Proceedings of 8th Transport Research Arena TRA 2020, April 27-30, 2020, Helsinki, Finland

Context-sensitive modeling of public transport data

Inês Leite, Anna Carolina Finamore, Rui Henriques

[a]*INESC-ID and Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal*

Abstract

Despite the efforts placed by major European cities to optimize public transportation, traffic data analytics often disregard vital situational context. This work proposes a methodology to integrate situational context (including public events, planned interventions and citizen notifications) in the analysis of public transport data. The major contributions are the: online consolidation and labeling of heterogeneous sources of context; calendar-driven statistical modeling of expected traffic behavior; and the integrative display of traffic and its situational context, accompanied by spatiotemporal navigation and zooming facilities. Preliminary results collected from the Lisbon's subway network system shows the relevance of these contributions to support context-sensitive decisions.

Keywords: situational context, sustainable mobility, subway network traffic data, integrative descriptive analytics

1. Introduction

Mobility in most European capital cities is not yet sustainable. In last years, the Lisbon City Council (CML) established acute efforts to collect all available traffic data and their situational context to tackle this issue. The situational context consists of public events (including sport events, large-scale congresses and cultural activities), interdictions (scheduled construction works, accidents and citizen notifications), city urban planning maps, amongst other activities with potential impact on mobility. The integrative analysis of traffic data with these sources of situational context data offers unique opportunities for understanding traffic dynamics and, under the knowledge of upcoming events or scheduled activities, both short- and long-term public transportation planning.

Despite the relevance of context-sensitive analysis of public traffic data, major drawbacks are typically observed. First, situational context is absent from traffic data analysis. Second, sources of situational context data are either private, dispersed or unavailable. Third, traffic data can be analyzed using multiple temporal-spatial-modal resolutions, leading to voluminous and hardly-usable results. Finally, ground-reference traffic behavior (to study meaningful traffic deviations) is not always adequately modeled. These challenges prevent a comprehensive, actionable and real-time understanding of traffic dynamics, leading to inefficiencies in the mobility system.

To address these challenges, this work develops a decision support system laid on five major contributions:

1) consolidation of traffic data from the Lisbon subway operator (METRO) with a comprehensive range of situational context data sources provided at the Open Data portal by the Lisbon City Council;
2) online analytics, grounded on the updatable retrieval and automatic labeling (in accordance with the spatial, temporal and modal footprint) of context data;
3) sound modeling of traffic behavior using calendar-guided statistical models for multivariate time series data;
4) integrative display of (both expected and observed) traffic and its situational context, accompanied by temporal navigation and zooming facilities; and
5) hierarchical views on traffic data analytics (aggregations available at station-route-region levels, for both entry and exit ticket validations) to support both citizen and operational decisions.

Context-sensitive analytics are critical to support public transport reinforcements, waiting time recommendations, sustainable transport planning, real-time messaging alerts, and data-centric coordination between authorities.

The proposed contributions ensure extensibility to other traffic modalities and scalability to other cities in Europe. Once the deployment stage is concluded at Lisbon, these contributions will be made available within an auditable decision support system, attemptively deployed over the Portuguese National Infrastructure for Distributed Computing (INCD), to be used by the municipalities and later available to the Lisbon citizens and visitors.

The manuscript is organized as follows. Section 2 surveys relevant related work. Section 3 describes the proposed methodology. Section 4 gathers preliminary results and implications. Finally, concluding remarks are drawn.

2. Related work

Public traffic data is complex, often being multivariate, spatial, temporal, and sparse. City traffic data – whether road, rail, river or pedestrian – is dependent on the situational context, including sport/cultural events, traffic interdictions, urban planning, weather, among others. To understand urban mobility patterns, such contextualizing factors should be integrated in traffic data analytics. In literature, advances come from both the dedicated analysis of traffic data and the analysis of context-enriched traffic data. In both research streams, contributions can be categorized according to the target task: descriptive, predictive and/or prescriptive. Descriptive analytics aim at summarizing and extracting non-trivial, meaningful and statistically significant relations from urban data, including deviant, periodic and emerging spatiotemporal patterns (Li, 2014); predictive analytics aim at forecasting mobility dynamics; and prescriptive analytics aim at supporting decisions using optimization and control views.

