



## ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ  
ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

### **Αναγνώριση Χωροχρονικών Συσχετίσεων για την Πρόβλεψη Κυκλοφορίας σε Αστικά Οδικά Δίκτυα με χρήση Τεχνικών Μηχανικής Μάθησης**



ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

**ΦΑΦΟΥΤΕΛΛΗ ΠΑΝΑΓΙΩΤΗ**

**Επιβλέπων :** Νεκτάριος Κοζύρης, Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2019





## ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ  
ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

### **Αναγνώριση Χωροχρονικών Συσχετίσεων για την Πρόβλεψη Κυκλοφορίας σε Αστικά Οδικά Δίκτυα με χρήση Τεχνικών Μηχανικής Μάθησης**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

**ΦΑΦΟΥΤΕΛΛΗ ΠΑΝΑΓΙΩΤΗ**

**Επιβλέπων :** Νεκτάριος Κοζύρης, Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 16<sup>η</sup> Οκτωβρίου 2019.

(Υπογραφή)

.....

Νεκτάριος Κοζύρης

Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....

Γεώργιος Στάμου

Αν. Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....

Ελένη Βλαχογιάννη

Αν. Καθηγήτρια Ε.Μ.Π.

Αθήνα, Οκτώβριος 2019

(Υπογραφή)

.....

**ΦΑΦΟΥΤΕΛΛΗΣ ΠΑΝΑΓΙΩΤΗΣ**

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

Copyright © Παναγιώτης Φαφουτέλλης, 2019

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

Η βραχυπρόθεσμη πρόβλεψη κυκλοφοριακών συνθηκών είναι ένα πεδίο έρευνας που συγκεντρώνει υψηλό ενδιαφέρον, λόγω της σπουδαιότητάς του στον έλεγχο και τη διαχείριση της κυκλοφορίας και στο σχεδιασμό έξυπνων μεταφορικών συστημάτων. Η πρόσφατη εισαγωγή τεχνικών Μηχανικής Μάθησης σε συνδυασμό με την πληθώρα διαθέσιμων δεδομένων που συλλέγονται από τις έξυπνες φορητές συσκευές των χρηστών του δικτύου έχει διευρύνει τους ορίζοντες των ερευνητών και έχει καταστήσει δυνατή την εφαρμογή καινοτόμων προσεγγίσεων. Παρ' όλα αυτά, οι ερευνητές δε θα πρέπει να υποτιμούν τη σημασία των χωρικών και χρονικών εξαρτήσεων ενός οδικού δικτύου και τις κλασικές στατιστικές μεθόδους πρόβλεψης, οι οποίες επίσης προσφέρουν καλύτερη επεξηγηματικότητα. Σε αυτή την εργασία, εντοπίζονται οι χωροχρονικές συσχετίσεις του οδικού δικτύου του κέντρου της κινεζικής πόλης Σιάν, με χρήση της συσχέτισης κατά Pearson, της Αμοιβαίας Πληροφορίας και του αλγορίθμου Dynamic Time Warping στις χρονοσειρές της ταχύτητας των τμημάτων του οδικού δικτύου. Από τις παραπάνω μεθοδολογίες, η Αμοιβαία Πληροφορία κρίνεται ως περισσότερο κατάλληλη στη συγκεκριμένη περίπτωση, καθώς οδηγεί σε ένα ακριβές πρότυπο πρόβλεψης Μπεϋζιανού Δικτύου. Τα συμπεράσματα της παραπάνω ανάλυσης αξιοποιούνται στην ανάπτυξη δύο δικτύων LSTM για την μεσοπρόθεσμη και βραχυπρόθεσμη πρόβλεψη των κυκλοφοριακών συνθηκών ενός οδικού τμήματος του οδικού δικτύου με ακρίβεια 89% και 86% αντίστοιχα. Η μεγάλη ακρίβεια των προβλέψεων αφενός τονίζει τη σπουδαιότητα της πραγματοποίησης ανάλυσης χωροχρονικών συσχετίσεων πριν την εφαρμογή αλγορίθμων πρόβλεψης και αφετέρου πιστοποιεί την καταλληλότητα του δικτύου LSTM στην πρόβλεψη χρονοσειρών.

Λέξεις κλειδιά: χωροχρονικές συσχετίσεις, πρόβλεψη κυκλοφορίας, αμοιβαία πληροφορία, dynamic time warping, δίκτυο LSTM

## **Abstract**

Short-term traffic forecasting is a field of research that has always attracted significant attention, due to its importance to Traffic Control and Management and the design of Intelligent Transportation Systems. The recent introduction of Machine Learning techniques in traffic forecasting and the vast amount of data collected using the network users' smart portable devices has broadened the researchers' horizons, making more innovative approaches possible. However, researchers should not disregard the importance of spatiotemporal relations of a road network and classic statistical modeling, which also provide better interpretation. In this work, we reveal the spatiotemporal relationships of the road network of Xian, China using Pearson's correlation, Mutual Information and Dynamic Time Warping on the network's speed time series. From the above methodologies, Mutual Information was proven to be more suitable for the specific case, as it leads to an accurate Bayesian Network prediction model. The findings of the above analysis are further exploited for the development of two LSTM networks for medium-term and short-term traffic prediction of a road section of the network, having 89% and 86% accuracy respectively. The high accuracy of the predictions highlights the significance of conducting spatiotemporal analysis prior to predictions generation and, on the other hand, demonstrates the suitability of LSTM networks in time series predictions.

Keywords: spatiotemporal relations, traffic forecasting, mutual information, dynamic time warping, LSTM network



## ΠΡΟΛΟΓΟΣ

Η πρόβλεψη κυκλοφοριακών συνθηκών είναι ένα πεδίο έρευνας που συγκεντρώνει μεγάλο ενδιαφέρον, λόγω της σπουδαιότητάς του στη διαχείριση της κυκλοφορίας και σε πλήθος πραγματικών εφαρμογών ευφών μεταφορικών συστημάτων. Ιδιαίτερα λόγω της εισόδου τεχνικών Μηχανικής Μάθησης και Δεδομένων Ευρείας Κλίμακας στη διαδικασία της πρόβλεψης, η έρευνα στο συγκεκριμένο πεδίο έχει σημειώσει σημαντική πρόοδο τα τελευταία χρόνια.

Στην παρούσα εργασία, παρουσιάζεται μία μέθοδος εντοπισμού των χωροχρονικών συσχετίσεων που διέπουν τη λειτουργία ενός αστικού οδικού δικτύου, με βάση την οποία, στη συνέχεια, αναπτύσσονται δύο πρότυπα πρόβλεψης των μεσοπρόθεσμων και των βραχυπρόθεσμων κυκλοφοριακών συνθηκών, αντίστοιχα, με χρήση του δημοφιλούς δικτύου LSTM. Στόχος της παρούσας εργασίας είναι να τονιστεί η χρησιμότητα της πραγματοποίησης ανάλυσης χωρικών και χρονικών συσχετίσεων για την αύξηση της ακρίβειας των προβλέψεων, κάτι που στην πρόσφατη βιβλιογραφία υποτιμάται, όπως και η εφαρμογή κλασικών στατιστικών μεθόδων.

Αντίθετα, η προσοχή των ερευνητών φαίνεται πως είναι στραμμένη στις πιο πρόσφατες μεθοδολογίες Μηχανικής Μάθησης, όπως είναι τα Βαθιά Νευρωνικά Δίκτυα, τα Συνελικτικά Δίκτυα και τα Επαναλαμβανόμενα Δίκτυα, οι οποίες ωστόσο συχνά δεν παρέχουν επεξηγηματικότητα των προβλέψεων. Παρά τα πλεονεκτήματα που παρουσιάζουν οι παραπάνω μεθοδολογίες, τονίζεται ότι σε πραγματικές εφαρμογές η γνώση των χωροχρονικών συσχετίσεων και γενικά της δυναμικής του οδικού δικτύου είναι εξίσου σημαντική με την ακρίβεια της πρόβλεψης. Για το λόγο αυτό, προσεγγίσεις «μαύρου κουτιού» θα πρέπει να αποφεύγονται.

Στη βιβλιογραφία μπορεί ακόμα κανείς να εντοπίσει δύο διακριτές κατηγορίες ερευνών ως προς τον τρόπο συλλογής των δεδομένων που αξιοποιούνται στις προβλέψεις. Παλιότερα, ο πιο συνηθισμένος τρόπος ήταν με τοποθέτηση αισθητήρων στο οδόστρωμα, ενώ τα τελευταία χρόνια κερδίζει συνεχώς έδαφος η συλλογή δεδομένων με χρήση του στίγματος GPS των έξυπνων κινητών τηλεφώνων των οδηγών. Η δεύτερη τεχνική είναι περισσότερο οικονομική και αποδοτική στη συλλογή μεγάλου όγκου δεδομένων. Ακόμα, παρατηρούνται δύο διαφορετικές προσεγγίσεις στην πρόβλεψη κυκλοφοριακών συνθηκών: πρόβλεψη σε επίπεδο οδού και πρόβλεψη σε επίπεδο δικτύου. Στην πρώτη περίπτωση η πρόβλεψη πραγματοποιείται για ένα οδικό τμήμα με δεδομένα που προέρχονται αποκλειστικά από αυτό, ενώ στη δεύτερη τα δεδομένα προέρχονται από ολόκληρο το οδικό δίκτυο και η πρόβλεψη αφορά ένα ή περισσότερα τμήματα.

Στην παρούσα διπλωματική εργασία, γίνεται πρόβλεψη σε επίπεδο δικτύου με χρήση δεδομένων που παρέχονται από την κινεζική εταιρεία μίσθωσης οχημάτων Didi Chuxing και αφορούν 3,2 εκατομμύρια διαδρομές των οχημάτων της κατά την περίοδο 2 – 30 Νοεμβρίου 2016 στην πόλη Σιάν. Στην αρχή πραγματοποιείται προεπεξεργασία των δεδομένων που αφορά τη μετατροπή των συντεταγμένων των



σημείων των διαδρομών στο Παγκόσμιο Γεωδαιτικό Σύστημα WGS 84, την αντιστοίχισή τους στα τμήματα του οδικού δικτύου της πόλης, τον υπολογισμό της ταχύτητας κίνησης των οχημάτων κατά τη διάρκεια των αντίστοιχων διαδρομών και, τέλος, την κατασκευή των χρονοσειρών των ταχυτήτων όλων των οδικών τμημάτων με χρονικά βήματα μίας ώρας και δεκαπέντε λεπτών αντίστοιχα.

Στη συνέχεια, εντοπίζονται οι συσχετίσεις των χρονοσειρών των ταχυτήτων των τμημάτων με χρήση τριών στατιστικών δεικτών: Γραμμική Συσχέτιση κατά Pearson, Αμοιβαία Πληροφορία και Dynamic Time Warping. Οι συγκεκριμένες τεχνικές είναι οι πλέον δημοφιλείς για την αναγνώριση των συσχετίσεων χρονοσειρών και όχι μόνο. Οι τρεις μεθοδολογίες παρουσιάζουν σημαντικές διαφορές μεταξύ τους, όσον αφορά τα χωροχρονικά μοτίβα εξαρτήσεων που προκύπτουν. Για να επιλεγεί η καταλληλότερη μεταξύ των μεθοδολογιών, αναπτύσσονται δύο Μπεϋζιανά Δίκτυα πρόβλεψης ταχύτητας επιλεγμένου οδικού τμήματος, ένα με αξιοποίηση μόνο των 20 τμημάτων με τη μεγαλύτερη συσχέτιση με βάση την Αμοιβαία Πληροφορία και ένα με αξιοποίηση μόνο των 20 τμημάτων με τη μεγαλύτερη συσχέτιση με βάση τον αλγόριθμο Dynamic Time Warping. Από αυτά προκύπτει πως ο υπολογισμός της Αμοιβαίας Πληροφορίας είναι η πλέον αξιόπιστη μέθοδος στη συγκεκριμένη εφαρμογή.

Τέλος, αξιοποιούνται τα παραπάνω συμπεράσματα για την ανάπτυξη δύο δικτύων LSTM με στόχο την πρόβλεψη των κυκλοφοριακών συνθηκών. Στα δύο δίκτυα αξιοποιούνται τα 20 τμήματα με τη μεγαλύτερη τιμή Αμοιβαίας Πληροφορίας ως προς το επιλεγμένο. Το πρώτο προβλέπει με ορίζοντα μίας ώρας και το δεύτερο με ορίζοντα 15 λεπτών, με ακρίβεια 89% και 86% αντίστοιχα. Τα αποτελέσματα αυτά αποδεικνύουν τη σπουδαιότητα της υλοποίησης ανάλυσης χωροχρονικών συσχετίσεων προτού προβεί κανείς σε προβλέψεις. Επιπλέον, η διαδικασία που περιεγράφηκε μειώνει τη διαστατικότητα του προβλήματος, κάτι που είναι πολύ σημαντικό για την καλή λειτουργία του αλγορίθμου και γενικότερα την αποδοτικότητα του υπολογιστικού συστήματος.

Η μελλοντική έρευνα θα περιλαμβάνει αξιοποίηση μεγαλύτερου όγκου δεδομένων, κάτι που οδηγεί και στη χρήση ισχυρότερου υπολογιστικού συστήματος και πιο εξειδικευμένων και σύγχρονων τεχνικών αναγνώρισης συσχετίσεων και πρόβλεψης των κυκλοφοριακών συνθηκών. Τέλος, θα πραγματοποιηθούν πιο βραχυπρόθεσμες προβλέψεις, έως και προβλέψεις σε πραγματικό χρόνο, που θα αφορούν χρονικό ορίζοντα μεγαλύτερο του ενός μόνο βήματος και το σύνολο των τμημάτων του δικτύου, και όχι μόνο ένα από αυτά.

## Περιεχόμενα

Κεφάλαιο 1 Εισαγωγή .....	1
1.1 Βραχυπρόθεσμη Πρόβλεψη Κυκλοφορίας .....	1
1.2 Αντικείμενο Διπλωματικής Εργασίας .....	3
1.2.1 Συνεισφορά.....	3
1.3 Οργάνωση Κειμένου.....	4
Κεφάλαιο 2 Βιβλιογραφική Ανασκόπηση .....	5
2.1 Εισαγωγή.....	5
2.2 Μεθοδολογίες Πρόβλεψης Κυκλοφοριακών Μεγεθών .....	5
2.3 Τεχνολογίες και Μέθοδοι Συλλογής Δεδομένων .....	7
2.4 Επιρροή Τύπου Οδού .....	9
2.5 Βασικά Συμπεράσματα .....	9
Κεφάλαιο 3 Μεθοδολογική Προσέγγιση.....	11
3.1 Προτεινόμενη Μεθοδολογία.....	11
3.2 Θεωρητικό υπόβαθρο .....	12
3.2.1 Γραμμική Συσχέτιση και Αμοιβαία Πληροφορία.....	12
3.2.2 Dynamic Time Warping .....	13
3.2.3 Ταξινομητής Μπεϋζιανού Δικτύου .....	14
3.2.4 Νευρωνικό Δίκτυο Μακράς και Βραχείας Μνήμης .....	15
Κεφάλαιο 4 Εφαρμογή Μεθοδολογίας .....	17
4.1 Περιγραφή και Προεπεξεργασία Δεδομένων .....	17
4.2 Εφαρμογή και Αποτελέσματα.....	22
4.2.1 Υπάρχουν χωρικές και χρονικές συσχετίσεις σε ένα οδικό δίκτυο; .....	22
4.2.2 Σύγκριση μεθοδολογιών .....	29
4.2.3 Ποια είναι η επιρροή των χωροχρονικών συσχετίσεων στη βραχυπρόθεσμη πρόβλεψη των κυκλοφοριακών συνθηκών;.....	31
4.2.4 Αξιοποίηση Τεχνικών Μηχανικής Μάθησης στην πρόβλεψη της κυκλοφορίας.....	32
4.2.5 Αξιολόγηση επιδόσεων δικτύου LSTM.....	34
4.2.6 Επέκταση δικτύου LSTM για βραχυχρόνια πρόβλεψη κυκλοφοριακών συνθηκών .....	35
Κεφάλαιο 5 Συμπεράσματα .....	39

5.1 Γενικά .....	39
5.2 Βασικά συμπεράσματα .....	39
5.3 Εισηγήσεις για περαιτέρω έρευνα.....	40
Βιβλιογραφία .....	41

# Κεφάλαιο 1

## Εισαγωγή

### 1.1 Βραχυπρόθεσμη Πρόβλεψη Κυκλοφορίας

Ένα από τα μεγαλύτερα προβλήματα που αντιμετωπίζουν οι σύγχρονες μεγαλουπόλεις είναι αυτό της κυκλοφοριακής συμφόρησης. Το μεγάλο κύμα αστικοποίησης των τελευταίων δεκαετιών σε συνδυασμό με την αύξηση της ιδιοκτησίας οχημάτων παγκοσμίως έχει εντείνει το πρόβλημα. Το φαινόμενο αυτό έχει πολλές συνέπειες. Κατ' αρχάς, οι μετακινήσεις έχουν άμεσο αντίκτυπο στην οικονομία. Ένα λάθος σχεδιασμένο οδικό δίκτυο και γενικότερα δίκτυο μεταφορών προκαλεί την άσκοπη σπατάλη χρόνου για τους χρήστες του. Γενικότερα, οι καλές συγκοινωνιακές υποδομές συνδέονται με μειωμένη ανεργία, αυξημένες επενδύσεις και καλύτερη πρόσβαση στις αγορές [1].

Επιπλέον, η κυκλοφοριακή συμφόρηση εντείνει το πρόβλημα της εκπομπής βλαβερών αερίων στην ατμόσφαιρα. Είναι γνωστό ότι ένα όχημα που κάνει συχνές στάσεις και κινείται σε πολύ χαμηλές ταχύτητες καταναλώνει μεγαλύτερη ποσότητα καυσίμου και απελευθερώνει μεγαλύτερη ποσότητα αερίων του θερμοκηπίου [2]. Εξάλλου, οι μεταφορές είναι γενικότερα η πρώτη αιτία μόλυνσης της ατμόσφαιρας, ειδικά στις μεγάλες πόλεις, σύμφωνα με στοιχεία της Ευρωπαϊκής Ένωσης [3]. Ένα λειτουργικό και βιώσιμο σύστημα μεταφορών αμβλύνει το παραπάνω πρόβλημα και τις συνέπειές του στο περιβάλλον και τον άνθρωπο.

Τέλος, μία ακόμα σημαντική συνέπεια της κυκλοφοριακής συμφόρησης είναι η επίδρασή της στην ψυχολογία του ατόμου. Σύμφωνα με πρόσφατες έρευνες, η οδήγηση σε συνθήκες κυκλοφοριακής συμφόρησης προκαλεί άγχος και εκνευρισμό στον οδηγό [4]. Ακόμα, έχει παρατηρηθεί και αύξηση των οδικών ατυχημάτων εξαιτίας των παραπάνω [5].

Οι αρμόδιες αρχές, καθώς και οι διαχειριστές του δικτύου, οφείλουν να λαμβάνουν τις κατάλληλες αποφάσεις ώστε το οδικό δίκτυο να λειτουργεί στο βέλτιστο επίπεδο της στάθμης εξυπηρέτησης. Οι αποφάσεις που πρέπει να λάβουν αφορούν στη ρύθμιση της σήμανσης και της σηματοδότησης, τη διοχέτευση της κυκλοφορίας από κατάλληλα οδικά τμήματα, την εκτροπή της κυκλοφορίας σε περίπτωση αναπάντεχων συμβάντων και καιρικών συνθηκών κ.ά.. Για τη λήψη των βέλτιστων αποφάσεων είναι απαραίτητη η γνώση του αναμενόμενου κυκλοφοριακού φόρτου του δικτύου καθώς και των κανόνων που τον καθορίζουν και τον κατανέμουν στο οδικό δίκτυο. Για τα τελευταία, απαιτείται ένας μεγάλος όγκος δεδομένων προηγούμενων περιόδων, καθώς και ένα ακριβές πρότυπο πρόβλεψης, ειδικά μάλιστα όταν πρέπει να ληφθούν αποφάσεις που αφορούν τις συνθήκες που θα επικρατήσουν βραχυπρόθεσμα στο δίκτυο, λίγα μόνο λεπτά αργότερα δηλαδή.

Η βραχυπρόθεσμη πρόβλεψη των κυκλοφοριακών συνθηκών αποτελούσε ανέκαθεν ένα πεδίο με έντονο ερευνητικό ενδιαφέρον λόγω της μεγάλης σπουδαιότητάς του στην αποτελεσματική διαχείριση της κυκλοφορίας καθώς και στην ανάπτυξη ευφυών συστημάτων μεταφορών (Intelligent Transportation Systems – ITS) και φιλικών προς το χρήστη εφαρμογών παροχής πληροφοριών [6]. Η ακριβής πρόβλεψη των κυκλοφοριακών συνθηκών μειώνει την αβεβαιότητα, διευκολύνοντας τους διαχειριστές του οδικού δικτύου κατά τη διαδικασία λήψης αποφάσεων και χάραξης πολιτικών. Αποτέλεσμα των παραπάνω είναι ο αποδοτικότερος έλεγχος των κυκλοφοριακών συνθηκών και η εξασφάλιση βιώσιμων συνθηκών στο οδικό δίκτυο.

