(Research Proposal)

Biometric Voice Recognition using Gaussian Mixture Model (GMM) and Radial Basis Function (RBF): A Comparative Study.

By

Anifowose Fatai Adesina

(260352)

Submitted to:

Dr. Lahouari Ghouti

On March 24, 2008

Term (072)

Independent Study (ICS – 591)

Abstract

A comparative study of the application of Gaussian Mixture Model (GMM) and Radial Basis Function (RBF) in biometric recognition of voice is proposed. The application of machine learning techniques to biometric authentication and recognition problems has gained a widespread acceptance. In this research, a GMM model will be trained, using Expectation Maximization (EM) algorithm, on a dataset containing 10 classes of vowels and the model will be used to predict the appropriate classes using a validation dataset. For experimental validity, the model will be compared to the performance of a RBF model using the same learning and validation dataset. A conclusion will be drawn as to which model performs better using classification rate and confusion matrix as criteria for performance evaluation.

1.0 Introduction

Biometrics is a measurable, physical characteristic or personal behavioral trait used to recognize the identity, or verify the claimed identity, of a candidate. Biometric recognition is a personal recognition based on “who you are or what you do” as opposed to “what you know” (password) or “what you have” (ID card) [17]. The goal of voice recognition in biometrics is to verify an individual's identity based on his or her voice. Because voice is one of the most natural forms of communication, identifying people by voice has drawn the attention of lawyers, judges, investigators, law enforcement agencies and other practitioners of forensics.

Computer forensics is the application of science and engineering to the legal problem of digital evidence. It is a synthesis of science and law [8]. A high level of accuracy is required in critical systems such as online financial transactions, critical medical records, preventing benefit fraud, resetting passwords, and voice indexing.

In view of the importance of accurate classification of vowels in a voice recognition system, the need for a well-trained computational intelligence model with an acceptable percentage of classification accuracy (hence a low percentage of misclassification error) is highly desired. Gaussian mixture models (GMMs) and radial basis function (RBF) networks have been identified in both practice and literature as two of the promising neural models for pattern classification.

2.0 Literature Survey

2.1 Voice Recognition

A good deal of effort has been made in the recent past by researchers in their attempt to come up with computational intelligence models with an acceptable level of classification accuracy.

A novel suspect-adaptive technique for robust forensic speaker recognition using Maximum A-Posteriori (MAP) estimation was presented by Ramos-Castro et al. [1]. The technique addressed Likelihood Ratio (LR) computation in limited suspect speech data conditions obtaining good calibration performance and robustness by allowing the system to weigh the relevance of the suspect specificities depending on the amount of suspect data available via MAP estimation. The results also showed that the proposed technique outperformed other previously proposed non-adaptive approaches.

Hongbin et al. [2] presented three mainstream approaches including parallel phone recognition language modeling (PPRLM), support vector machine (SVM) and the general Gaussian mixture models (GMMs). The experimental results showed that the SVM framework achieved an equal error rate (EER) of 4.0%, outperforming the state-of-art systems by more than 30% relative error reduction. Also, the performances of their proposed PPRLM and GMMs algorithms achieved an EER of 5.1% and 5.0% respectively.

Support vector machines (SVMs) were presented by Tianqiang et al. [3] by introducing a sequence kernel used in language identification. Then a Gaussian Mixture Model was developed to do the sequence mapping task of a variable length sequence of vectors to a fixed dimensional space. Their results demonstrated that the new system yielded a performance superior to those of a GMM classifier and a Generalized Linear Discriminant Sequence (GLDS) Kernel.

Using a vowel detection algorithm, Jean-Luc et al. [4] segmented rhythmic units related to syllables by extracting parameters such as consonantal and vowel duration, and cluster complexity and modeled with a Gaussian Mixture. Results reached up to $86 \pm 6\%$ of correct discrimination between stress-timed, mora-timed and syllable-timed classes of languages. These were then compared with that of a standard acoustic gaussian mixture modeling approach that yielded $88 \pm 5\%$ of correct identification.

Córdoba et al. [9] presented an additive and cumulative improvements over several innovative techniques that can be applied in a Parallel phone recognition followed by language modeling (PPRLM) system for language identification (LID), obtaining a 61.8% relative error reduction from the base system. They started from the application of a variable threshold in score computation with a 35% error reduction, then a random selection of sentences for the different sets and the use of silence models, then, compared the bias removal technique with up to 19% error reduction and a Gaussian classifier of up to 37% error reduction, then, included the acoustic score in the Gaussian classifier with 2% error reduction, increased the number of Gaussians to have a multiple-Gaussian classifier with 14% error reduction and finally, included additional acoustic HMMs of the same language with success gaining 18% relative improvement.

