C02029: Doctor of Philosophy *Subject NO.: 32903*
August 2020

Network-wide Spatio-Temporal Predictive Learning for the Intelligent Transportation System.

Yongshun Gong

School of Electrical and Data Engineering
Faculty of Engg. & IT
University of Technology Sydney
NSW - 2007, Australia

Network-wide Spatio-Temporal Predictive Learning for the Intelligent Transportation System.

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

by

Yongshun Gong

to

School of Electrical and Data Engineering
Faculty of Engineering and Information Technology

University of Technology Sydney
NSW - 2007, Australia

August 2020

AUTHOR'S DECLARATION

I, *Yongshun Gong* declare that this thesis, submitted in partial fulfillment of the requirements for the award of Doctor of Philosophy, in the *School of Electrical and Data Engineering*, *Faculty of Engineering and Information Technology* at the University of Technology Sydney, Australia, is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution. This research is supported by the Australian Government Research Training Program.

Production Note:

SIGNATURE: Signature removed prior to publication.

[Yongshun Gong]

DATE: 05th August, 2020

PLACE: Sydney, Australia

Acknowledgments

My deepest gratitude goes first and foremost to Prof. Jian Zhang, my principal supervisor, for his continuous support and guidance to my Ph.D. study. He has walked me through all the stages of the writing of this thesis, which greatly improved my understanding of academic writing and taught me a large amount of specific research skills. His consistent and illuminating instruction have helped me to grow as a research scientist.

Second, I would like to express my heartfelt gratitude to my co-supervisor Prof. Qiang Wu and collaborators: Dr. Wei Liu, Dr. Jinfeng Yi, Dr. Bei Chen, and Dr. Yu Zheng, for not only their constant encouragement and guidance but also for the thought-provoking questions, which have helped me to broaden my research horizon and develop the fundamental and essential academic competence.

I am also greatly indebted to all my labmates in the Global Big Data Technologies Centre: Zhibin Li, Lu Zhang, Junjie Zhang, Xiaoshui Huang, Muming Zhao, Huaxi Huang, Anan Du and Lingxiang Yao, for the weekly discussions around our research problems and enthusiastic atmosphere we worked together. I am very appreciate their encouragement and support in the past three years.

Last but not least, my heartfelt gratitude would go to my beloved family for their selfless supports and great confidence in me during my Ph.D. study.

LIST OF PUBLICATIONS

Journals:

1. **Yongshun Gong**, Zhibin Li, Jian Zhang, Wei Liu, Yu Zheng. Online Spatio-temporal Crowd Flow Distribution Prediction for Complex Metro System. *in* IEEE Transactions Knowledge and Data Engineering (TKDE), 2020.

2. Dong, Xiangjun, **Yongshun Gong***, and Longbing Cao. "e-RNSP: An efficient method for mining repetition negative sequential patterns." *in* IEEE transactions on cybernetics (TCYB), 2020: 2084-2096.

3. Dong, Xiangjun, **Yongshun Gong**, and Longbing Cao. "F-NSP+: A fast negative sequential patterns mining method with self-adaptive data storage." *in* Pattern Recognition (PR), (84) 2018: 13-27.

4. Xinming Gao, **Yongshun Gong**, Tiantian Xu, Jinhu Lv, etc.Towards to a Better Structure and Looser Constraint to Mine Negative Sequential Patterns. *in* IEEE transactions on Neural Networks and Learning Systems (TNNLS), 2020, accepted, Xinming Gao and Yongshun Gong contributed equally.

Conferences:

5. **Yongshun Gong**, Zhibin Li, Jian Zhang, Wei Liu, Bei Chen, Xiangjun Dong. A Spatial Missing Value Imputation Method for Multi-view Urban Statistical Data. *in* Proceedings of the International Joint Conferences on Artificial Intelligence (IJCAI20). pp. 1310-1316.

6. **Yongshun Gong**, Zhibin Li, Jian Zhang, Wei Liu, Jinfeng Yi. Potential Passenger Flow Prediction: A Novel Study for Urban Transportation Development. *in* Proceedings of the AAAI Conference on Artificial Intelligence (AAAI20). pp. 4020-4027.

7. **Yongshun Gong**, Zhibin Li, Jian Zhang, Wei Liu, Yu Zheng, Christina Kirsch.

