

Mixed-Initiative Conversational System using Question-Answer Pairs Mined from the Web

Wilson Wong^{*}, Lawrence Cavedon, John Thangarajah, Lin Padgham
RMIT University
Melbourne, Australia
{wilson.wong,lawrence.cavedon,john.thangarajah,lin.padgham}@rmit.edu.au

ABSTRACT

One of the biggest bottlenecks for conversational systems is large-scale provision of suitable content. Our approach readily provides this without the need for custom-crafting. In this demonstration, we present the use of question-answer (QA) pairs mined from online question-and-answer websites to construct system utterances for a conversational agent. Our system uses QA pairs to formulate utterances that drive a conversation as well as for answering user questions as has been done in previous work. We use a collection of strategies that specify how and when the different parts of our question-answer pairs can be used and augmented with a small number of generic hand-crafted text snippets to generate natural and coherent system utterances.

Categories and Subject Descriptors

I.2.7 [Artificial Intelligence]: Natural Language Processing—*Discourse*; H.5.2 [Information Interfaces and Presentation]: User Interfaces—*Natural language*

1. INTRODUCTION

The content used by chatbots such as ALICE [1] is custom-crafted to be as generic and deflective as possible to cover the needs of open-ended conversations. This approach reduces the amount of content that the system needs despite the relatively broad range of user inputs. For this reason, this class of conversational systems is only capable of content-free, small talk. As for systems (e.g., virtual nurse [2], intelligent toy [3]) that require content of reasonable depth and breadth to support content-rich conversations, the task of custom-crafting becomes impossible.

In this demonstration, we show the use of question-answer (QA) pairs from community-driven question-and-answer websites such as *AskKids.com* and *Answers.com* as content for our conversational system. By nature, the QA pairs extracted from a particular source on a certain topic are dis-

^{*}Corresponding author.

jointed, in that they do not have any temporal or structural information that could immediately lend themselves to straightforwardly building conversations. Our system innovatively uses the QA pairs to engage the users in coherent conversations using a range of different *conversational strategies* such as ‘question asking’, ‘fact telling’ and ‘question answering’. The system is able to share the initiative with the human users for determining conversation content and direction using these strategies.

This work is an extension of our interactive question answering (IQA) system using QA pairs, which was presented as a demo in CIKM 2011 [4]. The main distinctions between our new conversational system and the IQA system are that (1) the IQA system is driven solely by the user’s information seeking needs, and (2) the IQA system always interpret inputs as questions and service them with answers. Overall, the key contributions of this demonstration lie in the conversational strategies that we describe that specify how the different parts of the QA pairs can be used and combined with a small number of generic hand-crafted fragments to generate natural system utterances. These system contributions, when interplayed with cooperative user inputs, produce seemingly coherent conversations.

We will briefly present the system components and the conversational strategies in Section 2. We then move on to discuss how these components and strategies operate with the help of an actual conversation that took place between a human user with the system in Section 3. We conclude with a demonstration plan.

2. SYSTEM COMPONENTS

An overview of the main components of our conversational system is shown in Figure 1. The system is able to either respond to user initiated input, or alternatively, pro-actively make an utterance to initiate or maintain a conversation. A brief description of each component is provided below.

Input analysis performs shallow parsing of input to extract weighted terms and the domain(s) they belong to (if appropriate). The input is analysed to determine if a question has been asked, or a request to stop the conversation has been provided using a dictionary-based approach. Analysis of input utterances is relatively shallow, deliberately avoiding any attempt at sophisticated natural language understanding for computational efficiency. The input is tagged with parts of speech using FastTag¹ for its speed. Noun

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$10.00.

¹markwatson.com/opensource

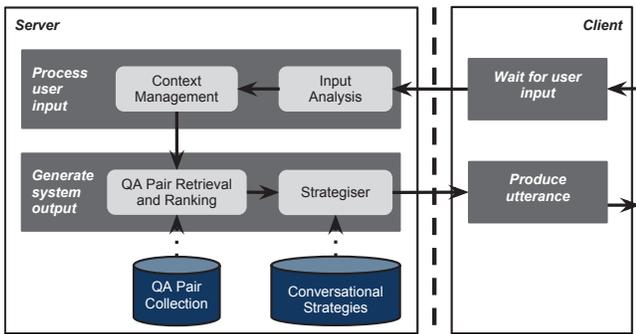


Figure 1: Conversational system architecture

phrases are then identified using simple regular expression patterns. Pronouns are resolved to the most prominent entities from previous inputs, using a method loosely based on the backward looking centering approach [5]. The phrases and words are then assigned weights to reflect their content bearing property using the deviation from Poisson approach [6]. Stopwords are removed by virtue of them being non-content bearing.

