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

List of Tables

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
MCFEC	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.