Network-wide Crowd Flow Prediction of Sydney Trains via customized Online Non-negative Matrix Factorization. *in* Proceedings of the Conference on Information and Knowledge Management (CIKM18), pp. 1243-1252.

8. Zhibin Li, Jian Zhang, Qiang Wu, **Yongshun Gong**, Jinfeng Yi, Christina Kirsch. Sample Adaptive Multiple Kernel Learning for Failure Prediction of Railway Points. *in* Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (SIGKDD19), pp. 2848-2856.

9. Zhibin Li, Jian Zhang, **Yongshun Gong**, Yazhou Yao, Qiang Wu. Field-wise Learning for Multi-field Categorical Data, NeurIPS-2020, accepted.

10. Lu Zhang, Jingsong Xu, Jian Zhang, **Yongshun Gong**. Information Enhancement for Travelogues via a Hybrid Clustering Model. *in* Proceedings of the Digital Image Computing: Techniques and Applications (DICTA18), pp. 1-8.

Under Review:

1. **Yongshun Gong**, Zhibin Li, Jian Zhang, Wei Liu, Yu Zheng. Missing Value Imputation for Multi-view Urban Statistical Data via Spatial Correlation Learning. Submitted to IEEE Transactions Knowledge and Data Engineering (TKDE).

2. **Yongshun Gong**, Jinfeng Yi, Dong-Dong Chen, etc. Inferring the Importance of Items Appearance: A Step towards the Screenless Retailing. Submitted to WWW-2021.

3. **Yongshun Gong**, Bei Chen, Jianguang Lou. An Exploratory Study on the Linguistic Structure in Text-to-SQL. Submitted to ACL-2021.

4. Zhibin Li, **Yongshun Gong**, Jian Zhang, Yazhou Yao, Qiang Wu. Missingness-pattern-adaptive Learning with Incomplete Data, Submitted to ICML-2021.

5. Ping Qiu, **Yongshun Gong**, Xiangjun Dong, Longbing Cao, Chengqi Zhang. An Efficient Method for Mining Negative Sequential Patterns Under Loose Constraints. Submitted to IEEE Trans. Neural Networks and Learning Systems (TNNLS).

6. Lu Zhang, Jingsong Xu, **Yongshun Gong**, Jian Zhang. Image and Text Fusion for Travel Information Enhancement via Multi-View Embeddings. Submit to IEEE Transactions on Multimedia (TMM).

Table of Contents

List of Figures

List of Tables

Table

Abstract

Large volumes of spatio-temporal data are increasingly collected and benefited to diverse domains, including transportation, urban optimization, community detection, climate science, etc. How to feed these large-scale data into a network-wide prediction model for the intelligent transportation system is a promising problem. Currently, even though a number of traffic prediction models have been proposed to enhance the travel services and improve operational performance of transit authorities, limited methods can be applied to forecast the network-wide traffic conditions afterward.

This thesis focuses on three problems in our predictive task. Firstly, the spatio-temporal data usually suffers from the missing data problem. Those missing values hide the useful information that may result in a distorted data analysis. In Chapter 3, a spatial missing data imputation method is proposed for multi-view urban statistical data. To address this problem, our method exploits an improved spatial multi-kernel clustering approach to guiding the imputation process cooperating with an adaptive-weight non-negative matrix factorization strategy. Secondly, in the crowd flow prediction, most existing techniques focus solely on forecasting entrance and exit flows of metro stations that do not provide enough useful knowledge for traffic management. In practical applications, managers desperately want to solve the problem of getting the potential passenger distributions to help authorities improve transport services, termed as crowd flow distribution (CFD) forecasts. Therefore, to improve the quality of transportation services, three spatiotemporal models are designed in Chapter 4 to effectively address the network-wide CFD prediction problem based on the online latent space (OLS) strategy. Our models take into account the various trending patterns and climate influences, as well as the inherent similarities among different stations that are able to predict both CFD and entrance and exit flows precisely. Lastly, with the development of urbanization, a real-world demand from transportation managers is to construct a new metro station in one city area that never planned before. Authorities are interested in the picture of the future volume of commuters before constructing a new station, and estimate how it would affect other areas. In this thesis, the specific problem is termed as potential passenger flow (PPF) prediction. Chapter 5 proposes a multi-view localized correlation learning method to provide a solution for the PPF prediction that can learn localized correlations via a multi-view learning process.