On a Chatbot Conducting Dialogue-in-Dialogue

Boris Galitsky[1], Dmitry Ilvovsky[2], and Elizaveta Goncharova[2]
[1]Oracle Inc. Redwood Shores CA
[2]National Research University Higher School of Economics
boris.galitsky@oracle.com; dilvovsky@hse.ru; egoncharova@hse.ru

Abstract

We demo a chatbot that delivers content in the form of virtual dialogues automatically produced from plain texts extracted and selected from documents. This virtual dialogue content is provided in the form of answers derived from the found and selected documents split into fragments, and questions are automatically generated for these answers.

1 Introduction

Presentation of knowledge in dialogue format is a popular way to communicate information effectively. It has been demonstrated in games, news, commercials, and educational entertainment. Usability studies have shown that for information acquirers dialogues often communicate information more effectively and persuade stronger than a monologue most of times (Cox et al., 1999, Craig et al., 2000).

We demo a chatbot that delivers content in the form of virtual dialogues automatically produced from plain texts extracted and selected from documents. Given an initial query, this chatbot finds documents, extracts topics from them, organizes these topics in clusters according to conflicting viewpoints, receives users clarification on which cluster is most relevant to them, and provides the content for this cluster. This content is presented in the form of a virtual dialogue where the answers are derived from the found and selected documents split into fragments, and questions are automatically generated for these answers.

A virtual dialogue is defined as a multi-turn dialogue between imaginary agents obtained as a result of content transformation. It is designed with the goal of effective information representation and is intended to look as close as possible to a genuine dialogue. Virtual dialogues as search results turn out to be more effective means of information access in comparison with original documents provided by a conventional chatbot or a search engine.

2 Dialogue Construction from Plain Text

To form a dialogue from text sharing information or explaining how to do things, we need to split it into parts which will serve as answers. Then for each answer a question needs to be formed. The cohesiveness of the resultant dialogue should be assured by the integrity of the original text; the questions are designed to "interrupt" the speaker similar to how journalists do interviews.

We employ a general mechanism of conversion of conversion a text paragraph of various styles and genres into a dialogue form. The paragraph is split into text fragments serving as a set of answers, and questions are automatically formed for some of these text fragments. The problem of building dialogue from text T is formulated as splitting it into a sequence of answers $A = [A_1...A_n]$ to form a dialogue $[A_1, <Q_1, A_2>, ..., <Q_{n-1}, A_n>]$, where A_i answers Q_{i-1} and possibly previous question, and $\cup A_i = T$. Q_{i-1} needs to be derived from the whole or a part of A_i by linguistic means and generalization; also some inventiveness may be required to make these questions sound natural. To achieve it, we try to find a semantically similar phrase on the web and merge it with the candidate question.

The main foundation of our dialogue construction algorithm is Rhetorical Structure Theory (RST, Mann and Thompson, 1988). RST represents the flow of entities in text via Discourse Tree – a hierarchical structure that sets inter-relations between text fragments (Elementary Discourse Units, EDU): what elaborates on what, what explains what, what is attributed to what, what contradicts what, etc.

Rhetorical relations between the EDUs are usually binary and anti-symmetric, which defines the main unites (nucleus) and the subordinate ones (satellite). Thus, once we split a text into EDUs, we know which text fragments will serve as answers to questions: satellites of all relations. *Elaboration* rhetorical relation is default and *What*-question to a verb phrase is formed. *Background* relation yields another *What*-question for the satellite '*...as <predicate>-<subject>*'. Finally, *Attribution* relation is a basis of "What/who is source" question.

A trivial approach to question generation is simple conversion of a satellite EDU into a question. But it would make it too specific and unnatural, such as '*the linchpin of its strategy handled just a small fraction of the tests then sold to whom*?'. Instead, a natural dialogue should be formed with more general questions like '*What does its strategy handle?*'.

An example of converting a text into a virtual dialogue is shown in Figure 1. First, the text is split into EDUs. They act as answers in the virtual dialogue. The questions generated on their basis are shown in angle brackets and bolded. Each leave of the discourse tree determining an EDU starts with 'TEXT'. Rhetorical relations (in italics) are followed by the tags 'LeftToRight' or 'RightToLeft' specifying dependency direction between the units, or which of the following unit is a nucleus and a satellite.

