

UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

**Development of Robust and Scalable Hyperbox
based Machine Learning Algorithms**

by

Thanh Tung KHUAT

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

Sydney, Australia

2021

Certificate of Original Authorship

I, Thanh Tung KHUAT, declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Computer Science, 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 Government Research Training Program.

Signature:

Production Note:

Signature removed prior to publication.

Date: 01 July 2021

Abstract

Together with the rapid development of digital information and the increase in amount of data, machine learning (ML) algorithms have been developed and evolved constantly to discover new information and knowledge from different data sources. The use of hyperbox fuzzy sets as fundamental representational and building blocks in learning algorithms forms an important branch of ML. Hyperbox-based algorithms have a huge potential for high scalability and incremental adaptation to applications working in the dynamically changing environments. Additionally, learning algorithms based on hyperbox representations can form interpretable models, which are highly desirable for areas with the requirement of safety and trust. This study aims to develop and expand robust, scalable, and transparent learning algorithms for hyperbox-based classification models with a specific focus on a general fuzzy min-max neural network (GFMMNN).

First of all, a comprehensive survey on hyperbox-based machine learning models together with empirical assessments of the GFMMNN on pattern classification problems were conducted. Next, a new online learning algorithm was proposed for the GFMMNN and improved the robustness of the whole family of GFMMNN learning algorithms to work effectively with mixed-attribute data by introducing a new learning mechanism for categorical features. In terms of scalability, the main steps of the learning algorithms were reformulated so they can be effectively executed on graphics processing units using matrix operations, simultaneously proposing mathematical lemmas to reduce the redundancies of hyperbox candidates in the learning process. This thesis also proposed a novel method to enhance the transparency of classifiers while maintaining a good classification performance by using hierarchical granular representations from hyperbox fuzzy sets. The last contribution was a simple but powerful ensemble model built from many individual hyperbox-based classifiers trained on random subsets of both sample and feature spaces. Extensive empirical analyses indicated that the proposed solutions are highly competitive with other evaluated learning algorithms.

To my loved ones

Acknowledgments

The Ph.D. study is a long journey, and this thesis would not have been possible without the support and encouragement of my supervisors, friends, and relatives during my Ph.D. research journey. Maybe the hardest part when I write this thesis is how to express my sincere gratitude to them.

First and foremost, I would especially like to express my sincerest gratitude to my principal supervisor, Professor Bogdan Gabrys for his continuous support, motivation, enthusiasm, inspiration to my Ph.D. study, together with his invaluable advice and discussion during our weekly meetings. His guidance is likely to positively affect me throughout my future career path. He will definitely be the first person to whom I seek advice for the difficult research problems that I have to deal with in the future. My Ph.D. research has faced a difficult time because of the spread of the COVID-19 pandemic. However, under his supervision and instructions, I have been quick to change research plans and adapt to the new ways of working using remote working tools. As a result, my research progress has not been negatively impacted by the pandemic. Those are precious experiences that will never be forgotten.

I would also like to thank my co-supervisors, Distinguished Professor Fang Chen and Dr. Dymitr Ruta for their valuable comments and discussion on the research manuscripts that I have written for my Ph.D. projects. My thanks also go to my Ph.D. assessment panel, Professor Paul Kennedy and Associate Professor Wei Liu, for their constructive feedback. Special thanks also to Dr. Hanh Le for encouraging me to pursue Ph.D. research. I wish to thank UTS for awarding me scholarships so that I can conduct my Ph.D. project.

A large part of this thesis comes from publications which have been peer-reviewed by anonymous reviewers. Therefore, I also wish to send a special thank to them who made valuable suggestions to enhance the quality of my research papers.

Additionally, I would like to thank all my best friends who have accompanied me throughout my Ph.D. journey: Tien-Dung Nguyen and his family, Thong Do, Thac Do, Cong Nguyen, Tung Huynh, An Le, Dung Duong, Xuan Yang, Joakim Skarding, Sunny Verma, and many others. Happy conversations and all the fun we have had during the past three years will be unforgotten memories in my life.

Finally, I would like to express my love and boundless gratitude to my parents and younger sister for their unconditional support and sacrifice. They have supported me both financially and emotionally so that I can completely concentrate on my Ph.D. research.

Thanh Tung Khuat
Sydney, Australia, 2021.

