Conversational/Multiturn Question Understanding

Gary Ren
Microsoft AI and Research
Sunnyvale, CA
gren@microsoft.com

Manish Malik
Microsoft AI and Research
Sunnyvale, CA
manisma@microsoft.com

Xiaochuan Ni
Microsoft AI and Research
Sunnyvale, CA
xiaon@microsoft.com

Qifa Ke
Microsoft AI and Research
Sunnyvale, CA
qke@microsoft.com

Nilesh Bhide
Microsoft AI and Research
Sunnyvale, CA
nileshb@microsoft.com

ABSTRACT

Existing research on question understanding and answering have focused on standalone questions. However, as interactions between humans and machines become increasingly conversational, there is a need for understanding conversational/multiturn questions, defined here as questions that depend on the context of the current conversation. This paper presents a novel architecture that leverages NLP techniques, deep learning, and search engine web knowledge to understand these multiturn questions by reformulating them into standalone questions that a downstream information retrieval system/dialogue agent expects. This paper also briefly explores the benefits of having a search powered system that can have guided conversations with users.

CCS CONCEPTS

• **Computing methodologies** → **Discourse, dialogue and pragmatics**; *Artificial intelligence*; *Natural language processing*;

KEYWORDS

NLP, conversational AI, conversational questions, multiturn questions, question understanding, deep learning, search engine, web knowledge

1 INTRODUCTION

The recent rise of technologies such as chatbots, digital personal assistants, and smart home devices has led to much more conversational interactions between humans and machines than ever before. In conversations, humans often ask questions that depend on the context of the current conversation, and will naturally expect machines to be able to understand such conversational questions.

A very basic example is asking "When was Microsoft founded?" followed by "Who founded it?" and "What is the stock price?", where both the follow up questions refer to Microsoft even though Microsoft is not explicitly stated. Existing research on question understanding and answering typically does not address these types of questions and only focus on standalone questions.

Therefore, this paper addresses the task of conversational question understanding (CQU), which consists of:

(1) Determining whether or not a question depends on the previous context.
(2) If so, reformulating the question to include the correct context and only the correct context.

We sought to create a CQU system that takes as input the previous context and current question, and outputs the reformulation of the current question (or the original question if no reformulation is needed). Another requirement is that our system must be generic and not restricted to individual domains.

2 RELATED WORK

The problem of CQU is related to the problem of coreference resolution, a well known NLP task which involves finding words that refer to the same entity in a text. Lots of research have been done on this problem, and recent attempts with deep learning have achieved state of the art results [1]. However, there are several challenges of CQU for which existing coreference resolution systems struggle with. These challenges are as follows, with example coreference resolution results from the Stanford CoreNLP toolkit [5].

(1) Grammatically incomplete/incorrect sentences (e.g. for "stanford tuition? where is it?", the "it" is not resolved to "stanford")
(2) Multiple possible entities (e.g. for "Is Microsoft in Seattle? Who is its mayor?", the "its" is incorrectly resolved to "Microsoft" instead of "Seattle")
(3) Knowing when a reformulation is actually needed (e.g. for "When was Microsoft founded? How long does it take bruised ribs to heal?", the "it" is incorrectly resolved to "Microsoft").
(4) When there isn't an explicit referring mention (e.g. for "When was Microsoft founded? Who is the founder?", there is no mention in the second question that explicitly refers to "Microsoft"). So even the perfect coreference resolution system will only be able to handle a subset of conversational questions.

Another related area that has been researched is the creation of dialogue agents using neural networks and reinforcement learning [4][2][3]. While these research show promising results, they are often limited to specific tasks, e.g. finding a movie.

In industry, many existing systems rely on whitelists of memorized conversational questions and patterns, so their coverage is very low.

3 ARCHITECTURE

The architecture of our CQU system consists of three main steps:

(1) Parse the context of the conversation.
(2) Generate possible reformulations.
(3) Select the best reformulation.

