Mining spatiotemporal features of city traffic

Panagiotis Fafoutellis[1*], Emmanouil Kampitakis[1], Eleni I. Vlahogianni[1], Nectarios Koziris[2], George Yannis[1], John C. Golias[1]

[1]School of Civil Engineering, National Technical University of Athens

[2]School of Electrical and Computer Engineering, National Technical University of Athens

*Email: panfaf@mail.ntua.gr

Abstract

Short-term traffic forecasting is a field of research that has always attracted significant attention. The recent introduction of Machine Learning techniques in traffic forecasting has broadened the researchers' horizons, making fresher approaches possible. However, researchers should not disregard the importance of spatiotemporal relations of a road network and classic statistical modeling, which also provide better interpretation. In this paper, we detect the spatiotemporal relationships of the extended 2nd ring road network of Xi'an, China using Pearson's Correlation, Mutual Information and Dynamic Time Warping on the network's speed time series. The first two give an indication of the spatial dependency between road sections by comparing their speeds' contributions, while Dynamic Time Warping takes also into account the temporal evolution of the phenomenon. Results show that, although the first approach leads to an accurate Bayesian Network prediction model, the second one leads to an improved accuracy using the same modeling structure.

Keywords: *spatiotemporal relations, time series, trajectories, mutual information, dynamic time warping*

1 Introduction

Short-term traffic forecasting has always been a field of high research interest due to its significant importance to traffic flow management and the development of intelligent transportation systems and user-friendly information providing applications (Vlahogianni, Karlaftis & Golias, 2014). Accurate traffic forecasting is also essential to efficient traffic control and sustainable road network conditions as it reduces the levels of uncertainty of the decision-making process.

Nowadays, the extended use of smart devices and systems (GPS, smartphone, in-vehicle telematics etc.), which are able to track a huge amount of real-time mobility data, gives researchers the opportunity to develop prediction models that are more accurate and constantly updated, as well as of high temporal resolution. This has been the turning point that moved researchers' attention from classical statistic approach to Machine Learning data-driven models with the assistance of Data Mining and Big Data algorithms (Vlahogianni, Karlaftis & Golias, 2014). The aspect that made the development of such models possible is the massive growth of the computational power of modern computers the previous two decades, which are able to cope with the high computational complexity calculations required within some seconds but can also handle large amounts of data efficiently.

A very popular approach for short term traffic forecasting is to identify the relations between traffic flow variables, such as speed, of different road sections of a road network over time, using statistical metrics, which have very solid mathematical foundations (Karlaftis & Vlahogianni, 2010). Nevertheless, with the overwhelming use of Deep Learning, researchers seem to disregard the importance of demystifying the spatiotemporal dynamics of traffic to the prediction and efficient management of traffic conditions. Neural networks produce very accurate predictions of traffic parameters, as they can approximate almost any function, regardless of its degree of nonlinearity and without prior knowledge of its functional form (Vlahogianni, Karlaftis & Golias, 2005). In contrast to traditional Neural Networks and Machine Learning techniques, models that also involve the spatiotemporal relations of a road network provide better interpretation and insight of the mechanisms creating the predictions.

The general understanding so far on how spatial data may improve traffic forecasting has been limited by the lack of network wide traffic information (Vlahogianni et al., 2014). Interestingly, most approaches so far either capture spatial dependency of adjacent upstream and downstream links with a study link using correlation analysis or develop forecasting methods in a corridor test sample, where all links are connected sequentially together, assume a similarity between the behaviour of both parallel and adjacent links, and overlook the competitive nature of traffic links (Ermagun & Levinson, 2018).

Moreover, the temporal-spatial features in forecasting are usually addressed internally in advanced deep learning structures (Laña et al., 2019) or by resorting to more sophisticated approaches, namely by constructing useful inputs for traffic flow predictors through the extraction of correlations in the data (Stathopoulos & Karlaftis, 2002), (Vlahogianni et al., 2005), (Sun, Huang & Gao, 2012), (Vlahogianni, 2015), (Zhao et al., 2017); considering each location as a module in a modular network (Vlahogianni et al., 2007), (Vlahogianni, 2009); considering each location as a task in a multitask DBN model (Huang et al., 2014). Evidently, methods that increase the explanatory power of the forecasting models are usually preferred against the so called “black box approaches”, in case we want to gain managerial knowledge not only what traffic conditions are expected, but also on why these conditions are most likely to occur.

The present paper attempts to introduce a much more sober analysis of spatiotemporal dependencies disengaged from deep learning with the aim to increase the understanding on the following research questions:

- Do spatial and temporal traffic dependencies exist in a road network?
- What are the impacts of spatiotemporal dependencies in short-term traffic forecasting?

To this end, this paper implements concepts spanning from classical correlation analysis, to Information Theory, Time Sequence Analysis and Bayesian Networks. The proposed methodological approach is implemented on the road network of the city of Xi’an, China using trajectory data provided by Didi Chuxing Technology Co, a Chinese taxi and private car-hailing company.

2 Identifying Spatiotemporal Patterns

2.1 Linear Correlation and Mutual Information

To answer the question on whether there are road sections that are correlated in terms of travel speed, we apply the concept of Mutual Information (MI) and compare the results with the classical correlation analysis (Pearson's correlation). Based on information theory, MI of two random variables is a metric that quantifies the amount of information obtained for the first random variable when observing the other random variable. Unlike the classical correlation analysis, the mutual information takes into account nonlinear correlations as well, because the computed measure is not connected to the linear or non-linear evolution rules of the quantities involved, but to Shannon Entropy (Abarbanel, 1996), (Kantz & Schreiber, 1997). Let x_n and y_n two equally spaced sets of random variables with joint probability density $p(x_n, y_n)$ and individual probability densities $p(x_n)$ and $p(y_n)$. The MI $I(x_n, y_n)$, which quantifies the expected information gained about x_n when observing y_n is given by:

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

This approach has two main limitations: first, it is static in a sense that time is not introduced in the analysis of travel speed interrelations between different network locations. Second, interrelations are assessed in a pairwise manner without letting understanding on how information from multiple locations may interact with each other and affect predictability.

