

“© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

Data Fusion for MaaS: Opportunities and Challenges

Jianqing Wu¹, Luping Zhou², Chen Cai³, Jun Shen¹, Sim Kim Lau¹

¹ School of Computing and Information Technology, University of Wollongong, Australia

² School of Electrical and Information Engineering, University of Sydney, Australia

³ Advanced Data Analytics in Transport, Data 61, CSIRO, Australia

Email: jw937@uowmail.edu.au, luping.zhou@sydney.edu.au, chen.cai@data61.csiro.au, {jshen, simlau}@uow.edu.au

Abstract—Computer Supported Cooperative Work (CSCW) in design is an essential facilitator for the development and implementation of smart cities, where modern cooperative transportation and integrated mobility are highly demanded. Owing to greater availability of different data sources, data fusion problem in intelligent transportation systems (ITS) has been very challenging, where machine learning modelling and approaches are promising to offer an important yet comprehensive solution. In this paper, we provide an overview of the recent advances in data fusion for Mobility as a Service (MaaS), including the basics of data fusion theory and the related machine learning methods. We also highlight the opportunities and challenges on MaaS, and discuss potential future directions of research on the integrated mobility modelling.

Keywords—data fusion; machine learning; mobility as a service.

I. INTRODUCTION

Intelligent transportation is one major focus for smart cities, where coordination and cooperation between different stakeholders in the value chain of transportation services provision are very important. In the era of big data, data fusion has attracted tremendous attention from both academia and industry, with many successful applications also found in the transportation field. In particular, the fast development of machine learning approaches is playing an increasingly crucial role in the development of data fusion systems. Specifically, the implementation of stage-based, feature-level based, and semantic meaning based data fusion reveals a strong dependence on the use of machine learning methods [1]. As inferences and predictions on future data are uncertain, probability theory can provide a framework to model uncertainty for supporting decision making in complex situations [2] [3].

With the assistance of accurate and timely traffic information, high quality prediction of transportation outcomes would help intelligent transportation systems to provide more reliable and stable management and operations. Accordingly, how to get more accurate and timely forecasts is raising unique opportunities and challenges for public organisations such as government agencies, and private companies such as toll road operators. Meanwhile, multiple disparate data sources can be considered as complementary to present a better interpretation of the observed transportation related behaviours by reducing the uncertainty [4].

In recent years, new transportation modes are transforming urban mobility, such as vehicle sharing, e-hailing, electric powertrains, and promisingly in near future, the autonomous vehicles [5]. The emergence of these transportation modes is changing people's travel behaviours, which results in generating massive data volumes by employing sensing technologies and large-scale computing infrastructures.

Thus, integrated mobility or mobility as a service (MaaS) systems of the next generation transportation will be more complex and diverse [6]. It is vital to unlock the power of data from integrated mobility to guide the strategy design and the construction of traffic infrastructure in order to encourage the use of mass transit and non-motorized transportation (e.g. bicycle riding and walking) while reducing the usage of private cars. For instance, to study the impact of weather factors on traffic road networks throughout a city, traffic data and weather data are combined into a weather-traffic index, which is used to conduct a key factor analysis with regional features. Specifically, such study verified the effectiveness of weather-traffic index against real-world observations, and also identified the correlation between the house age information and the weather-traffic index [7].

Hence, data fusion studies with machine learning approaches have been increasingly used to solve MaaS problems. Such approaches can leverage the characteristics of data heterogeneity to improve transport predictive modelling and mobility forecasting. In this paper, we will review the related literature in heterogeneous data fusion first. Then we point out the suggestions for future research directions.

