

# **Handling Concept Drift Using the Correlation between Multiple Data Streams**

**by Bin Zhang**

Thesis submitted in fulfilment of the requirements for  
the degree of

**Doctor of Philosophy**

under the supervision of Dist. Prof. Jie Lu and A/Prof.  
Guangquan Zhang

University of Technology Sydney  
Faculty of Engineering and Information Technology

October 2022

# CERTIFICATE OF ORIGINAL AUTHORSHIP

I, *Bin Zhang*, declare that this thesis is submitted in fulfillment of the requirements for the award of *Doctor of Philosophy*, in the *Faculty of Engineering and Information Technology* at the University of Technology Sydney

This thesis 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 Research Council.

Production Note:

Signature: Signature removed prior to publication.

Date: October 24th, 2022

# ABSTRACT

Machine learning has been widely applied to handle big data. In real-world applications, data are in the form of streams. Streaming data bring new challenges to machine learning. Concept drift is a major problem in handling streaming data. The newly arrived data have a different distribution from historical data. Hence, machine learning models do not work on the newly arrived data. Many methods have been proposed to detect whether concept drift occurs and to update the machine learning models to the drift.

To date, the research on concept drift considers data streams separately. That is, these methods handle data streams one by one, ignoring whether correlations exist between data streams. However, in real-world applications, correlations between data streams are widespread. If the correlations between data streams were available, this would support better decision making, rather than handling data streams separately. There is currently no research on how to model the correlations between data streams. Data streams are dynamic. Therefore, it is necessary to develop methods to track the correlations between data streams and to adapt to changes in correlations. In addition, after concept drift is detected, there is usually insufficient data to train a new model, which can lead to the over-fitting problem. How to deal with the over-fitting issue is still a challenge. Motivated by this, this research proposes several methods to overcome the aforementioned challenges.

To alleviate the over-fitting problem, a concept drift adaptation method, Drift Adaptation via Joint Distribution Alignment (DAJDA), is proposed. DAJDA performs a linear transformation to the drift instances instead of modifying the model. Instances are transformed into a common feature space, reducing the discrepancy of distributions before and after drift. Using additional historical data to train a new model can lead to better performance due to the increasing training set. Experimental studies show that DAJDA has the ability to improve the performance of the learning model under concept drift.

To model the correlations between data streams, we propose a novel Multi-stream Concept Drift Handling Framework via data sharing, containing fuzzy membership-based drift detection (FMDD) and fuzzy membership-based drift adaptation (FMDA) components, to train the new learning model for drifting streams by sharing weighted data from other non-drifting streams. A stream fuzzy set is defined with membership functions that measure the degree to which samples belong to a data stream. Our Concept Drift Handling Framework can detect when and in which streams concept drift occurs, and therefore, the over-fitting issue can be solved by adding the weighted data from non-drifting streams to train new learning models. Synthetic and real-world experiment results show that our method can help avoid the over-fitting issue caused by a lack of data and thereby significantly improve the prediction performance.

To track changes in correlations and adapt to correlation drift, we propose an ensemble chain-structured model, Evolutionary Regressor Chains. To provide the model with the ability to search the optimal order of the chain, we design a heuristic order searching strategy to be incorporated as part of the model’s ongoing process. The heuristic order searching strategy can also update the chains as time passes, to track the dynamicity of the correlations. A diversity pruning method is also proposed to reduce computation complexity while retaining the diversity of the ensemble. We undertake a theoretical analysis, and give the dynamic regret bound of our method. Our experiment results show that our Evolutionary Regressor Chains method can track data stream correlations accurately. Chains can update themselves to adapt to both concept drift and correlation drift. The performance of the machine learning models on data streams is improved.

To learn meta knowledge across multiple data streams, a concept drift adaptation strategy - Learning to Fast Adapt in the Evolving Environment - is proposed for neural network classifiers in the non-stationary environment. A meta-level LSTM recurrent neural network is used to learn a proper parameter updating rule instead of updating methods that are traditional gradient descent-based (e.g. stochastic

gradient descent) during fine-tuning. A suitable updating step will be generated according to current loss and its gradient for the parameter of a neural network classifier. Experiments on both synthetic and real-world data sets show that our method can quickly adapt the neural network classifier to concept drift and help improve the performance of the classifier in a dynamic environment.

# Dedication

I dedicate my dissertation work to my parents Xia Xiang and Xuxin Zhang.

