Hindawi Publishing Corporation
EURASIP Journal on Audio, Speech, and Music Processing
Volume 2008, Article ID 674859, 6 pages
doi:10.1155/2008/674859

Research Article

Using SVM as Back-End Classifier for Language Identification

Hongbin Suo, Ming Li, Ping Lu, and Yonghong Yan

ThinkIT Speech Laboratory, 109 DSP Building, No. 21, Bei-Si-Huan-Xi Road, Beijing 100190, China

Correspondence should be addressed to Yonghong Yan, yyan@hccl.ioa.ac.cn

Received 31 January 2008; Accepted 29 September 2008

Recommended by Woon-Seng Gan

Robust automatic language identification (LID) is a task of identifying the language from a short utterance spoken by an unknown speaker. One of the mainstream approaches named parallel phone recognition language modeling (PPRLM) has achieved a very good performance. The log-likelihood radio (LLR) algorithm has been proposed recently to normalize posteriori probabilities which are the outputs of back-end classifiers in PPRLM systems. Support vector machine (SVM) with radial basis function (RBF) kernel is adopted as the back-end classifier. But for the conventional SVM classifier, the output is not probability. We use a pair-wise posterior probability estimation (PPPE) algorithm to calibrate the output of each classifier. The proposed approaches are evaluated on the 2005 National Institute of Standards and Technology (NIST). Language recognition evaluation databases and experiments show that the systems described in this paper produce comparable results to the existing arts.

1. INTRODUCTION

Automatic spoken language identification without using deep knowledge of those languages is a challenging task. The variability of one spoken utterance can be incurred by its content, speakers, and environment. Normally, the training corpus and test corpus consist of unconstrained utterances from different speakers. Therefore, the core issue is how to extract the language differences regardless of content, speaker, and environment information [1, 2]. The clues that human use to identify languages are studied in [3, 4]. The sources of information used to discriminate one language from the others include phonetics, phonology, morphology, syntax, and prosody. At present, The most successful approach to LID uses phone recognizers of several languages in parallel. The analysis in [4] indicates that performance can be considerably improved in proportion to the number of front-end phone recognizers. Recently, a set of phone recognizers are used to transcribe the input speech into phoneme lattices [5, 6] which are later scored by n-gram language models.

Each spoken utterance is converted into a score vector with its components representing the statistics of the acoustic units. Vector space modeling approach [7] has been successfully applied to spoken language identification. Results in an anchor GMM system [8] show that it is able to achieve robust speaker independent language identification through compensation for intralanguage and interspeaker variability. However, the identity of a target language is not sufficiently described by the score vectors that are generated by the following language models in conventional PPRLM systems. To compensate these insufficiencies, it is a natural extension that multiple groups with similar speakers in one language are used to build the multiple target phonotactic language models. For example, the training data set for language modeling can be divided by genders. In our proposed framework, hierarchical clustering (HC) algorithm [9] and K-means clustering algorithm are used together to extract more information from the available training data. Here, generalized likelihood ratio (GLR) distance defined in [10] is chosen as the pair-wise distances between two clusters.

In PPRLM framework, back-end discriminative SVM classifiers are adopted to identify the spoken language. The SVM classifier has demonstrated superior performance over generative language modeling framework in [7, 11, 12]. SVM as a discriminative tool maps input cepstral feature vector into high-dimensional space and then separates classes with maximum margin hyperplane. In addition to its discriminative nature, its training criteria also balance the reduction of errors on the training data and the generalization on the unseen data. This makes it perform well on small dataset and suited for handling high-dimensional problem. In this paper,

a back-end radial basis function (RBF) kernel [13] SVM classifier is used to discriminate target languages based on the probability distribution in the discriminative vector space of language characterization scores. The choice of radial basis function kernel is based on its nonlinear mapping function and requirement of relatively small amount of parameters to tune. Furthermore, the linear kernel is a special case of RBF and the sigmoid kernel behaves like radial basis function for certain parameters [14]. Note that the training data of this back-end SVM classifier comes from development data rather than from the data used for training n-gram language models, and cross-validation is employed to select kernel parameters and prevent over-fitting. For testing, once the discriminative language characterization score vectors of a test utterance are generated, the back-end SVM classifier can estimate the posterior probability of each target language that is used to calibrate final outputs. As mentioned above, pair-wise posterior probability estimation (PPPE) algorithm is used to calibrate the output of each classifier. In fact, the multiclass classification problem refers to assigning each of the observations into one of k classes. As two-class problems are much easier to solve, many authors propose to use two-class classifiers for multiclass classification. PPPE algorithm is a popular multiclass classification method that combines all comparisons for each pair of classes. Furthermore, it focuses on techniques that provide a multiclass probability estimate by combining all pair-wise comparisons [15, 16].