II. MOBILITY AS A SERVICE AND DATA FUSION

MaaS is concerned with that people will use or be encouraged to use more than one mode of transport when traveling around in urban cities. Consequently, MaaS can be defined as: combining transport services from public and private transport providers through a unified gateway that creates and manages the trip, for which a user can pay with a single account. Users can pay per trip or a monthly fee for a limited distance [8]. The key concept behind MaaS is to provide users with mobility solutions in order to meet their diverse travel needs [6]. Accordingly, public and private sectors should have an integrated view on the future of mobility to understand the impacts of transportation mode changes [5]. From a data-driven perspective, MaaS data has evolved from small-scale survey datasets to large terabyte-scale datasets from diverse sources, such as smart card data, GPS

trajectory data, and mobile phone data. In contrast, in traditional approaches, a single data source or single domain dataset is usually used for the mobility research and there are very few studies considering the effect of heterogeneous factors (e.g. environmental pollution, housing prices, and individual physical and mental health) among datasets to optimize the integrated mobility and vice versa.

Data fusion, information fusion and knowledge fusion can be considered as three different levels of abstraction [9]. Nonetheless, they are tightly related. In data fusion, several sources of raw data are extracted to generate more useful information in order to remove noisy and redundant data; with information fusion, several sources of information are combined to create knowledge; for knowledge fusion, several heterogeneous sources of information and/or knowledge are merged to create a complete knowledge [10]. Both information and knowledge can be employed to support decision making [11]. Additionally, the knowledge fusion problem is regarded as to use multiple knowledge extractor to extract values from each datasets, and then to decide the degree of correctness of the extracted knowledge so that a knowledge base is generated [12].

III. MACHINE LEARNING METHODOLOGIES

The goal of integrated urban mobility research is to model large-scale heterogeneous data in order to find the underlying patterns of observed data or to predict future data, which leads to the performance enhancement of MaaS systems. Thus, the availability of MaaS data provides more useful information for understanding the mobility changes that have occurred in the urban areas. To understand integrated urban mobility, cluster analysis [13] [14] [15] and classification [16] [17] can be performed to identify the latent spatial-temporal patterns. With multi-source datasets, computational models are constructed for uncovering and optimizing urban mobility patterns. A few of them are discussed below.

A. Matrix Factorization

Matrix factorization decomposes an observed matrix M into a product of two matrices, which present the latent factors of user-feature and item-feature respectively [1]. Since the observed matrix M can be approximated by the production of

the two matrices, thus matrix factorization has potential to predict the missing or incomplete data for effectively fusing the information or knowledge from multiple heterogeneous data sources.

On the other hand, Singular Value Decomposition (SVD) and non-negative matrix factorization (NMF) are two main methods to deal with missing values and dimensionality reduction [1] [18]. Moreover, a data fusion approach with penalized matrix tri-factorization (DFMF) is proposed to simultaneously decompose data matrices reveal hidden associations. This approach identifies that matrix factorization based data fusion achieves the high accuracy result and time response in a particular scenario [19]. To tackle the issues with sparse data, a context-aware tensor factorization (CATF) model has employed high-order singular value decomposition (HOSVD) to integrate with contextual features (e.g. features of gas stations and weather conditions) [20].

B. Multi-View Learning

Multi-view learning is a learning paradigm, which uses one function to model each view and jointly optimizes all the functions to exploit redundant views of the same input data [21] [22]. For example, for predicting weekday ridership of a rail station, there are often two views representing the given rail station: its local temporal information (e.g. weather conditions) and spatial information (e.g. distribution of Points of Interests). According to [22], the existing multi-view learning algorithms can be divided into three groups: co-training style algorithms, co-regularization style algorithms and margin consistency style algorithms. Meanwhile, multi-view learning algorithms can also be grouped into three categories: co-training, multiple kernel learning and subspace learning [1][21].

The above algorithms can be mutually integrated; for the MaaS scenario, we herein only review subspace methods as main solutions. The main idea of subspace methods is to exploit the latent subspace for multi-view data [1]. Selected subspace methods from the literature are presented in Table 1. Their descriptions, properties and related representatives are also summarised in the table.

