Kurdish Dialects and Neighbor Languages Automatic Recognition

Abdulbasit K. Al-Talabani[1], Zrar K. Abdul[2], Azad A. Ameen[2]

[1]Department of Software Engineering, Faculty of Engineering, Koya University,
Daniel Mitterrand Boulevard, Koya KOY45, Kurdistan Region, Iraq

[2]Department of Computer Science, College of Basic Education, Charmo University,
Kurdistan Region, Iraq

Abstract–**Dialect recognition is one of the hottest topics in the speech analysis area. In this study, a system for dialect and language recognition is developed using phonetic and a style-based features. The study suggests a new set of feature using one-dimensional local binary pattern (LBP). The results show that the proposed LBP set of the feature is useful to improve dialect and language recognition accuracy. The acquired data involved in this study are three Kurdish dialects (Sorani, Badini, and Hawrami) with three neighbor languages (Arabic, Persian, and Turkish). The study proposed a new method to interpret the closeness of the Kurdish dialects and their neighbor languages using confusion matrix and a non-metric multi-dimensional visualization technique. The result shows that the Kurdish dialects can be clustered and linearly separated from the neighbor languages.**

Index Terms—**Dialect recognition, Language processing, Speech analysis, Machine learning, Local binary pattern**.

1. Introduction

Dialect is the language variation of a population established based on various real-life conditions (Chen, et al., 2010). Recently, dialect recognition (DR) has become a hot topic for its wide applications in speech recognition and forensic. Adapted speech recognition system needs different tools such as the recognition of the dialect or the accent to normalize the speech samples for speech recognition system. For example, Hirayama, et al. (2015) develop an automatic speech recognition system that accepts a mixture of various kinds of dialects.

There are several challenges in DR research area, such as the collection of speech data, which needs to model the diversity of the studied dialects and/or languages (Diakoloukas, et al., 1997). The conclusions made by the researches on DR are mostly restricted to the available collected data. Consequently, generalizing the developed algorithms starting with the set of the used feature or the classification methods is generally non-convincing. For this reason, some studies focus on using collected data under specific condition which "preserve" the real-life characteristic of the data. A study made by Huang and Hansen (2007) addresses novel advances in unsupervised spontaneous DR in English and Spanish. The problem considers the case where no transcripts are available for training and test data, and speakers are talking spontaneously. In this study, we adopt the use of spontaneous speech signals recorded from show and debate TV programs.

In the literature, some of the studies focus on investigating the nature of dialect speech signals. For example, in Bahari, et al. (2014) a non-negative factor analysis approach is developed for Gaussian mixture model (GMM) weight decomposition and adaptation. Their study show that GMM weights carry less, yet complimentary, information to GMM means for language and DR. In addition, in Patil and Basu, 2009, a new method of machine learning, called modified polynomial networks is proposed for the DR problem in an Indian language. The proposed algorithm for machine learning is interpreted as designing a neural network by viewing it as a curve fitting (approximation) problem in a high-dimensional space with the help of radial-basis functions.

The research of language and DR is widely using template based and/or phonetic based techniques. The template-based DR adopts the use of global parameters of the speech signal regardless the specific characteristics of the available phonemes related to each dialect. This kind of studies has been frequently used as in Choueiter, et al. (2008) which find that a purely acoustic approach based on a combination of heteroscedastic linear discriminant analysis and maximum mutual information training is very effective. However, phonetic-based DR is also adopted and compared with acoustic and token-based DR and also found to be effective as in Diakoloukas, et al. (1997).

Another approach that adopted for DR is phonetic based recognition of dialect. This approach adopts the use of local feature that reflects the presence of various phonemes in each language or dialect. For example, Chen, et al. propose supervised and unsupervised learning algorithms to extract

ARO-The Scientific Journal of Koya University
Volume V, No 1(2017), Article ID: ARO.10167, 4 pages
DOI: 10.14500/aro.10167
Received 27 October 2016; Accepted 04 April 2017
Regular research paper: Published 24 April 2017
Corresponding author's e-mail: abdulbasit.faeq@koyauniversity.org

Copyright © 2017 Abdulbasit K. Faeq, Zrar K. Abdul and Azad A. Ameen. This is an open-access article distributed under the Creative Commons Attribution License.

dialect discriminating phonetic rules and use these rules to adapt biphones to identify dialects. They discovered that dialect discriminating biphones compatible with the linguistic literature while outperforming a baseline monophone system by 7.5% (Chen, et al., 2010). While in Chen, et al. (2011), the authors propose an informative DR system that learns phonetic transformation rules and uses them to identify dialects. A hidden Markov model is used to align reference phones with dialect-specific pronunciations to characterize when and how often substitutions, insertions, and deletions occur.