List of Publications

Published Papers

1. **Thanh Tung Khuat**, and Bogdan Gabrys, “Random hyperboxes,” *IEEE Transactions on Neural Networks and Learning Systems* (Early Access). (*Chapter 8*)
2. **Thanh Tung Khuat**, and Bogdan Gabrys, “An in-depth comparison of methods handling mixed-attribute data for general fuzzy min-max neural network,” *Neurocomputing*, vol. 464, pp. 175-202, 2021. (*Chapter 5*)
3. **Thanh Tung Khuat**, Fang Chen, and Bogdan Gabrys, “An effective multi-resolution hierarchical granular representation based classifier using general fuzzy min-max neural network,” *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 2, pp. 427- 441, 2021 (**IEEE CIM Publication Spotlight paper**). (*Chapter 7*)
4. **Thanh Tung Khuat**, and Bogdan Gabrys, “Accelerated learning algorithms of general fuzzy min-max neural network using a novel hyperbox selection rule,” *Information Sciences*, vol. 547, pp. 887-909, 2021. (*Chapter 6*)
5. **Thanh Tung Khuat**, Dymitr Ruta, and Bogdan Gabrys, “Hyperbox based machine learning algorithms: A comprehensive survey,” *Soft Computing*, vol. 25, pp. 1325–1363, 2021. (*Chapter 2*)
6. **Thanh Tung Khuat**, and Bogdan Gabrys, “A comparative study of general fuzzy min-max neural networks for pattern classification problems,” *Neurocomputing*, vol. 386, pp. 110-125, 2020. (*Chapter 3*)
7. **Thanh Tung Khuat**, Fang Chen, and Bogdan Gabrys, “An improved online learning algorithm for general fuzzy min-max neural network,” in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, pp. 1-9,

2020 (**IEEE CIS Outstanding Student Paper Conference Registration Grant**). (*Chapter 4*)

8. **Thanh Tung Khuat**, and Bogdan Gabrys, “Accelerated training algorithms of general fuzzy min-max neural network using gpu for very high dimensional data,” in *Proceedings of the 26th International Conference on Neural Information Processing (ICONIP)*, pp. 583-595, 2019 (**Best paper award finalists**). (*Chapter 6*)

Under Reviewed Papers

- 9 **Thanh Tung Khuat**, and Bogdan Gabrys, “An online learning algorithm for a neuro-fuzzy classifier with mixed-attribute data,” Submitted to *IEEE Transactions on Fuzzy Systems* (Revised and Resubmit). (*Chapter 5*)

Contents

Certificate	ii
Abstract	iii
Dedication	iv
Acknowledgments	v
List of Publications	vii
List of Figures	xvi
List of Tables	xxii
List of Abbreviations	xxvii
List of Notations	xxix
1 Introduction	1
1.1 Background and Motivation	2
1.2 Research Aims and Objectives	8
1.3 Original Contributions	9
1.4 Thesis Organization	12
2 Literature Review	16
2.1 Introduction	16
2.2 Overview and Taxonomy of Hyperbox-based Learning Algorithms . . .	20
2.3 Fuzzy Min-Max Neural Network Architectures	23
2.3.1 Analysis of the Original Fuzzy Min-Max Neural Network . . .	26

2.3.2	Variants Using Specialized Neurons for Overlapping Areas . . .	28
2.3.3	Modified Variants Without Using Specialized Neurons for Overlapping Areas	33
2.4	Hyperbox based Hybrid Machine Learning Models	41
2.5	Other Hyperbox based Machine Learning Models	45
2.6	Discussion and Research Directions	50
2.6.1	Discussion	50
2.6.2	Research Directions	51
2.7	Summary	53
3	General Fuzzy Min-Max Neural Network: Practices and Existing Issues	54
3.1	Introduction	54
3.2	General Fuzzy Min-Max Neural Network and Learning Algorithms . .	56
3.2.1	An Overall Architecture	56
3.2.2	Online Learning Algorithm	58
3.2.3	Agglomerative Learning Algorithm	62
3.2.4	Accelerated Agglomerative Learning Algorithm	65
3.3	Existing Problems and Motivations	67
3.4	Experiments and Results	70
3.4.1	The Influence of the Maximum Hyperbox Size on the Performance of the Online Learning based GFMMNN	70
3.4.2	The Influence of the Similarity Threshold on the Performance of the Agglomerative Learning based GFMMNN using Different Similarity Measures	76