Traffic data analytics. In the area of *descriptive analytics*, contributions on modeling traffic behavior are divided along three major lines: classical statistical modeling, machine learning-based modeling, and hybrid modeling. In the first line, autoregressive, moving average and smoothing models are popular techniques used since 1970s with diverse applications on modeling (and forecasting) the volume and speed traffic data (Ahmed, 1979; Kumar, 2015; Williams, 2003; Wang, 2017; Pan, 2012). Their main drawback is the need of a significant amount of historical data to fit a good model. The second line, machine learning-based modeling, encompasses contributions from temporal pattern mining, multivariate time series analysis and subspace clustering (Higgs, 2014; Park, 2015; Lin, 2019; Cui, 2019; Yan, 2017). Despite their relevance, these models generally provide local views on traffic behavior. Research on the third line, hybrid modeling, explores synergies between approaches on the previous two lines (Wang, 2013; Li, 2018). Complementarily, visual analytic systems have been proposed for different mobility problems such as congestion patterns from mobile data, traffic dynamics from loop counters (Wang, 2014), or even social human activities (Sagl, 2012). We refer to (Zheng, 2016) for an overview of methods and applications of visual analytics on urban data. Beyond descriptive analytics, research on *predictive analytics* has seen recent breakthroughs from deep learning research, with increasing attempts able to explore the inherent spatial and temporal nature of traffic data (Bandara, 2017; Mehta, 2017; Tian, 2015; Lv, 2014; Song, 2016; Wang, 2017, Yu, 2017, Chen, 2018)). For a survey on this area, the reader is invited to consult (Lana, 2018; Suhas, 2017). Research on *prescriptive analytics* has seen recent advances led by the embracing of multi-agent system aspects, such as (deep) reinforcement learning, into the traditional simulation, control and optimization views (Aslani, 2017). Differently from our proposal, to our knowledge, none of the surveyed works consider situational context.

Context-enriched analytics. Different types of situational context have been considered in some previous works, namely: weather data (Soua, 2016); accident and weather context information derived from social media data recurring to natural language processing techniques (Tang, 2019; Wibisono, 2012); planned events such as festivals and sports matches (Rodrigues, 2016; Kwoczek, 2014); combinations of heterogeneous sources of context information such as planned events and weather data (Tempelmeier, 2019). Sagl (2015) discusses the challenge of integrating spatiotemporal contextual information, highlighting that the diverse nature of this data can be understood as a crucial factor in the development of smart cities. The field of context-sensitive traffic analytics has been less explored given the challenges it imposes (identification and sourcing relevant contexts, learning from heterogeneous data, etc.). However, the inclusion of this information is of paramount importance to explain mobility behavior and even reach better descriptive-and-predictive models and optimization routines. Contrasting to our approach, to our knowledge, none of these works support analyzes of deviant behavior along historical data, neither present a context-enriched display of traffic behavior with temporal and spatial navigation facilities.

3. Proposed Methodology

Traffic data. The traffic modality considered to motivate our contributions is Lisbon's subway usage (validation at the entry and exit stations) provided by METRO under an agreement between INESC-ID and the Lisbon City Council (CML) in the context of the ILU project. The subway network has 50 stations distributed in four lines, operating from 6am to 1am. Illustrating, when considering cumulative validations every 15 minutes, subway traffic data (passenger volume data) is seen as multivariate time series data with 75 data points per day (6am–1am) and multivariate order dependent on stations or station-groups of interest. Fig.1 depicts the total amount of entrances at Alameda station in October, 2018. It is noticeable calendar-specific aspects, such as different usage patterns between work days (blue) and weekends (green), and also a holyday (orange) in the Friday of the first week.