This study adopts a template-based DR from speech signal using global phonetic based features. It also introduces a new style-based feature (one-dimensional local binary pattern [1DLBP]), which is not used in DR so far. The study used data recorded from three Kurdish dialects (Sorani, Badini, and Hawrami). It also involves Arabic, Persian and Turkish as three neighbor languages to study how independent the Kurdish dialects from those languages, which supposed to have an influence on each other based on cultural and geographical interactions. The study proposes a method to visualize the recognizer confusion between different dialects and languages.

The rest of this paper is structured based on the following sections: In section, two feature extraction procedures are presented, followed by the description of the data used in section three, next to that the methodology is shown in section four, and finally discussion of the result and the conclusion are presented in sections five and six.

II. Feature Extraction

As any pattern recognition process, the DR includes some major steps starting with feature extraction. In DR, mel frequency cypstrum coefficients (MFCC) and linear prediction coefficients (LPC) based features are well known for its capability to model the phonetic characteristic of the speech signal (Choueiter, et al., 2008; Patil and Basu, 2009). In this study, global features (average and standard deviation) of 12 MFCC and 12 LPC on windows of length 30 and 15 ms overlap are computed. However, besides the MFCC and LPC, the study introduces a 1DLBP feature, which model the style of the speech, and investigates its benefit for DR. 1DLBP is adopted in many other applications such as Guo, et al. (2010) and Abdul, et al. (2016).

The 1DLBP operator labels every single value of the vibration signal by considering its neighborhoods and using the value of the center position as a threshold for the neighborhoods. If the neighbor value is less than the center value, the value of the neighbor will turn to 0; otherwise, it turns to 1. A LBP code for a neighborhood is then produced. The decimal value of the LBP binary code presents the local structural knowledge around the fixed value.

The histogram of the 1DLBP signal displays how often these various patterns appear in a given signal. The distribution of the patterns denotes the whole structure of the signal. The 1DLBP operation of a sample value can be defined as:

$$\mathrm{LBP}_P\left(x\left[i\right]\right)=\sum_{r=0}^{\frac{P}{2}-1}\left\{\begin{matrix} f\left[x\left[i+r-p/2\right]-x[i]\right]2^r + \\ s\left[x\left[i+r+1\right]-x\left[i\right]2^{r+\frac{p}{2}}\right] \end{matrix}\right\} \tag{1}$$

Where f is the sign function:

$$f\left(x\right)=\begin{cases}0, x<0\\ 1, x\geq 0\end{cases} \tag{2}$$

And x[i] is the signal and p is the number of considered neighbors. The Sign function f[x] transforms the differences to a P-bit binary code.

In this paper, only eight neighbors are considered (four to the left of the center and four to the right). Equation (1) illustrates how the 1DLBP is evaluated. Hence, the value range of the new signal is between 0 and 255. The obtained signal is discriminated into two parts, uniform and non-uniform number. The uniform number comprises the numbers with fewer than or equal to two transition bits from 1 to 0 or 0 to 1 in their circular bit patterns. The non-uniform numbers have more than two transition bits. For instance, the patterns 11111111 (0 transitions) and 10001111 (2 transitions) are uniform, while the patterns 10101 (4 transitions) and 01010111 (6 transitions) are non-uniform. There are 58 uniform numbers in the range 0–255 and the rest are non-uniform numbers. The histogram is computed such that an independent bin represents each uniform number, while all the non-uniform numbers are represented in one bin. Therefore, the set of features consists of 59 bins, 58 of them for each uniform number and one bin for all non-uniform numbers. These bins are utilized as features of the dialect speech signals. The number of bins in the histogram depends on how many neighbors are considered. Fig. 1 demonstrates a 1DLBP operator for number of neighbors (p=6), with the center sample as given. After processing 1DLBP, the 6-neighbor samples in the example above produce the 100101 codes. The code is then converted to a decimal system number (=37) and substituted in the same index of the center sample.

