχές να είναι προσανατολισμένα σχεδόν παράλληλα στον άξονα του financial sector στο προκύπτον διάγραμμα. Καθεμία από τις 10 παρατηρήσεις παριστάνεται από ένα σημείο στο διάγραμμά μας, και οι θέσεις τους ως προς τους άξονες καταδεικνύουν την βαθμολογία τους ως προς τον καθένα από τους τρεις factors, ήτοι τις τιμές των αντιστοίχων factor vectors.

Συμβατικά στην factor analysis, τα διανύσματα των παραγόντων (factor vectors) και του σφάλματος (error vectors) θεωρείται ότι ακολουθούν πολυμετάβλητες Gaussian κατανομές. Ωστόσο, οι κατανομές Student's- t προ-

σφέρουν μια ενδιαφέρουσα εναλλακτική στις Gaussians, εξασφαλίζουσα ισχυρή ανθεκτικότητα σε outliers. Προκειμένου να εκμεταλλευθούμε αυτά τα πλεονεκτήματα [25,26], εναλλακτικά το Student's- t factor analysis μοντέλο μπορεί να θεωρηθεί, όπου οι factor vectors \mathbf{y}_j θεωρούνται κατανεμημένοι κατά Student's- t ως

$$\mathbf{y}_j \sim t(\mathbf{0}, \mathbf{I}_q, \nu) \quad (1.2)$$

ανεξάρτητα από τα σφάλματα \mathbf{e}_j , που επίσης θεωρούνται κατανεμημένα κατά Student's- t ως

$$\mathbf{e}_j \sim t(\mathbf{0}, \Psi, \nu) \quad (1.3)$$

όπου Ψ είναι ένας διαγώνιος πίνακας, $\Psi = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$, \mathbf{I}_q είναι ο $q \times q$ μοναδιαίος πίνακας, και ν είναι οι βαθμοί ελευθερίας της κατανομής Student's- t . Τα σ_i^2 καλούνται μοναδικότητες (uniquenesses). Η συνάρτηση πυκνότητας πιθανότητας (σ.π.π.) της Student's- t κατανομής, $t(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, με μέσο $\boldsymbol{\mu}$, πίνακα συνδιακύμανσης $\boldsymbol{\Sigma}$, και ν βαθμούς ελευθερίας δίνεται από

$$t(\mathbf{x}_j; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma\left(\frac{\nu+p}{2}\right) |\boldsymbol{\Sigma}|^{-1/2}}{(\pi\nu)^{p/2} \Gamma(\nu/2) \{1 + \delta(\mathbf{x}_j, \boldsymbol{\mu}; \boldsymbol{\Sigma})/\nu\}^{\frac{\nu+p}{2}}} \quad (1.4)$$

όπου $\Gamma(s)$ είναι η συνάρτηση Gamma και $\delta(\mathbf{x}_j, \boldsymbol{\mu}; \boldsymbol{\Sigma})$ η τετραγωνική Mahalanobis distance μεταξύ $\mathbf{x}_j, \boldsymbol{\mu}$ με συνδιακύμανση $\boldsymbol{\Sigma}$

$$\delta(\mathbf{x}_j, \boldsymbol{\mu}; \boldsymbol{\Sigma}) = (\mathbf{x}_j - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_j - \boldsymbol{\mu}) \quad (1.5)$$

Τότε, από τον ορισμό του μοντέλου factor analysis (1) και τις κατανομές των factor vectors και των error vectors, συνάγεται ότι υπό την συνθήκη των αντιστοιχών τους factors \mathbf{y}_j , οι παρατηρήσεις \mathbf{x}_j είναι ανεξάρτητα κατανε-

μημένες ως $t(\boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{y}_j, \boldsymbol{\Psi}, \nu)$:

$$\mathbf{x}_j | \mathbf{y}_j \sim t(\boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{y}_j, \boldsymbol{\Psi}, \nu) \quad (1.6)$$

Ολοκληρώνοντας ως προς \mathbf{y}_j , η (6) δίνει ότι, άνευ συνθήκης, οι παρατηρήσεις \mathbf{x}_j είναι ανεξάρτητες και ισόνομες κατά μια κατανομή Student's- t με μέσο $\boldsymbol{\mu}$, και πίνακα συνδιακύμανσης ίσο με $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi}$, ήτοι

$$\mathbf{x}_j \sim t(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) \quad (1.7)$$

όπου το $\boldsymbol{\Sigma}$ δίνεται από

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi} \quad (1.8)$$

Ένας εναλλακτικός ορισμός της κατανομής των factor vectors και των error vectors μπορεί να εξαχθεί χρησιμοποιώντας μια ενδιαφέρουσα ιδιότητα της κατανομής Student's- t . Ακολουθώντας το [27], η Student's- t κατανομή μπορεί να αναπαρασταθεί ως μια άπειρη υπέρθεση από Gaussians του ιδίου μέσου και με διαβαθμισμένες ακρίβειες (αντίστροφοι των πινάκων συνδιακύμανσης), δίνοντας

$$t(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \int_0^\infty \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}/u) \mathcal{G}(u; \nu/2, \nu/2) du \quad (1.9)$$

όπου $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ είναι η κανονική κατανομή, και η τυχαία μεταβλητή $u > 0$ ακολουθεί κατανομή Gamma εξαρτώμενη μόνο από τους βαθμούς ελευθερίας, ν , της θεωρηθείσης κατανομής Student's- t , ήτοι

$$u \sim \mathcal{G}(\nu/2, \nu/2) \quad (1.10)$$

όπου η σ.π.π. της Gamma distribution, $\mathcal{G}(u; \alpha, \beta)$, δίνεται από

$$\mathcal{G}(u; \alpha, \beta) = u^{\alpha-1} \frac{\beta^\alpha e^{-\beta u}}{\Gamma(\alpha)} \quad (1.11)$$

Βάσει της (9), οι κατανομές των factor vectors και των error vectors, (2), (3), εναλλακτικά δίνουν

$$\mathbf{y}_j | u_j \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_q / u_j) \quad (1.12)$$

$$\mathbf{e}_j | u_j \sim \mathcal{N}(\mathbf{0}, \mathbf{\Psi} / u_j) \quad (1.13)$$

αντιστοίχως, όπου

$$u_j \sim \mathcal{G}(\nu/2, \nu/2) \quad (1.14)$$

Έτσι, η περιθώρια κατανομή των παρατηρήσεων, διδόμενη από την (7), μπορεί να ξαναγραφεί ως

$$\mathbf{x}_j | u_j \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma} / u_j) \quad (1.15)$$

όπου $\boldsymbol{\Sigma}$ δίνεται από την (8), ενώ οι κατανομή των παρατηρήσεων υπό την συνθήκη των αντιστοίχων τους factors, διδόμενη από την (6), μπορεί να ξαναγραφεί ως

$$\mathbf{x}_j | \mathbf{y}_j, u_j \sim \mathcal{N}(\boldsymbol{\mu} + \boldsymbol{\Lambda} \mathbf{y}_j, \boldsymbol{\Psi} / u_j) \quad (1.16)$$

1.4 Το προτεινόμενο FMSFA model

1.4.1 Διατύπωση Μοντέλου

Έστω μια c -συνιστωσών κατανομών υπέρθεση από Student's- t factor analyzes, με μέσους $\boldsymbol{\mu}_i$, πίνακες *factor loadings* $\boldsymbol{\Lambda}_i$, και διαγώνιους πίνακες *συνδιακύμανσης σφάλματος* $\boldsymbol{\Psi}_i$, σε ποσοστά υπέρθεσης (βάρη) $\pi_i, i = 1, \dots, c$.

Έχουμε δηλαδή

$$\boldsymbol{x}_j = \boldsymbol{\mu}_i + \boldsymbol{\Lambda}_i \boldsymbol{y}_{ij} + \boldsymbol{e}_{ij} \text{ with probability } \pi_i \quad (1.17)$$

όπου \boldsymbol{y}_{ij} είναι ο factor vector ο αντιστοιχών στην j οστη παρατήρηση δοθέντος ότι προήλθε από τον i οστο συνιστώντα analyzer

$$\boldsymbol{y}_{ij} \sim t(\mathbf{0}, \boldsymbol{I}_q, \nu_i) \quad (1.18)$$

και \boldsymbol{e}_{ij} είναι το σφάλμα (error) η αντιστοιχούσα στην j οστη παρατήρηση δοθέντος ότι προήλθε από τον i οστο συνιστώντα analyzer

$$\boldsymbol{e}_{ij} \sim t(\mathbf{0}, \boldsymbol{\Psi}_i, \nu_i) \quad (1.19)$$

Επιπροσθέτως, βάσει των (12)-(14), έχουμε

$$\boldsymbol{y}_{ij} | u_{ij} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_q / u_{ij}) \quad (1.20)$$

και

$$\boldsymbol{e}_{ij} | u_{ij} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_i / u_{ij}) \quad (1.21)$$

αντιστοίχως, όπου

$$u_{ij} \sim \mathcal{G}(\nu_i/2, \nu_i/2) \quad (1.22)$$

Τα βάρη της υπέρθεσης, π_i , είναι μη αρνητικά και αθροίζονται στην μονάδα.

Καθείς των συνιστώντων την θεωρηθείσα υπέρθεση Student's- t factor analyzers μπορεί να θεαθεί ως ορίζων/αντιπροσωπεύων ένα cluster στο χώρο \mathbf{X} των παρατηρήσεων. Ας ξεκινήσουμε την ανάλυσή μας από την πιθανότητα $p(i|\mathbf{x}_j)$, η οποία είναι η a posteriori πιθανότητα της παρατήρησης \mathbf{x}_j να ανήκει στο cluster που αντιπροσωπεύεται από τον ισοστό factor analyzer. Έχουμε τότε

$$0 \leq p(i|\mathbf{x}_j) \leq 1, \quad \sum_{i=1}^c p(i|\mathbf{x}_j) = 1 \quad (1.23)$$

($i = 1, \dots, c, j = 1, \dots, n$). Ακολουθώντας τα αποτελέσματα του Ruspini [3], η (23) υποδηλώνει ότι τα clusters αυτά μπορούν επίσης να θεωρηθούν ως ασαφή σύνολα (fuzzy sets), υπό τον ορισμό του Zadeh [1], αλλά και μια ισχυρή πιθανοθεωρητική αντίληψη.

Βάσει αυτών, μπορούμε να θεωρήσουμε ότι το υποτιθέν μοντέλο υπέρθεσης factor analyzers, (17), ορίζει μια ασαφή (fuzzy) c -διαμέριση του χώρου \mathbf{X} των παρατηρήσεων. Θα συμβολίζουμε αυτή την fuzzy διαμέριση as

$$R = \{r_{ij}\} \quad (1.24)$$

όπου r_{ij} ($i = 1, \dots, c, j = 1, \dots, n$) είναι ο βαθμός που η παρατήρηση \mathbf{x}_j ανήκει στο cluster που αντιπροσωπεύεται από τον ισοστό factor analyzer. Η συνάρτηση r_{ij} θα καλείται συνάρτηση ασαφούς μετοχής (fuzzy membership

function), και έχει τις ακόλουθες ιδιότητες:

$$0 \leq r_{ij} \leq 1, \quad \sum_{i=1}^c r_{ij} = 1, \quad 0 < \sum_{j=1}^n r_{ij} < n \quad (1.25)$$

Ένα σύνηθες clustering κριτήριο για τους αλγορίθμους fuzzy clustering FCM-τύπου [6] είναι η "απόσταση/διαφορά" μεταξύ ενός δεδομένου (παρατήρησης) και του προτύπου (αντιπροσωπευτικό διάνυσμα) του cluster, που εκφράζεται δια μέσου της *συνάρτησης ανομοιότητας (dissimilarity function)*, d_{ij} . Στον standard FCM αλγόριθμο [5], η *ασαφής αντικειμενική συνάρτηση (fuzzy objective function)* υπό βελτιστοποίηση, παίρνει, τότε, την μορφή

$$J_\phi \triangleq \sum_{i=1}^c \sum_{j=1}^n r_{ij}^\phi d_{ij} \quad (1.26)$$

όπου $\phi \geq 1$ είναι ένα εκθετικό βάρος σε κάθε fuzzy membership function r_{ij} , που ελέγχει τον βαθμό ασάφειας (*degree of fuzziness*) του αλγορίθμου, και καλείται ο *fuzzifier* του clustering αλγορίθμου. Καθώς η εισαγωγή του fuzzifier ϕ στα πλαίσια του *k-means algorithm*, η οποία και αποδίδει την FCM fuzzy objective function (26), έχει θεωρηθεί στερουμένη στερέας θεωρητικής ερμηνείας, μια διαφορετική προσέγγιση στο FCM-τύπου fuzzy clustering έχει προταθεί από τους Miyamoto και Mukaidono [9], όπου η ασαφοποίηση (fuzzification) επιτυγχάνεται μέσω μιας τεχνικής συστηματοποίησης (regularization technique) όπου ένας όρος εντροπίας εισάγεται στην fuzzy objective function. Τοιουτοτρόπως, η fuzzy objective function παίρνει την μορφή

$$J_\lambda^{en} \triangleq \sum_{i=1}^c \sum_{j=1}^n r_{ij} d_{ij} + \lambda \sum_{i=1}^c \sum_{j=1}^n r_{ij} \log r_{ij} \quad (1.27)$$

όπου ο όρος εντροπίας λειτουργεί ως ο *fuzzifier* και η παράμετρος λ είναι το *degree of fuzziness* του μοντέλου. Οι Ichihashi et al. [10] έχουν γενικεύσει την κανονικοποιημένη δια εντροπίας FCM-τύπου fuzzy objective function στα πλαίσια των *μοντέλων πεπερασμένων υπερθέσεων κατανομών (finite mixture models)*, εισάγοντας μια τεχνική regularization μέσω Kullback-Leibler information αντί εντροπίας. Υπό αυτή την θεώρηση

$$J_\lambda \triangleq \sum_{i=1}^c \sum_{j=1}^n r_{ij} d_{ij} + \lambda \sum_{i=1}^c \sum_{j=1}^n r_{ij} \log \left(\frac{r_{ij}}{\pi_i} \right) \quad (1.28)$$

όπου τα π_i είναι τα βάρη των υπερτιθέμενων κατανομών. Οι Ichihashi et al. [10] έχουν δείξει ότι με μια *κατάλληλη επιλογή* της dissimilarity function, d_{ij} , ο regularized μέσω KL information FCM-τύπου fuzzy clustering αλγόριθμος, με αντικειμενική συνάρτηση (28), μπορεί να θεωρηθεί ως μια ασαφοποίηση (fuzzification) του EM αλγορίθμου για GMMs, και ανάγεται σε αυτόν για συγκεκριμένη επιλογή της τιμής του λ .

Στην εργασία αυτή, εισάγουμε μια νέα προσέγγιση στο FCM-τύπου fuzzy clustering. Η μελέτη μας ξεκινάει υπό την υπόθεση της αναπαράστασης ενός δοθέντος συνόλου δεδομένων από μια πεπερασμένη υπέρθεση από factor analyzers (finite mixture of factor analyzers, MFA). Εν συνεχεία, προκειμένου να εκμεταλλευθούμε την ευρωστία της Student's- t κατανομής σε ατυπικά δεδομένα, υποθέτουμε ότι καθένας από τους θεωρηθέντες factor analyzers ακολουθεί μια Student's- t κατανομή, όπως εξηγήσαμε στο 1.3. Τελικά, θεωρούμε το εισαχθέν MFA μοντέλο ως ορίζων μια ασαφή διαμέριση (fuzzy partition) του χώρου παρατήρησης, όπως εξηγήσαμε προηγουμένως, απαιτούμε την εισαγωγή αυτού του μοντέλου στα πλαίσια ενός fuzzy clustering

αλγορίθμου. Αυτό μπορεί να γίνει εφικτό χρησιμοποιώντας την regularized μέσω KL information παραλλαγή του FCM αλγορίθμου, η οποία παρέχει ένα κατάλληλο πλαίσιο για την ασαφοποίηση (*fuzzification*) των *finite mixture models*, και δια της εισαγωγής μιας κατάλληλης dissimilarity function, ο ορισμός της οποίας να βασίζεται και να απορρέει κατάλληλα από την θεωρηθείσα αναπαράσταση των μοντελοποιουμένων δεδομένων με χρήση μιας πεπερασμένης υπέρθεσης Student's-*t* factor analyzers. Το προκύπτον υπό την θεώρηση αυτή, *μοντέλο ασαφούς υπερθέσεως κατά Student's-t αναλυτών παραγόντων* (*fuzzy mixture of Student's-t factor analyzers model, FMSFA*) είναι ένα fuzzy clustering μοντέλο παρέχον αυξημένη ανθεκτικότητα σε ατυπικά δεδομένα, και ταυτοχρόνως επιτρέπον την καλύτερη μοντελοποίηση εξαιρετικώς πολυδιαστάτων δεδομένων υπό το υπόδειγμα των fuzzy clustering αλγορίθμων, μέσω της χρήσης factor analysis.

Για να εξάγουμε, λοιπόν, την fuzzy objective function του FMSFA model, πρέπει να ορίσουμε μια κατάλληλη *dissimilarity function*, d_{ij} , που να εκφράζει την ανομοιότητα του j οστου δεδομένου, x_j , ως προς το i οστο cluster (συνιστώσα κατανομή της θεωρουμένης υπέρθεσης), επί την βάση του πιθανοτικού ορισμού του FMSFA model. Ακολουθώντας τον Hathaway [28], μια κατάλληλη συνάρτηση που ικανοποιεί αυτές τις απαιτήσεις, και συνεπώς, παρέχουσα μια κατάλληλη εκτίμηση της ανομοιότητας ενός δεδομένου, έστω x_j , και μιας θεωρουμένης, έστω της i οστής, συνιστώσας κατανομής του μοντέλου (factor analyzer), είναι η αρνητική λογαριθμο-πιθανοφάνεια (negative log-likelihood) της i οστής συνιστώσας κατανομής ως προς το δεδομένο x_j , ήτοι, ο αρνητικός λογάριθμος της σ.π.π. $p(x_j; \Theta_i)$. Αυτή η επιλογή της dissimilarity function έχει χρησιμοποιηθεί από πολλούς ερευνητές, ως το εργα-

λείο για την επιτυχή εισαγωγή της ρητής πληροφορίας σχετικά με την κατανομή των εξακτέων clusters στην διαδικασία του fuzzy clustering (δες, π.χ., [16,24,29]).

Ας αρχίσουμε την ανάλυσή μας με την σ.π.π. μιας υπέρθεσης από Student's- t factor analyzers. Από την (17) και τους ορισμούς των factor vectors, \mathbf{y}_{ij} , και των error vectors, e_{ij} ($i = 1, \dots, c$, $j = 1, \dots, n$), που δίδονται από τις (18) και (19), αντιστοίχως, έχουμε ότι η περιθώρια κατανομή κάθε παρατήρησης, \mathbf{x}_j , που μοντελοποιείται από μια c -συνιστωσών υπέρθεση Student's- t factor analyzers, δίνεται από

$$p(\mathbf{x}_j; \Theta) = \sum_{i=1}^c \pi_i p(\mathbf{x}_j; \Theta_i) \quad (1.29)$$

όπου

$$p(\mathbf{x}_j; \Theta_i) = t(\mathbf{x}_j; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \nu_i) \quad (1.30)$$

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i \boldsymbol{\Lambda}_i^T + \boldsymbol{\Psi}_i \quad (1.31)$$

$\Theta_i = \{\boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i, \boldsymbol{\Psi}_i, \nu_i\}$, και $\Theta = \{\pi_i, \Theta_i\}_{i=1}^c$. Ως παρατηρούμε, η $p(\mathbf{x}_j; \Theta_i)$ δεν εξαρτάται με άμεσο τρόπο από τις $\boldsymbol{\Lambda}_i$ and $\boldsymbol{\Psi}_i$, αλλά, αντιθέτως, από τους πίνακες συνδιακύμανσης $\boldsymbol{\Sigma}_i$, που αποτελούν συνάρτηση των $\boldsymbol{\Lambda}_i$ και $\boldsymbol{\Psi}_i$ ($i = 1, \dots, c$). Συνεπώς, η χρήση μιας άλλης, βολικότερης μορφής της κατανομής του μοντέλου θα ήταν επιθυμητή.

Μια καταλληλότερη έκφραση για την ανάπτυξη του FMSEFA μοντέλου μπορεί να εξαχθεί θεωρώντας την υπό συνθήκη κατανομή του \mathbf{x}_j επί του

κρυφού υποχώρου \mathbf{Y} . Έχουμε

$$p(\mathbf{x}_j | \{\mathbf{y}_{ij}\}_{i=1}^c; \Theta) = \sum_{i=1}^c \pi_i p(\mathbf{x}_j | \mathbf{y}_{ij}; \Theta_i) \quad (1.32)$$

όπου

$$p(\mathbf{x}_j | \mathbf{y}_{ij}; \Theta_i) = t(\mathbf{x}_j; \boldsymbol{\mu}_i + \Lambda_i \mathbf{y}_{ij}, \Psi_i, \nu_i) \quad (1.33)$$

και τα \mathbf{y}_{ij} είναι κατανομημένα όπως στην (18). Ωστόσο, μια κλειστού τύπου λύση για βελτιστοποίηση της log-likelihood μας Student's- t κατανομής δεν υφίσταται [27]. Στα πλαίσια του EM αλγορίθμου και των παραλλαγών του, αυτό το πρόβλημα έχει αντιμετωπιστεί με χρήση της (9), δηλ., χρησιμοποιώντας την έκφραση της Student's- t κατανομής ως μιας Gaussian με διαβαθμισμένη ακρίβεια, όπου η βαθμίδα είναι μια κατά Gamma κατανομημένη μεταβλητή. Υπό την θεώρηση αυτή, ο EM αλγόριθμος συνίσταται στην βελτιστοποίηση του τοιουτοτρόπως λαμβανομένου επαυξημένου μοντέλου, με χρήση της υπό την συνθήκη των δεδομένων προσδοκητής τιμής της λογαριθμο-πιθανοφανείας των πλήρων (επαυξημένων) δεδομένων (*conditional on the data expectation of the complete-data log-likelihood*) [27].

Με αυτά υπόψιν, και υπό τις (6) και (16), η υπό συνθήκη κατανομή της παρατήρησης \mathbf{x}_j επί του κρυφού υποχώρου \mathbf{Y} γράφεται ισοδυνάμως

$$p(\mathbf{x}_j | \{\mathbf{y}_{ij}, u_{ij}\}_{i=1}^c; \Theta) = \sum_{i=1}^c \pi_i p(\mathbf{x}_j | \mathbf{y}_{ij}, u_{ij}; \Theta_i) \quad (1.34)$$

όπου

$$p(\mathbf{x}_j | \mathbf{y}_{ij}, u_{ij}; \Theta_i) = \mathcal{N}(\mathbf{x}_j; \boldsymbol{\mu}_i + \Lambda_i \mathbf{y}_{ij}, \Psi_i / u_{ij}) \quad (1.35)$$

η κατανομή των βαθμίδων (scalars) u_{ij} δίνεται από την (22), και η υπό συνθήκη κατανομή $\mathbf{y}_{ij}|u_{ij}$ δίνεται από την (20). Τότε, η έκφραση της πιθανοφανείας του *i*οστού συνιστώσα factor analyzer ως προς την *j*οστή παρατήρηση, \mathbf{x}_j , τον αντίστοιχο factor vector, \mathbf{y}_{ij} , και το αντίστοιχο precision scalar, u_{ij} , δίνει

$$\begin{aligned} \mathcal{L}(\Theta_i; \mathbf{x}_j, \mathbf{y}_{ij}, u_{ij}) &\triangleq p(\mathbf{x}_j, \mathbf{y}_{ij}, u_{ij}; \Theta_i) = \\ &\mathcal{N}(\mathbf{x}_j; \boldsymbol{\mu}_i + \Lambda_i \mathbf{y}_{ij}, \Psi_i/u_{ij}) \mathcal{N}(\mathbf{y}_{ij}; \mathbf{0}, \mathbf{I}_q/u_{ij}) \mathcal{G}(u_{ij}; \nu_i/2, \nu_i/2) \end{aligned} \quad (1.36)$$

Από τα παραπάνω, συνάγεται ότι η likelihood function $\mathcal{L}(\Theta_i; \mathbf{x}_j, \mathbf{y}_{ij}, u_{ij})$ παρέχει μια κατάλληλη εκτίμηση ανομοιότητας για το FMSFA model και ταυτοχρόνως εξασφαλίζει μια υπολογιστικά αποδοτική διαδικασία εκτιμητικής. Ωστόσο, η μορφή της (36) υποδηλώνει επίσης ότι για την εκτιμητική του FMSFA απαιτείται και ο υπολογισμός εκτιμητριών για τα κρυφά δεδομένα, δηλ. τα factor vectors, \mathbf{y}_{ij} , και τα precision scalars, u_{ij} , η οποία, εν τούτοις, θα πρέπει να γίνεται λαμβανομένης υπόψιν της πληροφορίας σχετικά με τις posterior κατανομές των μεταβλητών αυτών, όπως συμβαίνει και στον EM algorithm για mixtures of Student's-*t* factor analyzers [25].

Υπό αυτά τα δεδομένα, εξάγουμε την έκφραση της regularized μέσω KL information FCM-τύπου fuzzy objective function, J_λ , για το FMSFA model ορίζοντας την dissimilarity function, d_{ij} , ως την υπό την συνθήκη των δεδομένων προσδοκητή τιμή της αρνητικής λογαριθμο-πιθανοφανείας (negative log-likelihood) του *i*οστού factor analyzer ως προς την *j*οστη παρατήρηση και τα αντίστοιχα factor vector και precision scalar, $-\log \mathcal{L}(\Theta_i; \mathbf{x}_j, \mathbf{y}_{ij}, u_{ij})$, δοθείσης της τρεχούσης εκτιμητριάς του μοντέλου, $\hat{\Theta}$. Με άλλα λόγια, ορίζουμε την

dissimilarity function ως

$$d_{ij}(\Theta_i; \hat{\Theta}) \triangleq -E_{\hat{\Theta}} (\log \mathcal{L}(\Theta_i; \mathbf{x}_j, \mathbf{y}_{ij}, u_{ij}) | \mathbf{x}_j) \quad (1.37)$$

Τότε, από τις (28) και (37), έχουμε ότι οι fuzzy objective functions του FMSFA model αποτελούν μια άπειρη οικογένεια με

$$\begin{aligned} J_\lambda(\Theta, R) \triangleq & - \sum_{i=1}^c \sum_{j=1}^n r_{ij} E_{\hat{\Theta}} (\log \mathcal{L}(\Theta_i; \mathbf{x}_j, \mathbf{y}_{ij}, u_{ij}) | \mathbf{x}_j) \\ & + \lambda \sum_{i=1}^c \sum_{j=1}^n r_{ij} \log \left(\frac{r_{ij}}{\pi_i} \right) \end{aligned} \quad (1.38)$$

όπου r_{ij} είναι τα fuzzy membership functions και λ το degree of fuzziness του FMSFA model.

Συγκρίνοντας την έκφραση της $J_\lambda(\Theta, R)$ με την αντικειμενική συνάρτηση του EM αλγορίθμου για mixtures of Student's- t factor analyzers, που δίδεται από [25]

$$\begin{aligned} J_{EM}(\Theta) = & - \sum_{i=1}^c \sum_{j=1}^n p(i | \mathbf{x}_j) E_{\hat{\Theta}} (\log \mathcal{L}(\Theta_i; \mathbf{x}_j, \mathbf{y}_{ij}, u_{ij}) | \mathbf{x}_j) \\ & - \sum_{i=1}^c \sum_{j=1}^n p(i | \mathbf{x}_j) \log \pi_i \end{aligned}$$

είναι προφανές ότι αυτή η επιλογή της dissimilarity function, d_{ij} , για το FMSFA model αποδίδει μια οικογένεια fuzzy objective functions που μπορούν να θεωρηθούν ως μια FCM-τύπου εναλλακτική του EM αλγορίθμου για mixtures of Student's- t factor analyzers [25], παρέχουσα μια νέα θεώρηση της χρησιμότητας του fuzzy clustering (καθώς αναδεικνύεται ως μια πλεονεκτούσα

εναλλακτική του EM αλγορίθμου για την εκπαίδευση finite mixture models).

1.4.2 Εκτιμητική του FMSFA Model

Η FCM-τύπου εκτιμητική των παραμέτρων του FMSFA model δοθείσης μιας συλλογής εκπαίδευσης είναι εφικτή μέσω της ελαχιστοποίησης της αντικειμενικής συνάρτησης (fuzzy objective function) $J_\lambda(\Theta, R)$, διδομένης από την (38), ως προς τα Θ και R , υπό ένα coordinate descent υπόδειγμα. Έστω η $(k+1)$ οστη επανάληψη του αλγορίθμου. Θεωρούμε δηλαδή ότι η τρέχουσα τιμή της εκτιμητικής του μοντέλου $\hat{\Theta}$, είναι η υπολογισμένη στην k οστή επανάληψη του αλγορίθμου, $\Theta^{(k)}$, ήτοι, $\hat{\Theta} = \Theta^{(k)}$. Αρχικά, πρέπει να εξάγουμε την έκφραση της υπό συνθήκη προσδοκητής τιμής των λογαριθμο-πιθανοφανειών των συνιστωσών το μοντέλο κατανομών. Τότε, από την (36) έχουμε

$$\log \mathcal{L}(\Theta_i; \mathbf{x}_j, \mathbf{y}_{ij}, u_{ij}) = \log \mathcal{L}_{1j}(\xi_i) + \log \mathcal{L}_{2j}(\nu_i) \quad (1.39)$$

όπου $\xi_i = \{\mu_i, \Lambda_i, \Psi_i\}$, και είναι

$$\begin{aligned} \log \mathcal{L}_{1j}(\xi_i) &\triangleq \log \mathcal{L}(\xi_i; \mathbf{x}_j, \mathbf{y}_{ij}, u_{ij}) \Rightarrow \\ \log \mathcal{L}_{1j}(\xi_i) &= -\frac{p}{2} \log(2\pi) - \frac{1}{2} \log |\Psi_i| + \frac{p}{2} \log u_{ij} \\ &\quad - \frac{u_{ij}}{2} (\mathbf{x}_j - \mu_i - \Lambda_i \mathbf{y}_{ij})^T \Psi_i^{-1} (\mathbf{x}_j - \mu_i - \Lambda_i \mathbf{y}_{ij}) \\ &\quad - \frac{q}{2} \log(2\pi) + \frac{q}{2} \log u_{ij} - \frac{u_{ij}}{2} \mathbf{y}_{ij}^T \mathbf{y}_{ij} \end{aligned} \quad (1.40)$$

$$\begin{aligned} \log \mathcal{L}_{2j}(\nu_i) &\triangleq \log \mathcal{L}(\nu_i; u_{ij}) = -\log \Gamma\left(\frac{\nu_i}{2}\right) + \frac{\nu_i}{2} \log\left(\frac{\nu_i}{2}\right) \\ &\quad + \frac{\nu_i}{2} (\log u_{ij} - u_{ij}) - \log u_{ij} \end{aligned} \quad (1.41)$$

Συνεπώς, η έκφραση της dissimilarity function, εκτιμώμενης στην $(k+1)$ οστη επανάληψη

$$d_{ij}(\Theta_i; \Theta^{(k)}) \triangleq -\mathbb{E}_{\Theta^{(k)}} (\log \mathcal{L}(\Theta_i; \mathbf{x}_j, \mathbf{y}_{ij}, u_{ij}) | \mathbf{x}_j) \quad (1.42)$$

δίνεται από

$$d_{ij}(\Theta_i; \Theta^{(k)}) = -Q_{1j}(\xi_i; \Theta^{(k)}) - Q_{2j}(\nu_i; \Theta^{(k)}) \quad (1.43)$$

όπου

$$\begin{aligned} Q_{1j}(\xi_i; \Theta^{(k)}) &= -\frac{1}{2} \log |\Psi_i| + \frac{p+q}{2} \mathbb{E}_{\Theta^{(k)}} (\log u_{ij} | \mathbf{x}_j) \\ &\quad - \frac{p+q}{2} \log(2\pi) + (\mathbf{x}_j - \boldsymbol{\mu}_i)^T \Psi_i^{-1} \boldsymbol{\Lambda}_i \mathbb{E}_{\Theta^{(k)}} (u_{ij} \mathbf{y}_{ij} | \mathbf{x}_j) \\ &\quad - \frac{1}{2} \text{trace} [\boldsymbol{\Lambda}_i^T \Psi_i^{-1} \boldsymbol{\Lambda}_i \times \mathbb{E}_{\Theta^{(k)}} (u_{ij} \mathbf{y}_{ij} \mathbf{y}_{ij}^T | \mathbf{x}_j)] \\ &\quad - \frac{1}{2} \mathbb{E}_{\Theta^{(k)}} (u_{ij} | \mathbf{x}_j) (\mathbf{x}_j - \boldsymbol{\mu}_i)^T \Psi_i^{-1} (\mathbf{x}_j - \boldsymbol{\mu}_i) \\ &\quad - \frac{1}{2} \text{trace} [\mathbb{E}_{\Theta^{(k)}} (u_{ij} \mathbf{y}_{ij} \mathbf{y}_{ij}^T | \mathbf{x}_j)] \end{aligned} \quad (1.44)$$

$$\begin{aligned} Q_{2j}(\nu_i; \Theta^{(k)}) &= -\log \Gamma \left(\frac{\nu_i}{2} \right) + \frac{\nu_i}{2} \log \left(\frac{\nu_i}{2} \right) + \\ &\quad + \frac{\nu_i}{2} \mathbb{E}_{\Theta^{(k)}} (\log u_{ij} - u_{ij} | \mathbf{x}_j) - \mathbb{E}_{\Theta^{(k)}} (\log u_{ij} | \mathbf{x}_j) \end{aligned} \quad (1.45)$$

από όπου έχουμε ότι ο υπολογισμός της dissimilarity function, $d_{ij}(\Theta_i; \Theta^{(k)})$, απαιτεί την εξαγωγή των posterior προσδοκητών τιμών των precision scalars u_{ij} , $\mathbb{E}_{\Theta^{(k)}} (u_{ij} | \mathbf{x}_j)$, και των λογαρίθμων αυτών, $\mathbb{E}_{\Theta^{(k)}} (\log u_{ij} | \mathbf{x}_j)$, ως και των προσδοκητών τιμών $\mathbb{E}_{\Theta^{(k)}} (u_{ij} \mathbf{y}_{ij} \mathbf{y}_{ij}^T | \mathbf{x}_j)$ και $\mathbb{E}_{\Theta^{(k)}} (u_{ij} \mathbf{y}_{ij} | \mathbf{x}_j)$. Μπορεί να δειχθεί (δες Παράρτημα Α), ότι ο υπολογισμός αυτών των προσδοκητών τιμών

ανάγεται, εν τέλει, στον υπολογισμό των ποσοτήτων

$$u_{ij}^{(k)} \triangleq \mathbb{E}_{\Theta^{(k)}}(u_{ij} | \mathbf{x}_j) = \frac{\nu_i^{(k)} + p}{\nu_i^{(k)} + \delta(\mathbf{x}_j, \boldsymbol{\mu}_i^{(k)}; \boldsymbol{\Sigma}_i^{(k)})} \quad (1.46)$$

$$\boldsymbol{\gamma}_i^{(k)} = (\boldsymbol{\Lambda}_i^{(k)} \boldsymbol{\Lambda}_i^{(k)T} + \boldsymbol{\Psi}_i^{(k)})^{-1} \boldsymbol{\Lambda}_i^{(k)} \quad (1.47)$$

$$\boldsymbol{\omega}_i^{(k)} = \mathbf{I}_q - \boldsymbol{\gamma}_i^{(k)T} \boldsymbol{\Lambda}_i^{(k)} \quad (1.48)$$

όπου τα $\boldsymbol{\Sigma}_i^{(k)}$ δίνονται από την (31).

Μπορούμε, πλέον, να προχωρήσουμε στην εκτιμητική των fuzzy membership functions r_{ij} , και των παραμέτρων $\pi_i, \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i, \boldsymbol{\Psi}_i$, και ν_i του μοντέλου. Ξεκινάμε με τα fuzzy membership functions, r_{ij} . Η ελαχιστοποίηση της $J_\lambda(\boldsymbol{\Theta}, R)$ επί των r_{ij} υπό τον περιορισμό $\sum_{i=1}^c r_{ij} = 1, \forall j = 1, \dots, n$, δίδει (δες Παράρτημα Β)

$$r_{ij}^{(k)} = \frac{\pi_i^{(k)} \exp\left(-\frac{1}{\lambda} d_{ij}^{(k)}\right)}{\sum_{h=1}^c \pi_h^{(k)} \exp\left(-\frac{1}{\lambda} d_{hj}^{(k)}\right)} \quad (1.49)$$

Εδώ, $d_{ij}^{(k)}$ είναι η τιμή της dissimilarity function του FMSFA model, (42), για $\boldsymbol{\Theta}_i = \boldsymbol{\Theta}_i^{(k)}$

$$d_{ij}^{(k)} \triangleq d_{ij}(\boldsymbol{\Theta}_i^{(k)}; \boldsymbol{\Theta}^{(k)}) \quad (1.50)$$

Περαιτέρω, η ελαχιστοποίηση της $J_\lambda(\boldsymbol{\Theta}, R)$ ως προς τα $\pi_i, i = 1, \dots, c$, υπό την συνθήκη $\sum_{i=1}^c \pi_i = 1$, δίδει (δες Παράρτημα Β)

$$\pi_i^{(k+1)} = \sum_{j=1}^n r_{ij}^{(k)} / n \quad (1.51)$$

Με παρόμοιο τρόπο (δες Παράρτημα Β) παίρνουμε

$$\boldsymbol{\mu}_i^{(k+1)} = \sum_{j=1}^n r_{ij}^{(k)} u_{ij}^{(k)} \left(\mathbf{x}_j - \boldsymbol{\Lambda}_i^{(k)} \mathbf{y}_{ij}^{(k)} \right) / \sum_{j=1}^n r_{ij}^{(k)} u_{ij}^{(k)} \quad (1.52)$$

όπου $\mathbf{y}_{ij}^{(k)}$ είναι η τρέχουσα τιμή της posterior προσδοκητής τιμής του factors vector που αντιστοιχεί στο \mathbf{x}_j , ως προς τον ισοστο συνιστώντα factor analyzer, $\mathbf{y}_{ij}^{(k)} \triangleq \mathbb{E}_{\Theta^{(k)}} (\mathbf{y}_{ij} | \mathbf{x}_j)$, και δίδεται από (see Appendix A)

$$\mathbf{y}_{ij}^{(k)} \triangleq \mathbb{E}_{\Theta^{(k)}} (\mathbf{y}_{ij} | \mathbf{x}_j) = \boldsymbol{\gamma}_i^{(k)T} (\mathbf{x}_j - \boldsymbol{\mu}_i^{(k)}) \quad (1.53)$$

Επιπλέον, λαμβάνουμε

$$\boldsymbol{\Lambda}_i^{(k+1)} = \mathbf{V}_i^{(k)} \boldsymbol{\gamma}_i^{(k)} \left(\boldsymbol{\gamma}_i^{(k)T} \mathbf{V}_i^{(k)} \boldsymbol{\gamma}_i^{(k)} + \boldsymbol{\omega}_i^{(k)} \right)^{-1} \quad (1.54)$$

όπου

$$\mathbf{V}_i^{(k)} = \frac{\sum_{j=1}^n r_{ij}^{(k)} u_{ij}^{(k)} (\mathbf{x}_j - \boldsymbol{\mu}_i^{(k)}) (\mathbf{x}_j - \boldsymbol{\mu}_i^{(k)})^T}{\sum_{j=1}^n r_{ij}^{(k)}} \quad (1.55)$$

και

$$\boldsymbol{\Psi}_i^{(k+1)} = \text{diag} \{ \mathbf{V}_i^{(k)} - \mathbf{V}_i^{(k)} \boldsymbol{\gamma}_i^{(k)} \boldsymbol{\Lambda}_i^{(k)T} \} \quad (1.56)$$

Τέλος, οι εκτιμήτριες $\nu_i^{(k+1)}$, $i = 1, \dots, c$ δίνονται από

$$\begin{aligned} & -\psi \left(\frac{\nu_i}{2} \right) + \log \left(\frac{\nu_i}{2} \right) + \psi \left(\frac{\nu_i^{(k)} + p}{2} \right) - \log \left(\frac{\nu_i^{(k)} + p}{2} \right) + \\ & + 1 + \frac{1}{\sum_{j=1}^n r_{ij}^{(k)}} \sum_{j=1}^n r_{ij}^{(k)} \left(\log u_{ij}^{(k)} - u_{ij}^{(k)} \right) = 0 \quad (1.57) \end{aligned}$$

όπου $\psi(s)$ είναι η digamma function, $\psi(s) = \partial \log \Gamma(s) / \partial s$.

Η λύση της (57) δεν υπάρχει σε κλειστή μορφή. Παρ' όλα αυτά, παρατηρούμε ότι η έκφρασή της είναι ίδια με αυτή της εκτιμητορίας των βαθμών ελευθερίας ενός SMM εάν αντικαταστήσουμε τις a posteriori probabilities των συνιστωσών κατανομών του SMM με τα fuzzy membership functions του μοντέλου μας. Όμως, για την περίπτωση των SMM, μια προσεγγιστική λύση κλειστού τύπου για την (57) έχει δοθεί από τον Shoham [30]. Υπό αυτό το πρίσμα, προτείνουμε μια προσαρμογή του τύπου του Shoham στο FMSFA model αντικαθιστώντας τις a posteriori probabilities των συνιστωσών κατανομών του SMM με τα fuzzy membership functions του μοντέλου μας, λαμβάνοντας [30]

$$\nu^{(k+1)} = 0.0416 \left(1 + \operatorname{erf} \left(0.6594 \log \left(\frac{2.1971}{\tau + \log \tau - 1} \right) \right) \right) + \frac{2}{\tau + \log \tau - 1} \quad (1.58)$$

όπου $\nu_\epsilon = \nu_\zeta = \nu \forall \zeta \neq \epsilon = 1, \dots, c$, και το τ δίνεται από

$$\tau \triangleq -\frac{1}{n} \sum_{i=1}^c \sum_{j=1}^n r_{ij}^{(k)} \left[\psi \left(\frac{\nu^{(k)} + p}{2} \right) + \log \left(\frac{2}{\nu^{(k)} + \delta(\mathbf{x}_j, \boldsymbol{\mu}_i^{(k)}; \boldsymbol{\Sigma}_i^{(k)})} \right) - u_{ij}^{(k)} \right] \quad (1.59)$$

Η ποιότητα αυτής της προσέγγισης θα επαληθευθεί στην πειραματική ενότητα αυτού του κεφαλαίου.

Όπως έχουμε ήδη συζητήσει, ο FCM-τύπου αλγόριθμος εκπαίδευσης του FMSFA model είναι κατ' ουσίαν μια fuzzy clustering εναλλακτική του expectation-maximization (EM) αλγορίθμου για mixtures of Student's- t factor

analyzers, όπου, καθώς συνάγεται από τις (46)-(59), σε κάθε βήμα, το E-step περιλαμβάνει τον υπολογισμό των fuzzy membership functions, r_{ij} , των precision scalars, u_{ij} , και των ποσοτήτων γ_i και ω_i , ενώ το M-step περιλαμβάνει την εκτιμητική των παραμέτρων, π_i , μ_i , Λ_i , Ψ_i and ν_i . Επιπλέον, παρατηρούμε ότι σε κάθε επανάληψη του αλγορίθμου κάθε μια από τις εκτιμήτριες των μ_i , Λ_i και Ψ_i υπολογίζονται θεωρώντας τις άλλες δυο σταθερές και ίσες με την εκτιμήτριά τους που ελήφθη στην προηγούμενη επανάληψη του αλγορίθμου. Βάσει αυτών, και προκειμένου να επιταχυνθεί η σύγκλιση του αλγορίθμου εκπαίδευσης του FMSFA model, προχωράμε σε μια περαιτέρω προσαρμογή του προτεινομένου FCM-τύπου αλγορίθμου υπό το υπόδειγμα του expectation-conditional maximization (ECM) αλγορίθμου [31]. Έτσι, ο αλγόριθμος εκπαίδευσης του FMSFA model διαμορφώνεται ως εξής

1. Υπολόγισε τα $u_{ij}^{(k)}$, $\gamma_i^{(k)}$, $\omega_i^{(k)}$, $r_{ij}^{(k)}$ ($i = 1, \dots, c$, $j = 1, \dots, n$) από τις (46)-(49) αντίστοιχα.
2. Υπολόγισε τις εκτιμήτριες $\pi_i^{(k+1)}$, $\mu_i^{(k+1)}$, $\nu_i^{(k+1)}$ από τις (51), (52) και (54), αντίστοιχα.
3. Υπολόγισε εκ νέου τις παραμέτρους u_{ij} , r_{ij} , παίρνοντας τις νέες τιμές $u_{ij}^{(k+1/2)}$, $r_{ij}^{(k+1/2)}$, με χρήση των (46) και (49), και των ενημερωμένων εκτιμητριών $\pi_i^{(k+1)}$, $\mu_i^{(k+1)}$, $\nu_i^{(k+1)}$ που βρέθηκαν στο βήμα 2. Υπολόγισε επίσης την ποσότητα

$$V_i^{(k+1/2)} = \frac{\sum_{j=1}^n r_{ij}^{(k+1/2)} u_{ij}^{(k+1/2)} (\mathbf{x}_j - \mu_i^{(k+1)}) (\mathbf{x}_j - \mu_i^{(k+1)})^T}{\sum_{j=1}^n r_{ij}^{(k+1/2)}}$$
4. Υπολόγισε τις εκτιμήτριες $\Lambda_i^{(k+1)}$ από την (57), με χρήση των $V_i^{(k+1/2)}$ αντί των $V_i^{(k)}$. Υπολόγισε επίσης τις εκτιμήτριες $\Psi_i^{(k+1)}$ από την (59), με χρήση των $\Lambda_i^{(k+1)}$ και $V_i^{(k+1/2)}$ αντί των $\Lambda_i^{(k)}$ και $V_i^{(k)}$, αντίστοιχα.

5. Σε περίπτωση σύγκλισης, δηλ. $\left| J_\lambda \left(\Theta^{(k+1)}, R^{(k+1)} \right) - J_\lambda \left(\Theta^{(k)}, R^{(k)} \right) \right| < T_c$ όπου, T_c το κατώφλι σύγκλισης, έξελθε. Αλλιώς, $k := k + 1$ και επίστρεψε στο 1.

1.4.3 Μεθοδολογίες Βέλτιστης Επιλογής Μοντέλου

Το πρόβλημα της επιλογής μοντέλου σε εφαρμογές probabilistic clustering δεν έχει πλήρως λυθεί. Στην πραγματικότητα, διαφορετικές μέθοδοι επιλογής μοντέλου, από μια μεγάλη γκάμα που έχει προταθεί στην διεθνή βιβλιογραφία, έχουν καταδειχθεί ότι δουλεύουν καλά για διαφορετικές εφαρμογές και δεδομένα υπό μελέτη [13].

Ας αρχίσουμε με το πρόβλημα του προσδιορισμού του αριθμού των υπερτιθεμένων συνιστωσών (clusters), c , και του αριθμού των factors, q , για το FMSFA model. Οι προτεινόμενες εναλλακτικές μέθοδοι περιλαμβάνουν: (i) τεχνικές cross-validation [32] και (ii) προσδιορισμό βάσει επίδοσης [33]. Οι cross-validation τεχνικές περιλαμβάνουν, βασικά, την k -fold cross-validation, την leave-one-out cross-validation, και την leave- v -out τεχνική. Στην k -fold cross-validation, το δείγμα εκπαίδευσης διαιρείται σε k υποσύνολα (περίπου) ίδιου μεγέθους. Εν συνεχεία, ένα FMSFA model (με συγκεκριμένη επιλογή για τις c και q παραμέτρους) εκπαιδεύεται k φορές, κάθε φορά παραλείποντας ένα από τα υποσύνολα από την εκπαίδευση, το οποίο όμως χρησιμοποιείται για τον υπολογισμό της ψευδο-πιθανο-φανείας του μοντέλου (FMSFA model fuzzy objective function, J_λ) η οποία και αποτελεί το κριτήριο επιλογής μοντέλου υπό την εν λόγω τεχνική. Αν το k ισούται με το μέγεθος του δείγματος, τότε έχουμε leave-one-out cross-validation. Η leave- v -out τεχνική είναι πιο πολύπλοκη και περιλαμβάνει παράλειψη των υποσυνόλων που αν-

τιστοιχούν σε ν περιπτώσεις μελέτης. Εναλλακτικά, προσδιορισμός βάσει επίδοσης μπορεί επίσης να εφαρμοστεί. Αυτός συνίσταται στην ανάθεση κάποιας εργασίας αναγνώρισης προτύπων σε FMSFA μοντέλα με διαφορετικές τιμές για τις c και q παραμέτρους και τον προσδιορισμό του συνδυασμού τιμών τους που βελτιστοποιεί την επίδοση του μοντέλου ως προς την ανατεθείσα εργασία.

Τέλος, σχετικά με την επιλογή του degree of fuzziness του μοντέλου, είναι γνωστό ότι ένα στέρεο θεωρητικό υπόβαθρο σχετικά με την επιλογή του degree of fuzziness του FCM algorithm και των παραλλαγών του δεν έχει ακόμα προταθεί. Η θεμελίωση μιας τέτοιας θεωρίας προφανώς είναι εκτός των στόχων αυτής της εργασίας. Επιπλέον, cross-validation τεχνικές δεν είναι εφαρμόσιμες σε αυτή την περίπτωση, καθώς είναι μάλλον χωρίς νόημα να συγκρίνουμε τις τιμές της fuzzy objective function του μοντέλου για διαφορετικές τιμές του degree of fuzziness. Συνεπώς, προς το παρόν η μόνη ορατή λύση είναι η ευρετική επιλογή του degree of fuzziness, λ , του FMSFA model, δια της εφαρμογής ενός κριτηρίου προσδιορισμού βάσει επίδοσης.

1.5 Πειραματική Αξιολόγηση

Στην ενότητα αυτή αξιολογούμε το προτεινόμενο μοντέλο σε clustering συνθετικών θορυβωδών δεδομένων, αναγνώριση χειρογράφων ψηφίων, σημασιολογική αναγνώριση αντικειμένων σε καταταμημένες εικόνες, και spam e-mail εντοπισμό. Πέραν του FMSFA model, προκειμένου να συγκρίνουμε την μεθόδό μας με υπάρχουσες fuzzy και μη-fuzzy clustering τεχνικές, στα πειράματά μας αξιολογούμε επιπλέον τις μεθόδους finite mixtures of Student's- t

factor analyzers (*t*MFA) [25], KLFCV [24], και MFA [22].

Στο πρώτο πείραμα, ένα 3-συνιστωσών μοντέλο από κάθε τύπο (FMSFA, *t*MFA, KLFCV και MFA) εκπαιδεύεται με δεδομένα από μια 3-συνιστωσών υπέρθεση διμετάβλητων Gaussian κατανομών, στα οποία προστίθεται αριθμός από outliers. Τα εκπαιδευμένα μοντέλα αξιολογούνται επί της ομοιότητας των εξαγομένων clusters με τα πραγματικά. Στα υπόλοιπα 3 πειράματα, οι αξιολογούμενες τεχνικές χρησιμοποιούνται σε εφαρμογές ταξινόμησης προτύπων, επί τη βάση του προκύπτοντος ρυθμού σφάλματος. Η πειραματική διαδικασία που ακολουθείται είναι η εξής: Έστω η τεχνική T υπό αξιολόγηση. Χρησιμοποιώντας την τεχνική T , εκπαιδεύουμε ένα μοντέλο για κάθε κλάση στο θεωρηθέν σύνολο δεδομένων εκπαίδευσης. Στην συνέχεια, για να εκτιμήσουμε την επίδοση της μεθόδου T , εφαρμόζουμε μια διαδικασία δυο σταδίων: Αρχικά για κάθε δεδομένο του συνόλου αξιολόγησης υπολογίζεται η posterior probability του ως προς καθένα από τα μοντέλα (κλάσεις) τα εκπαιδευμένα με την μέθοδο T . Αυτό επιτυγχάνεται (υπό τη υπόθεση ισότιμων prior probabilities για όλες τις κλάσεις) με υπολογισμό της πιθανοφάνειας των μοντέλων ως προς τις κλάσεις. Περαιτέρω, κάθε δεδομένο αντιστοιχίζεται στην κλάση που μεγιστοποιεί την πιθανοφάνειά του. Έτσι, υπολογίζεται για κάθε μέθοδο ο ρυθμός σφάλματος (δοθέντων των πραγματικών αντιστοιχίσεων σε κλάσεις των δεδομένων αξιολόγησης).

Τέλος, αναφορικά με την επιλογή του αριθμού των subclusters (συνιστωσών τα μοντέλα factor analyzers), c , και του αριθμού των factors, q , στα τελευταία πειράματα, είναι γνωστό (δες π.χ., [12, 13]) ότι οι cross-validation τεχνικές, δεν εγγυώνται πάντα την αξιόπιστη εκτίμησή τους, και, άρα, μπο-

ρεί να επηρεάσουν αρνητικά την επίδοση των εκπαιδευομένων μοντέλων. Για τον λόγο αυτό, και προκειμένου να εξασφαλίσουμε την δίκαιη σύγκριση των αξιολογούμενων μεθόδων, χρησιμοποιούμε προσδιορισμό των εν λόγω παραμέτρων βάσει επίδοσης. Με άλλα λόγια, επαναλαμβάνουμε τα πειράματά μας για διάφορες τιμές των c και q και επιλέγουμε τις τιμές τους που βελτιστοποιούν την επίδοσή τους.

1.5.1 Clustering Θορυβωδών Συνθετικών Δεδομένων

Ξεκινάμε με ένα πείραμα clustering θορυβωδών συνθετικών δεδομένων. Παρέχουμε, δε, μια γραφική απεικόνιση των clusters προκειμένου να συγκρίνουμε την επίδοση των αξιολογούμενων αλγορίθμων.

Θεωρούμε μια 3-συνιστωσών υπέρθεση διμετάβλητων Gaussian κατανομών με παραμέτρους

$$\begin{aligned} \boldsymbol{\mu}_1 &= (-6, 1.5), \boldsymbol{\mu}_2 = (0, 0)^T, \boldsymbol{\mu}_3 = (6, 1.5)^T, \\ \boldsymbol{\Sigma}_1 &= \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}, \boldsymbol{\Sigma}_2 = \begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix}, \\ \boldsymbol{\Sigma}_3 &= \begin{pmatrix} 1.56 & 0 \\ 0 & 1.56 \end{pmatrix} \end{aligned}$$

και $\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$. Εκατό παρατηρήσεις εξάγονται από κάθε κατανομή στο μοντέλο και σε αυτές προστίθενται 25% outliers. Τα outliers προέρχονται από μια ομοιόμορφη κατανομή στο διάστημα $[-10, 10]$ σε κάθε μεταβλητή. Με αυτά τα δεδομένα εκπαιδεύουμε ένα FMSFA, ένα t MFA, ένα KLFCV, και ένα MFA model με $c = 3$ και $q = 1$. Η επιλογή του degree of fuzziness λ στα FMSFA και KLFCV μοντέλα έγινε με κριτήριο την βελτιστοποίηση της επίδοσής τους.

Πίνακας 1.1: Clustering Θορυβωδών Συνθετικών Δεδομένων: Εκτιμήτριες των κέντρων των clusters

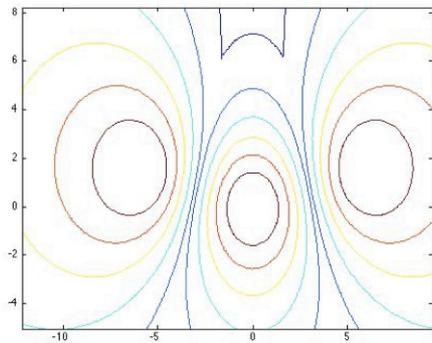
Model	μ_1	μ_2	μ_3
FMSFA ($\lambda = 0.9$)	$(-5.9083, 1.4285)^T$	$(5.9020, 1.4347)^T$	$(0.4191, -0.6677)^T$
t MFA	$(-6.6430, 0.6449)^T$	$(5.7948, 1.1307)^T$	$(-1.0158, 0.6960)^T$
KLFCV ($\lambda = 0.7$)	$(-3.9985, 1.7237)^T$	$(6.1850, 1.4079)^T$	$(2.1290, -2.3696)^T$
MFA	$(-8.1029, -0.8453)^T$	$(3.5751, 0.3388)^T$	$(-4.1815, 2.4472)^T$

Στον Πίνακα 1, δίνουμε τις εκτιμήτριες των κέντρων των clusters, όπως βρέθηκαν απο τις διάφορες μεθόδους. Στο Σχ. 2, δίνουμε τα περιγράμματα των πραγματικών clusters και των υπολογισμένων απο τις διάφορες μεθόδους. Είναι φανερό πως τα βασισμένα σε Student's- t μοντέλα δίνουν μεγάλη ανθεκτικότητα στα outliers, γεγονός που σε συνδυασμό με την ηυξημένη ευελιξία των τεχνικών fuzzy clustering οδηγεί στην προφανή ανωτερότητα του FMSFA model σε σχέση με τους ανταγωνιστές του, όπως φαίνεται στο Σχ. 2 και τον Πίνακα 1.

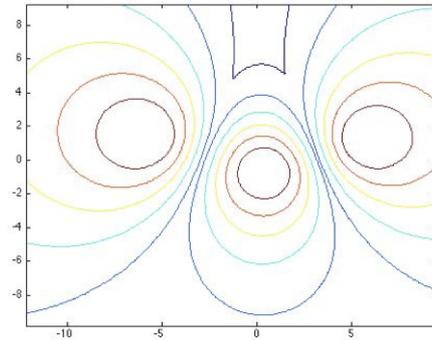
1.5.2 Αναγνώριση Χειρογράφων Ψηφίων

Εδώ εφαρμόζουμε το μοντέλο μας στην αναγνώριση χειρογράφων ψηφίων, χρησιμοποιώντας την *Optical Recognition of Handwritten Digits* database του UCI Repo-sitory of machine learning databases [34]. Για την δημιουργία της, 32×32 bitmaps χειρογράφων ψηφίων διαιρέθηκαν σε μη-επικαλυπτόμενα blocks 4×4 και ο αριθμός των pixels μετρήθηκε σε κάθε block. Έτσι ένα 64-διάστατο feature vector παρήχθη ανά δείγμα. Το σύνολο εκπαίδευσης περιλαμβάνει 3823 δείγματα, και το σύνολο αξιολόγησης 1797 δείγματα.

