

ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ  
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ  
ΤΜΗΜΑΤΟΣ ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ



ΕΞΟΥΥΞΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΙΔΙΩΤΙΚΟΤΗΤΑ ΣΕ ΚΙΝΗΤΑ ΠΕΡΙΒΑΛΛΟΝΤΑ

Διπλωματική Εργασία

Της

Χριστίνας Ραπασάνη

Θεσσαλονίκη, 08/2020

## Περίληψη

Τα κινητά τηλέφωνα έχουν αναδειχθεί αναπόσπαστο μέρος της καθημερινής ζωής της σύγχρονης κοινωνίας, ενώ ο αριθμός των χρηστών έξυπνων τηλεφώνων (smartphones) αυξάνεται με ραγδαίους ρυθμούς. Οι χρήστες των έξυπνων τηλεφώνων χρησιμοποιούν εφαρμογές οι οποίες καταγράφουν συνεχώς δεδομένα κάνοντας χρήση των αισθητήρων που διαθέτουν οι κινητές συσκευές. Η παρούσα εργασία αποσκοπεί στη διερεύνηση της σωστής εύρεσης της τοποθεσίας αρχικά μέσω ενός εξατομικευμένου πλαισίου εξόρυξης δεδομένων για κινητά το οποίο χρησιμοποιεί χωρικά και χρονικά δεδομένα αλλά και άλλα δεδομένα όπως το επιταχυνσιόμετρο, το bluetooth και το αρχείο κλήσεων και μηνυμάτων(sms). Επίσης θα γίνει μια μελέτη σχετικά με την εξόρυξη κινητών δεδομένων (mobile data mining) με σκοπό την πρόβλεψη της επόμενης τοποθεσίας (NextLocation) όπου αυτό επιτυγχάνεται πρώτον μέσω ενός πλαισίου που συλλέγει δεδομένα από τις κλήσεις , τα μηνύματα ,το Bluetooth και το επιταχυνσιόμετρο. Εφόσον το αντικείμενο αυτής της μελέτης είναι η επεξεργασία προσωπικών δεδομένων, θα γίνει αναφορά στο ζήτημα της ιδιωτικότητας που προκύπτει από την εξόρυξη των δεδομένων.

**ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ** : Αναγνώριση Δραστηριότητας, Κατηγοριοποίηση, Android, Χάρτες Google, NextLocation, ιδιωτικότητα

## **Abstract**

Mobile phones have become an integral part of the daily life of modern society, and the number of smart phone users is growing rapidly. Smartphone users use applications that constantly record data using the sensors available on mobile devices. The aim of this study is to investigate the correct location of the site initially through a personalized mobile data mining framework that uses spatial and temporal data as well as other data such as accelerometer, bluetooth and call and sms file. There will also be a study on mobile data mining in order to predict the next location (NextLocation) where this is first achieved through a framework that collects data from calls, messages, Bluetooth and accelerometer. Since the subject of this study is the processing of personal data, reference will be made to the issue of privacy arising from data mining.

**Keywords:** Activity Recognition, Categorization, Android, Google Maps, NextLocation, Privacy



## Περιεχόμενα

<b>Κεφάλαιο 1.....</b>	<b>6</b>
1.1 Εισαγωγή.....	6
1.2 Smartphones και λειτουργικό σύστημα.....	7
1.2.1 Smartphones.....	7
1.2.2 Αισθητήρες smartphones.....	7
1.2.3 Λειτουργικό σύστημα Android.....	9
1.2.3.1 Εκδόσεις Android.....	10
1.2.3.2 Αρχιτεκτονική Android.....	11
1.2.3.3 Χαρακτηριστικά του Android.....	12
1.3 Εξόρυξη δεδομένων και εξόρυξη δεδομένων σε κινητές συσκευές.....	14
1.3.1 Εξόρυξη δεδομένων(Data Mining).....	14
1.3.2 Μέθοδοι εξόρυξης δεδομένων.....	16
1.3.3 Εξόρυξη δεδομένων σε κινητές συσκευές (Mobile Data Mining).....	17
<b>Κεφάλαιο 2.....</b>	<b>19</b>
2.1 Δεδομένα μετακίνησης και χαρακτηριστικά.....	19
2.1.1 Βασικές Αρχές δεδομένων κίνησης.....	19
2.1.2 Χαρακτηριστικά κίνησης.....	20
2.2 Τεχνικές αναγνώρισης κίνησης.....	25
2.2.1 Κατηγοριοποίηση (Classification).....	26
2.2.2 Συσταδιοποίηση τροχιών κίνησης.....	27
2.2.3 Conditional Random Fields.....	28
<b>Κεφάλαιο 3.....</b>	<b>30</b>
3.1 Πρόβλεψη επόμενης τοποθεσίας.....	30

3.2 Σχετικές πληροφορίες.....	31
3.3 Προκλήσεις δεδομένων κινητής τηλεφωνίας.....	33
3.4 Next prediction Place: NextLocation.....	34
3.5 Πειραματική αξιολόγηση.....	36
3.5.1 Nokia MDC dataset.....	36
3.5.2 Μετασχηματισμός δεδομένων.....	37
3.5.3 Τεχνικές.....	39
3.6 Αποτελέσματα και συμπεράσματα.....	41
<b>Κεφάλαιο 4.....</b>	<b>42</b>
Ιδιωτικότητα.....	42
<b>Κεφάλαιο 5.....</b>	<b>45</b>
Συμπεράσματα.....	45
<b>Βιβλιογραφία.....</b>	<b>46</b>

# Κεφάλαιο 1

## 1.1 Εισαγωγή

Τα ασύρματα δίκτυα καλύπτουν την μεγαλύτερη έκταση των αστικών κέντρων, ενώ ο αριθμός των χρηστών έξυπνων τηλεφώνων αυξάνεται καθημερινά. Τα smartphones παρέχουν εφαρμογές καταγραφής διάφορων μορφών δεδομένων κάνοντας χρήση ενσωματωμένων αισθητήρων, όπως : Bluetooth, GPS, πυξίδα, φωτογραφική μηχανή και μικρόφωνο. Αυτοί οι αισθητήρες, κάνοντας χρήση ασύρματων δικτύων μπορούν να ανιχνεύσουν την θέση και την κίνηση του χρήστη κατά τη διάρκεια χρήσης του κινητού τηλεφώνου. Κατά συνέπεια, η συλλογή δεδομένων κίνησης από την καθημερινότητα του χρήστη αποτελεί μια απλή διαδικασία, η οποία έχει ως στόχο με την ανάλυση των δεδομένων να διευκολύνει τις διάφορες δραστηριότητες του, παρακολουθώντας τον τρόπο ζωής του. Επιπλέον, η παρακολούθηση και ο έλεγχος της καθημερινής δραστηριότητας του ατόμου βρίσκουν εφαρμογή στην υγειονομική περίθαλψη, καθώς με αυτόν τον τρόπο μπορούν να διαγνωστούν σοβαρά προβλήματα υγείας ή ακόμα να εποπτευθούν ασθενείς που χρειάζονται μετεγχειρητική παρακολούθηση

Η πρόβλεψη θέσης μέσω της εξόρυξης δεδομένων για κινητά εκμεταλλεύεται τόσο μεγάλα δεδομένα σε εφαρμογές όπως ο προγραμματισμός κίνησης, η διαφήμιση βάσει τοποθεσίας, η έξυπνη κατανομή πόρων καθώς και σε υπηρεσίες συμβούλων όπως η δημοφιλής Apple Siri ή το Google Now. Σε αυτή την αναφορά αναλύεται το δύσκολο πρόβλημα της πρόβλεψης της επόμενης θέσης ενός χρήστη κινητού που έχει δώσει δεδομένα σχετικά με την τρέχουσα τοποθεσία του. Σε αυτό το σημείο θα επικεντρωθούμε στο NextLocation που είναι ένα εξατομικευμένο πλαίσιο εξόρυξης δεδομένων για κινητά το οποίο χρησιμοποιεί χωρικά και χρονικά δεδομένα αλλά και άλλα δεδομένα όπως το επιταχυνσιόμετρο, το bluetooth και το αρχείο κλήσεων και μηνυμάτων(sms). Επιπροσθέτως το προτεινόμενο πλαίσιο αντιπροσωπεύει ένα νέο πρότυπο για την προστασία της ιδιωτικής ζωής με σκοπό την πρόβλεψη της επόμενης θέσης καθώς τα δεδομένα του κινητού τηλεφώνου δεν μοιράζονται χωρίς την άδεια του χρήστη. Πειράματα έχουν πραγματοποιηθεί χρησιμοποιώντας δεδομένα από το Nokia Mobile Data Challenge (MDC). Τα αποτελέσματα για τα

δεδομένα MDC δείχνουν μεγάλη μεταβλητότητα στην προβλεπτική ακρίβεια περίπου 17% μεταξύ των χρηστών. Για παράδειγμα, οι ακανόνιστοι χρήστες είναι πολύ δύσπιστοι να προβλέψουν, ενώ για πιο τακτικούς χρήστες είναι δυνατόν να επιτευχθεί ακρίβεια άνω του 80%.

## **1.2 Smartphones και Λειτουργικό Σύστημα Android**

### **1.2.1 Smartphones**

Με τον όρο έξυπνο τηλέφωνο (smartphone) αναφερόμαστε σε ένα κινητό τηλέφωνο βασισμένο σε ένα λειτουργικό σύστημα κινητής τηλεφωνίας με περισσότερη προηγμένη υπολογιστική ικανότητα και συνδεσιμότητα σε σχέση με ένα απλό κινητό τηλέφωνο. Τα σύγχρονα smartphones αποτελούν πολυχρηστικές συσκευές καθώς περιλαμβάνουν λειτουργίες media players, ψηφιακές φωτογραφικές μηχανές, πλοήγηση GPS, οθόνες αφής υψηλής ανάλυσης και web browsers που εμφανίζουν τυποποιημένες ιστοσελίδες, καθώς και βελτιστοποιημένες ιστοσελίδες για κινητά. Η πρόσβαση σε δεδομένα υψηλής ταχύτητας παρέχεται μέσω Wi-Fi και μέσω ασύρματων ευρυζωνικών υπηρεσιών. Τα λειτουργικά συστήματα (OS) των κινητών τηλεφώνων που χρησιμοποιούνται από τα σύγχρονα smartphones περιλαμβάνουν το Android της Google, το iOS της Apple, το Symbian της Nokia, το BlackBerry OS της RIM, το Firefox OS της Mozilla και το Ubuntu Phone της Canonical Ltd's[1].

### **1.2.2 Αισθητήρες Smartphones**

#### **Global Positioning System (GPS)**

Το GPS (Global Positioning System, Παγκόσμιο Σύστημα τοποθεσίας) είναι ένα παγκόσμιο σύστημα εντοπισμού θέσης, το οποίο βασίζεται σε εικοσιτέσσερις δορυφόρους της Γης, οι οποίοι διαθέτουν ειδικές συσκευές ονομάζονται δέκτες GPS. Οι δέκτες αυτοί παρέχουν ακριβείς πληροφορίες για τη θέση ενός σημείου, το υψόμετρό του, την ταχύτητα και την κατεύθυνση της κίνησης του. Το GPS παρέχει πληροφορίες τοποθεσίας και χρόνου σε όλες τις καιρικές συνθήκες, οπουδήποτε πάνω ή κοντά στη Γη, όπου υπάρχει ανεμπόδιστη οπτική επαφή με τέσσερις ή



περισσότερους δορυφόρους GPS. Συντηρείται από την κυβέρνηση των Ηνωμένων Πολιτειών και είναι ελεύθερα προσβάσιμο με ένα δέκτη GPS. Το GPS δημιουργήθηκε αναπτύχθηκε το 1973 από το Υπουργείο Άμυνας των ΗΠΑ με σκοπό να ξεπεράσει τους περιορισμούς. Αρχικά ονομάστηκε "NAVSTAR GPS" (Navigation Signal Timing and Ranging Global Positioning System) και τέθηκε σε πλήρη λειτουργία το 1994 [2]. Για τον προσδιορισμό της θέσης του χρήστη με γεωγραφικό μήκος (longitude) και γεωγραφικό πλάτος (latitude) απαιτείται σύνδεση σε τρεις δορυφόρους, ενώ για τον προσδιορισμό του υψόμετρου απαιτούνται τέσσερις δορυφόροι. Η πλειοψηφία των δεκτών GPS βρίσκονται σε κινητά τηλέφωνα, με ποικίλους βαθμούς κάλυψης και προσβασιμότητας των χρηστών. Στα περισσότερα smartphones υπάρχει διαθέσιμο λογισμικό πλοήγησης, καθώς και σε ορισμένα τηλέφωνα που διαθέτουν Java που τους επιτρέπει να χρησιμοποιούν εσωτερικό ή εξωτερικό δέκτη GPS. Ακόμα, μερικά κινητά τηλέφωνα χρησιμοποιούν A-GPS (assisted GPS), το οποίο όμως υπολειτουργεί όταν βρίσκεται εκτός του δικτύου κινητής τηλεφωνίας. Παράλληλα, μερικά άλλα χρησιμοποιούν ένα υβριδικό σύστημα εντοπισμού θέσης που μπορούν να χρησιμοποιήσουν άλλα σήματα, όταν τα σήματα GPS είναι ανεπαρκή [3]. Στα κινητά τηλέφωνα, η θέση μπορεί να οριστεί είτε με χρήση GPS, είτε με τον τριγωνισμό της απόστασης από την κεραία της κινητής τηλεφωνίας ή το ασύρματο δίκτυο (Wi-Fi) που χρησιμοποιεί ο χρήστης, είτε θεωρώντας ως τοποθεσία του χρήστη την κεραία της κινητής τηλεφωνίας ή του ασύρματου δικτύου. Τα smartphones έχουν τη δυνατότητα να επιλέγουν κάθε φορά τον βέλτιστο τρόπο για τον ορισμό της θέσης του χρήστη [4].

### **Επιταχυνσιόμετρο (accelerometer)**

Το επιταχυνσιόμετρο είναι μια συσκευή που μετρά την επιτάχυνση. Η επιτάχυνση που μετράει δεν είναι απαραίτητως η επιτάχυνση με βάση τις συντεταγμένες (ρυθμός μεταβολής της ταχύτητας). Αντίθετα, το επιταχυνσιόμετρο υπολογίζει την επιτάχυνση που συνδέεται με το φαινόμενο του βάρους οποιασδήποτε μάζας σε αδράνεια σε σχέση με τη διάταξη της συσκευής. Μερικά smartphones, περιέχουν επιταχυνσιόμετρα για τον έλεγχο της διεπαφής χρήστη. Συχνά το επιταχυνσιόμετρο χρησιμοποιείται για να παρουσιάσει θέα στο γύρω τοπίο ή πορτρέτο της οθόνης της συσκευής, με βάση την διάταξη της συσκευής. Η 5η και 6η γενιά Apple iPod Nano

διαθέτει ένα ενσωματωμένο επιταχυνσιόμετρο και μια εφαρμογή που ονομάζεται Fitness και μπορεί να χρησιμοποιηθεί για να καταγράψει τα βήματα κατά το περπάτημα ή το τρέξιμο [5].

### **Μαγνητόμετρο (magnetometer)**

Το μαγνητόμετρο είναι ένα όργανο μέτρησης της δύναμης και σε ορισμένες περιπτώσεις, της κατεύθυνσης των μαγνητικών πεδίων. Πολλά smartphones διαθέτουν μαγνητόμετρα και παρέχουν εφαρμογές που λειτουργούν ως πυξίδες. Άλλες συσκευές τηλεφώνων, χρησιμοποιούν μαγνητόμετρα τριών αξόνων, τα οποία δεν είναι ευαίσθητα στον προσανατολισμό ή την ανύψωση της συσκευής. Ερευνητές στην Deutsche Telekom έχουν χρησιμοποιήσει μαγνητόμετρα ενσωματωμένα σε κινητές συσκευές για να επιτρέπουν 3D αλληλεπίδραση του χρήστη χωρίς αφή. Η διεπαφή που δημιούργησαν, ονομάζεται MagiTact και παρακολουθεί τις αλλαγές στο μαγνητικό πεδίο γύρω από ένα κινητό τηλέφωνο για να εντοπίσει τις διάφορες χειρονομίες από ένα χέρι που κρατάει ή φοράει ένα μαγνήτη [6].