Στις μέρες μας, η ευρεία χρήση έξυπνων συσκευών και συστημάτων (smartphones, GPS, πληροφοριακά συστήματα εντός του οχήματος κ.α.), τα οποία καταγράφουν ένα τεράστιο όγκο δεδομένων σχετικών με τις μετακινήσεις, δίνει στους ερευνητές τη δυνατότητα να αναπτύξουν πρότυπα πρόβλεψης περισσότερο ακριβή, που ενημερώνονται σε πραγματικό χρόνο καθώς και με μεγαλύτερη χρονική ανάλυση. Τα δεδομένα κινητικότητας αφορούν, μεταξύ άλλων, στην ακριβή θέση των οχημάτων σε πραγματικό χρόνο, την ταχύτητα κίνησής τους, τον προορισμό τους και τη διαδρομή που πρόκειται να ακολουθήσουν. Η πληθώρα και η ποιότητα των διαθέσιμων δεδομένων υπήρξε ο καταλύτης για την στροφή του ενδιαφέροντος των ερευνητών από την παραδοσιακή στατιστική προσέγγιση για την πρόβλεψη των κυκλοφοριακών συνθηκών στα οδηγούμενα από τα δεδομένα (data-driven) πρότυπα μηχανικής μάθησης σε συνδυασμό με αλγορίθμους εξόρυξης δεδομένων (Data Mining) και αξιοποίησης δεδομένων ευρείας κλίμακας (Big Data) [6].

Η ανάπτυξη τέτοιων μοντέλων έγινε εφικτή τις τελευταίες δύο δεκαετίες χάρη στην αλματώδη αύξηση της υπολογιστικής ισχύος στους σύγχρονους υπολογιστές, οι οποίοι μπορούν να αντιμετωπίσουν τη μεγάλη υπολογιστική πολυπλοκότητα των απαιτούμενων υπολογισμών μέσα σε μερικά δευτερόλεπτα και να διαχειριστούν έναν τόσο μεγάλο όγκο δεδομένων με αποδοτικό τρόπο.

Μία πολύ δημοφιλής προσέγγιση για τη βραχυπρόθεσμη πρόβλεψη της κυκλοφορίας είναι ο εντοπισμός και η αναγνώριση των σχέσεων μεταξύ των θεμελιωδών μεταβλητών της κυκλοφοριακής τεχνικής (φόρτος, ταχύτητα, πυκνότητα) των διαφορετικών οδικών τμημάτων ενός οδικού δικτύου με την πάροδο του χρόνου. Η παραπάνω διαδικασία γίνεται με τον υπολογισμό στατιστικών μεγεθών και μετρικών, τα οποία έχουν ισχυρό μαθηματικό υπόβαθρο [7]. Παρ' όλα αυτά, λόγω της υπέρμετρης χρήσης τεχνικών Βαθιάς Μηχανικής Μάθησης, οι ερευνητές φαίνεται να υποτιμούν τη σπουδαιότητα της αξιοποίησης των χωροχρονικών εξαρτήσεων των μεγεθών της κυκλοφορίας σε ένα οδικό δίκτυο και της δυναμικής εξέλιξής τους στην πρόβλεψη και στην αποδοτική διαχείριση της κυκλοφορίας. Τα Τεχνητά Νευρωνικά Δίκτυα παράγουν πολύ ακριβείς προβλέψεις των συνθηκών της κυκλοφορίας, καθώς είναι σε θέση να προσεγγίσουν σχεδόν οποιαδήποτε συνάρτηση, ανεξαρτήτως του βαθμού της μη γραμμικότητάς της και χωρίς προηγούμενη γνώση της μορφής της [8]. Για το λόγο αυτό, συχνά χρησιμοποιούνται ως «μαύρο κουτί» για την πρόβλεψη των επιθυμητών μεγεθών,

με αποτέλεσμα να μη γίνεται σαφής ο μηχανισμός λειτουργίας του οδικού δικτύου και οι σχέσεις αιτίας-αποτελέσματος που τον καθορίζουν. Το τελευταίο είναι απαραίτητο για τη χάραξη πολιτικών και την αντιμετώπιση έκτακτων συνθηκών.

Σε αντίθεση με τα παραδοσιακά Νευρωνικά Δίκτυα και τις παραδοσιακές τεχνικές Μηχανικής Μάθησης, πρότυπα που περιλαμβάνουν επίσης και τις χωροχρονικές συσχετίσεις του οδικού δικτύου προσφέρουν καλύτερη επεξηγηματικότητα και κατανόηση του μηχανισμού που παράγει τις προβλέψεις.

## 1.2 Αντικείμενο Διπλωματικής Εργασίας

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη μιας μεθοδολογίας για τον εντοπισμό και την ανάλυση των χωρικών και χρονικών εξαρτήσεων μεταξύ των κυκλοφοριακών συνθηκών των τμημάτων ενός οδικού δικτύου με γραμμικές και μη γραμμικές μεθόδους, καθώς και η πρόβλεψη των μελλοντικών κυκλοφοριακών συνθηκών με αξιοποίηση τεχνικών μηχανικής μάθησης.

Στόχος της διπλωματικής εργασίας είναι να συνεισφέρει στην απάντηση των παρακάτω ερευνητικών ερωτημάτων:

- Υπάρχουν χωρικές και χρονικές συσχετίσεις σε ένα οδικό δίκτυο;
- Ποια είναι η επιρροή των χωροχρονικών συσχετίσεων στη βραχυπρόθεσμη πρόβλεψη των κυκλοφοριακών συνθηκών;

### 1.2.1 Συνεισφορά

Προκειμένου να απαντηθούν τα ερευνητικά ερωτήματα που διατυπώθηκαν, εφαρμόζονται μεθοδολογίες που εκτείνονται από την κλασσική Γραμμική Συσχέτιση του Pearson (Pearson's Correlation), τη Θεωρία της Πληροφορίας (Information Theory), την ανάλυση χρονοσειρών έως τα Μπεϋζιανά Δίκτυα (Bayesian Networks), τα Νευρωνικά Δίκτυα και τη Βαθιά Μηχανική Μάθηση. Η προτεινόμενη μεθοδολογία εφαρμόζεται στο οδικό δίκτυο της κινεζικής πόλης Σιάν (Χί'αν) χρησιμοποιώντας διαδρομές οχημάτων που παρέχονται από την Didi Chuxing Technology Co, μία κινεζική εταιρεία μίσθωσης ταξί και Ι.Χ. αυτοκινήτων.

Πιο συγκεκριμένα:

1. Γίνεται χρήση Συστήματος Γεωγραφικών Πληροφοριών (Geographic Information Systems – GIS) και σχετικού αλγορίθμου έτσι ώστε το στίγμα του GPS των οχημάτων να αντιστοιχιστεί στο οικείο οδικό τμήμα.
2. Υπολογίζεται η μέση ταχύτητα κίνησης σε κάθε οδικό τμήμα ανά μία ώρα.
3. Υπολογίζονται στατιστικά μεγέθη συσχέτισης των παραπάνω χρονοσειρών ανά δύο, ειδικότερα υπολογίζονται η συσχέτιση κατά Pearson, η Αμοιβαία Πληροφορία και η απόσταση του Dynamic Time Warping.
4. Με βάση τα παραπάνω μεγέθη, εντοπίζονται τα οδικά τμήματα που σχετίζονται περισσότερο μεταξύ τους.

- 
5. Αναπτύσσονται Μπεϋζιανό και Δίκτυο τύπου Long-Short Term Memory (LSTM), τα οποία αξιοποιούν τα συμπεράσματα της παραπάνω ανάλυσης στην πρόβλεψη της ταχύτητας σε μελλοντικές χρονικές στιγμές.

### 1.3 Οργάνωση Κειμένου

Η διάρθρωση της υπόλοιπης εργασίας έχει ως εξής: Στο Κεφάλαιο 2 γίνεται παρουσίαση της σχετικής βιβλιογραφίας και επισημαίνονται τα βιβλιογραφικά κενά. Στο Κεφάλαιο 3 δίνεται το θεωρητικό υπόβαθρο των μεθόδων που χρησιμοποιούνται στη συνέχεια. Το Κεφάλαιο 4 περιλαμβάνει την εφαρμογή της προτεινόμενης μεθοδολογίας και την παρουσίαση των αποτελεσμάτων. Το Κεφάλαιο 5 αποτελεί τον επίλογο, όπου αναλύονται τα βασικά συμπεράσματα και γίνονται προτάσεις για περαιτέρω έρευνα.

# Κεφάλαιο 2

## Βιβλιογραφική Ανασκόπηση

### 2.1 Εισαγωγή

Ο τομέας των τεχνικών προβλέψεων εξελίσσεται συνεχώς με γρήγορο ρυθμό ως αποτέλεσμα της συστηματικής ενασχόλησης ερευνητών διαφόρων ειδικοτήτων με αυτόν. Η σπουδαιότητα των εύστοχων προβλέψεων σε τομείς όπως τα χρηματοοικονομικά, η βιομηχανία και οι επιχειρήσεις είναι αδιαπραγμάτευτη. Ιδιαίτερα στην εποχή μας, που χαρακτηρίζεται από τη μαζικότητα των διαθέσιμων δεδομένων, οι προοπτικές του συγκεκριμένου τομέα γίνονται ακόμα περισσότερες. Σε αυτό βεβαίως έχει συμβάλει και η εξέλιξη της υπολογιστικής ισχύος των σύγχρονων υπολογιστών που καθιστούν δυνατή την αξιοποίηση των δεδομένων και την υλοποίηση των νέων μεθοδολογιών με ταχύτητα και ακρίβεια.

Ιδιαίτερα στον τομέα των μεταφορών, ανατρέχοντας στην πιο πρόσφατη βιβλιογραφία, μπορεί κανείς να εντοπίσει την εξέλιξη και τη χρησιμοποίηση συνεχώς πιο εξεζητημένων προσεγγίσεων όσον αφορά την πρόβλεψη κυκλοφοριακών μεγεθών. Εκτός της πολύ μεγάλης σημασίας της ακριβούς πρόβλεψης σε μια σειρά από εφαρμογές, το συγκεκριμένο αντικείμενο αποτελεί και έναν άτυπο διαγωνισμό μεταξύ των ερευνητών ως προς το ποιος θα παρουσιάσει την πιο επαναστατική μεθοδολογία και θα πετύχει την καλύτερη απόδοση.

Όπως παρουσιάζεται αναλυτικότερα στη συνέχεια, το σύνολο των μέχρι τώρα προσεγγίσεων μπορεί να ταξινομηθεί κατ' αρχήν με βάση τη μεθοδολογία που έχει χρησιμοποιηθεί για να παραχθεί η πρόβλεψη, αλλά και με βάση τον τύπο και το πλήθος των οδών από τις οποίες προέρχονται τα δεδομένα καθώς και τον τρόπο που αυτά έχουν συλλεχθεί.

### 2.2 Μεθοδολογίες Πρόβλεψης Κυκλοφοριακών Μεγεθών

Κατά την πρόβλεψη κυκλοφοριακών μεγεθών, ιδιαίτερη σημασία έχει η μεθοδολογία που θα επιλεγεί και το κατά πόσο είναι κατάλληλη για τα συγκεκριμένα μεγέθη και τις ιδιαιτερότητες που παρουσιάζουν. Ο αλγόριθμος πρόβλεψης βρίσκεται στο κέντρο της μεθοδολογίας που ακολουθείται και η επιλογή του είναι ιδιαίτερης σημασίας. Στη συνέχεια αναφέρονται οι δημοφιλέστερες προσεγγίσεις της σχετικής βιβλιογραφίας.

Μιας και τα κυκλοφοριακά μεγέθη που εξετάζονται είναι μεταβαλλόμενα με το χρόνο, συχνά χρησιμοποιούνται μεθοδολογίες ανάλυσης χρονοσειρών. Ευρέως έχει χρησιμοποιηθεί το πρότυπο ARIMA και συνεχίζει να χρησιμοποιείται στις μέρες μας. Το πρότυπο αυτό είναι κατάλληλο για την πρόβλεψη της μελλοντικής εξέλιξης μιας χρονοσειράς βασιζόμενο στις παρελθοντικές τιμές του μεγέθους που αυτή



παριστάνει. Ενδεικτικά έχει χρησιμοποιηθεί για την πρόβλεψη του φόρτου οδικών τμημάτων στις έρευνες [9], [10]. Ακόμα έχουν προταθεί προσεγγίσεις που εφαρμόζουν αποσύνθεση της χρονοσειράς πριν προβούν στην πραγματοποίηση προβλέψεων [11]. Στην ίδια κατηγορία ανήκει και η πρόβλεψη με το μοντέλο Φίλτρου Kalman (Kalman Filter), που είναι μια ακόμα μεθοδολογία πρόβλεψης χρονοσειρών [12].

Οι παραπάνω τεχνικές καθώς και οι παραλλαγές τους, σε αντίθεση με ό, τι επιχειρείται στη συγκεκριμένη διπλωματική εργασία, δε λαμβάνουν υπόψη τις κυκλοφοριακές συνθήκες του υπόλοιπου δικτύου, παρά μόνο του τμήματος για το οποίο γίνονται οι προβλέψεις. Συνήθως σε αυτές τις περιπτώσεις προκύπτουν ακριβείς προβλέψεις όταν η χρονοσειρά του οδικού τμήματος που εξετάζεται παρουσιάζει ένα σταθερό μοτίβο και δεν επηρεάζεται έντονα από τις συνθήκες που επικρατούν συνολικά στο δίκτυο.

Ως εξέλιξη των παραπάνω προσεγγίσεων, έχει προταθεί η χρήση Μπεϋζιανών Δικτύων για την πρόβλεψη των κυκλοφοριακών συνθηκών [13]. Η συγκεκριμένη προσέγγιση παρουσιάζει επιπλέον το πλεονέκτημα της αναγνώρισης συσχετίσεων μεταξύ των κυκλοφοριακών συνθηκών των οδικών τμημάτων, συνδυασμένης με κάποια από τα πλεονεκτήματα της Βαθιάς Μηχανικής Μάθησης.

Παλιότερα έχουν επίσης χρησιμοποιηθεί μοντέλα παλινδρόμησης και προσέγγισης/εκτίμησης συναρτήσεων [14]. Όλα τα παραπάνω μπορεί να θεωρηθεί ότι ανήκουν στην ευρύτερη κατηγορία των προσεγγίσεων που αντιμετωπίζουν το πρόβλημα της πρόβλεψης κυκλοφοριακών συνθηκών ως πρόβλημα πρόβλεψης χρονοσειρών και αξιοποιούν στατιστικά μοντέλα, με ισχυρό μαθηματικό υπόβαθρο.

Η δεύτερη μεγάλη κατηγορία τεχνικών προβλέψεων που χρησιμοποιείται σε μεγάλο βαθμό τα τελευταία χρόνια, εκτός από τα στατιστικά μοντέλα, είναι οι αλγόριθμοι Μηχανικής Μάθησης. Αυτοί οι αλγόριθμοι είναι πιο κατάλληλοι για την αναπαράσταση των σύνθετων εξαρτήσεων που συναντώνται συχνά στα αστικά οδικά δίκτυα [6]. Επιπλέον είναι περισσότερο αποδοτικοί στην αξιοποίηση μεγάλου όγκου δεδομένων [15]. Το βασικό πλεονέκτημα που παρουσιάζουν είναι η ικανότητά τους να προσαρμόζονται σε αναπάντεχα γεγονότα (π.χ. αύξηση κυκλοφοριακής συμφόρησης λόγω ατυχήματος) και να συμπεριλαμβάνουν στις προβλέψεις και εξωγενείς παράγοντες, όπως οι καιρικές συνθήκες. Τα τελευταία είναι σχεδόν αδύνατο να ληφθούν υπόψη σε κλασικά στατιστικά μοντέλα.

Από τις τεχνικές Μηχανικής Μάθησης, περισσότερο δημοφιλή είναι τα Βαθιά Νευρωνικά Δίκτυα. Η χρήση τους επιτρέπει την αξιοποίηση δεδομένων από πολλά οδικά τμήματα του δικτύου, χωρίς να είναι απαραίτητη η οποιαδήποτε παραμετροποίηση από το χρήστη ή γνώση για το μηχανισμό παραγωγής της πρόβλεψης [16].

Τα είδη των Νευρωνικών Δικτύων που έχουν προσελκύσει το μεγαλύτερο ενδιαφέρον είναι τα Συνελικτικά (Convolutional) Δίκτυα, τα οποία αντιμετωπίζουν το οδικό δίκτυο και τις επικρατούσες κυκλοφοριακές συνθήκες ως εικόνα και

αποδεικνύονται ιδιαίτερα αποτελεσματικά και ακριβή στις προβλέψεις τους, όπως ενδεικτικά διαφαίνεται στις έρευνες [17], [18].

Εξίσου δημοφιλής είναι τόσο τα Επαναλαμβανόμενα Νευρωνικά Δίκτυα (Recurrent Neural Networks) [19] όσο και τα Long-Short Term Memory (LSTM) δίκτυα, τα οποία είναι τα πλέον κατάλληλα σε περιπτώσεις όπου τα δεδομένα έχουν και χρονική διάσταση, διότι μεταφέρουν την επιρροή των προηγούμενων παρατηρήσεων σε αυτές που έπονται [20], [21]. Ειδικά τα τελευταία αποτελούν αυτή τη στιγμή μία από τις πλέον δημοφιλής και αποτελεσματικές τεχνικές και χρησιμοποιούνται ευρέως σε διάφορους επιστημονικούς κλάδους.

Τέλος, λιγότερο συχνές είναι προσεγγίσεις που αντιμετωπίζουν την πρόβλεψη κυκλοφοριακών συνθηκών ως πρόβλημα Αναγνώρισης Προτύπων (Pattern Recognition) ή Συσταδοποίησης (Clustering) [6]. Οι πιο εξεζητημένες προσεγγίσεις περιλαμβάνουν υβριδικά μοντέλα πρόβλεψης, τα οποία συνδυάζουν κάποιες από τις παραπάνω τεχνικές, καθώς και παραλλαγές τους [22], [23]. Αυτού του είδους η προσέγγιση, αν και έχει σύνθετη δομή, συνδυάζει κάποια από τα θετικά στοιχεία των Νευρωνικών Δικτύων και των στατιστικών μοντέλων.

Όσον αφορά τις χωροχρονικές συσχετίσεις, ο εντοπισμός των οποίων αποτελεί βασικό στόχο της παρούσας εργασίας, σε προσεγγίσεις με μοντέλα Βαθιάς Μηχανικής Μάθησης υλοποιούνται «εσωτερικά» στα κρυφά επίπεδα των Νευρωνικών Δικτύων και ως εκ τούτου δεν είναι διαθέσιμα στο χρήστη [24]. Προφανώς, μέθοδοι που αυξάνουν την κατανόηση των παραπάνω εξαρτήσεων και την ερμηνεία του τρόπου που προκύπτουν οι προβλέψεις είναι συνήθως προτιμότερες από προσεγγίσεις «μαύρου κουτιού», όπως αποκαλούνται, διότι στις περισσότερες πραγματικές εφαρμογές είναι επιθυμητή η γνώση όχι μόνο για τις συνθήκες που πρόκειται να επικρατήσουν, αλλά και της αιτίας που είναι πιθανό να παρουσιαστούν. Το παραπάνω επιτρέπει αποτελεσματική διαχείριση της κυκλοφορίας, ακόμα και σε περίπτωση αναπάντεχων συμβάντων και ατυχημάτων.

## 2.3 Τεχνολογίες και Μέθοδοι Συλλογής Δεδομένων

Ανατρέχοντας στη βιβλιογραφία μπορεί κανείς να παρατηρήσει ότι προκύπτουν διαφοροποιήσεις ως προς τον τρόπο συλλογής των δεδομένων σε άρθρα που εστιάζουν στην πρόβλεψη κυκλοφοριακών συνθηκών. Τα πρότυπα πρόβλεψης που χρησιμοποιούνται είναι τις περισσότερες φορές πολύ απαιτητικά ως προς το πλήθος και την ποιότητα των απαραίτητων δεδομένων [6], συνεπώς γίνεται φανερό ότι ο τρόπος συλλογής τους αποκτά κεντρικό ρόλο.

Οι έρευνες πρόβλεψης κυκλοφοριακών συνθηκών στοχεύουν στην πρόβλεψη είτε του πλήθους των οχημάτων που πρόκειται να διέλθουν από ένα συγκεκριμένο οδικό τμήμα, είτε στη μέση ταχύτητα κίνησης στο τμήμα αυτό κάποια συγκεκριμένη χρονική περίοδο. Τα δύο παραπάνω μεγέθη είναι δύο από τα τρία βασικά μεγέθη της Κυκλοφοριακής Τεχνικής, ο φόρτος και η ταχύτητα. Αναλόγως και των χαρακτηριστικών του οδικού τμήματος, οι τιμές των δύο μεγεθών μπορούν να

προσδιορίσουν το επίπεδο της στάθμης εξυπηρέτησης, δηλαδή την ποιότητα της ροής των οχημάτων.

Ο συνηθέστερος τρόπος συλλογής δεδομένων κινητικότητας είναι με χρήση αισθητήρων που τοποθετούνται επί των οδών. Οι αισθητήρες αυτοί μετρούν είτε το πλήθος των διερχόμενων οχημάτων σε επιλεγμένο χρονικό διάστημα (φόρτος), είτε την ταχύτητα διέλευσής τους. Είναι τοποθετημένοι επί του οδοστρώματος μίας ή περισσότερων οδών ή στο πλάι, δίπλα στο κράσπεδο. Επιπλέον μπορεί να υπάρχει ένας μόνο αισθητήρας ανά οδικό τμήμα ή περισσότεροι κατά μήκος του τμήματος, ανάλογα με την εφαρμογή στην οποία πρόκειται να χρησιμοποιηθούν τα δεδομένα και την επιθυμητή λεπτομέρεια των μετρήσεων [25].