2.2 Distance Based Time Series Similarity

To address the first point of criticism mentioned above, the present work implements the fast dynamic time warping (Fast DTW). Dynamic time warping (DTW) is a dynamic programming technique to find an optimal alignment between two given time series with the objective to minimize a specific distance measure (Berndt & Clifford, 1994). For the time series $X = x_1, x_2, \ldots, x_n$ and $Y = y_1, y_2, \ldots, y_n$, DTW distance is given by the following recurrent equation to the matrix $\gamma(i \ldots n, j \ldots n)$ using dynamic programming (Lee et al., 2017):

$$\gamma(i,j) = dist(x_i, y_j) + \min[\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1) \qquad (2)$$

The path that provides the optimum, namely minimum, distance is the warping path. The DTW distance $DTW(X,Y) = \gamma(n,n)$ is the Euclidean distance along the warping path. DTW has a quadratic time and space complexity that limits its use to only small size time series data sets.

To alleviate this limitation, an extension on classical DTW may be used, which first transforms high dimensional time series to low dimensional time series and then obtain DTW distances on the low dimensional time series. This extension known as Fast DTW operates on three steps (Salvador & Chan, 2007): coarsening to reduce the dimensionality, projection to calculate DTW distance in the lowest time series resolution, and refinement to project the warping path to an incrementally higher resolution. The last two steps repeat until the path is projected to the full time series resolution.

2.3 Bayesian Network Classifier

Finally, to address the limitation of pairwise time series comparison, we develop a Bayesian Network, which presents the relations between all the road sections and is based on the calculation of conditional probability between their speeds' contributions. A Bayesian Network (BN) is a directed acyclic graph whose nodes represent variables. The weights of the connections of the nodes are proportional to the relationship between the variables of the corresponding nodes. With the above model, it is possible to calculate the conditional probability of a variable getting a certain value when knowing the values of all the variables that are connected to it (child nodes) (Pearl, 2000).

The BN for a set of variables $X_i = \{X_1, ..., X_n\}$ also consists of a set $P_i = \{P_1, ..., P_n\}$ of local conditional probability distributions associated with each node and its parents. BN's causal interpretation is as follows: a directed edge from one variable to another Y, represents the claim that *X* is a direct cause of *Y* with respect to other variables in the DAG (Friedman et al., 1997). The joint distribution *p* can be factorized as a product of conditional probabilities, by specifying the distribution of each node conditional on its parents. For a given structure *B* of a BN, the joint probability distribution $P(X)$ for *X* can be written as:

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

where pa_i denotes the set of parents for X_i. The BN can be used as a classifier of X_i inputs to a set of classes, in our case, the travel speed classes (*C*), by the rule (Friedman et al., 1997):

$$classify(x_1,...,x_n) = \arg\max_n \, p(R) \prod_{i=1}^{n} p\left(X_i = x_i \,|\, C\right) \quad (2)$$

By the BN classification task, the influence of each variable (in our case the lagged information of volume and occupancy from both the upstream and downstream location and the location of interest) can be determined with respect to the prevailing speed class *C*. The selection of influential spatio-temporal patterns of travel speeds will be based on the mutual information criterion. Mutual information quantifies the amount of information flow between a node X_i and the knowledge of traffic speed levels *C*. The mutual information I(X,C) between a variable *X* and a class C measures the expected information gained about *C*, after observing the value of the variable *X*:

$$I\left(X,C\right) = \sum_{X,C} P\left(X \,|\, C\right) P(C) \log \frac{P\left(X \,|\, C\right)}{\sum_{c \in C} P\left(X \,|\, C\right) P\left(c\right)} \quad (3)$$

3. Implementation and Findings

3.1 Data Preprocessing

Data preprocessing is an essential procedure when conducting statistical analysis or applying machine learning techniques. Well-prepared input data lead to better performing and easier trained and tuned prediction models. The dataset used in this paper consists of the about 110000 trajectories per day of Didi's vehicles in Xi'an for 2nd to 30th of November of 2016. Each trajectory corresponds to the exact position of the vehicle per 2-4 seconds. More specifically, the attributes of the data are the latitude and the longitude of each point of the trajectory, an ID number specifying the route, a second ID number specifying the driver and the timestamp of the moment that the vehicle was at the particular position.

First, the coordinates of the points from the Chinese State Bureau of Surveying and Mapping coordinate system (GCJ-02) were transformed to the World Geodetic System (WGS 84) in order to depict properly on some of the most well-known web maps, such as Open Street Map, and to make any further processing possible using the “eviltransform” python library. Second, we apply map-matching to the points of our dataset using a nearest neighbor relationships (Tveite, 2014). In general, map matching is the procedure of matching recorded GPS traces to real-world road network edges, while at the same time correcting the system’s error at the time of recording. In our case, each of the recorded points was matched with a road section of Xi’an’s road network, as downloaded from Open Street Map. Last, we calculate the speed of the vehicle’s movement from each point to the next of the same route, as the Euclidean distance of the two points to the difference of their timestamps. This way, it is possible to generate the time series of the speed of each one of the road network’s sections.