TABLE 1
The outline of selected subspace methods

| Methods | Descriptions | Properties | Representatives |
|---------|---|---|--|
| PCA | Linear transformations of correlated variables into uncorrelated variables so that underlying factors or components have maximized variance [23]. | mutually orthogonal, Gaussian distributions | Probabilistic PCA [24] Sparse PCA [25] Kernel PCA [26] |
| ICA | Linear transformation of an m-dimensional feature space into an n-feature space so that underlying factors or components have maximized independence [27] [28]. | maximal mutual information, statistically independent, non-Gaussian distributions | Fast ICA [29] NMF [30] Kernel ICA [31] |
| CCA | Exploring two sets of multidimensional variables to maximise the correlation of two views in the subspace [1] [33]. | maximal mutual information, multivariate normal distributions | Sparse CCA[32] Kernel CCA[33] Tensor CCA[34] |

Particularly, Principal Component Analysis (PCA) is a powerful method to exploit the subspace for single-view data

[1]. For example, PCA can be utilized to explore the spatial and temporal structure of aggregated human mobility; the

predictions of most pixel population variations (PPVs) of PCA model can achieve superior performance than auto-regression moving average (ARMA) model [35].

Earlier in this section, we discussed different types of the matrix factorization model. Before discussing independent component analysis (ICA), herein we first explain a concrete mathematical meaning of the model. Specifically, a Matrix $M \in R^{m \times n}$ is factorized into the production of two matrices UV^T , where $U \in R^{m \times k}$ and $V \in R^{n \times k}$. According to [35], The equation can be written as

$$M = UV^T = (UX^{-1})(XV^T) \quad (1)$$

Moreover, $X = PD$ where P is a permutation matrix and D is a diagonal matrix. As the index of P is arbitrary ($1 \dots R$), the inherent indeterminacy of P in Formula (1) is inevitable. As one of Blind Source Separation (BSS) methods, ICA is widely used to solve indeterminacy in (1). As ICA can be commonly formulated as

$$x(t) = As(t), \quad t = 1, \dots, Z \quad (2)$$

where $s(t) = [s_1(t), \dots, s_k(t)]^T \in R^{k \times 1}$ is a vector of k statistically independent random processes (sources) and $x(t) \in R^{m \times 1}$ are observations. A is a full column rank. The link with Formula (1) is established via $M = [x(1), \dots, x(Z)]$, $n = Z$, and $V^T = [s(1), \dots, s(Z)]$. Therefore, the constraint or assumption of statistical independence on the sources helps to achieve essential uniqueness or diversity in data-driven models [36]. The use of ICA method to data fusion on MaaS, that has the potential to fix the indeterminacy of factor analysis (FA).

Canonical correlation analysis (CCA) is an efficient and powerful approach to find the correlation between two sets of variables. As CCA is concerned with seeking a pair of linear transformations associated to the two sets of variables, the projected variables on each view are maximally correlated [32]. Nevertheless, for capturing the nonlinear correlation among data, Kernel CCA is proposed to map each data point to a higher space [1][33]. Moreover, Tensor CCA is developed to generalize CCA to handle any number of views in a straightforward and yet natural way [34]. Particularly, CCA, Kernel CCA and Tensor CCA exploit the subspace in an unsupervised way. Since the generalization performances of unsupervised learning approaches may not be good enough for prediction tasks, Bayesian multi-view dimensionality reduction (BMDR) method is proposed to project data points into a unified subspace [37].

C. Probabilistic Graphical Model

A probabilistic graphical model (PGM) is to encode probability distributions or to express conditional dependencies among large numbers of random variables. Generally, a PGM is a declarative representation (or a graphical representation) that consists of nodes and edges, where nodes correspond to a group of random variables and edges express interactions among variables [38]. Additionally, there are two typical models of PGMs: directed graphical models (Bayesian Networks) and undirected graphical models (Markov Networks). Bayesian Networks (BNs) represent causality between variables; however, Markov Networks represent mutual relationships between variables [39].

Inference using PGMs is the process of predicting the status of latent variables from the probabilistic model. It can be regarded as an optimization problem. Approximate inference algorithms are derived to implement the optimization. There are two major methods in approximate inference algorithms: 1) variational methods are deterministic and 2) particle-based inference methods use stochastic numerical sampling from distributions such as forward sampling, importance sampling, and Gibbs sampling.

