



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ  
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ  
ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

**Τεχνικές Βαθιάς Μηχανικής Μάθησης και Γνώσης  
για Ανάλυση Συναισθήματος στην Αλληλεπίδραση  
Ανθρώπου Μηχανής**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**ΔΗΜΗΤΡΗΣ Σ. ΚΟΛΛΙΑΣ**

**Επιβλέπων :** Ανδρέας – Γεώργιος Ν. Σταφυλοπάτης  
Καθηγητής Ε.Μ.Π

Αθήνα, Ιούλιος 2015





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ  
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ  
ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

**Τεχνικές Βαθιάς Μηχανικής Μάθησης και Γνώσης  
για Ανάλυση Συναισθήματος στην Αλληλεπίδραση Ανθρώπου  
Μηχανής**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**ΔΗΜΗΤΡΗΣ Σ. ΚΟΛΛΙΑΣ**

**Επιβλέπων :** Ανδρέας – Γεώργιος Ν. Σταφυλοπάτης  
Καθηγητής Ε.Μ.Π

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 3<sup>η</sup> Ιουλίου 2015

.....  
Ανδρέας – Γεώργιος  
Σταφυλοπάτης  
Καθηγητής Ε.Μ.Π

.....  
Παναγιώτης Τσανάκας  
Καθηγητής Ε.Μ.Π

.....  
Γιώργος Στάμου  
Επίκουρος Καθηγητής Ε.Μ.Π

Αθήνα, Ιούλιος 2015

.....  
Δημήτρης Σ. Κόλλιας

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π

Copyright © Δημήτρης Σ. Κόλλιας, 2015

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.

## Περίληψη

Στη διπλωματική αυτή εργασία παρουσιάζονται νέες μεθοδολογίες μηχανικής μάθησης, οι οποίες συνδυάζουν τεχνικές υπολογιστικής νοημοσύνης, δίνοντας έμφαση σε βαθιά συνελκτικά νευρωνικά δίκτυα, και μηχανές διανυσμάτων υποστήριξης που εμπεριέχουν την προϋπάρχουσα γνώση και τις αντίστοιχες σημασιολογικές αναπαραστάσεις. Η περιγραφόμενη αρχιτεκτονική αποτελεί τη βάση για ανάπτυξη συστημάτων ανάλυσης ή κατηγοριοποίησης δεδομένων, τα οποία έχουν τη δυνατότητα να ενσωματώνουν διαθέσιμες οντολογίες και βάσεις γνώσης και να προσαρμόζουν τη γνώση τους σε προβλήματα του πραγματικού κόσμου, όπως είναι διαφορετικοί χρήστες και εννοιολογικό πλαίσιο αλληλεπίδρασης.

Στην παρούσα εργασία ο στόχος αυτός επιτυγχάνεται με την χρησιμοποίηση νέων αρχιτεκτονικών νευρωνικών δικτύων βαθιάς μάθησης για αποτελεσματική εξαγωγή συγκεκριμένων πληροφοριών ή χαρακτηριστικών από τα δεδομένα, όπως και με την μεταφορά της γνώσης, μέσω μεθόδων πυρήνα σε δομές διανυσμάτων υποστήριξης που υλοποιούν το στάδιο προσαρμοζόμενης ανάλυσης -κατηγοριοποίησης των δεδομένων. Στο πλαίσιο της εργασίας οι ανωτέρω μέθοδοι εφαρμόζονται με επιτυχία στο πρόβλημα της ανάλυσης συναισθήματος μέσω της παρακολούθησης των χαρακτηριστικών του προσώπου του χρήστη σε εφαρμογές αλληλεπίδρασης ανθρώπου μηχανής.

## Λέξεις Κλειδιά

Βαθιά Μηχανική Μάθηση, Αρχιτεκτονικές Συνελκτικών Νευρωνικών Δικτύων, Σημασιολογική Αναπαράσταση Γνώσης, Μέθοδοι Πυρήνα, Μηχανές Διανυσμάτων Υποστήριξης, Ανάλυση Συναισθήματος, Ανάλυση Χαρακτηριστικών Προσώπου, Αλληλεπίδραση Ανθρώπου Μηχανής.

## **Abstract**

New machine learning methodologies are presented in the current Dissertation, which are able to interweave computational intelligent techniques, focusing on deep convolutional neural networks and support vector machines based on pre-existing knowledge and respective semantic representation. The presented architecture is the basis for developing data analysis and data classification systems that are able, on the one hand to encompass available ontologies or knowledge bases and, on the other hand, to efficiently adapt their knowledge to different real life conditions, such as different users and context of usage.

In this work, the above goal is achieved by using deep convolutional neural networks for effective extraction of specific features from data and by transferring existing knowledge, through kernel methods, to support vector machines which are able to adaptively analyse and classify the data.

In this framework, the presented methodologies are successfully applied for user emotion analysis, based on facial feature extraction and analysis, in human machine interaction environments.

## **Key Words**

Deep Learning, Convolutional Networks, Semantics and Knowledge Representation, Kernel Methods, Support Vector Machines, Emotion Analysis, Facial Feature Analysis, Human Machine Interaction.

## Ευχαριστίες

Θέλω να ευχαριστήσω τον επιβλέποντα Καθηγητή μου κ. Ανδρέα-Γεώργιο Σταφυλοπάτη που μου εμπιστεύτηκε το θέμα και με βοήθησε όποτε χρειάστηκε στη διάρκεια εκπόνησης της διπλωματικής μου. Επίσης θέλω να ευχαριστήσω τον Επίκουρο Καθηγητή κ. Γεώργιο Στάμου, ο οποίος με βοήθησε στην κατανόηση πολλών από τις αναφερόμενες έννοιες. Επιπλέον θέλω να ευχαριστήσω τον Καθηγητή κ. Παναγιώτη Τσανάκα για την γενικότερη υποστήριξή του κατά τη διάρκεια της εκπόνησης της εργασίας μου και των σπουδών μου.

Τέλος, ευχαριστώ την οικογένειά μου, τους γονείς μου Στέφανο και Λούλα, και την αδελφή μου Ηλιάννα, για τη συνεχή ενθάρρυνση και υποστήριξη που μου παρείχαν στις σπουδές μου.





## Πίνακας περιεχομένων

---

1. Εισαγωγή.....	1
2. Βαθιά Συνελκτικά Νευρωνικά Δίκτυα.....	6
2.1 Τεχνητά Νευρωνικά Δίκτυα.....	6
2.1.1 Πολυεπίπεδα Νευρωνικά Δίκτυα.....	6
2.1.2 Μάθηση Πολυεπίπεδων Perceptrons .....	7
2.1.3 Προσαρμογή Πολυεπίπεδων Perceptrons .....	11
2.2 Βαθιές Αρχιτεκτονικές Νευρωνικών Δικτύων.....	17
2.2.1 Ορισμός και Ιδιότητες Βαθιών Αρχιτεκτονικών.....	17
2.2.2 Εξαγωγή Χαρακτηριστικών με Βαθιές Αρχιτεκτονικές.....	22
2.2.3 Εκπαίδευση του Συστήματος με Βαθιά Αρχιτεκτονική.....	24
2.3 Βαθιά Συνελκτικά Δίκτυα .....	29
3. Μεταφορά Συμβολικής Γνώσης σε Μηχανές Διανυσμάτων Υποστήριξης.....	34
3.1 Συναρτήσεις και Μηχανές Πυρήνα.....	34
3.1.1 Μέθοδοι Πυρήνα.....	34
3.1.2 Συναρτήσεις Πυρήνα .....	36
3.1.3 Είδη Πυρήνων.....	39
3.1.4 Στοιχεία Μεταβολής Πυρήνων .....	42
3.2 Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines) .....	43
3.2.1 Γραμμικά Διαχωρίσιμα Προβλήματα .....	43
3.2.2 Μη Γραμμικά Διαχωρίσιμα Προβλήματα .....	46
3.2.3 Χρήση Συναρτήσεων Πυρήνα .....	47
3.3 Μεταφορά Συμβολικής Γνώσης σε Πυρήνες Μηχανών Διανυσμάτων Υποστήριξης.....	51
4. Αναπαράσταση και Ανάλυση Συναισθήματος .....	55
4.1 Αναπαράσταση Συναισθήματος.....	55
4.2 Εξαγωγή Χαρακτηριστικών από Εικόνες Προσώπου.....	57
4.3 Περιγραφή Κανόνων για Ανάλυση Συναισθήματος.....	63
5. Αρχιτεκτονική Συστήματος Ανάλυσης Συναισθήματος.....	65
6. Πειραματικά Αποτελέσματα.....	69

6.1 Περιγραφή της Βάσης Δεδομένων .....	69
6.2 Τεκμηρίωση των Δεδομένων .....	72
6.3 Αποτελέσματα της Ανάλυσης Συναισθήματος .....	73
8. Συμπεράσματα και Μελλοντικές Επεκτάσεις .....	88
9. Βιβλιογραφία .....	90



## Κατάλογος Πινάκων

Πίνακας 1: Υπολογισμός των Χαρακτηριστικών FAPs.....	59
Πίνακας 2: Υπάρχουσα Γνώση (Κανόνες) για Ανάλυση Συναισθήματος.....	64



## Κατάλογος Σχημάτων

Σχήμα 2.1: Πολλαπλά επίπεδα λειτουργίας και ενδιάμεσες αναπαραστάσεις του συστήματος οπτικής αντίληψης.....	18
Σχήμα 2.2: Εκμάθηση ενδιάμεσων αναπαραστάσεων σε ένα πρόβλημα κατηγοριοποίησης/πρόβλεψης.....	18
Σχήμα 2.3: Μια μηχανή διανυσμάτων υποστήριξης με βάση συναρτήσεις πυρήνα.....	21
Σχήμα 2.4: Το βασικό δομικό στοιχείο των βαθιών αρχιτεκτονικών.....	22
Σχήμα 2.5: Πολλαπλά επίπεδα μιας βαθιάς αρχιτεκτονικής.....	23
Σχήμα 2.6: Μια πολυεπίπεδη αρχιτεκτονική.....	24
Σχήμα 2.7: Ένα παράδειγμα με βαθύ συνελκτικό δίκτυο.....	30
Σχήμα 2.8: Τοπικές συνδέσεις σε ένα βαθύ συνελκτικό δίκτυο.....	30
Σχήμα 2.9: Η βασική δομή του συνελκτικού δικτύου.....	31
Σχήμα 2.10: Κατηγοριοποίηση εικόνων αντικειμένων με βαθύ συνελκτικό δίκτυο.....	32
Σχήμα 2.11: Υπολογιστικό φορτίο σε ένα βαθύ συνελκτικό δίκτυο.....	33
Σχήμα 4.1: Αναπαράσταση Συναισθήματος.....	56
Σχήμα 4.2: Τα χαρακτηριστικά σημεία FDP.....	61
Σχήμα 5.1: Η αρχιτεκτονική που βασίζεται σε ένα βαθύ συνελκτικό δίκτυο.....	68
Σχήμα 5.2: Η αρχιτεκτονική των συστημάτων ανάλυσης συναισθήματος, με ή χωρίς ενδιάμεση εκτίμηση των παραμέτρων FAPs.....	68
Σχήμα 6.1: Χαρακτηριστικές Εικόνες Έκφρασης Συναισθήματος: α) (+,+) β) (+,-) γ) (-,+) δ) (-,-) ε) ουδέτερη.....	72
Σχήμα 6.2: 3 διαφορετικές εκφράσεις προσώπου (που αντιστοιχούν σε διαφορετικά σετ από FAPs) για κάθε μια από τις 5 κατηγορίες συναισθηματικής κατάστασης: α) (+,+) β) (+,-) γ) (-,+) δ) (-,-) ε) ουδέτερη.....	84
Σχήμα 6.3: Επιθυμητές και προβλεπόμενες έξοδοι του συστήματος ανάλυσης συναισθήματος.....	87



# 1. Εισαγωγή

---

Μεγάλο ενδιαφέρον έχει προσελκύσει τα τελευταία χρόνια η χρήση τεχνικών βαθιάς μάθησης νευρωνικών δικτύων ([12]) σε διάφορες εφαρμογές, συμπεριλαμβανομένης της ομιλίας και της επεξεργασίας γλώσσας ([13]-[15]), της ανάλυσης και της αναγνώρισης εικόνων ([16]-[20]). Τέτοιες τεχνικές περιλαμβάνουν συνελκτικά δίκτυα, τα οποία είναι αρχιτεκτονικές τεχνητών νευρωνικών δικτύων με τοπολογικό χώρο εισόδου, όπως είναι μια διδιάστατη εικόνα. Ένα σύνολο φίλτρων δύο διαστάσεων, ή τριών διαστάσεων (εάν οι εισοδοί είναι έγχρωμες εικόνες), εφαρμόζεται σε μικρές περιοχές της εικόνας, χρησιμοποιώντας συνελίξεις και δημιουργώντας μια τράπεζα χαρτών στην έξοδο του φίλτρου, δηλαδή, ένα χάρτη ανά φίλτρο, ο οποίος επίσης παρουσιάζει μια παρόμοια τοπολογία δύο διαστάσεων. Για να μειωθούν οι διαστάσεις των χαρτών των χαρακτηριστικών και η ανάλυση να είναι αναλλοίωτη ως προς μικρές μεταβολές της εικόνας εισόδου, τα συνελκτικά επίπεδα (layers) ακολουθούνται συνήθως από pooling επίπεδα, τα οποία υποδειγματολειπτούν τους χάρτες των χαρακτηριστικών, μειώνοντας μικρές περιοχές σε ενιαία στοιχεία (για παράδειγμα με την επιλογή της μέγιστης ή μέσης τιμής κάθε περιοχής). Δύο ή περισσότερα τέτοια επίπεδα, περιλαμβάνοντας και συνελκτικά και pooling υπο-επίπεδα, ακολουθούνται από ένα ή περισσότερα πλήρως συνδεδεμένα επίπεδα νευρωνικού δικτύου, τα οποία παρέχουν τελικά την έξοδο του συστήματος.

Η πρόσφατη επιτυχία των μεθόδων βαθιάς μάθησης και των συνελκτικών δικτύων σε απαιτητικές διεργασίες υπολογιστικής όρασης ([21], [22], [23]), μοντελοποίησης φυσικής γλώσσας ([24]) και αναγνώρισης ομιλίας ([25]),



οδήγησαν κάποιους ερευνητές να εφαρμόσουν τέτοιες τεχνικές και στο πρόβλημα της αναγνώρισης συναισθήματος ([31], [32]). Τα συστήματα αυτόματης αναγνώρισης συναισθήματος εξάγουν αναπαραστάσεις υψηλού επιπέδου (high level representations), όπως η συναισθηματική κατάσταση του χρήστη, με βάση την ανάλυση χαμηλού επιπέδου (low level cues) ανθρωποκεντρικών χαρακτηριστικών και σημάτων, όπως είναι οι μεταβολές των χαρακτηριστικών του προσώπου, οι κινήσεις του κεφαλιού και των χεριών, η στάση του σώματος, ο τρόπος και η χροιά της ομιλίας και άλλα. Η ακρίβεια της ανάλυσης ή και της αναγνώρισης των συναισθημάτων εξαρτάται σε μεγάλο βαθμό από την ικανότητα των συστημάτων να εξάγουν και να αναλύουν αντιπροσωπευτικά χαρακτηριστικά. Ωστόσο, αυτό είναι ένα πολύ απαιτητικό πρόβλημα. Οι καταστάσεις συναισθημάτων δεν έχουν σαφή χρονικά όρια και τα πρότυπα έκφρασης των συναισθημάτων συχνά διαφέρουν μεταξύ των ατόμων ([26]). Το πρόβλημα περιπλέκεται περαιτέρω από την υψηλή διάσταση του προκύπτοντος χώρου των χαρακτηριστικών. Ορισμένες προσεγγίσεις επιχειρούν να μειώσουν τον αρχικό χώρο χαρακτηριστικών εισόδου, μέσω επιλογής χαρακτηριστικών ([27]), το αποτέλεσμα, ωστόσο, που προκύπτει δεν είναι πάντοτε πλεονεκτικό για την ανάλυση των συναισθημάτων.

Εχει αναπτυχθεί ένας μεγάλος όγκος εργασιών με μεθόδους αναπαράστασης χαρακτηριστικών, εξόρυξης, και επιλογής τους για ανάλυση και αναγνώριση συναισθημάτων κατά την τελευταία δεκαετία. Τέτοια αποτελέσματα έχουν για παράδειγμα προκύψει στο πλαίσιο μιας ποικιλίας συνεργασιών του Ε. Μ. Πολυτεχνείου σε έργα έρευνας και ανάπτυξης, όπως τα έργα ERMIS, HUMAINE, CALLAS, CARE ([www.image.ntua.gr](http://www.image.ntua.gr)), έχοντας δείξει ότι είναι εφικτό να γίνει με επιτυχία εξαγωγή οπτικοακουστικών χαρακτηριστικών και στη συνέχεια ανάλυση και αναγνώριση συναισθημάτων. Σημσιολογικά αποτελέσματα, με τη μορφή συστημάτων που βασίζονται σε κανόνες, σε

οντολογίες και σε σχετικές αναπαραστάσεις γνώσης, έχουν επίσης εξαχθεί σε αυτό το πλαίσιο.

Τόσο τα συστήματα συμβολικής αναπαράστασης γνώσης, όσο και τα τεχνητά νευρωνικά δίκτυα έχουν ένα ιδιαίτερα σημαντικό ρόλο στο επιστημονικό πεδίο της Τεχνητής Νοημοσύνης και της Μηχανικής Μάθησης. Ένα ζήτημα μεγάλης σημασίας είναι η δυνατότητα συνδυασμού της υφιστάμενης γνώσης με τεχνικές μηχανικής μάθησης από τα δεδομένα. Τα ευφυή υπολογιστικά συστήματα που βασίζονται σε συμβολική επεξεργασία της γνώσης διαφέρουν σημαντικά από τα συστήματα υποσυμβολικής επεξεργασίας, όπως είναι τα τεχνητά νευρωνικά δίκτυα. Παρόλα αυτά και τα δύο αποτελούν τυπικές προσεγγίσεις της τεχνητής νοημοσύνης και είναι ιδιαίτερα επιθυμητό να μπορέσει να επιτευχθεί συνδυασμός της ευστάθειας των τεχνητών νευρωνικών δικτύων, π.χ., σε συνθήκες θορύβου ή/και μεγάλων όγων δεδομένων, και της δυνατότητας μάθησης από παραδείγματα που διαθέτουν, με την εκφραστικότητα της συμβολικής αναπαράστασης της γνώσης. Αυτό έχει αποδειχθεί ότι οδηγεί σε συστήματα μάθησης, που είναι πιο αποτελεσματικά από ότι αμιγώς συμβολικά ή αμιγώς υποσυμβολικά (connectionist) συστήματα, ειδικά όταν τα δεδομένα είναι θορυβώδη, ή εξαρτώνται από το περιεχόμενο και από το εννοιολογικό πλαίσιο στο οποίο δημιουργούνται. Αυτή είναι η αιτία που έχει διεθνώς αναγνωριστεί η ανάγκη έρευνας για σύνδεση των τεχνικών μηχανικής μάθησης και συμβολικής αναπαράστασης της γνώσης στον κλάδο της τεχνητής νοημοσύνης ([39], [40], [41]).

Ο όγκος των υβριδικών πληροφοριών, δηλαδή πληροφοριών που περιέχουν συμβολικά, αλλά και στατιστικά στοιχεία, όπως επίσης και θόρυβο, συνεχώς αυξάνεται, σε πολλές και διαφορετικές περιοχές, όπως στην βιοπληροφορική, στην επικοινωνία ανθρώπου μηχανής, στις ψηφιακές βιβλιοθήκες, στην

ανάλυση και διασύνδεση πληροφοριών κειμένου, αλλά και πολυμέσων, στον παγκόσμιο ιστό. Ως αποτέλεσμα της αύξησης αυτής, η νευρωνική (υποσυμβολική) – συμβολική μάθηση αποκτά μεγαλύτερη πρακτική σημασία.

Τυπικά, οι αλγόριθμοι μετατροπής από συμβολική αναπαράσταση γνώσης σε αρχιτεκτονικές τεχνητών νευρωνικών δικτύων, και αντίστροφα, στοχεύουν στο να δημιουργήσουν είτε μια νευρωνική υλοποίηση μιας λογικής, είτε ένα λογικό χαρακτηρισμό ενός νευρωνικού συστήματος, ή ένα υβριδικό σύστημα, το οποίο περιλαμβάνει στοιχεία τόσο μηχανικής μάθησης, όσο και συμβολικής τεχνητής νοημοσύνης. Εντούτοις, ενώ η αναπαράσταση συμβολικής γνώσης είναι αναδρομική και καλά ορισμένη, τα νευρωνικά δίκτυα κωδικοποιούν τη γνώση μέσω των βαρών των συνδέσεών τους, όπως αυτά προκύπτουν από τη διαδικασία της μάθησης από παραδείγματα (συνήθως διανύσματα χαρακτηριστικών των δεδομένων) και της γενίκευσης (generalisation) ([37], [38]). Δίκτυα που χρησιμοποιούν ειδικές συναρτήσεις σε σημασιολογικούς χώρους έχουν ιδιαίτερη σημασία, επειδή είναι ευκολότερο να ελεγχθούν και να αναπαρασταθούν με λογικούς κανόνες με βάση την ομοιότητα ([42], [43]).

Στην παρούσα διπλωματική εργασία παρουσιάζουμε και αναλύουμε αρχιτεκτονικές που περιλαμβάνουν βαθιά συνελκτικά νευρωνικά δίκτυα και μηχανές διανυσμάτων υποστήριξης, που έχουν τη δυνατότητα να εξάγουν πολύπλοκες μη γραμμικές συσχετίσεις χαρακτηριστικών των δεδομένων, να εκμεταλλεύονται την προ-υπάρχουσα γνώση και να επιτρέπουν την αποτελεσματική ανάλυση της συμπεριφοράς των χρηστών σε περιπτώσεις αλληλεπίδρασής τους με τα αντίστοιχα συστήματα. Η εργασία αποτελείται από την Εισαγωγή και 8 Κεφάλαια.

Το Κεφάλαιο 2 περιγράφει τις σύγχρονες αρχιτεκτονικές βαθιάς μηχανικής μάθησης, επικεντρώνοντας σε βαθιά συνελκτικά νευρωνικά δίκτυα, τα οποία αποτελούν ένα από τα βασικά εργαλεία της παρούσας εργασίας.

Το Κεφάλαιο 3 αναφέρεται στις μεθόδους πυρήνα και στις μηχανές διανυσμάτων υποστήριξης (SVM), επικεντρώνοντας σε μεθόδους μεταφοράς συμβολικής αναπαράστασης γνώσης στις μηχανές SVM.

Το Κεφάλαιο 4 περιγράφει το πρόβλημα της αναπαράστασης συναισθήματος, την εξαγωγή χαρακτηριστικών από εικόνες προσώπου και την δημιουργία κανόνων για την ανάλυση συναισθήματος.

Το Κεφάλαιο 5 περιγράφει την αρχιτεκτονική του προτεινόμενου συστήματος με βάση τις περιγραφές των Κεφαλαίων 2,3,4 για την ανάλυση συναισθήματος.

Στο Κεφάλαιο 6 γίνεται η περιγραφή των δεδομένων που παρήχθησαν και χρησιμοποιήθηκαν στην μελέτη, αλλά και η ανάλυση των αποτελεσμάτων που παρήχθησαν από την χρησιμοποίηση της αρχιτεκτονικής που υλοποιήθηκε σύμφωνα με το Κεφάλαιο 5.

Συμπεράσματα και προτάσεις για μελλοντική έρευνα δίδονται στο Κεφάλαιο 7 της παρούσας εργασίας.

Στο Κεφάλαιο 8 παρατίθεται η βιβλιογραφία της παρούσας εργασίας.

# 2. Βαθιά Συνελικτικά Νευρωνικά Δίκτυα

---

## 2.1 Τεχνητά Νευρωνικά Δίκτυα

### 2.1.1 Πολυεπίπεδα Νευρωνικά Δίκτυα

Τα τεχνητά νευρωνικά δίκτυα (ΤΝΔ) είναι αρχιτεκτονικές αποτελούμενες από πλήθος δια-συνδεδεμένων μονάδων - νευρώνων. Κάθε νευρώνας δέχεται στην είσοδό του σήματα, από το περιβάλλον αν είναι νευρώνας εισόδου, ή από τις εξόδους των άλλων νευρώνων, περιλαμβάνει μια συνάρτηση ενεργοποίησης, μέσω της οποίας παράγει ένα σήμα, με το οποίο τροφοδοτεί τους άλλους νευρώνες ή την έξοδο του δικτύου. Κάθε σύναψη μεταξύ δύο νευρώνων χαρακτηρίζεται από μια τιμή βάρους. Οι τιμές των βαρών των συνάψεων αποτελούν τη γνώση που είναι αποθηκευμένη στο δίκτυο και καθορίζουν τη λειτουργία του.

Το υπολογιστικό μοντέλο ενός νευρωνικού δικτύου πολλαπλασιάζει τα σήματα-εισόδους κάθε νευρώνα με τα συναπτικά βάρη των συνδέσεων από τις οποίες αυτά διέρχονται και υπολογίζει το άθροισμα των γινομένων αυτών. Το άθροισμα που προκύπτει αποτελεί την είσοδο της συνάρτησης ενεργοποίησης του νευρώνα. Εάν  $x_i$  είναι η  $i$ -οστή είσοδος του  $k$  νευρώνα,  $w_{ki}$  το  $i$ -οστό συναπτικό βάρος του  $k$  νευρώνα (από έναν αριθμό  $N$  συνδέσεων) και  $\phi$  η συνάρτηση ενεργοποίησης του νευρώνα, τότε η έξοδος  $y_k$  του νευρώνα δίδεται από την εξίσωση:

$$y_k = \phi\left(\sum_{i=0}^{N-1} x_i w_{ki}\right) \quad (2.1)$$

Οι πλέον συνήθεις μορφές της συνάρτησης ενεργοποίησης είναι η βηματική, η συνάρτηση προσήμου, η ταυτοτική συνάρτηση - στην περίπτωση γραμμικών νευρώνων – και η σιγμοειδής συνάρτηση.

