Technical Disclosure Commons

Defensive Publications Series

June 09, 2017

Conversational System Responses

Aleksandar Kracun

Austin Waters

CONVERSATIONAL SYSTEM RESPONSES

ABSTRACT

The disclosed system and method match verbosity of a machine's responses to the verbosity or brevity of the user's query. The system includes a semantic parser connected to an audio input and output device to process and respond to queries. The method includes building a simple statistical model of mean (M) and standard deviation (SD) for the lengths of the audio of the user’s query utterances and the multiple variants of the generated text-to-speech (TTS) output for each query type or action or intent. The system may extract fluff or slot filling words from verbose queries to use in formulating the response. The system then matches the M and SD between the query and the response to pick an optimum response. The system is more conversational and more dynamically reactive to the user's input. This system retrieves and presents relevant information faster and may be more user-friendly for accessibility users.

BACKGROUND

Currently, when the users interact with conversational systems, the responses presented are uniformly verbose and are not tailored to the expectations of the user. For example, if a user requests the present time of the day from a conversational system by issuing either the utterance "What time is it now?" and "Time, now!” the system presents the same long phrase "The time is now 11:39AM". While the response is valid for both queries, the second utterance may have received a less verbose response, such as "11:39AM". A parser may subtract needless words from queries to find matching actions. Short form or curt responses are ubiquitous in natural conversational speech and is required to be mirrored in artificial agents. Fluffy sentences may be seen as hints that the user may be expecting verbosity in the response.

DESCRIPTION

A system and method are disclosed that matches verbosity of a machine's responses to the verbosity or brevity of the user's query. The system includes a semantic parser connected to an audio input and output device to process and respond to queries. The method includes building a simple statistical model to rank the multiple variants of the generated text-to-speech (TTS) output for each query type or action or intent, given a specific utterance. The system extracts features from the query and uses the statistical model to pick an optimum response. One implementation would be to use the mean (M) and standard deviation (SD) as statistical parameters in the statistical model. The system may also extract fluff or slot filling words from verbose queries to use in formulating the response. The system then matches the statistical parameters between the query and the response to pick an optimum response as shown in FIG. 1.

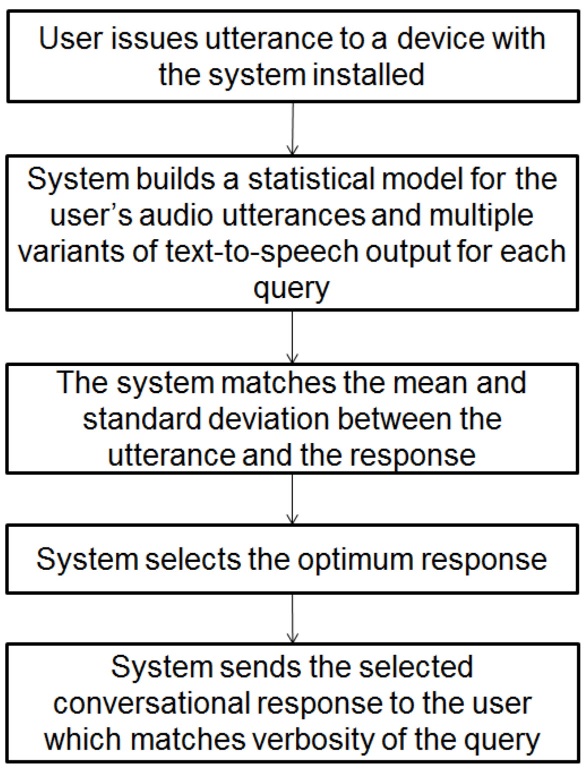

FIG. 1: Method to present conversational responses by the system

For example, when the user utterance is either “What time is it now?” or "Time, now!” the verbose response presented by the system may be “The time is now 11:39AM" for the first utterance. The curt response presented by the system may be "11:39AM" for the second utterance.

An alternative implementation of the disclosed system may match the number of "fluff" words in the grammar of the input query with the number for the response. Another implementation may train an end-to-end system that ranks a database of responses and queries that have been manually labeled by linguistic experts.

The system is more conversational and more dynamically reactive to the user's input. Generally, users with disability listen to TTS output at the maximum rate in order to process information. This system retrieves and presents relevant information faster and may be more user-friendly for accessibility users.