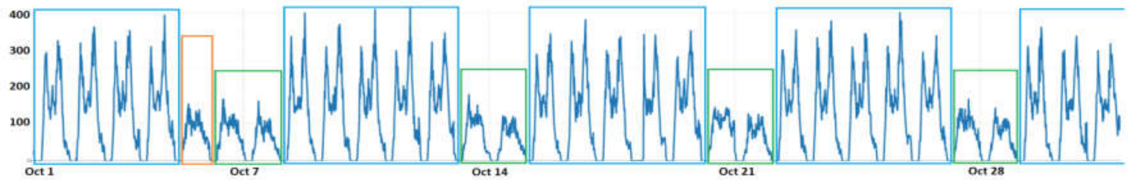

Fig.1 Passenger volume at the Alameda station (cumulative entry validations every 15 minutes) during October 2018.

Missing imputation was applied using interpolation weighted by calendar-driven auto-regression. Missings related with annotated events (e.g. underground strike in October 18th, 6-9h30am) were discarded (no passenger volume).

Online retrieval and labeling of situational context data. The proposed tool builds upon the already established efforts of the Lisbon City Council (CML) to collect all relevant events taking place in the city. These events are stored using semi-structured representations (JSON) at the *Lisboa Aberta* portal[*]. In this work, we consider: congresses*, sport events, Lisbon cultural agenda (concerts, exhibitions and other activities*), accidents, scheduled construction works, and citizen notifications at *Na minha rua* portal.
A periodic routine to read new information from these diverse and heterogeneous sources of situational context data is implemented for an automatic and integrative display of traffic data and their situational context.
In addition, the retrieved events are automatically annotated in accordance with their category and the duration of the event is placed in accordance with the information available or augmented in the presence of user rules. One illustrative rule is the specification that sport events approximately impacts entry validations at public transport 60min-5min before the game and 0-30mins after the game.

***Traffic behavior modeling*.** Observed versus modeled traffic behavior was assessed using: 1) different methods (Holt-Winders, SARIMAX and recurrent neural networks), 2) calendar-free *vs* specific (e.g. workday, day of week, holiday, undifferentiated), 3) small *vs.* lengthy partitions (from 72 to 2232 points), 4) diverse time granularities (windows of minutes to hours) and spatial granularities (station-specific, region-specific, line-specific and all-network). To guarantee proper generalization, Holt-Winters and neural networks were learned using a *k*-fold cross-validation schema on a rolling basis (*n* training + *h* testing subsequent points, shifted *h* time points per fold). Holt-Winters modeling/forecasting technique was observed to be the most competitive technique, inherently able to use triple exponential smoothing to deal with both seasonal and emerging trend components of subway traffic.

Fig.2 Implemented spatiotemporal navigation facilities.

Integrative display with spatiotemporal navigation. A tool able to incorporate previous requirements (traffic modeling, online context gathering and integrative display) was developed in Python ang augment with essential utilities to support spatiotemporal navigation. First, chart zoom and drag facilities are provided. Also, a timeline is available for a usable selection of calendar specific intervals (see Fig.2). Second, the tool allows the individual selection of stations to simultaneously display to establish comparisons. Stations can be alternatively selected using map-based facilities.

Others. The listed functionalities offer this system's stakeholders (whether municipalities or operator's operational and strategic teams) a robust way to detect deviant behavior and study correlations against traffic's situational context. In addition, the tool supports traffic analytics at different granularities, allowing aggregated traffic data analysis for station-sets, subway lines or at the full network level.

4. Results

We undertake a comprehensive analysis to identify the best statistical models of subway usage from multivariate time series data. Mean absolute error (MAE), root mean squared error (RMSE) and mean absolute percentage error (MAPE) were collected. In accordance with our methodology, situational context was dynamically integrated.

Fig.2 provides a graphical view of the observed versus estimated traffic using the best performing statistical model on RMSE (Holt-Winters, 20 calendar-specific days, 15 minutes granularity) for three stations and one line (red line). We select specific days where large sport events took place to understand their impact on subway usage.