Ένας από τους πλέον συνήθεις τύπους νευρωνικών δικτύων είναι τα πολυεπίπεδα δίκτυα εμπρόσθιας τροφοδότησης (feedforward multilayer neural networks). Στα δίκτυα αυτά οι νευρώνες κατανέμονται σε επίπεδα (στρώματα). Ένα δίκτυο N επιπέδων εμπρόσθιας τροφοδότησης είναι ένας ακυκλικός γράφος. Αποτελείται από N επίπεδα, κάθε ένα από τα οποία περιλαμβάνει κάποιο αριθμό από νευρώνες. Συνδέσεις υπάρχουν μόνο μεταξύ νευρώνων διαφορετικών (συνήθως διαδοχικών) επιπέδων, με το πρώτο επίπεδο να είναι το επίπεδο εισόδου και τους υπολογισμούς των εξόδων των νευρώνων να γίνονται από το επίπεδο εισόδου, στα ενδιάμεσα επίπεδα (κρυφά) και τελικά στο επίπεδο εξόδου (εμπρόσθια τροφοδότηση των πληροφοριών – από την είσοδο στην έξοδο, δίχως ανατροφοδότηση).

### 2.1.2 Μάθηση Πολυεπίπεδων Perceptrons

Η μηχανική μάθηση (machine learning), στην περίπτωση νευρωνικών δικτύων σταθερής δομής, δηλαδή με συγκεκριμένο αριθμό νευρώνων, επιπέδων, συνδέσεων, συνίσταται στον υπολογισμό των ανεξάρτητων παραμέτρων – βαρών των συνδέσεων που αυτά περιέχουν, ώστε να εκτελούν μια συγκεκριμένη λειτουργία. Ο πλέον συνήθης αλγόριθμος εκπαίδευσης των πολυεπίπεδων νευρωνικών δικτύων εμπρόσθιας τροφοδότησης είναι ο αλγόριθμος της ανάστροφης διάδοσης σφάλματος (Error Back-Propagation ή απλά back-propagation). Ο αλγόριθμος αυτός, προκύπτει εφαρμόζοντας τις τεχνικές της θεωρίας βελτιστοποίησης (optimisation theory) πάνω στην αρχιτεκτονική ενός πολυεπίπεδου δικτύου και αποτελεί επέκταση του αλγορίθμου μάθησης των

perceptrons στην περίπτωση των πολυεπίπεδων αρχιτεκτονικών (multilayer perceptrons). Η βελτιστοποίηση (ελαχιστοποίηση μιας συνάρτησης κόστους) επιτυγχάνεται στο στάδιο της εκπαίδευσης του δικτύου, οπότε αυτό μαθαίνει να προσεγγίζει, με τις εξόδους που παρέχει, ένα σύνολο από επιθυμητές τιμές (desired outputs) που του παρέχονται από ένα εξωτερικό δάσκαλο (supervised training) μαζί με το αντίστοιχο σύνολο από δεδομένα – πρότυπα εισόδου (training data set).

Έστω ότι έχουμε ένα σύνολο εκπαίδευσης που αποτελείται από  $N$  ζεύγη της μορφής  $[\bar{x}, \bar{d}]$ , όπου  $\bar{x}$  συμβολίζει το διάνυσμα (πρότυπο) εισόδου και  $\bar{d}$  το διάνυσμα της επιθυμητής εξόδου όταν εφαρμοσθεί ως είσοδος το πρότυπο  $\bar{x}$ . Ας αναφερθούμε σε έναν από τους νευρώνες του επιπέδου εξόδου του δικτύου. Το σφάλμα στην έξοδο του νευρώνα  $j$ , όταν στην είσοδο του δικτύου παρουσιάζεται το πρότυπο  $\bar{x}(n)$ , όπου  $n = 1, 2, \dots, N$ , δίδεται από τον τύπο:

$$e_j(n) = d_j(n) - y_j(n) \quad (2.2)$$

όπου  $d_j$  και  $y_j$  η επιθυμητή και η πραγματική έξοδος αντίστοιχα του νευρώνα  $j$  για είσοδο την  $\bar{x}(n)$ .

Η στιγμιαία τιμή του τετραγωνικού σφάλματος για τον νευρώνα  $j$  ορίζεται ως:

$$\frac{1}{2} e_j^2(n) \quad (2.3)$$

και το άθροισμα των τετραγωνικών σφαλμάτων όλων των νευρώνων εξόδου, στην  $n$ -οστή επανάληψη-παρουσίαση ενός προτύπου στην είσοδό του δίδεται από τη σχέση:

$$G(n) = \frac{1}{2} \sum_j e_j^2(n) \quad (2.4)$$

Η συνάρτηση  $G(n)$  αποτελεί τη συνάρτηση κόστους, η οποία ελαχιστοποιείται, συσσωρευτικά για όλα τα  $n$ , μέσω του αλγόριθμου back propagation. Ειδικότερα, ο αλγόριθμος της ανάστροφης διάδοσης προβαίνει, σε κάθε επανάληψη,  $n$ , σε μια ανανέωση (διόρθωση)  $\Delta w_{ji}(n)$  στο βάρος  $w_{ji}(n)$  της σύνδεσης μεταξύ του νευρώνα  $j$  κάποιου επιπέδου (έστω  $m$ ) του δικτύου και του νευρώνα  $i$  του προηγούμενου ( $m-1$ ) επιπέδου,  $m=1, \dots, N$ . Η ανανέωση γίνεται πρώτα για τα βάρη των συνδέσεων στο επίπεδο εξόδου, και στη συνέχεια, ανάστροφα, για τα βάρη των κρυμμένων επιπέδων, μέχρι το επίπεδο εισόδου. Η ανανέωση κάθε βάρους είναι ανάλογη της κλίσης  $\frac{\partial G(n)}{\partial w_{ji}(n)}$ ,

δηλαδή της παραγώγου του σφάλματος ως προς το συγκεκριμένο βάρος σύνδεσης, κατά την επεξεργασία του  $n$ -οστού προτύπου, και προσδιορίζεται, σύμφωνα με τη μέθοδο της βαθύτατης καθόδου (steepest descent) ([37], [38]) ως εξής :

$$\Delta w_{ji}(n) = -\eta \frac{\partial G(n)}{\partial w_{ji}(n)} \quad (2.5)$$

Επειδή η μέθοδος της βαθύτατης καθόδου είναι μια προσέγγιση της μεθόδου του Newton, η οποία δείχνει μεν την κατεύθυνση της ανανέωσης των βαρών, δηλαδή την κλίση της συνάρτησης σφάλματος, αλλά όχι το ακριβές βήμα της ανανέωσης (που το δείχνει η δεύτερη παράγωγος – ο πίνακας Hessian της συνάρτησης), η κλίση της συνάρτησης πολλαπλασιάζεται με μια παράμετρο,  $\eta$ , η οποία καθορίζει την ταχύτητα σύγκλισης, δηλαδή τον ρυθμό μάθησης (learning rate) του δικτύου.

Αποδεικνύεται ότι η ανανέωση των βαρών γράφεται ως εξής:

$$\Delta w_{ji}(n) = -\eta \delta_j(n) y_i(n) \quad (2.6)$$

όπου  $\delta_j$  είναι το σφάλμα στην έξοδο του νευρώνα  $j$  του επιπέδου - έστω  $m$  - και  $y_i$  είναι η έξοδος του νευρώνα  $i$  του επιπέδου ( $m-1$ ), ή το  $i$ -οστό στοιχείο του διανύσματος εισόδου, αν  $m=1$ .



Οι τιμές  $\delta_j$  υπολογίζονται με βάση τις τιμές των εξόδων  $y$  των νευρώνων του δικτύου και τις τιμές των βαρών των συνδέσεων του κατά την επανάληψη  $n$ . Ειδικότερα, αν το  $m$  αναφέρεται στο επίπεδο της εξόδου, ισχύει (για βηματική σιγμοειδή συνάρτηση ενεργοποίησης των νευρώνων) για τον  $k$ -οστό νευρώνα:

$$\delta_k(n) = y_k(n)[1 - y_k(n)][d_k(n) - y_k(n)] \quad (2.7)$$

Ενώ, αν το  $m$  αναφέρεται σε κρυμμένα επίπεδα, τότε ισχύει:

$$\delta'_k(n) = y_k(n)[1 - y_k(n)] \sum_j \delta_j(n) w_{jk} \quad (2.8)$$

Στον υπολογισμό των σφαλμάτων  $\delta'_k$ , στην περίπτωση αυτή, το άθροισμα του δεύτερου μέλους της εξίσωσης αφορά σε όλα τα σφάλματα  $\delta_j$  του ανωτέρου  $(m+1)$  επιπέδου. Αναλυτικές εξισώσεις για την ανανέωση των βαρών δίδονται, για παράδειγμα, στις αναφορές ([37], [38]).

Σε συνέχεια των ανωτέρω, η εφαρμογή του αλγορίθμου Back-Propagation κι αφού αρχικοποιηθούν τα βάρη των συνδέσεων του δικτύου σε μικρές τυχαίες τιμές, γίνεται, για κάθε πρότυπο εισόδου, σε δύο φάσεις, στην *ορθή φορά* (forward pass) και στην *ανάστροφη φορά* (reverse pass). Η πρώτη φάση ξεκινά από το επίπεδο εισόδου με την παρουσίαση του διανύσματος εισόδου στο δίκτυο, περιλαμβάνει υπολογισμό των εξόδων όλων των νευρώνων, με βάση τις εξισώσεις (2.1), προχωρώντας στα επόμενα επίπεδα του δικτύου, ώσπου να φτάσει τελικά στο τελευταίο επίπεδο (εξόδου) και να υπολογίσει το διάνυσμα εξόδου του δικτύου, καθώς και το σφάλμα του κάθε νευρώνα αυτού του επιπέδου από τις εξισώσεις (2.7).

Η δεύτερη φάση, στη συνέχεια, ξεκινά από το επίπεδο εξόδου, διαδίδοντας τα σήματα σφάλματος, ακολουθιακά, προς τα προηγούμενα (κρυμμένα) επίπεδα - με βάση τις

εξισώσεις (2.8) - και υπολογίζοντας την ανανέωση του κάθε βάρους του δικτύου, ώσπου να φτάσει τελικά στο επίπεδο εισόδου του δικτύου.

### 2.1.3 Προσαρμογή Πολυεπίπεδων Perceptrons

Το στάδιο της μάθησης των νευρωνικών δικτύων είναι ένα από τα σημαντικότερα για την μετέπειτα εφαρμογή τους σε πραγματικά προβλήματα κατηγοριοποίησης και ταξινόμησης, πρόβλεψης, ανάλυσης πληροφοριών. Το στάδιο της μάθησης πρέπει να αποφύγει προβλήματα υπερ-εκπαίδευσης (overtraining) και να συντελέσει στην καλή γενίκευση (generalisation) της συμπεριφοράς του δικτύου. Αυτόματη δόμηση του δικτύου με τεχνικές κλαδέματος (pruning) ή κατασκευαστικές (constructive) ([44]) χρησιμοποιούνται σε συνδυασμό με τον αλγόριθμο μάθησης για το σκοπό αυτό.

Παρά την προσεκτική σχεδίαση, η απόδοση των τεχνητών νευρωνικών δικτύων, όπως και όλων των μηχανών αναπαράστασης γνώσης, σε πραγματικές συνθήκες λειτουργίας θα επηρεάζεται από τα συγκεκριμένα χαρακτηριστικά – συνθήκες του περιβάλλοντος, χαρακτηριστικά των χρηστών – τα οποία δεν είναι δυνατόν να έχουν προβλεφθεί στο στάδιο της σχεδίασης. Στο πλαίσιο αυτό αποκτά ιδιαίτερη σημασία η προσαρμογή (adaptation) των αρχιτεκτονικών των νευρωνικών δικτύων στις συνθήκες λειτουργίας.

Ενα σημαντικό χαρακτηριστικό των πολυεπίπεδων νευρωνικών δικτύων εμπρόσθιας τροφοδότησης (multilayer, feedforward networks – perceptrons) είναι ότι μπορούν να προσαρμόσουν την γνώση τους στον χώρο λειτουργίας τους, με ημι-αυτόματο τρόπο. Δεν απαιτείται, δηλαδή, η προσεκτική και χρονοβόρα επανασχεδίασή τους, όπως, για παράδειγμα, θα χρειαζόταν μια μηχανή τεχνητής νοημοσύνης για την πιθανή επαν-διατύπωση των κανόνων, των αξιωμάτων που περιλαμβάνει. Απλά, στο συγκεκριμένο περιβάλλον, ο χρήστης, ή κάποια τοπική βάση γνώσης, ή κάποια μηχανή ανάλυσης των πληροφοριών, δίνει στο νευρωνικό δίκτυο την πληροφόρηση ότι σε κάποιες περιπτώσεις

αυτό έκανε λάθος εκτίμηση, αναφέροντας και την σωστή (επιθυμητή) απόκριση που θα περίμενε σε αυτές.

Ένα επιπλέον χαρακτηριστικό των multilayer perceptrons είναι ότι μπορούν να πετύχουν την προσαρμογή, ανεξάρτητα αν το πρόβλημα συνεχίζει να παραμένει γραμμικά διαχωρίσιμο, ή έχει γίνει μη-γραμμικά διαχωρίσιμο, σε αντίθεση με τις τεχνικές των απλών (kernel) perceptrons ή των support vector machines. Παραμένει βέβαια το πρόβλημα ότι η συμπεριφορά των δικτύων αυτών (απόδοση, προσαρμογή) παραμένει υπο-συμβολική, δηλαδή δεν επιδέχεται άμεσα σημασιολογική ερμηνεία.

Στη συνέχεια περιγράφουμε μια τεχνική προσαρμογής των multilayer perceptrons ([45], [46]) σε πραγματικές συνθήκες, οι οποίες απαιτούν επανεκπαίδευση του δικτύου, λαμβάνοντας υπόψη πρότυπα-δεδομένα από το συγκεκριμένο περιβάλλον και συνθήκες λειτουργίας.

Ας υποθέσουμε ότι προσπαθούμε να κατηγοριοποιήσουμε, σε μια από, έστω,  $p$  κλάσεις  $\omega$ , τα πρότυπα – διανύσματα εισόδου  $\bar{x}_i$ , τα οποία περιλαμβάνουν τα δεδομένα μας κάθε χρονική στιγμή. Το νευρωνικό δίκτυο δημιουργεί στην έξοδό του ένα διάνυσμα  $p$ -διάστατο, έστω  $\bar{y}(\bar{x}_i)$

$$\bar{y}(\bar{x}_i) = \left[ p_{\omega_1}^i \ p_{\omega_2}^i \ \dots \ p_{\omega_p}^i \right]^T \quad (2.9)$$

όπου  $p_{\omega_j}^i$  συμβολίζει την πιθανότητα η  $i$ -οστή έξοδος να ανήκει στην  $j$ -οστή κατηγορία.

Ας υποθέσουμε καταρχήν ότι το νευρωνικό δίκτυο έχει εκπαιδευτεί με ένα συγκεκριμένο – μεγάλου μεγέθους – σύνολο δεδομένων (training set), έστω,  $S_b = \left\{ (\bar{x}'_1, \bar{d}'_1), \dots, (\bar{x}'_{m_b}, \bar{d}'_{m_b}) \right\}$ , όπου τα διανύσματα  $\bar{x}'_i$  και  $\bar{d}'_i$  με  $i=1,2,\dots,m_b$  συμβολίζουν το  $i$ -ιστό διάνυσμα εκπαίδευσης στην είσοδο του δικτύου και το αντίστοιχο διάνυσμα επιθυμητής εξόδου που αποτελείται από  $p$  στοιχεία.

Στη συνέχεια, ας υποθέσουμε ότι  $\bar{y}(\bar{x}_i)$  συμβολίζει την έξοδο του δικτύου όταν αυτό εφαρμόζεται – σε συνθήκες λειτουργίας – στο νέο σύνολο δεδομένων. Εστω το  $i$ -οστό δείγμα – από το νέο σύνολο δεδομένων, που έχουν συλλεγεί για επανεκπαίδευση του δικτύου, και που πιθανόν αντιστοιχούν σε κάποιο νέο χρήστη, τα χαρακτηριστικά του οποίου διαφέρουν από αυτά των χρηστών που χρησιμοποιήθηκαν στο στάδιο της αρχικής εκπαίδευσης, ή σε κάποια αλλαγή των συνθηκών του περιβάλλοντος. Για να αντιμετωπιστούν τα νέα δεδομένα, το δίκτυο πρέπει να επανεκπαιδευτεί, διατηρώντας την προηγούμενη γνώση του και προσαρμόζοντάς την στα νέα δεδομένα.

Ας υποθέσουμε ότι το διάνυσμα  $\bar{w}_b$  περιλαμβάνει όλα τα βάρη του δικτύου, όπως ήταν πριν την επανεκπαίδευση, και ότι το διάνυσμα  $\bar{w}_a$  περιλαμβάνει τα νέα βάρη που προήλθαν μετά την επανεκπαίδευση.

Για να γίνει η επανεκπαίδευση, ένα σύνολο επανεκπαίδευσης  $S_c$  εξάγεται από τις νέες συνθήκες, αποτελούμενο από, έστω,  $m_c$  πρότυπα εισόδου:

$$S_c = \left\{ (\bar{x}_1, \bar{d}_1), \dots, (\bar{x}_{m_c}, \bar{d}_{m_c}) \right\} \quad (2.10)$$

όπου  $\bar{x}_i$  και  $\bar{d}_i$  με  $i=1,2,\dots,m_c$  αντιστοιχούν στο  $i$ -οστό πρότυπο εισόδου και επιθυμητή έξοδο από τα δεδομένα επανεκπαίδευσης.

Ο αλγόριθμος επανεκπαίδευσης – προσαρμογής του δικτύου – θα υπολογίσει τα νέα βάρη των συνδέσεων του δικτύου  $\bar{w}_a$ , ελαχιστοποιώντας την επόμενη συνάρτηση σφάλματος ως προς τα βάρη,

$$E_a = E_{c,a} + \eta E_{f,a} \quad (2.11)$$

$$E_{c,a} = \frac{1}{2} \sum_{i=1}^{m_c} \left\| \bar{z}_a(\bar{x}_i) - \bar{d}_i \right\|_2 \quad (2.12)$$

$$E_{f,a} = \frac{1}{2} \sum_{i=1}^{m_b} \left\| \bar{z}_a(\bar{x}'_i) - \bar{d}'_i \right\|_2 \quad (2.13)$$

Η συνάρτηση σφάλματος, που ελαχιστοποιείται, αποτελείται από δύο όρους: ο  $E_{c,a}$  συμβολίζει το σφάλμα που εμφανίζεται πάνω στο – τωρινό – σύνολο επανεκπαίδευσης  $S_c$  ενώ ο  $E_{f,a}$  συμβολίζει το αντίστοιχο σφάλμα που δημιουργείται στην αρχική γνώση του δικτύου, δηλαδή στο σύνολο εκπαίδευσης  $S_b$ .

Τα  $\bar{z}_a(\bar{x}_i)$  και  $\bar{z}_a(\bar{x}'_i)$  συμβολίζουν τις εξόδους του δικτύου μετά την επανεκπαίδευση, που αντιστοιχούν στα πρότυπα εισόδου  $\bar{x}_i$  και  $\bar{x}'_i$ , ενώ τα βάρη του δικτύου αποτελούν το διάνυσμα  $\bar{w}_a$ . Ομοίως το  $\bar{z}_b(\bar{x}_i)$  αναπαριστάνει την έξοδο του δικτύου με βάρη συνδέσεων  $\bar{w}_b$ , όταν το πρότυπο εισόδου είναι το  $\bar{x}_i$ . Την πρώτη φορά που γίνεται προσαρμογή της γνώσης, δηλ. των βαρών των συνδέσεων, του δικτύου, το  $\bar{z}_b(\bar{x}_i)$  είναι ίσο με το  $\bar{y}(\bar{x}_i)$ . Η παράμετρος  $\eta$  εδώ αποτελεί έναν παράγοντα που καθορίζει τη σημασία του τωρινού συνόλου δεδομένων (επανεκπαίδευσης) σε σύγκριση με το αρχικό σύνολο δεδομένων εκπαίδευσης, ενώ το  $\|\cdot\|_2$  συμβολίζει την  $L_2$  - νόρμα.

Ο σκοπός της επανεκπαίδευσης είναι η ελαχιστοποίηση της παραπάνω συνάρτησης σφάλματος και ο υπολογισμός των νέων βαρών  $\bar{w}_a$ . Ο αλγόριθμος έχει προταθεί στην αναφορά ([45]).

Ας υποθέσουμε ότι αρκεί μια μικρή μεταβολή των (αρχικών) βαρών των συνδέσεων του δικτύου  $\bar{w}_b$  για την επίτευξη καλής απόδοσης του δικτύου στις νέες συνθήκες που χρησιμοποιείται. Τότε θα είναι:

$$\bar{W}_a = \bar{W}_b + \Delta \bar{W} \quad (2.14)$$

όπου τα  $\Delta \bar{w}$  θα είναι μικρές ποσότητες. Αυτή η υπόθεση οδηγεί σε μια αναλυτική και βολική επίλυση του προβλήματος της εκτίμησης του διανύσματος  $\bar{w}_a$ , επειδή επιτρέπει γραμμικοποίηση της μη γραμμικής συνάρτησης σφάλματος γύρω από την τωρινή θέση των τιμών των βαρών στον χώρο ελαχιστοποίησης, με ανάπτυξη σε σειρά Taylor πρώτης τάξης.

Οι εξισώσεις (2.11) – (2.13) υποδηλώνουν ότι τα νέα βάρη του δικτύου υπολογίζονται λαμβάνοντας υπόψη τόσο την παρούσα κατάσταση, όσο και την προηγούμενη γνώση του δικτύου. Για να τονιστεί η σημασία του συνόλου δεδομένων επανεκπαίδευσης (το οποίο γενικά είναι μικρής διάστασης, πολύ μικρότερης από τη διάσταση του συνόλου των δεδομένων της αρχικής εκπαίδευσης του δικτύου), ο πρώτος όρος της συνάρτησης σφάλματος μπορεί να αντικατασταθεί από τον περιορισμό οι πραγματικές έξοδοι του δικτύου να είναι ίσες με τις επιθυμητές, δηλαδή:

$$z_a(\bar{x}_i) = d_i \quad i = 1, \dots, m_c \quad (2.15)$$

για όλα τα πρότυπα του συνόλου  $S_c$ .

Μέσω της γραμμικοποίησης, η επίλυση των εξισώσεων (3.64) ως προς τις ανανεώσεις των βαρών του δικτύου αποδεικνύεται ότι είναι ισοδύναμη με την επίλυση ενός συστήματος γραμμικών εξισώσεων

$$\bar{c} = \mathbf{A} \cdot \Delta \bar{w} \quad (2.16)$$

όπου το διάνυσμα  $\bar{c}$  και ο πίνακας  $\mathbf{A}$  περιλαμβάνουν εκφράσεις των προηγούμενων τιμών των βαρών των συνδέσεων του δικτύου.

Ειδικότερα,

$$\bar{c} = [d_1 \cdots d_{m_c}]^T - [z_b(\bar{x}_1) \cdots z_b(\bar{x}_{m_c})]^T \quad (2.17)$$

Επιπλέον, η ελαχιστοποίηση του δεύτερου όρου της συνάρτησης σφάλματος, ο οποίος εκφράζει την επίδραση των νέων βαρών του δικτύου στο αρχικό σύνολο εκπαίδευσης του δικτύου  $S_b$ , ισοδυναμεί με ελαχιστοποίηση του απολύτου της διαφοράς των σφαλμάτων - πάνω στο σύνολο δεδομένων εκπαίδευσης  $S_b$  - που προκύπτουν από την εφαρμογή σε αυτό του αρχικού και του επανεκπαιδευμένου δικτύου. Ισοδύναμα αυτό σημαίνει ότι οι ανανεώσεις των βαρών προκύπτουν από την ελαχιστοποίηση της επόμενης συνάρτησης σφάλματος

$$E_S = \|E_{f,a} - E_{f,b}\|_2 \quad (2.18)$$

Όπου το σφάλμα  $E_{f,b}$  ορίζεται όμοια με το σφάλμα  $E_{f,a}$ , με το σήμα  $\bar{z}_a$  να αντικαθίσταται με το  $\bar{z}_b$ .

Μπορεί να αποδειχτεί ότι η ανωτέρω συνάρτηση σφάλματος γράφεται ως εξής:

$$E_S = \frac{1}{2} (\Delta \bar{w})^T \cdot \mathbf{K}^T \cdot \mathbf{K} \cdot \Delta \bar{w} \quad (2.19)$$

όπου τα στοιχεία του πίνακα  $\mathbf{K}$  εκφράζονται με βάση τα προηγούμενα βάρη του δικτύου  $\bar{w}_b$  και τα πρότυπα εισόδου στο  $S_b$ . Η συνάρτηση σφάλματος που ορίζεται ανωτέρω είναι κυρτή, επειδή έχει τετραγωνική μορφή. Επομένως, οι ανανεώσεις των βαρών μπορούν να υπολογιστούν με ελαχιστοποίηση της (2.19), με τους περιορισμούς που τίθενται από την (2.16) στο πλαίσιο της ελαχιστοποίησης της (2.11). Η μέθοδος gradient projection ([45]) μπορεί να χρησιμοποιηθεί στο πλαίσιο αυτό για να υπολογίσει τις ανανεώσεις των βαρών.

