DOC2DIAL: a Framework for Dialogue Composition Grounded in Documents

Song Feng Kshitij Fadnis Q. Vera Liao Luis A. Lastras
IBM Thomas J. Watson Research Center.
sfeng,kpfadnis@us.ibm.com, vera.liao@ibm.com, lastrasl@us.ibm.com

Abstract

We introduce DOC2DIAL, an end-to-end framework for generating conversational data grounded in given documents. It takes the documents as input and generates the pipelined tasks for obtaining the annotations specifically for producing the simulated dialog flows. Then, the dialog flows are used to guide the collection of the utterances via the integrated crowdsourcing tool. The outcomes include the human-human dialogue data grounded in the given documents, as well as various types of automatically or human labeled annotations that help ensure the quality of the dialog data with the flexibility to (re)composite dialogues. We expect such data can facilitate building automated dialogue agents for goal-oriented tasks. We demonstrate DOC2DIAL system with the various domain documents for customer care.

Introduction

There has been growing interest in using automated dialogue agents for domains such as customer care. However, one bottleneck is the lack of chat logs that show how the agents use the given documents to assist end users. Meanwhile, enterprises and organizations often own a large amount of business documents that aim to address customers' requests Taken together, a promising solution is to build data for training the machine assisted agents that could perform task-oriented dialogues supported by the business documents.

In task-oriented dialogues for customer care, a recurrent theme is a diagnostic process – identifying the contextual conditions to retrieve the most relevant solutions. Meanwhile, the business documents often contain similar information with prior conditions. For instance, the sample document in STEP 1 in Figure 3 contains the information for an agent to use in the dialogue at STEP 2, where P (S) denotes text spans labeled a precondition (solution). Thus, we hypothesize that an essential capability for a dialogue agent to perform goal-oriented information retrieval tasks should be able to recognize the preconditions and their associated solutions covered by the given documents and then use them to carry out the diagnostic interactions. Towards this goal, we introduce DOC2DIAL, a novel end-to-end framework

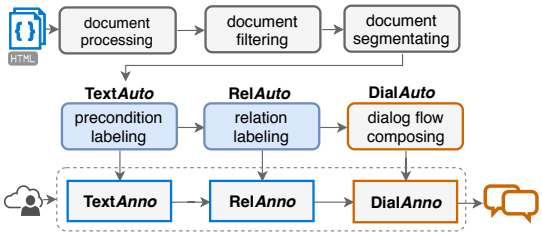

Figure 1: An Overview of DOC2DIAL Framework

for generating dialog flows for collecting task-oriented dialogues grounded in the given documents. We guide our investigation with the following principles: 1) aiming to identify the document content that corresponds the solution to a user's request as well as the prerequisites specified. 2) the generation of dialog flows should be closely related to the given documents without relying on heavily supervised or handcrafted work. 3) the data tasks should be easy to scale – feasible to crowdsourcing platforms and could be updated with respect to changes in the documents.

In this work, we propose a pipeline of three interconnected tasks: (1)TEXTAUTO: labeling text spans as preconditions or solutions in a given documents; (2) RELAUTO: identifying the relation(s) between these preconditions or solutions; (3)DIALAUTO: simulating dialog flows based on the linked preconditions/solutions and applying them to guide the collection of human generated utterances. The system is also integrated with the annotation tools TEXTANNO/RELANNO/DIALANNO with sophisticated quality control features. For the dialogue collection DIALANNO, it provides an asynchronized process that allows crowd workers to work on the creation and evaluation of individual turns without the constraints of timing or having a dialog partner.

To our knowledge, ours is the first end-to-end pipeline framework from extracting complex preconditions and solutions that present in the documents, to generating task-oriented dialogues that fit different scenarios. The most closely related work is ShARC (Saeidi et al. 2018) as it shares similar goals. However, it is only on asking boolean follow-up questions on precondition. Our work is also largely related to conversational QA such as CoQA (Reddy, Chen, and Manning 2019) and QuAC (Choi et al. 2018).