The remainder of this paper is organized as follows. The proposed PPRLM LID frameworks is stretched in Section 2. In Section 3, the proposed three basic classifiers are described. Besides, a score calibration method and a probability estimation algorithm are detailed in this section. In Section 4, a speech corpus used for this study is introduced. Experiments and results of the proposed method are given in Section 5. Finally, some conclusions are given in Section 6.

2. THE PPRLM LID FRAMEWORK

This section mainly introduces our PPRLM LID framework based on language characterization score vectors. The parallel phone recognizer with language modeling system is composed of four parts [17, 18]: feature extractor, language-dependent phone recognizers, score generators, and back-end classifier. The general system architecture for language identification task is given in Figure 1, where PRi and SGi are language-dependent phone recognizers and score generators for language i. Usually, two types of scores can be generated for using as the back-end features: acoustic scores and phonotactic scores. Acoustic scores (likelihood) are generated by a one pass forward-backward decoder. Phonotactic scores are generated by the following language models in score generators. Finally, the score vector that is composed of acoustic and phonotactic scores is sent to back-end classifier for identifying. The back-end system consists of three parts (applied in the listed order): a set of classifiers (equal to the number of target languages), probability estimation, and finally a log-likelihood ratio (LLR) normalization.

In feature extraction, speech data is parameterized every 25 milliseconds with 15 milliseconds overlap between contiguous frames. For each frame, a feature vector with 39 dimensions is calculated as follows: 13 Mel Frequency Perceptual Linear Predictive (MFPLP) [19, 20] coefficients, 13 delta cepstral coefficients, and 13 double delta cepstral coefficients. All the feature vectors are processed by cepstral mean subtraction (CMS) method.

A Mandarin score generator is shown in Figure 2. In the framework, training set of each target language is divided into multiple groups that are used to build corresponding language models. The language model subgroups are modeling based on the multiple training subsets. Thus, the dimension of score vector is increased. The total number of language models is $N_{\text{total}} = L \times N$, where L is the number of target languages and N is the number of target subgroups. So, taking no count of the acoustic scores, the dimension of discriminative language characterization score vectors (DLCSVs) in the PPRLM system is $N_{\text{DLCSV}} = L \times N \times P$, where P is denoted as the number of phone recognizers in parallel. Considering the amount of training data for language model building, N is limited to a small number. The main object of these measures is to derive the discriminative high-level feature vectors in LID tasks, while restraining the disturbance caused by the variability of speakers or channels in realistic systems. Thus, a discriminative classifier can be built in this score vector space to identify the target language.

3. THE BACK-END CLASSIFIER

The approach for classifying discriminative language characterization score vectors in LID system is demonstrated in this section. Three classifiers: Gaussian models (GMs), support vector machine (SVM), and feed-forward neural network (NN) are proposed to compartmentalize these high-level features, which are generated by n-gram language model scoring and parallel phone decoding. The architecture of different classifier is given in Figure 3, where Ci is the classifier i. Each of them estimates the posterior probability of target language, which is used to normalize final outputs with LLR method.

