Abstract

According to recent statistics, dementia in its various forms afflicts more than 7.5 million people in the United States, with Alzheimer's disease accounting for 60-80% of these cases. Doctors diagnose Alzheimer's and other types of dementia through a examination that includes assessing cognitive function as a key indicator, but the disease is often diagnosed late. Early intervention could improve quality of life helping alleviate the symptoms and slow down the progression of the disease.

This research aims to identify auditory biomarkers which could be used for early detection of dementia. As a base for our analysis we used audio data collected from the Framingham Heart Study cohort during neuropsychological testing exams and the associated high fidelity dementia status reports. Acoustic features have been computed on the raw audio, and language-based features have been computed over the annotated text obtained through both automatic and manual transcriptions. For the prediction task we trained a random forest classifier and used mean area under the receiver operating curve (AUC) scores across a 10-fold cross validation scheme as a performance metric.

Despite some issue processing the data due to the records' age and low quality audio, the results obtained are very encouraging: a model based only on health and demographic features obtained a 0.52 AUC; using simple acoustic features boosted the AUC to 0.81; and including part-of-speech features allowed the model to reach an AUC of 0.91.

1. Summary

The Framingham Heart Study (FHS) is a longitudinal, ongoing cohort study initially involving residents from the town of Framingham, Massachusetts. The study began in 1948 with 5,209 adult subjects from Framingham and an initial focus on cardiovascular disease and stroke. It has since expanded to an epidemiologic study of common chronic diseases, including dementia, and in 2002, enrolled its third generation of participants. FHS participants undergo routine health exams about every 4 years where updated demographic, lifestyle and other health-related behavior informations are collected and a battery of medical tests is performed. Starting in 1999, administration of a Neuropsychological Test Battery (NP) was offered to surviving members of the two oldest cohort about every 5 years; testing was done more frequently if they have shown signs of dementia or have had a stroke. NP testing of the third generation began in 2009. Starting in 2005 audio was recorded during most NP's creating a database of more than 7,000 sessions.

Unfortunately, not all of those records were usable during our research due to logistical, technical, and legal issues, which limited the availability of medical records for certain individuals. For example, the format of the digital medical record changed over the years and to maintain a high precision during data parsing (e.g., to extract age) we had to sacrifice some recall. Similarly, the sessions' audio was recorded using a mono tape recorder with a sub-optimal placement and later compressed digitally using lossy algorithm. This greatly complicated its analysis. State-of-the-art diarization tools failed and even well-known text-to-speech automated transcription services such as IBM Watson obtained low precision.

Some audio features have been computed over the entire audio sessions (which included the proctor) and others on the part where only the subject is speaking. Voicing probability techniques have been used to remove extended periods of silence. Spectral and temporal characteristics of a subject's voice (e.g., pitch) have been computed over 20ms frames of the audio and statistically aggregated (e.g., computing mean, standard deviation) to generate the final acoustic feature set. Other quantitative speech features were computed on the automatically transcribed audio files (e.g., number of words per response, average length of pauses between words).

When an accurate transcription was available, natural language processing (NLP) techniques have been applied to evaluate the complexity of the responses. Using well-known sentence parsers we broke down the subject's responses into a tree of production rules and computed features based on the frequency of occurrence of the most common ones. In addition, qualitative measures such as the fraction of nonverbal breaks (e.g., filler words, laughing,

crosstalk) over the total number of words spoken by each participant was computed and used as features.

Since our goal was to predict the dementia status of the participant as identified by the consensus dementia diagnosis panel, we posed the machine learning task as a classification problem where the model tries to identify the participant's cognitive status using the acoustic and lexicographical features computed as input. We trained a random forest classifier and used mean area under the receiver operating curve (AUC) scores across a 10-fold cross validation scheme as a performance metric. AUC score, ROC curve, and feature importance were averaged over the ten folds, and a rough estimate of 95th confidence intervals was computed using the t-distribution.

Due to the difficulties in obtaining accurate medical records and transcriptions, not all features were available for the entire dataset. For this reason, the dataset was split in 3 groups and various methodologies were applied and tested individually. In general, health and demographic information alone performed well in groups of diverse age (since dementia is very infrequent among young adults) but have a low predicting power (e.g., AUC 0.52) when used to distinguish the cognitive status among older subjects. Models based on acoustic features (AUC 0.81) performed slightly better than models based on quantitative speech features which also require automated transcription. Production-tree features based models performed best, achieving an AUC score of 0.91.

While the results of this project are encouraging, it is important to repeat the experiments with a larger dataset of higher quality medical and audio samples. Medical records should be obtained already in digital form and correctly labeled possibly from Electronic Medical Record systems. Similarly, subjects' voices should be captured in high-fidelity and saved directly in a lossless digital format. Finally, supplementing the data with other data sources (e.g., minute level heart rate and steps) is also highly recommended for future prospective studies since Alzheimer's disease is known to affect sleep. Monitoring subjects' nightly patterns could add useful signal to the system.