INSIST

Deliverable

D3.1.1 – High-Level Data Processing Algorithms

Editor:

Emine Ferraro

ITEA 3 Project 13021

Date: 07.08.2017

Document properties

Security	Confidential
Version	Final
Authors	Emine Ferraro, Argedor Özgür Devrim Orman, Bor Yazılım Mustafa Eren, Verisun Çağrı Kaçmaz, KoçSistem
Pages	

History of changes

Version	Author, Institution	Changes
0.1	Emine Ferraro, Argedor	Draft Document
0.2	Güven Fidan, Argedor	Finalisation of the document
0.3		
0.4		

Abstract

This document presents information about how to design and implement high-level data algorithms for smart traffic monitoring and advertising systems.

Table of contents

1. Executive Summary

Cities have a long history of traffic monitoring tools, ranging from traditional traffic surveys to analyzing tickets, traffic cameras and road surface inductive loops. For a large part, traffic analysis relies on knowing where vehicles (and people) reside at any given moment, that is, analysis of location data [01]. A challenge in measuring and modeling city-scale traffic is the rapid variability of the ''digital urban landscape,'' with wireless technologies and traffic monitoring techniques becoming more or less popular over time.

Recently, technologies such as GPS, WiFi, Bluetooth and mobile communication networks have facilitated collecting data on urban mobility. The constant evolution of the deployed technologies suggests that relying on any single urban technology for mapping traffic is likely to be short lived. For instance, while a lot of research has demonstrated the use of Bluetooth for mapping traffic [02, 03, 04], recent decisions of handset manufacturers to limit Bluetooth functionality suggest that this technique may soon be outdated. However, it is likely that location databases continue to grow with the emergence of new proximity technologies. Although proximity technologies offer opportunities for passive, infrastructure-centric monitoring, it should be noted that GPS data is valuable for accurate tracking in wide geographic areas and particularly when studying subjects that naturally lend themselves to continuous GPS tracking, such as taxis [05]. Therefore, instead of considering a single modality for capturing traffic, we require techniques for multi-modal traffic detection. In other words, tools are needed to systematically take advantage of multiple technologies, whatever those may be now or in the future, to effectively capture urban traffic.

2. Related work

Many projects have attempted to accurately reconstruct mobility patterns by exploiting people's mobile devices. In the past, mobile phone tracking has been used as an approach to measure the flows of passengers between parts of a city and for estimating speeds and travel times [06, 07]. The results typically have low spatial resolution and are most effective for long-distance segments such as highways. Lu [01] categorizes past research in geospatial analysis to three groups; first, in a data-driven approach, spatiotemporal patterns are mined from trajectory data. Another direction of research aims to analyze and model dynamic interactions between people. Third, ''urban study'' focuses on studying human and vehicular flows in cities. A very popular approach for mapping traffic has been the use of proximity-based technologies, such as Bluetooth or WiFi traces [08, 09, 10, 11, 12, 13, 14]. These studies suggest that due to its current popularity and widespread usage, Bluetooth technology is not only useful for capturing individual mobility traces, but can be also used to analyze the spatiotemporal behavior of masses. Gauging the popularity of a technology such as Bluetooth is challenging and is likely to be a moving target. It is important to note that Bluetooth devices may operate in non-discoverable mode, and hence not be detectable. This means that only a subset of existing Bluetooth devices is technically observable. Estimates show the ratio of observable

Bluetooth devices to range between 2 % for Bremen, Germany to 7 % for Bath, UK [15, 16]. These results indicate that while potentially a great subset of the population has Bluetooth-capable devices, ceteris paribus only a small portion keeps their Bluetooth devices in discoverable mode. While 7 % is not necessarily a big portion of the population, nevertheless it is potentially greater than the approximate 3 % of the population that traditional transport surveys cover in any region. However, mobile handset manufacturers have recently opted to substantially limit the functionality on Bluetooth on their handsets. For instance, iOS devices by Apple are typically not detectable by Bluetooth, while recent Android devices are by default limited to a small-time window of a few seconds when they are detectable. Despite these developments, Bluetooth remains heavily used for traffic monitoring in the context of highways and major transport arteries, where the deployment of Bluetooth scanners at strategic locations allows for the approximation of macro-travel behavior [02, 03, 04]. Similarly, Barcelo et al. [17] made use of statistical methods (e.g., Kalman filtering) in order to estimate traveling time and origin–destination (OD) matrices in highways. In [18], opportunities for signal timing improvement were studied by identifying time periods with long travel times using Bluetooth-based vehicle re-identification, and in [19], Bluetooth was used to estimate when passengers get on and off a public transport bus. An important challenge that urban traffic sensor systems face is the threat of any single technology, such as Bluetooth, declining in popularity. For this reason, it is important to have techniques and methods to complement multiple modalities and use each one’s strength for improving our overall understanding of traffic and mobility. Most previous work that has relied on proximity technologies such as Bluetooth has used it as the sole modality, and possibly relied on manual observations for verification, or post-hoc correlation with other aggregated traffic data for cross-validation.