Context management updates conversational context using information obtained from the parsed input. Context, which is essentially a collection of weighted terms decayed over time, is used to select QA pairs which are sufficiently relevant to the user inputs to generate system utterances. This mechanism ensures that outputs provided by the system contribute to the appearance of coherence in a conversation. Each time a new input is processed, the corresponding weighted terms are combined into the existing conversational context. If a term is already present in the context, the weight associated with the recent occurrence is added to the term’s existing decayed weight in the context to reinforce what we perceive as an important term. If the term is absent from the context, the new term and its weight are added to the context.

QA pair retrieval and ranking determines the most relevant QA pair from the **QA pair collection**, based on the context. Initially, a set of all candidate QA pairs containing at least one term from the context is retrieved. The relevance of the candidates with respect to the conversational context is then determined in terms of (1) the extent of the overlap of words in the question and answer parts of QA pairs with those in the conversational context, and (2) the string similarity based on edit distance [7] between the question component and the user input. Regarding keyword overlap, the more highly weighted terms from the conversational context that appear in a QA pair, the higher the pair’s score will be. The sum of the weights of terms from the context that appear in the question as well as the answer of a QA pair is determined. This sum is augmented by term-location, where term matches in the questions are scored twice as high as matches in the answers. At the end of retrieval and ranking, the top scoring QA pair is selected for utterance generation.

Strategiser uses the **Conversational strategies** and the

```

SYSM[1]: what is a dugong? do you know the answer?
USER[2]: a kind of fish?
SYSM[3]: erm... the dugong is related to the manatee.
         they look and act allot like them. they
         live in the coasts of africa and australia.
USER[4]: manatee?
SYSM[5]: can you tell me, are manatees and dugongs
         curious?
USER[6]: what are manatees?
SYSM[7]: adult manatees are about 10 feet long and
         look kind of like walruses...
USER[8]: well, in that case, manatees and dugongs
         should look quite alike.
SYSM[9]: are there dugongs in russia? do you know?
USER[10]: nope.
SYSM[11]: umm... dugongs and manatees are members of
          the order sirenia, and while today they have
          no representatives in russian waters, up
          until the latter part of...
USER[12]: that’s an interesting piece of fact. i’ve
          never really seen a dugong.
SYSM[13]: can you tell me, what eats the dugong?
USER[14]: sharks! bad ones...
SYSM[15]: well, guilt. the dugongs know what they did
          and they know the damage they caused...
USER[16]: that’s rubbish
SYSM[17]: do you know, what are sharks?
USER[18]: sharks are big bad fish
SYSM[19]: well, sharks are fish. they pretty much eat
          anything that they come across including
          seals, fish, humans and other sharks...
USER[20]: that’s why they are bad!
SYSM[21]: do you know that blue sharks, who are often
          called the wolves of the sea, primarily feed
          on smaller fish as well as squids, but can
          take on larger species of prey if needed?
USER[22]: no, i didn’t know that. so blue sharks are
          more kind i assume.
...
  
```

Figure 2: The conversation between the system and one of the 11 participants (truncated for brevity).

selected QA pair to build output utterances. These utterances are placed on a queue for use as system outputs for the current as well as subsequent iterations. A new QA pair is only selected to generate further utterances if the queue is empty or the user poses a question. An example of this is discussed in Section 3. The strategies attempt to recreate the types of exchange that occur in human-to-human communications. There are two *reactive* and five *proactive* strategies. They can also be divided into *progression* or *conclusion* depending on their roles, to either progress or conclude a conversation. The strategies also use a small number of generic *hand crafted fragments* which include such things as speech disfluencies, prefixes and suffixes, to produce more natural conversation [8]. A summary of the seven strategies is provided in Table 1.

3. EXAMPLE CONVERSATION

We selected a conversation between the system and one of the participants from our evaluation, shown in Figure 2, to illustrate the functioning of the conversational system. The system starts the conversation using the SAQ strategy and the randomly selected topic “*dugong*” at line 1. The SAQ strategy only generates two system utterances, one using the

Strategy	Type	Role	Purpose
UAQ (User Asks Question)	RA	P	To answer an explicit question posed by the user.
UEC (User Ends Conversation)	RA	C	To conclude a conversation as per user request.
SSK (System Shares Knowledge)	PA	P	To simulate a knowledge-sharing conversation.
SAQ (System Asks Question)	PA	P	The system poses questions to the users.
UNI (User Not Interested)	PA	C	To conclude a conversation due to the user’s loss of focus or interest.
LTT (Lost Train of Thought)	PA	P	To recover from the system’s inability to select relevant QA pairs.
SEC (System Ends Conversation)	PA	C	To conclude a conversation after multiple LTT.