Ο σημαντικότερος περιορισμός αυτού του τρόπου συλλογής δεδομένων είναι το κόστος αγοράς και εγκατάστασης των αισθητήρων. Λόγω αυτού, είναι συχνά αδύνατη η τοποθέτηση του απαραίτητου πλήθους αισθητήρων, με αποτέλεσμα τα δεδομένα που συλλέγονται να μην έχουν την απαιτούμενη ακρίβεια [25]. Ακόμα, για τον ίδιο λόγο, είναι συχνά αδύνατη και η τοποθέτηση αισθητήρων σε όλες τις οδούς ενός δικτύου, περιορίζοντας έτσι την περιοχή μελέτης.

Στην ίδια κατηγορία μπορούν να ταξινομηθούν και μέθοδοι συλλογής δεδομένων που βασίζονται στην επεξεργασία των εικόνων από κάμερες κυκλοφορίας που είναι ήδη εγκατεστημένες σε πολλά σημεία του οδικού δικτύου [26]. Η ανάπτυξη της τεχνολογίας των ασύρματων επικοινωνιών έχει ενισχύσει τις δύο αυτές μεθόδους, καθώς τα δεδομένα που συλλέγονται είναι δυνατόν να γίνουν διαθέσιμα σε πραγματικό χρόνο [6]. Ωστόσο, λόγω των περιορισμών που αναφέρθηκαν παραπάνω, πολλές φορές η συλλογή των δεδομένων γίνεται άπαξ, με αποτέλεσμα τα πρότυπα που αναπτύσσονται να μην ενημερώνονται με νέα δεδομένα.

Από την άλλη μεριά, τα τελευταία χρόνια η ανάπτυξη της τεχνολογίας έχει διευκολύνει και επιταχύνει τη συλλογή μεγάλου όγκου δεδομένων και μάλιστα μεγάλης ακρίβειας και σε πραγματικό χρόνο. Στις έρευνες πρόβλεψης κυκλοφοριακών συνθηκών είναι όλο και συχνότερη η χρήση δεδομένων που προέρχονται από τα συστήματα GPS των οχημάτων ή των έξυπνων κινητών τηλεφώνων των οδηγών ή των συστημάτων τηλεματικής των οχημάτων. Σε αυτή την περίπτωση είναι σημαντικό να έχουν παραχωρηθεί αυτά με τη σύμφωνη γνώμη των χρηστών, διότι σε αντίθετη περίπτωση ανακύπτουν πολλά νομικά ζητήματα [26].

Ο τρόπος αυτός υπερτερεί έναντι των υπόλοιπων, διότι κάνει δυνατή τη συλλογή μεγαλύτερου όγκου δεδομένων, τα δεδομένα που συλλέγονται έχουν μεγαλύτερη λεπτομέρεια και καλύπτουν συχνά το σύνολο του οδικού δικτύου. Επιπλέον η συλλογή τους είναι σχεδόν δωρεάν και δεν απαιτείται οποιαδήποτε υλικοτεχνική υποδομή [26].

Στην παρούσα εργασία αξιοποιούνται δεδομένα που έχουν συλλεγεί με τον τελευταίο τρόπο και έχουν γίνει διαθέσιμα από την εταιρεία Didi Chuxing Technology Co για ερευνητικούς σκοπούς.

## 2.4 Επιρροή Τύπου Οδού

Οι δημοσιευμένες έρευνες της βιβλιογραφίας σχετικά με την πρόβλεψη των κυκλοφοριακών συνθηκών και την ανίχνευση χωροχρονικών συσχετίσεων σε οδικά δίκτυα διακρίνονται ανάλογα με τον τύπο του οδικού δικτύου που μελετάται σε δύο κατηγορίες: αυτές που έχουν πραγματοποιηθεί σε υπεραστικές οδούς ή δίκτυο και μεγάλους αυτοκινητοδρόμους και αυτές που έχουν πραγματοποιηθεί σε οδούς αστικού οδικού δικτύου.

Στην πρώτη περίπτωση, εμφανίζονται, κατά κύριο λόγο, πιο ισχυρά μοτίβα στις εξεταζόμενες χρονοσειρές και η πρόβλεψη των μελλοντικών συνθηκών γίνεται ευκολότερη. Τα οχήματα γενικά κινούνται με υψηλές ταχύτητες και σπάνια εμφανίζονται συνθήκες συμφόρησης. Επίσης δεν υπάρχει συνήθως αλληλεπίδραση μεταξύ των οχημάτων, αλλά ούτε και με το υπόλοιπο οδικό δίκτυο. Σε τέτοιες μελέτες, είναι αρκετά συνηθισμένη η επίτευξη υψηλής ακρίβειας στις προβλέψεις με τις υπάρχουσες μεθοδολογίες [27], [28].

Από την άλλη μεριά, όταν οι προβλέψεις πρόκειται να γίνουν για κάποιο ή κάποια οδικά τμήματα ενός αστικού οδικού δικτύου, η αβεβαιότητα αυξάνεται λόγω της μεγαλύτερης ζήτησης σε συνδυασμό με τη μικρότερη ικανότητα εξυπηρέτησης των οδών ενός αστικού δικτύου σε σχέση με αυτές ενός υπεραστικού δικτύου. Επιπλέον, οι καιρικές συνθήκες και άλλα αναπάντεχα συμβάντα (π.χ. ατυχήματα) έχουν ισχυρότερη επίδραση στις οδούς ενός αστικού δικτύου, επίδραση που δεν είναι εύκολο να ληφθεί υπόψη από ένα πρότυπο πρόβλεψης [29]. Τέλος, η μελέτη ενός αστικού οδικού δικτύου είναι εν γένει πολυπλοκότερη και λόγω της ύπαρξης ισχυρότερης αλληλεπίδρασης μεταξύ των οχημάτων (π.χ. στους κόμβους), αλλά και λόγω της ύπαρξης σήμανσης και σηματοδότησης που ρυθμίζει την κυκλοφορία [6].

Στην παρούσα διπλωματική εργασία, το οδικό δίκτυο που εξετάζεται είναι αστικό και βρίσκεται στην πόλη Σιάν της Κίνας.

## 2.5 Βασικά Συμπεράσματα

Το αντικείμενο της πρόβλεψης των κυκλοφοριακών συνθηκών είναι ένα από τα δημοφιλέστερα του αντίστοιχου επιστημονικού πεδίου. Μπορεί κανείς να παρατηρήσει ότι οι πιο δημοφιλείς μεθοδολογίες Μηχανικής Μάθησης έχουν εισαχθεί εδώ και πολλά χρόνια, ενώ και οι νεότερες που παρουσιάζονται εισάγονται μέσα σε πολύ σύντομο χρονικό διάστημα. Ακόμα, έχουν αναπτυχθεί και πιο σύνθετα πρότυπα που συνδυάζουν τις παραπάνω μεθοδολογίες με τις κλασικές στατιστικές τεχνικές, με αποτέλεσμα να προκύπτει μεγάλη ακρίβεια στις προβλέψεις.

Πρόσφατα έχουν εισαχθεί και νέες τεχνικές συλλογής δεδομένων με χρήση έξυπνων κινητών τηλεφώνων και συστημάτων που βρίσκονται ενσωματωμένα στα οχήματα, μεταφέροντας το ενδιαφέρον και προς τους αλγορίθμους συγχώνευσης και επεξεργασίας δεδομένων ευρείας κλίμακας. Το γεγονός αυτό έχει ανοίξει νέους ορίζοντες στην πρόβλεψη κυκλοφοριακών συνθηκών μιας και η διαθεσιμότητα και

---

η ποιότητα των δεδομένων που είναι απαραίτητα για την ανάπτυξη των προτύπων είναι μεγαλύτερη από ποτέ.

Γενικά, λοιπόν, το πεδίο της πρόβλεψης των κυκλοφοριακών συνθηκών βρίσκεται αυτή τη στιγμή στο επίκεντρο και αναμένεται να αναπτυχθεί κι άλλο, λόγω του μεγάλου ενδιαφέροντος των ερευνητών συνεπικουρούμενο από την ανάπτυξη της τεχνολογίας.

Παρά τη συστηματική ενασχόληση των ερευνητών με την πρόβλεψη κυκλοφοριακών συνθηκών και την ανάπτυξη όλο και πιο εξειδικευμένων προτύπων, το κομμάτι της αναγνώρισης των χωροχρονικών συσχετίσεων περνάει σε δεύτερη μοίρα. Το κομμάτι αυτό είναι και ένα από τα βασικά αντικείμενα της παρούσας εργασίας, η παρουσίαση δηλαδή των χωρικών και χρονικών εξαρτήσεων και η σύνδεσή τους με την πρόβλεψη των κυκλοφοριακών συνθηκών.

Τα δεδομένα που χρησιμοποιούνται στις προβλέψεις έχουν οπωσδήποτε χωρική και χρονική συσχέτιση, οπότε ακόμα και όταν δε γίνεται προσπάθεια εντοπισμού της σίγουρα υλοποιείται εντός των μοντέλων πρόβλεψης. Επιπλέον, σε προσεγγίσεις όπου το οδικό δίκτυο αντιμετωπίζεται ως γράφος ή ως εικόνα γίνεται ακόμα εντονότερη η αξιοποίηση των χωρικών συσχετίσεων, ενώ σε προσεγγίσεις πρόβλεψης με μεθοδολογίες χρονοσειρών ή Επαναλαμβανόμενα Δίκτυα αυτή των χρονικών. Παρ' όλα αυτά, και σε εκείνες τις περιπτώσεις δεν τονίζεται η ύπαρξη των συσχετίσεων αυτών από τους ερευνητές.

Φαίνεται λοιπόν πως η επίτευξη μεγαλύτερης ακρίβειας στην πρόβλεψη και η εισαγωγή όλο και πιο εξεζητημένων προτύπων πρόβλεψης είναι πολλές φορές αυτοσκοπός για τους ερευνητές. Ωστόσο, η γνώση των χωροχρονικών συσχετίσεων και των δυναμικών σχέσεων που διέπουν τη λειτουργία του δικτύου είναι σε πολλές πραγματικές εφαρμογές και στη χάραξη πολιτικών περισσότερο χρήσιμη από την ίδια την πρόβλεψη. Συνεπώς στην παρούσα διπλωματική εργασία επιχειρείται η ανάδειξη της σπουδαιότητας των παραπάνω εξαρτίσεων καθώς και η παρουσίαση μιας απλής μεθοδολογίας για τον υπολογισμό τους και την εισαγωγή τους στη διαδικασία της πρόβλεψης.

# Κεφάλαιο 3

## Μεθοδολογική Προσέγγιση

### 3.1 Προτεινόμενη Μεθοδολογία

Στην πρόβλεψη κυκλοφοριακών συνθηκών, υπάρχουν δύο βασικές διαφορετικές προσεγγίσεις: η πρόβλεψη σε επίπεδο οδού ή οδικού τμήματος και η πρόβλεψη σε επίπεδο δικτύου. Πρέπει να τονιστεί ότι η διάκριση αυτή είναι πιο σημαντική από αυτές που παρουσιάστηκαν στο προηγούμενο κεφάλαιο. Μάλιστα, αυτή η επιλογή προσέγγισης καθορίζει και όλες αυτές που παρουσιάστηκαν στο προηγούμενο κεφάλαιο [6].

Πρόβλεψη σε επίπεδο δικτύου σημαίνει πρόβλεψη ενός κυκλοφοριακού μεγέθους, π.χ. της ταχύτητας, ενός ή περισσότερων οδικών τμημάτων λαμβάνοντας υπόψη δεδομένα και μετρήσεις από ένα μέρος ή το σύνολο του οδικού δικτύου στο οποίο ανήκουν. Αντίθετα, στην πρόβλεψη σε επίπεδο οδού ή οδικού τμήματος, αξιοποιούνται δεδομένα που προέρχονται μόνο από αυτό και φυσικά η πρόβλεψη αφορά επίσης το ίδιο τμήμα.

Η παραπάνω διάκριση καθορίζει και τον αλγόριθμο πρόβλεψης που πρόκειται να χρησιμοποιηθεί. Στην πρόβλεψη σε επίπεδο οδού ή οδικού τμήματος, εφαρμόζονται τεχνικές πρόβλεψης μεμονωμένων χρονοσειρών, είτε αυτές είναι στατιστικές είτε Μηχανικής Μάθησης. Από την άλλη μεριά, στην πρόβλεψη σε επίπεδο δικτύου χρησιμοποιούνται περισσότερο τεχνικές Μηχανικής Μάθησης, διότι καθιστούν ευκολότερη την εισαγωγή μεγαλύτερου πλήθους δεδομένων από διαφορετικές πηγές του οδικού δικτύου.

Επίσης, στην πρόβλεψη σε επίπεδο οδού είναι πιο συνηθισμένο, αλλά όχι απαραίτητο, η συλλογή δεδομένων να γίνεται από αισθητήρα ή αισθητήρες που έχουν τοποθετηθεί στην οδό ενδιαφέροντος [26]. Αντίθετα, στην πρόβλεψη σε επίπεδο δικτύου είναι πιο οικονομικό, αλλά και πιο αποτελεσματικό, η συλλογή δεδομένων να γίνεται μέσω των έξυπνων κινητών τηλεφώνων των οδηγών ή των συστημάτων των οχημάτων. Σε αντίθετη περίπτωση θα πρέπει να τοποθετηθεί ένας μεγάλος αριθμός αισθητήρων σε αρκετές οδούς του οδικού δικτύου.

Τέλος, η μεθοδολογία της πρόβλεψης σε επίπεδο δικτύου είναι δυνατό να εφαρμοστεί και έχει νόημα κυρίως σε αστικό οδικό δίκτυο, ενώ αυτή της πρόβλεψης σε επίπεδο οδού μπορεί να εφαρμοστεί εξίσου σε αστικές και υπεραστικές οδούς ή σε μεγάλους αυτοκινητοδρόμους [6].

Επιπλέον, η προσέγγιση της πρόβλεψης σε επίπεδο δικτύου προσφέρει το πλεονέκτημα ότι τα αποτελέσματά της μπορούν να αξιοποιηθούν σε μια πληθώρα εφαρμογών. Για παράδειγμα, επιτρέπουν την πρόβλεψη του χρόνου μετάβασης από

ένα σημείο του δικτύου σε ένα άλλο και την αποστολή ενημερωτικών μηνυμάτων προς τους χρήστες του δικτύου. Ακόμα, δίνουν τη δυνατότητα στους διαχειριστές του δικτύου να έχουν μια σφαιρική εικόνα της λειτουργίας του και να μπορούν έτσι να καταστρώσουν ένα γενικό σχέδιο διαχείρισης της κυκλοφορίας και να ορίσουν τις αντίστοιχες πολιτικές και εναλλακτικές επιλογές σε περίπτωση αναπάντεχων γεγονότων και συνθηκών.

Στην παρούσα εργασία γίνεται χρήση της προσέγγισης της πρόβλεψης της κυκλοφορίας σε επίπεδο δικτύου, διότι βασικός σκοπός της είναι και η ανάδειξη των χωρικών και χρονικών συσχετίσεων των κυκλοφοριακών μεγεθών, κάτι που έχει νόημα μόνο σε αυτή την προσέγγιση.

## 3.2 Θεωρητικό υπόβαθρο

Σε αυτή την ενότητα παρουσιάζεται το θεωρητικό υπόβαθρο των τεχνικών και μεθοδολογιών που χρησιμοποιούνται στη συνέχεια και έχουν ήδη αναφερθεί στο Κεφάλαιο 1.

### 3.2.1 Γραμμική Συσχέτιση και Αμοιβαία Πληροφορία

Για να εντοπιστούν οδικά τμήματα που παρουσιάζουν υψηλή συσχέτιση όσον αφορά την ταχύτητα κίνησης των οχημάτων στο καθένα, εφαρμόζεται η μεθοδολογία υπολογισμού της Αμοιβαίας Πληροφορίας (Mutual Information - MI) και τα αποτελέσματα συγκρίνονται με την κλασσική μεθοδολογία ανάλυσης συσχετίσεων του Pearson (Pearson's Correlation Analysis). Οι δύο παραπάνω μεθοδολογίες είναι οι πιο συχνά χρησιμοποιούμενες και μάλιστα σε δεδομένα που δεν είναι απαραίτητα χρονοσειρές. Όπως και στην παρούσα εργασία, είναι μέρος της διαδικασίας επιλογής των χαρακτηριστικών που θα χρησιμοποιηθούν στην πρόβλεψη (Feature Selection).

Σύμφωνα με τη Θεωρία της Πληροφορίας, η Αμοιβαία Πληροφορία δύο τυχαίων μεταβλητών είναι μία μετρική που ποσοτικοποιεί το μέγεθος της πληροφορίας που λαμβάνεται για την πρώτη τυχαία μεταβλητή, παρατηρώντας τη δεύτερη. Σε αντίθεση με την κλασσική ανάλυση συσχετίσεων, η αμοιβαία πληροφορία λαμβάνει υπόψη και μη γραμμικές συσχετίσεις, καθώς το μέγεθος που υπολογίζεται δε σχετίζεται με τους γραμμικούς ή μη γραμμικούς κανόνες εξέλιξης των δύο ποσοτήτων που εμπλέκονται, αλλά με την εντροπία κατά Shannon (Shannon Entropy) [30], [31].

Έστω  $x_n$  και  $y_n$  δύο τυχαίες μεταβλητές που ορίζονται στο ίδιο διάστημα με συναρτήσεις πυκνότητας πιθανότητας  $p(x_n)$  και  $p(y_n)$  και από κοινού συνάρτηση πυκνότητας πιθανότητας  $p(x_n, y_n)$ . Η Αμοιβαία Πληροφορία  $I(x_n, y_n)$ , η οποία ποσοτικοποιεί την αναμενόμενη πληροφορία που λαμβάνεται για τη  $x_n$  κατά την παρατήρηση της  $y_n$  δίνεται από την εξίσωση:

$$I(x_n, y_n) = - \sum_{x_n, y_n} p(x_n, y_n) \log_2 \frac{p(x_n, y_n)}{p(x_n)p(y_n)} \quad (1)$$

Αυτή η προσέγγιση έχει δύο κύριους περιορισμούς, όταν χρησιμοποιείται στη μελέτη χρονοσειρών: πρώτον, είναι στατική, με την έννοια ότι ο χρόνος δεν υπεισέρχεται στην ανάλυση των συσχετίσεων των ταχυτήτων ανάμεσα σε διαφορετικά τμήματα του οδικού δικτύου. Με άλλα λόγια, η χρονική διαδοχή των διαφορετικών τιμών της ταχύτητας των τμημάτων δε λαμβάνεται υπόψη, αλλά μόνο η στατιστική κατανομή των ταχυτήτων. Για να ξεπεραστεί αυτό το εμπόδιο πρέπει να χρησιμοποιούνται χρονοσειρές με ίδιο και σταθερό βήμα. Δεύτερον, οι συσχετίσεις ορίζονται μόνο μεταξύ ζευγαριών οδικών τμημάτων, μην επιτρέποντας τη διερεύνηση του κατά πόσο δεδομένα από περισσότερα τμήματα μπορούν να αλληλοεπιδράσουν μεταξύ τους και να επηρεάσουν την πρόβλεψη της ταχύτητας του ζητούμενου τμήματος.

### 3.2.2 Dynamic Time Warping

Το Dynamic Time Warping (DTW) είναι μία μεθοδολογία εκτίμησης της ομοιότητας δύο χρονοσειρών που βασίζεται στην (ευκλείδεια) απόστασή τους. Πρόκειται για μία τεχνική δυναμικού προγραμματισμού για την εύρεση μίας βέλτιστης μέσης χρονοσειράς ανάμεσα σε δύο δεδομένες χρονοσειρές, η οποία έχει την ελάχιστη απόσταση από τις άλλες δύο [32]. Η χρονοσειρά που προκύπτει ονομάζεται *warping path*.

Για δύο χρονοσειρές  $X = x_1, x_2, \dots, x_n$  και  $Y = y_1, y_2, \dots, y_n$ , υπολογίζεται η μεταξύ τους απόσταση από τον παρακάτω επαναληπτικό τύπο του πίνακα  $\gamma(n, n)$  του δυναμικού προγραμματισμού [33]:

$$\gamma(i, j) = \text{dist}(x_i, y_j) + \min [\gamma(i - 1, j - 1), \gamma(i - 1, j), \gamma(i, j - 1)] \quad (2)$$

Η απόσταση αυτή ονομάζεται απόσταση DTW και ισούται με την ευκλείδεια απόσταση κατά μήκος του *warping path*. Ο υπολογισμός της απόστασης DTW έχει τετραγωνική πολυπλοκότητα ως προς το χρόνο, κάτι που περιορίζει την εφαρμοσιμότητα της μεθόδου σε βάσεις δεδομένων με μικρές χρονοσειρές.

Για να ξεπεραστεί ο παραπάνω περιορισμός, μπορεί να εφαρμοστεί μία παραλλαγή της κλασσικής μεθόδου, σύμφωνα με την οποία πραγματοποιείται αρχικά μετασχηματισμός των χρονοσειρών μεγάλων διαστάσεων σε μικρότερων και στη συνέχεια υπολογίζεται η απόσταση στις χρονοσειρές μικρών διαστάσεων. Αυτή η παραλλαγή, γνωστή ως Fast DTW, λειτουργεί σε τρία βήματα [34]: μείωση της διαστατικότητας των χρονοσειρών, υπολογισμός της απόστασης DTW στις χρονοσειρές μειωμένης ανάλυσης, διόρθωση και προβολή του *warping path* στην αρχική ανάλυση.

