September 30, 2019

Explorative Dialogs And Open-Ended Conversations On Virtual Assistants

Allen Ni

Peng Yang

Explorative Dialogs And Open-Ended Conversations On Virtual Assistants

ABSTRACT

Virtual assistant responses are typically utilitarian in nature, e.g., answering a query, completing a pre-defined task, etc. However, a substantial amount of virtual assistant queries are exploratory in nature, e.g., they pertain to personality, knowledge, etc. An interactive dialog that enables extended conversation is suitable for such queries. This disclosure describes techniques that enable a virtual assistant to conduct an informational conversation or explore a knowledge domain with a user.

KEYWORDS

- Virtual assistant
- Voice assistant
- Open-ended conversation
- Explorative dialog
- Dialog manager
- Focus tracker
- Knowledge exploration
- Chatbot
- Smart speaker
- Smart appliance

BACKGROUND

Current virtual assistant features are utilitarian in nature, e.g., answering a query, completing a pre-defined task, filling a form, etc. To fulfill such tasks, appropriate rules are coded, and the user interface (or experience) is designed to enable users to efficiently receive

accurate answers or perform the requested tasks. However, a substantial amount of virtual assistant queries are exploratory in nature, e.g., pertaining to personality, knowledge, etc. An interactive dialog that enables extended conversation is suitable for such queries. Examples of such queries are as follows.

- What does a lion sound like?
- How tall is a giraffe?
- Do you have a dog?
- What's your favorite animal?

Although the first two queries can be answered by playing a sound and by returning a number respectively, the queries can also serve as entry points to longer conversations. The last two queries require creative writers to manually compose answers. It is difficult to write good answers to generic questions such as the last two queries. Therefore, writers typically find a way to guide the conversation such that users come back with a better search query.

For example, when children ask questions such as the above, they may be expressing healthy curiosity, or they may simply be bored and in search of stimulation. In either case, rather than providing a sound file or a height measurement, it is better to engage them with meaningful follow-on conversation that addresses their underlying needs. Conversation being a natural mode of interaction for children, attributes of a virtual assistant include the ability to provide companionship, the ability to conduct open-ended conversations, and the ability to enable children and more generally, all users to actively explore knowledge.

DESCRIPTION

This disclosure describes techniques that enable a virtual assistant to engage users in knowledge exploration via open-ended conversations. The techniques are specialized to specific

knowledge domains, e.g., the animal kingdom, the plant kingdom, space, education, discovery, etc.

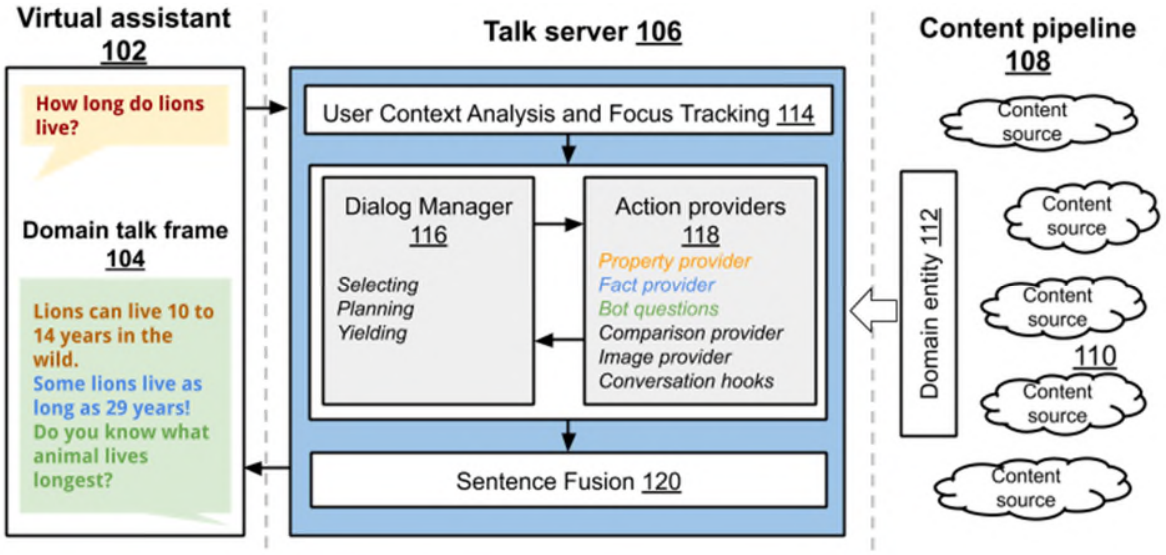

Fig. 1: Explorative dialogs and open-ended conversations with a virtual assistant

Fig. 1 illustrates an example architecture for explorative dialogs and open-ended conversations using a virtual assistant, per techniques of this disclosure. As illustrated in Fig. 1, a user has issued a domain-specific query, e.g., specific to the animal kingdom knowledge domain, such as “How long do lions live?” In response, the virtual assistant (102) spawns a domain talk frame (104) that is used to manage and display domain-specific conversations. In the example of the animal kingdom, the domain talk frame is an animal talk frame.