In order to calculate the time series of speeds of each road section, our data were grouped by the road section ID and by the chosen time-step. For the needs of this paper, a time-step of 1 hour was used, resulting in two time series for each road section. The value of the speed of each time-step of each road section was calculated as the average speed of all vehicles that passed from the road section the specific time period.

It is worth mentioning that road sections that did not have any record on any of the twenty-nine chosen days were excluded from further analysis, as it is clear that they do not play an important role in Xi’an’s transportation system. The same applies to road sections that do not have any record for more than an hour on any of the twenty-nine days. Figures 1 and 2 show the available full-length time series and one day time series of a specific road section; although a daily cyclicity is evident, there seem to be some short term features that may significantly affect the magnitude and evolution of speed and, consequently, the prediction accuracy.

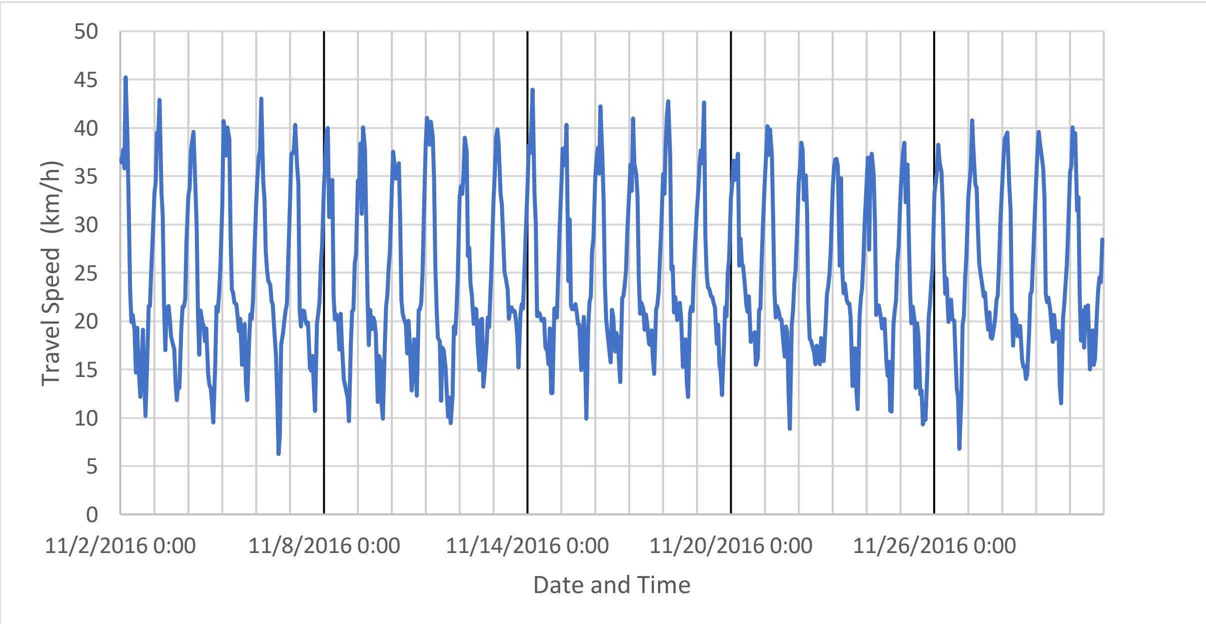

Figure 1: Sample 30-days time series

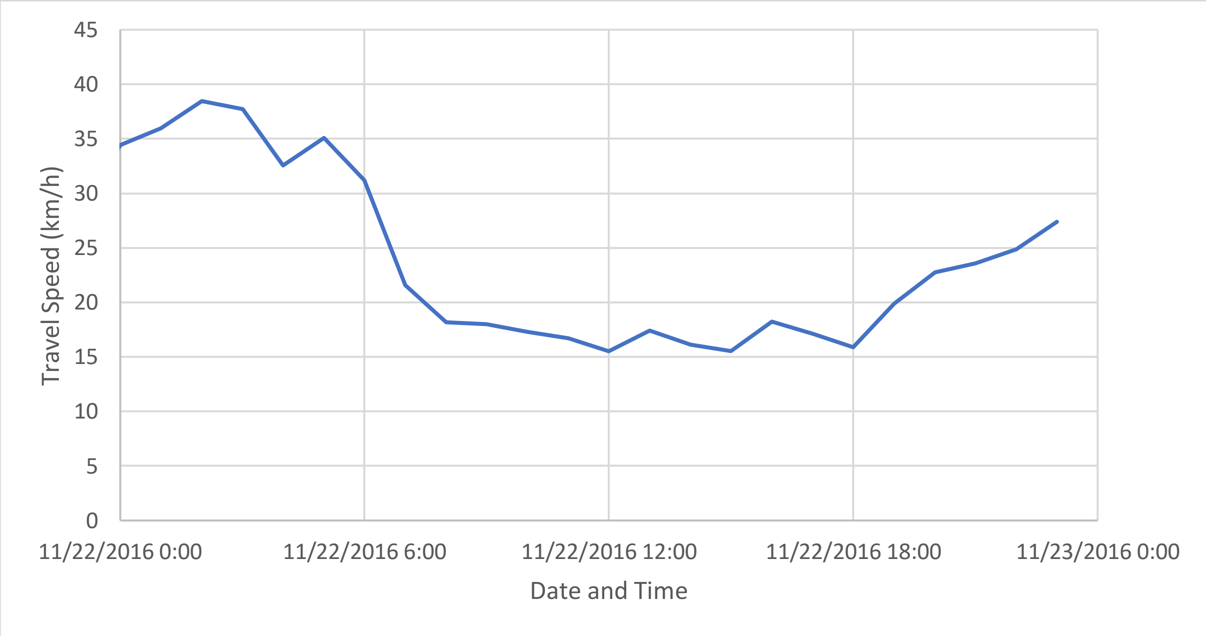

Figure 2: Sample 1-day time series

3.2 Do spatial and temporal traffic dependencies exist in a road network?

To address this research question, the Pearson's correlation coefficient between each pair of time series were calculated and are presented as a heatmap in Figure 3. This heatmap gives a clear indication of which sections are related to each other. In addition, mutual information of each pair of the series was calculated. The results are presented below as a heatmap in Figure 4. In both heatmaps, the lighter the color bands the stronger the relationship between road sections. In the two heatmaps, there are some common patterns that are clearly noticeable. However, MI criterion seems to produce lower values in terms of the strength of the identified spatial patterns. By examining each column (or row) separately, one can detect which sections are most related to the one of the columns.

