Diploma Theses Catalogue, 2020

Data Science Lab (www.datastories.org)

University of Piraeus

2020.01: Context-aware epidemic models for tracking the spread diseases via mobility networks. In the aftermath of epidemic or pandemic outbreaks like the latest SARS-nCoV-2, the previous SARS and MERS epidemics, or other known diseases like ebola, it has been established that one of the major factors in predicting and containing the spread is the mobility characteristics of the general population. The standard epidemic models like SIR or SEIR, even in their "compartmentalized" variants, are unable to capture the context-sensitive information about the transportation routes, the intensity of their flow, their temporal variability, etc. On the other hand, the field of spatial epidemiology usually incorporates poorer dynamics than SEIR-like modeling, as it is mostly related to health geography rather than mobility tracking. There are several datasets from various sources that can be used to fuse together these different approaches into more realistic and context-aware epidemic models, from incorporating international flight routes for continent-level analysis to urban traffic analytics for city-level analytics. This project aims at investigating these novel approaches, explore the mobility datasets that are available and propose hints on how the epidemic models can be augmented with detailed mobility analytics. **Contact person**: Dr. Harris Georgiou (hgeorgiou@unipi.gr)

2020.02: Spectral decomposition and sparse models for Big Data & robust data analytics. In modern biomedical imaging, such as fMRI for brain, the main processing pipeline includes proper registration, filtering and statistical characterization for a very large number of correlated random variables in the form of time series. These stages are often based on "spectral" component analysis in terms of blind source separation, i.e., "unmix" the original data and identify the (few) true underlying generating processes that produce this large sets of high dimensionality. The same problem is present in various other contexts and modalities, such as the identification of trends in "big data" tasks, graphs and flows in traffic/trajectory modeling, outlier events detection, etc. In practice, these approaches include algorithms from Machine Learning, such as PCA, ICA, Dictionary Learning and/or Compressed Sensing (depending on the exact problem). Furthermore, fractal dimension analysis and other approaches from adaptive filter theory & dynamic systems can be employed to assess the validity of features, clusters, dimensionality reduction, etc. In this project, we will study a specific big data and/or data mining problem under the scope of these approaches and provide a comparative assessment against other established algorithms used in the same task. **Contact person**: Dr. Harris Georgiou (hgeorgiou@unipi.gr)

2020.03: Driver behavior profiling from trajectory analytics using supervised and unsupervised models. Driver behavior profiling (DBP), specifically in relation to identifying

"good" versus "bad" driving patterns, is one of the most challenging problems in mobility data analytics. The core problem usually incorporates location and accelerometer data, which have to be combined, mutually "corrected" for noise and other artifacts, map-matched to the underlying road network, before extracting any useful mobility pattern or "signature" that can be used for modeling driver categories. While the long-term approach in DBP has been explored using location-only data, e.g. from single GPS sensors, the short-term approach is inherently more demanding in terms of spatio-temporal resolution, data quality and additional sensing modalities. In practice, tracking the movement of a single car or driver for an entire month to extract commonly used routes, visited POIs or risk of car crash within this context is inherently more straight-forward and well-studied than having to analyze movement patterns in the context of few minutes or seconds to distinguish between "good" and "bad" driving. The short-term case, being more challenging, is normally approached by employing multi-modal, high-resolution sensing, e.g. location tracking together with accelerometer measurements, while at the same time having pre-determined training routes and confirmed driver "events" as ground truth for model training. However, these pre-requisites cannot always be satisfied, as multiple sensing and/or ground truth may not be unavailable, sampling rates may be too low, etc. This project aims at investigating the DBP problem in the context of short-term mobility analytics, using location data with or without the availability of accelerometer data, in supervised and/or unsupervised modes (depending on the dataset), in order to identify robust and effective features and models that distinguish DBP categories for "good" and "bad" driving. **Contact person**: Dr. Harris Georgiou (hgeorgiou@unipi.gr)

2020.04: Discovering trends and outliers in biomedical sensory data for the general population. Modern wearable and IoT technology enables the creation of "always on" monitoring applications that can be active in dedicated devices or in a smartphone. Using non-invasive techniques and multi-modal sensory data, it is possible to monitor many important biomedical data from a person, either as part of a medical procedure (e.g. long-term heart monitoring) or normal daily activities and recreation (e.g. athletics). When these data are combined anonymously, it is possible to extract the predominant trends and statistics for the general population, but more importantly, to identify abnormal patterns and outliers that may be hints to medical conditions that require attention. This process can be investigated, optimized and streamlined, in order to be implementable as applications running in devices of very limited memory, processing and connectivity resources, such as in smartphones in offline/asynchronous mode. As an example, such datasets may include ECG (cardiac), SpO2 (oximetry), respiration, temperature, sweating, etc. **Contact person**: Dr. Harris Georgiou (hgeorgiou@unipi.gr)

