A TIME-SERIES PRE-PROCESSING METHODOLOGY WITH STATISTICAL AND SPECTRAL ANALYSIS FOR VOICE CLASSIFICATION

by

Lan Kun

Master of Science in E-Commerce Technology

2013



Department of Computer and Information Science

Faculty of Science and Technology

University of Macau

A TIME-SERIES PRE-PROCESSING METHODOLOGY WITH STATISTICAL AND SPECTRAL ANALYSIS FOR VOICE CLASSIFICATION

by

Lan Kun

A thesis submitted in partial fulfillment of the

requirements for the degree of

Master of Science in E-Commerce Technology

Department of Computer and Information Science

Faculty of Science and Technology

University of Macau

2013

Approved by _____

Supervisor

Date _____

In presenting this thesis in partial fulfillment of the requirements for a Master's degree at the University of Macau, I agree that the Library and the Faculty of Science and Technology shall make its copies freely available for inspection. However, reproduction of this thesis for any purposes or by any means shall not be allowed without my written permission. Authorization is sought by contacting the author at

Address: ROOM NLG201, Choi Kai Yau Building (N), University of Macau, Av. Padre Tomás Pereira Taipa, Macau, China

Telephone: 00853-62970832

E-mail: a624266080@126.com

Signature _____

Date _____

University of Macau

ABSTRACT

A TIME-SERIES PRE-PROCESSING METHODOLOGY WITH STATISTICAL AND SPECTRAL ANALYSIS FOR VOICE CLASSIFICATION

by

Lan Kun

Thesis Supervisor: Dr. Simon Fong

Voice biometrics is one kind of physical characteristics that differs from each individual. Due to this uniqueness, voice classification is found useful in classifying speakers' gender, mother tongue, ethnicity, emotion states, identity verification, verbal command control, and so forth.

In this study, we propose a pre-processing methodology named Statistical Feature Extraction (SFX) since we want to facilitate voice classification through Data Mining Methodology. Using SFX we can faithfully remodel statistical characteristics of the time series voice data via a sequence of piecewise transform functions. We focus on the comparison of effects of various popular data mining algorithms on multiple datasets. And the new methodology is tested through simulation experiments over four representative types of human voice data, namely Female and Male, Emotional Speech, Speaker Identification and Language Recognition.

The experiments yield encouraging results supporting the fact that SFX indeed produces better performance in voice classification than traditional signal processing techniques like Wavelets and LPC-to-CC.

KEY WORDS

Voice classification, Time series, Feature extraction, Pre-processing, Data mining