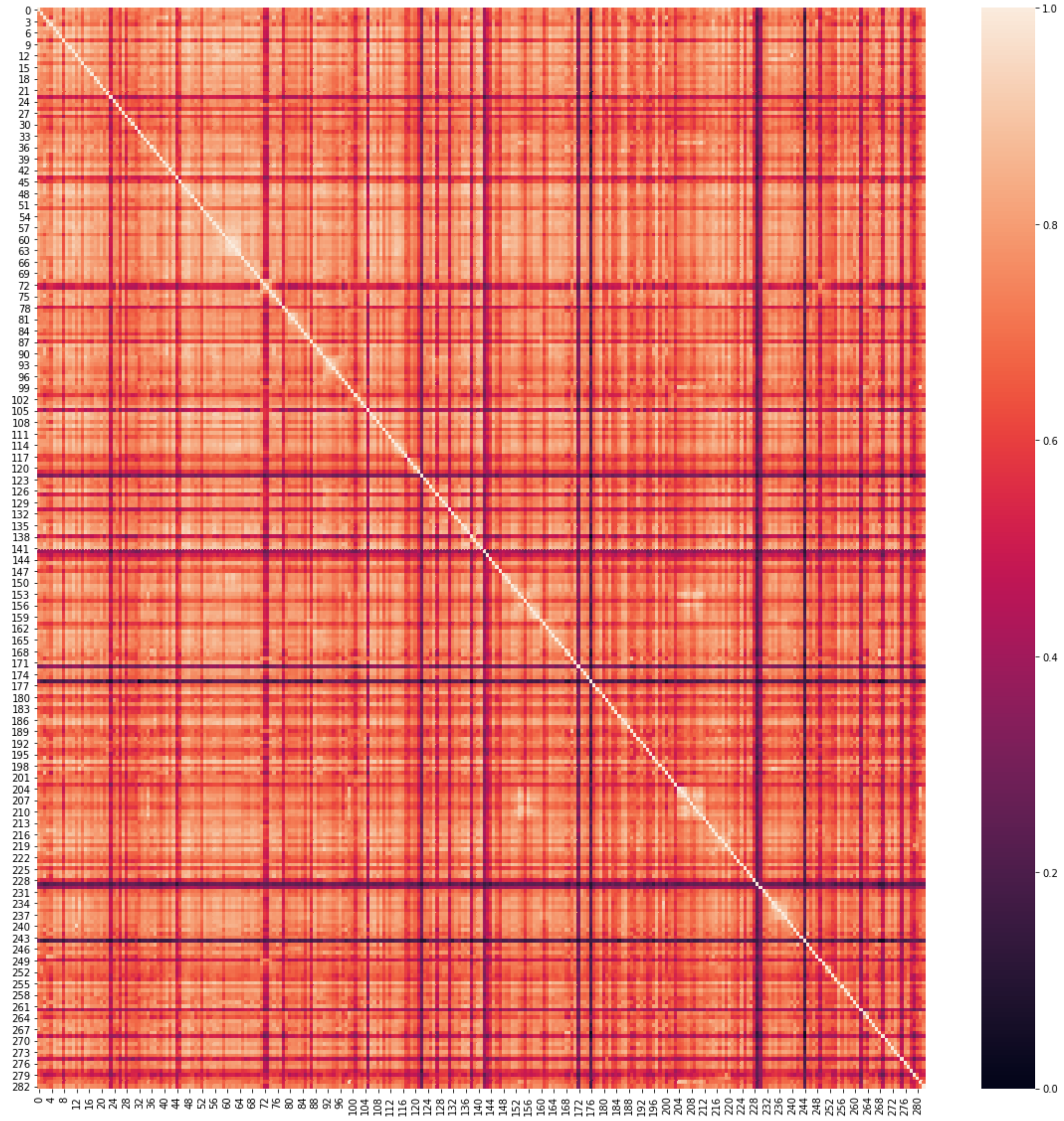

Figure 3: Correlation heatmap of time series

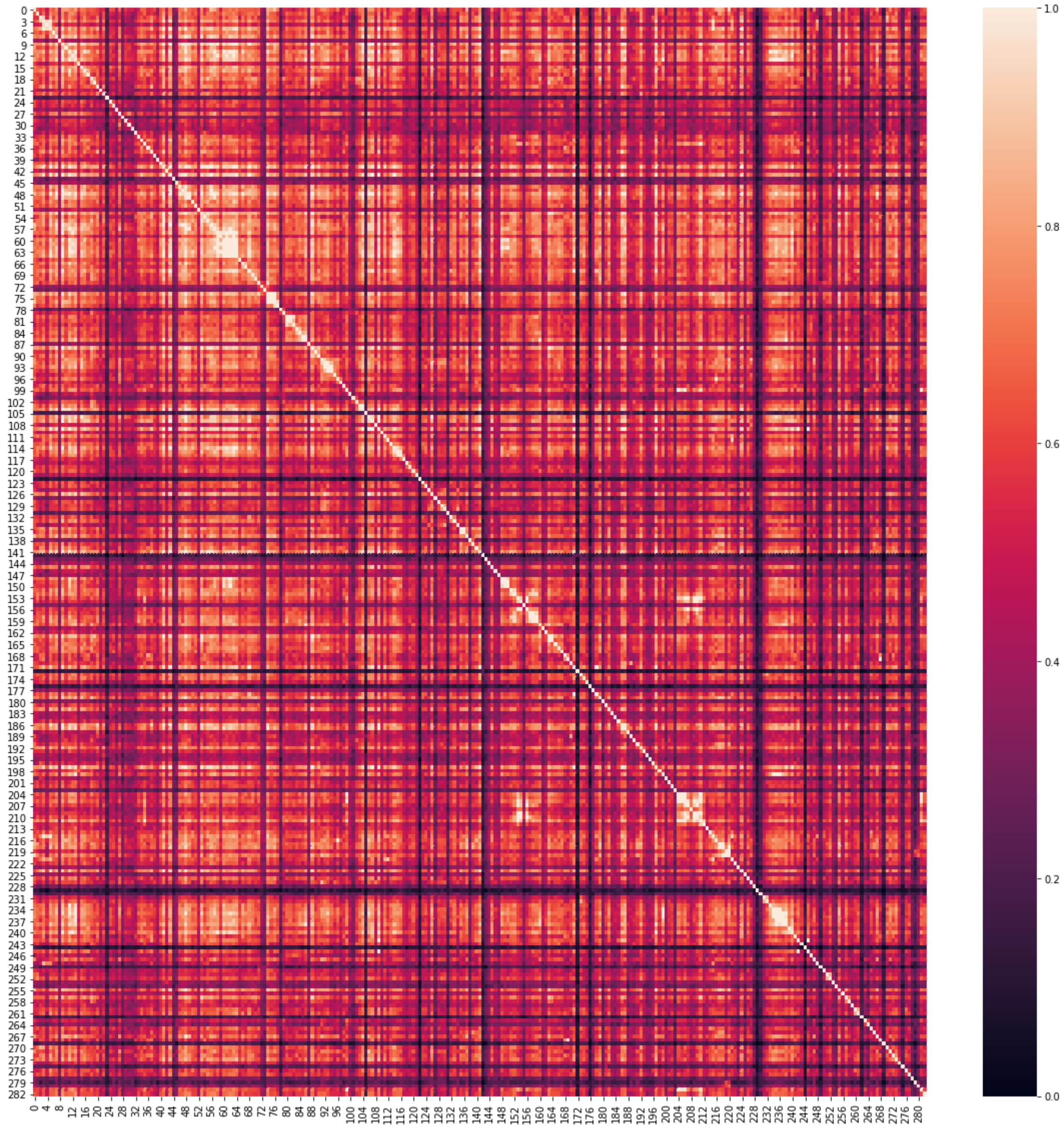

Figure 4: Mutual Information heatmap of time series

In order to compare the two metrics to each other, but also to the ones introduced later in this paper, the road section with ID number 28258922 on Open Street Map was selected to be presented in detail. The above road section is one of the most crowded road sections in Xi'an at the centre of the city. The exact position of the road section is shown on Figure 5.