---

elaboration (LeftToRight)
  attribution (RightToLeft)
<**who provided the evidence of responsibility?**>
TEXT: Dutch accident investigators say
  TEXT: that evidence points to pro-Russian rebels as being responsible for shooting down plane .
  contrast (RightToLeft)
    attribution (RightToLeft)
    TEXT: The report indicates
    joint
      TEXT: where the missile was fired from
      elaboration (LeftToRight)
        <**what else does report indicate?**>
        TEXT: and identifies
        TEXT: who was in control and pins the downing of the plane on the pro-Russian rebels .
      elaboration (LeftToRight)
        attribution (RightToLeft)
        TEXT: However , the Investigative Committee of the Russian Federation believes
        elaboration (LeftToRight)
          TEXT: that the plane was hit by a missile from the air
          <**where was it produced?**>
          TEXT: which was not produced in Russia .
        attribution (RightToLeft)
          TEXT: At the same time , rebels deny
          <**who denied about who controlled the territory?**>
          TEXT: that they controlled the territory from which the missile was supposedly fired
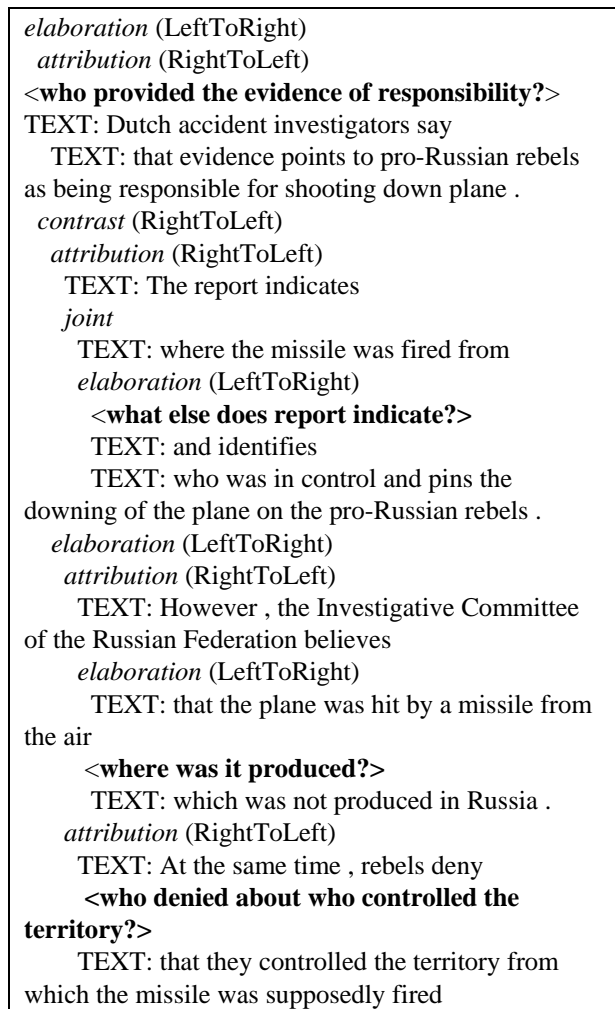

Figure 1: A discourse tree for a text paragraph with questions formulated for satellite EDUs as answers

The scheme of building a dialogue from text process is shown in Figure 2. Each paragraph of a document is converted into a dialogue via building a communicative discourse tree for it and then generating questions from its Satellite Elementary Discourse Units. Current chatbot is development of the previously built tool that conducted task-oriented conventional dialogues (Galitsky et al., 2017).
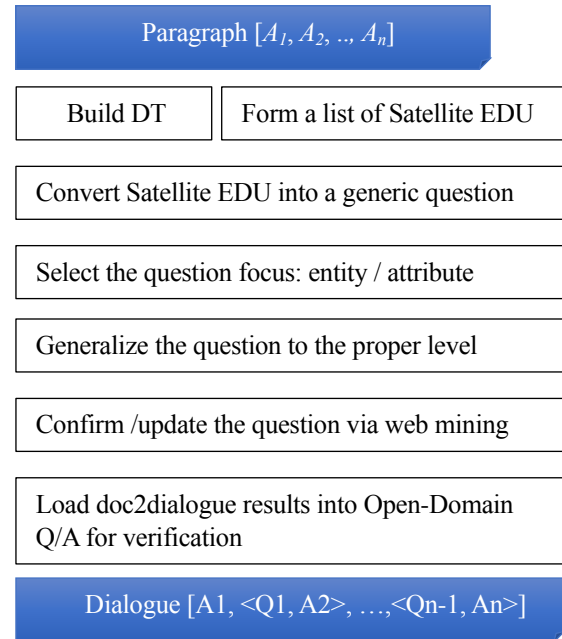

Figure 2: Scheme of dialog building process

3 Evaluation of Effectiveness

Evaluating the effectiveness of information delivery via virtual dialogues, we compare the conventional chatbot sessions where users were given plain-text answers, and the ones where users were given a content via virtual dialogues.