Οι επόμενες παρατηρήσεις δίνουν μερικές επιπλέον πληροφορίες για τη διαδικασία προσαρμογής των βαρών του νευρωνικού δικτύου:

- 1) Απαραίτητη προϋπόθεση για την προσαρμογή του δικτύου σε νέες συνθήκες είναι η δημιουργία ενός νέου συνόλου δεδομένων – επανεκπαίδευσης  $S_c$ .
- 2) Οι ανανεώσεις των βαρών των συνδέσεων του δικτύου υπολογίζονται, λαμβάνοντας υπόψη τόσο την τωρινή πληροφορία (τα δεδομένα στο σύνολο  $S_c$ ) όσο και την προηγούμενη γνώση (τα δεδομένα στο σύνολο  $S_b$ ).
- 3) Επειδή το σύνολο δεδομένων  $S_c$  περιέχει πληροφορίες μόνο για τις συγκεκριμένες (νέες) συνθήκες της εφαρμογής, δεν μπορεί να θεωρηθεί χαρακτηριστικό γενικά για το πρόβλημα το οποίο εξετάζεται. Μπορεί, επομένως, να χρησιμοποιηθεί για ανανέωση της γνώσης που έχουμε για το πρόβλημα - στις συγκεκριμένες συνθήκες - αλλά όχι να αποτελέσει προϋπάρχουσα γενική γνώση για την επίλυση του προβλήματος.
- 4) Αντιθέτως, το σύνολο των αρχικών δεδομένων εκπαίδευσης  $S_b$ , το οποίο έχει σχεδιαστεί με εκτενή πειραματισμό, μπορεί να αποτελεί αρχική συνθήκη για την επίλυση

του προβλήματος σε οποιεσδήποτε συνθήκες. Επομένως, σε κάθε νέα προσαρμογή του νευρωνικού δικτύου, εφόσον ελεγχθεί ότι είναι διαφορετική από τις προηγούμενες, δημιουργείται ένα νέο σύνολο επανεκπαίδευσης  $S_c$ , διαγράφοντας το παλαιό και τα νέα βάρη εκτιμώνται με βάση το νέο σύνολο δεδομένων  $S_c$  και το παλαιό  $S_b$ , το οποίο παραμένει σταθερό κατά τη λειτουργία του νευρωνικού δικτύου.

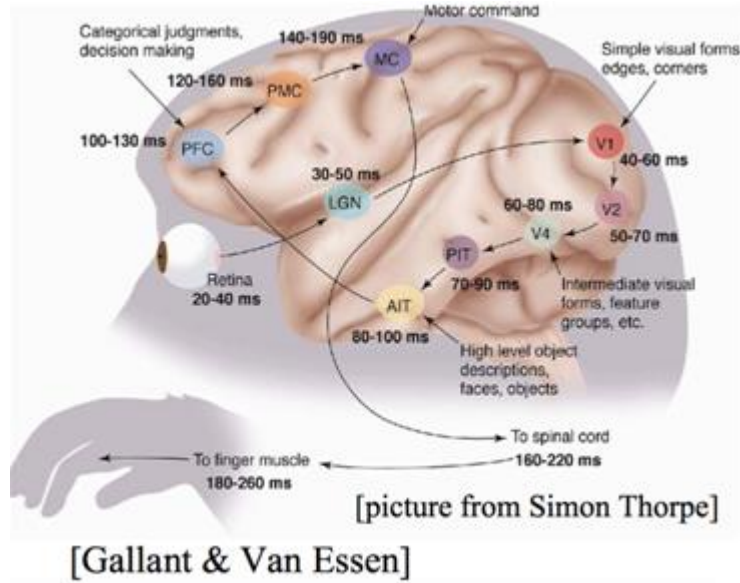
## 2.2 Βαθιές Αρχιτεκτονικές Νευρωνικών Δικτύων

### 2.2.1 Ορισμός και Ιδιότητες των Βαθιών Αρχιτεκτονικών

Μια βασική διαφορά των ρηχών (shallow) αρχιτεκτονικών νευρωνικών δικτύων από τις βαθιές (deep) αρχιτεκτονικές είναι η χρησιμοποίηση πολλών επιπέδων στην δομή του δικτύου και όχι μόνο δυο ή τριών, που, βεβαίως σύμφωνα με τα θεωρητικά αποτελέσματα επαρκούν για την αντιστοίχιση των δεδομένων εισόδου στις επιθυμητές αποκρίσεις του δικτύου στην έξοδο.

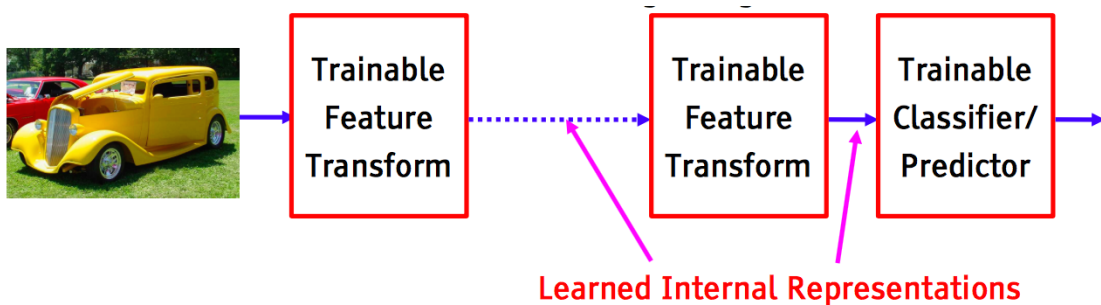
Τα δίκτυα αυτά προσομοιώνουν καλύτερα τις λειτουργίες της ανθρώπινης οπτικής αντίληψης, όπως φαίνεται στο Σχήμα 2.1, όπου αποτυπώνονται τα πολλαπλά επίπεδα λειτουργίας και οι ενδιάμεσες αναπαραστάσεις που δημιουργούνται, από την σύλληψη της εικόνας στον αμφιβληστροειδή χιτώνα μέχρι την ανθρώπινη αντίδραση στο ερέθισμα μέσω των μυών των χεριών ( Retina - LGN - V1 - V2 - V4 - PIT - AIT ...).





Σχήμα 2.1 Πολλαπλά επίπεδα λειτουργίας και ενδιάμεσες αναπαραστάσεις του συστήματος οπτικής αντίληψης.

Στο Σχήμα αυτό, κάθε μονάδα μετατρέπει την αναπαράσταση εισόδου της σε μια υψηλότερου επιπέδου. Τα χαρακτηριστικά υψηλού επιπέδου είναι πιο γενικά και πιο αμετάβλητα, ενώ τα χαρακτηριστικά χαμηλού επιπέδου βοηθούν στην κατηγοριοποίηση των εισόδων.



Σχήμα 2.2 Εκμάθηση ενδιάμεσων αναπαραστάσεων σε ένα πρόβλημα κατηγοριοποίησης/πρόβλεψης.

Τα τεχνητά συστήματα που προσομοιώνουν τέτοιες λειτουργίες έχουν ως στόχο να μαθαίνουν να δημιουργούν τις απαραίτητες ενδιάμεσες αναπαραστάσεις και να παράγουν με επιτυχία την τελική τους εκτίμηση, όπως φαίνεται στο Σχήμα 2.2.

Τα συστήματα αυτά μπορεί να βασίζονται σε μια από τις επόμενες τεχνικές:

- Επιβλεπόμενη μάθηση

Η αρχικοποίηση των παραμέτρων είναι τυχαία. η εκπαίδευση είναι επιβλεπόμενη, συνήθως τύπου στοχαστικής κλίσης καθόδου, με χρησιμοποίηση του αλγορίθμου backpropagation για υπολογισμό των κλίσεων, ενώ χρησιμοποιείται στα περισσότερα πρακτικά συστήματα για την αναγνώριση ομιλίας και εικόνας.

- Μη επιβλεπόμενη μάθηση με επιβλεπόμενο ταξινομητή στην έξοδο

Η μέθοδος εκπαιδεύει κάθε επίπεδο χωρίς επίβλεψη, το ένα μετά το άλλο, ενώ τελικά εκπαιδεύει έναν επιβλεπόμενο ταξινομητή στην κορυφή, διατηρώντας τα άλλα στρώματα σταθερά. Η μέθοδος έχει καλή απόδοση όταν είναι διαθέσιμα πολύ λίγα κατηγοριοποιημένα δείγματα.

- Μη επιβλεπόμενη μάθηση με επιβλεπόμενο εξομαλυντή

Η μέθοδος εκπαιδεύει κάθε επίπεδο χωρίς επίβλεψη, το ένα μετά το άλλο, προσθέτει ένα στρώμα ταξινομητή, και επανεκπαιδεύει το όλο πράγμα με επίβλεψη. Έχει καλή απόδοση όταν το σύνολο των ετικετών είναι φτωχό (π.χ. για ανίχνευση πεζών).

Ένα θεωρητικό δίλημμα που αφορά στην χρησιμοποίηση των βαθιών αρχιτεκτονικών νευρωνικών δικτύων αντί των ρηχών αφορά στο γεγονός ότι αφού μπορούμε να προσεγγίσουμε οποιαδήποτε συνάρτηση όσο περισσότερο/κοντά θέλουμε με χρήση ρηχών αρχιτεκτονικών, γιατί να χρησιμοποιήσουμε τις βαθιές.

Είναι γνωστό ότι οι μηχανές πυρήνα και τα νευρωνικά δίκτυα με 2 επίπεδα επιτυγχάνουν καθολική προσέγγιση των συναρτήσεων εισόδου-εξόδου.

$$y = \sum_{i=1}^P \alpha_i K(X, X^i) \quad (2.20)$$

$$y = F(W^1 . F(W^0 . X))$$

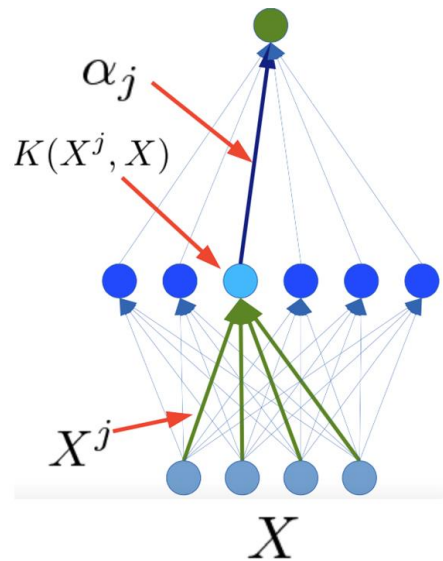
Οι μηχανές βαθιάς μάθησης

$$y = F(W^K . F(W^{K-1} . F(\dots F(W^0 . X) \dots))) \quad (2.21)$$

είναι πιο αποδοτικές για αναπαράσταση ορισμένων κατηγοριών συναρτήσεων, ιδιαίτερα εκείνων που συμμετέχουν στην οπτική αναγνώριση, καθόσον μπορούν να αντιπροσωπεύσουν περισσότερο πολύπλοκες συναρτήσεις με λιγότερο "υλικό"(hardware), σε πολλές εργασίες τεχνητής νοημοσύνης (υπολογιστική όραση, υπολογιστική ακοή, επεξεργασία φυσικής γλώσσας).

Τα μοντέλα 2 επιπέδων δεν είναι βαθιά (ακόμη και αν εκπαιδεύσουμε το πρώτο επίπεδο), επειδή δεν υπάρχει ιεραρχία χαρακτηριστικών. Ομοίως τα νευρωνικά δίκτυα με ένα κρυφό επίπεδο δεν είναι βαθιά, όπως και οι μηχανές διανυσμάτων υποστήριξης και οι μέθοδοι πυρήνα που εξετάζουμε στο επόμενο Κεφάλαιο. Τα ίδια ισχύουν και για τα δέντρα ταξινόμησης, όπου επίσης δεν υπάρχει ιεραρχία χαρακτηριστικών. Όλες οι αποφάσεις λαμβάνονται στο χώρο εισόδου. Μια τέτοια αρχιτεκτονική φαίνεται στο Σχήμα 2.3.

$$G(X, \alpha) = \sum_j \alpha_j K(X^j, X)$$



Σχήμα 2.3 Μια μηχανή διανυσμάτων υποστήριξης με βάση συναρτήσεις πυρήνα.

Η βαθιά μάθηση εμπεριέχει μη κυρτές συναρτήσεις σφάλματος. Η κυρτή μάθηση δεν εξαρτάται από την σειρά με την οποία παρουσιάζονται τα δείγματα (εξαρτάται μόνο από τις ασυμπτωτικές συχνότητες του δείγματος). Η ανθρώπινη μάθηση δεν είναι έτσι: μαθαίνουμε απλές έννοιες πριν μάθουμε πολύπλοκες. Η σειρά με την οποία μαθαίνουμε πράγματα έχει σημασία.

Όσον αφορά στη γενίκευση, στην πραγματικότητα, τα συνήθη όρια που μας δίνει η διάσταση Vapnik-Chervonenkis (VC) ισχύουν: τα περισσότερα συστήματα βαθιάς μάθησης έχουν μια πεπερασμένη διάσταση VC.

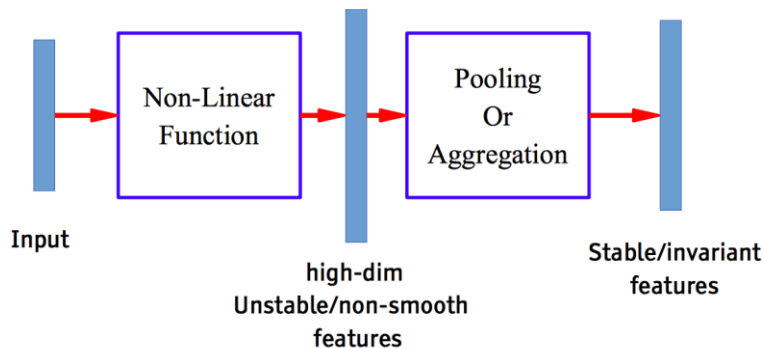
Σε σχέση με την απόδοση των τεχνικών βαθιάς μάθησης μπορούμε να αναφέρουμε τα επόμενα:

- Η βαθιά μάθηση έχει γίνει το πιο καυτό θέμα στην αναγνώριση ομιλίας κατά τα τελευταία 2 χρόνια.
- Η Microsoft και η Google έχουν αναπτύξει στα προϊόντα τους συστήματα αναγνώρισης ομιλίας βασισμένα σε βαθιά μάθηση.

- Η Microsoft, η Google, η IBM, η Nuance, η AT & T και όλοι οι μεγάλοι ακαδημαϊκοί και βιομηχανικοί φορείς στην αναγνώριση ομιλίας βασίζονται σε βαθιά μάθηση.
- Η βαθιά μάθηση είναι το πιο καυτό θέμα στην Υπολογιστική Όραση.

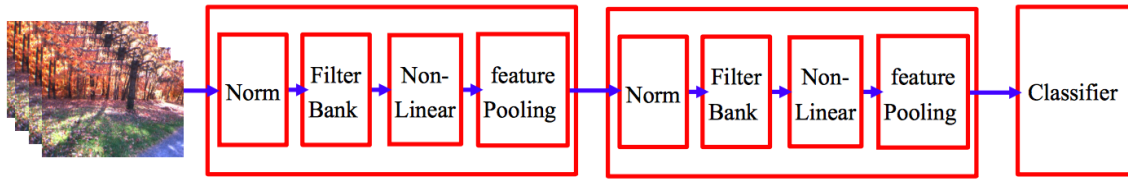
## 2.2.2 Εξαγωγή Χαρακτηριστικών με Βαθιές Αρχιτεκτονικές

Η λειτουργία των βαθιών αρχιτεκτονικών βασίζεται στην σταδιακή ενσωμάτωση της εισόδου με μη γραμμικό τρόπο σε χώρο μεγαλύτερων διαστάσεων. Στο νέο χώρο, τα πράγματα που ήταν μη διαχωρίσιμα μπορεί να γίνουν διαχωρίσιμα. Ακολουθεί pooling, δηλαδή, συγκέντρωση χαρακτηριστικών που είναι σημασιολογικά παρόμοια, όπως φαίνεται στο Σχήμα 2.4.



Σχήμα 2.4. Το βασικό δομικό στοιχείο των βαθιών αρχιτεκτονικών

Η συνολική αρχιτεκτονική συσσωρεύει πολλαπλά επίπεδα της επόμενης μορφής:



Σχήμα 2.5 Πολλαπλά επίπεδα μιας βαθιάς αρχιτεκτονικής.

Κάθε τέτοιο βασικό επίπεδο επιτελεί τα επόμενα:

- Κανονικοποίηση
- Εφαρμογή Τραπεζής Φίλτρων
- Μη-Γραμμικούς Μετασχηματισμούς
- Συγκέντρωση Χαρακτηριστικών.

Η Κανονικοποίηση αντιμετωπίζει παραλλαγές στη φωτεινότητα, υπολογίζοντας μέσους όρους, αφαιρέσεις, κάνοντας υπερατά φίλτραρίσματα (διέλευσης υψηλών συχνοτήτων) αλλά και τοπικές εξομαλύνσεις της αντίθεσης, με ομαλοποίηση των διακυμάνσεων.

Οι Τράπεζες Φίλτρων επεκτείνουν την διάσταση του προβλήματος, επιτελώντας προβολές σε πλήρεις βάσεις χαρακτηριστικών.

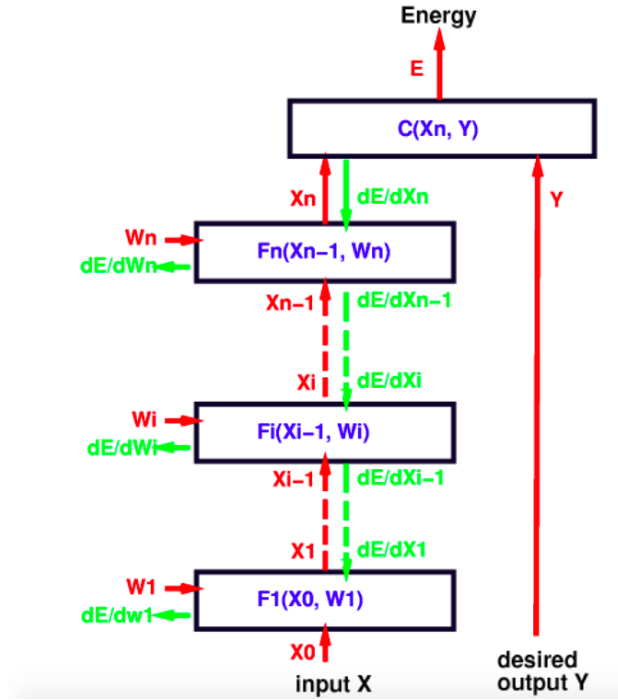
Οι Μη Γραμμικοί Μετασχηματισμοί επιτυγχάνουν αραίωση ή κορεσμό των δεδομένων, χρησιμοποιώντας συναρτήσεις ενεργοποίησης όπως η Relu, η υπερβολική εφαπτομένη (tanh), τη μέθοδο όπου ο νικητής τα παίρνει όλα (winner-takes-all).

Η Συγκέντρωση Χαρακτηριστικών επιτυγχάνει συνάθροιση πάνω σε κάποιο διάστημα ή τύπο χαρακτηριστικών, υπολογίζοντας νόρμες και πιθανότητες, όπως οι επόμενες:

$$L_p: \sqrt[p]{X_i^p}$$

$$PROB: \frac{1}{b} \log \left( \sum_i e^{bX_i} \right) \quad (2.22)$$

### 2.2.3 Εκπαίδευση του Συστήματος με Βαθιά Αρχιτεκτονική



Σχήμα 2.6 Μια πολυεπίπεδη αρχιτεκτονική.

Οι βαθιές αρχιτεκτονικές μπορούν να φτιαχτούν με τη συναρμολόγηση επιμέρους επιπέδων. Ένα απλό παράδειγμα είναι η διαδοχική πολυεπίπεδη feed-forward αρχιτεκτονική που φαίνεται στο Σχήμα 2.6.

$$\begin{aligned}
 X &= X_0 \\
 X_i &= F_i(X_{i-1}, W_i) \quad \forall i \in [1, n] \\
 E(Y, X, W) &= C(X_n, Y)
 \end{aligned} \quad (2.23)$$

Για να εκπαιδεύσουμε την αρχιτεκτονική αυτή πρέπει να υπολογίσουμε την παράγωγο της  $E$  ως προς όλες τις παραμέτρους του συστήματος (δηλαδή ως προς όλα τα βάρη  $W_i$ ). Ας σκεφτούμε το επίπεδο  $i$  του οποίου το forward propagation βήμα υπολογίζει το :

$$X_i = F_i(X_{i-1}, W_i) \quad (2.24)$$

Ας υποθέσουμε πως ξέρουμε ήδη το:

$$\frac{\partial E}{\partial X_i} \quad (2.25)$$

Μπορούμε να εφαρμόσουμε τον κανόνα αλυσίδας και έτσι :

$$\frac{\partial E}{\partial W_i} = \frac{\partial E}{\partial X_i} \frac{\partial F_i(X_{i-1}, W_i)}{\partial W_i} \quad (2.26)$$

με διαστάσεις

$$[1 \times N_w] = [1 \times N_x] \cdot [N_x \times N_w] \quad (2.27)$$

όπου :

$$\frac{\partial F_i(X_{i-1}, W_i)}{\partial W_i} \quad (2.28)$$

είναι ο Ιακωβιανός πίνακας του  $F_i$  ως προς το  $W_i$ .

$$\left[ \frac{\partial F_i(X_{i-1}, W_i)}{\partial W_i} \right]_{kl} = \frac{\partial [F_i(X_{i-1}, W_i)]_k}{\partial [W_i]_l} \quad (2.29)$$



Χρησιμοποιώντας το ίδιο τέχνασμα μπορούμε να υπολογίσουμε το :

$$\frac{\partial E}{\partial X_{i-1}} \quad (2.30)$$

Ας υποθέσουμε πάλι πως ξέρουμε ήδη το :

$$\frac{\partial E}{\partial X_i} \quad (2.31)$$

Μπορούμε να εφαρμόσουμε τον κανόνα αλυσίδας και έτσι :

$$\frac{\partial E}{\partial X_{i-1}} = \frac{\partial E}{\partial X_i} \frac{\partial F_i(X_{i-1}, W_i)}{\partial X_{i-1}} \quad (2.32)$$

όπου :

$$\frac{\partial F_i(X_{i-1}, W_i)}{\partial X_{i-1}} \quad (2.33)$$

είναι ο Ιακωβιανός πίνακας του  $F_i$  αναφορικά με το  $X_{i-1}$ .

Το  $F_i$  έχει δύο Ιακωβιανούς πίνακες επειδή έχει 2 ορίσματα. Η ανωτέρω εξίσωση είναι αναδρομική.

Αν κάποιος πάρει την παράγωγο ενός οριζόντιου διανύσματος, τότε προκύπτει ένα κάθετο διάνυσμα, διαστάσεων :

$$[1 \times N_{i-1}] = [1 \times N_i] * [N_i \times N_{i-1}] \quad (2.34)$$

Οπότε :

$$\frac{\partial E}{\partial X_{i-1}} = \frac{\partial E}{\partial X_i} \frac{\partial F_i(X_{i-1}, W_i)}{\partial X_{i-1}} \quad (2.35)$$

με διαστάσεις :

$$[1 \times N_{wi}] = [1 \times N_i] * [N_i \times N_{wi}] \quad (2.36)$$

$$\frac{\partial E}{\partial W_i} = \frac{\partial E}{\partial X_i} \frac{\partial F_i(X_{i-1}, W_i)}{\partial W} \quad (2.37)$$

Μπορούμε να γράψουμε τις εξισώσεις αυτές με κάθετα διανύσματα :

$$\frac{\partial E}{\partial X_{i-1}}' = \frac{\partial F_i(X_{i-1}, W_i)'}{\partial X_{i-1}} \frac{\partial E}{\partial X_i}' \quad (2.38)$$

$$\frac{\partial E}{\partial W_i}' = \frac{\partial F_i(X_{i-1}, W_i)'}{\partial W} \frac{\partial E}{\partial X_i}'$$

Για να υπολογίσουμε όλες τις παραγώγους χρησιμοποιούμε ένα backward sweep που βασικά είναι ο αλγόριθμος back-propagation που χρησιμοποιεί την επαναληπτική εξίσωση για το :

$$\frac{\partial E}{\partial X_{i-1}}$$

Οπότε έχουμε :

$$\frac{\partial E}{\partial X_n} = \frac{\partial C(X_n, Y)}{\partial X_n}$$

$$\frac{\partial E}{\partial X_{n-1}} = \frac{\partial E}{\partial X_n} \frac{\partial F_n(X_{n-1}, W_n)}{\partial X_{n-1}}$$

$$\frac{\partial E}{\partial W_n} = \frac{\partial E}{\partial X_n} \frac{\partial F_n(X_{n-1}, W_n)}{\partial W_n} \quad (2.39)$$

$$\frac{\partial E}{\partial X_{n-2}} = \frac{\partial E}{\partial X_{n-1}} \frac{\partial F_{n-1}(X_{n-2}, W_{n-1})}{\partial X_{n-2}}$$

$$\frac{\partial E}{\partial W_{n-1}} = \frac{\partial E}{\partial X_{n-1}} \frac{\partial F_{n-1}(X_{n-2}, W_{n-1})}{\partial W_{n-1}}$$

και ούτω καθεξής μέχρι να φτάσουμε στο πρώτο επίπεδο.

Άρα τώρα έχουμε όλα τα :

$$\frac{\partial E}{\partial W_i}$$

για όλα τα  $i$  που ανήκουν στο διάστημα  $[1, n]$ .