More works are abounding in literature.

2.2 Gaussian Mixture Model (GMM)

From a clustering perspective, most biometric data cannot be adequately modeled by a single-cluster Gaussian model. However, they can often be accurately modeled via a Gaussian mixture model (GMM) (i.e., data distribution can be expressed as a mixture of multiple normal distributions [7].

Torres-Carrasquillo et al. [5] presented a generalized technique by using GMM and obtained an error of 17%. A description of the major elements of MIT Lincoln Laboratory’s Gaussian mixture model (GMM)-based speaker verification system built around the likelihood ratio test for verification, using simple but effective GMMs for likelihood functions, a universal background model (UBM) for alternative speaker representation, and a form of Bayesian adaptation to derive speaker models from the UBM were presented by Reynolds et al. [6]. The results showed that the GMM-UBM system has proven to be very effective for speaker recognition tasks.

Other related works on GMM include [10, 11, 12, 13].

2.3 Radial Basis Function (RBF)

RBF model, with its mathematical properties of interpolation and design matrices, is one of the promising neural models for pattern classification [14] and has also gained popularity in voice recognition.

Li Guojie [15] presented a comparative study of the application of a minimal RBF Neural Network, the normal RBF and an elliptical RBF for speaker verification. The experimental results showed that the Minimal RBF outperforms the other techniques.

A work for explicitly modeling voice quality variance in the acoustic models using hidden Markov models, in order to improve word recognition accuracy, was demonstrated by Tae-Jin et al. They also presented SVM and concluded that voice quality can be classified using input features in speech recognition [16].

Other related works have been found in the fields of medicine [14], hydrology [18], computer security [19], petroleum engineering [20] and computer networking [21].

3.0 Data and Tools

3.1 Data

The training and testing data are obtained from an experimental 2-dimensional dataset available in [22]. The training data consists of 338 observations while the testing data consists of 333 observations. Each observation belongs to one of 10 classes of vowels to be classified using the trained models.

3.2 Tools

The GMM and RBF classifiers will be implemented in MATLAB with the support of Netlab toolbox obtained as freeware from [23].

4.0 Methodology

The methodology in this work shall be based on the standard Pattern Recognition approach to classification problem using GMM and RBF. For training the models, Expectation Maximization (EM) algorithm will be used for efficient optimization of the parameters. The parameters of the models will also be tuned and varied and those with maximum classification accuracy will be selected.

5.0 Criteria for Performance Evaluation

The most commonly used accuracy measures in classification tasks are classification rate and confusion matrix.

Classification rate is calculated by:

$$\frac{\text{Number of correctly classified points}}{\text{Total number of data points}} \times 100\%$$

The confusion matrix is a useful tool for analyzing how well a classifier can recognize tuples of different classes. For a classifier to have good accuracy, ideally most of the tuples would be represented along the diagonal of the confusion matrix, with the rest of the entries being close to zero. Given two classes, we can talk in terms of positive tuples (tuples of the main class of interest) versus negative tuples. True positives refer to the positive tuples that were correctly labeled by the classifier, while true negatives are the negative tuples that were correctly labeled by the classifier. False positives are the negative tuples that were incorrectly labeled. Similarly, false negatives are the positive tuples that were incorrectly labeled. These terms are useful when analyzing a classifier's ability [24]. A typical confusion matrix is shown below.

		Predicted Class	
Actual Class	Class 1	True Positives	False Negatives
	Class 2	False positives	True Negatives

6.0 Conclusion

A comparative study of the application of Gaussian Mixture Model (GMM) and Radial Basis Function (RBF) with parameters optimized with EM algorithm for biometric recognition of vowels is being proposed. At the end of the study, it is expected that a landmark classification accuracy will be obtained.