Table 1. Evaluation of comparative effectiveness of conventional and virtual dialogues

	Conventional dialogues			Virtual dialogues		
	# of iterations till found	# of iterations till decision	Coverage of exploration # of entities	# of iterations till found	# of iterations till decision	Coverage of exploration # of entities
Conv. only	4.6	6.3	10.8	-	-	-
Virtual only	-	-	-	4.1	6.0	13.7
Conventional followed by virtual	4.0	5.7	7.6	6.1	11.3	15.1
Virtual followed by convent.	5.6	7.1	12.3	3.7	7.0	11.5

The results on comparative usability of conventional dialogue and virtual dialogue are given in Table 1. We assess dialogues with respect to following usability properties averaged over the number of experiments:

The speed of arriving to the sought piece of information (first column). It is measured as a number

of iteration (a number of user utterances) preceding the final reply of the chatbot provided an answer wanted by the user. We measure the number of steps only if the user confirms that she accepts the answer.

The speed of arriving to decision to commit a transaction, such as purchase or reservation, or product selection (second column). A user is expected to accumulate sufficient information, and this information, such as reviews, should be convincing enough for making such decision. The less these values are the more relevant information was delivered via the dialogue.

We also measure **how many entities** (in linguistic sense) were explored during a session with the chatbot (third column). We are interested in how thorough and comprehensive the chatbot session is, how much a user actually learns from it. This assessment is sometimes opposite to the above two measures but is nevertheless important for understanding the overall usability of various conversational modes.

We do not compare precision and recall of search sessions with either dialogue mode since the same information is delivered, but in distinct modes.

In the first and second rows, we assess the stand-alone systems. One can observe that virtual dialogues take less iteration on average for information access (4.1 compared to 4.6) and a little less number of iterations for decisions than conventional dialogues do (6.0 and 6.3 respectively).

In the bottom two rows, we observe the usability of the hybrid system. Notice that the bottom row corresponds to the inverse architecture, where virtual dialog is followed by the conventional one. This scenario proceed from right to left, so, the first step's results are shown in three last columns of the table, then the values of the first three columns are calculated. When a conventional dialogue is followed by a virtual one, a lower portion of users is satisfied by the first step in comparison to the inverse architecture. Thus, the latter accounts for much less iteration required by user to be satisfied with the answer and make a final decision.

4 Sample ChatBot session

We present an exploratory session that combines information delivery in both the traditional textual answers (conventional dialogue) and a virtual dialogue form. The chatbot session is shown in Figure 3.

The dialogue starts from the user question, '*advantages and new features of 5G*'. The chatbot consults the sources (e.g. public URLs) and extracts the content from each page (or documents) expected to be relevant for the query. In this example seven URLs were processed, from domain-specific to general knowledge portals like Quora.com. Then the chatbot forms the list of topics extracted from these search results so that the user might select one of his interest.

Once the chatbot forms the topics for clarification of the user search intent, it shows them as a list. In Fig. 3 the list of topics proposed by the chatbot is underlined, the topics are numbered from 1to 5. The user selects his topic of interest and requests a specific answer via the topic number or the topic expression ('*next stage in technology*' or '*[5]*'). Once the answer is read, there are multiple options (*yes/more/ ... / virtual dialogue*):

- navigate to the next answer from the chatbot list;
- navigate to a specific answer from the chatbot list;
- reject this answer and attempt to reformulate the query;
- reduce search to a specified web domain (such as quota.com, for example);
- proceed in the same direction to more search results in the form of a virtual dialogue;
- accept the answer and conclude the session.

Bot > *ask a new question*
User > **advantages and new features of 5G**
Bot > exploring….
I believe these are the main topics of your query: is that what you meant? Please select
demonstrating the benefits of the technology [1].
wide range of people from student' [2]. 5g new radio nr specification' [3]. … next generation mobile networks alliance [4]. the next stage in mobile technology [5].

  User > **next stage in technology (or [5])**
  Bot > Put simply, it's the next stage in mobile technology. It follows 4G and 4G LTE and builds on what they offer, delivering everything at a faster speed.
  *Are you OK with this answer? **yes / more / no / specify [different topic] / reduce search to web domain / virtual dialogue***
User > **virtual dialogue**
Bot > This is what other people are chatting about your topic:

User1 > **What Are the Benefits of Technology?**
Bot1 > If you hopped over to Ericsson,
you 'd find a "5G avenue" with multiple exhibits demonstrating the benefits of the technology …
User1 > **How are these benefits leveraged by new tablets?**
Bot1 > Samsung kicked off its press conference not with its new tablets, but 5G is the next wireless network in telecom industry …
User2 > **Are the features right for them?**
Bot1 > It is important for people to start learning about 5G technology now so that they can decide if the features are right for them ...
User2 > **Which are the dual 4G VoLTE sim smartphones in India?**
Bot1 > The current 4G VoLTE smartphones with dual SIM slots that are available in the Indian markets may support 4G on single or both SIM slots.