Figure 5: Selected road section (green) on Xi'an City map.

Figure 6 depicts the 20 most related road sections (red) to the selected section, in terms of Pearson's Correlation (left) and Mutual Information (right). It seems that the two approaches capture different spatial patterns on the same dataset. The impacts of these differences should be further investigated in terms of prediction accuracy.

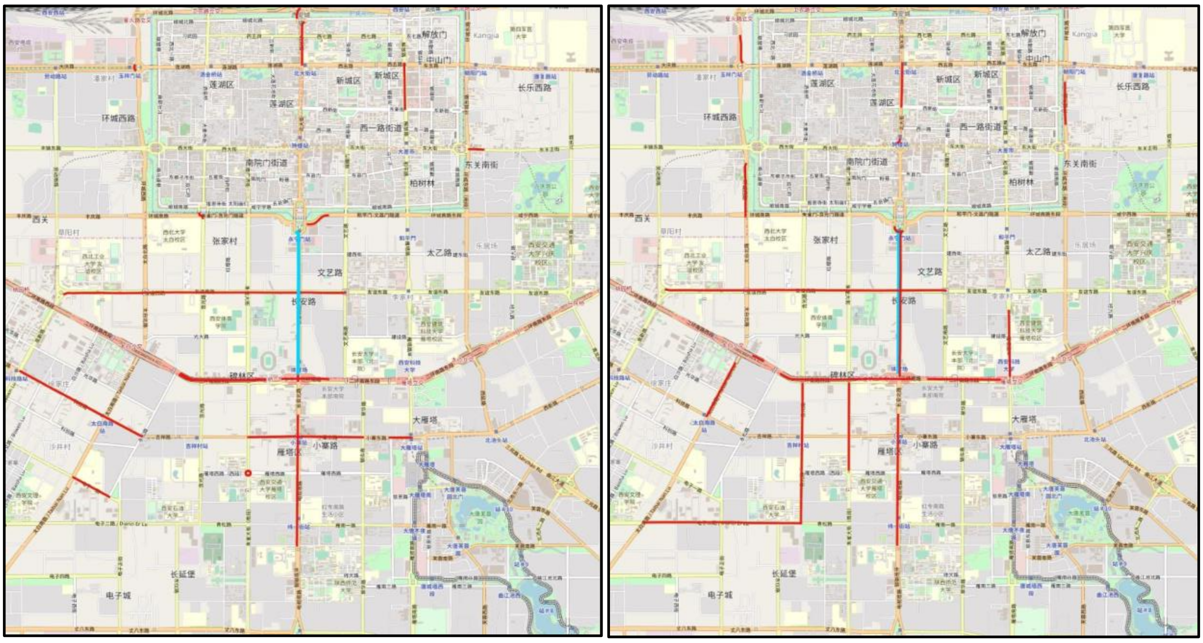

Figure 6: The 20 most related road sections (red) to the selected section (blue), in terms of Pearson's Correlation (left) and Mutual Information (right)

