

*C02029: Doctor of Philosophy
August 2020*

Subject NO.: 32903

*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

Title	i
Author's declaration	i
Acknowledgments	iii
List of Publications	v
List of Figures	xi
List of Tables	xv
Abstract	xvii
1 Introduction	1
1.1 Background	1
1.2 Research Challenges	3
1.2.1 Spatial Missing Data Imputation	3
1.2.2 Crowd Flow Distribution (CFD) Prediction	4
1.2.3 Potential Crowd Flow Prediction	6
1.3 Research Contributions	8
1.4 Thesis Structure	9
2 Literature Survey	11
2.1 Missing Data Completion	11
2.1.1 Spatial Missing Data Imputation	11
2.1.2 Multi-view Learning	12
2.1.3 Missing Data Imputation for Spatio-temporal Data	12
2.2 Traffic and Crowd Flow Prediction	14
2.2.1 Time-Series Models	14

TABLE OF CONTENTS

2.2.2	Deep Learning Model	15
2.2.3	Latent-Space Models	16
2.2.4	Network-wide Crowd Flow Prediction	18
2.3	Conclusion	20
3	Missing Value Imputation for Urban Statistical Data	21
3.1	Introduction	21
3.2	The Proposed Method	23
3.2.1	Problem Description and Preliminary	24
3.2.2	Multi-view Spatial Similarity Guidance	25
3.2.3	Adaptive-Weight NMF	27
3.2.4	Improved by Single-view and KNN Guidances	28
3.2.5	Learning Algorithm	29
3.2.6	Time complexity and convergence	31
3.3	Experiments	33
3.3.1	Datasets	33
3.3.2	Baselines & Measures	35
3.3.3	Results on Urban Statistical Datasets	37
3.3.4	Generalizability Test	38
3.3.5	The Sensitivity of Parameters	38
3.3.6	Initialization and Convergence	40
3.4	Conclusion	40
4	Crowd Flow Distribution Prediction	43
4.1	Introduction	43
4.2	Problem Description	47
4.2.1	Data Description	48
4.2.2	Problem Formulation	48
4.2.3	Exit Crowd Flow Prediction Problem	51
4.3	Online Latent Space Model: OLS-AO	52
4.3.1	The Basic Latent Space Model	52
4.3.2	Online Strategy	52
4.3.3	Learning From Side Information	55
4.3.4	Learning Process	57
4.3.5	Analysis of Complexity and Convergence	59
4.4	A Variant Model OLS-MR and a Dual-Track Model OLS-DT	60

4.4.1	Motivation	61
4.4.2	Learning the Most Recent Trend	61
4.4.3	Learning Process	62
4.4.4	A Dual-track Model	63
4.5	EXPERIMENTS	64
4.5.1	Datasets	64
4.5.2	Baselines & Measures & Parameters	65
4.5.3	Results on the State-wide Train Network	72
4.5.4	Results for the Major Stations	73
4.5.5	Results on Different Time Intervals	73
4.5.6	Visualization of Crowd Flow Distribution	73
4.5.7	Transferability Test on TaxiBJ Dataset	74
4.5.8	Ablation Study	74
4.5.9	The Sensitivity of Parameters	75
4.5.10	Scalability	77
4.6	Conclusions	78
5	Potential Passenger Flow Prediction	79
5.1	Introduction	80
5.2	Problem Statement	82
5.3	The Proposed Method	83
5.3.1	Localized Correlation Learning	85
5.3.2	Improvement by Cross-domain Learning Process	86
5.3.3	Learning and Prediction	87
5.4	Experiments	90
5.4.1	Data Description	90
5.4.2	Methods and Metrics	92
5.4.3	Comparisons on Different Time Periods	95
5.4.4	Comparisons on Various Missing Ratios	95
5.4.5	Ablation Study	95
5.4.6	Parameter Analysis	96
5.4.7	Case Study	97
5.4.8	Transfer to the Cold-start Problem	99
5.5	Conclusion	99
6	Conclusions and Future Work	101

TABLE OF CONTENTS

6.1	Conclusions	101
6.2	Future Work	102
	Bibliography	103

LIST OF FIGURES

FIGURE	Page
1.1 Regional similarity: the property of r_1 is similar to the ‘Sydney centre’ because they are neighboring each other. Although r_2 is closer to the park in terms of the physical distance, the attributes of r_2 are more analogous to ‘Sydney centre’ than the park because they have a similar functional property (business centre).	4
1.2 An example of the crowd flow distribution.	5
1.3 The example of PPF prediction problem. PPF aims to forecast the passenger flows of target areas (e.g., a_6, a_7, a_9) across the entire city network.	7
1.4 The illustration of the thesis structure.	10
2.1 Unified framework for traffic predictive model under current research review.	19
3.1 Problem description.	24
3.2 An example of building X_p^{mv} . Assume that regions \mathbf{x}_1 and \mathbf{x}_3 are falling into one cluster with the blue background, and \mathbf{x}_2 and \mathbf{x}_4 belong to another cluster with gray background. \mathbf{x}_2 and \mathbf{x}_3 are the centroid regions of two clusters, respectively. For a missing entry x_{12} , its corresponding value x_{32} is used as an imputation guide. Moreover, if the value in centroid region is missed, then a greedy strategy is implemented to find the nearest observed value (use x_{49} to fill x_{29}).	26
3.3 The example of ABS data and visualization.	34
3.4 Average RMSE with the variation of missing ratios.	37
3.5 The average RMSE in generalizability tests.	38
3.6 Effect of Parameters.	39
3.7 Convergence rate.	40
4.1 An example of crowd flow distribution.	45