Αξίζει να σημειωθεί ότι, σε αντίθεση με τις μεθοδολογίες που παρουσιάστηκαν νωρίτερα, το DTW αφορά αποκλειστικά χρονοσειρές και λαμβάνει υπόψη του και τη χρονική διάσταση των δεδομένων. Ωστόσο είναι πολύ ευαίσθητη στις τιμές της χρονοσειράς (στο μέτρο της ταχύτητας στη προκειμένη περίπτωση), ενώ η επιρροή των διακυμάνσεων και των μεταβολών, δηλαδή του μοτίβου της χρονοσειράς είναι μικρότερη. Για το λόγο αυτό χρησιμοποιείται περισσότερο στην αντιστοίχιση ίδιων χρονοσειρών μεταξύ τους, π.χ. σε εφαρμογές τηλεπικοινωνιών.



### 3.2.3 Ταξινομητής Μπεϋζιανού Δικτύου

Για να ξεπεραστεί ο περιορισμός της σύγκρισης των χρονοσειρών μόνο κατά ζεύγη, προσέγγιση που ακολουθούν όλες οι παραπάνω μεθοδολογίες, αναπτύσσεται στην παρούσα διπλωματική και ένα Μπεϋζιανό Δίκτυο, το οποίο αναπαριστά τις συσχετίσεις μεταξύ όλων των οδικών τμημάτων και βασίζεται στον υπολογισμό της δεσμευμένης πιθανότητας των κατανομών των ταχυτήτων τους.

Ένα Μπεϋζιανό Δίκτυο (Bayesian Network – BN) είναι ένας κατευθυνόμενος ακυκλικός γράφος του οποίου οι κόμβοι αναπαριστούν μεταβλητές. Τα βάρη των ακμών που συνδέουν δύο κόμβους είναι ανάλογα της συσχέτισης μεταξύ των μεταβλητών των αντίστοιχων κόμβων. Με το παραπάνω μοντέλο, είναι δυνατόν να υπολογιστεί η δεσμευμένη πιθανότητα να πάρει μια συγκεκριμένη τιμή μία μεταβλητή, όταν είναι γνωστές οι τιμές των υπόλοιπων μεταβλητών που συνδέονται με αυτή, δηλαδή είναι κόμβοι – παιδιά της [35].

Πιο αναλυτικά, έστω ένα Μπεϋζιανό δίκτυο με κόμβους τις μεταβλητές  $X_i = \{X_1, \dots, X_n\}$  και ακμές το σύνολο κατανομών δεσμευμένων πιθανοτήτων  $P_i = \{P_1, \dots, P_n\}$  που αντιστοιχούν σε κάθε κόμβο και στους κόμβους – γονείς του. Η συνηθέστερη υλοποίηση ενός Μπεϋζιανού Δικτύου είναι ως εξής: η κατευθυνόμενη ακμή από έναν κόμβο  $X_1$  σε έναν άλλο κόμβο  $X_2$  αναπαριστά την παραδοχή ότι η τιμή της μεταβλητής  $X_1$  είναι άμεση συνέπεια της τιμής της μεταβλητής  $X_2$  ( $X_1|X_2$ ), όπως βέβαια και των άλλων μεταβλητών του κατευθυνόμενου ακυκλικού γράφου [36]. Η από κοινού κατανομή  $P$  μπορεί να αναπαρασταθεί ως το γινόμενο των δεσμευμένων πιθανοτήτων κάθε κόμβου με τους κόμβους – γονείς του, όπως φαίνεται στην παρακάτω εξίσωση:

$$P(X) = \prod_{i=1}^n P_i(X_i | pa_i) \quad (3),$$

όπου  $pa_i$  είναι το σύνολο των κόμβων – γονέων του  $X_i$ .

Το Μπεϋζιανό Δίκτυο μπορεί να χρησιμοποιηθεί και ως ταξινομητής του  $X_i$ , αν το τελευταίο ανήκει σε κλάσεις  $C$ , στην προκειμένη περίπτωση κλάσεις ταχυτήτων κίνησης, σύμφωνα με τον παρακάτω κανόνα [36]:

$$classify(x_1, \dots, x_n) = \arg \max_n p(R) \prod_{i=1}^n p(X_i = x_i | C) \quad (4)$$

Χρησιμοποιώντας την παραπάνω τεχνική, η επίδραση κάθε μεταβλητής (στη συγκεκριμένη περίπτωση οι κυκλοφοριακές συνθήκες των τμημάτων του οδικού δικτύου) μπορεί να εκτιμηθεί ανάλογα με την προκύπτουσα κλάση  $C$ . Ο εντοπισμός των πλέον επιδραστικών χωροχρονικών μοτίβων ταχυτήτων βασίζεται στο κριτήριο της Αμοιβαίας Πληροφορίας. Όπως έχει ήδη αναφερθεί, η Αμοιβαία Πληροφορία ποσοτικοποιεί τη ροή πληροφορίας ανάμεσα σε ένα κόμβο  $X$  και τα γνωστά επίπεδα ταχύτητας  $C$ . Η αμοιβαία πληροφορία  $I(X, C)$  ανάμεσα σε μία μεταβλητή  $X$  και μία κλάση  $C$  αντιπροσωπεύει την αναμενόμενη πληροφορία που

μπορεί να αποκτηθεί για τη  $C$ , παρατηρώντας τη μεταβλητή  $X$  και δίνεται από τον παρακάτω τύπο:

$$I(X, C) = \sum_{x,c} P(X|C)P(C) \log \frac{P(X|C)}{\sum_{c \in C} P(X|C)P(c)} \quad (5)$$

### 3.2.4 Νευρωνικό Δίκτυο Μακράς και Βραχείας Μνήμης

Τα Νευρωνικά Δίκτυα Μακράς και Βραχείας Μνήμης (Long Short Memory Neural Networks) ανήκουν στην ευρύτερη κατηγορία των Επαναλαμβανόμενων (Recurrent) Νευρωνικών Δικτύων. Όπως και τα παραδοσιακά Νευρωνικά Δίκτυα, χρησιμοποιούνται κατά κύριο λόγο στην πρόβλεψη συνεχών τιμών και στην ταξινόμηση σε κλάσεις. Σε αντίθεση όμως με τα τελευταία, τα δίκτυα που ανήκουν στην κατηγορία των Επαναλαμβανόμενων Δικτύων περιλαμβάνουν ένα μηχανισμό ανατροφοδότησης ο οποίος καθιστά δυνατή την αξιοποίηση πληροφοριών που αφορούν προγενέστερες χρονικές στιγμές κατά την επεξεργασία δεδομένων που αφορούν μεταγενέστερες χρονικές στιγμές, σε περίπτωση βέβαια που τα δεδομένα είναι σε μορφή χρονοσειράς [37].

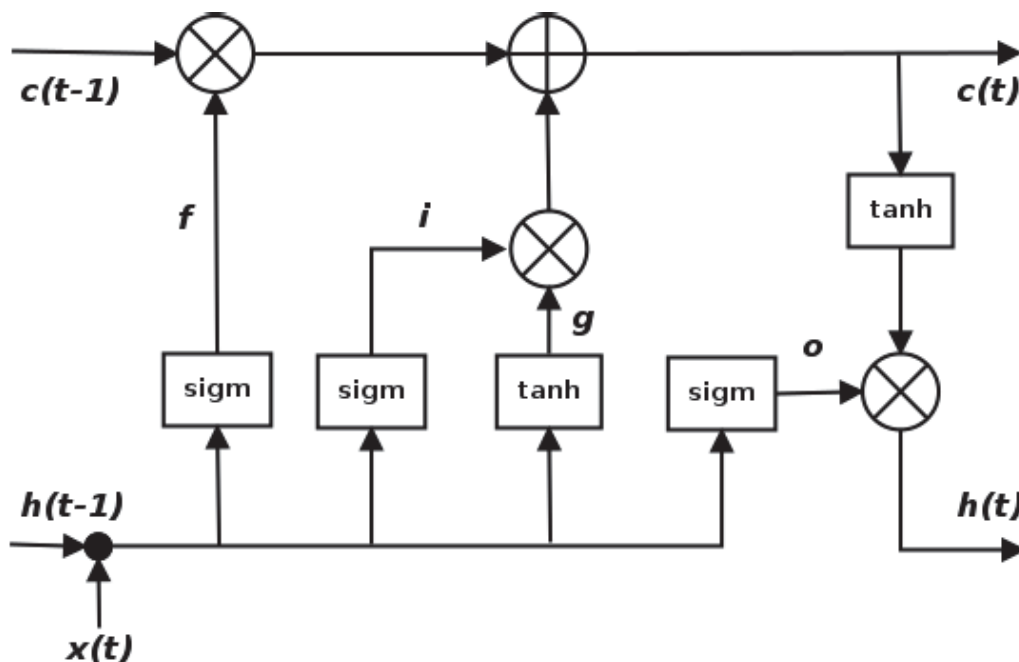
Όπως είναι γνωστό, στα απλά Νευρωνικά Δίκτυα δεν υπάρχει αυτή η δυνατότητα, αν και μπορούν να χρησιμοποιηθούν για την πρόβλεψη χρονοσειρών, χωρίς όμως το μοντέλο που προκύπτει να «αντιλαμβάνεται» και να αξιοποιεί τη χρονική διαδοχή των δεδομένων. Όπως είναι εύκολα κατανοητό, στη μεγάλη πλειοψηφία των περιπτώσεων, αυτή η χρονική διαδοχή έχει μεγάλη επίδραση στην ακρίβεια των προβλέψεων, με αποτελέσματα να χρησιμοποιούνται σχεδόν αποκλειστικά τα Επαναλαμβανόμενα Δίκτυα στην πρόβλεψη χρονοσειρών [38], [39]. Πράγματι, τα δίκτυα αυτής της κατηγορίας παρουσιάζουν πολύ καλή απόδοση και σε εργασίες όπως αναγνώριση φωνής, αναγνώριση εικόνας, μοντελοποίηση μιας γλώσσας, μετάφραση κ.α. [39].

Τα Επαναλαμβανόμενα Δίκτυα, έχουν έναν απλό μηχανισμό ανατροφοδότησης, ο οποίος κατά την επεξεργασία δεδομένων μιας χρονικής στιγμής μεταφέρει σε ένα νευρώνα πληροφορία από τα δεδομένα της αμέσως προηγούμενης χρονικής στιγμής και μόνο [38]. Αυτή η απλή πρακτική σε πολλές περιπτώσεις δεν επαρκεί, μιας και είναι πιθανό ένα φαινόμενο να επηρεάζεται περισσότερο από γεγονότα που συνέβησαν δύο ή περισσότερες χρονικές περιόδους πριν. Για παράδειγμα, η επίδραση της συμφόρησης σε ένα οδικό τμήμα να εμφανιστεί σε ένα άλλο, που απέχει αρκετά από το πρώτο, μετά από μεγάλο χρονικό διάστημα.

Φυσικά έχουν αναπτυχθεί αρκετές παραλλαγές του αρχικού προτύπου με σκοπό να αντιμετωπιστεί ο παραπάνω περιορισμός. Μία από τις πιο γνωστές και ευρέως χρησιμοποιούμενες είναι τα LSTM δίκτυα. Στα δίκτυα αυτά είναι δυνατή η μεταφορά της επίδρασης παρελθοντικών παρατηρήσεων κατά μήκος ακόμα και ολόκληρης της χρονοσειράς [38]. Όπως άλλωστε και όλα τα Επαναλαμβανόμενα Δίκτυα, τα LSTM αποτελούνται από μια αλυσίδα συνδεδεμένων κελιών που με τη σειρά τους περιέχουν κάποια μορφή νευρωνικών δικτύων. Στην πιο συνηθισμένη εκδοχή, η βασική ιδέα της υλοποίησης των LSTM δικτύων είναι η ύπαρξη μιας γραμμής

«μνήμης» που διατρέχει όλο το μήκος των διαδοχικών κελιών μεταφέροντας την πληροφορία από τα προηγούμενα στα επόμενα και αλληλοεπιδρώντας με αυτά εμπλουτίζοντας ή απλώς τροποποιώντας το περιεχόμενο της μνήμης [39].

Πιο συγκεκριμένα, σε κάθε κελί αρχικά προσδιορίζεται η πληροφορία που πρόκειται να αφαιρεθεί από τη μνήμη χρησιμοποιώντας ένα νευρωνικό δίκτυο ενός επιπέδου με σιγμοειδή ενεργοποίηση. Στη συνέχεια αποφασίζεται το ποσό της πληροφορίας αυτού του κελιού που πρόκειται να προστεθεί στη μνήμη χρησιμοποιώντας ένα επίπεδο σιγμοειδούς και ένα υπερβολικής εφασπτομένης. Τέλος, υπολογίζεται η έξοδος του κελιού συνδυάζοντας τα στοιχεία της νέας μνήμης με την είσοδο του κελιού και δύο επίπεδα σιγμοειδούς και υπερβολικής εφασπτομένης [39]. Αναλυτικότερα φαίνονται στην παρακάτω Εικόνα 1:



Εικόνα 1: Δομή κελιού LSTM, πηγή [38]

Οι ακριβείς υπολογισμοί γίνονται με χρήση των παρακάτω σχέσεων [38]:

$$i = \sigma(W_i h_{t-1} + U_i x_t) \quad (6)$$

$$f = \sigma(W_f h_{t-1} + U_f x_t) \quad (7)$$

$$o = \sigma(W_o h_{t-1} + U_o x_t) \quad (8)$$

$$g = \tanh(W_g h_{t-1} + U_g x_t) \quad (9)$$

$$c_t = (c_{t-1} * f) + (g * i) \quad (10)$$

$$h_t = \tanh(c_t) * o \quad (11)$$

# Κεφάλαιο 4

## Εφαρμογή Μεθοδολογίας

### 4.1 Περιγραφή και Προεπεξεργασία Δεδομένων

Η προεπεξεργασία των δεδομένων είναι μία πάρα πολύ σημαντική διαδικασία κατά την πραγματοποίηση στατιστικής ανάλυσης ή την εφαρμογή τεχνικών Μηχανικής Μάθησης. Καλά προετοιμασμένα δεδομένα εισόδου οδηγούν σε πρότυπα πρόβλεψης με καλύτερη απόδοση και ευκολότερη και γρηγορότερη εκπαίδευση.

Τα δεδομένα που χρησιμοποιούνται σε αυτή τη διπλωματική αποτελούνται από περίπου 3,2 εκατομμύρια διαδρομές οχημάτων της εταιρείας Didi στο οδικό δίκτυο της πόλης Σιάν κατά την περίοδο από 2 έως 30 Νοεμβρίου 2016. Η Σιάν είναι η 12<sup>η</sup> μεγαλύτερη πόλη της Κίνας με πληθυσμό πάνω από 6,5 εκατομμύρια κατοίκους και βρίσκεται στα κεντρικά της χώρας. Στην Εικόνα 2 που ακολουθεί φαίνεται η θέση της Σιάν στο χάρτη.



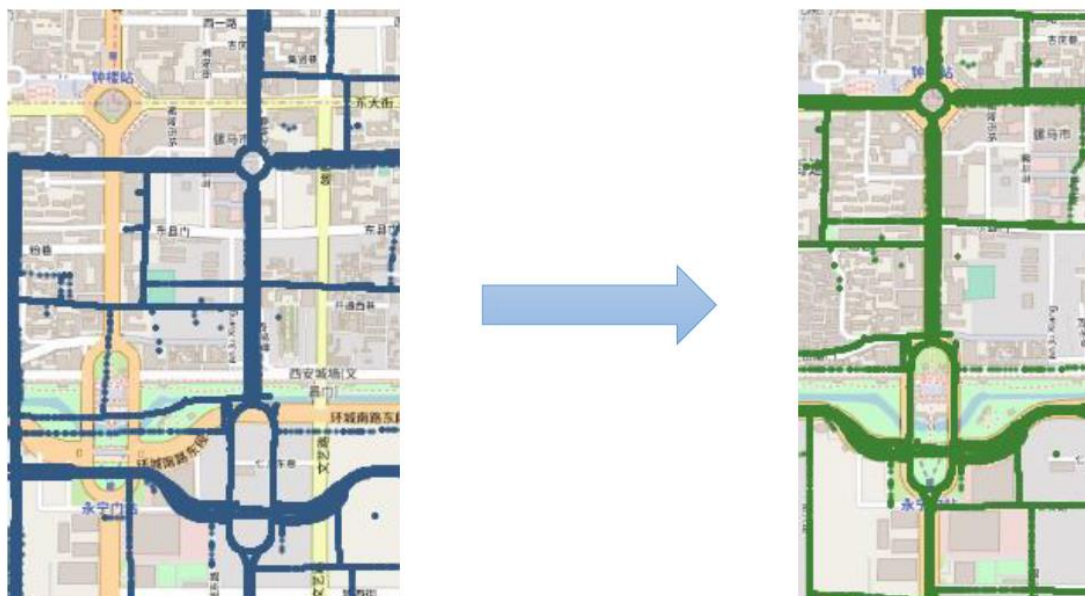
Εικόνα 2: Θέση της Σιάν στο χάρτη, πηγή Bing Maps

Η κάθε διαδρομή περιλαμβάνει την ακριβή τοποθεσία του οχήματος κάθε 2 έως 4 δευτερόλεπτα, που αντιστοιχεί σε περίπου 250 σημεία ανά διαδρομή, δηλαδή συνολικά περίπου 750 εκατομμύρια σημεία. Αναλυτικότερα, τα γνωρίσματα των δεδομένων είναι το γεωγραφικό πλάτος και το γεωγραφικό μήκος κάθε σημείου της διαδρομής, ένας μοναδικός κωδικός που προσδιορίζει τη διαδρομή, ένας δεύτερος

μοναδικός αριθμός που προσδιορίζει τον οδηγό και η χρονοσήμανση (timestamp), δηλαδή ένας αριθμός που προσδιορίζει την ακριβή χρονική στιγμή και ημερομηνία που το όχημα βρισκόταν στη συγκεκριμένη θέση.

Όπως έχει ήδη αναφερθεί, τα δεδομένα έχουν παραχωρηθεί από την εταιρεία για ερευνητικούς σκοπούς. Διαπιστώθηκε κατά την απεικόνισή τους σε χάρτη ότι εκτός από το μεγάλο όγκο τους, τα δεδομένα έχουν και μεγάλη ακρίβεια. Ωστόσο, για να μπορούν να αξιοποιηθούν απαιτήθηκε να ακολουθηθεί η διαδικασία που περιγράφεται στη συνέχεια. Να τονιστεί ότι, λόγω του μεγάλου όγκου των δεδομένων, η επεξεργασία τους απαιτεί ισχυρό υπολογιστικό σύστημα και πολύ χρόνο.

Αρχικά οι συντεταγμένες των σημείων μετατράπηκαν από το κινέζικο σύστημα συντεταγμένων (Chinese State Bureau of Surveying and Mapping coordinate system – GCJ-02) στο παγκόσμιο και ευρέως χρησιμοποιούμενο WGS 84 (World Geodetic System). Με αυτό τον τρόπο έγινε εφικτή η σωστή απεικόνιση των σημείων στις πιο δημοφιλείς διαδικτυακές υπηρεσίες χαρτών, όπως είναι το Open Street Maps, και η περαιτέρω επεξεργασία τους. Η μετατροπή έγινε χρησιμοποιώντας τη βιβλιοθήκη «eviltransform» της γλώσσας προγραμματισμού Python. Στην Εικόνα 3 φαίνεται η διόρθωση των συντεταγμένων ώστε να απεικονίζονται σωστά στο χάρτη.



Εικόνα 3: Ενδεικτική μετατροπή συντεταγμένων σε παγκόσμιο σύστημα

Στη συνέχεια εφαρμόζεται αντιστοίχιση των σημείων της βάσης δεδομένων στο χάρτη (map matching) χρησιμοποιώντας αλγόριθμο Κοντινότερου Γείτονα (Nearest Neighbor) [40]. Γενικά, map matching ονομάζεται η διαδικασία αντιστοίχισης καταγεγραμμένων ιχνών GPS στις ακμές του πραγματικού οδικού δικτύου, ενώ ταυτόχρονα πραγματοποιείται και η αντίστοιχη διόρθωση των συντεταγμένων που

οφείλεται σε σφάλμα του συστήματος κατά τη διάρκεια της καταγραφής. Στη συγκεκριμένη περίπτωση, κάθε ένα από τα καταγεγραμμένα σημεία αντιστοιχίζεται σε ένα οδικό τμήμα του οδικού δικτύου της Σιάν, όπως αυτό μεταφορτώθηκε από το Open Street Map. Η παραπάνω διαδικασία πραγματοποιείται σε QGIS, το οποίο είναι ένα λογισμικό επεξεργασίας γεωγραφικών δεδομένων ανοιχτού κώδικα.

Η διαδικασία περιλαμβάνει τη φόρτωση του δικτύου της Σιάν στο QGIS καθώς και των σημείων που πρέπει να αντιστοιχιστούν σε δύο διαφορετικά επίπεδα (layers) και την εκτέλεση στη συνέχεια του αλγορίθμου. Η πολυπλοκότητα του αλγορίθμου είναι  $n^2$ , όπου  $n$  το πλήθος των σημείων και  $m$  το πλήθος των οδικών τμημάτων.