3.4.3	Comparison of Different Agglomerative Learning Versions of GFMMNN	77
3.4.4	The Influence of Data Presentation Order on the Performance of GFMMNN	79
3.4.5	Comparison of GFMMNN and Other Types of Fuzzy Min-Max Neural Networks	80
3.4.6	Comparison of GFMMNN and Other Machine Learning Algorithms	86
3.5	Discussion	90
3.6	Summary	91
4	Improved Online Learning Algorithm for General Fuzzy Min-Max Neural Network	93
4.1	Introduction	93
4.2	Improved Online Learning Algorithm	97
4.2.1	Training Phase	97
4.2.2	Classification Phase	100
4.2.3	Time Complexity of the IOL-GFMM Algorithm	100
4.3	Experiments	100
4.3.1	Comparison of the Performance and Complexity of the Proposed and Original Learning Algorithms of GFMMNN	101
4.3.2	Evaluation the Robustness of the Proposed Method to Noise	104
4.3.3	Comparison of the Performance of the IOL-GFMM to Other Fuzzy Min-Max Classifiers	107
4.3.4	A Simple Ensemble Learning Approach to Tackle the Disadvantages of the IOL-GFMM	108

4.4	Summary	112
5	Mixed-Attribute Data Classification using General Fuzzy Min-Max Neural Network	113
5.1	Introduction	114
5.2	Extended Improved Online Learning Algorithm for Mixed-Attribute Data	117
5.2.1	Formal Description	117
5.2.2	Architecture of the GFMMNN for Mixed-Attribute Data . . .	118
5.2.3	Extended Improved Online Learning Algorithm for Mixed-Attribute Data Classification	119
5.2.4	Properties of the Change in Entropy of Categorical Features when Accommodating New Training Samples	123
5.2.5	Time Complexity Analysis for the EIOL-GFMM Algorithm . .	124
5.3	Experimental Results	125
5.3.1	Analyzing the Sensitivity of Parameters	126
5.3.2	Comparing the Performance of the EIOL-GFMM Algorithms with Other Methods using the Fixed-Parameter Settings . . .	130
5.3.3	Evaluating the Role of the Hyper-Parameter Tuning on the Performance of the EIOL-GFMM Algorithms	138
5.3.4	Discussion on Existing Issues of the Proposed Method	145
5.4	Summary	146
6	Accelerated Learning Algorithms for General Fuzzy Min- Max Neural Network	147
6.1	A Solution based on Matrix Operations and Advanced Engineering . .	148
6.1.1	Overview	148

6.1.2	Implementation of Learning Algorithms by Matrix Operations	149
6.1.3	Experiments for the First Solution	154
6.2	A Solution based on New Hyperbox Selection Rules	157
6.2.1	Overview	157
6.2.2	Accelerated Online Learning Algorithms	158
6.2.3	Accelerated Agglomerative Learning Algorithms	159
6.2.4	Experiments for the Second Solution	165
6.3	Summary	172
7	Learning at Different Granularity Levels using Hyper-	
	box Representations	174
7.1	Introduction	174
7.2	Multi-Resolution Hierarchical Granular Representation based Classifier	178
7.2.1	Overview	178
7.2.2	Formal Description	180
7.2.3	Missing Value Handling	184
7.2.4	Interpretability of the Proposed Classifier	184
7.3	Experiments	185
7.3.1	Performance of the Proposed Method on Synthetic Datasets .	188
7.3.2	Performance of the Proposed Method on Real Datasets	198
7.3.3	The Vital Role of the Pruning Process and the Use of Sample Centroids	201
7.3.4	Ability to Handling Missing Values	202
7.3.5	Comparison to State-of-the-art Studies	202
7.4	Summary	205

8	Ensemble Learning using Hyperboxed-based Classifiers	206
8.1	Introduction	207
8.2	Random Hyperboxes Model	210
8.2.1	Formal Description	210
8.2.2	Time Complexity	212
8.2.3	Properties of the Random Hyperboxes	212
8.3	Experimental Results	216
8.3.1	Analyzing the Random Hyperboxes Classifier	217
8.3.2	Comparing the Performance of the Random Hyperboxes to Other Classifiers	223
8.4	On the Estimation of Generalization Error Bounds and Open Problems	233
8.5	Summary	236
9	Conclusion and Future Work	237
9.1	Key Findings and Conclusions	237
9.2	Future Work	239
9.3	Final Summary	244
A	Additional Results for Chapter 3	247
B	Additional Results for Chapter 4	254
C	Additional Results for Chapter 5	255
D	Additional Results for Chapter 6	276
E	Additional Results for Chapter 7	289
F	Additional Results for Chapter 8	300

