Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

International Symposium of Transport Simulation (ISTS'18) and the International Workshop on Traffic Data Collection and its Standardization (IWTDCS'18)

Assessing spatiotemporal correlations from data for short-term traffic prediction using multi-task learning

Rafael Mena-Yedra[a,b,*], Jordi Casas[b], Ricard Gavaldà[a]

[a]*Universitat Politècnica de Catalunya, Campus Nord, Calle Jordi Girona, 1-3, 08034 Barcelona, Spain*
[b]*Aimsun SL, Ronda de la Universitat, 22b, 08007 Barcelona, Spain*

Abstract

Traffic flow prediction is a fundamental problem for efficient transportation control and management. However, most current data-driven traffic prediction work found in the literature have focused on predicting traffic from an individual task perspective, and have not fully leveraged the implicit knowledge present in a road-network through space and time correlations. Such correlations are now far easier to isolate due to the recent profusion of traffic data sources and more specifically their wide geographic spread.

In this paper, we take a multi-task learning (MTL) approach whose fundamental aim is to improve the generalization performance by leveraging the domain-specific information contained in related tasks that are jointly learned. In addition, another common factor found in the literature is that a historical dataset is used for the calibration and the assessment of the proposed approach, without dealing in any explicit or implicit way with the frequent challenges found in real-time prediction. In contrast, we adopt a different approach which faces this problem from a point of view of streams of data, and thus the learning procedure is undertaken online, giving greater importance to the most recent data, making data-driven decisions online, and undoing decisions which are no longer optimal. In the experiments presented we achieve a more compact and consistent knowledge in the form of rules automatically extracted from data, while maintaining or even improving, in some cases, the performance over single-task learning (STL).

Keywords: traffic prediction; multi-task learning (MTL); data stream learning; high-dimensionality;

A more comprehensive version of this paper with additional experimentation, that exceeds the conference's size limit, can be found at `https://www.dropbox.com/s/pog6805a0en9nzu/MenaYedraGavaldaCasas_MTL.pdf?dl=0` for the reviewers' convenience.

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mail address: rafael.mena@cs.upc.edu ; rafael.mena@aimsun.com

1. Background and Motivation

Solutions for traffic congestion include the provision of information (Traveler Information Services) to both the driver and the traffic manager. The former can make use of this predictive information to perform suitable travel decisions before the departure time (pre-trip information) and during the journey (on-trip information), while the traffic control agent in a Traffic Management system can take the appropriate decisions for accommodating traffic flow in an efficient, effective and safe manner whose assortment of techniques and strategies is known as Active Traffic Management (ATM). The use of this kind of information has a beneficial impact on the network performance in terms of throughput, congestion length and average network speeds.

One such tool that can be used in this process is Aimsun Live which is based on the Aimsun transport modeling software [Casas et al. (2010)], which includes a decision support system for real-time traffic management which is used by traffic control centers to make real-time decisions about the management of a road network, allowing a dynamic forecast of future traffic conditions based on the current state of the network and the evaluation of incident response or traffic management strategies.

In order to examine the utility of incorporating data generated constantly in a traffic network into such a system and evaluate its benefits, research has been conducted as part of the EU research project "SETA" (H2020-ICT-2015), which is setting out a ubiquitous data and service ecosystem for better metropolitan mobility, to examine how short term prediction can be improved through the use of multiple, highly diverse sources.

2. Methods and Technical Solutions

2.1. *Short-term traffic forecasting.*

Because road traffic is the visible result of the complex interplay between traffic demand (the amount of travelers making a trip at a particular place and time) and traffic supply (network infrastructure), when modeling it is usual to find that the input–output data relationship is noisy and that the relationships between these variables are multivariate and (highly) nonlinear van Lint and van Hinsbergen (2012), additionally the process is usually high-dimensional, non-stationary and tackled in real-time.

In the literature, there are two main approaches adopted for road traffic prediction: model-driven and data-driven. Model-driven approaches try to reproduce the road network behavior through simulation, and depending on the level of detail and the underlying traffic flow theory in which are based they could be distinguished among microscopic, mesoscopic, macroscopic and hybrid variants Barceló (2010); Treiber and Kesting (2013). One requirement of such model-driven approaches to obtain accurate predictions is to have a detailed knowledge about the network topology.