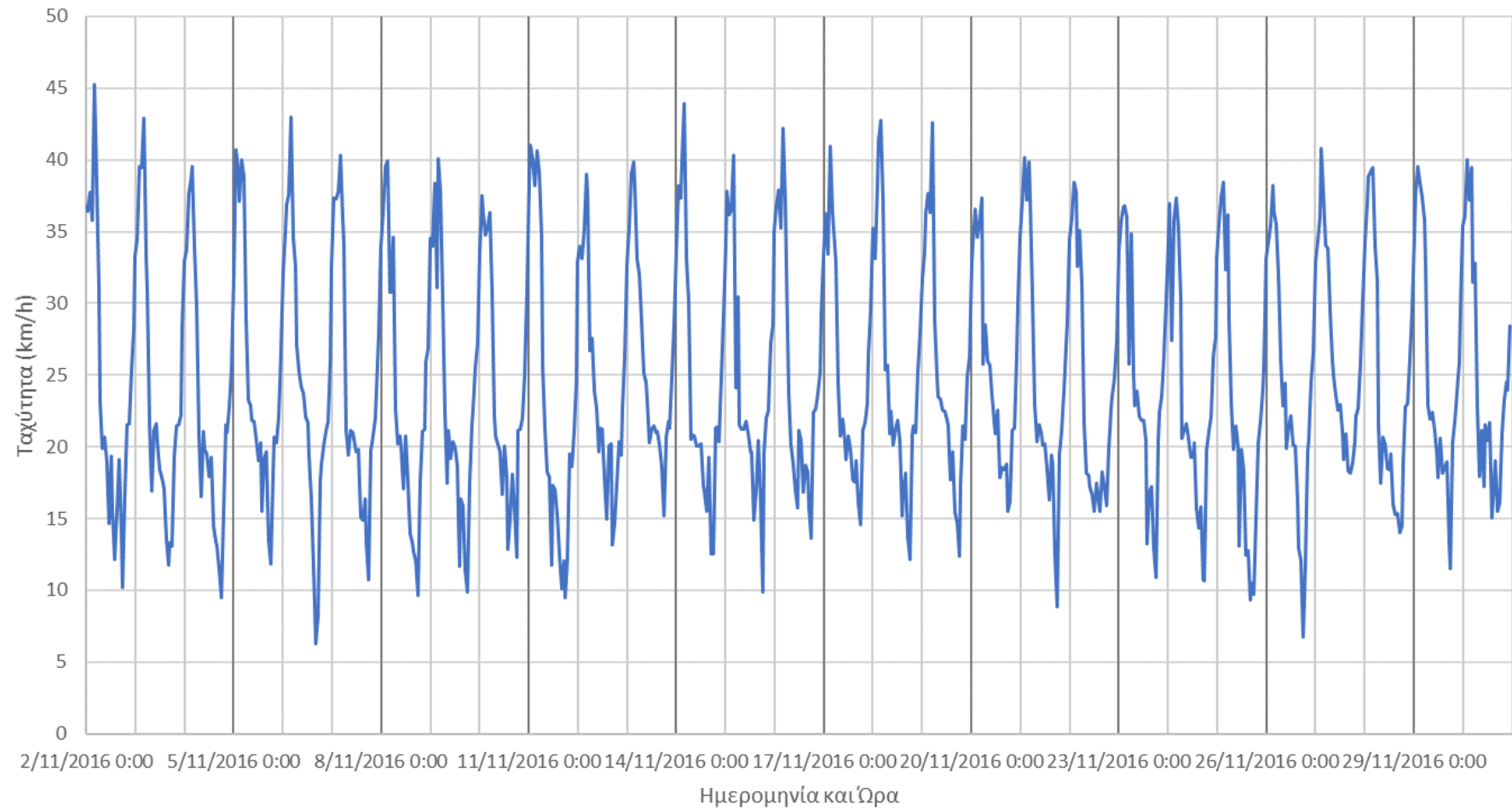
Στο τέλος αυτής της διαδικασίας η βάση δεδομένων έχει εμπλουτιστεί με ένα ακόμα πεδίο που αντιστοιχεί στον αριθμό που προσδιορίζει μοναδικά το οδικό τμήμα στο οποίο ανήκει το αντίστοιχο σημείο, όπως φαίνεται και στην Εικόνα 4, καθώς και με την περιγραφή του τύπου της οδού.

lat	lon	orderid	timestamp	join_osm_id	join_highway
34.20992214	108.906614	b08770fabd024e77497cd1e	1475500123	334648176	primary
34.20992213	108.906624	8c989417e0215661984567	1475472636	334648176	primary
34.21355086	108.9066737	5721bfa5ea988e33309d87	1475457628	334648176	primary
34.21347089	108.9066737	59ff40f07bbb2ccf2ac36fb7	1475466584	148975613	secondary
34.21347088	108.9066837	785e3183d0c7d44fc3448ac	1475453648	148975613	secondary
34.21355085	108.9067036	b08770fabd024e77497cd1e	1475499910	334648176	primary
34.21347088	108.9067036	8bbb73bbef1d02830cdda8	1475493861	148975613	secondary

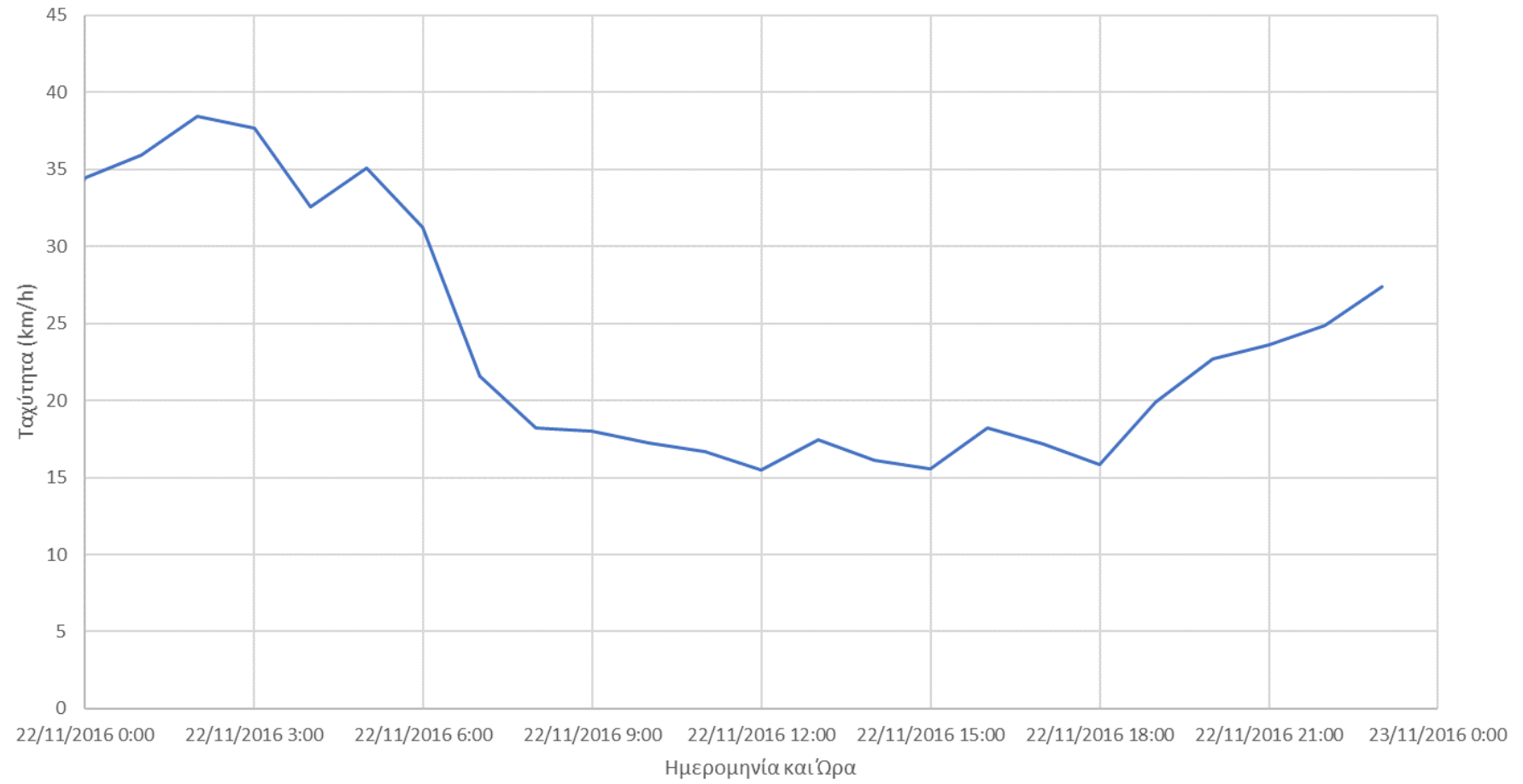
Εικόνα 4: Προσθήκη στοιχείων προσδιορισμού οδικού τμήματος

Τέλος, υπολογίζεται η ταχύτητα κίνησης του αντίστοιχου οχήματος από το κάθε σημείο κάθε διαδρομής στο επόμενο του ως το πηλίκο της ευκλείδειας απόστασής τους προς τη διαφορά της χρονοσήμανσής τους. Οι χρονοσειρές της ταχύτητας των τμημάτων του οδικού δικτύου προκύπτουν με ομαδοποίηση των δεδομένων κατά το μοναδικό αριθμό του τμήματος και στη συνέχεια κατά το επιλεγμένο χρονικό βήμα και υπολογισμό της μέσης τιμής των ταχυτήτων. Για τις απαιτήσεις της συγκεκριμένης διπλωματικής εργασίας, επιλέχθηκε χρονικό βήμα μίας ώρας. Η τιμή της ταχύτητας για κάθε χρονικό βήμα της χρονοσειράς υπολογίζεται ως ο μέσος όρος των ταχυτήτων όλων των οχημάτων που διέρχονται από το οδικό τμήμα, τη συγκεκριμένη χρονική περίοδο. Στα Διαγράμματα 1 και 2 που ακολουθούν παρατίθενται ενδεικτικά κάποιες από τις χρονοσειρές που προέκυψαν.





Διάγραμμα 1: Ενδεικτική χρονοσειρά τμήματος.



Διάγραμμα 2: Ενδεικτική χρονοσειρά μίας ημέρας



Η χρονοσειρά που ενδεικτικά παρουσιάστηκε, αλλά και οι υπόλοιπες σε μικρότερο ή μεγαλύτερο βαθμό, παρατηρήθηκε ότι παρουσιάζουν ένα αρκετά ισχυρό και διακριτό ημερήσιο μοτίβο. Αυτό αποτελεί μια ένδειξη ότι είναι πράγματι δυνατή η πρόβλεψη των κυκλοφοριακών συνθηκών και μάλιστα αναμένεται και υψηλή ακρίβεια. Παρατηρήθηκε ακόμα ότι υπάρχουν διακριτά μοτίβα και εντός της ημέρας, τα οποία επίσης πρέπει να αξιολογηθούν.

Αξίζει να σημειωθεί ότι τα οδικά τμήματα που δεν είχαν καμία εγγραφή για περισσότερο από μία ώρα, οποιαδήποτε από τις ημέρες που μελετήθηκαν, αποκλείστηκαν από την περαιτέρω ανάλυση, καθώς θεωρήθηκε πως δε διαδραματίζουν σημαντικό ρόλο στο οδικό δίκτυο της Σιάν. Τα οδικά τμήματα που τελικά αξιοποιήθηκαν είναι 283 και αφορούν αποκλειστικά κύριες και δευτερεύουσες οδούς, καθώς και τις συνδέσεις τους.

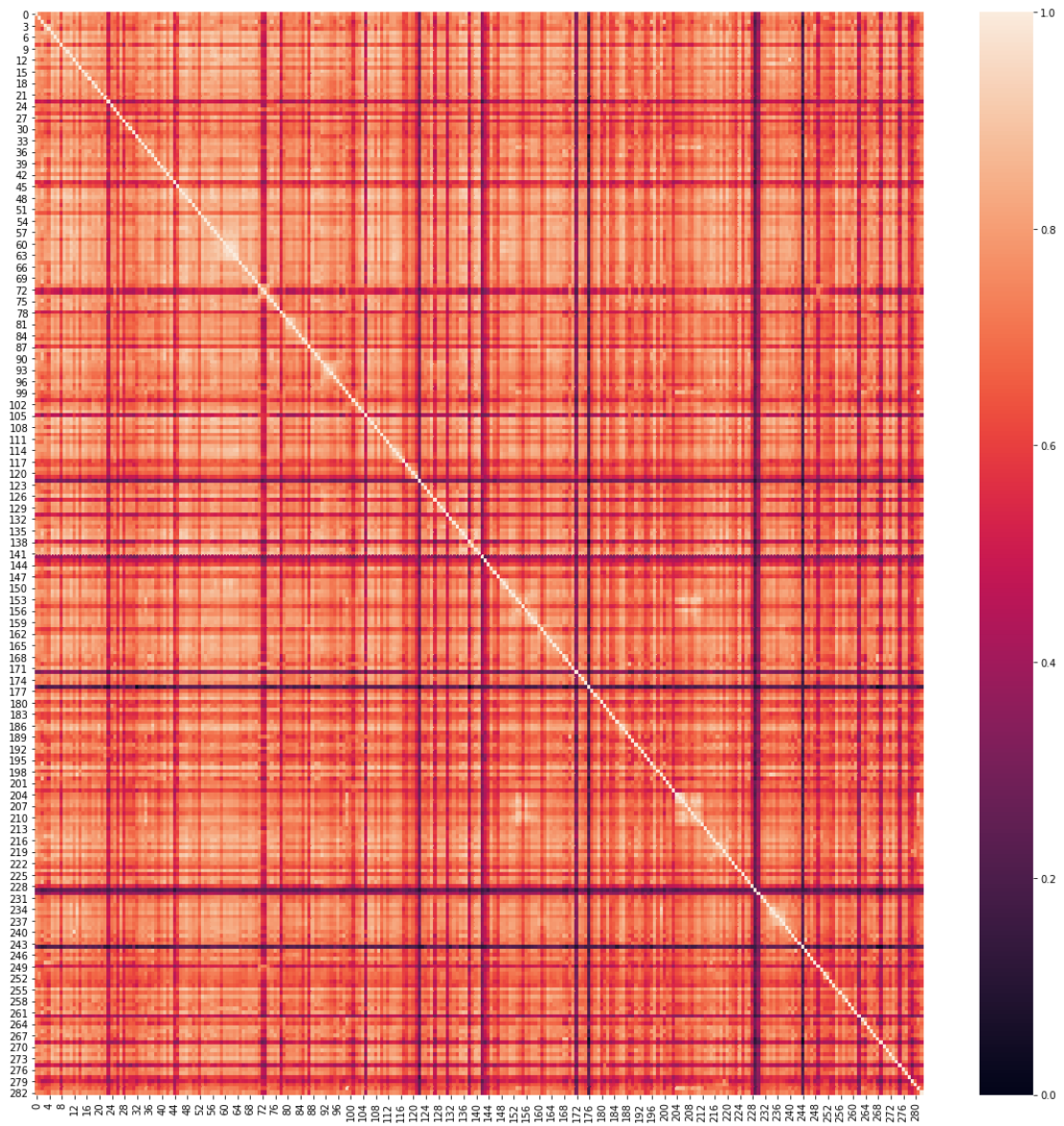
## 4.2 Εφαρμογή και Αποτελέσματα

### 4.2.1 Υπάρχουν χωρικές και χρονικές συσχετίσεις σε ένα οδικό δίκτυο;

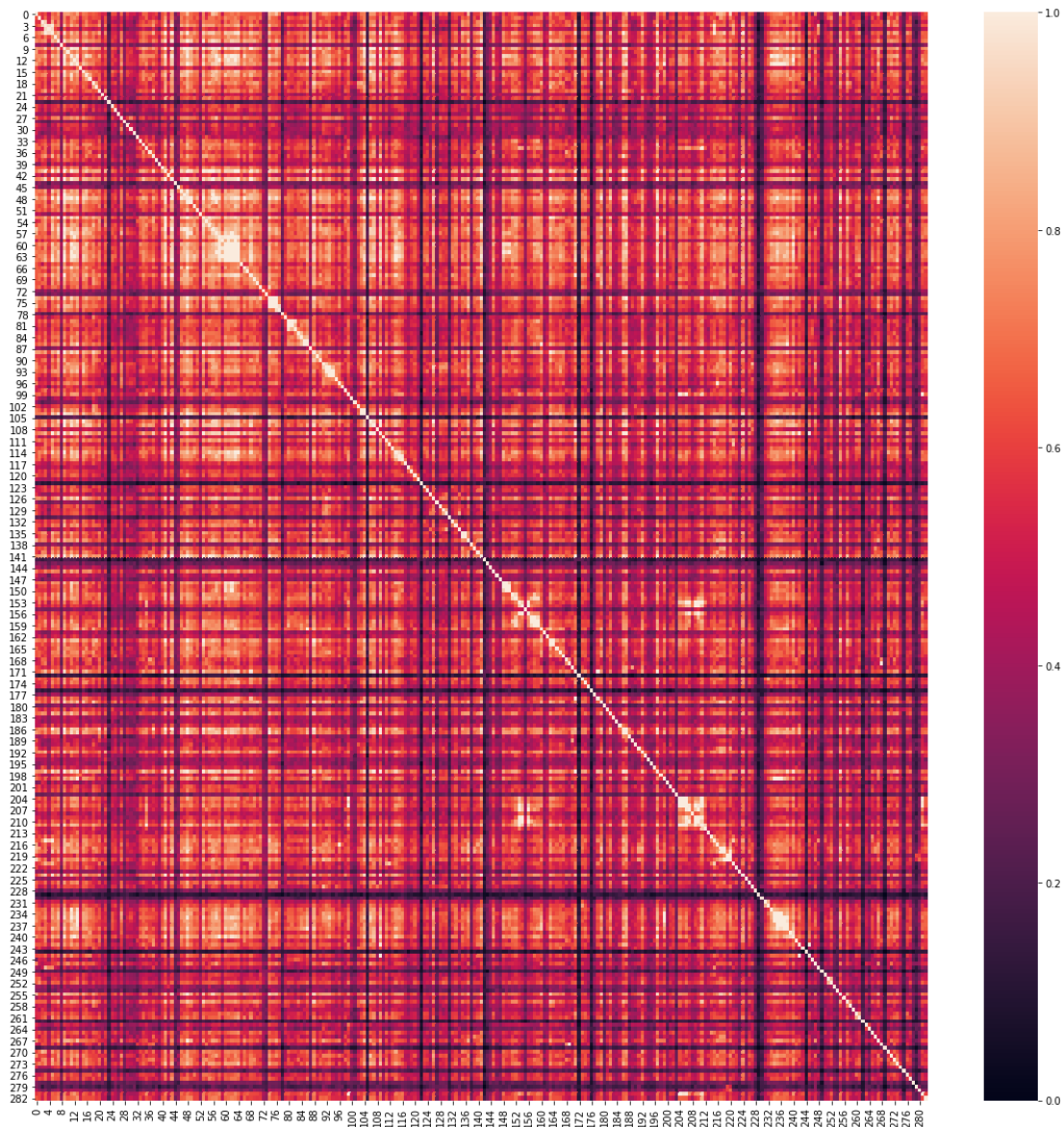
Για τον προσδιορισμό των χωροχρονικών συσχετίσεων, χρησιμοποιήθηκαν τόσο η συσχέτιση κατά Pearson, όσο και η Αμοιβαία Πληροφορία και ο υπολογισμός της απόστασης DTW. Θεωρήθηκε ότι, αν προκύψουν υψηλές τιμές των τριών μετρικών, μπορεί να συμπεράνει κανείς ότι πράγματι υπάρχουν τμήματα με κοινή συμπεριφορά.

Υπενθυμίζεται ότι με τις δύο πρώτες μεθοδολογίες που εφαρμόζονται σε αυτή την ενότητα δε λαμβάνεται υπόψη η χρονική διάσταση του προβλήματος, αλλά κατά κύριο λόγο οι πιθανοτικές κατανομές των ταχυτήτων των τμημάτων. Για το λόγο αυτό έχουν χρησιμοποιηθεί χρονοσειρές ίδιου χρονικού βήματος. Επίσης, δύο τμήματα που σχετίζονται με βάση αυτές τις μεθοδολογίες θα έχουν παραπλήσιες τιμές ταχυτήτων κατά τη διάρκεια της ημέρας, όμως όχι απαραίτητα πολύ κοντινές.

Τα δύο μεγέθη υπολογίστηκαν για όλα τα ζεύγη χρονοσειρών και παρουσιάζονται στους θερμικούς χάρτες που παρουσιάζονται στα Διαγράμματα 3 και 4 που ακολουθούν. Στους δύο θερμικούς χάρτες, όσο πιο ανοιχτή είναι η απόχρωση, τόσο ισχυρότερη είναι και η συσχέτιση των αντίστοιχων τμημάτων. Εξετάζοντας κάθε στήλη (ή γραμμή) ξεχωριστά, μπορεί κανείς να εντοπίσει ποια τμήματα έχουν τη μεγαλύτερη συσχέτιση με αυτό της στήλης.



Διάγραμμα 3: Θερμικός χάρτης συσχέτισης Pearson



Διάγραμμα 4: Θερμικός χάρτης Αμοιβαίας Πληροφορίας

Όπως μπορεί κανείς να παρατηρήσει, στους δύο χάρτες υπάρχουν σίγουρα κάποια κοινά μοτίβα, ωστόσο παρουσιάζουν και αρκετές διαφορές, παρ' όλο που τα δύο μεγέθη εκφράζουν σχεδόν το ίδιο πράγμα. Επίσης, ο θερμικός χάρτης της Αμοιβαίας Πληροφορίας έχει περισσότερο σκούρο χρώμα συνολικά, κάτι που υποδεικνύει λιγότερο ισχυρές συσχετίσεις. Υπενθυμίζεται ότι με τη συσχέτιση κατά Pearson υπολογίζεται η γραμμική συσχέτιση δύο μεταβλητών, οπότε η Αμοιβαία Πληροφορία μπορεί να θεωρηθεί περισσότερο γενική μέθοδος, μιας και λαμβάνει υπόψη και μη γραμμικές συσχετίσεις.