3.1. Gaussian mixture model

A Gaussian mixture model (GMM) is constructed by multiple K Gaussian components:

$$P(\vec{x} \mid \lambda) = \sum_{k=1}^{K} \omega_k b_k(\vec{x} \mid \vec{\mu}_k, \Sigma_k), \quad \sum_{k=1}^{K} \omega_k = 1, \tag{1}$$

where $\vec{x}$ is D-dimensional feature vector, λ: $\{\omega_k, \vec{\mu}_k, \Sigma_k\}$ are the parameters of the GMM, and ω_k is weight of an individual Gaussian component. $b_k(\cdot)$ is the individual Gaussian component defined in formula (2):

$$b_k(\cdot) = \frac{1}{(2\pi)^{D/2}|\Sigma_k|^{1/2}} \exp\left\{\frac{1}{2}(\vec{x} - \vec{\mu}_k)^{\tau} \Sigma_k^{-1} (\vec{x} - \vec{\mu}_k)\right\}. \tag{2}$$

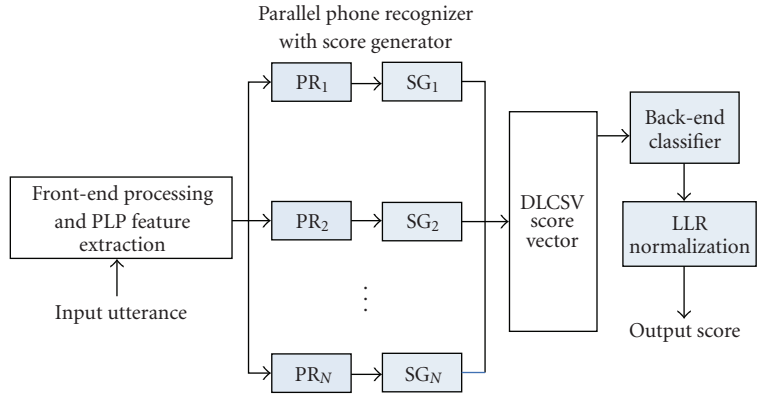

Figure 1: Structure of the proposed PPRLM system.

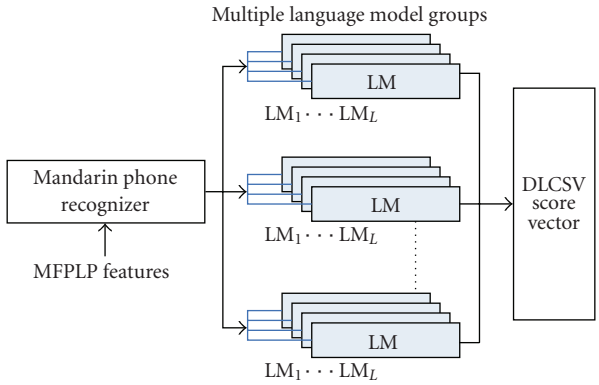

Figure 2: Structure of the mandarin score generator.

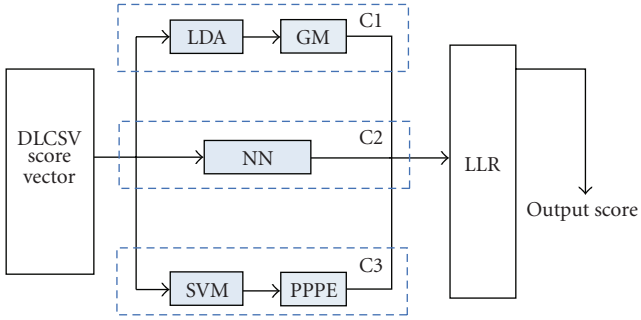

Figure 3: Types of the back-end classifiers.

The back-end procedure takes discriminative language characterization scores from all available classifiers and maps them to final target language post probabilities. Diagonal covariance Gaussian models that are used as the back-end classifiers are trained from the development data [21]. However, these models are hard to describe distribution of high-dimensional features. Usually, linear discriminant analysis (LDA) has been used for this task. As a last step in the back-end procedure, the score vectors are converted to log-likelihood ratios.

3.2. Feed-forward neural network

For feed-forward multilayer neural network training, many algorithms are based on the gradient descent algorithms, such as back propagation (BP). However, These algorithms usually have a poor convergence rate, because the gradient descent methods is using a linear function to approximate an object function. Conjugate gradient (CG), as a second derivative optimal method, has a better convergence rate than BP.