Particularly, Gibbs sampling can be employed equally well to both BNs and MNs [38]. Moreover, a topic model based on Latent Dirichlet Allocation (LDA) and Dirichlet Multinomial Regression (DMR) can infer the functional regions in a city by utilizing Gibbs sampling [40]. Inspired by PGMs in text mining, a Human Mobility Representation model (HuMoR) is proposed to infer latent patterns from anonymized sequences of user locations by using Collapsed Gibbs sampling; nevertheless, the Gibbs sampling is not an efficient parameter inference method [41].

Furthermore, Gaussian Bayesian Network (GBN) based graphical model was applied to generate the spatiotemporal casual pathways for air pollutants by combing pattern mining and Bayesian learning. The experiments identified that the model outperforms ARMA, linear regression model, and support vector machine for regression with a Gaussian radial basis function kernel (SVM-R) in both efficiency and inference accuracy [42].

According to [39], both BN and MN are used to develop a hybrid network for estimating multivariate Gaussian distribution so that computational complexity can be reduced and the probability of finding the best solution can be increased. As PGMs provide an approximation of exact distributions, large-scale heterogeneous data sets can describe complex relations by using statistical inference [38]. Additionally, the information in multiple heterogeneous information networks (HINs) can be fused to obtain a more comprehensive and consistent knowledge [43].

IV. EVALUATION MEASUREMENTS

For general estimation and prediction, conventional error metrics measures can be computed to evaluate the performance of different models in ITS, such as Mean Error (ME), Mean Relative Error (MRE), Mean Absolute Error (MAE), Mean Absolute Relative Error (MARE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE) and Root Mean Square Error (RMSE) [44][45][46].

However, for exploring the degree of predictability in individual mobility, the entropy measure is one of the most promising methods to characterize the limits of predictability \prod^{max} in one user mobility. To be specific, if a user with $\prod^{max} = 0.25$, it means that at least 75% of the time the user's location appears to be random, only in the remaining 25% of the time we can really predict the location that the user appears at. In other words, no matter how good the predictive algorithm is, the study cannot predict with better than 25% accuracy. Thus, \prod^{max} presents the fundamental theoretical limit for the potential predictability in user mobility [47].

Correspondingly, in [47], the study of actual prediction algorithms on 500,000 users was conducted to show how close they were to the maximum potential predictability. Markov chain (MC) based models were implemented to forecast the actual location visited by each user. A comparison of the results reveals that a higher order MC-based model does not bring significant improvement in the prediction accuracy when approaching the maximum predictability [48].

Moreover, predictive accuracy and explanation should be complementary to optimize the generalization error and to search for simpler and more interpretable model such as graph link prediction, from prediction results, so that trade-off between them can be achieved. After all, it is more important to

get close to optimal prediction and to gain insight into the interpretable causal mechanisms [49]. Apart from the prediction errors, the computation performance and domain knowledge from experts should also be considered to validate and evaluate the models [50].

Based on the above findings, we would propose a framework of data fusion for MaaS, using multiple data sources to validate and optimise the performance of the models, as shown in Figure 1. We believe that generating such a model will offer us more insights in understanding the MaaS-based problem and also provide decision support for the design of MaaS platform.