[*] http://dados.cm-lisboa.pt/tr/dataset, http://dados.cm-lisboa.pt/tr/dataset?q=espaco+publico, http://dados.cm-lisboa.pt/tr/dataset?q=agenda+cultural

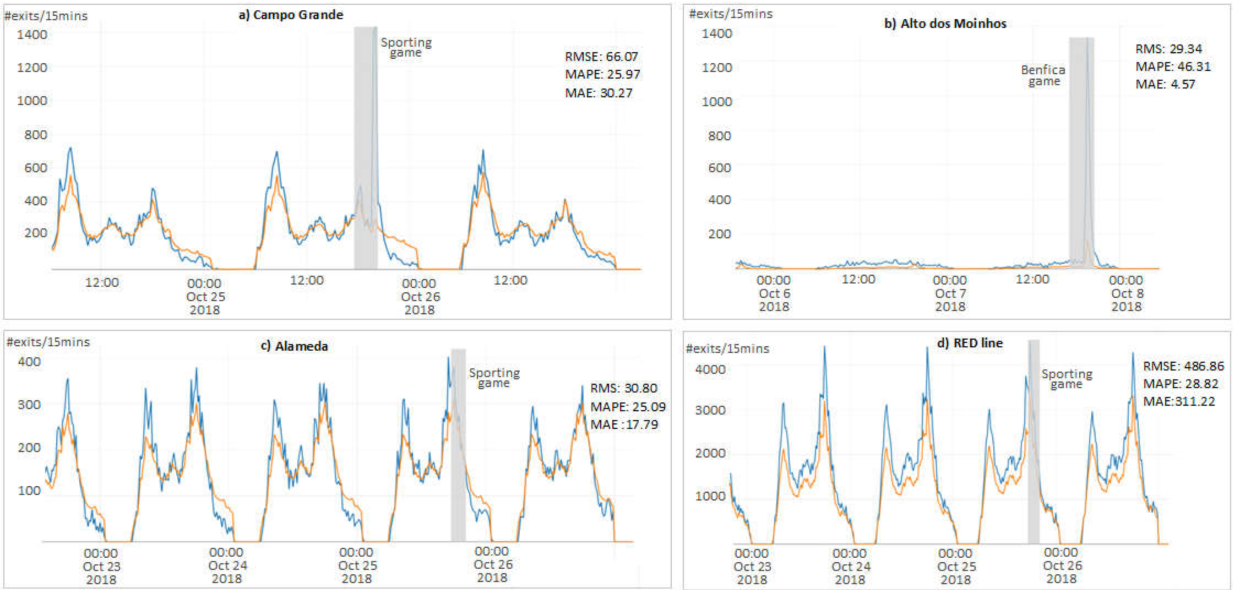

Fig. 3 Observed *versus* expected passenger volume (validations at exit stations every 15 minutes; calendar-driven Holt-Winters) around large sport events at: a) Campo Grande, b) Alto dos Moinhos, c) Alameda, and d) all RED line stations. RMSE, MAPE and MAE displayed.

The expected/reference traffic behavior appears to be robustly estimated given the high variability of the time series and the fact that we did not exclude context-specific variations from the estimations. In periods without situational context, RMSE and MAE generally account for the less than 7% of the real/observed traffic. We hypothesize the error to nearly fully explained by daily/irregular variations and context-specific variations.

Figures 3a and 3b are examples of the impact that sport events (events at the Benfica and Sporting stadiums, respectively) have in the subway traffic of specific station (Campo Grande and Alto dos Moinhos, respectively). On a contrary stance, Figures 3c and 3d show that this traffic disruption due to sport events is not observed at some stations (Alameda in Fig.3d) and metropolitan lines (red line in Fig.3c). These visuals pertain to ticket validations at the exiting station(s).

The gathered results show the relevance of integrative traffic data analytics where situational context is combined. In addition, the online retrieval and labeling of context data should be fully automated once rules pertaining to the timeline impact of specific events is specified. Finally, functional requirements guaranteed the usability of user experience while navigating throughout different types of situational context, timelines and regions (see Fig.2).

The introduced tool for context-sensitive traffic analysis supports the assessment of how deviant traffic behavior is correlated with targeted sources of situational context. Deviant traffic behavior that is not explained by the gathered situational context can be marked and its cause can be traced in order to identify whether any additional source of context is being neglected by our system and whether it can be incorporated. The spatiotemporal navigation utilities were shown to facilitate the retrieval of context-specific traffic behavior and thus support decisions pertaining the strengthening of public transportation during planned/upcoming events of interest. In the context of Lisbon's subway network, non-trivial context-specific patterns of passenger usage were observed.

