March 27, 2019

Simplified voice-based interaction by inferring user intent

Victor Cărbune

Daniel Keysers

Thomas Deselaers

Simplified voice-based interaction by inferring user intent

ABSTRACT

Carrying out tasks using voice-based user interfaces often requires navigating through a long list of options unless the user happens to know the precise phrase required to reach the desired option directly. Each option can require an interaction turn until all information required to carry out the task of sending the message is specified. This makes the user experience of carrying out the task via voice UI slow and awkward. Alternatively, the user needs to remember to include all relevant information in the initial command which is a difficult, unintuitive, and awkward user experience. This disclosure describes techniques to reduce the user interaction needed to have a voice-based assistant respond to voice commands. With user permission, a voice-based assistant resolves aspects not explicitly specified in a user's request based on analysis of the user's past interactions to determine options that the user is most likely to choose.

KEYWORDS

- Voice assistant
- Voice UI
- Audio UI
- Voice command
- Spoken command
- User context
- User intent
- Virtual assistant
- Smart speaker

BACKGROUND

Carrying out tasks using voice-based user interfaces often requires navigating through a long list of options unless the user happens to know the precise phrase required to reach the desired option directly. For example, consider a user Alice who wants to send a message to her friend Bob that she is on her way to his place. Alice may ask a voice-based assistant to "Send a message to Bob." However, such a request from Alice needs further clarification regarding which among multiple contacts named Bob in her contact list is to the intended recipient, which communication app is to be used for sending the message, which of Alice's user accounts should be used, what the message is, etc. Each clarification can require an interaction turn with the voice assistant until all information required to carry out the task of sending the message is specified. This makes the user experience of carrying out the task via voice UI slow and awkward. Alternatively, the user Alice needs to remember to include all relevant information in the initial command which is a difficult, unintuitive, and awkward user experience.

DESCRIPTION

This disclosure describes techniques to reduce the user interaction needed to have a voice-based assistant, e.g., provided via a smart speaker, smartphone, or other device, respond to voice commands. The techniques are based on the observation that most people communicate with only a small list of contacts with whom they communicate regularly, and they typically use the same communication channel when interacting with a specific person. As a result, if the user permits, a voice-based assistant resolves the aspects not explicitly specified in a user's requests based on analyzing the user's past interactions to determine the options the user is most likely to choose for these aspects along with a corresponding likelihood for each choice.

For example, if a user Alice issues a voice-based command: “Tell Bob that I am on the way,” then based on Alice’s conversation history (accessed with her permission), the voice-based assistant can infer that Alice most likely wishes to send a text message to Bob Smith from her personal phone informing him of her departure from work. The voice-based assistant can then seek confirmation from Alice with a message such as: “Do you wish to text Bob Smith from your personal phone that you have left work and are on your way to his place?” If Alice confirms, the task is performed. Otherwise, the interaction between Alice and the voice-based assistant falls back to the default multiple step interaction to seek clarification and determine the exact task to be performed, as described earlier.

The intended user action can be determined via a machine learning model, using data accessed with user permission. For example, the model can be based on a neural network that evaluates user-permitted relevant historical and contextual information, such as the user’s past communications, current time and day, location, the user’s recent actions, etc. With user permission, a simple model can be based only on analyzing past user actions while a complex model can be based on multiple factors as mentioned above. Alternatively, the described techniques can be implemented by automatic matching based on templates and rules. In such a case, all templates and rules need not be available at the outset, as these can be inferred during system operation and added to the existing set as appropriate.