Bibliography

317

List of Figures

1.1	A summary of motivation, objectives, and contributions in this thesis.	12
1.2	Thesis structure	15
2.1	3-D Hyperbox	21
2.2	A taxonomy of hyperbox-based learning algorithms	22
2.3	A demonstration of a hyperbox-based classification model.	23
2.4	Contraction of hyperboxes B_1 and B_2 with elimination along one coordinate	29
3.1	An architecture of general fuzzy min-max neural network	57
3.2	A hyperbox-based model is trained on the <i>Iris</i> dataset	68
3.3	A hyperbox-based model is trained on the <i>Iris</i> dataset	69
3.4	A hyperbox-based model is trained on the <i>Iris</i> dataset	70
3.5	The change in the number of hyperboxes, training time, and testing error of the <i>Waveform</i> dataset	71
3.6	Average rank of performance on 16 datasets using different values of θ	72
3.7	The change in the number of hyperboxes, training time, and testing error of the <i>Ringnorm</i> dataset	74
3.8	The influence of similarity threshold on the number of hyperboxes and testing errors for GFMMNN using agglomerative learning on the <i>Thyroid</i> dataset	76

4.1	A drawback of the contraction procedure.	94
4.2	A drawback of the current online learning algorithm for GFMMNN.	95
4.3	An example of hyperbox-based model and noise.	96
4.4	The performance and model complexity of the proposed and original online learning algorithm.	102
4.5	A limitation of the IOL-GFMM algorithm	109
5.1	The extended architecture of GFMMNN for mixed-attribute data	118
5.2	The change in the class balanced accuracy according to the different values of α for the <i>flag</i> dataset ($\theta = 1, \delta = 1$).	127
5.3	The change in the class balanced accuracy according to the different values of α for the <i>flag</i> dataset ($\theta = 0.1, \delta = 0.1$).	128
5.4	The change in the class balanced accuracy according to the different values of δ for the <i>flag</i> dataset ($\theta = 1, \alpha = n/(n + r)$).	129
5.5	Critical difference diagram for four learning algorithms ($\theta = 0.1$).	133
5.6	Critical difference diagram for four learning algorithms ($\theta = 0.7$).	134
5.7	Critical difference diagram for four learning algorithms ($\theta = 1$).	134
5.8	Critical difference diagram for the proposed method and the original algorithm using encoding methods ($\theta = 0.1$).	136
5.9	Critical difference diagram for the proposed method and the original algorithm using encoding methods ($\theta = 0.7$).	137
5.10	Critical difference diagram for the proposed method and the original algorithm using encoding methods ($\theta = 1$).	137
5.11	The distribution of the obtained α values for different methods used to find α and the CBA values for the <i>flag</i> dataset ($\theta = 1, \delta = 1$).	141
5.12	The distribution of the obtained α values for different methods used to find α and the CBA values for the <i>flag</i> dataset ($\theta = 0.1, \delta = 0.1$).	141

5.13	A CD diagram of four learning algorithms using the hyper-parameter tuning method.	145
6.1	The speedup factor of the AGGLO-2 and AGGLO-SM algorithms according to the number of samples and the number of features over 24 experimental datasets (using the “longest distance” similarity measure).	165
7.1	Pipeline of the training process of the proposed method	179
7.2	The error rate of classifiers on synthetic linear boundary datasets with the different number of samples	189
7.3	The error rate of classifiers on synthetic linear boundary datasets with the different number of classes	192
7.4	The error rate of classifiers on synthetic linear boundary datasets with the different number of features	192
7.5	The error rate of classifiers on synthetic non-linear boundary datasets with the different number of samples	196
7.6	The error rate of classifiers on the <i>Letter</i> datasets through data abstraction levels	200
8.1	The variances of RH models and their base learners (<i>plant species leaves margin</i> dataset).	218
8.2	Average weighted-F1 scores through 40 testing folds of a single model using training sets with top-k most used features (<i>plant species leaves margin</i> dataset).	219
8.3	The change in the average weighted-F1 scores when increasing the number of base learners (<i>plant species leaves margin</i> dataset).	220

