LETTER

Chinese Dialect Identification Based on Genetic Algorithm for Discriminative Training of Bigram Model

Wuei-He TSAI† *and* **Wen-Whei CHANG**†, ***Nonmembers***

SUMMARY A minimum classification error formulation based on genetic algorithm is proposed for discriminative training of the bigram language model. Results of Chinese dialect identification were reported which demonstrate performance improvement with use of the genetic algorithm over the generalized probabilistic descent algorithm.

key words: *Chinese dialect identification, minimum classification error, generalized probabilistic descent algorithm, genetic algorithm*

1. Introduction

Research in Chinese language processing was almost exclusively aimed at voice dictation of Mandarin Chinese [1]. However, hundreads of Chinese dialects exist and may be linguistically divided into seven major groups that sound mutually unintelligible. This motivated our research into building a system capable of recognizing three major Chinese dialects spoken in Taiwan, Mandarin, Holo, and Hakka. The baseline system performs phonotactic analysis after speech utterances have been tokenized into sequences of broad phonetic classes (BPCs). Our approach differs from similar approaches [2]–[4] in that BPC bigram probabilities were discriminatively trained to optimize the identification accuracy according to the minimum classification error (MCE) formulation [5]. The standard approach to MCE training is to apply the generalized probabilistic descent (GPD) algorithm [5], [6]. However, GPD approaches perform local searches and may fail to provide reliable results when used on complex optimization problem with multiple local optima. By contrast, the global properties of genetic algorithms have made them very effective at solving constrained as well as combinatorial optimization problems [7]. Recognizing this, we use a genetic algorithm to train the bigram model because the MCE formulation may be considered as searching for a global optimum in the parameter space.

2. System Configuraion

The baseline system we tested consisted of a BPC recognizer followed by phonotactically motivated language models, and included two subsystems as shown in Fig. 1. The first subsystem processed speech utterances to extract acoustic feature vectors, and these feature measurements were used to determine broad phonetic transcriptions of utterances. Five BPCs considered here are the stop, fricative, affricate, nasal, and vowel or diphthong. In our implementation, the BPC recognizer was designed using a 9-state left-to-right HMM with its state observation probability density modeled as a mixture of 15 Gaussian densities. Observations were streams of acoustic feature vectors consisting of the lowest 10 coefficients of the mel-scaled cepstrum. In the training phase, a phonetically labeled subset of the training speech is used to estimate the HMM parameters through the segmental k-means reestimation algorithm. The second subsystsem calculated the log-likelihood, $\log \Pr(W|D_i)$, that the phonetic language model for dialect D_i produced the BPC sequence $W = \{w_1, w_2, \ldots, w_T\}$. Language models designed to capture the phonotactic regularities of each dialect are constructed by running the training corpus into the BPC recognizer and computing bigram probabilities between consecutive symbols observed in the BPC sequence. The phonetic language model for each dialect D_i is characterized by means of bigram probabilities $\lambda_i = \{a_{mn}^{(i)}\}$, that gives the probability that BPC m is followed immediately by BPC n. When an unknown utterance is received, the language model receives it as input the recognized BPC sequence and produces as output the log-likelihood of the dialect being spoken. The dialect of the language model which predicts the utterance with the highest log-likelihood is hypothesized as the dialect of the test utterance.

3. Proposed Training Scheme

The effectiveness of the proposed system crucially depends on how the BPC bigram probabilities are estimated from a training corpus. Traditionally, bigram model training problems have been dealt with using the relative frequency approach [3], [4]. Discrimination can be improved if all the utterances spoken in J different dialects are jointly used in training the parameter set $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_J\}$. This task belongs to the class of discriminative training problems and has been studied by several researchers [5], [6]. To begin, define a discriminant function $g_i(W; \Lambda) = \log \Pr(W|\lambda_i)$ and let $W_n,\ n = 1, 2, \ldots, N$, denote the BPC sequence repre-

Manuscript received March 13, 2000.

†The authors are with the Department of Communication Engineering, National Chiao-Tung University, Hsinchu, Taiwan, R.O.C.

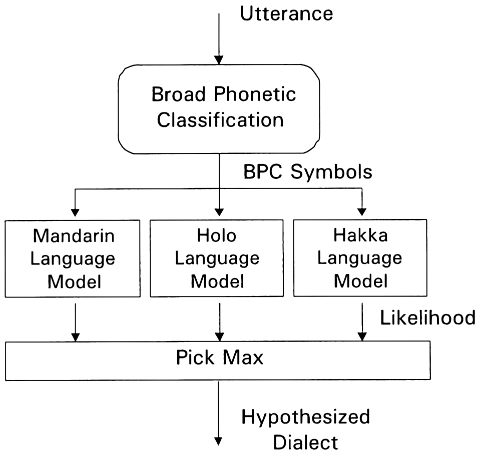

Fig. 1 The proposed dialect identification system.