Table 1: A summary of strategies (RA: reactive, PA: proactive, C: conclusion, P: progression).

question and the second using the answer of the selected QA pair. The first utterance from this strategy was produced at line 1 and the second remains in the queue. The first user input on line 2 may appear to be a question but was not detected as such by the input analysis module. As such, the remaining one utterance in the queue is produced as the system’s output at line 3. The QA pair selected to generate these two system utterances is shown below:

Q: What is a dugong?
A: The dugong is related to the manatee. They look and act allot like them. They live in the coasts of Africa and Australia.

The question is modified using the “\$X. Do you know the answer?” fragment as the system’s first utterance, while the answer in the QA pair is prefixed with fragment 7 to form the system’s second utterance. The second user input at line 4 is also not recognised as a question. The strategist then selects SAQ again as the next strategy. This time, the “Can you tell me, \$X?” fragment is selected to modify the question “Are manatees and dugongs curious?” from a new QA pair to create the system’s output in line 5. The user, not knowing what “manatees” are, posed a more explicit question at line 6, which is this time detected by the system. The remaining utterance in the queue from the SAQ strategy is dropped and the UAQ strategy is selected to generate an answer at line 7. From line 9 to 20, the exchanges between the system and the user are still shaped by the SAQ strategy. Next, the strategist picks the SSK strategy to manipulate the selected QA pair below:

Q: What do blue sharks eat
A: Blue sharks, who are often called the wolves of the sea, primarily feed on smaller fish as well as squids, but can take on larger species of prey if needed.

Since there is only one sentence in the answer, this strategy and QA pair combination result in only one system utterance, at line 21.

4. DEMONSTRATION PLAN

We will start the demonstration with a few examples to illustrate the use of the seven strategies to progress and conclude conversations. In particular, we will highlight the importance of context management and the diversity of strategies for generating coherent and natural system utterances from disjointed QA pairs. We will switch off the context management module to show to the audience the rapid deterioration of coherence. The choice of strategies and the generic hand-crafted fragments will be reduced to demonstrate the increase in artificiality of system generated utterances. After the guided introduction, we will allow the

audience to freely interact with the system using a desktop browser as well as an Android application, where both are equipped with automated speech recognition (ASR) and text-to-speech (TTS) to support the option of speech-based interaction. We have prepared three videos² as a preview of our demonstration. The system featured in these videos uses a text-based interface. ASR and TTS were not yet fully integrated into the system during the recording of these videos.

Acknowledgment

This work is partially supported by the Australian Research Council and Real Thing Entertainment Pty. Ltd. under Linkage grant number LP110100050.

5. REFERENCES

- [1] R. Wallace. The anatomy of a.l.i.c.e. In R. Epstein, G. Roberts, and G. Beber, editors, *Parsing the Turing Test*. Springer, 2009.
- [2] T. Bickmore, L. Pfeifer, and B. Jack. Taking the time to care: Empowering low health literacy hospital patients with virtual nurse agents. In *Proceedings of the 27th CHI*, Boston, USA, 2009.
- [3] Z. Chen, C. Liao, T. Chien, and T. Chan. Animal companions: Fostering children’s effort-making by nurturing virtual pets. *British Journal of Educational Technology*, 42(1):166–180, 2011.
- [4] W. Wong, J. Thangarajah, and L. Padgham. Health conversational system based on contextual matching of community-driven question-answer pairs. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM Demo)*, pages 2577–2580, Glasgow, UK, 2011.
- [5] R. Mitkov. Outstanding issues in anaphora resolution. In *Computational Linguistics and Intelligent Text Processing*, pages 110–125. Springer, 2001.
- [6] K. Church and W. Gale. Inverse document frequency (idf): A measure of deviations from poisson. In *Proceedings of the ACL 3rd Workshop on Very Large Corpora*, pages 121–130, 1995.
- [7] V. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8):707–710, 1966.
- [8] M. Marge, J. Miranda, A. Black, and A. Rudnicky. Towards improving the naturalness of social conversations with dialogue systems. In *Proceedings of the 11th SIGDIAL*, University of Tokyo, Japan, 2010.

²<http://www.youtube.com/watch?v=SCShDas7XDO>
<http://www.youtube.com/watch?v=BBSbmHmrONk>
http://www.youtube.com/watch?v=CrXXbQMbO_0