SRI LANKA
LIBRARY
INSTITUTE OF INFORMATION TECHNOLOGY

Analysis of Human Interpretability in Document Classification

P.K. Suriyaa Kumari

MS17903820

A THESIS
SUBMITTED TO
SRI LANKA INSTITUTE OF INFORMATION TECHNOLOGY
IN PARTIAL FULLFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

December 2018

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

2019/01/22

Mr. Prasanna S. Haddela

Approved for MSc. Research Project:

2019/1/22

MSc. Research Project Coordinator, SLIIT

Approved for MSc:

2019/1/02

MSc. Programme Coordinator, SLIIT

Declaration of originality

This is to certify that the work is entirely my own and not of any other person, unless explicitly acknowledged (including citation of published and unpublished sources). The work has not previously been submitted in any form to the Sri Lanka Institute of Information Technology or to any other institution for assessment for any other purpose.

Signed _____

Date _____

Acknowledgement

I would like to express my sincere gratitude to everyone who supported and guided me throughout this year to complete my research project. Without their help and guidance, I would not have been able to gain a successful research experience.

First, I would like to thank my supervisor Mr. Prasanna S. Haddela, for the guidance, he provided from very beginning to the end of my research. Despite his busy work schedule, he was always supportive and guided me on the correct path. His feedback, comments and insight provided a tremendous support during the research work.

I would also thank Mr. Samantha Rajapaksha, the M.Sc. Research coordinator for providing guidance and support and Prof.Samanatha Thelijjaoda for providing required knowledge about research concepts. Furthermore, I would like to thank all other lectures and staff members of the Department of Computing for the support they have given me in numerous ways.

I am grateful to my family for their support and love in many ways. Specially, I would like to convey my heartiest gratitude to my husband staying beside me an encouraging me to carry out this work, my parents who always supported me and provided me with a good education.

Finally, I would like to thank all my colleagues who assisted me throughout this study and it was a wonderful experience to be with them.

Table of Content

Abstract iv

Acknowledgement v

List of Figures ix

List of Tables xi

List of Equations xii

Chapter 1 1

1. Introduction 1

1.1 Study Background 1

1.2 Problem Definition 3

1.3 Significance of Study 4

1.4 Scope and Limitations 5

1.5 Research Questions 5

1.6 Research Aim and Objectives 6

Research Main Objectives 7

1.7 Dissertation Structure 8

Chapter 2 9

Literature Review 9

2.1 Introduction 9

2.1.1 Interpreting Predictions of a Classifier 9

2.1.2 Text Mining and Classification 10

2.1.2 Application Areas of Text Mining and Classification 11

2.2 Literature Relevant to Text Classification 11

2.2.1 Research's on Sinhala Language Document Classification 14

2.3 Data Pre-Processing 15

2.3.1 What is a feature vector? 16

2.3.1 Uses of Feature Vector 16

2.3.2 Feature identification Methods 17

Term Frequency (tf) 17

Inverse document frequency 17

Term frequency–Inverse document frequency 18

2.4 Text Classifiers 19

2.4.1 Naïve Bayes Classifier 19
2.4.2 Support Vector Machine 20
Advantages of SVM 20
2.4.3 K-nearest neighbor Classifier 20
Advantages of kNN 21
2.4.4 Radom Forest Classifier 21
Advantages of Random Forest Classification Algorithm 22
2.5 Text Classification Tools 22
2.5.1 R Studio 22
2.5.2 Weka (Waikato Environment for Knowledge Analysis) 23
2.5.3 Anaconda Jupyter Note Book 23
2.6 Performance Evaluation Matrices 24
2.6.1 Accuracy, Precision, Recall and F1-Measure 24
Accuracy 24
Precision, Recall and F1 Measure 24
Chapter 3 26
Application of Machine Learning Classifiers in Text Classification 26
3.1 Introduction 26
3.1.1 Data Used for Research Work 27
3.1.2 Software Tools 27
3.2 Experimental Method used for Text Classification 28
3.2.1 Data Preprocessing 29
3.2.2 Application of Classification Algorithms 33
Split Data into Training and Test Data Sets 35
Split Data into Training and Test Data Sets using WEKA Tool 36
Application of Machine Learning Classifiers 37
Working with Support Vector Machine Classification Algorithm 39
Working with Naïve Bayes Classification Algorithm 43
Working with k-nearset neighbour (kNN) Classification Algorithm 47
Application of kNN in WEKA Tool 48
Working with Random Forest Classification Algorithm 51
Application of Random Forest classifier in WEKA Tool 52
3.4 Evaluating Human Interpretability of Machine Learning Classifiers 57

3.4.1 Application of Prediction Interpretability libraries in Classification Models 58
3.4.2 Application of Lime library in text classification model with python 60
Application of lime library into Random Forest Classifier 63
Application of lime into DecisionTree Classifier 66
Application of Lime into kNN Classifier 68
Application of Lime into Support Vector Classifier 69
Chapter 4 71
Results and Evaluation 71
4.1 Introduction 71
4.2 Experimental Result Analysis 72
Choosing Best Classifier for Text Classification 72
Discuss about Evaluating Human Interpretability of Text Classifiers 74
Evaluating Machine Learning Predictions with Random Forest Classifier 75
Evaluating Machine Learning Predictions with SVM and kNN Classifier 77
Chapter 5 80
Research Discussion 80
5.1 Introduction 80
5.2 Limitations 80
5.3 Conclusion 81
5.4 Future Works 82
References 83

