March 11, 2019

Resolving user queries using contextual information

Marcos Calvo

Karol Stosiek

Resolving user queries using contextual information

ABSTRACT

When responding to a request from another person, human beings take into account non-verbal cues and relevant implicit contextual information to determine an appropriate response. In contrast, digital assistants do not currently take such contextual information into account. Thus, a digital assistant may provide a sub-optimal or contextually inappropriate response to the user's request, or require the user to modify the original request iteratively until the user's full intent is communicated. If the user permits, the techniques described in this disclosure enable a digital assistant to infer the user's intent by using device and environmental sensors to extract available contextual information, e.g., noise level, the user's tone of voice, the number of people in a room, objects in the room, temperature, time, location, weather, traffic, etc. The information is used, as permitted by the user, to determine an appropriate response to the user's request.

KEYWORDS

- digital assistant
- virtual assistant
- smart assistant
- query context
- user context
- context detection
- smart speaker
- environmental sensor

BACKGROUND

When responding to a request from another person, human beings take into account non-verbal cues and relevant implicit contextual information to determine an appropriate response. For example, when asked to play music, a human would make musical choices that fit the context, such as yoga, party, coffee with friends, etc. To that end, humans often consider implicit contextual cues including but not limited to: the number of people and objects in the vicinity of the requester, the tone of voice used when making the request, the status of surrounding devices and appliances, such as the oven, thermostat, etc., and so on.

In contrast, digital assistants do not currently take such contextual information into account. Thus, a digital assistant may provide a sub-optimal or contextually inappropriate response to the user's request, or require the user to modify the original request iteratively until the user's full intent is communicated.

DESCRIPTION

The techniques described in this disclosure enable a digital assistant, such as a voice-based assistive device or application, to infer the user's intent for a query by using available contextual information, if permitted by the user. With user permission, such contextual information is obtained from various sensors, such as a microphone, a camera, a thermometer, etc. The sensors may be present within a user's mobile device, such as a smartphone, or embedded within the user's physical environment.

When the user makes a request, data captured by the sensors is analyzed to extract relevant information regarding the context that can help infer the user's intent in making the request. If the user permits, such information can include, e.g., noise level in the room, the user's tone of voice, the number of people in the room, the objects in the room, the temperature of the room, time, location, weather, traffic, etc.

With permission from the user, a digital assistant can use relevant contextual information to resolve ambiguities in the user's request and/or infer the user's intent in cases where it may not be explicitly stated. The inference is then used to determine the appropriate response to the user's request. For example, with the permission of the people involved, the request to "play some music" can be served by playing relaxing music when only one person is in the room with a yoga mat on the floor, and by playing party music when the room is full of people with a high noise level. Similarly, with the permission of the people involved, if the user asks to "play a game," an educational game can be proposed in a family setting while a party-themed game can be recommended for a gathering of adults. As another example, if the user permits the use of contextual information, the query "what can I cook?" is answered with lunch recipes in the morning, and more sophisticated meal suggestions in case the user's calendar indicates that the user is expecting guests for dinner.

