Technical Disclosure Commons

Defensive Publications Series

March 2020

Contextual Error Correction in Automatic Speech Recognition

Manaal Faruqui

Janara Christensen

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation

Contextual Error Correction in Automatic Speech Recognition

ABSTRACT

This disclosure describes techniques that leverage the context of a conversation between a user and a virtual assistant to correct errors in automatic speech recognition (ASR). Once confirmed by the user, the correction event is used to augment the training data for ASR.

KEYWORDS

- Automatic speech recognition (ASR)

- ASR error correction

- Spoken query

- Misrecognized query

- Spoken command

- Smart speaker

- Virtual assistant

- Contextual ASR

BACKGROUND

ASR errors that occur in a conversation between a user and a virtual assistant can disrupt the conversation. Consequently, important information can be missed out or misunderstood by the virtual assistant, leading to user frustration.

DESCRIPTION

This disclosure describes techniques to identify potential ASR errors in virtual assistant queries and to seek clarification from the user. Per the techniques, the current query, the previous query, and the n-best hypotheses of the current query (produced by the ASR) are evaluated contextually to identify potential errors and to present potential corrections to the user. The

user's acceptance or rejection of a correction serves as explicit user feedback and is used as training data for ASR.
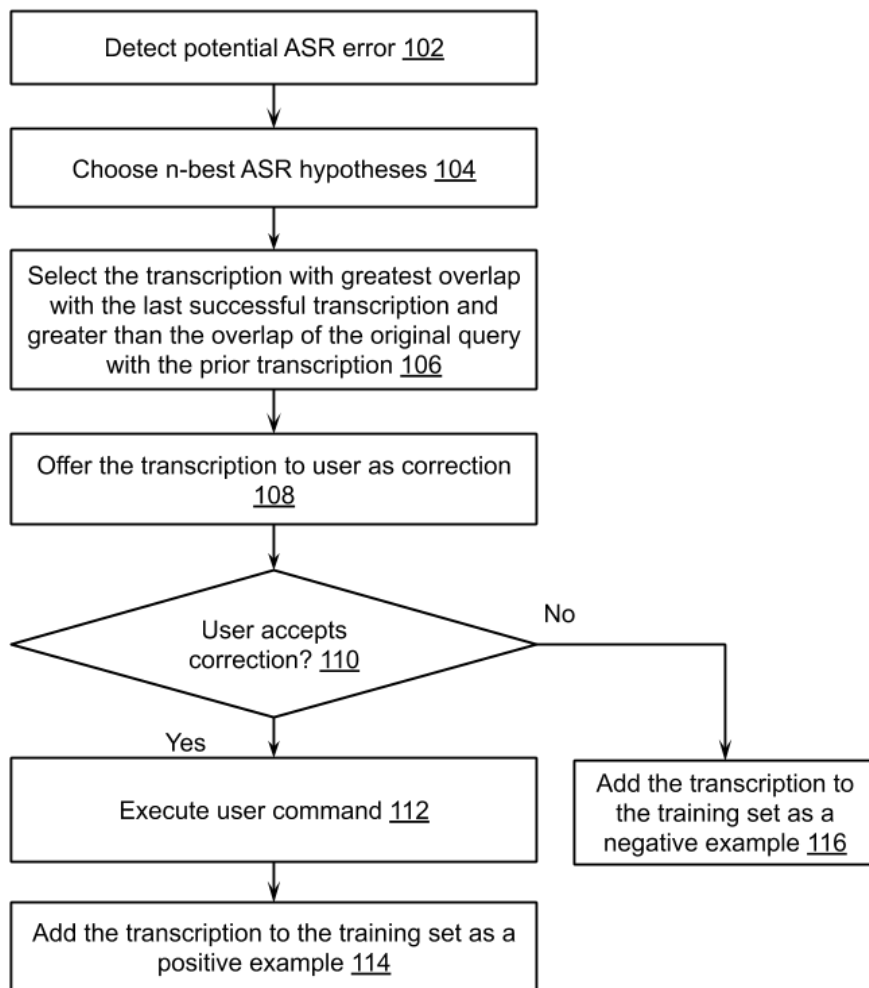
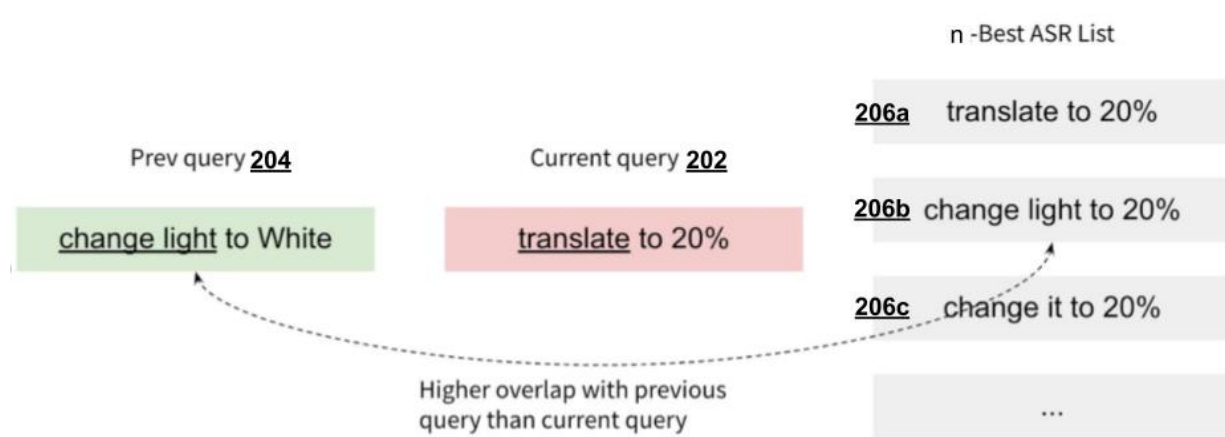**Fig. 1: Contextual error correction in automatic speech recognition**