Conversely, data-driven approaches aim to reproduce the input–output mapping but usually neglecting the underlying data generation process and disregarding, in general, the network topology. Despite this, integrating the network spatio-temporal information within the short-term traffic prediction task is of ultimate importance Ermagun and Levinson (2016). This branch has taken advantage from the fact that over time different measuring devices have been deployed within road networks to measure and verify road traffic conditions. For additional review references, see Ermagun and Levinson (2016); van Lint and van Hinsbergen (2012); Vlahogianni et al. (2014).

2.2. *Multi-task learning.*

Multi-task learning (MTL) is a paradigm in the realm of machine learning itself. The core idea, as was defined by Caruana (1997), is to act as an inductive bias causing a model to prefer the hypotheses best explaining the set of related tasks simultaneously. By doing this, the main goal is to improve generalization performance by leveraging the domain-specific information contained in the training signals of these related tasks. Among the benefits of MTL are: (1) biases the model to prefer hypotheses that other tasks also prefer, which improves the generalization for new tasks in the same domain, (2) implicitly does data augmentation by averaging the noise patterns among tasks, (3) it allows to differentiate between relevant and irrelevant features especially when the data is noisy or high-dimensional as other tasks will provide additional evidence for the relevance or irrelevance of those features, (4) having a regularization effect to avoid the risk of overfitting the random noise of a single task. For a recent overview about MTL, with special emphasis on the deep learning field, the reader is referred to Ruder (2017).

2.3. MTL to reconstruct traffic prediction.

Deciding which tasks are going to be grouped to be learnt together is a crucial decision in the MTL paradigm. Caruana (1997) demonstrates, with different experimental scenarios, that only when similar and related tasks are jointly learnt is the overall performance improved, while the opposite, learning unrelated tasks, can lead to performance degradation.

In the traffic research literature, most works using a MTL paradigm have been applied to traffic flow modeling for predicting multiple forecasting steps using a methodology based on neural networks Huang et al. (2014); Jin and Sun (2008); Sun (2009).

In this work, we assess if adopting a MTL paradigm into our rule-based system Mena-Yedra et al. (2017) has a beneficial impact in the prediction performance, interpretability, or efficiency. This is going to be evaluated by an approach that tries to leverage the fundamental relationship in traffic between traffic volume and occupancy for each specific detector or spatial point in the network. The basic idea is that the proposed predictive system enhanced with the MTL paradigm unveils spatiotemporal correlations in the road network or associated with qualitative variables such as the time to better perform in the current predictive task, while giving an interpretable reasoning of the most influential factors.

3. The Adarules model

Adarules Mena-Yedra et al. (2017) is a framework for predictive modeling, based on a multi-level decision-tree structure which is built online using streaming data. It is inspired by the works of Gama (2010) applied to data streaming scenarios Almeida et al. (2013), but tailored to the requirements for this application; i.e. the necessity to handle both structured and non-structured data as well as to take advantage from contextual information and data coming from multiple sources and technologies in a high-dimensional non-linear real-time scenario where noisy sources are common. The system is composed of rules (in the form: if antecedent(s) is satisfied, then [...]) that are automatically extracted from streaming data. These rules are self-contained since they contain modules for anomaly detection, change detection, data structures to collect statistics from the gathered samples and a customizable set of individual prediction models which are learnt from the historical samples covered by the rule but can give a response using real-time input data. This set of rules is both built and maintained in an autonomous manner, in a non-stationary system which can change over time.