Further, the implementation results of the Fast DTW algorithm on the available 1-h times series, which gives as an indication of which road sections' speed are related to each other- in terms of temporal evolution- is seen in Figure 7. Smaller value (darker color) means higher correlation. Figure 8 shows the 20 most related road sections to the chosen one, in terms of Fast DTW distance. Compared to the detected patterns in Figure 5, there are clear differences between the spatio-temporal correlations and the spatial correlation detected using MI or linear correlation.

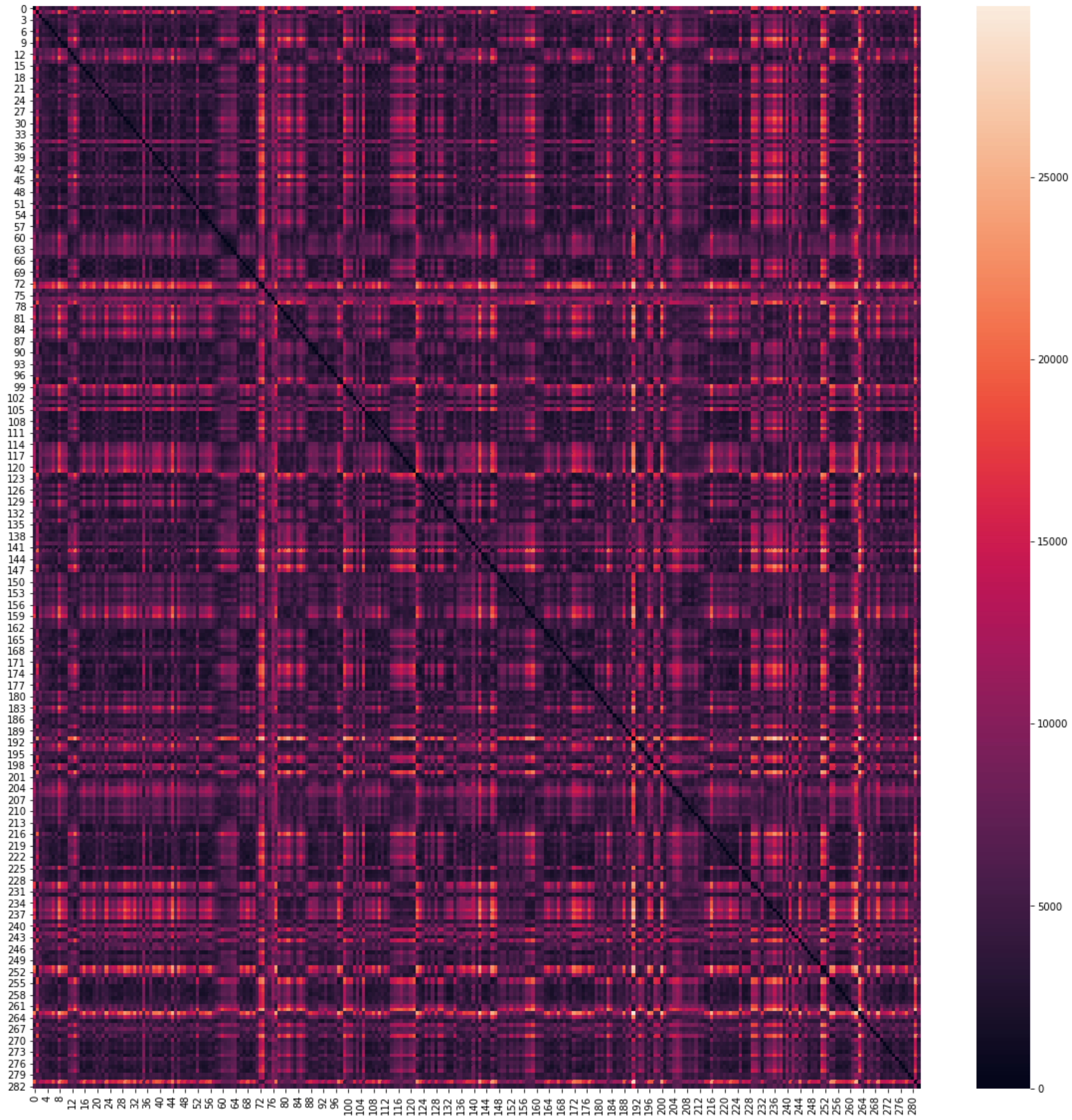

Figure 7: Dynamic Time Warping heatmap of 1-hour time step time series

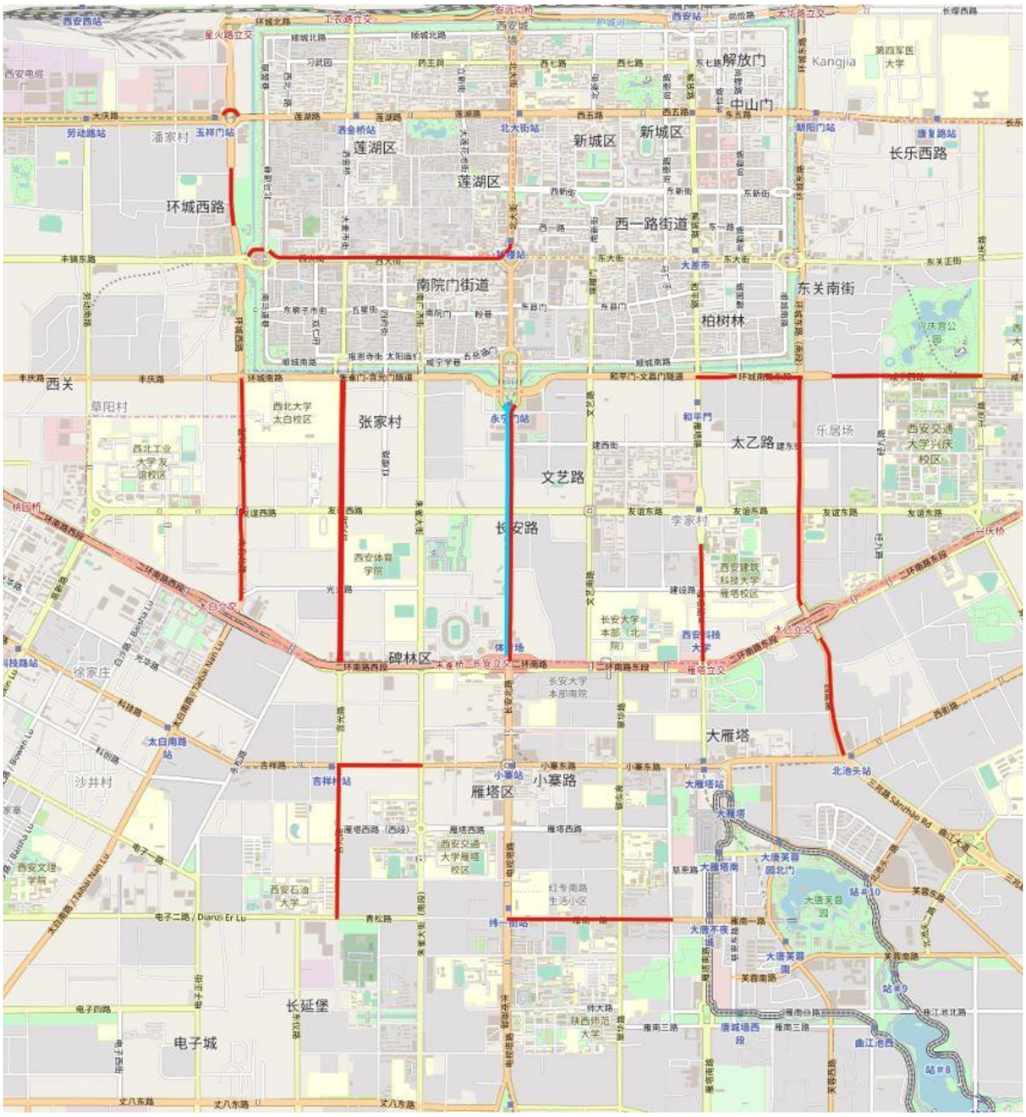

Figure 8: The 20 most related road sections (red) to the selected section, in terms of Dynamic Time Warping Distance

3.3 Comparison between Mutual Information and Dynamic Time Warping

In order to compare each method's results, we proceed to developing two prediction models of the speed of the target road section. Both models are Bayes Network Classifiers that assign the section's speed to three balanced classes: <20, 20-26, >26 km/h, which is a reasonable choice for signalized road sections, especially when we refer to travel speeds, including possible stops (e.g. upstream of a signalized intersection), as done in our case.

The first model was "forced" to create the Bayesian Network and predict using only data from the twenty road sections with the highest Mutual Information value, while the second using only the twenty road sections with the smaller values of DTW distance.