## Acknowledgements

It has been an indelible journey to complete my doctoral degree at the esteemed University of Technology Sydney. I am profoundly grateful to all those who have extended their assistance in myriad ways throughout this remarkable endeavor.

Foremost, I wish to express my sincerest appreciation to my principal supervisor, the highly esteemed Distinguished Professor Jie Lu. Her unwavering guidance has been an illuminating beacon, navigating me towards the successful culmination of my research. In addition to fostering an exceptional research environment, she bestowed upon me a generous scholarship that served as a solid bedrock for my four-year inquiry. Under her tutelage, I delved into uncharted research domains, identifying and addressing critical gaps with her expert counsel. Her unwavering dedication to scholarship served as a constant source of inspiration and solace during challenging times. Beyond research, her invaluable professional advice on life and potential career trajectories has shaped my path with utmost clarity. Moreover, her pursuit of innovation and unwavering commitment to scholarly rigor has left an indelible mark on my academic journey.

I also extend my deepest gratitude to my esteemed co-supervisor, Associate Professor Guangquan Zhang. His unwavering support and invaluable suggestions pertaining to literature review, manuscript writing, and paper revision have been instrumental in shaping the trajectory of my doctoral study. It is his trust, encouragement, and unwavering assistance that have played an integral role in my successful completion of this demanding program.

Joining the Decision Systems & e-Service Intelligence Lab has been an absolute pleasure, providing me with an enriching experience alongside numerous enthusiastic and dedicated colleagues. I wish to express my heartfelt gratitude to Dr. Anjin

Liu, Dr. Yiliao Song, Dr. Hang Yu, Dr. Bin Wang, and my fellow peers who have generously offered their guidance and suggestions throughout the course of my doctoral candidature.

Additionally, I would like to extend my sincerest appreciation to my parents, whose unwavering support has been a constant source of strength regardless of my location or pursuits. I am forever indebted to their encouragement and unwavering belief in my abilities. To my dear friends Donglai Yang and Shiyu Ma, I cherish the memories of our shared experiences and remain deeply grateful for your unwavering companionship during the challenging times imposed by the Covid-19 pandemic. I would like to express special appreciation to my best friend, Zihe Liu, whose presence has brought immeasurable joy and laughter to my life. Our bond is truly extraordinary, and I am immensely thankful for the camaraderie we share.

In summary, the successful completion of my doctoral journey would not have been possible without the support, guidance, and unwavering belief of those mentioned above. Their collective contributions have left an indelible mark on my academic and personal growth, and for that, I am eternally grateful.

Bin Zhang  
Sydney, Australia, 2022.



# Contents

Certificate	i
Dedication	v
Acknowledgments	vi
List of Figures	xiii
List of Tables	xvi
Abbreviation	xvii
Notation	xviii
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Research Questions and Objectives . . . . .	3
1.2.1 Research Questions . . . . .	4
1.2.2 Research Objectives . . . . .	4
1.3 Research Significance . . . . .	6
1.3.1 Theoretical Significance . . . . .	6
1.3.2 Practical Significance . . . . .	6
1.4 Research Contributions . . . . .	6
1.5 Thesis Organization . . . . .	9
<b>2 Literature Review</b>	<b>13</b>
2.1 Setting and Definitions . . . . .	13

2.2	Concept Drift Detection . . . . .	15
2.2.1	Detection Methods based on Error Rate . . . . .	15
2.2.2	Detection Methods based on Data Distribution . . . . .	16
2.2.3	Detection Methods based on Multiple Hypothesis Test . . . . .	19
2.3	Concept Drift Adaptation . . . . .	20
2.3.1	Retraining Models . . . . .	20
2.3.2	Adaptive Models . . . . .	21
2.3.3	Adaptive Ensembles . . . . .	24
2.4	Other Related Work . . . . .	28
2.4.1	Transfer Learning . . . . .	28
2.4.2	Fuzzy Systems in Machine Learning . . . . .	29
2.4.3	Multi-Output Learning . . . . .	30
2.4.4	Meta Learning . . . . .	31
2.5	Summary . . . . .	33
<b>3</b>	<b>Drift Adaptation via Joint Distribution Alignment</b>	<b>35</b>
3.1	Introduction . . . . .	35
3.2	Preliminaries . . . . .	37
3.2.1	Maximum Mean Discrepancy . . . . .	37
3.3	Proposed Method . . . . .	37
3.3.1	Marginal Distribution Discrepancy . . . . .	38
3.3.2	Conditional Distribution Discrepancy . . . . .	38
3.3.3	Preserve Data Variance . . . . .	39
3.3.4	Learning Algorithm . . . . .	41
3.4	Experimental Study . . . . .	44