Fig. 1 illustrates contextual error correction in automatic speech recognition, per the techniques of this disclosure. A potential ASR error is detected (102), e.g., when a semantic parsing of a user's query to a virtual assistant doesn't fit any template of intended user action.

The n-best ASR hypotheses, e.g., the n ASR transcriptions of the user's query with the highest confidence scores, are chosen (104). Of the n-best transcriptions, the one with the greatest word overlap with the last successful transcription is selected (106). The selected

transcription also has greater overlap than that of the original query (top-most transcription) with the prior transcription. The criterion for selecting corrections being word overlap, the technique works in a language-independent manner.

The selected transcription is offered to the user as a potential correction to the ASR error (108). Prior to issuing the potential correction to the user, a recursive ACE procedure can be run on the potential correction to ensure that it can be parsed. If the user accepts the correction (110), the command is executed. With user permission, the transcription and the corresponding speech waveform are added to the ASR training set as a positive training example (114). If the user rejects the correction, the transcription and the corresponding speech waveform are added to the ASR training set as a negative training example (116). If the user denies permission, the transcription and corresponding speech waveform are not used for training.

Example

Fig. 2: An example of contextual error correction in automatic speech recognition

Fig. 2 illustrates an example of contextual error correction in ASR, per the techniques of this disclosure. A query (202, "translate to 20%") made by a user to a virtual assistant is detected

as potentially erroneous. As mentioned before, the error may be detected due to a semantic parsing of the query not fitting any template of intended user action.

The n-best ASR transcriptions (206a-c) are chosen and compared with the last successfully transcribed query (204, "change light to white"). Of the n-best transcriptions, one that exhibits the greatest word overlap with the last successfully transcribed query (206b, "change light to 20%") is selected as a potential correction to the error and offered to the user as such. If the user accepts the correction, the user's command is executed, e.g., the virtual assistant takes an appropriate action and provides a response. Further, the user's acceptance or rejection of the correction serves as explicit user feedback and, if permitted by the user, is used as training data for the ASR.

Previous query	Current query	Corrected query
Turn on the TV.	Turn on mister Bolt EV.	Turn on Mr. Bull TV.
Turn on the pool pump.	Turn on the Pooh.	Turn on the pool.
White bed light on.	Chris Bellion.	Chris's bed light on.
How do you wash pillows?	How do you drive the list?	How do you dry pillows?
Turn yourself to the TV aux mode	Switch axe.	Switch to aux.
What's the legal age in South Carolina to get a tattoo?	What's in South Carolina what's the legal age going to Pearson?	What is in South Carolina what is the legal age to get a piercing?
Turn on the telly.	Unpost to tell us.	Unpause the telly.

Table 1: Examples of initially erroneous transcriptions (middle column) later corrected (right) based on word overlap with a previous query (left column).

Table 1 illustrates some examples of initially erroneous transcriptions later corrected based on word overlap with a previous query.

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 spoken queries or commands, corrections, interaction with virtual assistant, preferences), 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 leverage the context of a conversation between a user and a virtual assistant to correct errors in automatic speech recognition (ASR). Once confirmed by the user, the correction event is used to augment the training data for ASR.

REFERENCES

[1] Skobeltsyn, Gleb, Alexandru Ovidiu Dovlecel, Carl-Anton Ingmarsson, Martin Baeuml, Behshad Behzadi, and Dmitry Osmakov. "Contextual improvement of voice query recognition." U.S. Patent 9,576,578, issued February 21, 2017.