As it can be clearly seen in Table 1, the model using Mutual Information as the metric to choose which road sections to include outperforms the second one. The accuracy of the two models is 89% and 85.6% respectively. Hence, one can assume that using Mutual Information is a more accurate choice for the present application. This can be explained by considering again the definitions of the two metrics. Dynamic Time Warping is a measure of similarity between two time series, which is highly affected by the absolute size of the section's speed and only slightly by the time-series' pattern. On the other hand, the estimation of Mutual Information takes into account the trend of the timeseries and the proportional variation of speeds rather than their absolute prices.

Table 1: Classification metrics of the two models developed

Metrics	Model 1 (Mutual Info)	Model 2 (DTW)
Accuracy	89%	86%
Recall (Sensitivity)	89%	86%
Precision	89%	86%
F1 - score	89%	86%

3.4 What are the impacts of spatiotemporal dependencies in short-term traffic forecasting?

In order to identify if the analysis conducted in the previous chapters is relevant and improves the prediction task, we evaluate the findings by comparing them to the performance of a Bayesian classifier that is "free" to make predictions using data from any road section. In this case, each road section's contribution to forecasting is proportionate to its probabilistic relationship with the selected one. The classification results are summarized in Table 2. The accuracy of this model is 84.5%.

Table 2: Classification metrics of Naïve Bayes model

Metrics	Model 3 (Naïve Bayes)
Accuracy	84%
Recall (Sensitivity)	84%
Precision	85%
F1 - score	85%

Results indicate that the models presented at the previous chapter produce obviously better predictions. The first one, which uses the sections with the highest Mutual Information to the selected one, is performing noticeably better, while the second only slightly but still better. This result highlights the usefulness of performing spatiotemporal analysis.