3.4.1	Verification of the Effectiveness . . . . .	44
3.4.2	Evaluation on Real-world Datasets . . . . .	47
3.5	Summary . . . . .	51
<b>4</b>	<b>A Multi-stream Concept Drift Handling Framework via Data Sharing</b>	<b>52</b>
4.1	Introduction . . . . .	52
4.2	Proposed Method . . . . .	57
4.2.1	Problem Settings . . . . .	57
4.2.2	Basic Assumptions . . . . .	58
4.2.3	Stream Fuzzy Set . . . . .	58
4.2.4	Fuzzy Membership-based Drift Detection Method . . . . .	60
4.2.5	Fuzzy Membership-based Drift Adaptation Method . . . . .	61
4.3	Experimental Study . . . . .	64
4.3.1	Data Sets . . . . .	65
4.3.2	Evaluation of Stream Fuzzy Set . . . . .	68
4.3.3	Evaluation of FMDD . . . . .	69
4.3.4	Evaluation of FMDA on Synthetic Data Sets . . . . .	73
4.3.5	Evaluation of FMDA on Real-world Data Sets . . . . .	74
4.3.6	Batch Size Study on Real-world Data Sets . . . . .	80
4.3.7	Visualization of the Correlation between Data Streams . . . . .	84
4.4	Summary . . . . .	85
<b>5</b>	<b>Evolutionary Regressor Chains for Multi-stream Regres- sion</b>	<b>86</b>
5.1	Introduction . . . . .	86

5.2	Problem Settings . . . . .	90
5.3	Proposed Method . . . . .	92
5.3.1	Regressor Chains . . . . .	92
5.3.2	Heuristic Order Searching . . . . .	93
5.3.3	Online Training Procedure . . . . .	96
5.3.4	Diversity Pruning . . . . .	98
5.4	Theoretical Analysis . . . . .	102
5.4.1	Metrics . . . . .	102
5.4.2	Dynamic Regret Analysis . . . . .	103
5.5	Experimental Study . . . . .	106
5.5.1	Data Sets . . . . .	106
5.5.2	Experiment Setting . . . . .	107
5.5.3	Parameter Study of the Number of Chains . . . . .	108
5.5.4	Evaluation of the Diversity Pruning . . . . .	110
5.5.5	Evaluation of the Evolutionary Regressor Chains . . . . .	112
5.6	Summary . . . . .	114
<b>6</b>	<b>Learning to Fast Adapt in the Evolving Environment</b>	<b>115</b>
6.1	Introduction . . . . .	115
6.2	Proposed Method . . . . .	117
6.2.1	Model Description . . . . .	118
6.2.2	Coordinate-wise Parameter Sharing . . . . .	118
6.2.3	Preprocessing . . . . .	119
6.2.4	Training Procedure . . . . .	120
6.3	Experimental Study . . . . .	124

6.3.1	Datasets . . . . .	124
6.3.2	Experiment Settings . . . . .	124
6.3.3	Evaluation of LFAEE . . . . .	125
6.3.4	Evaluation of LFAEE with Cross-stream Meta Learners . . . .	126
6.4	Summary . . . . .	127
<b>7</b>	<b>Conclusions and Future Study</b>	<b>128</b>
7.1	Conclusions . . . . .	128
7.2	Future Study . . . . .	130

# List of Figures

1.1	Framework for handling concept drift . . . . .	3
1.2	The structure of the thesis. . . . .	12
2.1	Illustration of two types of concept drift. . . . .	14
2.2	Four patterns of changes in data distributions and outlier. . . . .	14
4.1	The paradigm of handling concept drift for single data stream. After new data arrives, the learning model gives its prediction. The true label is then available. A concept drift detection procedure is conducted to detect whether concept drift occurs. If yes, a concept drift adaptation method is used to update the old learning model. . .	53
4.2	Electricity price trend of NSW and VIC. The change in price in the two states shows a significant correlation. Negative data exist because the data set has been normalized . . . . .	55
4.3	Flow chart of Multi-stream Concept drift Handling Framework . . . .	64
4.4	Correlation between the MMD and the reciprocal of fuzzy membership. There exists a positive correlation between the reciprocal of our proposed fuzzy membership and MMD. However, the average computing time of fuzzy membership is 1.039e-4 seconds, while the average computing time of MMD is 2.419e-2 seconds. . . . .	76

