Tribhuvan University

Institute of Science and Technology

HYBRID FEATURE SELECTION AND FEATURE EXTRACTION BASED ENSEMBLE METHOD IN CLASSIFICATION

A Dissertation

Submitted to:

Central Department of Computer Science and Information Technology
Tribhuvan University, Kirtipur

In partial fulfillment of the requirements
for the Master's Degree in Computer Science & Information Technology

Submitted by:

Rajesh Pandey
Sept, 2015

Supervisor

Prof. Dr. Shashidhar Ram Joshi

Tribhuvan University

Institute of Science and Technology

Central Department of Computer Science and Information Technology

Student's Declaration

I hereby declare that I am the only author of this work and that no sources other than the listed here have been used in this work.

… … … … …

Rajesh Pandey

Date: 22th Sept, 2015

Tribhuvan University

Institute of Science and Technology

Central Department of Computer Science and Information Technology

Supervisor's Recommendation

I hereby recommend that the dissertation prepared under my supervision by **Mr. Rajesh Pandey** entitled "**Hybrid feature selection and feature extraction based ensemble method in classification**" be accepted as in fulfilling partial requirement for the completion of Masters Degree of Science in Computer Science & Information Technology.

Prof. Dr. Shashidhar Ram Joshi
Department of Electronics & Computer Engineering,
Institute of Engineering,
Pulchowk, Nepal
Date: 23th Sept, 2015

Tribhuvan University

Institute of Science and Technology

Central Department of Computer Science and Information Technology

LETTER OF APPROVAL

We certify that we have read this dissertation work and in our opinion it is appreciable for the scope and quality as a dissertation in the partial fulfillment of the requirements of Masters Degree of Science in Computer Science & Information Technology.

Evaluation Committee

Asst. Prof. Nawaraj Paudel
Head of Department
Central Department of Computer Science
& Information Technology
Tribhuvan University
Kirtipur

Prof. Dr. Shashidhar Ram Joshi
Department of Electronics & Computer Engineering,
Institute of Engineering,
Pulchowk, Nepal

(External Examiner)

(Internal Examiner)

Date: 23 Dec, 2015

Acknowledgement

I would never have been able to finish my dissertation without the guidance, support and encouragement of numerous people including my supervisor, my friends, colleagues and support from my family. At the end of my thesis I would like to thank all those people who made this thesis possible and an unforgettable experience for me.

First, I would like to express my gratitude to my supervisor **Professor Dr. Shashidhar Ram Joshi,** Department of Electronics and Computer Engineering, Institute of Engineering, Pulchowk for his continuous support without which the thesis wouldn't have been possible to complete. His suggestions, guidance, thorough knowledge and expertise helped me immensely in understanding and developing this thesis. I thank him immensely for his patience and generous time spent to guide me through the entire process.

Most importantly I would like to thank to respected Head of Department of Central Department of Computer Science and Information Technology, Asst. Prof. Nawaraj Paudel for his kind support, help and constructive suggestions. I am very much grateful and thankful to all the respected teachers Prof. Dr. Subarna Sakya, Dr. Arun Kumar Timilsina, Mr. Min Bahadur Khati, Mr. Dheeraj Kedar Pandey, Mr. Sarbin Sayami, Mrs. Lalita Sthapit, Mr. Arjun Singh Saud, Mr. Bikash Balami and Tej Bahadur Shahi for providing me such a broad knowledge and inspirations.

Special thanks to my family for their endless motivation, constant mental support and love which have been influential in whatever I have achieved so far. All my class fellows are worthy of my gratefulness for their direct or indirect support in completion of my dissertation. Finally, I would like to thank to respected teacher Mr. Jagdish Bhatt for his suggestions during my work.

I have done my best to complete this research work. Suggestions from the readers are always welcomed, which will improve this work.

Abstract

Ensemble methods have been called the most influential development in Data Mining and Machine Learning in the past decade. The idea of ensemble learning is to employ multiple learners and combine their predictions.

In this thesis, a novel method is proposed to build an ensemble of classifiers based on feature selection: Random selection, Relief and feature extraction: Principal component analysis method. The feature selection process chooses optimal subset of features according to objective function whereas feature extraction process maps the high dimensional dataset into lower dimensional dataset using the linear combination of original features. These feature selection and extraction method helps to produce diverse as well as accurate set of ensemble classifiers. A comparison of proposed method is made with the Bagging, AdaBoost, feature selection based NN, feature extraction based NN and also with plain NN using 22 benchmark dataset. The result obtained by the proposed method outperformed other algorithms with the following distribution: NN (14 cases), Random-NN (13 cases), Relief-NN (15 cases), PCA-NN (19 cases), AdaBoost (14 cases), Bagging (15 cases).

Keywords: Ensemble methods, feature selection, feature extraction, Relief, Principal component analysis, AdaBoost, Bagging, NN, Random-NN, Relief-NN, PCA-NN

Table of Contents

Contents **Page No.**

List of Figures

List of Tables

List of Abbreviations

Abbreviations	Full Form
NN	Nearest Neighbor
PCA	Principal Component Analysis
EOF	Empirical orthogonal function
PC	Principal component
CFS	Correlation feature selection
MDL	Minimum Description Length
HFEE	Hybrid feature selection & Extraction Ensemble
WEKA	Waikato Environment for Knowledge Analysis
SDK	Software development kit
IDE	Integrated development Environment
SWT	Standard Widget toolkit