View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by UUM Repository

COMPUSOFT, An international journal of advanced computer technology, 7(11), November-2018 (Volume-VII, Issue-XI)

Available online at: https://ijact.in

Date of Submission	18/10/2018
Date of Acceptance	16/11/2018
Date of Publication	30/11/2018
Page numbers	2862-2866 (5 Pages)

This work is licensed under Creative Commons Attribution 4.0 International License.

ISSN:2320-0790

COMPUSOFT
An International Journal of Advanced Computer Technology

DOUBLE-STAGE FEATURES EXTRACTION FOR MALAY VOWEL CLASSIFICATION USING MULTINOMIAL LOGISTIC REGRESSION

S A Mohd Yusof[1], N I Mahat[2]*, H Husni[3], A F Atanda[4]

[1,3,4]School of Computing, College of Arts and Sciences, Universiti Utara Malaysia, Sintok, Kedah, Malaysia
[2]Centre for Testing, Measurement, and Appraisal, School of Quantitative, College of Arts and Sciences, Universiti Utara Malaysia, Sintok, Kedah Malaysia, [2]noridayu@uum.edu.my

Abstract: Automatic speech recognition (ASR) has recorded enormous development in both research and implementation such as voice commands to control electronic appliances, video games, interface to voice dictation, assistive leaving for the elderly, and dialogue systems. Rapid development on ASR can be seen on English language, while duplicating the ASR framework for Malay language is possible, but the work demands for endlessly efforts. One of common tools that is able to classify Malay vowels is Multinomial Logistic Regression (MLR). However, careless on estimating the parameters of MLR may lead to producing biased classifier which inappropriate for future classification. Besides, the used on huge number of features for classification sometimes hinder MLR to perform well. This paper outlines a new idea for estimating the unknown MLR parameters with less number of features using a double-stage features extraction based on MLR (DSFE-MLR). The proposed DSFE-MLR extracted 39-MFCC from speech waveform and constructed an MLR using training set. Next, the MLR output of class membership probabilities were further extracted through MLR and evaluated using test set. Empirical evidence on Malay sample of students shows that the DSFE-MLR recorded the highest accuracy compared to other classifiers. Besides, the method is able to recognize each of five Malay vowels correctly. In general, DSFE-MLR provides increment of accuracy for Malay speech recognition.

Keywords: Malay vowels recognition; multinomial logistic regression; automatic speech recognition; accuracy.

I. INTRODUCTION

Existing work in speech recognition has considered to expand the automatic speech recognition (ASR) to Malay language (Mazenan& Tan, 2014; Tan, Goh, &Khaw, 2012; Rahman, Mohamed, Mustafa & Salim, 2014; Arifin &Tiun, 2013; Azmi, 2016), leading to endlessly efforts to diligently classify the language at high accuracy. In a simple form, speech pattern recognition involves some steps that process an input signal of voice into classes of vowels or consonants. The basis process starts by extracting acoustic feature using speech processing (i.e. Mel Frequency Cepstrum Coding- MFCC) as in Fig. 1 that resulted in a very large feature vector. Often, some feature selection method is employed to resize the size of MFCC so that the new size of features is sufficient on machine learning methods such as neural network, multinomial logistic regression, support vector machine, and others. Finally, the process is evaluated to ensure that it is reliable and valid for future speech classification.

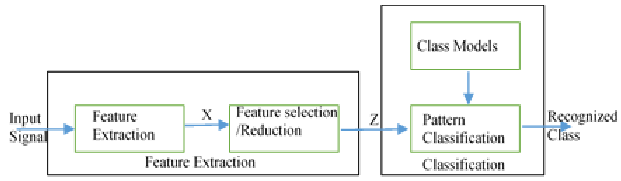

Figure 1: Standard speech recognition process

Similar to English language, the best classification strategy to Malay language is by separating the vowel segments and the consonant segments before the whole speech can be

understood. However, each of these segments demands for extensive research activities to have a complete Malay speech recognition.