In this paper, the feed-forward neural network with one hidden layer is used to learn the relations in DLCSV space. The classifier is built on both training set and development set with CG optimization [22]. Sigmoid function is chosen as the output function in the NN classifier. Suppose $\vec{S} = [S_1, S_2, \ldots, S_L]^t$ is observation output vector by NN classifier. Moreover, S_k is subject to the constraint in 3, which can be taken as posterior probability. Thus, LLR normalization method detailed in the following section can also be used:

$$\sum_{k=1}^{L} S_k = 1, \quad 0 \le S_k \le 1. \tag{3}$$

3.3. RBF support vector machine

An SVM is a two-class classifier constructed from the sum of a kernel function $K(\cdot, \cdot)$:

$$f(x) = \sum_{i=1}^{n} \alpha_i t_i K(x, x_i) + d,$$
$$\text{subject to } \alpha_i > 0, \quad \sum_{i=1}^{n} \alpha_i t_i = 0, \tag{4}$$

where n is the number of support vectors, t_i is the ideal outputs, and α_i is the weight for the support vectors x_i. A back-end radial basis function (RBF) [13] kernel is used to discriminate target languages. RBF kernel is defined as follows:

$$K(x_i, x_j) = \exp\left(-\gamma ||x_i - x_j||^2\right), \quad \gamma > 0, \tag{5}$$

where γ is the kernel parameter estimated from the training data.

3.4. Score calibration

The topic of calibrating confidence scores in the field of multiple-hypothesis language recognition has been studied in [23], and a detailed analysis of the information flow and the amount of information delivered to users through a language recognition system has been performed. The posterior probability of each of the M hypotheses is estimated and a maximum-a-posteriori (MAP) decision is made. In [21], log-likelihood ratio (LLR) normalization that has been proved to be useful is adopted as a simple back-end process. In the normalization, suppose $\vec{S} = [S_1, S_2, \ldots, S_L]^t$ is the vector of L relative log-likelihoods from L target languages for a particular message, then the posterior probabilities for original hypotheses can be denoted as

$$P_i = \frac{\pi_i e^{S_i}}{\left(\sum_{j=1}^{L} \pi_j e^{S_j}\right)}, \quad i = 1, 2, \ldots, L, \tag{6}$$

where $[\pi_1, \ldots, \pi_L]$ denotes the prior. Considering a flat prior, new log-likelihood ratio normalized score S_i' is denoted as

$$S_i' = S_i - \log\left(\frac{1}{M-1}\sum_{j \neq i} e^{S_j}\right). \tag{7}$$

However, the output scores of back-end RBF SVM are not log-likelihood values; thus, linear discriminant analysis (LDA) and diagonal covariance Gaussian models are used to calculate the log-likelihoods for each target language [24], and improvement has been achieved in detection performance [21].

In this paper, we proposed an alternative approach [14] to estimating the posterior probabilities. Given L classes of data, the goal is to estimate $p_i = p(y = i \mid x), \quad i = 1, \ldots, L$. In a pair-wise framework, firstly, pair-wise class probabilities are estimated as

$$r_{ij} \approx \frac{p(y = i \mid y = i \text{ or } j, x) \approx 1}{\left(1 + e^{A\hat{f}+B}\right)}, \tag{8}$$

where A and B are estimated by minimizing the negative log-likelihood function using known training data and their decision values $\hat{f}$. Then, the posterior probability p_i can be obtained by optimizing the following:

$$\begin{aligned} &\min \frac{1}{2}\sum_{i=1}^{L}\sum_{j \neq i}\left(r_{ji}p_j - r_{ij}p_i\right)^2, \\ &\text{subject to } \sum_{i=1}^{L} p_i = 1, \quad p_i > 0. \end{aligned} \tag{9}$$

Therefore, the estimated posterior probabilities are applicable to performance evaluation. The probability tools of LIBSVM [13] are used in our approach. Experiments in Section 5 show that this multiclass pair-wise posterior probability estimation algorithm is superior to commonly-used log-likelihood ratio normalization method.

