July 03, 2017

Prediction and Selection of Sequence of Actions Related to Voice Activated Computing Systems

John D. Lanza
Foley & Lardner LLP

PREDICTION AND SELECTION OF SEQUENCE OF ACTIONS RELATED TO VOICE ACTIVATED COMPUTING SYSTEMS

Voice activated computing systems provide a user with content or services in response to voice commands spoken by the user. Such systems can capture voice commands from a user, process the voice commands to determine requests and keywords in the voice commands, and provide the user with content or services related to the determined requests and keywords.

As discussed herein, a voice activated computing system processes the voice commands to not only determine explicitly requested content and services, but to also predict likely content and services that may be useful to the user in the context of the explicitly requested content and services. The computing system can generate actions related to the predicted content and services, and execute these actions prior to the execution of the actions related to the explicitly requested content and services. For example, the voice activated computing system receiving a voice command "OK, I would like to go to dinner and then a movie tonight," can determine actions related to "dinner" and "movie." In addition, the system can predict an action related to "transportation from the movie to home," even though this action was not explicitly requested in the voice command. The system may predict that the most likely order of actions would be actions related to "dinner" first, actions related to the "movie" next, followed by actions related to "transportation from the movie to home." However, instead of executing actions in this order, the system can bypass the actions related to "dinner" and "movie" and execute an action related to the transportation form the movie to home. For example, the system can send the user an audio file stating "Would you like a taxi waiting for you after the movie?" prior to the execution of the actions related to "dinner" and actions related to "movie." In this manner, the system not

only can predict content and services that the user may need, but also alter the execution sequence of the actions related to the content and services.

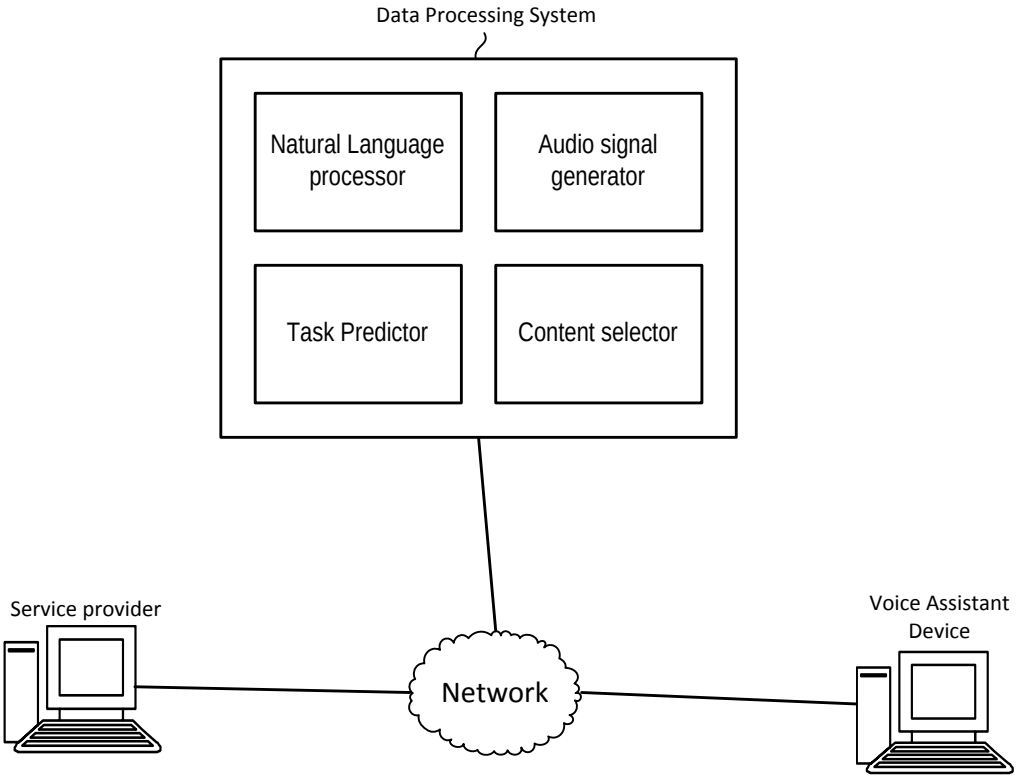

Figure 1

Figure 1 shows an example voice activated computing system. The system includes a voice assistant device, a service provider, and a data processing system communicating over a network. The voice assistant device can be a device accepts voice commands, and provides audio or visual output. The voice assistant can include one or more mics and cameras, such that voice commands received by the user are converted into corresponding audio signals. The voice assistant can send the audio signals to the data processing system and the service provider. The voice assistant device also can receive data such as audio signals or video signals from the data processing system or the service provider. The voice assistant device also can include audio

speakers that can convert the audio signals received from the data processing system or the service provider into sound.