List of Figures

Figure 2.1 Representation of feature vector in matrix form 16
Figure 2. 2: Visualization of ensemble Random Forest Classifier 21
Figure 3.1: Conceptual Framework of Text Classification prediction interpretation Process 39
Figure 3.2: Selected Business-Related Terminologies 41
Figure 3.3 : Selected Crime Related Terminologies 30
Figure 3.4: Selected Politics Related Terminologies 31
Figure 3.5: Selected Religion Related Terminologies 31
Figure 3.6:Selected Sports Related Terminologies 32
Figure 3.7:Generate Term Frequency Vector for sample news article 32
Figure 3.8:Comma Separated Vector(CSV) File including all term frequency vectors for selected set of files 33
Figure 3.9:Converted ARFF File use of WEKA Tool
34
Figure 3.10: Load ARFF file into WEKA Tool 34
Figure 3.11:Split Data Set into Training Set and Test Data Set 35
Figure 3.12:Machine Learning Training Process 35
Figure 3.13:Use Resampling Filter to Split Data Set into Training Set 36
Figure 3.14:Use Resampling Filter to Split Data Set into Test Data Set 37
Figure 3.15:News Corpus Classification by Learning Model 38
Figure 3.16:Representation of term-frequency counter vector 39
Figure 3.17:Separating Sample Vectors into n-dimensional space 40
Figure 3.18:Application of Support Vector Machine Classifier into Training Data Set 41
Figure 3.19:Support Vector Classifier Prediction Measures 41
Figure 3.20:Configuring Classifier into Test Data Set 42
Figure 3.21:Support Vector Machine Classifier Prediction Measures for Test Data Set 43
Figure 3.22: Naïve Bayes Classifier Class Separation Marginal Curve Representation for Test Data Set 56
Figure 3.23: Naïve Bayes Classifier Class Separation Marginal Curve Representation for Training Data Set 58
Figure 3.24: Prediction Measures of Support Vector Classifier for Test Data Set
59
Figure 3.25: Naïve Bayes Classifier Class Separation Marginal Curve Representation for Test Data Set 60
Figure 3.26: kNN Data Point Separation 61

Figure 3.27: Setting K-Value for kNN Classifier 61
Figure 3.28 : Prediction Measures of kNN Classifier for Training Data Set 62
Figure 3.29 : News Class Separation for kNN Classifier 63
Figure 3.30 : Prediction Measures of kNN Classifier for Test Data Set 64
Figure 3.31: Ensemble Approach of Random Forest Classifier 65
Figure 3.32: Classification Measures of Random Forest Classifier for Training Data Set 66
Figure 3.32: Classification Measures of Random Forest Classifier for Training Data Set 67
Figure 3.33: Margin Curve generated for Training Data Set by WEKA Random Forest Classifier 68
Figure 3.34: Prediction Measures for Test Data Set by Random Forest Classifier 69
Figure 3.35: Margin Curve generated for Test Data Set by WEKA Random Forest Classifier 70
Figure 3.36: Handling Machine Learning Prediction Interpretation 74
Figure 3.37: Lime Library Prediction Interpretations 75
Figure 3.38: Lime Library Learning Model Interpretation 76
Figure 4.1: Prediction Accuracy Variation among set of Algorithms 91
Figure 4.2: Algorithm Evaluation Metrics 91

List of Tables

Table 3.1:Categories of News Sources 31

Table 3.2:Software Tools & Techniques 32

Table 3.3: Classification Measures of Support Vector Classifiers for Training and Test Data Set 71

Table 3.4: Classification Measures of Naïve Bayes Classifiers for Training and Test Data Set 71

Table 3.5: Classification Measures of kNN Classifiers for Training and Test Data Set 72

Table 3.6: Classification Measures of Random Forest Classifiers for Training and Test Data Set 72

Table 3.7: Tagging Class Names into Integer Number 78

Table 3.1: Experimented Classification Algorithms with WEKA 88

Table 4.2: Classification Algorithms Measures for Training Data Set 89

Table 4.3: Classification Algorithms Measures for Test Data Set 90

List of Equations

Equation 1: Term Frequency 30
Equation 2: Inverse document frequency 30
Equation 3: Term frequency–Inverse document frequency 31
Equation 4: Naïve Bayes Classifier Probability Equation 32
Equation 5: Precision 38
Equation 6: Recall 38
Equation 7: F-measure 38

Abstract

Analysis of Human Interpretability in Document Classification

P.K.Suriyaa Kumari
MSc. in Information Technology
Supervisor: Mr. Prasanna S. Haddela
December 2018

With high use of computers, the collection of textual data generated, exchanged, stored and accessed increased in massive amount and became one of the richest sources of data for the organization. As a result, people are tending to use natural language processing application to categorize this large volume of data efficient and accurate manner. Their application of machine learning models. When it comes to Natural Language processing (NLP) applications where most of them follows supervised learning techniques, automatic document classification models developed to do content-based assignment where the materials are assigned into predefined categories. This makes it easier to find the relevant information at the right time and for filtering and routing documents directly to correct users.

Mostly these learning models are operating in black-box manner where there is no way to interpret how the model has decided which class an instance should assigned. understanding the reason behind how learning makes these predictions are very important to trust such learning models in real application. This thesis presents the work related to the experimental work been carried with set of text classifiers to interpret text classifiers predictions, so any classifier can be evaluated based on how well they support classification purpose.