The choice of machine learning method is vital as different methods are devoted for different characteristics of data. In this study, multinomial logistic regression (MLR) is used as a vowel classifier. However, MLR may suffers to biased estimators which due to improper design of resampling process. Ignoring such issue could contribute to incorrect future classification.

This paper proposes double stages resampling process in Malay vowel classification to combat biased MLR estimators hence improving MLR accuracy. The key idea of this work is to use different sets of sample for estimating the unknown parameters of MLR, constructing the MLR, and evaluating the constructed MLR by calculating the accuracy rate. The next Section 2 gives brief ideas on previous works that have been devoted to study the Malay speech recognition. Then, Section 3 outlines the methodology that has been executed to combat the weaknesses of MLR through the speech classification system and experimental setting. Section 4 provides evidence to indicate the performance of the proposed method in comparable to other classifiers and the last section summarises all discussion.

II. MALAY VOWELS RECOGNITION

Malay vowels

The interest on Malay vowel classification has recorded several attempts to understand the Malay language. The simplest attempts was studying a single Malay vowel /a/ as corpus (Ting & Zourmand, 2011), others considered five Malay vowels of /a/, /e/, /i/, /o/, and /u/ as corpus (Azmi, Siraj, Yaacob, Paulraj, & Nazri, 2010; Mohd Yusof & Yaacob, 2008; Azmi, Yaacob, & Paulraj, 2009), and six Malay vowels of /a/, /e/, /i/, /o/, /u/ and /ə/ (Azmi, 2016; Azmi, Idayu, Roshidi, Yaakob, & Yaacob, 2012; Lim, Woo, Loh, & Osman, 2000; Mohd Yusof, 2014; Siraj et al., 2009; Ting, Yong, &Mirhassani, 2013; Ting et al., 2011; Ting & Lam, 2009; Ting & Mark, 2008; Ting & Yunus, 2004; Yong & Ting, 2011). The different in the number of Malay vowels can be related to the strategy of segmenting the acoustic in the Malay language and the choice of classifiers or machine learning methods in hand.

Multinomial Logistic Regression

A statistical model MLR is a generalization of logistic regression for classification with multiclass problems. MLR predicts the class membership of vowels by calculating the probability of class membership using linear combination of p observed features where an individual will be classified to the class with the highest computed probability. Given a training set $\mathbf{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)\}$ consists n objects each with a vector of variables $\mathbf{x}_i = \left(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{ip}\right)^T$ for $i = 1,2,\ldots,n$. The values y_i takes the values $1,2,\ldots,g$ are class labels of the vector in the category. Ideally, the classification is in conditional probability model of the form

$$p(y = 1|\boldsymbol{\beta}, \mathbf{x}_i) = \psi(\boldsymbol{\beta}^T\mathbf{x}_i). \qquad \text{Eq. 1}$$

Using the logistic link function, then

$$\psi(\boldsymbol{\beta}^T\mathbf{x}_i) = \frac{\exp(\boldsymbol{\beta}^T\mathbf{x}_i)}{1 + \exp(\boldsymbol{\beta}^T\mathbf{x}_i)} \qquad \text{Eq. 2}$$

Producing a logistic regression for multinomial classes.

The unknown parameters $\boldsymbol{\beta}$can be estimated iteratively by maximizing the likelihood function

$$\prod_{i=1}^{n} \hat{p}(y_i = 1|\boldsymbol{\beta}, \mathbf{x}_i)^{c_i} \left(1 - \hat{p}(y_i = 1|\boldsymbol{\beta}, \mathbf{x}_i)\right)^{c_i}$$

where $c_i = 1,2,\ldots,g$. However, ones need to avoid overfitting the model by using only the training data. This can be done by splitting the classified data set into two mutual sets called training data set and test data set where the former is used to estimate the unknown parameters $\boldsymbol{\beta}$,while the latter is used to evaluate the constructed MLR. However, the quality of estimated parameters cannot be guaranteed if the size of training data is limited (Juang, Hou, & Lee, 1997) hence some studies split the data into training and test sets randomly each for parameters estimation and classifier assessment, and repeat the these processes number of times (Genkin. Lewis, & Madigan, 2007).This basic issue has driven this study to design a new strategy on estimating the parameters of MLR using limited size of training and test sets.

