

Partial least squares based speaker recognition system

Balaji Vasan Srinivasan and Ramani Duraiswami

Partial least squares (PLS) methods [1] have been successfully applied for several pattern recognition tasks in chemometrics, computer vision and medical imaging. PLS can be used for regression, classification and supervised dimensionality reduction. In this work, we present an efficient PLS framework for text-independent speaker recognition. Speaker recognition [2] deals with the task of verifying a speaker's identity based on the corresponding voice samples. The key challenge here is to model speaker characteristics with limited target-speaker training samples (ranging from a few seconds to a couple of minutes of data) that are susceptible to variabilities that are not speaker-specific (message, channel, noise). The features of choice for discriminative speaker training is a combination of features derived from the mel-frequency cepstral coefficients and cortical features, which are combined in to a high dimensional "super-vector" based on Gaussian Mixture Models (GMMs) [2]. We use these supervectors to learn a speaker-specific PLS-subspace where the target speaker is well-separated from non-target speaker. To treat the large datasets encountered in speaker recognition, we employ a novel extension of the PLS on graphical processors which obtains a ~30X speedup. The system was tested on NIST SRE 2008 data for various training and test conditions, and the results are promising.

References

- [1] R. Rosipal and N. Kramer (2006). "Overview and recent advances in partial least squares", in *Subspace, Latent Structure and Feature Selection Techniques: Lecture Notes in Computer Science*, pp. 34-51, Springer.
- [2] T. Kinnunen and H. Li (2010). "An overview of text-independent speaker recognition: From features to supervectors", *Speech Communication*, vol. 52, no. 1, pp. 12-40.

Topic: speech and audio

Preference: oral/poster