### **1.2.3 Λειτουργικό Σύστημα Android**

Το Android είναι λειτουργικό σύστημα το οποίο τρέχει τον πυρήνα του λειτουργικού Linux και έχει σχεδιαστεί για κινητές συσκευές με οθόνη αφής όπως smartphones και tablets. Αρχικά αναπτύχθηκε από την Google και αργότερα από την Open Handset Alliance. Το πρώτο κινητό τηλέφωνο με Android κυκλοφόρησε τον Οκτώβριο του 2008. Η Google δημοσίευσε το μεγαλύτερο μέρος του κώδικα του Android υπό τους όρους της Apache License, μιας ελεύθερης άδειας λογισμικού. Αυτό σημαίνει ότι το Android μπορεί να τροποποιηθεί και να διανεμηθεί ελεύθερα από τους κατασκευαστές συσκευών και τους προγραμματιστές. Επιπλέον, το Android διαθέτει μια μεγάλη κοινότητα προγραμματιστών που αναπτύσσουν εφαρμογές οι οποίες επεκτείνουν τη λειτουργικότητα των συσκευών, γραμμένες σε μια προσαρμοσμένη έκδοση της γλώσσας προγραμματισμού Java για Android. Αυτοί οι παράγοντες συνέβαλαν στο να γίνει το Android το πιο ευρέως χρησιμοποιούμενο λειτουργικό σύστημα για smartphone και το λογισμικό που επιλέγουν οι περισσότερες εταιρείες λογισμικού για συσκευές υψηλής τεχνολογίας,

καθώς έχει χαμηλό κόστος, είναι προσαρμόσιμο και ελαφρύ και δεν απαιτείται ανάπτυξη λογισμικού από το μηδέν. Αυτό έχει ως αποτέλεσμα, παρά το γεγονός ότι αρχικά σχεδιάστηκε για smartphones και tablets, να έχουν δημιουργηθεί εφαρμογές Android για τηλεοράσεις, κονσόλες παιχνιδιών και ψηφιακές κάμερες [7].

### **1.2.3.1 Εκδόσεις Android**

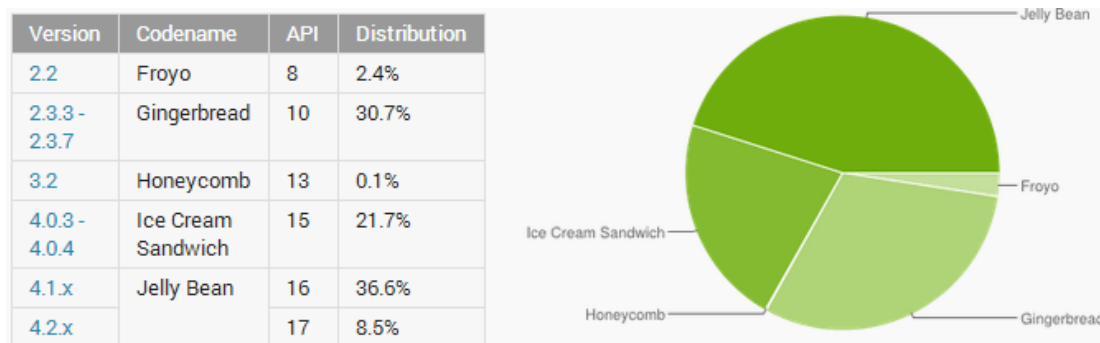
Η πρώτη beta έκδοση Android κυκλοφόρησε στις 5 Νοεμβρίου 2007. Από τότε έχουν κυκλοφορήσει πολλές εκδόσεις και έχουν γίνει ενημερώσεις οι οποίες περιγράφονται σε κάθε έκδοση με έναν αριθμό που ονομάζεται API Level και προσδιορίζει το API framework που υποστηρίζεται. Στο σχήμα 1 παρουσιάζονται όλες οι εκδόσεις Android που έχουν κυκλοφορήσει, με το API Level που

Platform Version	API Level	VERSION_CODE
Android 4.3	18	JELLY_BEAN_MR2
Android 4.2, 4.2.2	17	JELLY_BEAN_MR1
Android 4.1, 4.1.1	16	JELLY_BEAN
Android 4.0.3, 4.0.4	15	ICE_CREAM_SANDWICH_MR1
Android 4.0, 4.0.1, 4.0.2	14	ICE_CREAM_SANDWICH
Android 3.2	13	HONEYCOMB_MR2
Android 3.1.x	12	HONEYCOMB_MR1
Android 3.0.x	11	HONEYCOMB
Android 2.3.4 Android 2.3.3	10	GINGERBREAD_MR1
Android 2.3.2 Android 2.3.1 Android 2.3	9	GINGERBREAD
Android 2.2.x	8	FROYO
Android 2.1.x	7	ECLAIR_MR1
Android 2.0.1	6	ECLAIR_0_1
Android 2.0	5	ECLAIR
Android 1.6	4	DONUT
Android 1.5	3	CUPCAKE
Android 1.1	2	BASE_1_1
Android 1.0	1	BASE

υποστηρίζουν [8].

Σχήμα 1: Οι εκδόσεις Android που έχουν κυκλοφορήσει. (Πηγή : developer.android.com1 )

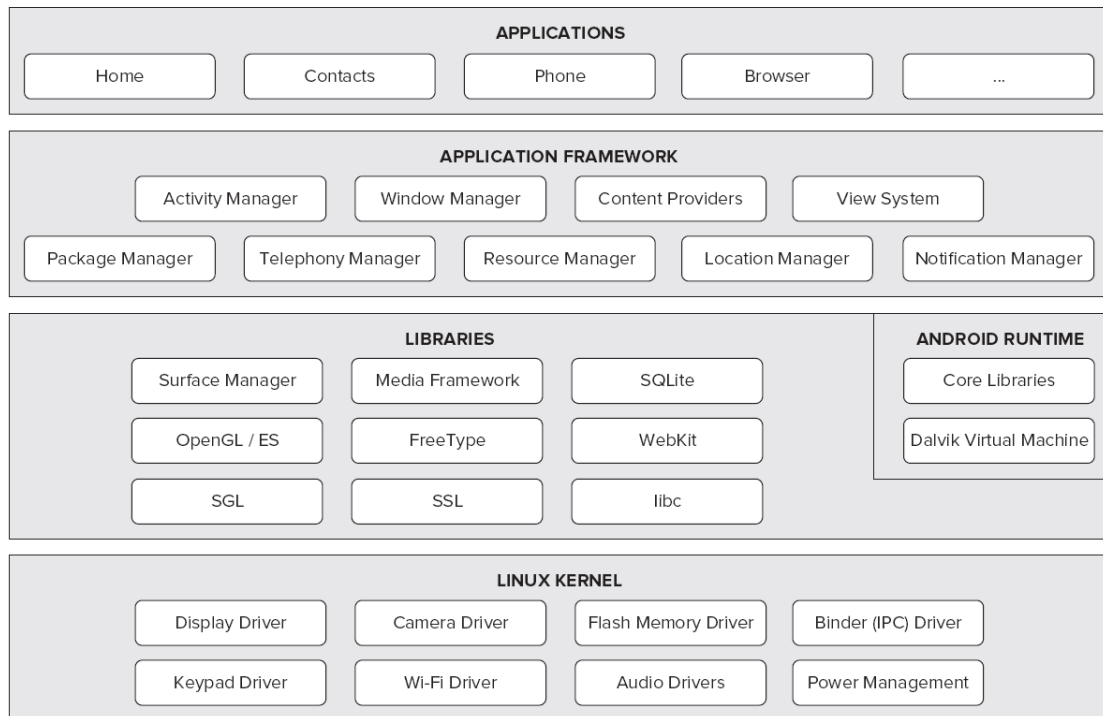
Είναι σημαντικό να αναφερθεί ότι σημαντικές βελτιώσεις του λειτουργικού συστήματος υπήρξαν στην έκδοση 2.2, καθώς υλοποιήθηκαν βελτιστοποιήσεις της ταχύτητας και της μνήμης αυξάνοντας την απόδοση του. Ακόμα, από την έκδοση 2.2 και έπειτα υποστηρίζεται η βιβλιοθήκη OpenGL 2.0 και η υπηρεσία Google Play, οι οποίες είναι απαραίτητες για την χρήση της έκδοσης 2.0 των χαρτών της Google για Android. Τα στατιστικά στοιχεία χρήσης των Android, τα οποία έχουν καταγραφεί μέχρι και την 4η Σεπτεμβρίου 2013, παρουσιάζονται παρακάτω. Οι εκδόσεις που κατέχουν ποσοστό χρήσης μικρότερο από 0,1% δεν φαίνονται στο σχήμα.



Σχήμα 2: Η κατανομή χρήσης των εκδόσεων Android.(Πηγή : developer.android.com2 )

### 1.2.3.2 Αρχιτεκτονική του Android

Το λειτουργικό σύστημα Android αποτελείται από τέσσερα βασικά επίπεδα, τα οποία χωρίζονται σε πέντε τμήματα, όπως φαίνεται στο Σχήμα 3 [9].



- Πυρήνας Linux - Αποτελεί τον πυρήνα στον οποίο είναι βασισμένο το Android. Αυτό το επίπεδο, περιλαμβάνει όλους τους χαμηλού επιπέδου οδηγούς συσκευών για όλα τα hardware εξαρτήματα της συσκευής Android.
- Βιβλιοθήκες (Libraries) - Περιλαμβάνει τον κώδικα ο οποίος παρέχει τα βασικά χαρακτηριστικά του Android. Για παράδειγμα, η βιβλιοθήκη WebKit παρέχει λειτουργικότητα για περιήγηση στον ιστό.
- Android Runtime - Στο ίδιο επίπεδο με τις βιβλιοθήκες, το Android Runtime διαθέτει βιβλιοθήκες πυρήνα που επιτρέπουν στους προγραμματιστές να αναπτύσσουν εφαρμογές Android κάνοντας χρήση της γλώσσας προγραμματισμού Java. Ακόμα, διαθέτει στην εικονική μηχανή Dalvik, η οποία επιτρέπει σε κάθε εφαρμογή να τρέχει σε ξεχωριστή διεργασία. Η Dalvik είναι μια εικονική μηχανή σχεδιασμένη για Android και βελτιστοποιημένη για να λειτουργεί σε κινητές συσκευές με μπαταρία και περιορισμένη μνήμη και επεξεργαστική ισχύ.
- Framework Εφαρμογών - Προσφέρει τις ποικίλες δυνατότητες του λειτουργικού Android στους προγραμματιστές εφαρμογών, ώστε να μπορούν να τις χρησιμοποιήσουν στις εφαρμογές που δημιουργούν.

- Εφαρμογές - Αποτελεί το υψηλότερο επίπεδο του λειτουργικού και διαθέτει όλες τις εφαρμογές που τρέχουν στη συσκευή όπως το τηλέφωνο, την διαχείριση επαφών και τον περιηγητή ιστού.

### **1.2.3.3 Χαρακτηριστικά του Android**

Λόγω του ότι το Android αποτελεί λειτουργικό σύστημα ανοιχτού κώδικα και είναι διαθέσιμο για οποιαδήποτε τροποποίηση, δεν περιλαμβάνει καθορισμένες ρυθμίσεις υλικού και λογισμικού. Παρόλα αυτά, υποστηρίζει από μόνο του τις λειτουργίες που περιγράφονται παρακάτω

- Αποθήκευση Δεδομένων - Χρήση της SQLite, μιας ελαφριάς βάσης δεδομένων για τις ανάγκες αποθήκευσης.
- Συνδεσιμότητα - Υποστηρίζει τεχνολογίες συνδεσιμότητας GSM/EDGE, CDMA, IDEN, EVDO, UMTS, Bluetooth, LTE, WiMAX και Wi-Fi.
- Υπηρεσίες Μηνυμάτων - Υποστηρίζει αποστολή και λήψη SMS και MMS.
- Περιήγηση στον Ιστό - Για την περιήγηση στον ιστό διαθέτει φυλλομετρητή βασισμένο στην ανοιχτή τεχνολογία WebKit.
- Υποστήριξη Πολυμέσων - Παρέχει υποστήριξη για τις ακόλουθες μορφές πολυμέσων : H.263, H.264 (σε 3GP ή MP4container), MPEG-4 SP, AMR, AMR-WB, AAC, HE-AAC, MP3, MIDI, OGG Vorbis, WAV, JPEG, PNG, GIF, BMP.
- Υποστήριξη υλικού - Υποστηρίζει κάμερες στατικής ή κινούμενης εικόνας, οθόνες αφής, GPS, αισθητήρες επιτάχυνσης, μαγνητόμετρα, καθώς και proximity sensors.
- Multitasking - Υποστηρίζει multitasking εφαρμογές.

Οι περισσότερες λειτουργίες του τηλεφώνου τρέχουν σαν εφαρμογές πάνω στο ενδιάμεσο λογισμικό (middleware) που διαθέτει. Το ενδιάμεσο λογισμικό είναι γραμμένο σε Java και C/C++. Οι εφαρμογές που τρέχουν σε Android είναι γραμμένες σε Java και μεταγλωττίζονται σε προσαρμοσμένο πηγαίο κώδικα που ονομάζεται Dalvik Executable (DEX) και εκτελούνται στην εικονική μηχανή Dalvik VM, η οποία είναι σχεδιασμένη για χρήση σε φορητές συσκευές. Οι εφαρμογές επικοινωνούν με τον μηχανισμό binder IPC, ο οποίος παρέχει διαφανή ανταλλαγή μηνυμάτων με

χρήση δεμάτων. Οι εφαρμογές Android αναπτύσσονται στη γλώσσα Java χρησιμοποιώντας το Android Software Development Kit (SDK). Το SDK περιλαμβάνει ένα ολοκληρωμένο σύνολο εργαλείων ανάπτυξης, όπως πρόγραμμα εντοπισμού σφαλμάτων (debugger), βιβλιοθήκες λογισμικού, emulator κινητού τηλεφώνου που βασίζεται σε QEMU, documentation, δείγματα κώδικα και tutorials. Το επίσημο περιβάλλον ανάπτυξης για εφαρμογές Android είναι το Eclipse (IDE) κάνοντας χρήση του Android Development Tools (ADT) plugin [7]. Ακόμα, το περιβάλλον προγραμματισμού Eclipse επιτρέπει την ενσωμάτωση εργαλείων και τεχνικών εξόρυξης δεδομένων χρησιμοποιώντας το API του WEKA. Το WEKA προσφέρει ένα πακέτο κλάσεων για προγραμματιστές που επιτρέπει την διαχείριση και την προεπεξεργασία δεδομένων (preprocessing), αλγόριθμους κατηγοριοποίησης (classification) και συσταδοποίησης (clustering) και την αξιολόγηση των επιδόσεων τους (evaluation). Έτσι, συμπεριλαμβάνοντας το πακέτο των κλάσεων του Weka και μέσω κώδικα Java είναι δυνατή η χρήση τεχνικών εξόρυξης γνώσης σε εφαρμογές Android [10].

### **1.3 Εξόρυξη Δεδομένων και Εξόρυξη Δεδομένων σε Κινητές Συσκευές**

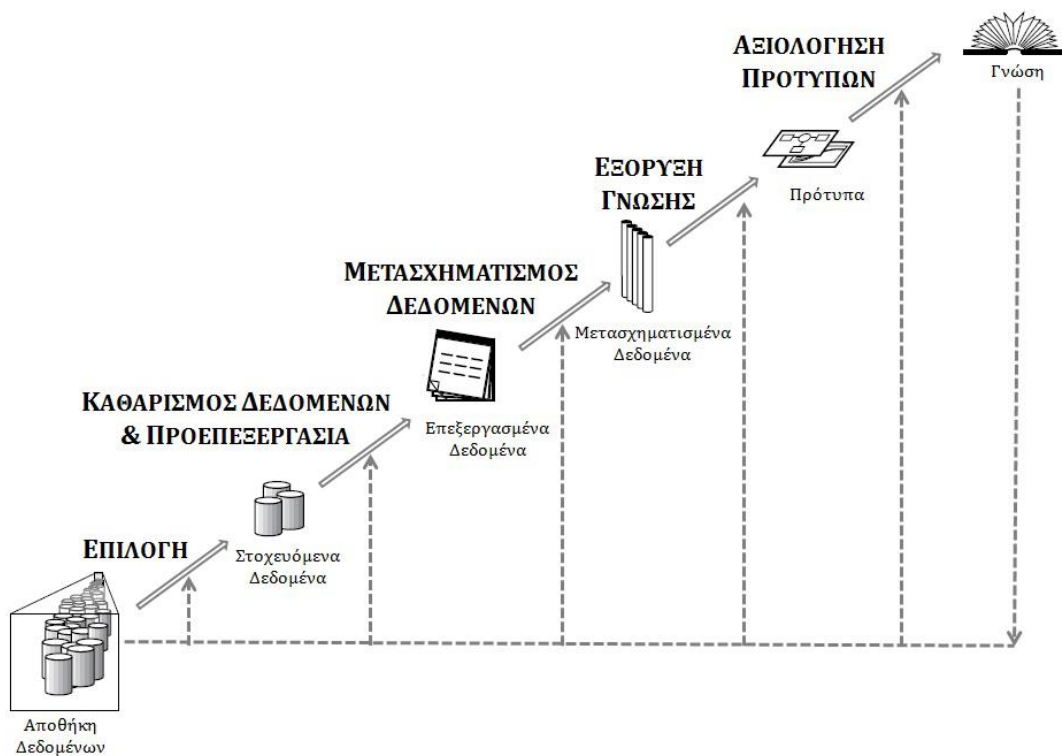
#### **1.3.1 Εξόρυξη Δεδομένων (Data Mining)**

Με τον όρο εξόρυξη δεδομένων αναφερόμαστε στη διαδικασία επιλογής, εξερεύνησης και μοντελοποίησης μεγάλου όγκου δεδομένων, με σκοπό την εξαγωγή συμπερασμάτων τα οποία είναι αρχικά άγνωστα, ώστε να προκύψουν σαφή και χρήσιμα αποτελέσματα για τον χρήστη της βάσης δεδομένων [11].