Στην περίπτωση ενός Γραμμικού Νευρώνα, το διάνυσμα εισόδου πολλαπλασιάζεται με τον πίνακα βαρών  $W$ :

$$X_{\text{out}} = W X_{\text{in}} \quad (2.40)$$

Εκτελώντας backward propagation στην είσοδο:

$$\frac{\partial E}{\partial X_{\text{in}}} = \frac{\partial E}{\partial X_{\text{out}}} \frac{\partial X_{\text{out}}}{\partial X_{\text{in}}} = \frac{\partial E}{\partial X_{\text{out}}} W \quad (2.41)$$

και παίρνοντας τον ανάστροφο, λαμβάνουμε τα κάθετα διανύσματα :

$$\frac{\partial E'}{\partial X_{in}} = W' \frac{\partial E'}{\partial X_{out}} \quad (2.42)$$

Ενώ εκτελώντας backward propagation στα βάρη λαμβάνουμε:

$$\frac{\partial E}{\partial W_{ij}} = \frac{\partial E}{\partial X_{outi}} \frac{\partial X_{outi}}{\partial W_{ij}} = X_{inj} \frac{\partial E}{\partial X_{outi}} \quad (2.43)$$

Αυτό μπορούμε να το γράψουμε και ως εξωτερικό γινόμενο :

$$\frac{\partial E'}{\partial W} = \frac{\partial E'}{\partial X_{out}} X'_{in} \quad (2.44)$$

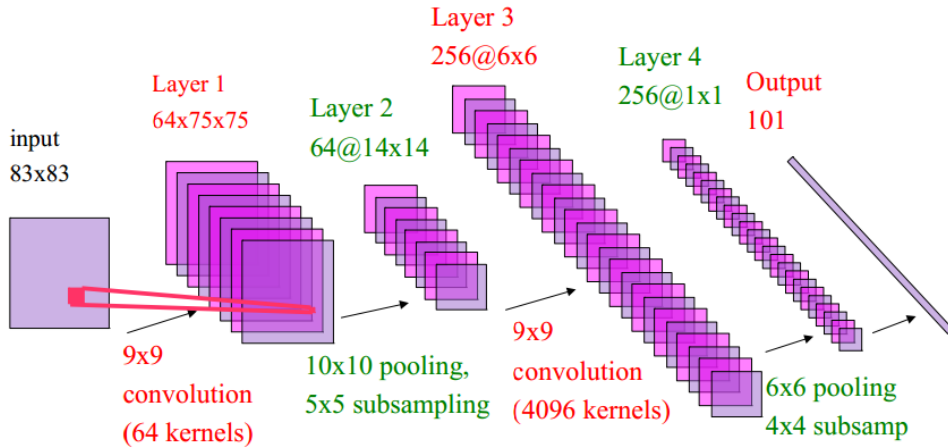
Αντίστοιχα ισχύουν και όταν χρησιμοποιείται η tanh ως συνάρτηση ενεργοποίησης.

## 2.3 Βαθιά Συνελικτικά Δίκτυα

Τα βαθιά συνελικτικά νευρωνικά δίκτυα έχουν αναπτυχθεί και εφαρμοστεί σε πολλές πρακτικές εφαρμογές, όπως στην αναγνώριση εικόνας, στην αναγνώριση ομιλίας, στον χαρακτηρισμό (tagging) φωτογραφιών της Google και της Baidu. Έχουν κερδίσει σε αρκετούς διαγωνισμούς, όπως στο IMAGENet, σε πολυτροπική μάθηση, σε αναγνώριση γερμανικών σημάτων κυκλοφορίας, χειρογράφων, κλπ.

Τα δίκτυα αυτά εφαρμόζονται σε δεδομένα σε μορφή πίνακα, στα οποία κοντινές (γειτονικά/γεωγραφικά) τιμές συσχετίζονται. Τέτοια δεδομένα περιλαμβάνουν εικόνες, ήχο, αναπαραστάσεις χρόνου-συχνότητας, βίντεο, ογκομετρικές εικόνες, εικόνες RGB-βάθους, κλπ.

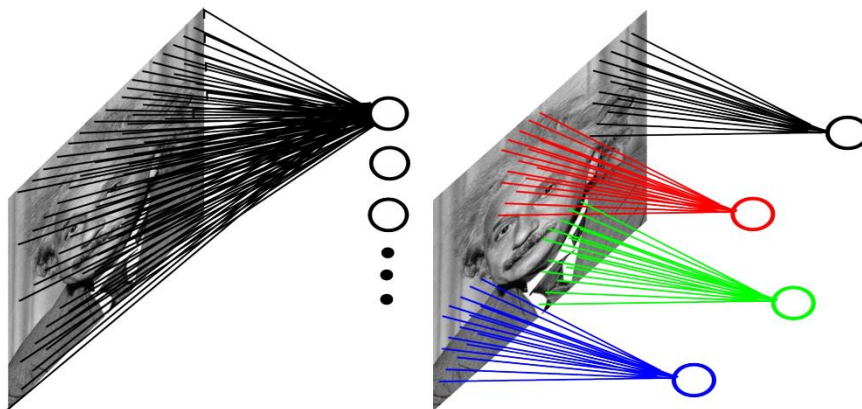
Ένα παράδειγμα με ένα βαθύ συνελκτικό δίκτυο το οποίο εκπαιδεύεται με αλγορίθμους επιβλεπόμενης μάθησης φαίνεται στο Σχήμα 2.7.



Σχήμα 2.7 Ένα παράδειγμα με βαθύ συνελκτικό δίκτυο

Αν, για παράδειγμα, θεωρήσουμε μια εικόνα διαστάσεων 200x200 εικονοστοιχείων, και ένα δίκτυο πλήρως συνδεδεμένο με 400.000 κρυφούς νευρώνες, προκύπτουν 16.000.000.000 παράμετροι. Αν το δίκτυο είναι τοπικά συνδεδεμένο με 400.000 κρυφούς νευρώνες 10x10 πεδίων, προκύπτουν 40 εκατομμύρια παράμετροι.

Τοπικές συνδέσεις μπορούν και συλλαμβάνουν τοπικές εξαρτήσεις



Σχήμα 2.8 Τοπικές συνδέσεις σε ένα βαθύ συνελκτικό δίκτυο

Όλοι οι νευρώνες μοιράζονται το ίδιο σύνολο των βαρών, επικεντρώνοντας σε διαφορετικά μέρη της εικόνας (weight sharing, regions of interest). Έτσι ανιχνεύονται σεν χαρακτηριστικών σε διάφορα μέρη της εικόνας. Ως αποτέλεσμα, όταν η είσοδος μετατοπίζεται, η έξοδος επίσης μετατοπίζεται αλλά παραμένει αμετάβλητη.

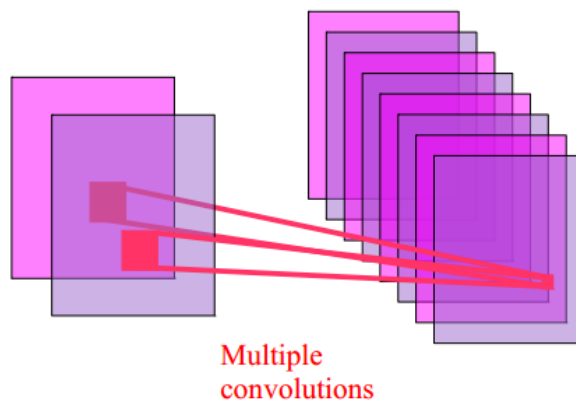
Κάθε νευρώνας εκτελεί μια συνέλιξη με έναν εκπαιδευμένο πυρήνα (ή φίλτρο). Ένας τέτοιος πυρήνας βασίζεται στη συνάρτηση Relu (περιορισμένη γραμμική)

$$A_{ij} = \sum_{kl} W_{kl} X_{i+j, k+l} \quad (2.45)$$

Η φιλτραρισμένη εικόνα  $Z$  θεωρείται ως ένας χάρτης χαρακτηριστικών (feature map).

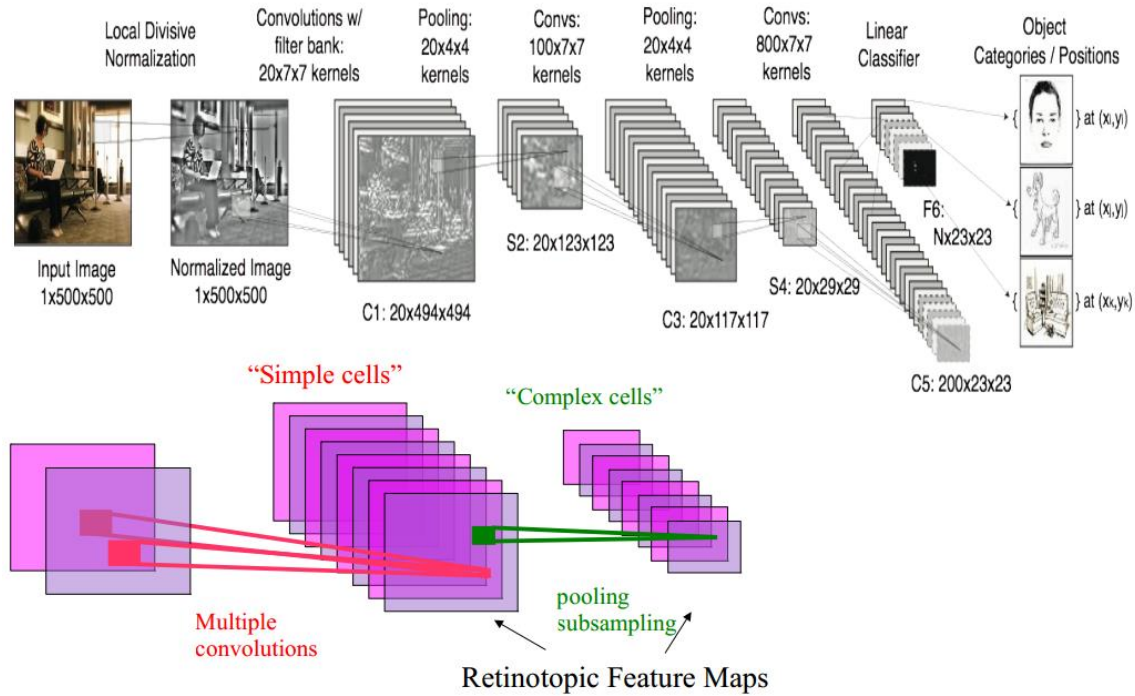
$$Z_{ij} = \max(0, A_{ij}) \quad (2.46)$$

Το αποτέλεσμα είναι ένας τρισδιάστατος (3-Δ) πίνακας, του οποίου κάθε 'φέτα' είναι ένας χάρτης χαρακτηριστικών, όπως φαίνεται στο Σχήμα 2.9.



Σχήμα 2.9 Η βασική δομή του συνελκτικού δικτύου.

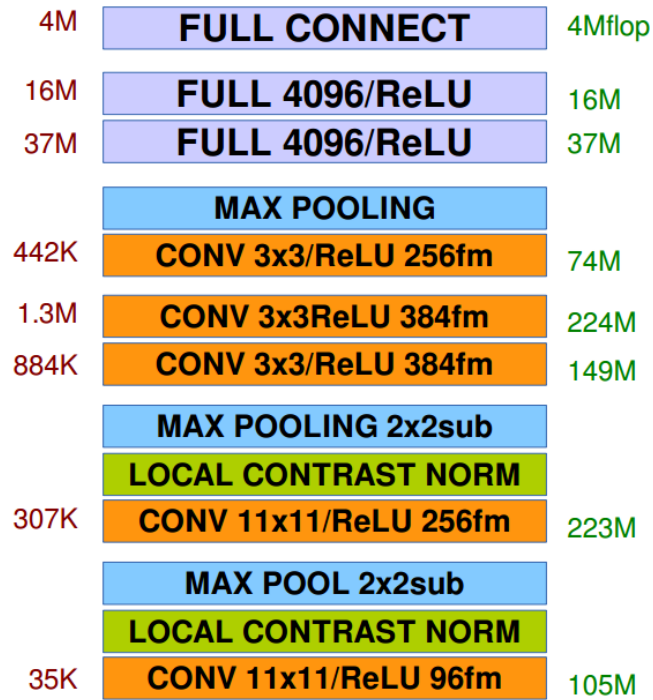
Στο επόμενο Σχήμα φαίνεται μια συνολική δομή συνελκτικού δικτύου, η οποία λαμβάνει μια εικόνα αντικειμένου στην είσοδό της και την κατηγοριοποιεί στην έξοδό της.



Σχήμα 2.10 Κατηγοριοποίηση εικόνων αντικειμένων με βαθύ συνελκτικό δίκτυο.

Τα βαθιά συνελκτικά δίκτυα έχουν επιλεγεί ως η καλύτερη μέθοδος σε μια σειρά από εφαρμογές, όπως η αναγνώριση χειρογράφων MNIST, αριθμών σπιτιών στο StreetView, αναγνώριση σημάτων κυκλοφορίας [2011] διαγωνισμός GTSRB (IDSIA, NYU), ανίχνευση πεζών σε σύνολα δεδομένων του INRIA και άλλα (NYU, 2013), σε ογκομετρική τμηματοποίηση εικόνων εγκεφάλου (IDSIA, MIT, 2009), σε αναγνώριση αντικειμένων IMAGEnet (2012), σε αναγνώριση ομιλίας και ακουστική μοντελοποίηση (IBM και Google, 2012), σε ανίχνευση καρκίνου του μαστού - μίωση των κυττάρων ΜΙΤΟΣ (IDSIA, 2011), κλπ.

Στο επόμενο Σχήμα φαίνεται το υπολογιστικό φορτίο μιας βαθιάς αρχιτεκτονικής που βασίζεται σε συνελκτικές μονάδες.



Σχήμα 2.11 Υπολογιστικό φορτίο σε ένα βαθύ συνελκτικό δίκτυο.



# 3.Μεταφορά Συμβολικής Γνώσης σε Μηχανές Διανυσμάτων Υποστήριξης

---

Στο παρόν Κεφάλαιο γίνεται αναφορά στις τεχνικές μηχανικής μάθησης, με έμφαση στις συναρτήσεις και μηχανές πυρήνα, ενώ περιγράφονται σύντομα οι Μηχανές Διανυσμάτων Υποστήριξης (support vector machines). Στο πλαίσιο αυτό παρουσιάζονται μεθοδολογίες για τη σύνδεση αναπαραστάσεων γνώσης με συστήματα μηχανικής μάθησης, με βασικό στόχο τον ορισμό πυρήνων που εμπεριέχουν την ομοιότητα μεταξύ των ατόμων μιας οντολογίας, με εκμετάλλευση των σημασιολογικών χαρακτηριστικών της γλώσσας που χρησιμοποιείται για αναπαράσταση.

## 3.1 Συναρτήσεις και Μηχανές Πυρήνα

### 3.1.1 Μέθοδοι Πυρήνα

Στη συνέχεια θα αναφερθούμε στις μεθόδους πυρήνα (kernel methods) και θα αναλύσουμε τα βασικότερα στοιχεία τους που θα μας χρειαστούν στη συνέχεια ([47-50]).

Οι μέθοδοι πυρήνα βασίζονται στον προσδιορισμό ενός θετικά ορισμένου πυρήνα στο χώρο προτύπων. Η χρήση του πυρήνα αυτού οδηγεί σε μετασχηματισμό του χώρου προτύπων σε ένα γραμμικό χώρο (συνήθως μεγάλης διάστασης), στον οποίο οι πυρήνες ζευγών προτύπων υλοποιούνται με εσωτερικά γινόμενα. Το βασικό υπολογιστικό πλεονέκτημα των μεθόδων πυρήνα είναι ότι μπορούν να εφαρμοστούν σε χώρους χαρακτηριστικών μεγάλων διαστάσεων, χωρίς να υποφέρουν από το κόστος του ακριβούς υπολογισμού των μετασχηματισμένων δεδομένων.

Οι μέθοδοι πυρήνα αποτελούνται από δυο συστατικά, την μηχανή πυρήνα (kernel machine) και τη συνάρτηση πυρήνα (kernel function).

Διαφορετικές μηχανές πυρήνα χρησιμοποιούν διαφορετικές διαδικασίες μάθησης, όπως π.χ., τα support vector machines εμπεριέχουν επιβλεπόμενη (supervised) μάθηση, ενώ το support vector clustering εμπεριέχει μη επιβλεπόμενη (unsupervised) μάθηση. Θα πρέπει να σημειωθεί ότι η μηχανή πυρήνα περιλαμβάνει τον αλγόριθμο μάθησης και τον τρόπο που αναζητείται η λύση, ενώ η συνάρτηση πυρήνα περιλαμβάνει την γλώσσα των υποθέσεων (hypothesis language), δηλαδή τον τρόπο που φτιάχνεται το σύνολο από τις πιθανές λύσεις. Διαφορετικές συναρτήσεις πυρήνα υλοποιούν διαφορετικούς χώρους υποθέσεων, ή ακόμα και διαφορετικές αναπαραστάσεις της γνώσης μας.

Πιο συγκεκριμένα, οι μέθοδοι πυρήνα ακολουθούν τα παρακάτω βήματα:

- Τα δεδομένα μεταφέρονται σε ένα νέο διανυσματικό χώρο που αναφέρεται και ως χώρος χαρακτηριστικών (feature space).
- Αναζητούνται γραμμικές σχέσεις ανάμεσα στις αναπαραστάσεις των δεδομένων στο χώρο χαρακτηριστικών.
- Οι υπολογισμοί γίνονται με τέτοιο τρόπο, ώστε δεν χρειάζονται οι συντεταγμένες των σημείων στον (μεγαλύτερης διάστασης) διανυσματικό χώρο, πάρα μόνο τα εσωτερικά γινόμενα των σημείων αυτών ανά δύο.
- Τα εσωτερικά γινόμενα των σημείων ανά δύο μπορούν να υπολογιστούν αποτελεσματικά (efficiently) από τα αρχικά δεδομένα μέσω της συνάρτησης πυρήνα.

Μια σημαντική παρατήρηση είναι ότι ο αλγόριθμος μάθησης (inductive bias) και η επιλογή της συνάρτησης πυρήνα (language bias) είναι ανεξάρτητα μεταξύ τους.

Ειδικότερα, οι μέθοδοι πυρήνα εφαρμόζουν στα δεδομένα εισόδου μια αντιστοίχιση

$$\Phi: \mathbf{x} \in \mathbb{R}^n \mapsto \Phi(\mathbf{x}) \in H \subseteq \mathbb{R}^n \quad (3.1)$$

με στόχο να μετατρέψουν τις μη γραμμικές σχέσεις στο χώρο εισόδου σε γραμμικές στο χώρο των χαρακτηριστικών, στον οποίο να μπορεί να χρησιμοποιηθεί το εσωτερικό γινόμενο ως μέτρο ομοιότητας.

Η αντιστοίχιση των δεδομένων στο χώρο  $H$  μέσω του μετασχηματισμού  $\Phi$  έχει τα ακόλουθα πλεονεκτήματα:

- Επιτρέπει τον καθορισμό ενός μέτρου ομοιότητας από το εσωτερικό γινόμενο στο χώρο  $H$ ,

$$k(\mathbf{x}, \mathbf{x}') := \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle = \Phi^T(\mathbf{x}) \Phi(\mathbf{x}') \quad (3.2)$$

- Επιτρέπει την γεωμετρική θεώρηση των δεδομένων, και με αυτόν τον τρόπο, την μελέτη των αλγορίθμων μάθησης με βάση τη γραμμική άλγεβρα και την αναλυτική γεωμετρία.
- Η ελευθερία της επιλογής της αντιστοίχισης  $\Phi$  επιτρέπει τη σχεδίαση μιας μεγάλης ποικιλίας μέτρων ομοιότητας και αλγορίθμων μάθησης. Η ελευθερία αυτή ισχύει και στην περίπτωση που οι είσοδοι  $x_i$  υπάρχουν ήδη σε χώρο εσωτερικού γινομένου. Στην τελευταία περίπτωση είναι μεν δυνατή η απευθείας χρήση του εσωτερικού γινομένου ως μέτρου ομοιότητας, όμως είναι δυνατή η κατ' αρχήν εφαρμογή ενός μη γραμμικού μετασχηματισμού  $\Phi$  για την αλλαγή της αναπαράστασης και την μετατροπή της σε μια που είναι πλέον κατάλληλη για το συγκεκριμένο πρόβλημα.

### 3.1.2 Συναρτήσεις Πυρήνα

Ανωτέρω περιγράψαμε πώς προκύπτει ένας πυρήνας ως μέτρο ομοιότητας που μπορεί να θεωρηθεί σαν ένα εσωτερικό γινόμενο στο χώρο χαρακτηριστικών. Στη συνέχεια, θα μελετήσουμε την κλάση των πυρήνων  $k$  που αντιστοιχούν σε εσωτερικά γινόμενα στο χώρο χαρακτηριστικών  $H$ , μέσω του μετασχηματισμού  $\Phi$ ,

$$\Phi: X \rightarrow H \quad (3.3)$$

Η κλάση αυτή ορίζεται ως:

$$k(\mathbf{x}, \mathbf{x}') := \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle = \Phi^T(\mathbf{x}) \Phi(\mathbf{x}') \quad (3.4)$$

Η χρήση των πυρήνων μας δίνει τη δυνατότητα να υπολογίσουμε το εσωτερικό γινόμενο ανάμεσα στις αναπαραστάσεις δύο προτύπων εισόδων στο χώρο χαρακτηριστικών μεγάλης διάστασης, χωρίς να υπολογίσουμε με ακρίβεια τις αναπαραστάσεις τους.

Η συνάρτηση  $k$ , η οποία επιστρέφει το εσωτερικό γινόμενο μεταξύ των αναπαραστάσεων δύο εισόδων στο χώρο των χαρακτηριστικών, λέγεται συνάρτηση πυρήνα.

Δύο απαιτήσεις υπάρχουν για την επιλογή της συνάρτησης πυρήνα  $k(\mathbf{x}, \mathbf{x}')$ . Καταρχάς, θα πρέπει να εμπεριέχει το μέτρο της ομοιότητας που είναι κατάλληλη για το εκάστοτε πρόβλημα. Κατά δεύτερο λόγο, ο υπολογισμός της θα πρέπει να είναι λιγότερο πολύπλοκος από τον υπολογισμό του αντίστοιχου μετασχηματισμού  $\Phi(\mathbf{x})$ .

Μια συνάρτηση πυρήνα, από την μορφή της, προσδιορίζει ένα χώρο χαρακτηριστικών, τον οποίο σε πολλές περιπτώσεις δεν χρειάζεται να κατασκευάσουμε αναλυτικά. Με άλλα λόγια, μπορούμε να ορίσουμε πυρήνες χωρίς να υπολογίσουμε πλήρως τον χώρο χαρακτηριστικών. Αυτό συμβαίνει όταν η διαθέσιμη γνώση μας παρέχει έναν τρόπο να συγκρίνουμε δύο εισόδους.

Το αν υπάρχει μια συνάρτηση  $k : X \times X \rightarrow \mathbb{R}$  και ένας μετασχηματισμός χαρακτηριστικών  $\Phi : \mathbf{x} \rightarrow H$  (σε ένα χώρο Hilbert  $H$ ), έτσι ώστε:

$$k(\mathbf{x}, \mathbf{x}') = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle = \Phi^T(\mathbf{x}) \Phi(\mathbf{x}') \quad (3.5)$$

για οποιαδήποτε πρότυπα  $\mathbf{x}, \mathbf{x}' \in X$ , μπορεί να πιστοποιηθεί ελέγχοντας ότι ο  $k$  είναι θετικά ορισμένος.

Για πληρότητα παραθέτουμε τους ορισμούς για το πότε ένας πυρήνας θεωρείται θετικά ορισμένος.

#### Ορισμός 1: Πίνακας Gram (Gram Matrix)

Δοθείσας μιας συνάρτησης  $k : X^2 \rightarrow \mathbb{R}$ , ή  $k : X^2 \rightarrow \mathbb{C}$  και των προτύπων  $x_1, \dots, x_m \in X$ , ο  $m \times m$  πίνακας  $K$  με στοιχεία  $K_{ij} := k(x_i, x_j)$  ονομάζεται πίνακας Gram ή πίνακας πυρήνα (kernel matrix) του  $k$  ως προς τα  $x_1, \dots, x_m$ .

#### Ορισμός 2: Θετικά ορισμένος πίνακας (Positive Definite Matrix)

Ένας μιγαδικός  $m \times m$  πίνακας  $K$  που πληροί τις σχέσεις:

$$\sum_{i,j} \overline{c_j} c_i K_{ij} \geq 0 \quad (3.6)$$

για όλα τα  $c_i \in \mathbb{C}$  ονομάζεται θετικά ορισμένος. Ομοίως, ένας πραγματικός συμμετρικός  $m \times m$  πίνακας  $K$  που πληροί την ως άνω σχέση, για όλα τα  $c_i \in \mathbb{R}$ , ονομάζεται θετικά ορισμένος.

#### Ορισμός 3: Θετικά ορισμένος Πυρήνας (Positive Definite Kernel)

Έστω  $X$  ένα μη κενό σύνολο. Μια συνάρτηση  $k$  στο  $X \times X$ , η οποία,  $\forall m \in \mathbb{N}$  και  $\forall x_1, \dots, x_m \in X$ , οδηγεί σε θετικά ορισμένο πίνακα Gram, ονομάζεται θετικά ορισμένος πυρήνας.

Δεν είναι πάντα εύκολο να αποδείξουμε ότι ένας πυρήνας είναι θετικά ορισμένος με βάση τον ορισμό. Για το λόγο αυτό, πολλές φορές χρησιμοποιούμε την ιδιότητα του κλεισίματος (closure) της οικογένειας των θετικά ορισμένων πυρήνων. Συγκεκριμένα, οι πυρήνες είναι κλειστοί ως προς το άθροισμα, τον πολλαπλασιασμό με μια σταθερά, το γινόμενο, την επέκταση με μηδενικά και την ύψωση σε δύναμη.

### 3.1.3 Είδη Πυρήνων

Πριν αναλύσουμε τους πυρήνες για δομημένα δεδομένα, θα αναφέρουμε περιληπτικά τους παραδοσιακά χρησιμοποιούμενους πυρήνες σε διανυσματικούς χώρους.

#### Πυρήνες σε διανυσματικούς χώρους

Έστω τα διανύσματα  $x, x' \in \mathbb{R}^n$  και έστω  $\langle \cdot, \cdot \rangle$  το εσωτερικό γινόμενο στο  $\mathbb{R}^n$ . Εκτός από τον γραμμικό πυρήνα

$$k(x, x') = \langle x, x' \rangle \quad (3.7)$$

και τον κανονικοποιημένο γραμμικό πυρήνα

$$k(x, x') = \frac{\langle x, x' \rangle}{\sqrt{\|x\| \|x'\|}} \quad (3.8)$$

οι δύο πιο συχνά χρησιμοποιούμενοι πυρήνες σε διανυσματικούς χώρους είναι ο πολυωνυμικός πυρήνας και ο γκαουσιανός RBF πυρήνας. Δεδομένων δύο παραμέτρων,  $c \in \mathbb{R}$ ,  $p \in \mathbb{N}^+$ , ο πολυωνυμικός πυρήνας ορίζεται ως

$$k(x, x') = (\langle x, x' \rangle + c)^p \quad (3.9)$$

Το σκεπτικό πίσω από αυτόν τον πυρήνα είναι ότι είναι χρήσιμο να παράγονται νέα χαρακτηριστικά σαν γινόμενα αρχικών χαρακτηριστικών. Η παράμετρος  $p$  είναι η

μέγιστη τάξη των μονώνυμων που δημιουργούν το νέο χώρο χαρακτηριστικών, ενώ το  $c$  μπορεί να χρησιμοποιηθεί ως ένα βήμα προς μονώνυμα χαμηλής τάξης – για παράδειγμα αν  $c = 0$ , ο χώρος χαρακτηριστικών αποτελείται μόνο από μονώνυμα τάξης  $p$  των αρχικών χαρακτηριστικών.