Moreover, the above procedure decreases the dimensionality of the problem, which is a very common issue when using Machine Learning algorithms. In the current case, we originally had 696 individuals (time periods) with 283 attributes (road sections' speeds) each, which is a really high value. After performing spatiotemporal analysis feature selection, we used only the 20 most related attributes. Furthermore, the above procedure reduces the computational resources needed, which is equally important.

4 Conclusions

Analyzing large scale spatiotemporal characteristics of a road network before proceeding to the development of prediction models is essential in order to produce accurate predictions. In addition, such models provide better interpretation of the spatio-temporal evolution of traffic.

General use metrics, such as Pearson's Correlation, as well as more specialized for time series such as Dynamic Time Warping's distance and Mutual Information provide a clear insight into the spatiotemporal relationships of the road sections of an urban road network. These relationships occur from the relative position of the road sections and the traffic flow its one serves; therefore, they provide explainable results.

In order to underline the importance of identifying spatio-temporal traffic patterns, a simple structured Bayesian Network was developed to classify speeds based on the identified spatio-temporal traffic patterns. Although the model that was used is not the best-suited choice for time series data, the improved performance by introducing the network level spatiotemporal data is noticeable and important for the standpoint of accuracy and computational efficiency.

Future research will certainly include processing of data from more days, in order to have more generalized conclusions. The above requires a more powerful processing system in terms of hardware, as well as software, because the vast amount of data involved lead to Big Data Analysis approach. Second, more specialized and cutting-edge Machine Learning techniques will be used, because there seems to be enough space for optimization to achieve accurate single step ahead and larger time horizons forecasting.

5 References

Abarbanel, H. D. (1996). Analysis of observed chaotic data.

Berndt, D. J., & Clifford, J. (1994). Using Dynamic Time Warping to Find Patterns in Time Series. AAA1-94 Workshop on Knowledge Discovery in Databases.

Dietrich, C. (1991). Uncertainty, Calibration and Probability: The Statistics of Scientific and Industrial Measurement.

Ermagun, A., & Levinson, D. (2018). Spatiotemporal traffic forecasting: review and proposed directions. Transport Reviews, 38(6), 786-814.

Friedman, N., Geiger, D. & Goldszmidt, M., (1997). Bayesian Network Classifiers. Machine Learning, Vol. 29, 131–163. Kluwer, Boston.

Huang, W., Song, G., Hong, H., & Xie, K. (2014). Deep architecture for traffic flow prediction: deep belief networks with multitask learning. In IEEE Transactions on Intelligent Transportation Systems, 2191-2201.

Hvistendahl, M. (2013). Foreigners Run Afoul of China's Tightening Secrecy Rules. SCIENCE.

Kantz, H., & Schreiber, T. (1997). Non-linear time series analysis.

Karlaftis, M. G., & Vlahogianni, E. I. (2010). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. Transportation Research Part C.

Kvålseth, T. O. (1991). The relative useful information measure: some comments.

Laña, I., Lobo, J. L., Capecci, E., Del Ser, J., & Kasabov, N. (2019). Adaptive long-term traffic state estimation with evolving spiking neural networks. Transportation Research Part C.

Lee, M., Lee, S., Choi, M.-J., Moon, Y.-S., & Lim, H.-S. (2017). HybridFTW: Hybrid Computation of Dynamic Time Warping Distances. IEEE Access.

Mangerman, D. M., & Mitchell, M. P. (n.d.). Parsing a Natural Language Using Mutual Information Statistics.

Pearl, J. (2000). Causality: Models, Reasoning and Inference. Cambridge University Press.

Ross, B. C. (2014). Mutual Information between Discrete and Continuous Data Sets.

Salvador, S., & Chan, P. (2007). FastDTW: Toward accurate dynamic time warping in linear time and space. In Intelligent Data Analysis, 561-580.

Silva, D., & Batista, G. (2015). Speeding Up All-Pairwise Dynamic Time Warping Matrix Calculation.

Stathopoulos, A., & Karlaftis, M. G. (2002). Modeling Duration of Urban Traffic Congestion. Journal of Transportation Engineering.

Sun, S., Huang, R., & Gao, Y. (2012). Network-scale traffic modeling and forecasting with graphical lasso and neural networks. Journal of Transportation Engineering.

Tveite, H. (2014). The QGIS NNJoin Plugin. Retrieved from http://arken.nmbu.no/~havatv/gis/qgisplugins/NNJoin/#

Vlahogianni, E. I. (2009). Enhancing Predictions in Signalized Arterials with Information on Short-Term Traffic Flow Dynamics. Journal of Intelligent Transportation Systems.

Vlahogianni, E. I. (2015). Computational intelligence and optimization for transportation big data: challenges and opportunities. Engineering and Applied Sciences Optimization.

Vlahogianni, E. I., Kalaftis, M. G., and Golias, J. C. (2005). Optimized and Meta-Optimized Neural Networks for Short-Term Traffic Flow Modeling: A Genetic Approach, Transportation Research Part C: Emerging Technologies, Volume 13, Issue 3, 211-234.

Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2005). Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach. Transportation Research Part C.

Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. Transportation Research Part C.

Vlahogianni, E. I., Karlaftis, M. G., Golias, J. C. (2007). Spatio-Temporal Short-Term Urban Traffic Volume Forecasting Using Genetically-Optimized Modular Networks, Computer-aided Civil and Infrastructure Engineering, in press.

Zhao, Z., Chen, W., Wu, X., Chen, P. C., & Liu, J. (2017). LSTM network: a deep learning approach for short-term traffic forecast. In IET Intelligent Transport Systems, 68-75.