TABLE OF CONTENTS

ABSTRACT....4

TABLE OF CONTENTS....5

LIST OF FIGURES....7

LIST OF TABLES....10

LIST OF ABBREVIATIONS....11

ACKNOWLEDGEMENTS....13

CHAPTER 1: INTRODUCTION....14

CHAPTER 2: RELATED WORK....18

2.1 Feature Extraction on Voice Data....18

2.2 Feature Extraction on Non-Voice Data....22

2.3 Classification Algorithms Review....23

CHAPTER 3: PROPOSED TIME-SERIES PRE-PROCESSING METHODOLOGY....24

3.1 Feature Extraction from the Frequency Domain....27

3.1.1 Linear Prediction Coefficients to Cepstral Coefficients....27

3.2 Feature Extraction from the Time Domain....30

3.2.1 Descriptive Statistics....30

3.2.2 Dynamic Time Warping Distance....33

3.2.3 Piecewise Transformation....35

CHAPTER 4: EXPERIMENT....39

4.1 Data Description....39

4.1.1 Data Source....39

4.1.2 Data Format....40

4.1.3 Data Visualization....41

4.2 Methods Comparison....51

4.3 Software Introduction....57

CHAPTER 5: RESULT AND ANALYSIS 58

5.1 Accuracy Comparison of Datasets 58

5.2 Accuracy Comparison of Pre-Processing Methods 62

5.3 Detailed Accuracy Comparison 64

5.3.1 Kappa Statistic 64
5.3.2 Precision 67
5.3.3 Recall 70
5.3.4 F-measure 73
5.3.5 ROC 76

5.4 Overall Averaged Performance Comparison 81

CHAPTER 6: CONCLUSION 83

6.1 Conclusion 83

6.2 Future Work 84

APPENDIX 85

BIBLIOGRAPHY 93

UNIVERSIDADE DE MACAU
仁義禮知信
澳門大學

LIST OF FIGURES

Figure 3.1 Proposed pre-processing methodology as a part of the classification model learning process....24
Figure 3.2 The detailed illustration about SFX without FS....25
Figure 3.3 The detailed illustration about SFX with Ensemble FS....25
Figure 3.4 A sample time series voice data represented in LPC coefficients....28
Figure 3.5 Cepstral Coefficients computation steps....29
Figure 3.6 Quartile....31
Figure 3.7 DTW calculation illustration....34
Figure 3.8 An example of sampled time series voice data and its partition....36
Figure 3.9 The amplified view of piecewise linear regression (partly)....36
Figure 3.10 Calibration Curve for segmentation selection on FM....38

Figure 4.1 Visualization of FM dataset, that belongs to the "Female" group....42
Figure 4.2 Visualization of FM dataset, that belongs to the "Male" group....42
Figure 4.3 Visualization of ES dataset, that belongs to the "Anger" group....42
Figure 4.4 Visualization of ES dataset, that belongs to the "Happiness" group....43
Figure 4.5 Visualization of ES dataset, that belongs to the "Neutral" group....43
Figure 4.6 Visualization of ES dataset, that belongs to the "Sadness" group....43
Figure 4.7 Visualization of SI dataset, that belongs to the "Speaker 1" group....44
Figure 4.8 Visualization of SI dataset, that belongs to the "Speaker 2" group....44
Figure 4.9 Visualization of SI dataset, that belongs to the "Speaker 3" group....44
Figure 4.10 Visualization of SI dataset, that belongs to the "Speaker 4" group....45
Figure 4.11 Visualization of SI dataset, that belongs to the "Speaker 5" group....45
Figure 4.12 Visualization of SI dataset, that belongs to the "Speaker 6" group....45
Figure 4.13 Visualization of SI dataset, that belongs to the "Speaker 7" group....46
Figure 4.14 Visualization of SI dataset, that belongs to the "Speaker 8" group....46
Figure 4.15 Visualization of SI dataset, that belongs to the "Speaker 9" group....46
Figure 4.16 Visualization of SI dataset, that belongs to the "Speaker 10" group....47

Figure 4.17 Visualization of SI dataset, that belongs to the “Speaker 11” group.....47
Figure 4.18 Visualization of SI dataset, that belongs to the “Speaker 12” group 47
Figure 4.19 Visualization of SI dataset, that belongs to the “Speaker 13” group 48
Figure 4.20 Visualization of SI dataset, that belongs to the “Speaker 14” group 48
Figure 4.21 Visualization of SI dataset, that belongs to the “Speaker 15” group 48
Figure 4.22 Visualization of SI dataset, that belongs to the “Speaker 16” group 49
Figure 4.23 Visualization of LR dataset, that belongs to the “Cantonese” group 49
Figure 4.24 Visualization of LR dataset, that belongs to the “English” group.....49
Figure 4.25 Visualization of LR dataset, that belongs to the “Mandarin” group 50
Figure 4.26 MD visulaization of FM Figure 4.27 MD visulaization of ES 50
Figure 4.28 MD visulaization of SI Figure 4.29 MD visulaization of LR 50
Figure 4.30 A snapshot of WSA searching.....54
Figure 4.31 Pseudo code of WSA 54

Figure 5.1 FM boxplot and accuracy table.....59
Figure 5.2 ES boxplot and accuracy table.....59
Figure 5.3 SI boxplot and accuracy table.....59
Figure 5.4 LR boxplot and accuracy table 60
Figure 5.5 Comparison of average accuracy for different voice datasets and different pre-processing methods.....61
Figure 5.6 Accuracy comparison of Wavelet pre-processing method 62
Figure 5.7 Accuracy comparison of LPC-to-CC pre-processing method.....63
Figure 5.8 Accuracy comparison of SFX pre-processing method.....63
Figure 5.9 Accuracy comparison of SFX+FS pre-processing method 63
Figure 5.10 Kappa statistic comparison of Wavelet pre-processing method.....66
Figure 5.11 Kappa statistic comparison of LPC-to-CC pre-processing method 66
Figure 5.12 Kappa statistic comparison of SFX pre-processing method 66
Figure 5.13 Kappa statistic comparison of SFX+FS pre-processing method 66
Figure 5.14 FM precision comparison 68
Figure 5.15 ES precision comparison 68