Δεδομένης μιας παραμέτρου  $\gamma$ , ο γκαουσιανός RBF πυρήνας ορίζεται ως εξής :

$$k(x, x') = e^{-\gamma \|x-x'\|^2} \quad (3.10)$$

Αν χρησιμοποιήσουμε αυτόν τον πυρήνα σε μηχανές, όπως τα Support Vector Machines, είναι σαν να χρησιμοποιούμε ένα (νευρωνικό) δίκτυο με γκαουσιανούς πυρήνες, κεντραρισμένους στα διανύσματα υποστήριξης (support vectors). Οι αναπαραστάσεις των σημείων ξεκινώντας από τον διανυσματικό χώρο  $\mathbb{R}^n$  μέσω του μετασχηματισμού  $\Phi : \mathbb{R}^n \rightarrow H$  με  $k(x, x') = \langle \Phi(x), \Phi(x') \rangle$  κείνται όλες στην επιφάνεια μιας υπερσφαίρας στον χώρο Hilbert  $H$ .

Ο ακόλουθος πυρήνας μπορεί να χρησιμοποιηθεί με διακριτά δεδομένα. Ο πυρήνας ταιριάσματος (matching)  $k_d : X \times X \rightarrow \mathbb{R}$  ορίζεται ως ακολούθως:

$$k_d(x, x') = \begin{cases} 1 & \text{αν } x = x' \\ 0 & \text{αλλιώς} \end{cases} \quad (3.11)$$

Η αναπαράσταση κάθε στοιχείου του  $X$  μέσω του μετασχηματισμού  $\Phi : X \rightarrow H$ , με  $k_d(x, x') = \langle \Phi(x), \Phi(x') \rangle$ , είναι ένα διάνυσμα ορθογώνιο προς όλες τις άλλες αναπαραστάσεις των στοιχείων του  $X$  μέσω του  $\Phi$ .

## Πυρήνες σε δομημένα δεδομένα

Η ιδέα για τον ορισμό πυρήνων σε δομημένα δεδομένα είναι πρώτα να ορίσουμε έναν τρόπο να συγκρίνουμε συστατικά των δεδομένων και στη συνέχεια να συνδυάσουμε τους υπο-πυρήνες που αντιστοιχούν σε ένα σύνολο από διαμερίσεις των δεδομένων. Ένας τύπος δεδομένων θεωρείται δομημένος (structured), αν είναι δυνατόν να τον διαχωρίσουμε σε μικρότερα τμήματα. Η ιδέα είναι να υπολογίσουμε το γινόμενο των υπο-πυρήνων, συγκρίνοντας τα τμήματα, και να αθροίσουμε στο σύνολο όλων των επιτρεπόμενων διαμερίσεων.

Ο πιο γνωστός πυρήνας, για χώρους αναπαράστασης που δεν είναι διανυσματικής (attribute-value) μορφής, είναι ο συνελκτικός πυρήνας (convolution kernel) που προτάθηκε από τον Haussler το 1999. Η βασική ιδέα του είναι ότι η σημασιολογία σύνθετων αντικειμένων μπορεί να εκφραστεί μέσα από μια σχέση  $R$  ανάμεσα στο αντικείμενο και στα τμήματά του. Έστω ότι  $x, x' \in X$  είναι τα αντικείμενα, και  $\mathbf{x}, \mathbf{x}' \in X_1 \times \dots \times X_D$  είναι τμήματα αυτών των αντικειμένων. Δεδομένης της σχέσης  $R : (X_1 \times \dots \times X_D) \times X$ , μπορούμε να ορίσουμε τις διαμερίσεις  $R^{-1}(x) = \{\mathbf{x} : R(\mathbf{x}, x)\}$ . Για θετικά ορισμένους πυρήνες  $k_d : X_d \times X_d \rightarrow \mathbb{R}$ , ο συνελκτικός πυρήνας ορίζεται ως εξής:

$$k_{conv}(x, x') = \sum_{\mathbf{x} \in R^{-1}(x), \mathbf{x}' \in R^{-1}(x')} \prod_{d=1}^D k_d(x_d, x'_d) \quad (3.12)$$

Ο όρος ‘συνελκτικός πυρήνας’ αναφέρεται σε μια κλάση πυρήνων που μπορούν να σχηματιστούν με τον παραπάνω τρόπο. Το πλεονέκτημα των συνελκτικών πυρήνων είναι ότι είναι πολύ γενικοί και μπορούν να εφαρμοστούν σε πολλές διαφορετικές περιπτώσεις. Παρόλα αυτά εξαιτίας αυτής τους της γενικότητάς τους, απαιτείται αρκετή προσπάθεια για να προσαρμοστούν σε ένα συγκεκριμένο πρόβλημα, πράγμα το οποίο κάνει την επιλογή της σχέσης  $R$  σε εφαρμογές του πραγματικού κόσμου αρκετά δύσκολη εργασία.



### Πυρήνες για Βασικούς Όρους

Στη συνέχεια θα ορίσουμε έναν πυρήνα για σύνολα και τον πυρήνα γινομένου.

Οι παρακάτω πυρήνες ορίζονται στα σύνολα  $A, B \subseteq X$  των αντικειμένων  $x \in X$

*Πυρήνας τομής συνόλων ή Πυρήνας Συνόλου (set intersection kernel or set kernel)*

$$k_{\text{set}}(A, B) = |A \cap B| \quad (3.13)$$

όπου με  $|A \cap B|$  συμβολίζουμε τον πληθάρημο της τομής των δύο συνόλων.

*Πυρήνας Γινομένου (crossproduct kernel)*

$$k_x(A, B) = \sum_{x_A \in A} \sum_{x_B \in B} k_{\text{base}}(x_A, x_B) \quad (3.14)$$

όπου  $k_{\text{base}}(\cdot, \cdot)$  είναι κάθε έγκυρος πυρήνας που ορίζεται στο  $X$ .

#### 3.1.4 Στοιχεία Μεταβολής Πυρήνων

Προκειμένου να εισάγουμε γνώση για κάποιο πεδίο στον ορισμό ενός πυρήνα, συχνά είναι απαραίτητο να μεταβάλουμε τους πυρήνες κάποιου τύπου. Παρακάτω περιγράφουμε τυπικά αυτές τις αλλαγές με τη χρήση ενός πυρήνα

$$K_{\text{modifier}} : P \rightarrow (X \times X \rightarrow \mathbb{R}) \rightarrow (X \times X \rightarrow \mathbb{R}), \quad (3.15)$$

ο οποίος δεχόμενος ένα στοιχείο μεταβολής (modifier) και τις παραμέτρους του (ένα στοιχείο του χώρου παραμέτρων  $P$ ), αντιστοιχίζει έναν οποιοδήποτε πυρήνα στον μεταβληθέντα πυρήνα.

$$\kappa_{\text{polynomial}}(\mathbf{p}, l)(\mathbf{k})(x, x') = (\kappa(x, x') + 1)^p \quad (l \geq 0, p \in \mathbb{Z}^+) \quad (3.16)$$

$$\kappa_{\text{gaussian}}(\gamma)(\mathbf{k})(x, x') = e^{-\gamma[k(x, x) - 2k(x, x') + k(x', x')]} \quad (\gamma > 0) \quad (3.17)$$

$$\kappa_{\text{normalised}}(\mathbf{k})(x, x') = \frac{k(x, x')}{\sqrt{k(x, x)k(x', x')}} \quad (3.18)$$

## 3.2 Μηχανές Διανυσμάτων Υποστήριξης

Οι Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines) είναι μοντέλα τα οποία λύνουν προβλήματα ταξινόμησης προτύπων. Η μορφή τους δόθηκε στο Σχήμα 2.3 του Κεφαλαίου 2. Σε αντίθεση με τα δίκτυα Perceptron που υποφέρουν από το πρόβλημα της βραδείας εκπαίδευσης, τα μοντέλα αυτά έχουν μικρούς χρόνους εκπαίδευσης.

### 3.2.1 Γραμμικά Διαχωρίσιμα Προβλήματα

Ας υποθέσουμε ότι αντιμετωπίζουμε το πρόβλημα ταξινόμησης δύο κλάσεων  $C_0$  και  $C_1$  που είναι γραμμικά διαχωρίσιμες. Έτσι υπάρχουν ένα διάνυσμα  $w$  και ένα κατώφλι  $w_0$  τέτοια ώστε:

$$\mathbf{w}^T \mathbf{x} + w_0 = \begin{cases} < 0 & \text{αν } \mathbf{x} \in C_0 \\ > 0 & \text{αν } \mathbf{x} \in C_1 \end{cases} \quad (3.19)$$

Επειδή υπάρχουν πολλές λύσεις, δηλαδή άπειρα ζεύγη  $(\mathbf{w}, w_0)$  για το διαχωρισμό των κλάσεων, θεωρούμε κάποιο κριτήριο αξιολόγησης των λύσεων.

Το κριτήριο αυτό είναι το περιθώριο ταξινόμησης (margin)  $\gamma$  μεταξύ των δύο κλάσεων το οποίο ορίζεται ως το άθροισμα  $\gamma = \gamma_0 + \gamma_1$  μεταξύ των δύο παρακάτω περιθωρίων (το  $\gamma_0$  για την κλάση  $C_0$  και το  $\gamma_1$  για την κλάση  $C_1$ ), όπου

$$\gamma_0 = \min_{\mathbf{x} \in C_0} \frac{-(\mathbf{w}^T \mathbf{x} + w_0)}{\|\mathbf{w}\|} \quad (3.20)$$

$$\gamma_1 = \min_{\mathbf{x} \in C_1} \frac{(\mathbf{w}^T \mathbf{x} + w_0)}{\|\mathbf{w}\|} \quad (3.21)$$

Τα πρότυπα  $\mathbf{x}$  της κλάσης  $C_0$  καθώς και τα πρότυπα  $\mathbf{x}'$  της κλάσης  $C_1$ , για τα οποία επιτυγχάνεται η ελάχιστη απόσταση (η ελάχιστη απόλυτη τιμή των  $\gamma_0$  και  $\gamma_1$  στους παραπάνω τύπους), λέγονται διανύσματα υποστήριξης (support vectors).

Στη συνέχεια θεωρούμε το «κανονικό διαχωριστικό υπερεπίπεδο» για το οποίο

- α) το κατώφλι  $w_0$  τοποθετείται ακριβώς στη μέση ανάμεσα στις δύο κλάσεις ( $\gamma_0 = \gamma_1$ )
- β) η κλιμάκωση των  $\mathbf{w}$  και  $w_0$  είναι τέτοια, ώστε:

$$\mathbf{w}^T \mathbf{x} + w_0 = \begin{cases} \leq -1 & \text{αν } \mathbf{x} \in C_0 \\ \geq 1 & \text{αν } \mathbf{x} \in C_1 \end{cases} \quad (3.22)$$

και για το οποίο προκύπτει  $\gamma_0 = \gamma_1 = \frac{1}{\|\mathbf{w}\|}$  και άρα  $\gamma = \frac{2}{\|\mathbf{w}\|}$ .

Το πρόβλημα συνεπώς είναι ο προσδιορισμός του ελαχίστου της συνάρτησης

$$J(\mathbf{w}, \mathbf{w}_0) = \frac{1}{2} \|\mathbf{w}\|^2 \quad (3.23)$$

υπό τους περιορισμούς των P ανισοτήτων που δίνονται παρακάτω

$$d_i(\mathbf{w}^T \mathbf{x}_i + w_0) \geq 1 \quad i = 1, \dots, P \quad (3.24)$$

Για την ελαχιστοποίηση της παραπάνω συνάρτησης χρησιμοποιούνται πολλαπλασιαστές Lagrange δημιουργώντας τη συνάρτηση κόστους  $L^d(\lambda_1, \dots, \lambda_p) = -L(\lambda_1, \dots, \lambda_p)$ , η οποία πρέπει να ελαχιστοποιηθεί ως προς τα  $\lambda_i$ , όπου:

$$L(\lambda_1, \dots, \lambda_p) = \sum_{i=1}^P \lambda_i - \frac{1}{2} \sum_{i=1}^P \sum_{j=1}^P \lambda_i \lambda_j d_i d_j \mathbf{x}_i^T \mathbf{x}_j \quad (3.25)$$

Έτσι έχουμε να λύσουμε το δυϊκό του αρχικού προβλήματος, που είναι ο υπολογισμός του ελαχίστου της συνάρτησης  $L^d$  ως προς τα  $\lambda_1, \dots, \lambda_p$  υπό τους περιορισμούς:

$$\begin{aligned} \sum_{i=1}^P \lambda_i d_i &= 0 \\ \lambda_i &\geq 0, \quad i=1, \dots, P \end{aligned} \quad (3.26)$$

Αποδεικνύεται ότι το διάνυσμα λύσης  $\mathbf{w}$  είναι ένας θετικός γραμμικός συνδυασμός των διανυσμάτων υποστήριξης.

$$\mathbf{w} = \sum_{i \in I_{sv}} \lambda_{o,i} d_i \mathbf{x}_i \quad (3.27)$$

όπου  $I_{sv}$  είναι το σύνολο που αποτελείται από τα διανύσματα υποστήριξης, και  $\lambda_{o,i}$  είναι οι βέλτιστοι πολλαπλασιαστές Lagrange που βρίσκονται από τη λύση της (3.26).

Η βέλτιστη διαχωριστική επιφάνεια είναι η ακόλουθη :

$$\mathbf{g}^*(\mathbf{x}) = \sum_{i \in I_{SV}} \lambda_{o,i} d_i \mathbf{x}_i^T \mathbf{x} + w_0 \quad (3.28)$$

και το κατώφλι υπολογίζεται ως εξής:

$$w_0 = \frac{1}{|I_{SV}|} \sum_{i \in I_{SV}} \frac{1}{d_i} - \mathbf{w}^T \mathbf{x}_i \quad (3.29)$$

### 3.2.2 Μη Γραμμικά Διαχωρίσιμα Προβλήματα

Όταν έχουμε μη γραμμικά προβλήματα, δηλαδή προβλήματα στα οποία οι κλάσεις δεν είναι γραμμικά διαχωρίσιμες, το πρόβλημα, είναι ο προσδιορισμός του ελαχίστου της συνάρτησης

$$J_{ns}(\mathbf{w}, w_0) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^P \xi_i \quad (3.30)$$

υπό τους περιορισμούς των παρακάτω P ανισοτήτων

$$\begin{aligned} d_i(\mathbf{w}^T \mathbf{x}_i + w_0) &\geq 1 - \xi_i \\ \xi_i &\geq 0 \end{aligned} \quad (3.31)$$

όπου  $i = 1, \dots, P$ , η μεταβλητή  $\xi_i$  λέγεται μεταβλητή χαλαρότητας και η παράμετρος  $C$  επιλέγεται από το χρήστη και είναι το βάρος του κόστους των λάθος ταξινομήσεων.

Το παραπάνω πρόβλημα μετασχηματίζεται στο ακόλουθο δυϊκό πρόβλημα υπολογισμού του ελαχίστου της συνάρτησης  $L^d$  ως προς τα  $\lambda_1, \dots, \lambda_P$ , όπως στην (3.25) υπό τους περιορισμούς

$$\sum_{i=1}^P \lambda_i d_i = 0 \text{ και}$$

$$0 \leq \lambda_i \leq C, \quad i=1, \dots, P \quad (3.32)$$

Το διάνυσμα λύσης βρίσκεται ότι είναι το ίδιο όπως στην (3.27).

Σε αυτή την περίπτωση, επειδή χρησιμοποιούνται γραμμικές συναρτήσεις για διαχωρισμό μη γραμμικών κλάσεων, είναι πιθανόν πολλά πρότυπα να ταξινομηθούν λανθασμένα.

### 3.2.3 Χρήση Συναρτήσεων Πυρήνα

Στην ενότητα αυτή θα ασχοληθούμε με την εύρεση του βέλτιστου διαχωριστικού υπερεπιπέδου, στην περίπτωση που τα πρότυπα είναι διαχωρίσιμα από μη γραμμικές επιφάνειες, αποφεύγοντας τις λάθος ταξινομήσεις. Για το σκοπό αυτό:

- Αρχικά χρησιμοποιείται κάποιος κατάλληλος μη γραμμικός μετασχηματισμός  $\Phi(\cdot)$ , αντιστοιχίζοντας το διάνυσμα εισόδου σε ένα χώρο χαρακτηριστικών μεγαλύτερης διάστασης στον οποίο τα πρότυπα γίνονται γραμμικά διαχωρίσιμα.
- Στη συνέχεια κατασκευάζεται στο χώρο αυτό το βέλτιστο διαχωριστικό υπερεπίπεδο, χρησιμοποιώντας τα μετασχηματισμένα πρότυπα.

Η ιδέα για το πρώτο βήμα έρχεται από το θεώρημα του Cover το οποίο είναι το ακόλουθο:

#### **Θεώρημα Cover**

Ένα σύνθετο πρόβλημα ταξινόμησης προτύπων το οποίο μετασχηματίζεται μη γραμμικά σε ένα χώρο μεγάλης διάστασης είναι πιο πιθανό να είναι στο χώρο αυτό γραμμικά διαχωρίσιμο από ότι σε ένα χώρο μικρής διάστασης.

Με άλλα λόγια, το θεώρημα του Cover λέει ότι ο αρχικός χώρος μπορεί να μετασχηματιστεί σε ένα νέο χώρο χαρακτηριστικών, όπου τα πρότυπα είναι γραμμικά διαχωρίσιμα με μεγάλη πιθανότητα δεδομένου ότι ικανοποιούνται δύο συνθήκες:

- ο μετασχηματισμός είναι μη γραμμικός
- η διάσταση του χώρου χαρακτηριστικών είναι αρκετά μεγάλη.

Το δεύτερο βήμα είναι σε αντιστοιχία με τη θεωρία που αναφέραμε για γραμμικά δεδομένα, με τη διαφορά ότι το διαχωριστικό υπερεπίπεδο τώρα ορίζεται σαν γραμμική συνάρτηση των διανυσμάτων που παίρνουμε από το χώρο χαρακτηριστικών και όχι από τον αρχικό χώρο εισόδου.

Ας συμβολίσουμε με  $\mathbf{x}$  ένα διάνυσμα από το χώρο εισόδου, που έστω ότι είναι διάστασης  $m_0$  και με  $\{\varphi_j(\mathbf{x})\}_{j=1}^{m_1}$  το σύνολο των μη γραμμικών μετασχηματισμών από τον αρχικό χώρο στο χώρο χαρακτηριστικών διάστασης  $m_1$ . Ορίζουμε το ακόλουθο υπερεπίπεδο:

$$\sum_{j=1}^{m_1} w_j \varphi_j(\mathbf{x}) = 0 \quad (\text{έχουμε υποθέσει } \varphi_0(\mathbf{x}) = 1 \text{ για κάθε } \mathbf{x}) \quad (3.33)$$

Η παραπάνω εξίσωση γράφεται σε συμπαγή μορφή

$$\mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}) = 0 \quad (3.34)$$

Προσαρμόζοντας τον τύπο της περίπτωσης των γραμμικά διαχωρίσιμων προτύπων παίρνουμε:

$$\mathbf{w} = \sum_{i=1}^P \lambda_i d_i \boldsymbol{\varphi}(\mathbf{x}_i) \quad (3.35)$$

όπου το διάνυσμα  $\boldsymbol{\varphi}(\mathbf{x}_i)$  αντιστοιχεί στο πρότυπο εισόδου  $\mathbf{x}_i$ .

Η διαχωριστική επιφάνεια στο χώρο χαρακτηριστικών είναι η ακόλουθη :

$$\sum_{i=1}^P \lambda_i d_i \boldsymbol{\varphi}^T(\mathbf{x}_i) \boldsymbol{\varphi}(\mathbf{x}) = 0 \quad (3.36)$$

Ο όρος  $\boldsymbol{\varphi}^T(\mathbf{x}_i) \boldsymbol{\varphi}(\mathbf{x})$  είναι το εσωτερικό γινόμενο δύο διανυσμάτων στο χώρο χαρακτηριστικών που δημιουργήθηκε από τον μετασχηματισμό του διανύσματος εισόδου  $\mathbf{x}$  και του προτύπου εισόδου  $\mathbf{x}_i$ .

Ορίζουμε τον πυρήνα εσωτερικού γινομένου (inner-product kernel) ως εξής:

$$k(\mathbf{x}, \mathbf{x}_i) = \boldsymbol{\varphi}^T(\mathbf{x}_i) \boldsymbol{\varphi}(\mathbf{x}) = \sum_{j=0}^{m_i} \varphi_j(\mathbf{x}) \varphi_j(\mathbf{x}_i) \quad \text{για } i = 1, 2, \dots, P \quad (3.37)$$

Από τον ορισμό αυτό βλέπουμε ότι ο πυρήνας εσωτερικού γινομένου είναι μια συμμετρική συνάρτηση ως προς τα ορίσματά της, δηλαδή:

$$k(\mathbf{x}, \mathbf{x}_i) = k(\mathbf{x}_i, \mathbf{x}) \quad \text{για κάθε } i \quad (3.38)$$

Με βάση τα παραπάνω, το βέλτιστο διαχωριστικό υπερεπίπεδο δίνεται από την ακόλουθη εξίσωση:

$$\sum_{i=1}^P \lambda_i d_i k(\mathbf{x}_i, \mathbf{x}) = 0 \quad (3.39)$$

Η εξίσωση (3.37) επιτρέπει την κατασκευή ενός διαχωριστικού υπερεπίπεδου που είναι μη γραμμικό στο χώρο εισόδου, αλλά η εικόνα του στο χώρο χαρακτηριστικών είναι γραμμική. Το δυικό πρόβλημα για την εύρεση της βέλτιστης διαχωριστικής επιφάνειας από μια μηχανή διανυσμάτων υποστήριξης ορίζεται ως εξής :



Δεδομένων των προτύπων  $\{(\mathbf{x}_i, d_i)\}$  για  $i=1, \dots, P$ , υπολογίζονται οι πολλαπλασιαστές Lagrange  $\lambda_i$  για  $i=1, \dots, P$ , που μεγιστοποιούν την αντικειμενική συνάρτηση

$$Q(\lambda) = \sum_{i=1}^P \lambda_i - \frac{1}{2} \sum_{i=1}^P \sum_{j=1}^P \lambda_i \lambda_j d_i d_j k(\mathbf{x}_i, \mathbf{x}_j) \quad (3.40)$$

υπό τους περιορισμούς

- 1)  $\sum_{i=1}^P \lambda_i d_i = 0$
- 2)  $0 \leq \lambda_i \leq C$  για  $i=1, 2, \dots, P$

Θα πρέπει να σημειωθεί ότι μπορούμε να δούμε το  $k(\mathbf{x}_i, \mathbf{x}_j)$  σαν το  $ij$ -στοιχείο ενός συμμετρικού  $N \times N$  πίνακα  $\mathbf{k}$  ως εξής:

$$\mathbf{k} = \left\{ k(\mathbf{x}_i, \mathbf{x}_j) \right\}_{(i,j)=1}^P \quad (3.41)$$

Έχοντας βρει τους βέλτιστους πολλαπλασιαστές Lagrange  $\lambda_{o,i}$  η βέλτιστη λύση για το διάνυσμα βαρών  $\mathbf{w}$  είναι

$$\mathbf{w} = \sum_{i=1}^P \lambda_{o,i} d_i \varphi(\mathbf{x}_i) \quad (3.42)$$

Μπορεί να αναφερθεί ότι η διάσταση του χώρου χαρακτηριστικών καθορίζεται από το πλήθος των διανυσμάτων υποστήριξης τα οποία λαμβάνονται από τα πρότυπα εκπαίδευσης με τη λύση του ανωτέρω προβλήματος βελτιστοποίησης.

### 3.3 Μεταφορά Συμβολικής Γνώσης σε Πυρήνες Μηχανών Διανυσμάτων Υποστήριξης

Η ιδέα πίσω από τη μεταφορά συμβολικής γνώσης σε πυρήνες ([38]) είναι ότι σε σημασιολογικό επίπεδο, παρόμοια άτομα θα συμπεριφέρονται παρόμοια ως προς τις ίδιες σημασιολογικές έννοιες.

Έστω ότι με  $K = \langle T, A \rangle$  συμβολίζουμε τη βάση γνώσης μας, όπου  $T$  είναι το TBox και  $A$  το ABox αυτής και έστω ότι  $\text{Ind}(A)$  είναι το σύνολο των ατόμων στο  $A$ . Δεδομένου του ακόλουθου συνόλου περιγραφών εννοιών  $F = \{ F_1, F_2, \dots, F_m \}$  στο  $T$ , μια οικογένεια συναρτήσεων πυρήνα  $k_p^F : \text{Ind}(A) \times \text{Ind}(A) \mapsto [0,1]$  ορίζεται ως ο  $L_p$  μέσος απλούστερων συναρτήσεων, ως ακολούθως:

$$\forall a, b \in \text{Ind}(A) \quad k_p^F(a,b) := \left[ \sum_{i=1}^m \left| \frac{\kappa_i(a,b)}{m} \right|^p \right]^{1/p} \quad (3.43)$$

όπου  $p > 0$  και,  $\forall i \in \{1, \dots, m\}$ , η αντίστοιχη (απλούστερη) συνάρτηση πυρήνα  $\kappa_i$  ορίζεται ως εξής,  $\forall a, b \in \text{Ind}(A)$ :

$$\kappa_i(a,b) = \begin{cases} 1 & (F_i(a) \in A \wedge F_i(b) \in A) \vee (\neg F_i(a) \in A \wedge \neg F_i(b) \in A) \\ 0 & (F_i(a) \in A \wedge \neg F_i(b) \in A) \vee (\neg F_i(a) \in A \wedge F_i(b) \in A) \\ 1/2 & \text{Αλλιώς} \end{cases} \quad (3.44)$$

ή μοντελο-θεωρητικά:

$$\kappa_i(a,b) = \begin{cases} 1 & (K \models F_i(a) \wedge K \models F_i(b)) \vee (K \models \neg F_i(a) \wedge K \models \neg F_i(b)) \\ 0 & (K \models F_i(a) \wedge K \models \neg F_i(b)) \vee (K \models \neg F_i(a) \wedge K \models F_i(b)) \\ \frac{1}{2} & \text{Αλλιώς} \end{cases} \quad (3.45)$$

Από τον παραπάνω τύπο βλέπουμε ότι, επιλέγοντας τις βέλτιστες περιγραφές εννοιών, δηλαδή αυτές που συμβάλλουν στη διαφοροποίηση των ατόμων της Βάσης Γνώσης, για κάθε συνδυασμό δυο ατόμων από το ABox η συνάρτηση ομοιότητας  $\kappa_i$  για την  $F_i$  περιγραφή έννοιας επιστρέφει 1 (μονάδα): αν και τα δύο άτομα είναι στιγμιότυπα της έννοιας  $F_i$ , ή της άρνησης της έννοιας  $F_i$ , είτε αυτό προκύπτει ρητά από το ABox, είτε προκύπτει χρησιμοποιώντας τα αξιώματα του TBox μέσα από τη διαδικασία συλλογιστικής. Η συνάρτηση ομοιότητας επιστρέφει 0 (μηδέν): αν το ένα άτομο, π.χ., το  $a$  είναι στιγμιότυπο της έννοιας  $F_i$ , ενώ το  $b$  δεν είναι (είναι στιγμιότυπο της έννοιας  $\neg F_i$ ), ή και το αντίστροφο.

