
Concept Drift Adaptation for Real-time Prediction

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

Doctor of Philosophy
in
Computer Science

by

Yiliao Song

to

School of Computer Science, Faculty of Engineering and
Information Technology
Australian Artificial Intelligence Institute

University of Technology Sydney
NSW - 2007, Australia

August 2020

© 2020 by Yiliao Song
All Rights Reserved

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, *Yiliao Song* declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy in the School of Computer Science at the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference 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.

DATE: August, 2020

ABSTRACT

Concept drift refers to the phenomenon of distribution changes in a data stream. Using concept drift adaptation techniques to predict the target variable(s) of real-time data streams has gained the ever-increasing attention of researchers in recent years.

This research aims to develop a set of concept drift adaptation methods for predicting the target variable of real-time data streams. The literature review reveals two issues in the area of concept drift: i) how the concept drift problem limits the learning capability; ii) how to make adaptation in more realistic scenarios that data streams have uncertainties other than concept drift.

To address the issue i), this research discovers three root causes of limited learning capability when concept drift occurs. It is found that when concept drift occurs in a data stream, the prediction accuracy is decreased because 1) the training set contains more than one patterns so that the predictor cannot be well-learned; 2) a newly arrived data instance may present old patterns but an old instance presents the new pattern; and 3) few data instances are available when a new concept is identified at its early stage. Three concept drift adaptation methods are designed to address the three situations separately. Situation 1) is solved by developing a *fuzzy clustering-based adaptive regression* (FUZZ-CARE) approach. FUZZ-CARE can learn how many patterns exist in the training set and the membership degree of each instance belonging to each pattern; To learn the predictor with the most relevant data rather than the newest arrived data, a *segment-based drift adaptation* (SEGA) method to sequentially pick out the best segments in the training data to update the predictors. This addresses the situation 2). An *adaptive fuzzy network* (AFN) is designed to address the situation 3) through generating samples of the new concept with the previous data instances.

To address the issue ii), this research discusses the concept drift phenomenon under two scenarios that are more realistic. One is to solve the concept drift problem when data is noisy. A *noise-tolerant drift adaptation* (NoA) method is designed for handling concept drift when the data stream contains signal

noise; the other is to solve the concept drift problem when data also contains temporal dependency. A theoretical study is conducted for the regression of data streams with concept drift and temporal dependency, and based on this study, a *drift-adapted regression* (DAR) framework is established.

To conclude, this thesis not only provides a set of effective drift adaptation methods for real-time prediction, but also contributes to the development of concept drift area.

DEDICATION

To my loving husband and parents.

ACKNOWLEDGMENTS

It is a memorial and exciting journey at University of Technology Sydney (UTS) for pursuing my Ph.D. degree in the past four years. I am sincerely grateful to the people who inspired and helped me in many ways.

I would like to express my foremost and deepest gratitude to my principal supervisor, Distinguished Professor Jie Lu. Without her patience and encouragement, I would not have been able to complete this Ph.D. program. She led me into a new academic research direction, and convincingly guided and encouraged me to think and work as a professional scientist. She placed considerable trust in my research ability and unconditionally support me in pursuing my own research interests. Her wisdom and immense knowledge always enlightened me to go further and deeper in my research. Her decisiveness and sharp insights continuously motivated me when I got lost or afraid about the future. Her confidence and enthusiasm inspired me to do the right thing even when the road got tough. I felt extremely honored to be guided by such a rigorous researcher as well as an enthusiastic mentor. What she taught me and what I learned from her in the past four years has benefited my Ph.D. study and will be a great treasure throughout my life.

Meanwhile, I am greatly indebted to my co-advisor, A./Professor Guangquan Zhang. He taught me step by step how to become a qualified researcher from its beginning. He always led me to the right research direction with his expert knowledge of theory and abundant research experience. Without his critical comments, I would waste my time on trivial research ideas. Discussion with him greatly improves the scientific aspect and quality of my research. He helped me to build my confidence in my research outcomes and to be hopeful when faced with any difficulty, from academic to living.

I would like to truly thank my co-advisor, A./Professor Haiyan Lu. Without her endless trust and help, I would never have this valuable opportunity to pursue my Ph.D. research in UTS. She encouraged me to participate in different research activities which have significantly broaden my research horizons. She gave me constant care over my conditions. Her kindness helped me to go through

the tough period in my Ph.D. program. It is always pleasant to share all kinds of opinions with her.

During my Ph.D. period, I am very fortunate to join the Department of Computer Science at Southern University of Science and Technology as a visiting scholar, working with Prof. Xin Yao and Dr. Changwu Huang for three months. I would like to express my thankfulness to Prof. Xin Yao and Dr. Changwu Huang that they led me to the evolutionary computation area. Discussion and cooperation with them helped me to fill my knowledge blank of evolutionary computation efficiently and enlightened me on my own research at the same time.