A. Data Discription

Data acquisition is an important task in any classification process. The data collected in this paper consists of three Kurdish dialects (Sorani, Badini, and Hawrami), and three different languages (Arabic, Persian, and Turkish) recorded from TV broadcasts. For each dialect and individual

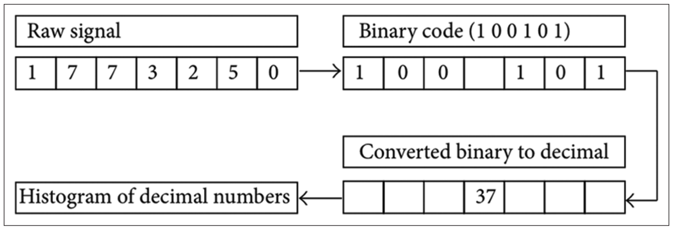
Raw signal
1 7 7 3 2 5 0
Binary code (1 0 0 1 0 1)
1 0 0 1 0 1
Converted binary to decimal
37
Histogram of decimal numbers

Fig. 1. One-dimensional local binary pattern, number of neighbors (p=6).

language, 15 different speakers are involved; each speaker has three different 2 s duration recordings. Consequently, 45 samples are recorded for each dialect and each language. The total length is 6 s for one speaker and 90 s for each individual dialect or language.

III. Methodology

The procedure of DR in this study adopts the use of three sets of features, which are MFCC, LPC, and LBP. Individual feature sets and their fusions at the feature level feed a pairwise based support vector machine classifier, with linear kernel function and sequential minimal optimization optimization method. The protocol used in this study use the whole set of the data and validate them using leave one sample out validation approach. Fusion at the feature level for a couple sets of feature and the whole sets of features is computed. To visualize the relation between the classes, confusion matrix (CM) a no-metric multi-dimensional scaling (NMDS) is adopted. NMDS is an optimization procedure that aims to estimate the non-metric relations between different objects. To show the significance of the improvement, a chi-square test is used and p value is computed for each comparison made between the results.

IV. Result and Discussion

Using the method presented in the past section, experiments are conducted for the three different types of feature (MFCC, LPC, and LBP) and their fusion at the feature level. Table I shows the obtained recognition accuracy for both Kurdish dialects and the involved languages. Based on the phonetic characteristic of the MFCC and LPC features, we can observe how both of these features are similarly contribute in dialect and language recognition. While the pattern regarding the LPB feature, which reflects the style characteristic of the speech signal, is totally different between the Kurdish dialects recognition from one side and the languages recognition from the other sides (76-46%). This could be interpreted by the observation that the dialects of the same language are mostly different in style of the speech, while the languages are phonetically different.

In the other hand, from the fusion-based experiments, it can be clearly observed how the LBP fusion with both MFCC and LPC can significantly improve the recognition accuracy for Kurdish dialects (from 71% to 88.9% with p=5.1E-9, and from 78% to 89.6% with p=0.001, for both MFCC and LPC, respectively) and also for Language recognition (from 67.8% to 73% with p=0.02 and from 71.8% to 81.1% with p=0.002 with both of MFCC and LPC, respectively). This improvement reflects the complementarity characteristic of the LBP feature to the widely used phonetic based features (MFCC and LPC in this study). This complementarity of LBP to both MFCC and LPC is also supported by the non-improved recognition accuracy when MFCC and LPC are fused. The best result obtained for dialect and language recognition obtained by fusing LPC and LBP features.

The second aim of this study is to show how close each Kurdish dialect to the neighbor languages as an attempt to study the influence of the neighbor languages and the Kurdish dialects on each other from a phonetic and style based of view. For this purpose, CM of the accuracy results is used and visualized by an NMDS technique using SPSS software. The CM of the highest result obtained by fusing LPC and LBP and the visualized form using NMDS are shown in Table II and Fig. 2, respectively.

This study suggests to interpret the relations and the influence of different languages and dialects through the CM of the recognition procedure.

From Fig. 2, we can clearly observe that the Kurdish dialects are clustered in the top of the graph such that it can

TABLE I
Recognition accuracy (%) of experiments using various features and their fusions

Feature sets	Kurdish DR accuracy	Languages DR accuracy
MFCC	71	67.8
LPC	78	71.8
LBP	76	46
LBP-MFCC	88.9	73
LBP-LPC	89.6	81.1
LPC-MFCC	74.8	70
ALL	88.2	80

MFCC: Mel frequency cypstrum coefficients, LPC: Linear prediction coefficients, LBP: Local binary pattern, DR: Dialect recognition

