A BROADCAST NEWS PROCESSING CHAIN FOR SEVERAL VARIETIES OF PORTUGUESE

Isabel Trancoso

INESC-ID Lisboa / IST
R. Alves Redol, 9, 1000-029 Lisboa, Portugal
http://www.l2f.inesc-id.pt/

The Broadcast News (BN) processing system developed at the Spoken Language Systems Lab of INESC-ID integrates several core technologies, in a pipeline architecture: jingle detection (JD) for excluding areas with publicity; audio pre-processing (APP) which aims at speech/non-speech classification, gender and background conditions classification, speaker clustering (diarization), and speaker identification; automatic speech recognition (ASR) that converts the segments classified as speech into text; punctuation and capitalization (Pu/Ca); topic segmentation (TS) which splits the broadcast news show into constituent stories and topic indexation (TI) which assigns one or multiple topics to each story; and summarization (Su), which assigns a short summary to each story.

The first modules of this system were optimized for on-line performance, given their deployment in the fully automatic subtitling system that is running on the main news shows of the public TV channel in Portugal since March 2008. That was one of the major outcomes of the national project Tecnovoz.

In order to improve the performance of the recognition system, its lexical and language models are daily adapted, with text material retrieved from online newspapers, and acoustic models are specially trained for very frequent speakers (anchors).

Although the integration of a spoken language identification module was originally planned, the current subtitling system is currently blocking off speech transcriptions for which the confidence measure scores are below a given threshold, thus effectively excluding speech in other languages.

The system takes advantage of the gender information produced by the APP module to change the color of the subtitles for female speakers, an action that is greatly appreciated by the representatives of the deaf community.

The current latency of the subtitling system is close to 6s, half of which are due to the speech processing chain above described, and half to the need to fill two lines of the Teletex system that appears on the top of the TV screen.

The broadcast news processing chain was originally developed for European Portuguese (EP). One of the goals of the national PoSTPort project is porting spoken language technologies to other varieties of Portuguese. Although ideally we would like to cover all the varieties spoken in CPLP countries (Community of Portuguese-speaking Countries), we had to exclude East Timor from our list, because of the difficulties in collecting corpora from this variety. The project therefore covers only 2 broad varieties besides EP: Brazilian Portuguese (BP), the variety spoken in South America, with the largest number of speakers, and African Portuguese (AP), the generic name that covers all the varieties spoken in African countries that have Portuguese as official language: Angola (AN), Cape Verde (CV), Guinea-Bissau (GB), Mozambique (MO) and São Tomé and Príncipe (ST).

Our EP ASR system dramatically fails when it is used for transcribing BP data. After several stages of model adaptation, including lexical, language, and acoustic modeling, we were able to port our EP speech recognition system to the specific characteristics of BP achieving reasonable good recognition results even with a considerable limitation of available corpora, both audio and textual. A potential cause is the fact that EP recognition systems have to deal with much more pronounced vowel reduction phenomena that create very large consonant clusters and render EP much more difficult than BP for foreign speakers. We expect that the future use of additional resources will allow remarkable improvements in the BP recognizer, including unsupervised training with unlabeled corpora.

The first experiments conducted with AP data have shown us that, even without any sort of model adaptation, the ASR system trained for EP does not significantly degrade for most BN segments. In fact, many frequent speakers such as anchors, reporters, and politicians show a very slight accent, which justifies the overall good performance. Other speakers show a pronounced accent with a great deal of variability, as expected from speakers that do not have Portuguese as a first language. Our current target is to take advantage of our automatic accent identifier module to separate the BN segments which are most confidently recognized as AP in order to retrain specific AP models.

It is interesting to notice that the punctuation and capitalization modules which were originally developed for EP, seem robust enough to be used for other varieties, although the formal evaluation was not yet conducted.

The on-line modules which were optimized for the subtitling application are also available for European Spanish and for English.

The deployment of the subtitling system was recently taken over by VoiceInteraction, the spin-off company of INESC-ID. The automatic speech-to-text translation of BN data is one of the goals of the current project in cooperation with Carnegie Mellon University.

Acknowledgments

The work described in this abstract resulted from the cooperation of a very large team, with special emphasis on the PhD theses of D. Caseiro (ASR), H. Meinedo (APP/ASR), C. Martins (ASR), R. Amaral (TS/TD), F. Batista (Pu/Ca) and R. Ribeiro (Su), in chronological order, supervised by J. Neto, I. Trancoso, N. Mamede, and D. Matos. We would also like to acknowledge the very significant contribution of A. Abad in project PoSTPort.