Το σκεπτικό για τα ανωτέρω είναι ότι η ομοιότητα μεταξύ των ατόμων καθορίζεται από την ομοιότητά τους με βάση κάθε έννοια από ένα δεδομένο αριθμό χαρακτηριστικών. Δύο άτομα είναι όμοια - σε μέγιστο βαθμό - με βάση μια έννοια  $F_i$ , αν έχουν την ίδια συμπεριφορά, με άλλα λόγια είτε και τα δύο είναι στιγμιότυπα της έννοιας  $F_i$ , ή και τα δύο είναι στιγμιότυπα της άρνησης της έννοιας αυτής. Αντίστοιχα, δύο άτομα είναι όμοια - σε ελάχιστο βαθμό - όταν το ένα είναι στιγμιότυπο της έννοιας  $F_i$  και το άλλο της άρνησης αυτής της έννοιας και αντίστροφα.

Τέλος εξαιτίας της υπόθεσης ανοικτού κόσμου, όταν η μηχανή συλλογιστικής δεν μπορεί να συμπεράνει για το αν ένα άτομο είναι στιγμιότυπο μιας έννοιας - επομένως και οι δύο δυνατότητες είναι ανοιχτές - η συνάρτηση ομοιότητας  $\kappa_i$  παίρνει την τιμή  $\frac{1}{2}$ , μια ενδιάμεση τιμή, δηλαδή, που αντανακλά αυτήν την αβεβαιότητα.

Όπως αναφέρθηκε, ο έλεγχος στιγμιότυπου (instance checking), ο οποίος αφορά την εύρεση αν ένα άτομο είναι στιγμιότυπο μιας έννοιας, χρησιμοποιείται για να βρούμε την τιμή των απλών συναρτήσεων ομοιότητας  $k_i$ . Ο έλεγχος αυτός, όμως, είναι υπολογιστικά πολύπλοκος (εξαρτάται και από την συγκεκριμένη περιγραφική γλώσσα που χρησιμοποιείται). Εναλλακτικά, ειδικά για οντολογίες οι οποίες είναι πλούσιες σε ισχυρισμούς που δηλώνονται ρητά στο ABox, ένα απλό κοίταγμα μπορεί να είναι αρκετό, όπως προτείνεται από τον πρώτο ορισμό (3.44) των  $k_i$  συναρτήσεων.

Αξίζει να σημειωθεί ότι η παραπάνω οικογένεια πυρήνων είναι παραμετροποιημένη ως προς την επιλογή των χαρακτηριστικών. Συγκεκριμένα, οι πυρήνες αυτοί μπορούν να παραμετροποιηθούν σε κάθε σύνολο σύνθετων περιγραφών εννοιών, συμπεριλαμβανομένων των εννοιών με άρνηση. Επίσης λαμβάνουν υπόψη τους την περίπτωση που δεν είναι σίγουρο αν ένα άτομο είναι στιγμιότυπο μιας έννοιας ή όχι, όπως φαίνεται στις (3.44) - (3.45). Η επιλογή των εννοιών που θα συμπεριληφθούν στην ομάδα χαρακτηριστικών  $F$  είναι σημαντική και μπορεί να είναι το αντικείμενο ενός αρχικού προβλήματος μάθησης (επιλογή χαρακτηριστικών).

Θα πρέπει να σημειωθεί ότι οι παραπάνω πυρήνες δεν εξαρτώνται από τη γλώσσα αναπαράστασης που χρησιμοποιείται, αφού είναι ο έλεγχος στιγμιότυπου που αποφαίνεται για το αν ένα άτομο είναι στιγμιότυπο κάποιας έννοιας.

Η πιο σημαντική ιδιότητα μιας συνάρτησης πυρήνα είναι - όπως έχει προαναφερθεί - η εγκυρότητά της (αν είναι δηλαδή θετικά ορισμένη).

### **Πρόταση**

Δεδομένου ενός ακεραίου  $p > 0$  και μιας ομάδας χαρακτηριστικών  $F$ , η συνάρτηση  $k_p^F$  είναι ένας έγκυρος (θετικά ορισμένος) πυρήνας.

### Απόδειξη

Ένας τρόπος απόδειξης είναι να δείχτεί ότι η συνάρτηση  $k_p^F$  είναι θετικά ορισμένη. Ένας πιο εύκολος τρόπος για να δείχτεί αυτή η ιδιότητα, είναι να δείχτεί ότι η συνάρτηση μπορεί να σχηματιστεί ως σύνθεση απλούστερων έγκυρων πυρήνων, χρησιμοποιώντας τελεστές που εγγυώνται το κλείσιμο (closure) όσον αφορά σε αυτήν την ιδιότητα. Συγκεκριμένα, δεδομένου ότι οι απλές συναρτήσεις πυρήνα  $k_i$  ( $i = 1, \dots, n$ ) αντιστοιχούν σε πυρήνες ταιριάσματος (matching kernels), η συνάρτηση  $k_p^F$  είναι θετικά ορισμένη αφού προκύπτει από αυτές χρησιμοποιώντας τους τελεστές του αθροίσματος, του πολλαπλασιασμού με μια σταθερά και του πολλαπλασιασμού μεταξύ τους έγκυρων πυρήνων.

Οι μηχανές SVM που στηρίζονται σε συναρτήσεις πυρήνα μπορούν, με αποτελεσματικό τρόπο, να αντιστοιχίσουν τα στιγμιότυπα σε ένα χώρο χαρακτηριστικών, στον οποίο αυτά μπορούν να διαχωριστούν με τη βοήθεια ενός γραμμικού ταξινομητή ([52], [53]). Με τη χρήση της συνάρτησης πυρήνα, που ορίστηκε ανωτέρω, με εύκολο τρόπο μπορεί να σχεδιαστεί μια μηχανή SVM η οποία θα λύνει το πρόβλημα που διατυπώνεται τυπικά ως εξής:

Έστω ότι  $K = \langle T, A \rangle$  είναι η βάση γνώσης,  $Ind(A)$  το σύνολο όλων των ατόμων που εμφανίζονται στο  $A$  και  $C = \{C_1, \dots, C_s\}$  είναι το σύνολο (ατομικών και σύνθετων) εννοιών στο  $K$ .

Το πρόβλημα ταξινόμησης που θα επιλυθεί είναι:

$$\begin{aligned} & \text{Δεδομένου ενός ατόμου } a \in Ind(A), \\ & \text{προσδιόρισε τα } \{C_1, \dots, C_t\} \subseteq C, \text{ ούτως ώστε } K \models C_i(a) \forall i \in \{1, \dots, t\} \end{aligned} \quad (3.46)$$

## 4. Αναπαράσταση και Ανάλυση Συναισθήματος

---

Στο παρόν Κεφάλαιο περιγράφουμε σύντομα τις βασικές θεωρίες που θα χρησιμοποιήσουμε στη συνέχεια για την αναπαράσταση συναισθήματος, όπως και τα χαρακτηριστικά προσώπου στα οποία θα βασιστούμε για την ανάλυση των εκφράσεων του προσώπου για διερεύνηση της συμπεριφοράς και της συναισθηματικής κατάστασης των χρηστών κατά την αλληλεπίδρασή τους με υπολογιστικά συστήματα ([62], [69], [72]).

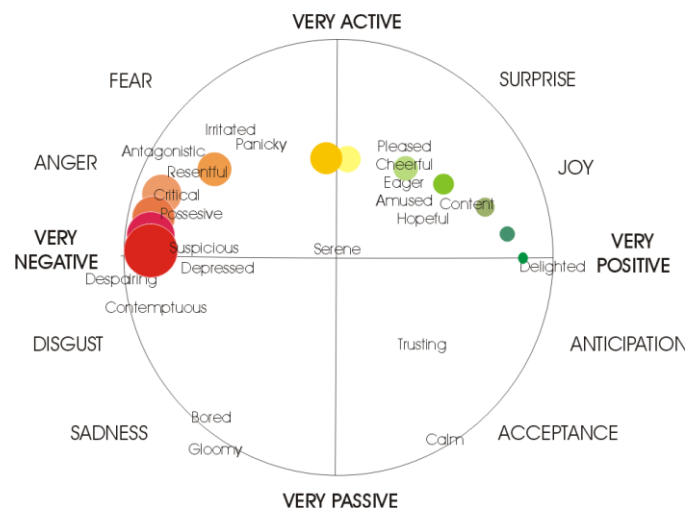
### 4.1 Αναπαράσταση Συναισθήματος

Η ψυχολογία έχει εξετάσει αναλυτικά την αναπαράσταση των συναισθημάτων, είτε σε κατηγορική μορφή, όπου επικρατούν τα 6 βασικά συναισθήματα, είτε σε συνεχή ανάλυση και αξιολόγηση των συναισθημάτων. Στο πλαίσιο αυτό, πολύ λίγες μελέτες και λίγα αποτελέσματα μπορούν να αξιοποιηθούν στον τομέα της αλληλεπίδρασης ανθρώπου μηχανής. Μία από αυτές τις μελέτες, που πραγματοποιήθηκε από τη Whissel ([1]), υποδηλώνει ότι τα συναισθήματα είναι σημεία σε ένα χώρο που μπορεί να περιγραφεί από ένα σχετικά μικρό αριθμό διαστάσεων.

Μια ιδιαίτερα ενδιαφέρουσα αναπαράσταση γίνεται σε ένα διδιάστατο χώρο που ορίζεται με βάση την ενεργοποίηση (activation) και την αξιολόγηση (evaluation). Ο χώρος ενεργοποίησης - αξιολόγησης δίνει μια αναπαράσταση που είναι απλή και ικανή να συλλάβει ένα ευρύ φάσμα σημαντικών αποχρώσεων του συναισθήματος. Αυτή η αναπαράσταση είναι πολύ χρήσιμη όταν θέλουμε να δείξουμε τη φύση του συναισθήματος, χωρίς να το προσδιορίζουμε επακριβώς.

Στηρίζεται στις επόμενες δυο παραμέτρους, όπως φαίνεται στο Σχήμα 4.1:

- Επίπεδο αξιολόγησης (Valence): ένα βασικό κοινό στοιχείο των συναισθηματικών καταστάσεων είναι ότι το άτομο επηρεάζεται από τα συναισθήματα «σθένους», δηλαδή το άτομο εκφράζει συνήθως θετικές ή αρνητικές εκτιμήσεις για τα συμβαίνοντα γύρω του. Η σχέση μεταξύ του συναισθήματος και της αξιολόγησης – που κυμαίνεται μεταξύ των τιμών positive/negative είναι ευρέως αποδεκτή (οριζόντιος άξονας).
- Επίπεδο ενεργοποίησης: η έρευνα έχει δείξει ότι οι συναισθηματικές καταστάσεις περιλαμβάνουν προδιαθέσεις του ατόμου να δρα με συγκεκριμένους τρόπους. Ένας βασικός τρόπος που αντανακλά αυτό το θέμα είναι το αντίστοιχο επίπεδο ενεργοποίησης, δηλαδή η δύναμη της διάθεσης του ατόμου να λάβει κάποια μέτρα ή να μην λάβει κανένα – ήτοι να έχει (σε κάποιο βαθμό) ενεργητική ή παθητική συμμετοχή (κάθετος άξονας).



Σχήμα 4.1 Αναπαράσταση Συναισθήματος

Ένα ιδιαίτερα σημαντικό ποσοστό της έκφρασης των συναισθημάτων μπορεί να συλληφθεί από την αναπαράσταση της ενεργοποίησης - αξιολόγησης. Τα

αντιλαμβανόμενα συναισθήματα δεν είναι ομοιόμορφα κατανεμημένα στο διάστημα επίπεδο ενεργοποίησης - αξιολόγησης, αλλά έχουν την τάση να σχηματίζουν ένα σχεδόν κυκλικό μοτίβο. Στο πλαίσιο αυτό, συνήθως γίνεται προσδιορισμός του κέντρου του κύκλου ως κατάσταση ηρεμίας (neutral). Ο βαθμός της συναισθηματικής έκφρασης μπορεί να μετρηθεί ως η απόσταση από το κέντρο του κύκλου έως ένα συγκεκριμένο σημείο στο χώρο ενεργοποίησης - αξιολόγησης.

Ένα ενδιαφέρον συμπέρασμα είναι ότι οι τα πιο «δυνατά» συναισθήματα είναι πιο έντονα διακριτά μεταξύ τους, από ό,τι είναι τα λιγότερο δυνατά συναισθήματα με τον ίδιο συναισθηματικό προσανατολισμό. Μια σχετική επέκταση οδηγεί στη θεώρηση των πρωτογενών ή βασικών συναισθημάτων ως κύριων σημείων στην περιφέρεια του κύκλου του συναισθήματος.

## *4.2 Εξαγωγή Χαρακτηριστικών από Εικόνες Προσώπου*

Οι περισσότερες από τις τεχνικές αυτόματης αναγνώρισης συναισθημάτων βασίζονται στο γνωστό Σύστημα Κωδικοποίησης Αντιδράσεων Προσώπου (FACS). Το FACS είναι ένα ανατομικά προσανατολισμένο σύστημα κωδικοποίησης εκφράσεων προσώπου, βασισμένο στον ορισμό των «μονάδων δράσης» ενός προσώπου που προκαλούν κινήσεις του προσώπου ([56], [60], [61]). Το μοντέλο FACS έχει εμπνεύσει τα μοντέλα σύνθεσης εκφράσεων προσώπου και τον καθορισμό των παραμέτρων στο πλαίσιο του προτύπου ISO MPEG-4 ([68], [71]).

Μία από τις σημαντικότερες εξελίξεις για την ενοποίηση των αναπαραστάσεων τόσο για σκοπούς ανάλυσης όσο και σύνθεσης υπήρξε το πρότυπο ISO MPEG-4, και ιδιαίτερα το τμήμα της υβριδικής κωδικοποίησης (synthetic natural hybrid coding, SNHC). Η υποομάδα SNHC του MPEG-4 κυκλοφόρησε τρεις εκδόσεις:

- MPEG-4 Έκδοση 1η : κίνηση του προσώπου . Ο ορισμός και η κίνηση συνθετικών χαρακτήρων (avatars) βασίζεται στο μοντέλο FBA (κινήσεων



προσώπου και σώματος) και έχει υιοθετηθεί από το πρότυπο MPEG-4 από το 1998. Το μοντέλο προσώπου περιγράφεται από την παραμόρφωση των χαρακτηριστικών του προσώπου και ελέγχεται από τυποποιημένες μονάδες χαρακτηριστικών. Ειδικότερα, τα Facial Definition Points (FDPs) χρησιμοποιούνται για τον ορισμό του μοντέλου του προσώπου και τα Facial Animation Parameters (FAPS) ελέγχουν την κίνησή του.

- MPEG-4 Έκδοση 2η : κίνηση του σώματος. Αντίστοιχα με τα FDPs και τα FAPS, τα BDPs περιγράφουν τις εγγενείς ιδιότητες του κάθε ανατομικού τμήματος του σώματος του avatar και τα BAPs ελέγχουν την κίνησή του.
- MPEG-4 Μέρος 16: κίνηση γενικών εικονικών αντικειμένων. Το MPEG-4 ενέκρινε στην τελευταία τροπολογία του ένα γενικευμένο μοντέλο παραμόρφωσης, το οποίο ονομάστηκε BBA (Bone-based Animation). Όπως ορίζεται στο μέρος 16 του προτύπου MPEG-4, αυτό το γενικευμένο μοντέλο οστών και μυών οδήγησε σε μια αναπαράσταση Σκελετού, Μυών και Δέρματος (SMS).

Στο πλαίσιο του προτύπου MPEG-4, οι παράμετροι που έχουν καθοριστεί για την κίνηση προσώπου και σώματος (FBA) καθορίζουν συγκεκριμένους κόμβους του προσώπου και του σώματος στο γράφημα μιας σκηνης. Ο στόχος του ορισμού FBA είναι η κίνηση των χαρακτήρων να μπορεί να χρησιμοποιείται τόσο σε κανονικές ταινίες όσο και σε cartoons. Έτσι, το MPEG-4 έχει ορίσει ένα μεγάλο σύνολο παραμέτρων και ο χρήστης μπορεί να επιλέξει υποσύνολα αυτών των παραμέτρων ανάλογα με την εφαρμογή.

Ειδικότερα, το πρότυπο MPEG-4 ορίζει 84 σημεία χαρακτηριστικών FDPs στο ουδέτερο πρόσωπο, τα οποία παρέχουν την χωρική αναφορά που είναι απαραίτητη για τον ορισμό των FAPS. Ειδικότερα, το σετ των σημείων του προσώπου (FDP) και το σετ παραμέτρων κίνησης του προσώπου (FAP) σχεδιάστηκαν στο πλαίσιο MPEG-4 για να επιτρέψουν τον καθορισμό της τοπολογίας της υποβόσκουσας γεωμετρίας

(μέσω των FDPs) και την κίνηση των χαρακτηριστικών του προσώπου που αναπαριστούν εκφράσεις, συναισθήματα και προφορά ομιλίας (μέσω των FAPs).

Παρά το γεγονός ότι τα FAPs παρέχουν όλα τα απαραίτητα στοιχεία για να είναι η κίνηση συμβατή με το MPEG-4, η χρησιμοποίησή τους για ανάλυση εκφράσεων από σκηνές βίντεο είναι δύσκολη, λόγω έλλειψης σαφούς ποσοτικού πλαισίου. Για να μετρήσουμε τα FAPs σε πραγματικές ακολουθίες εικόνων, πρέπει να ορίσουμε μια αντιστοίχιση μεταξύ αυτών και της κίνησης συγκεκριμένων σημείων FDPs, δηλαδή, χαρακτηριστικών σημείων του προσώπου (FPs), τα οποία αντιστοιχούν σε σημαντική πληροφορία για τις κινήσεις και παραμορφώσεις του ανθρώπινου προσώπου.

Πίνακας 1: Υπολογισμός των Χαρακτηριστικών FAPs

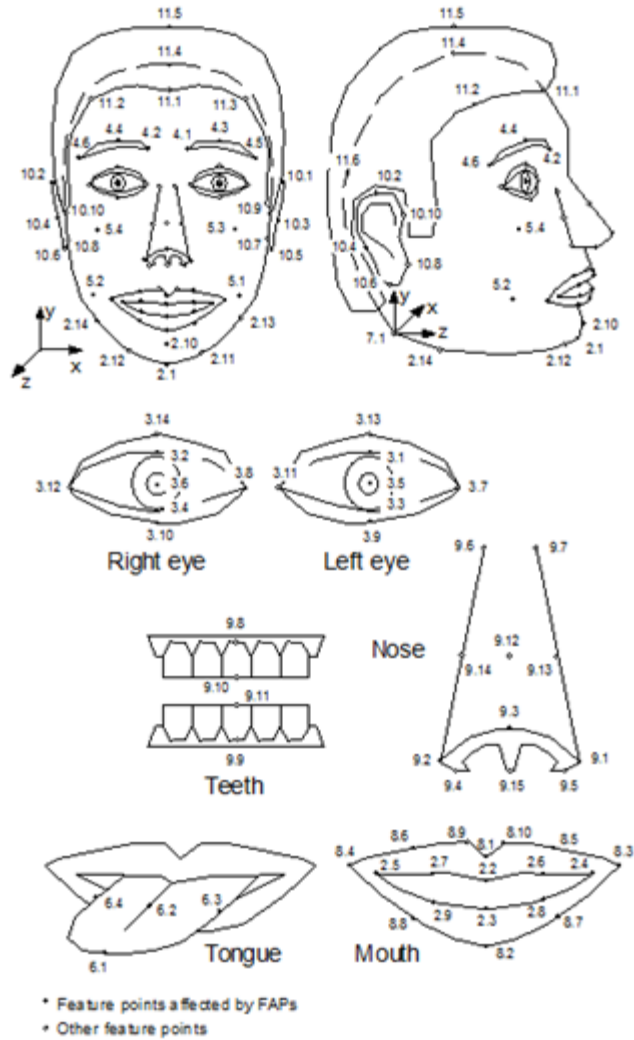
<b>FAP name</b>	<b>Feature for the description</b>	<b>Utilized feature</b>
<i>Squeeze_l_eyebrow (F<sub>37</sub>)</i>	$D_1=s(4.5,3.11)$	$f_1= D_{1-NEUTRAL} -D_1$
<i>Squeeze_r_eyebrow (F<sub>38</sub>)</i>	$D_2=s(4.6,3.8)$	$f_2= D_{2-NEUTRAL} -D_2$
<i>Lower_t_midlip (F<sub>4</sub>)</i>	$D_3=s(9.3,8.1)$	$f_3= D_3 -D_{3-NEUTRAL}$
<i>Raise_b_midlip (F<sub>5</sub>)</i>	$D_4=s(9.3,8.2)$	$f_4= D_{4-NEUTRAL} -D_4$
<i>Raise_l_l_eyebrow (F<sub>31</sub>)</i>	$D_5=s(4.1,3.11)$	$f_5= D_5 -D_{5-NEUTRAL}$
<i>Raise_r_l_eyebrow (F<sub>32</sub>)</i>	$D_6=s(4.2,3.8)$	$f_6= D_6 -D_{6-NEUTRAL}$
<i>Raise_l_o_eyebrow (F<sub>33</sub>)</i>	$D_7=s(4.5,3.7)$	$f_7= D_7 -D_{7-NEUTRAL}$
<i>Raise_r_o_eyebrow (F<sub>36</sub>)</i>	$D_8=s(4.6,3.12)$	$f_8= D_8 -D_{8-NEUTRAL}$
<i>Raise_l_m_eyebrow (F<sub>33</sub>)</i>	$D_9=s(4.3,3.7)$	$f_9= D_9 -D_{9-NEUTRAL}$
<i>Raise_r_m_eyebrow (F<sub>34</sub>)</i>	$D_{10}=s(4.4,3.12)$	$f_{10}= D_{10} -D_{10-NEUTRAL}$
<i>Open_jaw (F<sub>3</sub>)</i>	$D_{11}=s(8.1,8.2)$	$f_{11}= D_{11} -D_{11-NEUTRAL}$
<i>close_t_l_eyelid (F<sub>19</sub>) - close_b_l_eyelid (F<sub>21</sub>)</i>	$D_{12}=s(3.1,3.3)$	$f_{12}= D_{12} -D_{12-NEUTRAL}$
<i>close_t_r_eyelid (F<sub>20</sub>) - close_b_r_eyelid (F<sub>22</sub>)</i>	$D_{13}=s(3.2,3.4)$	$f_{13}= D_{13} -D_{13-NEUTRAL}$

$stretch\_l\_cornerlip (F_6)$ $(stretch\_l\_cornerlip\_o)(F_{33}) -$ $stretch\_r\_cornerlip (F_7)$ $(stretch\_r\_cornerlip\_o)(F_{34})$	$D_{14}=s(8.4, 8.3)$	$f_{14}= D_{14} - D_{14-NEUTRAL}$
$squeeze\_l\_eyebrow (F_{37})$ AND $squeeze\_r\_eyebrow (F_{38})$	$D_{15}=s(4.6, 4.5)$	$f_{15}= D_{15-NEUTRAL} - D_{15}$

Ο Πίνακας 1 παρέχει την ποσοτική μοντελοποίηση των FAPs που έχουμε εφαρμόσει χρησιμοποιώντας 15 χαρακτηριστικά σημεία που αναφαίρονται ως  $f_i$  ( $i = 1..15$ ).

Αυτό το σύνολο χαρακτηριστικών περιλαμβάνει σημεία χαρακτηριστικών που βρίσκονται κύρια στις περιοχές του στόματος και των ματιών και, σε περιβάλλοντα αλληλεπίδρασης ανθρώπου - μηχανής, μπορούν αυτόματα να ανιχνευθούν και να παρακολουθούνται. Αποτελείται από αποστάσεις, που σημειώνονται ως  $s(x, y)$ , μεταξύ των σημείων  $x$  και  $y$ , δηλαδή των χαρακτηριστικών σημείων που φαίνονται στο Σχήμα 4.2. Ορισμένα από αυτά τα σημεία είναι σταθερά κατά τη διάρκεια των εκφράσεων και μπορούν να χρησιμοποιηθούν ως σημεία αναφοράς.

Στον Πίνακα 1 τα  $s(x,y)$  συμβολίζουν τις Ευκλείδειες Αποστάσεις μεταξύ των σημείων FPs, ενώ τα  $D_{i-NEUTRAL}$  αναφέρονται στις αποστάσεις  $D_i$  όταν το πρόσωπο είναι στην ουδέτερη κατάσταση.



Σχήμα 4.2 Τα χαρακτηριστικά σημεία FDP.