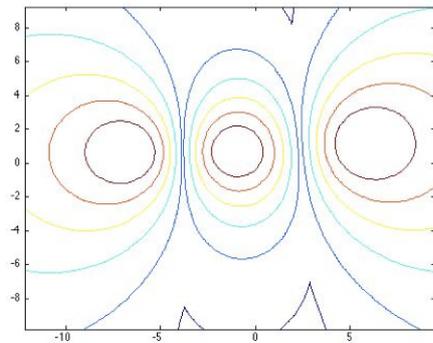
Στον Πίνακα 2 δίνουμε το ρυθμό σφάλματος ανά ψηφίο για κάθε μέθοδο, και για βέλτιστες c , q , και λ τιμές. Τα αποτελέσματα αυτά είναι μέσοι επί



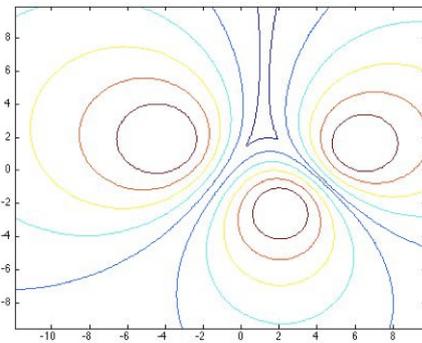
(i) Original clusters



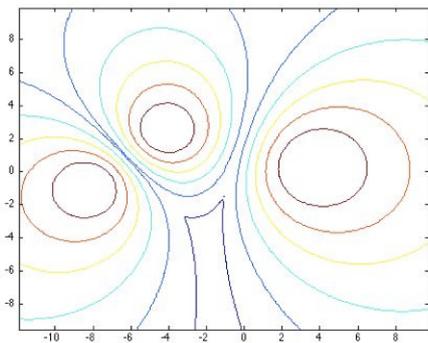
(ii) FMSFA



(iii) *t*MFA



(iv) KLFCV



(v) MFA

Σχήμα 1.2: Clustering Θορυβωδών Συνθετικών Δεδομένων: Περιγράμματα των λαμβανομένων clusters

Πίνακας 1.2: Αναγνώριση Χειρογράφων Ψηφίων: Ρυθμοί σφάλματος για βέλτιστη επιλογή των εκπαιδευομένων μοντέλων.

Model	c	q	λ	Μέσος Ρυθμός Σφάλματος
FMSFA	9	8	4.7	0.67
t MFA	9	8	-	1.32
KLFCV	12	9	18.6	1.20
MFA	12	9	-	2.10

30 εκτελέσεων των αλγορίθμων εκπαίδευσης από διαφορετικές εκκινήσεις. Διαπιστώνουμε ότι τα βασισμένα σε Student's- t μοντέλα έχουν καλύτερη επίδοση για μικρότερο μέγεθος μοντέλου, σε σχέση με τα GMM-βασισμένα μοντέλα. Ειδικότερα, το προτεινόμενο μοντέλο FMSFA έχει 68.10% καλύτερη επίδοση από το MFA και 44.17% καλύτερη επίδοση από το KLFCV. Τέλος, όπως παρατηρούμε, η μεγάλη ευελιξία του FMSFA model, λόγω της fuzzy φύσης του, οδηγεί σε μια μεγάλη βελτίωση σε σχέση με το μη-fuzzy αντίστοιχό του, t MFA μοντέλο, βελτιώνοντας την επίδοσή του κατά 49.24%.

1.5.3 Σημασιολογική Αναγνώριση Αντικειμένων σε Κατατμημένες Εικόνες

Εδώ εφαρμόζουμε το μοντέλο μας στην σημασιολογική αναγνώριση αντικειμένων σε κατατμημένες εικόνες, χρησιμοποιώντας τα *Image Segmentation data* του UCI Repository of machine learning databases [34]. Οι θεωρούμενες σημασιολογικές κλάσεις είναι *brickface*, *sky*, *foliage*, *cement*, *window*, *path*, και *grass*. Τα δεδομένα μας αντιστοιχούν σε 3×3 περιοχές, και περιλαμβάνουν 19 attributes σχετικά με χρωματικά και χωρικά χαρακτηριστικά των αντικει-

Πίνακας 1.3: Σημαιολογική Αναγνώριση Αντικειμένων σε Κατατμημένες Εικόνες: Ρυθμοί σφάλματος για βέλτιστη επιλογή των εκπαιδευομένων μοντέλων.

Model	c	q	λ	Μέσος Ρυθμός Σφάλματος
FMSFA	7	10	28.2	6.46
t MFA	7	10	-	9.72
KLFCV	7	12	14.4	11.57
MFA	7	12	-	16.73

μένων. Το σύνολο εκπαίδευσης περιλαμβάνει 150 δείγματα, και το σύνολο αξιολόγησης 1500 δείγματα.

Στον Πίνακα 3 δίνουμε το ρυθμό σφάλματος ανά κλάση για κάθε μέθοδο, και για βέλτιστες c , q , και λ τιμές. Τα αποτελέσματα αυτά είναι μέσοι επί 30 εκτελέσεων των αλγορίθμων εκπαίδευσης από διαφορετικές εκκινήσεις. Διαπιστώνουμε και εδώ ότι τα βασισμένα σε Student's- t μοντέλα έχουν καλύτερη επίδοση για μικρότερο μέγεθος μοντέλου, σε σχέση με τα GMM-βασισμένα μοντέλα. Ειδικότερα, το προτεινόμενο μοντέλο FMSFA έχει 61.39% καλύτερη επίδοση από το MFA και 44.17% καλύτερη επίδοση από το KLFCV. Τέλος, όπως παρατηρούμε, η μεγάλη ευελιξία του FMSFA model, λόγω της fuzzy φύσης του, οδηγεί σε μια σημαντική βελτίωση σε σχέση με το μη-fuzzy αντίστοιχό του, t MFA μοντέλο, βελτιώνοντας την επίδοσή του κατά 33.54%.

1.5.4 Εντοπισμός Spam E-Mail

Στην τελευταία αυτή εφαρμογή, εφαρμόζουμε το FMSFA και τους ανταγωνιστές του στον εντοπισμό spam e-mail. Στην απλή αυτή εφαρμογή μας

Πίνακας 1.4: Εντοπισμός Spam E-Mail: Ρυθμοί σφάλματος για βέλτιστη επιλογή των εκπαιδευομένων μοντέλων.

Model	c	q	λ	False-Positives (%)	False-Negatives (%)
FMSFA	10	23	32.1	12.67	7.24
t MFA	10	23	-	13.70	11.29
KLFCV	12	38	13.2	12.57	12.73
MFA	12	38	-	13.88	17.22

ενδιαφέρει η ελαχιστοποίηση των false-negatives, με διατήρηση των false-positives κάτω από 15%. Βάσει αυτών έγινε ο βέλτιστος προσδιορισμός των τιμών των c , q , και λ παραμέτρων για τα διάφορα μοντέλα.

Για το πείραμά μας χρησιμοποιούμε την *Spam E-mail Database* του UCI Repository of machine learning databases [34]. Περιλαμβάνει 4601 δείγματα, με το spam να αποτελεί το 39.4%. Υπάρχουν 57 attributes για κάθε δεδομένο, που προέκυψε από την γλωσσολογική ανάλυση ενός (spam ή non-spam) e-mail. Το σύνολο εκπαίδευσης περιλαμβάνει το 30% των δεδομένων, και το σύνολο αξιολόγησης το υπόλοιπο 70%.

Στον Πίνακα 4, παρέχουμε την επίδοση του κάθε μοντέλου. Τα αποτελέσματα αυτά είναι μέσοι επί 30 εκτελέσεων των αλγορίθμων εκπαίδευσης από διαφορετικές εκκινήσεις. Διαπιστώνουμε ότι τα fuzzy μοντέλα έχουν καλύτερη επίδοση από τα μη-fuzzy και τα Student's- t μοντέλα έχουν καλύτερη επίδοση από τα Gaussian.

1.5.5 Αξιολόγηση του Προσεγγιστικού Τύπου Υπολογισμού των Βαθμών Ελευθερίας

Τέλος, αξιολογούμε τον προσεγγιστικό τύπο υπολογισμού των βαθμών ελευθερίας, ν_i , των συνιστώντων analyzers το FMSFA model. Για τον σκοπό αυτό εκτελούμε τα προηγούμενα πειράματα δυο φορές, μια με χρήση της (57) υπό τον περιορισμό $\nu_\epsilon = \nu_\zeta = \nu \forall \zeta \neq \epsilon = 1, \dots, c$, από την οποία και παίρνουμε τις ακριβείς τιμές του ν , και μια με χρήση της (58), από την οποία και παίρνουμε τις προσεγγιστικές τιμές ν^* . Συγκρίνοντας τα αποτελέσματα παρατηρούμε ότι η προσέγγισή μας είναι πολύ ακριβής, καθώς παίρνουμε $|\nu^* - \nu| < 10^{-3}$, ενώ η επίδοση αναγνώρισης προτύπων του αλγορίθμου δεν αλλάζει με την αλλαγή μεταξύ των (57) και (58). Συνεπώς, η προσέγγιση (58) ενδείκνυται για χρήση, αρκεί ο περιορισμός $\nu_\epsilon = \nu_\zeta = \nu \forall \zeta \neq \epsilon = 1, \dots, c$ να είναι ανεκτός στην υπό εξέταση εφαρμογή.

Παράρτημα Α. Υπολογισμός Προσδοκητών Τιμών

Αρχικά εξάγουμε την posterior κατανομή των factor vectors, $\mathbf{y}_j | \mathbf{x}_j, u_j$. Από τον ορισμό του Student's- t factor analysis model, προκύπτει

$$p \left(\begin{bmatrix} \mathbf{x}_j \\ \mathbf{y}_j \end{bmatrix} | u_j \right) = \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\mu} \\ \mathbf{0} \end{bmatrix}, \frac{1}{u_j} \begin{bmatrix} \boldsymbol{\Lambda} \boldsymbol{\Lambda}^T + \boldsymbol{\Psi} & \boldsymbol{\Lambda} \\ \boldsymbol{\Lambda}^T & \mathbf{I}_q \end{bmatrix} \right) \quad (1.60)$$

Από την (60), προκύπτει απευθείας ότι η υπό συνθήκη κατανομή $\mathbf{y}_j | \mathbf{x}_j, u_j$ δίδεται απο

$$\mathbf{y}_j | \mathbf{x}_j, u_j \sim \mathcal{N}(\boldsymbol{\gamma}^T (\mathbf{x}_j - \boldsymbol{\mu}), \boldsymbol{\omega} / u_j) \quad (1.61)$$

όπου

$$\begin{aligned}\boldsymbol{\gamma} &= (\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi})^{-1}\boldsymbol{\Lambda} \\ \boldsymbol{\omega} &= \mathbf{I}_q - \boldsymbol{\gamma}^T\boldsymbol{\Lambda} \\ u_j &\sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)\end{aligned}$$

Από την (61), συνάγεται ότι για ένα mixture of Student's- t factor analyzers της μορφής (17), ισχύει

$$\mathbf{y}_{ij}|\mathbf{x}_j, u_{ij} \sim \mathcal{N}(\boldsymbol{\gamma}_i^T(\mathbf{x}_j - \boldsymbol{\mu}_i), \boldsymbol{\omega}_i/u_{ij}) \quad (1.62)$$

όπου

$$\begin{aligned}\boldsymbol{\gamma}_i &= (\boldsymbol{\Lambda}_i\boldsymbol{\Lambda}_i^T + \boldsymbol{\Psi}_i)^{-1}\boldsymbol{\Lambda}_i \\ \boldsymbol{\omega}_i &= \mathbf{I}_q - \boldsymbol{\gamma}_i^T\boldsymbol{\Lambda}_i \\ u_{ij} &\sim \mathcal{G}\left(\frac{\nu_i}{2}, \frac{\nu_i}{2}\right)\end{aligned}$$

Τότε, από την (62), έχουμε

$$\begin{aligned}\mathbf{E}_{\boldsymbol{\Theta}^{(k)}}(u_{ij}\mathbf{y}_{ij}|\mathbf{x}_j) &= \mathbf{E}_{\boldsymbol{\Theta}^{(k)}}(u_{ij}|\mathbf{x}_j)\mathbf{E}_{\boldsymbol{\Theta}^{(k)}}(\mathbf{y}_{ij}|\mathbf{x}_j, u_{ij}) = \\ &= u_{ij}^{(k)}\boldsymbol{\gamma}_i^{(k)T}(\mathbf{x}_j - \boldsymbol{\mu}_i^{(k)})\end{aligned} \quad (1.63)$$

$$\begin{aligned}\mathbf{E}_{\boldsymbol{\Theta}^{(k)}}(u_{ij}\mathbf{y}_{ij}\mathbf{y}_{ij}^T|\mathbf{x}_j) &= \mathbf{E}_{\boldsymbol{\Theta}^{(k)}}(\mathbf{y}_{ij}\mathbf{y}_{ij}^T|\mathbf{x}_j, u_{ij})\mathbf{E}_{\boldsymbol{\Theta}^{(k)}}(u_{ij}|\mathbf{x}_j) \\ &= \boldsymbol{\omega}_i^{(k)} + u_{ij}^{(k)}\boldsymbol{\gamma}_i^{(k)T}(\mathbf{x}_j - \boldsymbol{\mu}_i^{(k)})\left(\mathbf{x}_j - \boldsymbol{\mu}_i^{(k)}\right)^T\boldsymbol{\gamma}_i^{(k)}\end{aligned} \quad (1.64)$$

Επιπλέον χρειαζόμαστε την κατανομή $u_{ij}|\mathbf{x}_j$. Η posterior κατανομή των scalars u_{ij} , με prior κατανομή κατά Gamma, όπως περιγράφεται από την (22), δοθέντων των κατά Student's- t δεδομένων \mathbf{x}_j , με περιθώρια σ.π.π. που δίνε-

ται απο τις (29) και (30), έχει εξαχθεί στο [19], όπου αποδεικνύεται ότι

$$u_{ij}|\mathbf{x}_j \sim \mathcal{G}(\alpha_i, \beta_{ij})$$

όπου

$$\alpha_i = \frac{\nu_i + p}{2}$$

$$\beta_{ij} = \frac{\nu_i + \delta(\mathbf{x}_j, \boldsymbol{\mu}_i^{(k)}; \boldsymbol{\Sigma}_i^{(k)})}{2}$$

και το $\boldsymbol{\Sigma}_i^{(k)}$ δίνεται από την (31). Τότε, παίρνουμε [19]

$$u_{ij}^{(k)} \triangleq \mathbf{E}_{\Theta^{(k)}}(u_{ij}|\mathbf{x}_j) = \frac{\nu_i^{(k)} + p}{\nu_i^{(k)} + \delta(\mathbf{x}_j, \boldsymbol{\mu}_i^{(k)}; \boldsymbol{\Sigma}_i^{(k)})}$$

$$\mathbf{E}_{\Theta^{(k)}}(\log u_{ij}|\mathbf{x}_j) = \log u_{ij}^{(k)} - \log \left(\frac{\nu_i^{(k)} + p}{2} \right) +$$

$$+ \psi \left(\frac{\nu_i^{(k)} + p}{2} \right)$$

Παράρτημα Β. Εκτιμητική Παραμέτρων

Η εκτιμητική των fuzzy membership functions γίνεται με ελαχιστοποίηση της αντικειμενικής συνάρτησης $J_\lambda(\Theta, R)$ επί των r_{ij} υπό την συνθήκη $\sum_{i=1}^c r_{ij} = 1 \forall j = 1, \dots, n$. Εισάγοντας ένα πολλαπλασιαστή Lagrange η_j για κάθε δεδομένο ώστε να επιβάλλουμε την συνθήκη, έχουμε

$$\frac{\partial}{\partial r_{ij}} \left[J_\lambda - \sum_{j=1}^n \eta_j \left(\sum_{h=1}^c r_{hj} - 1 \right) \right] = 0$$

που δίνει την (49).

Ομοίως, ελαχιστοποιώντας την $J_\lambda(\Theta, R)$ ως προς π_i υπό την συνθήκη $\sum_{i=1}^c \pi_i = 1$, ήτοι, εισάγοντας ένα πολλαπλασιαστή Lagrange κ ώστε να επιβάλλουμε την συνθήκη, έχουμε

$$\frac{\partial}{\partial \pi_i} \left[J_\lambda - \kappa \left(\sum_{h=1}^c \pi_h - 1 \right) \right] = 0$$

που δίνει την (51).

Για τον υπολογισμό των μ_i έχουμε ότι η έκφραση του $J_\lambda(\Theta, R)$ συναρτηθεί μόνο του μ_i δίνει

$$J_\lambda(\Theta, R) = -\frac{1}{2} \sum_{i=1}^c \sum_{j=1}^n r_{ij} u_{ij}^{(k)} \left[-2\mathbf{x}_j^T \Psi_i^{-1} \mu_i + 2\mu_i^T \Psi_i^{-1} \Lambda_i \mathbf{y}_{ij}^{(k)} + \mu_i^T \Psi_i^{-1} \mu_i \right]$$

Καθώς λοιπόν

$$\frac{\partial \mu_i^T \Psi_i^{-1} \mu_i}{\partial \mu_i} = 2\Psi_i^{-1} \mu_i$$

$$\frac{\partial \mathbf{x}_j^T \Psi_i^{-1} \mu_i}{\partial \mu_i} = \Psi_i^{-1} \mathbf{x}_j$$

$$\frac{\partial \mu_i^T \Psi_i^{-1} \Lambda_i \mathbf{y}_{ij}^{(k)}}{\partial \mu_i} = \Psi_i^{-1} \Lambda_i \mathbf{y}_{ij}^{(k)}$$

εύκολα συνάγεται ότι η $\partial J_\lambda(\Theta, R) / \partial \mu_i = 0$ δίνει την (52).

Ομοίως για το Λ_i έχουμε

$$\frac{\partial(\mathbf{x}_j - \boldsymbol{\mu}_i)^T \boldsymbol{\Psi}_i^{-1} \Lambda_i \mathbf{E}_{\Theta^{(k)}}(u_{ij} \mathbf{y}_{ij} | \mathbf{x}_j)}{\partial \Lambda_i} = \boldsymbol{\Psi}_i^{-1} (\mathbf{x}_j - \boldsymbol{\mu}_i) \mathbf{E}_{\Theta^{(k)}}(u_{ij} \mathbf{y}_{ij} | \mathbf{x}_j)^T \quad (1.65)$$

$$\frac{\partial \text{trace} [\Lambda_i^T \boldsymbol{\Psi}_i^{-1} \Lambda_i \mathbf{E}_{\Theta^{(k)}}(u_{ij} \mathbf{y}_{ij} \mathbf{y}_{ij}^T | \mathbf{x}_j)]}{\partial \Lambda_i} = 2 \boldsymbol{\Psi}_i^{-1} \Lambda_i \mathbf{E}_{\Theta^{(k)}}(u_{ij} \mathbf{y}_{ij} \mathbf{y}_{ij}^T | \mathbf{x}_j) \quad (1.66)$$

από όπου η $\partial J_\lambda(\Theta, R)/\partial \Lambda_i = 0$ δίδει

$$\Lambda_i^{(k+1)} = \left[\sum_{j=1}^n r_{ij}^{(k)} (\mathbf{x}_j - \boldsymbol{\mu}_i^{(k)}) \mathbf{E}_{\Theta^{(k)}}(u_{ij} \mathbf{y}_{ij} | \mathbf{x}_j)^T \right] \times \left[\sum_{j=1}^n r_{ij}^{(k)} \mathbf{E}_{\Theta^{(k)}}(u_{ij} \mathbf{y}_{ij} \mathbf{y}_{ij}^T | \mathbf{x}_j) \right]^{-1} \quad (1.67)$$

Αντικαθιστώντας στην (67) τις εκφράσεις των $\mathbf{E}_{\Theta^{(k)}}(u_{ij} \mathbf{y}_{ij} | \mathbf{x}_j)$, από την (63), και $\mathbf{E}_{\Theta^{(k)}}(u_{ij} \mathbf{y}_{ij} \mathbf{y}_{ij}^T | \mathbf{x}_j)$, από την (64), αμέσως προκύπτει η (57).

Εν συνεχεία, για το $\boldsymbol{\Psi}_i$ έχουμε

$$\frac{\partial \log |\boldsymbol{\Psi}_i|}{\partial \boldsymbol{\Psi}_i^{-1}} = -\boldsymbol{\Psi}_i \quad (1.68)$$

$$\frac{\partial(\mathbf{x}_j - \boldsymbol{\mu}_i)^T \boldsymbol{\Psi}_i^{-1} (\mathbf{x}_j - \boldsymbol{\mu}_i)}{\partial \boldsymbol{\Psi}_i^{-1}} = \frac{\partial \text{trace} [\boldsymbol{\Psi}_i^{-1} (\mathbf{x}_j - \boldsymbol{\mu}_i) (\mathbf{x}_j - \boldsymbol{\mu}_i)^T]}{\partial \boldsymbol{\Psi}_i^{-1}} = (\mathbf{x}_j - \boldsymbol{\mu}_i) (\mathbf{x}_j - \boldsymbol{\mu}_i)^T \quad (1.69)$$

$$\frac{\partial(\mathbf{x}_j - \boldsymbol{\mu}_i)^T \boldsymbol{\Psi}_i^{-1} \boldsymbol{\Lambda}_i \mathbb{E}_{\Theta^{(k)}}(u_{ij} \mathbf{y}_{ij} | \mathbf{x}_j)}{\partial \boldsymbol{\Psi}_i^{-1}} = \quad (1.70)$$

$$(\mathbf{x}_j - \boldsymbol{\mu}_i) \mathbb{E}_{\Theta^{(k)}}(u_{ij} \mathbf{y}_{ij} | \mathbf{x}_j)^T \boldsymbol{\Lambda}_i^T$$

$$\frac{\partial \text{trace}(\boldsymbol{\Lambda}_i^T \boldsymbol{\Psi}_i^{-1} \boldsymbol{\Lambda}_i \mathbb{E}_{\Theta^{(k)}}(u_{ij} \mathbf{y}_{ij} \mathbf{y}_{ij}^T | \mathbf{x}_j))}{\partial \boldsymbol{\Psi}_i^{-1}} = \quad (1.71)$$

$$\boldsymbol{\Lambda}_i \mathbb{E}_{\Theta^{(k)}}(u_{ij} \mathbf{y}_{ij} \mathbf{y}_{ij}^T | \mathbf{x}_j) \boldsymbol{\Lambda}_i^T$$

από όπου, με χρήση των (63) - (64), συνάγεται ότι η $\partial J_\lambda(\boldsymbol{\Theta}, R) / \partial \boldsymbol{\Psi}_i^{-1} = 0$ έχει λύση την (59).

Ομοίως λαμβάνονται και οι εκτιμήτριες των $\nu_i^{(k)}$.

Βιβλιογραφία

- [1] L. Zadeh, "Fuzzy sets," *Inform. Contr.*, vol. 8, pp. 338--353, 1965.
- [2] R. Bellman, R. Kalaba, and L. Zadeh, "Abstraction and pattern classification," *J. Math. Anal. and Appl.*, vol. 2, pp. 581--586, 1966.
- [3] E. Ruspini, "A new approach to clustering," *Inform. Contr.*, vol. 15, pp. 22--32, 1969.
- [4] J. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated cluster," *J. Cybernetics*, vol. 3, pp. 32--57, 1974.
- [5] J. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York and London: Plenum Press, 1981.
- [6] D. E. Gustafson and W. C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," in *Proc. IEEE Conf. Decision Control*, vol. 2, 1979, pp. 761--766.
- [7] I. Gath and A. B. Geva, "Unsupervised optimal fuzzy clustering," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 7, pp. 773--781, 1989.
- [8] R. Li and M. Mukaidono, "A maximum entropy approach to fuzzy clustering estimation," in *Conf. on Fuzzy Systems (Fuzz'IEEE)*, 1995.

- [9] S. Miyamoto and M. Mukaidono, "Fuzzy c -means as a regularization and maximum entropy approach," in *Proceedings of the Seventh International Fuzzy Systems Association World Congress*, vol. 2, 1997, pp. 86--92.
- [10] H. Ichihashi, K. Honda, and N. Tani, "Gaussian mixture pdf approximation and fuzzy c -means clustering with entropy regularization," in *Proceedings of the Fourth Asian Fuzzy System Symposium*, 2000, pp. 217--221.
- [11] M.-T. Gan, M. Hanmandlu, and A. H. Tan, "From a Gaussian mixture model to additive fuzzy systems," *IEEE Trans. Fuzzy Systems*, vol. 13, no. 3, pp. 303--316, 2005.
- [12] G. McLachlan and D. Peel, *Finite Mixture Models*. New York: Wiley Series in Probability and Statistics, 2000.
- [13] J. C. Bezdek, N. R. Pal, J. Keller, and R. Krishnapuram, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. Norwell, MA, USA: Kluwer Academic Publishers, 1999.
- [14] A. Kosinski, "A procedure for the detection of multivariate outliers," *Computational Statistics and Data Analysis*, vol. 29, pp. 145--161, 1999.
- [15] R. N. Dave and R. Krishnapuram, "Robust clustering methods: A unified approach," *IEEE Trans. Fuzzy Syst.*, vol. 5, no. 2, pp. 270--293, 1997.
- [16] K. Honda and H. Ichihashi, "Component-wise robust linear fuzzy clustering for collaborative filtering," *Int. J. Approx. Reason.*, vol. 37, no. 2, pp. 127--144, 2004.

- [17] S. Miyamoto and A. C. Alanzado, "Fuzzy c -means and mixture distribution models in the presence of noise cluster," *Int. J. Image Graph.*, vol. 2, no. 4, pp. 573--586, 2002.
- [18] A. Alanzado and S. Miyamoto, "Fuzzy c -means and a mixture distribution for clustering of noisy data," in *Proceedings of Joint SCIS & ISIS Conference*, Tsukuba, Japan, 2002.
- [19] D. Peel and G. J. McLachlan, "Robust mixture modelling using the t distribution," *Statistics and Computing*, vol. 10, no. 4, pp. 339--348, 2000.
- [20] M. Tipping and C. Bishop, "Mixtures of probabilistic principal component analyzers," *Neural Computation*, vol. 11, no. 2, pp. 443--482, 1999.
- [21] A. Hyvärinen, "Survey on independent component analysis," *Neural Computing Surveys*, vol. 2, pp. 94--128, 1999.
- [22] Z. Ghahramani and G. Hinton, "The EM algorithm for mixtures of factor analyzers," Department of Computer Science, University of Toronto, Toronto, Canada, M5S 1A4, Tech. Rep. CRGTR- 96-1, 1997.
- [23] J. C. Bezdek, C. Coray, R. Gunderson, and J. Watson, "Detection and characterization of cluster substructure: Fuzzy c -varieties and convex combinations thereof," *SIAM J. Appl. Math.*, vol. 40, no. 2, pp. 358--372, 1981.
- [24] K. Honda and H. Ichihashi, "Regularized linear fuzzy clustering and probabilistic PCA mixture models," *IEEE Trans. Fuzzy Systems.*, vol. 13, no. 4, pp. 508--516, 2005.

- [25] G. McLachlan, R. Bean, and L. B.-T. Jones, "Extension of the mixture of factor analyzers model to incorporate the multivariate t -distribution," *Comp. Stat. Data Analysis*, vol. 51, no. 11, pp. 5327--5338, 2007.
- [26] Z. Khan and F. Dellaert, "Robust generative subspace modeling: The subspace t distribution," Georgia Institute of Technology, Atlanta, GA, Tech. Rep. GIT-GVU-04-11, 2004.
- [27] C. Liu and D. Rubin, "ML estimation of the t distribution using EM and its extensions, ECM and ECME," *Statistica Sinica*, vol. 5, no. 1, pp. 19--39, 1995.
- [28] R. J. Hathaway, "Another interpretation of the EM algorithm for mixture distributions," *Statist. Probab. Lett.*, vol. 4, pp. 53--56, 1986.
- [29] D. Tran and M. Wagner, "A robust clustering approach to fuzzy Gaussian mixture models for speaker identification," in *Proc. Third International Conference on Knowledge-Based Intelligent Information Engineering Systems*, 1999, pp. 337--340.
- [30] S. Shoham, "Robust clustering by deterministic agglomeration EM of mixtures of multivariate t distributions," *Pattern Recognition*, vol. 35, no. 55, pp. 1127--1142, 2002.
- [31] X. Meng and D. van Dyk, "The EM algorithm-an old folk song sung to a fast new tune (with discussion)," *Journal of the Royal Statistical Society B*, vol. 59, no. 3, pp. 511--567, 1997.
- [32] S. Weiss and C. Kulikowski, *Computer Systems That Learn*. San Francisco, CA, USA: Morgan Kaufmann, 1991.

- [33] E. Backer and A. Jain, "A clustering performance measure based on fuzzy set decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 3, pp. 66-75, 1981.
- [34] P. Murphy and D. Aha. (1994) UCI repository of machine learning databases. [Online]. Available: <http://mllearn.ics.uci.edu/databases>

Κεφάλαιο 2

Μια Εύρωστη Μπεϋζιανή Τεχνική Μοντελοποίησης Πολυδιάστατων Σημάτων

2.1 Προοίμιο

Στο προηγούμενο κεφάλαιο μελετήσαμε την χρήση της factor analysis υπό το υπόδειγμα του fuzzy clustering. Προτείνοντας μια καινοτόμα προσέγγιση στην διατύπωση του fuzzy clustering αλγορίθμου, που επιτρέπει την εισαγωγή στη διαδικασία του fuzzy clustering των υποθέσεων για την πιθανοτική κατανομή των εξακτέων clusters, συνάγαμε ένα πρωτότυπο fuzzy clustering αλγόριθμο που στην ουσία αποτελεί μια fuzzy clustering μεθοδολογία εκπαίδευσης, εναλλακτική του EM αλγορίθμου, για finite mixtures of Student's- t factor analyzers (ο οποίος, τελικά, αποτελεί μια statistical clustering μεθοδολογία εκπαίδευσης).

Στην εργασία αυτή συνεχίζουμε την ενασχόλησή μας με τις πεπερασμένες υπερθέσεις κατά Student's- t κατανεμημένων αναλυτών παραγόντων, υπό ένα διαφορετικό όμως πρίσμα. Συγκεκριμένα, προτείνουμε μια καινοτόμο Μπεϋζιανή (Bayesian) μεθοδολογία εκτιμητικής του μοντέλου αυτού. Η συγκεκριμένη μεθοδολογία στηρίζεται στην διατύπωση ενός υπολογιστικά αποδοτικού variational inference αλγορίθμου για το υπό μελέτη μοντέλο, που γίνεται εφικτή δια της διατύπωσης των κατά Student's- t κατανεμημένων factor analyzers ως μιας περιθωριοποίησης (marginalization) επί προσθέτων κρυφών μεταβλητών. Η καινοτόμος προσέγγισή μας είναι πιο εύρωστη από τον EM αλγόριθμο, επιλύει τα singularity και overfitting προβλήματά του, και καθιστά εύκολη τον αυτόματο προσδιορισμό του καταλλήλου μεγέθους για το μοντέλο.

Τα αποτελέσματα της εργασίας αυτής έχουν δημοσιευθεί υπό το άρθρο μου:

Sotirios Chatzis, Dimitrios Kosmopoulos, Theodora Varvarigou, "Signal Modeling and Classification Using a Robust Latent Space Model Based on t Distributions," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 949-963, March 2008.

2.2 Εισαγωγή

Το μοντέλο finite mixture of Student's- t factor analyzers και τα πλεονεκτήματά του έχει ήδη παρουσιαστεί στο Κεφ. 1. Στην εργασία αυτή, παρέχουμε έναν πρωτότυπο αλγόριθμο statistical-clustering τύπου εκτιμητικής του finite mixture of Student's- t factor analyzers model υπό το Bayesian υπόδειγμα με χρήση μιας variational προσέγγισης, τοιουτοτρόπως λαμβάνοντας το

Variational Bayes Mixture of Student's-t Factor Analyzers (VB-MSFA) model. Σε σχέση με τις μεθόδους εκτιμητικής υπό το υπόδειγμα μεγίστης πιθανοφανείας, με χρήση του *EM* αλγορίθμου ή παραλλαγών του (δες, π.χ., [1]), αξίζει να σημειωθεί ότι, λόγω της ανεπιθύμητης ιδιότητας της εκτιμητικής μεγίστης πιθανοφανείας (και άρα του *EM* αλγορίθμου) να βελτιστοποιεί μια μη-φραγμένη αντικειμενική συνάρτηση, καθώς η συνάρτηση πιθανοφανείας (likelihood function) είναι μη φραγμένη [2--4] έχει πολύ σημαντικές παρενέργειες. Κατ' αρχήν, ο *EM* αλγόριθμος εύκολα μπορεί να παγιδευτεί σε τοπικά μέγιστα, και συχνά πολλές δοκιμές και εφαρμογή των υπολογιστικά απαιτητικών τεχνικών cross-validation απαιτούνται προκειμένου να βρεθεί ένα καλό μέγιστο. Μια περαιτέρω δυσκολία αφορά τα infinities που "λυμαίνονται" την likelihood function, τα οποία έχουν να κάνουν με την κατάρρευση των παραβολοειδών-εκθετικών συνιστωσών το μοντέλο κατανομών σε μεμονωμένα σημεία (δεδομένα) εκπαίδευσης, και τα οποία κατά συνέπεια οδηγούν σε singular ή σχεδόν singular πίνακες συνδιακύμανσης [4]. Τέλος, ένα εξίσου σημαντικό θέμα είναι η αυτόματη επιλογή του πλήθους των συνιστωσών κατανομών: οι μέθοδοι μεγίστης πιθανοφανείας αποτυγχάνουν σε αυτό τον τομέα, καθώς η μη-φραγμένη μορφή της λογαριθμο-πιθανοφανείας τους καθιστά επιρρεπείς στην εύνοια μοντέλων με αυξημένη πολυπλοκότητα, και συνεπώς, με σημαντικά προβλήματα overfitting.

Στην εργασία αυτή παρέχουμε μια *Bayesian επίλυση* για το mixture of Student's-*t* factor analyzers μοντέλο, που υπερβαίνουν τα ως άνω προβλήματα με κομψό τρόπο, μέσω της περιθωριοποίησης (marginalization) επί των παραμέτρων του μοντέλου ως προς κατάλληλες prior κατανομές, και μεγιστοποίησης της προκύπτουσας περιθώριας πιθανοφανείας (marginal likeli-

hood) του μοντέλου σε σχέση με τον αριθμό των συνιστωσών του κατανομών, για υπολογισμό του βελτίστου μεγέθους του μοντέλου. Η προσέγγισή μας βασίζεται σε μια μέθοδο *variational προσέγγισης* [5], οι οποίες έχουν πρόσφατα αναδειχθεί ως μια deterministic εναλλακτική των Markov chain Monte-Carlo (MCMC) αλγορίθμων Bayesian εκτιμητικής [6,7], με πολύ καλύτερες επιδόσεις σε ότι αφορά τα υπολογιστικά τους κόστη [8]. Η Variational Bayesian inference έχει προηγουμένως εφαρμοστεί στα GMMs (e.g. [9]), αυτοπαλινδρομικά μοντέλα [10,11], και πεπερασμένες υπερθέσεις κατανομών Student's-*t* (SMMs) [12,13], παρέχοντας μια αποδοτική και επιτυχημένη λύση στα singularity και overfitting προβλήματα του EM αλγορίθμου.

Το υπόλοιπο του παρόντος κεφαλαίου έχει οργανωθεί ως εξής: Στην ενότητα 2.3, αφού παρέχουμε μια σύντομη ανασκόπηση των εξισώσεων του παρουσιασθέντος στο προηγούμενο κεφάλαιο mixture of Student's-*t* factor analyzers (MSFA) μοντέλου, για την διευκόλυνση του αναγνώστη, εισάγουμε την διαμόρφωση του MSFA model υπό το Bayesian υπόδειγμα. Στην ενότητα 2.4, εξάγουμε την Bayesian εκτιμητική του MSFA model με χρήση μιας προσέγγισης *variational inference*, τοιουτοτρόπως λαμβάνοντας το προτεινόμενο στην εργασία αυτή *Variational Bayes-MSFA* (VB-MSFA) model. Στην ενότητα 2.5, το προτεινόμενο μοντέλο αξιολογείτε πειραματικά σε ένα πλήθος εφαρμογών μοντελοποίησης σήματος από διαφορετικά πεδία εφαρμογών.

2.3 Διατύπωση Προβλήματος

2.3.1 Mixtures of Student's- t Factor Analyzers

Παρακάτω παρέχουμε μια σύντομη ανασκόπηση των εξισώσεων του MSFA μοντέλου για διευκόλυνση του αναγνώστη και σύνδεση με το υπόλοιπο κεφάλαιο. Έστω $\mathbf{x}_1, \dots, \mathbf{x}_n$ τυχαίο δείγμα μεγέθους n από ένα P -διάστατο χώρο \mathbf{X} . Η factor analysis μοντελοποιεί τις παρατηρήσεις $\mathbf{x}_j, j = 1, \dots, n$, ως

$$\mathbf{x}_j = \boldsymbol{\mu} + \boldsymbol{\Lambda} \mathbf{y}_j + \mathbf{e}_j \quad (2.1)$$

όπου \mathbf{y}_j είναι το Q -διάστατο ($Q < P$) διάνυσμα των factors, $\boldsymbol{\mu}$ είναι ο μέσος των παρατηρήσεων \mathbf{x}_j , $\boldsymbol{\Lambda}$ είναι ο $P \times Q$ πίνακας των factor loadings, και \mathbf{e}_j το σφάλμα. Στην Student's- t factor analysis [1], ισχύει

$$\mathbf{y}_j \sim t(\mathbf{0}, \mathbf{I}_Q, \nu) \quad (2.2)$$

για τα factor vectors, και

$$\mathbf{e}_j \sim t(\mathbf{0}, \boldsymbol{\Psi}, \nu) \quad (2.3)$$

όπου $\boldsymbol{\Psi}$ είναι ένας διαγώνιος πίνακας, $\boldsymbol{\Psi} = \text{diag}(\sigma_1^2, \dots, \sigma_P^2)$, και ν οι βαθμοί ελευθερίας της t κατανομής.

Η σ.π.π. της t κατανομής $t(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ με μέσο $\boldsymbol{\mu}$, πίνακα συνδιακύμανσης $\boldsymbol{\Sigma}$, και ν βαθμούς ελευθερίας δίνεται από

$$t(\mathbf{x}_j | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma\left(\frac{\nu+P}{2}\right) |\boldsymbol{\Sigma}|^{-1/2}}{(\pi\nu)^{P/2} \Gamma(\nu/2) \{1 + d(\mathbf{x}_j, \boldsymbol{\mu} | \boldsymbol{\Sigma})/\nu\}^{(\nu+P)/2}} \quad (2.4)$$

όπου $\Gamma(s)$ η Gamma function και $d(\mathbf{x}_j, \boldsymbol{\mu} | \boldsymbol{\Sigma})$

$$d(\mathbf{x}_j, \boldsymbol{\mu} | \boldsymbol{\Sigma}) = (\mathbf{x}_j - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_j - \boldsymbol{\mu}) \quad (2.5)$$

Επιπλέον [15], ισχύει

$$t(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \int_0^\infty \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma}/u) \mathcal{G}(u | \nu/2, \nu/2) du \quad (2.6)$$

όπου

$$u \sim \mathcal{G}(\nu/2, \nu/2) \quad (2.7)$$

και

$$\mathcal{G}(u | \alpha, \beta) = u^{\alpha-1} \frac{\beta^\alpha e^{-\beta u}}{\Gamma(\alpha)} \quad (2.8)$$

Έτσι, χρησιμοποιώντας την (6) παίρνουμε

$$\mathbf{y}_j | u_j \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_Q / u_j) \quad (2.9)$$

και

$$\mathbf{e}_j | u_j \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi} / u_j) \quad (2.10)$$

όπου $u_j \sim \mathcal{G}(\nu/2, \nu/2)$, από τις οποίες και συνάγεται ότι η κατανομή των παρατηρήσεων δοθέντων των factor vectors δίνεται από

$$\mathbf{x}_j | \mathbf{y}_j \sim t(\boldsymbol{\mu} + \boldsymbol{\Lambda} \mathbf{y}_j, \boldsymbol{\Psi}, \nu) \quad (2.11)$$

ή, εναλλακτικά

$$\mathbf{x}_j | \mathbf{y}_j, u_j \sim \mathcal{N}(\boldsymbol{\mu} + \boldsymbol{\Lambda} \mathbf{y}_j, \boldsymbol{\Psi} / u_j) \quad (2.12)$$

Ολοκληρώνοντας ως προς \mathbf{y}_j , λαμβάνεται η περιθώρια κατανομή των παρατηρήσεων, που δίνεται από

$$\mathbf{x}_j \sim t(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) \quad (2.13)$$

ή, εναλλακτικά

$$\mathbf{x}_j | u_j \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}/u_j) \quad (2.14)$$

όπου το $\boldsymbol{\Sigma}$ δίνεται από

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi} \quad (2.15)$$

Καθ' αυτό τον τρόπο, αν θεωρήσουμε μια g -συνιστωσών υπέρθεση από Student's- t factor analyzers, με μέσους $\boldsymbol{\mu}_i$, factor loading πίνακες $\boldsymbol{\Lambda}_i$, συνδιακυμάνσεις σφάλματος $\boldsymbol{\Psi}_i$, και βάρη $\pi_i, i = 1, \dots, g$, παίρνουμε

$$\mathbf{x}_j = \boldsymbol{\mu}_i + \boldsymbol{\Lambda}_i \mathbf{y}_{ij} + \mathbf{e}_{ij} \text{ with probability } \pi_i \quad (2.16)$$

όπου \mathbf{y}_{ij} είναι ο factor vector που αντιστοιχεί στην j οστη παρατήρηση ως προς τον i οστο analyzer

$$\mathbf{y}_{ij} \sim t(\mathbf{0}, \mathbf{I}_Q, \nu_i) \quad (2.17)$$

και \mathbf{e}_{ij} είναι το σφάλμα που αντιστοιχεί στην j οστη παρατήρηση ως προς τον i οστο analyzer

$$\mathbf{e}_{ij} \sim t(\mathbf{0}, \boldsymbol{\Psi}_i, \nu_i) \quad (2.18)$$

Έτσι, συνάγεται ότι

$$p(\mathbf{x}_j | \{\pi_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \nu_i\}_{i=1}^g) = \sum_{i=1}^g \pi_i t(\mathbf{x}_j | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \nu_i) \quad (2.19)$$

ή, εναλλακτικά

$$p(\mathbf{x}_j | \mathbf{u}_j; \{\pi_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}_{i=1}^g) = \sum_{i=1}^g \pi_i \mathcal{N}(\mathbf{x}_j | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i / u_{ij}) \quad (2.20)$$

όπου

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i \boldsymbol{\Lambda}_i^T + \boldsymbol{\Psi}_i \quad (2.21)$$

$$u_{ij} \sim \mathcal{G}(\nu_i/2, \nu_i/2) \quad (2.22)$$

και $\mathbf{u}_j = (u_{ij})$. Εδώ, u_{ij} είναι το κατά Gamma κατανεμημένο scalar της j οστης παρατήρηση ως προς τον i οστο analyzer. Τέλος, υπό την συνθήκη των factor vectors \mathbf{y}_{ij} λαμβάνεται

$$p(\mathbf{x}_j | \mathbf{Y}_j; \{\pi_i, \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i, \boldsymbol{\Psi}_i, \nu_i\}_{i=1}^g) = \sum_{i=1}^g \pi_i t(\mathbf{x}_j | \boldsymbol{\mu}_i + \boldsymbol{\Lambda}_i \mathbf{y}_{ij}, \boldsymbol{\Psi}_i, \nu_i) \quad (2.23)$$

$$p(\mathbf{x}_j | \mathbf{Y}_j, \mathbf{u}_j; \{\pi_i, \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i, \boldsymbol{\Psi}_i\}_{i=1}^g) = \sum_{i=1}^g \pi_i \mathcal{N}(\mathbf{x}_j | \boldsymbol{\mu}_i + \boldsymbol{\Lambda}_i \mathbf{y}_{ij}, \boldsymbol{\Psi}_i / u_{ij}) \quad (2.24)$$

όπου \mathbf{Y}_j είναι ο πίνακας των factors, $\mathbf{Y}_j = (\mathbf{y}_{1j} \dots \mathbf{y}_{gj})$.

2.3.2 To Variational Bayes Mixture of Student's- t Factor Analyzers

Έστω ένα σύνολο παρατηρήσεων $X = \{\mathbf{x}_j\}_{j=1}^n$ ειλημμένων από ένα mixture of Student's- t factor analyzers (MSFA model) με κατανομή που δίνεται από την (19). Από τις (19) ή (23) είναι φανερό ότι ένα MSFA model έχει στην ουσία την μορφή ενός SMM. Όπως έχει συζητηθεί και προηγουμένα [4], δεν υπάρχει κλειστού τύπου μεγιστοποίηση της likelihood ενός SMM σε αυτή την μορφή. Ωστόσο, αυτό μπορεί να γίνει εφικτό για ένα Student's- t mixture,

και, άρα, για το MSFA model, με αξιοποίηση της (6) [13], ήτοι με χρήση των εκφράσεων (20) και (24) για την κατανομή του MSFA model, και, συνεπώς, θέτοντας τις scaling μεταβλητές u_{ij} ως κρυφές μεταβλητές του μοντέλου όπου μια Gamma prior έχει επιβληθεί.

Προκειμένου να εξάγουμε μια κομψή variational μέθοδο εκτιμητικής του μοντέλου MSFA, θα χρησιμοποιήσουμε την υπό συνθήκη των factor vectors, \mathbf{y}_{ij} , και των scale vectors, \mathbf{u}_j , έκφραση της κατανομής του MSFA model, που δίνεται από την (24), και θα την επανεκφράσουμε ως μια περιθωριοποίηση (marginalization) ως προς ένα σύνολο από επιπρόσθετες δυαδικές μεταβλητές δηλώνουσες από ποιόν συνιστώντα το μοντέλο factor analyzer έχει προέλθει (παρήχθη) καθεμία από τις παρατηρήσεις \mathbf{x}_j , $j = 1, \dots, n$. Έστω $Z = \{\mathbf{z}_j\}_{j=1}^n$ των διανυσμάτων ενδεικτών προέλευσης, $\mathbf{z}_j = (z_{ij})$, τ.ω. $z_{ij} \in \{0, 1\}$ και $z_{ij} = 1$ αν το \mathbf{x}_j θεωρείται ότι παρήχθη από τον i οστο συνιστώντα analyzer, $z_{ij} = 0$ ειδάλλως. Συμβολίζουμε ακόμα ως $U = \{\mathbf{u}_j\}_{j=1}^n$ το σύνολο των scale vectors, και ως $Y = \{\mathbf{Y}_j\}_{j=1}^n$ το σύνολο των πινάκων των factors. Κατά συνέπεια, για κάθε παρατήρηση, \mathbf{x}_j , $j = 1, \dots, n$, οι αντίστοιχες κρυφές μεταβλητές περιλαμβάνουν τους πίνακες των factors \mathbf{Y}_j , το scale vector \mathbf{u}_j , και το διάνυσμα ενδεικτών προέλευσης \mathbf{z}_j . Τότε, από την (24) και τις (17), (22), προκύπτει ότι για ένα δεδομένο πλήθος από συνιστώντες analyzers, g , η κρυφών μεταβλητών διαμόρφωση του MSFA model διατυπώνεται ως εξής:

$$p(\mathbf{z}_j | \boldsymbol{\pi}) = \prod_{i=1}^g \pi_i^{z_{ij}} \quad (2.25)$$

$$p(\mathbf{u}_j | \mathbf{z}_j, \boldsymbol{\nu}) = \prod_{i=1}^g \mathcal{G}(u_{ij} | \nu_i/2, \nu_i/2)^{z_{ij}} \quad (2.26)$$

$$p(\mathbf{Y}_j | \mathbf{u}_j, \mathbf{z}_j) = \prod_{i=1}^g \mathcal{N}(\mathbf{y}_{ij} | \mathbf{0}, \mathbf{I}_Q / u_{ij})^{z_{ij}} \quad (2.27)$$

$$p(\mathbf{x}_j | \mathbf{Y}_j, \mathbf{u}_j, \mathbf{z}_j; \{\boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i, \boldsymbol{\Psi}_i\}_{i=1}^g) = \prod_{i=1}^g \mathcal{N}(\mathbf{x}_j | \boldsymbol{\mu}_i + \boldsymbol{\Lambda}_i \mathbf{y}_{ij}, \boldsymbol{\Psi}_i / u_{ij})^{z_{ij}} \quad (2.28)$$

όπου $\boldsymbol{\pi} = (\pi_i)$, $\boldsymbol{\nu} = (\nu_i)$. Πλέον τούτων, η Bayesian διατύπωση του MSFA model μπορεί να ολοκληρωθεί επιβάλλοντας κατάλληλες prior κατανομές στις παραμέτρους του μοντέλου. Είναι βολικό και δόκιμο να επιλέξουμε conjugate priors ως προς τους όρους (25)-(28) [5]. Υπό αυτό το πρίσμα, το διάνυσμα των βαρών ανάμειξης επιλέγεται να ακολουθεί κατανομή Dirichlet (δηλ. conjugate στην πολυωνυμική κατανομή $p(\mathbf{z}_j | \boldsymbol{\pi})$, που δίνεται από την (25))

$$p(\boldsymbol{\pi} | \mathbf{a}) = \mathcal{D}(\boldsymbol{\pi} | \mathbf{a}) = \frac{\Gamma(a_0)}{\prod_{i=1}^g \Gamma(a_i)} \prod_{i=1}^g \pi_i^{a_i - 1} \quad (2.29)$$

όπου $a_0 = \sum_i a_i$. Σχετικά με τα factor loading matrices, $\boldsymbol{\Lambda}_i$, επιλέγουμε μια ιεραρχική conjugate prior κατανομή, ώστε να επιτρέψουμε την διενέργεια *automatic relevance determination (ARD)* [16]. Για τον λόγο αυτό, σε κάθε στήλη κάθε factor loading matrix επιβάλλεται μια Gaussian prior με μηδενικό μέσο και μια διαφορετική υπερπαραμέτρο ακριβείας (precision, αντίστροφη της συνδιακύμανσης), ήτοι

$$p(\boldsymbol{\Lambda}_i | \boldsymbol{\phi}_i) = \prod_{k=1}^Q \mathcal{N}(\boldsymbol{\lambda}_{ik} | \mathbf{0}, \mathbf{I}_P / \phi_{ik}) \quad (2.30)$$

όπου $\boldsymbol{\lambda}_{ik}$ είναι η k οστή στήλη του $\boldsymbol{\Lambda}_i$ και ϕ_{ik} η precision παράμετρος για κάθε στοιχείο της αντιστοίχου στήλης. Εφόσον ο αριθμός των υπερπαραμέτρων που περιλαμβάνονται στο precision vector, $\boldsymbol{\phi}_i = (\phi_{ik})$, ενός factor loading

matrix αυξάνεται με τον αριθμό των factors, Q , επιλέγουμε να επιβάλλουμε στις υπερπαραμέτρους ϕ_{ik} μια Gamma υπερ-prior. Με άλλα λόγια, επιλέγουμε

$$p(\boldsymbol{\phi}_i|\gamma_0, \omega_0) = \prod_{k=1}^Q \mathcal{G}(\phi_{ik}|\gamma_0, \omega_0) \quad (2.31)$$

όπου γ_0 και ω_0 είναι οι υπερ-υπερπαραμέτροι της Gamma prior που επιβάλλεται σε κάθε precision vector $\boldsymbol{\phi}_i$. Ας σημειωθεί ότι, καθόσον η σφαιρική Gaussian prior (30) is είναι διαμερίσιμη στις P διαστάσεις της, μπορεί ισοδύναμα να αναλυθεί ως προς τις γραμμές της λαμβάνοντας

$$p(\boldsymbol{\Lambda}_i|\boldsymbol{\phi}_i) = \prod_{l=1}^P \mathcal{N}(\boldsymbol{\lambda}_{il}|\mathbf{0}, \text{diag}(\boldsymbol{\phi}_i)^{-1}) \quad (2.32)$$

όπου $\boldsymbol{\lambda}_{il}$ είναι η l οστη γραμμή της $\boldsymbol{\Lambda}_i$. Σημειώνεται επίσης ότι περιοριζώντας τις $\boldsymbol{\Lambda}_i$ επί των $\boldsymbol{\phi}_i$, με χρήση των (30) και (31), προκύπτει ότι η περιοθώριες priors των factor loading matrices, $\boldsymbol{\Lambda}_i$, είναι επίσης κατανεμημένες κατά Student's- t . Στην ενότητα 2.4.4 θα δείξουμε πως η ARD μέθοδος [16], που συνίσταται ακριβώς στην επιβολή της prior (30) επί των factor loading matrices, με υπερπαραμέτρους κατανεμημένες κατά την (31), παρέχει μια υπολογιστικά αποδοτική μέθοδο για την αυτόματη εύρεση του καταλλήλου αριθμού factors για κάθε factor analyzer. Τέλος, οι conjugate priors για τους μέσους των factor analyzers επιλέγονται ως

$$p(\boldsymbol{\mu}_i|\mathbf{m}_0, \mathbf{s}_0) = \mathcal{N}(\boldsymbol{\mu}_i|\mathbf{m}_0, \text{diag}(\mathbf{s}_0)^{-1}) \quad (2.33)$$

Υπογραμμίζουμε ότι στην ανωτέρω ανάλυση δεν επιβάλλαμε prior στον αριθμό των συνιστωσών analyzers, g , ο οποίος θεωρήθηκε ως δεδομένος.

Η παράμετρος αυτή θα εξαχθεί μέσω μεγιστοποίησης του variational κάτω φράγματος της marginal likelihood του MSFA model, όπως θα συζητηθεί στην ενότητα 2.4.4. Ομοίως δεν επιβάλλαμε prior στον αριθμό των factors, Q , κάθε factor analyzer, αφού, όπως εξηγήσαμε, αυτός θα ελέγχεται μέσω της ARD μεθόδου. Για ευκολία δεν επιβάλλουμε prior στους πίνακες συνδιακύμανσης των σφαλμάτων Ψ_i . Τέλος, δεν υφίσταται conjugate prior για τα διανύσματα βαθμών ελευθερίας, ν . Οι τελευταίες δυο παράμετροι του MSFA model θα βελτιστοποιούνται ως υπερπαράμετροι του μοντέλου, όπως θα εξηγηθεί στην συνέχεια.

2.4 Variational Bayesian Inference

Έχοντας εισάγει και prior κατανομές επί των παραμέτρων του MSFA model, η διατύπωση του *Variational Bayes-MSFA* (VB-MSFA) model έχει ολοκληρωθεί. Συνεπώς, μπορούμε να προχωρήσουμε στην εύρεση της έκφρασης της marginal likelihood του μοντέλου. Η ακριβής έκφραση της marginal likelihood του μοντέλου μας είναι υπολογιστικά απαγορευτική. Ωστόσο, η επιλογή conjugate exponential prior κατανομών για τις παραμέτρους του μοντέλου μας επιτρέπει την διατύπωση ενός κομψού variational υποδείγματος επίλυσης.

Έστω $\theta_S = (\pi, \mu_1, \dots, \mu_g, \Lambda_1, \dots, \Lambda_g, \phi_1, \dots, \phi_g)$ το σύνολο των παραμέτρων του μοντέλου, και θ το σύνολο όλων των στοχαστικών μεταβλητών που σχετίζονται με το VB-MSFA model, ήτοι $\theta = (\theta_S, Y, U, Z)$. Η variational Bayesian επίλυση του VB-MSFA model μπορεί, λοιπόν, να γίνει με την εισαγωγή μιας αυθαίρετης κατανομής $q(\theta) = q(Y, U, Z, \theta_S)$ και την θεώρηση

της γνωστής ιδιότητας της log marginal likelihood (log evidence), $\log p(X)$ [8]

$$\log p(X) = \mathcal{L}(q) + \text{KL}(q||p) \quad (2.34)$$

όπου

$$\mathcal{L}(q) = \int q(\boldsymbol{\theta}) \log \frac{p(X, \boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \quad (2.35)$$

Στην (34), $\text{KL}(q||p)$ είναι η Kullback-Leibler (KL) divergence μεταξύ της αυθαίρετης κατανομής $q(\boldsymbol{\theta})$, η οποία και αποτελεί την variational (προσεγγιστική) posterior επί των παραμέτρων του μοντέλου, και της $p(\boldsymbol{\theta}|X)$ που είναι η πραγματική posterior επί των παραμέτρων του μοντέλου. Ισχύει

$$\text{KL}(q||p) = - \int q(\boldsymbol{\theta}) \log \frac{p(\boldsymbol{\theta}|X)}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \quad (2.36)$$

Καθόσον η KL divergence είναι μη αρνητική ποσότητα, εκ της (36) συνάγεται ότι η $\mathcal{L}(q)$ αποτελεί ένα κάτω φράγμα της log evidence, i.e.

$$\log p(X) \geq \mathcal{L}(q) \quad (2.37)$$

Συνεπώς, μεγιστοποιώντας αυτό το κάτω φράγμα της log evidence, $\mathcal{L}(q)$, ώστε να γίνει το δυνατόν πιο σφικτό, ήτοι, ελαχιστοποιώντας την KL divergence μεταξύ της πραγματικής και της variational posterior, μια καλή variational προσέγγιση στο VB-MSFA model μπορεί να εξαχθεί.

Προκειμένου να διαμορφώσουμε μια υπολογιστικά αποδοτική έκφραση για το κάτω φράγμα της log evidence του VB-MSFA model, υποθέτουμε ότι η από κοινού variational posterior κατανομή των στοχαστικών μεταβλητών του VB-MSFA model, $q(\boldsymbol{\theta}) = q(Y, U, Z, \boldsymbol{\theta}_S)$, παραγοντοποιείται επί των

κρυφών μεταβλητών και των παραμέτρων του μοντέλου, ήτοι

$$q(\boldsymbol{\theta}) = q(Y, U, Z, \boldsymbol{\theta}_S) \approx q(Y, U, Z)q(\boldsymbol{\theta}_S) \quad (2.38)$$

Επιπλέον δεχόμαστε ότι

$$q(\boldsymbol{\theta}_S) \approx q(\boldsymbol{\pi}) \prod_{i=1}^g q(\boldsymbol{\mu}_i)q(\boldsymbol{\phi}_i)q(\boldsymbol{\Lambda}_i) \quad (2.39)$$

οπότε, τελικά, η υπόθεση μας διαμορφώνεται ως

$$q(\boldsymbol{\theta}) = q(Y, U, Z, \boldsymbol{\theta}_S) \approx q(Y, U, Z)q(\boldsymbol{\pi}) \prod_{i=1}^g q(\boldsymbol{\mu}_i)q(\boldsymbol{\phi}_i)q(\boldsymbol{\Lambda}_i) \quad (2.40)$$

Η παραγοντοποίηση του $q(\boldsymbol{\theta})$ στην μορφή (40) είναι μια κοινότυπη προσέγγιση στο χώρο του variational Bayesian inference, και μπορεί ναδειχθεί ότι είναι σωστή, ως απορρέουσα από την προηγούμενη θεώρηση μας για την μορφή της από κοινού conjugate prior κατανομής (που επίσης έχει στην ουσία παραγοντισθεί ομοίως με την (40)) [12].

Έχοντας επιλέξει μια οικογένεια προσεγγιστικών (variational) posterior κατανομών, μπορούμε να αναζητήσουμε το βέλτιστο μέλος αυτής της οικογένειας δια της μεγιστοποίησης του κάτω φράγματος της log marginal likelihood (variational lower bound), $\mathcal{L}(q)$. Το υπόλοιπο της παρούσας ενότητας έχει ως ακολούθως: Στην επόμενη ενότητα, θα δώσουμε την πλήρη έκφραση του variational lower bound για το VB-MSFA model, $\mathcal{L}(q)$. Στην ενότητα 2.4.2, οι εκφράσεις των variational posteriors των στοχαστικών μεταβλητών του μοντέλου model (παράμετροι και κρυφές μεταβλητές) θα εξαχθούν δια της μεγιστοποίησης του variational lower bound. Στην ενότητα 2.4.3, θα

συζητηθεί το πρόβλημα της επιλογής τιμών για τις υπερπαραμέτρους του μοντέλου. Στην ενότητα 2.4.4, θα περιγραφεί η διαδικασία επιλογής μεγέθους μοντέλου, και η λειτουργία του ARD μηχανισμού για τον προσδιορισμό του αριθμού των factors. Τέλος, στην ενότητα 2.4.5, θα εξαχθεί η έκφραση της κατανομής προβλέψεων για το VB-MSFA model.