Οι δύο παραπάνω χάρτες περιλαμβάνουν μεγάλο όγκο πληροφορίας, ωστόσο είναι δύσκολο να διαβαστούν λόγω του μεγέθους τους. Για το λόγο αυτό και για να είναι δυνατή η σύγκριση των δύο μετρικών μεταξύ τους, αλλά και με αυτές που

παρουσιάζονται στη συνέχεια, επιλέχθηκε να παρουσιαστεί με περισσότερη λεπτομέρεια το τμήμα με κωδικό 28258922 στο Open Street Map. Το συγκεκριμένο οδικό τμήμα είναι ένα από τα πιο πολυσύχναστα και βρίσκεται στο κέντρο της πόλης Σιάν. Η ακριβής θέση του φαίνεται παρακάτω στην Εικόνα 5.

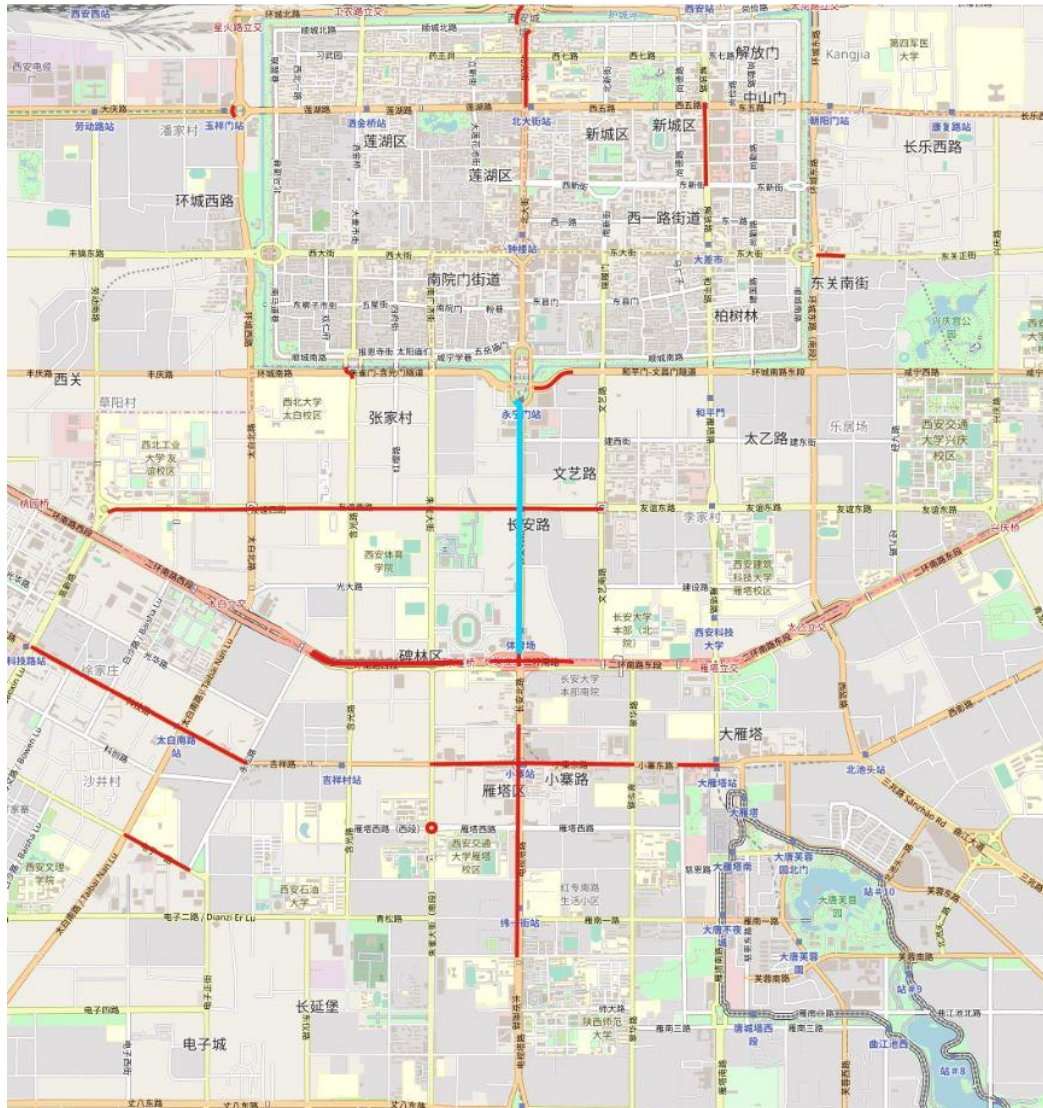


Εικόνα 5: Απεικόνιση επιλεγμένου τμήματος στο χάρτη

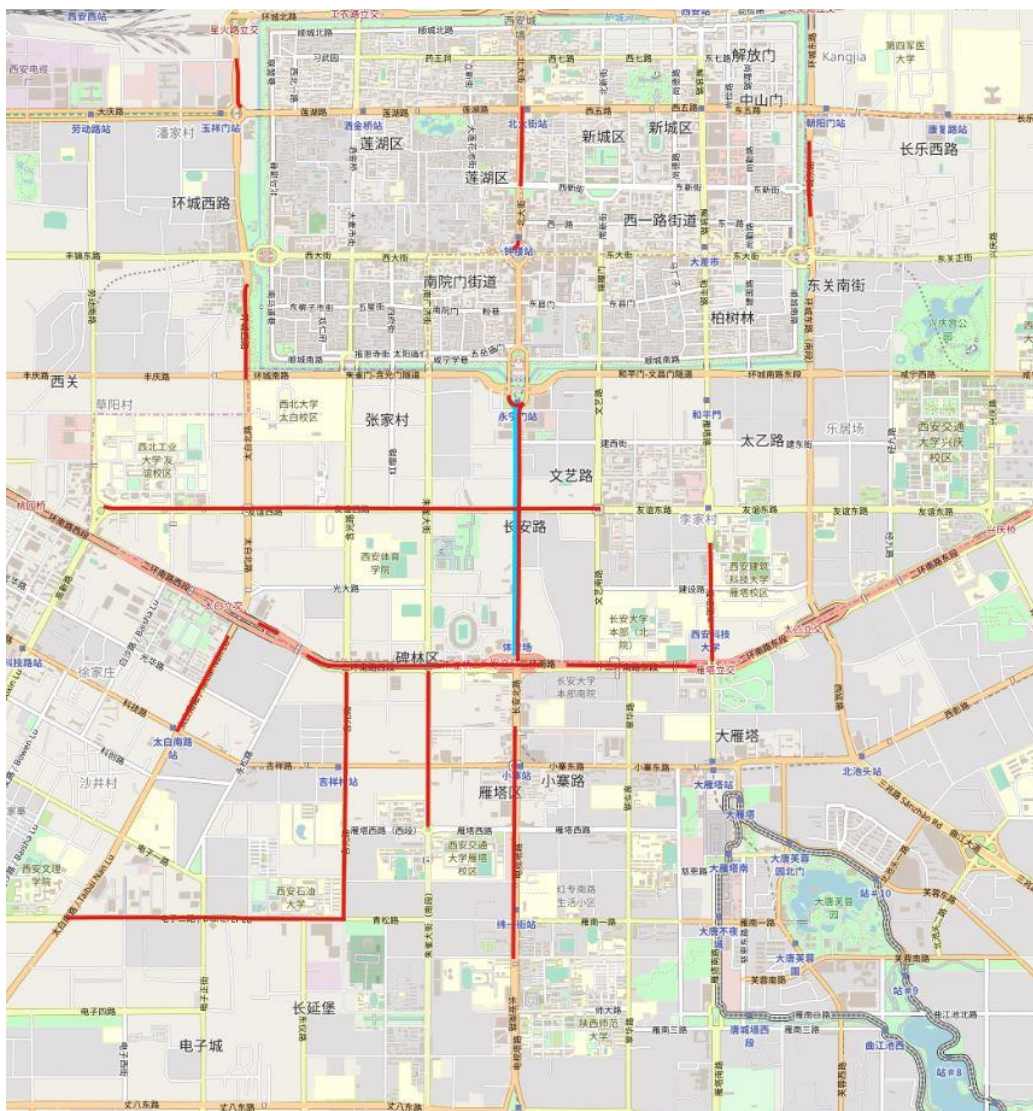
Στην Εικόνα 6 παρουσιάζονται τα 20 οδικά τμήματα με την μεγαλύτερη συσχέτιση με το επιλεγμένο, σύμφωνα με τον υπολογισμό της συσχέτισης Pearson και στην Εικόνα 7 τα 20 οδικά τμήματα με την μεγαλύτερη συσχέτιση με το επιλεγμένο, σύμφωνα με τον υπολογισμό της Αμοιβαίας Πληροφορίας. Όπως προκύπτει, οι δύο προσεγγίσεις καταγράφουν διαφορετικά μοτίβα χωρικών συσχετίσεων στην ίδια βάση δεδομένων, η επίδραση των οποίων θα αξιολογηθεί περαιτέρω στη συνέχεια σε σχέση με την ακρίβεια πρόβλεψης. Ωστόσο, φαίνεται πως υπάρχουν πράγματι πολύ ισχυρές χωρικές συσχετίσεις.

Προφανώς αντίστοιχοι χάρτες μπορούν να κατασκευαστούν για όλα τα τμήματα του οδικού δικτύου και να αξιοποιηθούν για την πρόβλεψη των κυκλοφοριακών τους συνθηκών. Για παράδειγμα, όταν παρουσιαστεί ένα συμβάν σε ένα οδικό τμήμα, με χρήση ενός τέτοιου χάρτη μπορεί να προβλεφθεί ποια άλλα τμήματα θα επηρεαστούν πιο άμεσα.





Εικόνα 6: Χάρτης των 20 τμημάτων με την υψηλότερη συσχέτιση κατά Pearson ως προς το επιλεγμένο τμήμα.



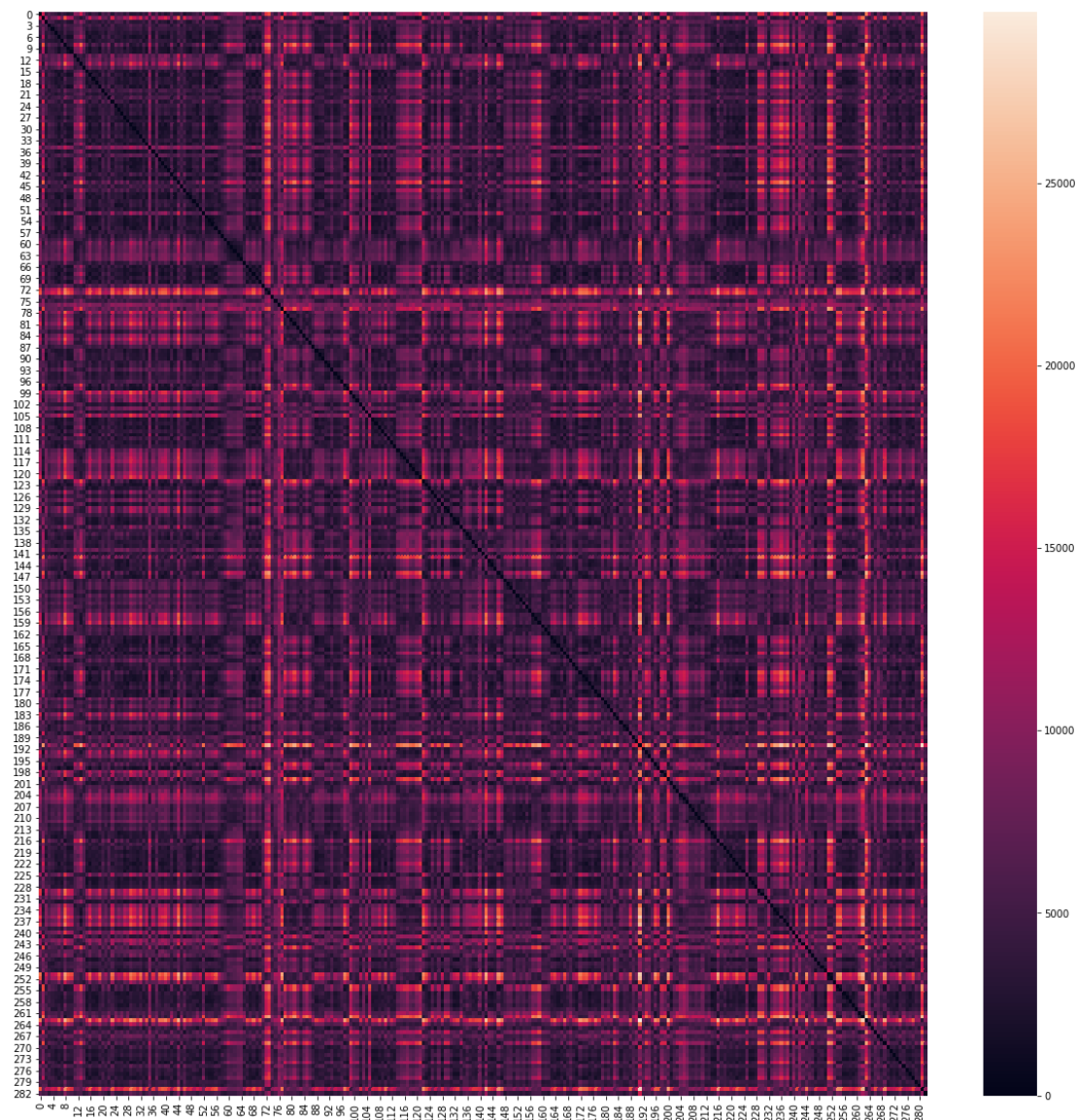
Εικόνα 7: Χάρτης των 20 τμημάτων με την υψηλότερη Αμοιβαία Πληροφορία ως προς το επιλεγμένο τμήμα.

Για την περαιτέρω διερεύνηση των χωρικών και των χρονικών συσχετίσεων ανάμεσα στις ταχύτητες των διαφορετικών οδικών τμημάτων, εφαρμόστηκε ο αλγόριθμος Fast DTW στις διαθέσιμες χρονοσειρές. Ομοίως με τα προηγούμενα μεγέθη, η απόσταση που υπολογίζεται από τον αλγόριθμο Fast DTW αποτελεί μία ένδειξη του ποιων οδικών τμημάτων οι ταχύτητες σχετίζονται μεταξύ τους. Παρ' όλα αυτά, αυτός ο αλγόριθμος λαμβάνει υπόψη ότι τα δεδομένα που επεξεργάζεται είναι διαδοχικά, δηλαδή χρονοσειρές.

Πιο συγκεκριμένα, αντιστοιχίζει χρονοσειρές οι οποίες αφενός παρουσιάζουν κοντινές τιμές ταχυτήτων και αφετέρου παρουσιάζουν παρόμοια συμπεριφορά και ως προς τις διακυμάνσεις κατά τη διάρκεια της ημέρας, με την επιρροή των ακριβών τιμών να είναι εδώ πιο ισχυρή. Με αυτό τον τρόπο λαμβάνεται υπόψη και η χρονική διάσταση του προβλήματος.



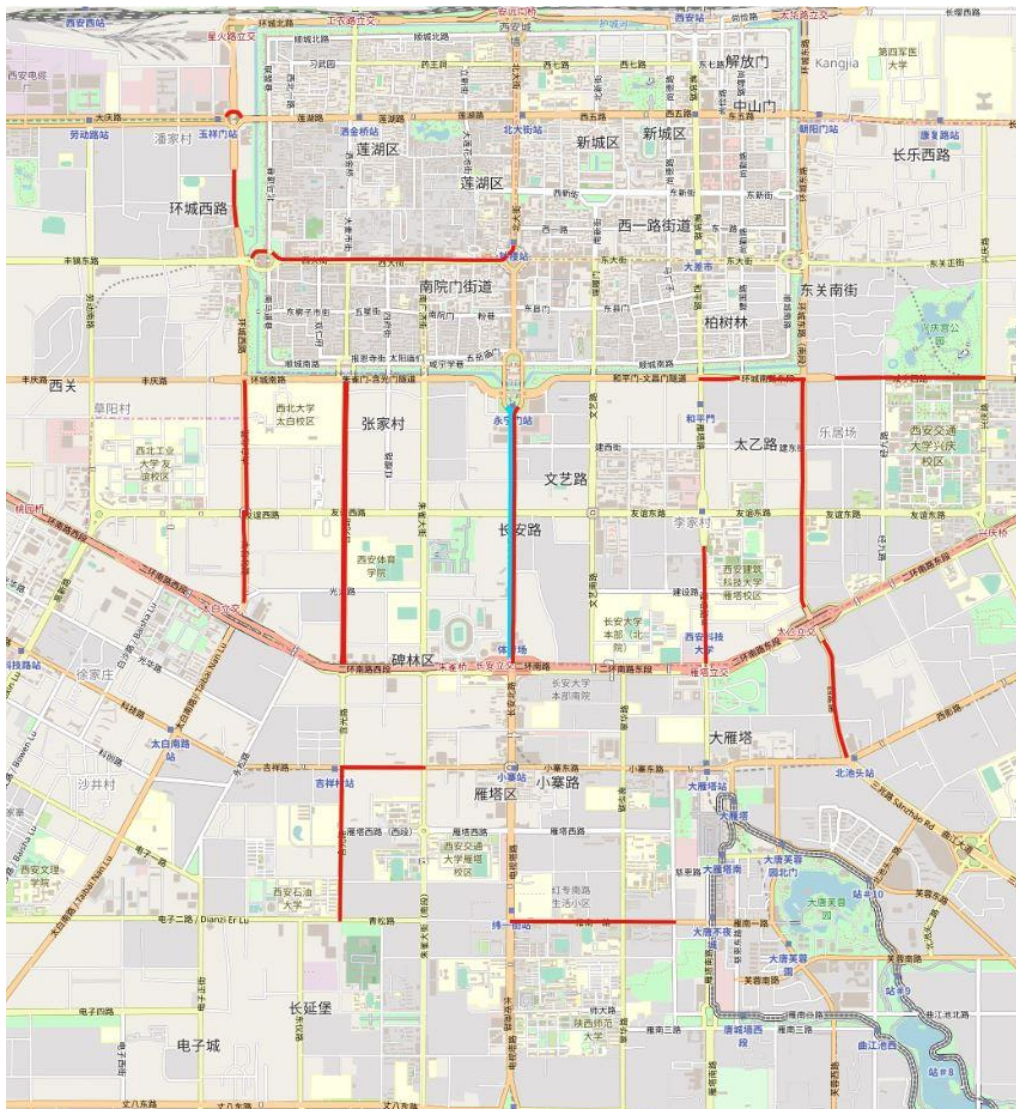
Ο θερμικός χάρτης που απεικονίζεται στο Διάγραμμα 5 δείχνει τις τιμές της απόστασης DTW μεταξύ των χρονοσειρών όλων των οδικών τμημάτων. Μικρότερη τιμή (πιο σκούρο χρώμα) ισοδυναμεί με μεγαλύτερη συσχέτιση.



Διάγραμμα 5: Θερμικός χάρτης Dynamic Time Warping

Στην Εικόνα 8 φαίνονται τα 20 τμήματα με τη μεγαλύτερη συσχέτιση με το επιλεγμένο, σύμφωνα με τον αλγόριθμο Fast DTW. Σε σύγκριση με τα μοτίβα που εντοπίστηκαν στην προηγούμενη ενότητα, υπάρχουν σαφείς διαφορές μεταξύ της χωροχρονικής συσχέτισης που προκύπτει μέσω του DTW και των χωρικών συσχετίσεων που εκτιμώνται μέσω Αμοιβαίας Πληροφορίας και Γραμμικής Συσχέτισης. Συνεπώς μπορεί κανείς να συμπεράνει ότι η χρήση διαφορετικών μετρικών στον εντοπισμό των εξαρτίσεων σε ένα οδικό δίκτυο οδηγεί σε

διαφορετικά συμπεράσματα, τα οποία θα πρέπει να αξιολογηθούν και να εξεταστεί στη συνέχεια αν βελτιώνουν την πρόβλεψη των κυκλοφοριακών συνθηκών.



Εικόνα 8: Χάρτης των 20 τμημάτων με τη χαμηλότερη απόσταση DTW ως προς το επιλεγμένο

#### 4.2.2 Σύγκριση μεθοδολογιών

Για να συγκριθεί η ακρίβεια των μεθόδων που παρουσιάστηκαν στην προηγούμενη ενότητα, αναπτύχθηκαν δύο πρότυπα πρόβλεψης της ταχύτητας του επιλεγμένου τμήματος. Και τα δύο πρότυπα είναι ταξινομητές Μπεύζιανού Δικτύου που κατατάσσουν την ταχύτητα του συγκεκριμένου τμήματος - στόχου σε τρεις περίπου ισομεγέθεις κατηγορίες: <20, 20-26 και >26 km/h, που είναι μια λογική επιλογή για σηματοδοτούμενο οδικό τμήμα, ειδικά αφού ο υπολογισμός της ταχύτητας έχει γίνει συμπεριλαμβάνοντας και πιθανές στάσεις εξαιτίας μεγάλου φόρτου ή σηματοδότησης και σήμανσης. Ειδικά στο συγκεκριμένο τμήμα, η κατανομή ης



ταχύτητας έχει τα περιγραφικά στατιστικά που φαίνονται στον Πίνακα 1 για τις 696 ώρες για τις οποίες υπάρχουν δεδομένα.