8.4	The change in the average weighted-F1 scores when increasing the maximum number of used dimensions (<i>plant species leaves margin</i> dataset).	220
8.5	Average correlation scores between base learners when increasing the maximum number of used dimensions (<i>plant species leaves margin</i> dataset).	221
8.6	The change in the average weighted-F1 scores when increasing the maximum hyperbox size (<i>plant_species_leaves_margin</i> dataset).	223
8.7	Comparison of average weighted-F1 scores of the random hyperboxes and the best value from single FMNNs.	225
8.8	Critical difference diagram for the performance of the RH classifier and other FMNNs ($\theta = 0.1$).	225
8.9	Critical difference diagram for the performance of the RH classifier and other FMNNs ($\theta = 0.7$).	228
8.10	Critical difference diagram for the performance of the RH classifier and other ensemble models.	230
8.11	Critical difference diagram for the performance of the RH classifier and other popular learning algorithms.	232
8.12	The relationship of the difference in the estimated upper error bound and actual testing error with respect to the ratio of the average number of training samples per class and the number of features.	235
C.1	A demonstration for the methods used to estimate the values for parameter α	262
C.2	Class balanced accuracy values for different values of α ($\theta = 1, \delta = 1$).	263
C.3	Class balanced accuracy values for different values of α using the IOL-GFMM-v1 algorithm ($\theta = 0.1, \delta = 0.1$).	264

C.4	Class balanced accuracy values for different values of α using the IOL-GFMM-v2 algorithm ($\theta = 0.1, \delta = 0.1$).	265
C.5	Class balanced accuracy values for different values of δ using the IOL-GFMM-v1 algorithm ($\theta = 1, \alpha = n/(n + r)$).	266
C.6	Class balanced accuracy values for different values of δ using the IOL-GFMM-v2 algorithm ($\theta = 1, \alpha = n/(n + r)$).	267
C.7	Class balanced accuracy values and range of obtained α for different ways of estimating α ($\theta = 1, \delta = 1$).	273
C.8	Class balanced accuracy values and range of obtained α for different ways of estimating α (using EIOL-GFMM-v1 with $\theta = 0.1, \delta = 0.1$).	274
C.9	Class balanced accuracy values and range of obtained α for different ways of estimating α (using EIOL-GFMM-v2 with $\theta = 0.1, \delta = 0.1$).	275
E.1	Demonstration of the training process proposed in Chapter 7	290
E.2	An dendrogram showing the changes in the number of hyperboxes through different levels of granularity for the example in Figure E.1	291
E.3	The error rate of classifiers on several real testing datasets with the change in the data abstraction levels	297
F.1	The variances of the random hyperboxes models and their base learners for different datasets.	305
F.2	The probability of the number of used features for all base learners over different datasets.	306
F.3	The probability of each feature used for all base learners over different datasets.	308
F.4	Average weighted-F1 scores over 40 testing folds of a single model using training sets with top-k most used features over different datasets.	309

F.5	Weighted-F1 score of the random hyperboxes and IOL-GFMM for the <i>PEMS database</i> dataset	310
F.6	Weighted-F1 score of the random hyperboxes and IOL-GFMM for the <i>Complex Hydraulic System</i> dataset	310
F.7	The change in the average weighted-F1 scores when increasing the number of base learners for different datasets.	312
F.8	The change in the average weighted-F1 scores when increasing the maximum number of used dimensions for different datasets.	315
F.9	The change in the average weighted-F1 scores when increasing the maximum hyperbox size threshold for different datasets.	316

List of Tables

2.1	Summary of the main hyperboxed-based machine learning algorithms	24
3.1	Comparison of fixed and adaptive maximum hyperbox size parameters ($\theta = 0.26$)	73
3.2	Comparison of fixed and adaptive maximum hyperbox size parameters ($\theta = 0.56$)	75
3.3	The comparison of the full similarity matrix based agglomerative learning and accelerated agglomerative learning, $\theta = 0.26, \sigma = 0.8$	78
3.4	Standard deviation on results of different versions of GFMMNNs due to the impact of presentation order	79
3.5	Testing error ranking of the different FMNN variants	85
3.6	Comparison of the average testing errors of the GFMMNN with other machine learning algorithms	87
3.7	Testing error ranking of GFMMNN and other machine learning algorithms	87
3.8	Outcomes of Holm post-hoc test for AGGLO-2	89
3.9	Outcomes of Holm post-hoc test for incremental learning based GFMMNN	89
4.1	Average training time (in seconds) of GFMMNNs using the original and improved learning algorithms (smaller values are better and highlighted in bold)	104

