November 16, 2018

Empathetic and mindful virtual assistant

Sammy El Ghazzal

Migle Padegimaite

Empathetic and mindful virtual assistant

ABSTRACT

Virtual digital assistants generally do not account for the user's emotional context. As a result, the responses of a virtual assistant may not better the user's mood.

This disclosure describes techniques by which an assistant detects with user permission the current emotional context of the user, and uses it to modulate its responses. For instance, if the assistant detects that the user is in a sad mood, it focuses its utterances on positive developments. A request for news by the user results in positive news items, e.g., "unemployment is down," or "USA won the Olympics." A request for playing a song results in joyful tunes being played. Per the techniques, the assistant detects, with user consent, the user's emotional context by using machine-learning models to analyze the user's speech. The assistant tailors its response to user moods by running a sentiment analyzer on possible responses, and picking a response that is appropriate to the user's mood.

KEYWORDS

Virtual assistant; digital assistant; emotion detection; sentiment analysis; machine learning; mood classification

BACKGROUND

Virtual digital assistants generally do not account for the user's emotional context. As a result, the responses of a virtual assistant may not better the user's mood. A mood-appropriate, human-like response from a virtual assistant makes interactions with the assistant smooth and natural.

DESCRIPTION

This disclosure describes techniques that enable a virtual assistant to adapt its responses to the user's emotional context. For example, if the assistant detects that the user is in a sad mood it focuses its utterances on positive developments. A request for world news by the user results in positive news items, e.g., "unemployment is down," or "USA won the Olympics." A request for playing a song or movie results in joyful tunes/movies being played. A request for the weather forecast results in an emphasis on the sunny parts of the day and a de-emphasis on the cloudy parts of the day.

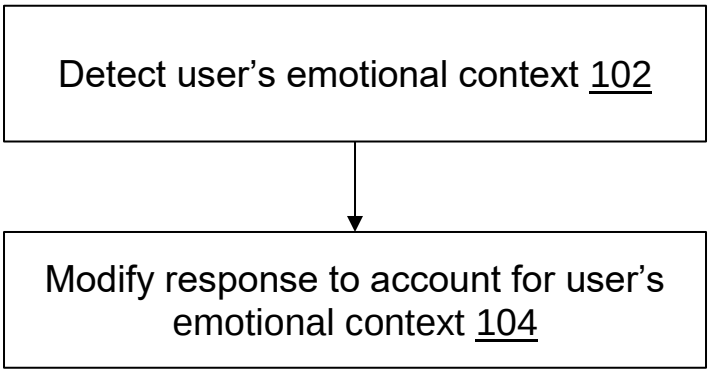

Fig. 1: Tailoring the assistant's response to the user's emotional context

Fig. 1 illustrates tailoring the assistant's response to the user's emotional context, per techniques of this disclosure. The virtual assistant detects the user's emotional context (102) by e.g., simply asking the user how they are feeling today. Recognizing that the user may not provide any more than a stock answer, e.g., "I'm fine," to such a question, the virtual assistant uses machine learning techniques to detect the user's emotional state, as further described below. The virtual assistant modifies its response to account for the user's emotional context (104) using, e.g., sentiment analysis, to pick out an appropriate response from a family of possible responses.