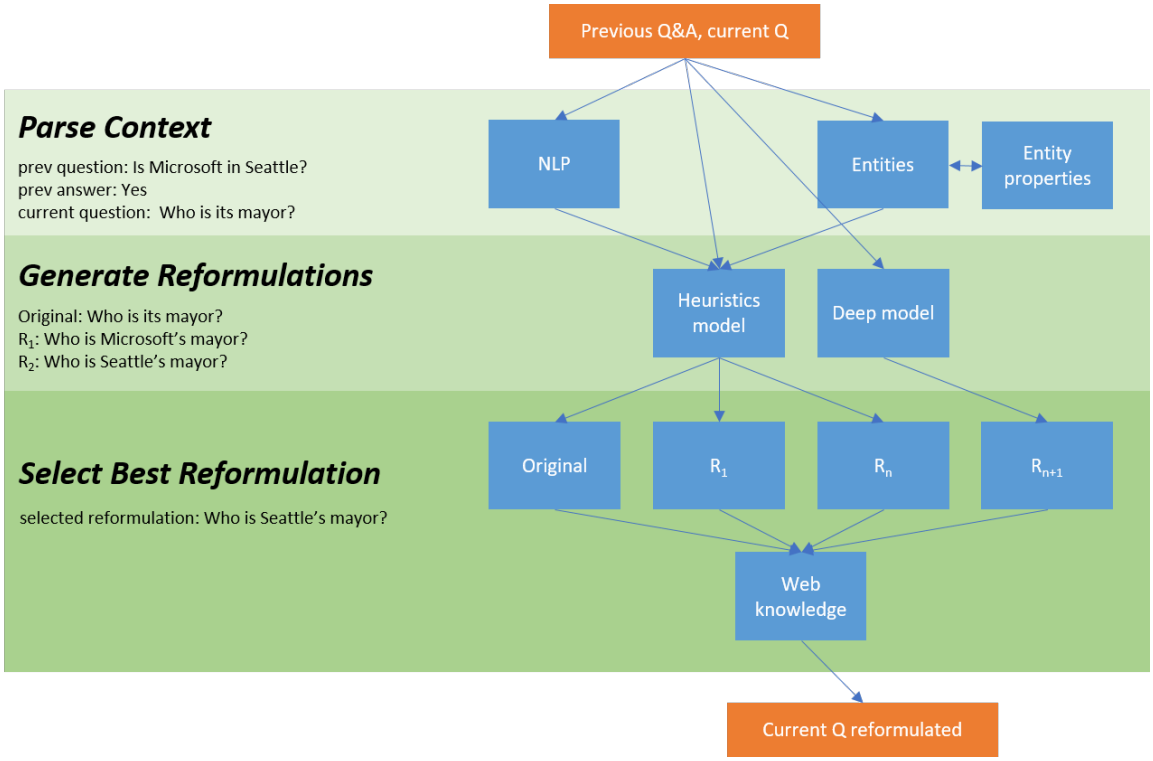

Figure 1: Architecture of multiturn reformulation system.

A high level diagram of this architecture with its various components is shown in figure 1, along with an example input/output.

3.1 Parse Context

Assuming that the user's conversation session is stored and retrievable, the previous question and answer can be retrieved and, along with the current user question, passed in as input to our CQU system. The first step is to parse these inputs to extract useful information that will help with generating reformulations, such as the detected entities and various NLP properties. Relevant properties of each entity are obtained by searching a knowledge graph for that entity.

3.2 Generate Reformulations

The next step is to generate a list of possible reformulations given the context and the information extracted from it. This generation is done by both a heuristics model and a deep model.

3.3 Heuristics Model

The heuristics model uses the context as well as the extracted NLP and entities information to generate reformulations via a set of rules. These rules go beyond a whitelist of memorized questions and text patterns, and are more broad and generic, allowing for much greater coverage/recall. Other benefits are that these rules don't require grammatically correct sentences and don't require explicit reference mentions, addressing a couple of the challenges of CQU mentioned earlier. Allowing for multiple reformulations to be generated further increases the recall. This model can focus solely on recall because of the precision provided by the reformulation selection step, which will be discussed in more detail in section 3.5.