III. METHODOLOGY

Double stage features extraction

The motivation of this work is to obtain a good estimate of MLR parameters **β,** in which the re sampling process was designed to use different sets of training and test data, but still able to compute correct classification, an assessment for the constructed classifier, using the same size of sample to the original data, **D**. In conventional approach, often the 39-MFCC were split randomly into two, called training and test sets, where the former was used for constructing a classifier and the latter was used to assess the constructed classifier. Using this approach, the assessment was based on the size of the test set which was less than the size of the original data.

Alternatively, this study proposed on similar idea to split the original data set, $\mathbf{D}$, into two, but the classification process involves with two features extraction phases:
(i) The first feature extraction was executed when speech signals was detected. Acoustic features from the speech signal were extracted using MFCC which leads to 39-MFCC features.
(ii) The second features extraction was executed when 39-MFCC features were classified by MLR which leads to g number of features that explains behavior of g classes of Malay vowels.
Thus, the classification on Malay vowels was performed on the new data sets that have gone through the double-stage features extraction with full descriptions as in **Error! Reference source not found.**.

Algorithm 1: Double re-sampling for unbiased MLR estimators

Step 1: Extract 39 important information from speech waveform using MFCC. Label the data with extracted 39-MFCC features ($\mathbf{x}$) and class membership (y) as

$\mathbf{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)\}$which consists ofn objects.

Step 2: For each iteration k where $k = 1,2, \ldots, K$

2.1 Split $\mathbf{D}$randomly into 50% training set, $\mathbf{D}_{Tr}$, and 50% test set, $\mathbf{D}_{Te}$.
2.2 Construct MLR using $\mathbf{D}_{Tr}$ and get the estimators of $\boldsymbol{\beta}$label as $\mathbf{b}$.
2.3 Predict the probability of class membership of objects in $\mathbf{D}_{Te}$, label as $\hat{\mathbf{y}}_i = \{\hat{\mathbf{y}}_{i1}, \hat{\mathbf{y}}_{i2}, \ldots, \hat{\mathbf{y}}_{ig}\}$ where g is the number of classes.
2.4 Construct MLR using
$\mathbf{D}^*_{Te} = \{(\hat{\mathbf{y}}_1, y_1), (\hat{\mathbf{y}}_2, y_2), \ldots, (\hat{\mathbf{y}}_{n_{Te}}, y_{n_{Te}})\}$
where $\mathbf{D}^*_{Te}$ consists of the new g features of MLR class membership and the original information of object's class membership, y_i.
2.5 Predict the class of objects in $\mathbf{D}^*_{Te}$.
2.6 If the prediction is correct, then $\text{error}_{ki} = 0$, otherwise $\text{error}_{ki} = 1$ for $i = 1,2, \ldots, n$.
2.7 Switch $\mathbf{D}_{Tr}$ and $\mathbf{D}_{Te}$, each now as test set and training set respectively.
2.8 Repeat Steps 2.2 to 2.6 for the switched $\mathbf{D}_{Tr}$ and $\mathbf{D}_{Te}$.
2.9 Repeat Steps 2.1 to 2.8 at K times.
2.10 Compute correct classification as $\text{correct}_k = 1 - \sum_{i=1}^{n} \text{error}_{ki} / n$.

Step 3: Compute the correct classification as $\sum_{k=1}^{K} \text{correct}_k / K$

In **Error! Reference source not found.**, the size of iteration Kmay vary but this study set the value$K = 10$.