4.2 The topology example of metro network. 47

4.3 A sample of our data. 49

4.4 An example of delayed data collection. Suppose there are two stations (v_1 and v_2), and we will only focus on the OD pair from v_1 to v_2 . At the current timestamp T , the data in X_T and X_{T-1} are increasing until all passengers have reached their destinations. The blue box illustrates the data we can collect at T . Can we use the collected data “3” in X_T as a complete data? No, because there are a large number of passengers still on their journeys. Does “22” indicate the complete number of travels in X_{T-1} ? Possible but uncertain, because there are many routes (or express and local train) between v_1 and v_2 , the faster one may have arrived in one time interval, but the slower one maybe not. Make our attention at X_{T-2} . Is the number “75” complete? Much more possible, because two time intervals passed. 50

4.5 The flowchart of OLS-AO. In the learning process, given a set of previous CFD matrices $\{X_t\}$ with the time window T (use $T = 4$ as an example), OLS-AO learns the latent spaces W_t and H_t of each X_t and the transitions matrices A and B by an average optimization method in section 4.3.2.3. The side information is utilized to guide the updating of W_t , H_t , A and B during the learning process. Predicted latent spaces W_{T+1} and H_{T+1} can be inferred by the Algorithm 2 shown in section 4.3.4. 51

4.6 The latent space example. It represents how to build the static latent space model for our CFD problem in each timestamp. As shown in subfigure(a), crowd flow (x_{14}) is determined by two sets of latent attributes. These attributes might illustrate many factors, such as time spans, business region, station size, etc. It is remarkable that subfigure(b) provides an example for these latent attributes when $k = 3$, and these latent attributes can be any factors without existing a strict explanation. The dimension of latent space k is a hyper-parameter. 53

4.7 An example of building indication matrix P_t . We take the entries x_{14} and p_{14} as the example. If the values meet the condition of data completion, then we can use these values as the guidance, $p_{14} = 1$ as shown in the red solid line box; if not, set p_{14} to 0 which means the collected data are incomplete yet as shown in the blue dotted line box. 53

4.8 Crowd flow changes in different scenarios. 62

4.9 CFD prediction on the entire trains network. 67

4.10	The visualization of CFD prediction.	74
4.11	Effect of parameters.	76
4.12	Comparisons between running time and various k, T	77
4.13	Convergence rate.	78
5.1	The example of PPF prediction problem. We aim to forecast the passenger flows of target areas (e.g., a_6, a_7, a_9) across the entire city network.	80
5.2	The flowchart of our proposed model. In the learning process, given a set of previous PPF matrices $\{F_d\}$, MLC-PPF learns the localized correlation matrix C and adaptive-weight W via a k -nearest indicator matrix H . The cross-domain knowledge is utilized to guide the updating of C . Then, the target prediction can be inferred by Algorithm 5.	84
5.3	Examples of transactional data.	90
5.4	Examples of ABS data.	91
5.5	City partition and station mapping.	92
5.6	Effect of parameters.	96
5.7	The case study. This figure shows the passenger flow prediction that departure from “Homebush” to other areas. To keep figure clear, we only draw our method and the ground-truth because other baselines perform far worse than the MLC-PPF.	98

LIST OF TABLES

TABLE

3.1	Symbol description.	24
3.2	The average MRE and RMSE of all missing ratios on four urban statistical datasets. Best results are bold.	36
3.3	Generalizability test. We report the average MRE and RMSE of all missing ratios and best results are bold.	39
3.4	Effects of different initialization methods.	40
4.1	Symbol description.	47
4.2	Parameters.	66
4.3	Comparisons on different time spans. We report the average mean relative errors (MRE) through all test data and best results are bold. The time spans are M-rush (7:30-9:00 AM), Non-rush (14:00-15:30 PM), A-rush (16:45-18:15 PM).	68
4.4	Overall results. We report the average errors among different methods between 6:00 AM and 10:00 PM. Best results are bold.	68
4.5	Comparisons on major stations. We report the average mean relative errors (MRE) of major stations. Best results are bold.	69
4.6	Comparisons with different time intervals. We report the average errors with different time interval between 6:00 AM and 10:00 PM. Best results are bold.	70
4.7	Transferability test on TaxiBJ dataset. We report the average mean relative errors (MRE) of city crowd flow prediction. Best results are bold.	71
4.8	Ablation studies on models. We report the average mean relative errors (MRE) of entrance CFD prediction on the entire trains network. Best results are bold.	75
4.9	Scalability test. OLS-AO/OLS-MR completed each prediction step in a reasonable time span (about 5.5 seconds) with the highest accuracy.	77
5.1	Symbol description.	83

LIST OF TABLES

5.2	Comparisons with different time periods. We report the average mean absolute errors (MAE) and normalized root mean square error (NRMSE) among various methods. The target areas occupied 20% of the total set. Best results are bold.	94
5.3	Comparisons with different removing ratios. We report MAE and NRMSE through all test data.	94
5.4	Ablation Studies on our method. We report how the adaptive matrix W and ABS guidance affect the performance. The average MAE and NRMSE conducted on the morning rush period are shown below.	96
5.5	Transfer to the cold-start problem. We report the MAE of all test methods. . .	99

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.