4.2	Testing errors on noisy datasets in the case of using $\theta = 0.1$	105
4.3	Testing errors on noisy datasets in the case of using $\theta = 0.7$	106
4.4	The lowest average testing errors (%) of models on nine experimental datasets (smaller values are better and shown in bold) .	108
4.5	The average rank of the performance of models through nine datasets (the best value is highlighted in bold)	108
4.6	The average standard deviation (%) of algorithms	110
4.7	Average testing errors (%) of the IOL-GFMM and the ensemble method	111
5.1	The average rank for the algorithms using their best settings	132
5.2	The average ranks for the proposed method and the original IOL-GFMM using different encoding techniques	136
5.3	Average rank for different methods used to find values for parameter α	140
5.4	Average ranks for the learning algorithms using the hyper-parameter tuning approach	143
6.1	Summary of high dimensional datasets for experiments in Chapter 6 .	154
6.2	Training time in seconds of the <i>PEMS Database</i> dataset	155
6.3	Training time in seconds of the <i>Complex Hydraulic System</i> dataset . .	156
6.4	Testing time in seconds of the <i>Complex Hydraulic System</i> dataset . .	156
6.5	Speed-up factor and the number of hyperbox candidates considered during the training process of online learning algorithms	166
6.6	Speed-up factor of the AGGLO-2 algorithm	168
6.7	The number of hyperboxes considered during the training process of the AGGLO-2 algorithm	169

6.8	Speed-up factor of the AGGLO-SM algorithm	170
6.9	Number of hyperbox candidates considered during the training process of the AGGLO-SM algorithm	171
7.1	The lowest error rates and training time of classifiers on synthetic linear boundary datasets with different number of samples ($n = 2, C = 2$)	190
7.2	The lowest error rates and training time of classifiers on synthetic linear boundary datasets with different classes ($N = 10K, n = 2$) . . .	193
7.3	The lowest error rates and training time of classifiers on synthetic linear boundary datasets with different features ($N = 10K, C = 2$) . .	194
7.4	The lowest error rates and training time of classifiers on synthetic non-linear boundary datasets with different number of samples ($n = 2, C = 2$)	197
7.5	The real datasets and their statistics for experiments in Chapter 7 . .	198
7.6	The change in the number of generated hyperboxes through different levels of granularity of the proposed method	199
7.7	The role of the pruning process and the use of sample centroids . . .	201
7.8	The training time and the lowest error rates of the proposed method on the datasets with missing values	203
7.9	The AUC value of the proposed method and other methods on the <i>susy</i> dataset	204
7.10	The accuracy of the proposed method and other methods on the <i>PhysioNet MIT-BIH Arrhythmia</i> dataset	204
8.1	The average weighted-F1 scores of the random hyperboxes and other fuzzy min-max neural networks ($\theta = 0.1$)	226

8.2	The average weighted-F1 scores of the random hyperboxes and other fuzzy min-max neural networks ($\theta = 0.7$)	227
8.3	The average weighted-F1 scores of the random hyperbox model and other ensemble models	229
8.4	The average weighted-F1 scores of the random hyperboxes and other machine learning algorithms	232
8.5	Estimated upper generalization error bounds (%), real testing error (%), and their standard deviations computed from different assessment methods	234
A.1	Datasets were used for experiments in Chapter 3	247
A.2	Average performance of different variants of FMNN	248
A.3	Average performance of variants of FMNN using a pruning procedure	250
B.1	Descriptions of datasets for experiments in Chapter 4	254
C.1	Experimental dataset used in Chapter 5	261
C.2	Average class balanced accuracy for the proposed EIOL-GFMM algorithms and two existing algorithms with the mixed-attribute learning ability	268
C.3	The Average Number of Generated Hyperboxes from the Proposed Method and Other Existing Algorithms with Mixed-Attribute Learning Ability	269
C.4	The average class balanced accuracy for the proposed method and the IOL-GFMM using encoding techniques	270
C.5	The Average Class Balanced Accuracy for Different Methods Used to Find the Values for Parameter α	271

C.6	The average class balanced accuracy for the learning algorithms using the hyper-parameter tuning method	272
C.7	The rank for the learning algorithms using the hyper-parameter tuning method	272
D.1	The summary of the used datasets for the second solution of accelerating learning algorithms	285
D.2	Training time of online learning algorithms in Chapter 6	286
D.3	Training time of the AGGLO-2 algorithm in Chapter 6	287
D.4	Training time of the AGGLO-SM algorithm in Chapter 6	288
E.1	The real datasets and their statistics for experiments in Chapter 7 . .	296
E.2	The lowest error rates and training time of classifiers on real datasets	298
F.1	Training time (s) of the IOL-GFMM and random hyperboxes model on the high dimensional datasets	311
F.2	Testing time (s) of the IOL-GFMM and random hyperboxes model on the high dimensional datasets	311
F.3	The descriptions of the used datasets in Chapter 8	314