Among the main features and motivation to build up this predictive system are:

1. The model complexity grows up to adapt itself to that of the modeled problem. This means that the number of rules is not fixed beforehand (i.e. non-parametric approach) and they are extracted automatically from streaming data. This certainly require some control and regularization to avoid overfitting the data and keeping a good generalization performance.
2. Adaptation to change: considering gradual seasonal changes (seasonality), and sudden changes through change detection (concept drift detection in the machine learning jargon).
3. The number of assumptions have been kept to a minimum to rely solely on data to make decisions and associations discovery. In this way, it is possible to start modeling a network even when there is a scarce amount of historical data or there is still limited knowledge about the road network geometry. This has been intended to consider the inherent volatility in traffic supply/demand information. To this end, an empirical approach where the rules discovery is made through maximizing the outcome probability and the forecasting modeling is made through the minimization of the squared error.
4. Autonomous approach to minimize human intervention to decide the amount of historical data to use and when to train again the models.
5. Interpretability to be able to extract and analyze which factors are most influential in the traffic prediction task.
6. Robustness facing outliers and missing data due to the inherent noise and failures found in traffic data.

Because the methodology is explained in Mena-Yedra et al. (2017), we name here the main differences referred to the MTL setting. More specifically, to scale up the individual scores in the split decision mechanism for each task into a single score for the tasks group being modeled in the MTL setting, we do a combination to avoid bias towards situations with a small group of high scores. We first order the n individual scores x, and obtain a robust measure of central tendency using the median $m(x)$, i.e.:

$$m(x) = \begin{cases} x_{\frac{n+1}{2}} & n \text{ odd} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) & n \text{ even} \end{cases}$$

Then, we weight these scores using a penalty factor $\rho(x)$ that takes into account the dispersion of the scores and their range using the first and last element from the ordered set:

$$\rho(x) = \frac{1 - \sigma(x)}{x_n - x_1}$$

In the case of applying MTL into the individual predictors contained in the rules, we have adopted the following strategies. In the case of the mean predictor, the solution is simply a vectorized version of such mean values. While in the case of the penalized linear regression, we have adopted a block-sparse regularization ℓ_1/ℓ_2 as in Yuan and Lin (2006):

$$\min_{\beta \in \mathbb{R}^{p \times K}} \frac{1}{2N} \|y - X\beta\|_F^2 + \lambda \sum_{j=1}^{p} \|\beta\|_2$$

4. Experimental Results

For the validation of the proposed experiments, data from the City of Santander (Spain) is used, an urban network with more than 300 individual inductive loop detectors collecting data in real-time on traffic vehicle counts, occupancies and speeds (Figure 1). The network has been chosen as the sensors span the entire City, there is a rich information, and given that it is an urban network it makes a challenging scenario with more than 4000 links in the Aimsun model. Collected data for experiments spans an entire year ranging from January to December 2016 with a 15-min aggregation resolution. This data aggregation is convenient for mitigating the inherent noise in road network measuring devices without compromising the validity of the results.

A 60-min forecasting horizon has been chosen to evaluate the different approaches because it is a challenging and interesting horizon for real world purposes. Lastly, since we explore the applicability of local vehicle traffic prediction in a real-time scenario, a time component must be set to use the new collected data to update system knowledge, statistics, parameters and the rules. Thus, the developed method treats data as continuous incoming data flow in real-time in contrast to more conventional machine learning methods that rely on a data split in training and validation sets. This evaluation procedure is called prequential evaluation, because there is a virtual buffer with a size corresponding to one-week data; such size has been selected as a trade-off decision between computational cost and react-timing to mimic the real-time conditions. In this way traffic observations are collected in a virtual buffer for one week and eventually fed into the learning process of the developed method. For this reason and for the sake of evaluating the proposed experiments in this research work, we have decided to show the prediction performance using solely the output from the last two months of data (November and December) because the system's learning is more steady.

Finally, before presenting the specific experiments for the MTL approach, it is interesting to get a preview of the system's performance for a single specific detector in the Figure 2. For the experimentation, a subset of detection stations (Figure 1) from the road network has been chosen to focus on a small set of locations. We chose eight detection