Πίνακας 1: Περιγραφικά στατιστικά επιλεγμένου οδικού τμήματος

ΠΛΗΘΟΣ ΤΙΜΩΝ	696
ΜΕΣΗ ΤΙΜΗ	24.39 km/h
ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ	8.38
ΕΛΑΧΙΣΤΗ ΤΙΜΗ	6.24
ΤΕΤΑΡΤΗΜΟΡΙΟ 25%	18.82
ΤΕΤΑΡΤΗΜΟΡΙΟ 50%	21.71
ΤΕΤΑΡΤΗΜΟΡΙΟ 75%	32.27
ΜΕΓΙΣΤΗ ΤΙΜΗ	45.23

Ως δεδομένα εισόδου στα δύο πρότυπα δίνονται οι χρονοσειρές των οδικών τμημάτων, και ζητείται να εκπαιδευτεί ώστε να προβλέπει την κατηγορία της ταχύτητας του επιλεγμένου τμήματος στην ίδια χρονική στιγμή. Το πρώτο πρότυπο «υποχρεώθηκε» να δημιουργήσει το Μπεϋζιανό Δίκτυο και να κάνει τις προβλέψεις χρησιμοποιώντας μόνο τα δεδομένα των οδικών τμημάτων με τη μεγαλύτερη τιμή Αμοιβαίας Πληροφορίας, ενώ το δεύτερο μόνο των οδικών τμημάτων με τις μικρότερες τιμές της απόστασης DTW.

Όπως φαίνεται στον παρακάτω πίνακα, το πρότυπο που χρησιμοποιεί την Αμοιβαία Πληροφορία ως μετρική για τη συσχέτιση των τμημάτων υπερτερεί έναντι του δεύτερου. Η ακρίβεια των δύο προτύπων είναι 89% έναντι 85.6%. Όπως αναφέρθηκε και νωρίτερα, ο υπολογισμός της απόστασης DTW είναι στην ουσία ένα μέτρο της ομοιότητας των χρονοσειρών, με την επιρροή του μέτρου της ταχύτητας να είναι σημαντικότερη από το συνολικό μοτίβο της. Αντίθετα, η Αμοιβαία Πληροφορία αξιολογεί περισσότερο τις τάσεις και τις διακυμάνσεις εντός της χρονοσειράς και την αναλογία (και όχι απόλυτη τιμή) των ταχυτήτων σε συνδυασμό με την πιθανοτική κατανομή τους. Στη συγκεκριμένη εφαρμογή λοιπόν, αποδεικνύεται η τελευταία ως ένα πιο αξιόπιστο μέτρο των χωροχρονικών συσχετίσεων του οδικού δικτύου.

Να σημειωθεί ότι χρησιμοποιήθηκε στις προβλέψεις ο απλός Μπεϋζιανός Ταξινομητής (Naïve Bayesian Classifier), χωρίς άλλη παραμετροποίηση ή βελτιστοποίηση υπερπαραμέτρων.

Ο Πίνακας 2 που ακολουθεί παρουσιάζει συγκριτικά τα μέτρα αξιολόγησης της προβλεπτικής ικανότητας των δύο προτύπων.

Πίνακας 2: Συγκριτικός πίνακας επιδόσεων Προτύπων 1 και 2

ΜΕΤΡΙΚΗ	ΠΡΟΤΥΠΟ 1 (MUTUAL INFO)	ΠΡΟΤΥΠΟ 2 (DTW)
ΑΚΡΙΒΕΙΑ (ACCURACY)	0.89	0.86
RECALL (SENSITIVITY)	0.89	0.86
PRECISION	0.89	0.86
F1 – SCORE	0.89	0.86

Σημειώνεται ότι οι μετρικές που χρησιμοποιούνται για την αξιολόγηση των προτύπων είναι μερικές από τις πιο συχνά χρησιμοποιούμενες. Δοθέντος του πίνακα σύγχυσης των προβλέψεων (Confusion Matrix), υπολογίζονται ως εξής:

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (12)$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (13)$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (14)$$

όπου True Positive είναι το πλήθος των δειγμάτων κάθε κατηγορίας που ταξινομήθηκαν σωστά, False Negative το πλήθος των δειγμάτων κάθε κατηγορίας που ταξινομήθηκαν λανθασμένα και False Positive το πλήθος των δειγμάτων άλλων κατηγοριών που ταξινομήθηκαν σε αυτή που εξετάζεται. Τα παραπάνω μεγέθη υπολογίζονται για όλες τις κατηγορίες και υπολογίζεται στη συνέχεια η μέση τιμή τους.

4.2.3 Ποια είναι η επιρροή των χωροχρονικών συσχετίσεων στη βραχυπρόθεσμη πρόβλεψη των κυκλοφοριακών συνθηκών;

Σε αυτή την ενότητα αναπτύσσεται ένα πρότυπο που δημιουργεί τον Μπεϋζιανό ακυκλικό γράφο όντας «ελεύθερο» να αξιοποιήσει δεδομένα από όσα και όποια οδικά τμήματα προκύψουν από το σχετικό αλγόριθμο. Σε αυτή την περίπτωση, η επιρροή κάθε οδικού τμήματος στην πρόβλεψη είναι ανάλογη της δεσμευμένης πιθανότητάς του με το τμήμα για το οποίο γίνονται οι προβλέψεις.

Η πρόβλεψη αφορά την ταξινόμηση της ταχύτητας του επιλεγμένου τμήματος σε μία από τις τρεις κατηγορίες που αναφέρθηκαν νωρίτερα, αξιοποιώντας δεδομένα από τα υπόλοιπα τμήματα. Τα αποτελέσματα των προβλέψεων συνοψίζονται παρακάτω (Πίνακας 3). Η ακρίβεια αυτού του προτύπου είναι 84.5%.

Πίνακας 3: Πίνακας επιδόσεων Προτύπου 3

ΜΕΤΡΙΚΗ	ΠΡΟΤΥΠΟ 3 (ΝΑΪΒΕ BAYES)
ΑΚΡΙΒΕΙΑ (ACCURACY)	0.84
RECALL (SENSITIVITY)	0.84
PRECISION	0.85
F1 – SCORE	0.85

Όπως μπορεί κανείς να παρατηρήσει, τα πρότυπα που παρουσιάστηκαν στην προηγούμενη ενότητα βελτιώνουν την ακρίβεια των προβλέψεων αισθητά. Το μεν πρώτο, που αξιοποιεί τα τμήματα με τη μεγαλύτερη Αμοιβαία Πληροφορία με αυτό για το οποίο γίνεται η πρόβλεψη, έχει σημαντικά καλύτερη απόδοση, ενώ το δεύτερο επίσης βελτιώνει, αν και οριακά, την ακρίβεια των προβλέψεων. Αυτό το αποτέλεσμα τονίζει την αξία της πραγματοποίησης της παραπάνω ανάλυσης χωροχρονικών συσχετίσεων προτού προβεί κανείς σε προβλέψεις, ακόμα και κατά τη χρήση ενός τόσο δημοφιλούς αλγορίθμου πρόβλεψης.

Η διαδικασία αυτή μειώνει επίσης τη διαστατικότητα του προβλήματος. Το πρόβλημα της διαστατικότητας είναι αρκετά συνηθισμένο στην εφαρμογή αλγορίθμων Μηχανικής Μάθησης, και όχι μόνο, και αφορά την ύπαρξη πολλών γνωρισμάτων σε σχέση με το πλήθος των ατόμων της βάσης δεδομένων, με αποτέλεσμα την ανομοιόμορφη κατανομή των ατόμων στο χώρο των γνωρισμάτων. Έχουν αναπτυχθεί αρκετές μεθοδολογίες για την αντιμετώπιση του προβλήματος, με συνηθέστερη πρακτική την αφαίρεση ή τη συγχώνευση μερικών γνωρισμάτων. Η ύπαρξη διαστατικότητας ευθύνεται πολλές φορές για τη μη σύγκλιση των αλγορίθμων και την κακή συμπεριφορά στις προβλέψεις.

Στην προκειμένη περίπτωση, είναι διαθέσιμα 696 άτομα (χρονικές στιγμές) με 283 γνωρίσματα (πλήθος οδικών τμημάτων) αρχικά, αριθμός υπερβολικά μεγάλος. Μετά την ανάλυση που περιγράφεται παραπάνω ο αριθμός των γνωρισμάτων γίνεται 20, και μάλιστα αυτά που παρουσιάζουν τη μεγαλύτερη συσχέτιση σύμφωνα με κάποια μετρική. Η διαδικασία αυτή επίσης μειώνει και τις απαιτήσεις σε υπολογιστικούς πόρους, κάτι που είναι εξίσου σημαντικό.

#### 4.2.4 Αξιοποίηση Τεχνικών Μηχανικής Μάθησης στην πρόβλεψη της κυκλοφορίας

Στην ενότητα αυτή αναπτύσσεται ένα LSTM δίκτυο με στόχο την πρόβλεψη των κυκλοφοριακών συνθηκών, αξιοποιώντας και τα συμπεράσματα των προηγούμενων ενοτήτων ως προς τη χωροχρονική συσχέτιση μεταξύ των τμημάτων του οδικού δικτύου. Το δίκτυο που αναπτύσσεται ακολουθεί τη δομή που περιεγράφηκε στο Κεφάλαιο 3.

Για τη μοντελοποίηση του δικτύου χρησιμοποιήθηκε η βιβλιοθήκη Keras της γλώσσας προγραμματισμού Python. Το Keras παρέχει ένα φιλικό περιβάλλον χρήσης, όπου μπορεί κανείς να υλοποιήσει τεχνικές και αλγορίθμους Μηχανικής Μάθησης [41]. Λόγω της απλότητας και της αμεσότητας της βιβλιοθήκης έχει εξελιχθεί σε μία από τις δημοφιλέστερες στην ανάπτυξη των συγκεκριμένων προτύπων.

Η λογική που ακολουθήθηκε είναι παρόμοια με αυτή των Μπεϋζιανών Δικτύων των προηγούμενων ενοτήτων ως προς το αντικείμενο της πρόβλεψης και τα δεδομένα που παρέχονται στο πρότυπο. Συγκεκριμένα, η πρόβλεψη αφορούσε την κατηγορία στην οποία θα ανήκει η ταχύτητα κίνησης του επιλεγμένου τμήματος με αξιοποίηση των ταχυτήτων άλλων τμημάτων. Ωστόσο υπάρχει μια σημαντική διαφορά: Στην προηγούμενη ενότητα, η πρόβλεψη γινόταν για την ίδια χρονική στιγμή που

αφορούσαν τα δεδομένα. Σε αυτή την ενότητα, ζητείται η πρόβλεψη των κυκλοφοριακών συνθηκών στο τμήμα-στόχο μία ώρα μετά τη χρονική στιγμή στην οποία αντιστοιχούν τα δεδομένα. Η παραπάνω διαφοροποίηση τονίζει ότι στην προηγούμενη ενότητα το πρότυπο αναπτύχθηκε για να επιβεβαιωθεί η ισχύς των εξαρτίσεων και να συγκριθούν οι μεθοδολογίες, ενώ σε αυτή την ενότητα γίνεται πρόβλεψη που θα μπορούσε να αξιοποιηθεί και σε πραγματικές εφαρμογές.

Αρχικά λοιπόν, αντιστοιχίστηκε η χρονοσειρά του επιλεγμένου τμήματος με αυτές των υπολοίπων με χρονική υστέρηση μίας ώρας. Στη συνέχεια αποφασίστηκε να χρησιμοποιηθούν οι πρώτες 24 από τις 29 ημέρες στην εκπαίδευση του μοντέλου, ενώ οι υπόλοιπες 5 στην αξιολόγηση της απόδοσής του.

Επίσης πρέπει να τονιστεί ότι για την πρόβλεψη αξιοποιήθηκαν τα 20 τμήματα με τη μεγαλύτερη συσχέτιση σύμφωνα με τον υπολογισμό της Αμοιβαίας Πληροφορίας. Η επιλογή αυτή εναρμονίζεται με τα συμπεράσματα του προηγούμενου κεφαλαίου σχετικά με τη χρησιμότητα πραγματοποίησης ανάλυσης χωροχρονικών συσχετίσεων, επιπλέον όμως μειώνει τη διαστατικότητα του προβλήματος και κάνει εφικτή τη γρήγορη σύγκλιση του αλγορίθμου που εφαρμόζεται, χωρίς υπερβολικές απαιτήσεις υπολογιστικής ισχύος.

Το δίκτυο που αναπτύχθηκε αποτελείται από ένα επίπεδο LSTM, όπως αυτό που περιεγράφηκε νωρίτερα στο Κεφάλαιο 3 και ένα επίπεδο εξόδου, απαραίτητο για την ταξινόμηση σε τρεις κατηγορίες. Ακολουθεί ο κώδικας για την υλοποίηση του δικτύου, καθώς και οι διευκρινίσεις σχετικά με τις επιλογές των υπερπαραμέτρων.

```
model = Sequential()  
model.add(LSTM(120, input_shape=(x_train.shape[1], x_train.shape[2])))  
model.add(Dense(3, activation='softmax'))  
model.compile(loss='categorical_crossentropy', optimizer='adam',  
metrics=['accuracy'])  
model.fit(x_train, y_train, epochs=120, batch_size=144)
```

Κατ' αρχάς, όπως είναι προφανές, στους πίνακες `x_train` και `y_train` περιέχονται οι ταχύτητες των 20 τμημάτων και του ενός τμήματος-στόχου αντίστοιχα. Για το επίπεδο του LSTM, επιλέχθηκαν μετά από εκτεταμένη αναζήτηση της βέλτιστης τιμής (grid search) 120 νευρώνες ως μέγεθος των Νευρωνικών Δικτύων που υλοποιούνται εντός του κελιού του LSTM. Ακόμα, το μέγεθος εισόδου δίνεται (1,20), διότι η πρόβλεψη αφορά το επόμενο 1 χρονικό βήμα και ο αριθμός των χαρακτηριστικών των δεδομένων (ταχύτητες τμημάτων με μεγαλύτερη συσχέτιση) είναι 20.

Το επόμενο επίπεδο (εξόδου) έχει 3 νευρώνες, όσες και οι κατηγορίες ταχύτητας, και η ενεργοποίησή του είναι η 'Softmax', που αποτελεί τυπική επιλογή σε προβλήματα ταξινόμησης σε κλάσεις.

Όσον αφορά την εκπαίδευση του προτύπου, ως συνάρτηση μέτρησης σφάλματος χρησιμοποιείται η 'Categorical Crossentropy', που επίσης είναι η τυπική επιλογή για προβλήματα ταξινόμησης, ενώ η βελτιστοποίησή της έγινε με έναν από τους πιο δημοφιλείς αλγορίθμους, τον Adam, ο οποίος αποτελεί μία παραλλαγή του κλασικού Stochastic Gradient Descent. Η επιλογή του έγινε διότι είχε καλύτερα αποτελέσματα από όλες τις εναλλακτικές. Οι εποχές εκπαίδευσης ορίστηκαν σε 120, που αποτελεί μια λογική επιλογή για να αποφευχθεί η υπερ-εκπαίδευση του μοντέλου και το πλήθος των δειγμάτων (batch size) που εκπαιδεύονται μαζί σε κάθε εποχή ορίστηκε σε 144, που αποτελεί τη βέλτιστη επιλογή που προέκυψε μετά από αναζήτηση, και μάλιστα επιλέχθηκε να είναι πολλαπλάσιο του πλήθους των τιμών που αντιστοιχούν σε μία ημέρα.

#### 4.2.5 Αξιολόγηση επιδόσεων δικτύου LSTM

Σε αυτή την ενότητα παρουσιάζονται τα αποτελέσματα που προέκυψαν από την πρόβλεψη των κυκλοφοριακών συνθηκών με χρήση του δικτύου LSTM. Ο Πίνακας 4 είναι ο πίνακας σύγχυσης (confusion matrix) της ταξινόμησης σε κλάσεις, ενώ στη συνέχεια (Πίνακας 5) παρατίθενται και οι σχετικές μετρικές. Η ακρίβεια της πρόβλεψης είναι 89,2%, που αποτελεί μια ικανοποιητική τιμή, ανάλογη των εργασιών της βιβλιογραφίας, που πιστοποιεί την επάρκεια της συγκεκριμένης τεχνικής στην πρόβλεψη χρονοσειρών.

Πίνακας 4: Πίνακας Σύγχυσης (Confusion Matrix) δικτύου LSTM

		ΠΡΟΒΛΕΨΗ		
		<20 km/h	20-26 km/h	>26 km/h
ΠΡΑΓΜΑΤΙΚΗ ΤΙΜΗ	<20 km/h	29	6	0
	20-26 km/h	6	36	0
	>26 km/h	0	1	42

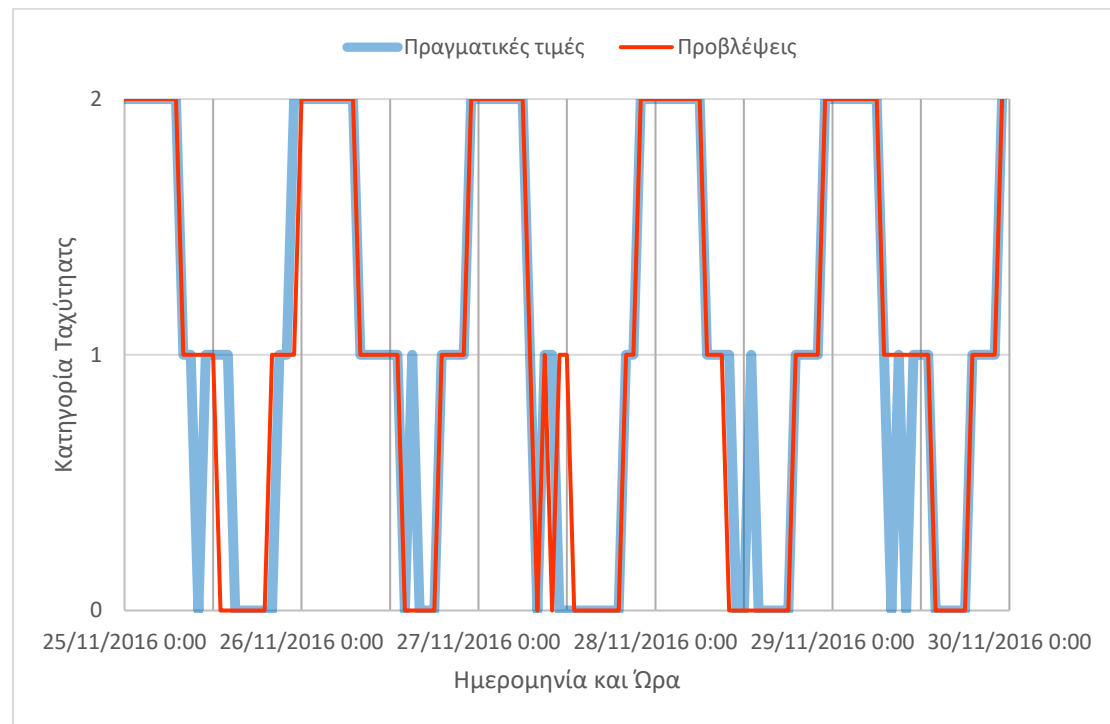
Πίνακας 5: Πίνακας επιδόσεων δικτύου LSTM

ΜΕΤΡΙΚΗ	ΔΙΚΤΥΟ LSTM
ΑΚΡΙΒΕΙΑ (ACCURACY)	0.89
RECALL (SENSITIVITY)	0.89
PRECISION	0.89
F1 – SCORE	0.89

Όπως φαίνεται στον Πίνακα 4, δεν υπάρχουν χονδροειδή σφάλματα στις προβλέψεις, δηλαδή ταξινόμηση στην τρίτη κατηγορία στοιχείων που ανήκουν στην πρώτη ή το αντίστροφο, κάτι που αποτελεί μία ακόμα ένδειξη καλής συμπεριφοράς

του μοντέλου. Ακόμα, μπορεί κανείς να παρατηρήσει ότι η τρίτη κατηγορία (>26 km/h) είναι αυτή που προβλέπεται με τη μεγαλύτερη ακρίβεια.

Τέλος, στο Διάγραμμα 6: παρουσιάζονται συγκριτικά οι πραγματική και η προβλεπόμενη χρονοσειρά.



Διάγραμμα 6: Συγκριτική αποτύπωση πραγματικής και προβλεπόμενης χρονοσειράς

#### 4.2.6 Επέκταση δικτύου LSTM για βραχυχρόνια πρόβλεψη κυκλοφοριακών συνθηκών

Τα οδικά δίκτυα είναι ένα σύστημα που συχνά χαρακτηρίζεται από μεγάλη αστάθεια, με την έννοια ότι ένα αναπάντεχο συμβάν (ατύχημα, κακές καιρικές συνθήκες) μπορεί να διαταράξει την καλή τους λειτουργία και να οδηγήσει σε συνθήκες κυκλοφοριακής συμφόρησης. Για να καταφέρουν να εξασφαλίσουν οι διαχειριστές του δικτύου την εύρυθμη λειτουργία του θα πρέπει να έχουν γνώση των σχέσεων μεταξύ των τμημάτων και ένα εργαλείο που αποφέρει ακριβείς προβλέψεις των κυκλοφοριακών συνθηκών.