2.4.1 Variational Lower Bound

Εκ της (35) και υπό την υπόθεση (40) λαμβάνουμε

$$\begin{aligned}
\mathcal{L}(q) = & \int d\boldsymbol{\pi} q(\boldsymbol{\pi}) \log \frac{p(\boldsymbol{\pi}|\mathbf{a})}{q(\boldsymbol{\pi})} + \sum_{i=1}^g \left\{ \int d\boldsymbol{\phi}_i q(\boldsymbol{\phi}_i) \left[\log \frac{p(\boldsymbol{\phi}_i|\boldsymbol{\gamma}_0, \boldsymbol{\omega}_0)}{q(\boldsymbol{\phi}_i)} \right. \right. \\
& + \left. \int d\boldsymbol{\Lambda}_i q(\boldsymbol{\Lambda}_i) \log \frac{p(\boldsymbol{\Lambda}_i|\boldsymbol{\phi}_i)}{q(\boldsymbol{\Lambda}_i)} \right] + \int d\boldsymbol{\mu}_i q(\boldsymbol{\mu}_i) \log \frac{p(\boldsymbol{\mu}_i|\mathbf{m}_0, \mathbf{s}_0)}{q(\boldsymbol{\mu}_i)} \left. \right\} \\
& + \sum_{i=1}^g \sum_{j=1}^n q(z_{ij} = 1) \left\{ \int d\boldsymbol{\pi} q(\boldsymbol{\pi}) \log \frac{p(z_{ij} = 1|\boldsymbol{\pi})}{q(z_{ij} = 1)} \right. \\
& + \int du_{ij} q(u_{ij}|z_{ij} = 1) \left[\log \frac{p(u_{ij}|z_{ij} = 1)}{q(u_{ij}|z_{ij} = 1)} \right. \\
& + \int d\mathbf{y}_{ij} q(\mathbf{y}_{ij}|u_{ij}, z_{ij} = 1) \left(\log \frac{p(\mathbf{y}_{ij}|u_{ij})}{q(\mathbf{y}_{ij}|u_{ij}, z_{ij} = 1)} \right. \\
& \left. \left. \left. + \int d\boldsymbol{\Lambda}_i q(\boldsymbol{\Lambda}_i) \int d\boldsymbol{\mu}_i q(\boldsymbol{\mu}_i) \log p(\mathbf{x}_j|\boldsymbol{\Lambda}_i, \boldsymbol{\mu}_i, \boldsymbol{\Psi}_i, \nu_i, \mathbf{y}_{ij}, u_{ij}, z_{ij} = 1) \right) \right] \right\} \\
& \left. \right\} \tag{2.41}
\end{aligned}$$

Η αναλυτική έκφραση για τα ολοκληρώματα (posterior προσδοκητές τιμές) που περιέχονται στην ως άνω έκφραση του variational lower bound $\mathcal{L}(q)$ μπορεί να βρεθεί στο Παράρτημα.

Εκ της (35) γίνεται φανερό ότι το variational lower bound της log evidence, $\mathcal{L}(q)$, είναι μια μη-convex συνάρτηση της variational posterior κατανομής $q(\boldsymbol{\theta})$

[17]. Συνεπώς, θα υπάρχουν, εν γένει, πολλαπλά τοπικά μέγιστα της $\mathcal{L}(q)$, και συνεπώς, η λύση που βρίσκεται από την διαδικασία του variational inference θα εξαρτάται από την αρχικοποίηση. Παρ' όλα αυτά, το γεγονός αυτό δεν είναι πρόβλημα, καθώς, υπό το variational Bayesian υπόδειγμα, αρκεί η εκτέλεση της variational inference procedure από διαφορετικές τυχαίες εκκινήσεις και η λήψη εκείνης των προκυπτουσών λύσεων που αποδίδει την μεγαλύτερη τιμή για το variational bound, $\mathcal{L}(q)$. Υπογραμμίζουμε ότι το μεγάλο πλεονέκτημα της προτεινομένης variational Bayesian προσεγγίσεως σε σχέση με τον EM αλγόριθμο είναι ότι αυτή η διαδικασία δεν απαιτεί την χρήση του υπολογιστικά κοστοβόρου cross-validation, καθώς η άνω φραγμένη φύση της εδώ βελτιστοποιούμενης συνάρτησης (variational bound) δεν απαιτεί την χρήση τέτοιων μεθόδων, αλλά αρκεί η χρήση όλου του συνόλου εκπαίδευσης σε ένα μόνο πέραςμα [13].

2.4.2 Variational Posteriors

Οι εκφράσεις των variational posteriors για το VB-MSFA model συνάγονται, ως είπαμε, δια της μεγιστοποίησης του $\mathcal{L}(q)$ ως προς καθένα από τους παράγοντες της $q(\theta)$, διατηρώντας κάθε φορά τους υπολοίπους σταθερούς, υπό μια επαναληπτική διαδικασία [18]. Στο τέλος κάθε επανάληψης, η τιμή του variational lower bound, $\mathcal{L}(q)$, υπολογίζεται και χρησιμοποιείται για την εφαρμογή ενός κριτηρίου σύγκλισης του variational inference. Σημειώνεται ότι, ως συνέπεια της conjugate exponential δομής του μοντέλου μας, οι συναγόμενοι βέλτιστοι παράγοντες της variational posterior κατανομής, $q(\theta)$, θα έχουν την ίδια μορφή με τις αντίστοιχες prior κατανομές τις αποτελούσες την $p(X, \theta)$ [17]. Επιπλέον, εκ κατασκευής, το variational lower bound

δεν μπορεί να μειωθεί μετά από μια ενημέρωση του $q(\boldsymbol{\theta})$. Αντιθέτως, η διαδοχική, επαναληπτική διαδικασία ενημέρωσης των επί μέρους variational posterior κατανομών που συνιστούν την $q(\boldsymbol{\theta})$ εγγυάται την μονοτονική αύξηση του lower bound $\mathcal{L}(q)$ [14].

Έστω $\langle \chi \rangle_\xi$ η προσδοκητή τιμή της έκφρασης χ ως προς την κατανομή ξ . Οι αναλυτικές εκφράσεις των παρακάτω χρησιμοποιούμενων προσδοκητών τιμών, μπορούν να βρεθούν στο Παράρτημα. Από την έκφραση του κάτω φράγματος της log evidence του MSFA model, που δίνεται από την (41), έχουμε

$$q(\mathbf{y}_{ij}|u_{ij}, z_{ij} = 1) = \mathcal{N}(\mathbf{y}_{ij}|\bar{\mathbf{y}}_{ij}, \boldsymbol{\Sigma}_i^y/u_{ij}) \quad (2.42)$$

όπου

$$\boldsymbol{\Sigma}_i^y = \left(\mathbf{I}_Q + \langle \boldsymbol{\Lambda}_i^T \boldsymbol{\Psi}_i^{-1} \boldsymbol{\Lambda}_i \rangle_{q(\boldsymbol{\Lambda}_i)} \right)^{-1} \quad (2.43)$$

$$\bar{\mathbf{y}}_{ij} = \boldsymbol{\Sigma}_i^y \langle \boldsymbol{\Lambda}_i \rangle_{q(\boldsymbol{\Lambda}_i)}^T \boldsymbol{\Psi}_i^{-1} \left(\mathbf{x}_j - \langle \boldsymbol{\mu}_i \rangle_{q(\boldsymbol{\mu}_i)} \right) \quad (2.44)$$

Με χρήση και της (42), συνάγεται

$$q(u_{ij}|z_{ij} = 1) = \mathcal{G}(u_{ij}|\alpha_{ij}, \beta_{ij}) \quad (2.45)$$

όπου

$$\alpha_{ij} = \frac{\nu_i + P}{2} \quad (2.46)$$

$$\begin{aligned} \beta_{ij} = & \frac{1}{2} \left\{ \nu_i + \langle \mathbf{y}_{ij} \rangle_{q(\mathbf{y}_{ij}|u_{ij}, z_{ij}=1)}^T \langle \mathbf{y}_{ij} \rangle_{q(\mathbf{y}_{ij}|u_{ij}, z_{ij}=1)} - 2 \left(\mathbf{x}_j - \langle \boldsymbol{\mu}_i \rangle_{q(\boldsymbol{\mu}_i)} \right)^T \right. \\ & \times \boldsymbol{\Psi}_i^{-1} \langle \boldsymbol{\Lambda}_i \rangle_{q(\boldsymbol{\Lambda}_i)} \langle \mathbf{y}_{ij} \rangle_{q(\mathbf{y}_{ij}|u_{ij}, z_{ij}=1)} + \left\langle (\mathbf{x}_j - \boldsymbol{\mu}_i)^T \boldsymbol{\Psi}_i^{-1} (\mathbf{x}_j - \boldsymbol{\mu}_i) \right\rangle_{q(\boldsymbol{\mu}_i)} \\ & \left. + \text{tr} \left[\boldsymbol{\Psi}_i^{-1} \text{tr} \left(\langle \mathbf{y}_{ij} \rangle_{q(\mathbf{y}_{ij}|u_{ij}, z_{ij}=1)} \langle \mathbf{y}_{ij} \rangle_{q(\mathbf{y}_{ij}|u_{ij}, z_{ij}=1)}^T \langle \boldsymbol{\Lambda}_i^T \boldsymbol{\Lambda}_i \rangle_{q(\boldsymbol{\Lambda}_i)} \right) \right] \right\} \quad (2.47) \end{aligned}$$

και tr είναι το ίχνος ενός πίνακα. Χρησιμοποιώντας και την (45), λαμβάνουμε για τα διανύσματα ενδεικτών $Z = \{z_j\}_{j=1}^n$ ότι

$$\begin{aligned}
q(z_{ij} = 1) = \exp \left\{ - \left\langle \frac{u_{ij}}{2} \text{tr} \left[\Psi_i^{-1} (\mathbf{x}_j - \boldsymbol{\mu}_i - \mathbf{\Lambda}_i \mathbf{y}_{ij}) (\mathbf{x}_j - \boldsymbol{\mu}_i - \mathbf{\Lambda}_i \mathbf{y}_{ij})^T \right] \right\rangle_{q(\boldsymbol{\theta})} \right. \\
- \frac{1}{2} \log |\Psi_i| + \log \Gamma(\alpha_{ij}) - \log \Gamma \left(\frac{\nu_i}{2} \right) + \frac{\nu_i}{2} \log \left(\frac{\nu_i}{2} \right) - \alpha_{ij} \log(\beta_{ij}) \\
- \left(\alpha_{ij} - \frac{\nu_i}{2} \right) \langle \log u_{ij} \rangle_{q(u_{ij}|z_{ij}=1)} + \left(\beta_{ij} - \frac{\nu_i}{2} \right) \langle u_{ij} \rangle_{q(u_{ij}|z_{ij}=1)} \\
\left. + \frac{1}{2} \log |\Sigma_i^y| + \frac{P-Q}{2} \langle \log u_{ij} \rangle_{q(u_{ij}|z_{ij}=1)} + \langle \log \boldsymbol{\pi} \rangle_{q(\boldsymbol{\pi})} \right\} \exp(\text{const.})
\end{aligned} \tag{2.48}$$

Ωστόσο, πρέπει να ληφθεί επιπλέον υπόψιν ότι η κατανομή $q(z_j)$ πρέπει να είναι κανονικοποιημένη για κάθε \mathbf{x}_j ώστε να ισχύει $\sum_{i=1}^g q(z_{ij} = 1) = 1$. Υπό αυτό το πρίσμα η έκφραση της variational posterior $q(z_{ij} = 1)$ τελικά γίνεται

$$q(z_{ij} = 1) = \frac{r_{ij}}{\sum_{i=1}^g r_{ij}} \tag{2.49}$$

όπου

$$\begin{aligned}
r_{ij} \triangleq \exp \left\{ - \left\langle \frac{u_{ij}}{2} \text{tr} \left[\Psi_i^{-1} (\mathbf{x}_j - \boldsymbol{\mu}_i - \mathbf{\Lambda}_i \mathbf{y}_{ij}) (\mathbf{x}_j - \boldsymbol{\mu}_i - \mathbf{\Lambda}_i \mathbf{y}_{ij})^T \right] \right\rangle_{q(\boldsymbol{\theta})} \right. \\
- \frac{1}{2} \log |\Psi_i| + \log \Gamma(\alpha_{ij}) - \log \Gamma \left(\frac{\nu_i}{2} \right) + \frac{\nu_i}{2} \log \left(\frac{\nu_i}{2} \right) - \alpha_{ij} \log(\beta_{ij}) \\
- \left(\alpha_{ij} - \frac{\nu_i}{2} \right) \langle \log u_{ij} \rangle_{q(u_{ij}|z_{ij}=1)} + \left(\beta_{ij} - \frac{\nu_i}{2} \right) \langle u_{ij} \rangle_{q(u_{ij}|z_{ij}=1)} \\
\left. + \frac{1}{2} \log |\Sigma_i^y| + \frac{P-Q}{2} \langle \log u_{ij} \rangle_{q(u_{ij}|z_{ij}=1)} + \langle \log \boldsymbol{\pi} \rangle_{q(\boldsymbol{\pi})} \right\}
\end{aligned} \tag{2.50}$$

Αναφορικά, τώρα, με τα βάρη υπέρθεσης π , από την (41) λαμβάνουμε

$$q(\boldsymbol{\pi}) = \mathcal{D}(\boldsymbol{\pi}|\hat{\mathbf{a}}) \quad (2.51)$$

όπου

$$\hat{a}_i = a_i + \sum_{j=1}^n q(z_{ij} = 1) \quad (2.52)$$

Για τις variational posteriors επί των precision parameters των factor loading πινάκων, $q(\phi_{ik})$, συνάγεται

$$q(\phi_{ik}) = \mathcal{G}(\phi_{ik}|\gamma_{ik}, \omega_{ik}) \quad (2.53)$$

όπου

$$\gamma_{ik} = \gamma_0 + \frac{P}{2} \quad (2.54)$$

$$\omega_{ik} = \omega_0 + \frac{\langle \boldsymbol{\lambda}_{ik}^T \boldsymbol{\lambda}_{ik} \rangle_{q(\boldsymbol{\lambda}_{ik})}}{2} \quad (2.55)$$

Επισημαίνουμε ότι ο ως άνω υπολογισμός των variational posteriors $q(\phi_{ik})$ αποτελεί το βασικό συστατικό του ARD μηχανισμού που επιβάλλαμε επί των στηλών των πινάκων factor loadings.

Σχετικά με την variational posterior των πινάκων factor loadings, $q(\boldsymbol{\Lambda}_i)$, ορίζοντας ως $(\boldsymbol{M})_{m,n}$ το (m, n) στοιχείο του πίνακα \boldsymbol{M} , κι ως $(\boldsymbol{v})_l$ το l οστο στοιχείο του διανύσματος \boldsymbol{v} , η εξ. (41) δίδει

$$q(\boldsymbol{\Lambda}_i) = \prod_{l=1}^P q(\boldsymbol{\lambda}_{il}) \quad (2.56)$$

και

$$q(\lambda_{il}) = \mathcal{N}(\lambda_{il} | \mathbf{m}_{il}^*, \mathbf{S}_{il}^*) \quad (2.57)$$

όπου

$$\mathbf{S}_{il}^* = \left[(\Psi_i^{-1})_{ll} \sum_{j=1}^n q(z_{ij} = 1) \langle u_{ij} \rangle_{q(u_{ij})} \langle \mathbf{y}_{ij} \mathbf{y}_{ij}^T \rangle_{q(\mathbf{y}_{ij} | u_{ij}, z_{ij}=1)} + \langle \text{diag}(\phi_i) \rangle_{q(\phi_i)} \right]^{-1} \quad (2.58)$$

$$\mathbf{m}_{il}^* = \mathbf{S}_{il}^* (\Psi_i^{-1})_{ll} \sum_{j=1}^n q(z_{ij} = 1) \langle u_{ij} \rangle_{q(u_{ij})} \left[(\mathbf{x}_j)_l - (\langle \boldsymbol{\mu}_i \rangle_{q(\boldsymbol{\mu})})_l \right] \langle \mathbf{y}_{ij} \rangle_{q(\mathbf{y}_{ij} | u_{ij}, z_{ij}=1)} \quad (2.59)$$

και λ_{il} είναι η *l*οστη γραμμή του *i*οστου πίνακα factor loadings, Λ_i . Τέλος, για την έκφραση των $q(\boldsymbol{\mu}_i)$ έχουμε

$$q(\boldsymbol{\mu}_i) = \mathcal{N}(\boldsymbol{\mu}_i | \mathbf{m}_i, \mathbf{S}_i) \quad (2.60)$$

όπου,

$$\mathbf{S}_i = \left[\text{diag}(\mathbf{s}_0) + \Psi_i^{-1} \sum_{j=1}^n q(z_{ij} = 1) \langle u_{ij} \rangle_{q(u_{ij} | z_{ij}=1)} \right]^{-1} \quad (2.61)$$

$$\begin{aligned} \mathbf{m}_i = \mathbf{S}_i & \left[\text{diag}(\mathbf{s}_0) \mathbf{m}_0 + \Psi_i^{-1} \sum_{j=1}^n q(z_{ij} = 1) \langle u_{ij} \rangle_{q(u_{ij} | z_{ij}=1)} \right. \\ & \left. \times \left(\mathbf{x}_j - \langle \Lambda_i \rangle_{q(\Lambda_i)} \langle \mathbf{y}_{ij} \rangle_{q(\mathbf{y}_{ij} | u_{ij}, z_{ij}=1)} \right) \right] \end{aligned} \quad (2.62)$$

2.4.3 Επιλογή Υπερπαραμέτρων

Έχοντας πλέον βρει τις εκφράσεις των variational posterior κατανομών για τις μεταβλητές (παραμέτρους όπου επιβάλλαμε prior κατανομές και κρυφές μεταβλητές) του MSFA model, πρέπει τώρα να προχωρήσουμε στην κατάλ-

ληλη επιλογή των τιμών των υπερπαραμέτρων του μοντέλου, που περιλαμβάνουν τις $\{\mathbf{a}, \gamma_0, \omega_0, \mathbf{m}_0, \mathbf{s}_0, \{\Psi_i\}_{i=1}^g, \nu\}$.

Ας ξεκινήσουμε με την εύρεση της τιμής των παραμέτρων Ψ_i του μοντέλου, που, για ευκολία, δεν περιελήφθησαν στην variational inference procedure που περιεγράψαμε προηγουμένως, αλλά την χειριζόμαστε ως υπερπαραμέτρο του μοντέλου. Παίρνοντας παραγώγους της $\mathcal{L}(q)$ with ως προς Ψ_i^{-1} συνάγεται

$$\Psi_i^{-1} = \text{diag} \left\{ \frac{1}{\sum_{j=1}^n q(z_{ij} = 1)} \sum_{j=1}^n q(z_{ij} = 1) \times \left\langle u_{ij} (\mathbf{x}_j - \boldsymbol{\mu}_i - \mathbf{\Lambda}_i \mathbf{y}_{ij}) (\mathbf{x}_j - \boldsymbol{\mu}_i - \mathbf{\Lambda}_i \mathbf{y}_{ij})^T \right\rangle_{q(\boldsymbol{\theta})} \right\} \quad (2.63)$$

όπου diag είναι ο τελεστής που διατηρεί μόνο τα διαγώνια στοιχεία ενός τετραγωνικού πίνακα, θέτοντας τα λοιπά σε μηδέν. Καθ' όμοιον τρόπο για τους βαθμούς ελευθερίας ν_i λαμβάνουμε

$$\log \frac{\nu_i}{2} + 1 - \psi \left(\frac{\nu_i}{2} \right) + \frac{1}{\sum_{j=1}^n q(z_{ij} = 1)} \sum_{j=1}^n q(z_{ij} = 1) \times \left(\langle \log u_{ij} \rangle_{q(u_{ij}|z_{ij}=1)} - \langle u_{ij} \rangle_{q(u_{ij}|z_{ij}=1)} \right) = 0 \quad (2.64)$$

Για τις λοιπές των υπερπαραμέτρων του VB-MSFA model, αντί της βελτιστοποίησής τους ως προς το variational lower bound του μοντέλου, επιλέγουμε κάποιες *ad hoc* τιμές. Αυτό είναι προτιμότερο καθώς το όφελος από την εύρεση των τιμών τους δια βελτιστοποίησης του variational lower bound δεν κρίνεται σημαντικό, όταν μια καλή *ad hoc* επιλογή τιμής είναι δυνατή. Αντιθέτως, ο υπολογιστικός φόρτος που επάγει η βελτιστοποίηση του variational

lower bound ως προς τις υπερπαραμέτρους αυτές είναι πολύ σημαντικός, καθώς απαιτεί την επίλυση εξισώσεων ανοιχτού τύπου (δες, π.χ., [13,14]).

Για τους λόγους αυτούς, αναφορικά με τις υπερπαραμέτρους των μ_i , επιλέγουμε $m_0 = 0$, $s_0 = 10^{-3} \mathbf{I}_P$, ώστε να πάρουμε ευρείες κατανομές. Ομοίως για factor loading matrix precision parameters, ϕ_{ik} , επιλέγουμε ευρείες κατανομές, θέτοντας $\gamma_0 = \omega_0 = 10^{-3}$. Τέλος, για τις υπερπαραμέτρους a των βαρών υπέρθεσης, υιοθετούμε την συνήθη επιλογή παίρνοντας $a_i = 10^{-3} \forall i$ [13].

2.4.4 Επιλογή Μοντέλου και Χρήση της ARD Μεθόδου

Ας αρχίσουμε με το πρόβλημα επιλογής αριθμού συνιστωσών factor analyzers g . Όπως είπαμε, αυτό μπορεί να γίνει δια της μεγιστοποίησης του variational lower bound του VB-MSFA model. Ειδικότερα, η υιοθέτηση της προτεινομένης Bayesian προσέγγισης επιτρέπει την εύρεση του βελτίστου αριθμού των συνιστωσών κατανομών, g , με την απλή επανάληψη της variational inference procedure για διαφορετικές τιμές της παραμέτρου g και την λήψη εκείνης των τιμών που μεγιστοποιεί το variational bound, $\mathcal{L}(g)$. Σημειώνεται ότι, εν αντιθέσει με τα της προτεινομένης μεθόδου, οι μέθοδοι βασισμένες στο EM αλγόριθμο, εξ αιτίας της μη φραγμένης μορφής της αντικειμενικής τους συνάρτησης, απαιτούν χρήση cross-validation απέναντι σε ανεξάρτητα σύνολα δεδομένων ώστε να γίνει δυνατός ο κατάλληλος προσδιορισμός του αριθμού των συνιστωσών κατανομών, μια μέθοδος που είναι γνωστό ότι έχει μεγάλη υπολογιστική πολυπλοκότητα και είναι επιρρεπής σε προβλήματα over-fitting [4].

Σχετικά με τον αριθμό των factors, ως αναλύθηκε προηγουμένως, για

τον προσδιορισμό του χρησιμοποιούμε automatic relevance determination (ARD) [16]. Η ιδέα της μεθόδου ARD είναι να δημιουργούμε συνέχεια επιπλέον στοιχεία ελέγχοντας παράλληλα πότε ένα στοιχείο αρχίζει να υπερεκπαιδεύεται. Η υπερεκπαίδευση εκδηλώνεται ως μια *posterior κατανομή υπερ-παραμέτρου ακριβείας (precision)* που τείνει στο άπειρο, δεικνύουσα ότι το στοιχείο αυτό αντιστοιχεί σε μια μόνο παρατήρηση εκπαίδευσης. Βάσει αυτών, την περίπτωση του VB-MSFA model, ο ARD μηχανισμός υλοποιήθηκε επιβάλλοντας μια prior επί των πινάκων factor loadings, ώστε να αποτρέψουμε μεγάλα factor loadings, με το εύρος κάθε prior να ελέγχεται από μια κατά *Gamma κατανεμημένη υπερπαραμέτρο ακριβείας (precision)*, ϕ_{ik} , ως φαίνεται στις εξ. (30), (31). Αν κάποια από αυτές τις ακρίβειες τείνει στο άπειρο, π.χ. $\phi_{ik} \rightarrow \infty$, τότε τα εξερχόμενα βάρη (στήλες λ_{ik}) του κοστού factor του ιοστού analyzer θα πρέπει να είναι πολύ κοντά στο μηδέν, ώστε να διατηρηθεί υψηλή η likelihood, και συνεπώς, αυτός ο factor στην πραγματικότητα αυτο-ακυρώνεται.

Συμπερασματικά, ο προσδιορισμός του αριθμού των factors για το VB-MSFA model, υπό την εφαρμογή της ARD τεχνικής, συνίσταται στην αρχική εκτέλεση της εκτιμητικής του μοντέλου θέτοντας το Q στην μεγίστη δυνατή τιμή του, $Q = P - 1$, και την περαιτέρω απαλοιφή εκείνων των factors με μεγάλο μέσο (τείνοντα στο άπειρο) για την variational posterior την υπερπαραμέτρου ακριβείας τους. Σημειωτέον ότι, εφόσον οι περιττοί factors στην πραγματικότητα δεν μοντελοποιούν δεδομένα, η απαλοιφή τους θα πρέπει να αποδώσει μια αύξηση στο log evidence του μοντέλου.

2.4.5 Εκτίμηση της Κατανομής Πρόβλεψης

Έστω ένα εκπαιδευμένο VB-MSFA model. Προκειμένου να κάνουμε εκτίμηση της πιθανότητας ή ταξινόμηση ενός δείγματος $X' = \{\mathbf{x}'_j\}_{j=1}^{n'}$ ως προς το μοντέλο αυτό, χρειάζεται να υπολογίσουμε την κατανομή πρόβλεψης

$$p(X'|X) = \frac{p(X', X)}{p(X)} = \int d\boldsymbol{\theta} p(\boldsymbol{\theta}|X) p(X'|\boldsymbol{\theta}) \quad (2.65)$$

όπου X τα δεδομένα εκπαίδευσης. Χρησιμοποιώντας την variational posterior αντί της πραγματικής, συνάγεται

$$p(X'|X) \approx \int d\boldsymbol{\theta} q(\boldsymbol{\theta}) p(X'|\boldsymbol{\theta}) \quad (2.66)$$

Τότε, από την (40) και θέτοντας $\boldsymbol{\Lambda} = \{\boldsymbol{\Lambda}_i\}_{i=1}^g$ και $\boldsymbol{\mu} = \{\boldsymbol{\mu}_i\}_{i=1}^g$, λαμβάνουμε

$$\begin{aligned} \log p(X'|X) &\approx \log \int d\boldsymbol{\pi} q(\boldsymbol{\pi}) \int d\boldsymbol{\Lambda} q(\boldsymbol{\Lambda}) \int d\boldsymbol{\mu} q(\boldsymbol{\mu}) \\ &\times \left[\prod_{j=1}^{n'} \sum_{i=1}^g p(z_{ij} = 1 | \boldsymbol{\pi}) p(\mathbf{x}'_j | \boldsymbol{\Lambda}_i, \boldsymbol{\mu}_i, \boldsymbol{\Psi}_i, \nu_i, z_{ij} = 1) \right] \\ &\geq \sum_{j=1}^{n'} \int d\boldsymbol{\pi} q(\boldsymbol{\pi}) \int d\boldsymbol{\Lambda} q(\boldsymbol{\Lambda}) \int d\boldsymbol{\mu} q(\boldsymbol{\mu}) \\ &\times \left[\log \sum_{i=1}^g p(z_{ij} = 1 | \boldsymbol{\pi}) p(\mathbf{x}'_j | \boldsymbol{\Lambda}_i, \boldsymbol{\mu}_i, \boldsymbol{\Psi}_i, \nu_i, z_{ij} = 1) \right] \\ &= \sum_{j=1}^{n'} \int d\boldsymbol{\pi} q(\boldsymbol{\pi}) \int d\boldsymbol{\Lambda} q(\boldsymbol{\Lambda}) \int d\boldsymbol{\mu} q(\boldsymbol{\mu}) \\ &\times \left[\log \sum_{i=1}^g q(z_{ij} = 1) \frac{p(z_{ij}=1 | \boldsymbol{\pi}) p(\mathbf{x}'_j | \boldsymbol{\Lambda}_i, \boldsymbol{\mu}_i, \boldsymbol{\Psi}_i, \nu_i, z_{ij}=1)}{q(z_{ij}=1)} \right] \\ &\geq \sum_{j=1}^{n'} \int d\boldsymbol{\pi} q(\boldsymbol{\pi}) \int d\boldsymbol{\Lambda} q(\boldsymbol{\Lambda}) \int d\boldsymbol{\mu} q(\boldsymbol{\mu}) \sum_{i=1}^g q(z_{ij} = 1) \\ &\quad \times \log \frac{p(z_{ij}=1 | \boldsymbol{\pi}) p(\mathbf{x}'_j | \boldsymbol{\Lambda}_i, \boldsymbol{\mu}_i, \boldsymbol{\Psi}_i, \nu_i, z_{ij}=1)}{q(z_{ij}=1)} \end{aligned}$$

$$\begin{aligned}
\Rightarrow \log p(X'|X) &\geq \sum_{j=1}^{n'} \sum_{i=1}^g q(z_{ij} = 1) \left\{ \int d\boldsymbol{\pi} q(\boldsymbol{\pi}) \log \frac{p(z_{ij} = 1 | \boldsymbol{\pi})}{q(z_{ij} = 1)} \right. \\
&+ \int du_{ij} q(u_{ij} | z_{ij} = 1) \left[\log \frac{p(u_{ij} | z_{ij} = 1)}{q(u_{ij} | z_{ij} = 1)} + \int d\mathbf{y}_{ij} q(\mathbf{y}_{ij} | u_{ij}, z_{ij} = 1) \right. \\
&\times \left(\int d\boldsymbol{\Lambda}_i q(\boldsymbol{\Lambda}_i) \int d\boldsymbol{\mu}_i q(\boldsymbol{\mu}_i) \log p(\mathbf{x}'_j | \boldsymbol{\Lambda}_i, \boldsymbol{\mu}_i, \boldsymbol{\Psi}_i, \nu_i, \mathbf{y}_{ij}, u_{ij}, z_{ij} = 1) \right. \\
&\left. \left. \left. + \log \frac{p(\mathbf{y}_{ij} | u_{ij})}{q(\mathbf{y}_{ij} | u_{ij}, z_{ij} = 1)} \right) \right] \right\} \\
&\hspace{15em} (2.67)
\end{aligned}$$

Όπως παρατηρούμε από την ανωτέρω ανάλυση, η κατανομή πρόβλεψης είναι κάτω φραγμένη από μια ποσότητα που ομοιάζει αρκετά στο variational lower bound (αν αφαιρέσουμε τους όρους που αναφέρονται στις κατανομές των παραμέτρων του μοντέλου). Συνεπώς, για την εκτίμηση της κατανομής πρόβλεψης και την χρήση της σε εφαρμογές ταξινόμησης προτύπων, αρκεί ο υπολογισμός του ως άνω κάτω φράγματος. Οι αναλυτικές εκφράσεις των ολοκληρωμάτων (προσδοκητών τιμών) στην έκφρασή του μπορούν να βρεθούν στο Παράρτημα.

2.5 Εφαρμογές

Παρέχουμε παρακάτω την επίδοση του προτεινομένου μοντέλου σε μια συνθετική και δυο πραγματικές εφαρμογές. Η επίδοσή του συγκρίνεται με την επίδοση της μεθόδου VB-MFA [14], μιας προσέγγισης μεγίστης πιθανοφανείας στο MSFA model (ML-MSFA) [1], μιας προσέγγισης μεγίστης πιθανοφανείας στο MFA model (ML-MFA) [19], και τις μεθόδους MPPCA [20] και tPPCA [21]. Υπογραμμίζουμε ότι σε όλα τα ως άνω πειράματα, οι χρόνοι εκπαίδευσης και αξιολόγησης των ως άνω μεθόδων ήσαν της ίδιας τάξης

μεγέθους, αν δεν εφαρμόσουμε cross validation (συχνά απαραίτητη για τις τεχνικές μεγίστης πιθανοφανείας). Πιο λεπτομερής αξιολόγηση κρίθηκε με εφικτή, καθώς είναι σαφές ότι οι διαφορετικές επιλογές υλοποίησης θα είχαν επίδραση στα αποτελέσματα. Ωστόσο, καθώς η δική μας μέθοδος δεν απαιτεί cross validation, μπορούμε να ισχυριστούμε ότι σε αυτό το πεδίο υπερτερεί των αντιστοίχων μεθόδων μεγίστης πιθανοφανείας.

2.5.1 Εύρωση Επιλογή Μεγέθους Μοντέλου

Στο πείραμα αυτό επιδεικνύουμε την διαδικασία επιλογής μεγέθους μοντέλου με ένα συνθετικό πείραμα. Θεωρούμε μια 3-συνιστωσών υπέρθεση διμεταβλήτων Gaussian κατανομών με $\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$, $\boldsymbol{\mu}_1 = (0, 3)^T$, $\boldsymbol{\mu}_2 = (3, 0)^T$, $\boldsymbol{\mu}_3 = (-3, 0)^T$, και

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}, \boldsymbol{\Sigma}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0.1 \end{pmatrix}, \boldsymbol{\Sigma}_3 = \begin{pmatrix} 2 & -0.5 \\ -0.5 & 0.5 \end{pmatrix}$$

Από κάθε μια από τις συνιστώσες αυτές κατανομές τραβάμε 800 δεδομένα, λαμβάνοντας έτσι συνολικά 2400 εξομοιωμένα σημεία. Επιπλέον, για να αξιολογήσουμε την ανθεκτικότητα των θεωρούμενων μοντέλων σε θόρυβο και outliers, στο ως άνω δείγμα προσθέτουμε και 600 θορυβογενή σημεία (outliers) ληφθέντα από μια ομοιόμορφη κατανομή στο διάστημα $[-\delta, \delta]$ σε κάθε διάσταση. Με χρήση των ως άνω δεδομένων, εκπαιδεύουμε ένα μοντέλο από καθένα από τους θεωρούμενους τύπους, επιλέγοντας $Q = 1$. Για την προτεινόμενη μέθοδο VB-MSFA, και την μέθοδο VB-MFA, η επιλογή μεγέθους μοντέλου γίνεται όπως περιεγράφη στην ενότητα 2.4.4. Για τις μεθόδους ML-MSFA, ML-MFA, MPPCA, και *t*PPCA, χρησιμοποιήσαμε μιας

Πίνακας 2.1: Εύρωστη Επιλογή Μεγέθους Μοντέλου: Τιμές για διάφορα εύρη θορύβου

Μοντέλο	Μέγεθος Μοντέλου (g)			
	$\delta = 0$	$\delta = 5$	$\delta = 10$	$\delta = 20$
VB-MSFA	3	3	3	3
VB-MFA	3	3	5	5
ML-MSFA	3	3	3	4
ML-MFA	3	3	5	7
t PPCA	3	3	3	5
MPPCA	3	3	5	7

cross-validation τεχνική [4]. Τα ληφθέντα αποτελέσματα για τιμές της παραμέτρου δ ίσες με $\delta = 0$ (i.e. no outliers), 5, 10, 20 δίδονται στον Πίνακα 1. Παρατηρούμε ότι τα Gaussian μοντέλα δουλεύουν καλά απουσία θορύβου. Καθώς, όμως, το εύρος του θορύβου, δ , αυξάνει, η επίδοση των Gaussian μοντέλων εμφανίζει μείωση. Αντίθετα, τα κατά Student's- t κατανεμημένα μοντέλα αποδεικνύονται εξαιρετικά εύρωστα.

2.5.2 Ταξινόμηση σημάτων ηλεκτροεγκεφαλογραφήματος

Η ταξινόμηση σημάτων ηλεκτροεγκεφαλογραφήματος (HEΓ) έχει ευρέως χρησιμοποιηθεί σε διαπροσωπείες ανθρώπου-μηχανής (Brain-Computer Interfaces) [22]. Ωστόσο, τα HEΓ συνίστανται από την υπέρθεση πολλών ταυτοχρόνως ενεργών σημάτων με σημαντικά ποσοστά θορύβου, ο οποίος δύναται να έχει σημαντική αρνητική επίδραση στην επίδοση μοντέλων εκπαιδευομένων για την ταξινόμηση αυτών των σημάτων [22]. Το γεγονός αυτό σε συνδυασμό με την υψηλή διάσταση του χώρου παρατήρησης καθιστούν

την συγκεκριμένη εφαρμογή κατάλληλη για την αξιολόγηση του προτεινομένου μοντέλου.

Τα χρησιμοποιούμενα ΗΕΓ δεδομένα αποτελούν τμήμα της συλλογής που παρουσιάστηκε στο [23]. Περιλαμβάνει ΗΕΓ σήματα από 4 υποκείμενα εκτελούντα 2 πνευματικές διεργασίες, στις οποίες είχαν (α) να ηρεμήσουν μη σκεπτόμενα τίποτα συγκεκριμένο (μέτρηση βάσης) (β) να επιλύσουν μη τετριμμένα προβλήματα πολλαπλασιασμού με το μυαλό τους. Τα εν λόγω ΗΕΓ σήματα περιλαμβάνουν 6 κανάλια, και κατεγράφησαν για 10 sec. κάθε φορά, με μια συχνότητα δειγματοληψίας 250 Hz.

Στο πείραμά μας, κάθε ΗΕΓ σήμα διαιρείται σε παράθυρα 1/4 sec. επικαλυπτόμενα κατά 1/8 sec και, εν συνεχεία, για κάθε κανάλι υπολογίζουμε ένα βαθμού-6 αυτοπαλινδρομικό (AR) μοντέλο με χρήση της Yule-Walker approach [24]. Έτσι λαμβάνεται ένα 36×1 feature vector από κάθε παράθυρο. Από τα δεδομένα αυτά, το 30%, ήτοι 200 δείγματα ανά κλάση, χρησιμοποιούνται για εκπαίδευση των μοντέλων, και τα υπόλοιπα για δοκιμές. Στον Πίνακα 2 παρέχουμε τους προκύπτοντες ρυθμούς σφάλματος για τα αξιολογούμενα μοντέλα, και για βέλτιστη επιλογή μεγέθους. Τα αποτελέσματα των EM-τύπου μεθόδων είναι μέσοι επί 30 εκτελέσεων του EM αλγορίθμου. Επιπλέον, παραθέτουμε και τα αποτελέσματα από την εργασία των Anderson et al. in [25].

Πίνακας 2.2: Ταξινόμηση ΗΕΓ σημάτων: Ρυθμοί σφάλματος για βέλτιστη επιλογή μοντέλου

Model	g	Q	Βασική Μέτρηση	Πολλ/μός	Μέσος
VB-MSFA	3	22	1.428%	8.575%	5.001%
VB-MFA	4	29	2.860%	11.420%	7.140%
ML-MSFA	3	18	4.57%	25.64%	15.10%
ML-MFA	4	18	9.15%	32.32%	20.74%
t PPCA	3	18	5.14%	25.68%	15.86%
MPPCA	4	18	8.88%	36.54%	22.71%
Anderson et al. [25]	-	-	-	-	18.7%

2.6 Συμπεράσματα

Στο κεφάλαιο αυτό προτείνουμε μια νέα θεώρηση στα μοντέλα πεπερασμένων υπερθέσεων factor analyzers, καταναμημένων κατα Student's- t , εισάγοντας μια καινοτόμο προσέγγισή τους υπό το υπόδειγμα της Bayesian εκτιμητικής με χρήση μιας variational τεχνικής. Το προκύπτον, VB-MSFA model, παρέχει σημαντικά πλεονεκτήματα εν σχέσει με εναλλακτικές μεγίστης πιθανοφαινείας προσεγγίσεις (π.χ. [1]): τα προβλήματα βελτίστου προσδιορισμού του αριθμού των συνιστωσών κατανομών (factor analyzers) και του πλήθους των factors μπορεί να λυθεί με έναν εύκολο, αξιόπιστο, και υπολογιστικά αποδοτικό τρόπο, το πρόβλημα των singularities που συνδέεται με την φύση της μεγίστης πιθανοφαινείας δεν υφίσταται, ενώ τυχόν επιπλέον συνιστώσες κατανομές περιληφθείσες στο εκπαιδευμένο μοντέλο εν τέλει θα έχουν βάρη υπέρθεσης π_i με υπερπαραμέτρους των variational posterior κατανομών τους συμπίπτουσες με τις υπερπαραμέτρους των prior κατανομών τους, $\hat{a}_i = a_i$ (π.χ. $\hat{a}_i = 10^{-3}$), και άρα δεν παίζουν κανένα ρόλο στην υπολογισμό των κατανομών πρόβλεψης [13]. Επισημαίνουμε ακόμη ότι η προ-

ταθείσα variational προσέγγιση επάγει υπολογιστικά κόστη συγκρίσιμα με αυτά των τεχνικών μεγίστης πιθανοφανείας (εφόσον δεν χρησιμοποιείται από τις τελευταίες και cross-validation) καθώς, σε κάθε περίπτωση, τα βασικά υπολογιστικά κόστη οφείλονται στην εκτιμητική και την αντιστροφή εμπειρικών πινάκων συνδιακύμανσης. Στο πειραματικό μέρος της εργασίας αυτής καταδείχθηκαν και με απτό τρόπο τα πλεονεκτήματα της προταθείσης μεθόδου. Όπως δείξαμε, το VB-MSFA model είχε καλύτερη επίδοση στην αναγνώριση προτύπων, για μικρότερα μεγέθη μοντέλου σε σχέση με τους ανταγωνιστές του, παρουσιάζοντας υψηλή ανθεκτικότητα στον θόρυβο.

Παράρτημα

Ας ξεκινήσουμε με την εύρεση των variational posteriors του VB-MSFA model.

Από την (41), έχουμε για τα factor vectors \mathbf{y}_{ij} ότι

$$\begin{aligned} \frac{\partial \mathcal{L}(q)}{\partial q(\mathbf{y}_{ij} | u_{ij}, z_{ij} = 1)} = 0 \Rightarrow \\ q(z_{ij} = 1) \int d u_{ij} q(u_{ij} | z_{ij} = 1) \left\{ \log p(\mathbf{y}_{ij} | u_{ij}) - \log q(\mathbf{y}_{ij} | u_{ij}, z_{ij} = 1) \right. \\ \left. + \int d \Lambda_i q(\Lambda_i) \int d \boldsymbol{\mu}_i q(\boldsymbol{\mu}_i) \log p(\mathbf{x}_j | \Lambda_i, \boldsymbol{\mu}_i, \Psi_i, \nu_i, \mathbf{y}_{ij}, u_{ij}, z_{ij} = 1) \right\} + \text{const.} = 0 \Rightarrow \\ 0 = \log p(\mathbf{y}_{ij} | u_{ij}) - \log q(\mathbf{y}_{ij} | u_{ij}, z_{ij} = 1) + \int d \Lambda_i q(\Lambda_i) \int d \boldsymbol{\mu}_i q(\boldsymbol{\mu}_i) \\ \times \log p(\mathbf{x}_j | \Lambda_i, \boldsymbol{\mu}_i, \Psi_i, \nu_i, \mathbf{y}_{ij}, u_{ij}, z_{ij} = 1) \end{aligned}$$

$$\begin{aligned}
&\Rightarrow \log q(\mathbf{y}_{ij} | u_{ij}, z_{ij} = 1) \propto \\
&\quad - \frac{u_{ij}}{2} \left\{ \mathbf{y}_{ij}^T \mathbf{y}_{ij} + \left\langle (\mathbf{x}_j - \boldsymbol{\mu}_i - \boldsymbol{\Lambda}_i \mathbf{y}_{ij})^T \boldsymbol{\Psi}_i^{-1} (\mathbf{x}_j - \boldsymbol{\mu}_i - \boldsymbol{\Lambda}_i \mathbf{y}_{ij}) \right\rangle_{q(\boldsymbol{\mu}_i), q(\boldsymbol{\Lambda}_i)} \right\} \\
&\quad \propto - \frac{u_{ij}}{2} \left\{ \mathbf{y}_{ij}^T \left(\mathbf{I}_Q + \langle \boldsymbol{\Lambda}_i^T \boldsymbol{\Psi}_i^{-1} \boldsymbol{\Lambda}_i \rangle_{q(\boldsymbol{\Lambda}_i)} \right) \mathbf{y}_{ij} - 2 \mathbf{y}_{ij}^T \langle \boldsymbol{\Lambda}_i^T \boldsymbol{\Psi}_i^{-1} (\mathbf{x}_j - \boldsymbol{\mu}_i) \rangle_{q(\boldsymbol{\mu}_i), q(\boldsymbol{\Lambda}_i)} \right\}
\end{aligned} \tag{2.68}$$

εκ της οποίας απευθείας απορρέει η (42). Σχετικά με την variational posterior των scaling variables, από την (41) λαμβάνουμε

$$\begin{aligned}
&\int d\boldsymbol{\Lambda}_i q(\boldsymbol{\Lambda}_i) \int d\boldsymbol{\mu}_i q(\boldsymbol{\mu}_i) \int d\mathbf{y}_{ij} q(\mathbf{y}_{ij} | u_{ij}, z_{ij} = 1) \log p(\mathbf{x}_j | \boldsymbol{\Lambda}_i, \boldsymbol{\mu}_i, \boldsymbol{\Psi}_i, \nu_i, \mathbf{y}_{ij}, u_{ij}, z_{ij} = 1) \\
&\quad + \int d\mathbf{y}_{ij} q(\mathbf{y}_{ij} | u_{ij}, z_{ij} = 1) [\log p(\mathbf{y}_{ij} | u_{ij}) - \log q(\mathbf{y}_{ij} | u_{ij}, z_{ij} = 1)] \\
&\quad + \log p(u_{ij} | z_{ij} = 1) - \log q(u_{ij} | z_{ij} = 1) + \text{const.} = 0
\end{aligned} \tag{2.69}$$

Χρησιμοποιώντας την ταυτότητα

$$\left\langle \frac{u_{ij}}{2} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_{ij})^T (\boldsymbol{\Sigma}_i^y)^{-1} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_{ij}) \right\rangle_{q(\mathbf{y}_{ij} | u_{ij}, z_{ij} = 1)} = \frac{u_{ij}}{2} \text{tr} [(\boldsymbol{\Sigma}_i^y)^{-1} \boldsymbol{\Sigma}_i^y / u_{ij}] = \frac{Q}{2} \tag{2.70}$$

η (69) δίνει

$$\begin{aligned}
\log q(u_{ij}|z_{ij}=1) \propto & -\frac{u_{ij}}{2} \text{tr} \left[\Psi_i^{-1} \text{tr} \left(\langle \mathbf{y}_{ij} \rangle_{q(\mathbf{y}_{ij}|u_{ij}, z_{ij}=1)} \langle \mathbf{y}_{ij} \rangle_{q(\mathbf{y}_{ij}|u_{ij}, z_{ij}=1)}^T \langle \Lambda_i^T \Lambda_i \rangle_{q(\Lambda_i)} \right) \right] \\
& + u_{ij} \left(\mathbf{x}_j - \langle \boldsymbol{\mu}_i \rangle_{q(\boldsymbol{\mu}_i)} \right)^T \Psi_i^{-1} \langle \Lambda_i \rangle_{q(\Lambda_i)} \langle \mathbf{y}_{ij} \rangle_{q(\mathbf{y}_{ij}|u_{ij}, z_{ij}=1)} \\
& - \frac{u_{ij}}{2} \left\langle (\mathbf{x}_j - \boldsymbol{\mu}_i)^T \Psi_i^{-1} (\mathbf{x}_j - \boldsymbol{\mu}_i) \right\rangle_{q(\boldsymbol{\mu}_i)} \\
& - \frac{u_{ij}}{2} \langle \mathbf{y}_{ij} \rangle_{q(\mathbf{y}_{ij}|u_{ij}, z_{ij}=1)}^T \langle \mathbf{y}_{ij} \rangle_{q(\mathbf{y}_{ij}|u_{ij}, z_{ij}=1)} \\
& + \frac{P}{2} \log u_{ij} + \left(\frac{\nu_i}{2} - 1 \right) \log u_{ij} - \frac{\nu_i}{2} u_{ij}
\end{aligned} \tag{2.71}$$

εκ της οποίας απευθείας συνάγεται η (45). Καθ' όμοιον τρόπο, για τις z_{ij} λαμβάνουμε

$$\begin{aligned}
& \int d\boldsymbol{\pi} q(\boldsymbol{\pi}) \log p(z_{ij}=1|\boldsymbol{\pi}) - \log q(z_{ij}=1) + \int du_{ij} q(u_{ij}|z_{ij}=1) \left[\log p(u_{ij}|z_{ij}=1) \right. \\
& \quad - \log q(u_{ij}|z_{ij}=1) + \int d\Lambda_i q(\Lambda_i) \int d\boldsymbol{\mu}_i q(\boldsymbol{\mu}_i) \int d\mathbf{y}_{ij} q(\mathbf{y}_{ij}|u_{ij}, z_{ij}=1) \\
& \quad \times \log p(\mathbf{x}_j | \Lambda_i, \boldsymbol{\mu}_i, \Psi_i, \nu_i, \mathbf{y}_{ij}, u_{ij}, z_{ij}=1) \\
& \quad \left. - \int d\mathbf{y}_{ij} q(\mathbf{y}_{ij}|u_{ij}, z_{ij}=1) \log q(\mathbf{y}_{ij}|u_{ij}, z_{ij}=1) \right] \\
& + \text{const.} = 0
\end{aligned} \tag{2.72}$$

που, χρησιμοποιώντας την (70), δίνει την (48). Για τις posteriors του mixing proportions vector $\boldsymbol{\pi}$, η (41) δίνει

$$\frac{\partial \mathcal{L}(q)}{\partial q(\boldsymbol{\pi})} = \log p(\boldsymbol{\pi}|\mathbf{a}) + \sum_{i=1}^g \sum_{j=1}^n q(z_{ij}=1) \log p(z_{ij}=1|\boldsymbol{\pi}) - \log q(\boldsymbol{\pi}) + \text{const.} = 0 \Rightarrow$$

$$\log q(\boldsymbol{\pi}) \propto \sum_{i=1}^g \left[(a_i - 1) + \sum_{j=1}^n q(z_{ij} = 1) \right] \log \pi_i \quad (2.73)$$

εκ της οποίας αμέσως συνάγεται η (51). Για τις $q(\phi_{ik})$, από την (41) και βάσει της (30) έχουμε

$$\begin{aligned} \frac{\partial \mathcal{L}(q)}{\partial q(\phi_{ik})} &= \log p(\phi_{ik} | \gamma_0, \omega_0) + \int d\boldsymbol{\lambda}_{ik} q(\boldsymbol{\lambda}_{ik}) \log p(\boldsymbol{\lambda}_{ik} | \phi_{ik}) - \log q(\phi_{ik}) + \text{const.} \\ &= 0 \Rightarrow \\ \log q(\phi_{ik}) &\propto (\gamma_0 - 1) \log \phi_{ik} - \omega_0 \phi_{ik} + \frac{1}{2} \left[P \log \phi_{ik} - \phi_{ik} \langle \boldsymbol{\lambda}_{ik}^T \boldsymbol{\lambda}_{ik} \rangle_{q(\boldsymbol{\lambda}_{ik})} \right] \end{aligned}$$

εκ της οποίας απορρέει ότι κατά Gamma κατανεμημένες αυτές υπερπαράμετροι έχουν την variational posterior που δίδεται από την (53). Αναφορικά, τώρα, με τα factor loading matrices, $q(\boldsymbol{\Lambda}_i)$, η (41) δίδει

$$\begin{aligned} \frac{\partial \mathcal{L}(q)}{\partial q(\boldsymbol{\Lambda}_i)} &= \sum_{j=1}^n q(z_{ij} = 1) \int du_{ij} q(u_{ij} | z_{ij} = 1) \int d\boldsymbol{\mu}_i q(\boldsymbol{\mu}_i) \int d\mathbf{y}_{ij} q(\mathbf{y}_{ij} | u_{ij}, z_{ij} = 1) \\ &\quad \times \log p(\mathbf{x}_j | \boldsymbol{\Lambda}_i, \boldsymbol{\mu}_i, \boldsymbol{\Psi}_i, \nu_i, \mathbf{y}_{ij}, u_{ij}, z_{ij} = 1) + \int d\boldsymbol{\phi}_i q(\boldsymbol{\phi}_i) \log p(\boldsymbol{\Lambda}_i | \boldsymbol{\phi}_i) \\ &\quad - \log q(\boldsymbol{\Lambda}_i) + \text{const.} = 0 \end{aligned} \quad (2.74)$$

Χρησιμοποιώντας την (32) και ορίζοντας ως $(M)_{m,n}$ το (m, n) στοιχείο του πίνακα M , ως $(\mathbf{v})_l$ το l οστο στοιχείο το διανύσματος \mathbf{v} , και ως λ_{il} την l οστη γραμμή του i οστου πίνακα factor loading, η (74) γράφεται

$$\begin{aligned}
\log q(\Lambda_i) \propto & \sum_{l=1}^P \left\{ -\frac{1}{2} \boldsymbol{\lambda}_{il}^T \left((\Psi_i^{-1})_{ll} \sum_{j=1}^n q(z_{ij} = 1) \langle \mathbf{u}_{ij} \rangle_{q(u_{ij})} \langle \mathbf{y}_{ij} \mathbf{y}_{ij}^T \rangle_{q(\mathbf{y}_{ij}|u_{ij}, z_{ij}=1)} \right) \boldsymbol{\lambda}_{il} \right. \\
& - \frac{1}{2} \boldsymbol{\lambda}_{il}^T \langle \text{diag}(\phi_i) \rangle_{q(\phi_i)} \boldsymbol{\lambda}_{il} + \boldsymbol{\lambda}_{il}^T (\Psi_i^{-1})_{ll} \sum_{j=1}^n q(z_{ij} = 1) \langle \mathbf{u}_{ij} \rangle_{q(u_{ij})} \\
& \left. \times \left[(\mathbf{x}_j)_l - (\langle \boldsymbol{\mu}_i \rangle_{q(\boldsymbol{\mu})})_l \right] \langle \mathbf{y}_{ij} \rangle_{q(\mathbf{y}_{ij}|u_{ij}, z_{ij}=1)} \right\}
\end{aligned} \tag{2.75}$$

εκ της οποίας ευθέως συνάγεται η μορφή της variational posterior $q(\Lambda_i)$, όπως δίδεται από τις (56)-(57). Τέλος, σχετικά με την έκφραση των variational posterior επί των μέσων των factor analyzers, εκ της (41) έχουμε

$$\begin{aligned}
\frac{\partial \mathcal{L}(q)}{\partial q(\boldsymbol{\mu}_i)} &= \sum_{j=1}^n q(z_{ij} = 1) \int d u_{ij} q(u_{ij}|z_{ij} = 1) \int d \Lambda_i q(\Lambda_i) \int d \mathbf{y}_{ij} q(\mathbf{y}_{ij}|u_{ij}, z_{ij} = 1) \\
&\quad \times \log p(\mathbf{x}_j | \Lambda_i, \boldsymbol{\mu}_i, \Psi_i, \nu_i, \mathbf{y}_{ij}, u_{ij}, z_{ij} = 1) + \log p(\boldsymbol{\mu}_i | \mathbf{m}_0, \mathbf{s}_0) \\
&\quad - \log q(\boldsymbol{\mu}_i) + \text{const.} = 0 \Rightarrow \\
\log q(\boldsymbol{\mu}_i) \propto & -\frac{1}{2} \boldsymbol{\mu}_i^T \left[\text{diag}(\mathbf{s}_0) + \Psi_i^{-1} \sum_{j=1}^n q(z_{ij} = 1) \langle \mathbf{u}_{ij} \rangle_{q(u_{ij}|z_{ij}=1)} \right] \boldsymbol{\mu}_i \\
& + \boldsymbol{\mu}_i^T \left[\text{diag}(\mathbf{s}_0) \mathbf{m}_0 + \Psi_i^{-1} \sum_{j=1}^n q(z_{ij} = 1) \langle \mathbf{u}_{ij} \rangle_{q(u_{ij}|z_{ij}=1)} \right. \\
& \left. \times \left(\mathbf{x}_j - \langle \Lambda_i \rangle_{q(\Lambda_i)} \langle \mathbf{y}_{ij} \rangle_{q(\mathbf{y}_{ij}|u_{ij}, z_{ij}=1)} \right) \right]
\end{aligned} \tag{2.76}$$

από την οποία απορρέει η (60).

