Data Visualization By Voice: Syntax, Parsing, and Recognition Techniques

Maurice Diesendruck
Department of Statistics and Data Sciences
University of Texas
Austin, TX
momod@utexas.edu

Honghe Zhao
Department of Mathematics
University of Texas
Austin, TX
joehonghe@utexas.edu

February 16, 2016

Abstract

This research identifies techniques that enable a computer system to perform automated data visualization by actively "listening" to a user's natural spoken language, in an interactive and real-time session. This work coins the name *ggspeak* as the "grammar of graphics" for speech, and presents a Python library that incorporates speech recognition and a domain-specific entity extraction algorithm that respects and resolves errors of mistranscription.

1 Introduction

As mobile and wearable computers become more powerful, interactions with such devices are likely to become more frequent, more personal, and more casual in style. Speech, in contrast to typing, presents an appropriate and comfortable way to communicate with such devices.

With computing becoming increasingly voice-driven, so too should applications and interfaces for data manipulation and visualization. Today, one can visualize data using (among other things) Python's matplotlib, R's ggplot, or Microsoft's Excel; but these interfaces typically require long work-flows of option selection and formatting, and can include cumbersome syntax. To the best of the researcher's knowledge, there is no other publicly available system that targets a fluid, interactive, voice-driven data visualization interface.

The closest related works by Harada et al. (2007), and Levin and Lieberman (2004), utilize speech to produce animations and other art.

Project code for the system mentioned here is made available on the author's GitHub page at `https://github.com/diesendruck/ggspeak`.

2 Application

2.1 Speech Recognition

The system utilizes the popular Python module SpeechRecognition (Zhang, 2016) to manage audio capture and transcription, and currently runs locally on the user's laptop.

2.2 Architecture

The user first selects a comma-separated values (CSV) file to be the data for the session. The system then initializes a Graphic object, which will contain all relevant details for the graph. At this point, the system begins to listen, and at each utterance, evaluates whether words relating to graphical syntax are present. For example, did the words "scatterplot" or "histogram" or "line graph" appear? Similarly, did the user say any names that match the CSV column headers?

Meanwhile, with each completed utterance, the system determines whether the user has asked to quit, and if not, whether the current graph details produce a valid graph. The interaction begins when the system produces a valid graph that can be further edited by the user in subsequent spoken commands.

2.3 Disambiguation and Mistranscription

While the scope of graphing vocabulary and syntax may seem small, it is difficult to generalize the task of recognizing header names. Homophones, rhyming words, and non-English concatenations will typically be mistranscribed. Consider, for example, the headers "X", "Y", and "Under10". A user that says "plot X versus Y and group by Under10", may produce a transcription of "plot ex versus why and group by under ten".

To resolve this, the researchers propose a fuzzy comparison or distance metric between phonetic encodings, e.g. distance between Soundex or Metaphone encodings.

References

Anthony Zhang. 2016. Speech Recognition (Version 3.1) [Software]. Available from `https://github.com/Uberi/speech_recognition#readme`.

Golan Levin and Zachary Lieberman. 2004. *In-Situ* Speech Visualization in Real-Time Interactive Installation and Performance. *Proceedings of The 3rd International Symposium on Non-Photorealistic Animation and Rendering*. Annecy, France.

Susumu Harada, Jacob O. Wobbrock, and James A. Landay. 2007. Voicedraw: a hands-free voice-driven drawing application for people with motor impairments. *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*. Tempe, Arizona, USA.