

CERTIFICATE

It is certified that the work contained in the thesis titled “*Learning Optimal Decision Criteria for Early Classification*” by *Anshul* has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all requirements of Comprehensive Examination, Candidacy, and SOTA for the award of Ph.D. Degree.

Supervisor

Prof. Sanjay Kumar Singh

Professor and HOD,

Department of Computer Science and Engineering,

Indian Institute of Technology (BHU) Varanasi,

Uttar Pradesh, INDIA 221005.

DECLARATION BY THE CANDIDATE

I, *Anshul*, certify that the work embodied in this Ph.D. thesis is my own bonafide work carried out by me under the supervision of *Prof. Sanjay Kumar Singh* from *January 2017* to *March 2021* at *Department of Computer Science and Engineering*, Indian Institute of Technology (BHU) Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, *etc.* reported in journals, books, magazines, reports, dissertations, theses, *etc.*, or available at websites and have not included them in this thesis and have not cited as my own work.

Date:

Place: Varanasi

(Anshul)

CERTIFICATE BY THE SUPERVISOR

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

(Prof. Sanjay Kumar Singh)

Professor and HOD,

Dept. of Computer Science and Engineering,

Indian Institute of Technology (BHU) Varanasi

Signature of Head of Department

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: Learning Optimal Decision Criteria for Early Classification

Name of the Student: Anshul

Copyright Transfer

The undersigned hereby assigns to the Institute of Technology (Banaras Hindu University) Varanasi all rights under copyright that may exist in and for the above thesis submitted for the award of the *Doctor of Philosophy*.

Date:

Place: Varanasi

(Anshul)

Note: However, the author may reproduce or authorize others to reproduce material extracted verbatim from the thesis or derivative of the thesis for author's personal use provided that the source and the Institute's copyright notice are indicated.

Dedicated

to

My family

ACKNOWLEDGEMENT

First of all, I would like to take this opportunity to express my profound gratitude and sincere regards to my supervisor Dr. Sanjay Kumar Singh, Professor & HOD, Department Computer Science and Engineering, Indian Institute of Technology (BHU), Varanasi, for his excellent guidance, monitoring, and constant encouragement throughout this doctorate program. Also, special thanks for creating such a great working atmosphere in the lab.

I am obliged to faculty members and staff members of the Computer Science and Engineering Department, Indian Institute of Technology (BHU), Varanasi. Also, I want to express my deepest gratitude to Mr. Aneesh G Nath, Mr. Ritesh Sharma, and Dr. Rishav Singh for providing moral support throughout my thesis work. Further, I extend my special gratitude to my colleagues and friends, Dr. Sandeep Sambhaji Udamle, Mr. Abhinav, and Miss. Vandana Bharti supported me throughout my thesis work and provided me technical, moral, and emotional support during my research work. I have also spent some of the craziest and most memorable moments with them.

Finally, I would like to express my deepest gratitude to my parents, Mr. Prem Shankar and Mrs. Rama Devi, for their constant support and encouragement. I am incredibly grateful to My wife Sushma and my daughter Arushi for their tremendous understanding and encouragement in the past few years. Without their support, it would be impossible for me. I am also grateful to my other family members who have supported me along the way.

(Anshul)

Contents

List of Figures	xi
List of Tables	xiv
List of Symbols	xv
List of Abbreviations	xvii
Preface	xix
1 Introduction	1
1.1 Motivation	4
1.2 Challenges and main objectives of this thesis	6
1.3 Contributions of the thesis	7
1.4 Organization of the thesis	8
2 Background and Related Works	11
2.1 Background	11
2.1.1 Time series representation	11
2.1.2 Time series distance/similarity measures	13
2.1.3 Time series classification	14
2.1.4 Early classification on time series	16
2.2 Related works	18
2.2.1 Instance-based methods	19
2.2.2 Shapelet-based methods	20
2.2.3 Model-based methods	22
2.2.4 Other methods	24
2.2.5 Research Gap	29

3	Early classification on UTS	31
3.1	3.1 Introduction	31
3.2	3.2 Preliminaries	32
3.3	3.3 Early classification by measuring uncertainty	33
3.3.1	3.3.1 Model description	34
3.3.2	3.3.2 Experimental evaluation	39
3.4	3.4 Early classification by learning optimal decision rule	45
3.4.1	3.4.1 Model description	45
3.4.2	3.4.2 Classification and optimization method	51
3.4.3	3.4.3 Experimental evaluation	53
3.5	3.5 Summary	66
4	Early classification on MTS	67
4.1	4.1 Introduction	67
4.2	4.2 Motivation and significant contributions:	68
4.3	4.3 Model description	69
4.3.1	4.3.1 Training phase	70
4.3.2	4.3.2 Prediction phase	76
4.4	4.4 Experimental evaluation	76
4.4.1	4.4.1 Evaluation metrics	76
4.4.2	4.4.2 Dataset description	77
4.4.3	4.4.3 Parameter selection	78
4.4.4	4.4.4 Results analysis	79
4.5	4.5 Summary	89
5	Early classification on time series by using deep learning approach	91
5.1	5.1 Introduction	91
5.2	5.2 Motivation and significant contributions	93
5.3	5.3 Preliminaries	94
5.4	5.4 Model description	95
5.4.1	5.4.1 Training phase	95
5.4.2	5.4.2 Prediction phase	102
5.5	5.5 Experimental evaluation	102
5.5.1	5.5.1 Datasets description	102
5.5.2	5.5.2 Results analysis	104
5.6	5.6 Summary	110