2020.05: Απεικόνιση μεγάλων κυκλοφοριακών δεδομένων μέσων μεταφοράς στην Αττική. Η συλλογή και διαχείριση δεδομένων σχετικών με τα ΜΜΜ στην περιοχή της Αττικής, γίνεται από τον ΟΑΣΑ. Ο ΟΑΣΑ συλλέγει δεδομένα μέσω του συστήματος «Τηλεματική» και μέσω των ηλεκτρονικών εισιτηρίων, τα οποία μπορούν να χαρακτηριστούν ως Μεγάλα Δεδομένα (Big Data), λόγω του όγκου, της ποικιλομορφίας, του ρυθμού παραγωγής και μετάδοσης, της μεταβαλλόμενης ακρίβειας, καθώς και της επιχειρησιακής αξίας τους («5Vs»). Η εγγενής δυσκολία διαχείρισης αυτών των δεδομένων, δημιουργεί προκλήσεις όσον αφορά στην αποθήκευσή τους σε

πλατφόρμες διαχείρισης Μεγάλων Δεδομένων, στην περαιτέρω ανάλυσή τους καθώς και στην απεικόνισή τους. Σκοπός της πτυχιακής είναι η χρήση ή/και ανάπτυξη κατάλληλων εργαλείων για τη διαχείριση και απεικόνιση μεγάλων κυκλοφοριακών δεδομένων. **Contact person**: Dr. Eva Chondrodima (evachon@unipi.gr)

2020.06: Βελτιστοποίηση των δρομολογίων των μέσων μεταφοράς. Τα δημόσια ΜΜΜ επηρεάζουν σημαντικά τον οικονομικό προϋπολογισμό μιας χώρας. Από τη μία ενισχύεται η οικονομία, διότι ένα αποτελεσματικό σύστημα δημόσιων μεταφορών ελκύει περισσότερους επιβάτες, οι οποίοι μέσω των εισιτηρίων που εκδίδουν συμβάλλουν στην αύξηση των εσόδων του κράτους. Από την άλλη, οι κυβερνήσεις των χωρών έχουν εκδηλώσει έντονη ανησυχία για θέματα, όπως η υψηλή τιμή των καυσίμων, τα έξοδα συντήρησης των οχημάτων, τα ατυχήματα που μπορεί να συμβαίνουν σε καθημερινή βάση, τα οποία, όμως, με τις κατάλληλες ενέργειες θα μπορούσαν να αντιμετωπιστούν με επιτυχία. Σκοπός της πτυχιακής είναι η χρήση ή/και ανάπτυξη κατάλληλων τεχνικών για τη βελτιστοποίηση των δρομολογίων των μέσων μεταφοράς με στόχο τη μείωση του αριθμού των επιβατών ανά στάση, ή/και τη μείωση της κατανάλωσης καυσίμου, ή/και τη μείωση του ρίσκου των ατυχημάτων. **Contact person**: Dr. Eva Chondrodima (evachon@unipi.gr)

2020.07: Πρόβλεψη του αριθμού των επιβατών στις στάσεις των μέσων μεταφοράς ή/και των γραμμών τους. Στόχος των ΜΜΜ είναι η παροχή αποτελεσματικών, αξιόπιστων, ασφαλών και βιώσιμων υπηρεσιών μεταφοράς υψηλής ποιότητας, ώστε να αποτελούν μια ελκυστική εναλλακτική λύση έναντι άλλων μέσων μεταφοράς (ιδιωτικά αυτοκίνητα, ταξί, κτλ.). Ο σχεδιασμός τέτοιων «ελκυστικών» και ταυτόχρονα «οικονομικών» ΜΜΜ αντιμετωπίζει πολλές προκλήσεις. Τα ΜΜΜ αποτελούν εξαιρετικά περίπλοκα συστήματα, λόγω της τεχνικής και οργανωτικής τους πολυπλοκότητας, αλλά και λόγω του μεγάλου αριθμού επιβατών που εξυπηρετούν καθημερινά. Η πρόβλεψη του αριθμού των επιβατών στις στάσεις των λεωφορείων μπορεί να βοηθήσει στο να επανασχεδιαστούν τα δρομολόγια των λεωφορείων ώστε για παράδειγμα να περνούν πιο συχνά από στάσεις που παρουσιάζουν συνωστισμό. Σκοπός της πτυχιακής είναι η χρήση ή/και ανάπτυξη αλγορίθμων μηχανικής μάθησης, για την πρόβλεψη του αριθμού των επιβατών στις στάσεις των μέσων μεταφοράς ή/και των γραμμών τους, κατάλληλων για μεγάλα κυκλοφοριακά δεδομένα. **Contact person**: Dr. Eva Chondrodima (evachon@unipi.gr)