3. Multi Modal Traffic Data Management

We will explore the urban traffic analysis based on the paper "Urban traffic analysis through multi-modal sensing", Mikko Perttunen, Vassilis Kostakos, Jukka Riekki, Timo Ojala. Springer DOI 10.1007/s00779-015-0833-4. In this paper techniques to analyze traffic in the urban areas from sensor data are investigated with inductive loop traffic detectors and Bluetooth sensing. We will also go in exploring data fusion with additional modalities, such as WiFi and GPS, in order to study to which extent each additional modality improves the ability to reliably reconstruct movement trajectories in urban areas. Throughout this project and based on the paper given above, data mining techniques based on neural networks will be used for urban traffic analysis.

3.1 Data Mining

Data Mining is an analytic process designed to explore data (usually large amounts of data - typically business or market related - also known as "big data") in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. The ultimate goal of data mining is prediction - and predictive data mining is the most common type of data mining and one that has the most direct business applications. The process of data mining consists of three stages: (1) the initial exploration, (2) model building or pattern identification with validation/verification, and (3) deployment (i.e., the application of the model to new data in order to generate predictions).

Stage 1: Exploration. This stage usually starts with data preparation which may involve cleaning data, data transformations, selecting subsets of records and - in case of data sets with large numbers of variables ("fields") - performing some preliminary feature selection operations to bring the number of variables to a manageable range (depending on the statistical methods which are being considered). Then, depending on the nature of the analytic problem, this first stage of the process of data mining may involve anywhere between a simple choice of straightforward predictors for a regression model, to elaborate exploratory analyses using a wide variety of graphical and statistical methods in order to identify the most relevant variables and determine the complexity and/or the general nature of models that can be taken into account in the next stage.

Stage 2: Model building and validation. This stage involves considering various models and choosing the best one based on their predictive performance (i.e., explaining the variability in question and producing stable results across samples). This may sound like a simple operation, but in fact, it sometimes involves a very elaborate process. There are a variety of techniques developed to achieve that goal - many of which are based on so-called "competitive evaluation of models," that is, applying different models to the same data set and then comparing their performance to choose the best. These techniques - which are often considered the core of predictive data mining - include: Bagging (Voting, Averaging), Boosting, Stacking (Stacked Generalizations), and Meta-Learning.

Stage 3: Deployment. That final stage involves using the model selected as best in the previous stage and applying it to new data in order to generate predictions or estimates of the expected outcome.

The concept of Data Mining is becoming increasingly popular as a business information management tool, similar like Smart Cities, where it is expected to reveal knowledge structures that can guide decisions in conditions of limited certainty. Recently, there has been increased interest in developing new analytic techniques specifically designed to address the issues relevant to business Data Mining (e.g., Classification Trees), but Data Mining is still based on the conceptual principles of statistics including the traditional Exploratory Data Analysis (EDA) and modeling and it shares with them both some components of its general approaches and specific techniques.

However, an important general difference in the focus and purpose between Data Mining and the traditional Exploratory Data Analysis (EDA) is that Data Mining is more oriented towards applications than the basic nature of the underlying phenomena. In other words, Data Mining is relatively less concerned with identifying the specific relations between the involved variables. For example, uncovering the nature of the underlying functions or the specific types of interactive, multivariate dependencies between variables are not the main goal of Data Mining. Instead, the focus is on producing a solution that can generate useful predictions. Therefore, Data Mining accepts among others a "black box" approach to data exploration or knowledge discovery and uses not only the traditional Exploratory Data Analysis (EDA) techniques, but also such techniques as Neural Networks which can generate valid predictions but are not capable of identifying the specific nature of the interrelations between the variables on which the predictions are based.

3.2. Neural Networks

Neural Networks are analytic techniques modeled after the (hypothesized) processes of learning in the cognitive system and the neurological functions of the brain and capable of predicting new observations (on specific variables) from other observations (on the same or other variables) after executing a process of so-called learning from existing data. Neural Networks is one of the Data Mining techniques.