For informational conversation and domain exploration, a talk server (106) is provided. The talk server includes:

- A focus-tracking and context-analysis mechanism (114).
- A dialog manager (116) that selects content from content providers that is relevant to the current focus. For example, the dialog manager can be implemented using machine learning techniques.

- An action provider (118) that offers actions based on current focus and conversation history.
- A sentence-fusion layer (120) that fuses the content into a coherent response.

A content pipeline (108) extracts domain-related data from various vetted content sources (110) and stores the data in storage, referred to as a domain entity (112).

When a user issues a domain-triggering query, the domain talk frame in the virtual assistant creates a new talk session and forwards the request to the talk server. The talk server analyzes the user input and conversation history and infers the user focus for the current turn of the virtual assistant.

Focus can be thought of as a representation of the user's intent as well as a representation of bot actions. Focus captures the topic of the conversation. The virtual assistant (or bot) is designed to satisfy users by providing interesting talking points on the current focus; however, either of the virtual assistant/bot or the user can influence and change the focus. Through the focus mechanism, the virtual assistant can guide users to explore the domain, e.g., by suggesting interesting and relevant fun facts, and possible follow-up questions.

A bot turn comprises multiple bot steps. In a bot step, the dialog manager asks action providers to offer actions given the current focus, the conversation history, and, for personalization, a cross-session of the memory of the virtual assistant pertaining to the user, obtained with user permission. An action can be an utterance, a sound, an emoji, an image, a video, or a combination thereof. An action can also include a property or fact relating to an object, e.g., lion, from the current knowledge domain, e.g., the animal kingdom. An action can also be a bot question, a comparison between objects in the knowledge domain, or a conversational hook. An action can set a new conversation focus if it is selected. Based on a

model trained on rated conversations, the dialog manager picks the most suitable action and updates the conversation focus and history. At the end of each bot step, the dialog manager determines whether to add an additional step or to end the current bot turn.

Once the current bot turn ends, the sentence fusion layer fuses utterances from bot steps to form a cohesive response and returns it to the domain talk frame. Fusion includes, among other things, pronominalization and adding discourse markers, e.g., “well,” “however,” “in comparison,” “also,” etc. The domain talk frame saves the talk conversation state and converts the talk server response to a speech synthesizer response. This completes a talk conversation cycle.

If the user is detected as being engaged, the domain talk frame forwards the follow-on query to the talk server together with the saved conversation state to continue the talk session. The domain talk frame automatically tries to answer follow-on queries even if those are not a triggering query, e.g., “tell me more,” “yes,” etc. The talk server applies a domain classifier to detect user inputs that are out of the domain (e.g., “set a timer”) and yields the dialog accordingly. If the talk server does not yield, a bot turn is constructed as described above. Otherwise, the talk session is terminated. It is restarted when the user issues another triggering query.

Yielding a dialog can happen in a number of ways, as follows.

- **Hard stop:** when the user says “stop” or a phrase expressing negativity, the talk server immediately ends the conversation and returns no response.
- **Soft stop:** when users show no strong interest in continuing, the talk server acknowledges the user’s apparent intent to move on, e.g., by saying “OK, come back anytime; I always like to talk about animals,” and ends the conversation.

- **Lack of good actions:** when the talk server cannot find actions with sufficiently high scores, it ends the conversation so that other candidates provided by the virtual assistant can take over. These may be queries where there is insufficient domain content for the talk server to proceed.
- **Out-of-domain queries:** when the user asks out-of-domain queries, the talk server offers no response but keeps the domain talk frame in the background so that the dialog can continue if users follow up with relevant queries.

In this manner, the described techniques use a dialog manager and a focus tracking mechanism to create open-ended virtual assistant conversations with users. The dialog manager uses the content pipeline to load large amounts of knowledge content to generate a variety of conversation candidates. The dialog manager ranks and selects candidates to provide the user with suitable responses to advancing the dialog.

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 that enable a virtual assistant to conduct an informational conversation or explore a knowledge domain with a user.

REFERENCES

[1] Razavi, S. Zahra, Lenhart K. Schubert, Benjamin Kane, Mohammad Rafayet Ali, Kimberly Van Orden, and Tianyi Ma. "Dialogue Design and Management for Multi-Session Casual Conversation with Older Adults." *arXiv preprint arXiv:1901.06620* (2019).

[2] Emotix, "Introducing Miko's classroom" http://emotix.in:8300/emotix/?p=332 accessed on Sep. 24, 2019.