ICSI System Description for SRE 2010 Submission

Howard Lei

The International Computer Science Institute, Berkeley, CA

hlei@icsi.berkeley.edu

1. Introduction

The two ICSI systems involved in the SRE10 evaluation are the GMM-UBM system with simplified factor analysis, and the GMM supervector system with simplified factor analysis. The ALIZE implementation [1] for UBM training, factor analysis training [2], and likelihood-ratio scoring is used, and SVM-Light [3] is used for the SVM implementation.

2. Data and feature extraction

We've used separate development data for UBM and inter-session variability matrix training (for simplified factor analysis) for four separate speaker model vs. test utterance channel conditions - telephone and telephone, telephone and interview, interview and interview along with interview and microphone, and microphone and microphone. The number of conversation sides and corpora used for each channel condition is shown in table 1. Note that the same data used for UBM training is also used for inter-session variability matrix training.

Channel condition	Corpora	Conversation sides
Tel-Tel	SRE04	2,500
Int-Tel	SRE08	3,000
Int-Int + Int-Mic	SRE08	2,800
Mic-Mic	SRE06+SRE08	2,900

Table 1: Data used for UBM and inter-session variability matrix training.

Table 2 shows the data used for the development of scoring thresholds and optimal system combinations for each channel condition.

Channel condition	Corpora	Conversation sides
Tel-Tel	SRE06+SRE08	6,100
Int-Tel	SRE08	1,700
Int-Int + Int-Mic	SRE08	4,500
Mic-Mic	SRE08	1,300

Table 2: Data used for development of scoring thresholds and optimal system combinations.

We are provided with implicit speech/non-speech decodings via the DECIPHER recognizer [4]. We've used MFCC features (C0-C12 plus deltas and double-deltas) with Wiener filtering and mean-variance normalization for all systems. Feature extraction is performed using HTK [5].

3. System 1: GMM-UBM system

A 512-mixture GMM-UBM system with likelihood-ratio scoring is implemented via ALIZE [1], and the simplified factor analysis model [2] is used for inter-session variability compensation. The SFA scoring technique is used [1], giving us the optimal speaker verification results while making score normalizations (z-norm, t-norm, zt-norm) unnecessary. A separate system is implemented for each of the telephone-telephone, interview-telephone, interview-interview along with interview-microphone, and microphone-microphone channel conditions. The single-CPU execution times for all SRE10 trials, including UBM and inter-session variability matrix training, is approximately 120 hours.

4. Systems 2: GMM supervector system

The ALIZE framework is used for 512-mixture UBM and GMM supervector training along with simplified factor analysis training, and SVM-Light [3] is used for SVM training. The standard GMM supervector SVM kernel is used [6]. A separate system is implemented for each of the telephone-telephone, interview-telephone, interview-interview along with interview-microphone, and microphone-microphone channel conditions. The single-CPU execution times for all SRE10 trials, including UBM and inter-session variability matrix training, is approximately 80 hours.

5. System combination

We've used the LNKnet [7] MLP implementation with 2 hidden nodes and 1 hidden layer for score-level system combinations. For all but the telephone-telephone channel condition, however, the GMM supervector system performs poorly, and is left out of the combinations.

6. SRE08 submissions

We've submitted results for the core-core condition. Our primary submission involves the MLP combination of the GMM-UBM and GMM supervector systems for the telephone-telephone channel condition trials, and the GMM-UBM system for all remaining trials. Our second submission involves using the GMM-UBM system for all trials.

7. References

[1] Bonastre, J.F., Wils, F., Meignier, S.,"ALIZE, a free Toolkit for Speaker Recognition", in Proc. of ICASSP, 2005. and Machine Intelligence, 24(12):1667-1671, 2002.

[2] Kenny, P., Boulianne, G., Ouellet, P., and Dumouchel, P.,"Factor analysis simplified", in Proc. of ICASSP, 2005.

[3] Joachims, T., "Making Large Scale SVM Learning Practical",

in Advances in kernel methods - support vector learning, B. Schoelkopf, C. Burges, and A. Smola, Eds. MIT-press, 1999.

[4] Kajarekar, S., Ferrer, L., Venkataraman, A., Sonmez, K., Shriberg, E., Stolcke, A., Gadde, R.R., "Speaker Recognition using Prosodic and Lexical features", in Proc. of the IEEE Automatic Speech Recognition and Understanding Workshop, pp. 19-24, 2003.

[5] HMM Toolkit (HTK): http://htk.eng.cam.ac.uk

[6] Campbell, W.M., Sturim, D.E., Reynolds, D.A., Solomonoff, A., "SVM Based Speaker Verification using a GMM Supervector Kernel and NAP Variability Compensation", in Proc. of ICASSP, Vol. 1, pp I-97-I -100, 2006.

[7] Lippmann, R.P., Kukolich, L.C., Singer, E., "LNKnet: Neural Network, Machine Learning, and Statistical Software for Pattern Classification", in Lincoln Laboratory Journal, Vol. 6, pp 249-268, 1993.