Technical Disclosure Commons

Defensive Publications Series

January 23, 2019

Voice interface responses based on prior user interactions

Minkoo Seo

Brian Spilsbury

Gadi Ben Amram

Joo Hyung Lee

Togan Gencoglu

See next page for additional authors

Inventor(s)
Minkoo Seo, Brian Spilsbury, Gadi Ben Amram, Joo Hyung Lee, Togan Gencoglu, and Youngchan Park

Voice interface responses based on prior user interactions

ABSTRACT

Voice interfaces, such as those provided by virtual assistant applications on smartphones, smart speakers, and other devices utilize natural language when providing responses to user queries. Per techniques of this disclosure, user interaction patterns are used to generate natural language responses to queries posed over a voice user interface. With user permission and express consent, data that pertain to prior user interaction, e.g., text entry such as search terms entered; other input operations such as clicks, taps, swipes, mouse hovers; viewports for the user, etc. are obtained. For ease of interaction, such permission is obtained, e.g., at initial setup, and is modifiable. Such data is analyzed to derive insight into context-based topics of interest to users. User preferences gathered across multiple contexts are utilized to generate natural language responses for voice-based user interaction.

KEYWORDS

- Virtual assistant
- Voice assistant
- Voice UI
- Personalized UI
- Spoken response
- Natural language
- User interaction
- Smart speaker
- Home speaker

BACKGROUND

Natural language generation techniques are utilized in virtual assistant applications to produce text-to-speech (TTS) responses for user queries. For example, a user query regarding weather can be interpreted as a request for information about any of multiple weather data elements such as a predicted highest or lowest temperatures of the day, sky condition, current temperature, current humidity, hourly forecasts, daily forecasts, etc.

Generation of a natural language response to user queries can be based on an approach that uses a predefined template. For example, utilizing such templates provides a similar response at each instance of the query, and can include a subset of available information, e.g., local sky conditions, high/low temperatures, and a current temperature. An alternative approach is to provide a response that includes all available information pertaining to the query. In another approach, heuristics are utilized to prioritize certain pieces of information in the response. For example, if rain or snow is expected, information relating to the rain or snow is prioritized and presented earlier in the response.

However, the above approaches do not take into account user preferences. When utilizing a graphical user interfaces (GUI), users can navigate relatively easily to content of their interest (for example, by scrolling and/or swiping). However, when a query response is provided via voice UI, users are presented information in a serial manner. When the sequencing is not personalized, users sometimes have to listen to information of less interest or relevance before being presented with the most useful information.

DESCRIPTION

This disclosure describes generation of text to speech (TTS) responses personalized for the user. Per techniques of this disclosure, user interaction patterns, e.g., determined from user

interaction with a page of search results delivered via a graphical user interface, are used to generate voice responses to user queries.

With user permission and express consent, data that pertain to prior user interaction, e.g., text entry such as search terms entered; other input operations such as clicks, taps, swipes, mouse hovers; viewports for the user, etc. are obtained e.g., from user interaction with search results, websites, applications on a user device, etc. For ease of interaction, such permission is obtained, e.g., at initial setup, and is modifiable. User interaction data is categorized using parameters such as location, time, user model data, etc. Data can be obtained from a large number of users and combined to determine context-based topics of interest to users.