Η ανακάλυψη γνώσης από μια βάση δεδομένων (KDD - Knowledge Discovery from Database), αναφέρεται σε ολόκληρη τη διαδικασία ανακάλυψης χρήσιμης πληροφορίας από μεγάλα σύνολα δεδομένων. Η εξόρυξη δεδομένων αποτελεί το βήμα της KDD διαδικασίας, στο οποίο οι αλγόριθμοι εκμάθησης εφαρμόζονται στα δεδομένα [11].

Η διαδικασία KDD είναι μια διαλογική και επαναληπτική διαδικασία που περιλαμβάνει τα ακόλουθα βήματα (Σχήμα 4)

1. Ανάπτυξη και κατανόηση της περιοχής της εφαρμογής, της υπάρχουσας γνώσης στον τομέα έρευνας και τους τελικούς στόχους.
2. Ολοκλήρωση των δεδομένων, συνδυάζοντας πολλαπλές πηγές δεδομένων ώστε να καθοριστεί το σύνολο στο οποίο τελικά η διαδικασία εξόρυξης πρόκειται να εφαρμοστεί.
3. Δημιουργία του στόχου – συνόλου δεδομένων, επιλέγοντας το σύνολο δεδομένων (μεταβλητές, δείγματα δεδομένων) στο οποίο η διαδικασία εξόρυξης πρόκειται να εκτελεστεί.
4. Καθαρισμός και προεπεξεργασία δεδομένων, εφαρμόζοντας βασικές διαδικασίες όπως η αφαίρεση θορύβου, η συλλογή των απαραίτητων πληροφοριών για τη διαμόρφωση ή τη μέτρηση του θορύβου, η απόφαση σχετικά με τις στρατηγικές διαχείρισης των ελλειπόντων πεδίων δεδομένων.



Σχήμα 4 Η διαδικασία ανακάλυψης γνώσης από μια βάση δεδομένων (KDD)

5. Μετασχηματισμός των δεδομένων. Τα δεδομένα μετασχηματίζονται σε μορφές κατάλληλες για εξόρυξη, κάνοντας χρήση μεθόδων μείωσης διαστάσεων για τη



μείωση των υπό εξέταση μεταβλητών ή την εύρεση κατάλληλης αντιπροσώπευσης των δεδομένων χωρίς μεταβλητές.

6. Επιλογή των στόχων και των αλγορίθμων εξόρυξης δεδομένων. Σε αυτό το βήμα αποφασίζεται ο στόχος της KDD διαδικασίας, επιλέγοντας τους στόχους εξόρυξης δεδομένων που πρέπει να επιτευχθούν. Ακόμα, επιλέγονται οι μέθοδοι εξόρυξης δεδομένων που θα χρησιμοποιηθούν, λαμβάνοντας υπόψη τις απαιτήσεις και τα γενικά κριτήρια της KDD διαδικασίας.

7. Εξόρυξη δεδομένων, εφαρμόζοντας ευφυείς μεθόδους με στόχο την εύρεση προτύπων γνώσης. Τα πρότυπα μπορεί να είναι κανόνες κατηγοριοποίησης, δέντρα, παλινδρόμηση, συσταδοποίηση, κτλ.

8. Αξιολόγηση των προτύπων. Τα πρότυπα που προέκυψαν από την εξόρυξη δεδομένων αξιολογούνται με κάποια μέτρα, προκειμένου να προσδιοριστούν τα πρότυπα που αντιπροσωπεύουν καλύτερα την εξαγόμενη γνώση.

9. Σταθεροποίηση και παρουσίαση της γνώσης, ενσωματώνοντας τη γνώση στο σύστημα ή απεικονίζοντας την χρησιμοποιώντας τεχνικές αντιπροσώπευσης γνώσης, ώστε να μπορεί να παρουσιαστεί η εξορυγμένη γνώση στον χρήστη.

### **1.3.2 Μέθοδοι Εξόρυξης Δεδομένων**

Η εξόρυξη δεδομένων έχει ως βασικό στόχο την εφαρμογή τεχνικών περιγραφής και πρόβλεψης σε μεγάλα σύνολα δεδομένων. Η πρόβλεψη έχει ως στόχο την πρόβλεψη της συμπεριφοράς κάποιων μεταβλητών που παρουσιάζουν ενδιαφέρον και οι οποίες βασίζονται στην συμπεριφορά άλλων μεταβλητών. Η περιγραφή στοχεύει στην ανακάλυψη προτύπων και αναπαριστά τα δεδομένα μιας πολύπλοκης βάσης δεδομένων με κατανοητό και αξιοποιήσιμο τρόπο. Όλες οι υπάρχουσες μέθοδοι εξόρυξης δεδομένων έχουν ως βασικό στόχο να προσδιορίσουν και να περιγράψουν τα πρότυπα γνώσης που εξάγονται από ένα σύνολο δεδομένων. Οι κυριότερες μέθοδοι εξόρυξης δεδομένων περιγράφονται παρακάτω.

- Κατηγοριοποίηση (Classification). Αποτελεί μια από τις βασικές εργασίες εξόρυξης δεδομένων. Βασίζεται στην εξέταση των χαρακτηριστικών ενός μη-

κατηγοριοποιημένου αντικειμένου, το οποίο με βάση αυτά τα χαρακτηριστικά αντιστοιχίζεται σε κάποια από τις κατηγορίες που έχουν προκαθοριστεί. Η βασική εργασία κατηγοριοποίησης, είναι η δημιουργία ενός μοντέλου το οποίο θα μπορούσε να εφαρμοστεί για να κατηγοριοποιεί μη κατηγοριοποιημένα δεδομένα. Χρησιμοποιεί ένα σύνολο κατηγοριοποιημένων δεδομένων για την εκπαίδευση του μοντέλου και απαιτεί έναν καλά καθορισμένο ορισμό των κατηγοριών. Οι αλγόριθμοι κατηγοριοποίησης διακρίνονται σε δυο βασικές κατηγορίες τεχνικών, τα Δέντρα Απόφασης (Decision Trees) και τα Νευρωνικά Δίκτυα (Neural Networks).

- Συσταδοποίηση (Clustering). Είναι η διαδικασία καταμερισμού ετερογενών δεδομένων σε ένα σύνολο περισσότερων ετερογενών συστάδων. Σε αντίθεση με την κατηγοριοποίηση, η συσταδοποίηση δεν βασίζεται σε προκαθορισμένες κατηγορίες. Οι εγγραφές των δεδομένων ομαδοποιούνται σε σύνολα με βάση την ομοιότητα που παρουσιάζουν μεταξύ τους.

- Κανόνες Συσχέτισης (Association Rules). Οι κανόνες συσχέτισης στοχεύουν στην ανακάλυψη κρυμμένων συσχετίσεων μεταξύ των γνωρισμάτων ενός συνόλου δεδομένων. Παρέχουν έναν συνοπτικό τρόπο για να εκφραστούν οι ενδεχομένως χρήσιμες πληροφορίες και να γίνουν κατανοητές από τους τελικούς χρήστες.

- Πρότυπα ακολουθιών (Sequential Patterns). Η εξόρυξη πρότυπων ακολουθιών αναφέρεται στην ανίχνευση των συχνά εμφανιζόμενων προτύπων σχετικών με τον χρόνο ή άλλες ακολουθίες. Οι περισσότερες έρευνες στα πρότυπα ακολουθιών επικεντρώνονται σε συμβολικά πρότυπα.

- Παλινδρόμηση (Regression). Αναφέρεται στην εκμάθηση μιας λειτουργίας που εκχωρεί τα δεδομένα σε μια μεταβλητή πρόβλεψης, η οποία παίρνει πραγματικές τιμές.

### **1.3.3 Εξόρυξη Δεδομένων σε Κινητές Συσκευές (Mobile Data Mining)**

Ο στόχος της εξόρυξης γνώσης σε κινητές συσκευές (mobile data mining) είναι να παρέχει προηγμένες τεχνικές για την ανάλυση και την παρακολούθηση δεδομένων που καταγράφονται μέσω κινητών συσκευών [12].

Η εξόρυξη δεδομένων σε κινητές συσκευές, έχει να αντιμετωπίσει τα τυπικά ζητήματα της εξόρυξης δεδομένων σε ένα κατανεμημένο περιβάλλον ενώ ταυτόχρονα και τους τεχνολογικούς περιορισμούς ενός κινητού τηλεφώνου, όπως δίκτυα με χαμηλό bandwidth, περιορισμένη χωρητικότητα αποθήκευσης, μικρή ισχύς μπαταρίας, αργούς επεξεργαστές και μικρές οθόνες για την απεικόνιση των αποτελεσμάτων [13].

Ο τομέας της εξόρυξης δεδομένων σε κινητές συσκευές περιλαμβάνει διάφορα σενάρια εφαρμογών, στις οποίες ένα κινητό τηλέφωνο μπορεί να καταγράφει δεδομένα, να τα αναλύει, να αποτελεί τον client ενός απομακρυσμένου server που υλοποιεί την εξόρυξη δεδομένων ή να υλοποιεί τον συνδυασμό αυτών των λειτουργιών. Πιο συγκεκριμένα, μπορούμε να διακρίνουμε τρία βασικά σενάρια για την εξόρυξη δεδομένων σε κινητές συσκευές [12].

1. Η κινητή συσκευή χρησιμοποιείται σαν τερματικό για την πρόσβαση σε έναν απομακρυσμένο server ο οποίος παρέχει υπηρεσίες εξόρυξης δεδομένων. Ο server αναλύει τα δεδομένα που είναι αποθηκευμένα σε μια τοπική ή κατανεμημένη βάση δεδομένων και στέλνει τα αποτελέσματα της διαδικασίας εξόρυξης στην συσκευή για να τα απεικονίσει.

2. Τα δεδομένα που καταγράφηκαν μέσω μιας κινητής συσκευής, αποστέλλονται σε έναν απομακρυσμένο server και αποθηκεύονται σε μια τοπική βάση δεδομένων. Τα δεδομένα αναλύονται κάνοντας χρήση συγκεκριμένων αλγορίθμων εξόρυξης δεδομένων και τα αποτελέσματα χρησιμοποιούνται για να ληφθούν αποφάσεις για κάποιο συγκεκριμένο σκοπό.

3. Η ανάλυση των δεδομένων υλοποιείται στις κινητές συσκευές. Βέβαια λόγω της μειωμένης επεξεργαστικής ισχύος και χωρητικότητας των σύγχρονων συσκευών, δεν είναι πάντα εφικτή η υλοποίηση όλων των αλγορίθμων εξόρυξης δεδομένων σε μια κινητή συσκευή. Ωστόσο, κάποια βήματα της διαδικασίας εξόρυξης, όπως η επιλογή και η προεπεξεργασία των δεδομένων μπορούν να εκτελεστούν σε μια τέτοια συσκευή, καθώς επίσης και μερικοί αλγόριθμοι εξόρυξης γνώσης που δεν απαιτούν μεγάλη επεξεργαστική ισχύ.

Μια πιο σύγχρονη προσέγγιση εξόρυξης δεδομένων σε κινητές συσκευές παρουσιάζουν οι Steinbauer et al.[14], η οποία βασίζεται σε τεχνολογία cloud. Ο στόχος τους είναι να παρέχουν εργαλεία και τεχνικές για την ανάλυση δεδομένων κοινωνικών δικτύων σε σχεδόν πραγματικό χρόνο. Τα δεδομένα παράγονται και καταγράφονται από τους αισθητήρες της κινητής συσκευής και έπειτα αποστέλλονται σε ένα δυναμικό μοντέλο cloud για περαιτέρω επεξεργασία. Στην εφαρμογή της παρούσας πτυχιακής εργασίας, η διαδικασία της εξόρυξης δεδομένων βασίζεται στο τρίτο σενάριο που περιγράφηκε, καθώς η ανάλυση των δεδομένων υλοποιείται στην κινητή συσκευή. Κατά την εκκίνηση της εφαρμογής, γίνεται η εκπαίδευση του μοντέλου με δεδομένα εκπαίδευσης που βρίσκονται αποθηκευμένα στην εφαρμογή. Έπειτα, ξεκινά η καταγραφή των δεδομένων ενώ ταυτόχρονα γίνεται πρόβλεψη της κίνησης για κάθε εγγραφή χρησιμοποιώντας το μοντέλο κατηγοριοποίησης που έχει δημιουργηθεί

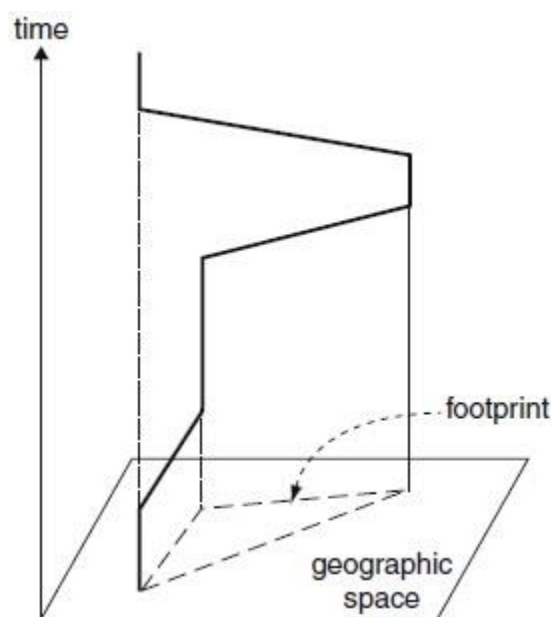
## **Κεφάλαιο 2**

### **2.1 Δεδομένα Κίνησης και Χαρακτηριστικά**

Στις μέρες μας, οι καθημερινές δραστηριότητες των ανθρώπων δημιουργούν συνεχώς ψηφιακά ίχνη μέσω των ασύρματων δικτύων των κινητών τηλεφώνων. Οι τεχνολογίες προσδιορισμού θέσης όπως το GPS, το UMTS και το GSM, συνεχώς βελτιώνονται ως προς την ακρίβεια της θέσης του χρήστη. Αυτό έχει ως αποτέλεσμα, οι ανθρώπινες δραστηριότητες μπορούν να αναγνωριστούν χρησιμοποιώντας τα δεδομένα κίνησης που καταγράφονται μέσω των κινητών τηλεφώνων. Όμως, με ποιο τρόπο μπορεί να καταγραφεί η κίνηση του χρήστη και τι είδους δεδομένα απαιτούνται για τον καλύτερο προσδιορισμό της κίνησης .

#### **2.1.1 Βασικές Αρχές Δεδομένων Κίνησης**

Μια από τις πρώτες προοπτικές ανάλυσης των προτύπων ανθρώπινης δραστηριότητας και της κίνησης στον χωροχρόνο είναι η χρονο-γεωγραφία (time geography). Αναπτύχθηκε από μια ομάδα Σουηδών γεωγράφων το 1970, με σημαντικότερο εκπρόσωπο τον Torsten Hagerstrand. Η χρονο-γεωγραφία έχει χρησιμοποιηθεί από γενιές κοινωνικών επιστημόνων, ιδιαίτερα γεωγράφων και ερευνητών μεταφορών, για την περιγραφή και την ανάλυση των ανθρώπινων δραστηριοτήτων στον χώροχρόνο. Θεωρεί και απεικονίζει τις δραστηριότητες ενός ατόμου σε ένα 24-ωρο ως μία συνεχή χρονική ακολουθία στον γεωγραφικό χώρο. Ο αριθμός και οι τοποθεσίες των καθημερινών δραστηριοτήτων ενός ατόμου περιορίζονται από τον διαθέσιμο χρόνο και από ποικίλες υποχρεωτικές δραστηριότητες (π.χ. εργασία). Η τροχιά της κίνησης του χρήστη θεωρείται μια διαδρομή στον χωρόχρονο και αναπαρίσταται στον τρισδιάστατο χώρο χρησιμοποιώντας τον οριζόντιο άξονα για την αναπαράσταση του γεωγραφικού χώρου και τον κάθετο άξονα για τον χρόνο. Η αναπαράσταση αυτή ονομάζεται κύβος χωροχρόνου (space-time cube) και απεικονίζεται στο Σχήμα 5 [15].