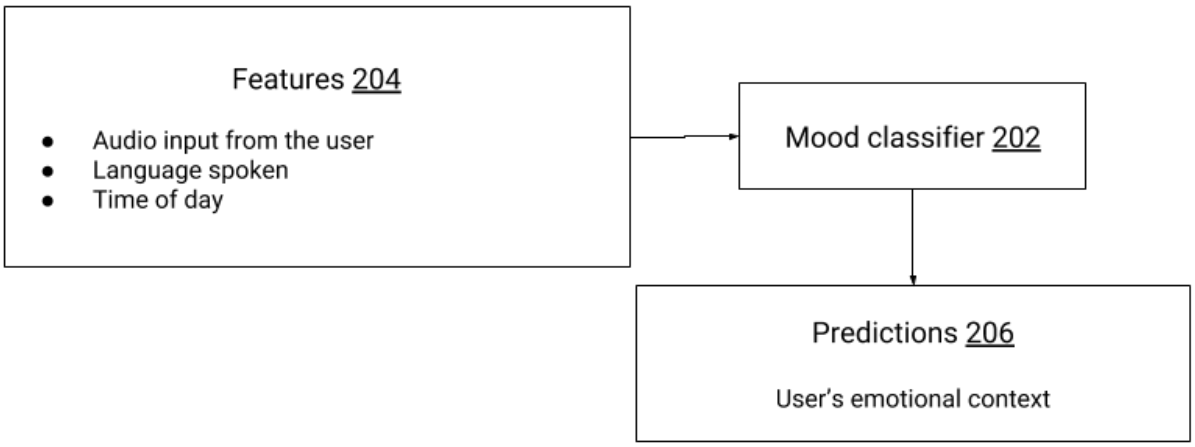

Fig. 2: Detection of the user's emotional context

Fig. 2 illustrates detection of emotional context, per the techniques of this disclosure. With user permission, a mood classifier (202) accepts as input features (204) such as audio input from the user, language spoken (as inferred from audio input), time-of-day, etc. in order to predict the user's emotional context (206), e.g., happy, enthusiastic, sad, bored, moody, etc. The mood classifier is a machine learning model that is trained on annotated pairs of input features and optimal predictions. Audio can be processed prior to classification using, e.g., time- or frequency-domain convolution. The machine learning model can be, for example, a regression learning model, a neural network model, a deep learning model, etc. Example types of neural networks include recurrent neural networks (e.g., long short-term memory networks), convolutional neural networks, etc. Other machine learning models, e.g., support vector machines, random forests, boosted decision trees, etc., can also be used. Reinforcement learning can be used to advantageously incorporate past predictions as feedback to improve machine-learning performance. The machine learning model can have a final softmax layer.

A fine tuning of the model can be performed based on recently acquired data, e.g., the user's response to an initial question such as "how are you feeling?" Machine classification of

the user's emotional context can be done on the user's device, at a server, or in a combination of device and server. In addition, federated learning can be used in a privacy-preserving fashion to further refine the model, for example by aggregating learnings over multiple users.

Once the user's emotional context is determined by the virtual assistant, it modulates its response appropriately, using, e.g., sentiment analysis, to pick out an appropriate response from a family of possible responses. If the virtual assistant determines that emphasis is to be given to positive developments, it can also replace words/phrases with positive synonyms, use positive sentence constructions, etc. However, depending on the importance of a negative statement, e.g., news, the virtual assistant may present it in unvarnished form. For example, news such as "volcano erupted; flights delayed and diverted" may be presented neutrally, e.g., without attempt towards positive emphasis. The virtual assistant can also access content metadata, e.g., of music or movies, in order to determine a mood-appropriate response. For example, if user is detected as being sad, the virtual assistant leans towards recommending movies with metadata indicative of comedy or action genre. If the user is detected as being moody, the virtual assistant leans towards recommending a playlist with metadata indicative of calming or soothing genre. Depending on probabilities of events, the virtual assistant may determine if a potentially negative event is worthy of mention. For example, if rain is predicted with a relatively low 5% chance, the virtual assistant may not mention it at all, especially if the user is already detected as being melancholy. Alternately, the virtual assistant can focus on the sunnier aspects of the weather while balancing out a weather report by stating, e.g., "there is a small chance of rain but the temperatures will be pleasant." The virtual assistant tracks over time the match of its response to the user's mood, using, e.g., sentiment analysis, such that it betters its performance over time.

In this manner, the techniques of this disclosure use machine learning to provide an enhanced user experience with virtual assistants. By modulating the assistant-user interactions to account for the user's emotional context, the assistant becomes more human-like, leading to smoother, more natural conversations and an overall improvement in the well-being of the user. Alternative to machine learning, heuristics, e.g., time-of-day, calendar content, etc. can be used to predict the user's mood.

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 by which an assistant detects with user permission the current emotional context of the user, and uses it to modulate its responses. For instance, if the assistant detects that the user is in a sad mood, it focuses its utterances on positive developments. A request for news by the user results in positive news items, e.g., "unemployment is down," or "USA won the Olympics." A request for playing a song results in

joyful tunes being played. Per the techniques, the assistant detects, with user permission, the user's emotional context by using machine-learning models to analyze the user's speech. The assistant tailors its response to user moods by running a sentiment analyzer on possible responses, and picking a response that is appropriate to the user's mood.