5. Conclusion

A novel methodology for the online analysis of traffic data against its situational context is proposed. This work explores the unique opportunities arising from current public efforts at Lisbon of gathering historic and prospective city events (including large-scale public events and interdictions) in semi-structured repositories. To this end, we provide a tool able to gather and label these events in real-time and integrate their display with traffic data analytics.

The gathered results and conducted validation near planning experts confirm the criticality of context-sensitive analytics and their augmentation with robustly modeled reference traffic behavior, as well as both geographical and temporal navigation facilities to support both short-term and long-term planning decisions.

Four major future directions are identified: 1) the extension of the proposed methodology for multi-modal traffic data analysis; 2) the context-sensitive modeling of expected behavior based on the selective recollection of historical data with analogous context; 3) the automatic detection of anomalous behavior; and 4) the context-sensitive forecasting of traffic data in the presence of planned events.

Acknowledgments. The authors thank the support of METRO and *Câmara Municipal de Lisboa* given specially by the *Centro de Operações Integrado* and *Gabinete da Mobilidade*. This work is futher supported by national funds through FCT under project ILU DSAIPA/DS/0111/2018 and INESC-ID pluriannual UID/CEC/50021/2019.

References

Chen, M., Yu, X. and Liu, Y., 2018. PCNN: Deep convolutional networks for short-term traffic congestion prediction. IEEE Transactions on Intelligent Transportation Systems, (99), pp.1-10.

Cui, H., Wu, L., He, Z., Hu, S., Ma, K., Yin, L. and Tao, L., 2019. Exploring Multidimensional Spatiotemporal Point Patterns Based on an Improved Affinity Propagation Algorithm. International journal of environmental research and public health, 16(11), p.1988.

Higgs, B. and Abbas, M.M., 2014. Multi-resolution comparison of car-following models using naturalistic data (No. 14-4528).

Kumar, S.V. and Vanajakshi, L., 2015. Short-term traffic flow prediction using seasonal ARIMA model with limited input data. European Transport Research Review, 7(3), p.21.

Kwoczek, S., Di Martino, S. and Nejdl, W., 2014. Predicting and visualizing traffic congestion in the presence of planned special events. Journal of Visual Languages & Computing, 25(6), pp.973-980.

Lana, I., Del Ser, J., Velez, M. and Vlahogianni, E.I., 2018. Road traffic forecasting: recent advances and new challenges. IEEE Intelligent Transportation Systems Magazine, 10(2), pp.93-109.

Li, L., Wang, Y., Zhong, G., Zhang, J. and Ran, B., 2018. Short-to-medium term passenger flow forecasting for metro stations using a hybrid model. KSCE Journal of Civil Engineering, 22(5), pp.1937-1945.

Li, Z., 2014. Spatiotemporal pattern mining: algorithms and applications. In Frequent Pattern Mining (pp. 283-306). Springer, Cham.

Lin, X., 2019. A Road Network Traffic State Identification Method Based on Macroscopic Fundamental Diagram and Spectral Clustering and Support Vector Machine. Mathematical Problems in Engineering, 2019.

Lv, Y., Duan, Y., Kang, W., Li, Z. and Wang, F.Y., 2014. Traffic flow prediction with big data: a deep learning approach. IEEE Transactions on Intelligent Transportation Systems, 16(2), pp.865-873.

Mehta, V. and Chana, I., 2017. Urban traffic state estimation techniques using probe vehicles: A review. In Computing and Network Sustainability (pp. 273-281). Springer, Singapore.

Pan, B., Demiryurek, U. and Shahabi, C., 2012, December. Utilizing real-world transportation data for accurate traffic prediction. In 2012 IEEE 12th International Conference on Data Mining (pp. 595-604). IEEE.

Park, E. and Oh, H., 2015. Automatic identification of spatio-temporal highway congestion patterns using historic database. In Computer Science and its Applications (pp. 491-498). Springer, Berlin, Heidelberg.

Rodrigues, F., Borysov, S.S., Ribeiro, B. and Pereira, F.C., 2016. A bayesian additive model for understanding public transport usage in special events. IEEE transactions on pattern analysis and machine intelligence, 39(11), pp.2113-2126.