Figure 3: Chatbot session with conventional and virtual dialogue

In the example the user selects the last option and the chatbot builds a virtual dialogue. It is a conversation among imaginary people whereas the topic stays the same, matching the original query. The virtual dialog is shown in the bottom frame (Fig. 3). As long as an imaginary chatbot responds to the same person, the dialog is intended to stay cohesive; coreferences in the follow-up questions are maintained. The main dialogue can be viewed as a one in the meta-level, and the object-level dialogue is naturally embedded into the meta-level one.

Now the user can either browse the built virtual dialogue or search it to find a fragment of conversation which is relevant to the user current exploration intent. If the user types the query '*Are the features right for me?*', he is directed to the virtual dialogue fragment where some other users are discussing if the technology is '*right for them*'. The search matches the query either against the fragments of an original text, generated questions, or both.

5 Related Work and Conclusions

(Piwek et al 2007) were pioneers of automated construction of dialogues, proposing Text2Dialogue system. The authors provided a theoretical foundation of the mapping that the system performs from RST structures to Dialogue representation structures. The authors introduced a number of requirements for a dialogue generation system (robustness, extensibility, and variation and control) and reported on the evaluation of the mapping rules.

An important body of work concerns tutorial dialogue systems. Some of the work in that area focuses on authoring tools for generating questions, hints, and prompts. Typically, these are, however, single utterances by a single interlocutor, rather than an entire conversation between two agents. Some researchers have concentrated on generating questions together with possible answers such as multiple choice test items, but this work is restricted to a very specific type of Q/A pairs (Mitkov et al 2006).

Dialogue acts are an important source which differentiates between a plain text and a dialogue. Proposed algorithm of virtual dialogues can assist with building domain-specific chatbot training datasets. Recently released dataset, DailyDialog (Li et al., 2017), is the only dataset that has utterances annotated with dialogue acts and is large enough for learning conversation models.

We proposed a novel mode of chatbot interaction via virtual dialogue. It addresses sparseness of dialogue data on the one hand and convincingness, perceived authenticity of information presented via dialogues on the other hand. We quantitatively evaluated improvement of user satisfaction with virtual dialogue in comparison to regular chatbot replies and confirmed the strong points of the former. We conclude that virtual dialogue is an important feature related to social search to be leveraged by a chatbot.

Chatbot demo videos (please, check 10 min video) and instructions on how to use it are available at https://github.com/bgalitsky/relevance-based-on-parse-trees in the "**What is new?**" section.

References

Mann, William and Sandra Thompson. 1988. *Rhetorical structure theory: Towards a functional theory of text organization. Text* - Interdisciplinary Journal for the Study of Discourse, 8(3):243–281.

Joty, Shafiq R, Giuseppe Carenini, Raymond T Ng, and Yashar Mehdad. 2013. Combining intra-and multi- sentential rhetorical parsing for document-level discourse analysis. In *ACL (1)*, pages 486–496.

Galitsky, B, Ilvovsky, D. and Kuznetsov SO. 2015. Rhetoric Map of an Answer to Compound Queries. ACL-2, 681–686.

Kipper, K. Korhonen, A., Ryant, N. and Palmer, M. 2008. A large-scale classification of English verbs. Language Resources and Evaluation Journal, 42, pp. 21-40.

Cox, R J. McKendree, R. Tobin, J. Lee, and T. Mayes. Vicarious learning from dialogue and discourse: A controlled comparison. Instructional Science, 27:431– 458, 1999.

Craig, S, B. Gholson, M. Ventura, A. Graesser, and the Tutoring Research Group. Overhearing dialogues and monologues in virtual tutoring sessions: Effects on questioning and vicarious learning. International Journal of Artificial Intelligence in Education, 11:242–253, 2000.

Yi, L., Ji Y, and Mari Ostendorf. 2016. LSTM based conversation models. arXiv preprint arXiv:1603.09457 .

Li Y, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. Dailydialog: A manually labelled multi-turn dialogue dataset. arXiv preprint arXiv:1710.03957.

Piwek, Paul; Hernault, Hugo; Prendinger, Helmut and Ishizuka, Mitsuru (2007). T2D: Generating Dialogues Between Virtual Agents Automatically from Text. Lecture Notes in Artificial Intelligence, Springer, Berlin Heidelberg, pp. 161–174.

Mitkov R, L. A. Ha, and N. Karamanis. A computer-aided environment for generating multiple-choice test items. Natural Language Engineering: Special Issue on using NLP for Educational Applications, 12(2):177–194, 2006.

Boris Galitsky and Dmitry Ilovsky. 2017. Chatbot with a discourse structure-driven dialogue management. EACL System Demonstrations.