June 2020

Providing Contextual Recommendations Based On Prior Interactions Across Apps

Victor Carbune

Matthew Sharifi

Providing Contextual Recommendations Based On Prior Interactions Across Apps

ABSTRACT

A user's interaction with a virtual assistant is typically limited to the context of a single app. As such, the context and history of a user's interaction with the virtual assistant within one app is completely isolated from the context of history of the user's interaction with the virtual assistant when using another app. Such compartmentalization reduces the seamlessness with which users can use a virtual assistant for tasks and recommendations that span across multiple apps. With user permission, the techniques described in this disclosure integrate the knowledge of user interaction across apps and contexts. The information in turn enables enhancement of content or action recommendations.

KEYWORDS

- Contextual recommendation
- Proactive suggestion
- Interaction data
- Interaction history
- On-screen content
- Content understanding
- Virtual assistant
- Entity extraction
- Topicality score
- Cross-app operation

BACKGROUND

People typically use virtual assistants for a variety of purposes across a number of apps on their devices. Typically, virtual assistant functions can be invoked using voice input or other mechanisms. To support such mechanisms, virtual assistants need to integrate with various apps on the device and provide interaction support at the device operating system (OS) level.

However, a user's interaction with the virtual assistant is typically limited to the context of a single app. As such, the context and history of a user's interaction with the virtual assistant within one app is isolated from the context of history of the user's interaction with the virtual assistant when using another app, e.g., on the same device. Such compartmentalization reduces the seamlessness with which users can use a virtual assistant for tasks and recommendations that span across multiple apps.

DESCRIPTION

Techniques described in this disclosure integrate the knowledge of user interaction across apps and contexts. The integrated interaction information enhancement of content or action recommendations. The interaction integration and/or the recommendation delivery can be performed by a virtual assistant. The techniques are implemented such that any use of interaction data is performed with specific user permission. Users are provided with options to select the interaction data that can be utilized for such purposes, to restrict data from certain apps, or to turn off the feature entirely.

The integration of interactions across the device can enable a more complete understanding of the user's needs and tasks based on the content consumed across apps and online content sources, such as web pages. If the user permits, the improved understanding can be leveraged in subsequent interactions in other apps and online content sources. For example, a

user that is interested in learning to bake bread may watch an online video with instructions on breadmaking. The instructions typically include a list of essential bread making ingredients, such as yeast. When the user invokes the shopping app later on to make weekly purchases, the virtual assistant can suggest adding yeast to the shopping list.

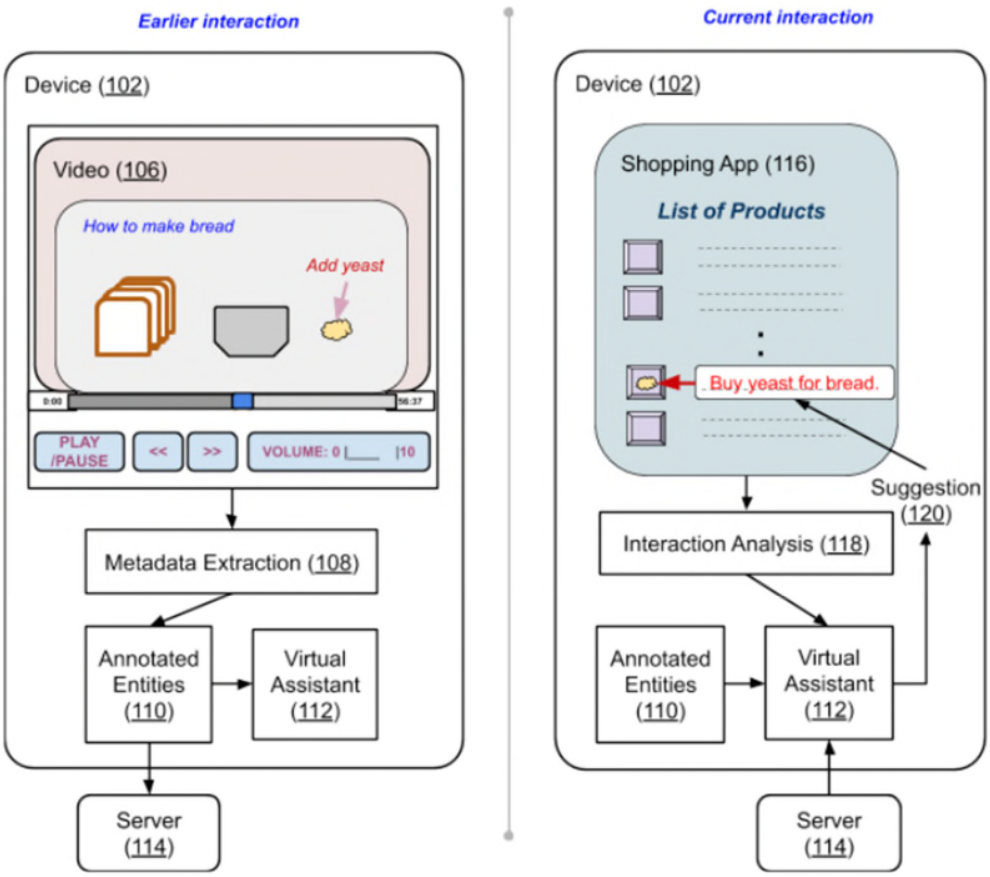

Fig. 1: Offering a recommendation based on prior content consumption in another app

Fig. 1 shows an example of implementation of the operation described above. The left side of Fig. 1 shows an earlier interaction when a user watches a video (106) related to bread making on a device (102). With user permission, relevant metadata from the video is extracted (108) and analyzed to identify and annotate the various entities (110) present in the metadata. Alternatively, or in addition, the video itself may be analyzed to obtain the entity information.