4. SPEECH CORPUS

In phone recognizer framework, the Oregon Graduate Institute Multi-Language Telephone Speech (OGI-TS) Corpus [25] is used. It contains 90 speech messages in each of the following 11 languages: English, Farsi, French, German, Hindi, Japanese, Korean, Mandarin, Spanish, Tamil, and Vietnamese. Each message is spoken by a unique speaker and comprises responses to 10 prompts. Besides, phonetically transcribed training data is available for six of the OGI Languages (English, German, Hindi, Japanese, Mandarin, and Spanish). Otherwise, the labeled Hong Kong University of Science and Technology (HKUST) Mandarin Telephone Speech Part 1 [26] is used to accurately train an acoustic model for another Mandarin phone recognizer. A telephone speech database in common use for back-end language modeling is the Linguistic Data Consortium's CallFriend corpus. The corpus comprises two-speaker, unprompted, and conversational speech messages between friends. Hundred North-American long-distance telephone conversations are recorded in each of twelve languages (the same as 11 languages as OGI-TS plus Arabic). There are three sets in this corpus including training, development, and test set, each set consists of 20 two-sided conversations from each language, approximately 30-minute long.

In this paper, experiments are performed on the 2005 NIST LRE [27] 30 s test set. Comparing to the last evaluation, the account of test utterances is rapidly increased. Martin has summarized the numbers of utterances in each language from the primary evaluation data used in this task [28]. Note that in addition to the seven target languages, NIST also collected some conversations in German that are used as evaluation test utterances, though the trials involving these are not considered part of the primary evaluation condition. Moreover, development data which can be used to tune the parameters of back-end classifiers is obtained from the 2003 NIST LRE evaluation sets. Thus, the data comprises 80 development segments, for each of the 7 target languages as given in [28]. All of the training, development and evaluation data is in standard 8-bit 8 kHz mu-law fromat from digital telephone channel.

5. EXPERIMENTS AND RESULTS

The performance of a detection system is characterized by its miss and false alarm probabilities. The primary evaluation metric is based upon 2005 NIST language recognition evaluation [27]. The task of this evaluation is to detect the presence of a hypothesized target language, given a segment of conversational speech over the telephone. Submitted scores are given in the form of equal error rates (EER). EER is the point where miss probability and false alarm probability are equal. Experiments of the proposed application are explained in the following sections.

5.1. Performance of proposed systems

A Mandarin phone recognizer is built from HKUST Telephone data in a PRLM system. There are 68 mono-phones

Table 1: PRLM systems results on 2005 NIST 30-second tasks.

PRLM system	1	2	3	4	5
DLCSV12	√	√			
DLCSV24			√	√	√
NN			√		
LDA + GM				√	
SVM + PPPE					√
LLR		√	√	√	√
EER(%)	17.8	15.4	13.7	13.6	12.8

Table 2: PPRLM systems results on 2005 NIST 30-second tasks.

PPRLM system	1	2	3	4	5
DLCSV72	√	√	√		
DLCSV144				√	√
NN	√				
LDA + GM		√		√	
SVM + PPPE			√		√
LLR	√	√	√	√	√
EER(%)	7.2	6.3	6.4	5.9	5.7

and a three-state left-to-right hidden Markov model (HMM) is used for each tri-phone in each language. Thus, the acoustic model can be described in more detail. But, PPRLM system is mainly composed of six phone recognizers. Acoustic model for each phone recognizer is initialized on OGI-TS corpus and retrained on CallFriend training set corpus. Since the amount of labeled data is limited, monophone is chosen as the acoustic modeling unit. The outputs of all recognizers are phone sequences that are used to build the following 3-gram phone language models. Phonotactic scores are only composed of DLCSV for classifying.

The equal error rate performances of ten systems with the phone recognizer algorithm are given in Tables 1 and 2. In the main frameworks, the discriminative language characterization score vectors and the following different back-end classifiers are checked with marks. Firstly, the baseline systems are denoted as DLCSV12 and DLCSV72 for no speaker clustering in the phone recognizer framework. Then, the 12-dimensional scores of PRLM-DLCSV12 can be used to identify the target language without any back-end classifiers. Besides, the high-dimensional scores can be generated by multiple language models with subgroups. Considering the amount of training data for language modeling, the target number of subgroups is set to 2 (female and male). Thus, the dimension of the DLCSV is 24 in the PRLM framework and 144 in the PPRLM framework. Secondly, when using SVM to be the back-end classifier, the PPPE algorithm is proposed to calibrate output scores. Besides, diagonal-covariance Gaussian model (GM) classifier is also evaluated for comparison. In the mean time, a feed-forward neural network (NN) is used as the back-end classifier for another competent system [22]. Finally, LLR method is adopted to normalize the posteriori probabilities generated by each type of classifiers.