The Set up Experiment

The proposed classification method was tested on speech samples of Azmi (2010). The data consist of samples from the three main ethnic group in Malaysia – Chinese, India, and Malay who are students of Universiti Utara Malaysia (UUM) and Universiti Malaysia Perlis (UNIMAP) respectively. Details of the corpus formation are obtainable from Azmi, (2010). For this study, a sampling frequency of 8 kHz is used to record the vowels. A conventional microphone and a laptop were used to capture the vowel utterances in the noise level of around 40dB room environment measured using noise measuring equipment. The microphone is placed just below the lips of the speaker to avoid blowing into the microphone which gave a higher quality recording. The speakers are given ample time of proper explanation on the recording and its purposes. They are also given a few practice sessions uttering the vowels before the actual recording was done for more consistent recording by reducing the speaker's nervousness during the recording session. The vowel words are recorded continuously in each of the session and saved into a .wav file. A total of 1953 vowel samples made up of five vowels of /a/, /e/, /i/, /o/, and /u/ were used resulted in obtaining

39-MFCC feature vectors (39-MFCC FVs) from the speech samples.

IV. RESULTS AND DISCUSSION

Performance of the proposed classification process (termed as DSFE-MLR) was assessed using correct vowel classification rate where a rate that closes to 100% means all objects have been correctly classified to the correct class of vowels which are /a/, /e/, /i/, /o/ and /u/. The assessments were made on overall correct classification rate that represents the number of misclassified objects over the sample size and the correct classification rate for each class of vowels. Table records the correct classification rate for the proposed classification DSFE-MLR, together with other classifiers which were executed in similar number of random splits of training and test sets, but with only one features extraction, 39-MFCC.

Overall, the proposed classifier has shown promising result with the highest rate compared to the other classifiers. Details observation to each Malay vowel also show that DSFE-MLR outperform other classifiers. It is also interesting to highlight that the best two classifiers are DSFE-MLR and MLR, which both are linear based classifier. Linear discriminant analysis comes as the fourth best classifier also records more than 90% correct. Such results indicate that acoustic features could be best classified using a linear based classifier rather than a non-linear classifier.

TABLE 1. CORRECT CLASSIFICATION RATE OF MALAY VOWELS ACCORDING TO CLASSIFIERS

Classifier	***Malay vowels***					Overall
	/a/	***/e/***	***/i/***	***/o/***	***/u/***	
DSFE–MLR	99.46	100.00	100.00	95.99	97.31	98.55
Multinomial Logistic Regression	98.00	97.82	99.71	94.43	92.42	96.48
Linear Discriminant Analysis	97.64	94.06	97.99	90.14	88.17	93.60
k-Nearest Neighbours	88.39	88.40	90.27	86.48	72.69	85.25
Levenberg Marquart	98.29	96.93	99.56	91.94	91.37	95.62

V. CONCLUSION

Double-stage features extraction of MLR provides state-of-the-art speech recognition effectiveness with greater accuracy even for each Malay class of vowel. The proposed classification process reduces the size of features used in the classification and suitable to be considered when the size of speech sample is limited. Results show significant improvement in terms of vowel recognition performance across all vowels with overall classification rate increasing by 2.07% over the controlled Multinomial Logistic Regression rate. The process could be extended to other classifiers, but further extensive on fitting a new classifier into the designed structure need to be further explored.

VI. REFERENCES

[1] A Genkin, D D Lewis, and D Madigan. (2007). Large-scale Bayesian Logistic Regression for text categorization. Technometrics, 49(3), 291-304. doi: 10.1198/004017007000000245

[2] Arifin, N. ', &Tiun, S. (2013). Predicting Malay Prominent Syllable Using Support Vector Machine. Procedia Technology,11, 861-869. doi:10.1016/j.protcy.2013.12.269

[3] Azmi, M. S. (2016). Development of Malay Word Pronunciation Application using Vowel Recognition. International Journal of U- and E- Service, Science and Technology,9(1), 221-234. doi:10.14257/ijunesst.2016.9.1.24

[4] Azmi, M. Y. S., Idayu, M. N., Roshidi, D., Yaakob, A., &Yaacob, S. (2012). Noise Robustness of Spectrum Delta (SpD) Features in Malay Vowel Recognition Computer Applications for Communication, Networking, and Digital Contents (pp. 270-277): Springer.