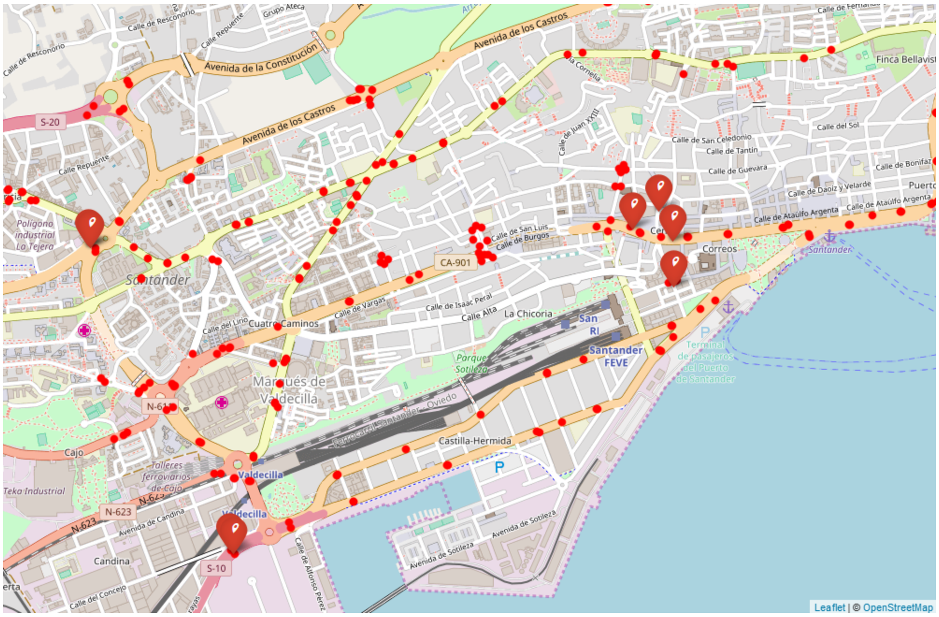

Fig. 1. Partial picture of the sensor network with more than 300 inductive loop detectors (the red points) in the City of Santander. Selected candidates for experimentation appear as red markers, where some of them are located at urban freeways and others in urban streets.

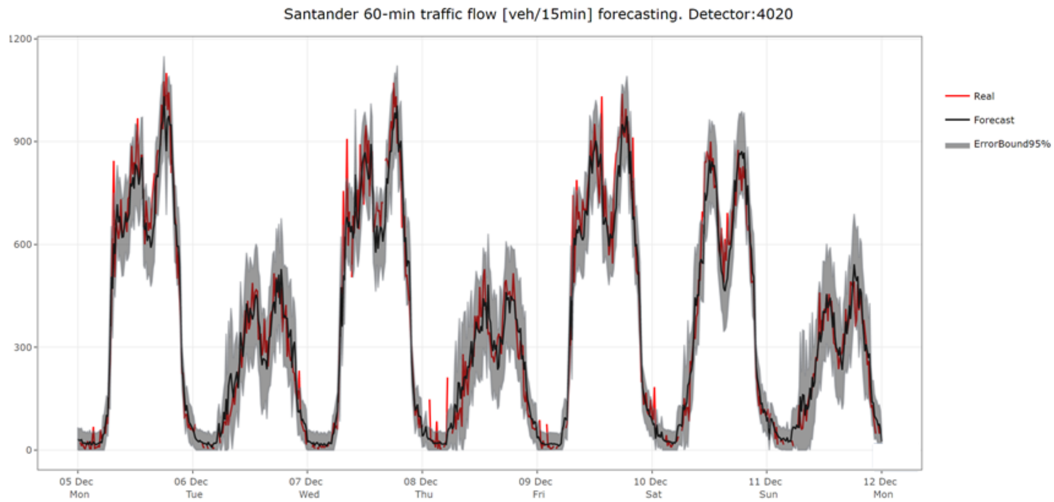

Fig. 2. 60-min traffic flow forecast for a single detector in Santander. The time period has been specifically chosen as it represents a challenging week in Spain with two public holidays in it.

stations which are far apart with the purpose of achieving a representative picture over the entire road network instead of focusing on a small area; and having two of them (1021, 1023) in one of the main entrance/exit to the City, two of them (3078, 3079) in another main entrance/exit, two of them (2016, 2019) in one of the main arterials, and finally another two (2057, 2070) in the City center, with the purpose of capturing sufficiently different dynamics from the network.