List of Abbreviations

2lv-CSWCE	Two-level classification system with testing in dynamically changing environment
2-pHC	Two-phase hyperbox classifier
ACO	Ant colony optimization
Adaboost	Adaptive boosting
AGGLO-SM	Agglomerative learning algorithm using the full similarity matrix
AGGLO-2	Agglomerative algorithm version 2
ANNs	Artificial Neural Networks
ARC	Adaptive resolution classifier
ART	Adaptive Resonance Theory
AUC	Area under the curve
CACL	Concept adapting contraction less
CBA	Class balanced accuracy
CD	Critical difference
CLFMNN	Contraction-less fuzzy min-max neural network
CNs	Compensatory Neurons
DCFMMN	Data core based fuzzy min-max neural network
DPS	Density-preserving sampling
ECTs	Ensemble model of clustering trees
EFMNN	Enhanced fuzzy min-max neural network
EFMNN-ACO	Enhanced fuzzy min-max neural network with ant colony optimization
EFMMDT	Evolving fuzzy min-max decision tree
EFMNN-II	Enhanced fuzzy min-max neural network with K-nearest hyperbox expansion rule and pruning
EFMNN-UL	Evolved fuzzy min-max neural network for unknown labeled data
EIOL-GFMM	Extended improved online learning algorithm for general fuzzy min-max neural network
EMILP	A enhanced version of mixed integer linear programming model-based hyperbox classifier
Esb-GFMMNN	Ensemble of neuro-fuzzy classifiers
Ens-IOL-GFMM(DL)	Ensemble of GFMMNNs using the IOL-GFMM algorithm at the decision level
Ens-IOL-GFMM(ML)	Ensemble of GFMMNNs using the IOL-GFMM algorithm at the model level
FMCN	Fuzzy min-max neural network with compensatory neurons
FMM-CART	Offline and online fuzzy min-max neural network and classification and regression trees
FMM-CT	Fuzzy min-max clustering neural network with the clustering tree
FMM-GA	Fuzzy min-max neural network with genetic algorithms
FMMDT	Fuzzy min-max neural network based decision tree
FMM-PSO	Fuzzy min-max neural network with the particle swarm optimization
FMM-ECT	Fuzzy Min-Max neural network with an ensemble of clustering trees
FMNWSM	Fuzzy min-max neural network with symmetric margin
FMNN	Fuzzy min-max neural network for classification
FMNN-clu	Fuzzy min-max neural network for clustering
FPGAs	Field-programmable gate arrays

GA	Genetic algorithms
GFMM	General fuzzy min-max
GFMMNN	General fuzzy min-max neural network
GFMMNN-CD1	General fuzzy min-max neural networks for categorical data
GFMMNN-CD2	Enhanced general fuzzy min-max neural networks for categorical data
GNB	Gaussian Naive Bayes
GPU	Graphics Processing Unit
GRFMN	General reflex fuzzy min-max neural
HACO	Hyperbox based clustering with ant colony optimization
HACO2	Hyperbox classifier with ant colony optimization
HFC	Hyperbox fuzzy classifier
HNN	Hyperbox neural network algorithm
IEFCN	Inclusion/exclusion fuzzy hyperbox classification network
IGs	Information granules
IOL-GFMM	Improved online learning algorithm for general fuzzy min-max neural network
KNEFMN	Enhanced fuzzy min-max neural network with K-nearest hyperbox expansion rule
KNN	K-nearest neighbors
LightGBM	Light Gradient Boosting Machines
LOO	Leave-One-Out
MILP	Mixed integer linear programming model-based hyperbox classifier
MDCFMM	Modified data-core-based fuzzy min-max neural network
MEFMN	Modified-enhanced fuzzy min-max neural network
MFMC	Modified fuzzy min-max neural network for clustering
MFMMN	Modified fuzzy min-max neural network for two-stage pattern classification
EFMNNC	Enhanced fuzzy min-max neural network for clustering
MFMMN-GA	Modified fuzzy min-max neural network with genetic algorithms
MLF	Multi-level fuzzy min-max neural network
MMM-BL	Modified fuzzy min-max neural network with a new batch learning algorithm
MRHGRC	Multi-resolution hierarchical granular representations based classifier
Onln-GFMM	Original online learning algorithm for general fuzzy min-max neural network
PARC	Pruning Adaptive resolution classifier
RBF	Radial Basis function
ReFMN	Reflex Fuzzy Min Max Neural Network
ReFMN-FN	Reflex Fuzzy Min Max Neural Network with floating neurons
RFMMN	Refined Fuzzy Min-Max Neural Network
RH	Random Hyperboxes
RMILP	Refined mixed integer linear programming model-based hyperbox classifier
SAS	Smart adaptive systems
SFMN	Stochastic fuzzy min-max neural network
SS-FMM	Semi-supervised classification method based on fuzzy min-max neural network
SVM	Support Vector Machines
TDFMM	Top-down fuzzy min-max
TDFMMR	Top down fuzzy min-max regressor
TEH-GFMMNN	Tree ensemble hyperboxes via general fuzzy min-max neural network
XGBoost	Extreme Gradient Boosting
WFMM	Weighted fuzzy min-max neural network