3.4 Deep Model

The deep model uses the raw text of the context to generate a single additional reformulation. It uses a sequence to sequence recurrent neural network (RNN) model with attention mechanism [6], a commonly used model in NLP for tasks involving text generation. A high level diagram of the model is shown in figure 2. Long short term memory (LSTM) is the RNN used, and the input text is first converted into word embeddings. More details about the model can be found in the cited paper, though our parameters, preprocessing, and training procedure differ from the paper.

The dataset used for training and evaluation was created from search engine logs. It contains user-issued conversational questions along with their user-issued reformulations, which are used as labels. It also contains non-conversational, single turn questions so that the model learns to not reformulate such questions. Because this dataset comes from search engine logs, it is large enough for deep learning, and not restricted to specific domains.

Through word embeddings and the inherent capability of neural networks to learn features/representations, the deep model is able to generalize to words and patterns beyond what it saw during training. On the test set, the model is able to achieve 44.8 exact match on conversational questions with a BLEU score of 76.7. For

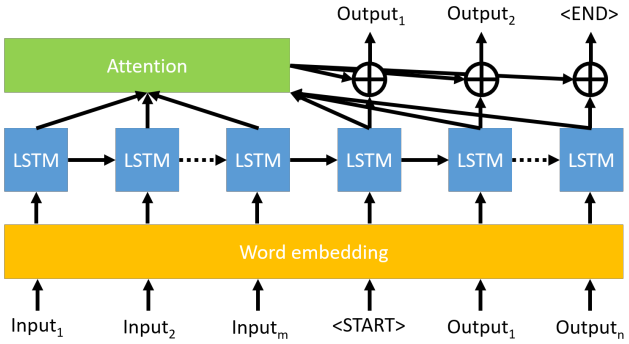

Figure 2: Sequence to sequence with attention.

the non-conversational questions, it was able to maintain a 75.3 exact match with a BLEU score of 87.7.

Here is a sample question session that was inputted to the deep model. The arrow indicates the output of the model. No arrow means that the model outputted the original question without any reformulation. The answers were simply obtained from a commercial search engine's search results.

Q: Where is amsterdam?
A: Netherlands
Q: What is its weather? → *What is amsterdam weather?*
A: 61°F
Q: Who is the mayor? → *Who is the mayor of amsterdam?*
A: Eberhard van der Laan
Q: How to split string in python?
A: split()
Q: How to read file? → *How to read file in python?*
A: open()
Q: How tall is kobe bryant?
A: 6'6"
Q: When was he born? → *When was kobe bryant born?*
A: August 23, 1978
Q: What are differences between bacteria and virus?
A: ...
Q: What are the similarities? → *What are the similarities between bacteria and virus?*
A: ...

This sample session shows that the model is able to perform reformulations that involve coreference resolution and also reformulations where there isn't a mention to be resolved. The model is also able to determine when the topic changes and the question no longer depends on the previous context, e.g. when the topic switched from Amsterdam to Python, the context of Amsterdam is no longer passed to the new question.

The deep model further increases the recall of the generate reformulations step by adding a reformulation that the heuristics model might have missed due to either a missing rule or a mistake in the context parsing step.

3.5 Select Best Reformulation

The final step is to select the best reformulation by leveraging search engine web knowledge. Each reformulation is issued as a query to a commercial search engine and the returned results and signals are used to select the best one.

Search engine web knowledge helps to solve the challenges of CQU mentioned earlier. Search engines do not require grammatically correct queries as inputs and are already adept at handling grammatically incorrect ones.

When there are multiple possible entities, the search results and signals will indicate which entity makes the most sense for the reformulation. In the example stated earlier, "Who is Seattle's mayor?" will have better search results than "Who is Microsoft's mayor?", allowing the correct reformulation to be selected.

The search results are also a great indicator for whether a reformulation is needed. Consider the example of "When was Microsoft founded? Who founded it?", there aren't good search results for "Who founded it?" and the search results for "Who founded Microsoft?" are much better, so the "it" will be replaced with "Microsoft". But for the example of "When was Microsoft founded? How long does it take bruised ribs to heal?", the original question "How long does it take bruised ribs to heal?" has better search results than "How long does Microsoft take bruised ribs to heal?", so the "it" will not be replaced.