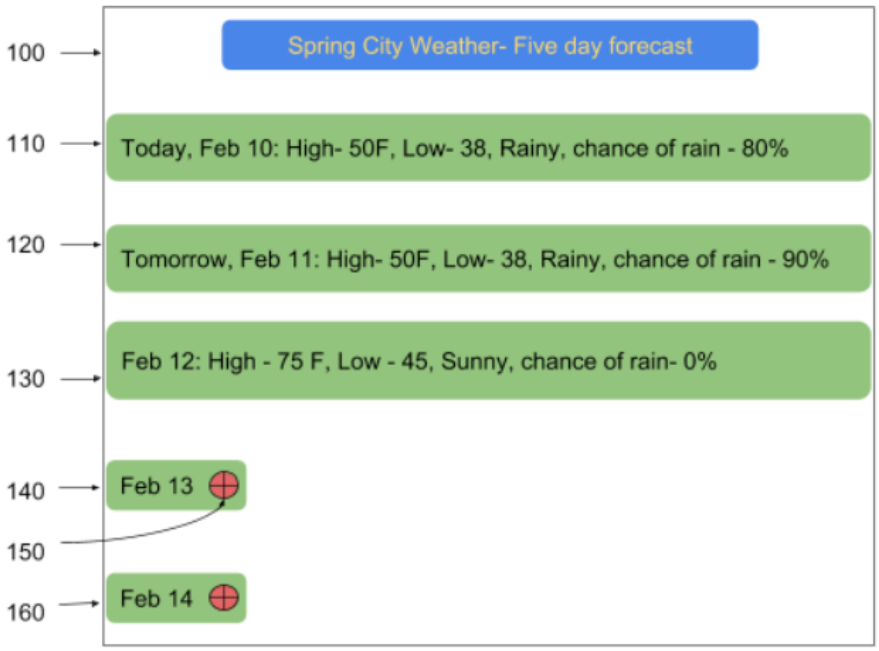

Fig. 1: User interaction with content provides insight into user preferences

Fig. 1 illustrates an example of user interaction with a weather-related website (100). In this illustrative example, a user views a weather forecast of a current day (110) and two following days (120, 130) by clicking on (navigating to) corresponding links provided on the

site. In this illustrative example, the user does not proceed to view content for additional days (140, 160) beyond the third day (February 12, in this example) and does not utilize the navigation button (150) to access the expanded information available for February 13.

When users consent to use of such user interaction data, it is inferred that an important weather feature of interest of the user is the presence/absence of rainy days. A user interaction preference may be stored that is indicative of this feature of interest. User preferences gathered for multiple contexts and from a large number of users are utilized to design and generate natural language responses during subsequent voice-based user interactions.

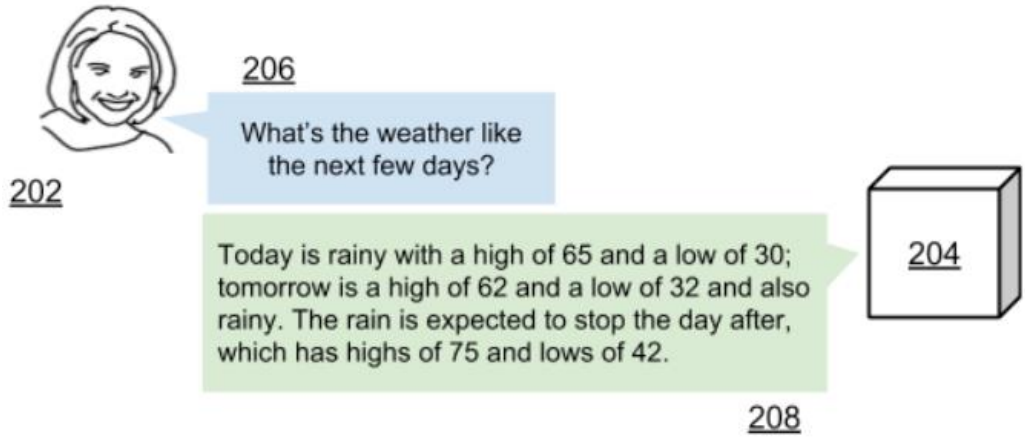

Fig. 2: Providing natural language responses to queries

Fig. 2 illustrates an example of a user (202) interacting with a device (204), e.g., a smart speaker, a smartphone, or other device. The user asks a question about the weather (206). The inference regarding user interest in rain and the stoppage of the rain as described above is utilized to generate a natural language response (208), "Today is rainy with a high of 65 and a low of 30, tomorrow is a high of 62 and a low of 32 and also rainy. The rain is expected to stop the day after, which has highs of 75 and lows of 42."

In another example, a user is observed to pose a query using a text-based interface regularly at a certain time, e.g., at or near 7 am, regarding the weather forecast for two different locations (a home location, and a work location, for example). Such query patterns are analyzed to infer that a user preference is to obtain a forecast of the weather at the two locations every morning. In this example, when the user asks the same question via a voice-based UI, the provided response is tailored to include both home and work locations, e.g., "It is sunny today; carry an umbrella as rain is expected near your office at lunchtime."

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

Voice interfaces, such as those provided by virtual assistant applications on smartphones, smart speakers, and other devices utilize natural language when providing responses to user queries. Per techniques of this disclosure, user interaction patterns are used to

generate natural language responses to queries posed over a voice user interface. With user permission and express consent, data that pertain to prior user interaction, e.g., text entry such as search terms entered; other input operations such as clicks, taps, swipes, mouse hovers; viewports for the user, etc. are obtained. For ease of interaction, such permission is obtained, e.g., at initial setup, and is modifiable. Such data is analyzed to derive insight into context-based topics of interest to users. User preferences gathered across multiple contexts are utilized to generate natural language responses for voice-based user interaction.