Βάσει αυτών, η αναλυτική έκφραση του variational lower bound παίρνει

την μορφή

$$\begin{aligned}
\mathcal{L}(q) = & \langle \log p(\boldsymbol{\pi} | \mathbf{a}) \rangle_{q(\boldsymbol{\pi})} - \langle \log q(\boldsymbol{\pi}) \rangle_{q(\boldsymbol{\pi})} + \sum_{i=1}^g \left\{ \langle \log p(\boldsymbol{\phi}_i | \gamma_0, \omega_0) \rangle_{q(\boldsymbol{\phi}_i)} - \langle \log q(\boldsymbol{\phi}_i) \rangle_{q(\boldsymbol{\phi}_i)} \right. \\
& + \langle \log p(\boldsymbol{\Lambda}_i | \boldsymbol{\phi}_i) \rangle_{q(\boldsymbol{\Lambda}_i), q(\boldsymbol{\phi}_i)} - \langle \log q(\boldsymbol{\Lambda}_i) \rangle_{q(\boldsymbol{\Lambda}_i)} + \langle \log p(\boldsymbol{\mu}_i | \mathbf{m}_0, \mathbf{s}_0) \rangle_{q(\boldsymbol{\mu}_i)} \\
& - \langle \log q(\boldsymbol{\mu}_i) \rangle_{q(\boldsymbol{\mu}_i)} + \sum_{j=1}^n q(z_{ij} = 1) \left[\langle \log p(z_{ij} = 1 | \boldsymbol{\pi}) \rangle_{q(\boldsymbol{\pi})} - \log q(z_{ij} = 1) \right. \\
& + \langle \log p(\mathbf{x}_j | \boldsymbol{\Lambda}_i, \boldsymbol{\mu}_i, \boldsymbol{\Psi}_i, \nu_i, \mathbf{y}_{ij}, u_{ij}, z_{ij} = 1) \rangle_{q(\boldsymbol{\theta})} + \langle \log p(u_{ij} | z_{ij} = 1) \rangle_{q(u_{ij} | z_{ij} = 1)} \\
& - \langle \log q(u_{ij} | z_{ij} = 1) \rangle_{q(u_{ij} | z_{ij} = 1)} + \langle \log p(\mathbf{y}_{ij} | u_{ij}) \rangle_{q(\mathbf{y}_{ij} | u_{ij}, z_{ij} = 1), q(u_{ij} | z_{ij} = 1)} \\
& \left. \left. - \langle \log q(\mathbf{y}_{ij} | u_{ij}, z_{ij} = 1) \rangle_{q(\mathbf{y}_{ij} | u_{ij}, z_{ij} = 1), q(u_{ij} | z_{ij} = 1)} \right] \right\}
\end{aligned} \tag{2.77}$$

όπου,

$$\langle \log p(\boldsymbol{\pi} | \mathbf{a}) \rangle_{q(\boldsymbol{\pi})} = \log \Gamma\left(\sum_{i=1}^g a_i\right) + \sum_{i=1}^g [(a_i - 1) \langle \log \pi_i \rangle_{q(\boldsymbol{\pi})} - \log \Gamma(a_i)] \tag{2.78}$$

$$\langle \log q(\boldsymbol{\pi}) \rangle_{q(\boldsymbol{\pi})} = \log \Gamma\left(\sum_{i=1}^g \hat{a}_i\right) + \sum_{i=1}^g [(\hat{a}_i - 1) \langle \log \pi_i \rangle_{q(\boldsymbol{\pi})} - \log \Gamma(\hat{a}_i)] \tag{2.79}$$

$$\langle \log \pi_i \rangle_{q(\boldsymbol{\pi})} = \psi(\hat{a}_i) - \psi\left(\sum_{i=1}^g \hat{a}_i\right) \tag{2.80}$$

$$\langle \log p(\boldsymbol{\phi}_i | \gamma_0, \omega_0) \rangle_{q(\boldsymbol{\phi}_i)} = \sum_{k=1}^Q \left[(\gamma_0 - 1) \langle \log \phi_{ik} \rangle_{q(\boldsymbol{\phi}_i)} + \gamma_0 \log \omega_0 - \omega_0 \langle \phi_{ik} \rangle_{q(\boldsymbol{\phi}_i)} - \log \Gamma(\gamma_0) \right] \tag{2.81}$$

$$\langle \log q(\boldsymbol{\phi}_i) \rangle_{q(\boldsymbol{\phi}_i)} = \sum_{k=1}^Q \left[(\gamma_{ik} - 1) \langle \log \phi_{ik} \rangle_{q(\boldsymbol{\phi}_i)} + \gamma_{ik} \log \omega_{ik} - \omega_{ik} \langle \phi_{ik} \rangle_{q(\boldsymbol{\phi}_i)} - \log \Gamma(\gamma_{ik}) \right] \tag{2.82}$$

$$\langle \phi_{ik} \rangle = \frac{\gamma_{ik}}{\omega_{ik}} \quad (2.83)$$

$$\langle \log \phi_{ik} \rangle_{q(\phi_i)} = \psi(\gamma_{ik}) - \log \omega_{ik} \quad (2.84)$$

$$\begin{aligned} \langle \log p(\Lambda_i | \phi_i) \rangle_{q(\Lambda_i), q(\phi_i)} &= \sum_{l=1}^P \left\{ -\frac{1}{2} \langle \lambda_{il}^T \text{diag}(\phi_i) \lambda_{il} \rangle_{q(\Lambda_i), q(\phi_i)} - \frac{Q}{2} \log 2\pi \right. \\ &\quad \left. + \frac{1}{2} \langle \log |\text{diag}(\phi_i)| \rangle_{q(\phi_i)} \right\} \end{aligned} \quad (2.85)$$

$$\begin{aligned} \langle \log q(\Lambda_i) \rangle_{q(\Lambda_i)} &= \sum_{l=1}^P \left\{ -\frac{1}{2} \langle (\lambda_{il} - \mathbf{m}_{il}^*)^T (\mathbf{S}_{il}^*)^{-1} (\lambda_{il} - \mathbf{m}_{il}^*) \rangle_{q(\Lambda_i)} - \frac{Q}{2} \log 2\pi \right. \\ &\quad \left. - \frac{1}{2} \log |\mathbf{S}_{il}^*| \right\} \end{aligned} \quad (2.86)$$

$$\begin{aligned} \langle \log p(\boldsymbol{\mu}_i | \mathbf{m}_0, \mathbf{s}_0) \rangle_{q(\boldsymbol{\mu}_i)} &= -\frac{1}{2} \langle (\boldsymbol{\mu}_i - \mathbf{m}_0)^T \text{diag}(\mathbf{s}_0) (\boldsymbol{\mu}_i - \mathbf{m}_0) \rangle_{q(\boldsymbol{\mu}_i)} - \frac{P}{2} \log 2\pi \\ &\quad + \frac{1}{2} \log |\text{diag}(\mathbf{s}_0)| \end{aligned} \quad (2.87)$$

$$\langle \log q(\boldsymbol{\mu}_i) \rangle_{q(\boldsymbol{\mu}_i)} = -\frac{1}{2} \langle (\boldsymbol{\mu}_i - \mathbf{m}_i)^T \mathbf{S}_i^{-1} (\boldsymbol{\mu}_i - \mathbf{m}_i) \rangle_{q(\boldsymbol{\mu}_i)} - \frac{P}{2} \log 2\pi - \frac{1}{2} \log |\mathbf{S}_i| \quad (2.88)$$

$$\langle \log p(z_{ij} = 1 | \boldsymbol{\pi}) \rangle_{q(\boldsymbol{\pi})} = \langle \log \pi_i \rangle_{q(\boldsymbol{\pi})} \quad (2.89)$$

$$\begin{aligned} \langle \log p(u_{ij} | z_{ij} = 1) \rangle_{q(u_{ij} | z_{ij}=1)} &= \frac{\nu_i}{2} \log \frac{\nu_i}{2} - \log \Gamma\left(\frac{\nu_i}{2}\right) + \left(\frac{\nu_i}{2} - 1\right) \langle \log u_{ij} \rangle_{q(u_{ij} | z_{ij}=1)} \\ &\quad - \frac{\nu_i}{2} \langle u_{ij} \rangle_{q(u_{ij} | z_{ij}=1)} \end{aligned} \quad (2.90)$$

$$\begin{aligned} \langle \log q(u_{ij} | z_{ij} = 1) \rangle_{q(u_{ij} | z_{ij}=1)} &= (\alpha_{ij} - 1) \langle \log u_{ij} \rangle_{q(u_{ij} | z_{ij}=1)} + \alpha_{ij} \log \beta_{ij} \\ &\quad - \beta_{ij} \langle u_{ij} \rangle_{q(u_{ij} | z_{ij}=1)} - \log \Gamma(\alpha_{ij}) \end{aligned} \quad (2.91)$$

$$\langle u_{ij} \rangle = \frac{\alpha_{ij}}{\beta_{ij}} \quad (2.92)$$

$$\langle \log u_{ij} \rangle_{q(u_{ij} | z_{ij}=1)} = \psi(\alpha_{ij}) - \log \beta_{ij} \quad (2.93)$$

$$\begin{aligned} \langle \log p(\mathbf{y}_{ij} | u_{ij}) \rangle_{q(\mathbf{y}_{ij} | u_{ij}, z_{ij}=1), q(u_{ij} | z_{ij}=1)} &= -\frac{\langle u_{ij} \rangle_{q(u_{ij} | z_{ij}=1)}}{2} \langle \mathbf{y}_{ij}^T \mathbf{y}_{ij} \rangle_{q(\mathbf{y}_{ij} | u_{ij}, z_{ij}=1), q(u_{ij} | z_{ij}=1)} \\ &\quad - \frac{Q}{2} \log 2\pi + \frac{Q}{2} \langle \log u_{ij} \rangle_{q(u_{ij} | z_{ij}=1)} \end{aligned} \quad (2.94)$$

$$\begin{aligned} \langle \log q(\mathbf{y}_{ij} | u_{ij}, z_{ij} = 1) \rangle_{q(\mathbf{y}_{ij} | u_{ij}, z_{ij}=1), q(u_{ij} | z_{ij}=1)} &= -\frac{Q}{2} - \frac{Q}{2} \log 2\pi + \frac{Q}{2} \langle \log u_{ij} \rangle_{q(u_{ij} | z_{ij}=1)} \\ &\quad - \frac{1}{2} \log |\Sigma_i^y| \end{aligned} \quad (2.95)$$

$$\begin{aligned} \langle \log p(\mathbf{x}_j | \Lambda_i, \mu_i, \Psi_i, \nu_i, \mathbf{y}_{ij}, u_{ij}, z_{ij} = 1) \rangle_{q(\theta)} &= -\frac{P}{2} \log 2\pi + \frac{P}{2} \langle \log u_{ij} \rangle_{q(u_{ij} | z_{ij}=1)} - \frac{1}{2} \log |\Psi_i| \\ &\quad - \frac{\langle u_{ij} \rangle_{q(u_{ij} | z_{ij}=1)}}{2} \text{tr} \left[\Psi_i^{-1} \left\langle (\mathbf{x}_j - \mu_i - \Lambda_i \mathbf{y}_{ij}) (\mathbf{x}_j - \mu_i - \Lambda_i \mathbf{y}_{ij})^T \right\rangle_{q(\mu_i), q(\Lambda_i), q(\mathbf{y}_{ij} | u_{ij}, z_{ij}=1)} \right] \end{aligned} \quad (2.96)$$

και $\psi(\cdot)$ είναι η digamma function.

Τέλος, σχετικά με την κατανομή πρόβλεψης $\log p(X'|X)$, από την (67) είναι

$$\log p(X'|X) \geq \text{pred}(X') \quad (2.97)$$

όπου

$$\begin{aligned}
\text{pred}(X') = & \sum_{j=1}^{n'} \sum_{i=1}^g q(z_{ij} = 1) \left[\langle \log p(z_{ij} = 1 | \boldsymbol{\pi}) \rangle_{q(\boldsymbol{\pi})} - \log q(z_{ij} = 1) \right. \\
& - \langle \log q(\mathbf{y}_{ij} | u_{ij}, z_{ij} = 1) \rangle_{q(\mathbf{y}_{ij} | u_{ij}, z_{ij} = 1), q(u_{ij} | z_{ij} = 1)} \\
& - \langle \log q(u_{ij} | z_{ij} = 1) \rangle_{q(u_{ij} | z_{ij} = 1)} + \langle \log p(u_{ij} | z_{ij} = 1) \rangle_{q(u_{ij} | z_{ij} = 1)} \\
& + \langle \log p(\mathbf{y}_{ij} | u_{ij}) \rangle_{q(\mathbf{y}_{ij} | u_{ij}, z_{ij} = 1), q(u_{ij} | z_{ij} = 1)} \\
& \left. + \langle \log p(\mathbf{x}'_j | \boldsymbol{\Lambda}_i, \boldsymbol{\mu}_i, \boldsymbol{\Psi}_i, \nu_i, \mathbf{y}_{ij}, u_{ij}, z_{ij} = 1) \rangle_{q(\boldsymbol{\theta})} \right]
\end{aligned} \tag{2.98}$$

Βιβλιογραφία

- [1] G. McLachlan, R. Bean, and L. B.-T. Jones, "Extension of the mixture of factor analyzers model to incorporate the multivariate t -distribution," *Comp. Stat. Data Analysis*, vol. 51, no. 11, pp. 5327--5338, 2007.
- [2] K. Yamazaki and S. Watanabe, "Singularities in mixture models and upper bounds of stochastic complexity," *Neural Networks*, vol. 16, no. 7, pp. 1029--1038, 2003.
- [3] C. Archambeau, J. Lee, and M. Verleysen, "On the convergence problems of the EM algorithm for finite Gaussian mixtures," in *Eleventh European symposium on artificial neural networks*, 2003, pp. 99--106.
- [4] G. McLachlan and D. Peel, *Finite Mixture Models*. New York: Wiley Series in Probability and Statistics, 2000.
- [5] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.
- [6] J. Diebolt and C. Robert, "Estimation of finite mixture distributions through Bayesian sampling," *J. Roy. Statist. Soc. B*, vol. 56, pp. 363--375, 1994.

- [7] S. Richardson and P. Green, "On Bayesian analysis of mixtures with unknown number of components," *J. Roy. Statist. Soc. B*, vol. 59, pp. 731--792, 1997.
- [8] M. Jordan, Z. Ghahramani, T. Jaakkola, and L. Saul, "An introduction to variational methods for graphical models," in *Learning in Graphical Models*, M. Jordan, Ed. Dordrecht: Kluwer, 1998, pp. 105--162.
- [9] C. C. A. Likas, "Unsupervised learning of Gaussian mixtures based on variational component splitting," *IEEE Trans. Neural Networks*, vol. 18, pp. 745--755, 2007.
- [10] S. Roberts and W. Penny, "Variational Bayes for generalized autoregressive models," *IEEE Trans. Signal Processing*, vol. 50, pp. 2245--2257, 2002.
- [11] V. Smidl and A. Quinn, "Mixture-based extension of the AR model and its recursive Bayesian identification," *IEEE Trans. Signal Processing*, vol. 53, pp. 3530--3542, 2005.
- [12] C. Archambeau and M. Verleysen, "Robust Bayesian clustering," *Neural Networks*, vol. 20, pp. 129--138, 2007.
- [13] M. Svensén and C. M. Bishop, "Robust Bayesian mixture modelling," *Neurocomputing*, vol. 64, pp. 235--252, 2005.
- [14] Z. Ghahramani and M. Beal, "Variational inference for Bayesian mixture of factor analysers," *Advances Neural Information Processing Systems*, vol. 12, 1999.

- [15] C. Liu and D. Rubin, "ML estimation of the t distribution using EM and its extensions, ECM and ECME," *Statistica Sinica*, vol. 5, no. 1, pp. 19--39, 1995.
- [16] E. Fokoue, "Stochastic determination of the intrinsic structure in Bayesian factor analysis," Statistical and Applied Mathematical Sciences Institute, Tech. Rep. TR-2004-17, 2004.
- [17] J. Winn and C. Bishop, "Variational message passing," *J. Machine Learning Research*, vol. 6, pp. 661--694, 2005.
- [18] D. Chandler, *Introduction to Modern Statistical Mechanics*. New York: Oxford University Press, 1987.
- [19] Z. Ghahramani and G. Hinton, "The EM algorithm for mixtures of factor analyzers," Department of Computer Science, University of Toronto, Toronto, Canada, M5S 1A4, Tech. Rep. CRGTR- 96-1, 1997.
- [20] M. Tipping and C. Bishop, "Mixtures of probabilistic principal component analyzers," *Neural Computation*, vol. 11, no. 2, pp. 443--482, 1999.
- [21] J. Zhao and Q. Jiang, "Probabilistic PCA for t distributions," *Neurocomputing*, vol. 69, no. 16-18, pp. 2217--2226, Oct. 2006.
- [22] K.-R. Mueller, C. W. Anderson, and G. E. Birch, "Linear and nonlinear methods for brain-computer interfaces," *IEEE Trans. Neural Systems Rehabilitation Engineering*, vol. 11, no. 2, pp. 165--169, 2003.

- [23] Z. Keirn and J. Aunon, "A new mode of communication between man and his surroundings," *IEEE Trans. Biomedical Engineering*, vol. 37, no. 12, pp. 1209--1214, 1990.
- [24] S. M. Kay, *Modern Spectral Estimation: Theory and Application*. Prentice-Hall, 1988.
- [25] C. W. Anderson, E. A. Stolz, and S. Shamsunder, "Multivariate autoregressive models for classification of spontaneous electroencephalogram during mental tasks," *IEEE Trans. Biomedical Engineering*, vol. 45, no. 3, pp. 277--286, 1998.
- [26] T. Giannakopoulos, D. Kosmopoulos, A. Aristidou, and S. Theodoridis, "Violence content classification using audio features," in *Advances in Artificial Intelligence*, 2006, pp. 502--507.

Κεφάλαιο 3

Μια Fuzzy Clustering Τύπου

Προσέγγιση στα Hidden Markov

Random Field Models

3.1 Προοίμιο

Τα μοντέλα Hidden Markov random field (HMRF) έχουν ευρέως χρησιμοποιηθεί σε εφαρμογές κατάτμησης εικόνων, καθώς αποτελούν μια ιδανική μέθοδο σε προβλήματα όπου το ζητούμενο είναι ένα χωρικά-εξαρτώμενο σχήμα clustering, λαμβάνον υπόψιν τις αμοιβαίες αλληλεπιδράσεις γειτονικών θέσεων. Το Fuzzy c -means (FCM) clustering έχει επίσης εφαρμοσθεί με επιτυχία σε εφαρμογές κατάτμησης εικόνων. Στην παρούσα εργασία, συνδυάζουμε τα πλεονεκτήματα των δυο αυτών προσεγγίσεων, προτείνοντας μια καινοτόμο θεώρηση των μοντέλων HMRF, επί τη βάση μιας fuzzy clustering θεάσεως. Κατ' αυτήν, προσεγγίζουμε το πρόβλημα εκπαίδευσης

του μοντέλου HMRF ως ένα FCM-τύπου clustering πρόβλημα, που μπορεί να επιλυθεί δια της εισαγωγής των υποθέσεων του HMRF μοντέλου στην διαδικασία του fuzzy clustering. Η προσέγγισή μας αξιοποιεί την regularized μέσω Kullback-Leibler (KL) divergence information fuzzy objective function, και γίνεται εφικτή χρησιμοποιώντας την προσέγγιση μέσου-πεδίου (mean-field approximation) της Markov random field prior κατανομής. Η εφαρμογή του προτεινομένου μοντέλου σε κατάτμηση εικόνων καταδεικνύει την ανωτερότητά του επί των ανταγωνιστών του.

Η εργασία μου αυτή έχει δημοσιευθεί υπό το άρθρο μου:

Sotirios Chatzis, Theodora Varvarigou, "A Fuzzy Clustering Approach Towards Hidden Markov Random Field Models for Enhanced Spatially-Constrained Image Segmentation," *IEEE Transactions on Fuzzy Systems*, Accepted for Future Publication, 2008.

3.2 Εισαγωγή

Τα μοντέλα Markov random field (MRF) έχουν ευρέως χρησιμοποιηθεί με επιτυχία σε ένα πλήθος εφαρμογών επεξεργασίας εικόνας. Μεταξύ των σημαντικότερων αυτών, αναφέρουμε την κατάτμηση εικόνων, ανασκευή εικόνων, εντοπισμό ακμών, και edge preserving smoothing (για μια εκτενή ανασκόπηση τέτοιων εφαρμογών ανατρέξτε, π.χ., στο [1]). Ένας ιδιαίτερος τύπος MRF μοντέλων, εμπνευσμένος από την βιβλιογραφία των hidden Markov μοντέλων, είναι το hidden Markov random field (HMRF) model [2]. Ένα HMRF model είναι στην πραγματικότητα μια στοχαστική διαδικασία παραχθείσα από ένα MRF μοντέλο, η ακολουθία καταστάσεων της οποίας είναι μη πα-

ρατηρήσιμη και μπορεί μόνο να συμπεραστεί από ένα τυχαίο πεδίο (random field) παρατηρήσεων. Τα HMRF models έχουν αναδειχθεί σε ένα πολύτιμο εργαλείο κατάτμησης εικόνων, καθώς παρέχουν ένα ισχυρό και αυστηρό τρόπο εισαγωγής της πληροφορίας σχετικά με τις αμοιβαίες αλληλεπιδράσεις των γειτονικών pixels στην διαδικασία της κατάτμησης [3,4]. Αυτό γίνεται εφικτό κωδικοποιώντας την χωρική πληροφορία σε μια εικόνα δια της επιβολής χωρικών περιορισμών στην κατανομή των ετικετών (αντιστοιχούντα τμήματα εικόνας) των pixels, που χαρακτηρίζονται από MRF κατανομές.

Ωστόσο, αυτές οι χωρικές εξαρτήσεις, που αποτελούν και το πλεονέκτημα του HMRF model αναφορικά με την επίδοση των εφαρμογών κατάτμησης εικόνων, είναι επίσης υπεύθυνες για έναν βαρύ υπολογιστικό φόρτο επαγόμενο επί της διαδικασίας εκτιμητικής του μοντέλου. Μια απλή και δημοφιλής λύση σε αυτό το πρόβλημα αποτελεί η pseudo-likelihood approximation της κατανομής του Markov random field [5]. Επί τη βάση αυτής της προσέγγισης, ποικίλες μέθοδοι εκτιμητικής του HMRF model έχουν προταθεί. Μια κοινότυπη προσέγγιση περιλαμβάνει την εκτέλεση ενός expectation-maximization (EM)-τύπου αλγορίθμου, όπου, στο E-step, μια στοχαστική ανάκτηση της άγνωστης κατάτμησης της εικόνας ανακτάται, δια εφαρμογής ενός Bayesian κριτηρίου βελτιστοποίησης, ενώ, στο M-step, οι παράμετροι του HMRF μοντέλου εκτιμούνται βάσει των παρατηρήσεων και των ανακτηθέντων δεδομένων [6]. Ένα βασικό χαρακτηριστικό αυτής της προσέγγισης είναι ότι στο M-step του EM-τύπου αλγορίθμου οι ετικέτες των image pixel (ήτοι, τα τμήματα της εικόνας όπου ανήκουν) θεωρούνται γνωστές ποσότητες. Αναφορικά με το βήμα στοχαστικής ανάκτησης των ετικετών,

πολλές εναλλακτικές μέθοδοι θα μπορούσαν να χρησιμοποιηθούν, περιλαμβανομένων των ICM [3,4], και marginal posterior modes (MPM) [7]. Παρότι η εν λόγω προσέγγιση συνήθως δίδει καλά αποτελέσματα, είναι γνωστό ότι παράγει μεροληπτικές εκτιμήτριες [6], λόγω της εκτέλεσης του M-step του αλγορίθμου θεωρώντας τις ετικέτες των pixels ως γνωστές ποσότητες. Για την επίλυση αυτών των προβλημάτων, μια δημοφιλής λύση βασίζεται στην χρήση Gibbsian sampling και Monte-Carlo τεχνικών για τον υπολογισμό της posterior κατανομής του Markov random field [32], και την χρήση των ποσοτήτων αυτών στο M-step του αλγορίθμου. Εναλλακτικά, μια mean-field προσέγγιση του Markov random field έχει προσφάτως προταθεί [8,9], προσφέροντας καλύτερη υπολογιστική πολυπλοκότητα σε σχέση με τις sampling τεχνικές.

Από την άλλη πλευρά, μεθοδολογίες fuzzy clustering, και, ιδίως, ο fuzzy *c*-means (FCM) αλγόριθμος [11], τον οποίο παρουσιάσαμε διεξοδικά στο Κεφ. 1, έχουν επίσης ευρέως χρησιμοποιηθεί σε εφαρμογές κατάτμησης εικόνων [12]. Η επιτυχία τους φαίνεται ότι οφείλεται στο ότι η fuzzy φύση τους, επιτρέπει στην διαδικασία του clustering να ανακτήσει καλύτερα τα χωρικά εξαρτημένα πρότυπα σε σχέση με τις μη fuzzy clustering τεχνικές [13, 14]. Ωστόσο, παρότι ο FCM αλγόριθμος συνήθως δουλεύει καλά με ελεύθερες θορύβου εικόνες, η επίδοσή του μπορεί να γίνει αρκετά φτωχή στην περίπτωση εικόνων υποβαθμισμένης ποιότητας εξαιτίας θορύβου, ατυπικών δεδομένων, ή άλλων παραμορφώσεων, που όμως αποτελούν τον κανόνα σε εικόνες του πραγματικού κόσμου (δηλ., μη συνθετικές) [13]. Αυτή η ανεπιθύμητη συμπεριφορά αποδίδεται κυρίως στην συνήθη χρήση της μη-εύρωστης Ευκλείδειας dissimilarity function, και στην μη αξιοποίηση της

πληροφορίας σχετικά με τις αμοιβαίες αλληλεπιδράσεις μεταξύ των γειτονικών θέσεων (sites), δηλ. των pixels της εικόνας. Πολλές προσπάθειες έχουν καταβληθεί ώστε να επιλυθούν αυτές οι αδυναμίες του FCM αλγορίθμου. Επί παραδείγματι, στις εργασίες [13, 15--22], διάφορες πιο εύρωστες επιλογές για το dissimilarity function του FCM algorithm έχουν προταθεί. Ωστόσο, ένα μείζον μειονέκτημα αυτών των μεθόδων, σε σχέση με το πρόβλημα του υπό χωρική συνθήκη clustering (όπως είναι το πρόβλημα της κατάτμησης εικόνων), είναι ότι οι μέθοδοι αυτές δεν λαμβάνουν υπόψη τις χωρικές εξαρτήσεις μεταξύ των δεδομένων. Στις εργασίες [23--26], ποικίλα FCM-τύπου clustering σχήματα, λαμβάνοντα υπόψη αυτές τις χωρικές εξαρτήσεις στην διατύπωση της fuzzy objective function, έχουν προταθεί. Ωστόσο, στα μειονεκτήματα αυτών των μεθόδων συγκαταλέγονται κατά περίπτωση η αποτυχία να εξάγουν clusters αυθαιρέτων σχημάτων, μη ευρωστία σε ατυπικά δεδομένα, και μεγάλη υπολογιστική πολυπλοκότητα.

Προκειμένου να επιλύσουμε αυτά τα προβλήματα, στην εργασία αυτή προτείνουμε μια καινοτόμο fuzzy θεώρηση στο HMRF model. Εισάγοντας τα αποτελεσματικά μέσα για την επιτυχή εισαγωγή των υποθέσεων του HMRF model στην διαδικασία του fuzzy clustering, παρέχουμε μια αποδοτική μεθοδολογία εκτιμητικής του HMRF μοντέλου υπό το fuzzy clustering υπόδειγμα. Η προτεινόμενη προσέγγιση διαμορφώνεται επί τη βάση του regularized μέσω Kullback-Leibler (KL) divergence information, FCM-τύπου αλγορίθμου (δες Κεφ. 1), και γίνεται εφικτή χρησιμοποιώντας την mean-field προσέγγιση της Markov random field (prior) κατανομής. Ο τοιουτοτρόπως λαμβανόμενος, HMRF-FCM αλγόριθμος, προσφέρει μια FCM-τύπου εκτιμητική του HMRF μοντέλου, συνδυάζοντας τα πλεονεκτήματα για τον αλ-

γόριθμο κατάτμησης εικόνων που απορρέουν από τις εξαιρετικές δυνατότητες μοντελοποίησης χωρικής δομικότητας που παρέχει το HMRF model, και την ευελιξία των fuzzy clustering αλγορίθμων. Η επίδειξη των προτερημάτων της προτεινόμενης προσεγγίσεως σε σχέση με ανταγωνιστικές μεθόδολογίες κατάτμησης εικόνων καταδεικνύεται στην πειραματική ενότητα αυτού του κεφαλαίου. Το υπόλοιπο του τρέχοντος κεφαλαίου είναι δομημένο ως εξής: Στην ενότητα 3.3, παρέχεται μια εισαγωγή στα HMRF models. Στην ενότητα 3.4, διατυπώνεται ο προτεινόμενος HMRF-FCM αλγόριθμος. Στην ενότητα 3.5, διεξάγεται η πειραματική αξιολόγηση του διατυπωμένου αλγορίθμου σε εφαρμογές κατάτμησης εικόνων.

3.3 Hidden Markov Random Field Models

Έστω ένα αλφάβητο $Q = \{1, \dots, q\}$. Έστω S ένα πεπερασμένο σύνολο δεικτών, $S = \{1, \dots, s\}$, το οποίο θα καλούμε ως το σύνολο των θέσεων (sites). Έστω ότι για κάθε site $j \in S$ ορίζεται ένας πεπερασμένος χώρος καταστάσεων, \mathcal{X}_j , τέτοιος ώστε

$$\mathcal{X}_j = \{x_j : x_j \in Q\}$$

Ο χώρος

$$\mathcal{X} = \prod_{j=1}^s \mathcal{X}_j = Q^s$$

θα καλείται ο χώρος των διαμορφώσεων των τιμών των καταστάσεων του θεωρηθέντος συνόλου sites, $\mathbf{x} = (x_j)_{j \in S}$. Μια αυστηρά θετική κατανομή

πιθανότητας, $p(\mathbf{x})$, $\mathbf{x} \in \mathcal{X}$, επί του χώρου \mathcal{X} , ήτοι

$$p(\mathbf{x}) > 0 \quad \forall \mathbf{x} \in \mathcal{X} \quad (3.1)$$

καλείται ως τυχαίο πεδίο, ή random field [27]. Μια διαμόρφωση τιμών \mathbf{x} του random field συχνά αναφέρεται και ως μια υλοποίηση του random field.

Έστω ∂ ένα σύστημα γειτνίασης στο S , δηλ. μια συλλογή $\partial = \{\partial_j : j \in S\}$ συνόλων, τέτοια ώστε $j \notin \partial_j$ και $l \in \partial_j$ αν και μόνο αν $j \in \partial_l \quad \forall l, j \in S$. Τότε, το προηγουμένως θεωρηθέν random field, $p(\mathbf{x})$, θα είναι ένα Markov random field ως προς το σύστημα γειτνίασης ∂ εφόσον [4]

$$p(x_j | \mathbf{x}_{S-\{j\}}) = p(x_j | \mathbf{x}_{\partial_j}) \quad \forall j \in S \quad (3.2)$$

Έχει αποδειχθεί (θεώρημα Hammersley-Clifford) ότι ένα δοσμένο random field είναι ένα Markov random field αν και μόνο αν η (από κοινού) κατανομή πιθανότητάς του, $p(\mathbf{x})$, είναι μια Gibbs κατανομή, ήτοι [28]

$$p(\mathbf{x} | \beta) \triangleq \frac{1}{W(\beta)} \exp(-U(\mathbf{x} | \beta)) \quad (3.3)$$

όπου $W(\beta)$ είναι μια σταθερά κανονικοποίησης (η συνάρτηση διαμέρισης), τέτοια ώστε

$$W(\beta) = \sum_{\mathbf{x} \in \mathcal{X}} \exp(-U(\mathbf{x} | \beta)) \quad (3.4)$$

$U(\mathbf{x} | \beta)$ είναι η συνάρτηση ενεργείας, με έκφραση

$$U(\mathbf{x} | \beta) = \sum_{c \in C} V_c(\mathbf{x} | \beta) \quad (3.5)$$

C είναι μια κλάση από υποσύνολα των sites, τις κλίκες, όπου κάθε κλίκα $c \in C$ περιλαμβάνει sites που είναι όλα γείτονες μεταξύ τους, V_c είναι μια συνάρτηση δυναμικού κλικών συσχετισμένη με την κλίκα c , και β είναι μια παράμετρος των δυναμικών κλικών καλουμένη ως αντίστροφη υπερκριτική θερμοκρασία.

Έστω τώρα ένα δεύτερο random field, $p(\mathbf{y})$, ο χώρος καταστάσεων \mathcal{Y} του οποίου επίσης ορίζεται επί του υποτιθέντος συνόλου sites S και δίνεται από την

$$\mathcal{Y} = \prod_{j=1}^s \mathcal{Y}_j$$

$$\mathcal{Y}_j = \{\mathbf{y}_j : \mathbf{y}_j \in \mathbb{R}^w\}$$

Θα συμβολίζουμε ως \mathbf{y} μια υλοποίηση αυτού του field, και θα ισχύει προφανώς $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_s^T)^T$. Το ζεύγος του ανωτέρω random field, $p(\mathbf{y})$, και του υποτιθέντος Markov random field, $p(\mathbf{x})$, τέτοιο ώστε

$$p(\mathbf{y}, \mathbf{x}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \quad (3.6)$$

καλείται ως hidden Markov random field model [29]. Συνήθως υποτίθεται ότι η κατανομή $p(\mathbf{y}|\mathbf{x})$ παραγοντοποιείται επί των sites, δηλ.

$$p(\mathbf{y}|\mathbf{x}) = \prod_{j=1}^s p(\mathbf{y}_j|x_j) \quad (3.7)$$

υπόθεση που υιοθετούμε και εδώ, καθώς παρέχει μια βολική έκφραση του poste-rior field $p(\mathbf{x}|\mathbf{y})$ χωρίς να επηρεάζει την Μαρκοβιανή του φύση, παρότι πιο πολύπλοκες υποθέσεις θα μπορούσαν επίσης να υιοθετηθούν [29].

Το πεδίο $p(\mathbf{y})$ καλείται το παρατηρήσιμο (ή εκπεμπόμενο) random field των παρατηρήσεων που συσχετίζονται με τα sites $j \in S$, το \mathcal{Y} καλείται ο χώρος των παρατηρήσεων, και το field $p(\mathbf{x})$ καλείται το Markov random field των κρυφών μεταβλητών κατάστασης που συσχετίζονται με τα sites $j \in S$ [5]. Συνήθως, οι κατανομές των παρατηρήσεων υπό τη συνθήκη των καταστάσεων του HMRF μοντέλου από τις οποίες προέρχονται, $p(\mathbf{y}_j|x_j)$, λαμβάνονται ως έχουσες μια Gaussian μορφή [30]. Με άλλα λόγια

$$p(\mathbf{y}|\mathbf{x}; \boldsymbol{\theta}) = \prod_{j=1}^s p(\mathbf{y}_j|x_j; \boldsymbol{\theta}_{x_j}) = \prod_{j=1}^s \mathcal{N}(\mathbf{y}_j|\boldsymbol{\mu}_{x_j}, \boldsymbol{\Sigma}_{x_j}) \quad (3.8)$$

όπου $\boldsymbol{\mu}_{x_j}$ και $\boldsymbol{\Sigma}_{x_j}$ είναι ο μέσος και ο πίνακας συνδιακύμανσης της (Gaussian) κατανομής εκπομπής της x_j οστης κρυφής κατάστασης του HMRF model, αντίστοιχα, το $\boldsymbol{\theta}_{x_j}$ περιλαμβάνει τα $\boldsymbol{\mu}_{x_j}$ και $\boldsymbol{\Sigma}_{x_j}$, και $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_q^T)^T$.

Ο υπολογισμός του όρου $W(\beta)$, που ορίζεται στην (4), περιλαμβάνει όλες τις πιθανές υλοποιήσεις \mathbf{x} του (hidden) Markov random field, και συνεπώς, είναι εν γένει μη εφικτός λόγω υπερβολικού υπολογιστικού κόστους. Το γεγονός αυτό είναι αρκετά περιοριστικό καθώς καθιστά απαγορευτικούς πολλούς υπολογισμούς όπου εμπλέκεται η εύρεση της από κοινού (prior) κατανομής του Markov field, $p(\mathbf{x}|\beta)$. Μια απλή και επιτυχής λύση στο πρόβλημα αυτό παρέχεται από την pseudo-likelihood προσέγγιση της prior κατανομής του Markov field [5]. Υπό την προσέγγιση αυτή, λαμβάνουμε

$$p(\mathbf{x}|\beta) = \prod_{j=1}^s p(x_j|\mathbf{x}_{\partial_j}; \beta) \quad (3.9)$$

όπου οι όροι $p(x_j|\mathbf{x}_{\partial_j}; \beta)$ δίνονται από [5]

$$p(x_j|\mathbf{x}_{\partial_j}; \beta) = \frac{\exp(-\sum_{c \ni j} V_c(\mathbf{x}|\beta))}{\sum_{x_j=1}^q \exp(-\sum_{c \ni j} V_c(\mathbf{x}|\beta))} \quad (3.10)$$

Όπως έχει δειχθεί στην σχετική βιβλιογραφία (δες, π.χ., [31]), η βελτιστοποίηση της Markov field prior υπό την pseudo-likelihood approximation (9) δίδει συνεπείς εκτιμήτριες υπό ήπιες συνθήκες.

Ένα άλλο σημαντικό ζήτημα σχετικά με το HMRF model αφορά στον υπολογισμό (και όχι απλά την βελτιστοποίηση) των posterior κατανομών $p(x_j|\mathbf{y})$ και $p(\mathbf{x}|\mathbf{y})$. Συνήθως, αυτές οι ποσότητες υπολογίζονται μέσω Bayesian sampling, π.χ., χρησιμοποιώντας μεθόδους Markov chain Monte Carlo [32]. Ωστόσο, τέτοιες μέθοδοι απαιτούν, ως γνωστόν, μεγάλα υπολογιστικά φορτία. Μια εναλλακτική των μεθόδων αυτών, είναι η προσέγγιση μέσου-πεδίου (mean-field approximation) [9,33]. Η mean-field approximation αρχικά αναπτύχθηκε ως μια μέθοδος για τον προσεγγιστικό υπολογισμό του μέσου ενός Markov random field. Προέρχεται από τον χώρο της στατιστικής μηχανικής (π.χ. [34]), όπου χρησιμοποιείται ως ένα εργαλείο για την μελέτη φαινομένων μετάπτωσης φάσης. Η εφαρμογή της στην προσέγγιση κατανομών Markov random fields βασίζεται στην ιδέα της παράλειψης των διακυμάνσεων των γειτονικών sites που αλληλεπιδρούν με ένα θεωρούμενο site. Έτσι, το προκύπτον σύστημα συμπεριφέρεται ως ένα αποτελούμενο από ανεξάρτητες μεταβλητές για το οποίο οι υπολογισμοί γίνονται αποδοτικοί.

Έτσι λοιπόν, δοθείσης μας εκτιμητριάς \hat{x} του αγνώστου διανύσματος ετικετών (καταστάσεων) των sites, \mathbf{x} , υπολογισμένου, π.χ., με χρήση των ICM, MPM, ή κάποιου άλλου καταλλήλου κριτηρίου (δες, π.χ., [4, 9, 35]),

κάνουμε την υπόθεση [6,9]

$$p(\mathbf{x}|\beta) = \prod_{j=1}^s p(x_j|\hat{\mathbf{x}}_{\partial_j}; \beta) \quad (3.11)$$

Ως έχει δειχθεί, η mean-field approximation (11) του MRF παρέχει μια ενδιαφέρουσα εναλλακτική των μεθόδων Bayesian sampling, παρέχουσα εξίσου καλές εκτιμήτριες των Markov posteriors για κατά τάξεις μεγέθους καλύτερα υπολογιστικά κόστη [9,33,35].

3.4 Προτεινόμενη Προσέγγιση

3.4.1 Μια FCM-τύπου θεώρηση του HMRF model

Έστω ένα q -καταστάσεων hidden Markov random field model, περιλαμβάνων ένα παρατηρήσιμο random field εκπεμπομένων δεδομένων, $p(\mathbf{y})$, και ένα Markov random field κρυφών μεταβλητών καταστάσεων, $p(\mathbf{x})$, ορισμένο ως περιεγράφη στην ενότητα 3.3. Υποθέτουμε ότι η υπό συνθήκη κατανομή εκπομπής του θεωρηθέντος HMRF model, $p(\mathbf{y}|\mathbf{x})$, έχει Gaussian μορφή, ως δίδεται από την (8). Στην ανάλυσή μας, υιοθετούμε την mean-field approximation (11) του MRF, υπό την οποία λαμβάνονται οι pointwise posteriors [33]

$$p(x_j = i|\mathbf{y}_j) \triangleq p(x_j = i|\mathbf{y}_j, \hat{\mathbf{x}}_{\partial_j}) \quad (3.12)$$

Μια εκάστη των καταστάσεων του υποτιθέντος HMRF model μπορεί να θεωρηθεί ως ορίζουσα ένα cluster επί του χώρου \mathcal{Y} των παρατηρήσεων. Έστω η κατανομή $p(x_j = i|\mathbf{y}_j)$, ορισμένη από την (12), που αποτελεί την

posterior πιθανότητα η παρατήρηση \mathbf{y}_j , η οποία συσχετίζεται με την j οστη κατάσταση, να προέρχεται από το cluster που αντιστοιχεί στην i οστη κατάσταση του μοντέλου. Έχουμε τότε

$$0 \leq p(x_j = i | \mathbf{y}_j) \leq 1, \quad \sum_{i=1}^q p(x_j = i | \mathbf{y}_j) = 1 \quad (3.13)$$

($i = 1, \dots, q, j = 1, \dots, s$). Ακολουθώντας τα αποτελέσματα του Ruspini [38], η (13) υπονοεί ότι τα clusters αυτά μπορούν επίσης να θεωρηθούν ως κατά Zadeh [10] ασαφή σύνολα (fuzzy sets), με μια ισχυρή πιθανοτική ερμηνεία. Βάσει αυτής της παρατήρησης, μπορούμε να θεωρήσουμε το υποτιθέν HMRF model ως ορίζον μια ασαφή (fuzzy) q -διαμέριση του χώρου παρατήρησης \mathcal{Y} . Συμβολίζουμε αυτή την διαμέριση ως

$$R = \{r_{ij}\} \quad (3.14)$$

όπου r_{ij} ($i = 1, \dots, q, j = 1, \dots, s$) είναι ο βαθμός που η παρατήρηση \mathbf{y}_j ανήκει στο cluster που αναπαρίσταται από την i οστη κατάσταση του HMRF model, ήτοι το fuzzy membership function του μοντέλου, με ιδιότητες (δες και Κεφ. 1):

$$0 \leq r_{ij} \leq 1, \quad \sum_{i=1}^q r_{ij} = 1, \quad 0 < \sum_{j=1}^s r_{ij} < s \quad (3.15)$$

Υπό την θεώρηση αυτή, αναζητούμε μια fuzzy clustering-τύπου μέθοδο εκτιμητικής του θεωρηθέντος HMRF model, στα πλαίσια ενός FCM-type algorithm, και με χρήση της mean-field approximation (11) της κατανομής του MRF. Επί τη βάσει όσων συζητήσαμε στο Κεφ. 1 για τον FCM αλγόριθμο, είναι εύκολα αντιληπτό ότι η αναζητούμενη fuzzy clustering-τύπου προσέγ-

γηση του HMRF model μπορεί να θεμελιωθεί επί τη βάση της regularized μέσω KL information fuzzy objective function, εισάγοντας μια HMRF-τύπου παραλλαγή της, υπό το πρίσμα της υιοθετούμενης mean-field approximation, βάσει της οποίας συνάγεται η αντικειμενική συνάρτηση

$$Q_\lambda \triangleq \sum_{i=1}^q \sum_{j=1}^s r_{ij} d_{ij} + \lambda \sum_{i=1}^q \sum_{j=1}^s r_{ij} \log \left(\frac{r_{ij}}{\pi_{ij}} \right) \quad (3.16)$$

όπου, π_{ij} είναι οι σημειακές (pointwise) prior πιθανότητες των καταστάσεων του HMRF model, συναγόμενες από την mean-field approximation (11 του MRF ως

$$\pi_{ij} \triangleq p(x_j = i | \hat{\mathbf{x}}_{\partial_j}; \beta) = \frac{\exp(-\sum_{c \ni j} V_c(\tilde{\mathbf{x}}_{ij} | \beta))}{\sum_{h=1}^q \exp(-\sum_{c \ni j} V_c(\tilde{\mathbf{x}}_{hj} | \beta))} \quad (3.17)$$

και $\tilde{\mathbf{x}}_{ij} \triangleq (x_j = i, \hat{\mathbf{x}}_{\partial_j})$, και ορίζοντας την dissimilarity function, d_{ij} , του αλγορίθμου ως την αρνητική λογαριθμο-πιθανοφάνεια της *i*οστης κατάστασης του μοντέλου ως προς την *j*οστη παρατήρηση. Με άλλα λόγια, έχουμε

$$d_{ij}(\boldsymbol{\theta}_i) \triangleq -\log p(\mathbf{y}_j | x_j = i; \boldsymbol{\theta}_i) \quad (3.18)$$

εκ της οποίας η αντικειμενική συνάρτηση του HMRF-FCM αλγορίθμου παίρνει τελικά την μορφή

$$Q_\lambda(\Psi) = - \sum_{i=1}^q \sum_{j=1}^s r_{ij} \log p(\mathbf{y}_j | x_j = i; \boldsymbol{\theta}_i) + \lambda \sum_{i=1}^q \sum_{j=1}^s r_{ij} \log \left(\frac{r_{ij}}{\pi_{ij}} \right) \quad (3.19)$$

όπου, $\Psi = \{R, \theta, \beta\}$, και η $p(\mathbf{y}_j | x_j = i; \theta_i)$ δίδεται από την (8).

3.4.2 Εκτιμητική των παραμέτρων του HMRF model με χρήση του HMRF-FCM αλγορίθμου

Για να εξάγουμε τις εκφράσεις των εκτιμητριών του HMRF model, δοθέντος ενός συνόλου εκπαίδευσης, υπό τη θεώρηση του HMRF-FCM αλγορίθμου, πρέπει να ελαχιστοποιήσουμε με επαναληπτικό τρόπο την fuzzy objective function $Q_\lambda(\Psi)$, που δίδεται από την (19), ως προς τα R , θ , και β , υπό το coordinate descent υπόδειγμα. Έστω $V^{(k)}$ η εκτιμήτρια της ποσότητας V που ελήφθη στην k οστη επανάληψη του αλγορίθμου. Έστω η $(k+1)$ οστη επανάληψη του αλγορίθμου. Έχουμε δηλαδή ότι η τιμή της $\hat{\Psi}$ είναι η ληφθείσα από την k οστη επανάληψη του αλγορίθμου, $\Psi^{(k)}$, ήτοι $\hat{\Psi} = \Psi^{(k)}$. Για να περαιώσουμε την $(k+1)$ οστη επανάληψη του αλγορίθμου, αρχικά πρέπει να βρούμε τις εκτιμήτριες των σημειακών (pointwise) Markov field priors, $\pi_{ij}^{(k)}$. Αυτό μπορεί να γίνει προσεγγιστικά, χρησιμοποιώντας την mean-field approximation της Markov field prior, $p(\mathbf{x})$. Από την (11) λαμβάνουμε [6,9]

$$p(\mathbf{x} | \beta^{(k)}) = \prod_{j=1}^s p(x_j | \mathbf{x}_{\partial_j}^{(k)}; \beta^{(k)}) \quad (3.20)$$

όπου [6,9]

$$\pi_{ij}^{(k)} \triangleq p(x_j = i | \mathbf{x}_{\partial_j}^{(k)}; \beta^{(k)}) = \frac{\exp(-\sum_{c \in \partial j} V_c(\tilde{\mathbf{x}}_{ij}^{(k)} | \beta^{(k)}))}{\sum_{h=1}^q \exp(-\sum_{c \in \partial j} V_c(\tilde{\mathbf{x}}_{hj}^{(k)} | \beta^{(k)}))} \quad (3.21)$$

$\tilde{\mathbf{x}}_{ij}^{(k)} \triangleq (x_j = i, \mathbf{x}_{\partial_j}^{(k)})$, και $\mathbf{x}_{\partial_j}^{(k)}$ είναι η τρέχουσα εκτιμήτρια (ληφθείσα από την κοστη επανάληψη του αλγορίθμου) της γειτονίας του j οστου site. Στον προτεινόμενο HMRF-FCM αλγόριθμο, η εκτιμήτρια $\mathbf{x}^{(k)}$ των ετικετών των pixels στην κοστη επανάληψη του αλγορίθμου λαμβάνεται με χρήση μιας προφανούς μεθοδολογίας υπό το προτεινόμενο fuzzy context, η οποία περιλαμβάνει την αποασαφοποίηση (defuzzification) των υπολογισμένων fuzzy memberships $r_{ij}^{(k)}$, δια της αντιστοιχίσεως κάθε site στην MRF κατάσταση που βελτιστοποιεί την fuzzy membership function του εν λόγω site. Με άλλα λόγια, λαμβάνουμε

$$x_j^{(k)} = \operatorname{argmax}_{i=1}^q r_{ij}^{(k)} \quad (3.22)$$

Αναφορικά με την έκφραση των εκτιμητριών των fuzzy membership functions, $r_{ij}^{(k+1)}$, από την (19), έχουμε

$$r_{ij}^{(k+1)} = \frac{\pi_{ij}^{(k)} \exp\left(-\frac{1}{\lambda} d_{ij}^{(k)}\right)}{\sum_{h=1}^q \pi_{hj}^{(k)} \exp\left(-\frac{1}{\lambda} d_{hj}^{(k)}\right)} \quad (3.23)$$

όπου

$$d_{ij}^{(k)} \triangleq d_{ij}(\boldsymbol{\theta}_i^{(k)}) = -\log p(\mathbf{y}_j | x_j = i; \boldsymbol{\theta}_i^{(k)}) \quad (3.24)$$

που, βάσει της (8), δίδει

$$d_{ij}^{(k)} = \frac{w}{2} \log(2\pi) + \frac{1}{2} \log |\boldsymbol{\Sigma}_i^{(k)}| + \frac{1}{2} (\mathbf{y}_j - \boldsymbol{\mu}_i^{(k)})^T \boldsymbol{\Sigma}_i^{(k)-1} (\mathbf{y}_j - \boldsymbol{\mu}_i^{(k)}) \quad (3.25)$$

Σχετικά με τις εκτιμήτριες των $\boldsymbol{\mu}_i$, από την (19) έχουμε

$$\boldsymbol{\mu}_i^{(k+1)} = \sum_{j=1}^s r_{ij}^{(k)} \mathbf{y}_j / \sum_{j=1}^s r_{ij}^{(k)} \quad (3.26)$$

Καθ' όμοιον τρόπο, για τους πίνακες συνδιακύμανσης Σ_i λαμβάνουμε

$$\Sigma_i^{(k+1)} = \frac{\sum_{j=1}^s r_{ij}^{(k)} (\mathbf{y}_j - \boldsymbol{\mu}_i^{(k)}) (\mathbf{y}_j - \boldsymbol{\mu}_i^{(k)})^T}{\sum_{j=1}^s r_{ij}^{(k)}} \quad (3.27)$$

ενώ, για την παράμετρο θερμοκρασίας β ισχύει

$$\beta^{(k+1)} = \operatorname{argmax}_{\beta} \sum_{i=1}^q \sum_{j=1}^s r_{ij}^{(k)} \log p(x_j = i | \mathbf{x}_{\partial_j}^{(k)}; \beta) \quad (3.28)$$

Σημειώνεται ότι η ακριβής έκφραση της εκτιμητριας $\beta^{(k+1)}$ εξαρτάται από την συγκεκριμένη επιλογή της συνάρτησης ενεργείας της Markov random field prior. Αυτή η επιλογή υποβάλλεται από την εκάστοτε εφαρμογή και εξαρτάται από το επίπεδο πολυπλοκότητας του μοντέλου που κάθε φορά είναι επιθυμητό [1].

Συνοπτικά, ο προτεινόμενος HMRF-FCM αλγόριθμος για την fuzzy επίλυση του HMRF model, επί τη βάση της mean-field προσέγγισης της MRF prior, περιλαμβάνει τα ακόλουθα βήματα:

1. Βρες μια εκτιμήτρια του $\mathbf{x}^{(k)}$ με χρήση του κανόνα (22).
2. Χρησιμοποιώντας την εκτιμήτρια $\mathbf{x}^{(k)}$, και επί τη βάση της mean-field approximation (20) της MRF κατανομής, βρες τις σημειακές prior πιθανότητες του Markov random field, $\pi_{ij}^{(k)}$, που δίδονται από την (21).
3. Υπολόγισε τις fuzzy membership functions, $r_{ij}^{(k+1)}$, χρησιμοποιώντας την (23).
4. Υπολόγισε τις εκτιμήτριες $\boldsymbol{\mu}_i^{(k+1)}$ και $\Sigma_i^{(k+1)}$ από τις (26) και (27), αντιστοίχως.

5. Υπολόγισε την $\beta^{(k+1)}$ από την (28).
6. Σε περίπτωση σύγκλισης, ήτοι $\left| Q_\lambda(\Psi^{(k+1)}) - Q_\lambda(\Psi^{(k)}) \right| / Q_\lambda(\Psi^{(k)}) < T_c$ όπου, T_c το κατώφλι σύγκλισης, έξελθε. Αλλιώς, $k := k + 1$ και επίστρεψε στο 1.

Συγκρίνοντας την έκφραση του $Q_\lambda(\Psi)$ για τον HMRF-FCM αλγόριθμο με την αντικειμενική συνάρτηση, $Q(\Psi)$, του EM αλγορίθμου για το HMRF μοντέλο, του επίσης διαμορφωμένου επί τη βάση της mean-field approximation της κατανομής του MRF, όπως έχει επί παραδείγματι χρησιμοποιηθεί στις μεθόδους PPL-EM [6], simulated field (SF), και mean field (MF) [9], με

$$Q(\Psi) = - \sum_{i=1}^q \sum_{j=1}^s p(x_j = i | \mathbf{y}_j) \log p(\mathbf{y}_j | x_j = i; \boldsymbol{\theta}_i) - \sum_{i=1}^q \sum_{j=1}^s p(x_j = i | \mathbf{y}_j) \log \pi_{ij}$$

μπορούμε εύκολα να παρατηρήσουμε ότι ο HMRF-FCM αλγόριθμος αποδίδει μια οικογένεια αντικειμενικών συναρτήσεων οι οποίες, κατ' ουσίαν, αποτελούν μια FCM-τύπου εναλλακτική του EM αλγορίθμου για το HMRF model, βασισμένη στην mean-field approximation (11) της Markov random field prior κατανομής. Σημειώνεται τέλος ότι για μοναδιαίο degree of fuzziness, $\lambda = 1$, ο HMRF-FCM αλγόριθμος ανάγεται στον βασισμένο στην mean-field approximation (11) EM αλγόριθμο για το HMRF model, καθώς

$$r_{ij}^{(k+1)} = \frac{\pi_{ij}^{(k)} p(\mathbf{y}_j | x_j = i; \boldsymbol{\theta}_i^{(k)})}{\sum_{h=1}^q \pi_{hj}^{(k)} p(\mathbf{y}_j | x_j = h; \boldsymbol{\theta}_h^{(k)})}$$

που, κατ' ουσίαν, αποτελεί την έκφραση των σημειακών posterior κατανο-

μών του Markov random field, $p(x_j = i | \mathbf{y}_j, \mathbf{x}_{\partial_j}^{(k)}; \Psi^{(k)})$, των λαμβανομένων επί τη βάση της mean-field approximation (11) του MRF [6].

Τέλος, σχετικά με τα υπολογιστικά κόστη του προτεινομένου αλγορίθμου, επισημαίνουμε πως αυτά είναι συγκρίσιμα με τα υπολογιστικά κόστη EM-τύπου προσεγγίσεων στο HMRF model, που βασίζονται επίσης στην mean-field approximation της MRF prior κατανομής, καθώς τα βασικά υπολογιστικά κόστη σε όλους αυτούς τους αλγορίθμους οφείλονται στην εκτίμηση και αντιστροφή εμπειρικών πινάκων συνδιακύμανσης. Επιπλέον, επισημαίνεται ότι ο αλγόριθμός μας είναι τάξεις μεγέθους πιο αποδοτικός σε σχέση με Bayesian sampling μεθόδους εκτιμητικής του HMRF model.

3.5 Πειραματική Μελέτη

Ο αλγόριθμος HMRF-FCM παρέχει μια καινοτόμο FCM-τύπου μεθοδολογία ειδικά κατάλληλη για clustering χωρικών εξηρημένων δεδομένων, π.χ., σε εφαρμογές επεξεργασίας εικόνας [7]. Στην παρούσα εργασία, προκειμένου να αξιολογήσουμε πειραματικά την μέθοδό μας, ασχολούμεθα με το πρόβλημα της κατάτμησης εικόνων. Εκτελούμε την διαδικασία κατάτμησης εικόνων επιβάλλοντας το (hidden) Markov random field στις ετικέτες των pixels της εικόνας (ήτοι τα τμήματα της εικόνας όπου αντιστοιχίζονται τα διάφορα pixels). Στα πειράματά μας, υιοθετούμε ένα τάξης-2 (8-γείτονες) σύστημα γειννίας [4], με συνάρτηση ενέργειας διδόμενη από την έκφραση

$$U(\mathbf{x}|\beta) = -\beta \sum_{j \in S} \sum_{l \in \partial_j} \delta(x_j - x_l) \quad (3.29)$$

τοιουτοτρόπως λαμβάνοντας

$$\pi_{ij}^{(k)} = \frac{\exp(\beta^{(k)} \sum_{l \in \partial_j} \delta(i - x_l^{(k)}))}{\sum_{h=1}^q \exp(\beta^{(k)} \sum_{l \in \partial_j} \delta(h - x_l^{(k)}))} \quad (3.30)$$

για τις σημειακές MRF prior κατανομές, όπου $\delta(\cdot)$ είναι η συνάρτηση Kronecker's delta, οριζόμενη ως

$$\delta(x_j - x_l) = \begin{cases} 1, & \text{if } x_j = x_l \\ 0, & \text{otherwise} \end{cases}$$

Αυτός ο τύπος συστήματος γειτνίασης και συνάρτησης ενεργείας χρησιμοποιείται στην αξιολόγηση τόσο του HMRF-FCM αλγορίθμου όσο και των θεωρουμένων HMRF-τύπου ανταγωνιστών του. Επισημαίνουμε επίσης ότι στα ακόλουθα πειράματα όλοι οι αξιολογούμενοι αλγόριθμοι αρχικοποιούνται καθ' όμοιον τρόπο, μέσω απλής επιβολής ενός κατωφλίου (thresholding) στις υπό μελέτη εικόνες. Τέλος, σημειώνουμε ότι ο βαθμός ασαφείας (degree of fuzziness) του HMRF-FCM αλγορίθμου και των θεωρουμένων FCM-τύπου ανταγωνιστών του προσδιορίζεται ευρετικά, δια βελτιστοποίησης της επίδοσής τους.

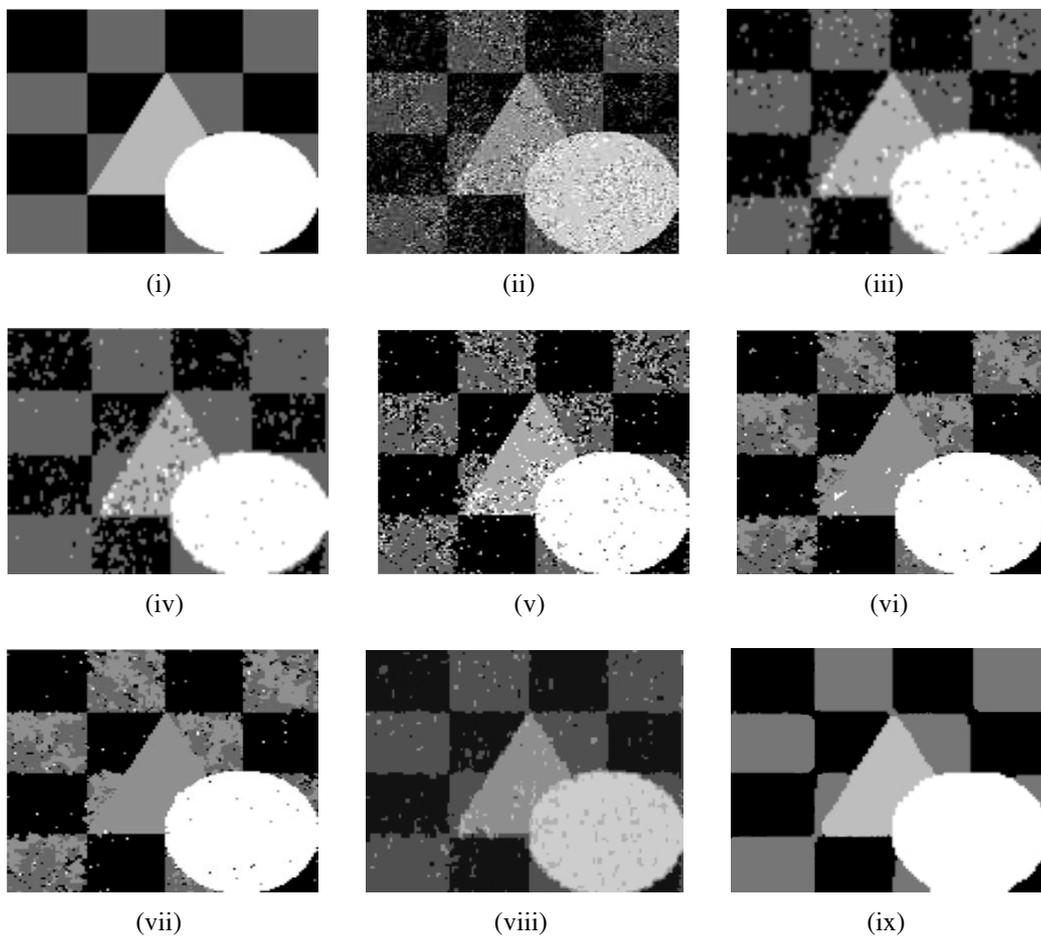
3.5.1 Κατάτμηση θορυβωδών εικόνων

Στο πρώτο αυτό πείραμα, δείχνουμε την ευρωστία του προτεινομένου αλγορίθμου στον θόρυβο. Για τον σκοπό αυτό, θεωρούμε μια εφαρμογή του HMRF-FCM αλγορίθμου στην κατάτμηση μιας διαβρωμένης δια θορύβου εικόνας. Για σύγκριση, εκτός της μεθόδου μας, παραθέτουμε επίσης την επίδοση των αλγορίθμων FCM [12], KL-FCM [37], ICM [3], MF [9], SF [9],

και CA-SVFMM [39]. Οι αλγόριθμοι MF και SF είναι EM-τύπου προσεγγίσεις στο HMRF βασισμένες στην mean-field approximation (11) της MRF κατανομής, και χρησιμοποιούν διαφορετικές μεθοδολογίες κτήσης των εκτιμητριών \hat{x} . Ο CA-SVFMM αλγόριθμος είναι βασισμένος σε μια μεθοδολογία MAP εκτιμητικής ενός Gaussian mixture model με χωρικούς περιορισμένους, οριζομένους δια της επιβολής ενός απλοποιημένου Gauss–Markov random field.