2020.08: Detecting super-spreaders by utilizing big mobility data. With the recent Covid-19 outbreak, where severe restriction in the movement of individuals were imposed, the need for identifying, in time, individuals who are more likely to infect other is more imperative than ever. A factor that plays a significant role in this is human mobility and interaction. In this thesis, we are going to utilize large scale human mobility data in order to detect super-spreaders, who may play a significant role in the outbreak of a disease. **Contact person**: Dr. Panagiotis Tampakis (ptampak@unipi.gr)

2020.09: Online trajectory join over big mobility data. Joining trajectory datasets is a significant operation in mobility data analytics and the cornerstone of various methods that aim to extract knowledge out of them.In the era of Big Data, the production of mobility data has become massive in terms of volume and velocity and, consequently, performing such an operation in a centralized way is not feasible. In this thesis, we are going to tackle the problem of trajectory join in a streaming environment where the goal is, given two large sets of trajectories, to identify in a latent way all pairs of trajectories that are "close enough" in space and time for at least some duration. **Contact person**: Dr. Panagiotis Tampakis (ptampak@unipi.gr)

2020.10: Detecting statistically significant spatial clusters in vessel data. Maritime surveillance operations are of great importance in terms of determining the exact location and activity of vessels, among others. Monitoring traffic patterns from various surveillance means (e.g., AIS, VMS), provides a wealth of information useful for maritime safety, security and efficiency. In this thesis, we will explore the application of distributed clustering algorithms in order to discover statistically significant spatial clusters from large maritime surveillance data. In particular, we will use real datasets containing surveillance information, i.e., navigation-related data regarding fishing activity, to provide insights into the movements of individual fishing vessels. The exploitation of these data can discover locations where vessels concentrate, indicating highly populated areas (e.g., anchorages and ports) as well as allow us to understand fishing patterns. **Contact person**: Yannis Kontoulis (ikontoulis@unipi.gr)

2020.11: Sport analytics via mobility analytics. Recently analytics in professional sports has experienced a huge growth due to the wide deployment of player and ball tracking systems. The massive amount of such fine-grained data can provide insights on players' and team's performance. The goal of this thesis is to apply mobility analytics techniques to such data, exploring offline or online methods of classification and/or clustering patterns that could be useful for predictive analytic purposes (e.g. prediction of a game's outcome). **Contact person**: Prof. Nikos Pelekis (npelekis@unipi.gr)

2020.12: Online network-aware trajectory clustering. Nowadays, large amounts of tracking data are generated via location enabled devices and other advanced tracking technologies. In this thesis, we will design and implement an online and distributed clustering algorithm for such streaming trajectory data moving on urban networks. **Contact person**: Prof. Nikos Pelekis (npelekis@unipi.gr)

2020.13: Comparative analysis of tools for big trajectory datasets. Lately, a number of big data tools has been introduced to facilitate the analysis and processing of big trajectory datasets. The goal of this thesis will be to develop a comparative analysis of state-of-the-art systems such as TrajSpark, UlTraMan, VIPTRA, STHadoop, STARK, etc. **Contact person**: Prof. Nikos Pelekis (npelekis@unipi.gr)

2020.14: Indexing & mining single- and multi-dimensional data series. In recent years, with the enormous growth of WWW and the advances in hardware and software technologies, we have the ability to track in real time any kind of transactions, such as customer click data, patient health data, TCP/IP traffic, GPS data etc., in order to support real time decision making. Typically, this kind of data arrives as data series (timeseries), therefore, searching and mining through this information requires efficient techniques for data series indexing. Recently, a number of effective methods for indexing data series have been proposed, with iSAX family being the most notable[1], reaching up to the efficient indexing of billion data series. In this project, we will study these techniques, implement the most promising ones, and validate them using large real datasets. **Contact person**: Prof. Yannis Theodoridis (ytheod@unipi.gr)

For more info about the Data Science Lab activities, please visit: www.datastories.org.

Piraeus, April 2020.

[1] iSAX page: http://www.cs.ucr.edu/~eamonn/iSAX/iSAX.html.