I would like to express my thankfulness to every member of the Decision Systems & e-Service Intelligence Lab (DeSI) in the Centre for Artificial Intelligence (CAI) for their careful participation in my presentation and valuable comments for my research. It was a wonderful experience to spend four years with these dedicated researchers. I especially thank Dr. Ning Lu, Dr. Anjin Liu, Dr. Fan Dong and Dr. Feng Gu who helped me greatly to deeply understand my research problem during my Ph.D. candidature; Dan Shang, Hang Yu, Bin Wang and Bin Zhang who have shared their opinions and comments with me, Dr. Hua Zuo, Dr. Qian Zhang, Dr. Shan Xue who shared my joys and sadness, Dr. Dianshuang Wu who helped me in the living.

I genuinely thank Sue Felix and Robyn Barden for polishing the language of my publications. They are always patient to all my emails of questioning revised sentences. I have learned much about academic writing from them.

Meanwhile, I must thank Dr. Xin Wang and Dr. Bo Han for the valuable suggestions at the beginning of my Ph.D. study. Although they are working in different research areas, their persistence and passion on research impressed and inspired me. I thank all my wonderful friends, classmates and colleagues for every enjoyable moment.

Last, I would like to express my heartfelt appreciation and gratitude to my husband, parents, and families for their love and support.

LIST OF PUBLICATIONS

1. Y. Song, G. Zhang, H. Lu, J. Lu, "Fuzzy Clustering-based Adaptive Regression for Drifting Data Streams", *IEEE Transactions on Fuzzy Systems* , (2019). [CORE A*]
2. Y. Song, G. Zhang, H. Lu and J. Lu, "A Fuzzy Drift Correlation Matrix for Multiple Data Stream Regression", *IEEE International Conference on Fuzzy Systems* , Glasgow, Scotland (2020). [CORE A]
3. Y. Song, G. Zhang, J. Lu, H. Lu, "A Noise-tolerant Fuzzy c-Means based Drift Adaptation Method for Data Stream Regression", *IEEE International Conference on Fuzzy Systems* , New Orleans, USA (2019). [CORE A]
4. Y. Song, G. Zhang, H. Lu, J. Lu, "A Self-adaptive Fuzzy Network for Prediction in Non-stationary Environments", *IEEE International Conference on Fuzzy Systems* , Rio de Janeiro, Brazil (2018). [CORE A]
5. Y. Song, G. Zhang, J. Lu, H. Lu, "A fuzzy kernel c-means clustering model for handling concept drift in regression", *IEEE International Conference on Fuzzy Systems* , Naples, Italy (2017). [CORE A]
6. A. Liu, Y. Song, G. Zhang, and J. Lu, "Regional concept drift detection and density synchronized drift adaptation", *in the 26th International Joint*

Conference on Artificial Intelligence , Melbourne, Australia (2017). [CORE A*]

7. J. Lu, A. Liu, Y. Song, G. Zhang, "Data-driven Decision Support under Concept Drift in Streamed Big Data", *Complex & Intelligent Systems* , (2019).
8. Y. Song, J. Lu, H. Lu, G. Zhang, "A Segment-based Drift Adaptation Method for Data Streams", *IEEE Transactions on Neural Networks and Learning Systems* , (submitted). [CORE A*]
9. Y. Song, J. Lu, H. Lu, G. Zhang, "A drift-adapted regression framework for data streams with temporal dependency", *Artificial Intelligence* , (submitted). [CORE A*]
10. F. Dong, J. Lu, Y. Song, F. Liu, G. Zhang, "A Drift Region-Based Data Sample Filtering Method", *IEEE Transactions on Cybernetics* , (submitted). [CORE A]

TABLE OF CONTENTS

List of Publications	ix
List of Figures	xvii
List of Tables	xxi
1 Introduction	1
1.1 Background and Motivations	1
1.2 Research Questions and Objectives	4
1.3 Research Contributions	10
1.4 Research Significance	12
1.5 Thesis Structure	14
2 Literature Review	17
2.1 Data Stream Mining	17
2.1.1 Evolving data streams	18
2.1.2 Real-time prediction	20
2.2 A General Introduction of Concept Drift	22
2.2.1 Problem description	22
2.2.2 Types of concept drift	24