Search engine web knowledge essentially provides a world model that informs our system about which questions make sense, allowing for precise reformulations. On a human created evaluation set that includes more natural language conversational questions than what is typically found in the search engine logs dataset, our CQU system correctly reformulated 46% of conversational questions while correctly preserving 97% of non-conversational questions.

4 SEARCH POWERED CONVERSATIONS

The system described so far in the paper helps with one way conversational interactions between humans and machines, where the human is the questioner and the machine is the respondent. However, for many information seeking tasks, it is more natural to have two way conversations where the respondent can ask questions back to the questioner in order to better satisfy the questioner's information need. This sort of interaction feels even more natural when the information seeking task is carried out through the aforementioned conversational technologies (chatbots, digital personal assistants, smart home devices).

For example, consider the case where the user simply asks "Microsoft", how do we know what information the user is seeking about Microsoft? The stock price, the customer service number, a general description, etc.? A good default answer is to show a short description of Microsoft, but ideally it would be better to ask the user to clarify what he/she is asking for. To enable such conversations with the user, we can again leverage the power of search engine web knowledge. The results for the query "Microsoft" will include a variety of information about Microsoft, which we can use to figure out what clarification questions to ask the user.

Another example is if the user asks "hiking trails near me". A commercial search engine will show a list of relevant hiking trails. On screen based experiences such as laptops and phones, the user can easily navigate through and explore this list, but on voice only experiences or experiences with limited screen space, such as personal assistants and smart home devices, it would be necessary to

explore this list via a conversation. An example of this exploration can start by the system asking the user which one of the hiking trails he/she is interested in finding out more about, and then follow up by asking the user which attribute of that trail he/she is interested in.

These examples demonstrate the importance of having guided conversations in order to convert the knowledge into bits of information and share them with users. Enabling these guided conversations using search engine web knowledge is an area that we will continue to investigate as search becomes more conversational.

5 CONCLUSION

We created a system that takes the first step towards generic conversational question understanding. Our CQU system is agnostic to the storage solution used to store the context, and also agnostic to the information retrieval system/dialogue agent used to respond to the questions. Therefore, it can be easily plugged into any scenario that is conversational and involves question answering. We also presented the benefit of having two way conversations in order to better satisfy users' information seeking needs during conversational scenarios. As can be seen from recent technology trends, these scenarios will become more and more prevalent, making such a CQU system and guided conversations to be crucial pieces for improved interactions between humans and machines.

REFERENCES

[1] K. Clark and C. D. Manning. 2016. Deep Reinforcement Learning for Mention-Ranking Coreference Models. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Austin, Texas, 2256–2262.

[2] B. Dhingra, L. Li, X. Li, J. Gao, Y. N. Chen, F. Ahmed, and L. Deng. 2016. End-to-End Reinforcement Learning of Dialogue Agents for Information Access. *CoRR* abs/1609.00777 (2016). http://arxiv.org/abs/1609.00777

[3] S. P. Singh, M. J. Kearns, D. J. Litman, and M. A. Walker. 2000. Reinforcement Learning for Spoken Dialogue Systems. In *Advances in Neural Information Processing Systems 12*, S. A. Solla, T. K. Leen, and K. Müller (Eds.). MIT Press, 956–962. http://papers.nips.cc/paper/1775-reinforcement-learning-for-spoken-dialogue-systems.pdf

[4] A. Sordoni, M. Galley, M. Auli, C. Brockett, Y. Ji, M. Mitchell, J. Y. Nie, J. Gao, and B. Dolan. 2015. A Neural Network Approach to Context-Sensitive Generation of Conversational Responses. *CoRR* abs/1506.06714 (2015). http://arxiv.org/abs/1506.06714

[5] Stanford University. [n. d.]. Stanford CoreNLP. ([n. d.]). Retrieved August 14, 2017 from http://corenlp.run/

[6] O. Vinyals, L. Kaiser, T. Koo, S. Petrov, I. Sutskever, and G. E. Hinton. 2014. Grammar as a Foreign Language. *CoRR* abs/1412.7449 (2014). http://arxiv.org/abs/1412.7449