Σχήμα 5: Απεικόνιση του space-time cube (Πηγή : Giannotti et al.,2008).

Η γραμμή αναπαριστά τις κινήσεις μιας οντότητας, για παράδειγμα, ας θεωρήσουμε έναν εργαζόμενο, ο οποίος αρχικά ήταν στο σπίτι, μετά πήγε στη δουλειά του και έμεινε εκεί για ένα χρονικό διάστημα, μετά πήγε στο σουπερμάρκετ για ψώνια και αφού έμεινε κι εκεί για λίγο, γύρισε σπίτι. Οι κάθετες γραμμές ερμηνεύονται σαν

την παραμονή του χρήστη σε συγκεκριμένο μέρος (σπίτι, εργασία, σουπερμάρκετ). Τα κεκλιμένα τμήματα της γραμμής υποδεικνύουν κινήσεις. Όσο πιο αργή είναι η κίνηση, τόσο πιο απότομη είναι η γραμμή. Η ευθεία γραμμή δείχνει ότι το άτομο κινείται με σταθερή ταχύτητα, η οποία είναι συνήθως μια προσέγγιση της πραγματικής συμπεριφοράς. Η απεικόνιση αυτής της αναπαράστασης σε χάρτη είναι το αποτύπωμα της διαδρομής που ακολούθησε ο χρήστης.

### **2.1.2 Χαρακτηριστικά Κίνησης**

Για την αναγνώριση της δραστηριότητας του χρήστη πρέπει να εξεταστούν πολλοί παράγοντες της κίνησης. Αρχικά, τα χαρακτηριστικά της κάθε κίνησης όπως για παράδειγμα η ταχύτητα και η κατεύθυνση, η τροχιά της κίνησης που ακολουθεί ο χρήστης αλλά και τα χαρακτηριστικά του χρήστη. Παράλληλα, πολλά συμπεράσματα μπορούν να εξαχθούν για την φύση της κίνησης παρατηρώντας το περιβάλλον της και τα συμβάντα που εξελίσσονται σε αυτό [16].

Η έννοια της κίνησης μπορεί να οριστεί σαν την μεταβολή της φυσικής θέσης μιας οντότητας σε σχέση με ένα σύστημα αναφοράς στο οποίο μπορεί να προσδιοριστεί η θέση της οντότητας. Το σύστημα αναφοράς συνήθως είναι ένας γεωγραφικός χώρος.

Η κίνηση μιας οντότητας αποτελείται από τροχιές, δηλαδή από τις διαδρομές που ακολουθεί η οντότητα στον χώρο κατά την κίνηση της. Μια διαδρομή δεν δημιουργείται ποτέ στιγμιαία, αντίθετα απαιτεί ένα χρονικό διάστημα. Για αυτό τον λόγο, η τροχιά στην οποία κινείται μια οντότητα πρέπει να συνδέεται με τον χρόνο και μπορεί να περιγραφεί από μια σειρά εγγραφών της μορφής (χρονική στιγμή, θέση).

#### **Χώρος :**

Ως χώρος μπορεί να θεωρηθεί το σύνολο από τοποθεσίες και μέρη, με την ιδιότητα να μπορεί να οριστεί η απόσταση μεταξύ τους. Για τη διάκριση των διαφορετικών τοποθεσιών στον χώρο πρέπει να θεωρηθεί ένα σύστημα αναφοράς, όπως ένα σύστημα συντεταγμένων. Ανάλογα με τις απαιτήσεις του προβλήματος, ο χώρος

μπορεί να θεωρηθεί ως δισδιάστατος ή τρισδιάστατος, δηλαδή ένα σημείο μπορεί να περιγραφεί από δύο ή τρεις συντεταγμένες αντίστοιχα. Ο φυσικός χώρος είναι συνεχής, το οποίο σημαίνει ότι αποτελείται από άπειρο αριθμό από τοποθεσίες. Από την άλλη πλευρά, κάποιες φορές μπορεί να είναι χρήσιμο να θεωρούμε τον χώρο διακριτό ή από πεπερασμένο αριθμό από τοποθεσίες. Σε κάποιες περιπτώσεις, η διακριτοποίηση του χώρου μπορεί να είναι απαραίτητη, όταν οι θέσεις των οντοτήτων δεν μπορούν να προσδιοριστούν με ακρίβεια αλλά με την έννοια ευρύτερης περιοχής όπως για παράδειγμα περιοχές δικτύου κινητής τηλεφωνίας, περιοχές μιας πόλης ή χώρες.

Ο χώρος μπορεί να δομηθεί με διάφορους τρόπους για να επιτευχθεί καλύτερη καταγραφή της κίνησης μιας οντότητας. Ένας τρόπος είναι η ιεραρχική διαίρεση, για παράδειγμα μια χώρα χωρίζεται σε επαρχίες, οι επαρχίες σε δήμους και οι κοινότητες σε επαρχίες. Περιοχές μπορούν να προκύψουν ακόμα από γεωγραφική διαίρεση, δηλαδή δημιουργώντας κελιά στον χώρο με συγκεκριμένο μέγεθος (π.χ. 1 km<sup>2</sup>).

Επίσης, ένας πολύ κοινός τρόπος δόμησης του φυσικού χώρου είναι με βάση το οδικό δίκτυο. Γενικά, οι πιθανοί τρόποι με τους οποίους μπορεί να προσδιοριστεί η θέση μιας οντότητας στον γεωγραφικό χώρο περιγράφονται παρακάτω.

- Αναφορά με βάση συντεταγμένες. Οι θέσεις ορίζονται με αριθμούς που αντιστοιχούν στη γραμμική ή γωνιακή απόσταση από καθορισμένους άξονες ή γωνίες.
- Αναφορά με βάση τη διαίρεση του χώρου. Αναφέρεται σε περιοχές μιας γεωμετρική ή σημασιολογική διαίρεση του χώρου, ενδεχομένως ιεραρχική.
- Γραμμική αναφορά. Αναφέρονται σε σχετικές θέσεις κατά μήκος γραμμικών αντικειμένων όπως δρόμοι, ποτάμια, αγωγοί. Για παράδειγμα, ονόματα δρόμων καθώς και αριθμούς σπιτιών ή κωδικούς δρόμων και αποστάσεις από κάποιο από τα άκρα.

## **Χρόνος:**

Από μαθηματική άποψη, ο χρόνος είναι μια συνεχής σειρά στοιχείων με γραμμική διάταξη και αποστάσεις μεταξύ των στοιχείων, όπου τα στοιχεία είναι στιγμές ή θέσεις στο χρόνο. Όπως συμβαίνει και με τον χώρο, απαιτείται ένα σύστημα αναφοράς για τον προσδιορισμό των χρονικών στιγμών στα δεδομένα. Στις περισσότερες περιπτώσεις, ως σύστημα αναφοράς χρησιμοποιείται το Γρηγοριανό ημερολόγιο (Gregorian calendar), και η διαίρεση του χρόνου σε ημέρες, μήνες, ώρες, λεπτά και δευτερόλεπτα. Η ώρα της ημέρας μπορεί να προσδιοριστεί με τη ζώνη ώρας της περιοχής στην οποία καταγράφονται τα δεδομένα ή με βάση την ζώνη ώρας του Greenwich (GMT). Υπάρχουν βέβαια περιπτώσεις, στις οποίες τα δεδομένα αναφέρονται σε σχετικές χρονικές στιγμές, για παράδειγμα ο χρόνος που έχει παρέλθει από την έναρξη μιας διαδικασίας ή της παρατήρησης.

Ο πραγματικός χρόνος περιλαμβάνει επίσης χρονικούς κύκλους που προκύπτουν από την καθημερινή και την ετήσια περιστροφή της γης. Αυτοί οι φυσικοί χρονικοί κύκλοι γίνονται αντιληπτοί με τη φυσική ροή του χρόνου, για παράδειγμα οι ημερομηνίες επαναλαμβάνονται κάθε χρόνο και οι ώρες κάθε μέρα. Εκτός από τους φυσικούς χρονικούς κύκλους, υπάρχουν και αυτοί που σχετίζονται με τις δραστηριότητες των ανθρώπων, για παράδειγμα μια δραστηριότητα μπορεί να γίνεται μια φορά τη μέρα ή τη βδομάδα. Για τη ανάλυση των κινήσεων, είναι πολύ σημαντικό να είναι γνωστοί οι χρονικοί κύκλοι που σχετίζονται με την κίνηση. Οι ιδιότητες κάθε χρονικού κύκλου μπορεί να διαφέρουν και αυτές οι διαφορές μπορεί να έχουν σημαντική επίδραση στις κινήσεις. Για παράδειγμα, οι κινήσεις των ανθρώπων τις καθημερινές διαφέρουν από τα Σαββατοκύριακα.

Όμως, η ετερογένεια των ιδιοτήτων κάθε χρονικής στιγμής δεν μπορεί να εκφραστεί σαφώς στα δεδομένα και για αυτό τον λόγο δεν είναι δυνατό να ληφθούν υπόψη αυτόματα στην ανάλυση των δεδομένων. Η ανάλυση των ιδιοτήτων του χώρου εξαρτάται σημαντικά από την ικανότητα του αναλυτή να χρησιμοποιήσει τις γνώσεις του και είναι απαραίτητο οι μέθοδοι και τα εργαλεία που θα χρησιμοποιήσει για την ανάλυση τους να του δίνουν αυτή τη δυνατότητα.

Κινούμενες οντότητες και τα χαρακτηριστικά τους



Εκτός από τα χαρακτηριστικά της κίνησης που αναφέρθηκαν, οι οντότητες που κινούνται έχουν τα δικά τους χαρακτηριστικά, τα οποία επηρεάζουν την κίνηση τους και πρέπει να ληφθούν υπόψη κατά την ανάλυση των δεδομένων κίνησης. Έτσι, οι κινήσεις των ανθρώπων μπορεί να επηρεάζονται σημαντικά από το επάγγελμα, την ηλικία, την κατάσταση υγείας, την οικογενειακή κατάσταση και από άλλα χαρακτηριστικά. Ακόμα, μπορεί να αναλυθεί ο σκοπός της κίνησης ενός ανθρώπου, καθώς μπορεί έτσι να καθοριστεί η διαδρομή που θα ακολουθήσει και η ταχύτητα κίνησης του. Τα χαρακτηριστικά της κίνησης μπορεί επίσης να εξαρτώνται από τις δραστηριότητες των ανθρώπων ενώ κινούνται. Για παράδειγμα, η κίνηση ενός ανθρώπου σε ένα κατάστημα, διαφέρει από την κίνηση στον δρόμο. Επίσης, τα χαρακτηριστικά της κίνησης είναι διαφορετικά όταν ένας άνθρωπος απλά περπατάει από όταν περπατάει ενώ ταυτόχρονα μιλάει στο τηλέφωνο.

#### **Σχετικά φαινόμενα και συμβάντα :**

Οι κινήσεις που εμφανίζονται σε ένα περιβάλλον, επηρεάζονται από τα διάφορα συμβάντα αυτού του περιβάλλοντος. Οι κινήσεις των ανθρώπων επηρεάζονται από το κλίμα και τις καιρικές συνθήκες, από τις αθλητικές και πολιτιστικές εκδηλώσεις, από τις τρέχουσες νομοθετικές ρυθμίσεις, από τα τέλη διοδίων και τις τιμές των καυσίμων, από τα τροχαία ατυχήματα και ούτω καθεξής. Για την ανίχνευση τέτοιων επιρροών κατά την ανάλυση των δεδομένων, ο αναλυτής πρέπει να χρησιμοποιήσει πρόσθετα στοιχεία και γνώσεις.

#### **Χαρακτηριστικά της κίνησης:**

Όσον αφορά στα χαρακτηριστικά της κίνησης που μελετούνται, μπορούν να διακριθούν σε δυο κατηγορίες ανάλογα με το αν η κίνηση εξετάζεται σε επίπεδο στιγμιότυπου ή σε επίπεδο τροχιάς κίνησης. Τα χαρακτηριστικά της κίνησης που μπορούν να καταγραφούν σε μια συγκεκριμένη χρονική στιγμή και να περιγράψουν ένα στιγμιότυπο είναι τα ακόλουθα :

- Χρονοσφραγίδα (timestamp) της συγκεκριμένης χρονικής στιγμής
- Θέση της οντότητας στον χώρο
- Κατεύθυνση

- Ταχύτητα
- Αλλαγή κατεύθυνσης
- Επιτάχυνση (αλλαγή ταχύτητας)
- Συνολικός χρόνος κίνησης
- Συνολική απόσταση που έχει διανυθεί

Ωστόσο, όταν η κίνηση εξετάζεται σε μεγαλύτερο χρονικό παράθυρο παρατηρώντας την τροχιά της κίνησης, τα χαρακτηριστικά της διαφοροποιούνται καθώς αναφέρονται σε ένα χρονικό διάστημα της κίνησης. Τα γνωρίσματα που περιγράφουν την κίνηση σε ένα χρονικό διάστημα παρατίθενται παρακάτω :

- Γεωμετρικό σχήμα της κίνησης στον χώρο
- Απόσταση που έχει διανυθεί στον χώρο
- Χρονική διάρκεια της κίνησης
- Διάνυσμα της κίνησης (από την αρχική στην τελική θέση)
- Μέση και μέγιστη τιμή της ταχύτητας
- Δυναμική συμπεριφορά της ταχύτητας (σταθερή ταχύτητα, επιτάχυνση, επιβράδυνση, μηδενική ταχύτητα)
- Δυναμική συμπεριφορά της κατεύθυνσης (ευθεία, καμπυλόγραμμη, κυκλική κίνηση)

Τρόποι καταγραφής κίνησης

Η καταγραφή της τροχιάς μιας κινούμενης οντότητας μπορεί να γίνει με ποικίλους τρόπους, ανάλογα με τον τρόπο παρατήρησης της κίνησης :

- Χρονική. Η καταγραφή της θέσης της οντότητας γίνεται ανά τακτά χρονικά διαστήματα (π.χ. ανά λεπτό).
- Με βάση την μεταβολή θέσης. Γίνεται καταγραφή όταν η θέση της οντότητας διαφέρει από την προηγούμενη.

- Με βάση τη θέση κοντά σε συγκεκριμένη τοποθεσία. Καταγραφή της θέσης όταν η οντότητα πλησιάζει σε μια συγκεκριμένη τοποθεσία.
- Με βάση κάποιο συμβάν. Η θέση της οντότητας καταγράφεται όταν συμβεί κάποιο συγκεκριμένο γεγονός, συνήθως αφορά δραστηριότητες των ίδιων οντοτήτων.

## **2.2 Τεχνικές Αναγνώρισης Κίνησης**

Η αναγνώριση της ανθρώπινης δραστηριότητας έχει συγκεντρώσει μεγάλο ενδιαφέρον και έχουν πραγματοποιηθεί πολλές έρευνες για τον εντοπισμό πληροφοριών, χρήσιμων για την αναγνώριση κίνησης. Η αναγνώριση κίνησης μπορεί εύκολα να υλοποιηθεί σε κινητά τηλέφωνα χρησιμοποιώντας τους ενσωματωμένους αισθητήρες της συσκευής και γι' αυτό τον λόγο έχουν αναπτυχθεί πολλές εφαρμογές για smartphone. Οι κυριότερες τεχνικές εξόρυξης δεδομένων που χρησιμοποιούν οι υπάρχουσες εφαρμογές για την αναγνώριση της ανθρώπινης δραστηριότητας αναλύονται παρακάτω.

### **2.2.1 Κατηγοριοποίηση (Classification)**

Οι περισσότεροι ερευνητές που χρησιμοποιούν την τεχνική της κατηγοριοποίησης για να ανιχνεύσουν την δραστηριότητα του χρήστη, καταγράφουν χαρακτηριστικά (attributes) της κίνησης τα οποία μπορούν να διακριθούν σε τρεις βασικές κατηγορίες. Παρακάτω περιγράφονται τα χαρακτηριστικά που ανήκουν σε κάθε κατηγορία.

- Μεγέθους. Αυτή η κατηγορία αναφέρεται σε χαρακτηριστικά που βασίζονται σε τιμές που καταγράφουν οι αισθητήρες της συσκευής. Συνήθως οι τιμές είναι συντεταγμένες, μέση τιμή, τυπική απόκλιση, ελάχιστη και μέγιστη τιμή, μέση απόλυτη διαφορά και άλλες.
- Συχνότητας. Είναι τα χαρακτηριστικά που βασίζονται στις τιμές συχνότητας των αισθητήρων. Το πιο συνηθισμένο μέγεθος είναι από τον μετασχηματισμό Fourier

(FFT). Άλλα χαρακτηριστικά αυτής της κατηγορίας είναι η εντροπία της συχνότητας, η μέγιστη συχνότητα, η μέση τιμή, η τυπική απόκλιση και η ενέργεια του μετασχηματισμού Fourier.