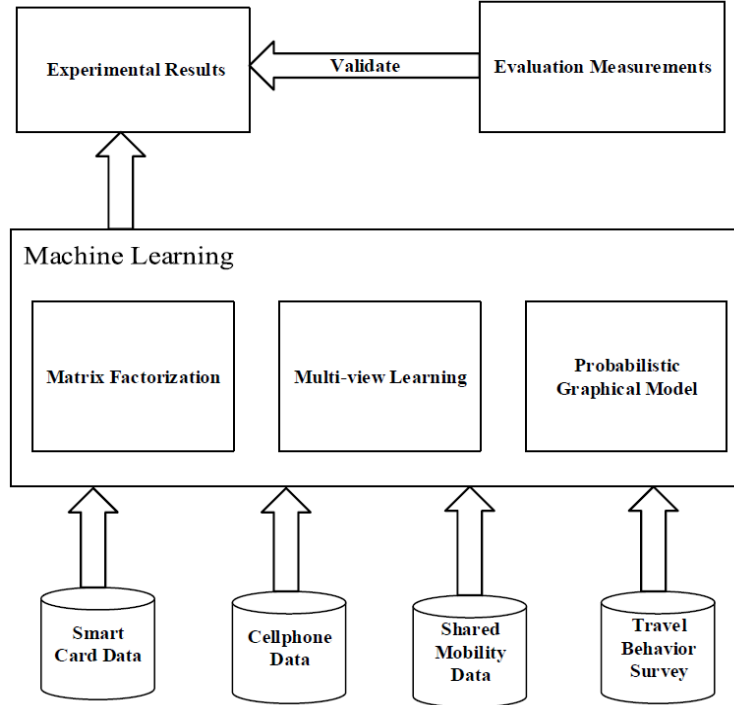


Figure 1: A framework of data fusion on MaaS

V. CHALLENGES AHEAD

Machine learning offers a variety of metrics for performance evaluation, such as approximation, accuracy and prediction error [51]. Obviously, the combination of multi-source datasets is able to create more specific, comprehensive, unified information about urban mobility. However, fusing heterogeneous data to understand MaaS is not a simple combination of multiple types of metrics. To achieve a better and more explainable solution, bridging the trade-offs between these definitions and metrics could be an option [51]. Therefore, there are ongoing challenges for MaaS data fusion research in many areas, including:

a) *Demand Estimation for MaaS*: Shared mobility provides environmentally-friendly travel modes to fill the gap in existing urban transportation networks. It includes carsharing, scooter sharing, bike sharing, ridesplitting, e-hail

services, on-demand micro transit, and other modes[52]. The motivation for facilitating first mile and last mile public transit connections is to increase ridership. Furthermore, it also brings opportunities to estimate the demands on first and last-mile trips, which can be used to better guide infrastructure investments in proximity areas to public transportation stations. Concretely, smart card data can be collected by the Automated Fare Collection (AFC) systems that imply valuable knowledge on the travel patterns [14][53] while travel surveys can provide socio-economic information [54]. Extracting useful information from multi-source data is vital to infer unmet mobility demand patterns and enhance decision-making processes in ITS.

b) *Modeling Travel Behaviour Using MaaS*: Motivated by the need to capture users' historical travel behaviour to better predict future travel patterns, we can develop models to forecast how user travel in time and space, and to understand

the factors that influence on travel-related choices. Specifically, the emerging trend is to incorporate more data sources to identify behavioural factors that contain the underlying causal mechanisms, to predict user movement behaviours, and to mine and uncover user mobility patterns [55].

c) *Optimization Algorithms for Organizing MaaS*: Recent researches on MaaS tends to become increasingly attractive for both research and system development communities, as heavily supported by governments, NGOs and business. Here is an example but real scenario explaining why it is essential to optimize algorithms for MaaS. For instance, on 13 July 2017, Melbourne Metro trains were shut down after computer failure during peak-hour; massive crowds of passengers were stuck on crowded trains, experiencing great frustration getting home. To tackle the problem, one of the emergency plans could be scheduling alternative transport modes to reduce traffic stress and congestion. More specifically, the main solution was to optimize multi-site scheduling algorithms for organizing

MaaS, which should be based on the ridership prediction of related stations and the estimation of origin-destination (OD) matrices from the passengers' historical journeys.

VI. CONCLUSION

In this paper, we presented an overview of opportunities and challenges of data fusion for MaaS. There is great potential for discovering and exploiting valuable knowledge for improving MaaS. Also, it is recommended that further research is conducted to investigate the related open issues through using causal inference in machine learning.

It is worth to point out that MaaS may not reduce congestion or cause congestion getting worse if it is car centric without protecting the future of public transport [56] [57]. Nevertheless, MaaS has not a commonly agreed definition yet, is now becoming a mainstream transport paradigm with the initial developments of theory and practice around the world [58][59]. It has great potential to enhance the quality and efficiency of the existing transport services.