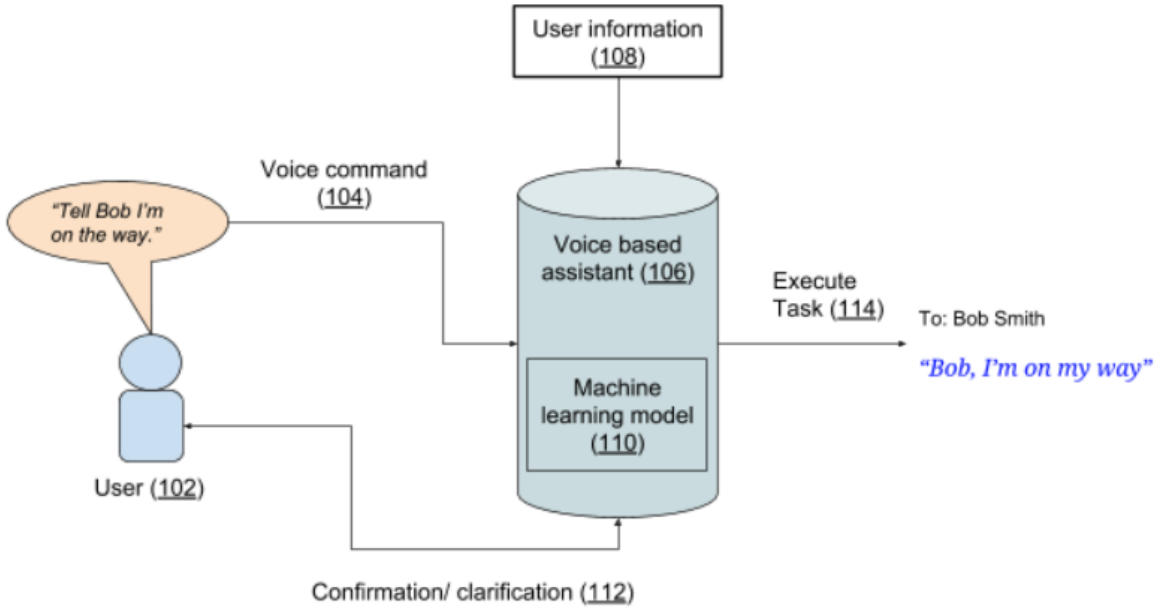

Fig. 1: Example operation

Fig. 1 shows an example implementation of the techniques described in this disclosure. A user (102) issues a voice command (104) to a device (106) that provides a voice-based assistant user interface. A machine learning model (110) is used to analyze the voice command and infer the user's intended task based on historical and contextual user information (108), accessed with the user's permission. If the user permits, such information can be obtained from the user's devices, the user's accounts with relevant service providers, device sensors, etc.

The inferred task is then presented to the user to seek confirmation (112). If the user confirms that the inference is correct, the task is executed on the user's behalf (114). In the example illustrated in Fig. 1, the user's message "I'm on my way" is sent to Bob Smith. Alternatively, if the user's intent cannot be determined with a certainty level that meets a threshold, or if the user rejects the inferred intended task, one or more clarification steps (112) are performed until all aspects required to determine the user's intended task have been specified.

Whether to infer the user's intent and seek confirmation instead of engaging in the typical multiple step interaction to capture the user's full intent is decided based on the likelihood associated with the inferred intent. If the likelihood of the inferred intent does not meet a specified threshold, the probability that the user will reject the inferred intent is high. In such cases, the default mode of multiple interactive steps is used to select the gather the correct task options from the user.

For example, in a messaging context, if a user is new or lacks sufficient communication history with the recipient, there may not be enough information available to determine the user's intent with a sufficiently high likelihood. As a result, the default multiple-step interaction with several clarification questions is used to handle the voice requests made by such users. The default interaction is also used if the user denies access to the user's past communication history and/or other relevant user information.

In contrast, the likelihood of inferring the user's intended action is high for a user who has engaged in similar communication patterns repeated and allows access to such information. For such users, the faster interaction based on intent inferred as described in this disclosure can be used.

The threshold values of likelihood used to determine whether the voice-based assistant is to use the inferred user intent for the task can be specified by the software developers, device manufacturers, and/or users. Alternatively or in addition, the threshold values can be computed dynamically during system operation, e.g., based on the success rate of past inferences.

Application of the techniques described in this disclosure can reduce the number of voice-based interaction steps needed to determine and execute a task. This enhances the user experience of voice-based assistant applications. In case the inference does not match the user's

intent, a single extra step is needed to fall back to the default multiple-step voice interaction, thus keeping overhead to a minimum.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user's social network, social actions or activities, profession, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure describes techniques to reduce the user interaction needed to have a voice-based assistant respond to voice commands. With user permission, a voice-based assistant resolves aspects not explicitly specified in a user's request based on analysis of the user's past interactions to determine options that the user is most likely to choose. Such inference can be performed by a machine-learning model that utilizes user permitted information such as the user's past communications, current time and day, location, the user's recent actions, etc. Application of the techniques described in this disclosure can reduce the number of voice-based interaction steps needed to determine and execute a task. This enhances the user experience of voice-based assistant applications.