- Συσχέτισης. Πολλοί ερευνητές χρησιμοποιούν ως χαρακτηριστικά, τις συσχετίσεις μεταξύ των υπολοίπων χαρακτηριστικών.

Όσον αφορά στον αλγόριθμο κατηγοριοποίησης που μπορούμε να χρησιμοποιήσουμε για να κατηγοριοποιήσουμε διαφορετικές δραστηριότητες του χρήστη, η επιλογή του εξαρτάται από την επεξεργαστική ισχύ του συστήματος που θα εκτελεστεί ο αλγόριθμος. Δηλαδή αν ο αλγόριθμος εκτελεστεί σε server είναι προφανές ότι η επεξεργαστική ισχύς του μηχανήματος θα είναι αρκετά πιο μεγάλη από την επεξεργαστική ισχύ ενός smartphone.

Οι περισσότεροι ερευνητές χρησιμοποιούν αλγόριθμους κατηγοριοποίησης με επίβλεψη. Αυτοί οι αλγόριθμοι εκπαιδεύονται με κατηγοριοποιημένα δείγματα για τη δημιουργία μοντέλου κατηγοριοποίησης, το οποίο στη συνέχεια θα χρησιμοποιηθεί για την κατηγοριοποίηση των δεδομένων εισόδου. Οι πιο συνηθισμένοι αλγόριθμοι που χρησιμοποιούνται για αναγνώριση κίνησης είναι τα Δέντρα απόφασης (Decision Trees), ο κ – κοντινότερος γείτονας (kNN), ο Naive Bayes, ο SVM και τα Νευρωνικά δίκτυα. Όμως, η κατηγοριοποίηση με επίβλεψη χρειάζεται μεγάλη επεξεργαστική ισχύ για να δημιουργήσει μοντέλο από δεδομένα εκπαίδευσης και για αυτόν τον λόγο οι περισσότερες υλοποιήσεις έχουν γίνει σε servers. Για αυτό τον λόγο μερικοί ερευνητές, δημιουργούν το μοντέλο κατηγοριοποίησης εκτελώντας τον αλγόριθμο σε ένα μηχάνημα και έπειτα μεταφέρουν το μοντέλο στο τηλέφωνο για την κατηγοριοποίηση των δεδομένων εισόδου.

Ακόμα, η αξιολόγηση των αλγορίθμων κατηγοριοποίησης είναι πολύ σημαντική, καθώς δείχνει ποιος αλγόριθμος αποδίδει καλύτερα. Οι δημοφιλέστερες μέθοδοι αξιολόγησης που χρησιμοποιούνται είναι οι n-fold cross validation (συνήθως 10-fold) και τα μέτρα precision, recall, F-measure και accuracy [17].

### **2.2.2 Συσταδοποίηση τροχιών κίνησης (Spatemporal/Trajectory Clustering)**

Η συσταδοποίηση των τροχιών των κινούμενων αντικειμένων, παρουσιάζει μεγάλο ενδιαφέρον στην ερευνητική κοινότητα και για αυτόν τον λόγο υπάρχει μεγάλος αριθμός ερευνών που εξετάζουν διαφορετικές τεχνικές συσταδοποίησης με σκοπό τη βελτίωση των αποτελεσμάτων τους. Η τροχιά (trajectory) κίνησης ενός χρήστη, αναφέρεται σε μια χρονική ακολουθία τοποθεσιών με τη σειρά που ο χρήστης τις επισκέφτηκε [18]. Η συσταδοποίηση των τροχιών έχει ως στόχο την ομαδοποίηση των τροχιών σε συστάδες όμοιων τροχιών [19]. Οι τροχιές των κινούμενων οντοτήτων πολύ συχνά περιέχουν επαναλαμβανόμενα μοτίβα κίνησης από τα οποία με την τεχνική της συσταδοποίησης είναι δυνατόν να ανιχνευτούν όμοια μοτίβα με σκοπό την αναγνώριση της δραστηριότητας [20].

Οι μέθοδοι που υπάρχουν μέχρι σήμερα στη βιβλιογραφία ακολουθούν δυο βασικές προσεγγίσεις. Η πρώτη προσέγγιση έχει ως στόχο να βρει ένα μέτρο ομοιότητας μεταξύ των τροχιών. Ορίζοντας την απόσταση μεταξύ αντικειμένων καθορίζεται ποιες τροχιές πρέπει να βρίσκονται στην ίδια συστάδα και στη συνέχεια ποιες είναι οι συστάδες που πρέπει να ανιχνευτούν. Ένας βασικός τρόπος για οριστεί η απόσταση είναι να εξεταστούν τροχιές χρηστών που είναι παρόμοιες, για παράδειγμα τροχιές στις οποίες κάθε χρονική στιγμή οι χρήστες βρίσκονται σχεδόν στην ίδια τοποθεσία. Μια απλή προσέγγιση για τη μοντελοποίηση αυτής της σύγκρισης είναι να θεωρηθούν οι τροχιές διανύσματα των συντεταγμένων και να γίνει σύγκριση των διανυσμάτων χρησιμοποιώντας ένα μέτρο απόστασης, όπως η Ευκλείδεια απόσταση [16].

Η δεύτερη μέθοδος συσταδοποίησης των τροχιών κίνησης, δεν κάνει χρήση μέτρων ομοιότητας αλλά αντιμετωπίζει κάθε τροχιά ως ένα αντικείμενο και αναζητά ομάδες τροχιών που κινούνται μαζί [16]. Μια αποτελεσματική προσέγγιση χρησιμοποιούν οι Lee et al.[21], με τον αλγόριθμο TRACCLUS. Αρχικά χωρίζουν τις τροχιές σε μια σειρά υποσυνόλων κάνοντας χρήση του MDL (Minimum Description Length) και έπειτα δημιουργούν συστάδες όμοιων υποσυνόλων λαμβάνοντας υπόψη την πυκνότητα τους. Για κάθε συστάδα, η τροχιά που περιγράφει την συνολική κίνηση των υποσυνόλων που ανήκουν στην ίδια συστάδα, είναι η τροχιά που αντιπροσωπεύει τη συστάδα. Αντίθετα, οι Giannotti et al.[18] προτείνουν την έννοια

των μοτίβων κίνησης και παρουσιάζουν έναν αλγόριθμο για την ανίχνευση τους στις τροχιές. Τα μοτίβα κίνησης αντιπροσωπεύουν σύνολα τοποθεσιών ενδιαφέροντος που σχετίζονται χρονικά, και μπορούν να είναι προκαθορισμένα από τον χρήστη ή να ανακαλυφθούν με κάποιο αλγόριθμο συσταδοποίησης βασισμένο στην πυκνότητα των συστάδων. Ταυτόχρονα, οι Sung, Feldman και Rus[20], παρουσιάζουν έναν αλγόριθμο για την εξαγωγή μοτίβων κίνησης από τις τροχιές και χρησιμοποιούν τα αποτελέσματα για να δημιουργήσουν ένα Μαρκοβιανό μοντέλο το οποίο θα κάνει πρόβλεψη της κίνησης για την κάθε οντότητα.

### **2.2.3 Conditional Random Fields (CRF)**

Μια διαφορετική προσέγγιση για την αναγνώριση δραστηριότητας παρουσιάζουν οι Liao et al. [22] κάνοντας χρήση ιεραρχικών μοντέλων CRF σε δεδομένα GPS για την εξαγωγή σημαντικών και δραστηριοτήτων του χρήστη.

Αρχικά, χωρίζουν τα δεδομένα GPS ομαδοποιώντας τα με βάση τη σχέση τους στον χώρο και κάθε GPS στίγμα το αντιστοιχούν στον πιο κοντινό δρόμο. Για να δημιουργήσουν μια σωστή συσχέτιση μεταξύ των τοποθεσιών, δημιουργούν ένα μοντέλο CRF με βάση τη χωρική σχέση μεταξύ τους.

Αφού χωρίσουν τις τοποθεσίες, ο αλγόριθμος τους κάνει εκτίμηση για την δραστηριότητα που εκτελείται σε κάθε σύνολο και ξεχωρίζει τις σημαντικές τοποθεσίες κάθε χρήστη. Για να επιτευχθεί αυτό, δημιουργεί ένα νέο μοντέλο CRF το οποίο περιέχει έναν κρυμμένο κόμβο που αντιστοιχεί στη δραστηριότητα σε κάθε σύνολο που δημιουργήθηκε από τις καταγεγραμμένες GPS συντεταγμένες. Κάθε κόμβος δραστηριότητας είναι συνδεδεμένος με μερικά χαρακτηριστικά που προκύπτουν από τις πληροφορίες που προέρχονται από τη διαίρεση των δεδομένων, όπως ημερομηνία και ώρα, μέση ταχύτητα, πληροφορίες για τα σημεία ενδιαφέροντος που βρίσκονται κοντά. Έπειτα, στο μοντέλο CRF που δημιουργήθηκε, εφαρμόζει τον αλγόριθμο Naive Bayes για να ανιχνεύσει την δραστηριότητα του χρήστη σε κάθε τοποθεσία.

Στη συνέχεια, χρησιμοποιεί τις ακολουθίες στις οποίες ο χρήστης πραγματοποιεί τον ίδιο τύπο κίνησης για να αναζητήσει τις σημαντικές τοποθεσίες του χρήστη. Αυτό γίνεται κατηγοριοποιώντας ξεχωριστές δραστηριότητες στην ίδια ακολουθία, με βάση το αν ανήκουν σε μια συγκεκριμένη τοποθεσία και θεωρώντας ότι όλες οι εγγραφές στις οποίες παρατηρείται μια συγκεκριμένη δραστηριότητα αποτελούν μια σημαντική τοποθεσία.

Η προσέγγιση τους παρουσιάζει ενδιαφέρον, καθώς ο αλγόριθμος που χρησιμοποιούν έχει αρκετά μεγάλο ποσοστό ακρίβειας (86%), ενώ η ταυτόχρονη εκτίμηση δραστηριότητας και τοποθεσίας αυξάνει την ποιότητα του αποτελέσματος. Αυτό συμβαίνει γιατί η κάθε θέση συνδέει τις δραστηριότητες που συμβαίνουν στην χωρική περιοχή της και οι δραστηριότητες αναγνωρίζονται με πιο συνεπή τρόπο [21].

## **Κεφάλαιο 3**

### **3.1 Πρόβλεψη επόμενης τοποθεσίας**

Η πρόβλεψη της επόμενης θέσης είναι ένα ιδιαίτερο πρόβλημα, όπου η πρόκληση συνίσταται στην πρόβλεψη της επόμενης θέσης ενός χρήστη κινητού, δεδομένης της τρέχουσας τοποθεσίας του [23, 24]. Τα περισσότερα υπάρχοντα μοντέλα εργασίας προβλέπουν την επόμενη θέση ως πρόβλημα ταξινόμησης, όπου τα χωρικά και χρονικά δεδομένα χρησιμοποιούνται για εκπαίδευση.

Ωστόσο, ζητήματα όπως η ενσωμάτωση άλλων πλούσιων δεδομένων, που διατίθενται σε smartphone σήμερα, όπως επιταχυνσιόμετρο, bluetooth κλήσεις και sms logs δεν έχουν διερευνηθεί σοβαρά. Επιπλέον, οι περισσότερες υπάρχουσες προσεγγίσεις επικεντρώνονται κυρίως στο πρόβλημα ταξινόμησης, ενώ άλλα προβλήματα, και πιο συγκεκριμένα προβλήματα που σχετίζονται με αλλαγές συμπεριφοράς του χρήστη, το απόρρητο, η διαχείριση δεδομένων και επεκτασιμότητα δεν έχουν διερευνηθεί σε βάθος. Για την αντιμετώπιση αυτών των ζητημάτων σύμφωνα με τους αρθρογράφους έχει προταθεί το NextLocation - ένα νέο ολοκληρωμένο πλαίσιο για το πρόβλημα πρόβλεψης της επόμενης θέσης - που

προβλέπει την επόμενη τοποθεσία χρησιμοποιώντας μόνο τρέχουσες τοποθεσίες και δεδομένα περιβάλλοντος για κάθε χρήστη κινητού τηλεφώνου. Το NextLocation μαθαίνει ένα μοντέλο ταξινόμησης, το οποίο ενσωματώνει δεδομένα μετάδοσης για την πρόβλεψη της επόμενης θέσης με αυξητικό τρόπο. Επιτρέπει μεγαλύτερη εξατομίκευση και προστασία της ιδιωτικής ζωής ενώ φέρνει όλη τη διαδικασία μάθησης στην κινητή συσκευή. Επιπλέον, εκτός από τις χωρικές και χρονικές πληροφορίες, η προτεινόμενη προσέγγιση συνδυάζει άλλες πληροφορίες περιβάλλοντος που είναι διαθέσιμες στο κινητό. Τα κύρια πλεονεκτήματα της NextLocation για την πρόβλεψη της επόμενης θέσης είναι:

- Η διατήρηση της ιδιωτικής ζωής καθώς αυτός είναι βασικός χρήστης πρέπει να έχει τον έλεγχο των προσωπικών δεδομένων κινητής τηλεφωνίας του [24, 25]. Χρησιμοποιώντας το NextLocation για εξόρυξη δεδομένων κινητής τηλεφωνίας, τα δεδομένα προσωπικού κινητού τηλεφώνου δεν κοινοποιούνται σε εξωτερικό μέρος χωρίς άδεια.

- Μειωμένη επιβάρυνση επικοινωνίας όσον αφορά το εύρος ζώνης καθώς και την εξάντληση της μπαταρίας, καθώς η τοπική επεξεργασία είναι συνήθως λιγότερο δαπανηρή από την ασύρματη μεταφορά δεδομένων [26].

- Η δυναμική αντί της στατικής δημιουργίας μοντέλων διευκολύνει την προσαρμογή του μοντέλου, ώστε να επιφέρει ενημερωμένη συμπεριφορά των χρηστών.

- Επιτρέπει διαδικτυακή εκτίμηση εξατομικευμένης ακρίβειας επόμενης θέσης.

- Ενεργοποιεί ένα εναλλακτικό επιχειρηματικό μοντέλο όπου οι πάροχοι διαφημίσεων μπορούν να προωθήσουν περιεχόμενο που σχετίζεται με μια συγκεκριμένη τοποθεσία και ο χρήστης θα το λάβει μόνο όταν πρόκειται να την επισκεφθεί

### **3.2 Σχετικές πληροφορίες**

Η πρόβλεψη τοποθεσίας προϋποθέτει ότι είναι διαθέσιμες παρατηρήσεις αισθητήρων κινητής τηλεφωνίας από ασύρματο τοπικό δίκτυο (Wi-Fi), το Παγκόσμιο Σύστημα Κινητών Επικοινωνιών (GSM), και το Παγκόσμιο Σύστημα Εντοπισμού



(GPS). Η εργασία πρόβλεψης συνίσταται στη χρήση τέτοιων δεδομένων που γνωρίζουν και κατανοούν την τρέχουσα τοποθεσία του χρήστη. Η έρευνα σχετικά με τη συμπεριφορά των χρηστών, μπορεί να προσφέρει πρόσθετη αξία σε διαφορετικούς τομείς, όπως διαφήμιση [27], κατανομή πόρων [28] και ανακούφιση από καταστροφές [29]. Έχει προταθεί ένα γενικό μοντέλο για σημασιολογικές τροχιές και εισήγαγε την αντίληψη των στάσεων και των κινήσεων. Οι τοποθεσίες ενδιαφέροντος είναι οι τοποθεσίες όπου ο χρήστης σταματά για μια χρονική περίοδο και η σημασιολογική τροχιά αντιπροσωπεύει την ιστορική επίσκεψη των σημασιολογικών τόπων (για παράδειγμα, εργασία, σπίτι, εστιατόρια) [19] καθώς και μια προσέγγιση που βασίζεται σε clusters για να ανακαλύψει τα ενδιαφέροντα σημασιολογικά μέρη στις τροχιές. [30]

Σύμφωνα με την βιβλιογραφία, υπάρχει το ενδιαφέρον για ένα σχετικό αλλά πιο δύσκολο πρόβλημα πρόβλεψης τοποθεσίας, το οποίο στοχεύει στην πρόβλεψη της επόμενης τοποθεσίας χωρίς να υπάρχει η γνώση εκ των προτέρων στοιχεία από μελλοντικά δεδομένα αισθητήρων. Σε γενικές γραμμές, τα δεδομένα κινητής τηλεφωνίας που χρησιμοποιήθηκαν για το πρόβλημα της επόμενης τοποθεσίας συνίστανται στις ιστορικές πληροφορίες σχετικά με τις επαφές και τις σχετικές πληροφορίες περιβάλλοντος (για παράδειγμα, χρονικές σημάνσεις, επιταχυνσιόμετρο, bluetooth και call log / sms log) από αυτές τις επισκέψεις. Υπάρχει εκτεταμένη έρευνα σχετικά με το πρόβλημα της πρόβλεψης της μελλοντικής τοποθεσίας. Οι περισσότερες από αυτές τις εργασίες δημιουργούν ένα μοντέλο που βασίζεται σε συχνά μοτίβα και κανόνες συσχέτισης από ένα ιστορικό διαδρομών χρηστών / συλλογών ως μια ταξινομημένη ακολουθία τοποθεσιών που εκτυπώθηκαν [23]. Άλλα μοντέλα διαδοχικής μάθησης όπως τα Hidden Markov Models [31], τα Conditional Random Fields [32], το Particle Filter [33] έχουν επίσης εφαρμοστεί σε αυτό το πρόβλημα. Ωστόσο, το πρόβλημα που αντιμετωπίζεται είναι διαφορετικό επειδή, για οποιονδήποτε χρήστη, η πρόβλεψη της επόμενης θέσης προϋποθέτει μόνο γνώση σχετικά με την τρέχουσα τοποθεσία (χωρίς δεδομένα σχετικά με προηγούμενες τοποθεσίες). Αυτή η περιορισμένη / μειωμένη ιστορία κάνει το πρόβλημά μας πιο γενικό, καθώς δεν είναι ασυνήθιστο να έχουμε κενά στα δεδομένα του κινητού αισθητήρα. Τα κενά αναφέρονται σε σημαντικές χρονικές

περιόδους όπου το κινητό τηλέφωνο δεν συλλέγει δεδομένα (για παράδειγμα, όταν στο κινητό τηλέφωνο έχει εξαντληθεί η μπαταρία).