Figure 5.16 SI precision comparison 69
Figure 5.17 LR precision comparison 69
Figure 5.18 FM recall comparison 71
Figure 5.19 ES recall comparison 71
Figure 5.20 SI recall comparison 72
Figure 5.21 LR recall comparison 72
Figure 5.22 FM F-measure comparison 74
Figure 5.23 ES F-measure comparison 74
Figure 5.24 SI F-measure comparison 75
Figure 5.25 LR F-measure comparison 75
Figure 5.26 FM ROC comparison 77
Figure 5.27 ES ROC comparison 77
Figure 5.28 SI ROC comparison 78
Figure 5.29 LR ROC comparison 78
Figure 5.30 Comparison of average Kappa statistic for different voice datasets and different pre-processing methods 79
Figure 5.31 Comparison of average precision for different voice datasets and different pre-processing methods 79
Figure 5.32 Comparison of average recall for different voice datasets and different pre-processing methods 79
Figure 5.33 Comparison of average F-measure for different voice datasets and different pre-processing methods 80
Figure 5.34 Comparison of average ROC AUC for different voice datasets and different pre-processing methods 80

LIST OF TABLES

Table 2.1 G.Peeters's feature extraction table 19

Table 2.2 G.Peeters's feature extraction table 2 19

Table 2.3 Classification algorithms mostly used on voice classification 23

Table 3.1 An example of the relation of instances, class labels and target values 34

Table 3.2 The piecewise segment statistics feature extraction 37

Table 4.1 Distributions of classes in different datasets 41

Table 4.2 The numbers of attributes associated with datasets and instances for training and testing by various pre-processing methods 41

Table 4.3 List of standard classification algorithms 51

Table 4.4 Optimal FS methods for each dataset 52

Table 5.1 Numbers of various classification algorithms applied on different datasets 58

Table 5.2 Strength of agreement of Kappa statistic 65

Table 5.3 Definitions of precision and recall terms 67

Table 5.4 Overall Averaged Performance Comparison of Pre-processing Methods 81

Table 5.5 Overall Averaged Performance Comparison of Ensemble Feature Selections 81

Table 5.6 Overall Averaged Time Cost Comparison 82

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
AR	Auto Regressive
ARCH	Auto Regressive Conditional Heteroskedasticity
ARMA	Auto Regressive Moving Average
AUC	Area Under the ROC Curve
BF	Best First
CC	Cepstral Coefficient or Cepstrum Coefficient
CFS	Correlation Feature Selection
DFT	Discrete Fourier Transform
DSP	Digital Signal Processing
DT	Decision Tree
DTW	Dynamic Time Warping
ES	Emotional Speech
FFT	Fast Fourier Transform
FM	Female and Male
FPR	False Positive Rate
FS	Feature Selection
FT	Functional Tree
FURIA	Fuzzy Unordered Rule Induction Algorithm
FX	Feature Extraction
GARCH	Generalized Auto Regressive Conditional Heteroskedasticity
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
IDFT	Inverse Discrete Fourier Transform
IQR	Interquartile Range
JRIP/RIPPER	Repeated Incremental Pruning to Produce Error Reduction
LMT	Logistic Model Tree
LPC	Linear Prediction Coding or Linear Prediction Coefficient

LR	Language Recognition
MA	Moving Average
MFCC	Mel Frequency Cepstral Coefficient
MRMR	Minimum Redundancy Maximum Relevance
NB	Naive Bayes
NNGE	Nearest Neighbor Like
PLF	Piecewise Linear Function
PLP	Perceptual Linear Prediction
REP	Reduced Error Pruning
ROC	Receiver Operating Characteristic
RSS	Residual Sum of Squares
SFX	Statistical Feature Extraction
SI	Speaker Identification
SMO	Sequential Minimal Optimization
SVM	Support Vector Machine
TPR	True Positive Rate
WSA	Wolf Search Algorithm

UNIVERSIDADE DE MACAU
仁義禮知信
澳門大學

ACKNOWLEDGEMENTS

I wish to express my deep appreciation to my supervisor, Dr. Simon Fong, for your patience, support, guidance and encouragement in all the time of research and writing of this thesis. I would also like to thank my classmate, Tang Rui, for your nice help on my fulfillment of WSA feature selection method.

I also thank my friends and my family who have provided constant support and encouragement to me through this process.

UNIVERSIDADE DE MACAU
仁義禮知信
澳門大學