March 14, 2018

Action prediction from conversations

Victor Carbune

Alexandru Damian

Action prediction from conversations

ABSTRACT

This disclosure describes techniques to automatically extract actions and reminders from conversations between users, when users permit analysis of conversations. With user permission, speech recognition, speaker disambiguation, and action extraction, e.g., implemented using trained machine learning models, are utilized to analyze conversations, detect when the conversations include user actions, and determine the owner of the actions. Actions are grouped by user and/or by type. A user interface displays follow-up actions identified from a conversation for review by users and enables users to make corrections.

KEYWORDS

- Virtual assistant
- Voice instructions
- Speech recognition
- Speaker disambiguation
- Action extraction
- Natural language processing

BACKGROUND

Speech recognition is used to recognize voice instructions, e.g., when explicitly triggered by users. For example, in response to a voice instruction from a user such as “Virtual Assistant, set a reminder for tomorrow,” a virtual assistant uses speech recognition to parse the instruction and create a reminder for the user, e.g., via an application and/or a user device.

Conversations between users can also include references that lead users to generate a set of actions for one or more of the participants, but such actions are usually not identified or

executed by the virtual assistant. User experience can be enhanced by the automatic identification of user actions from a conversation.

DESCRIPTION

Techniques described herein enable users to generate actionable tasks and reminders from conversations. With permission and express consent of all participating users, conversation between users is analyzed to determine a set of actionable tasks (if any) for one or more of the users. Extracted actions are then assigned to individual users.

Users can be prompted with follow-up actions, utilizing for example, reminders using virtual assistants, calendar reminders, email drafts, to-do lists, search queries, etc. In addition to determining the set of actions and their owners from the conversation, actions can be combined and organized by owner and/or logical type.

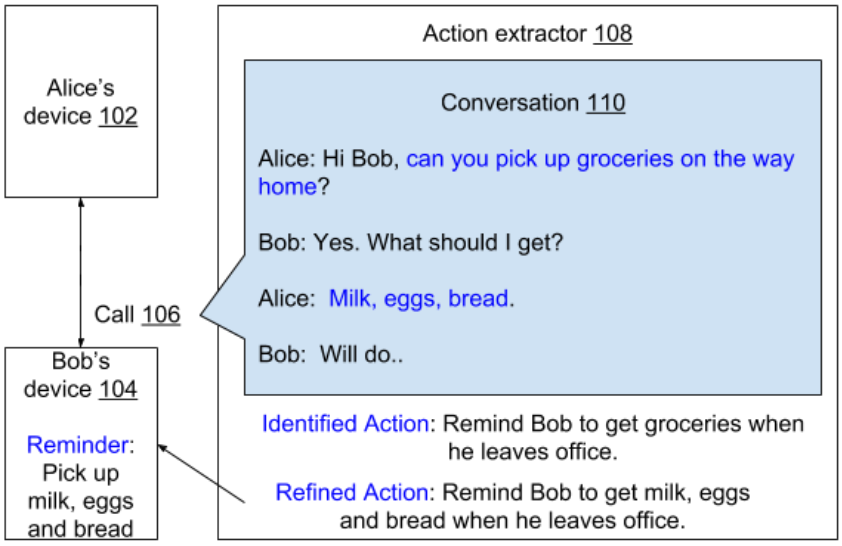

Fig. 1: Extracting actions from user conversation

Fig. 1 illustrates an ongoing] conversation (110) between two users, Alice and Bob, using respective devices (102 and 104), e.g., a voice call (106). The users have provided consent for a virtual assistant to automatically extract actions. In the example illustrated in Fig. 1, a user Alice

asks another user Bob to pick up groceries. An action extractor module (108), e.g., that is part of a virtual assistant, identifies the action "pick up groceries" from the conversation (110) and determines, with a probability that exceeds a predetermined threshold, that Bob is responsible for completing the action. The conversation is further analyzed to determine that specific items (milk, eggs, and bread) are listed as the groceries to be picked up. With permission from Bob, the refined action is set as a reminder on Bob's device (104).

Actions can be grouped by user, as in the example, and/or by type. For example, in a discussion where different users add items to a grocery list, references by all users to requests for groceries are analyzed to generate a shopping list. Corrections or modifications to previous actions as determined from the conversation by users are also incorporated. Continuing with the example of Fig. 1, if a user adds 6 eggs to the grocery list initially, but later corrects the quantity to 12 eggs, the shopping list is revised to reflect the 12 eggs.

To accommodate situations where actions are modified by users during conversation, the status of each action is recorded and updated accordingly such that different versions of the same action are not generated. For example, in a conversation between users that pertains to a shopping list, the quantity of each item requested in the list is maintained separately. Revisions are made to the shopping list if a user revises the quantity during the conversation.

Recognition of a new action that is related to an existing action is used to update the existing action. For example, in response to a user's comment: "oh, also remember to buy some chocolate when you go to the shop," the item chocolate is automatically appended to an existing shopping list.

Actions from a conversation are determined using one or more machine learning models. With user permission and express consent, speech recognition is used to generate a transcript of

the conversation. A speaker identification module ascribes the content in the transcript to the respective users. An assignment identification module assigns the actions that are identified and extracted from the transcript to the determined owners of the actions. Natural language processing methods are used to predict the most likely owner of actions that are identified and extracted from a conversation.

Further, a user interface displays follow-up actions identified from a conversation for review by users. Examples of follow-up actions include calendar reminders, email drafts, to-do lists, search queries, etc. that are executed upon user confirmation. The user interface enables users to refine, modify, or correct the proposed follow-up actions. When users permit use of corrections input for training, such corrections also serve to provide feedback to the machine learning models. Incorrect classification of actions from a user conversation and/or inaccurate attribution of extracted actions to wrong users, are examples of feedback provided to the machine learning models.

Users can choose to receive follow-up action confirmation requests and related prompts on additional user devices, e.g., a mobile phone or a tablet. If a user operates multiple devices during a conversation, for example, video chatting on a laptop while working on a phone, both devices are configured to be able receive and edit prompts with follow-up actions.

The described techniques seamlessly integrate different technologies such as speech recognition, speaker disambiguation, and action extraction. The techniques can be utilized in combination with different applications to extract and implement actions predicted from conversations. With user permission and express consent, user actions and responses to the proposed follow-up actions are utilized by the machine learning models. For example, a reminder originally phrased as “buy some eggs” can be phrased as “seems like the usual place

you are buying groceries from is still open; do you want to buy some eggs on your home?" if the models can determine the locations user typically buys groceries, based on the user's past actions. Such determination is made using data, e.g., location data, that the user has provided consent for use in setting reminders.

The present techniques can be incorporated into communication applications, e.g., messaging and chat applications, voice calling applications, video calling applications, etc. Speech recognition, speaker disambiguation, and action extraction can be implemented on a server or on user devices, e.g., based on user permission, available network connectivity and processing capacity, etc. Virtual assistant applications on user devices can obtain user confirmation for actions that are extracted from conversations and to receive follow-up prompts. These techniques can also be implemented at a platform level, e.g., integrated into a device operating system, to improve the multi-user environment experience, e.g., shopping recommendations can be generated from a multi-user conversation.

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 automatically extract actions and reminders from conversations between users, when users permit analysis of conversations. With user permission, speech recognition, speaker disambiguation, and action extraction, e.g., implemented using trained machine learning models, are utilized to analyze conversations, detect when the conversations include user actions, and determine the owner of the actions. Actions are grouped by user and/or by type. A user interface displays follow-up actions identified from a conversation for review by users and enables users to make corrections.