The goal of this experiment is to check if jointly learning both traffic variables that have a well understood relation helps to the predictive task by letting the system identifying different situations (free flow, bound flow, congestion) through rules discovery in the data. To this aim, we have learnt different rulesets for each detector. On one hand, we have learnt a ruleset to predict the traffic flows on each location and another ruleset to predict the respective occupancies, which means 16 rulesets overall. On the other hand, we have learnt a ruleset for each detector for both variables at the same time, which means 8 rulesets.

Results for 60-min traffic flow prediction can be seen in Figure 3, where it is evident that the prediction performance is practically the same during the four periods. The Empirical cumulative distribution function (ECDF) plots show the same curve which means there is no significant difference in the traffic flow prediction performance between learning solely traffic flows or jointly learning both variables together. This is confirmed in Table 1, where more disaggregated results for each detector corroborate that almost identical MAPE values are obtained for both approaches. More specifically, the averaged MAPE for both approaches are around the 15% for all the detectors and periods excluding the off-peak period that obviously has a higher averaged MAPE, around 40%, due to the low traffic volume and its impact on this relative metric. However, it is interesting to note that one of the detector with highest error using the single-task approach (2019) obtains a slight reduction of almost a 2% using the MTL approach for the morning, noon and evening periods.

Predicting occupancies is clearly a harder task due to the small scale, the high non-linearity and the presence of noisy data or even binned data. For that reason, higher MAPE values are obtained (Figure 3) compared to predicting traffic flows. In this case, there is a tiny improvement in all the periods, but especially in the morning period with a decrement of around 3% in the averaged MAPE (Table 1).

Besides the comparison of the prediction performance, which is a critical aspect of any predictive system, there is another important factor regarding the interpretability and efficiency of the system. Interpretability was one of the aim this traffic prediction system was built for, thus any factor that can improve this aspect is especially important for traffic engineers and managers that use this tool. As can be observed in Table 1, the number of rules is reduced noticeably in the case of using the MTL paradigm, going from a total of 236 to 163 rules, and going from an average of 32 rules per ruleset to an average of 23. This is done without sacrificing performance, or even improving it at some points.

A final remark is that Santander is not a City with a high presence of congested links so probably this kind of study should deserve another analysis in a kind of network with a higher presence of congestion events.

Table 1. Statistics showing the averaged MAPE predicting traffic flow and occupancy for each time period.

Flow	**Off-peak**		**Morning**		**Noon**		**Evening**		**No. of rules**	
Source	**Single**	**MTL**	**Single**	**MTL**	**Single**	**MTL**	**Single**	**MTL**	**Single**	**MTL**
1021	22.36	22.37	10.73	11.84	6.99	6.94	7.43	7.39	22	23
1023	41.25	41.50	20.81	21.83	11.04	9.97	11.06	12.47	14	23
2016	32.65	32.83	12.62	13.03	10.07	10.72	11.58	12.04	24	23
2019	56.61	57.22	25.20	24.04	23.29	21.72	26.15	24.28	28	27
2057	39.23	39.58	18.60	19.11	14.76	15.68	17.20	17.91	25	23
2070	37.98	36.67	16.42	16.10	11.88	11.83	12.59	13.04	21	23
3078	44.65	45.70	15.75	15.32	12.58	12.61	14.30	14.57	18	21
3079	52.43	52.74	17.75	16.12	13.88	13.66	17.05	16.58	17	19
Mean	**40.67**	**40.79**	**17.21**	**17.13**	**13.08**	**12.91**	**14.67**	**14.79**	—	

Occupancy	**Off-peak**		**Morning**		**Noon**		**Evening**		**No. of rules**	
Source	**Single**	**MTL**	**Single**	**MTL**	**Single**	**MTL**	**Single**	**MTL**	**Single**	**MTL**
1021	24.86	20.44	15.49	15.31	11.93	11.31	13.21	11.47	9	
1023	42.45	38.34	26.04	24.05	15.38	15.87	20.80	19.12	3	
2016	49.53	46.48	41.72	36.44	38.10	35.31	41.18	38.00	9	
2019	56.08	57.04	26.20	23.89	21.43	21.74	24.72	21.94	8	
2057	63.27	64.75	40.59	39.64	32.77	32.80	29.98	30.73	11	—
2070	55.11	55.21	25.73	25.75	21.70	21.94	25.88	27.87	14	
3078	68.37	67.80	19.94	17.67	19.05	18.23	21.27	19.64	6	
3079	70.11	68.94	38.56	27.41	28.77	26.83	24.43	27.89	7	
Mean	**52.83**	**51.43**	**29.35**	**26.33**	**23.65**	**23.01**	**25.18**	**24.58**	—	
Total	—								**236**	**163**
Median									**32**	**23**