Πρόσφατα το Nokia Mobile Data Challenge (MDC) κυκλοφόρησε ένα μεγάλο σύνολο δεδομένων για έρευνα και μία από τις ειδικές εργασίες αποτελούταν από πρόβλεψη επόμενης θέσης [34]. Από αυτήν την πρόκληση του MDC, αρκετές προσεγγίσεις μπόρεσαν να προβλέψουν την επόμενη θέση με υψηλή ακρίβεια [3, 35]. Οι προτεινόμενες προσεγγίσεις επικεντρώθηκαν στην εκμάθηση ενός μοντέλου για κάθε χρήστη, που καταγράφει τη χωροχρονική πορεία των επισκέψεων των χρηστών. Το μεγαλύτερο μέρος της προσπάθειας ήταν αφιερωμένο στη κατασκευή χαρακτηριστικών για κάθε προσέγγιση.

Ωστόσο, δύο βασικά ζητήματα παραμένουν σχετικά ανεξερεύνητα στη βιβλιογραφία της επόμενης πρόβλεψης τοποθεσίας. Πρώτον, ζητήματα απορρήτου προκύπτουν από τη χρήση τέτοιων δεδομένων, παρόλο που υπάρχουν επιχειρήματα προς την κατεύθυνση της ανωνυμοποίησης [34]. Δεύτερον, πλούσιες πληροφορίες περιβάλλοντος μπορούν να αξιοποιηθούν για εξατομίκευση.

Για την απόκρυψη των δεδομένων και πιο συγκεκριμένα αριθμών, ονομάτων, επαφών κινητών συσκευών, αναγνωριστικά δικτύου WLAN, αναγνωριστικά συσκευών Bluetooth και διευθύνσεις Mac χρησιμοποιήθηκαν τεχνικές ανωνυμοποίησης. Για την εύρεση της επόμενης τοποθεσίας για open track χρησιμοποιήθηκε η μέθοδος k-anonymity όπου για να μπορεί να σεβαστεί και να προστατέψει την ιδιωτική ζωή καθώς και να αντλήσει σωστά δεδομένα περικόπτεται τα δεδομένα τοποθεσία όπως γεωγραφικό πλάτος και μήκος και προσπαθεί να οριστεί το ανώνυμο ορθογώνιο. Για να μπορέσουν οι ειδικοί να ορίσουν το μέγεθος του ανώνυμου ορθογωνίου χρειάστηκε να παρακολουθήσουν οπτικά τους συμμετέχοντες στο LDCC στα πιο επισκέψιμα μέρη. Έπειτα το ανώνυμο ορθογώνιο εφαρμόστηκε σε όλα τα δεδομένα των χρηστών. Επίσης χρησιμοποιήθηκαν και οι πληροφορίες από το WLAN όπου και φιλτραριστήκαν μέσω του ανώνυμου ορθογωνίου. Για την εύρεση της επόμενης τοποθεσίας για dedicated track χρησιμοποιείται παρόμοια μέθοδος όπου και εδώ δεν χρησιμοποιήθηκε η γεωγραφική τοποθεσία αλλά τα δεδομένα κατηγοριοποιήθηκαν σε 10 κατηγορίες μερικές από αυτές όπως διακοπές σπίτι ,δουλεία . Η επόμενη μέθοδος που

χρησιμοποιήθηκε για ανωνυμοποίηση διευθύνσεων Mac , επαφών κινητού, καταχωρήσεις κειμένου είναι ο κατακερματισμός δηλαδή γίνεται διάσπαση των στοιχείων για να μην αναγνωρίζονται. Παραδείγματος χάρη για τους αριθμούς τηλεφώνου ο κατακερματισμός γίνεται ως εξής διασπάται στον κωδικό της χώρας του χρήστη και ο υπόλοιπος αριθμός ως κείμενο [34].

Η διατήρηση της ιδιωτικής ζωής καθώς αυτός είναι βασικός χρήστης πρέπει να έχει τον έλεγχο των προσωπικών δεδομένων κινητής τηλεφωνίας του [25]. Χρησιμοποιώντας το NextLocation για εξόρυξη δεδομένων κινητής τηλεφωνίας, τα δεδομένα προσωπικού κινητού τηλεφώνου δεν κοινοποιούνται σε εξωτερικό μέρος χωρίς άδεια.

### **3.3 Προκλήσεις δεδομένων κινητής τηλεφωνίας**

Σε αυτήν την ενότητα, θα αναφερθούν οι προκλήσεις που προκύπτουν από τη συλλογή δεδομένων κινητής τηλεφωνίας για το πρόβλημα πρόβλεψης της επόμενης θέσης

#### **Ανίχνευση τοποθεσίας :**

Τα ανεπεξέργαστα δεδομένα της τοποθεσίας κάθε χρήστη υπολογίζονται συνήθως με βάση το GPS και το Wi-Fi που στη συνέχεια μετατρέπεται σε βάση που συλλέγει τις περισσότερες από τις πληροφορίες που βασίζονται στην κινητικότητα / τοποθεσία χωρίς να περιλαμβάνει τις πραγματικές γεωγραφικές συντεταγμένες / σημεία πρόσβασης. Επιπλέον, οι πληροφορίες από υπηρεσίες κοινωνικής δικτύωσης που υποστηρίζουν τοποθεσία, όπως το Facebook , επιτρέπουν ήδη στο χρήστη να «κάνει check in». Αυτές οι υπηρεσίες περιλαμβάνουν ήδη αυτόματο εντοπισμό θέσης που μπορεί να αξιοποιηθεί για τη δημιουργία ή τον εμπλουτισμό των χρονικών σειρών των σημασιολογικών τοποθεσιών, που απαιτούμε για την πρόβλεψη της επόμενης θέσης.

#### **Ειδικότητα χρηστών:**

Η πρόβλεψη επόμενης θέσης είναι ένα πρόβλημα σχετιζόμενο με το χρήστη, καθώς το σύνολο των τοποθεσιών που επισκέφτηκε είναι προσωπικό και ακόμη και αν αυτό το σύνολο αλληλεπικαλύπτεται μεταξύ διαφορετικών χρηστών, η πορεία των επισκέψεων χρηστών σε διαφορετική τοποθεσία είναι πιθανότατα μοναδική. Επομένως, είναι δύσκολο να δημιουργηθούν με ακρίβεια κοινά μοντέλα σε πολλούς χρήστες, όπως μπορεί να πραγματοποιηθεί σε άλλες εργασίες ταξινόμησης, όπως η αναγνώριση δραστηριότητας / ομιλίας. Οι προκλήσεις των συγκεκριμένων χρηστών δημιουργούν την ανάγκη για χρήση ενός εξατομικευμένου μοντέλου.

#### **Εξέλιξη δεδομένων :**

Η συμπεριφορά κίνησης των χρηστών μπορεί να αλλάξει με την πάροδο του χρόνου. Για παράδειγμα, η αλλαγή σπιτιού / πόλης / χώρας / χώρου εργασίας μπορεί να έχει βαθύ αντίκτυπο στο πιο πρόσφατο μοτίβο κίνησης. Επομένως, προτάθηκε ότι η μοντελοποίηση πρέπει να είναι προσαρμοσμένη, και ακόμη τη χρήση ενός μοντέλου που μπορεί ανά πάσα στιγμή να ενσωματώνει νέες πληροφορίες και να ξεχνά τις παλιές. Επιπλέον, το μοντέλο θα πρέπει να ενσωματώνει απρόσκοπτα νέες τοποθεσίες.

#### **Αραιά και ελλείποντα δεδομένα :**

Είναι πιθανό να λείπουν δεδομένα σε αυτήν την ακολουθία επισκέψεων σε συγκεκριμένες τοποθεσίες. Αυτός είναι ο κύριος λόγος που οδήγησε του ερευνητές να διατυπώσουν το πρόβλημα πρόβλεψης της επόμενης θέσης λαμβάνοντας υπόψη μόνο την τρέχουσα τοποθεσία και όχι μια ακολουθία προηγούμενων τοποθεσιών.

### **3.4 Next Prediction Place: NextLocation**

Το πλαίσιο που προτείνεται στην βιβλιογραφία , ονομάζεται NextLocation, διαμορφώνει την πρόβλεψη ως πρόβλημα ταξινόμησης. Ωστόσο, αντί της εκτέλεσης της παραδοσιακής διαδικασίας μάθησης (δηλαδή, συλλογή δεδομένων, μεταφορά δεδομένων, δημιουργία μοντέλων, ανάπτυξη μοντέλου), όπου δημιουργήθηκε ένα ολοκληρωμένο πλαίσιο που εκτελείται στην ίδια την κινητή συσκευή. Μια βασική καινοτομία του NextLocation είναι ότι διατηρεί το απόρρητο των χρηστών καθώς

επιτρέπει την κατασκευή ενός μοντέλου για την πρόβλεψη της επόμενης θέσης χωρίς αποκάλυψη ιδιωτικών δεδομένων χρήστη. Ένα τέτοιο πλαίσιο δίνει στον χρήστη τον έλεγχο για το ποιος μπορεί να χρησιμοποιήσει τα αποτελέσματα αυτού του μοντέλου (δηλαδή, τις επόμενες προβλέψεις τοποθεσίας) χωρίς να αποκαλύψει τις πραγματικές τοποθεσίες που επισκέφτηκαν και συσχετίστηκαν δεδομένα περιβάλλοντος. Το συστατικό προεπεξεργασίας, το μοντέλο και ο εκτιμητής ακρίβειας παίζουν κεντρικό ρόλο στο προτεινόμενο πλαίσιο. Κάθε ένα από αυτά τα στοιχεία εκτελεί τα ακόλουθα:

- Προεπεξεργασία : τα ανεπεξέργαστα δεδομένα πρέπει να είναι προεπεξεργασμένα / μετασχηματισμένα για την επόμενη προγνωστική θέση. Εδώ, τα δεδομένα τοποθεσίας από μια επίσκεψη εμπλουτίζονται με άλλες πληροφορίες περιβάλλοντος. Το στοιχείο προεπεξεργασίας απαιτεί μόνο τη διατήρηση βραχυπρόθεσμου παραθύρου δεδομένων. Κατά την ενημέρωση του μοντέλου, τα δεδομένα αντιπροσωπεύουν την προηγούμενη τοποθεσία επίσκεψης και τις πληροφορίες περιβάλλοντος, και η μεταβλητή στόχος (για πρόβλεψη) είναι η τρέχουσα τοποθεσία.

- Anytime Model : πρέπει να μπορεί να ενσωματώνει νέες πληροφορίες όποτε αυτές είναι διαθέσιμες (όπως νέες επισκέψεις) και πρέπει επίσης να είναι σε θέση να προβλέψει την επόμενη τοποθεσία. Ένας αλγόριθμος που μαθαίνει σταδιακά μπορεί κατ 'αρχήν να χρησιμοποιηθεί σε αυτό το στοιχείο για τη δημιουργία / ενημέρωση του μοντέλου. Επιπλέον, αυτοί οι αλγόριθμοι είναι ελαφροί και μπορούν να εκτελεστούν χρησιμοποιώντας τους υπολογιστικούς πόρους που συνήθως διατίθενται σε smartphone. Ωστόσο, είναι χρήσιμο εάν ο αλγόριθμος μπορεί να προσαρμόσει το μοντέλο ανά πάσα στιγμή, όταν υπάρχει εξέλιξη στα παρατηρούμενα δεδομένα.

- Εκτιμητής ακρίβειας : η σύγκριση της πρόβλεψης μοντέλου με τον πραγματικό προσδιορισμό επιτρέπει τη διατήρηση μιας εκτίμησης για την ακρίβεια της πρόβλεψης της επόμενης θέσης.

Τα προηγούμενα δεδομένα μπορούν να απορριφθούν μόλις ενσωματωθούν στο μοντέλο, κατά συνέπεια, η κατανάλωση μνήμης της διαδικασίας εκμάθησης

NextLocation είναι πολύ χαμηλή σε σύγκριση με προσεγγίσεις που απαιτούν τη συλλογή όλων των δεδομένων και την επεξεργασία τους σε batchmode.

Προσαρμογή του μοντέλου : Δεδομένου του ζητήματος της εξέλιξης των δεδομένων, είναι σημαντικό να μπορεί να προσαρμοστεί το μοντέλο ανά πάσα στιγμή. Η απλούστερη λύση για την επίτευξη προσαρμογής μοντέλου είναι ότι το μοντέλο αντιπροσωπεύει μόνο τις πιο πρόσφατες εγγραφές

Απευθείας αξιολόγηση μοντέλου: Ως μέρος του προτεινόμενου πλαισίου διατηρήθηκε η αναθεώρηση ή διαδικτυακή ακρίβεια μοντέλου ανά πάσα στιγμή.

### **3.5 Πειραματική αξιολόγηση**

Αυτή η ενότητα περιγράφει τα πειράματα που πραγματοποιήθηκαν για την αξιολόγηση της σκοπιμότητας και της ακρίβειας του NextLocation. Τα δεδομένα που χρησιμοποιήθηκαν στα πειράματα κυκλοφόρησαν για το Nokia Mobile Data Challenge (MDC) [34] και συλλέχθηκαν από τα smartphones σχεδόν 200 συμμετεχόντων κατά τη διάρκεια ενός έτους παγκοσμίως.

Υπήρξε σημαντική προσπάθεια στον μετασχηματισμό δεδομένων και στη λειτουργία χαρακτηριστικών. Χρησιμοποιήθηκαν συνολικά 70 δυνατότητες, συμπεριλαμβανομένων: 11 χρονικών λειτουργιών, 8 επιταχυνσιόμετρων, 2 bluetooth χαρακτηριστικών, 23 call logs (κλήσεις / sms), 20 χαρακτηριστικών σχετικά με επισκέψεις τοποθεσίας, και 6 χαρακτηριστικά του συστήματος.

#### **3.5.1 Nokia MDC Dataset**

Τα δεδομένα MDC συλλέχθηκαν 24 ώρες το 24ωρο, 7 ημέρες τη βδομάδα για μήνες. Οι χρήστες των δεδομένων MDC χωρίζονται σε τρία ξεχωριστά σύνολα. Το σύνολο δεδομένων εκπαίδευσης (που ονομάζεται set A) αποτελείται από δεδομένα κινητής τηλεφωνίας που συλλέγονται από 80 άτομα κατά τη διάρκεια μιας περιόδου που κυμαίνεται από μερικές εβδομάδες έως δύο χρόνια. Τα αόρατα δεδομένα για κάθε συμμετέχοντα στο set A χρησιμοποιείται για τη δημιουργία του συνόλου

δεδομένων δοκιμής (που ονομάζεται set C), όπου το κρυφά δεδομένα αντιστοιχούν στη συνέχιση του set A (στο χρόνο). Τα αληθινά στοιχεία του set C δεν κοινοποιήθηκαν ποτέ. Ωστόσο, το σύνολο επικύρωσης (που ονομάζεται set B) κυκλοφόρησε και χρησιμοποιείται για την αξιολόγηση των αποτελεσμάτων. Το σύνολο επικύρωσης περιέχει επισκέψεις που επιλέχθηκαν τυχαία από το τελευταίο μέρος του set A (στο χρόνο). Το σύνολο επικυρώσεων δημιουργήθηκε με την προσθήκη δεδομένων στο set A με χρονικά διαστήματα που αντιστοιχούν σε τυχαία επιλεγμένες επισκέψεις. Για την πρόκληση MDC, οι συμμετέχοντες ήταν ελεύθεροι να εκτιμήσουν το περιεχόμενο από όλα τα διαθέσιμα δεδομένα εντός καθορισμένου χρονικού διαστήματος (δηλαδή, την τρέχουσα τοποθεσία που αντιστοιχεί σε μια επίσκεψη σε ένα μέρος). Οι επισκέψεις είχαν χρονική σήμανση με την έναρξη / τελικό σημείο εισόδου / εξόδου από την τοποθεσία στην οποία πραγματοποιήθηκε επίσκεψη. Για αυτήν την εργασία υπολογίζονται μόνο επισκέψεις όπου ο χρήστης του κινητού μένει σε αυτήν την τοποθεσία για 20 λεπτά ή περισσότερο.