Η θεωρούμενη εικόνα, επιδεικνυόμενη στο Σχ. 1(ii), ελήφθη προσθέτοντας χωρικά συσχετισμένο θόρυβο στην 4-χρωμάτων εικόνα σκακιέρας που δίδεται στο Σχ. 1(i). Για να παράγουμε αυτόν τον θόρυβο, δειγματοληφθήθηκε μια διαμόρφωση δυαδικών ενδεικτών από ένα Potts-model MRF [4], με χρήση της τεχνικής Gibbs sampling. Αν ο εκάστοτε λαμβανόμενος ενδείκτης για κάποιο pixel ήταν μοναδιαίος, προσθέταμε θόρυβο ακολουθώντας την τυποποιημένη Gaussian κατανομή, αλλιώς δεν προσθέσαμε κανένα θόρυβο. Είναι ξεκάθαρο ότι κατά αυτό τον τρόπο λαμβάνεται ένα χωρικά συσχετισμένο είδος θορύβου, προσφέροντας μια πολύ απαιτητική περίπτωση πειραματικής αξιολόγησης.

Τα λαμβανόμενα αποτελέσματα δίδονται στα Σχ. 1(iii)-(ix). Παρατηρούμε ότι ο προτεινόμενος αλγόριθμος παρέχει αξιοσημείωτη ευρωστία υπό τις θεωρηθείσες συνθήκες βαρέος θορύβου. Έτσι, παρέχει μια εξαιρετικώς ανώτερη, σε σχέση με τον ανταγωνισμό, επίδοση, προσφέροντας μια εξαιρετικά ακριβή κατάτμηση της θεωρηθείσας εικόνας, εν αντιθέσει με την φτωχή επίδοση των ανταγωνιστών του.



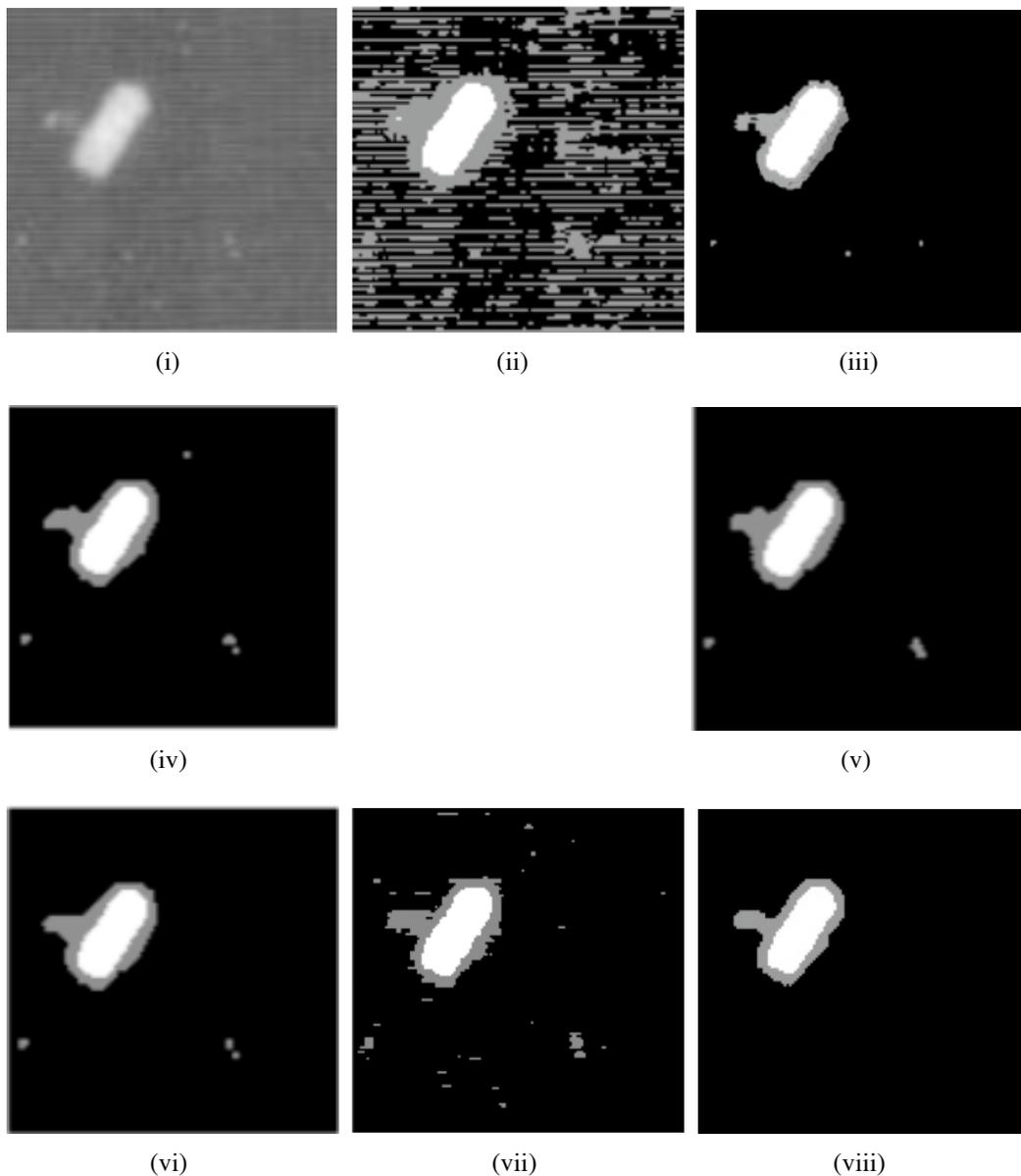
Σχήμα 3.1: (i) Αρχική 4-χρωμάτων εικόνα σκακιέρας. (ii) Διαβρωμένη δια θορύβου εικόνα. (iii) Κατάτμηση της εικόνας δια του αλγορίθμου FCM. (iv) Κατάτμηση της εικόνας δια του αλγορίθμου KL-FCM. (v) Κατάτμηση της εικόνας δια του αλγορίθμου ICM. (vi) Κατάτμηση της εικόνας δια του αλγορίθμου SF. (vii) Κατάτμηση της εικόνας δια του αλγορίθμου MF. (viii) Κατάτμηση της εικόνας δια του αλγορίθμου CA-SVFMM. (ix) Κατάτμηση της εικόνας δια του αλγορίθμου HMRF-FCM.

3.5.2 Κατάτμηση φυσικώς ληφθεισών εικόνων

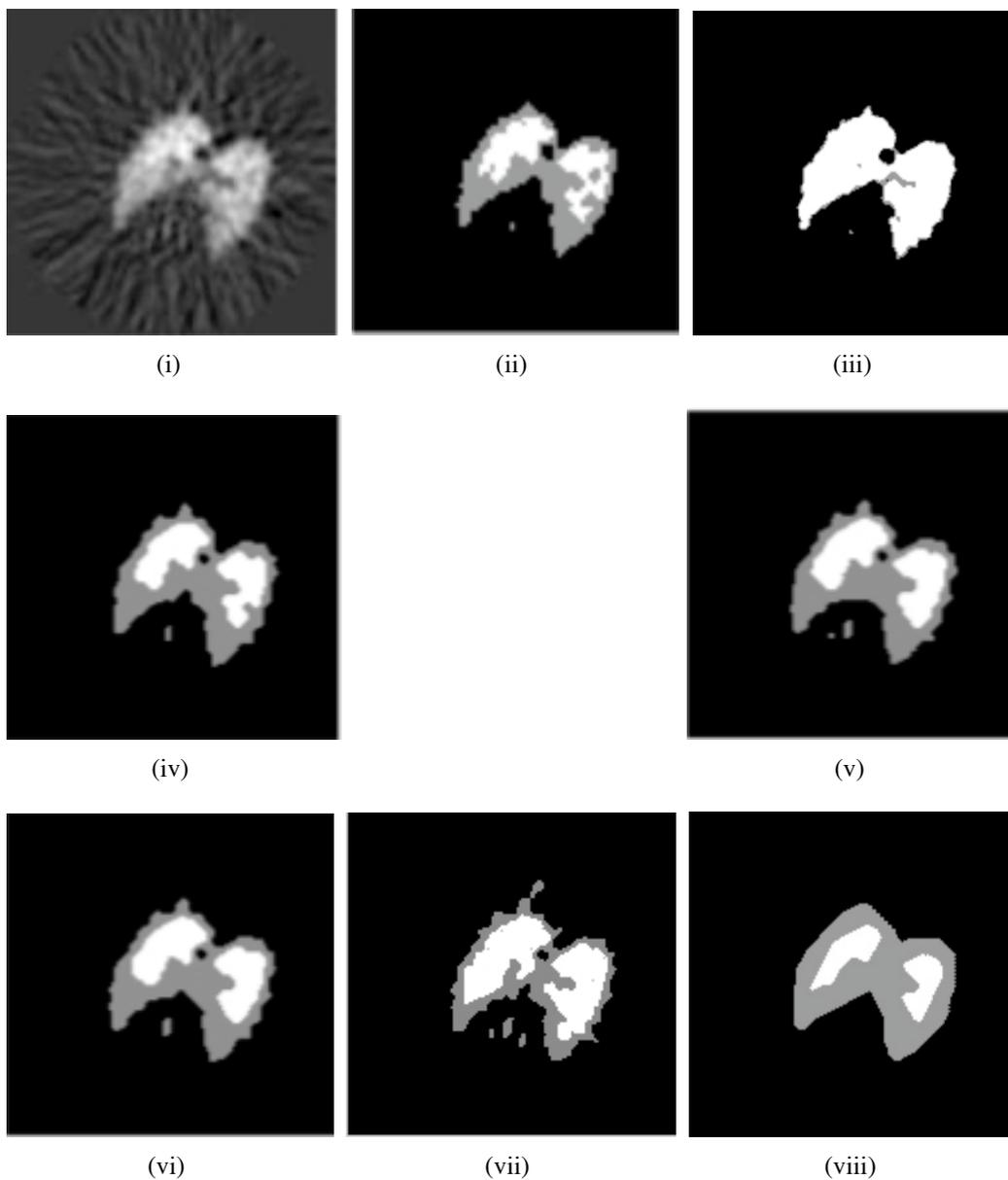
Σε φυσικώς ληφθείσες εικόνες, ένα πλήθος διαφορετικών εκφάνσεων του προβλήματος κατάτμησης εικόνων έρχονται στο προσκήνιο, όπως θόρυβος, παραμορφώσεις, διαταραχές φωτισμού, κ.ο.κ. Συνεπώς, η κατάτμηση φυσικών εικόνων είναι ένα αξιόλογο benchmark για την αξιολόγηση των αλγορίθμων κατάτμησης. Στο πείραμα αυτό, θεωρούμε δυο φυσικώς ληφθείσες εικόνες, προταθείσες στις εργασίες [35, 40]. Η πρώτη, απεικονιζόμενη στο Σχ. 2(i), είναι μια 100×100 αεροφωτογραφία ενός σημαντήρα (buoy) επιπλέοντος σε βαθύ μπλε νερό, ενώ η δεύτερη, απεικονιζόμενη στο 3(i), είναι μια 128×128 απεικόνιση ποζιτρονιογράφου (PET) του πνεύμονα ενός σκυλιού. Οι κατατμήσεις των εικόνων αυτών δια των μεθόδων FCM, KL-FCM, ICM, SF, MF, CA-SVFMM, και HMRF-FCM, θέτοντας $q = 3$, δίνονται στα Σχ. 2(ii)-(viii) και 3(ii)-(viii), αντιστοίχως. Παρατηρούμε ότι ο προτεινόμενος αλγόριθμος δίνει πιο ομογενείς περιοχές, και άρα, προσφέρει καλύτερη επίδοση σε σχέση με τον ανταγωνισμό.

3.5.3 Αξιολόγηση με το Berkeley image segmentation benchmark

Στο πείραμα αυτό, αξιολογούμε την επίδοση του αλγορίθμου HMRF-FCM, χρησιμοποιώντας ένα υποσύνολο του *Berkeley image segmentation data set and bench-mark* [41], το οποίο περιλαμβάνει ένα σύνολο εγχρώμων φυσικώς ληφθεισών εικόνων καθώς και τις ground truth κατατμήσεις τους από διαφορετικά άτομα (ειδικούς). Δοθέντων των πολλαπλών κατατμήσεων για κάθε εικόνα, για να λάβουμε μια αντικειμενική εκτίμηση της επίδοσης του κάθε αλγορίθμου, χρησιμοποιούμε το probabilistic rand (PR) index [42, 43].



Σχήμα 3.2: (i) Εικόνα σημαντήρα (buoy). (ii) Κατάτμηση της εικόνας δια του αλγορίθμου FCM. (iii) Κατάτμηση της εικόνας δια του αλγορίθμου KL-FCM. (iv) Κατάτμηση της εικόνας δια του αλγορίθμου ICM. (v) Κατάτμηση της εικόνας δια του αλγορίθμου SF. (vi) Κατάτμηση της εικόνας δια του αλγορίθμου MF. (vii) Κατάτμηση της εικόνας δια του αλγορίθμου CA-SVFMM. (viii) Κατάτμηση της εικόνας δια του αλγορίθμου HMRF-FCM.



Σχήμα 3.3: (i) PET απεικόνιση του πνεύμονα ενός σκυλιού. (ii) Κατάτμηση της εικόνας δια του αλγορίθμου FCM. (iii) Κατάτμηση της εικόνας δια του αλγορίθμου KL-FCM. (iv) Κατάτμηση της εικόνας δια του αλγορίθμου ICM. (v) Κατάτμηση της εικόνας δια του αλγορίθμου SF. (vi) Κατάτμηση της εικόνας δια του αλγορίθμου MF. (vii) Κατάτμηση της εικόνας δια του αλγορίθμου CA-SVFMM. (viii) Κατάτμηση της εικόνας δια του αλγορίθμου HMRF-FCM.

Το PR index μετράει το ποσοστό των ζευγών pixels οι ετικέτες των οποίων είναι οι ίδιες μεταξύ μιας υπολογιζόμενης κατάτμησης και μιας δοθείσης ground truth κατάτμησης, αθροίζοντας επί πολλαπλών ground truth κατατμήσεων ώστε να λάβει υπόψιν τις διαφορετικές εκτιμήσεις που θα μπορούσαν να προκύψουν υπό την ανθρώπινη αντίληψη. Ορίζοντας ως $G = \{G_1, G_2, \dots, G_M\}$ ένα σύνολο ground truth εικόνων, και ως G_{eval} μια εξ αλγορίθμου ειλημμένη κατάτμηση υπό αξιολόγηση, ισχύει

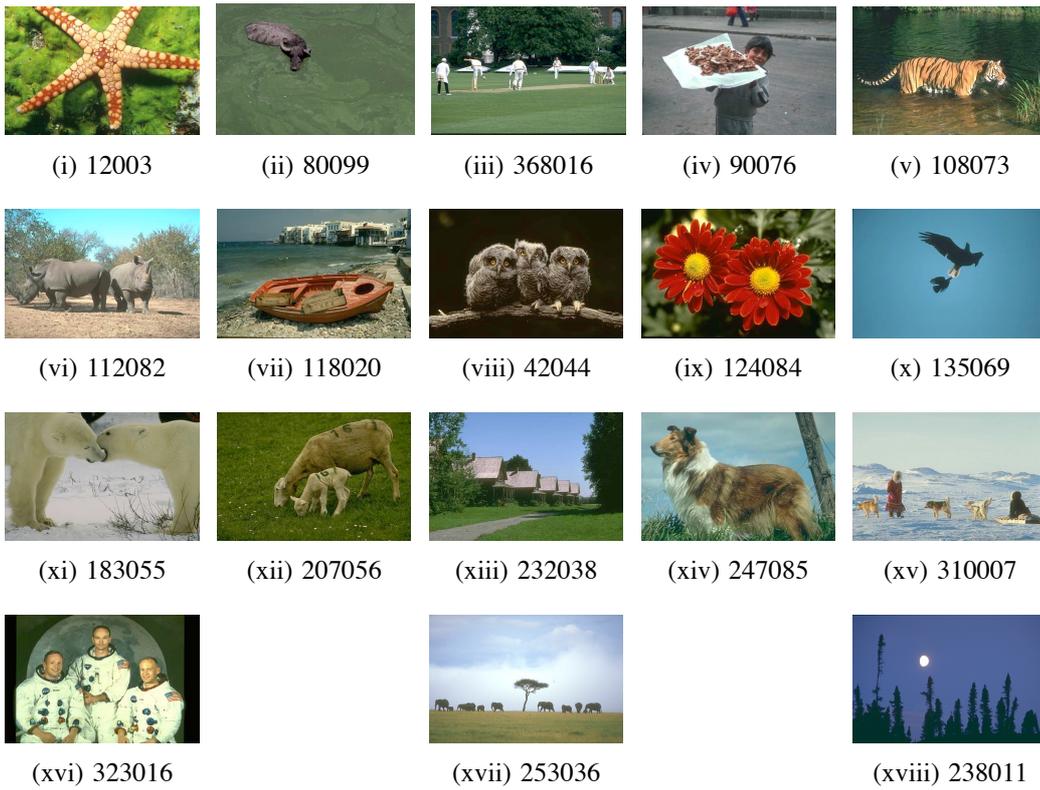
$$PR(G_{eval}, G) = \frac{2}{s(s-1)} \sum_i \sum_{j>i} [c_{ij}p_{ij} + (1-c_{ij})(1-p_{ij})] \quad (3.31)$$

όπου $c_{ij} = 1$ αν τα pixels i και j ανήκουν στο ίδιο segment στην κατάτμηση G_{eval} , $c_{ij} = 0$ ειδάλως, s είναι ο αριθμός των pixels στην εικόνα, και p_{ij} είναι η πιθανότητα βάσει του ground truth τα pixels i και j να ανήκουν στο ίδιο segment, υπολογιζόμενη ως το κλάσμα των διαθέσιμων ground truths όπου τα pixels i και j έχουν αντιστοιχηθεί στο ίδιο segment. Το PR index παίρνει τιμές μεταξύ 0 και 1, με τις τιμές που βρίσκονται κοντά στο 0 να σηματοδοτούν ένα κακό αποτέλεσμα (ήτοι μια φτωχή κατάτμηση), και τις τιμές που βρίσκονται κοντά στο 1 να σηματοδοτούν ένα καλό αποτέλεσμα

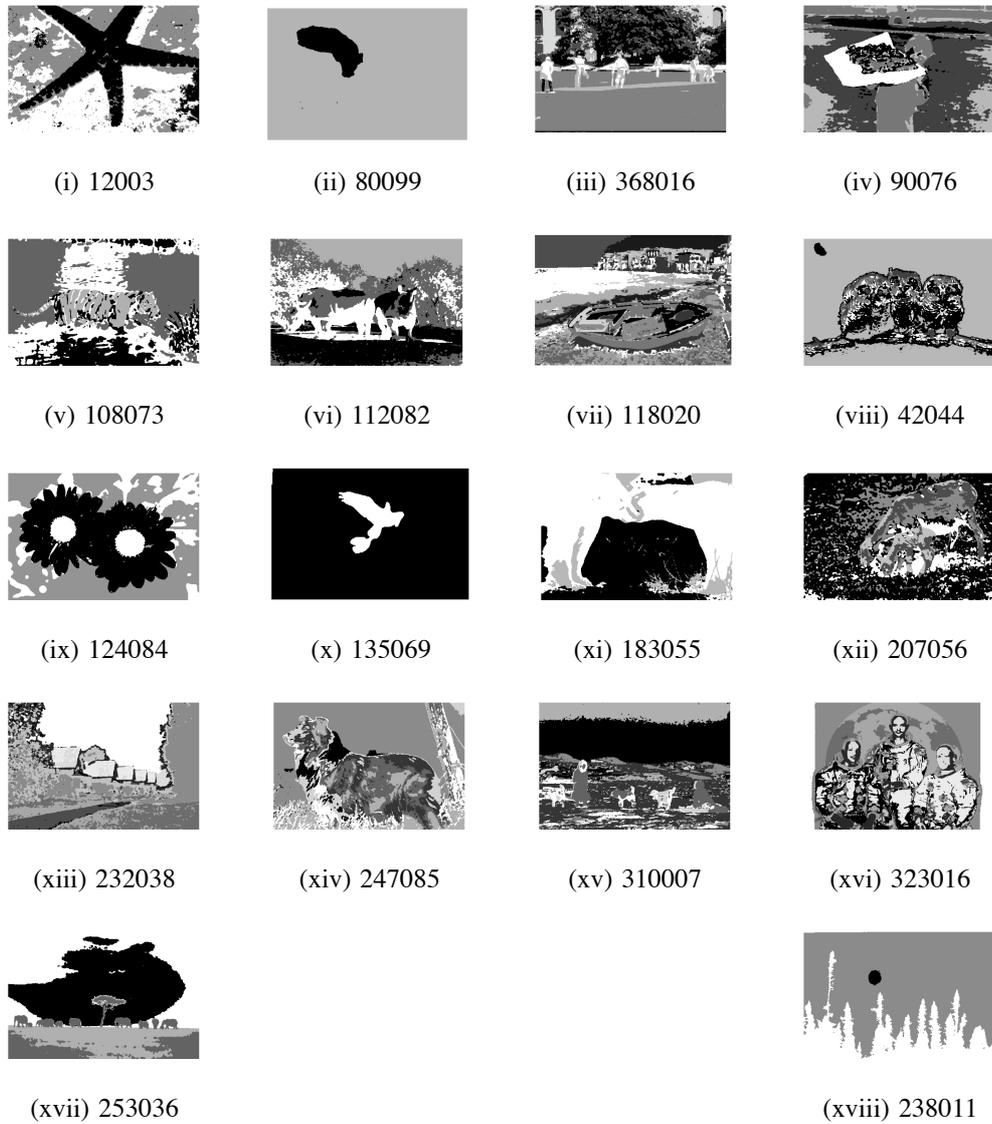
Για λόγους σύγκρισης, πλέον της δικής μας μεθόδου, αξιολογούμε την επίδοση και των μεθόδων FCM, KL-FCM, SF, MF, και CA-SVFMM. Στον πίνακα 1, δίνουμε τις τιμές του PR δείκτη τις ληφθείσες για καθεμία από τις θεωρούμενες εικόνες, από καθένα των αξιολογουμένων αλγορίθμων, θέτοντας κάθε φορά τον αριθμό των εξακτέων segments, q , με τον ίδιο τρόπο όπως στην εργασία [39]. Στο Σχ. 4, δίνουμε μερικές από τις εικόνες του χρησιμοποιούμενου συνόλου δεδομένων. Στο Σχ. 5, παρέχουμε τις αντίστοιχες

Πίνακας 3.1: Berkeley Image Data Set: Ληφθέντες PR Δείκτες για τις Θεωρηθείσες Εικόνες

Εικόνα #	q	FCM	KL-FCM	SF	MF	CA-SVFMM	HMRP-FCM
102061	15	0.69	0.69	0.76	0.75	0.73	0.78
80099	2	0.51	0.77	0.85	0.87	0.81	0.91
108073	4	0.51	0.54	0.58	0.58	0.49	0.62
118035	3	0.80	0.76	0.85	0.84	0.65	0.89
134008	2	0.47	0.47	0.52	0.50	0.63	0.58
135069	2	0.58	0.95	0.98	0.98	0.87	0.98
207056	5	0.55	0.64	0.67	0.67	0.64	0.71
232038	10	0.81	0.82	0.86	0.85	0.62	0.91
310007	7	0.65	0.69	0.74	0.74	0.7	0.78
323016	10	0.80	0.81	0.83	0.84	0.67	0.88
238011	3	0.77	0.86	0.89	0.91	0.84	0.95
299091	5	0.58	0.56	0.67	0.69	0.58	0.75
314016	4	0.47	0.48	0.50	0.50	0.52	0.55
368016	6	0.77	0.77	0.80	0.80	0.77	0.86
271031	5	0.62	0.55	0.71	0.69	0.63	0.74
253036	4	0.63	0.63	0.67	0.68	0.72	0.74
247085	7	0.69	0.68	0.76	0.76	0.64	0.80
28075	6	0.49	0.64	0.69	0.69	0.63	0.75
113044	5	0.67	0.61	0.76	0.78	0.62	0.78
12003	3	0.50	0.48	0.64	0.63	0.46	0.72
106024	4	0.65	0.65	0.68	0.68	0.68	0.70
183055	3	0.59	0.51	0.73	0.75	0.51	0.79
118020	12	0.77	0.77	0.84	0.84	0.7	0.84
90076	5	0.61	0.55	0.72	0.73	0.52	0.76
42044	5	0.65	0.68	0.73	0.71	0.74	0.76
163014	5	0.67	0.70	0.75	0.77	0.6	0.79
42049	7	0.69	0.69	0.70	0.71	0.62	0.73
112082	4	0.66	0.68	0.73	0.73	0.62	0.76
124084	3	0.52	0.45	0.72	0.73	0.56	0.77
Μέσος	-	0.63	0.66	0.74	0.74	0.64	0.78
Τυπ. Απόκλ.	-	0.11	0.13	0.1	0.11	0.09	0.1
Διάμεσος	-	0.65	0.68	0.73	0.73	0.65	0.77



Σχήμα 3.4: Berkeley image segmentation data set and benchmark: Αρχικές Εικόνες



Σχήμα 3.5: Berkeley image segmentation data set and benchmark: Κατατμήσεις Ληφθείσες από τον HMRF-FCM Αλγόριθμο.

κατατημήσεις τις ληφθείσες με χρήση του HMRF-FCM αλγορίθμου. Όπως παρατηρούμε, ο προτεινόμενος αλγόριθμος δίνει αξιοσημείωτα καλύτερα αποτελέσματα εν συγκρίσει με τους HMRF-τύπου και FCM-τύπου ανταγωνιστές του.

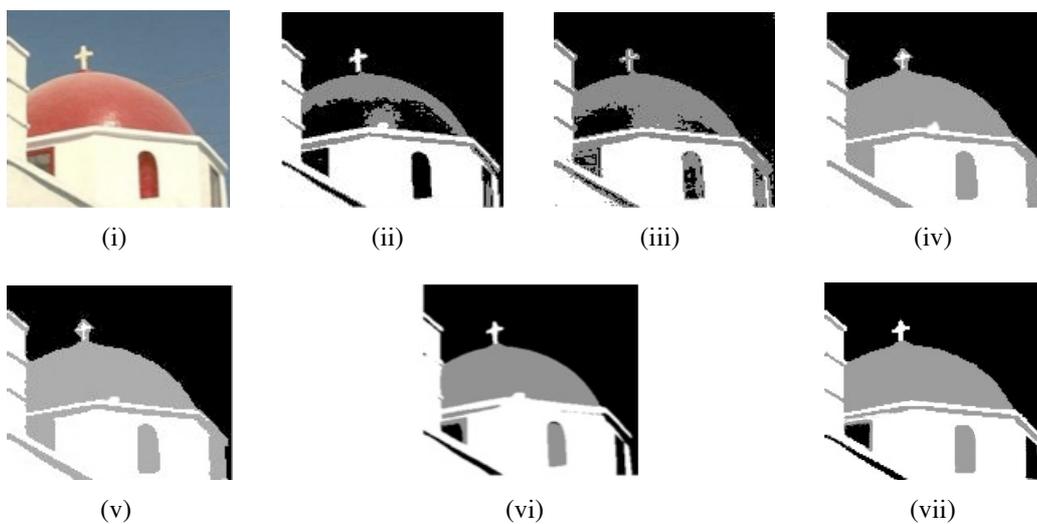
3.5.4 Κατάτμηση πολυδιαστάτων εικόνων

Τέλος, αξιολογούμε τον προτεινόμενο αλγόριθμο στην κατάτμηση πολυδιάστατων εικόνων. Αυτός ο τύπος εικόνων μπορεί να ληφθεί είτε δια της εξαγωγής σημαντικών features χαρακτηριζόντων το περιεχόμενο της εικόνας (π.χ. πληροφορία υφής - textural features), ή απευθείας μέσω αισθητήρων, όπως συμβαίνει με δορυφορικές ή RGB έγχρωμες εικόνες από τηλεοράσεις. Εδώ, ασχολούμαστε με ένα παράδειγμα από την δεύτερη περίπτωση, θεωρούμε την RGB εικόνα που δίνεται στο Σχ. 6(i). Η εικόνα αυτή περιλαμβάνει κατά βάση 3 χρωματικά segments: το σκούρο ουρανό, τον ερυθρό τρούλο, και τον λευκό τοίχο της εκκλησίας. Επιπλέον, υπάρχουν ακόμη μερικά ελάχιστα segments κυρίως στα παράθυρα.

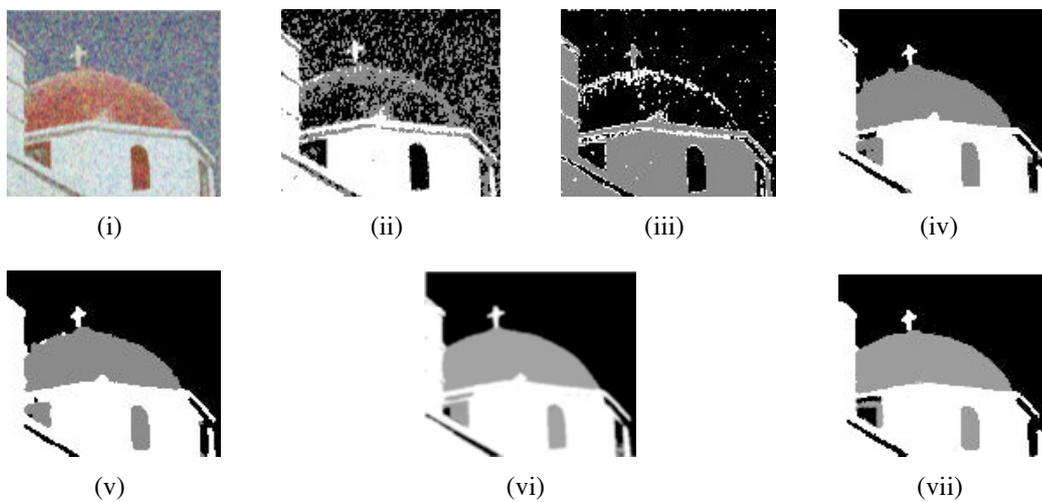
Οι κατατημήσεις της εν λόγω εικόνας με χρήση του HMRF-FCM αλγορίθμου καθώς και τον FCM, KL-FCM, SF, MF, και CA-SVFMM αλγορίθμων, με $q = 3$ clusters (HMRF καταστάσεις), δίδονται στο Σχ. 6(ii)-(vii). Παρατηρούμε ότι η προτεινόμενη μέθοδος κατατμεί καλύτερα το ερυθρό component διατηρώντας την λεπτή ακμή μεταξύ του τρούλου και του ουρανού, καθώς και την λεπτή ακμή μεταξύ του τρούλου και του λευκού τοίχου της εκκλησίας. Είναι αξιοσημείωτο ότι ο αλγόριθμος HMRF-FCM είναι πολύ εύρωστος στις ανακλάσεις στην μέση του τρούλου, σε αντίθεση με τους ανταγωνιστές του. Επιπλέον, διαπιστώνουμε ότι οι λεπτές σκιές στον τοίχο κάτω από την

βάση του τρούλου είναι καλύτερα διατηρημένοι από την δική μας μέθοδο.

Περαιτέρω, επαναλαμβάνουμε το προηγούμενο πείραμα χρησιμοποιώντας μια θορυβική εκδοχή της θεωρηθείσης RGB εικόνας, που ελήφθη προσθέτοντας σε καθεμία των R, G, και B χρωματικών συνιστωσών λευκό θόρυβο Gauss διαφορετικών τυπικών αποκλίσεων, καθεξής λαμβάνοντας σηματοθορυβικούς λόγους (SNRs) ίσους με 2, 4, και 3 dB, αντιστοίχως. Η τοιούτοτρόπως παραχθείσα εικόνα δίδεται στο Σχ. 7(i). Οι κατατμήσεις της εν λόγω εικόνας με χρήση των μεθόδων FCM, KL-FCM, SF, MF, CA-SVFMM, και HMRF-FCM δίδονται στο Σχ. 7(ii)-(vii). Ως παρατηρούμε, οι clustering αλγόριθμοι οι μη λαμβάνοντες υπόψιν τις χωρικές αλληλοεξαρτήσεις των γειτονικών pixels δίνουν ένα φτωχό αποτέλεσμα, εν αντιθέσει με τις HMRF-τύπου μεθόδους. Τέλος, παρατηρούμε ακόμη ότι ο HMRF-FCM αλγόριθμος ξανά αποδίδει καλύτερη διατήρηση των ακμών του τρούλου, πλήρως αγνοώντας τις επ' αυτού ανακλάσεις.



Σχήμα 3.6: Κατάτμηση RGB εικόνας: (i) Αρχική εικόνα. (ii) Κατάτμηση της εικόνας δια του αλγορίθμου FCM. (iii) Κατάτμηση της εικόνας δια του αλγορίθμου KL-FCM. (iv) Κατάτμηση της εικόνας δια του αλγορίθμου SF. (v) Κατάτμηση της εικόνας δια του αλγορίθμου MF. (vi) Κατάτμηση της εικόνας δια του αλγορίθμου CA-SVFMM. (vii) Κατάτμηση της εικόνας δια του αλγορίθμου HMRF-FCM.



Σχήμα 3.7: Κατάτμηση θορυβώδους RGB εικόνας: (i) Αρχική εικόνα. (ii) Κατάτμηση της εικόνας δια του αλγορίθμου FCM. (iii) Κατάτμηση της εικόνας δια του αλγορίθμου KL-FCM. (iv) Κατάτμηση της εικόνας δια του αλγορίθμου SF. (v) Κατάτμηση της εικόνας δια του αλγορίθμου MF. (vi) Κατάτμηση της εικόνας δια του αλγορίθμου CA-SVFMM. (vii) Κατάτμηση της εικόνας δια του αλγορίθμου HMRF-FCM.

Βιβλιογραφία

- [1] G. Winkler, *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*, 2nd ed., ser. Stochastic Modeling and Applied Probability. Springer, 2006.
- [2] W. Qian and D. Titterington, "Parameters estimation for hidden Gibbs chains," *Statistics and Probability Letters*, vol. 10, pp. 49--58, 1990.
- [3] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society B*, vol. 48, pp. 259--302, 1986.
- [4] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721--741, 1984.
- [5] J. Besag, "Statistical analysis of non-lattice data," *Statistician*, vol. 24, pp. 179--195, 1975.
- [6] W. Qian and D. Titterington, "Estimation of parameters in hidden Markov models," *Philosophical Transactions of the Royal Society of London A*, vol. 337, pp. 407--428, 1991.

- [7] N. Hjort, E. Mohn, and G. Storvik, "A simulation study of some contextual classification methods for remotely sensed data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 25, no. 6, pp. 796--804, 1987.
- [8] A. P. Dunmur and D. Titterton, "Parameter estimation in latent profile models," *Computational Statistics and Data Analysis*, pp. 371--388, 2007.
- [9] G. Celeux, F. Forbes, and N. Peyrard, "EM procedures using mean field-like approximations for Markov model-based image segmentation," *Pattern Recognition*, vol. 36, no. 1, pp. 131--144, 2003.
- [10] L. Zadeh, "Fuzzy sets," *Inform. Contr.*, vol. 8, pp. 338--353, 1965.
- [11] J. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York and London: Plenum Press, 1981.
- [12] J. C. Bezdek, N. R. Pal, J. Keller, and R. Krishnapuram, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. Norwell, MA, USA: Kluwer Academic Publishers, 1999.
- [13] S. Chen and D. Zhang, "Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure," *IEEE Trans. Syst., Man, Cybern. B*, vol. 34, no. 4, pp. 1907--1916, 2004.
- [14] D. Pham and J. L. Prince, "An adaptive fuzzy c -means algorithm for image segmentation in the presence of intensity inhomogeneities," *Pattern Recognit. Lett.*, vol. 20, pp. 57--68, 1999.
- [15] J. Leski, "Toward a robust fuzzy clustering," *Fuzzy Sets Syst.*, vol. 137, no. 2, pp. 215--233, 2003.

- [16] R. J. Hathaway and J. C. Bezdek, "Generalized fuzzy c -means clustering strategies using L_p norm distances," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 5, pp. 576--582, 2000.
- [17] J. Leski, "An ϵ -insensitive approach to fuzzy clustering," *Int. J. Applicat. Math. Comp. Sci.*, vol. 11, no. 4, pp. 993--1007, 2001.
- [18] S. Chatzis and T. Varvarigou, "Factor analysis latent subspace modeling and robust fuzzy clustering using t distributions," *IEEE Transactions on Fuzzy Systems*, Accepted for future publication.
- [19] N. R. Pal, K. Pal, J. M. Keller, and J. Bezdek, "A possibilistic fuzzy c -means clustering algorithm," *IEEE Trans. Fuzzy Systems*, vol. 13, no. 4, pp. 517--530, 2005.
- [20] F. Masulli and S. Rovetta, "Soft transition from probabilistic to possibilistic fuzzy clustering," *IEEE Trans. Fuzzy Systems*, vol. 14, no. 4, pp. 516--527, 2006.
- [21] J. Yu and M. S. Yang, "A generalized fuzzy clustering regularization model with optimality tests and model complexity analysis," *IEEE Trans. Fuzzy Systems*, vol. 15, no. 5, pp. 904--915, 2007.
- [22] S. Nefti, M. Oussalah, and U. Kaymak, "A new fuzzy set merging technique using inclusion-based fuzzy clustering," *IEEE Trans. Fuzzy Systems*, vol. 16, no. 1, pp. 145--161, 2008.
- [23] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag, and T. Moriarty, "A modified fuzzy c -means algorithm for bias field estimation and segmentation

- of MRI data," *IEEE Transactions on Medical Imaging*, vol. 21, pp. 193--199, 2002.
- [24] Y. A. Tolias and S. M. Panas, "Image segmentation by a fuzzy clustering algorithm using adaptive spatially constrained membership functions," *IEEE Trans. Syst., Man, Cybern. A*, vol. 28, no. 359-369, 1998.
- [25] D. L. Pham, "Fuzzy clustering with spatial constraints," in *IEEE Int. Conf. Image Processing*, vol. 2, 2002, pp. 65--68.
- [26] A.-C. Liew, H. Yan, and N. F. Law, "Image segmentation based on adaptive cluster prototype estimation," *IEEE Trans. Fuzzy Systems*, vol. 13, no. 4, pp. 444--453, 2005.
- [27] J. Maroquin, S. Mitte, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *Journal of The American Statistical Association*, vol. 82, pp. 76--89, 1987.
- [28] P. Clifford, "Markov random fields in statistics," in *Disorder in physical systems. A volume in honour of John M. Hammersley on the occasion of his 70th birthday*, G. Grimmett and D. Welsh, Eds. Clarendon Press, Oxford: Oxford Science Publication, 1990.
- [29] G. McLachlan and D. Peel, *Finite Mixture Models*. Wiley Series in Probability and Statistics, 2000.
- [30] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 45--57, 2001.

- [31] S. Geman and C. Graffigne, "Markov random field image models and their applications to computer vision," in *Proceedings of the International Conference of Mathematicians*. American Mathematical Society, 1987, pp. 1498--1517.
- [32] B. Chalmond, "An iterative Gibbsian technique for reconstruction of m -ary images," *Pattern Recognition*, vol. 22, no. 6, pp. 747--761, 1989.
- [33] J. Zhang, "The mean field theory in EM procedures for Markov random fields," *IEEE Transactions on Image Processing*, vol. 2, no. 1, pp. 27--40, 1993.
- [34] D. Chandler, *Introduction to Modern Statistical Mechanics*. New York: Oxford University Press, 1987.
- [35] F. Forbes and N. Peyrard, "Hidden Markov random field model selection criteria based on mean field-like approximations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1089--1101, 2003.
- [36] S. Miyamoto and M. Mukaidono, "Fuzzy c -means as a regularization and maximum entropy approach," in *Proceedings of the Seventh International Fuzzy Systems Association World Congress*, vol. 2, 1997, pp. 86--92.
- [37] H. Ichihashi, K. Miyagishi, and K. Honda, "Fuzzy c -means clustering with regularization by K-L information," in *10th IEEE Int. Conf. Fuzzy Systems*, 2001, pp. 924--927.
- [38] E. Ruspini, "A new approach to clustering," *Inform. Contr.*, vol. 15, pp. 22--32, 1969.

- [39] C. Nikou, N. Galatsanos, and A. Likas, "A class-adaptive spatially variant mixture model for image segmentation," *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 1121--1130, 2007.
- [40] D. C. Stanford, "Fast automatic unsupervised image segmentation and curve detection in spatial point patterns," Ph.D. dissertation, Department of Statistics, University of Washington, 1999.
- [41] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proceedings of 8th IEEE International Conference on Computer Vision*, vol. 2, 2001, pp. 416--423.
- [42] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "A measure for objective evaluation of image segmentation algorithms," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 3, 2005, pp. 34--41.
- [43] R. Unnikrishnan and M. Hebert, "Measures of similarity," in *Proc. IEEE Workshop Computer Vision Applications*, 2005, pp. 394--400.

Κεφάλαιο 4

Μια Καινοτόμος Μεθοδολογία Εύρωστης Μοντελοποίησης Ακολουθιακών Δεδομένων

Κρυπτο-Μαρκοβιανά μοντέλα (Hidden Markov models, HMM) χρησιμοποιούνται πεπερασμένες υπερθέσεις κανονικών κατανομών (Gaussian mixture models, GMM) ως τις κατανομές των κρυφών καταστάσεών τους, έχουν με επιτυχία χρησιμοποιηθεί σε αναρίθμητες εφαρμογές μοντελοποίησης και ταξινόμησης ακολουθιακών δεδομένων. Ωστόσο, τα GMM είναι γνωστό ότι είναι εξαιρετικά ευπαθή και δυσανεκτικά στην παρουσία ατυπικών δεδομένων (outliers) εντός των δειγμάτων εκπαίδευσής τους. Όπως έχουμε ήδη συζητήσει, οι πεπερασμένες υπερθέσεις κατανομών Student's- t (SMM) έχουν εσχάτως προταθεί ως μια πιο εύρωστη εναλλακτική στα GMM, υπερβαίνουσα αυτές τις ευπάθειες. Προκειμένου να εκμεταλλευθούμε αυτά τα πλεονεκτήματα των SMM, εισάγουμε στην παρούσα εργασία ένα πρωτότυπο

HMM όπου οι κατανομές των κρυφών καταστάσεων του μοντελοποιούνται με χρήση SMM. Εξάγουμε δε έναν αλγόριθμο για την εκτιμητική των παραμέτρων του μοντέλου υπό το υπόδειγμα της μεγίστης πιθανοφανείας, με χρήση του expectation-maximization (EM) αλγορίθμου.

Η εργασία μου αυτή έχει δημοσιευθεί υπό το άρθρο μου:

Sotirios Chatzis, Dimitrios Kosmopoulos, Theodora Varvarigou, " Robust Sequential Data Modeling Using an Outlier Tolerant Hidden Markov Model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Accepted for Future Publication, 2008.

4.1 Εισαγωγή

Τα HMM μοντέλα χρησιμοποιούνται σήμερα ευρέως σε προβλήματα αναγνώρισης προτύπων, καθώς παρέχουν μία εξαιρετικά επιτυχημένη μέθοδο μοντελοποίησης παρατηρήσεων που εμφανίζονται με ένα ακολουθιακό τρόπο και τείνουν να ταξινομούνται ή να εναλλάσσονται μεταξύ διαφορετικών πιθανών υποπληθυσμών. Ειδικότερα, HMM με συνεχείς κατανομές παρατηρήσεων έχουν χρησιμοποιηθεί με επιτυχία σε πλείστες εφαρμογές στην οικολογία, κρυπτογραφία, κατανόηση εικόνων και βίντεο, αναγνώριση ομιλίας και ομιλητή [1].

Οι κρυφές κατανομές εκπομπής παρατηρήσεων οι σχετιζόμενες με τις καταστάσεις ενός HMM συνεχών κατανομών πρέπει να δύνανται να προσεγγίζουν τυχαίες, αυθαίρετα πολύπλοκες κατανομές πιθανότητας. Τα GMM αποτελούν μια κοινή επιλογή για τις κατανομές των κρυφών καταστάσεων των HMM μοντέλων, αποδίδουσα τα συχνά αποκαλούμενα, Gaussian HMM

(GHMM) μοντέλα [2]. Η μεγάλη δημοφιλία των GHMM απορρέει από την γνωστή δυνατότητα των GMM να προσεγγίζουν με επιτυχία άγνωστες απείρως πολύπλοκες κατανομές πιθανότητας, προσφέροντας παραλλήλως ένα απλό και υπολογιστικά αποδοτικό πλαίσιο εκτιμητικής των παραμέτρων τους υπό το υπόδειγμα της μεγίστης πιθανοφανείας, με χρήση του EM αλγορίθμου [3]. Ωστόσο, τα GMM είναι γνωστό ότι υποφέρουν από ένα κρίσιμο μειονέκτημα αναφορικά με την εκτιμητική των παραμέτρων τους, η αποδοτικότητα της οποίας είναι γνωστό ότι μπορεί να επιδεινωθεί ανεπανόρθωτα από την παρουσία ατυπικών δεδομένων στα χρησιμοποιούμενα δείγματα εκπαίδευσης.

Το πρόβλημα της προστασίας από ατυπικά πολυδιάστατα δεδομένα είναι μεγάλης δυσκολίας πρόβλημα, αυξανόμενης με την διάσταση των δεδομένων. Πολλοί ερευνητές έχουν προσπαθήσει να δώσουν διαφορετικές λύσεις στο πρόβλημα τις εύρωστης εκτιμητικής για μοντέλα πεπερασμένων υπερθέσεων κατανομών [4, 5]. Στην εργασία [6] προτάθηκαν οι πεπερασμένες υπερθέσεις κατανομών Student's- t (SMM) ως μια εξαιρετικά εύρωστη σε ατυπικά δεδομένα εναλλακτική στα GMM. Όπως έχουμε ήδη συζητήσει (δες Κεφ. 1), η κατανομή Student's- t είναι εκθετική κατανομή με μεγαλύτερες ουρές από ότι η κατανομή Gauss. Έτσι, τα SMM προσφέρουν μια πολύ πιο εύρωστη εναλλακτική σε σχέση με τα GMM, καθώς οι μακρύτερες ουρές των συνιστωσών τους κατανομών έχουν σαν αποτέλεσμα παρατηρήσεις εκπαίδευσης ατυπικές μιας συνιστώσας κατανομής να παίρνουν μικρότερο βάρος στην εκτιμητική των παραμέτρων της, υπό μια ενδογενή στο μοντέλο, καλά θεμελιωμένη, στατιστική διαδικασία. Η σημαντική ανθεκτικότητα των SMM σε ατυπικά δεδομένα εκπαίδευσης έχει πειραμα-

τικά επιδειχθεί σε πολλά πρόσφατα άρθρα (π.χ. [6, 7]), όπου έχει καταδειχθεί ότι δύνανται να μοντελοποιούν κρυφά πρότυπα το ίδιο καλά ακόμα και υπό την παρουσία σημαντικών ποσοστών ατυπικών δεδομένων, περιπτώσεις όπου τα GMM είτε αποτυγχάνουν πλήρως είτε απαιτούν την εκπαίδευση πολύ μεγάλων μοντέλων (για να πιάνουν τις ουρές της προσεγγιζόμενης κατανομής), οδηγώντας σε σημαντική μείωση της υπολογιστικής τους αποδοτικότητας.

Φυσική συνεπεία της δυσανεξίας των GMM σε ατυπικά δεδομένα είναι η ανάλογη ευπάθεια να εμφανίζεται και στα GHMM μοντέλα. Στην σύγχρονη βιβλιογραφία των HMM πολλές προσπάθειες έχουν γίνει για την επίλυση αυτών των προβλημάτων των GHMM. Επί παραδείγματι, στο [8], προτείνεται μια μέθοδος επιλεκτικής εκπαίδευσης. Στο [9], ένα άνω-προς-κάτω φίλτρο επιλεκτικής έμφασης εφαρμόζεται στην εκπαίδευση των GHMM υπό το υπόδειγμα της μεγίστης πιθανοφανείας, με εφαρμογή στην εύρωστη αναγνώριση ομιλίας. Στο [10], ένα GHMM μεγίστης εμπιστοσύνης προτείνεται για την εύρωστη αναγνώριση δισδιάστατων προτύπων. Ωστόσο, οι μέθοδοι αυτές έχουν ποικίλα σημαντικά μειονεκτήματα, μεταξύ των οποίων περιλαμβάνονται η ευρετική φύση της πλειονότητάς τους, καθώς και η περιορισμένη σε συγκεκριμένες εφαρμογές διατύπωσή τους. Αντιθέτως, αυτό το οποίο χρειαζόμαστε σήμερα είναι μια μαθηματικώς στέρεα θεώρηση θεμελιώνουσα ένα γενικό μοντέλο για την εύρωστη σε ατυπικά δεδομένα μοντελοποίηση ακολουθιακών δεδομένων με χρήση HMM μοντέλων. Κατά την γνώση μας επί της σημερινής βιβλιογραφίας, μια τέτοια μέθοδος ακόμα λείπει από τον χώρο της αναγνώρισης προτύπων.

Το κίνητρο της εργασίας αυτής συνίσταται στις ως άνω παρατηρήσεις:

Σήμερα, τα HMM συνεχών κατανομών τα προταθέντα για εφαρμογές αναγνώρισης προτύπων δεν παρέχουν μια συστηματική και γενική μεθοδολογία εύρωστης σε ατυπικά δεδομένα εκτιμητικής παραμέτρων. Στην εργασία αυτή προσπαθούμε να επιλύσουμε αυτά τα ζητήματα προτείνοντας ένα καινοτόμο HMM όπου οι κατανομές των κρυφών καταστάσεων του είναι SMM. Το καθεξής προκύπτει, Student's- t hidden Markov model (SHMM) νέο αυτό μοντέλο, επιτρέπει την εκμετάλλευση της ευρωστίας των SMM στα πλαίσια των τεχνικών μοντελοποίησης ακολουθιακών δεδομένων με χρήση HMM, παρέχοντας έτσι ένα εξαιρετικών επιδόσεων HMM μοντέλο για την εύρωστη σε ατυπικά δεδομένα αναπαράσταση και ταξινόμηση ακολουθιακών δεδομένων.

4.2 Διατύπωση Μοντέλου

4.2.1 Μοντέλα Πεπερασμένων Υπερθέσεων Κατανομών Student's- t (SMM)

Ας ξεκινήσουμε με μια σύντομη ανασκόπηση των μοντέλων πεπερασμένων υπερθέσεων κατανομών Student's- t (SMM). Η υιοθέτηση της πολυμετάβλητης κατανομής Student's- t παρέχει μια μέθοδο διεύρυνσης της Κανονικής κατανομής για πιθανά ατυπικά δεδομένα. Η συνάρτηση πυκνότητας πιθανότητας (σ.π.π.) της κατανομής Student's- t με μέση τιμή μ , πεπερασμένα θετικό πίνακα εσωτερικού γινομένου (πολλαπλάσιο του πίνακα συνδιακύ-

μανσης) Σ , και ν βαθμούς ελευθερίας δίνεται από [11]

$$t(\mathbf{y}_t; \boldsymbol{\mu}, \Sigma, \nu) = \frac{\Gamma\left(\frac{\nu+p}{2}\right) |\Sigma|^{-1/2} (\pi\nu)^{-p/2}}{\Gamma(\nu/2) \{1 + d(\mathbf{y}_t, \boldsymbol{\mu}; \Sigma)/\nu\}^{(\nu+p)/2}} \quad (4.1)$$

όπου p είναι η διάσταση των παρατηρήσεων \mathbf{y}_t , $d(\mathbf{y}_t, \boldsymbol{\mu}; \Sigma)$ είναι η τετραγωνική απόσταση Mahalanobis μεταξύ των $\mathbf{y}_t, \boldsymbol{\mu}$ με συνδιακύμανση Σ , και $\Gamma(s)$ είναι η συνάρτηση Γάμμα, $\Gamma(s) = \int_0^\infty e^{-z} z^{s-1} dz$. Επιπλέον, ισχύει ότι [11] η κατανομή μιας τυχαίας μεταβλητής \mathbf{y}_t κατανεμημένης κατά Student's- t με μέσο $\boldsymbol{\mu}$, πίνακα εσωτερικού γινομένου Σ , και ν βαθμούς ελευθερίας

$$\mathbf{y}_t \sim t(\boldsymbol{\mu}, \Sigma, \nu) \quad (4.2)$$

ισοδύναμα μπορεί να γραφεί ως

$$\mathbf{y}_t | u_t \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma/u_t) \quad (4.3)$$

όπου η βαθμίδα συνδιακύμανσης u_t , είναι κατανεμημένη κατά

$$u_t \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \quad (4.4)$$

$\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ είναι η Κανονική κατανομή με μέσο $\boldsymbol{\mu}$ και πίνακα συνδιακύμανσης Σ , και $\mathcal{G}(\alpha, \beta)$ είναι η κατανομή Γάμμα.

Η σ.π.π. ενός SMM με J -το πλήθος συνιστώσες κατανομές και βάρη συνιστωσών κατανομών (πρότερες (prior) πιθανότητες) c_1, \dots, c_J , δίνεται,

κατά συνέπεια, από την [6]

$$p(\mathbf{y}_t; \Theta) = \sum_{j=1}^J c_j t(\mathbf{y}_t; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j, \nu_j) \quad (4.5)$$

όπου το διάνυσμα παραμέτρων της κατανομής Θ περιλαμβάνει τα στοιχεία των $\boldsymbol{\mu}_j$ και $\boldsymbol{\Sigma}_j$, καθώς και τους βαθμούς ελευθερίας ν_j και τις πρότερες πιθανότητες c_j , $j = 1, \dots, J$. Στο κείμενο αυτό χρησιμοποιούμε τον γενικό συμβολισμό $p(\cdot)$ για να αποδώσουμε μια συνάρτηση πιθανότητας.

4.2.2 Το Student's- t hidden Markov model (SHMM)

Ο ορισμός του προτεινομένου Student's- t hidden Markov model (SHMM) εξάγεται δια της θεωρήσεως ενός πεπερασμένου χώρου καταστάσεων μοντέλου κρυπτο-Μαρκοβιανής αλυσίδας και της παραδοχής ότι οι κατανομές εκπομπής παρατηρήσεων των κρυφών του καταστάσεων μοντελοποιούνται με χρήση SMM και άρα με σ.π.π. κατά την (5).

Έστω λοιπόν ένα SHMM περιλαμβάνον I καταστάσεις. Έστω ότι $\mathbf{y}_1, \dots, \mathbf{y}_T$ είναι μια ακολουθία από δεδομένα μοντελοποιούμενα με χρήση του θεωρηθέντος SHMM. Ορίζουμε ως $\mathbf{s}_t = (s_{1t}, \dots, s_{It})$, $t = 1, \dots, T$, τα διανύσματα ενδεικτών καταστάσεως, όπου s_{it} είναι ένα ή μηδέν, ανάλογα με το αν η παρατήρηση \mathbf{y}_t θεωρείται εκπεμφθείσα ή μη από την i -οστή κατάσταση του μοντέλου ($i = 1, \dots, I$). Δεχόμαστε για ευκολία, και χωρίς καμία απώλεια γενικότητας, ότι όλες οι κρυφές καταστάσεις του μοντέλου έχουν κατανομές παρατηρήσεων μοντελοποιούμενες από SMM με τον ίδιο αριθμό συνιστωσών κατανομών, J . Τότε, βάσει της ιδιότητας υπό συνθήκη ανεξαρτησίας της Μαρκοβιανής αλυσίδας [1, 2] συνάγεται αμέσως ότι οι

παρατηρήσεις \mathbf{y}_t οι προερχόμενες από την ίδια κατάσταση ενός SHMM είναι μεταξύ τους ανεξάρτητες και ισόνομες. Συνεπώς, η σ.π.π. τους υπό την συνθήκη $s_{it} = 1$ δίνεται από

$$p(\mathbf{y}_t; \Theta_i) = \sum_{j=1}^J c_{ij} t(\mathbf{y}_t; \boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}, \nu_{ij}) \quad (4.6)$$

όπου c_{ij} , $\boldsymbol{\mu}_{ij}$, $\boldsymbol{\Sigma}_{ij}$ και ν_{ij} είναι αντίστοιχα ο συντελεστής βάρους, ο μέσος, ο πίνακας εσωτερικού γινομένου και οι βαθμοί ελευθερίας της j -οστης συνιστώσας της κατανομής της i -οστης καταστάσεως του μοντέλου ($i = 1, \dots, I$, $j = 1, \dots, J$), και το διάνυσμα παραμέτρων της κατανομής της i -οστης καταστάσεως του μοντέλου, Θ_i , περιλαμβάνει τα c_{ij} , ν_{ij} , και τα στοιχεία των $\boldsymbol{\mu}_{ij}$ και $\boldsymbol{\Sigma}_{ij}$. Από τις (2) - (4), παίρνουμε επίσης ότι, ισοδυνάμως, ισχύει

$$p(\mathbf{y}_t | \{u_{ijt}\}_{j=1}^J; \Theta_i) = \sum_{j=1}^J c_{ij} \mathcal{N}(\mathbf{y}_t; \boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij} / u_{ijt}) \quad (4.7)$$

όπου οι βαθμίδες συνδιακύμανσης, u_{ijt} είναι κατανεμημένες ως

$$u_{ijt} \sim \mathcal{G}\left(\frac{\nu_{ij}}{2}, \frac{\nu_{ij}}{2}\right) \quad (4.8)$$

Τέλος, εκ των ιδιοτήτων της Μαρκοβιανής αλυσίδας [1], συνάγεται ότι η πιθανοφάνεια ενός SHMM μοντέλου με διάνυσμα παραμέτρων Ψ ως

προς την ακολουθία παρατηρήσεων $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T$ δίνεται από

$$L(\Psi) = p(\mathbf{y}_1, \dots, \mathbf{y}_T; \Psi) = \sum_{s\{1\}=1}^I \dots \sum_{s\{T\}=1}^I \pi_{s\{1\}} p(\mathbf{y}_1; \Theta_{s\{1\}}) \quad (4.9)$$

$$\times \pi_{s\{1\}, s\{2\}} p(\mathbf{y}_2; \Theta_{s\{2\}}) \dots \pi_{s\{T-1\}, s\{T\}} p(\mathbf{y}_T; \Theta_{s\{T\}})$$

όπου, η μεταβλητή $s\{t\}$ συμβολίζει τη κατάσταση εκπομπής της t -οστης παρατήρησης, \mathbf{y}_t , π_i είναι οι πιθανότητες αρχικής κατάστασης της Μαρκοβιανής αλυσίδας, π_{hi} είναι οι πιθανότητες μετάβασης ενός βήματος της Μαρκοβιανής αλυσίδας, και το Ψ περιλαμβάνει τις παραμέτρους των κρυφών καταστάσεων Θ_i , καθώς και τις πιθανότητες της Μαρκοβιανής αλυσίδας, π_i και π_{hi} .