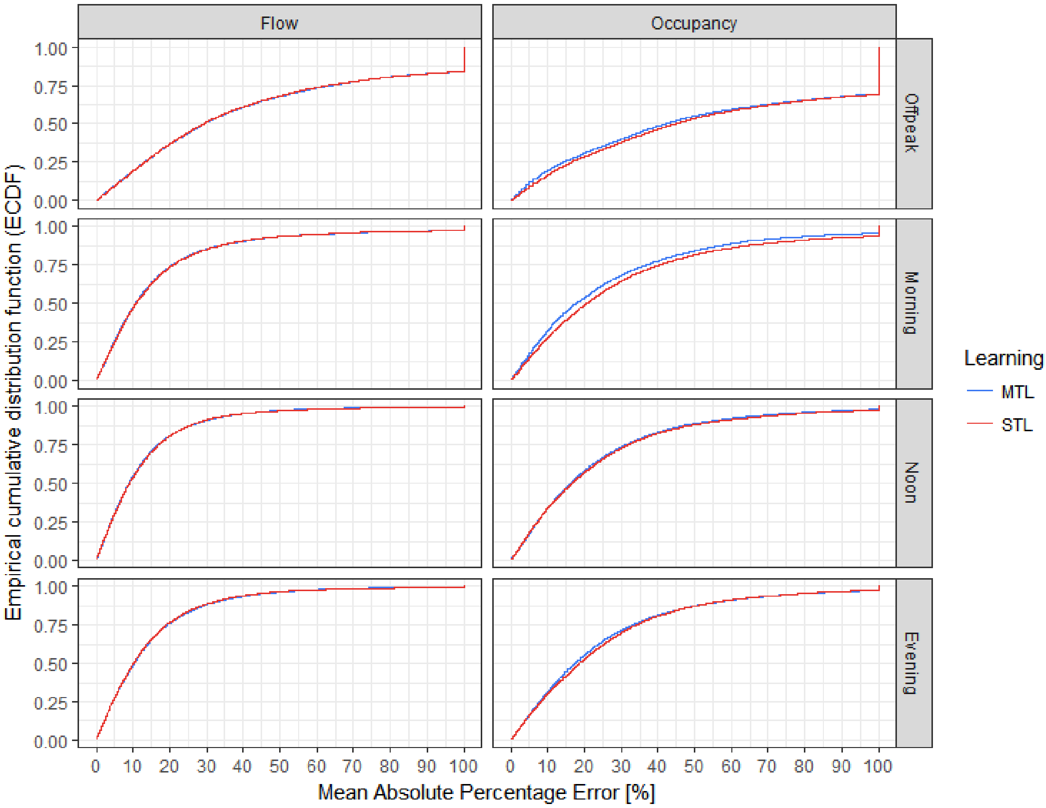

Fig. 3. ECDF showing the cumulative distribution of the MAPE predicting traffic flow and occupancy for each period in experiment 1.

5. Conclusions

In this paper, a MTL paradigm has been tested for traffic prediction. The MTL has been integrated with an existing traffic prediction system, 'Adarules', whose fundamental idea is based on a multi-level decision-tree structure which is built online using streaming data. The main objective of this predictive system, is to find patterns in the data using spatiotemporal correlations for the best purpose of the predictive modeling task. This has motivated the idea of including a MTL paradigm, as there is a lot of shared information and patterns in a transportation network.