6 Conclusion and Future Directions	113
6.1 Conclusion	113
6.2 Future directions	116
References	117
List of Publications	130

List of Figures

1.1	Illustration of (a) Traditional TSC and (b) Early classification approach.	3
1.2	A general framework of early classification model	4
2.1	Taxonomy of time series representation and dimensionality reduction methods	12
2.2	Traditional TSC approach.	14
2.3	Early classification on time series.	16
3.1	Models Comparison: CD diagram for accuracy using Nemenyi post-hoc procedure ($\alpha = 0.05$). The goodness of models is considered as best to worse (<i>left to right</i>). The bold line shows that methods do not yield statistically significance differences.	44
3.2	Models Comparison: CD diagram for earliness using Nemenyi post-hoc procedure ($\alpha = 0.05$). The goodness of models is considered as best to worse (<i>right to left</i>). The bold line shows that methods do not yield statistically significance differences.	44
3.3	Block diagram of the proposed early classification model	46
3.4	Learning of classifier using truncated training set at every time point t	47
3.5	Prediction Process	51
3.6	CD diagram for accuracy. The goodness of the model is considered as best to worse (left to right), and the bold line shows that methods do not yield statistical significance differences.	58
3.7	CD diagram for earliness. The goodness of the model is considered as best to worse (left to right), and the bold line shows that methods do not yield statistical significance differences.	58
3.8	Accuracy and earliness plot for ESR1, ESR2, and ESR3	59
3.9	Accuracy and earliness plot for five sample datasets includes Coffee, ECG200, Gun_Point, Synthetic Control and Wafer	60

3.10	Accuracy and earliness plot over different values of α for ESR1, ESR2 and ESR3	61
3.11	Effect of accuracy, earliness, and execution time for PSO and GA	63
4.1	Block diagram of the proposed model for early classification on MTS	69
4.2	Training a set of classifiers for MTS	71
4.3	Cost evaluation process for a MTS \mathbf{X}	75
4.4	Prediction process for a incoming MTS \mathbf{X}	76
4.5	Effect of α parameter : (a) Scattered plot between Accuracy and Earliness by taking average over all the datasets (b) Earliness vs. Accuracy plot for individual dataset.	79
4.6	Accuracy plot at increasing length of MTS on different datasets.	81
4.7	Effect of α parameter on different combination of ESRs and cost function.	82
4.8	Regularization effect on ESRs $\mathcal{R}1_{\Theta}$ and $\mathcal{R}2_{\Theta}$	83
4.9	Accuracy and earliness plot of ESR $\mathcal{R}1$ for $\alpha \in \{0.8, 0.9\}$	84
4.10	Accuracy and earliness plot of ESR $\mathcal{R}2$ for $\alpha \in \{0.8, 0.9\}$	85
4.11	Comparison of the proposed model with GP-full by considering $\alpha = 0.9$	87
5.1	Classification model for sensory time series data.	93
5.2	The Block diagram of the proposed ETMD model is divided into two phases: training and prediction. Part (a) denotes the training process of base classifier \mathcal{H} and its architecture. Part (b) represents the confidence threshold selection for early decision making. Part (c) represents the prediction process on incoming time series data.	96
5.3	Accuracy vs. earliness curve for α	105
5.4	Effect of α on accuracy and earliness over individual transportation mode.	106
5.5	Confusion matrix (a) TMD dataset (b) SHL dataset	107
5.6	Comparison between proposed ETMD and traditional approach. The last <i>bar</i> , named as average, denotes the mean accuracy or earliness over all transportation modes.	108
5.7	Effect of optimizers SGD and Adam	109
5.8	Accuracy vs. epoch and loss vs. epoch learning curve of hybrid-DL classifier.	109

List of Tables

2.1	Comparative analysis of early classification approaches	25
3.1	Intraclass accuracy at different time points	36
3.2	Accuracy of five classifiers on fifteen dataests: GP(DotProduct), GP(rbf), SVM(linear), SVM(rbf), Naive Bayes	41
3.3	Earliness of five classifiers on fifteen dataests: GP(DotProduct), GP(rbf), SVM(linear), SVM(rbf), Naive Bayes	41
3.4	Accuracy values on fifteen datasets for ECTS, EDSC , RelClass, ECDIRE and Proposed method	42
3.5	Earliness values on fifteen datasets for ECTS, EDSC, RelClass, ECDIRE and Proposed.	43
3.6	Parameter setting for other early classification models	55
3.7	Accuracy values for ECTS, EDSC, RelClass, ECDIRE, ECTS_OAE, and Proposed (ESR1, ESR2, ESR3) methods.	56
3.8	Earliness values for ECTS, EDSC, RelClass, ECDIRE, ECTS_OAE, and Proposed (ESR1, ESR2, ESR3) methods.	57
3.9	Domination counts for the proposed ESR1, ESR2 and ESR3 compared to other methods. The first entry tells about how many times the proposed model dominates other methods, and third entry tells how many times others dominate the proposed model. In between, second entry refers to the draw condition.	59
3.10	The performance of the proposed early classification model (ESR3) on malware dataset by considering different values of α	65
4.1	Datasets description	77
4.2	Comparison of proposed models with other methods	86
4.3	Accuracy and earliness of proposed model over six MTS datasets	88
5.1	Comparison of base classifiers in ETMD model	110