The first step is to design a specific network architecture (that includes a specific number of "layers" each consisting of a certain number of "neurons"). The size and structure of the network needs to match the nature (e.g., the formal complexity) of the investigated phenomenon. Because the latter is obviously not known very well at this early stage, this task is not easy and often involves multiple "trials and errors." (Now, there is, however, neural network software that applies artificial intelligence techniques to aid in that tedious task and finds "the best" network architecture.)

The new network is then subjected to the process of "training." In that phase, neurons apply an iterative process to the number of inputs (variables) to adjust the weights of the network in order to optimally predict (in traditional terms, we could say find a "fit" to) the sample data on which the "training" is performed. After the phase of learning from an existing data set, the new network is ready and it can then be used to generate predictions.

The resulting "network" developed in the process of "learning" represents a pattern detected in the data. Thus, in this approach, the "network" is the functional equivalent of a model of relations between variables in the traditional model building approach. However, unlike in the traditional models, in the "network," those relations cannot be articulated in the usual terms used in statistics or methodology to describe relations between variables (such as, for example, "A is positively correlated with B but only for observations where the value of C is low and D is high"). Some neural networks can produce highly accurate predictions; they represent, however, a typical a-theoretical (one can say, "a black box") research approach. That approach is concerned only with practical considerations, that is, with the predictive validity of the solution and its applied relevance and not with the nature of the underlying mechanism or its relevance for any "theory" of the underlying phenomena.

However, it should be mentioned that Neural Network techniques can also be used as a component of analyses designed to build explanatory models because Neural Networks can help explore data sets in search for relevant variables or groups of variables; the results of such explorations can then facilitate the process of model building. Moreover, now there is neural network software that uses sophisticated algorithms to search for the most relevant input variables, thus potentially contributing directly to the model building process.

One of the major advantages of neural networks is that, theoretically, they are capable of approximating any continuous function, and thus the researcher does not need to have any hypotheses about the underlying model, or even to some extent, which variables matter. An important disadvantage, however, is that the final solution depends on the initial conditions of the network, and, as stated before, it is virtually impossible to "interpret" the solution in traditional, analytic terms, such as those used to build theories that explain phenomena.

3.3. Deep Learning

Deep learning, while sounding flashy, is really just a term to describe certain types of neural networks and related algorithms that consume often very *raw* input data. They process this data through many layers of nonlinear transformations of the input data in order to calculate a target output.

Unsupervised *feature extraction* is also an area where deep learning excels. Feature extraction is when an algorithm is able to automatically derive or construct meaningful features of the data to be used for further learning, generalization, and understanding. The burden is traditionally on the data scientist or programmer to carry out the feature extraction process in most other machine learning approaches, along with feature selection and engineering.

Feature extraction usually involves some amount dimensionality reduction as well, which is reducing the amount of input features and data required to generate meaningful results. This has many benefits, which include simplification, computational and memory power reduction, and so on.

More generally, deep learning falls under the group of techniques known as *feature learning* or *representation learning*. As discussed so far, feature extraction is used to 'learn' which features to focus on and use in machine learning solutions. The machine learning algorithms themselves 'learn' the optimal parameters to create the best performing model.

Paraphrasing Wikipedia, *feature learning* algorithms allow a machine to both learn for a specific task using a well-suited set of features, and also learn the features themselves. In other words, these algorithms *learn how to learn*!

Deep learning has been used successfully in many applications, and is considered to be one of the most cutting-edge machine learning and AI techniques at the time of this writing. The associated algorithms are often used for *supervised*, *unsupervised*, and *semi-supervised* learning problems.

For neural network-based deep learning models, the number of layers are greater than in so-called *shallow learning* algorithms. Shallow algorithms tend to be less complex and require more up-front knowledge of optimal features to use, which typically involves feature selection and engineering.

In contrast, deep learning algorithms rely more on optimal model selection and optimization through model tuning. They are more well suited to solve problems where prior knowledge of features is less desired or necessary, and where labeled data is unavailable or not required for the primary use case.

In addition to statistical techniques, neural networks and deep learning leverage concepts and techniques from signal processing as well, including nonlinear processing and/or transformations. You may recall that a nonlinear function is one that is not characterized simply by a straight line. It therefore requires more than just a slope to model the relationship between the input, or independent variable, and the output, or dependent variable. Nonlinear functions can include polynomial, logarithmic, and exponential terms, as well as any other transformation that isn't linear.

Many phenomena observed in the physical universe are actually best modeled with nonlinear transformations. This is true as well for transformations between inputs and the target output in machine learning and AI solutions. As mentioned, input data is transformed throughout the layers of a deep learning neural network by artificial neurons or processing units. The chain of transformations that occur from input to output is known as the credit assignment path, or CAP.

