November 2019

Automatic correction of disfluent spoken queries

Manaal Faruqui

Siddharth Gopal

Automatic correction of disfluent spoken queries

ABSTRACT

A user's interaction with a virtual assistant typically involves spoken requests, queries, and commands which often includes disfluencies. This disclosure describes techniques to automatically correct disfluent queries. Per techniques of this disclosure, a disfluency correction machine learning model is utilized to convert a disfluent query to a corresponding fluent query. Lexical features extracted from the disfluent query are utilized to determine a portion of the query that is removed from the disfluent query to convert it to a fluent query. The model is trained using pairs of <disfluent, fluent> queries.

KEYWORDS

- Disfluency
- Speech transcript
- Query interpretation
- Noisy channel model
- Reparandum
- Natural language processing (NLP)
- Voice assistant
- Smart speaker

BACKGROUND

A user's interaction with a virtual assistant typically involves spoken requests, queries, and commands. Since users are speaking spontaneously, their speech often includes disfluencies. For example, a user that issues a query to request an airfare quote to Chicago may say "Get me the price for a flight to New York, … uh, I mean to Chicago for next weekend." In such a circumstance, it is necessary that a virtual assistant that is to respond to the query make an inference that the user’s query is for flights to Chicago (and not to New York).

Correction (repair) of such disfluent queries to fluent queries can enable correct inference of actual user intent and a subsequent correct response. In this example, the corrected version of the query is “Get me the price for a flight to Chicago for next weekend.” While disfluencies such as the occurrence of words like “oh”, “umm,” “uh,” etc. can be identified and corrected relatively easily, semantic disfluencies, e.g., when uttered words are actual words, but are not part of the intended query, can pose a challenge in query interpretation using natural language processing techniques.

DESCRIPTION

This disclosure describes techniques for the correction of disfluent speech, e.g., spoken commands or queries. Per techniques of this disclosure, a machine learned model is utilized to convert a disfluent query to a corresponding fluent query. The techniques can be used in any context, e.g., by a virtual assistant, that responds to spoken commands or queries. Fig. 1 illustrates an example of correction of disfluent speech, per techniques of this disclosure.

