Information Retrieval Performance on Automatic Speech-To-Text Queries

Max Robinson   MAX.ROBINSON@JHU.EDU

Johns Hopkins University,
Baltimore, MD 21218 USA

1. Introduction

In modern day devices, speech-to-text is becoming a more and more common tool. While advances in speech-to-text have come a long ways over the past few years, it is still not perfect. Since these systems are not perfect and still have errors, it is likely these errors will be translated into the areas in which the translated text is used. One of those areas is in user queries of information retrieval systems.

Automatic Speech-To-Text systems also have errors different than user typing a query. Where a user might be likely to misspell a word, automatic speech-to-text systems might replace entire words with different words. Understanding the effect that these errors can have on the performance of IR systems could be crucial to building better IR systems in the future.

2. Background

In a typical IR system, errors typically are caused by a human mistyping a word in a query resulting in a spelling mistake. Spelling mistakes have been studied for sometime and there are existing solutions such as edit distance, context sensitive spelling corrections, or phonetic corrections shown in books like Manning's "Introduction to Information Retrieval".

There has also been research into detecting when ASR makes errors. Techniques such as using Decision Trees or Naïve Bayes classifiers have been tried (Zavareh, Zukerman, Kim, & Kleinbauer, 2013). However, little research has been done regarding how these errors can effect IR systems.

3. Data Sources

The proposed document data source to use is the 1.3 GB test collection (CDS14) sample from the TREC 2014 Clinical Decision Support (TREC, 2014) task, provided for programming assignment 3 of this course. The baseline queries for performance measurements will be the queries provided along with this sample of data.

The speech-to-text queries will be generated by using a Mozilla Foundation implementation of DeepSpeech (Mozilla, 2014) with their pre-trained English language model. I will speak the same queries as described above, and record the text generated from the speech-to-text translation. These generated queries will then become the test dataset.

4. Experimental Design

The experiment proposed is as follows. Documents from the corpus will be indexed into an open source search engine, Elasticsearch, to run queries against. Documents will be indexed twice. One will be into an index that uses word term indexing, and another will use character 3-gram indexing for terms.

Once the data is indexed, the queries provided for the corpus will be run against Elasticsearch, on both indices, retrieving the top 100 documents for each query. These top 100 documents will become the "ground truth".

After the ASR translated queries have been created, each query will be run against Elasticsearch indices to produce the top 100 documents for each query. This data is the result data from the automatic speech-to-text queries.

To evaluate the performance of the ASR queries, the top 100 documents for the ground truth and ASR queries will be compared. The metric calculated will be the precision of the ASR queries compared to the ground truth, or the number of documents in common between the ground truth and the ASR queries over 100. This will be done for both the word indexing and the character 3-gram indexing.

The queries will then be compared to calculate the word error in the ASR query. This word error will then be analyzed with the performance of the ASR queries to look for correlations.

References

Manning, C. D., Raghavan, P., & Schtze, H. (2008). *Introduction to Information Retrieval* (1 edition). Cambridge University Press.

Mozilla (2014). https://github.com/mozilla/deepspeech..

TREC (2014). http://www.trec-cds.org/2014.html..

Zavareh, F., Zukerman, I., Kim, S. N., & Kleinbauer, T. (2013). Error detection in automatic speech recognition. *Proceedings of Australasian Language Technology Association Workshop*, 101–105.