Sagl, G., Resch, B. and Blaschke, T., 2015. Contextual sensing: Integrating contextual information with human and technical geo-sensor information for smart cities. Sensors, 15(7), pp.17013-17035.

Sagl, G., Resch, B., Hawelka, B. and Beinat, E., 2012, July. From social sensor data to collective human behaviour patterns: Analysing and visualising spatio-temporal dynamics in urban environments. In Proceedings of the GI-Forum (pp. 54-63). Berlin: Herbert Wichmann Verlag.

Song, X., Kanasugi, H. and Shibasaki, R., 2016, July. DeepTransport: Prediction and Simulation of Human Mobility and Transportation Mode at a Citywide Level. In IJCAI (Vol. 16, pp. 2618-2624).

Soua, R., Koesdwiady, A. and Karray, F., 2016, July. Big-data-generated traffic flow prediction using deep learning and dempster-shafer theory. In 2016 International Joint Conference on Neural Networks (IJCNN) (pp. 3195-3202). IEEE.

Suhas, S., Kalyan, V.V., Katti, M., Prakash, B.A. and Naveena, C., 2017, March. A comprehensive review on traffic prediction for intelligent transport system. In 2017 International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT) (pp. 138-143). IEEE.

Tang, L., Duan, Z. and Zhao, Y., 2019. Toward using social media to support ridesharing services: challenges and opportunities. Transportation Planning and Technology, 42(4), pp.355-379.

Tempelmeier, N., Rietz, Y., Lishchuk, I., Kruegel, T., Mumm, O., Carlow, V.M., Dietze, S. and Demidova, E., 2019. Data4UrbanMobility: Towards Holistic Data Analytics for Mobility Applications in Urban Regions. arXiv preprint arXiv:1903.12064.

Tian, Y. and Pan, L., 2015, December. Predicting short-term traffic flow by long short-term memory recurrent neural network. In 2015 IEEE international conference on smart city/SocialCom/SustainCom (SmartCity) (pp. 153-158). IEEE.

Wang, J. and Shi, Q., 2013. Short-term traffic speed forecasting hybrid model based on chaos-wavelet analysis-support vector machine theory. Transportation Research Part C: Emerging Technologies, 27, pp.219-232.

Wang, Z., Ye, T., Lu, M., Yuan, X., Qu, H., Yuan, J. and Wu, Q., 2014. Visual exploration of sparse traffic trajectory data. IEEE transactions on visualization and computer graphics, 20(12), pp.1813-1822.

Wang, J., Hu, F. and Li, L., 2017, November. Deep Bi-directional Long Short-Term Memory Model for Short-Term Traffic Flow Prediction. In International Conference on Neural Information Processing (pp. 306-316). Springer, Cham.

Wang, Z.H., Lu, C.Y., Pu, B., Li, G.W. and Guo, Z.J., 2017. Short-term Forecast Model of Vehicles Volume Based on ARIMA Seasonal Model and Holt-Winters. In ITM Web of Conferences (Vol. 12, p. 04028). EDP Sciences.

Wibisono, A., Sina, I., Ihsannuddin, M.A., Hafizh, A., Hardjono, B., Nurhadiyatna, A. and Jatmiko, W., 2012, December. Traffic intelligent system architecture based on social media information. In 2012 International Conference on Advanced Computer Science and Information Systems (ICACSIS) (pp. 25-30). IEEE.

Williams, B.M. and Hoel, L.A., 2003. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. Journal of transportation engineering, 129(6), pp.664-672.

Yan, Y., Zhang, S., Tang, J. and Wang, X., 2017. Understanding characteristics in multivariate traffic flow time series from complex network structure. Physica A: Statistical Mechanics and its Applications, 477, pp.149-160.

Yu, H., Wu, Z., Wang, S., Wang, Y. and Ma, X., 2017. Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. Sensors, 17(7), p.1501.

Zheng, Y., Wu, W., Chen, Y., Qu, H. and Ni, L.M., 2016. Visual analytics in urban computing: An overview. IEEE Transactions on Big Data, 2(3), pp.276-296.