The data processing system can process voice commands received from the voice assistant device. The data processing system includes a natural language processor, an audio signal generator, a task predictor, and a content selector. The natural language processor is capable of processing voice commands included in the audio signals received from the voice assistant device. The natural language processor can convert the audio signals into recognized text by comparing the audio signals against a stored, representative set of audio waveforms, and choosing the closest matches. The representative waveforms are generated across a large set of users, and can be augmented with speech samples. After the audio signals are converted into recognized text, the natural language processor can match the text to words that are associated, for example via training across users or through manual specification, with actions that the data processing system can serve. Basically, the natural language processor identifies requests and trigger words in the converted text, based on which the natural language processor can determine the content and actions to be carried out. The task predictor can predict tasks or actions based on the converted text, and in particular by identifying requests and trigger keywords in the converted text. The task predictor also can predict the most likely sequence in which the tasks would be executed. The content selector can select content, such as services to be offered to the user based on the actions identified by the task predictor. In addition, the content selector also can alter the sequence or the order in which the actions related to the services offered to the user are executed. The audio signal generator can generate audio signals based on the services selected by the content selector. The audio signals can be representative of voice responses or voice instructions provided to the user in response to the voice commands.

The service provider can provide one or more service to the user. For example, the service provider can be a taxi or car sharing service provider, dining or reservation service provider, and the like. The service provider can communicate with the voice assistant device independently of the data processing system and provide the user the ability to request a ride, do a dinner reservation, or avail of other services provided by the service provider. The service provider can also include a natural language processor, similar to the one discussed above in relation to the data processing system, to convert user voice commands into text, and identify requests and keywords to determine the services requested by the user.

Referring again to the voice command example mentioned above, the user can speak the voice command "OK, I would like to go to dinner and then a movie tonight," to the voice assistant device. The mics at the voice assistant device can convert the voice commands into corresponding audio signals, which are be transmitted by the voice assistant device to the data processing system over the network.

At the data processing system, the natural language processor processes the audio signal received form the voice assistant device and identify requests for "dinner" and "movie." The natural language processor also can identify a trigger keyword "go" or "to go to," which can indicate a need for transportation. Even though the user's voice command does not directly express an intent for transportation, the trigger keyword indicates that transportation may be needed.

The task predictor, based on the requests for "dinner" and "movie" and on the trigger keywords, can determine a most likely sequence of actions related to the voice command. For example, one sequence of action is shown below in Figure 2.

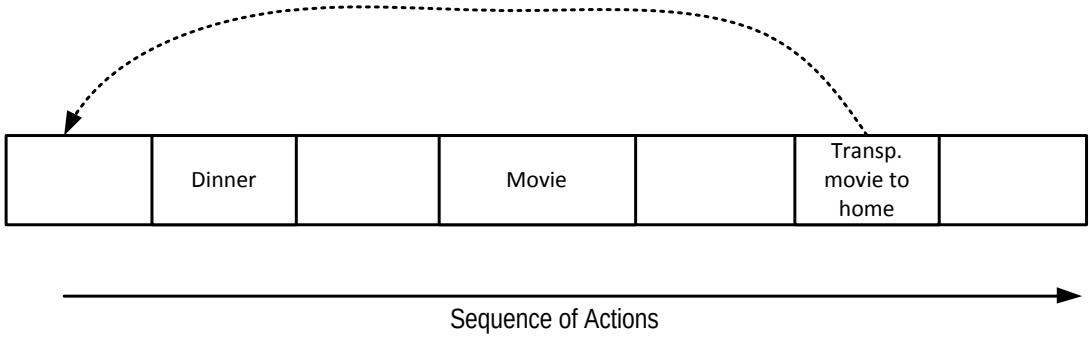

Figure 2

Figure 2 shows a predicted sequence of actions from time T-1 to time T-7. The task predictor determines that the most likely sequence of actions is actions related to "dinner," followed by actions related to "movie," and then actions related to "transportation from movie to home." The actions related to "dinner" can include making reservations at a restaurant, while the actions related to "movie" can include purchasing tickets to a movie. The action of "transportation from the movie to home," in turn, can include requesting a taxi from the movie to the user's home. While the task predictor predicts the above sequence of actions as the most likely sequence, the content selector can alter this sequence of actions. In particular, the content selector can execute actions related to "transportation from movie to home" prior to the actions related to "dinner." For example, the content selector can change the sequence of actions such that the user is send a voice request "Would you like a taxi waiting for you after the movie?" prior to sending voice requests regarding making a dinner reservation or purchasing movie tickets. While Figure 2 shows that the actions related to "transportation from movie to home" are executed prior to the actions related to "dinner," it is understood that the actions related to "transportation from movie to home" can be executed prior to any other action in the sequence of actions predicted by the task predictor. For example, the actions related to "transportation from

movie to home" may be executed after actions related to making dinner reservations but before the actions related to purchasing movie tickets.

In the predicted sequence of actions, the system would send voice requests to the user regarding "transportation from movie to home" after the movie has ended. Thus, the user would potentially spend time after the movie has ended to interact with the service provider and wait for the taxi to arrive. By executing the actions related to "transportation from movie to home," earlier in the sequence of actions, the taxi can arrive at the movie before or as soon as the movie ends, saving the user valuable time.

Abstract

This document describes a technique for processing voice commands in a voice activated computing system. In particular, the system processes the voice commands to not only determine explicitly requested content and services, but to also predict likely content and services that may be useful to the user in the context of the explicitly requested content and services. The computing system can generate actions related to the predicted content and services, and execute these actions prior to the execution of the actions related to the explicitly requested content and services.