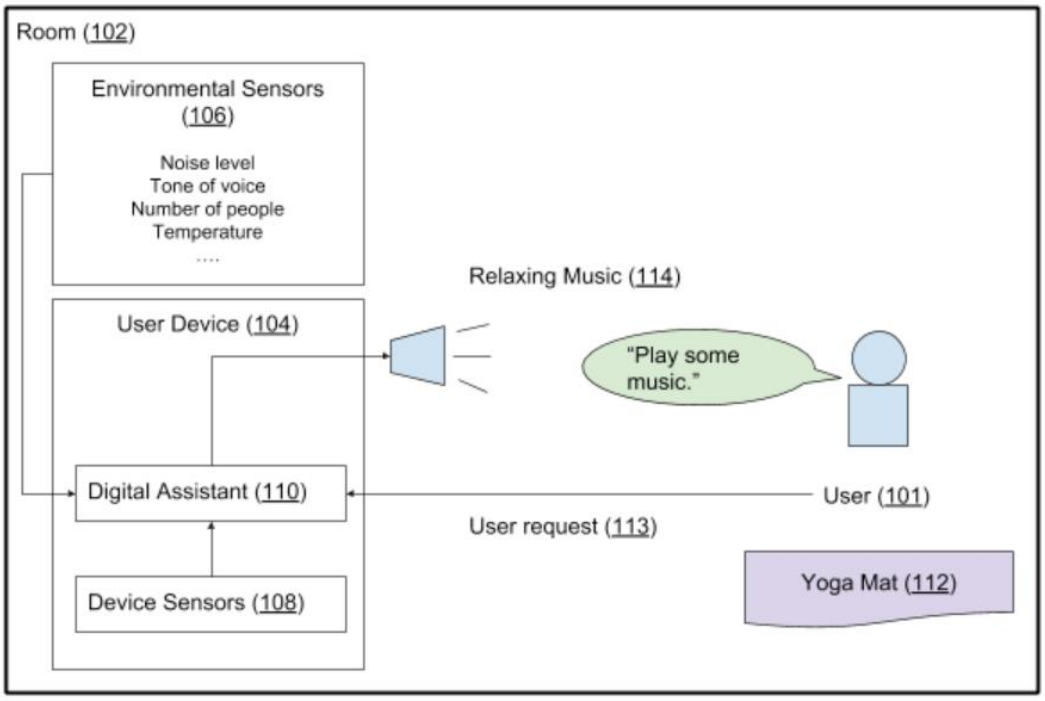

Fig. 1

Fig. 1 shows an operational example using the described techniques. A user (101) is alone in a room (102) and about to do yoga on a yoga mat (112). Prior to starting yoga, the user issues a request (113) to "play some music" to digital assistant (110) provided via a user device (104). If the user permits, the digital assistant obtains input from various device sensors (108) and other sensors in the room (106). The various environmental sensors in the room can detect contextual factors such as, e.g., noise level, tone of voice, number of people, objects present in the room, temperature, clothing worn by the user, etc. The device sensors can include microphone, camera, etc. as well as relevant online information sources, e.g., for information such as the weather.

Input from the device and environmental sensors is analyzed to determine relevant contextual information that associates the user's request with the user's context (e.g., performing yoga). Based on this determination, it is inferred that the user's request is to play music appropriate for yoga, and the digital assistant provides playback of relaxing music (114).

The described techniques can be utilized in any device or application that incorporates digital assistant features. With user permission, the techniques of this disclosure enable implementation of digital assistants that can provide answers or perform actions tailored to a user's specific situation and intent, thus enhancing the user experience and efficiency of interacting with such assistive technologies and improving the quality of the responses. Taking relevant contextual information into account with the user's permission can help the digital assistant determine contextually appropriate responses to ambiguous queries without requiring the user to engage in a clarification dialog or to reformulate the request or to reduce the need for further clarifications from the user. For example, if the user is detected to be wearing sports clothes, and is known to do two types of exercises (cardio + yoga), a much simpler clarification

question can be asked, e.g., “music for cardio or yoga workout?” instead of "what type of music?".

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

When responding to a request from another person, human beings take into account non-verbal cues and relevant implicit contextual information to determine an appropriate response. In contrast, digital assistants do not currently take such contextual information into account. Thus, a digital assistant may provide a sub-optimal or contextually inappropriate response to the user’s request, or require the user to modify the original request iteratively until the user’s full intent is communicated. If the user permits, the techniques described in this disclosure enable a digital assistant to infer the user’s intent by using device and environmental sensors to extract available contextual information, e.g., noise level, the user’s tone of voice, the number of people in a room, objects in the room, temperature, time, location, weather, traffic, etc. The information is

used, as permitted by the user, to determine an appropriate response to the user's request. The described techniques can be utilized in any device or application that incorporates digital assistant features. With user permission, the techniques of this disclosure enable implementation of digital assistants that can provide answers or perform actions tailored to a user's specific situation and intent, thus enhancing the user experience and efficiency of interacting with assistive technologies.