REFERENCES

- [1] Y. Zheng, "Methodologies for Cross-Domain Data Fusion: An Overview," *IEEE Transactions on Big Data*, vol. 1, pp. 16-34, 2015.
- [2] Z. Ghahramani, "Probabilistic Machine Learning and Artificial Intelligence," *Nature*, vol. 521, pp. 452-459, 2015.
- [3] J. Zhu, J. Chen, W. Hu, and B. Zhang, "Big Learning with Bayesian Methods," *National Science Review*, vol. 4, pp. 627-651, 2017.
- [4] N.-E. E. Faouzi and L. A. Klein, "Data Fusion for ITS: Techniques and Research Needs," *Transportation Research Procedia*, vol. 15, pp. 495-512, 2016.
- [5] E. Hannon et al. "An Integrated Perspective on the Future of Mobility," 2016. [Online]. Available: <http://www.mckinsey.com/business-functions/sustainability-and-resource-productivity/our-insights/an-integrated-perspective-on-the-future-of-mobility?cid=other-eml-alt-mip-mck-oth-1610>
- [6] D. A. Hensher, "Future Bus Transport Contracts under a Mobility as a Service (MaaS) Regime in the Digital Age: Are They Likely to Change?," *Transportation Research Part A*, vol. 98, pp. 86-96, 2017.
- [7] Y. Ding, Y. Li, K. Deng, H. Tan, M. Yuan, and L. M. Ni, "Detecting and Analyzing Urban Regions with High Impact of Weather Change on Transport," *IEEE Transactions on Big Data*, vol. 3, pp. 126-139, 2017.
- [8] Wikipedia. "Transportation as a Service," 2017. [Online]. Available: https://en.wikipedia.org/wiki/Transportation_as_a_Service
- [9] H. Fan, F. Wang, and M. Zheng, Research on Knowledge Fusion Connotation and Process Model, in *Knowledge Graph and Semantic Computing: Semantic, Knowledge, and Linked Big Data: First China Conference, CCKS 2016, Beijing, China, September 19-22, 2016, Revised Selected Papers*, H. Chen, H. Ji, L. Sun, H. Wang, T. Qian, and T. Ruan, Eds., ed Singapore: Springer Singapore, 2016, pp. 184-195.
- [10] L. Snidaro, J. Garcia, J. Llinas, E. Blasch, "Context-Enhanced Information Fusion: Boosting Real-World Performance with Domain Knowledge," *Advances in Computer Vision and Pattern Recognition*, Springer International Publishing, 2016.
- [11] C. Zins, "Conceptual Approaches for Defining Data, Information, and Knowledge," *Journal of the American Society for Information Science and Technology*, vol. 58, pp. 479-493, 2007.
- [12] X. L. Dong, E. Gabrilovich, G. Heitz, W. Horn, K. Murphy, S. Sun, and W. Zhang, "From Data Fusion to Knowledge Fusion," *Proceedings of the VLDB Endowment*, vol. 7, no. 10, pp. 881-892, 2014.
- [13] M. K. El Mahrsi, E. Come, L. Oukhellou, and M. Verleysen, "Clustering Smart Card Data for Urban Mobility Analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, 2016.
- [14] Y. Yuan and M. Raubal, "Extracting Clustered Urban Mobility and Activities from Georeferenced Mobile Phone Datasets," in *ISPRS Workshop on Spatio-Temporal Data Mining and Analysis (STDM'11)*, pp. 21-23, 2011.
- [15] X. Yang, Z. Zhao, and S. Lu, "Exploring Spatial-Temporal Patterns of Urban Human Mobility Hotspots," *Sustainability*, vol. 8, no. 7, p. 674, 2016.
- [16] L. Cagliero, T. Cerquitelli, S. Chiusano, P. Garza, and X. Xin, "Predicting Critical Conditions in Bicycle Sharing Systems," *Computing*, vol. 99, pp. 39-57, 2017.
- [17] X. Guo, R. Zhang, X. Liu, and J. Huai, "Human Mobility Semantics Analysis: a Probabilistic and Scalable Approach," *Geoinformatica*, April 2017.
- [18] C. Boutsidis and E. Gallopoulos, "SVD Based Initialization: a Head Start for Nonnegative Matrix Factorization," *Pattern Recognition*, vol. 41, no. 4, pp. 1350-1362, 2008.
- [19] M. Zitnik and B. Zupan, "Data Fusion by Matrix Factorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no.1, pp. 41-53, January 2015.
- [20] F. Zhang, N. J. Yuan, Y. Zheng, X. Xie, and D. Wilkie, "Sensing the Pulse of Urban Refueling Behavior: a Perspective from Taxi Mobility," *ACM Transactions on Intelligent Systems and Technology*, vol. 6, 2015.
- [21] C. Xu, D. Tao, and C. Xu. "A Survey on Multi-view Learning," 2013. [Online]. Available: <http://arxiv.org/abs/1304.5634>
- [22] J. Zhao, X. Xie, X. Xu, and S. Sun, "Multi-view Learning Overview: Recent Progress and New Challenges," *Information Fusion*, vol. 38, pp. 43-54, 2017.
- [23] J. Shlens. "A Tutorial on Principal Component Analysis," 2014. [Online]. Available: <http://arxiv.org/abs/1404.1100>
- [24] M.E. Tipping and C.M. Bishop, "Probabilistic Principal Components Analysis," *J. Royal Statistical Soc. B*, vol. 61, no. 3, pp. 611- 622, 1999.