TABLE OF CONTENTS

2.2.3	Concept drift applications	25
2.3	Concept Drift Adaptation	27
2.3.1	Classification of concept drift adaptation methods	28
2.3.2	Blind concept drift adaptation	30
2.3.3	Drift detection	32
2.3.4	Informed concept drift adaptation	33
2.3.5	Concept drift adaptation methods in regression cases	35
2.3.6	Drift adaptation using fuzzy techniques	37
2.3.7	Drift adaptation in a noisy environment	38
3	Drift Adaptation by Identifying Concepts by Fuzzy Clustering	
	Process	41
3.1	Introduction	42
3.2	Definitions and Notations	45
3.3	A Fuzzy Clustering-based Drift Adaptation Method—FUZZ-CARE	49
3.3.1	Embedding clustering in objective function	49
3.3.2	Finding an optimum predictor with the minimum loss	51
3.3.3	The kernel clustering-based version	53
3.3.4	The general procedure and pseudocode of FUZZ-CARE	56
3.4	Experimental Evaluations	59
3.4.1	Validation on synthetic data	62
3.4.2	Comparison on real-world data streams	70
3.4.3	Comparison on stationary datasets	75
3.4.4	Statistical test on data streams	78
3.4.5	Parameter analysis and computation complexity	81

3.5	Summary	83
4	Drift Adaptation by Sequentially Updated Data Segments	85
4.1	Introduction	86
4.2	Definitions and Notations	88
4.3	A Segment-based Drift Adaptation Method—SEGA	89
4.3.1	The segmented symmetric degree (SSD)	90
4.3.2	Drift-gradient and how to sequentially update it	94
4.3.3	The general procedure and pseudocode of SEGA	99
4.4	Experimental Evaluations	102
4.4.1	Evaluation on synthetic data	105
4.4.2	Evaluation on real-world data streams	112
4.4.3	Statistical test of real-world data streams	120
4.4.4	Parameter analysis and computation complexity	122
4.5	Summary	125
5	Drift Adaptation by Generating Samples of New Concept	127
5.1	Introduction	128
5.2	Definitions and Notations	129
5.2.1	Generative adversarial nets (GAN)	130
5.2.2	Adaptive neuro-fuzzy inference system	132
5.3	Drift Adaptation by an Adaptive Fuzzy Network—AFN	134
5.3.1	Detection by the adversarial model in GAN	135
5.3.2	Generating data by the generative model in GAN	136
5.3.3	The general procedure and pseudocode of AFN	137
5.4	Experimental Evaluations	138

TABLE OF CONTENTS

5.4.1	Experiment setup	140
5.4.2	Experimental results	142
5.5	Summary	146
6	Drift Adaptation with Noisy Data	147
6.1	Introduction	148
6.2	Definitions and Notations	150
6.3	A Noise-tolerant Drift Adaptation Method—NoA	153
6.4	Experimental Evaluations	155
6.4.1	Evaluation on synthetic data	157
6.4.2	A case study: solar radiation prediction	162
6.5	Summary	163
7	Drift Adaptation with Temporal Dependency	165
7.1	Introduction	166
7.2	Definitions and Notations	167
7.3	Drift Adaptation with Temporal Dependency	171
7.3.1	Analysis of testing error when the real drift exists	171
7.3.2	Drift adaptation procedure in DAR	179
7.4	Experimental Evaluations	186
7.4.1	Experiments on synthetic data	189
7.4.2	Experiments on real data streams	197
7.4.3	Statistical test on data Streams	200
7.5	Summary	202
8	Conclusion and Future Research	203

8.1	Conclusions	203
8.2	Future Study	206
A	Appendix	209
A.1	Derivation step from Symmetric Degree (SD) to Segmented Sym- metric Degree(SSD)	209
A.2	Estimation of model parameters for the mixed training set	211
A.3	The locally weighted regression	211
	Bibliography	213

LIST OF FIGURES

FIGURE	Page
1.1 Standard procedure of conventional machine learning methods. . . .	2
1.2 Learning with concept drift adaption.	4
1.3 Thesis structure	16
2.1 Three main challenges in evolving data streams	19
2.2 Four types of concept drift.	25
2.3 Learning components and their classification in concept drift adapta- tion methods.	29
3.1 The training set for a data stream.	42
3.2 Learn a predictor by the OLS method using the training set that contains two patterns.	43
3.3 The difference between FUZZ-CARE and other learning methods when the same input maps to two outputs. Black dots represent a 2-dimensional dataset, where two patterns exist in the same training set. The red line on the left is the fitted line of the linear regression method. The red and blue lines on the right are the fitted lines of FUZZ-CARE.	57