Το δίκτυο LSTM που παρουσιάστηκε στην προηγούμενη ενότητα θα μπορούσε να αξιοποιηθεί για τη χάραξη γενικών πολιτικών όσον αφορά τη διαχείριση της κυκλοφορίας, όμως δεν επαρκεί για τη λήψη αποφάσεων σε πραγματικό χρόνο. Υπενθυμίζεται ότι ο ορίζοντας πρόβλεψής του είναι μία ώρα. Σε περίπτωση, λοιπόν, κάποιου απρόοπτου συμβάντος σε κάποιο τμήμα του δικτύου, το οποίο θα προκαλέσει κυκλοφοριακή συμφόρηση (πολύ χαμηλή ταχύτητα κίνησης) στο τμήμα αυτό, η διαθέσιμη πρόβλεψη θα αφορά τις συνθήκες του δικτύου μία ώρα μετά. Το

πιθανότερο είναι η αιτία της συμφοράς να έχει αρθεί μετά από μία ώρα και η επίδρασή της να μην είναι πλέον ορατή στο δίκτυο. Για την αντιμετώπιση τέτοιου είδους προβλημάτων, οι διαχειριστές πρέπει να στραφούν σε προσεγγίσεις με κοντινότερο ορίζοντα.

Στις μέρες μας, λόγω της πληθώρας των διαθέσιμων δεδομένων και της μεγάλης υπολογιστικής ισχύος των σύγχρονων ηλεκτρονικών υπολογιστών, είναι δυνατή η ανάπτυξη προτύπων που καθιστούν δυνατή τη βραχυχρόνια πρόβλεψη των κυκλοφοριακών συνθηκών. Στο πλαίσιο αυτό, παρουσιάζεται σε αυτή την ενότητα ένα LSTM δίκτυο με ορίζοντα πρόβλεψης 15 λεπτών.

Για να είναι αυτό δυνατό, ακολουθήθηκε ξανά η διαδικασία που περιεγράφηκε στα προηγούμενα κεφάλαια της εργασίας, με τις προφανείς αλλαγές ώστε να προκύψουν χρονοσειρές με βήμα 15 λεπτών και οι αντίστοιχες μετρικές συσχέτισης που τις αφορούν. Ακόμα, προέκυψε ένα ελαφρώς διαφοροποιημένο δίκτυο LSTM ως το βέλτιστο για την πραγματοποίηση προβλέψεων για τα συγκεκριμένα δεδομένα. Ο κώδικας υλοποίησης του δικτύου έχει ως εξής:

```
model = Sequential()
model.add(LSTM(100,input_shape=(x_train.shape[1],x_train.shape[2])))
model.add(Dense(3, activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer='adam',
metrics=['accuracy'])
model.fit(x_train, y_train, epochs=120, batch_size=480)
```

Όπως στην προηγούμενη ενότητα, οι επόμενοι πίνακες (Πίνακας 6, Πίνακας 7) παρουσιάζουν την επίδοση του δικτύου.

Πίνακας 6: Πίνακας Σύγχυσης (Confusion Matrix) δικτύου LSTM βραχυχρόνιας πρόβλεψης

		ΠΡΟΒΛΕΨΗ		
		<20 km/h	20-26 km/h	>26 km/h
ΠΡΑΓΜΑΤΙΚΗ ΤΙΜΗ	<20 km/h	105	35	0
	20-26 km/h	18	143	6
	>26 km/h	0	6	167

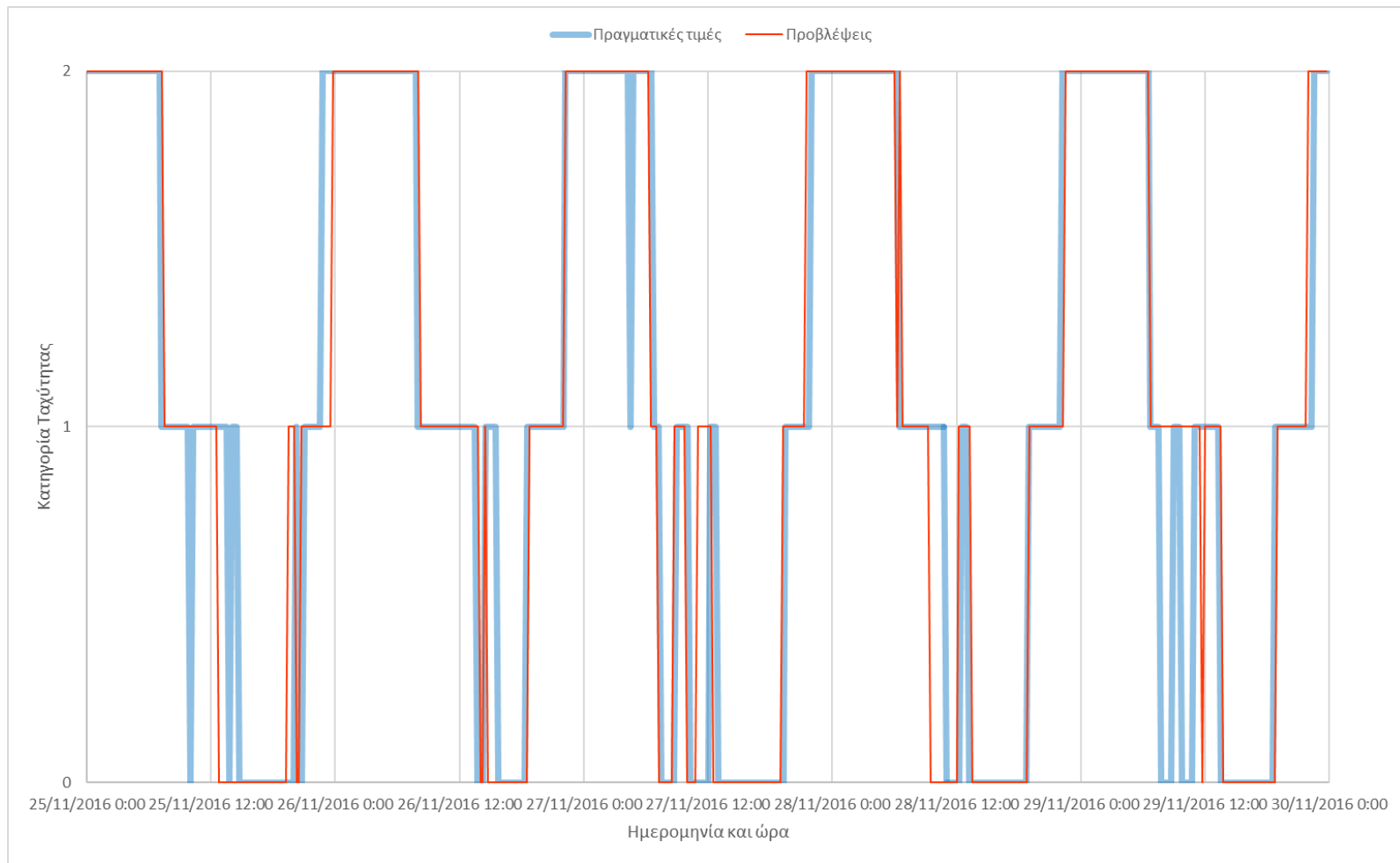
Πίνακας 7: Πίνακας επιδόσεων δικτύου LSTM βραχυχρόνιας πρόβλεψης

ΜΕΤΡΙΚΗ	ΔΙΚΤΥΟ LSTM 2
ΑΚΡΙΒΕΙΑ (ACCURACY)	0.86
RECALL (SENSITIVITY)	0.86
PRECISION	0.86
F1 – SCORE	0.86

Η απόδοση αυτού του δικτύου είναι ελαφρώς μειωμένη σε σχέση με αυτή του δικτύου της προηγούμενης ενότητας. Ωστόσο κρίνεται αρκετά ικανοποιητική, αν λάβει κανείς υπόψη τη μεγαλύτερη διακύμανση των ταχυτήτων λόγω της μείωσης του χρονικού βήματος της χρονοσειράς.

Ακολουθεί η συγκριτική αποτύπωση της πραγματικής και της προβλεπόμενης χρονοσειράς (Διάγραμμα 7).





Διάγραμμα 7: Συγκριτική αποτύπωση πραγματικής και προβλεπόμενης χρονοσειράς βήματος 15 λεπτών

# Κεφάλαιο 5

## Συμπεράσματα

### 5.1 Γενικά

Ο κλάδος της πρόβλεψης κυκλοφοριακών συνθηκών προσελκύει έντονο ερευνητικό ενδιαφέρον, διότι η ακριβής πρόβλεψη των κυκλοφοριακών συνθηκών είναι απαραίτητη στην αποτελεσματική διαχείριση της κυκλοφορίας και στην αντιμετώπιση αναπάντεχων συμβάντων στο οδικό δίκτυο. Η είσοδος των αλγορίθμων Μηχανικής Μάθησης και η διαθεσιμότητα δεδομένων ευρείας κλίμακας έχει δώσει νέα πνοή στο συγκεκριμένο τομέα.

Η ανάλυση των χωροχρονικών συσχετίσεων ενός οδικού δικτύου πριν την ανάπτυξη προτύπων πρόβλεψης είναι πολύ σημαντική για την ακρίβεια των προβλέψεων. Επιπλέον, τα συγκεκριμένα πρότυπα προσφέρουν καλύτερη επεξηγηματικότητα της χωροχρονικής εξέλιξης των κυκλοφοριακών συνθηκών και, για το λόγο αυτό, μπορούν να αξιοποιηθούν σε μια πληθώρα πραγματικών εφαρμογών.

### 5.2 Βασικά συμπεράσματα

Μετρικές ευρύτερης χρησιμότητας, όπως η γραμμική συσχέτιση κατά Pearson, όπως επίσης και πιο εξειδικευμένες μετρικές χρονοσειρών, όπως η Αμοιβαία Πληροφορία και το Dynamic Time Warping, παρέχουν μία ξεκάθαρη εικόνα των χωροχρονικών συσχετίσεων των οδικών τμημάτων ενός αστικού οδικού δικτύου. Αυτές οι συσχετίσεις πηγάζουν από τις σχετικές θέσεις των τμημάτων και του φόρτου που αυτά εξυπηρετούν. Συνεπώς, τα αποτελέσματα που προκύπτουν είναι λογικά και εξηγήσιμα.

Για να διαπιστωθεί η σημασία του εντοπισμού των χωροχρονικών συσχετίσεων, αλλά και για να αξιολογηθούν οι μετρικές που αναφέρονται παραπάνω, αναπτύχθηκε ένα Μπεϋζιανό Δίκτυο με απλή δομή, το οποίο ταξινομεί την ταχύτητα ενός οδικού τμήματος, βασιζόμενο στις χωροχρονικές συσχετίσεις που έχουν εντοπιστεί. Παρ' όλο που ο συγκεκριμένος αλγόριθμος δεν είναι η πλέον ενδεδειγμένη επιλογή για δεδομένα σε μορφή χρονοσειράς, η βελτιωμένη απόδοση που παρατηρήθηκε κατά την εισαγωγή των χωροχρονικών συσχετίσεων του δικτύου είναι εμφανής και σημαντική τόσο για την ακρίβεια των προβλέψεων, όσο και για την υπολογιστική αποδοτικότητα του συστήματος. Τα αποτελέσματα έδειξαν επίσης ότι η Αμοιβαία Πληροφορία είναι η πιο κατάλληλη μετρική για τη συγκεκριμένη εφαρμογή, γιατί εστιάζει στην ύπαρξη γραμμικών και μη γραμμικών συσχετίσεων, ενώ η επιρροή του μέτρου της ταχύτητας είναι λιγότερο σημαντική.

Τέλος, τα δίκτυα LSTM είναι μία από τις ισχυρότερες και αποτελεσματικότερες τεχνικές Μηχανικής Μάθησης για την πρόβλεψη χρονοσειρών. Τα δίκτυα που αναπτύχθηκαν για την πρόβλεψη των κυκλοφοριακών συνθηκών ενός οδικού τμήματος αποδείχθηκαν ικανοποιητικά ακριβή, ενώ η μείωση της διαστατικότητας μέσω την ανάλυσης χωροχρονικών συσχετίσεων έκανε εφικτή την υλοποίησή τους, ακόμα και με χρήση των περιορισμένων διαθέσιμων υπολογιστικών πόρων.

### 5.3 Εισηγήσεις για περαιτέρω έρευνα

Η παρούσα εργασία αφήνει πολλά ανοιχτά παράθυρα για περαιτέρω εξέλιξή της. Η μελλοντική έρευνα θα περιλαμβάνει σίγουρα την επεξεργασία δεδομένων από περισσότερες ακόμα ημέρες, με σκοπό την εξαγωγή πιο γενικών συμπερασμάτων. Το παραπάνω απαιτεί ένα ισχυρότερο υπολογιστικό σύστημα όσον αφορά τόσο το υλικό, όσο και το λογισμικό που θα χρησιμοποιηθεί, καθώς το πλήθος των δεδομένων μοιραία οδηγεί σε προσεγγίσεις επεξεργασίας δεδομένων ευρείας κλίμακας (Big Data).

Επίσης, θα χρησιμοποιηθούν πιο εξειδικευμένες και πρωτοποριακές τεχνικές Μηχανικής Μάθησης, τόσο για την πρόβλεψη, όσο και για την αναγνώριση των χωροχρονικών συσχετίσεων, καθώς φαίνεται πως υπάρχει αρκετά μεγάλο περιθώριο για βελτιστοποίηση και επίτευξη μεγάλης ακρίβειας προβλέψεων ενός ή ακόμα και περισσότερων βημάτων στο μέλλον.

Ακόμα, όσα αναφέρθηκαν παραπάνω επιτρέπουν την περαιτέρω μείωση του χρονικού βήματος των χρονοσειρών και την ακόμα πιο βραχυχρόνια πρόβλεψη της κυκλοφορίας, έως και την πρόβλεψη κυκλοφοριακών συνθηκών σε πραγματικό χρόνο (online learning).

Τέλος, αναμένεται να καταστεί δυνατή και η πρόβλεψη για όλα τα τμήματα του δικτύου ταυτόχρονα, και όχι μόνο ενός, με χρήση διαφορετικής συνολικά προσέγγισης όσον αφορά το στήσιμο του προβλήματος και την αξιοποίηση των δεδομένων.

## Βιβλιογραφία

- [1] J.-P. Rodrigue and T. Notteboom, *The Geography of Transport Systems*. 2017.
- [2] E. R. Jayaratne, L. Wang, D. Heuff, L. Morawska, and L. Ferreira, "Increase in particle number emissions from motor vehicles due to interruption of steady traffic flow," *Transp. Res. Part D Transp. Environ.*, vol. 14, no. 7, pp. 521–526, 2009.
- [3] European environment agency, "Total greenhouse gas emissions from transport," *Indic. Assess. - Data maps*, p. 13, 2018.
- [4] A. K. Emo, G. Matthews, and G. J. Funke, "The slow and the furious: Anger, stress and risky passing in simulated traffic congestion," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 42, pp. 1–14, 2016.
- [5] M. A. Quddus, C. Wang, and S. G. Ison, "Road traffic congestion and crash severity: Econometric analysis using ordered response models," *J. Transp. Eng.*, vol. 136, no. 5, pp. 424–435, 2010.
- [6] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transp. Res. Part C Emerg. Technol.*, vol. 43, pp. 3–19, 2014.
- [7] M. G. Karlaftis and E. I. Vlahogianni, "Statistical methods versus neural networks in transportation research: Differences, similarities and some insights," *Transp. Res. Part C Emerg. Technol.*, vol. 19, no. 3, pp. 387–399, 2010.
- [8] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach," *Transp. Res. Part C Emerg. Technol.*, vol. 13, no. 3, pp. 211–234, 2005.
- [9] T. Alghamdi, K. Elgazzar, M. Bayoumi, T. Sharaf, and S. Shah, "Forecasting Traffic Congestion Using ARIMA Modeling," *2019 15th Int. Wirel. Commun. Mob. Comput. Conf.*, pp. 1227–1232, 2019.
- [10] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *Eur. Transp. Res. Rev.*, vol. 7, no. 3, pp. 1–9, 2015.
- [11] F. Xu *et al.*, "Big Data Driven Mobile Traffic Understanding and Forecasting: A Time Series Approach," *IEEE Trans. Serv. Comput.*, vol. 9, no. 5, pp. 796–805, 2016.
- [12] Y. Xie, Y. Zhang, and Z. Ye, "Short-Term Traffic Volume Forecasting Using Kalman Filter with Discrete Wavelet Decomposition," vol. 22, pp. 326–334, 2007.
- [13] J. Wang, W. Deng, and Y. Guo, "New Bayesian combination method for short-term traffic flow forecasting," *Transp. Res. PART C*, pp. 1–16, 2014.
- [14] S. Dunne and B. Ghosh, "Regime-Based Short-Term Multivariate Traffic

- Condition Forecasting Algorithm," no. April, pp. 455–466, 2012.
- [15] F. Schimbinschi, X. V. Nguyen, J. Bailey, C. Leckie, H. Vu, and R. Kotagiri, "Traffic forecasting in complex urban networks: Leveraging big data and machine learning," *Proc. - 2015 IEEE Int. Conf. Big Data, IEEE Big Data 2015*, pp. 1019–1024, 2015.
- [16] L. Qu, W. Li, W. Li, D. Ma, and Y. Wang, "Daily long-term traffic flow forecasting based on a deep neural network," *Expert Syst. Appl.*, vol. 121, pp. 304–312, 2019.
- [17] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors (Switzerland)*, vol. 17, no. 4, 2017.
- [18] H. Yu, Z. Wu, S. Wang, Y. Wang, and X. Ma, "Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks," *Sensors (Switzerland)*, vol. 17, no. 7, pp. 1–16, 2017.
- [19] Z. Cui, K. Henrickson, R. Ke, and Y. Wang, "Traffic Graph Convolutional Recurrent Neural Network: A Deep Learning Framework for Network-Scale Traffic Learning and Forecasting," 2019.
- [20] Z. Zhao, W. Chen, X. Wu, P. C. V. Chen, and J. Liu, "LSTM network: A deep learning approach for short-term traffic forecast," *IET Image Process.*, vol. 11, no. 1, pp. 68–75, 2017.
- [21] Z. Cui, R. Ke, and Y. Wang, "Deep Stacked Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction," pp. 1–12, 2018.
- [22] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, "A hybrid deep learning based traffic flow prediction method and its understanding," *Transp. Res. Part C Emerg. Technol.*, vol. 90, no. 5, pp. 166–180, 2018.
- [23] Y. Wu and H. Tan, "Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework," pp. 1–14, 2016.
- [24] I. Laña, J. L. Lobo, E. Capecci, J. Del Ser, and N. Kasabov, "Adaptive long-term traffic state estimation with evolving spiking neural networks," *Transp. Res. Part C Emerg. Technol.*, vol. 101, no. April 2018, pp. 126–144, 2019.
- [25] J. Lopes, J. Bento, E. Huang, C. Antoniou, and M. Ben-Akiva, "Traffic and mobility data collection for real-time applications," *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC*, pp. 216–223, 2010.
- [26] C. Antoniou, R. Balakrishna, and H. N. Koutsopoulos, "A Synthesis of emerging data collection technologies and their impact on traffic management applications," *Eur. Transp. Res. Rev.*, vol. 3, no. 3, pp. 139–148, 2011.
- [27] C. Goves, R. North, R. Johnston, and G. Fletcher, "Short term traffic prediction on the UK motorway network using neural networks," *Transp. Res. Procedia*, vol. 13, pp. 184–195, 2016.
- [28] A. Sánchez, A. L. Lara, and F. Cabero, "Measuring uncertainty of traffic volume

- 
- on motorway concessions: a time-series analysis," *Transp. Res. Procedia*, vol. 27, pp. 3–10, 2017.
- [29] A. Ermagun and D. Levinson, "Spatiotemporal traffic forecasting: review and proposed directions," *Transp. Rev.*, vol. 0, no. 0, pp. 1–29, 2018.
- [30] H. D. Abarbanel, *Analysis of observed chaotic data*. 1996.
- [31] H. Kantz and T. Schreiber, *Non-linear time series analysis*. 1997.
- [32] D. Berndt and J. Clifford, "Using Dynamic Time Warping to Find Patterns in Time Series," *Proc. AAAI-94 Work. Knowl. Discov. Databases*, pp. 229–248, 1994.
- [33] M. Lee, S. Lee, M. J. Choi, Y. S. Moon, and H. S. Lim, "HybridFTW: Hybrid computation of dynamic time warping distances," *IEEE Access*, vol. 6, pp. 2085–2096, 2017.
- [34] S. Salvador and P. Chan, *FastDTW: Toward accurate dynamic time warping in linear time and space*. 2007.
- [35] J. Pearl, *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2000.
- [36] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian Network Classifiers," *Mach. Learn.*, vol. 29, pp. 131–163, 1997.
- [37] A. Geron, *Hands-on Machine Learning with Scikit-Learn, Keras and Tensorflow*. 2017.
- [38] A. Gulli and S. Pal, *Deep Learning with Keras*. 2017.
- [39] C. Olah, "Understanding LSTM Networks," 2015. [Online]. Available: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
- [40] H. Tveite, "The QGIS NNJoin Plugin," 2014. [Online]. Available: <http://arken.nmbu.no/~havatv/gis/qgisplugins/NNJoin/#>.
- [41] F. Chollet, "Keras: The Python Deep Learning library," 2014. [Online]. Available: <https://keras.io/>.