4.3 Επίλυση Μεγίστης Πιθανοφανείας του SHMM Μοντέλου με Χρήση Πολλαπλών Ακολουθιών

Η πολλαπλών ακολουθιών επίλυση μέγιστης πιθανοφανείας του SHMM μοντέλου μπορεί να περαιωθεί με χρήση του EM αλγορίθμου [3]. Συνίσταται στον υπολογισμό της εκτιμήτριας μέγιστης πιθανοφανείας $\hat{\Psi}$ του διανύσματος παραμέτρων του μοντέλου, Ψ , δοθέντος ενός συνόλου δεδομένων εκπαίδευσης. Ο EM αλγόριθμος είναι μια ισχυρή, γενική επαναληπτική διαδικασία για την υπολογιστική επίλυση μέγιστης πιθανοφανείας στατιστικών μοντέλων και έχει ευρέως χρησιμοποιηθεί για την εκτιμητική μέγιστης πιθανοφανείας HMM μοντέλων σε πλείστες εφαρμογές [2]. Ο αλγόριθμος EM βασίζεται στην επαναληπτική, διαδοχική βελτιστοποίηση μιας ενδια-

μέσου ποσότητας, της υπό συνθήκη προσδοκητής τιμής της λογαριθμο-πιθανοφανείας πλήρων δεδομένων του υπό εκτίμηση μοντέλου, δοθέντος ενός συνόλου εκπαίδευσης [12].

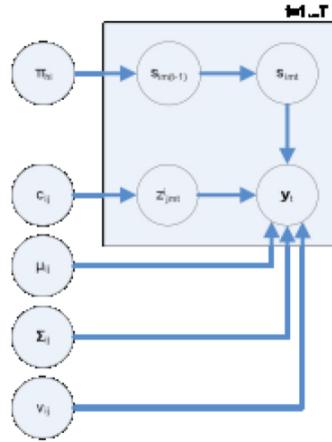
Έστω M ανεξάρτητες ακολουθίες από δεδομένα εκπαίδευσης. Δεχόμαστε για απλότητα ότι όλες οι ακολουθίες έχουν το ίδιο μήκος T , ήτοι περιλαμβάνουν απο T δεδομένα η καθεμία, χωρίς βλάβη της γενικότητας. Έστω ότι η m -οστη ακολουθία είναι $\mathbf{y}_m = \{\mathbf{y}_{mt}\}_{t=1}^T$, $m = 1, \dots, M$, όπου \mathbf{y}_{mt} είναι το t -οστο δεδομένο της m -οστης ακολουθίας. Ορίζουμε ως \mathbf{s}_{mt} τα διανύσματα ενδεικτών καταστάσεως των παρατηρήσεων, όπου $\mathbf{s}_{mt} = (s_{1mt}, \dots, s_{gmt})$ και $s_{imt} = 1$ αν \mathbf{y}_{mt} προέρχεται από την i -οστη κατάσταση του μοντέλου, ειδήλλως $s_{imt} = 0$. Έστω ακόμα \mathbf{z}_{mt}^i οι ενδείκτες συνιστωσών κατανομών των παρατηρήσεων, όπου $\mathbf{z}_{mt}^i = (z_{1mt}^i, \dots, z_{nmt}^i)$, και $z_{jmt}^i = 1$ αν, δοθέντος ότι η παρατήρηση \mathbf{y}_{mt} προήλθε από την i -οστη κατάσταση του μοντέλου, έχουμε ότι ειδικότερα παρήχθη από την j -οστη συνιστώσα της κατανομής της i -οστης κρυφής καταστάσεως του μοντέλου, $z_{jmt}^i = 0$ διαφορετικά. Τότε, από τις (6) - (8) συνάγεται ότι

$$p(\mathbf{y}_{mt}; \Theta_i) = \sum_{j=1}^J c_{ij} t(\mathbf{y}_{mt}; \boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}, \nu_{ij}) \quad (4.10)$$

ή, ισοδυνάμως

$$p(\mathbf{y}_{mt} | \{u_{ijmt}\}_{j=1}^J; \Theta_i) = \sum_{j=1}^J c_{ij} \mathcal{N}(\mathbf{y}_{mt}; \boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}/u_{ijmt}) \quad (4.11)$$

όπου u_{ijmt} είναι η βαθμίδα συνδιακύμανσης της παρατήρησης \mathbf{y}_{mt} δοθέντος ότι παρήχθη από την j -οστη συνιστώσα της κατανομής της i -οστης κα-



Σχήμα 4.1: Γραφική αναπαράσταση του θεωρουμένου SHMM για T παρατηρήσεις, $\{\mathbf{y}_t\}_{t=1}^T$.

ταστάσεως του μοντέλου

$$u_{ijmt} \sim \mathcal{G}\left(\frac{\nu_{ij}}{2}, \frac{\nu_{ij}}{2}\right) \quad (4.12)$$

Μια γραφική παράσταση του θεωρούμενου SHMM δίνεται στο Σχ. 1.

Όπως έχει συζητηθεί στα άρθρα [6,11], μια κλειστού τύπου λύση για την βελτιστοποίηση λογαριθμο-πιθανοφανείας της κατανομής Student's- t στην μορφή (10) δεν υφίσταται. Ωστόσο, εκμεταλλευόμενοι τις (11) - (12), ήτοι θεωρώντες την εναλλακτική έκφραση της κατανομής Student's- t ως μιας Κανονικής κατανομής με επιβαθμισμένη συνδιακύμανση, όπου η βαθμίδα είναι μια κατανεμημένη κατά Γάμμα κρυφή μεταβλητή, ένα υπολογιστικά προσεγγίσιμο υπόδειγμα βελτιστοποίησης μπορεί να εξαχθεί. Υπό αυτές τις παρατηρήσεις, η εφαρμογή του EM αλγορίθμου για την επίλυση μεγίστης πιθανοφανείας του SHMM μοντέλου με χρήση πολλαπλών ακολουθιών είναι δυνατή υπό την θεώρηση των πλήρων δεδομένων των αντιστοιχούντων

στην m -οστη ακολουθία, \mathbf{y}_m^{comp} , ως περιλαμβανόντων

1. τα παρατηρήσιμα δεδομένα \mathbf{y}_{mt} , $t = 1, \dots, T$, $m = 1, \dots, M$
2. τα διανύσματα ενδεικτών καταστάσεων, \mathbf{s}_{mt}
3. τα διανύσματα ενδεικτών συνιστωσών κατανομών, \mathbf{z}_{mt}^i
4. τις βαθμίδες συνδιακύμανσης, u_{ijmt} .

Έστω $\mathbf{y}^{comp} = \{\mathbf{y}_m^{comp}\}_{m=1}^M$ τα πλήρη δεδομένα για αυτή την εφαρμογή του EM αλγορίθμου. Τότε, από την (9) και χρησιμοποιώντας τις (11) - (12), συνάγεται ότι η λογαριθμο-πιθανοφάνεια πλήρων δεδομένων του SHMM μοντέλου γράφεται ως

$$\begin{aligned} \log L_c(\Psi) &= \sum_{m=1}^M \sum_{h=1}^I \left[s_{hm1} \log \pi_h + \sum_{i=1}^I \sum_{t=1}^{T-1} s_{hmt} s_{im,t+1} \log \pi_{hi} \right] \\ &+ \sum_{m=1}^M \sum_{i=1}^I \sum_{t=1}^T s_{imt} \log p(\mathbf{y}_{mt}^{comp}; \Theta_i) \end{aligned} \quad (4.13)$$

όπου \mathbf{y}_{mt}^{comp} είναι τα πλήρη δεδομένα τα αντιστοιχούντα στην t -οστη παρατήρηση της m -οστης ακολουθίας, \mathbf{y}_{mt} , και $\log p(\mathbf{y}_{mt}^{comp}; \Theta_i)$ είναι η λογαριθμο-πιθανοφάνεια πλήρων δεδομένων της κατανομής εκπομπής παρατηρήσεων (SMM) της i -οστης κατάστασης του SHMM μοντέλου, ως προς την παρατήρηση \mathbf{y}_{mt} . Αναφορικά με την τελευταία αυτή ποσότητα, εκ των (11) - (12) έχουμε

$$p(\mathbf{y}_{mt}^{comp}; \Theta_i) = \prod_{j=1}^J [c_{ij} p(\mathbf{y}_{mt} | u_{ijmt}; \Theta_i) p(u_{ijmt}; \Theta_i)]^{z_{jmt}^i}$$

εκ του οποίου συνάγεται (αγνοώντας τις σταθερές)

$$\begin{aligned} \log p(\mathbf{y}_{mt}^{comp}; \Theta_i) = \sum_{j=1}^J z_{jmt}^i \left\{ -\log \Gamma\left(\frac{\nu_{ij}}{2}\right) + \frac{\nu_{ij}}{2} \left[\log\left(\frac{\nu_{ij}}{2}\right) \right. \right. \\ \left. \left. + \log u_{ijmt} - u_{ijmt} \right] - \frac{u_{ijmt}}{2} d(\mathbf{y}_{mt}, \boldsymbol{\mu}_{ij}; \boldsymbol{\Sigma}_{ij}) - \frac{1}{2} \log |\boldsymbol{\Sigma}_{ij}| + \log c_{ij} \right\} \end{aligned} \quad (4.14)$$

Ο EM αλγόριθμος περιλαμβάνει ένα E-βήμα και ένα M-βήμα. Στο E-βήμα της $(k+1)$ -οστης επαναλήψεως του EM αλγορίθμου απαιτείται ο υπολογισμός της ποσότητας

$$Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)}) = E_{\boldsymbol{\Psi}^{(k)}}(\log L_c(\boldsymbol{\Psi}) | \mathbf{y}) \quad (4.15)$$

η οποία είναι η υπό συνθήκη προσδοκητή τιμή της λογαριθμο-πιθανοφάνειας πλήρων δεδομένων του μοντέλου ως προς τα παρατηρήσιμα δεδομένα $\mathbf{y} = \{\mathbf{y}_m\}_{m=1}^M$, όπου $\boldsymbol{\Psi}^{(k)}$ είναι ο τρέχουσα τιμή της εκτιμητορίας (υπολογισμένη στην k -οστη επανάληψη του EM αλγορίθμου) του $\boldsymbol{\Psi}$. Χρησιμοποιώντας την (13), η έκφραση (15) του $Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)})$ δίνει

$$\begin{aligned} Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)}) = \sum_{m=1}^M \sum_{h=1}^I \left[\gamma_{hm1}^{(k)} \log \pi_h + \sum_{i=1}^I \sum_{t=1}^{T-1} \gamma_{himt}^{(k)} \log \pi_{hi} \right] \\ + \sum_{m=1}^M \sum_{i=1}^I \sum_{t=1}^T \gamma_{imt}^{(k)} E_{\boldsymbol{\Psi}^{(k)}}(\log p(\mathbf{y}_{mt}^{comp}; \Theta_i) | \mathbf{y}) \end{aligned} \quad (4.16)$$

όπου $\gamma_{imt}^{(k)}$ είναι οι εκτιμητορίες των υστέρων (posterior) πιθανοτήτων εκπομπής των καταστάσεων του μοντέλου, ως υπολογίζονται στην k -οστη επα-

νάληψη του αλγορίθμου, οριζόμενες ως

$$\gamma_{imt} \triangleq p(s_{imt} = 1 | \mathbf{y}) = p(s_{imt} = 1 | \mathbf{y}_m) \quad (4.17)$$

($t = 1, \dots, T$) και $\gamma_{himt}^{(k)}$ είναι οι εκτιμήτριες των υστέρων (posterior) πιθανοτήτων μετάβασης ενός βήματος της Μαρκοβιανής αλυσίδας, ως υπολογίζονται στην k -οστη επανάληψη του αλγορίθμου, οριζόμενες ως

$$\gamma_{himt} \triangleq p(s_{im,t+1} = 1, s_{hmt} = 1 | \mathbf{y}) \quad (4.18)$$

($t = 1, \dots, T - 1$) και $m = 1, \dots, M, h, i = 1, \dots, I$.

Ας ξεκινήσουμε με τις εκφράσεις των υστέρων πιθανοτήτων γ_{imt} και γ_{himt} . Οι ενημερώσεις των ποσοτήτων αυτών στην $(k+1)$ -οστη επανάληψη του EM αλγορίθμου υπολογίζονται με χρήση του εμπροσθεν-όπισθεν (forward-backward) αλγορίθμου. Έχουμε [1,2]

$$\gamma_{himt}^{(k)} = \frac{a_{hmt}^{(k)} \pi_{hi}^{(k)} p(\mathbf{y}_{m,t+1}; \Theta_i^{(k)}) b_{im,t+1}^{(k)}}{\sum_{v=1}^I \sum_{\phi=1}^I a_{vmt}^{(k)} \pi_{v\phi}^{(k)} p(\mathbf{y}_{m,t+1}; \Theta_\phi^{(k)}) b_{\phi m,t+1}^{(k)}} \quad (4.19)$$

($t = 1, \dots, T - 1$) και

$$\gamma_{imt}^{(k)} = \frac{a_{imt}^{(k)} b_{imt}^{(k)}}{\sum_{h=1}^I a_{hmt}^{(k)} b_{hmt}^{(k)}} \quad (t = 1, \dots, T) \quad (4.20)$$

όπου

$$a_{im1}^{(k)} = \pi_i^{(k)} p(\mathbf{y}_{m1}; \Theta_i^{(k)}) \quad (4.21)$$

$$a_{im,t+1}^{(k)} = p(\mathbf{y}_{m,t+1}; \Theta_i^{(k)}) \sum_{h=1}^I a_{hmt}^{(k)} \pi_{hi}^{(k)} \quad (t = 1, \dots, T - 1) \quad (4.22)$$

$$b_{hmT}^{(k)} = 1 \quad (4.23)$$

$$b_{hmt}^{(k)} = \sum_{i=1}^I \pi_{hi}^{(k)} p(\mathbf{y}_{m,t+1}; \Theta_i^{(k)}) b_{im,t+1} \quad (t = T - 1, \dots, 1) \quad (4.24)$$

και η έκφραση του $p(\mathbf{y}_{mt}; \Theta_i)$ δίνεται από την (10).

Τέλος, σχετικά με την έκφραση του όρου $E_{\Psi^{(k)}}(\log p(\mathbf{y}_{mt}^{comp}; \Theta_i) | \mathbf{y})$, από την (14) μπορεί να δειχθεί (δες Παράρτημα Α) ότι η εξαγωγή αυτής της έκφρασης ανάγεται στον υπολογισμό των υπό συνθήκη πιθανοτήτων των \mathbf{y}_{mt} να παράγονται από την j -οστη συνιστώσα κατανομή της i -οστης κατάστασης του SHMM μοντέλου, δοθείσης της εκπομπής τους από την i -οστη κατάσταση του μοντέλου

$$\begin{aligned} \xi_{ijmt}^{(k)} &\triangleq E_{\Psi^{(k)}}(z_{jmt}^i | \mathbf{y}_{mt}, s_{imt} = 1) \\ &= \frac{c_{ij}^{(k)} t(\mathbf{y}_{mt}; \boldsymbol{\mu}_{ij}^{(k)}, \boldsymbol{\Sigma}_{ij}^{(k)}, \nu_{ij}^{(k)})}{\sum_{h=1}^J c_{ih}^{(k)} t(\mathbf{y}_{mt}; \boldsymbol{\mu}_{ih}^{(k)}, \boldsymbol{\Sigma}_{ih}^{(k)}, \nu_{ih}^{(k)})} \end{aligned} \quad (4.25)$$

καθώς και τον υπολογισμό των υστέρων προσδοκητών τιμών των βαθμικών συνδιακύμανσης, u_{ijmt} , των παρατηρήσεων

$$\begin{aligned} u_{ijmt}^{(k)} &\triangleq E_{\Psi^{(k)}}(u_{ijmt} | \mathbf{y}_{mt}) \\ &= \frac{\nu_{ij}^{(k)} + p}{\nu_{ij}^{(k)} + d(\mathbf{y}_{mt}, \boldsymbol{\mu}_{ij}^{(k)}; \boldsymbol{\Sigma}_{ij}^{(k)})} \end{aligned} \quad (4.26)$$

Περαιτέρω, το Μ-βήμα της επίλυσης μεγίστης πιθανοφαινείας του SHMM μοντέλου με χρήση πολλαπλών ακολουθιών υπό τον EM αλγόριθμο, συνί-

σταται στην εκτέλεση των υπολογισμών (ενημερώσεις παραμέτρων)

$$\pi_i^{(k+1)} = \frac{1}{M} \sum_{m=1}^M \gamma_{im1}^{(k)} \quad (4.27)$$

$$\pi_{hi}^{(k+1)} = \frac{\sum_{m=1}^M \sum_{t=1}^{T-1} \gamma_{himt}^{(k)}}{\sum_{m=1}^M \sum_{t=1}^{T-1} \gamma_{hmt}^{(k)}} \quad (4.28)$$

$$c_{ij}^{(k+1)} = \frac{\sum_{m=1}^M \sum_{t=1}^T r_{ijmt}^{(k)}}{\sum_{m=1}^M \sum_{t=1}^T \gamma_{ijmt}^{(k)}} \quad (4.29)$$

$$\boldsymbol{\mu}_{ij}^{(k+1)} = \frac{\sum_{m=1}^M \sum_{t=1}^T r_{ijmt}^{(k)} u_{ijmt}^{(k)} \mathbf{y}_{mt}}{\sum_{m=1}^M \sum_{t=1}^T r_{ijmt}^{(k)} u_{ijmt}^{(k)}} \quad (4.30)$$

$$\begin{aligned} \boldsymbol{\Sigma}_{ij}^{(k+1)} &= \sum_{m=1}^M \sum_{t=1}^T r_{ijmt}^{(k)} u_{ijmt}^{(k)} (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij}^{(k+1)}) (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij}^{(k+1)})^T \\ &\times \left[\sum_{m=1}^M \sum_{t=1}^T r_{ijmt}^{(k)} \right]^{-1} \end{aligned} \quad (4.31)$$

ενώ, οι βαθμοί ελευθερίας, ν_{ij} , δίνονται από την αριθμητική επίλυση της μη γραμμικής εξίσωσης

$$\begin{aligned} 1 - \psi\left(\frac{\nu_{ij}}{2}\right) + \log\left(\frac{\nu_{ij}}{2}\right) + \psi\left(\frac{\nu_{ij}^{(k)} + p}{2}\right) - \log\left(\frac{\nu_{ij}^{(k)} + p}{2}\right) \\ + \frac{1}{\sum_{m=1}^M \sum_{t=1}^T r_{ijmt}^{(k)}} \sum_{m=1}^M \sum_{t=1}^T r_{ijmt}^{(k)} \left(\log u_{ijmt}^{(k)} - u_{ijmt}^{(k)} \right) = 0 \end{aligned} \quad (4.32)$$

όπου, $\psi(s)$ είναι η συνάρτηση Δίγαμμα και $r_{ijmt}^{(k)}$ είναι η ύστερη πιθανότητα η παρατήρηση \mathbf{y}_{mt} να παράγεται από την i -οστη κατάσταση του μοντέλου

Πίνακας 4.1: EM Αλγόριθμος για το SHMM

$k := 0$

1. Εκτέλεσε των forward-backward algorithm για να βρεις τις ποσότητες $a_{imt}^{(k)}$ and $b_{imt}^{(k)}$.
2. Εκτέλεσε το E-step υπολογίζοντας τις $\gamma_{hit}^{(k)}$, $\gamma_{imt}^{(k)}$, $\xi_{ijmt}^{(k)}$, $r_{ijmt}^{(k)}$, και $u_{ijmt}^{(k)}$, βάσει των (19), (20), (25), (34), και (26), αντιστοίχως.
3. Εκτέλεσε το M-step υπολογίζοντας τις $\pi_i^{(k+1)}$, $\pi_{hi}^{(k+1)}$, $c_{ij}^{(k+1)}$, $\mu_{ij}^{(k+1)}$, $\Sigma_{ij}^{(k+1)}$, και $\nu_{ij}^{(k+1)}$, βάσει των (27)-(32), αντιστοίχως.
4. Αν ο EM αλγόριθμος συγκλίνει, **έξελθε**; ειδάλλως, θέσε ($k := k + 1$) και επέστρεψε στο 1.

και ειδικότερα από την j -οστη συνιστώσα κατανομή αυτής, ήτοι έχουμε

$$r_{ijmt} \triangleq p(s_{imt} = 1, z_{jmt}^i = 1 | \mathbf{y}) = \gamma_{imt} \xi_{ijmt} \quad (4.33)$$

$$r_{ijmt}^{(k)} = \gamma_{imt}^{(k)} \xi_{ijmt}^{(k)} \quad (4.34)$$

(για περαιτέρω επί των ανωτέρω αποτελεσμάτων, δες Παράρτημα Α). Ένα περίγραμμα του εξαχθέντος αλγορίθμου δίνεται στον Πίνακα 1.

Αναφορικά με τα υπολογιστικά κόστη του EM αλγορίθμου για το SHMM, σημειώνουμε ότι, κατ' ουσίαν, δυο είναι οι επιπλέον υπολογισμοί του EM αλγορίθμου για το SHMM σε σχέση με τον EM αλγόριθμο για το GHMM: (α) η εύρεση των scalars u_{ijmt}^k και (β) η εκτίμηση των βαθμών ελευθερίας ν_{ij}^k . Σχετικά με την εκτιμητική των scalars, επισημαίνουμε ότι το μεγαλύτερο ποσοστό των υπολογιστικών τους δαπανών οφείλεται στον υπολογισμό των Mahalanobis distances, που, όμως, είναι γνωστό ότι θα υπολογίζονταν

ούτως ή άλλως για την εύρεση της σ.π.π. των SHMM αλλά και των GHMM. Συνεπώς, αυτός ο υπολογισμός επάγει έναν αμελητέο υπολογιστικό φόρτο. Αναφορικά τέλος με την εκτίμηση των βαθμών ελευθερίας, μπορούμε να υποστηρίξουμε ότι η διαδικασία αυτή δεν επάγει καθοριστικά επιβαρυντικό υπολογιστικό φόρτο. Στην πειραματική ενότητα θα καταδείξουμε την αλήθεια αυτού του ισχυρισμού μέσα από ένα πλήθος εφαρμογών.

Ένα εξίσου σημαντικό θέμα αφορά στην εισαγωγή μιας υπολογιστικά αποδοτικής λύσης στο πρόβλημα του υπολογισμού της πιθανοφανείας $p(\mathbf{y}'_1, \dots, \mathbf{y}'_T; \hat{\Psi})$ ενός εκπαιδευμένου SHMM ως προς την ακολουθία δεδομένων $\mathbf{y}' = \{\mathbf{y}'_t\}_{t=1}^T$, καθώς η (9) είναι υπολογιστικά μη αποδοτική. Ένα τέτοιο αποτέλεσμα, ισχύον για κάθε τύπου HMM πεπερασμένου χώρου καταστάσεων, ανεξαρτήτως κατανομών παρατηρήσεων, μπορεί να υπολογισθεί με την χρήση του έμπροσθεν-όπισθεν αλγορίθμου. Συγκεκριμένα, ισχύει [1,2]

$$p(\mathbf{y}'_1, \dots, \mathbf{y}'_T; \hat{\Psi}) = \sum_{i=1}^I \hat{a}_{iT} \quad (4.35)$$

όπου, τα \hat{a}_{iT} υπολογίζονται με χρήση του έμπροσθεν-όπισθεν αλγορίθμου για την ακολουθία \mathbf{y}' , με χρήση της ήδη εξαχθείσης εκτιμητορίας παραμέτρων, $\hat{\Psi}$.

Τέλος, ένα σημαντικό πρόβλημα είναι και η εύρεση της ακολουθίας καταστάσεων ενός SHMM η οποία παρήγαγε μια παρατηρήσιμη ακολουθία. Επισημαίνουμε ότι και αυτό το πρόβλημα, όπως και το πρόβλημα του υπολογισμού πιθανοφανείας, έχει λύση (αλγόριθμος Viterbi) που δεν εξαρτάται από την επιλογή της μορφής της κατανομής εκπομπής παρατηρήσεων και, άρα, εφαρμόζεται και στο SHMM κατά ταυτόσημο τρόπο και με την ίδια

υπολογιστική αποδοτικότητα όπως και για τα μοντέλα GHMM.

4.4 Μια Τεχνική Μοντελοποίησης Συνδιακύμανσης για Χειραφέτηση του SHMM στην Περίπτωση Πολυδιάστατων Δεδομένων

Μια σημαντική πρακτική πρόνοια κατά την χρήση πιθανοτικών παραγωγικών μοντέλων, όπως τα μοντέλα GMM και SMM, ως τις κατανομές των κρυφών καταστάσεων HMM μοντέλων, αφορά στον αριθμό των παραμέτρων του υπό εκτίμηση μοντέλου. Ο αριθμός των παραμέτρων των συνιστωσών κατανομών ενός τέτοιου μοντέλου δεν πρέπει να είναι εξαιρετικά υψηλός, ώστε να επιτρέπει την γρήγορη και εύρωστη εκτιμητική των παραμέτρων με χρήση του EM αλγορίθμου. Χρήση πλήρων πινάκων συνδιακύμανσης συνεπάγεται δυνητικά δραματική αύξηση του πλήθους των παραμέτρων του μοντέλου, ειδικά σε περιπτώσεις όπου τα μοντελοποιούμενα δεδομένα είναι εξαιρετικά πολυδιάστατα. Το γεγονός αυτό, σε συνδυασμό με την απαίτηση συγκριτικά μεγάλων δειγμάτων εκπαίδευσης ώστε να γίνει τελικά εφικτή η αξιόπιστη εκτιμητική του HMM μοντέλου, καθώς αυξάνει το πλήθος των παραμέτρων του, υπαγορεύουν την ανάπτυξη τεχνικών μείωσης των παραμέτρων των HMM μοντέλων, ειδικά σε περιπτώσεις όπου τα παρατηρήσιμα δεδομένα έχουν μεγάλο αριθμό διαστάσεων.

Για την επίλυση αυτού του προβλήματος, πολλοί ερευνητές έχουν προτείνει μεθόδους συνδυασμού παραμέτρων, ήτοι οι ίδιοι πίνακες συνδιακύμανσης μοιράζονται από τις συνιστώσες κατανομές πολλές καταστάσε-

ων του μοντέλου [13]. Ωστόσο, αυτή η μέθοδος έχει πολλά μειονεκτήματα, το σημαντικότερο εκ των οποίων είναι η επιφερόμενη πολυπλοκότητα στην διαδικασία εκπαίδευσης και η επιδεξιότητα που απαιτείται για τον σχεδιασμό του HMM. Μια άλλη κοινότυπη λύση με μεγάλη δημοφιλία μεταξύ των ερευνητών είναι η χρήση διαγωνίων πινάκων συνδιακύμανσης. Οι διαγώνιοι πίνακες συνδιακύμανσης σε GMM μοντέλα έχουν την δυνατότητα να προσεγγίζουν επαρκώς τις πραγματικές συνδιακυμάνσεις των διανυσμάτων παρατήρησης (πολυμεταβλήτων δεδομένων). [14]. Θα ήταν, ωστόσο, μεγάλο πλεονέκτημα να έχουμε ασυσχέτιστα διανύσματα παρατηρήσεων όταν χρησιμοποιούμε διαγωνίους πίνακες συνδιακύμανσης [15].

Η ανάλυση βασικών συνθετημάτων (Principal component analysis, PCA) [16] είναι μια αρκετά προφανής λύση σε αυτό το πρόβλημα, και έχει χρησιμοποιηθεί από πολλούς ερευνητές, έχει όμως δυο βασικά προβλήματα: (α) δεν ορίζει μια κατάλληλη σ.π.π. εκτός του υποχώρου των βασικών συνθετημάτων, (β) οι διακυμάνσεις εκτός αυτού του υποχώρου μοντελοποιούνται ομοιόμορφα, ακόμα και όταν τα δεδομένα δεν εγγυώνται τέτοια υπόθεση. Συνεπώς, εναλλακτικές λύσεις πρέπει να αναζητηθούν. Όπως έχουμε αναπτύξει και σε προηγούμενα κεφάλαια, μια καλά θεμελιωμένη τεχνική, επιλύουσα τα προβλήματα του PCA [16] και προσφέροντας τις απαιτούμενες λειτουργικότητες, ως περιεγράψαμε ανωτέρω, είναι η ανάλυση παραγόντων (factor analysis) [17]. Όπως έχουμε συζητήσει, το μοντέλο υπέρθεσης αναλυτών παραγόντων (mixture of factor analyzers (MFA) model) προσφέρει την δυνατότητα επαρκώς καλής μοντελοποίησης της κατανομής τυχαίων πολυμεταβλήτων παρατηρήσεων, διατηρώντας την επίδοση αναγνώρισης προτύπων των GMM μοντέλων, εξασφαλίζοντας ωστόσο ταυτόχρονα ση-

μαντική μείωση των εκτιμητέων παραμέτρων του μοντέλου [17].

4.4.1 Πεπερασμένες Υπερθέσεις Student's- t Αναλυτών Παραγόντων

Παρακάτω παρέχουμε μια σύντομη ανασκόπηση των βασικών εξισώσεων ενός J -συνιστωσών mixture of Student's- t factor analyzers (MSFA) model. Θεωρούμε p -διάστατες παρατηρήσεις $\{\mathbf{y}_t\}_{t=1}^T$. Έστω ότι τα διανύσματα των q -διάστατων παραγόντων είναι τα \mathbf{x}_{jt} , \mathbf{B}_j οι πίνακες των factor loadings, και \mathbf{D}_j οι διαγώνιοι πίνακες συνδιακύμανσης των σφαλμάτων. Τότε, υπό την συνθήκη των \mathbf{x}_{jt} , η κατανομή του \mathbf{y}_t γράφεται [18]

$$p(\mathbf{y}_t | \{\mathbf{x}_{jt}\}_{j=1}^J; \Theta) = \sum_{j=1}^J c_j t(\mathbf{y}_t; \boldsymbol{\mu}_j + \mathbf{B}_j \mathbf{x}_{jt}, \mathbf{D}_j, \nu_j) \quad (4.36)$$

ή, εναλλακτικά

$$p(\mathbf{y}_t | \{\mathbf{x}_{jt}, u_{jt}\}_{j=1}^J; \Theta) = \sum_{j=1}^J c_j \mathcal{N}(\mathbf{y}_t; \boldsymbol{\mu}_j + \mathbf{B}_j \mathbf{x}_{jt}, \mathbf{D}_j / u_{jt}) \quad (4.37)$$

Απαλείφοντας δια ολοκληρώσεως το \mathbf{x}_{jt} από τις (36) - (37), μπορεί ναδειχθεί [7, 18] ότι, άνευ συνθηκών, η κατανομή κάθε παρατήρησης, \mathbf{y}_t , είναι μια J -το πλήθος συνιστωσών υπέρθεση Student's- t κατανομών, σε αναλογίες c_1, \dots, c_J , της μορφής

$$p(\mathbf{y}_t; \Theta) = \sum_{j=1}^J c_j t(\mathbf{y}_t; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j, \nu_j) \quad (4.38)$$

ή, εναλλακτικά

$$p(\mathbf{y}_t | \{u_{jt}\}_{j=1}^J; \Theta) = \sum_{j=1}^J c_j \mathcal{N}(\mathbf{y}_t; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j / u_{jt}) \quad (4.39)$$

όπου

$$\boldsymbol{\Sigma}_j = \mathbf{B}_j \mathbf{B}_j^T + \mathbf{D}_j \quad (4.40)$$

και το διάνυσμα παραμέτρων Θ περιλαμβάνει τις συνιστώσες των $\boldsymbol{\mu}_j$, \mathbf{B}_j , και \mathbf{D}_j καθώς και τα ν_j και c_j ($j = 1, \dots, J$).

Ως έχει πειραματικώς καταδειχθεί [7,18], η υψηλή ανεκτικότητα της κατανομής Student's- t σε ατυπικά δεδομένα έχει σαν αποτέλεσμα την εξόχως ανώτερη επίδοση των πεπερασμένων υπερθέσεων Student's- t αναλυτών παραγόντων σε εφαρμογές αναγνώρισης προτύπων εν συγκρίσει με το συμβατικό (Gaussian) MFA.

4.4.2 Κρυφού Υποχώρου Μοντελοποίηση Συνδιακύμανσης για το SHMM Μοντέλο με Χρήση Ανάλυσης Παραγόντων

Από την προηγηθείσα συζήτηση, είναι εύκολα αντιληπτό ότι προκειμένου να επιτύχομε την χειραφέτηση της υπολογιστικής αποδοτικότητας του προταθέντος, SHMM μοντέλου, για την περίπτωση εξαιρετικά πολυδιαστάτων δεδομένων, χωρίς επίδραση επί της δεινότητας του στην αναγνώριση προτύπων, μια αναλυμένη κατά παράγοντες (factor analyzed) μορφή του SHMM θα μπορούσε να υιοθετηθεί.

Ορίζουμε ως factor analyzed SHMM ένα HMM μοντέλο οι κατανομές

των κρυφών καταστάσεων του οποίου μοντελοποιούνται με χρήση πεπερασμένων υπερθέσεων Student's- t αναλυτών παραγόντων, τοιουτοτρόπως αποδίδοντας ένα SHMM μοντέλο με ανελυμένους κατά παράγοντες (factor analyzed) πίνακες συνδιακύμανσης.

Ας θεωρήσουμε ένα σύνολο δεδομένων εκπαίδευσης, \mathbf{y} , περιλαμβάνον M ανεξάρτητες ακολουθίες, \mathbf{y}_m , και ας υποθέσουμε ότι όλες οι ακολουθίες έχουν το ίδιο μήκος T , δηλ. $\mathbf{y} = \{\mathbf{y}_m\}_{m=1}^M$, $\mathbf{y}_m = \{\mathbf{y}_{mt}\}_{t=1}^T$. Τότε, από την (38), συνάγεται ότι η έκφραση της πιθανότητας της παρατήρησης \mathbf{y}_{mt} υπό την συνθήκη $s_{it} = 1$ είναι ταυτόσημη με την αντίστοιχη έκφραση για το απλό SHMM μοντέλο, ήτοι, δίνεται από την (10):

$$p(\mathbf{y}_{mt}; \Theta_i) = \sum_{j=1}^J c_{ij} t(\mathbf{y}_{mt}; \boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}, \nu_{ij}) \quad (4.41)$$

ή, εναλλακτικά, βάσει της (39), έχουμε

$$p(\mathbf{y}_{mt} | \{u_{ijmt}\}_{j=1}^J; \Theta_i) = \sum_{j=1}^J c_{ij} \mathcal{N}(\mathbf{y}_{mt}; \boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij} / u_{ijmt}) \quad (4.42)$$

όπου, τώρα, οι ανελυμένοι κατά παράγοντες πίνακες εσωτερικού γινομένου δίνονται από

$$\boldsymbol{\Sigma}_{ij} = \mathbf{B}_{ij} \mathbf{B}_{ij}^T + \mathbf{D}_{ij} \quad (4.43)$$

ενώ η κατανομή των βαθμδών συνδιακύμανσης u_{ijmt} δίνεται από την (12). Εδώ, τα Θ_i περιέχουν τα c_{ij} , και ν_{ij} , καθώς και τα στοιχεία των $\boldsymbol{\mu}_{ij}$, \mathbf{B}_{ij} και \mathbf{D}_{ij} . Επιπλέον, υπό την συνθήκη των διανυσμάτων παραγόντων, \mathbf{x}_{ijmt} , η (36) δίνει

$$p(\mathbf{y}_{mt} | \{\mathbf{x}_{ijmt}\}_{j=1}^J; \Theta_i) = \sum_{j=1}^J c_{ij} t(\mathbf{y}_{mt}; \boldsymbol{\mu}_{ij} + \mathbf{B}_{ij} \mathbf{x}_{ijmt}, \mathbf{D}_{ij}, \nu_{ij}) \quad (4.44)$$

ή, εναλλακτικά

$$p(\mathbf{y}_{mt} | \{\mathbf{x}_{ijmt}, u_{ijmt}\}_{j=1}^J; \Theta_i) = \sum_{j=1}^J c_{ij} \mathcal{N}(\mathbf{y}_{mt}; \boldsymbol{\mu}_{ij} + \mathbf{B}_{ij} \mathbf{x}_{ijmt}, \mathbf{D}_{ij}/u_{ijmt}) \quad (4.45)$$

όπου το \mathbf{x}_{ijmt} είναι το διάνυσμα παραγόντων το αντιστοιχόν στην παρατήρηση \mathbf{y}_{mt} δοθέντος ότι προέρχεται από την j -οστή συνιστώσα κατανομή της i -οστής καταστάσεως του μοντέλου, και ισχύει

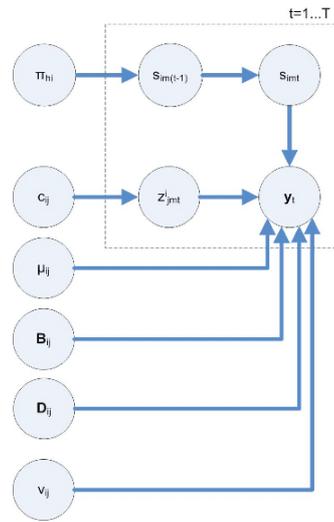
$$\mathbf{x}_{ijmt} \sim t(\mathbf{0}, \mathbf{I}_q, \nu_{ij}) \quad (4.46)$$

και

$$\mathbf{x}_{ijmt} | u_{ijmt} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_q/u_{ijmt}) \quad (4.47)$$

Μια γραφική αναπαράσταση του factor analyzed SHMM δίνεται στο Σχ. 2.

Προκειμένου να εξάγουμε μια πολλαπλών ακολουθιών επίλυση του factor analyzed SHMM, επιστρατεύουμε τον alternating expectation-conditional maximization algorithm (AECM) [12]. Κάθε επανάληψη του AECM αλγορίθμου διαιρείται σε δυο κύκλους, και σε κάθε κύκλο θεωρείται μια διαφορετική διατύπωση των πλήρων δεδομένων του μοντέλου, επιτρέπουν μια σημαντική ενίσχυση της υπολογιστικής αποδοτικότητας του αλγορίθμου εκτιμητικής μεγίστης πιθανοφανείας σε σχέση με τον τετριμμένο EM αλγόριθμο.



Σχήμα 4.2: Γραφική αναπαράσταση του factor analyzed SHMM για μια ακολουθία T παρατηρήσεων, $\{\mathbf{y}_t\}_{t=1}^T$.

Πρώτος Κύκλος

Στον πρώτο κύκλο του αλγορίθμου, εξάγουμε τις εκτιμήτριες των πιθανοτήτων της Μαρκοβιανής αλυσίδας, π_i και π_{hi} , ως και τις εκτιμήτριες των βαρών υπέρθεσης, c_{ij} , των μέσων, μ_{ij} , και των βαθμών ελευθερίας, ν_{ij} , του μοντέλου. Για τον σκοπό αυτό, δεχόμαστε ότι τα πλήρη δεδομένα περιλαμβάνουν

1. τα παρατηρήσιμα δεδομένα \mathbf{y}_{mt} , $t = 1, \dots, T$, $m = 1, \dots, M$
2. τα διανύσματα ενδεικτών καταστάσεων, \mathbf{s}_{mt}
3. τα διανύσματα ενδεικτών συνιστωσών κατανομών, \mathbf{z}_{mt}^i
4. τις βαθμίδες συνδιακύμανσης, u_{ijmt} .

Έτσι, η απορρέουσα έκφραση της λογαριθμο-πιθανοφανείας πλήρων δεδομένων του factor analyzed SHMM είναι ταυτόσημη με την ανάλογη έκφραση για το απλό SHMM, και δίνεται από τις (13) και (14). Συνεπώς, ταυτόσημα με τον EM αλγόριθμο για το απλό SHMM, το E-βήμα του πρώτου κύκλου του AECM αλγορίθμου για το factor analyzed SHMM συνίσταται απλά στην εφαρμογή των (19)-(26), προκειμένου να υπολογισθούν οι ύστερες πιθανότητες της Μαρκοβιανής αλυσίδας γ_{imt} και γ_{himt} , οι υπό την συνθήκη των εκπεμπουσών καταστάσεων ύστερες πιθανότητες των συνιστωσών κατανομών, ξ_{ijmt} , και οι ύστερες προσδοκητές τιμές των βαθμίδων συνδιακύμανσης u_{ijmt} , ενώ, το M-βήμα περαιώνεται με εκτέλεση των υπολογισμών που υπαγορεύονται από τις (27)-(30) και (32).

Δεύτερος Κύκλος

Στο δεύτερο κύκλο, εξάγονται οι εκτιμήτριες των παραμέτρων B_{ij} και D_{ij} του μοντέλου. Για τον σκοπό αυτό, και βάσει των (44)-(47), μια βολική επιλογή των πλήρων δεδομένων είναι η περιλαμβάνουσα

1. τα παρατηρήσιμα δεδομένα \mathbf{y}_{mt} , $t = 1, \dots, T$, $m = 1, \dots, M$
2. τα διανύσματα ενδεικτών καταστάσεων, \mathbf{s}_{mt}
3. τα διανύσματα ενδείκτών συνιστωσών κατανομών, \mathbf{z}_{mt}^i
4. τις βαθμίδες συνδιακύμανσης, u_{ijmt} .
5. τα διανύσματα παραγόντων, \mathbf{x}_{ijmt} .

Τότε, από την (9) συνάγεται ότι, λογαριθμο-πιθανοφάνεια πλήρων δεδομένων του factor analyzed SHMM δίνεται από την

$$\begin{aligned} \log L_c(\Psi) = & \sum_{m=1}^M \sum_{h=1}^I \left[s_{hm1} \log \pi_h + \sum_{i=1}^I \sum_{t=1}^{T-1} s_{hmt} s_{im,t+1} \log \pi_{hi} \right] \\ & + \sum_{m=1}^M \sum_{i=1}^I \sum_{t=1}^T s_{imt} \log p(\mathbf{y}_{mt}^{comp}; \Theta_i) \end{aligned} \quad (4.48)$$

όπου το \mathbf{y}_{mt}^{comp} συμβολίζει τα πλήρη δεδομένα τα αντιστοιχούντα στην παρατήρηση \mathbf{y}_{mt} , ενώ, αναφορικά με την λογαριθμο-πιθανοφάνεια πλήρων δεδομένων της κρυφής κατανομής εκπομπής της i -οστής κατάστασης του μοντέλου εν σχέση με την παρατήρηση \mathbf{y}_{mt} , $\log p(\mathbf{y}_{mt}^{comp}; \Theta_i)$, από τις (45), (47), και (12), και αγνοώντας σταθερές, προκύπτει

$$\begin{aligned} p(\mathbf{y}_{mt}^{comp}; \Theta_i) & \triangleq \prod_{j=1}^J [c_{ij} p(\mathbf{y}_{mt} | \mathbf{x}_{ijmt}, u_{ijmt}; \Theta_i) p(\mathbf{x}_{ijmt} | u_{ijmt}) p(u_{ijmt}; \Theta_i)]^{z_{jmt}^i} \\ \Rightarrow \log p(\mathbf{y}_{mt}^{comp}; \Theta_i) & = \sum_{j=1}^J z_{jmt}^i \left[-\log \Gamma \left(\frac{\nu_{ij}}{2} \right) + \frac{\nu_{ij}}{2} \log \left(\frac{\nu_{ij}}{2} \right) + \frac{\nu_{ij}}{2} (\log u_{ijmt} - u_{ijmt}) \right. \\ & \quad - \frac{u_{ijmt}}{2} (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij} - \mathbf{B}_{ij} \mathbf{x}_{ijmt})^T \mathbf{D}_{ij}^{-1} (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij} - \mathbf{B}_{ij} \mathbf{x}_{ijmt}) \\ & \quad \left. - \frac{1}{2} \log |\mathbf{D}_{ij}| + \log c_{ij} \right] \end{aligned} \quad (4.49)$$

Έτσι, βάσει της (48), η (15) δίνει

$$\begin{aligned}
Q(\Psi; \Psi^{(k)}) &= \sum_{m=1}^M \sum_{h=1}^I \left[\gamma_{hm1}^{(k)} \log \pi_h + \sum_{i=1}^I \sum_{t=1}^{T-1} \gamma_{himt}^{(k)} \log \pi_{hi} \right] \\
&+ \sum_{m=1}^M \sum_{i=1}^I \sum_{t=1}^T \gamma_{imt}^{(k)} E_{\Psi^{(k)}} (\log p(\mathbf{y}_{mt}^{comp}; \Theta_i) | \mathbf{y})
\end{aligned} \tag{4.50}$$

Ας ξεκινήσουμε με το E-βήμα αυτού του κύκλου του αλγορίθμου. Αναφορικά με τις ύστερες πιθανότητες της Μαρκοβιανής αλυσίδας, γ_{imt} και γ_{himt} , η εκτιμητική τους μπορεί εύκολα να περαιωθεί με χρήση του forward-backward αλγορίθμου [1], βάσει των (19) - (24). Επιπλέον, η (50) υπονοεί ότι προκειμένου να ολοκληρωθεί το E-βήμα του δευτέρου κύκλου του AECM αλγορίθμου πρέπει επίσης να υπολογισθεί ο όρος $E_{\Psi^{(k)}} (\log p(\mathbf{y}_{mt}^{comp}; \Theta_i) | \mathbf{y})$. Από την (49), μπορεί να αποδειχθεί (δες Παράρτημα Β) ότι αυτό ανάγεται στον υπολογισμό των υπό την συνθήκη των καταστάσεων εκπομπής υστέρων πιθανοτήτων των συνιστωσών κατανομών

$$\xi_{ijmt}^{(k)} = \frac{c_{ij}^{(k)} t(\mathbf{y}_{mt}; \boldsymbol{\mu}_{ij}^{(k)}, \boldsymbol{\Sigma}_{ij}^{(k)}, \nu_{ij}^{(k)})}{p(\mathbf{y}_{mt}; \Theta_i^{(k)})} \tag{4.51}$$

των υστέρων προσδοκητών τιμών των βαθμιδών συνδιακύμανσης u_{ijmt}

$$u_{ijmt}^{(k)} = \frac{\nu_{ij}^{(k)} + p}{\nu_{ij}^{(k)} + d(\mathbf{y}_{mt}, \boldsymbol{\mu}_{ij}^{(k)}; \boldsymbol{\Sigma}_{ij}^{(k)})} \tag{4.52}$$

καθώς και των εκ της ανάλυσης παραγόντων απορρεουσών ποσοτήτων

$$\mathbf{v}_{ij}^{(k)} = (\mathbf{B}_{ij}^{(k)} \mathbf{B}_{ij}^{(k)T} + \mathbf{D}_{ij}^{(k)})^{-1} \mathbf{B}_{ij}^{(k)} \tag{4.53}$$

$$\boldsymbol{\omega}_{ij}^{(k)} = \mathbf{I}_q - \mathbf{v}_{ij}^{(k)T} \mathbf{B}_{ij}^{(k)} \quad (4.54)$$

όπου, το $\Sigma_{ij}^{(k)}$ δίνεται από την (43), και το $p(\mathbf{y}_{mt}; \Theta_i^{(k)})$ από την (41).

Βάσει αυτών των αποτελεσμάτων, η μεγιστοποίηση της (50) επί των \mathbf{B}_{ij} και \mathbf{D}_{ij} , όπου και συνίσταται το M-βήμα του δευτέρου κύκλου του αλγορίθμου, δίνει (δες Παράρτημα Β)

$$\mathbf{B}_{ij}^{(k+1)} = \mathbf{V}_{ij}^{(k)} \mathbf{v}_{ij}^{(k)} \left(\mathbf{v}_{ij}^{(k)T} \mathbf{V}_{ij}^{(k)} \mathbf{v}_{ij}^{(k)} + \boldsymbol{\omega}_{ij}^{(k)} \right)^{-1} \quad (4.55)$$

$$\mathbf{D}_{ij}^{(k+1)} = \text{diag}\{\mathbf{V}_{ij}^{(k)} - \mathbf{V}_{ij}^{(k)} \mathbf{v}_{ij}^{(k)} \mathbf{B}_{ij}^{(k+1)T}\} \quad (4.56)$$

όπου, το r_{ijmt} δίνεται από την (33), και

$$\mathbf{V}_{ij}^{(k)} = \frac{\sum_{m=1}^M \sum_{t=1}^T r_{ijmt}^{(k)} u_{ijmt}^{(k)} (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij}^{(k+1)}) (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij}^{(k+1)})^T}{\sum_{m=1}^M \sum_{t=1}^T r_{ijmt}^{(k)}} \quad (4.57)$$

Ένα περίγραμμα του προτεινομένου αλγορίθμου δίνεται στον Πίνακα 2.

4.5 Εφαρμογές

Στο κεφάλαιο αυτό, παρέχουμε μια εκτενή πειραματική αξιολόγηση των μοντέλων SHMM και factor analyzed SHMM, σε μια σειρά εφαρμογών μοντελοποίησης ακολουθιακών δεδομένων από διαφορετικές περιοχές. Για λόγους σύγκρισης, πλέον των προτεινομένων μεθόδων, αξιολογούμε επίσης τις μεθόδους εύρωστης μοντελοποίησης, ST-HMM [8], και large margin CHMM (LM-HMM) [19], καθώς και το baseline GHMM. Η υλοποίηση των θεωρουμένων μεθόδων έγινε σε Matlab R2008a, και εξετελέσθη σε μια Macintosh πλατφόρμα με μια Intel Core 2 Duo 2 GHz CPU, και 2 GB RAM, που έτρεχε

Πίνακας 4.2: AECM αλγόριθμος για το factor analyzed SHMM
 $k := 0$

• **Πρώτος Κύκλος**

1. Εκτέλεσε των forward-backward algorithm για να βρεις τις ποσότητες $a_{imt}^{(k)}$ and $b_{imt}^{(k)}$.
2. Εκτέλεσε το E-step υπολογίζοντας τις $\gamma_{hit}^{(k)}$, $\gamma_{imt}^{(k)}$, $\xi_{ijmt}^{(k)}$, $r_{ijmt}^{(k)}$, και $u_{ijmt}^{(k)}$, βάσει των (19), (20), (25), (34), και (26), αντιστοίχως.
3. Εκτέλεσε το M-step υπολογίζοντας τις $\pi_i^{(k+1)}$, $\pi_{hi}^{(k+1)}$, $c_{ij}^{(k+1)}$, $\mu_{ij}^{(k+1)}$, και $\nu_{ij}^{(k+1)}$, βάσει των (27)-(30) και (32), αντιστοίχως.

• **Δεύτερος Κύκλος**

1. Εκτέλεσε των forward-backward algorithm για να βρεις τις ποσότητες $a_{imt}^{(k)}$ and $b_{imt}^{(k)}$.
2. Υπολόγισε τις ποσότητες $\gamma_{hit}^{(k)}$, $\gamma_{imt}^{(k)}$, $\xi_{ijmt}^{(k)}$, $r_{ijmt}^{(k)}$, and $u_{ijmt}^{(k)}$, από τις (19), (20), (25), (34), και (26), αντιστοίχως. Ολοκλήρωσε το E-step του δευτέρου κύκλου της τρέχουσας επανάληψης του αλγορίθμου υπολογίζοντας τις ποσότητες $\nu_{ij}^{(k)}$ and $\omega_{ij}^{(k)}$, από τις (53) και (54), αντιστοίχως.
3. Εκτέλεσε το M-step υπολογίζοντας τις ποσότητες $B_{ij}^{(k+1)}$ και $D_{ij}^{(k+1)}$, από τις (55) και (56), αντιστοίχως.
4. Αν ο AECM αλγόριθμος συγκλίνει, **έξελθε**; αλλιώς θέσε $(k := k + 1)$ και επέστρεψε στο *Πρώτο Κύκλο*-1.

το λειτουργικό Mac OS X 10.5.2 (Leopard).

4.5.1 Εύρωστη Ταξινόμηση Ακολουθιών με Χρήση του SHMM

Αρχίζουμε με μια αξιολόγηση της επίδοσης του SHMM στην ταξινόμηση ακολουθιακών δεδομένων. Το πρώτο από τα θεωρούμενα προβλήματα συνίσταται στην αναγνώριση αμφιδέξιων χειρονομιών της American Sign Language για τις λέξεις: *against, aim, balloon, bandit, cake, chair, computer, concentrate, cross, deaf, explore, hunt, knife, relay, reverse*, και *role*.

Η χρησιμοποιούμενη συλλογή δεδομένων έχει υιοθετηθεί από την εργασία [20], και εκτήθη από 4 πρόσωπα που εξετέλεσαν μια εκάστη των εν λόγω μεθόδων. Περιλαμβάνει ένα σύνολο εκπαίδευσης, περιλαμβάνον 30 videos μεταβλητής διάρκειας ανά χειρονομία, και ένα σύνολο δοκιμής περιλαμβάνον 10 videos μεταβλητής διάρκειας ανά χειρονομία¹. Από τα δεδομένα αυτά εξήγαγαμε ένα σύνολο από features αντιστοιχούντα στην σχετική θέση προσώπου-χειρών, καθώς και το σχήμα καταλλήλων περιοχών δέρματος του προσώπου, αποδιδόμενων με χρήση ροπών Zernike [21], ως περιγράφεται στο [20]. Τοιουτοτρόπως, ένα 12-διάστατο feature vector παρήχθη για κάθε ακολουθία. Για τους σκοπούς της εφαρμογής ταξινόμησης, χρησιμοποιούμε 3 καταστάσεων μοντέλα. Σε κάθε περίπτωση, δοκιμάζουμε διάφορες τιμές του πλήθους συνιστωσών κατανομών ανά κατάσταση, J , και προσδιορίζουμε την τιμή εκείνη που αποδίδει την καλύτερη επίδοση για το κάθε μοντέλο. Οι προκύπτοντες ρυθμοί σφάλματος δίνονται στον Πίνακα 3. Τα αποτελέσματα των μεθόδων GHMM, SHMM, και ST-HMM

¹Τα εν λόγω δεδομένα μπορούν να ανακτηθούν δωρεάν από την ιστοθέση: http://www.iit.demokritos.gr/~dkosmo/downloads/gesture/help_gestureDB.htm

Πίνακας 4.3: Αναγνώριση Χειρονομιών: Ρυθμοί σφάλματος (%) για βέλτιστες τιμές J .

Gesture	SHMM ($J = 7$)	GHMM ($J = 9$)	ST-HMM ($J = 9$)	LM-HMM ($J = 9$)
<i>against</i>	0.19	0.97	0.51	0.28
<i>aim</i>	0.25	0.69	0.50	0.31
<i>balloon</i>	4.47	7.02	5.33	4.61
<i>bandit</i>	0.14	0.83	0.71	0.20
<i>chair</i>	11.33	23.64	16.06	13.18
<i>cake</i>	11.08	18.25	14.82	11.95
<i>computer</i>	6.4	6.88	5.50	4.13
<i>concentrate</i>	10.03	16.21	12.85	10.76
<i>cross</i>	15.04	28.13	19.54	16.88
<i>deaf</i>	0.42	4.37	2.19	1.62
<i>explore</i>	4.41	6.62	5.53	4.97
<i>hunt</i>	15.26	28.56	20.83	16.64
<i>knife</i>	11.65	24.12	18.13	13.47
<i>relay</i>	0.42	6.43	3.69	3.36
<i>reverse</i>	3.82	8.50	6.21	5.10
<i>role</i>	0.48	1.84	1.84	1.10
Μέσος	5.96	11.44	8.39	6.79

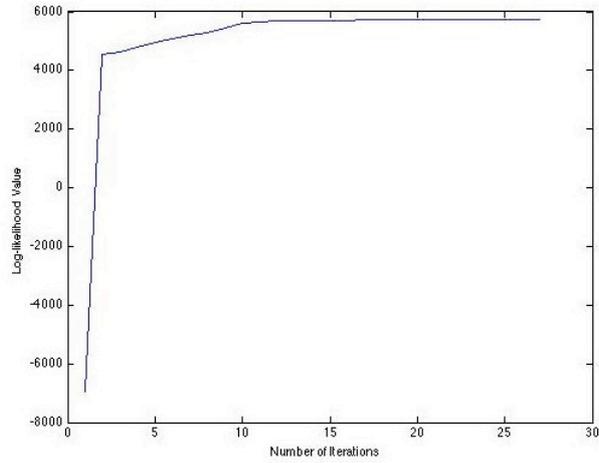
είναι μέσοι επί 30 εκτελέσεων του αλγορίθμου εκπαίδευσης από διαφορετικές εκκινήσεις Παρατηρούμε ότι το SHMM δίνει καλύτερη επίδοση από τους ανταγωνιστές του για μικρότερο μέγεθος μοντέλου (και, άρα, καλύτερη υπολογιστική αποδοτικότητα).

Εν συνεχεία, ασχολούμαστε με μια εφαρμογή βάσει κειμένου αναγνώρισης ομιλητή, χρησιμοποιώντας την συλλογή Japanese Vowels Data Set [22] από το UCI machine learning repository [23]. Στην συλλογή αυτή, η λέξη κλειδί που χρησιμοποιείται για αναγνώριση ομιλητή περιλαμβάνει δυο Ιαπωνικά σύμφωνα, /ae/, προφερόμενα διαδοχικά από 9 άνδρες ομιλητές. Για

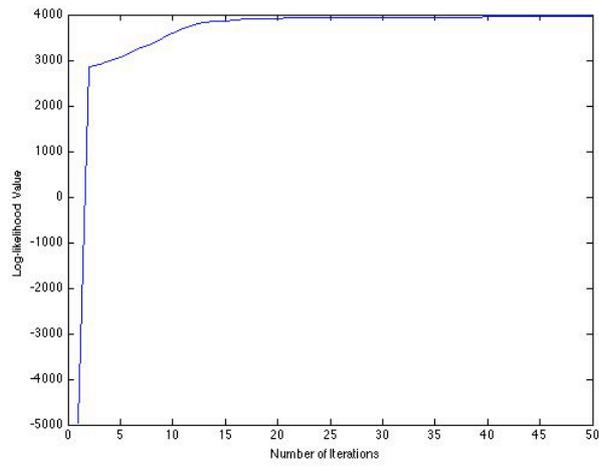
κάθε ακολουθία, μια βαθμού-12 LPC ανάλυση εφηρμόσθη και τα αντίστοιχα cepstrum coefficients ελήφθησαν, αποδίδοντας τα χρησιμοποιούμενα feature vectors. Ο αριθμός των διαθέσιμων χρονοσειρών είναι 640 στο σύνολο. Από αυτές, χρησιμοποιούμε τις 270 για εκπαίδευση και τις υπόλοιπες 370 για αξιολόγηση των εκπαιδευμένων μοντέλων. Χρησιμοποιούμε CHMMs με 2 καταστάσεις ($I = 2$), αντιστοιχούσες στα 2 προφερόμενα φωνήεντα, και με διάφορα πλήθη συνιστωσών κατανομών, J . Οι προκύπτοντες ρυθμοί αναγνώρισης δίνονται στο Πίνακα 4. Τα αποτελέσματα των μεθόδων GHMM, SHMM, και ST-HMM είναι μέσοι επί 30 εκτελέσεων του αλγορίθμου εκπαίδευσης από διαφορετικές εκκινήσεις, συνοδευόμενοι και από τις αντίστοιχες τυπικές αποκλίσεις. Παρατηρούμε για άλλη μια φορά ότι το SHMM δίνει καλύτερη επίδοση από τους ανταγωνιστές του για μικρότερο μέγεθος μοντέλου (και, άρα, καλύτερη υπολογιστική αποδοτικότητα). Τέλος, στα Σχ. 3 και Σχ. 4 παρέχουμε τους ρυθμούς σύγκλισης του EM αλγορίθμου για τις μεθόδους GHMM και SHMM, με $J = 3$, και $J = 5$ συνιστώσες κατανομές. Τα αποτελέσματα αυτά είναι μέσοι επί των 30 εκτελέσεων του EM αλγορίθμου και επί των 9 ομιλητών. Παρατηρούμε ότι ο EM αλγόριθμος συγκρίνει γρήγορα για αμφότερα τα μοντέλα, με τον ρυθμό σύγκλισης για το SHMM να είναι ελαφρώς καλύτερος σε σχέση με το GHMM, για όλες τις τιμές του J .