Ιδιαίτερο ενδιαφέρον έχει η αυτόματη εξαγωγή των ανωτέρω χαρακτηριστικών σημείων και ο υπολογισμός των χαρακτηριστικών FAPs. Στη συνέχεια αναφέρουμε μια συνήθη διαδικασία που ακολουθείται και είχαμε διαθέσιμη στο Εργαστήριο Εικόνων του Ε.Μ. Πολυτεχνείου ([57]).

- Το πρώτο στάδιο που εκτελείται είναι η ανίχνευση/αναγνώριση του προσώπου του ατόμου - χρήστη ([51], [59], [65]). Μετά από αυτό, τα κύρια χαρακτηριστικά

του προσώπου, όπως τα μάτια, το στόμα και η μύτη, αντιμετωπίζονται ως κύριες ασυνέχειες του προσώπου του ([54], [58]).

- Στο πρώτο βήμα/στάδιο της μεθόδου ανίχνευσής του, το σύστημα εκτελεί μια βελτιστοποιημένη διαδικασία τμηματοποίησης ([64]). Οι αρχικές εκτιμήσεις των τμημάτων, που ονομάζονται σπόροι, προσεγγίζονται μέσω ανάλυσης μεγίστου-ελαχίστου και προσδιορίζονται με τη μεγιστοποίηση μιας υπό όρους συνάρτησης πιθανότητας. Στη συνέχεια ακολουθεί μια διαδικασία ενίσχυσης των σημαντικών σπόρων με ανάλυση των χρωματικών συνιστωσών της εικόνας του προσώπου.
- Οι ενισχυμένοι σπόροι δημιουργούν ένα σύνολο αντικειμένων, το οποίο περιλαμβάνει τις τοπικές παραμορφώσεις του προσώπου μέσω της χρήσης ενεργών περιγραμμάτων ([63]) που εφαρμόζονται σε όλα τα αντικείμενα του συνόλου.
- Το σύνολο αυτό, τελικά, περιορίζεται σε ένα μικρότερο σύνολο, όπου τα χαρακτηριστικά σημεία ανιχνεύονται σύμφωνα με ένα κριτήριο ελαχιστοποίησης σφάλματος .
- Σε μια απλοποιημένη εκδοχή αυτής της προσέγγισης, χρησιμοποιούνται μορφολογικές λειτουργίες (διαβρώσεις και διαστολές) που λαμβάνουν υπόψη τυχόν συμμετρίες, για να καθορίσουν πρώτα τα πιο πιθανά blobs της περιοχής του προσώπου που συμπεριλαμβάνουν τα μάτια και το στόμα.
- Η αναζήτηση μέσω φίλτρων κλίσης πάνω από τα μάτια και ανάμεσα στα μάτια και το στόμα, παρέχει εκτιμήσεις των θέσεων των φρυδιών και της μύτης.
- Με βάση τις ανιχνευόμενες θέσεις υπολογίζονται και αξιολογούνται τα χαρακτηριστικά σημεία FPs.

### 4.3 Περιγραφή Κανόνων για Ανάλυση Συναισθήματος

Το πρότυπο MPEG-4 δείχνει έναν εναλλακτικό τρόπο μοντελοποίησης των εκφράσεων του προσώπου και των εκφραζόμενων συναισθημάτων. Τα χαρακτηριστικά FAPs, που χρησιμοποιούνται στο πλαίσιο του MPEG-4 για μοντελοποίηση των εκφράσεων του προσώπου ([2]), είναι στενά συνδεδεμένα με τις Μονάδες Δράσης (AUs), που αποτελούν το βασικό συστατικό του Συστήματος Κωδικοποίησης Δράσης Προσώπου (FACS) ([3], [67]).

Με βάση την αναπαράσταση των συναισθημάτων στα τεταρτημόρια του τροχού της Whissel (Σχήμα 4.1), έχουν δημιουργηθεί κανόνες για κάθε ένα από τα τρία πρώτα τεταρτημόρια (στο τέταρτο τεταρτημόριο δε βρίσκεται κανένα ενδιαφέρον - που να συναντάται σε συνήθεις εφαρμογές αλληλεπίδρασης ανθρώπου μηχανής - συναίσθημα).

Για να επιτευχθεί ο στόχος αυτός, μεταφράστηκαν οι κινήσεις των μυών του προσώπου σε FAPs ([4]). Αξίζει να αναφέρουμε αυτό επιτεύχθηκε αντικαθιστώντας το εύρος της διακύμανσης των τιμών των FAPs με τους όρους υψηλή, μεσαία, χαμηλή, αφότου πρώτα κανονικοποιήθηκαν τα αντίστοιχα διαστήματα ([5]).

Ένα υποσύνολο των κανόνων που χρησιμοποιήσαμε ως υπάρχουσα Γνώση στην παρούσα εργασία απεικονίζεται στον Πίνακα 2.

Το πλήρες σύνολο των κανόνων περιγράφεται στην εργασία ([5]).

Πίνακας 2: Υπάρχουσα Γνώση (Κανόνες) για Ανάλυση Συναισθήματος

Κανόνας	FAPs	Τεταρτημόριο
2	F3_M+F4_L+F5_L+[F53+F54]_H +[F19+F21]_H+[F20+F22]_H	(+,+)
7	F3_L+F4_L+F5_H+[F53+F54]_H+ [F19+F21]_H+[F20+F22]_H+[F37 +F38]_M+F59_H+F60_H	(+,+)
13	F3_L+F4_M+F5_H+F31_L+F32_L +F33_L+F34_L+F37_H+F38_H+ F59_M+F60_M	(-,+)
16	F3_L+F4_M+F5_L+F31_L+F32_L +F33_L+F34_L+F37_H+F38_H+ F59_M+F60_M	(-,+)
26	F3_M+F5_L+[F19+F21]_H+[F20+ F22]_H+F31_H+F32_H+F33_M+ F34_M+F35_M+F36_M+[F37+F3 8]_H	(-,-)
34	F3_L+F4_L+[F53+F54]_L+F31_M +F32_M+F33_M+F34_M+F35_M +F36_M+F37_M+F38_M	(-,-)
41	F3_L+F4_M+F31_M+F32_M+F33 _M+F34_M+F35_M+F36_M+F37 _M+F38_M+ F59_M+F60_M	Neutral

# 5. Αρχιτεκτονική Συστήματος

## Ανάλυσης Συναισθήματος

---

Σύμφωνα με όσα αναπτύχθηκαν στα προηγούμενα Κεφάλαια, στην παρούσα διπλωματική έχουμε τους ακόλουθους στόχους:

- Να διερευνηθεί η δυνατότητα των νευρωνικών δικτύων βαθιάς μάθησης να αναλύουν τη συμπεριφορά των χρηστών σύμφωνα με την περιγραφείσα μοντελοποίηση συναισθήματος, δηλαδή να διερευνηθεί αν είναι σε θέση να κατατάσσουν με ακρίβεια τις εκφράσεις του προσώπου του χρήστη κατά την αλληλεπίδρασή του με τον υπολογιστή στον διδιάστατο χώρο (κύκλο) ενεργοποίησης - αξιολόγησης.

Είναι, ωστόσο, γνωστό ότι η ανάλυση των συναισθημάτων που βασίζονται σε εκφράσεις του προσώπου μπορεί να είναι πολύ διαφορετική για κάθε άνθρωπο, υπό την έννοια ότι διαφορετικοί άνθρωποι εκφράζουν τα συναισθήματά τους με διαφορετικούς τρόπους όσον αφορά στις μεταβολές των παραμέτρων του προσώπου τους. Ως εκ τούτου είναι απαραίτητη μια μικρή ή μεγάλη προσαρμογή ενός δικτύου, που είναι εκπαιδευμένο για να προβλέπει γενικά τα ανθρώπινα συναισθήματα από τις εκφράσεις του προσώπου, όταν το εφαρμόζουμε σε ένα συγκεκριμένο άτομο.

- Να διερευνηθεί η δυνατότητα μιας τέτοιας προσαρμογής νευρωνικών ταξινομητών που αναλύουν την συναισθηματική κατάσταση των χρηστών τους, μεταφέρεται η υπάρχουσα γνώση για τη μοντελοποίηση του συναισθήματος, με βάση τις παραμέτρους ενεργοποίησης του προσώπου, σε μια μηχανή διανυσμάτων υποστήριξης (SVM). Στη συνέχεια χρησιμοποιούμε το SVM αυτό για ανάλυση της συναισθηματικής κατάστασης ενός συγκεκριμένου χρήστη, αφού πρώτα επαν-εκπαιδύσουμε τη μηχανή SVM με λίγα δεδομένα που θα συλλεγούν από τις εκφράσεις του προσώπου του συγκεκριμένου χρήστη.



Τα αποτελέσματα που λαμβάνονται με τον τρόπο αυτό εξαρτώνται σε μεγάλο βαθμό από την ακρίβεια της εξαγωγής των παραμέτρων κίνησης του προσώπου, η οποία είναι ένα σημαντικό θέμα της ανάλυσης εικόνων του προσώπου. Θα πρέπει, ωστόσο, να αναφερθεί ότι οι παράμετροι κίνησης του προσώπου είναι πολύ λιγότερο εξαρτημένες από κάθε χρήστη, από ότι οι εκφράσεις και τα συναισθήματα του χρήστη.

- Να διερευνηθεί η ακρίβεια της εξαγωγής των εν λόγω παραμέτρων κίνησης του προσώπου από τις εικόνες του προσώπου, από νευρωνικά δίκτυα βαθιάς μάθησης.
- Να συνδυάσει την εξαγωγή των παραμέτρων FAPs με την υπάρχουσα γνώση - δηλαδή το σετ κανόνων για ανάλυση συναισθήματος με βάση τα FAPs που περιγράφηκε στο Κεφάλαιο 4 – για εύρωστη και προσαρμοζόμενη στον χρήστη ανάλυση συναισθήματος.

Αυτό γίνεται τροφοδοτώντας την έξοδο του νευρωνικού δικτύου βαθιάς μάθησης στην είσοδο της μηχανής διανύσματος υποστήριξης στην οποία έχουν ενσωματωθεί οι ανωτέρω κανόνες για ανάλυση συναισθήματος. Η προσαρμογή αναφέρεται μόνο στο τελευταίο συστατικό και μπορεί να επιτευχθεί εύκολα, υποθέτοντας ότι μια μικρή φάση επανεκπαίδευσης μπορεί να υλοποιηθεί, όταν ένας συγκεκριμένος άνθρωπος χρησιμοποιήσει το προτεινόμενο σύστημα.

Ένα συνελκτικό δίκτυο ήταν η βάση για την ανάλυση των εισερχόμενων εικόνων των χρηστών στην παρούσα διπλωματική. Η αρχιτεκτονική του δικτύου που χρησιμοποιείται εδώ παρουσιάζεται στο σχήμα 5.1. Τα καρέ μιας ακολουθίας βίντεο (σε μορφή PAL) αρχικά επεξεργάστηκαν μέσω ενός αλγόριθμου ανίχνευσης προσώπου και οι εικόνες προσώπων που προέκυψαν περικόπηκαν και παρουσιάστηκαν ως είσοδοι (με μέγεθος 250x250 εικονοστοιχεία) στο δίκτυο, μαζί με την αντίστοιχη κατηγορία-ετικέτα του είδους του συναισθήματος.

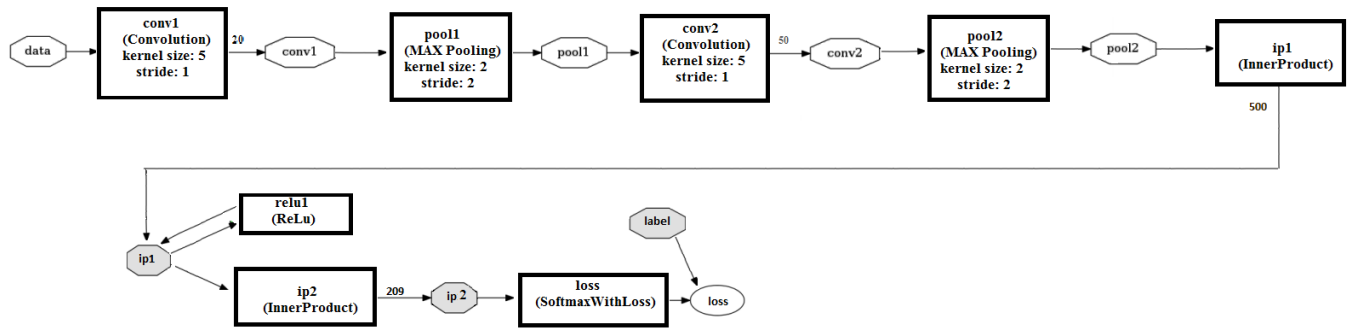
Η αρχιτεκτονική του δικτύου έχει 4 στάδια που περιέχουν διαφορετικά επίπεδα. Τα δύο πρώτα στάδια περιλαμβάνουν ένα συνελκτικό επίπεδο που ακολουθείται από ένα επίπεδο pooling. Τα δύο τελευταία στάδια είναι πλήρως συνδεδεμένα επίπεδα εμπρόσθιας τροφοδότησης, με το τελευταίο στάδιο να αποτελείται από softmax μονάδες, οι οποίες εξάγουν την πιθανότητα επιτυχίας για κάθε έξοδο του δικτύου. Η συνάρτηση ενεργοποίησης που χρησιμοποιείται στο προηγούμενο εσωτερικό επίπεδο του δικτύου (πλήρως συνδεδεμένο επίπεδο) είναι η διορθωμένη συνάρτηση γραμμικής μονάδας (Relu).

Αρχικά εξετάζεται η ικανότητα της αρχιτεκτονικής βαθιάς μάθησης (που φαίνεται στο Σχήμα 5.2) να μάθει να εκτελεί με επιτυχία την ανάλυση της έκφρασης του προσώπου βασισμένη στη μοντελοποίηση της ενεργοποίησης - αξιολόγησης συναισθήματος.

Αυτό το σύστημα, το οποίο συμπληρώνει ένα άλλο σύστημα που παρουσιάζεται στο ([33]) για κατηγοριοποίηση συναισθήματος (στα βασικά έξι συναισθήματα) εξαρτάται, ωστόσο, σε μεγάλο βαθμό από τον τρόπο που τα άτομα συμπεριφέρονται και εκφράζουν τα συναισθήματα τους. Κατά συνέπεια, χρειάζεται επανεκπαίδευση του συνόλου των βαρών της βαθιάς αρχιτεκτονικής, με την απαίτηση συλλογής νέων μεγάλων ποσοτήτων δεδομένων για κάθε νέο χρήστη, οι οποίες να αντιπροσωπεύουν τους τρόπους έκφρασης των συναισθημάτων του.

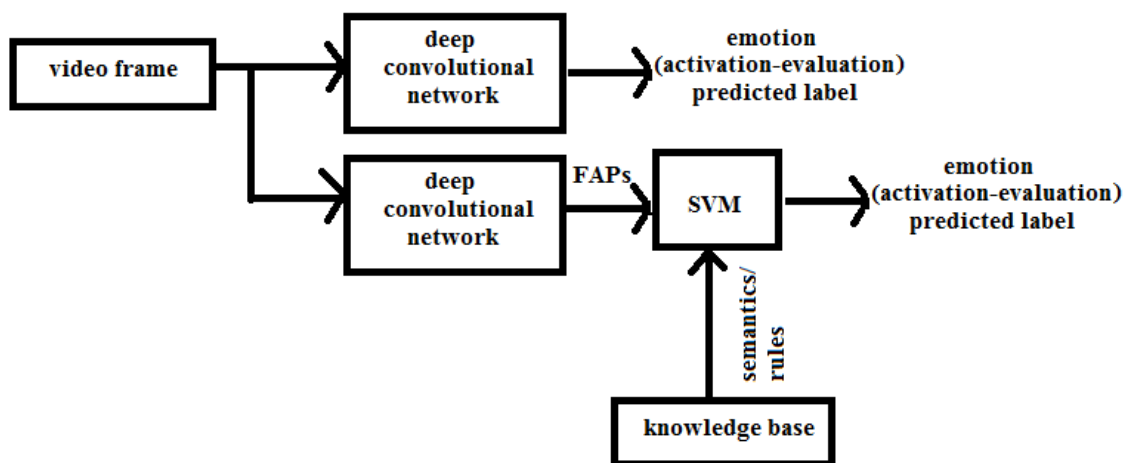
Η πρωτοτυπία της προσέγγισής μας έγκειται στο ότι αντιμετωπίζει το πρόβλημα της ανάλυσης συναισθήματος με το να το χωρίζει σε δύο υποπροβλήματα, κάθε ένα από τα οποία εστιάζει σε διαφορετικό στόχο.

Το πρώτο χρησιμοποιεί μια αρχιτεκτονική δικτύου βαθιάς μάθησης με στόχο την εξαγωγή των παραμέτρων κίνησης του προσώπου (FAPs), οι οποίες, όπως περιγράφηκε στο Κεφάλαιο 4, μπορούν να χρησιμοποιηθούν για ανάλυση της συναισθηματικής κατάστασης του ατόμου στην αλληλεπίδραση ανθρώπου - μηχανής. Σε αυτό το πλαίσιο, εξετάζεται η ικανότητα μιας τέτοιας αρχιτεκτονικής να μπορεί να εξάγει αποτελεσματικά τα FAPs από τα καρέ του βίντεο που καταγράφει την εικόνα του προσώπου του χρήστη.



Σχήμα 5.1 Η αρχιτεκτονική που βασίζεται σε ένα βαθύ συνελκτικό δίκτυο.

Το δεύτερο στάδιο βασίζεται σε ένα SVM που έχει σχεδιαστεί με βάση τη μεταφορά της γνώσης σε πυρήνα και την μοντελοποίηση ενεργοποίησης - αξιολόγησης των εκφράσεων του προσώπου. Μεταφέρουμε τις υπάρχουσες γνώσεις - σε μορφή κανόνων -- για την εκτίμηση των συναισθημάτων (η εκτίμηση βασίζεται στις παραμέτρους ενεργοποίησης του προσώπου, FAPs) , στην αρχιτεκτονική SVM και χρησιμοποιούμε το SVM για ανάλυση του συναισθήματος. Θα πρέπει να προστεθεί ότι είναι μόνο το μέρος του SVM που θα πρέπει να προσαρμοστεί όταν επικεντρωνόμαστε σε ένα συγκεκριμένο άτομο και στον τρόπο έκφρασης των συναισθημάτων του. Το Σχήμα 5.2 παρουσιάζει τις δύο προσεγγίσεις που χρησιμοποιούμε για την ανάλυση του συναισθήματος.



Σχήμα 5.2 Η αρχιτεκτονική των συστημάτων ανάλυσης συναισθήματος, με ή χωρίς ενδιάμεση εκτίμηση των παραμέτρων FAPs.

# 6. Πειραματικά Αποτελέσματα

---

## 6.1 Περιγραφή της Βάσης Δεδομένων

Τα δεδομένα που χρησιμοποιήθηκαν για τη διερεύνηση της απόδοσης των μεθόδων που παρουσιάζονται στην παρούσα διπλωματική εργασία, ελήφθησαν από την naturalistic βάση δεδομένων που έχει δημιουργηθεί από το Πανεπιστήμιο Queens του Μπέλφαστ (QUB) στο πρόγραμμα EC FP5 IST EPMHS και επεκτάθηκε περαιτέρω στο EC FP6 IST Humaine Δίκτυο Αριστείας (Network of Excellence) και στο πρόγραμμα EC FP7 IST Semaine.

Για τους σκοπούς της εργασίας, όπως αναφέρθηκαν προηγούμενα στο Κεφάλαιο 5, επικεντρωθήκαμε στην ανάλυση του συνόλου δεδομένων ενός χρήστη, έτσι ώστε να απεικονίσουμε τα αποτελέσματα της προτεινόμενης ανάλυσης της έκφρασης και του συστήματος αναγνώρισης συναισθημάτων σε ένα συγκεκριμένο χρήστη.

Το Σχήμα 6.1 δείχνει 4 χαρακτηριστικές εικόνες κάθε μια από τις οποίες αντιστοιχεί σε έκφραση συναισθήματος που ανήκει σε διαφορετικό τεταρτημόριο του κύκλου ενεργοποίησης - αξιολόγησης.

Παρά το γεγονός ότι η βάση δεδομένων είναι οπτικοακουστική, στην παρούσα εργασία χρησιμοποιήσαμε ένα μεγάλο αριθμό από οπτικά καρέ, προσεκτικά επιλεγμένα μετά την τεκμηρίωση της βάσης δεδομένων.



( $\alpha$ )



( $\beta$ )



(γ)



(δ)



(ε)

Σχήμα 6.1 Χαρακτηριστικές Εικόνες Έκφρασης Συναισθήματος: α) (+,+) β) (+,-)  
γ) (-,+) δ) (-,-) ε) ουδέτερη

## 6.2 Τεκμηρίωση των Δεδομένων

Το σύνολο δεδομένων που χρησιμοποιήθηκε σε αυτή τη διπλωματική, δημιουργήθηκε με τη συμμετοχή ενός εμπειρογνώμονα στο πεδίο (Dr Roddie Cowie, Professor QUB), έτσι ώστε οι εκφράσεις του προσώπου σε αυτές τις ακολουθίες βίντεο να μην είναι προσποιητές, ούτε ακραίες, αλλά φυσιολογικές.

Η εργασία βασίστηκε στη συνεργασία δύο εμπειρογνομώνων, οι οποίοι επέλεξαν ένα σημαντικό αριθμό από καρέ του δημιουργηθέντος βίντεο και τον τεκμηρίωσαν, αξιολογώντας τόσο τις παραμέτρους κίνησης (FAPs) του προσώπου, όσο και τη συναισθηματική κατάσταση του χρήστη σε δύο ξεχωριστές φάσεις.

Κατά τη διάρκεια της πρώτης φάσης, οι εμπειρογνώμονες τεκμηρίωσαν τις επιλεγείσες εικόνες του βίντεο ως προς τις τιμές των παραμέτρων FAPs, λαμβάνοντας υπόψιν τη σχέση τους με τις κινήσεις των χαρακτηριστικών του προσώπου. Κατά τη δεύτερη φάση, αξιολόγησαν τη συναισθηματική κατάσταση (θετική ή αρνητική, ενεργητική ή παθητική) του χρήστη σε όλες τις εικόνες και έτσι δημιουργήθηκε η τεκμηρίωση των εικόνων του προσώπου σε σχέση με την διδιάστατη αναπαράσταση συναισθήματος.

Τα επιλεγμένα καρτέ ήταν τα πιο εκφραστικά ανάμεσα στα καρτέ εκφράσεων του προσώπου (με διακριτές κινήσεις του προσώπου και ένα υψηλό επίπεδο εκτίμησης), τα οποία στη συνέχεια επεκτάθηκαν σε ένα ευρύτερο σύνολο δεδομένων μέσω μιας προσέγγισης ομοιότητας των FAPs που βασίστηκε στα αποτελέσματα του υφιστάμενου στο Εργαστήριο Εικόνων του Ε.Μ.Πολυτεχνείου συστήματος εξαγωγής χαρακτηριστικών του προσώπου ([6], [55], [66], [70]).

### *6.3 Αποτελέσματα της Ανάλυσης Συναισθήματος*

Στην πρώτη φάση των πειραμάτων μελετήθηκε η απόδοση που έχει το βαθύ συνελκτικό δίκτυο των Κεφαλαίων 2 και 5 για την ανάλυση συναισθήματος στο σύνολο των οπτικών δεδομένων, που αποτελείται από 11.152 καρτέ της ακολουθίας των εικόνων. Περίπου το 40% από αυτά αντιπροσωπεύει συναισθήματα του (+, +) τεταρτημορίου του Σχήματος 4.1, το 30% συναισθήματα του (-, +) τεταρτημορίου, το 15% συναισθήματα του (+, -), ένα 10% συναισθήματα του (-, -) και ένα 5 % την ουδέτερη έξοδο.

Στην αρχή εφαρμόσαμε μια τεχνική ανίχνευσης του προσώπου σε κάθε επιλεγμένο καρτέ του βίντεο ([51], μέσω της υλοποίησής της που υπάρχει στην πλατφόρμα OpenCV) και στη συνέχεια οι εικόνες που προέκυψαν περικόπηκαν (πάλι με τη βοήθεια της πλατφόρμας OpenCV) σε διάσταση 256 x 256 εικονοστοιχείων έκαστη.



Για κάθε μια από τις εικόνες αυτές έγινε - από τους δυο εμπειρογνώμονες -  
α) κατηγοριοποίησή τους στο αντίστοιχο τεταρτημόριο του κύκλου αναπαράστασης με  
βάση την ενεργοποίηση - αξιολόγηση της έκφρασης, β) καθορισμός των τιμών κάθε  
FAP του προσώπου (Low, Medium, ή High κίνηση).

Στο Σχήμα 6.2 φαίνεται ένα σετ από περικεκομμένες εικόνες που δείχνουν τρεις  
χαρακτηριστικές εκφράσεις (που αντιστοιχούν σε διαφορετικά σετ από FAPs) για κάθε  
μια από τις 5 κατηγορίες συναισθηματικής κατάστασης (θετικής/αρνητικής,  
ενεργητικής/παθητικής, ή ουδέτερης).





(a)





(β)





(γ)







(8)





(ε)

Σχήμα 6.2 3 διαφορετικές εκφράσεις προσώπου (που αντιστοιχούν σε διαφορετικά σετ από FAPs) για κάθε μια από τις 5 κατηγορίες συναισθηματικής κατάστασης: α) (+,+) β) (+,-) γ) (-,+) δ) (-,-) ε) ουδέτερη.

Οι περικεκομμένες εικόνες δόθηκαν ως είσοδοι στο βαθύ δίκτυο, το οποίο αποτελείται από 4 στάδια, όπως περιγράφηκε στο κεφάλαιο 5. Η πλατφόρμα Caffe χρησιμοποιήθηκε για την υλοποίηση του δικτύου αυτού.

