Meeting Assistant Application

Michel Assayag¹, Jonathan Huang², Jonathan Mamou¹, Oren Pereg¹, Saurav Sahay², Oren Shamir², Georg Stemmer³, Moshe Wasserblat¹

¹Intel Corporation Israel, PO Box 1659, Matam Industrial Park, Haifa 31015, Israel
²Intel Labs, Santa Clara, CA, 95054 USA
³Intel Corporation Germany, Dornacher Strasse 1, D-85622 Feldkirchen/Muenchen, Germany

firstname.lastname@intel.com

Abstract

This paper describes the Meeting Assistant application developed at Intel. Unlike existing human-to-machine solutions, the challenges induced by human-to-human conversations are currently poorly addressed by the industry. In this paper, we describe the capabilities of in-house speech and NLP technologies: automatic speech recognition, speaker diarization, keyphrase extraction and sentiment detection. These technologies have been adapted to conversational speech domain and integrated into the Meeting Assistant.

Index Terms: automatic speech recognition, ASR, speaker diarization, NLP, sentiment detection, keyphrase extraction, meeting assistant

1. Introduction

In the last decade, several solutions have been developed for human-to-machine interaction. However, the challenges induced by human-to-human conversations are less dealt with than human-to-machine. State-of-the-art transcription systems typically achieve 70% accuracy in transcription for conversational speech. In other words for this kind of data, approximately one out of three words is mis-recognized. In some circumstances like noisy environments, foreign accent, under-trained engines etc., accuracy may fall to 50% or even less. The low accuracy of the transcription can have a dramatic effect on the performance of the speech and NLP (Natural Language Processing) technologies. In this paper, we describe the Meeting Assistant application developed at Intel that demonstrates how speech and NLP technologies have been used on low accuracy conversational speech data. The Meeting Assistant application allows the user to transcribe free speech directly from microphone, transcribe a meeting with multiple participants, save and retrieve saved transcriptions and recordings. During the transcription, keyphrases are extracted; negative and positive sentiments are detected and marked. We use also third party software to translate the transcript of the conversation, and to display the extracted keyphrase as a word cloud.

The paper is organized as follows: we describe briefly the speech and NLP technologies in Section 2; in Section 3, we present how these technologies have been integrated into the Meeting Assistant.

2. Speech and NLP Technologies

In this section, we describe briefly speech and NLP technologies used by the Meeting Assistant.
Sentiment Detection: sentiment analysis is an NLP application that aims to identify and extract subjective information, namely attitudes and opinions from textual documents. The input to the sentiment analysis module in the Meeting Assistant is a textual document that originates from automatically transcribed spoken human to human meeting interaction. The quality of sentiment analysis algorithms in terms of recall and precision is highly dependent on the quality and coverage of the sentiment lexicon. The sentiment lexicon is essentially a list of sentiment words and phrases along with their sentiment polarities. The generation of a sentiment lexicon is a delicate task. Some sentiment terms may convey positive opinion in one topical domain but neutral or negative opinion in another topical domain. That is, lexicon-based sentiment analysis is sensitive to the topical domain it operates in. The variety of topics that may be conveyed in human to human meetings is large. In order to achieve high sentiment analysis quality in every domain, there is a need to generate a domain-specific sentiment lexicon per domain. Manual acquisition of sentiment lexicons is a costly labor-intensive task and therefore impractical for the industry. Semi-supervised sentiment lexicon adaptation methods, such as the method proposed by [6], enable cost-effective generation of domain-specific sentiment lexicons. The Meeting Assistant uses such methods for achieving high precision and recall percentages of sentiment classification across different domains.

3. Integration
We show in Figure 1 the User Interface of the application. The Meeting Assistant transcribes the audio data and the transcript is displayed along with the NLP metadata. Keyphrases are marked in bold; sentiments are marked – red for negative, green for positive; the summary based on the extracted keyphrases is available. On the right-hand side, a word cloud, based on WordCloudControl1, displays frequent extracted keyphrases.

Translation service is provided by MS Translator2 cloud service. The application filters the transcript according to the NLP metadata, e.g., display only detected sentiments. If multiple speakers are recorded on the same audio stream, the speaker diarization process labels the transcribed speech according to its speaker. If multiple clients participate to the meeting, a client-server model is implemented for synchronization. Microsoft Lync is our meeting provider. Each client records its speakers. The new transcript with its NLP metadata is distributed to all the clients connected to the meeting and is displayed according to its timestamp.

4. References

1http://www.codeproject.com/Articles/224231/Word-Cloud-Tag-Cloud-Generator-Control-for-NET-Win
2http://www.microsoft.com/translator/