The experiments have been carried out using the network of the City of Santander and 60-min short-term traffic predictions have been performed to assess the feasibility of the approach, measuring the impact on the predictive performance along with other criteria such as interpretability and efficiency. More specifically, the experiments tested a classical STL approach for traffic flow and occupancy prediction separately and a MTL approach that jointly learns both. The results showed that there was no significant improvement in the traffic flow prediction performance, and only a slight improvement in the occupancy prediction task. Efficiency and interpretability were improved by reducing 40% of the rules created in the STL approach. However, this requires more testing in other networks with more congestion events.

In general, the MAPE result for 60-min traffic flow prediction was 15% during the noon and evening periods, 18% during the morning period, and +30% during the off-peak period. These results for 60-min prediction are positive and satisfy the needs taking into account Santander is an urban network. Furthermore, the results were consistent among the different tests carried out. However, from the experiments it can be seen the main benefit obtained from integrating a MTL paradigm in the current traffic prediction system has not been an improvement in the prediction quality, but a huge improvement in the efficiency of the method and its interpretability by letting the system choose the spatiotemporal correlations taking into account multiple related tasks which gives a more compacted model. This boost in interpretability is especially interesting and useful for traffic engineers and managers that eventually use this prediction tool. Assessing this interpretability and the relationship between the spatiotemporal correlations found in data and the causality is one of the authors' future research direction. In addition, it would be interesting to assess the proposed MTL paradigm in a kind of network with a higher presence of congestion events, and also to use a more intelligent automated method that associates related detectors for the sake of jointly learning.

6. Acknowledgements

The authors would like to thank EU research project SETA (H2020-ICT-2015) for providing the traffic data from the City of Santander. Work by R. Mena-Yedra is funded by the PhD grant: Pla de Doctorats Industrials (DI-2014) by AGAUR. Work by R. Gavaldà is partially funded by AGAUR project 2014 SGR-890 (MACDA) and by MINECO project TIN2014-57226-P (APCOM). The authors would also like to thank Mark Brackstone for his helpful comments.

References

Almeida, E., Ferreira, C., Gama, J., 2013. Adaptive Model Rules from Data Streams, in: Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases - Volume 8188, Springer-Verlag New York, Inc., New York, NY, USA. pp. 480–492.

Barceló, J., 2010. Fundamentals of traffic simulation. volume 145 of *International series in operations research & management science*. Springer.

Caruana, R., 1997. Multitask Learning. Machine Learning 28, 41–75.

Casas, J., Ferrer, J.L., Garcia, D., Perarnau, J., Torday, A., 2010. Traffic simulation with Aimsun, in: Fundamentals of traffic simulation. Springer, pp. 173–232.

Ermagun, A., Levinson, D.M., 2016. Spatiotemporal Traffic Forecasting: Review and Proposed Directions .

Gama, J., 2010. Knowledge Discovery from Data Streams. Chapman & Hall/CRC.

Huang, W., Song, G., Hong, H., Xie, K., 2014. Deep architecture for traffic flow prediction: deep belief networks with multitask learning. IEEE Transactions on Intelligent Transportation Systems 15, 2191–2201.

Jin, F., Sun, S., 2008. Neural network multitask learning for traffic flow forecasting, in: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), IEEE. pp. 1897–1901.

van Lint, H., van Hinsbergen, C., 2012. Short-Term Traffic and Travel Time Prediction Models. Transportation Research E-Circular .

Mena-Yedra, R., Gavaldà, R., Casas, J., 2017. Adarules: Learning rules for real-time road-traffic prediction. Transportation Research Procedia 27, 11 – 18.

Ruder, S., 2017. An Overview of Multi-Task Learning in Deep Neural Networks. arXiv preprint arXiv:1706.05098 Http://sebastianruder.com/multi-task/.

Sun, S., 2009. Traffic flow forecasting based on multitask ensemble learning, in: Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation, ACM. pp. 961–964.

Treiber, M., Kesting, A., 2013. Traffic flow dynamics. Traffic Flow Dynamics: Data, Models and Simulation, Springer-Verlag Berlin Heidelberg .

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: Emerging Technologies 43, Part 1, 3–19.

Yuan, M., Lin, Y., 2006. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68, 49–67.