senting the phonetic transcription of the nth training utterance. The basic idea is to formulate the discriminative training process as a constrained optimination problem that can be stated as follows:

$$\min_{\{\lambda_i\}} L(\Lambda) = \frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{J} \frac{1}{1 + e^{-\gamma d_i(W_n)}} \mathbf{1}(W_n \in D_i), \quad (1)$$

where $\mathbf{1}(\cdot)$ is the indicator function, and the misclassification measure in the form

$$d_i(W_n) = -g_i(W_n;\Lambda) + \log\left\{\frac{1}{J-1}\sum_{k,k\neq i}\exp[g_k(W_n;\Lambda)\eta]\right\}^{1/\eta}. \quad (2)$$

The GPD algorithm is often used to identify optimal parameter setting for which the empirical loss $L(\Lambda)$ is minimal. However, it may suffer from the problem of convergence toward local minima that are critically dependent upon the starting point. An alternative approach to function optimization is based on genetic algorithm (GA). The main attraction of GAs arises from the fact that the given search space is explored in parallel by means of iterative modifications a population of potential solutions (chromosomes). Choosing an appropriate chromosome representation is the first step in applying genetic algorithm to solving optimization problems. Here BPC bigram probabilities define the solution and hence, can be encoded into a chromosome as a list of real numbers, that is: $\mathbf{S} = \{a_{mn}^{(i)}, 1 \leq m, n \leq 5, 1 \leq i \leq J\}$. Starting with an initial population, the fitness measures of all chromosomes were ranked with respect to the objective function $F(\mathbf{S}) = 1/L(\Lambda)$. As a result of this evaluation, a particular group of chromosomes were selected from the population to generate offspring by subsequent recombination. Natural evolution gives chromosomes with a higher fitness level more chance to survive in the next generation. Crossover among the selected chromosomes then proceeded when the value of a random number generated between 0 and 1 is less than the specified probability p_c. After crossover, with a probability of p_m, the mutation operator was applied to each chromosome by inverting its substring between two randomly selected positions. When the maximum number of generations was reached, the best chromosome in the final population was taken as GA's solution for functional optimization. In the GA implementation, the parameter values used for the p_c, p_m, the population size, and the maximum number of generation were empirically determined to be 0.6, 0.1, 200, and 5000, respectively.

Table 1 Results of Chinese dialect identification.

(a) MCE/GPD-trained bigram model.

Actual	Recognition		
	Mandarin	Holo	Hakka
Mandarin	**0.95**	0.02	0.03
Holo	0.00	**0.90**	0.10
Hakka	0.03	0.08	**0.89**

(b) MCE/GA-trained bigram model.

Actual	Recognition		
	Mandarin	Holo	Hakka
Mandarin	**0.95**	0.02	0.03
Holo	0.00	**0.94**	0.06
Hakka	0.02	0.03	**0.95**

4. Experimental Results

Experiments were carried out to investigate the potential advantages of using genetic algorithms to improve the language-discriminating power of MCE-trained BPC bigram models. The training database consisted of 120 sentential utterances, and the database for use in testing consisted of 60 utterances that did not include those used for training. Table 1 presents dialect identification results that compare the case where the bigram language model was discriminatively trained using the GPD algorithm against the case that was trained using the genetic algorithm. The rows of the confusion matrix correspond to the dialects actually being spoken and the columns indicate the dialects identified. As the table shows, Chinese dialects can be identified by computing BPC bigram probabilities that are likely to capture the relevant phonetic and phonotactic aspects of individual dialects. Simulation results also indicate that the proposed system using a BPC bigram model trained by the genetic algorithm with the MCE formulation yields better performance with 94.7% accuracy, compared to 91.3% for the system using the GPD method.

5. Conclusions

This Letter explores the benefits of a genetic algorithm for use in training bigram language models that have the most discriminating power. While we only addressed the Chinese dialect identification problem, the proposed approach can be applied to discriminate among other languages as well.

References

[1] L.S. Lee, "Voice dictation of Mandarin Chinese," IEEE Signal Processing Magazine, vol.14, no.4, pp.63–101, 1997.

[2] A.S. House and E.P. Neuburg, "Toward automatic identification of the language of an utterance. I. Preliminary methodological considerations," J. Acoust. Soc. Am., vol.62, no.3, pp.708–713, 1977.

[3] T.J. Hazen and V.W. Zue, "Segment-based automatic language identification," J. Acoust. Soc. Am., vol.101, no.4, pp.2323–2331, 1997.

[4] M.A. Zissman, "Comparison of four approaches to automatic language identification of telephone speech," IEEE Trans. Speech & Audio Processing, vol.4, no.1, pp.31–44, 1996.

[5] B.H. Juang, W. Chou, and C.H. Lee, "Minimum classification error rate methods for speech recognition," IEEE Trans. Speech & Audio Processing, vol.5, no.3, pp.257–265, 1997.

[6] S. Katagiri, C.H. Lee, and B.H. Juang, "New discriminative algorithms based on the generalized probabilistic descent method," Proc. IEEE-SP Workshop on Neural Network for Signal Processing, Sept. 1991.

[7] D.E. Goldberg, Genetic Algorithm in Search, Optimization and Machine Learning, Addison-Wesley, New York, 1989.