Table 3: The computational cost of back-end classifiers.

Systems	Real time (RT)
PPRLM system 1	0.743
PPRLM system 2	0.728
PPRLM system 3	0.716
PPRLM system 5	0.739

The experiment results of phone recognizing systems show that discriminative score vector modeling method improves system performance in most cases. As mentioned above, the main reason is that multiple discriminative classifiers based on hierarchically clustered speaker groups are employed to map the speech utterance into discriminative language characterization score vector space, which not only represents enhanced language information but also compensates for intralanguage and interspeaker variability. Moreover, by using back-end classifiers, this speaker group specific variability can be compensated sufficiently and make system less speaker dependent. Furthermore, as shown in Tables 1 and 2, the proposed SVM classifier with the PPPE method adopted in the improved systems is comparable to the other classifier. Because the output scores of back-end classifiers are not real log-likelihood values, this alternative language score calibration method performs better.

5.2. Computational cost

Compared with conventional systems, computational cost of the proposed algorithm is not visibly improved. The main reasons can be explained as follows. Firstly, the improved back-end SVM classification with the PPPE algorithm requires a low computational cost. Secondly, the increment of computational cost is focused on generating the discriminative language characterization score vectors. Thus, in the PPRLM system, the time cost of language model scoring is much lower than phone recognizing. Table 3 shows the computational cost of the most PPRLM systems in this paper. The evaluations are carried out on a machine with 3.4G Hz Intel Pentium CPU and 1 G Byte memory.

6. CONCLUSIONS

In this paper, we have presented our basic PPRLM system and three classifiers for processing the high-level score features. The progressive use of groups' training data for building 3-gram language models is exploited to map spoken utterance into discriminative language characterization score vector space efficiently. The proposed method enhances language information and compensates the disturbances caused by intralanguage and interspeaker variability. After comparing the results of the different back-end classifying algorithms, discriminative SVM classifier with pairwise posterior probability achieves the most performance improvement. Furthermore, log-likelihood normalization method is adopted to further improve the performance of language identification task.

ACKNOWLEDGMENTS

This work is partially supported by the Ministry of Science and Technology of the People's Republic of China (973 Program, 2004CB318106), National Natural Science Foundation of China (10574140, 60535030), and The National High Technology Research and Development Program of China (863 Program, 2006AA010102, 2006AA01Z195).

REFERENCES

[1] K.-P. Li, "Automatic language identification using syllabic spectral features," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '94)*, vol. 1, pp. 297–300, Adelaide, Australia, April 1994.

[2] T. Nagarajan and H. A. Murthy, "Language identification using acoustic log-likelihoods of syllable-like units," *Speech Communication*, vol. 48, no. 8, pp. 913–926, 2006.

[3] Y. K. Muthusamy, N. Jain, and R. A. Cole, "Perceptual benchmarks for automatic language identification," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '94)*, vol. 1, pp. 333–336, Adelaide, Australia, April 1994.

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

[5] J. L. Gauvain, A. Messaoudi, and H. Schwenk, "Language recognition using phone lattices," in *Proceeding of the International Conference on Spoken Language Processing (ICSLP '04)*, pp. 1283–1286, Jeju Island, South Korea, October 2004.

[6] W. Shen, W. Campbell, T. Gleason, D. Reynolds, and E. Singer, "Experiments with lattice-based PPRLM language identification," in *Proceedings of IEEE Odyssey on Speaker and Language Recognition Workshop*, pp. 1–6, San Juan, Puerto Rico, June 2006.

[7] H. Li, B. Ma, and C.-H. Lee, "A vector space modeling approach to spoken language identification," *IEEE Transaction on Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 271–284, 2006.

[8] E. Noor and H. Aronowitz, "Efficient language identification using anchor models and support vector machines," in *Proceedings of IEEE Odyssey on Speaker and Language Recognition Workshop*, pp. 1–6, San Juan, Puerto Rico, June 2006.