List of Notations

If there is no specific definition in each section, the following are the default meanings of notations used in this thesis.

Symbol	Meaning
B_i	The i -th hyperbox in the list of hyperboxes
$X = [X^l, X^u]$	An input sample in the form of lower and upper bounds
\mathbf{X}	A sample space contains all input samples
$\mathcal{T}_N = \{(X_i, c_i)\}_{i=1}^N$	A training data set with N samples
c_i	Class label of sample X_i
c_X	Class label of sample X
\mathcal{C}	A set of categorical variables denoting classes to which the observations fall into
n	Number of features of input samples. If a input sample contains both continuous and categorical features, then n denotes the number of continuous features
r	Number of categorical features
$d(A, B)$	Euclidean distance between two vectors A and B
$X^l = (x_1^l, \dots, x_n^l)$	Lower bound of an input sample in n -dimensional space
$X^u = (x_1^u, \dots, x_n^u)$	Upper bound of an input sample in n -dimensional space
$X^d = (x_1^d, \dots, x_r^d)$	A vector contains r categorical attributes for an input sample
$V_i = (v_{i1}, \dots, v_{in})$	A vector represents the minimum point of the hyperbox B_i in n -dimensional space
$W_i = (w_{i1}, \dots, w_{in})$	A vector represents the maximum point of the hyperbox B_i in n -dimensional space
G_i	Sample centroid of the hyperbox B_i
n_i	The number of current samples included in the hyperbox B_i
$b_i(X, V_i, W_i)$	Membership value between the input sample X and the hyperbox B_i
\mathcal{V}	A matrix contains all minimum points V_i of all hyperboxes B_i generated in the learning process
\mathcal{W}	A matrix contains all minimum points W_i of all hyperboxes B_i generated in the learning process
\mathcal{L}	A vector contains all class labels for all hyperboxes B_i generated in the learning process
θ	Maximum hyperbox size
Θ	A list of maximum hyperbox sizes
σ	Minimum similarity threshold so that two hyperboxes B_i and B_k can be aggregated
γ	Sensitivity parameter to control the decreasing speed of membership values
$H_j(B_i)$	The current entropy of hyperbox B_i for the j -th categorical feature
Ω_{ij}	A set of categorical values on the j -th categorical attribute of the hyperbox B_i
s_{ik}	Middle gap similarity measure between two hyperboxes B_i and B_k
\tilde{s}_{ik}	shortest gap similarity measure between two hyperboxes B_i and B_k
\hat{s}_{ik}	longest gap similarity measure between two hyperboxes B_i and B_k

Symbol	Meaning
Φ	A randomizing vector
Φ_i	An independent and identically distributed random vector
M	Number of base learners in an ensemble model
$\mathbb{1}(\cdot)$	Indicator function
\mathbb{E}	Expectation
r_s	Sampling rate to find the number of training samples for base learners in an ensemble model
m_f	The maximum number of used features for each base learner in an ensemble model
\mathcal{E}^*	Upper bound of the generalization error
$\bar{\rho}$	Average correlation between base learners in an ensemble model
\bar{K}	The average number of hyperbox candidates with the same class as the input sample
\bar{R}	The average number of hyperboxes representing classes different from the class of the input pattern in each iteration