LIST OF FIGURES

3.4	FUZZ-CARE flowchart showing different updating strategies.	58
3.5	Generated synthetic data with different types of drift. We generate one no-drift data on which we base five examples of drifting data. The red and black dots represent two different patterns. In the sub-figures 1)–4), the dots are drawn in a 2-dimensional plane with axis of X_t and Y_t axis; while in 5) and 6) subplots, they are drawn in a 3-dimensional sphere with an additional axis of time T	65
3.6	Estimated parameters and errors of FUZZ-CARE ^{1,2} for Rec-Drift-Mix.	69
3.7	Parameter analysis of FUZZ-CARE.	82
4.1	Relationship between the definitions proposed in this chapter.	89
4.2	An example to show the drawback of a one-sided measurement.	92
4.3	Flowchart of SEGA.	102
4.4	Parameter analysis for real-world data streams of regression tasks.	123
4.5	Parameter analysis for real-world data streams of classification tasks. Usenet1 and Usenet2 are not included because they have few instances.	123
5.1	Generative Adversarial Net (GAN).	131
5.2	Adaptive neuro-fuzzy inference system (ANFIS).	132
5.3	The flowchat of AFN.	137
6.1	Synthetic noisy data stream with different types of drift.	159

- 7.1 Error decreasing process. Circles represent the scatter plots of inputs and outputs before drift occurs, while dots show the results after drift has occurred. Colors denote the different estimation results: blue is for real values, red is for LWR-LDD⁺, and black is for DAR-linear. The testing errors of 950th-1150th instances are given in the subplot E, where the gray shadow represents LWR-NoAdapt, the red dotted line represents LWR-LDD⁺ and the black line represents DAR-linear. . . 194
- 7.2 The improved percentage of MAE/MAPE of DAR-linear from LWR-LDD⁺(linear). Both methods have drift adaptation process. The difference is that DAR-linear is implemented on the reconstructed space. . 197

LIST OF TABLES

TABLE	Page
3.1 Experimental design in this section.	60
3.2 A comprehensive comparison of different editions of FUZZ-CARE method on different types of drift.	66
3.3 Validation on real-world data streams. MAE and its corresponding rank are used as the evaluations	73
3.4 Comparison results on 13 stationary datasets.	77
3.5 Friedman test and its post-hoc test of all the methods over all eight data streams.	80
3.6 Friedman test and its post-hoc test of all the methods over House, Sensor20, Sensor46, SMEAR and Solar data streams.	80
4.1 Validation of the effectiveness of drift-gradient	100
4.2 Type I and Type II error of drift-gradient under three distributions .	100
4.3 Experimental design in this section.	104
4.4 Validation of Regression Tasks on Synthetic Data (MAE as the Evalu- ation Criterion).	109
4.5 Validation on Synthetic Data of Classification Tasks (Acc as the Eval- uation Criterion).	111

LIST OF TABLES

4.6	Validation on Real-world Data of Regression Tasks (MAE as the Evaluation Criterion).	115
4.7	Validation on Real-world Data of Classification Tasks (Acc as the Evaluation Criterion).	118
4.8	Friedman test and its post-hoc test of all the methods over real-world regression data streams (no CCPP), where "Friedman Test" is the result for Friedman test and "Friedman - post-hoc test after Conover" is for the pairwise comparison. "+", "*", "**", and "***" means this value is significant at the level of 0.1, 0.05, 0.01 and 0.001 respectively. "df" denotes the freedom degree.	121
4.9	Friedman test and its post-hoc test of all the methods over real-world classification data streams, where "Friedman Test" is the result for the Friedman test and "Friedman - post-hoc test after Conover" shows the pairwise comparison. "+", "*", "**", "***" and "df" have the same meaning as they are in Table 4.8	121
4.10	Run-time on real-world data streams (s CPU-time)	126
5.1	Experimental design in this section.	139
5.2	Main results table.	144
5.3	Robustness test of AFN with different parameters.	145
6.1	The uniqueness of NFA.	150
6.2	Experimental design in this section.	156
6.3	Evaluation results on synthetic data.	161
6.4	Evaluation on the SMEAR data	163

7.1	Experimental Design	188
7.2	Parameters for generating multi-dimensional non-linear data in (7.32)	190
7.3	Comparisons between NoAdapt and adaptation by LDD ⁺ approaches on different spaces, and DAR with different predictors.	192
7.4	Comparisons of MAE between NoAdapt and adaptation by LDD ⁺ approaches on different spaces, and DAR with different predictors (multi-dimensional non-linear cases).	195
7.5	Comparisons of MAPE between NoAdapt and adaptation by LDD ⁺ approaches on different spaces, and DAR with different predictors (multi-dimensional non-linear cases).	196
7.6	MAE comparisons between different methods on real-world data streams. The DAR methods outperform the compared methods on data streams where concept drift occurs.	199
7.7	Friedman test and its post-hoc test of all the methods over all seven data streams, where "Friedman Test" is the result for Friedman test and "Friedman - post-hoc test after Conover" is for the pairwise com- parison. "*", "**", and "***" means this value is significant at the level of 0.05, 0.01 and 0.001 respectively. "df" denotes the freedom degree.	201