References

1. Ramos-Castro, D.; Gonzalez-Rodriguez, J.; Montero-Asenjo, A.; Ortega-Garcia, J., "Suspect-Adapted MAP Estimation of Within-Source Distributions in Generative Likelihood Ratio Estimation," Speaker and Language Recognition Workshop, 2006. IEEE Odyssey 2006: The , vol., no., pp.1-5, June 2006.
2. Hongbin SUO1,2, Ming LI1, Ping LU1 and Yonghong YAN, Automatic Language Identification with Discriminative Language Characterization Based on SVM, IEICE-Transactions on Info and Systems, Volume E91-D, Number 3 , Pp. 567-575, 2008.
3. Tianqiang Peng, Wenlin Zhang, Bicheng Li, "SVM-UBM Based Automatic Language Identification Using a Vowel-guided Segmentation," Third International Conference on Natural Computation (ICNC 2007), icnc, pp. 310-314, 2007.
4. Jean-Luc Rouas, Jerome Farinas, Francois Pellegrino and Regine Andre-Obrecht, Rhythmic unit extraction and modeling for automatic language identification, Speech Communication, Volume 47, Issue 4, December 2005, Pages 436-456.
5. Torres-Carrasquillo, P.A.; Reynolds, D.A.; Deller, J.R., Jr., "Language identification using Gaussian mixture model tokenization," Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02). IEEE International Conference on , vol.1, no., pp. I-757-I-760 vol.1, 2002.
6. Douglas A. Reynolds, Thomas F. Quatieri, and Robert B. Dunn, Speaker Verification Using Adapted Gaussian, Mixture Models, Digital Signal Processing, Vol. 10, 19–41 (2000).
7. S. Y. Kung; M. W. Mak; S. H. Lin, Biometric Authentication: A Machine Learning Approach, Prentice Hall, September 14, 2004, Pages: 496
8. Tony Sammes and Brian Jenkinson, Forensic Computing: A Practitioner's Guide, Second Edition, Springer-Verlag, 2007, Pp. 10.
9. R. Córdoba, L.F. D'Haro, R. San-Segundo, J. Macías-Guarasa, F. Fernández, J.C. Plaza, A multiple-Gaussian classifier for Language Identification using acoustic information and PPRLM scores, IV Jornadas en Tecnologia del Habla, 2006, Pp. 45-48.
10. Pedro A. Torres-Carrasquillo1, 2, Elliot Singer2, Mary A. Kohler3, Richard J. Greene2, Douglas A. Reynolds2, and J.R. Deller, Jr., Approaches to Language Identification using

Gaussian Mixture Models and Shifted Delta Cepstral Features, Proceedings of International Conference on Spoken Language Processing, 2002.

11. Too Chen; Chao Huang; Chang, E.; Jingehan Wang, "Automatic accent identification using Gaussian mixture models," Automatic Speech Recognition and Understanding, 2001. ASRU '01. IEEE Workshop on , vol., no., pp. 343-346, 9-13 Dec. 2001
12. Torres-Carrasquillo, P. A., Gleason, T. P., and Reynolds, D. A., Dialect Identification Using Gaussian Mixture Models. In Proc. Odyssey: The Speaker and Language Recognition Workshop in Toledo, Spain, ISCA, pp. 297-300, 31 May - 3 June 2004.
13. Wuei-He Tsai and Wen-Whei Chang, Discriminative training of Gaussian mixture bigram models with application to Chinese dialect identification, Speech Communication, Volume 36, Issue 3, March 2002, Pages: 317 - 326
14. Miyoung Shin and Cheehang Park, A Radial Basis Function Approach to Pattern Recognition and Its Applications, ETRI Journal, Volume 22, Number 2, June 2000.
15. Li Guojie, Radial Basis Function Neural Network for Speaker Verification, A Master of Engineering thesis submitted to the Nanyang Technological University, 2004.
16. Yoon, T., Zhuang, X., Cole, J., Hasegawa-Johnson, M., Voice quality dependent speech recognition, In Tseng, S. (Ed.), Linguistic Patterns of Spontaneous Speech, Special Issue of Language and Linguistics, Academica Sinica, 2007.
17. Anil K. Jain, Multimodal User Interfaces: Who’s the User? International Conference on Multimodal Interfaces, Documents in Computing and Information Science, 2003.
18. Gwo-Fong Lin, Lu-Hsien Chen, A non-linear rainfall-runoff model using radial basis function network, Journal of Hydrology 289, 2004.
19. Azzedine Boukerche, Behavior-Based Intrusion Detection in Mobile Phone Systems, Journal of Parallel and Distributed Computing 62, 1476–1490, 2002.
20. Anna Ilce´a Fischetti, Andre´ Andrade, Porosity images from well logs, Journal of Petroleum Science and Engineering 36, 2002, 149– 158.
21. Dimitris Gavrilis, Evangelos Dermatas, Real-time detection of distributed denial-of-service attacks using RBF networks and statistical features, Computer Networks 48, 2005, 235–245.
22. http://www.eie.polyu.edu.hk/~mwmak/Book
23. http://www.ncrg.aston.ac.uk/netlab

24. Jiawei Han and Micheline Kamber, Data Mining Concepts and Techniques, Second Edition, Morgan Kaufmann, 2006, Pp. 361.