Review of Ph.D dissertation

Candidate: Ondrej Glembek
Title: OPTIMIZATION OF GAUSSIAN MIXTURE SUBSPACE MODELS AND RELATED SCORING ALGORITHMS IN SPEAKER VERIFICATION
Reviewer: Niko Brummer
15 October 2012

Analysis

This review addresses the following questions:

- *Is the topic appropriate to the particular area of dissertation and is it up-to-date from the viewpoint of the present level of knowledge?*

  Yes. Automatic speaker recognition is a very active research field, supported by a substantial research community as well as both commercial and government funding in several countries. The algorithms (JFA and ivectors) researched in this dissertation are an important part of the state-of-the-art.

- *Is the work original and does it mean a contribution to the area - specify where the original contribution lies?*

  One the one hand this dissertation reports on research done by the candidate from 2008 to the present, in collaboration with other researchers. It contributes to this research by providing a novel summary, with comparative analyses of several variants of JFA and ivectors algorithms. I believe in particular that the analysis of different ways to evaluate GMM likelihoods is a valuable contribution and could be used as a reference by future researchers to understand this problem.

  On the other hand, the dissertation reports on innovations where the candidate himself made a substantial contribution to their success. This includes:

  a. Linear scoring for JFA, which gave orders of magnitude improvement in speed and thereby provided an important contribution towards progress in the field.
  b. Discriminative training of ivector systems. Although the work described here by the candidate did not lead to accuracy that improved on the existing state of the art, it nevertheless contributes to the groundwork for further research in this area.

- Has the core of the doctoral thesis been published at an appropriate level?

  Yes. The work described here (and much else besides on related topics, like automatic spoken language recognition) has been published by this candidate and co-authors in three peer-reviewed journal papers, many peer reviewed conference proceedings and workshop reports.

- Does the list of the candidate's publications imply that he is a person with an outstanding research erudition?

  Yes. I believe this candidate has demonstrated that he has acquired a wide knowledge and deep understanding of this area in automatic speaker recognition.

Conclusion

In my opinion, the doctoral thesis meets the requirements of the proceedings leading to PhD title conferment. I would however ask the candidate to make a few minor corrections to the dissertation as listed below.

Corrections

Abstract: “dramatically *reducing* computation speed” should be *increasing*.

Chapter 1: Introduction.

- First paragraph, second sentence. “It is assumed that the process is independent of the channel, i.e. language, communication channel, content, etc.” This sentence is unclear to me. Please rephrase this.
- Second paragraph. The term *voiceprint* is frowned upon by many speaker recognition researchers. Please use *speaker model* instead.

Chapter 2:

- section 2.1: Typos: *unseed, developped, corelate*
- section 2.3.3: Typo: *tripplet*

Chapter 3:

- section 3.1, equation 3.9: Sloppy notation. You should use a different index, say c’, in the sum in the denominator.
- section 3.2: The first paragraph contains a statement about sufficient statistics. The statistics referred to here are sufficient for one component, but not for the whole GMM. Please clarify this.
- section 3.5. The MAP adaptation recipe is attributed to Reynolds 2000. It may be appropriate here to reference also the original work by Gauvain.

Chapter 4:

- section 4.2.1: refers to the *average frame likelihood.* The theory does not call for the average, but the sum of the frame log-likelihoods.Please clarify that (i) we are working with *log* likelihoods here and (ii) that taking the average, rather than the sum is an empirically motivated expedient, that probably helps to compensate for the unrealistic GMM frame independence assumption.
- table 4.3 (and maybe elsewhere): faulty notation. The engineering notation for e.g. 1.60E-1 or 1.6e-1 should *not* be written with a superscript: e^{-1}. See for example http://en.wikipedia.org/wiki/Engineering_notation.

Chapter 5:

- introduction (and maybe elsewhere): please remove the term *total-variability.* Your model does not account for all variability.

Chapter 6:

- introduction “maximize the system’s cross-entopy” should be *minimize*.
- between equations 6.3 and 6.4: the term *log-posterior-ratio*, could be better expressed as *posterior log odds*.
- section 6.4.2, results: There is reference here to *the regularization coefficient* \lambda. Where is this coefficient and yoyr regularization penalty defined?
- Please make it clear that (if I understand correctly), you retrained the extractor parameters, while keeeping the PLDA parameters *fixed*.
- figure 6.9: please define the two traces in this plot.