List of Symbols

Symbol	Description
\mathcal{D}	A training dataset with labeled instance
X	A time series
M	A number of instances in training set
T	Length of a complete time series
\mathcal{Y}	A set of class labels
X_t	A incomplete time series of length t
X_T	A complete time series of length T
\mathcal{D}_t	A truncated training dataset in which each $X \in \mathcal{R}^t$
\mathbf{X}	A Multivariate Time series
h	A classifier
\mathcal{H}	A set of classifiers
C_f	A cost function
α	A balancing parameter between accuracy and earliness
K	Number of classes in training set
π	A vector of class probabilities for UTS
Π	A matrix of class probabilities for MTS
\hat{y}_t	A predicted class label at time step t
t^*	A time step at which decision of class prediction is made.
δ	Confidence threshold
λ	A regularization parameter
h_t^v	A classifier at time step t for v^{th} variable of MTS

Abbreviations

Abbreviation	Description
CD	Critical Difference
ECTS	Early Classification on Time Series
EDSC	Early Distinctive Shapelet Classification
ESR	Early Stopping Rule
ETMD	Early Transportation Mode Detection
GA	Genetic Algorithm
GP	Gaussian Process
ITS	Intelligent Transportation Systems
MC FEC	Mining Core Feature method for Early Classification
MPL	Minimum Predictive Length
MTS	Multivariate Time Series
PC	Probabilistic Classifier
PSO	Particle Swarm Optimization
QDA	Quadratic Discriminant Analysis
REACT	Reliable EARly ClassificaTion
RNN	Recurrent Neural Network
SGD	Stochastic Gradient Descent
SHL	Sussex-Huawei Locomotion and Transportation
SVM	Support Vector Machine
TMD	Transportation Mode Detection
TSC	Time Series Classification
UTS	Univariate Time Series

Preface

Early classification of time series is valuable in many real-world applications where data is generated over time. The aim of early classification is to predict the class label of incoming time series as early as possible before observing its complete sequence. In general, whenever early prediction time improves, the prediction accuracy decreases. In other words, one can achieve better accuracy by waiting for more data points in the series, but it will delay the response time. In time-sensitive applications, it is worth sacrificing some classification accuracy in favour of early predictions, preferably early enough for taking actionable decisions. Thus, there exists a trade-off between earliness and accuracy. However, existing approaches do not consider trade-off optimization well in their decision criteria.

Time Series Classification (TSC) is one of the major research areas that developed over the past few years, mainly due to its practical applicability in various domains such as agriculture, healthcare, medicine, finance, and industries. The main objective of TSC is to maximize prediction accuracy. In contrast, an early classification of time series has two conflicting objectives, i.e., accuracy and earliness. Nowadays, the early classification of time series attracts researchers more due to its useful applications in various domains such as early disease prediction, early gas leakage prediction, drought prediction, etc.

This thesis focuses on the problem of early classification of time series by learning optimal decision criteria. The problem of early classification has been identified as

the composition of two sub-problems. The first one is to design the early classifier that can label the incomplete time series. The second is to define the decision criteria that can estimate the right time for making an online decision. Initially, we propose an early classification model for Univariate Time Series (UTS), which relies on two factors (i) a set of probabilistic classifier and (ii) a confidence threshold. The confidence threshold ensures the reliability of class prediction defined by measuring the uncertainty in predicted output. In this method, decision policy is more inclined toward accuracy and does not take trade-off optimization into consideration. In this regard, a further optimization-based approach has been adapted for early classification and defines the early stopping rules for optimal decision making, which have been learned through optimization between accuracy and earliness simultaneously.

Furthermore, this optimization-based approach has been extended for Multivariate Time Series (MTS), which is more challenging than UTS because of the multiple variables involved in decision making. An ensemble-based system has been designed to label the incomplete MTS, and collective output from all the variables has been utilized for decision making. These proposed methods are highly effective for small training data sets, but feature transformation is required for training the classifiers. Finally, a deep learning-based hybrid classifier has been proposed that can capture the temporal information from the raw sensory data effectively to perform the classification task. Moreover, the optimal confidence threshold has been defined by balancing the trade-off between accuracy and earliness. The proposed approaches have been evaluated on publicly available datasets and they demonstrated effective solutions for early classification on time series.