- [25] H. Zou, T. Hastie, and R. Tibshirani, "Sparse Principal Component Analysis," *J. Comput. Graphical Statist.*, vol. 15, no. 2, pp. 265–286, 2006.
- [26] B. Schölkopf, A. Smola, and K. Müller, "Kernel Principal Component Analysis," in *Advances in Kernel Methods-Support Vector Learning*, B. Schölkopf, C. Burges, and A. Smola, Eds. Cambridge, MA: MIT Press, 1999, pp. 327–352.
- [27] H. B. Mitchell, *Data Fusion: Concepts and Ideas*. Springer, 2nd edition, 2012.
- [28] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Wiley, 2001.
- [29] A. Hyvärinen, "Fast and Robust Fixed-Point Algorithms For Independent Component Analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.
- [30] D. Lee and S. Seung, "Learning the Parts of Objects by Nonnegative Matrix Factorization," *Nature*, 401, 1999.
- [31] F. R. Bach and M. I. Jordan, "Kernel Independent Component Analysis," *Journal of Machine Learning Research*, vol. 3, pp. 1–48, 2002.
- [32] D. Chu, L. Z. Liao, M. K. Ng, and X. Zhang, "Sparse Canonical Correlation Analysis: New Formulation and Algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 3050–3065, 2013.
- [33] D. Hardoon, S. Szedmak, and J.S. Taylor, "Canonical Correlation Analysis; An Overview with Application to Learning Methods," *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, 2004.
- [34] Y. Luo, D. Tao, K. Ramamohanarao, C. Xu, and Y. Wen, "Tensor Canonical Correlation Analysis for Multi-view Dimension Reduction," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 11, pp. 3111–3124, Nov 2015.
- [35] J. Sun, Y. Wang, H. Si, J. Yuan, and X. Shan, "Aggregate Human Mobility Modeling Using Principal Component Analysis," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 1, no. 2-3, p. 83, 2010.
- [36] D. Lahat, T. Adali, and C. Jutten, "Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–77, 2015.
- [37] M. Gönen et al. "Bayesian Multiview Dimensionality Reduction for Learning Predictive Subspaces," *presented at the Proceedings of the Twenty-first European Conference on Artificial Intelligence*, Prague, Czech Republic, 2014.
- [38] S. N. Srihari, "Probabilistic Graphical Models," 2017. [Online]. Available: <http://www.cedar.buffalo.edu/~srihari/papers/PGM-ESNAM.pdf>
- [39] L. PourMohammadBagher, M. M. Ebadzadeh, and R. Safabakhsh, "Graphical Model Based Continuous Estimation of Distribution Algorithm," *Applied Soft Computing*, Article vol. 58, pp. 388–400, September 2017.
- [40] N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong, "Discovering Urban Functional Zones Using Latent Activity Trajectories," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 3, pp. 712–725, Mar. 2015.
- [41] B. Alharbi, A. Qahtan, and X. Zhang, "Minimizing User Involvement for Learning Human Mobility Patterns from Location Traces," in *AAAI*, 2016.
- [42] J. Y. Zhu et al., "pg-Causality: Identifying Spatiotemporal Causal Pathways for Air Pollutants with Urban Big Data," 2016. [Online]. Available: <https://www.microsoft.com/en-us/research/wp-content/uploads/2017/07/pg-Causality-yuzheng.pdf>