[9] H. Jin, F. Kubala, and R. Schwartz, "Automatic speaker clustering," in *Proceedings of the DARPA Speech Recognition Workshop*, pp. 108–111, Chantilly, Va, USA, February 1997.

[10] H. Gish, M.-H. Siu, and R. Rohlicek, "Segregation of speakers for speech recognition and speaker identification," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '91)*, vol. 2, pp. 873–876, Toronto, Canada, May 1991.

[11] C. White, I. Shafran, and J.-L. Gauvain, "Discriminative classifiers for language recognition," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '06)*, vol. 1, pp. 213–216, Toulouse, France, May 2006.

[12] L.-F. Zhai, M.-H. Siu, X. Yang, and H. Gish, "Discriminatively trained language models using support vector machines for language identification," in *Proceedings of IEEE Odyssey on Speaker and Language Recognition Workshop*, pp. 1–6, San Juan, Puerto Rico, June 2006.

[13] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," 2001, http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[14] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *The Journal of Machine Learning Research*, vol. 5, pp. 975–1005, 2004.

[15] D. Price, S. Knerr, L. Personnaz, and G. Dreyfus, "Pairwise neural network classifiers with probabilistic outputs," in *Neural Information Processing Systems*, vol. 7, pp. 1109–1116, MIT Press, Cambridge, Mass, USA, 1995.

[16] P. Refregier and F. Vallet, "Probabilistic approach for multiclass classification with neural networks," in *Proceedings of International Conference on Artificial Networks*, pp. 1003–1007, Espoo, Finland, June 1991.

[17] Y. Yan and E. Barnard, "An approach to automatic language identification based on language-dependent phone recognition," in *Proceedings of the 20th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '95)*, vol. 5, pp. 3511–3514, Detroit, Mich, USA, May 1995.

[18] E. Barnard and Y. Yan, "Toward new language adaptation for language identification," *Speech Communication*, vol. 21, no. 4, pp. 245–254, 1997.

[19] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," *The Journal of the Acoustical Society of America*, vol. 87, no. 4, pp. 1738–1752, 1990.

[20] A. Zolnay, R. Schlüter, and H. Ney, "Acoustic feature combination for robust speech recognition," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, vol. 1, pp. 457–460, Philadelphia, Pa, USA, March 2005.

[21] W. M. Campbell, J. P. Campbell, D. A. Reynolds, E. Singer, and P. A. Torres-Carrasquillo, "Support vector machines for speaker and language recognition," *Computer Speech & Language*, vol. 20, no. 2-3, pp. 210–229, 2006.

[22] E. Barnard and R. A. Cole, "A neural-net training program based on conjugate gradient optimization," Tech. Rep. CSE 89-014, Department of Computer Science, Oregon Graduate Institute of Science and Technology, Portland, Ore, USA, 1989.

[23] N. Brümmer and D. A. van Leeuwen, "On calibration of language recognition scores," in *Proceedings of IEEE Odyssey on Speaker and Language Recognition Workshop*, pp. 1–8, San Juan, Puerto Rico, June 2006.

[24] E. Singer, P. A. Torres-Carrasquillo, T. P. Gleason, W. M. Campbell, and D. A. Reynolds, "Acoustic, phonetic and discriminative approaches to automatic language recognition," in *Proceedings of the European Conference on Speech Communication Technology (Eurospeech '03)*, pp. 1345–1348, Geneva, Switzerland, September 2003.

[25] Y. K. Muthusamy, R. A. Cole, and B. T. Oshika, "The OGI multilanguage telephone speech corpus," in *Proceeding of the International Conference on Spoken Language Processing (ICSLP '92)*, pp. 895–898, Banff, Canada, October 1992.

[26] http://www.ldc.upenn.edu/Catalog.

[27] http://www.nist.gov/speech/tests.

[28] A. F. Martin and A. N. Le, "The current state of language recognition: NIST 2005 evaluation results," in *Proceedings of IEEE Odyssey on Speaker and Language Recognition Workshop*, pp. 1–6, San Juan, Puerto Rico, June 2006.