The CAP value is a proxy for the measurement or concept of 'depth' in a deep learning model architecture. According to Wikipedia, most researchers in the field agree that deep learning has multiple nonlinear layers with a CAP greater than two, and some consider a CAP greater than ten to be very deep learning.

While a detailed discussion of the many different deep-learning model architectures and learning algorithms is beyond the scope of this article, some of the more notable ones include:

- Feed-forward neural networks
- Recurrent neural network
- Multi-layer perceptrons (MLP)
- Convolutional neural networks
- Recursive neural networks
- Deep belief networks
- Convolutional deep belief networks
- Self-Organizing Maps
- Deep Boltzmann machines
- Stacked de-noising auto-encoders

It's worth pointing out that due to the relative increase in complexity, deep learning and neural network algorithms can be prone to overfitting. In addition, increased model and algorithmic complexity can result in very significant computational resource and time requirements. It's also important to consider that solutions may represent local minima as opposed to a global optimal solution. This is due to the complex nature of these models when combined with optimization techniques such as gradient descent.

Given all of this, proper care must be taken when leveraging artificial intelligence algorithms to solve problems, including the selection, implementation, and performance assessment of algorithms themselves. While out of scope for this article, the field of machine learning includes many techniques that can help with these areas.

References

01. Lu Y, Liu Y (2012) Pervasive location acquisition technologies: opportunities and challenges for geospatial studies. Comput Environ Urban Syst 36(2):105–108

02. Haghani A, Hamedi M, Sadabadi KF, Young S, Tarnoff P (2009) Data collection of freeway travel time ground truth with bluetooth sensors. Transp Res Rec 2160(–1):60–68

03. Martchouk M, Street NC, Suite NE (2010) Analysis of freeway travel time variability using bluetooth detection. J Transp Eng 2051(July):1–30

04. Quayle S, Koonce P (2010) Arterial performance measures using MAC Readers-Portland's experience. Kittelson & Associates, Inc., Portland

05. Liu L, Andris C, Ratti C (2010) Uncovering cabdrivers' behavior patterns from their digital traces. Comput Environ Urban Syst 34(6):541–548

06. Caceres N, Wideberg JP, Benitez FG (2007) Deriving origin– destination data from a mobile phone network. Eng Technol 1(1):15–26

07. Gonzalez MC, Hidalgo CA, Barabasi A-L (2008) Understanding individual human mobility patterns. Nature 453(7196):779–782

08. Balazinska M, Castro P (2003) Characterizing mobility and network usage in a corporate wireless local-area network. ACM Press, New York, pp 303–316

09. Chaintreau A, Hui P, Crowcroft J, Diot C, Gass R, Scott J (2007) Impact of human mobility on opportunistic forwarding algorithms. IEEE Trans Mob Comput 6(6):606–620

010. Eagle N, Pentland A (2006) Reality mining: sensing complex social systems. Pers Ubiquitous Comput 10:255–268

011. Kostakos V, Nicolai T, Yoneki E, O'Neill E, Kenn H, Crowcroft J (2009) Understanding and measuring the urban pervasive infrastructure. Pers Ubiquitous Comput 13:355–364

012. McNett M, Voelker GM (2005) Access and mobility of wireless PDA users. ACM SIGMOBILE Mob Comput Commun Rev 9(2):40

013. Nicolai T, Kenn H (2007) About the relationship between people and discoverable bluetooth devices in urban environments. In: Mobility. ACM, New York, pp 72–78

014. Versichele M, Neutens T, Delafontaine M, de Weghe NV (2012) The use of bluetooth for analysing spatiotemporal dynamics of human movement at mass events: a case study of the ghent festivities. Appl Geogr 32(2):208–220

015. Nicolai T, Yoneki E, Behrens N, Kenn H (2006) Exploring social context with the wireless rope, vol 4277. Springer, Berlin

016. O'Neill E, Kostakos V, Kindberg T, Schieck AF gen, Penn, Fraser ADS, Jones T (2006) Instrumenting the city: developing methods for observing and understanding the digital cityscape. In: Ubicomp, vol 4206. Springer, pp 315–332

017. Barcelo J, Montero L, Marques L, Carmona C (2010) Travel time forecasting and dynamic od estimation in freeways based on bluetooth traffic monitoring. Transp Res Rec 2175(1):19–27

018. Day CM, Haseman R, Premachandra H, Brennan TM, Wasson JS, Sturdevant JR, Bullock DM (2010) Visualization and assessment of arterial progression quality using high resolution signal event data and measured travel time. Transp Res Rec 10–0039(2192):1–30