[5] Azmi, S. (2010). Feature extraction and classification of malay speech vowels. Universiti Malaysia Perlis.

[6] Azmi, S., Siraj, F., Yaacob, S., Paulraj, M., &Nazri, A. (2010). Improved Malay Vowel Feature Extraction Method Based on First and Second Formants. Paper presented at the Computational Intelligence, Modelling and Simulation (CIMSiM), 2010 Second International Conference on.

[7] B H Juang, W Chou, and C H Lee. (1997). Minimum classification error rate methods for speech classification. IEEE Transactions on Speech and Audio Processing, 5(3), 257-265. doi: 10.1109/89.568732

[8] Lim, C. P., Woo, S. C., Loh, A. S., & Osman, R. (2000). Speech recognition using artificial neural networks. Paper presented at the WebInformation Systems Engineering, 2000. Proceedings of the First International Conference on.

[9] Mazenan, M. N., & Tan, T. (2014). Malay Wordlist Modeling for Articulation Disorder Patient by Using Computerized Speech Diagnosis System. Research Journal of Applied Sciences, Engineering and Technology,7(21), 4535-4540. doi:10.19026/rjaset.7.830

[10] Mohd Yusof, S. A. (2014). An improved feature extraction method for Malay vowel recognition based on spectrum delta. International Journal of Software Engineering and Its Applications, 8(1), 413-426.

[11] Mohd Yusof, S. A., Yaacob, S., &Paulraj, M. (2009). Vowel recognition using First Formant Feature.

[12] Mohd Yusof, S., &Yaacob, S. (2008). Classification of Malaysian vowels using formant based features. Journal of ICT, 7, 27-40.

[13] Rahman, F. D., Mohamed, N., Mustafa, M. B., & Salim, S. S. (2014). Automatic speech recognition system for Malay speaking children. 2014 Third ICT International Student Project Conference (ICT-ISPC). doi:10.1109/ict-ispc.2014.6923222

[14] Singh, A., & Deep, A. (2011). Piperine: a bioenhancer. International Journal of Pharmacy Research and Technology, 1(1), 01-05.

[15] Tan, T., Goh, S., &Khaw, Y. (2012). A Malay Dialect Translation and Synthesis System: Proposal and Preliminary System. 2012 International Conference on Asian Language Processing. doi:10.1109/ialp.2012.14

[16] Ting, H. N., & Lam, Y. (2009). Speaker-Independent Malay Vowel Recognition of Children Using Neural Networks. Paper presented at the World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany.

[17] Ting, H. N., & Mark, K. (2008). Speaker-dependent Malay Vowel Recognition for a child with articulation disorder using multi-layer perceptron. Paper presented at the 4th Kuala Lumpur International Conference on Biomedical Engineering 2008.

[18] Ting, H. N., &Yunus, J. (2004). Speaker-independent Malay vowel recognition of children using multi-layer perceptron. Paper presented at the TENCON 2004. 2004 IEEE Region 10 Conference.

[19] Ting, H. N., Chia, S. Y., Hamid, B. A., &Mukari, S. Z.-M. S. (2011). Acoustic characteristics of vowels by normal Malaysian Malay young adults. Journal of Voice, 25(6), e305-e309.

[20] Ting, H., &Zourmand, A. (2011). Gender Identification by Using Fundamental and Formant Frequency for Malay Children. Paper presented at the 5th Kuala Lumpur International Conference on Biomedical Engineering 2011.

[21] J., Yakubu H., Aboiyar T., and Zirra P. B.. "An improved RSA image encryption algorithm using 1-D logistic map." International Journal of Communication and Computer Technologies 6 (2018), 1-6.