

On a Chatbot Providing Virtual Dialogues

Boris Galitsky¹, Dmitry Ilvovsky², and Elizaveta Goncharova²

¹Oracle Inc. Redwood Shores CA

²National Research University Higher School of Economics

boris.galitsky@oracle.com; dilvovsky@hse.ru; egoncharova@hse.ru

Abstract

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

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 from the user clarification on which cluster is most relevant to her opinion, and provides the content for this cluster. This content is provided 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.

Once the proper piece of content is identified, users frequently like to consume it in the form of frequently asked question pages, discussion forums and blogs, rather than formal lengthy document. However, for the majority of knowledge domains, from legal and medical to

engineering, most reliable information is only available as documents and web pages. Hence we convert plain documents into dialogues, imitating multiple people conversing on the specific topic of interest.

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 Related Systems

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 question–answer pairs (Mitkov et al., 2006).

Conversion a text into a dialogue is different from the dialogue generation problem; the former is a training set-based foundation for the latter. Response generation for dialogue can be viewed as a source-to-target transduction problem. (Sordoni et al., 2015) rescues the outputs of a phrasal machine translation-based conversation system with a neural model incorporating prior context. Recent progress in sequence-to-sequence models has been leveraged to build end-to-end dialogue systems that firstly applies an utterance message to a distributed vector representation using an encoder, then generates a response from this representation.

(Li et al., 2016) simulate dialogues between two virtual agents, using policy gradient methods to reward sequences that display three useful conversational properties: “informativity”, coherence, and ease of answering.

We measured comparable dialogue effectiveness, properties such as the speed of arrival to a search result, a decision and domain coverage, in the current study.

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.

3 Demo Description

3.1 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 a paragraph of text 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 from some of these text fragments. The problem of building dialogue from text T is formulated as splitting it

into a sequence of answers $A = [A_1 \dots A_n]$ to form a dialogue

$$[A_1, \langle Q_1, A_2 \rangle, \dots, \langle Q_{n-1}, A_n \rangle],$$

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: what elaborates on what, what explains what, what is attributed to what, what is contradicting what, etc. Such text fragments are called elementary discourse units (EDUs). Most rhetorical relations are binary anti-symmetric, specifying which EDU has more important (nucleus) compared to less important (satellite).

A dialogue is formed from text by the following rule: once nucleus EDU is finished, and before satellite EDU starts, questions against this satellite EDU is inserted. In terms of dialogue flow between a text author and a person asking question, the latter “interrupts” the author to ask his question such that the satellite EDU and possibly consecutive text would be an answer to this question. The question is supposed to be about the entity from the nucleus, but this nucleus does not contain an answer to this question. The person asking questions only interrupts the text author when his question sounds suitable; it does not have to be asked for any nucleus-satellite transition.

Once we split a text into EDUs, we know which text fragments will serve as answer 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 $\langle \text{predicate} \rangle$ - $\langle \text{subject} \rangle$ ’. Finally, *Attribution* relation is a basis of “What/who is source” question.

A trivial approach to question generation would be to just convert 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 Fig. 1. Answers are obtained by splitting text into EDUs, and questions are inserted in text before satellite EDUs. Questions are shown in angle brackets and bolded. Each rhetorical relation in this example such as contrast ranges over a nucleus and a satellite. Each leave of this discourse tree starts with 'TEXT' (Fig. 1).

The reader of a virtual dialogue might feel that the interviewer is guessing what the speaker is going to answer for each question. A discourse – tree based approach does not deliver most natural dialogues however it is a systematic method of building ones without distorting the logical flow of answers.

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 .

Fig. 1. A discourse tree for a paragraph of text with questions formulated for satellite EDUs as answers

3.2 System Architecture

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

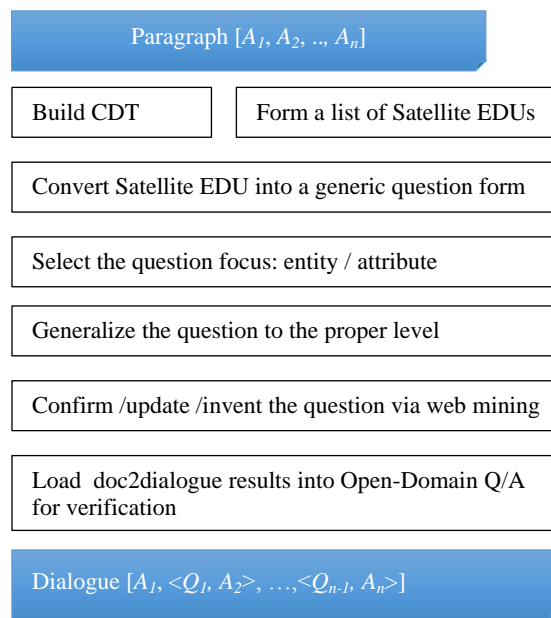


Fig. 2. System architecture

4 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 content via virtual dialogues.

We present the results on comparative usability of conventional dialogue and virtual dialogue. We assess dialogues with respect to following usability properties.

The speed of arriving to the sought piece of information. It is measured as a number of iteration (a number of user utterances) preceding the final reply of the chatbot that gave 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 a decision to commit a transaction such as purchase or reservation or product selection. A user is expected to accumulate sufficient information,

and this information such as reviews should be convincing enough for making such decision;

We also measure **how many entities** (in linguistic sense) were explored during a session with the chatbot. 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.

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

	Conventional dialogues			Virtual dialogues		
	# of iterations till found	# iterations till decision	Coverage of exploration # of entities	# of iterations till found	# iterations till decision	Coverage exploration # of entities
Conventional 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 conventional	5.6	7.1	12.3	3.7	7.0	11.5

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 and about the same number of iterations for decisions as conventional dialogues do.

Notice that the bottom row, the chat scenario proceeds from right to left. In the bottom two rows, we observe the usability of the hybrid system. 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, where virtual is followed by conventional.

5 Sample Chatbot Session

We present an exploratory session that combines information delivery in the form of traditional textual answers (conventional dialogue) and also in a form of a virtual dialogue.

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.

Fig. 3. Chatbot session with conventional and virtual dialogue

The dialogue starts from the user question, 'advantages and new features of 5G'. The chatbot consults the sources (in this case, 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. 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.

The user selects his topic of interest and requests a specific answer via the topic number of the topic expression. Once the answer is read, there are multiple options:

- 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.

The user selects the last option and the chatbot builds a virtual dialogue. It is a conversation between an imaginary people but the topic stays the same, matching the original query. Virtual dialogues are shown in frames. 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. The user now types the query ‘*Are the features right for me?*’ and gets 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.

6 Conclusions

We proposed a novel mode of chatbot interaction via virtual dialogue. It addresses sparseness of dialogue data on 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.

7 Acknowledgements

The article was prepared within the framework of the HSE University Basic Research Program and funded by the Russian Academic Excellence Project '5-100'.

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.
- Alessandro Sordani, Michel Galley, Michael Auli, Chris Brockett, Yangfeng Ji, Meg Mitchell, Jian-Yun Nie, Jianfeng Gao, and Bill Dolan. 2015. A neural network approach to context-sensitive generation of conversational responses. In *Proc. of NAACL-HLT*, May–June.

- 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.
- 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 Ilvovsky. 2017. Chatbot with a discourse structure-driven dialogue management. *EACL System Demonstrations*.