Instituto de Ciencias de la Computación

August 23rd, 2021

Review of the doctoral thesis by Ing. ONDREJ NOVOTNÝ

Dear Committee Members,

It is my pleasure to act as reviewer for Ing. Novotny's doctoral thesis entitled "Improving Robustness of Speaker Recognition using Discriminative Techniques".

The issue of robustness in speaker verification (SV) systems is still an open research problem. While SV systems usually perform quite well on conditions that are perfectly matched to those seen during training and development, they degrade significantly when there is a mismatch in conditions. Ondrej Novotny's thesis deals with this very important problem, approaching the problem from several different perspectives. I believe Ondrej's work was an important part of the community's body of work that resulted in the current state of the art for speaker verification and language recognition.

Generative methods (i-vector followed by PLDA) were the standard approach a few years ago when Ondrej started his thesis work. Ondrej's work starts with that approach and proposes several changes to improve performance and robustness to the system. These approaches include many of the methods that are currently part of the state of the art for robust SV systems, like multi-condition training, data augmentation and discriminative training.

Ondrej's work was published on prestigious journal and conferences in our area, like Interspeech, ICASSP, Odyssey and Computer Speech and Language. These are some of the main venues in which the SV community publishes their work. Further, he has a large number of publications as co-author, showing that he is an important member of the group who has contributed to a number of different projects.

One aspect that I believe is central to the issue of robustness in SV is calibration. While, as shown in the thesis, discrimination metrics are often degraded in mismatched conditions, the degree of degradation in these metrics is generally relatively small compared to the degree of degradation in terms of metrics that are sensitive to misscalibration (e.g., actual DCF, Cllr). Calibration results are shown in Chapter 8, where it can be seen that miss-calibration is indeed a very big problem, leading to virtually unusable systems in some conditions. I would have been very interested to see calibration sensitive metrics in chapters 9 and 10, where the methods proposed have the potential to also make the scores better calibrated across conditions. I believe this could be a very useful analysis for the community.

The thesis is generally well-written though I have made notes in the pdf about some grammatical errors that could be fixed, as well as some places were clarifications are needed. I have also made a number of minor comments and suggestions. In particular, I believe that

the presentation of results could greatly benefit from turning all (or most) tables into bar plots. Other than that, the thesis is well organized, easy to follow, and contains a good review of the technology that is important for understanding the proposed approaches.

In summary, I believe Ondrej's thesis work was an important contribution for the community, focusing on one of the main open problems for this task. His work follows the ever-changing state of the art in SV, proposing some of the techniques that are currently being used in many modern systems. His publications are at top conferences and journals in our area. For these reasons, I believe Ondrej's work should be accepted for the Doctor of Philosophy degree.

Best regards,

Luciana Ferrer
Associate Researcher
Computer Science Institute
University of Buenos Aires / CONICET, Argentina