4.5	Results of evaluation of FMDA on synthetic data sets. The x-axis is the size of training set and the y-axis is the average mean absolute percentage error. The box plot show the medians and quartiles of average mean absolute percentage errors of twelve data sets. The medians are connected by dashed lines. . . . .	77
4.6	Results of batch study on Train data set. Each sub-figure shows result of a data stream. There are eight streams in total. The x-axis is the size of batch and the y-axis is the average mean squared error.	81
4.7	Results of batch study on Weather data set. Each sub-figure shows result of a data stream. There are ten streams in total. The x-axis is the size of batch and the y-axis is the average mean squared error.	82
4.8	Visualization of the correlation between data streams. The color of block means the correlation between two data streams represented by fuzzy membership. . . . .	85
5.1	An illustration of a basic chain structure built on four data streams .	93
5.2	Performance of models with different numbers of chains on five real-world datasets. The first subfigure show the mean squared error with different number of chains. The second subfigure show the average running time to process one sample. . . . .	109
5.3	Performance of models with 3/5/7 chains left after pruning and without pruning. The first row show the mean squared error of models with different number of chains left after pruning of three datasets. The second row show the average running time of models with different number of chains left after pruning of three datasets. 10 chains left means no pruning procedure is conducted. . . . .	111
6.1	The framework of our strategy . . . . .	119
6.2	The mechanism of coordinatewise parameter sharing . . . . .	120

6.3	The computational graph for calculation the loss of a episode . . . .	122
-----	---	-----



# List of Tables

3.1	Synthetic Data Streams to Evaluate DAJDA . . . . .	46
3.2	Parameters of the Synthetic Data Streams to Evaluate DAJDA . . . . .	46
3.3	Experiment Results of Verification of the Effectiveness of DAJDA . . . . .	48
3.4	Real-world Data Streams to Evaluate DAJDA . . . . .	49
3.5	Average Rank of DAJDA on All Three Datasets . . . . .	49
3.6	Experiment Results of Evaluation of DAJDA on Real-world Dataset . . . . .	50
4.1	Summary of Datasets to Evaluate the Multi-stream Concept Drift Handling Framework . . . . .	68
4.2	Results of evaluation of FMDD on Synthetic Streams . . . . .	70
4.3	Results of evaluation on Train . . . . .	78
4.4	Results of evaluation on Weather . . . . .	79
5.1	Real-world datasets . . . . .	107
5.2	Evaluation of Evolutionary Regressor Chains on Three Multi-Stream dataset . . . . .	113
6.1	Real-world Data Stream to Evaluate LFAEE . . . . .	125
6.2	Evaluation of LFAEE on Four Real-world Datasets . . . . .	126
6.3	Evaluation of LFAEE with Cross-stream Meta Learners . . . . .	126

## Abbreviation

ARF: Adaptive Random Forest

DAJDA: Drift Adaptation via Joint Distribution Alignment

DDM: Drift Detection Method

ECDD: EWMA for Concept Drift Detection

ERC: Ensemble Regressor Chains

FMDA: Fuzzy Membership-based Drift Adaptation

FMDD: Fuzzy Membership-based Drift Detection

HAT: Hoeffding Adaptive Tree

HDDM: Hoeffding's inequality based Drift Detection Method

HT: Hoeffding Tree

JDA: Joint Distribution Adaptation

MMD: Maximum Mean Discrepancy

MOA: Massive Online Analysis

MTRS: Multi-target Regression Stacking

NSW: New South Wales

PAC: Probably approximately Correct learning

PCA: Principal Component Analysis

RC: Regressor Chains

SEA: SEA Moving Hyperplane Concepts

ST: Single Target

VIC: Victoria

## Nomenclature and Notation

$X_t^{(i)}$  is the feature at time  $t$  of the  $i$ th stream.

$y_t^{(i)}$  is the label at time  $t$  of the  $i$ th stream.

$S_i$  is the  $i$ th stream.

$P(\cdot)$  is the probability.

$E(\cdot)$  is the expect.