RESEGMENTATION OF SWITCHBOARD

Janna Shaffer, Neeraj Deshmukh, Aravind Ganapathiraju, Jonathan Hamaker, Joseph Picone

Institute for Signal and Information Processing
Department of Electrical and Computer Engineering
Mississippi State University,
Mississippi State, MS 39762
{shaffer, deshmukh, ganapath, hamaker, picone}@isip.msstate.edu

SUMMARY

The growing interest in performing accurate and automatic large vocabulary conversational speech recognition (LVCSR) has made the SWITCHBOARD (SWB) corpus [1] one of the most important benchmarks in recent times. Analysis of current recognition performance on SWB reveals two significant issues — monosyllabic words dominate the overall error rate, and the transcriptions for monosyllabic words (and compound-word phrases) are frequently erroneous [2]. It is clear that, to achieve better acoustic modeling, the quality of the transcriptions needs to be improved. The high speech rates of spontaneous conversations; the variation in pronunciations due to different speaking styles, dialect and context; poor coarticulation and a vocabulary dominated by monosyllabic words pose unique challenges to accurate transcription, and hence effective training of acoustic models.

Segmentation of conversational speech into relatively short phrases enhances transcription accuracy, helps in reducing the computational requirements for training and testing each utterance, and simplifies the language model (LM) application during recognition. Traditional techniques for automatic segmentation rely on energy levels, phone-level recognition etc. [3]. However, these introduce unnatural breakpoints in the utterances, thus decreasing the effectiveness of the LM. Linguistically motivated segmentation [4] often results in extremely short phrases that do not provide sufficient acoustic context. We seek to balance this trade-off by resegmenting the automatic segmentations and ensuring ample context for both acoustic and language modeling applications.

We have developed a Tcl-Tk based platform-independent graphical-interface tool that assists in performing the resegmentation at less than 10x real-time. This primarily involves splitting or merging segments into utterances buffered with approximately 0.5 seconds of silence on either side, keeping the typical utterance duration under 10 seconds; and correcting any transcription errors. As a further check, the resegmented data is used to generate automatic word-alignments, which are then reviewed using the segmenter (at less than 5x real-time). We have observed less than 10% error in the automatic word alignments, and this is rectified during the review process.

We believe that resegmentation of the database is an important step in improving overall recognition performance — as is evident from our resegmentation work on a dev-test set that resulted in an over 2% absolute reduction in WER [2]. We have already released this resegmented dev set in the public domain [5]. Resegmentation of the entire corpus is expected to be complete soon, and we anticipate a significant gain in performance by the conference time. Our final paper will contain a detailed analysis of the effect of resegmentation and the resulting modalities in WER.

1 J. Godfrey, E. Holliman and J. McDaniel, “SWITCHBOARD: Telephone Speech Corpus for Research and Development”, in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 517-520, San Francisco, California, USA, March 1992.

2 A. Ganapathiraju, V. Goel, J. Picone, A. Corrada, G. Doddington, K. Kirchoff, M. Ordowski, and B. Wheatley, “Syllable - A Promising Recognition Unit for LVCSR”, *Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop*, pp. 207-214, Santa Barbara, California, USA, December 1997.

3 B. Wheatley, G. Doddington, C. Hemphill, J. Godfrey, E. Holliman, J. McDaniel and D. Fisher, “Robust Automatic Time Alignment of Orthographic Transcriptions with Unconstrained Speech”, in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 533-536, San Francisco, California, USA, March 1992.

4 S. Greenberg, “The Switchboard Transcription Project”, *1996 LVCSR Summer Research Workshop*, Research Notes 24, Center for Language and Speech Processing, Johns Hopkins University, Baltimore, Maryland, USA, April 1997.

5 *http://www.isip.msstate.edu/resources/technology/projects/1998/switchboard/.*