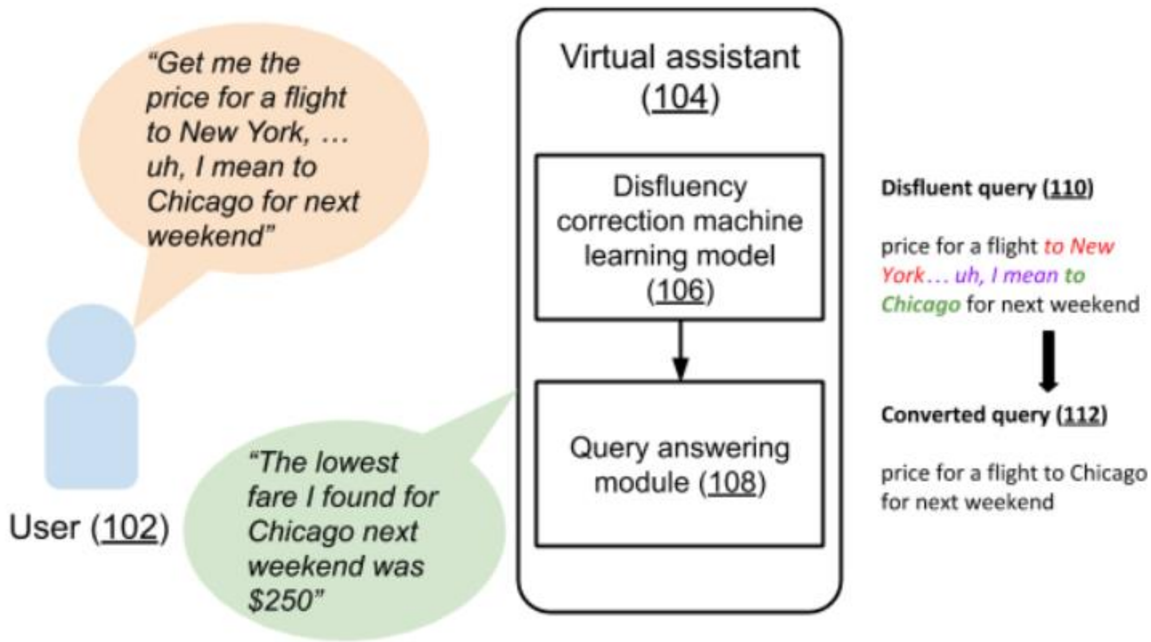

Fig. 1: Correction of a disfluent query using a machine learned model

In this illustrative example, a user (102) intending to obtain a price quote for a flight to Chicago poses instead a disfluent query (“Get me the price for a flight to New York, … uh, I mean to Chicago for next weekend”) to virtual assistant (104).

A disfluency correction machine learning model (106) is provided as part of the virtual assistant application. Using the disfluency correction model, it is determined that the portion of the query “to New York” (depicted in red in Fig. 1) is the reparandum, the portion to be replaced; the portion of the query “uh, I mean” (depicted in purple in Fig. 1) is the interregnum, the portion that marks the interval between the disfluent portion and the correct portion; and the portion of the query “to Chicago” (depicted in green in Fig. 1) is the repair, the correct portion that is the intended phrase in the query.

Based on the determination, the disfluent query (“price for a flight to New York… uh, I mean to Chicago for next weekend”) is corrected (converted) to the fluent query. For example, A decision tree model (or other suitable model) is utilized that uses lexical features extracted from the disfluent query (110) and determines the portion of the query that is to be removed from the query to convert it to a fluent query. In this illustrative example, the disfluent span is determined to be "to New York… uh, I mean" and the corresponding converted query (112) is determined as “price for a flight to Chicago for next weekend.” The query answering module (108) utilizes on the converted query and provides a response (“The lowest fare I found for Chicago next weekend was $250”) to the intended query posed by the user.

The machine learning model can be trained to detect disfluent queries and determine the corresponding fluent versions by using pairs, e.g., 2-tuples, ordered pairs, etc. of <disfluent, fluent> queries as training data. If users permit, the pairs of <disfluent, fluent> queries can be obtained from logs of virtual assistant data. For example, such pairs can be obtained from previous virtual assistant sessions where a failed disfluent query by a user (for example, a query that did not evoke a response or that evoked an incorrect response from a virtual assistant due to the disfluency) was corrected by the user in a subsequent attempt.

Alternatively, or in addition, the pairs of <disfluent, fluent> queries can also be obtained from human users who rewrite the fluent form of a disfluent query. For example, the pairs of queries generated by the human users can be used to evaluate models trained using the virtual assistant data logs.

The trained machine learning model can be any suitable type of machine learning model, e.g., regression learning models, neural networks, etc. Example types of neural networks that can be used include long short-term memory (LSTM) neural networks, recurrent neural networks, convolutional neural networks, etc. Other machine learning models, e.g., support vector machines, random forests, boosted decision trees, etc. can also be used.

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

A user's interaction with a virtual assistant typically involves spoken requests, queries, and commands which often includes disfluencies. This disclosure describes techniques to automatically correct disfluent queries. Per techniques of this disclosure, a disfluency correction machine learning model is utilized to convert a disfluent query to a corresponding fluent query. Lexical features extracted from the disfluent query are utilized to determine a portion of the query that is removed from the disfluent query to convert it to a fluent query. The model is trained using pairs of <disfluent, fluent> queries.

REFERENCES

[1] Kim, Joungbum, Sarah E. Schwarm, and Mari Ostendorf. "Detecting structural metadata with decision trees and transformation-based learning." In Proceedings of the Human Language

Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004. 2004.

[2] Lou, Paria Jamshid, Peter Anderson, and Mark Johnson. "Disfluency Detection using Auto-Correlational Neural Networks." In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 4610-4619. 2018.