Τα δύο πρώτα στάδια περιελάμβαναν ένα συνελικτικό επίπεδο που ακολουθείτο από ένα pooling επίπεδο, που χρησιμοποιούσε την τεχνική max-pooling. Στο πρώτο συνελικτικό επίπεδο χρησιμοποιήθηκαν 20 φίλτρα και στο δεύτερο 50, όλα μεγέθους 5x5 εικονοστοιχείων. Κάθε συνελικτικό επίπεδο είχε τις ίδιες παραμέτρους μάθησης: την τιμή 0,001 ως βασική παράμετρο μάθησης για τα φίλτρα και την τιμή 0.002 για τα bias,

την τιμή 0,9 ορμή τόσο για τα δύο φίλτρα όσο και για τα bias, όπως και παράμετρο για την αποσύνθεση βαρών (weight decay) με τιμή 0,004 ανά εποχή. Στα pooling επίπεδα το μέγεθος του παραθύρου ήταν 2x2.

Τα πλήρως διασυνδεδεμένα επίπεδα είχαν τις ίδιες υπερ-παραμέτρους εκτός από την αυτήν της αποσύνθεσης των βαρών, η οποία είχε οριστεί σε 1. Τα δύο τελευταία επίπεδα περιείχαν 100 γραμμικές μονάδες (Relu) το πρώτο και 5 εξόδους (σε αντιστοιχία με τις κατηγορίες του μοντέλου των συναισθημάτων) με μονάδες Softmax το τελευταίο.

Μια μέση ακρίβεια 83% επιτεύχθηκε με τη χρήση υποσυνόλων μάθησης και επικύρωσης (με διάσπαση 80-20%) των διαθέσιμων δεδομένων.

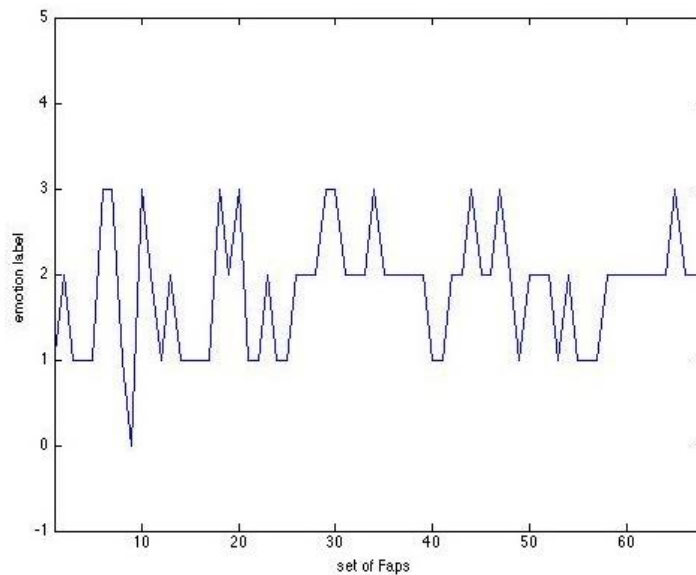
Στη συνέχεια εξετάστηκε η ικανότητα ενός παρόμοιου δικτύου - η βασική διαφορά ήταν η αύξηση των μονάδων Relu του πρώτου πλήρως διασυνδεδεμένου επιπέδου σε 500 και του αριθμού των εξόδων σε 210, όσοι ήταν οι πιο συχνοί συνδυασμοί των τιμών των FAPs που εμφανίζονται στο σύνολο δεδομένων.

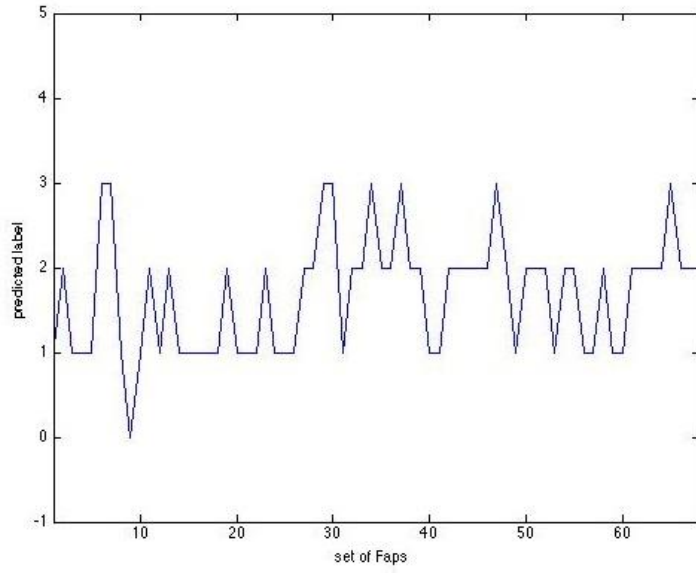
Το δίκτυο ήταν σε θέση να μάθει να εκτιμάει σωστά τα σύνολα των FAPs σχεδόν στο 95% του συνόλου των περιπτώσεων, πράγμα το οποίο δείχνει την ακρίβεια της εκτίμησης των FAPs που επιτυγχάνεται.

Στη συνέχεια, εξετάσαμε πρώτα την ικανότητα του γενικού συνόλου κανόνων για ανάλυση συναισθήματος βασισμένου στα FAPs, από το ([5]), για να παράγει μια πρόβλεψη συναισθήματος (σύμφωνα με το μοντέλο ενεργοποίησης - αξιολόγησης) για καθεμιά από τις επιλεγμένες 210 περιπτώσεις FAPs. Μόνο 67 από αυτές (δηλαδή περίπου το 32%) ικανοποίησε έναν από τους κανόνες που περιλαμβάνονται στη βάση γνώσης μας, πράγμα το οποίο δείχνει τη μεγάλη μεταβλητότητα της σχέσης μεταξύ της αναπαράστασης συναισθήματος και του συνόλου των FAPs όταν εφαρμόζονται σε συγκεκριμένα άτομα.

Στη συνέχεια χρησιμοποιήσαμε τη διαδικασία που παρουσιάστηκε στο κεφάλαιο 5 και μεταφέραμε το ανωτέρω σύνολο κανόνων από το ([5]) στην αρχιτεκτονική SVM χρησιμοποιώντας τις εξισώσεις (3.43)-(3.45) και εκπαιδεύσαμε το SVM με μόνο τις παραπάνω 67 περιπτώσεις που ταίριαζαν σωστά με το σύνολο των κανόνων. Κατά τη δοκιμή του SVM με το υπόλοιπο των συνόλων FAP επιτύχαμε ακρίβεια σχεδόν 60%, πράγμα το οποίο δείχνει την ικανότητα του SVM να γενικεύει καλά τη διαθέσιμη γνώση.

Επιπλέον, επιλέξαμε 50 ακόμη σετ χαρακτηριστικών FAPs και χρησιμοποιήσαμε το προκύπτον σύνολο από FAP σετ για να εκπαιδεύσουμε το SVM. Η απόδοση του SVM αυξήθηκε στο 85%, πράγμα που δείχνει την αποτελεσματική προσαρμογή του σημασιολογικά εκπαιδευμένου SVM σε συγκεκριμένα χαρακτηριστικά και συμπεριφορές των χρηστών. Ένα μέρος της εξόδου του SVM (σωστές και εκτιμώμενες επισημάνσεις) φαίνεται στο Σχήμα 6.3.





Σχήμα 6.3: Επιθυμητές και προβλεπόμενες έξοδοι του συστήματος ανάλυσης συναισθήματος

# 7. Συμπεράσματα και Μελλοντικές Επεκτάσεις

---

Στην παρούσα εργασία μελετήθηκε μια σύγχρονη ερευνητική περιοχή της Τεχνητής Νοημοσύνης, η ανάπτυξη συστημάτων που εμπεριέχουν τόσο βαθιές αρχιτεκτονικές νευρωνικών δικτύων, οι οποίες δίδουν τη δυνατότητα ακριβούς εκτίμησης των παραμέτρων ενός προβλήματος, όσο και μηχανές διανυσμάτων υποστήριξης που αποτελούν αποτελεσματικούς κατηγοριοποιητές των δεδομένων οι οποίοι μπορούν να ενσωματώσουν την υπάρχουσα γνώση για το πρόβλημα και να προσαρμοστούν στα συγκεκριμένα χαρακτηριστικά του προβλήματος.

Οι ανωτέρω μέθοδοι εφαρμόστηκαν με επιτυχία για την ανάλυση συναισθήματος σε περιβάλλοντα αλληλεπίδρασης ανθρώπου μηχανής. Η περιγραφόμενη αρχιτεκτονική, όμως, αποτελεί μια γενική δομή για ανάπτυξη συστημάτων που έχουν τη δυνατότητα να ενσωματώνουν διαθέσιμες οντολογίες και βάσεις γνώσης, και να προσαρμόζουν τη γνώση τους σε προβλήματα του πραγματικού κόσμου όπως είναι μεταβλητές συνθήκες χρήσης και περιβάλλοντος, διαφορετικοί χρήστες και εννοιολογικό πλαίσιο αλληλεπίδρασης.

Η διπλωματική αυτή εργασία αφορά σε ένα ερευνητικό πεδίο ιδιαίτερης σημασίας για την τεχνητή νοημοσύνη και γι' αυτό η βασική της αρχιτεκτονική, όπως και τα αποτελέσματά της υποβλήθηκαν ήδη σε διεθνές συνέδριο με θέμα την ανάπτυξη ευφών συστημάτων ([73]).

Μια σειρά από ανοικτά θέματα αποτελούν μελλοντικά πεδία έρευνας, όπως είναι τα εξής:

- Επεκτάσεις των τρόπων διασύνδεσης βαθιών συνελκτικών αρχιτεκτονικών νευρωνικών δικτύων με μεθόδους πυρήνα.

- Επεκτάσεις με διαφορετικά μοντέλα βαθιών αρχιτεκτονικών (μη επιβλεπόμενα, υβριδικά).
- Η επέκταση των τρόπων μεταφοράς της αναπαράστασης γνώσης στις αρχιτεκτονικές βαθιάς μηχανικής μάθησης.
- Η εφαρμογή των μεθόδων σε μεγάλες βάσεις δεδομένων ανάλυσης συναισθήματος.
- Η εφαρμογή της προτεινόμενης αρχιτεκτονικής σε πραγματικά προβλήματα ανάλυσης συμπεριφοράς σε περιβάλλοντα αλληλεπίδρασης ανθρώπου μηχανής.

Ιδιαίτερο ενδιαφέρον παρουσιάζουν οι πιθανές επεκτάσεις στις αρχιτεκτονικές με βαθιά νευρωνικά δίκτυα. Στο πεδίο αυτό η έρευνα περιλαμβάνει την ανάπτυξη αρχιτεκτονικών με διάδοση της πληροφορίας προς τα εμπρός αλλά και ανατροφοδότηση (integrating feed-forward and feedback transfer of information). Η βαθιά μηχανή Boltzmann το κάνει αυτό, αλλά υπάρχουν θέματα κλιμάκωσης. Σημαντικό ενδιαφέρον παρουσιάζει η ενσωμάτωση επιβλεπόμενης και μη επιβλεπόμενης μάθησης σε ένα ενιαίο αλγόριθμο, όπως και η ενσωμάτωση βαθιάς μάθησης και δομημένης πρόβλεψης («συλλογιστικής»), όπως "Αναδρομικά" δίκτυα που λειτουργούν πάνω σε αναπαραστάσεις γνώσης στο χώρο διανυσμάτων, αλλά και καλύτερη θεωρητική κατανόηση της βαθιάς μάθησης και των συνελκτικών δικτύων, π.χ. μέσω του «μετασχηματισμού διασκορπισμού» του Stephane Mallat, ή την ανάπτυξη αραιών αναπαραστάσεων.

Η παρούσα διπλωματική εργασία θέτει τα θεμέλια για την επέκταση της περιγραφόμενης μεθοδολογίας και των τεχνικών που αναπτύχθηκαν στις ανωτέρω περιοχές έρευνας, οι οποίες αναμένεται να παρουσιάσουν πληθώρα νέων αποτελεσμάτων τα επόμενα χρόνια.



## 8. Βιβλιογραφία

---

- [1] C. M. Whissel, The dictionary of affect in language, R. Plutchnik and H. Kellerman (Eds) Emotion: Theory, research and experience: vol 4, The measurement of emotions. Academic Press, New York, 1989.
- [2] M. Tekalp, "Face and 2-D Mesh Animation in MPEG-4", Tutorial Issue On The MPEG-4 Standard, Image Communication Journal, Elsevier, 1999.
- [3] P. Ekman and W. Friesen, The Facial Action Coding System, Consulting Psychologists Press, San Francisco, CA, 1978.
- [4] Raouzaïou, A., Tsapatsoulis, N., Karpouzis, K., & Kollias, S. (2002). Parameterized facial expression synthesis based on MPEG-4. EURASIP Journal on Applied Signal Processing, 2002(10), 1021–1038. Hindawi Publishing Corporation.
- [5] S. Ioannou, A. Raouzaïou, V. Tzouvaras, T. Mailis, K. Karpouzis, S. Kollias, "Emotion recognition through facial expression analysis based on a neurofuzzy network", Special Issue on Emotion: Understanding & Recognition, Neural Networks, Elsevier, Volume 18, Issue 4, May 2005, Pages 423-435.
- [6] S. Asteriadis, P. Tzouveli, K. Karpouzis, S. Kollias, "Non-verbal feedback on user interest based on gaze direction and head pose", 2nd International Workshop on Semantic Media Adaptation and Personalization (SMAP 2007), London, United Kingdom, 17-18 December 2007.
- [7] Fanizzi, N., d Amato, C., Esposito, F.: Statistical learning for inductive query answering on owl ontologies. In: Proceedings of the 7th International Semantic Web Conference (ISWC), pp. 195–212 (2008).
- [8] Avila Garcez, A.S., Broda, K., Gabbay, D.: Symbolic knowledge extraction from trained neural networks: A sound approach. Artificial Intelligence 125, 155–207 (2001).

- [9] Avila Garcez, A.S., Broda, K., Gabbay, D.: The connectionist inductive learning and logic programming system. *Applied Intelligence, Special Issue on Neural networks and Structured Knowledge* 11, 59–77 (1999).
- [10] Hitzler, P., Holldobler, S., Seda, A.: Logic programs and connectionist networks. *Journal of Applied Logic*, 245–272 (2004).
- [11] Pinkas, G.: Propositional non-monotonic reasoning and inconsistency in symmetric neural networks. In: *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pp. 525–530 (1991).
- [12] Y. Bengio, “Learning deep architectures for AI,” *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.
- [13] N. Morgan, “Deep and wide: Multiple layers in automatic speech recognition,” *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 1, pp. 7–13, 2012.
- [14] A. Mohamed, G.E. Dahl, and G. Hinton, “Acoustic modeling using deep belief networks,” *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 1, pp. 14–22, 2012.
- [15] G. Sivaram and H. Hermansky, “Sparse multilayer perceptron for phoneme recognition,” *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 1, pp. 23–29, 2012.
- [16] H. Lee, C. Ekanadham, and A. Ng, “Sparse deep belief net model for visual area v2,” *Advances in neural information processing systems*, vol. 20, pp. 873–880, 2008.
- [17] Y. Tang and C. Eliasmith, “Deep networks for robust visual recognition,” in *International Conference on Machine Learning*. Citeseer, 2010, vol. 28.
- [18] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, “Unsupervised learning of hierarchical representations with convolutional deep belief networks,” *Communications of the ACM*, vol. 54, no. 10, pp. 95–103, 2011.
- [19] K. Sohn, D.Y. Jung, H. Lee, and A.O. Hero, “Efficient learning of sparse, distributed, convolutional feature representations for object recognition,” in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2643–2650.

- [20] A. Krizhevsky, I. Sutskever, and G.E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems* 25, 2012, pp. 1106–1114.
- [21] Krizhevsky, A., Sutskever, I., Hinton, G.: Imagenet classification with deep convolutional neural networks. In: *NIPS*, pp. 1106–1114 (2012). B. Schoelkopf, A. Smola *Learning with Kernels*, The MIT Press, Cambridge, 2002.
- [22] Neverova, N., Wolf, C., Taylor, G.W., Nebout, F.: Moddrop: adaptive multi-modal gesture recognition. *arXiv:1501.00102* (2014).
- [23] Kahou, S.E., Froumenty, P., Pal, C.: Facial expression analysis based on high dimensional binary features. In: *ECCV Workshop on Computer Vision with Local Binary Patterns Variants*. Zurich, Switzerland (2014).
- [24] Kalchbrenner, N., Grefenstette, E., Blunsom, P.: A convolutional neural network for modelling sentences. *arXiv:1404.2188* (2014).
- [25] Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N., et al.: Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Sig. Proc. Magazine*, 29(6), 82–97 (2012).
- [26] C.N. Anagnostopoulos, T. Iliou, and I. Giannoukos, “Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011,” *Artificial Intelligence Review*, pp. 1–23, 2012.
- [27] M. El Ayadi, M.S. Kamel, and F. Karray, “Survey on speech emotion recognition: Features, classification schemes, and databases,” *Pattern Recognition*, vol. 44, no. 3, pp. 572–587, 2011.
- [28] D. Ververidis and C. Kotropoulos, “Fast and accurate sequential floating forward feature selection with the bayes classifier applied to speech emotion recognition,” *Signal Processing*, vol. 88, no. 12, pp. 2956–2970, 2008.
- [29] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C.-M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan, “Analysis of emotion recognition using facial expressions, speech and multimodal information,” in *Proceedings of the 6th international conference on Multimodal interfaces*. ACM, 2004, pp. 205–211.

- [30] G.W. Taylor, G.E. Hinton, and S.T. Roweis, “Modeling human motion using binary latent variables,” *Advances in neural information processing systems*, vol. 19, pp. 1345, 2007.
- [31] S.E. Kahou, et al, “Facial expression analysis based on high dimensional binary features,” *ECCV Workshop on Computer Vision*, Zurich, Switzerland, 2014.
- [32] Y. Kim, H. Lee and E.M. Provost, “Deep learning for robust feature generation in audiovisual emotion recognition,” *ICASSP*, Vancouver, British Columbia, Canada, May 2013. N. Fanizzi, C. d’ Amato. A declarative kernel for ALC concept descriptions. In : F. Esposito, Z. W. Ras, D. Malerba, G. Semeraro (eds) *ISMIS 2006*. LNCS (LNAI) Springer, vol. 4203, pp. 322-331, 2006.
- [33] S.E. Kahou, et al, “Emonets: Multimodal deep learning approaches for emotion recognition in video,” *arXiv:1503.01800v2*, 30 March 2015.
- [34] Y. Li, S. Wang, Y. Zhao and Q. Ji, “Simultaneous facial feature tracking and facial expression recognition”, *IEEE Transaction on Image Processing*, vol.22, pp. 2559-2573, 2013
- [35] A. Saeed, A. Al-Hamadi, R. Niese and M. Elzobi, “Frame-based facial expression recognition using geometric features”, *Advances in Human-Computer Interaction*, vol. 2014, pp. 1-13, 2014
- [36] A. C. Cruz, B. Bhanu and N. S. Thakoor, “Vision and attention theory base esampling for continuous facial emotion recognition”, *IEEE Transaction on Affective Computing*, vol. 5, pp. 418-431, 2014
- [37] S. Haykin. *Neural Networks: A Comprehensive Foundation*, Prentice Hall International, Inc, 1999.
- [38] K. Διαμαντάρας. *Τεχνητά Νευρωνικά Δίκτυα, Κλειδάριθμος*, 2007.
- [39] Sebastian Bader, Pascal Hitzler, Steffen Hölldobler. ‘Connectionist Model Generation: A First-Order Approach,’ *Neurocomputing*, Elsevier, 2008.

- [40] B. Hammer, P. Hitzler (eds.). Perspectives of Neural-Symbolic Integration. *Studies in Computational Intelligence*, Vol. 77. Springer, ISBN 978-3-540-73952-1, 2007.
- [41] A. Chortaras, G. Stamou, A. Stafylopatis, Adaptation of Weighted Fuzzy Programs, *Lecture Notes in Computer Science*, 4132, pp. 45-54, 2006.
- [42] W. Duch. Similarity based methods: a general framework for classification, approximation and association, *Control and Cybernetics* 29 (4), pp. 937-968, 2000.
- [43] M. Ehrig, P. Haase, M. Hefke, N. Stojanovic. Similarity for ontologies – a comprehensive framework. In : Proceedings of the 13th European Conference on Information Systems, ECIS 2005.
- [44] N. Fanizzi, C. d Amato, and F. Esposito. Randomized metric induction and evolutionary conceptual clustering for semantic knowledge bases. In Proceedings of CIKM 07, 2007.
- [45] N. Simou, Th. Athanasiadis, S. Kollias, G. Stamou, and A. Stafylopatis. Semantic adaptation of neural network classifiers in image segmentation. 18th International Conference on Artificial Neural Networks (ICANN 2008), pp. 907-916, September 2008, Prague, Czech Republic, 2008.
- [46] N. Doulamis, A. Doulamis, and S. Kollias. On-line retrainable neural networks: Improving performance of neural networks in image analysis problems. *IEEE Transactions on Neural Networks* 11, pp. 1-20, 2000.
- [47] B. Schoelkopf, A. Smola *Learning with Kernels*, The MIT Press, Cambridge, 2002.
- [48] T. Gartner, J. Lloyd, P. Flach. Kernels and distances for structured data. *Machine Learning*, 57, pp. 205-232, 2004.
- [49] J. Shawe-Taylor, N. Cristianini. *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.

- [50] C. Bishop. *Pattern Recognition and Machine Learning*, Springer, 2006.
- [51] Ming-Hsuan Yang, David J. Kriegman, and Narendra Ahuja, “Detecting Faces in Images: A Survey”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 1, January 2002.
- [52] C. J. C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition”, *Knowledge Discovery and Data Mining*, 2(2), 1998.
- [53] E. Osuna, R. Freund, and F. Girosi, “Training support vector machines: An application to face detection”, , *In Proceedings of CVPR'97, Puerto Rico*, 1997.
- [54] J. Ahlberg, “An Optimisation Approach to facial feature extraction”, *Proc. of IEEE Workshop on Real-Time Analysis and Tracking of Face and Gesture in Real-Time Systems*, Kerkyra, Greece, September 1999.
- [55] A.L. Yuille, D.S. Cohen and P.W. Hallinan, “Feature Extraction from Faces Using Deformable Templates”, *International Journal Computer Vision*, vol. 8, no. 2, pp. 99-111, August 1992.
- [56] R. Chellappa, C. L. Wilson and S. Sirohey, “Human and Machine Recognition of Faces: A Survey”, *Proc .of the IEEE*, vol. 83, no. 5, pp. 705-740, May 1995.
- [57] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz and J. G. Taylor, “Emotion Recognition in Human Computer Interaction”, *IEEE Signal Processing Magazine*, vol.18, no. 1, Jan.2001.
- [58] P. Eisert and B. Girod, “Model-Based Estimation of Facial Expression Parameters from Image Sequences”, *Proc. of IEEE International Conference on Image Processing*, 1997.
- [59] Y. Tian, T. Kanade and J. F. Cohn, “Multi-State Based Facial Feature Tracking and Detection”, tech. report CMURITR9918, Robotics Institute, Carnegie Mellon University, Aug.1999.

- [60] L.D. Harmon, M.K. Khan, R. Lasch and P.F. Ramig, "Machine Identification of Human Faces", *Pattern Recognition*, vol. 13, no. 2, pp. 97-110, 1981.
- [61] X. Jia and M.S. Nixon, "Extending the Feature Vector for Automatic Face Recognition", *IEEE Tans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 12, pp. 1167-1176, December 1995.
- [62] K. Karpouzis, G. Votsis, N. Tsapatsoulis and S. Kollias, "Compact 3D Model Generation based on 2D Views of Human Faces: Application to Face Recognition", *Machine Graphics and Vision*, vol. 7, no.1-2, pp. 75-85, 1998.
- [63] M. Kass, A. Witkin and D. Terzopoulos, "Snakes: Active Contour Models", *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321-331, 1988.
- [64] B.K. Low and M.K. Ibrahim, "A Fast and Accurate Algorithm for Facial Feature Segmentation", *Proc. of IEEE International Conference on Image Processing*, 1997.
- [65] K. Sobottka and I. Pitas, "Looking for Faces and Facial Features in Color Images", *Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications*, vol. 7, no. 1, 1997.
- [66] D. Metaxas, "Deformable Model and HMM-Based Tracking, Analysis and Recognition of Gestures and Faces", *Proc. of IEEE Workshop on Real-Time Analysis and Tracking of Face and Gesture in Real-Time Systems*, Kerkyra, Greece, September 1999.
- [67] A. Samal and P.A. Iyengar, "Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey", *Pattern Recognition*, vol. 25, no. 1, pp. 65-76, 1992.
- [68] M. Preda & F. Prêteux, "Advanced animation framework for virtual characters within the MPEG-4 standard", *Proc. of the Intl Conference on Image Processing*, Rochester, NY, 2002.

- [69] Y. Votsis, N. Drosopoulos and S. Kollias, "Facial feature segmentation: a modularly optimal approach on real sequences" *Signal Processing: Image Communication*, vol. 18 , no 1, pp. 67-89, 2003.
- [70] G. Tsechpenakis, N. Tsapatsoulis and S. Kollias, "Probabilistic Boundary-Based Contour Tracking with Snakes in Natural Cluttered Video Sequences", *International Journal of Image and Graphics: Special Issue on Deformable Models for Image Analysis and Pattern Recognition*, to appear, 2003.
- [71] N. Tsapatsoulis, A. Raouzaïou, S. Kollias, R. Cowie and E. Douglas-Cowie, "Emotion Recognition and Synthesis based on MPEG-4 FAPs", in *MPEG-4 Facial Animation*, Igor Pandzic, R. Forchheimer (eds), John Wiley & Sons, UK, 2002.
- [72] K. Karpouzis, A. Raouzaïou, A. Drosopoulos, S. Ioannou, T. Balomenos, N. Tsapatsoulis and S. Kollias, "Facial Expression and Gesture Analysis for Emotionally-rich Man-machine Interaction", N. Sarris, M. Strintzis, (eds.), "3D Modeling and Animation: Synthesis and Analysis Techniques", *Idea Group Publ.*, 2003.
- [73] D. Kollias, G. Marandianos, A. Raouzaïou and A. Stafylopatis, "Interweaving Deep Learning and Semantic Techniques for Emotion Analysis in Human-Machine Interaction", SMAP2015, Italy, November 2015
- [74] Y. LeCun, M. Ranzato, "Deep Learning Tutorial", ICML, Atlanta, 2013