TABLE II
CM for the whole involved classes using LBP and LPC features

LBP_LPC	Sorani	Hawrami	Badini	Arabic	Persian	Turkish
Sorani	36	3	5	1	0	0
Hawrami	2	39	0	2	2	0
Badini	2	0	40	3	0	0
Arabic	4	0	2	36	0	3
Persian	1	2	0	0	38	4
Turkish	7	0	0	2	6	30

CM: Confusion matrix, LPC: Linear prediction coefficients, LBP: Local binary pattern

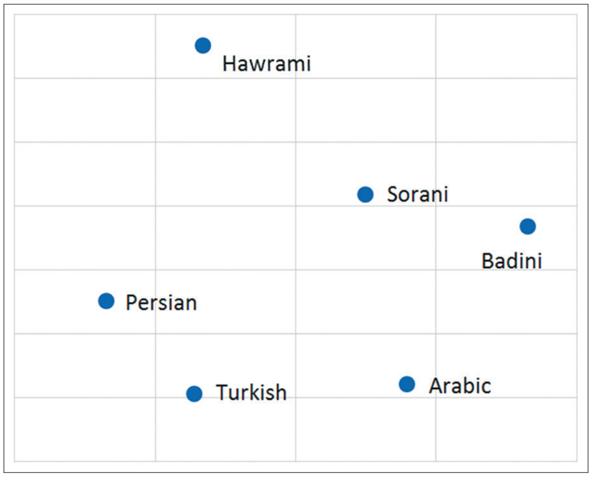

Fig. 2. No-metric multi-dimensional scaling figure for the confusion matrix shown in Table II.

be separated linearly from the involved languages. Another observation is that the Sorani and Badiny dialects are closer to each other than the Hawrami dialect and the nearest language to these two dialects is the Arabic language. While the closest language to the Hawrami dialect is the Persian Language.

V. Conclusion

The result obtained in this study shows that the LBP features for DR are useful especially when fused with phonetic based feature like the LPC. The LBP characterizes the speech style, and therefore it is useful for DR more than language recognition. The first contribution of this study is the use of the LBP set of feature for DR, which has not been used so far. The study also contributes in using NMDS to visualize CM to interpret the relations among different languages for future works. For future work, it might be useful to investigate the fusion for more models at the decision level.

VI. Acknowledgment

Authors would like to thank the Koya University and Charmo University for their support during the preparation of this work.

References

Abdul, Z.K., Al-Talabani, A. and Abdulrahman, A.O., 2016. A new feature extraction technique based on 1D local binary pattern for gear fault detection. *Shock and Vibration*, 2016, pp.6.

Bahari, M.H., Dehak, N., Burget, L., Ali, A.M. and Glass, J., 2014. Non negative factor analysis of gaussian mixture model weight adaptation for language and dialect recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(7), pp.1117-1129.

Chen, N.F., Shen, W. and Campbell, J.P., 2010. A linguistically-informative approach to dialect recognition using dialect-discriminating context-dependent phonetic models. *In: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, pp.5014-5017.

Chen, N.F., Shen, W., Campbell, J.P. and Torres-Carrasquillo, P.A., 2011. Informative dialect recognition using context-dependent pronunciation modeling. *In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp.4396-4399.

Choueiter, G., Zweig, G. and Nguyen, P., 2008. An empirical study of automatic accent classification. *In: 2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, pp.4265-4268.

Diakoloukas, V., Digalakis, V., Neumeyer, L. and Kaja, J., 1997. April. Development of dialect-specific speech recognizers using adaptation methods. *In: Acoustics, Speech, and Signal Processing, 1997. ICASSP-97. 1997 IEEE International Conference.* Vol. 2. IEEE, pp.1455-1458.

Guo, Z., Zhang, L. and Zhang, D., 2010. Rotation invariant texture classification using LBP variance (LBPV) with global matching. *Pattern Recognition*, 43(3), pp.706-719.

Hirayama, N., Yoshino, K., Itoyama, K., Mori, S. and Okuno, H.G., 2015. Automatic speech recognition for mixed dialect utterances by mixing dialect language models. *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* 23(2), pp.373-382.

Huang, R. and Hansen, J.H., 2007. Unsupervised discriminative training with application to dialect classification. *IEEE Transactions on Audio, Speech, and Language Processing,* 15(8), pp.2444-2453.

Patil, H.A. and Basu, T.K., 2009. A novel modified polynomial network design for dialect recognition. *In: Advances in Pattern Recognition, 2009. ICAPR'09. Seventh International Conference.* IEEE, pp.175-178.