4.5.2 Εύρωστη Κατάτμηση Ακολουθιών

Στο πείραμα αυτό ασχολούμαστε με κατάτμηση ακολουθιών. Ειδικότερα, εκτελούμε ένα πείραμα αναγνώρισης φθόγγων, κάνοντας χρήση του TIMIT

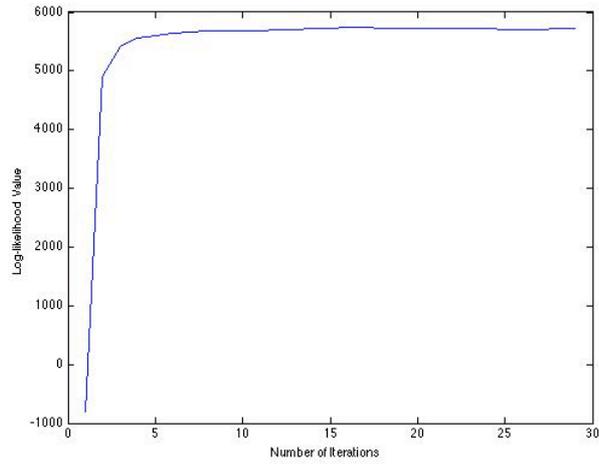


(i) GHMM with $J = 3$

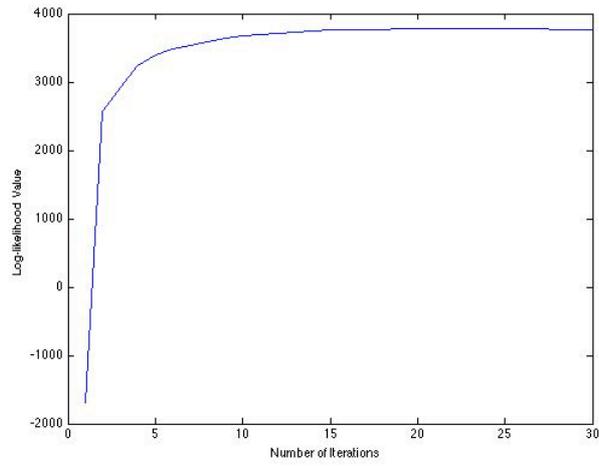


(ii) GHMM with $J = 5$

Σχήμα 4.3: Βάσει Κειμένου Αναγνώριση Ομιλητή: Ρυθμοί σύγκλισης της μεθόδου GHMM για $J = 3$ και $J = 5$.



(i) SHMM with $J = 3$



(ii) SHMM with $J = 5$

Σχήμα 4.4: Βάσει Κειμένου Αναγνώριση Ομιλητή: Ρυθμοί σύγκλισης της μεθόδου SHMM για $J = 3$ και $J = 5$.

Πίνακας 4.4: Βάσει Κειμένου Αναγνώριση Ομιλητή: Ρυθμοί σφάλματος (%) των διαφόρων μεθόδων για βέλτιστη επιλογή του J .

Model	$J = 2$	$J = 3$	$J = 4$	$J = 5$
SHMM	1.29 ± 0.17	1.21 ± 0.12	1.25 ± 0.15	1.27 ± 0.14
GHMM	2.07 ± 0.24	2.06 ± 0.25	1.97 ± 0.21	2.06 ± 0.22
ST-HMM	1.82 ± 0.21	1.78 ± 0.16	1.73 ± 0.18	1.77 ± 0.15
LM-HMM	1.46	1.41	1.38	1.39

Πίνακας 4.5: Κατάτμηση Ακολουθιών TIMIT: Ρυθμοί σφάλματος αναγνώρισης φθόγγων (%) για διάφορες τιμές του αριθμού συνιστωσών, J .

Μοντέλο	$J = 1$	$J = 2$	$J = 4$	$J = 8$
διαγώνιο SHMM	44.6	41.7	38.3	36.1
διαγώνιο GHMM	45.1	43.5	42.7	41.3
ST-HMM	44.9	43.8	42.3	39.4
LM-HMM	44.6	43.3	40.9	37.7

Πίνακας 4.6: Κατάτμηση Ακολουθιών TIMIT: Χρόνοι εκτέλεσης των αλγορίθμων εκπαίδευσης των αξιολογούμενων μοντέλων.

Μοντέλο	διαγ. SHMM	διαγ. GHMM	ST-HMM	LM-HMM
Χρόνος εκτέλεσης (sec)	1.65	1.23	1.33	904.71

speech corpus² [24,25]. Η συλλογή αυτή δεδομένων περιλαμβάνει 39-διάστατα ακουστικά feature vectors που περιλαμβάνουν 13 mel-frequency cepstral coefficients και τις πρώτης και δεύτερης τάξης παραγώγους τους. Στο πείραμά μας εκπαιδεύουμε ένα CHMM με χρήση καθεμίας των μεθόδων SHMM, GHMM, ST-HMM, and LM-HMM. Σε κάθε περίπτωση, τα διαθέσιμα 39-διάστατα ακουστικά feature vectors είναι σημασμένα βάσει 48 φωνητικών κλάσεων, καθεμία εκ των οποίων αναπαρίσταται με μια CHMM κατάσταση. Για κάθε μέθοδο, συγκρίνουμε την ακολουθία φωνητικών καταστάσεων που δίδει ο Viterbi αλγόριθμος με το “ground-truth” της φθογγολογικής ανάλυσης που παρέχει το TIMIT corpus. Για να λάβουμε ρυθμούς σφάλματος, αντιστοιχούμε τους 48 φθόγγους σε 39 ευρύτερες κατηγορίες, χρησιμοποιώντας την ίδια κοινότυπη σύμβαση όπως στο [19]. Έτσι, ένας κατάλληλος ρυθμός σφάλματος μπορεί να ληφθεί ευθυγραμμίζοντας τις Viterbi και ground truth ακολουθίες, με χρήση δυναμικού προγραμματισμού [25], και αθροίζοντας τα σφάλματα αντικατάστασης, προσθήκης και παράλειψης. Τα λαμβανόμενα αποτελέσματα δίδονται στον Πίνακα 5. Σημειώνεται ότι το SHMM δουλεύει καλύτερα από τον ανταγωνισμό, ιδίως για μεγαλύτερα πλήθη συνιστωσών κατανομών. Στον Πίνακα 6, παρέχουμε τους χρόνους εκτέλεσης του αλγορίθμου εκπαίδευσης για τις διάφορες μεθόδους, με $J = 1$. Παρατηρούμε ότι οι μέθοδοι SHMM, GHMM, και ST-HMM έχουν συγκρίσιμα υπολογιστικά κόστη σε σχέση με την μέθοδο LM-HMM, που έχει μεγάλες υπολογιστικές απαιτήσεις.

²Τα εν λόγω δεδομένα μπορούν να ανακτηθούν δωρεάν από την ιστοθέση http://www.cs.berkeley.edu/~feisha/codes/lm_cdhmm/ όπου παρέχονται για την αξιολόγηση της μεθόδου LM-HMM [19].

4.5.3 Εύρωστη Ταξινόμηση Πολυδιάστατων Ακολουθιών με Χρήση του Factor Analyzed SHMM

Τέλος, εξετάζουμε μια περίπτωση εξαιρετικά πολυδιάστατης ακολουθίας για να επιδείξουμε τα πλεονεκτήματα της μεθόδου factor analyzed SHMM σε σχέση με την διαγώνια επιλογή. Για τον σκοπό αυτό ασχολούμαστε με το πρόβλημα της εύρεσης της νοητικής κατάστασης ατόμων, με χρήση ακολουθιών από εγκεφαλικές fMRI απεικονίσεις [26]. Ο στόχος σε αυτή την εφαρμογή είναι να μπορέσουμε να εντοπίσουμε τις μεταβατικές νοητικές καταστάσεις με χρήση fMRI απεικόνισης εγκεφάλου, δια της εκπαίδευσης καταλλήλων HMM ταξινομητών. Αυτό το πρόβλημα είναι πολύ ενδιαφέρον από πλευράς μηχανικής μάθησης, καθώς τα δεδομένα εισόδου του προβλήματος είναι εξαιρετικά πολυδιάστατα, αραιά (ήτοι, μόνο περιορισμένα δεδομένα εκπαίδευσης είναι διαθέσιμα), και αρκετά θορυβώδη. Για λόγους σύγκρισης, στο πείραμα αυτό, πλέον του factor analyzed SHMM, αξιολογούνται επίσης το diagonal SHMM και το diagonal GHMM.

Τα δεδομένα που χρησιμοποιούμε για αυτό μας το πείραμα³ είναι διαιρεμένα σε δοκιμές. Σε κάποιες από αυτές, το υποκείμενο απλά ηρεμούσε κοιτάζοντας αφηρημένα σε ένα προκαθορισμένο σημείο σε μια οθόνη. Στις υπόλοιπες, παρουσιαζόταν στο υποκείμενο μια πρόταση και μια εικόνα (διαδοχικά) και ζητείτο από το υποκείμενο να πατήσει ένα κατάλληλο κουμπί αν η εικόνα συννφωνούσε με την πρόταση ή όχι. Το πλήθος αυτών των δοκιμών ήταν 40 ανα υποκείμενο. Τα δεδομένα εκτήθησαν κατά τις δοκιμές αυτές ως εξής: Μια απεικόνιση fMRI ελαμβάνετο κάθε 500msec.

³Τα εν λόγω δεδομένα μπορούν να ανακτηθούν δωρεάν από την ιστοθέση: <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/>

Πίνακας 4.7: Αναγνώριση Νοητικής Κατάστασης με Χρήση Απεικόνιση Εγκεφάλου: Μέσοι ρυθμοί σφάλματος και η τυπική τους απόκλιση (%) επί 30 εκτελέσεων των αλγορίθμων εκπαίδευσης.

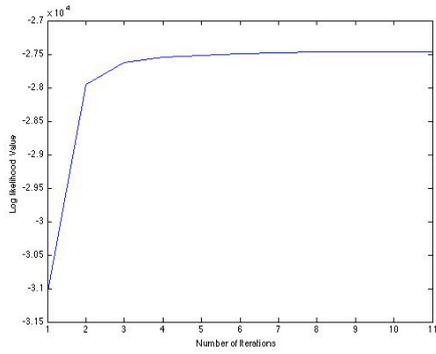
Μοντέλο	factor analyzed SHMM	διαγ. SHMM	διαγ. GHMM
p	70	70	70
q	9	-	-
Ρυθμός Σφάλματος	23.44 ± 2.22	29.87 ± 3.17	38.25 ± 4.79

Πίνακας 4.8: Αναγνώριση Νοητικής Κατάστασης με Χρήση Απεικόνιση Εγκεφάλου: Μέσοι χρόνοι εκτέλεσης των αλγορίθμων εκπαίδευσης των αξιολογουμένων μοντέλων.

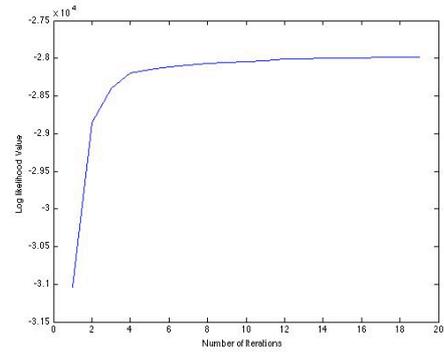
Model	factor analyzed SHMM	διαγ. SHMM	διαγ. GHMM
Χρόνος Εκτέλεσης (sec)	3.36	3.33	2.48

Μόνο ένα τμήμα του εγκεφάλου απεικονίζετο κάθε φορά. Τα δεδομένα μαρκάρωνταν σε 25-30 ανατομικές περιοχές του εγκεφάλου, τις Περιοχές Ενδιαφέροντος ("Regions of Interest", ROIs). Ο χρονισμός σε κάθε δοκιμή ήταν ο ακόλουθος: Το πρώτο ερέθισμα (εικόνα ή πρόταση) εμφανιζόταν στην αρχή της δοκιμής (εικόνα=1). Τέσσερα δευτερόλεπτα αργότερα (εικόνα=9), το ερέθισμα αφαιρείτο αντικαθιστούμενο με μια μαύρη οθόνη. Τέσσερα μετέπειτα δευτερόλεπτα (εικόνα=17) το δεύτερο ερέθισμα (πρόταση ή εικόνα) εμφανιζόταν στην οθόνη και παρέμενε εκεί τέσσεσερα δευτερόλεπτα ή μέχρι το υποκείμενο να πατήσει το κουμπί. Τέλος μια περίοδος χαλάρωσης 15 δευτερολέπτων (30 εικόνες) προσετίθετο στο τέλος κάθε δοκιμής. Έτσι κάθε δοκιμή διαρκούσε περίπου 27 δευτερόλεπτα (54 εικόνες).

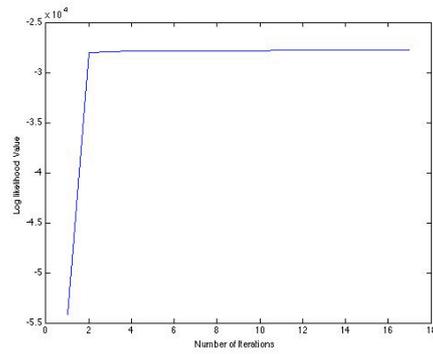
Στην εφαρμογή μας, εκπαιδεύουμε CHMMs ώστε να διακρίνουμε αν τα υποκείμενα βλέπουν μια εικόνα ή μια πρόταση με χρήση των δοθέντων ακολουθιών fMRI sequences. Για το σκοπό αυτό, εκπαιδεύουμε δυο 4×4 CHMMs (ένα για τα ερεθίσματα προτάσεων και ένα για τα ερεθίσμα-



(i)



(ii)



(iii)

Σχήμα 4.5: Αναγνώριση Νοητικής Κατάστασης με Χρήση Απεικόνισης Εγκεφάλου: Ρυθμοί σύγκλισης του αλγορίθμου εκπαίδευσης για τα μοντέλα (i) factor analyzed SHMM, (ii) διαγώνιο SHMM, και (iii) διαγώνιο GHMM.

τα εικόνων), με χρήση καθεμίας των αξιολογούμενων μεθόδων. Ειδικότερα, διαιρούμε τα δεδομένα σε ακολουθίες προτάσεων και εικόνων, και εκπαιδεύουμε τα μοντέλα χρησιμοποιώντας μόνο τα features που αντιστοιχούν στα p πιο ενεργά φωτεινά σημεία (voxels) από τα εξής Regions of Interest (ROI's): {'CALC' 'LIPL' 'LT' 'LTRIA' 'LOPER' 'LIPS' 'LDLPFC'}.

Κατ' αρχάς, παρατηρούμε ότι για τιμές $p > 70$, η εκπαίδευση των μοντέλων συχνά έχει προβλήματα καθώς καταλήγουμε σε singular ή σχεδόν singular πίνακες συνδιακύμανσης. Για αυτό, θέτουμε $p = 70$. Σχετικά με τον προσδιορισμό του βελτίστου πλήθους factors, q , για το factor analyzed SHMM, η χρήση του δημοφιλούς BIC κριτηρίου έδωσε $q = 9$. Σημειώνεται ότι η εφαρμοσιμότητα του κριτηρίου αυτού στο προσδιορισμό των factors σε mixtures of factor analyzers έχει θεωρητικά αποδειχθεί, δοθέντος ότι ο αριθμός των factors είναι γνωστός [27]. Στον Πίνακα 7 δίνονται οι ρυθμοί σφάλματος των θεωρουμένων μεθόδων και οι αντίστοιχες τυπικές αποκλίσεις επί 30 εκτελέσεων του EM αλγορίθμου. Είναι φανερό ότι το factor analyzed SHMM έχει καταφανώς καλύτερη επίδοση από το διαγώνιο αντίστοιχό του για ελάχιστο περισσότερο χρόνο εκτέλεσης. Τέλος, στο Σχ. 5 δίνονται οι ρυθμοί σύγκλισης των διαφόρων μεθόδων. Τα μεγέθη αυτά είναι μέσοι επί των 30 εκτελέσεων του EM αλγορίθμου και των δυο εκπαιδευμένων μοντέλων (για τις δυο νοητικές καταστάσεις). Παρατηρούμε και εδώ συγκρίσιμους ρυθμούς σύγκλισης για τα ποικίλα μοντέλα.

4.6 Συμπεράσματα

Στο κεφάλαιο αυτό προτείναμε την χρήση των πεπερασμένων υπερθέσεων της κατανομής Student's- t ως των κατανομών εκπομπής των κρυφών καταστάσεων των HMM μοντέλων, αποβλέποντες στην εισαγωγή μιας πιο εύρωστης σε ατυπικά δεδομένα μεθοδολογίας μοντελοποίησης ακολουθιακών δεδομένων εν συγκρίσει με τις τετριμμένες HMM-βασισμένες λύσεις οι οποίες βασίζονται σε χρήση Κανονικών κατανομών. Παρείχαμε δε μια πολλαπλών ακολουθιών επίλυση του μοντέλου με χρήση του EM αλγορίθμου, τόσο για την απλή περίπτωση της χρήσης πλήρων ή διαγωνίων πινάκων συνδιακύμανσης, όσο και για την περίπτωση αναλυμένων κατά παράγοντες πινάκων συνδιακύμανσης. Τα πειραματικά μας αποτελέσματα κατέδειξαν την αποτελεσματικότητα των προτεινομένων μεθόδων.

Παράρτημα Α. Επίλυση του SHMM

Ας ξεκινήσουμε με το Ε-βήμα του αλγορίθμου. Από την (13), έχουμε

$$\begin{aligned}
 Q(\Psi; \Psi^{(k)}) &= \sum_{m=1}^M \sum_{i=1}^I E_{\Psi^{(k)}}(s_{im1} | \mathbf{y}) \log \pi_i + \\
 &+ \sum_{m=1}^M \sum_{h=1}^I \sum_{i=1}^I \sum_{t=1}^{T-1} E_{\Psi^{(k)}}(s_{hmt} s_{im,t+1} | \mathbf{y}) \log \pi_{hi} + \\
 &+ \sum_{m=1}^M \sum_{i=1}^I \sum_{t=1}^T E_{\Psi^{(k)}}(s_{imt} | \mathbf{y}) E_{\Psi^{(k)}}(\log p(\mathbf{y}_{mt}^{comp}; \Theta_i) | \mathbf{y})
 \end{aligned} \tag{4.58}$$

εκ της οποίας συνάγεται η (16). Περαιτέρω, από την (14) και αγνοώντας τις σταθερές έχουμε

$$E_{\Psi^{(k)}}(\log p(\mathbf{y}_{mt}^{comp}; \Theta_i) | \mathbf{y}) = \sum_{j=1}^J \xi_{ijmt}^{(k)} \left[\log c_{ij} + Q_{1mt}(\nu_{ij}; \Psi^{(k)}) + Q_{2mt}(\boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}; \Psi^{(k)}) \right] \quad (4.59)$$

όπου

$$Q_{1mt}(\nu_{ij}; \Psi^{(k)}) = \frac{\nu_{ij}}{2} E_{\Psi^{(k)}}(\log u_{ijmt} - u_{ijmt} | \mathbf{y}_{mt}) - \log \Gamma\left(\frac{\nu_{ij}}{2}\right) + \frac{\nu_{ij}}{2} \log\left(\frac{\nu_{ij}}{2}\right) \quad (4.60)$$

$$Q_{2mt}(\boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}; \Psi^{(k)}) = -\frac{1}{2} \log |\boldsymbol{\Sigma}_{ij}| - \frac{1}{2} E_{\Psi^{(k)}}(u_{ijmt} | \mathbf{y}_{mt}) (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij})^T \boldsymbol{\Sigma}_{ij}^{-1} (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij}) \quad (4.61)$$

Αντικαθιστώντας τις (60)-(62) στην (59) παίρνουμε

$$\begin{aligned} & \sum_{m=1}^M \sum_{i=1}^I \sum_{t=1}^T \gamma_{imt}^{(k)} E_{\Psi^{(k)}}(\log p(\mathbf{y}_{mt}^{comp}; \Theta_i) | \mathbf{y}) \\ &= \sum_{i=1}^I \sum_{j=1}^J \sum_{m=1}^M \sum_{t=1}^T r_{ijmt}^{(k)} \left[\log c_{ij} + Q_{1mt}(\nu_{ij}; \Psi^{(k)}) \right. \\ & \quad \left. + Q_{2mt}(\boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}; \Psi^{(k)}) \right] \end{aligned} \quad (4.62)$$

Συνεπώς, για την περαιώση του Ε-βήματος της επίλυσης του SHMM, χρειάζεται να υπολογισθούν οι ποσότητες $\gamma_{imt}^{(k)}$, $\gamma_{himt}^{(k)}$, $\xi_{ijmt}^{(k)}$, $u_{ijmt}^{(k)}$, καθώς και $E_{\Psi^{(k)}}(\log u_{ijmt} | \mathbf{y}_{mt})$. Οι εκφράσεις των πιθανοτήτων της Μαρκοβιανής αλυσίδας, $\gamma_{imt}^{(k)}$ και $\gamma_{himt}^{(k)}$ δίνονται από τον εμπροσθεν-όπισθεν αλγόριθμο [1], εξ. (19) - (24). Η έκφραση των $\xi_{ijmt}^{(k)}$ μπορεί να βρεθεί απλά με χρήση του νόμου του Bayes, και δίνεται από την (25). Τέλος, οι δυο τελευταίες ποσότητες μπορούν να βρεθούν απευθείας με γενίκευση των αντιστοίχων αποτελεσμάτων

από τη SMM βιβλιογραφία [6], δίνοντας την (26) και

$$E_{\Psi^{(k)}}(\log u_{ijmt} | \mathbf{y}_{mt}) = \log u_{ijmt}^{(k)} - \log \left(\frac{\nu_{ij}^{(k)} + p}{2} \right) + \psi \left(\frac{\nu_{ij}^{(k)} + p}{2} \right) \quad (4.63)$$

από όπου συνάγεται ότι ο υπολογισμός του $E_{\Psi^{(k)}}(\log u_{ijmt} | \mathbf{y}_{mt})$ ανάγεται στον υπολογισμό των $u_{ijmt}^{(k)}$, με τον οποίο και ολοκληρώνεται το Ε-βήμα του αλγορίθμου.

Σχετικά με το Μ-βήμα, ξεκινάμε από τις πιθανότητες της Μαρκοβιανής αλυσίδας. Από την έκφραση (16), παρατηρούμε εύκολα ότι οι πιθανότητες αυτές δεν εξαρτώνται από τις υποθέσεις του μοντέλου σχετικά με την μορφή των κατανομών των παρατηρήσεων. Έτσι, ακολουθώντας την οικεία βιβλιογραφία (π.χ. [1,2]) παίρνουμε τις (27), (28). Οι εκτιμήτριες $c_{ij}^{(k+1)}$ υπολογίζονται με μεγιστοποίηση της $Q(\Psi; \Psi^{(k)})$ επί αυτών, υπό τον περιορισμό $\sum_{j=1}^J c_{ij} = 1$. Χρησιμοποιώντας τις (16) και (63), και εισάγοντας ένα Lagrange πολλαπλασιαστή για επιβολή του περιορισμού, παίρνουμε την (29). Οι εκτιμήτριες $\mu_{ij}^{(k+1)}$, υπολογίζονται με μεγιστοποίηση της $Q(\Psi; \Psi^{(k)})$ επί αυτών. Έχουμε

$$Q^*(\mu_{ij}; \Psi^{(k)}) = -\frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J \sum_{m=1}^M \sum_{t=1}^T r_{ijmt}^{(k)} u_{ijmt}^{(k)} [-2\mathbf{y}_{mt}^T \Sigma_{ij}^{-1} \mu_{ij} + \mu_{ij}^T \Sigma_{ij}^{-1} \mu_{ij}] \quad (4.64)$$

Καθώς

$$\frac{\partial \mu_{ij}^T \Sigma_{ij}^{-1} \mu_{ij}}{\partial \mu_{ij}} = 2 \Sigma_{ij}^{-1} \mu_{ij} \quad (4.65)$$

$$\frac{\partial \mathbf{y}_{mt}^T \Sigma_{ij}^{-1} \mu_{ij}}{\partial \mu_{ij}} = \Sigma_{ij}^{-1} \mathbf{y}_{mt} \quad (4.66)$$

εύκολα δείχνεται ότι η εξίσωση $\frac{\partial Q^*(\boldsymbol{\mu}_{ij}; \boldsymbol{\Psi}^{(k)})}{\partial \boldsymbol{\mu}_{ij}} = 0$ δίνει την (30). Καθ' όμοιον τρόπο, οι εκτιμήτριες των $\boldsymbol{\Sigma}_{ij}$ υπολογίζονται με μεγιστοποίηση της $Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)})$ επί αυτών. Έχουμε

$$\frac{\partial \log |\boldsymbol{\Sigma}_{ij}|}{\partial \boldsymbol{\Sigma}_{ij}^{-1}} = -\boldsymbol{\Sigma}_{ij} \quad (4.67)$$

$$\frac{\partial (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij})^T \boldsymbol{\Sigma}_{ij}^{-1} (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij})}{\partial \boldsymbol{\Sigma}_{ij}^{-1}} = (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij})(\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij})^T \quad (4.68)$$

από όπου και προκύπτει η (31). Τέλος, αναλόγως $\nu_{ij}^{(k+1)}$ η μεγιστοποίησης της $Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)})$ επί των ν_{ij} , δίνει την (32).

Παράρτημα Β. Επίλυση του Factor Analyzed SHMM

Αναφορικά με το όρο $E_{\boldsymbol{\Psi}^{(k)}}(\log p(\mathbf{y}_{mt}^{comp}; \boldsymbol{\Theta}_i) | \mathbf{y})$ στο δεύτερο κύκλο του αλγορίθμου, από την (50) έχουμε

$$E_{\boldsymbol{\Psi}^{(k)}}(\log p(\mathbf{y}_{mt}^{comp}; \boldsymbol{\Theta}_i) | \mathbf{y}) = \sum_{j=1}^J \xi_{ijmt}^{(k)} \left[\log c_{ij} + Q_{1mt}(\nu_{ij}; \boldsymbol{\Psi}^{(k)}) + Q_{2mt}(\boldsymbol{\mu}_{ij}, \mathbf{B}_{ij}, \mathbf{D}_{ij}; \boldsymbol{\Psi}^{(k)}) \right] \quad (4.69)$$

όπου

$$Q_{1mt}(\nu_{ij}; \boldsymbol{\Psi}^{(k)}) = -\log \Gamma\left(\frac{\nu_{ij}}{2}\right) + \frac{\nu_{ij}}{2} \log\left(\frac{\nu_{ij}}{2}\right) + \frac{\nu_{ij}}{2} E_{\boldsymbol{\Psi}^{(k)}}(\log u_{ijmt} - u_{ijmt} | \mathbf{y}_{mt}) \quad (4.70)$$

$$\begin{aligned} Q_{2mt}(\boldsymbol{\mu}_{ij}, \mathbf{B}_{ij}, \mathbf{D}_{ij}; \boldsymbol{\Psi}^{(k)}) &= -\frac{1}{2} \log |\mathbf{D}_{ij}| - \frac{1}{2} E_{\boldsymbol{\Psi}^{(k)}}(u_{ijmt} | \mathbf{y}_{mt}) (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij})^T \mathbf{D}_{ij}^{-1} (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij}) \\ &\quad + (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij})^T \mathbf{D}_{ij}^{-1} \mathbf{B}_{ij} E_{\boldsymbol{\Psi}^{(k)}}(u_{ijmt} \mathbf{x}_{ijmt} | \mathbf{y}_{mt}) \\ &\quad - \frac{1}{2} \text{tr}(\mathbf{B}_{ij}^T \mathbf{D}_{ij}^{-1} \mathbf{B}_{ij} E_{\boldsymbol{\Psi}^{(k)}}(u_{ijmt} \mathbf{x}_{ijmt} \mathbf{x}_{ijmt}^T | \mathbf{y}_{mt})) \end{aligned} \quad (4.71)$$

Αντικαθιστώντας τις (70)-(72) στην (51), παίρνουμε

$$\begin{aligned}
& \sum_{i=1}^I \sum_{m=1}^M \sum_{t=1}^T \gamma_{imt}^{(k)} E_{\Psi^{(k)}} (\log p(\mathbf{y}_{mt}^{comp}; \Theta_i) | \mathbf{y}) \\
= & \sum_{i=1}^I \sum_{j=1}^J \sum_{m=1}^M \sum_{t=1}^T r_{ijmt}^{(k)} \left[\log c_{ij} + Q_{1mt}(\nu_{ij}; \Psi^{(k)}) + Q_{2mt}(\boldsymbol{\mu}_{ij}, \mathbf{B}_{ij}, \mathbf{D}_{ij}; \Psi^{(k)}) \right]
\end{aligned} \tag{4.72}$$

Συνεπώς, για την περαιώση του Ε-βήματος στον δεύτερο κύκλο του αλγορίθμου, εκτός των $\gamma_{imt}^{(k)}$, $\gamma_{himt}^{(k)}$, $\xi_{ijmt}^{(k)}$, και $u_{ijmt}^{(k)}$, χρειάζεται επίσης η εκτιμητική των

$$E_{\Psi^{(k)}} (u_{ijmt} \mathbf{x}_{ijmt} | \mathbf{y}_{mt}) = u_{ijmt}^{(k)} E_{\Psi^{(k)}} (\mathbf{x}_{ijmt} | u_{ijmt}, \mathbf{y}_{mt}) \tag{4.73}$$

και

$$E_{\Psi^{(k)}} (u_{ijmt} \mathbf{x}_{ijmt} \mathbf{x}_{ijmt}^T | \mathbf{y}_{mt}) = u_{ijmt}^{(k)} E_{\Psi^{(k)}} (\mathbf{x}_{ijmt} \mathbf{x}_{ijmt}^T | u_{ijmt}, \mathbf{y}_{mt}) \tag{4.74}$$

η οποία ευθέως ανάγεται στην εκτιμητική των υστέρων προσδοκητών τιμών των διανυσμάτων παραγόντων \mathbf{x}_{ijmt} δοθέντων των παρατηρήσεων \mathbf{y}_{mt} και τω αντιστοίχων βαθμιδών συνδιακύμανσης, u_{ijmt} , ήτοι, $\mathbf{x}_{ijmt} | \mathbf{y}_{mt}, u_{ijmt}$. Τα αποτελέσματα αυτά έχουν εξαχθεί στο [18] για την περίπτωση μιας απλής πεπερασμένης υπέρθεσης κατανομών Student's- t και μπορούν εύκολα να γενικευθούν για το factor analyzed SHMM. Βάσει του [18], έχουμε

$$\mathbf{x}_{ijmt} | \mathbf{y}_{mt}, u_{ijmt} \sim \mathcal{N}(\mathbf{v}_{ij}^T (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij}), \boldsymbol{\omega}_{ij} / u_{ijmt}) \tag{4.75}$$

από όπου προκύπτει

$$E_{\Psi^{(k)}}(\mathbf{x}_{ijmt} | \mathbf{y}_{mt}, u_{ijmt}) = \mathbf{v}_{ij}^{(k)T} \left(\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij}^{(k)} \right) \quad (4.76)$$

$$\begin{aligned} E_{\Psi^{(k)}}(u_{ijmt} \mathbf{x}_{ijmt} \mathbf{x}_{ijmt}^T | \mathbf{y}_{mt}) &= \boldsymbol{\omega}_{ij}^{(k)} + \\ &+ u_{ijmt} \mathbf{v}_{ij}^{(k)T} \left(\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij}^{(k)} \right) \left(\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij}^{(k)} \right)^T \mathbf{v}_{ij}^{(k)} \end{aligned} \quad (4.77)$$

όπου

$$\mathbf{v}_{ij}^{(k)} = (\mathbf{B}_{ij}^{(k)} \mathbf{B}_{ij}^{(k)T} + \mathbf{D}_{ij}^{(k)})^{-1} \mathbf{B}_{ij}^{(k)} \quad (4.78)$$

$$\boldsymbol{\omega}_{ij}^{(k)} = \mathbf{I}_q - \mathbf{v}_{ij}^{(k)T} \mathbf{B}_{ij}^{(k)} \quad (4.79)$$

και συνεπώς η εκτιμητική των (74) - (75) ανάγεται στον υπολογισμό των $\mathbf{v}_{ij}^{(k)}$ και $\boldsymbol{\omega}_{ij}^{(k)}$.

Αναφορικά με την εκτιμήτρια των \mathbf{B}_{ij} , μεγιστοποιώντας την $Q(\Psi; \Psi^{(k)})$ επί αυτών έχουμε υπολογισμό των παραγώγων

$$\frac{\partial (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij})^T \mathbf{D}_{ij}^{-1} \mathbf{B}_{ij} E_{\Psi^{(k)}}(u_{ijmt} \mathbf{x}_{ijmt} | \mathbf{y}_{mt})}{\partial \mathbf{B}_{ij}} = \mathbf{D}_{ij}^{-1} (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij}) E_{\Psi^{(k)}}(u_{ijmt} \mathbf{x}_{ijmt} | \mathbf{y}_{mt})^T \quad (4.80)$$

$$\frac{\partial \text{tr}(\mathbf{B}_{ij}^T \mathbf{D}_{ij}^{-1} \mathbf{B}_{ij} E_{\Psi^{(k)}}(u_{ijmt} \mathbf{x}_{ijmt} \mathbf{x}_{ijmt}^T | \mathbf{y}_{mt}))}{\partial \mathbf{B}_{ij}} = 2 \mathbf{D}_{ij}^{-1} \mathbf{B}_{ij} E_{\Psi^{(k)}}(u_{ijmt} \mathbf{x}_{ijmt} \mathbf{x}_{ijmt}^T | \mathbf{y}_{mt}) \quad (4.81)$$

Βάσει των (81) - (82) και (77) - (78), εύκολα συνάγεται ότι η $\partial Q(\Psi; \Psi^{(k)}) / \partial \mathbf{B}_{ij} = 0$ δίνει την (59).

Τέλος, αναφορικά με την εκτιμήτρια των D_{ij} , έχουμε

$$\frac{\partial \log |D_{ij}|}{\partial D_{ij}^{-1}} = -D_{ij} \quad (4.82)$$

$$\frac{\partial (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij})^T D_{ij}^{-1} (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij})}{\partial D_{ij}^{-1}} = (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij})(\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij})^T \quad (4.83)$$

$$\frac{\partial (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij})^T D_{ij}^{-1} \mathbf{B}_{ij} E_{\Psi^{(k)}} (u_{ijmt} \mathbf{x}_{ijmt} | \mathbf{y}_{mt})}{\partial D_{ij}^{-1}} = (\mathbf{y}_{mt} - \boldsymbol{\mu}_{ij}) E_{\Psi^{(k)}} (u_{ijmt} \mathbf{x}_{ijmt} | \mathbf{y}_{mt})^T \mathbf{B}_{ij}^T \quad (4.84)$$

$$\frac{\partial \text{tr} (\mathbf{B}_{ij}^T D_{ij}^{-1} \mathbf{B}_{ij} E_{\Psi^{(k)}} (u_{ijmt} \mathbf{x}_{ijmt} \mathbf{x}_{ijmt}^T | \mathbf{y}_{mt}))}{\partial D_{ij}^{-1}} = \mathbf{B}_{ij} E_{\Psi^{(k)}} (u_{ijmt} \mathbf{x}_{ijmt} \mathbf{x}_{ijmt}^T | \mathbf{y}_{mt}) \mathbf{B}_{ij}^T \quad (4.85)$$

από όπου συνάγεται ότι η εξίσωση $\partial Q(\Psi; \Psi^{(k)}) / \partial D_{ij}^{-1} = 0$ δίνει την (59).

Βιβλιογραφία

- [1] O. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models*. New York: Springer Series in Statistics, 2005.
- [2] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, pp. 245--255, 1989.
- [3] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, B*, vol. 39, no. 1, pp. 1--38, 1977.
- [4] R. Dave and R. Krishnaparum, "Robust clustering methods: a unified view," *IEEE Transactions on Fuzzy Systems*, vol. 5, pp. 270--293, 1997.
- [5] A. Kosinski, "A procedure for the detection of multivariate outliers," *Computational Statistics and Data Analysis*, vol. 29, pp. 145--161, 1999.
- [6] D. Peel and G. McLachlan, "Robust mixture modeling using the t distribution," *Statistics and Computing*, vol. 10, pp. 335--344, 2000.
- [7] S. Chatzis, D. Kosmopoulos, and T. Varvarigou, "Signal modeling and classification using a robust latent space model based on t distributions," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 949--963, 2008.

- [8] L. Arslan and J. Hansen, "Selective training for hidden Markov models with applications to speech classification," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 1, pp. 46--54, 1999.
- [9] C.-H. Lee and S.-Y. Lee, "Noise-robust speech recognition using top-down selective attention with an HMM classifier," *IEEE Signal Processing Letters*, vol. 14, no. 7, pp. 489--491, 2007.
- [10] J.-T. Chien and C.-P. Liao, "Maximum confidence hidden Markov modeling for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 4, pp. 606--616, 2008.
- [11] C. Liu and D. Rubin, "ML estimation of the t distribution using EM and its extensions, ECM and ECME," *Statistica Sinica*, vol. 5, no. 1, pp. 19--39, 1995.
- [12] X. Meng and D. van Dyk, "The EM algorithm - an old folk song sung to a fast new tune (with discussion)," *Journal of the Royal Statistical Society B*, vol. 59, no. 3, pp. 511--567, 1997.
- [13] S. Dharanipragada and K. Visweswariah, "Gaussian mixture models with covariances or precisions in shared multiple subspaces," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1255--1266, 2006.
- [14] J. Lindblom and J. Samuelsson, "Bounded support Gaussian mixture modeling of speech spectra," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 1, pp. 88--99, 2003.

- [15] J. Frankel and S. King, "Speech recognition using linear dynamic models," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 1, pp. 246--256, 2007.
- [16] M. Tipping and C. Bishop, "Mixtures of probabilistic principal component analyzers," *Neural Computation*, vol. 11, no. 2, pp. 443--482, 1999.
- [17] Z. Ghahramani and G. Hinton, "The EM algorithm for mixtures of factor analyzers," Department of Computer Science, University of Toronto, Toronto, Canada, M5S 1A4, Tech. Rep. CRGTR- 96-1, 1997.
- [18] G. McLachlan, R. Bean, and L. B.-T. Jones, "Extension of the mixture of factor analyzers model to incorporate the multivariate t -distribution," *Comp. Stat. Data Analysis*, vol. 51, no. 11, pp. 5327--5338, 2006.
- [19] F. Sha and L. K. Saul, "Large margin hidden Markov models for automatic speech recognition," in *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. Platt, and T. Hoffman, Eds. Cambridge, MA: MIT Press, 2007, pp. 1249--1256.
- [20] D. Kosmopoulos and I. Maglogiannis, "Hand tracking for gesture recognition tasks using dynamic Bayesian network," *International Journal of Intelligent Systems and Applications*, vol. 1, no. 3/4, pp. 359--375, 2006.
- [21] R. Mukundan and K. R. Ramakrishnan, *Moment Functions in Image Analysis: Theory and Applications*. World Scientific, 1998.
- [22] M. Kudo, J. Toyama, and M. Shimbo, "Multidimensional curve classification using passing-through regions," *Pattern Recognition Letters*, vol. 20, no. 11-13, pp. 1103--1111, 1999.

- [23] A. Asuncion and D. Newman, "UCI machine learning repository," 2007. [Online]. Available: <http://www.ics.uci.edu/~mllearn/MLRepository.html>
- [24] L. F. Lamel, R. H. Kassel, and S. Seneff, "Speech database development: design and analysis of the acoustic-phonetic corpus," in *Proceedings of the DARPA Speech Recognition Workshop*, 1986, pp. 100--109.
- [25] K. F. Lee and H. W. Hon, "Speaker-independent phone recognition using hidden Markov models," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 11, pp. 1641--1648, 1988.
- [26] T. M. Mitchell, R. Hutchinson, R. S. Niculescu, F. Pereira, X. Wang, M. Just, and S. Newman, "Learning to decode cognitive states from brain images," *Machine Learning*, vol. 57, no. 1, pp. 145--175, 2004.
- [27] G. McLachlan and D. Peel, *Finite Mixture Models*. Wiley Series in Probability and Statistics, 2000.

Κεφάλαιο 5

Ταυτοποίηση Ομιλητή με Χρήση Gaussian Process Ταξινομητών

5.1 Προοίμιο

Οι σημερινές κυριαρχούσες εν τη βιβλιογραφία και πρακτική μεθοδολογίες βασίζονται σε πιθανοτικά παραγωγικά μοντέλα ομιλητών, όπως τα Gaussian mixture models (GMMs) και τα hidden Markov models (HMMs). Ένας σημαντικός περιορισμός αυτών των μοντέλων είναι η δειγματική, frame-level φύση τους, καθώς παρέχουν την δυνατότητα διάκρισης μεταξύ πλήρων σημάτων εκφράσεων (ακολουθιακά σήματα) με έναν έμμεσο τρόπο, δια του χειρισμού τους ως αλληλουχιών μεμονωμένων δειγμάτων (frames). Προκειμένου να επιλύσουμε αυτό το μειονέκτημα, στο τελευταίο αυτό κεφάλαιο του παρόντος πονήματος, παρουσιάζουμε μια καινοτόμο, διακριτική πλήρων ακολουθιών μεθοδολογία ταυτοποίησης ομιλητή, βασισμένη σε μια πρωτότυπη μορφή Gaussian process ταξινομητών, κατάλληλη για την

διάκριση πλήρων ακολουθιών ως ενιαίων ενοτήτων με αλληλουχικές ιδιότητες. Οι στοχαστικές διαδικασίες Gauss (Gaussian processes, GPs) είναι μια σημαντική Bayesian μεθοδολογία μηχανικής μάθησης, γνωστή για την πολύ καλή της επίδοση σε ποικίλα προβλήματα ταξινόμησης και παλινδρόμησης σε πολυδιάστατα δεδομένα. Είναι δε σχεδιασμένες για διάκριση πολυδιάστατων δεδομένων. Η διαχείριση ακολουθιακών δεδομένων με χρήση GPs είναι εύκολα εφικτή, χρησιμοποιώντας τους λεγόμενους score-space kernels ως τις συναρτήσεις συνδιακύμανσης των GP μοντέλων, διαμορφούμενους επί τη βάση υποθεμελιωνώντων παραγωγικών μοντέλων. Ειδικότερα, προτείνουμε την χρήση πυρήνων λογαριθμο-πιθανοφανείας (log likelihood-ratio kernels) βασισμένων σε GMM μοντέλα για την περίπτωση της ανεξάρτητης από το κείμενο ταυτοποίησης ομιλητή, και την χρήση log likelihood-ratio kernels βασισμένων σε HMM μοντέλα για την περίπτωση της βάσει κειμένου ταυτοποίησης ομιλητή. Η εξαντλητική αξιολόγηση της προτεινόμενης μεθόδου, με χρήση δημοσίως διαθέσιμων δεδομένων, καταδεικνύει τα πλεονεκτήματα της προτεινόμενης μεθόδου σε σχέση με μερικές από τις πιο δημοφιλείς σήμερα τεχνικές.

Η εργασία μου αυτή έχει γίνει δεκτή για δημοσίευση υπό το άρθρο μου: Sotirios Chatzis, Theodora Varvarigou, "Speaker Verification Using Sequence Discriminant Gaussian Process Classifiers," *IEEE Transactions on Audio, Speech, and Language Processing*, Accepted for Future Publication, 2008.

5.2 Εισαγωγή

Τα πιθανοτικά παραγωγικά μοντέλα αποτελούν την βάση των συγχρόνων μεθοδολογιών ταυτοποίησης ομιλητή [1]. Συνήθως, Gaussian mixture models (GMMs) χρησιμοποιούνται για μια ανεξάρτητη κειμένου ταυτοποίηση, [2], ενώ hidden Markov models (HMMs) χρησιμοποιούνται για μια ταυτοποίηση βάσει κειμένου [3]. Στην πιο αποτελεσματική τους μορφή, τέτοια μοντέλα λειτουργούν στην βάση ενός υποδείγματος διακριτικής ταξινόμησης, με χρήση μιας συνάρτησης βαθμού (score function) εφαρμοζόμενης επί των τιμών πιθανοφαιίας δυο ανταγωνιστικών μοντέλων ως προς ένα δοσμένο σήμα έκφρασης. Μια βασική ιδιότητα αυτών των προσεγγίσεων είναι ότι παρέχουν διάκριση μεταξύ πλήρων εκφράσεων-ακολουθιών με έναν έμμεσο τρόπο. Αυτό οφείλεται στην δειγματικού επιπέδου (frame level) λειτουργία τους, καθώς οι τιμές της πιθανοφαιίας ως προς μια ακολουθία με χρήση τέτοιων μοντέλων εξάγονται είτε ως ο μέσος των πιθανοφαιιών ως προς τα συνιστώμενα τη δείγματα, ή με χρήση ενός HMM. Συνεπώς, διάκριση με χρήση αυτών των προσεγγίσεων γίνεται χρησιμοποιώντας τις ακολουθίες εισόδου ως σύνολα ανεξαρτήτων δειγμάτων, ή δειγμάτων που συσχετίζονται υπό τις συνήθως περιοριστικές προδιαγραφές της Μαρκοβιανής αλυσίδας, σε αντίθεση με την πραγματική φύση του προβλήματος ταυτοποίησης ομιλητή που έχει να κάνει με την διάκριση πλήρων ακολουθιών ως ενός όλου. Το γεγονός αυτό έχει τελικά ως αντίκτυπο την απώλεια σχετικής πληροφορίας και, συνεπώς, την υποβάθμιση της επίδοσης του τελικού μοντέλου.

Προκειμένου να επιλύσουμε αυτά τα προβλήματα, στην εργασία αυτή παρέχουμε μια πρωτότυπη μεθοδολογία ταυτοποίησης ομιλητή, βασι-

σμένη σε μια ενιαίως διακρίνουσα πλήρεις ακολουθίες μορφή δυαδικών ταξινομητών Gaussian process (Gaussian process binary classifiers). Οι Gaussian processes (GPs) είναι βασισμένες σε μια εξαιρετικά αποδοτική μέθοδο τοποθέτησης μιας prior κατανομής επί του χώρου των δυνατών συναρτήσεων ταξινόμησης ή παλινδρόμησης. Έχουν ένα μικρό αριθμό ρυθμιστών παραμέτρων, μπορούν να εκτιμηθούν με χρήση ενός σχετικά μικρού συνόλου εκπαίδευσης, διαθέτουν σημαντική ευρωστία σε outliers και την ικανότητα να ταξινομούν αραιά δεδομένα χωρίς να ρέπουν προς υπερεκπαίδευση (καθώς στην πραγματικότητα δεν έχουμε διαδικασία εκπαίδευσης υπό την αυστηρή του όρου έννοια) [4]. Συγκριτικά με μια άλλη ευρέως δημοφιλή μορφή διακριτικών ταξινομητών πυρήνα (discriminative kernel classifiers), το support vector machine (SVM) [5], τα GPs διαθέτουν πολλά πλεονεκτήματα, το σημαντικότερο των οποίων είναι ότι ο ταξινομητής GP παράγει μια έξοδο με καθαρή πιθανοτική ερμηνεία, παρέχοντας ένα μέτρο αβεβαιότητας για τις λαμβανόμενες προβλέψεις, σε αντίθεση με τα SVMs που περιορίζονται στην παροχή των καθαυτό προβλέψεων ταξινόμησης και μόνο. Ως ε συζητήθη στο [6], αυτή η ελκυστική ιδιότητα των GPs οδηγεί γενικά σε καλύτερες καμπύλες σφάλματος, δια της λήψης υπόψιν της επίδρασης της λαμβανόμενης αβεβαιότητας (για μια εκτενή παράθεση των πλεονεκτημάτων και των διαφορών των δυαδικών ταξινομητών GP σε σχέση με τα SVMs, ανατρέξτε στο [4]).

Στην βασική τους διατύπωση, τα GPs θεωρούν διανυσματικές εισόδους σταθερής διάστασης. Ωστόσο, μια διακριτική ακολουθιών μορφή Gaussian processes, κατάλληλη για την αποδοτική διεξαγωγή της ταυτοποίησης ομιλητή, μπορεί εύκολα να ληφθεί χρησιμοποιώντας ένα score-space kernel ως

τη συνάρτηση συνδιακύμανσης του GP μοντέλου. Οι score-space kernels [7], περιλαμβάνοντας τον δημοφιλή Fisher kernel [8] ως μια ειδική περίπτωση τους, επιτρέπουν την ταξινόμηση ακολουθιών με χρήση διακριτικών ταξινομητών όπως Gaussian processes και SVMs. Αυτό γίνεται εφικτό αξιοποιώντας ένα σύνολο από υποθεμελιώνοντα πιθανοτικά παραγωγικά μοντέλα για την δημιουργία μιας αντιστοιχίας των μεταβλητού μήκους ακολουθιών σε πολυδιάστατα σημεία σε ένα σταθερής διάστασης χώρο, που θα καλείται ως χώρος βαθμού (score-space). Στην εργασία αυτή, οι χρησιμοποιούμενοι score-space kernels είναι log likelihood-ratio kernels, βασισμένοι σε GMMs για την περίπτωση της ανεξάρτητης κειμένου ταυτοποίησης, και βασισμένοι σε HMMs για την περίπτωση της βάσει κειμένου ταυτοποίησης.

Το υπόλοιπο του ενθάδε κεφαλαίου έχει ως εξής: Η ενότητα 5.3 παρέχει μια σύνοψη των δημοφιλέστερων σήμερα τεχνικών ταυτοποίησης ομιλητή, επί τη βάσει πιθανοτικών παραγωγικών μοντέλων ομιλητή. Στην ενότητα 5.4, παρέχουμε μια εισαγωγή στα Gaussian process binary classifiers. Στην ενότητα 5.5, αρχικά εισάγουμε τους διακριτικούς ακολουθιών GP classifiers. Εν συνεχεία, παρέχουμε την προτεινόμενη μεθοδολογία ταυτοποίησης ομιλητή, επί τη βάσει διακριτικών ακολουθιών ταξινομητών GP διαμορφούμενων με χρήση καταλλήλων log likelihood-ratio kernels. Τέλος, στην ενότητα 5.6, η πειραματική αξιολόγηση της προτεινόμενης μεθόδου διεξάγεται με χρήση των συλλογών PolyVar database [9], και YOHO database [10].

5.3 Κυρίαρχες Τεχνικές Ταυτοποίησης Ομιλητή

5.3.1 Ανεξάρτητη Κειμένου Ταυτοποίηση με Χρήση GMM

Ένα g -συνιστωσών GMM H με εισόδους από ένα d -διάστατο χώρο παρατήρησης \mathcal{X} έχει την μορφή

$$p(\mathbf{x}|H, \boldsymbol{\theta}_H) \triangleq \sum_{j=1}^g w_j \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \Sigma_j) \quad (5.1)$$

όπου $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \Sigma_j)$ είναι μια πολυμετάβλητη Gaussian με μέσο $\boldsymbol{\mu}_j$ και συνδιακύμανση Σ_j

$$\begin{aligned} \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \Sigma_j) &\triangleq (2\pi)^{-d/2} |\Sigma_j|^{-1/2} \\ &\times \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{x} - \boldsymbol{\mu}_j)\right) \end{aligned} \quad (5.2)$$

$\boldsymbol{\theta}_H = \{\boldsymbol{\mu}_j, \Sigma_j, w_j | j = 1, \dots, g\}$ είναι οι παράμετροι του μοντέλου, και $p(\mathbf{x}|H, \boldsymbol{\theta}_H)$ η πιθανοφάνεια του μοντέλου H για είσοδο $\mathbf{x} \in \mathcal{X}$. Τα Gaussian mixture models αποτελούν τα δημοφιλέστερα παραγωγικά μοντέλα ομιλητή που χρησιμοποιούνται στην ανεξάρτητη κειμένου ταυτοποίηση ομιλητή, λαμβάνοντας υψηλούς ρυθμούς αναγνώρισης για χαμηλές υπολογιστικές πολυπλοκότητες [1]. Αν και η γενική μορφή του μοντέλου, ως δίδεται από την (1), υποστηρίζει πλήρεις πίνακες συνδιακύμανσης, ήθισται διαγώνιοι πίνακες συνδιακύμανσης να προτιμούνται σε εφαρμογές ανεξάρτητης κειμένου ταυτοποίησης ομιλητή, καθώς προσφέρουν καλύτερη υπολογιστική αποδοτικότητα εκπαίδευσης, εφόσον συνεχείς αντιστροφές πινάκων $d \times d$ δεν είναι αναγκαίες.

Η πιο επιτυχημένη GMM-βασισμένη μεθοδολογία ανεξάρτητης κειμένου ταυτοποίησης ομιλητή είναι η μέθοδος *Gaussian mixture model -- universal background model (GMM-UBM)* [2]. Υπό αυτήν, η ταυτοποίηση ομιλητή για έναν ομιλητή-πελάτη γίνεται με χρήση δυο διαφορετικών GMMs: (α) ενός *οικουμενικού* μοντέλου Ω , αναπαριστώντος τις κοινές ιδιότητες των σημάτων ομιλίας μεταξύ όλων των ομιλητών, και με εκτιμητική των παραμέτρων του από ένα μεγάλο αριθμό ομιλητών με χρήση του EM algorithm [12], και (β) ενός μοντέλου *πελάτη* M , υπολογιζομένου με προσαρμογή του οικουμενικού μοντέλου Ω με χρήση της ομιλίας του πελάτη και μιας Bayesian προσαρμοστικής τεχνικής (maximum a posteriori (MAP) estimation) [13].

Τότε, δοθέντος του ακολουθιακού σήματος $\mathbf{x} = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$, μια *βαθμολογία διάκρισης (utterance score)* $S(\mathbf{x})$ λαμβάνεται παίρνοντας το log likelihood-ratio του μοντέλου πελάτη, M , και του οικουμενικού μοντέλου, Ω , για καθένα από τα frames \mathbf{x}_t τα συνιστώντα την ακολουθία \mathbf{x} , και ακολουθώντας αθροίζοντας κατά μήκος της ακολουθίας, ήτοι

$$\begin{aligned} S(\mathbf{x}) &\triangleq \frac{1}{T} \sum_{t=1}^T \log \frac{p(\mathbf{x}_t|M, \boldsymbol{\theta}_M)}{p(\mathbf{x}_t|\Omega, \boldsymbol{\theta}_\Omega)} \\ &= \frac{1}{T} \sum_{t=1}^T \log p(\mathbf{x}_t|M, \boldsymbol{\theta}_M) - \frac{1}{T} \sum_{t=1}^T \log p(\mathbf{x}_t|\Omega, \boldsymbol{\theta}_\Omega) \\ &= \frac{1}{T} \log p(\mathbf{x}|M, \boldsymbol{\theta}_M) - \frac{1}{T} \log p(\mathbf{x}|\Omega, \boldsymbol{\theta}_\Omega) \end{aligned} \quad (5.3)$$

Αυτό το utterance score $S(\mathbf{x})$ αποτελεί και το κριτήριο λήψης απόφασης υπό την GMM-UBM προσέγγιση, δια της συγκρίσεως της τιμής του με ένα κατώφλι κατάλληλα επιλεγμένο ώστε να αποδίδει μια επιθυμητή ισορροπία μεταξύ των διαφορετικών τύπων σφαλμάτων (i.e. false positives και false

negatives).

5.3.2 Βάσει Κειμένου Ταυτοποίηση με Χρήση HMM

Στις βάσει κειμένου εφαρμογές, όπου το περιεχόμενο της ομιλίας είναι από πριν γνωστό, επιπλέον χρονική πληροφορία μπορεί να περιληφθεί στην διαδικασία της ταυτοποίησης ομιλητή δια της χρήσης hidden Markov models ως της βάσης των συναρτήσεων πιθανοφαινείας, αντί των GMMs. Όπως έχει καταδειχθεί σε πολλές πρόσφατες μελέτες (δες, π.χ., [1]), αν και τα HMMs δεν φαίνεται να προσφέρουν αξιολογικά οφέλη στην ανεξάρτητη κειμένου ταυτοποίηση ομιλητή, εμφανίζονται πράγματι να προσφέρουν υψηλότερους ρυθμούς αναγνώρισης σε σχέση με τα GMMs όταν έχουμε να κάνουμε με βάσει κειμένου ταυτοποίηση. Αυτό οφείλεται προφανώς στην ικανότητα των HMMs να συλλαμβάνουν καλύτερα τις χρονικές ιδιότητες μας εκ των προτέρων γνωστής φράσης-κλειδιού ταυτοποίησης, πληροφορία που είναι καταδικασμένη στην απώλεια όταν χρησιμοποιούμε GMMs.

Ας συμβολίσουμε ένα v -καταστάσεων HMM H με το σύνολο παραμέτρων $\theta_H = \{\pi_h, a_{h\xi}, b_h(\mathbf{x}_t) | h, \xi = 1, \dots, v\}$, όπου π_h είναι η κατανομή των αρχικών καταστάσεων, $a_{h\xi}$ οι πιθανότητες μετάβασης ενός βήματος, και $b_h(\mathbf{x}_t)$ η πιθανότητα εκπομπής της παρατήρησης \mathbf{x}_t από την h οστή κατάσταση, μοντελοποιούμενη ως ένα GMM [14]. Τύποις, αριστερά-προς-δεξιά HMMs χωρίς υπερβάσεις καταστάσεων, και με διαγώνιους πίνακες συνδιακύμανσης προτιμούνται, καθώς έχει καταδειχθεί ότι είναι κατάλληλα για χρήση με είσοδο βασισμένη σε mel frequency cepstral coefficients (MFCC), που ήθισται να χρησιμοποιούνται για αναπαράσταση των σημάτων ομιλίας [15]. Με άλ-

λα λόγια, έχουμε $\pi_1 = 1$, $\pi_{h \neq 1} = 0$, $a_{h\xi} = 0 \forall \xi > h + 1 \wedge \xi < h$, και

$$b_h(\mathbf{x}_t) = \sum_{j=1}^g w_{hj} \mathcal{N}(\mathbf{x}_t | \boldsymbol{\mu}_{hj}, \Sigma_{hj}) \quad (5.4)$$

όπου g είναι ο αριθμός των συνιστωσών κατανομών ανά κατάσταση, και $\Sigma_{hj} = \text{diag}(\Sigma_{hj}^1, \dots, \Sigma_{hj}^d)$. Συνήθως, προκειμένου να εκτελέσουμε την βάσει κειμένου ταυτοποίηση ενός πελάτη ομιλητή, δυο διαφορετικά μοντέλα χρησιμοποιούνται, ένα *οικουμενικό* μοντέλο, Ω , αναπαριστών τις κοινές ιδιότητες των σημάτων ομιλίας διαφορετικών ομιλητών, εκπαιδευόμενο με χρήση του Baum-Welch αλγορίθμου [14], και ένα HMM μοντέλο *πελάτη*, M , αναπαριστών την συνάρτηση πιθανοφανείας του ομιλητή, λαμβανόμενο με χρήση μιας τεχνικής MAP προσαρμογής του οικουμενικού μοντέλου Ω στα δεδομένα ομιλίας του πελάτη, παρομοίως με την περίπτωση του GMM-UBM αλγορίθμου [3]. Τότε, δοθείσης μιας ακολουθίας (έκφρασης) $\mathbf{x} = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$, ένα διακριτικό score $S(\mathbf{x})$ μπορεί να υπολογισθεί λαμβάνοντας το log likelihood-ratio του μοντέλου πελάτη, M , προς το οικουμενικό μοντέλο, Ω , για την πλήρη ακολουθία \mathbf{x} , ήτοι

$$S(\mathbf{x}) \triangleq \log \frac{p(\mathbf{x} | M, \boldsymbol{\theta}_M)}{p(\mathbf{x} | \Omega, \boldsymbol{\theta}_\Omega)} = \log p(\mathbf{x} | M, \boldsymbol{\theta}_M) - \log p(\mathbf{x} | \Omega, \boldsymbol{\theta}_\Omega) \quad (5.5)$$

Το score $S(\mathbf{x})$ χρησιμοποιείται για την λήψη απόφασης ως προς την εργασία της βάσει κειμένου ταυτοποίησης ομιλητή, δια της συγκρίσεως με κάποιο κατάλληλα προκαθορισμένο κατώφλι. Τέλος, αναφορικά με την log likelihood $\log p(\mathbf{x} | H, \boldsymbol{\theta}_H)$, αυτή μπορεί εύκολα να υπολογισθεί με χρήση του τοις πάσι γνωστού *forward-backward algorithm* (δες, π.χ., [14]). Στο υπόλοιπο του παρόντος κεφαλαίου, θα αναφερόμεθα σε αυτή την βάσει κειμένου

μεθοδολογία ταυτοποίησης ομιλητή ως η *HMM-UBM* μέθοδος.