- [43] C. Shi, Y. Li, J. Zhang, Y. Sun, and S. Y. Philip, "A Survey of Heterogeneous Information Network Analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 1, pp. 17–37, Jan. 2017.
- [44] R. Li, G. Rose, H. Chen, and J. Shen, "Effective Long-term Travel Time Prediction with Fuzzy Rules for Tollway," *Neural Computing and Applications*, 2017.
- [45] X. Song, H. Kanasugi, and R. Shibasaki, "DeepTransport: Prediction and Simulation of Human Mobility and Transportation Mode at a Citywide Level," in *IJCAI*, New York, USA, July 2016.
- [46] Z. Junbo, Z. Yu, and Q. Dekang, "Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction," *AAAI Conference on Artificial Intelligence; Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
- [47] C. Song, Z. Qu, N. Blumm, and A.-L. Barabasi, "Limits of predictability in Human Mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.
- [48] X. Lu, E. Wetter, N. Bharti, A. J. Tatem, and L. Bengtsson, "Approaching the Limit of Predictability in Human Mobility," *Scientific Reports*, vol. 3, 2013.
- [49] J. M. Hofman, A. Sharma, and D. J. Watts, "Prediction and Explanation in Social Systems," *Science*, Article vol. 355, no. 6324, pp. 486–488, 2017.
- [50] K. Zhao, S. Tarkoma, S. Liu, and H. Vo, "Urban Human Mobility Data Mining: An Overview," in *2016 IEEE International Conference on Big Data*, pp. 1911–1920, 2016.
- [51] L. Zhou, S. Pan, J. Wang, and A. V. Vasilakos, "Machine Learning on Big Data: Opportunities and Challenges," *Neurocomputing*, vol. 237, pp. 350–361, 2017.
- [52] S. Shaheen and N. Chan, "Mobility and the Sharing Economy: Potential to Facilitate the First- and Last-Mile Public Transit Connections," *Built Environment*, vol. 42, no. 4, pp. 573, 2016.
- [53] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban Computing: Concepts, Methodologies, and Applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 5, 2014.
- [54] Y. Long, J.C. Thill, "Combining Smart-card Data and Household Travel Survey to Analyze Jobs-housing Relationships in Beijing," *Computers, Environment and Urban Systems*, vol. 53, pp. 19–35, 2015.
- [55] C. Chen, J. Ma, Y. Susilo, Y. Liu, and M. Wang, "The Promises of Big Data and Small Data for Travel Behavior (aka Human Mobility) Analysis," *Transportation Research: Part C*, vol. 68, pp. 285–299, 2016.
- [56] G. McCabe, "Is Mobility as a Service a Fraud," *Presentation at the Bus Industry Confederation (BIC) National Conference*, 2017.
- [57] D. A. Hensher, "Tackling Road Congestion—What Might It Look Like in the Future under a Collaborative and Connected Mobility Model?" *Transport Policy*, 2018.
- [58] P.E. Holmberg, M. Collado, S. Sarasini, and M. Willander, "Mobility as a Service-MaaS: Describing the Framework," *Viktoria Swedish ICT*, 2016.
- [59] P. Jittrapirom, V. Caiati, A. M. Feneri, S. Ebrahimiagharehbaghi, M. J. A. González, and J. Narayan, "Mobility as a Service: a Critical Review of Definitions, Assessments of Schemes, and Key Challenges," *Urban Planning*, vol. 2, no. 2, p. 13, 2017.