Classifier Combination Schemes In Speech Impediment Therapy Systems

Dénes Paczolay, László Felföldi, and András Kocsor

Research Group on Artificial Intelligence
of the Hungarian Academy of Sciences and University of Szeged
H-6720 Szeged, Aradi vértanúk tere 1., Hungary
{pdenes, lfelfold, kocsor}@inf.u-szeged.hu
http://www.inf.u-szeged.hu/speech

Extended Abstract

In the therapy of the hearing impaired one of the central problems is the handling of the lack of proper auditive feedback which impedes the development of intelligible speech. Our Phonological Awareness Teaching System, the "SpeechMaster" package, seeks to apply speech recognition technology to speech therapy [7, 8]. It provides a visual phonetic feedback for replacing the insufficient auditive feedback of the hearing impaired. We designed and implemented computer-aided training software that uses an effective phoneme recognizer and provides a real-time visual feedback in the form of flickering letters on calling pictures. The brightness of the letters is proportional to the speech recognizers output.

The effectiveness of the therapy relies heavily on accurate phoneme recognition. Phoneme recognition is a special pattern recognition problem [1, 2, 11] where the continuously varying speech signal has to be mapped to a symbol of a phoneme. Because of the environmental conditions, simple recognition algorithms may have a weak classification performance, so various techniques such as normalization and classifier combination are applied to increase the recognition accuracy.

Speaker normalization reduces the variance in the speech data of different speakers caused by their different vocal tract lengths. Vocal Tract Length Normalization techniques [3, 10] transform the speech data to the space of the "standard" speaker. This transformation is determined by a warp factor correlated with the speaker's vocal track length. In an earlier paper [9] we demonstrated how to estimate this warp factor in real-time.

Classifier combinations [6, 12] aggregate the results of many classifiers, overcoming the possible local weakness of the individual inducers, thus producing a more robust classification performance. In this paper the traditional (*Prod*, *Sum*, *Min*, *Max*, etc.) [5], linear (*simple-*, *weighted-*, and *AHP-based* [4] *averaging*), nonlinear (*kernel*) and stacked combination rules are examined.

From experimental tests we found that classifier combinations did prove effective in real-time speech recognition, fulfilling the special requirements of the task of therapy.

References

1. C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
2. R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley and Son, New York, 2001.
3. E. Eide and H. Gish. A parametric approach to vocal tract length normalization. In *ICASSP*, pages 1039–1042, Munich, 1997.
4. L. Felföldi and A. Kocsor. Ahp-based classifier combination. In *The 4th International Workshop on Pattern Recognition in Information Systems (PRIS-2004)*, Porto, 2004.
5. L. Felföldi, A. Kocsor, and L. Tóth. Classifier combination in speech recognition. *Periodica Polytechnica*. Accepted for publication.
6. Anil K. Jain, Robert P. W. Duin, and Jianchang Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, 2000.
7. A. Kocsor and K. Kovács. Kernel springy discriminant analysis and its application to a phonological awareness teaching system. In *Text Speech and Dialogue*, volume 2448, pages 325–328. Springer, 2002.
8. A. Kocsor, L. Tóth, and D. Paczolay. A nonlinearized discriminant analysis and its application to speech impediment therapy. In *Text Speech and Dialogue*, volume 2166, pages 249–257, Czech Republic, 2001. Springer.
9. D. Paczolay, A. Kocsor, and L. Tóth. Real-time vocal tract length normalization in a phonolgical awareness teaching system. In *Text Speech and Dialogue*, volume 2807, pages 4–37, Czech Republic, 2003. Springer.
10. P. Pitz, S. Molau, R. Schlter, and H. Ney. Vocal tract normalization equals linear transformation in cepstral space. In *EUROSPEECH*, volume 4, pages 2653–2656, Denmark, 2001.
11. V. N. Vapnik. *Statistical Learning Theory*. John Wiley and Son, 1998.
12. L. Xu, A. Krzyzak, and C.Y. Suen. Method of combining multiple classifiers and their application to handwritten numeral recognition. *IEEE Trans. on SMC*, 22(3):418–435, 1992.