Η βάση δεδομένων MDC διαθέτει 5 κύριους τύπους δεδομένων: περιβαλλοντικά, προσωπικά, τηλεφωνικά, κατάσταση τηλεφώνου και δεδομένα επισκέψεων. Αυτά τα δεδομένα παρουσιάζονται σε 18 πίνακες, με περισσότερα από 130 βασικά χαρακτηριστικά και έχουν μέγεθος περίπου 50 GB. Μια λεπτομερής περιγραφή της καμπάνιας συλλογής δεδομένων είναι διαθέσιμη στο [34].

Μια σημαντική πρόκληση που παρατηρήθηκε, ήταν το γεγονός ότι ενώ ορισμένοι χρήστες είχαν πολύ τακτικά πρότυπα κίνησης, για ορισμένους χρήστες υπήρχε σημαντική μεταβλητότητα. Αυτό επιδεινώνεται περαιτέρω από το γεγονός ότι ορισμένοι χρήστες κινητής τηλεφωνίας έχουν πολύ περισσότερα δεδομένα από τους άλλους (αν και πρέπει να σημειωθεί ότι περισσότερα δεδομένα δεν σημαίνει απαραίτητα υψηλότερη προβλεψιμότητα)

### **3.5.2 Μετασηματισμός δεδομένων**

Στο σύνολο δεδομένων MDC, ο μετασηματισμός δεδομένων σε επισκέψεις ακολουθίας σε συμβολικά μέρη έχει ήδη υποβληθεί σε επεξεργασία. Ωστόσο, άλλες

πληροφορίες περιβάλλοντος που θα μπορούσαν να χρησιμοποιηθούν για να βελτιώσουν την προγνωστική απόδοση του πρέπει να προέρχονται από τα ανεπεξέργαστα δεδομένα που σχετίζονται με αυτές τις επισκέψεις. Τα πειράματα κατέληξαν με 70 χαρακτηριστικά. Σε αυτήν την ενότητα σύμφωνα με την βιβλιογραφία θα αναλυθεί η διαδικασία εξαγωγής χαρακτηριστικών. Τονίζεται ότι αυτές οι λειτουργίες ήταν όλες υπολογισμένες ανά χρήστη και τα ανεπεξέργαστα δεδομένα υποβάλλονται σε επεξεργασία τοπικά και στη συνέχεια απορρίπτονται χωρίς την ανάγκη να διατηρηθούν όλες οι πληροφορίες στην κύρια μνήμη.

- Χρονικά χαρακτηριστικά : Από τις χρονικές σημάνσεις έναρξης και λήξης μιας συγκεκριμένης επίσκεψης, χρησιμοποιήθηκαν μερικές χρονικά χαρακτηριστικά. Η διάρκεια της επίσκεψης, η ημέρα της εβδομάδας, το Σαββατοκύριακο ή η εργάσιμη ημέρα, η ώρα της ημέρας (0h-24h). Αυτές οι λειτουργίες μπορούν να υπολογιστούν τόσο από την ώρα έναρξης όσο και από τη χρονική σήμανση.

- Λειτουργίες κατάστασης τηλεφώνου : Καταγράφηκαν αρκετοί τύποι δεδομένων σχετικά με την κατάσταση του τηλεφώνου και το λειτουργικό σύστημα τηλεφώνου. Από αυτά τα δεδομένα αντλήθηκαν χαρακτηριστικά για την καταγραφή της κατάστασης του τηλεφώνου που ήταν χαρακτηριστικό της επίσκεψης σε μια συγκεκριμένη τοποθεσία. Το πιο συχνό πρόγραμμα (γενικό, αθόρυβο), ο πιο συχνός ήχος κλήσης που χρησιμοποιείται (κανονικός, αθόρυβος), ελάχιστο και μέγιστο επίπεδο μπαταρίας, κατάσταση φόρτισης τηλεφώνου, μέγιστος χρόνος ενεργοποίησης.

-Λειτουργίες χρήσης τηλεφώνου: Από τη χρήση του τηλεφώνου, θεωρήθηκαν διαθέσιμες πληροφορίες στο αρχείο καταγραφής κλήσεων, συγκεκριμένα, ο πιο συχνός αριθμός. Οι ερευνητές φιλοδοξούσαν ότι αυτό θα μπορούσε να τους βοηθήσει να καταγράψουν καταστάσεις όπου ο επόμενος προορισμός να είναι πολύ συσχετισμένος με τη λήψη συγκεκριμένης κλήσης ή κειμένου. Συνήθως, πριν ένας χρήστης κινητής τηλεφωνίας εγκαταλείψει την τρέχουσα τοποθεσία για τον επόμενο προορισμό, η τελευταία κλήση ή SMS είναι αρκετά προβλέψιμη για τον επόμενο προορισμό (για παράδειγμα, ο χρήστης κινητής τηλεφωνίας καλεί το άτομο που θα συναντήσει αργότερα στον επόμενο προορισμό).



- Περιβαλλοντικά χαρακτηριστικά: Για τα περιβαλλοντικά χαρακτηριστικά διερευνήθηκαν δεδομένα από 4 διαφορετικούς αισθητήρες, επιταχυνσιόμετρο, bluetooth, wlan και gsm. Ωστόσο, δεδομένου ότι τα δεδομένα είναι ανώνυμα ανά χρήστη, ήταν αδύνατο να συλλεχθούν πληροφορίες μεταξύ των χρηστών. Για παράδειγμα, εάν δύο χρήστες έρχονται σε επαφή με τον ίδιο πύργο GSM ή σημείο πρόσβασης Wi-Fi, οι κατακερματισμένες τιμές ή το αντίστοιχο αναγνωριστικό πύργου κυψελών και τα σημεία πρόσβασης σημείου πρόσβασης θα εμφανίζονται διαφορετικά παρά το ότι είναι το ίδιο φυσικό αντικείμενο. Επομένως, χρησιμοποιήθηκαν πληροφορίες που είναι εξατομικευμένες και για τις οποίες το κλειδί κατακερματισμού αντιστοιχεί στο ίδιο αντικείμενο που μπορεί να συλλάβει κάποιες χρήσιμες πληροφορίες για τον προορισμό του χρήστη για κινητά. Ως περιβαλλοντικά χαρακτηριστικά χρησιμοποιήθηκαν:

- Bluetooth: Ομοίως με το κίνητρο πίσω από τις δυνατότητες που δημιουργήθηκε από το αρχείο καταγραφής κλήσεων, οι ερευνητές προσπάθησαν να καταλάβουν αν υπάρχει για μια συγκεκριμένη Bluetooth συσκευή κοντά που επηρεάζει την επόμενη θέση. Δημιουργήσαν μια δυνατότητα που απαιτούσε ορισμένα στατιστικά στοιχεία για την τρέχουσα τοποθεσία και παρατηρήθηκαν οι διευθύνσεις mac Bluetooth για την τοποθεσία. Η διαδικασία προσπαθεί να υπολογίσει την πιθανότητα ενός συγκεκριμένου mac address να συσχετίζεται με έναν συγκεκριμένο προορισμό.

- Επιταχυνσιόμετρο: Αυτό καταγράφει έναν διαφορετικό τύπο δραστηριότητας σε σύγκριση με τη λειτουργία αδράνειας κατάστασης τηλεφώνου. Για παράδειγμα, μπορεί να υπάρχουν καταστάσεις χωρίς αλληλεπίδραση με το τηλέφωνο, αλλά επειδή το τηλέφωνο μεταφέρεται από τον χρήστη κινητού, το επιταχυνσιόμετρο καταγράφει κίνηση. Σε άλλες καταστάσεις, το επιταχυνσιόμετρο δεν καταγράφει καθόλου κίνηση. Τα χαρακτηριστικά που χρησιμοποιήθηκαν είναι: η ελάχιστη, μέγιστη, μέση και τυπική απόκλιση του επιταχυνσιόμετρου 3 αξόνων που καταγράφηκε καθ' όλη τη διάρκεια της περιόδου επίσκεψης και κατά τη διάρκεια των τελευταίων 10 λεπτών.

### 3.5.3 Τεχνικές

Σε αυτήν την εργασία, αξιολογήθηκαν διαφορετικές τεχνικές ταξινόμησης για να συγκριθούν τα διάφορα μοντέλα που έχουν κατασκευαστεί. Στην έρευνά που πραγματοποίησαν οι ειδικοί, χρησιμοποίησαν το WEKA [36], ένα δημοφιλές λογισμικό εξόρυξης δεδομένων, για να συγκρίνουμε τις διαφορετικές τεχνικές ταξινόμησης. Διερεύνησαν επίσης το Massive Online Analysis (MOA) [37] - ένα πλαίσιο ανοιχτού κώδικα για εξόρυξη ροής δεδομένων γραμμένο στην Java. Το WEKA [36] περιλαμβάνει μια συλλογή αλγορίθμων μηχανικής μάθησης και εργαλείων αξιολόγησης (π.χ., προ-διαδοχικό σφάλμα ) ειδικά για προβλήματα μάθησης ροής δεδομένων.

Χρησιμοποιώντας το Weka, πραγματοποιήθηκε αξιολόγηση της προγνωστικής ακρίβειας χρησιμοποιώντας το σύνολο επικύρωσης στα δεδομένα. Αποφασίσθηκε η διερεύνηση διάφορων ταξινομητικών αλγορίθμων και τα προκαταρκτικά αποτελέσματα έδειξαν μια ελαφρώς ανώτερη απόδοση του αλγορίθμου J48 για την εισαγωγή δέντρων αποφάσεων.

- Επιλογή λειτουργιών - Ως τελικό βήμα μετά τη μηχανική χαρακτηριστικών, πραγματοποιήθηκε η επιλογή χαρακτηριστικών. Αυτό περιελάμβανε τη χρήση γνωστών τεχνικών για τον προσδιορισμό των χαρακτηριστικών που έχουν την καλύτερη ικανότητα να προβλέψουν την επόμενη τοποθεσία με βάση την τρέχουσα τοποθεσία του υποκειμένου. Έγινε δυναμική επιλογή των καλύτερων χαρακτηριστικών για κάθε χρήστη (από τους 70 που ανακατασκευάστηκαν / χρησιμοποιήθηκαν) χρησιμοποιώντας δύο γνωστές τεχνικές επιλογής χαρακτηριστικών από το WEKA. Πρώτον, συλλογή πληροφοριών και δεύτερον, επικυρωμένη αξιολόγηση καλύτερων δυνατοτήτων (CfsSubsetEval). Επομένως, το σύνολο των δυνατοτήτων που επιλέγεται για κάθε χρήστη διαφέρει ανάλογα με την παραγωγικότητά τους για το συγκεκριμένο χρήστη / περιβάλλον.

- Στιγμιαία στάθμιση - Ένα άλλο ζήτημα που αντιμετωπίστηκε στην επόμενη θέση πρόβλεψη είναι η ποιότητα των παρατηρήσεων, δηλαδή η αβεβαιότητα που σχετίζεται με αυτές (λόγω αβεβαιότητας / μη διαθεσιμότητας ανάγνωσης αισθητήρα) και επίσης πόσο σχετικά παρουσιάζουν τα μοτίβα κινητικότητας των

χρηστών. Δεδομένου ότι τα δεδομένα έχουν πληροφορίες σχετικά με την αβεβαιότητα (μια σχέση που σχετίζεται με μια αξιόπιστη επίσκεψη, ώρα έναρξης και λήξης), αποφασίστηκε να διερευνηθούν αυτές οι πληροφορίες και να εκτελεστεί στάθμιση συντελεστών σε συνάρτηση με τη διαβεβαίωση για την αξιόπιστη ώρα έναρξης και λήξης.

Χρησιμοποιώντας MOA : Στο σύστημα MOA [37] διερευνήθηκαν προκαταρκτικά διαφορετικοί αλγόριθμοι και αποκτήθηκαν καλά αποτελέσματα με τα Hoeffding Trees. Λόγω του προβλήματος της εξέλιξης των δεδομένων που περιγράφηκε, αποφασίστηκε να διερευνηθεί με μια τεχνική ανίχνευσης μετατόπισης (SingleClassifierDrift). Αυτός ο αλγόριθμος εφαρμόζει μια γνωστή μέθοδο ανίχνευσης μετατόπισης

### **3.6 Αποτελέσματα και συμπεράσματα**

Σε αυτό το άρθρο προτείνεται το πλαίσιο NextLocation, που είναι μια προσέγγιση εξόρυξης δεδομένων κινητής τηλεφωνίας στο πρόβλημα πρόβλεψης της επόμενης θέσης. Το κύριο πλεονέκτημα του NextLocation είναι ότι είναι μια λύση διατήρησης της ιδιωτικής ζωής που εκτελείται πλήρως στην ίδια την κινητή συσκευή. Δεν αποκαλύπτονται ευαίσθητα δεδομένα σχετικά με τις τοποθεσίες και το περιβάλλον χρήστη. Επιπλέον, το NextLocation χρησιμοποιεί ένα προσαρμοστικό μοντέλο και επιτρέπει την προσαρμογή στις αλλαγές στα πρότυπα κινητικότητας των χρηστών. Τέλος, διατηρεί μια εκτίμηση της ακρίβειας του μοντέλου ανά πάσα στιγμή σε πραγματικό χρόνο. Τα αποτελέσματα στα δεδομένα MDC δείχνουν μεγάλη μεταβλητότητα απρόβλεπτης ακρίβειας μεταξύ των χρηστών, όπου οι περιστασιακοί χρήστες είναι πολύ δύσκολο να προβλεφθούν, ενώ για πιο τακτικούς χρήστες είναι δυνατόν να επιτευχθεί ακρίβεια άνω του 80%.. Επιπλέον, προτάθηκε να εναλλακτικό μοντέλο διαφήμισης για κινητά που μπορεί να εφαρμοστεί χρησιμοποιώντας το NextLocation.

## **Κεφάλαιο 4**

### **Ιδιωτικότητα**

Στο κεφάλαιο 3 αναφερθήκαμε στο πρόβλημα της ιδιωτικότητας που προκύπτει από τη χρήση δεδομένων για την πρόβλεψη της επόμενης τοποθεσίας αλλά και από τη γενικότερη εξόρυξη δεδομένων από κινητές συσκευές. Στο κεφάλαιο 4 θα δούμε πως είναι δυνατόν να προστατευθούν αυτά τα προσωπικά δεδομένα.

Μια σημαντική πτυχή των εφαρμογών MCS (Mobile crowdsensing applications) είναι ότι συλλέγουν δυνητικά ευαίσθητα δεδομένα αισθητήρων που αφορούν άτομα. Για παράδειγμα, οι αναγνώσεις αισθητήρων GPS μπορούν να χρησιμοποιηθούν για να εξαχθούν προσωπικές πληροφορίες σχετικά με το άτομο, όπως οι διαδρομές που ακολουθεί κατά τη διάρκεια των καθημερινών μετακινήσεων, καθώς και οι θέσεις του σπιτιού και της εργασίας [38]. Από την άλλη πλευρά, αυτές οι μετρήσεις αισθητήρα GPS (από καθημερινές μετακινήσεις) που μοιράζονται σε μια ευρύτερη κοινότητα μπορεί να χρησιμοποιηθούν για την επίτευξη επιπέδων συμφόρησης κυκλοφορίας σε μια πόλη [39]. Επομένως, είναι σημαντικό να διατηρηθεί η ασφάλεια και το απόρρητο ενός ατόμου, αλλά ταυτόχρονα να ενεργοποιούνται οι εφαρμογές MCS. Είναι επίσης απαραίτητο να διασφαλιστεί ότι τα δεδομένα αισθητήρων ενός ατόμου δεν αποκαλύπτονται σε αναξιόπιστα τρίτα πρόσωπα. Ένα πρόβλημα που προκύπτει από τη δυνατότητα επιλογής των εφαρμογών crowdsensing είναι όταν κακόβουλα άτομα συνεισφέρουν λανθασμένα δεδομένα αισθητήρων (π.χ. ψευδείς αναγνώσεις GPS). Ως εκ τούτου, η διατήρηση της ακεραιότητας των δεδομένων των αισθητήρων που συλλέγονται είναι ένα σημαντικό πρόβλημα.