The annotated entities can be provided to an on-device virtual assistant (112). Optionally, the annotated entities can be stored on a server (114) external to the device, if permitted by the user.

At a later time, shown on the right side of Fig. 1, when the user is using a shopping app (116) to purchase groceries, the user's interactions can be examined (118). Based on a match between the user's shopping context and the entities extracted from the prior consumption of the bread making video, the virtual assistant (112) can offer a suggestion (120) to add yeast to the list of products in the shopping app.

The operations described above can be realized with user permission by extracting relevant metadata from the content or app in the user's focus. With permission, the extraction can be performed by analyzing various pieces of on-screen content for relevant information. Such on-screen content can include, e.g., view hierarchy, on-screen text and photos, subtitles in video content, images of video frames, parameters in content links, etc. The extraction and analysis can involve one or more suitable techniques such as optical character recognition (OCR), computer vision, annotation of entities in text using standard trained machine learning models, etc. Each extracted entity can be associated with a topicality score computed to indicate the degree to which the entity is central to the content on screen. The extracted information along with results of the analyses can be stored on the user's device.

The stored entities along with the corresponding annotations and topicality scores can be used to offer contextually appropriate suggestions for the user's later interactions with apps and content, regardless of which app or content source is in use during such later interactions. For instance, if a user asks the virtual assistant to open a shopping app after having watched a video in the recent past, the user can be offered product recommendations based on the contents of the recently watched video, as illustrated in Fig. 1.

The recommendations can be offered at an appropriate time during the user's interaction with an app or content. For instance, product recommendations can be made when the user is likely to be using the app for making purchases by browsing a list of products that includes a product encountered in previous content consumption. The contextually appropriate time for offering recommendations can be determined by checking the repository of saved annotated entities from the user's previous interactions to identify the entities that can be leveraged to offer helpful suggestions relevant to the current on-screen interactions. Alternatively, or in addition, matching the saved annotated entities and the user's current interaction context can be based on relevant off-screen information. For instance, the saved annotated entities can be matched with the product catalog of the shopping app such that relevant products can be recommended even if they are not currently displayed on the user's screen.

The suggestions generated as above can be delivered to the user in any suitable form. For instance, an on-screen suggestion for adding an item to the shopping cart can be displayed next to a recommended product. Alternatively, the suggestions can be delivered as contextual notifications that serve as hints such as "remember to buy yeast." Alternatively, or in addition, the user can be provided options to permit a virtual assistant to automate high-confidence tasks by operating the relevant app on behalf of the user. For instance, the user can request the virtual assistant to use the shopping app and automatically purchase recommended products that are deemed highly likely to serve the user's needs.

With user permission, the techniques described above can support extracting annotated entities from any user interaction regardless of whether the interaction involved the use of a virtual assistant. Similarly, the delivery of suggested actions or content can occur in any interaction regardless of the involvement of a virtual assistant. However, the types and modes of

the suggestions can differ if a virtual assistant is involved. For instance, the virtual assistant can be used to proactively provide recommendations.

The user can also choose to receive on-demand recommendations from the virtual assistant. For instance, when shopping, the users can ask the virtual assistant for product suggestions or inquire if they forgot to add something to the shopping cart. The on-demand recommendations can be requested via any suitable mechanism, such as voice commands to the virtual assistant. The on-demand recommendations can be delivered in any suitable manner such as voice response, on-screen notification or hint, on-screen tappable annotation that performs the corresponding action, link to related content, deep link to the relevant entity within a previous interaction, etc.

Where possible, content analysis for metadata extraction can be done offline in advance. In such cases, precomputed information can be obtained from the relevant backend. If the user permits, the extracted annotated entities can be stored on a server external to the device which can support cross-device operation. The extracted annotated entities can be aggregated or boosted across multiple pieces of consumed content to improve the confidence in matching and prediction. Various threshold values used for matching and prediction can be set by the developers, specified by the user, and/or determined dynamically at runtime.

The techniques described in this disclosure can be integrated within any operating system or virtual assistant. The entity information serves as additional memory for a virtual assistant. Implementation of the techniques can enhance the user experience of using the OS and the virtual assistant, and can make recommendations more relevant and useful to the users.

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 interactions with various apps on a user device, web pages visited via a browser, 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

With user permission, the techniques described in this disclosure integrate the knowledge of user interaction across apps and contexts. The information in turn enables enhancement of content or action recommendations, e.g., by a virtual assistant. The integration of information of user interaction can be realized by extracting relevant entities from the content or app that is the subject of the current interaction. The extracted entities along with the corresponding annotations and topicality scores are used to offer contextually appropriate suggestions for the user's later interactions with apps and content, regardless of which app or content source is involved in these later interactions. Recommendations can be offered at an appropriate time (or on-demand) and in a form suitable for the user's current interactions. The described features can be implemented in a device operating system and/ or a virtual assistant. The techniques are implemented such that any use of interaction data is performed with specific user permission.