5.4 Δυαδική Ταξινόμηση με Χρήση Gaussian Processes

Έστω ένας χώρος παρατήρησης \mathcal{X} . Μια στοχαστική διαδικασία Gauss (Gaussian process) $f(\mathbf{x})$, $\mathbf{x} \in \mathcal{X}$, ορίζεται ως μια συλλογή τυχαίων μεταβλητών, οιοδήποτε τυχαίο πλήθος των οποίων έχει μια από κοινού κατανομή Gauss [4]. Μια Gaussian process ορίζεται πλήρως δια της συνάρτησης μέσης τιμής της και της συνάρτησης συνδιακύμανσής της. Ορίζουμε την μέση τιμή $m(\mathbf{x})$ και την συνάρτηση συνδιακύμανσης $k(\mathbf{x}, \mathbf{x}')$ μιας πραγματικής διαδικασίας $f(\mathbf{x})$ ως

$$\begin{aligned} m(\mathbf{x}) &= \mathbb{E}[f(\mathbf{x})] \\ k(\mathbf{x}, \mathbf{x}') &= \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))] \end{aligned} \tag{5.6}$$

και γράφουμε

$$f(\mathbf{x}) \sim \mathcal{N}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x})) \tag{5.7}$$

Συνήθως, για απλότητα, και χωρίς απώλεια γενικότητας, ο μέσος της διαδικασίας λαμβάνεται ως μηδενικός, $m(\mathbf{x}) = 0$, παρότι αυτό δεν είναι απαραίτητο. Αναφορικά με την επιλογή της συνάρτησης συνδιακύμανσης, μια μεγάλη γκάμα από kernel functions $k(\mathbf{x}, \mathbf{x}')$ θα μπορούσε να θεωρηθεί, ανάλογα με την εκάστοτε εφαρμογή [4].

Ας θεωρήσουμε ένα πρόβλημα δυαδικής ταξινόμησης επί του χώρου παρατήρησης \mathcal{X} . Στην συνέχεια, θα χρησιμοποιούμε τις ετικέτες $y = +1$ και $y = -1$ για να διακρίνουμε τις δυο κλάσεις. Μια κομψή λύση στο πρόβλημα αυτό μπορεί να ληφθεί με χρήση των Gaussian processes [4]. Η βασική ιδέα στην χρήση των Gaussian processes για διενέργεια δυαδικής ταξινόμησης εί-

ναι αρκετά απλή -- επιβάλλουμε μια GP prior κατανομή επί της κρυφής συνάρτησης $f = f(\mathbf{x})$ την οποία και τροφοδοτούμε στην α σιγμοειδή συνάρτηση πιθανοφαιίας $\sigma(f(\mathbf{x}))$ για να λάβουμε μια prior επί της πιθανότητας $\pi(\mathbf{x}) \triangleq p(y = +1|\mathbf{x}) = \sigma(f(\mathbf{x}))$. Σημειώσετε ότι η π είναι ντετερμινιστική συνάρτηση της f , και καθώς η f είναι στοχαστική, το ίδιο θα ισχύει και για την π . Έστω, τώρα, $\mathcal{D} = \{(\mathbf{x}_i, y_i) | i = 1, \dots, n\}$ ένα σύνολο δεδομένων, όπου \mathbf{x}_i είναι τα παρατηρήσιμα δεδομένα και y_i οι αντίστοιχες μεταβλητές κλάσεως (ετικέτες). Ορίζουμε επίσης τις ποσότητες $X = [\mathbf{x}_1 \dots \mathbf{x}_n]$ και $\mathbf{y} = [y_1 \dots y_n]^T$. Η ταξινόμηση με χρήση Gaussian processes περιλαμβάνει δυο στάδια: (α) υπολογισμό της (posterior) κατανομής της κρυφής μεταβλητής f_* της αντιστοιχούσας στην περίπτωση δοκιμής \mathbf{x}_* , δοθέντος του συνόλου δεδομένων εκπαίδευσης \mathcal{D}

$$p(f_*|X, \mathbf{y}, \mathbf{x}_*) = \int p(f_*|X, \mathbf{x}_*, \mathbf{f})p(\mathbf{f}|X, \mathbf{y})d\mathbf{f} \quad (5.8)$$

όπου $p(\mathbf{f}|X, \mathbf{y}) = p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|X)/p(\mathbf{y}|X)$ είναι η posterior επί των κρυφών μεταβλητών των αντιστοιχουσών στα δεδομένα εκπαίδευσης, και, (β) την εν συνεχεία χρήση της posterior κατανομής επί της κρυφής μεταβλητής f_* προκειμένου να λάβουμε μια πρόβλεψη ταξινόμησης, δια μέσου της συνάρτησης κατανομής πρόβλεψης π_* , ήτις ορίζεται ως

$$\pi_* \triangleq p(y_* = +1|f_*) = \sigma(f_*) \quad (5.9)$$

Είναι προφανές ότι το ολοκλήρωμα στην (8) είναι μη υπολογίσιμο αναλυτικά. Συνεπώς, προκειμένου να περαιώσουμε το πρόβλημα δυαδικής ταξινόμησης με χρήση του υποτιθέντος GP model, πρέπει να χρησιμοποιήσου-

με είτε αναλυτικές προσεγγίσεις ολοκληρωμάτων, είτε λύσεις βασισμένες σε τεχνικές Monte Carlo [4]. Στην εργασία αυτή, υιοθετούμε¹ την δημοφιλή *Laplace approximation* [17], ήπερ χρησιμοποιεί μια Gaussian προσέγγιση $q(\mathbf{f}|X, \mathbf{y})$ στην posterior $p(\mathbf{f}|X, \mathbf{y})$ στο ολοκλήρωμα (8), επί της βάσει μιας δευτέρας τάξεως ανάπτυξης κατά Taylor της $\log p(\mathbf{f}|X, \mathbf{y})$ γύρω από το μέγιστο της posterior. Έστω \mathbf{k}_* το διάνυσμα των συνδιακυμάνσεων μεταξύ του σημείου δοκιμής \mathbf{x}_* και n σημείων εκπαίδευσης περιλαμβανόντων το X

$$\mathbf{k}_* \triangleq [k(\mathbf{x}_1, \mathbf{x}_*) \dots k(\mathbf{x}_n, \mathbf{x}_*)]^T \quad (5.10)$$

και K ο πίναξ των συνδιακυμάνσεων μεταξύ n δεδομένων εκπαίδευσης, ήτοι

$$K \triangleq \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & k(\mathbf{x}_1, \mathbf{x}_2) \dots & k(\mathbf{x}_1, \mathbf{x}_n) \\ k(\mathbf{x}_2, \mathbf{x}_1) & k(\mathbf{x}_2, \mathbf{x}_2) \dots & k(\mathbf{x}_2, \mathbf{x}_n) \\ \vdots & \vdots & \vdots \\ k(\mathbf{x}_n, \mathbf{x}_1) & k(\mathbf{x}_n, \mathbf{x}_2) \dots & k(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix} \quad (5.11)$$

Τότε, υπό την Laplace προσέγγιση, η προσεγγιστική posterior κατανομή της

¹ Προτιμούμε τις αναλυτικές προσεγγίσεις από τις Monte Carlo τεχνικές λόγω του πολύ καλύτερου υπολογιστικού τους κόστους. Μεταξύ των δημοφιλέστερων αναλυτικών τεχνικών, που είναι η προσέγγιση Laplace και expectation-propagation (EP) [16] αλγόριθμος, προτιμούμε την πρώτη μέθοδο, λόγω της απλουστέρας μορφής του προκύπτοντος αλγορίθμου, σε συνδυασμό με το ότι απαιτούν υπολογιστικά της ίδιας τάξης μεγέθους, κι έχουν βρεθεί σε ένα αρχικό σύνολο πειραμάτων μας να δίνουν την ίδια απόδοση ανανώρισης ομλητή.

κρυφής συνάρτησης f_* του δυαδικού ταξινομητή GP λαμβάνει [17]

$$q(f_*|X, \mathbf{y}, \mathbf{x}_*) = \mathcal{N}(f_*|\mathbb{E}_q[f_*|X, \mathbf{y}, \mathbf{x}_*], \mathbb{V}_q[f_*|X, \mathbf{y}, \mathbf{x}_*]) \quad (5.12)$$

όπου

$$\mathbb{E}_q[f_*|X, \mathbf{y}, \mathbf{x}_*] = \mathbf{k}_*^T K^{-1} \hat{\mathbf{f}} = \mathbf{k}_*^T \nabla \log p(\mathbf{y}|\hat{\mathbf{f}}) \quad (5.13)$$

$$\mathbb{V}_q[f_*|X, \mathbf{y}, \mathbf{x}_*] = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^T (K + W^{-1})^{-1} \mathbf{k}_* \quad (5.14)$$

$$W \triangleq -\nabla^2 \log p(\mathbf{y}|\hat{\mathbf{f}}) \quad (5.15)$$

και η $\hat{\mathbf{f}}$ λαμβάνεται με εφαρμογή της μεθόδου Newton's με επανάληψη οριζόμενη από την

$$\mathbf{f}^{new} = (K^{-1} + W)^{-1} (W \mathbf{f} + \nabla \log p(\mathbf{y}|\mathbf{f})) \quad (5.16)$$

Λαμβάνοντας την posterior κατανομή της κρυφής συνάρτησης f_* του δυαδικού ταξινομητή GP, $q(f_*|X, \mathbf{y}, \mathbf{x}_*)$, μπορούμε πλέον να κάνουμε προβλέψεις για την άγνωστη ετικέτα κλάσεως y_* . Αυτές οι προβλέψεις είναι απλό να γίνουν με χρήση ενός *MAP κανόνα πρόβλεψης*, ός βασίζεται σε μια εκτιμήτρια της προβλεπτικής πιθανότητας, ληφθείσης με χρήση της posterior προσδοκητής τιμής της f_* , διδόμενης από την

$$\hat{\pi}_* = \sigma(\mathbb{E}_q[f_*|X, \mathbf{y}, \mathbf{x}_*]) \quad (5.17)$$

Αν $\hat{\pi}_* > P$, όπου $P = 1/2$, θεωρείται ότι $y_* = +1$, ειδάλως $y_* = -1$. Μια εναλλακτική προσέγγιση, λαμβάνουσα ένα ταυτόσημο αποτέλεσμα δυαδικής ταξινόμησης [18], αλλά παρέχουσα ταυτοχρόνως μια εκτίμηση εμπι-

στοσύνης για την εκάστοτε πρόβλεψη, περιλαμβάνει την αντικατάσταση της MAP εκτιμητορίας $\hat{\pi}_*$ με την προσδοκητή τιμή της π_* , η οποία αναφέρεται συνήθως ως η *μέση προβλεπτική πιθανότητα*, και ορίζεται ως

$$\bar{\pi}_* \triangleq p(y_* = +1|X, \mathbf{y}, \mathbf{x}_*) = \int \sigma(f_*)p(f_*|X, \mathbf{y}, \mathbf{x}_*)df_* \quad (5.18)$$

Δοθέντος του προσεγγιστικού (υπό την προσέγγιση Laplace) posterior μέσου και συνδιακύμανσης της f_* , η έκφρασης της μέσης προβλεπτικής κατανομής $\bar{\pi}_*$ λαμβάνει

$$\bar{\pi}_* \approx \mathbb{E}_q[\pi_*|X, \mathbf{y}, \mathbf{x}_*] = \int \sigma(f_*)q(f_*|X, \mathbf{y}, \mathbf{x}_*)df_* \quad (5.19)$$

Έτσι, με την χρήση της μέσης προβλεπτικής πιθανότητας, οι ταξινομητές GP παρέχουν σημαντικά πλεονεκτήματα σε σχέση με τους SVM ταξινομητές, οι οποίοι περιορίζονται στην παροχή προβλέψεων για τα υπό μελέτη δεδομένα, χωρίς να παρέχουν και μια εκτίμηση εμπιστοσύνης στην πρόβλεψη [5]. Στην παρούσα εργασία, εκμεταλλευόμεθα τα πλεονεκτήματα αυτά των δυαδικών ταξινομητών GP, χρησιμοποιώντας την μέση προβλεπτική πιθανότητα $\bar{\pi}_*$ για τις προβλέψεις, λαμβάνοντας ένα καλύτερο σύνολο απόφασης για την διαδικασία της ταυτοποίησης, σε σχέση με την MAP εκτιμητική, με κατάλληλη επιλογή του κατωφλίου απόφασης. Σημειώνουμε ότι μια παρόμοια προσέγγιση έχει επίσης υιοθετηθεί και στο [3], όπου οι προσδοκητές τιμές των πιθανοτήτων εξόδου ενός HMM χρησιμοποιούνται στην διαδικασία της ταυτοποίησης ομιλητή, αντί των καθαυτό πιθανοτήτων, τοιουτοτρόπως βελτιώνοντας τα σύνορα απόφασης, ήτοι τους ρυθμούς επιτυχίας του ταυτοποιητή ομιλητή.

Τέλος, σχετικά με την συνάρτηση πιθανοφανείας $p(y|f)$ του δυαδικού ταξινομητή GP, υπάρχουν πολλές συναρτησιακές μορφές του, αποδίδουσες την πιθανότητα της κάθε κλάσης στόχου ως συνάρτηση της μεταβλητής f . Εδώ, υιοθετούμε την *αθροιστική Gaussian (probit)* συνάρτηση $\Phi(\cdot)$, λαμβάνοντας

$$p(y|f) = \Phi(yf) \quad (5.20)$$

και, συνεπώς

$$\sigma(f) = \Phi(f) \quad (5.21)$$

ήτις επιτρέπει την λήψη μιας αναλυτικά υπολογίσιμης έκφρασης της μέσης προβλεπτικής πιθανότητας, $\bar{\pi}_*$, δίνοντας [4]

$$\bar{\pi}_* = \Phi(z_*) \quad (5.22)$$

όπου

$$z_* = \frac{\mathbb{E}_q[f_*|X, \mathbf{y}, \mathbf{x}_*]}{\sqrt{1 + \mathbb{V}_q[f_*|X, \mathbf{y}, \mathbf{x}_*]}} \quad (5.23)$$

Οι εκφράσεις των $\nabla \log p(\mathbf{y}|\mathbf{f})$ και $\nabla^2 \log p(\mathbf{y}|\mathbf{f})$ για αυτή την επιλογή πιθανοφανείας δίνονται στο Παράρτημα.

5.5 Ταυτοποίηση Ομιλητή με Χρήση Διακρινοσών Ακολουθίες Gaussian Processes

Μέχρι στιγμής έχουμε υποθέσει ότι η είσοδος \mathbf{x} σε ένα ταξινομητή Gaussian process είναι ένα απλό διάνυσμα, με τις τιμές ενός πλήθους μετρήσιμων χαρακτηριστικών. Ωστόσο, σε πολλά προβλήματα μάθησης, οι είσοδοι ενός

ταξινομητή δεν είναι απλά διανύσματα, αλλά δομημένα αντικείμενα όπως ακολουθίες ή ακόμα και γράφοι. Αυτή είναι η περίπτωση και στην ταυτοποίηση ομιλητή, όπου οι εισοδοί είναι ακολουθιακές εκφράσεις, περιλαμβάνουσες ένα μη σταθερό αριθμό πολυδιάστατων δεδομένων, $\mathbf{x} = \{\mathbf{x}_t\}_{t=1}^T$, όπως εξηγήσαμε και προηγούμενα. Συνεπώς, για να περαιώσουμε την ταυτοποίηση ομιλητή με χρήση ταξινομητών Gaussian process, χρειάζεται να εισάγουμε μια μεθοδολογία απεικόνισης των μεταβλητού μήκους ακολουθιών σε σημεία ένα σταθερής διάστασης χώρου, επί του οποίου και θα χτίσουμε τον ταξινομητή Gaussian process.

Μια κομψή λύση στο πρόβλημα αυτό μπορεί να ληφθεί χρησιμοποιώντας ένα *score-space kernel* ως τη συνάρτηση συνδιακύμανσης του GP ταξινομητή. Οι *score-space kernels* [7, 19], εξαγόμενοι ως μια γενίκευση του *Fisher kernel* [8], επιτρέπουν διακριτικοί ταξινομητές όπως GPs και SVMs να ταξινομούν πλήρεις ακολουθίες, με αξιοποίηση ενός συνόλου παραμετρικών παραγωγικών μοντέλων προκειμένου να λάβουν μια απεικόνιση των μεταβλητού μήκους ακολουθιών σε μεμονωμένα σημεία σε ένα σταθεράς διάστασης χώρο, τον αποκαλούμενο *score-space*. Η απεικόνιση αυτή λαμβάνεται εφαρμόζοντας ένα κατάλληλο τελεστή σε μια συνάρτηση βαθμολογίας πιθανοφανείας (likelihood score) την οριζόμενη επί ενός συνόλου παραγωγικών μοντέλων. Έστω ένα σύνολο από R παραγωγικά μοντέλα με συναρτήσεις πυκνότητας πιθανότητας $p(\mathbf{x}|M_r, \boldsymbol{\theta}_r)$, $r = 1, \dots, R$, όπου $\boldsymbol{\theta}_r$ είναι το σύνολο παραμέτρων του μοντέλου M_r . Η γενική έκφραση της αντιστοίχισης της ακολουθίας $\mathbf{x} = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$ εις ένα *score-space* δίδεται από

$$\phi_{\hat{\mathcal{G}}, \mathcal{L}}(\mathbf{x}) = \hat{\mathcal{G}}[\mathcal{L}(\{p(\mathbf{x}|M_r, \boldsymbol{\theta}_r)|r \in R\})] \quad (5.24)$$

όπου $\phi_{\hat{\mathcal{G}}, \mathcal{L}}(\mathbf{x})$ είναι το *score-vector* όπου αντιστοιχίζεται η δοσμένη ακολουθία \mathbf{x} , $\mathcal{L}(\{p(\mathbf{x}|M_r, \boldsymbol{\theta}_r)|r \in R\})$ είναι η *score function* η εφαρμοζόμενη επί των πιθανοφανειών του θεωρηθέντος συνόλου παραγωγικών μοντέλων, καλουμένη ως το *score-argument*, και $\hat{\mathcal{G}}$ είναι ο *score-operator* που απεικονίζει το βαθμωτό *score-argument* στο *score-space*, λαμβάνοντας το *score-vector* της ακολουθίας \mathbf{x} .

Ως συνεξηγήθη στο [4], σχεδόν κάθε συνάρτηση θα μπορούσε να χρησιμοποιηθεί στην (24) ως το *score-argument*, $\mathcal{L}(\cdot)$, και αυτή η επιλογή εξαρτάται μόνο από την εκάστοτε εφαρμογή. Εδώ, επί τη βάση των προηγουμένως παρουσιασθέντων αποτελεσμάτων σχετικά με τις κυρίαρχες σήμερα τεχνικές βάσει κειμένου και ανεξαρτήτως κειμένου ταυτοποίησης ομιλητή, χρησιμοποιούμε το *log likelihood-ratio* δυο ανταγωνιστικών μοντέλων, δηλ. ενός μοντέλου πελάτη, M , και ενός *οικουμενικού* μοντέλου, Ω , ως προς μια δοσμένη ακολουθία. Αναφορικά με τα θεωρούμενα παραγωγικά μοντέλα, χρησιμοποιούμε GMMs για την ανεξάρτητη κειμένου ταυτοποίηση, και HMMs για την βάσει κειμένου ταυτοποίηση. Τέλος, σχετικά με την επιλογή του *score-operator*, $\hat{\mathcal{G}}$, κατάλληλες εναλλακτικές περιλαμβάνουν τον τελεστή πρώτης παραγώγου, $\hat{\mathcal{G}} = \nabla_{\boldsymbol{\theta}}$, και τον τελεστή πρώτης παραγώγου και ορίσματος, $\hat{\mathcal{G}} = [\nabla_{\boldsymbol{\theta}}, 1]^T$, όπου $\boldsymbol{\theta} = \{\boldsymbol{\theta}_M, \boldsymbol{\theta}_\Omega\}$ [20]. Αμφότερες αυτές οι εναλλακτικές θα αξιολογηθούν στο πειραματικό μέρος του ενθάδε κεφαλαίου.

Ένα σημαντικό ζήτημα σχετικά με την *score-space* απεικόνιση (24) αναφέρεται στην γενικά μη Ευκλείδεια φύση του προκύπτοντος χώρου απεικόνισης ως έχει, π.χ., αποδειχθεί στις εργασίες [20] και [21]. Είναι προφανές ότι το γεγονός αυτό επιβάλλει την εφαρμογή μιας διαδικασίας κανο-

νικοποίησης (whitening) ώστε να λάβουμε εν τέλει έναν Ευκλείδειο χώρο. Euclidean one. Ειδικότερα, θεωρούμε τον μετασχηματισμό [22]

$$\psi_{\hat{g},\mathcal{L}}(\mathbf{x}) = \Lambda^{-1}\phi_{\hat{g},\mathcal{L}}(\mathbf{x}) \quad (5.25)$$

όπου

$$\Lambda = \mathbb{E} \left\{ \left(\phi_{\hat{g},\mathcal{L}}(\mathbf{x}) - \mathbb{E} \left[\phi_{\hat{g},\mathcal{L}}(\mathbf{x}) \right] \right) \left(\phi_{\hat{g},\mathcal{L}}(\mathbf{x}) - \mathbb{E} \left[\phi_{\hat{g},\mathcal{L}}(\mathbf{x}) \right] \right)^T \right\} \quad (5.26)$$

Υπό τον μετασχηματισμό αυτό, οι συνιστώσες των score-vectors κανονικοποιούνται σε μηδενικό μέσο και μοναδιαία διακύμανση (ήτοι, τα διανύσματα βάσης του score-space απεικονίζονται σε ένα ορθοκανονικό χώρο). Ωστόσο, είναι φανερό ότι ο μετασχηματισμός (26) περιλαμβάνει την αντιστροφή ενός πίνακα Λ και άρα μπορεί να επάγει ανεπιθυμήτως σημαντικά υπολογιστικά κόστη. Σε τέτοιες περιπτώσεις, κανείς μπορεί να καταφύγει σε μια διαγώνια ή μπλοκ διαγώνια προσέγγιση του Λ (δες, π.χ., [23]).

Χρησιμοποιώντας τους ληφθέντες score-vectors $\psi_{\hat{g},\mathcal{L}}(\mathbf{x})$ των παρατηρούμενων πολυδιαστάτων ακολουθιών, ένας κοινότυπος score-space kernel μπορεί να εξαχθεί με χρήση μιας θεώρησης εσωτερικού γινομένου, ως [20]

$$k_{\text{dot}}(\mathbf{x}, \mathbf{x}') = \psi_{\hat{g},\mathcal{L}}^T(\mathbf{x})\psi_{\hat{g},\mathcal{L}}(\mathbf{x}') \quad (5.27)$$

Ωστόσο, και άλλες μορφές score-space kernels θα μπορούσαν επίσης να θεωρηθούν (δες, π.χ., [4]). Για λόγους επίδειξης, εμείς θα θεωρήσουμε και την τύπου νευρωνικού δικτύου kernel function, η οποία αποδεικνύεται ότι λαμβάνεται από ένα νευρωνικό ενός κρυφού επιπέδου, στο όριο καθώς το

πλήθος κρυφών κόμβων τείνει στο άπειρο [24--26]. Ισχύει

$$k_{\text{NN}}(\psi_{\hat{\mathcal{G}},\mathcal{L}}(\mathbf{x}), \psi_{\hat{\mathcal{G}},\mathcal{L}}(\mathbf{x}')) = \frac{2}{\pi} \sin^{-1} \left(\frac{2\tilde{\psi}_{\hat{\mathcal{G}},\mathcal{L}}(\mathbf{x})^T \Xi \tilde{\psi}_{\hat{\mathcal{G}},\mathcal{L}}(\mathbf{x}')}{\sqrt{(1 + 2\tilde{\psi}_{\hat{\mathcal{G}},\mathcal{L}}(\mathbf{x})^T \Xi \tilde{\psi}_{\hat{\mathcal{G}},\mathcal{L}}(\mathbf{x}))(2\tilde{\psi}_{\hat{\mathcal{G}},\mathcal{L}}(\mathbf{x}')^T \Xi \tilde{\psi}_{\hat{\mathcal{G}},\mathcal{L}}(\mathbf{x}'))}} \right) \quad (5.28)$$

όπου $\Xi = \text{diag}(\rho_0^2, \rho^2, \dots, \rho^2)$, και $\tilde{\psi}_{\hat{\mathcal{G}},\mathcal{L}}(\mathbf{x}) \triangleq [1, \psi_{\hat{\mathcal{G}},\mathcal{L}}(\mathbf{x})^T]^T$. Αμφότερες οι εν λόγω εναλλακτικές, k_{dot} και k_{NN} , θα αξιολογηθούν στο πειραματικό μέρος αυτού του κεφαλαίου.

Στην συνέχεια δίνουμε τις αναλυτικές εκφράσεις των score-arguments, $\mathcal{L}(p(\mathbf{x}|M, \boldsymbol{\theta}_M), p(\mathbf{x}|\Omega, \boldsymbol{\theta}_\Omega))$, και των προ την κανονικοποίηση score-vectors, $\phi_{\hat{\mathcal{G}},\mathcal{L}}(\mathbf{x})$, για τις ποικίλες μορφές του score-operator, $\hat{\mathcal{G}}$, για αμφότερα τα προβλήματα της βάσει κειμένου και της ανεξάρτητης κειμένου ταυτοποίησης.

5.5.1 Ανεξάρτητη Κειμένου Ταυτοποίηση

Όπως ήδη ανεπτύξαμε, για τους σκοπούς της ανεξάρτητης κειμένου ταυτοποίησης ομιλητή με χρήση του διακριτικού ακολουθιών Gaussian process ταξινομητή, χρησιμοποιούμε δυο ανταγωνιστικά διαγώνια GMMs, ένα μοντέλο πελάτη, M , και ένα οικουμενικό μοντέλο, Ω , εκπαιδευόμενα ως περιγράφεται στην Ενότητα 5.3.1 για λήψη των score-vectors. Έτσι, δοθείσης της ακολουθίας (έκφρασης) $\mathbf{x} = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$, υπολογίζουμε την αντίστοιχη βαθμολογία ακολουθίας (utterance score) $S(\mathbf{x})$, οριζόμενη στην (3), και το

χρησιμοποιούμε ως το score-argument του ταξινομητή μας. Έχουμε

$$\begin{aligned}
& \mathcal{L}(\{p(\mathbf{x}|M_r, \boldsymbol{\theta}_r) | r \in \{M, \Omega\}\}) \\
&= \frac{1}{T} \sum_{t=1}^T \log p(\mathbf{x}_t | M, \boldsymbol{\theta}_M) - \frac{1}{T} \sum_{t=1}^T \log p(\mathbf{x}_t | \Omega, \boldsymbol{\theta}_\Omega) \quad (5.29) \\
&= \frac{1}{T} \log p(\mathbf{x} | M, \boldsymbol{\theta}_M) - \frac{1}{T} \log p(\mathbf{x} | \Omega, \boldsymbol{\theta}_\Omega)
\end{aligned}$$

Έτσι, θέτοντας τον τελεστή πρώτης παραγώγου ως τον χρησιμοποιούμενο score-operator, $\hat{\mathcal{G}} = \nabla_{\boldsymbol{\theta}}$, λαμβάνουμε το score-vector

$$\begin{aligned}
\phi(\mathbf{x}) &= \nabla_{\boldsymbol{\theta}} \left(\frac{1}{T} \log p(\mathbf{x} | M, \boldsymbol{\theta}_M) - \frac{1}{T} \log p(\mathbf{x} | \Omega, \boldsymbol{\theta}_\Omega) \right) \\
&= \begin{bmatrix} \frac{1}{T} \sum_{t=1}^T \nabla_{\boldsymbol{\theta}_M} \log p(\mathbf{x}_t | M, \boldsymbol{\theta}_M) \\ -\frac{1}{T} \sum_{t=1}^T \nabla_{\boldsymbol{\theta}_\Omega} \log p(\mathbf{x}_t | \Omega, \boldsymbol{\theta}_\Omega) \end{bmatrix} \quad (5.30)
\end{aligned}$$

ενώ, θεωρώντας τον τελεστή πρώτης παραγώγου και ορίσματος, $\hat{\mathcal{G}} = [\nabla_{\boldsymbol{\theta}}, 1]^T$, έχουμε

$$\phi_+(\mathbf{x}) = \begin{bmatrix} \phi(\mathbf{x}) \\ \frac{1}{T} \log p(\mathbf{x} | M, \boldsymbol{\theta}_M) - \frac{1}{T} \log p(\mathbf{x} | \Omega, \boldsymbol{\theta}_\Omega) \end{bmatrix} \quad (5.31)$$

Η έκφραση του $\nabla_{\boldsymbol{\theta}_H} \log p(\mathbf{x}_t | H, \boldsymbol{\theta}_H)$ για ένα g -συνιστωσών διαγώνιο GMM H με εισόδους από ένα d -διάστατο χώρο παρατήρησης \mathcal{X} , και με σύνολο παραμέτρων $\boldsymbol{\theta}_H = \{w_j, \boldsymbol{\mu}_j^l, \boldsymbol{\Sigma}_j^l | j = 1, \dots, g, l = 1, \dots, d\}$, όπου $\boldsymbol{\mu}_j = [\mu_j^l]$ και $\boldsymbol{\Sigma}_j = \text{diag}(\Sigma_j^1, \dots, \Sigma_j^d)$, δίνονται στο Παράρτημα.

5.5.2 Βάσει Κειμένου Ταυτοποίηση

Ακολουθώντας την συζήτηση της ενότητας 5.3.2, για την διεξαγωγή της βάσει κειμένου ταυτοποίησης με χρήση των διακριτικών ακολουθιών Gaussian process δυαδικών ταξινομητών, τα score-vectors ευρίσκονται με χρήση ενός παραγωγικού μοντέλου πελάτη ομιλητή, M , και ενός οικουμενικού μοντέλου ομιλητή, Ω , που τώρα επιλέγονται να είναι αριστερά-προς-δεξιά hidden Markov models χωρίς υπερβάσεις καταστάσεων με διαγώνιους πίνακες συνδιακύμανσης (εξ. (4)). Συνεπώς, δοθείσης μιας ακολουθίας $\mathbf{x} = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$, το αντίστοιχο score $S(\mathbf{x})$, υπολογίζεται από την (5), και χρησιμοποιείται ως το score-argument του ταξινομητή μας. Με άλλα λόγια

$$\mathcal{L}(\{p(\mathbf{x}|M_r, \boldsymbol{\theta}_r)|r \in \{M, \Omega\}\}) = \log p(\mathbf{x}|M, \boldsymbol{\theta}_M) - \log p(\mathbf{x}|\Omega, \boldsymbol{\theta}_\Omega) \quad (5.32)$$

Από την (32), έχουμε ότι θέτοντας τον τελεστή πρώτης παραγώγου ως το score-operator του ταξινομητή GP, $\hat{\mathcal{G}} = \nabla_{\boldsymbol{\theta}}$, λαμβάνουμε το score-vector

$$\phi(\mathbf{x}) = \begin{bmatrix} \nabla_{\boldsymbol{\theta}_M} \log p(\mathbf{x}|M, \boldsymbol{\theta}_M) \\ -\nabla_{\boldsymbol{\theta}_\Omega} \log p(\mathbf{x}|\Omega, \boldsymbol{\theta}_\Omega) \end{bmatrix} \quad (5.33)$$

ενώ, θεωρώντας τον τελεστή πρώτης παραγώγου και ορίσματος, $\hat{\mathcal{G}} = [\nabla_{\boldsymbol{\theta}}, 1]^T$, έχουμε

$$\phi_+(\mathbf{x}) = \begin{bmatrix} \phi(\mathbf{x}) \\ \log p(\mathbf{x}|M, \boldsymbol{\theta}_M) - \log p(\mathbf{x}|\Omega, \boldsymbol{\theta}_\Omega) \end{bmatrix} \quad (5.34)$$

Η έκφραση του $\nabla_{\boldsymbol{\theta}_H} \log p(\mathbf{x}|H, \boldsymbol{\theta}_H)$ για ένα v -καταστάσεων, αριστερά-προς-δεξιά HMM H άνευ υπερβάσεων καταστάσεων και με χρήση διαγώνιων, g -συνιστωσών, GMMs οριζόμενων επί ενός d -διάστατου χώρου, με $\boldsymbol{\theta}_H =$

$\{a_{h\xi}, w_{hj}, \mu_{hj}^l, \Sigma_{hj}^l | h, = 1, \dots, v, h \leq \xi \leq h + 1, j = 1, \dots, g, l = 1, \dots, d\}$, όπου $\mu_j = [\mu_j^l]_l$ και $\Sigma_j = \text{diag}(\Sigma_j^1, \dots, \Sigma_j^d)$, δίνονται στο Παράρτημα.

5.6 Πειράματα

5.6.1 Ανεξάρτητη Κειμένου Ταυτοποίηση

Η προτεινόμενη μεθοδολογία ανεξάρτητης κειμένου ταυτοποίησης ομιλητή αξιολογείται με χρήση της συλλογής PolyVar [9]. Η PolyVar περιλαμβάνει 38 ομιλητές πελάτες, 24 άνδρες και 14 γυναίκες, καταγεγραμμένους μέσω ενός τηλεφωνικού δικτύου. 85 ακολουθίες κατεγράφησαν από καθένα σε 5 συνόδους, με 17 ακολουθίες ανά σύνοδο. Υπάρχουν επίσης 952 ξένες ακολουθίες από 56 άγνωστους ομιλητές, καθείς συνεισφέρων 17 ακολουθίες ανά σύνοδο. Οι λήψεις έγιναν μέσω τηλεφώνου, άρα, τα δεδομένα περιέχουν ενδογενώς θόρυβο. Οι προφερόμενες φράσεις περιλαμβάνουν 5 επαναλήψεις 17 ουσιαστικών και καθημερινών εκφράσεων Γαλλικής, με 3 ως 12 φθόγγους. Η PolyVar παρέχει ένα αυστηρό πρωτόκολλο καθορισμού των ακολουθιών που θα περιλαμβάνονται σε καθένα των συνόλων εκπαίδευσης και αξιολόγησης της εκάστοτε εφαρμοζόμενης μεθοδολογίας.

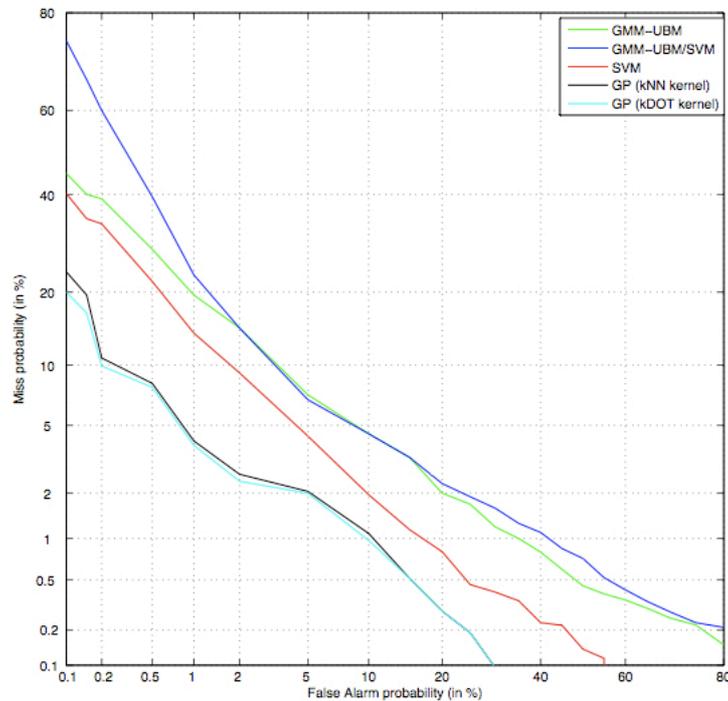
Στα πειράματά μας, τα σήματα ομιλίας αναπαρίστανται με την χρήση των συντελεστών 12ου-βαθμού perceptual linear prediction (PLP) cepstral coefficients, οι οποίοι υπολογίζονται με χρήση ενός 32-ms παραθύρου με 10-ms ολίσθηση. Επιπλέον, η ενέργεια του σήματος, καθώς και οι πρώτης και δεύτερης τάξης παράγωγοι υπολογίζονται, αποδίδοντας feature vectors με συνολικά 39 συνιστώσες. Περαιτέρω, cepstral mean subtraction χρησιμοποιήθηκε για να αφαιρεθεί η επίδραση του διαύλου επικοινωνίας. Εκτός του

Πίνακας 5.1: Ανεξάρτητη Κειμένου Ταυτοποίηση

Μοντέλο	% EER	% min HTER
GMM-UBM	6.12	5.53
GMM-UBM/SVM	5.94	5.37
SVM / first derivative k_{dot} kernel	5.55	5.13
SVM / first derivative plus argument k_{dot} kernel	5.55	5.03
GP / first derivative k_{dot} kernel	3.32	2.95
GP / first derivative k_{NN} kernel	3.66	3.15
GP / first derivative plus argument k_{dot} kernel	3.32	2.95
GP / first derivative plus argument k_{NN} kernel	3.64	3.15

προτεινομένου μοντέλου, αξιολογούμε επίσης τα συστήματα GMM-UBM (3), ένα υβριδικό GMM-UBM/SVM όπου ο λόγος πιθανοφανειών παραμετροποιείται με χρήση ενός SVM για εκτιμητική των παραμέτρων [27], και το βασισμένο σε GMM διακριτικό ακολουθιών SVM [23]. Το βέλτιστο μέγεθος των GMMs (ήτοι, ο αριθμός των συνιστωσών τους κατανομών) ηυρέθη με χρήση cross-validation, όπως και στα [27] και [23], λαμβάνοντας ένα οικουμενικό μοντέλο με 1000 συνιστώσες κατανομές, και μοντέλα πελατών με 200 συνιστώσες κατανομές, ακριβώς όπως ελήφθη και στα [27] και [23]. Οι SVM- και GP-βασισμένες μεθοδολογίες αξιολογήθηκαν με χρήση αμφοτέρων των score-operators πρώτης παραγώγου και πρώτης παραγώγου και ορίσματος. Τα χρησιμοποιούμενα SVMs ήταν γραμμικά όπως και στο [23]. Η προτεινόμενη μέθοδος αξιολογήθηκε με χρήση αμφοτέρων των πυρήνων k_{dot} και k_{NN} . Οι υπερπαραμέτροι το πυρήνα k_{NN} ηυρέθησαν με βελτιστοποίηση του log marginal likelihood του μοντέλου [4].

Στον Πίνακα 1, παρέχουμε τον ρυθμό σφάλματος ισότητας (equal error rate, EER), και το ελάχιστο ήμισυ συνολικού ρυθμού σφάλματος (minimum half total error rate, min HTER) για τα θεωρηθέντα μοντέλα. Επισημαίνε-



Σχήμα 5.1: Ανεξάρτητη Κειμένου Ταυτοποίηση: Ληφθήσες DET καμπύλες

ται ότι τα αποτελέσματα αυτά ελήφθησαν χρησιμοποιώντας ένα σταθερό κατώφλι απόφασης για όλους τους ομιλητές. Παρατηρούμε ότι τα προτεινόμενα μοντέλα δίνουν HTER μέχρι και 46.65% καλύτερο από το GMM-UBM model, και μέχρι και 42.5% καλύτερο από τα SVM-τύπου μοντέλα. Παρατηρούμε επίσης ότι ο πυρήνας k_{NN} είχε χειρότερη επίδοση από τον πυρήνα εσωτερικού γινομένου, τόσο σε σχέση με το EER όσο και με το min HTER. Τέλος, στο Σχ. 1, δίδονται οι ληφθήσες DET καμπύλες των διαφόρων μοντέλων (θεωρώντας score-operators πρώτης παραγώγου).

5.6.2 Βάσει Κειμένου Ταυτοποίηση

Για να παρέχουμε μια αξιολόγηση της προτεινόμενης βάσει κειμένου μεθοδολογίας ταυτοποίησης ομιλητή, χρησιμοποιούμε την συλλογή YOHO [10]. Η YOHO περιλαμβάνει ένα λεξιλόγιο αποτελούμενο από διψήφιους αριθμούς συνεχώς προφερόμενους σε σύνολα των τριών (π.χ., “62-31-53” ή “sixty-two thirty-one fifty-three”). Διαιρείται σε τμήματα εκπαίδευσης (*enrollment*) και αξιολόγησης μοντέλου (*verification*). Κάθε τμήμα περιλαμβάνει δεδομένα από 138 ομιλητές (108 άνδρες και 30 γυναίκες). Υπάρχουν 4 enrollment σύνοδοι ανά ομιλητή και κάθε σύνοδος περιλαμβάνει 24 ακολουθίες, ενώ υπάρχουν και 10 verification σύνοδοι ανά ομιλητή και κάθε σύνοδος περιλαμβάνει 4 ακολουθίες. Από τους 138 ομιλητές, στο πείραμά μας, οι 41 εξ αυτών χρησιμοποιούνται ως ξένοι - άγνωστοι, και οι λοιποί ως γνωστοί - πελάτες. Τα σήματα ομιλίας μοντελοποιούνται με 39-διάστατα διανύσματα, όπως και στο προηγούμενο πείραμα.

Η επίδοση του μοντέλου μας συγκρίνεται με το HMM-UBM σύστημα (5), και το HMM-βασισμένο διακριτικό ακολουθιών SVM που προετάθη στο [7]. Κάθε διψήφιος αριθμός του διαθέσιμου λεξικού διαιρείται σε δυο λέξεις: μια δεκάδα (“30,” “40,” “50,” “60,” “70,” “80” ή “90”) και μια μονάδα (“1,” “2,” “3,” “4,” “5,” “6,” “7,” “8,” ή “9”). Κατά συνέπεια, έχουμε συνολικά 16 λέξεις. Στο πείραμά μας, κάθε λέξη αναπαρίσταται με ένα αριστερά-προς-δεξιά HMM, χωρίς υπερβάσεις καταστάσεων, με 8 καταστάσεις και μια απλή Gaussian κατανομή εκπομπής ανά κατάσταση, με διαγώνιο πίνακα συνδιακύμανσης. Τα HMM μοντέλα πελατών (16 ανά ομιλητή) εκτιμούνται με χρήση 96 ακολουθιών (4 enrollment sessions and 24 utterances per session), ενώ τα 16 οικουμενικά HMMs (ένα ανά λέξη) εκπαιδεύονται με χρήση 3936

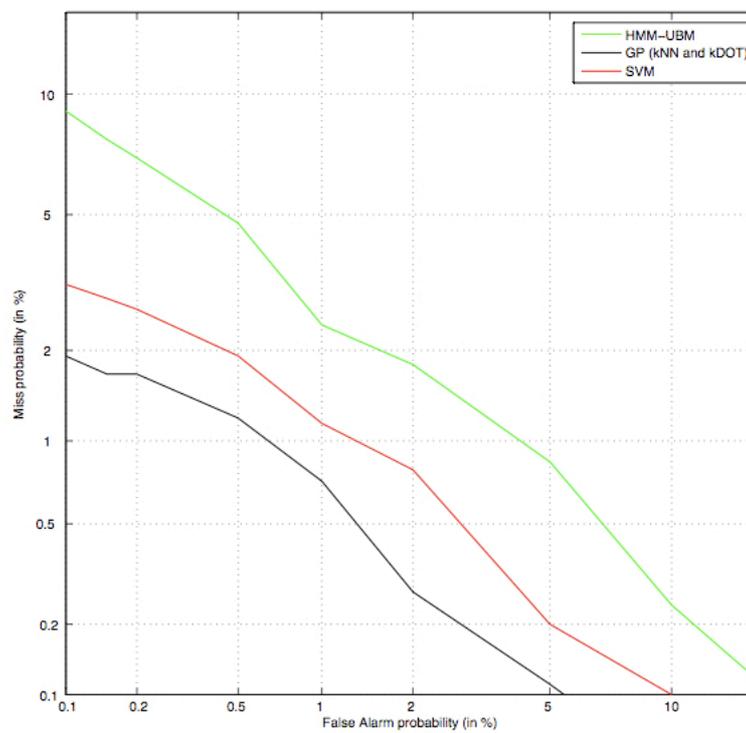
Πίνακας 5.2: Βάσει Κειμένου Ταυτοποίηση

Μοντέλο	EER	% min HTER
HMM-UBM	1.44	1.35
SVM / first derivative k_{dot} kernel	0.96	0.88
SVM / first derivative plus argument k_{dot} kernel	0.96	0.86
GP / first derivative k_{dot} kernel	0.60	0.55
GP / first derivative k_{NN} kernel	0.60	0.55
GP / first derivative plus argument k_{dot} kernel	0.60	0.55
GP / first derivative plus argument k_{NN} kernel	0.60	0.55

ακολουθιών (41 ομιλητές και 96 ακολουθίες εκπαίδευσης ανά ομιλητή).

Οι SVM- και GP-βασισμένες μεθοδολογίες αξιολογήθηκαν με χρήση αμφοτέρων των score-operators πρώτης παραγώγου και πρώτης παραγώγου και ορίσματος. Τα χρησιμοποιούμενα SVMs ήταν γραμμικά όπως και στο [23]. Η προτεινόμενη μέθοδος αξιολογήθηκε με χρήση αμφοτέρων των πυρήνων k_{dot} και k_{NN} . Οι υπερπαραμέτροι το πυρήνα k_{NN} ηυρέθησαν με βελτιστοποίηση του log marginal likelihood του μοντέλου [4], όπως και στο προηγούμενο πείραμα. Στο Πίνακα 2, δίδουμε τις μετρικές EER και min HTER για τα ποικίλα μοντέλα. Επισημαίνεται ότι τα αποτελέσματα αυτά ελήφθησαν χρησιμοποιώντας ένα σταθερό κατώφλι απόφασης για όλους τους ομιλητές. Παρατηρούμε ότι τα προτεινόμενα μοντέλα δίνουν HTER μέχρι και 59.26% καλύτερο από το HMM-UBM model, και μέχρι και 37.5% καλύτερο από τα SVM-τύπου μοντέλα. Παρατηρούμε επίσης ότι η επίδοση του προτεινόμενου μοντέλου δεν άλλαξε αλλάζοντας τον kernel ή το score-operator. Τέλος, στο Σχ. 2, δίδονται οι ληφθήσες DET καμπύλες των διαφόρων μοντέλων (θεωρώντας score-operators πρώτης παραγώγου).

Σχετικά με τις εκφράσεις των $\nabla \log p(\mathbf{y}|\mathbf{f})$ και $\nabla^2 \log p(\mathbf{y}|\mathbf{f})$ για την επι-



Σχήμα 5.2: Βάσει Κειμένου Ταυτοποίηση: Ληφθήσες DET καμπύλες

λεχθείσα σιγμοειδή πιθανοφάνεια, έχουμε

$$\nabla \log p(\mathbf{y}|\mathbf{f}) = \left[\frac{\partial}{\partial f_1} \log p(y_1|f_1) \dots \frac{\partial}{\partial f_n} \log p(y_n|f_n) \right]^T \quad (35)$$

όπου

$$\frac{\partial}{\partial f_i} \log p(y_i|f_i) = \frac{y_i \mathcal{N}(f_i)}{\Phi(y_i, f_i)} \quad (36)$$

$\mathcal{N}(f_i)$ είναι η σ.π.π. της τυποποιημένης Gaussian κατανομής, και

$$\nabla^2 \log p(\mathbf{y}|\mathbf{f}) = \begin{bmatrix} \frac{\partial^2}{\partial f_1^2} \log p(y_1|f_1) & 0 & \dots & 0 \\ 0 & \frac{\partial^2}{\partial f_2^2} \log p(y_2|f_2) & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & \frac{\partial^2}{\partial f_n^2} \log p(y_n|f_n) \end{bmatrix} \quad (37)$$

όπου

$$\frac{\partial^2}{\partial f_i^2} \log p(y_i|f_i) = -\frac{\mathcal{N}(f_i)^2}{\Phi(y_i, f_i)^2} - \frac{y_i f_i \mathcal{N}(f_i)}{\Phi(y_i, f_i)} \quad (38)$$

Σχετικά με την έκφραση της $\nabla_{\boldsymbol{\theta}_H} \log p(\mathbf{x}_t|H, \boldsymbol{\theta}_H)$ για διαγώνια GMMs, με χρήση της (1) έχουμε

$$\nabla_{\boldsymbol{\theta}_H} \log p(\mathbf{x}_t|H, \boldsymbol{\theta}_H) = \left[\frac{\partial}{\partial w_{j^*}}, \dots, \frac{\partial}{\partial \mu_{j^*}^*}, \dots, \frac{\partial}{\partial \Sigma_{j^*}^*} \right]^T \quad (39)$$

$$\times \log p(\mathbf{x}_t|H, \boldsymbol{\theta}_H)$$

όπου

$$\frac{\partial}{\partial w_{j^*}} \log p(\mathbf{x}_t|H, \boldsymbol{\theta}_H) = \frac{\mathcal{N}(t, j^*)}{\sum_{j=1}^g w_j \mathcal{N}(t, j)} \quad (40)$$

$$\frac{\partial}{\partial \mu_{j^*}^{l^*}} \log p(\mathbf{x}_t | H, \boldsymbol{\theta}_H) = \frac{w_{j^*} \mathcal{N}(t, j^*)}{\sum_{j=1}^g w_j \mathcal{N}(t, j)} \frac{x_t^{l^*} - \mu_{j^*}^{l^*}}{\Sigma_{j^*}^{l^*}} \quad (41)$$

$$\begin{aligned} & \frac{\partial}{\partial \Sigma_{j^*}^{l^*}} \log p(\mathbf{x}_t | H, \boldsymbol{\theta}_H) \\ &= \frac{w_{j^*} \mathcal{N}(t, j^*)}{2 \sum_{j=1}^g w_j \mathcal{N}(t, j)} \left\{ \left(\frac{x_t^{l^*} - \mu_{j^*}^{l^*}}{\Sigma_{j^*}^{l^*}} \right)^2 - \frac{1}{\Sigma_{j^*}^{l^*}} \right\} \end{aligned} \quad (42)$$

και $\mathcal{N}(t, j)$ είναι η πιθανοφάνεια της j οστης συνιστώσας κατανομής του μοντέλου H ως προς το δεδομένα \mathbf{x}_t , ήτοι

$$\mathcal{N}(t, j) \triangleq \mathcal{N}(\mathbf{x}_t | \boldsymbol{\mu}_j, \Sigma_j) \quad (43)$$

Τέλος, σχετικά με την έκφραση της $\nabla_{\boldsymbol{\theta}_H} \log p(\mathbf{x} | H, \boldsymbol{\theta}_H)$ για τα του θεωρηθέντος τύπου HMMs, αγνοώντας τις priors $a_{h,h+1}$ δια της ιδιότητας

$$a_{h,h+1} = 1 - a_{hh}$$

έχουμε [20]

$$\begin{aligned} \nabla_{\boldsymbol{\theta}_H} \log p(\mathbf{x} | H, \boldsymbol{\theta}_H) &= \left[\frac{\partial}{\partial a_{h^*h^*}}, \dots, \frac{\partial}{\partial w_{h^*j^*}}, \dots, \frac{\partial}{\partial \mu_{h^*j^*}^{l^*}}, \dots, \frac{\partial}{\partial \Sigma_{h^*j^*}^{l^*}} \right]^T \\ &\times \log p(\mathbf{x} | H, \boldsymbol{\theta}_H) \end{aligned} \quad (44)$$

όπου

$$\frac{\partial}{\partial a_{h^*h^*}} \log p(\mathbf{x} | H, \boldsymbol{\theta}_H) = \sum_{t=1}^T \left\{ \frac{\gamma_{h^*}(t)}{a_{h^*h^*}} - \frac{1}{T a_{h^*h^*} (1 - a_{h^*h^*})} \right\} \quad (45)$$

$$\frac{\partial}{\partial w_{h^*j^*}} \log p(\mathbf{x}|H, \boldsymbol{\theta}_H) = \sum_{t=1}^T \gamma_{h^*j^*}(t) \left\{ \frac{1}{w_{h^*j^*}} - \frac{\gamma_{h1}(t)}{w_{h^*1} \gamma_{h^*j^*}(t)} \right\} \quad (46)$$

$$\frac{\partial}{\partial \mu_{h^*j^*}^{l^*}} \log p(\mathbf{x}|H, \boldsymbol{\theta}_H) = \sum_{t=1}^T \gamma_{h^*j^*}(t) \frac{x_t^{l^*} - \mu_{h^*j^*}^{l^*}}{\Sigma_{h^*j^*}^{l^*}} \quad (47)$$

$$\frac{\partial}{\partial \Sigma_{h^*j^*}^{l^*}} \log p(\mathbf{x}|H, \boldsymbol{\theta}_H) = \sum_{t=1}^T \frac{1}{2} \gamma_{h^*j^*}(t) \left\{ \left(\frac{x_t^{l^*} - \mu_{h^*j^*}^{l^*}}{\Sigma_{h^*j^*}^{l^*}} \right)^2 - \frac{1}{\Sigma_{h^*j^*}^{l^*}} \right\} \quad (48)$$

$$\gamma_{h^*j^*}(t) = \tilde{\gamma}_{h^*}(t) \frac{w_{h^*j^*} \mathcal{N}(\mathbf{x}_t | \boldsymbol{\mu}_{h^*j^*}, \Sigma_{h^*j^*})}{\sum_{j=1}^g w_{h^*j} \mathcal{N}(\mathbf{x}_t | \boldsymbol{\mu}_{h^*j}, \Sigma_{h^*j})} \quad (49)$$

$\tilde{\gamma}_h(t)$ είναι η κανονικοποιημένη posterior πιθανότητα ρίψης της παρατήρησης \mathbf{x}_t από την *hoστη* κατάσταση του μοντέλου

$$\tilde{\gamma}_h(t) = \frac{\gamma_h(t)}{\sum_{\tau=1}^T \gamma_h(\tau)} \quad (50)$$

και η έκφραση των $\gamma_h(t)$ μπορεί να βρεθεί στο Κεφ. 4. Σημειώνεται ότι η εν λόγω κανονικοποίηση, η αποδίδουσα τις ποσότητες $\tilde{\gamma}_h(t)$ έναντι των $\gamma_h(t)$, έγινε υποκινούμενη από ενθαρρυντικά αποτελέσματα από την σχετική βιβλιογραφία επεξεργασίας σήματος (δες, π.χ., [7]).

Βιβλιογραφία

- [1] G. R. Doddington, M. A. Przybocki, A. F. Martin, and D. A. Reynolds, "The NIST speaker recognition evaluation - overview, methodology, systems, results, perspective," *Speech Communication*, vol. 31, no. 2-3, pp. 225--254, 2000.
- [2] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, pp. 19--41, 2000.
- [3] N. B. Yoma and M. Villar, "Speaker verification in noise using a stochastic version of the weighted Viterbi algorithm," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 3, pp. 158--166, 2002.
- [4] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- [5] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
- [6] M. Seeger, "Bayesian Gaussian process models: PAC-Bayesian generalisation error bounds and sparse approximations," Ph.D. dissertation, School of Informatics, University of Edinburgh, 2003.

- [7] N. Smith and M. Gales, "Speech recognition using SVMs," in *Advances in Neural Information Processing Systems*, vol. 14. Cambridge, MA: MIT Press, 2002.
- [8] T. S. Jaakola, M. Diekhans, and D. Haussler, "A discriminative framework for detecting remote protein homologies," *Journal of Computational Biology*, vol. 7, pp. 95--114, 2000.
- [9] G. Chollet, J. Cochard, A. Constantinescu, C. Jaboulet, and P. Langlais, "Swiss French polyphone and PolyVar: Telephone speech databases to model inter- and intra-speaker variability," in *Linguistic Databases*, J. Nerbonne, Ed., 1997, pp. 117--135.
- [10] J. P. Campbell, "Testing with the YOHO CD-ROM voice verification corpus," in *ICASSP*, vol. 1, 1995, pp. 341--344.
- [11] G. McLachlan and D. Peel, *Finite Mixture Models*. New York: Wiley Series in Probability and Statistics, 2000.
- [12] X. Meng and D. van Dyk, "The EM algorithm--an old folk song sung to a fast new tune (with discussion)," *Journal of the Royal Statistical Society B*, vol. 59, no. 3, pp. 511--567, 1997.
- [13] J. L. Gauvain and C.-H. Lee, "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains," *IEEE Trans. Speech Audio Process.*, vol. 2, pp. 291--298, 1994.
- [14] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, pp. 245--255, 1989.

- [15] S. Young, "A review of large-vocabulary continuous-speech recognition," *IEEE Signal Processing Magazine*, vol. 13, no. 5, pp. 45--56, 1996.
- [16] T. Minka, "A family of algorithms for approximate Bayesian inference," Ph.D. dissertation, MIT, 2001.
- [17] C. K. I. Williams and D. Barber, "Bayesian classification with Gaussian processes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1342--1351, 1998.
- [18] C. M. Bishop, *Neural Networks for Pattern Recognition*. Clarendon Press, 1995.
- [19] N. Smith and M. Gales, "Using SVM's and discriminative models for speech recognition," in *Proc. ICASSP*, vol. 1, 2002, pp. 77--80.
- [20] N. Smith, M. Gales, and M. Niranjan, "Data-dependent kernels in SVM classification of speech patterns," Engineering Dept., Cambridge Univ., U.K., Tech. Rep., 2001.
- [21] S. Amari, *Differential-Geometrical Methods in Statistics*. Berlin: Springer-Verlag, 1985.
- [22] N. Smith and M. Gales, "Using SVMs to classify variable length speech patterns," University of Cambridge, Tech. Rep., 2002.
- [23] V. Wan and S. Renals, "Speaker verification using sequence discriminant support vector machines," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 2, pp. 203--210, 2005.

- [24] K. Hornik, "Some new results on neural network approximation," *Neural Networks*, vol. 6, no. 8, pp. 1069--1072, 1993.
- [25] R. M. Neal, "Bayesian learning for neural networks," in *Lecture Notes in Statistics*. New York: Springer, 1996, vol. 118.
- [26] C. K. I. Williams, "Computation with infinite neural networks," *Neural Computation*, vol. 10, no. 5, pp. 1203--1216, 1998.
- [27] S. Bengio and J. Mariéthoz, "Learning the decision function for speaker verification," in *ICASSP*, vol. 1, 2001, pp. 425--428.