Μια δημοφιλής προσέγγιση για τη διατήρηση του απορρήτου των δεδομένων είναι η ανωνυμοποίηση [40], η οποία καταργεί οποιαδήποτε αναγνώριση πληροφοριών από τα δεδομένα του αισθητήρα προτού αυτά διαμοιραστούν σε τρίτα πρόσωπα. Το μειονέκτημα μιας τέτοιας προσέγγισης είναι ότι οι ανώνυμες μετρήσεις αισθητήρα GPS (ή τοποθεσίας) μπορούν ακόμα να χρησιμοποιηθούν για να συναγάγουν τις συχνές επισκέψεις ενός ατόμου και να αντλήσουν τα προσωπικά τους στοιχεία. Μια άλλη προσέγγιση για τη διατήρηση του απορρήτου είναι ο ασφαλής υπολογισμός πολλαπλών μερών [41], όπου χρησιμοποιούνται κρυπτογραφικές τεχνικές για τη μετατροπή των δεδομένων προκειμένου να διατηρηθεί το απόρρητο.

Τέτοιες κρυπτογραφικές τεχνικές είναι εντατικές και δεν είναι επεκτάσιμες, επειδή απαιτούν τη δημιουργία και συντήρηση πολλαπλών κλειδιών, γεγονός που οδηγεί επίσης σε υψηλότερη κατανάλωση ενέργειας. Πιστεύουμε ότι οι προσεγγίσεις που βασίζονται στη διαταραχή δεδομένων, οι οποίες προσθέτουν θόρυβο στα δεδομένα αισθητήρων πριν από την κοινή χρήση τους με την κοινότητα για τη διατήρηση της ιδιωτικής ζωής ενός ατόμου, είναι κατάλληλες. Οι προσεγγίσεις διαταραχής δεδομένων [42, 43] βασίζονται στην προσθήκη θορύβου με τέτοιο τρόπο ώστε να διατηρείται το απόρρητο ενός ατόμου, αλλά ταυτόχρονα είναι δυνατόν να υπολογιστούν τα στατιστικά στοιχεία ενδιαφέροντος με υψηλή ακρίβεια (λόγω της φύσης του θορύβου που προστίθεται)

Η ακεραιότητα των δεδομένων, διασφαλίζοντας την ακεραιότητα των δεδομένων αισθητήρα που δημιουργούνται από άτομα, πρέπει να αντιμετωπιστεί από εφαρμογές MCS. Μερικές προσεγγίσεις έχουν προταθεί στην υπάρχουσα βιβλιογραφία [44, 45], οι οποίες βασίζονται σε συνεγκατεστημένες υποδομές ως μάρτυρες. Μια τέτοια προσέγγιση βασίζεται σε εισροές από την εγκατάσταση ακριβής υποδομής, που μπορεί να είναι απαγορευτικό και μη διαθέσιμο κατά καιρούς. Μια άλλη προσέγγιση είναι η υπογραφή του αισθητήρα δεδομένων (από αξιόπιστο υλικό εγκατεστημένο σε κινητά τηλέφωνα). Ωστόσο, αυτή η προσέγγιση είναι δυνητικά προβληματική καθώς πρέπει να γίνει η επαλήθευση ακόμη και στο λογισμικό.

Κάνουμε μερικές παρατηρήσεις σχετικά με το απόρρητο, την ασφάλεια και την ακεραιότητα δεδομένων για εφαρμογές MCS. Πρώτον, παρατηρήθηκε ότι το απόρρητο είναι πολύ συγκεκριμένο για τον χρήστη, δηλαδή κάθε άτομο έχει διαφορετική αντίληψη για το απόρρητο. Για παράδειγμα, ένα άτομο μπορεί να είναι πρόθυμο να κοινοποιεί συνεχώς τις πληροφορίες τοποθεσίας του, ενώ ένα άλλο μπορεί να μην το κάνει.

Απαιτείται η ανάπτυξη τεχνικών απορρήτου που να αντιμετωπίζουν τις διαφορές στις μεμονωμένες προτιμήσεις. Επιπλέον, μια γενική τεχνική διαταραχής ή ένα πλαίσιο τεχνικών διαταραχής πρέπει να αναπτυχθεί έτσι ώστε η ιδιωτικότητα και η ασφάλεια να μπορούν να επιτευχθούν σε ένα γενικό περιβάλλον ανεξάρτητα από τη φύση των δεδομένων που κοινοποιούνται. Τέλος, οι πραγματικές εφαρμογές

MCS πρέπει να αντιμετωπίσουν το πρόβλημα ακεραιότητας δεδομένων, προκειμένου να παράσχουν ουσιαστικά συμπεράσματα από τα συγκεντρωτικά δεδομένα αισθητήρων.

Προδιαγεγραμμένες πολιτικές ασφάλειας, ως μέτρα προστασίας, τίθενται σε ισχύ από τους κατασκευαστές συσκευών και τους παρόχους, αλλά αυτό που χρειάζεται να γίνει με επιτυχία είναι η σωστή εκπαίδευση των χρηστών κινητής τηλεφωνίας, προκειμένου να προστατεύσουν την ιδιωτική τους ζωή και τα δεδομένα που αποθηκεύουν σε συσκευές (είτε πρόκειται για προσωπικά δεδομένα ή για αρχεία που σχετίζονται με το περιβάλλον εργασίας). Στα πλαίσια της εκπαιδευτικής μεθόδου, θα πρέπει να υπάρχει μια σαφέστατη και πλήρως κατανοητή μέθοδος επικοινωνίας με τους εκπαιδευόμενους .

## **Κεφάλαιο 5**

### **Συμπεράσματα**

Είναι πλέον σαφές ότι η Εξόρυξη Δεδομένων είναι ένας από τους πιο σημαντικούς τομείς της εποχής μας. Τα δεδομένα είναι ασταμάτητα και η άχρηστη πληροφορία κατέχει το μεγαλύτερο ποσοστό μέσα σε αυτά. Η έξυπνη ανάλυση τους αποτελεί πολύτιμο πόρο αλλά είναι ένα δύσκολο έργο. Στις περισσότερες περιπτώσεις, η χρήση μιας τεχνικής, μιας μεθόδου ή ενός αλγορίθμου δεν είναι αρκετή. Οι δομές δεδομένων είναι περίπλοκες και οι βάσεις δεδομένων ξεχειλίζουν από ακατέργαστη πληροφορία, ώστε είναι αρκετά προκλητική και σκληρή διαδικασία για να αναλυθούν αποτελεσματικά. Η ανάγκη όμως να κατανοήσουμε αυτά τα πλούσια σε γνώση σύνολα δεδομένων, είναι κοινός στόχος σε όλους τους τομείς, συμπεριλαμβανομένων και των επιχειρήσεων, της επιστήμης και της μηχανικής.

Με την χρήση τεχνικών, αλγορίθμων και άλλων εργαλείων, η εξαγωγή γνώσης καθίστανται δυνατή. Δεν υπάρχει καλύτερος αλγόριθμος ή κάποια καλύτερη τεχνική που χρησιμοποιείται ώστε να λάβουμε το επιθυμητό και καλύτερο δυνατό αποτέλεσμα. Συγκριμένα, η επιλογή των αλγορίθμων γίνεται ανάλογα με την γνώση που θέλουμε να εξάγουμε, τα δεδομένα που διαθέτουμε, την προτίμηση ανάμεσα

στον ρυθμό και την ακρίβεια, καθώς και πάνω σε ποιες εφαρμογές θέλουμε να τους χρησιμοποιήσουμε. Ουσιαστικά, υπάρχουν πολλές παράμετροι και πολλοί παράγοντες που πρέπει να ληφθούν υπόψη στην επιλογή του κατάλληλου – για εμάς – αλγορίθμου.

## Βιβλιογραφία

[1] Wikipedia, The Free Encyclopedia. Smartphone. <http://en.wikipedia.org/wiki/Smartphone>.

[2] Wikipedia, The Free Encyclopedia. Global Positioning System. [http://en.wikipedia.org/wiki/Global\\_Positioning\\_System](http://en.wikipedia.org/wiki/Global_Positioning_System)

[3] Wikipedia, The Free Encyclopedia. GPS navigation device. [http://en.wikipedia.org/wiki/GPS\\_navigation\\_device#Mobile\\_phones\\_with\\_GPS\\_capability](http://en.wikipedia.org/wiki/GPS_navigation_device#Mobile_phones_with_GPS_capability).

[4] Hagen Hopfner . Smartphone Hardware Sensors. <http://www.uni-weimar.de/medien/wiki/images/Zeitmaschinen-smartphonesensors.pdf>.

[5] Wikipedia, The Free Encyclopedia. Accelerometer. <http://en.wikipedia.org/wiki/Accelerometer>.

[6] Wikipedia, The Free Encyclopedia. Magnetometer. <http://en.wikipedia.org/wiki/Magnetometer>.

[7] Wikipedia, The Free Encyclopedia. Android(operating system). [http://en.wikipedia.org/wiki/Android\\_\(operating\\_system\)](http://en.wikipedia.org/wiki/Android_(operating_system)).

[8] Developers Android. Android Platform Versions Dashboard. <http://developer.android.com/about/dashboards/index.html>.

[9] Wei-Meng Lee. Beginning Android Application Development. Wrox Press Ltd., Birmingham, UK, UK, 1st edition, 2011.

[10] The University of Waikato. Use WEKA in your Java code. <http://weka.wikispaces.com/Use+WEKA+in+your+Java+code>.

- [11] Paolo Giudici and Silvia Figini. *Applied Data Mining for Business and Industry*. Wiley Publishing, 2nd edition, 2009.
- [12] Domenico Talia and Paolo Trunfio. *Mobile Data Mining on Small Devices through Web Services*, pages 264–276. John Wiley & Sons, Inc., 2010.
- [13] Sweta Pittie, Hillol Kargupta, and Byung-Hoon Park. Dependency detection in Mobimine: a systems perspective. *Inf. Sci. Inf. Comput. Sci.*, 155(3-4):227–243, 2003.
- [14] Matthias Steinbauer, Ismail Khalil, and Gabriele Kotsis. *Reality Mining at the Convergence of Cloud Computing and Mobile Computing*. ERCIM News, 2013
- [15] Mei-po Kwan, Jiyeong Lee, In Michael, F. Goodchild, and Donald G. Janelle. *Geovisualization of Human Activity Patterns Using 3D GIS: A Time-Geographic Approach*. In *Spatially Integrated Social Science*, pages 48–66. University Press, 2004.
- [16] Gianni Giannotti, Fosca Giannotti, and Dino Pedreschi. *Mobility, data mining and privacy: Geographic knowledge discovery*. Springer, 2008.
- [17] Mohd Fikri Azli bin Abdullah, Ali Fahmi Perwira Negara, Md Shohel Sayeed, DeokJai Choi, and Kalaiarasi Sonai Muthu. *Classification algorithms in human activity recognition using smartphones*. In *Proceedings of World Academy of Science, Engineering and Technology*, volume 68. World Academy of Science, Engineering and Technology, 2012.
- [18] Fosca Giannotti, Mirco Nanni, Dino Pedreschi, Fabio Pinelli, Chiara Renso, Salvatore Rinzivillo, and Roberto Trasarti. *Mobility data mining: discovering movement patterns from trajectory data*. In *Proceedings of the Second International Workshop on Computational Transportation Science, IWCTS '10*, pages 7–10. ACM, 2010.
- [19] Gerasimos Marketos and Y Theodoridis. *Mobility Data Warehousing and Mining*. In *Proceedings of VLDB PhD Workshop*, 2009.
- [20] Cynthia Sung, Dan Feldman, and Daniela Rus. *Trajectory clustering for motion prediction*. In *IROS*, pages 1547–1552. IEEE, 2012.



- [21] Jae-Gil Lee, Jiawei Han, and Kyu-Young Whang. Trajectory clustering: a partition-and-group framework. In Proceedings of the 2007 ACM SIGMOD international conference on Management of data, SIGMOD '07, pages 593–604. ACM, 2007.
- [22] Lin Liao, Dieter Fox, and Henry Kautz. Extracting Places and Activities from GPS Traces Using Hierarchical Conditional Random Fields. *Int. J. Rob. Res.*, 26:119–134, 2007.
- [23] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti. Wherenext: a location predictor on trajectory pattern mining. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 637–646. ACM, 2009.
- [24] J. K. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T. M. T. Do, O. Dousse, J. Eberle, and M. Miettinen. The mobile data challenge: Big data for mobile computing research. In *Mobile Data Challenge by Nokia Workshop*, in conjunction with *Int. Conf. on Pervasive Computing*, Newcastle, UK, 2012
- [24] B. Ly. *Mobile data challenge 2012: Unlocking the secrets of smartphone data*, 2012.
- [25] R. K. Ganti, F. Ye, and H. Lei. Mobile crowdsensing: current state and future challenges. *IEEE Communications Magazine*, 49(11):32–39, 2011.
- [26] W. Sherchan, P. P. Jayaraman, S. Krishnaswamy, A. Zaslavsky, S. Loke, and A. Sinha. Using on-the-move mining for mobile crowdsensing. In *Mobile Data Management (MDM), 2012 IEEE 13th International Conference on*, pages 115–124. IEEE, 2012
- [27] P. Barwise and C. Strong. Permission-based mobile advertising. *Journal of interactive Marketing*, 16(1):14–24, 2002
- [28] G. Yavas, D. Katsaros, O. Ulusoy, and Y. Manolopoulos. A data mining approach for location prediction in mobile environments. *Data & Knowledge Engineering*, 54(2):121–146, 2005

- [29] M. Zook, M. Graham, T. Shelton, and S. Gorman. Volunteered geographic information and crowdsourcing disaster relief: a case study of the haitian earthquake. World Medical
- [30] A. T. Palma, V. Bogorny, B. Kuijpers, and L. O. Alvares. A clustering-based approach for discovering interesting places in trajectories. In Proceedings of the 2008 ACM symposium on Applied computing, pages 863–868. ACM, 2008
- [31] W. Mathew, R. Raposo, and B. Martins. Predicting future locations with hidden markov models. 2012
- [32] R. Pan, J. Zhao, V. W. Zheng, J. J. Pan, D. Shen, S. J. Pan, and Q. Yang. Domain-constrained semi-supervised mining of tracking models in sensor networks. In Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining. P 1023-1027. ACM, 2007
- [33] V. Fox, J. Hightower, L. Liao, D. Schulz, and G. Borriello. Bayesian filtering for location estimation. Pervasive Computing, IEEE, 2(3):24–33, 2003
- [34] J. K. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T. M. T. Do, O. Dousse, J. Eberle, and M. Miettinen. The mobile data challenge: Big data for mobile computing research. In Mobile Data Challenge by Nokia Workshop, in conjunction with Int. Conf. on Pervasive Computing, Newcastle, UK, 2012
- [35] H. Gao, J. Tang, and H. Liu. Mobile location prediction in spatio-temporal context.
- [36] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update. ACM SIGKDD Explorations Newsletter, 11(1):10–18, 2009.
- [37] A. Bifet, G. Holmes, R. Kirkby, and B. Pfahringer. Moa: Massive online analysis. The Journal of Machine Learning Research, 11:1601–1604, 2010.
- [38] J. Krumm, “A Survey of Computational Location Privacy,” Personal and Ubiquitous Computing, vol. 13, no. 6, 2009, pp. 391–99.

[39] B. Hull et al., "Cartel: A Distributed Mobile Sensor Computing System," Proc. SenSys, 2006, pp. 125–38.

[40] L. Sweeney, "k-Anonymity: A Model for Protecting Privacy," Int'l. J. Uncertainty, Fuzziness, and KnowledgeBased Systems, vol. 10, no. 5, Oct. 2002, pp. 557–70.

[41] A. Yao, "Protocols for Secure Computations," Proc. IEEE Symp. Foundations of Comp. Sci., 1982, pp. 160–64.

[42] R. Agrawal and R. Srikant, "Privacy Preserving Data Mining," Proc. ACM Conf. Mgmt. of Data, May 2000, pp. 439–50

[43] R. K. Ganti et al., "Poolview: Stream Privacy for Grassroots Participatory Sensing," Proc. SenSys '08, 2008, pp. 281–94

[44] V. Lenders et al., "Location-based Trust for Mobile User-Generated Content: Applications, Challenges, and Implementations," Proc. HotMobile, 2008, pp. 60–64.

[45] S. Saroiu and A. Wolman, "Enabling New Mobile Applications With Location Proofs," Proc. HotMobile, 2009, pp. 1–6.