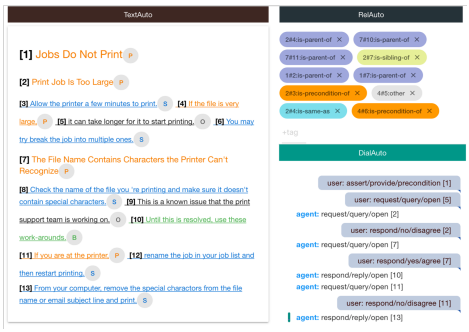

Figure 2: An illustration of DOC2DIAL-Auto UI

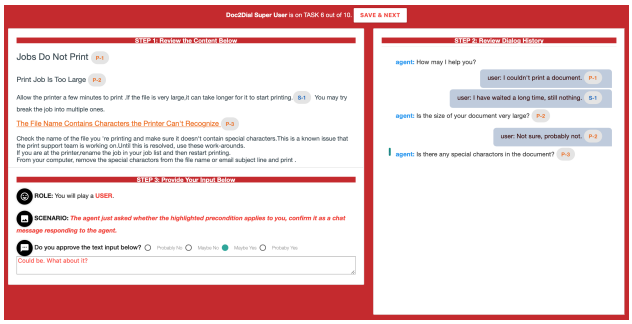

Figure 3: UI of DIALANNO

System

DOC2DIAL in Figure 1 consists of document processing and text labeling modules for analyzing the textual content and structure of the given documents. The labeled texts are used to generate the downstream tasks used in the three annotation tasks of DOC2DIAL pipeline for collecting dialogues.

Document Processing and Automated Labeling It first prepares the given documents for the pipeline tasks. We obtain various syntactic-semantic analysis ranging from computational linguistic statistics to HTML-based tree structures. For instance, we apply constituency parsing results for splitting long sentences to text spans. We also extract sub-clauses with certain discourse connectives (e.g. "if", "as long as") via (Das et al. 2018) for identifying linguistic indicators of preconditions. We also try to capture the outline patterns embedded in document structures (Mukherjee et al. 2003). The ones that are well structured and clearly written with descriptive sub-titles and the discourse connectives are considered as good candidates for generating dynamic dialog flows. The system is equipped with TEXTAUTO that automatically labels text spans based on the syntactic-semantic indicators mentioned above. It also employs heuristics based on the HTML tree structures and text proximity for the relation linking via RELAUTO. Such labels are used to generate dialog flows without human labels; also as pseudo-gold labels for quality control for the crowdsourced tasks TEXTANNO and RELANNO. The human labels could be used to further improve TEXTAUTO and RELAUTO.

Dialog Flow Composition generates the dialog flows for DIALANNO from the labels of precondition/solution text and their relations obtained via previous two tasks. For each turn, the dialog scene is determined by three factors, i.e., selected text span content, role and dialog act, which are determined sequentially. The dynamics of the dialog flows are introduced by varying the three aforementioned factors that are constrained by the relations collected via RELANNO. First, we randomly select content from a candidate pool of preconditions and solutions identified in the document, which is updated after every scenario generated. The general rule for updating the candidate pool is to eliminate preconditions/solutions that are already verified or eliminated. Then role is randomly selected between agents and end users. For our pilot study, we mainly consider dialog acts corresponding to the preconditions/solutions such as `request/query/open`, `respond/no/disagree`. The dialogue ends when the candidate pool is empty, no solution is found, or it reaches the preassigned maximum turns.

Discussions

For our pilot study, we demonstrate how to apply DOC2DIAL to about 2000 proprietary documents on topics from technical trouble shooting to policy guidance for customer care. We generated 5 dialog flows with more than 5 turns per document. Next, we evaluated sampled dialog flows by asking the crowd worker if the highlighted text matched the given dialogue scenario of a turn in a dialog flow. 67% of the turns were labeled as "match". Most of the mismatches were due to disagreement on the precondition/solution labeled in earlier steps. Figure 3 shows the sample task on collecting the 6-th turn of a dialog flow of 7 turns by providing the dialogue scenario and the chat history. The sample dialogues show that when the crowd contributors were able to understand the selected text in the context of the document, they could properly interpret the assigned dialog scene to produce utterances.

References

Choi, E.; He, H.; Iyyer, M.; Yatskar, M.; Yih, W.-t.; Choi, Y.; Liang, P.; and Zettlemoyer, L. 2018. Quac: Question answering in context. *arXiv preprint arXiv:1808.07036*.

Das, D.; Scheffler, T.; Bourgonje, P.; and Stede, M. 2018. Constructing a lexicon of english discourse connectives. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, 360–365.

Mukherjee, S.; Yang, G.; Tan, W.; and Ramakrishnan, I. 2003. Automatic discovery of semantic structures in html documents. In *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.*, 245–249. IEEE.

Reddy, S.; Chen, D.; and Manning, C. D. 2019. Coqa: A conversational question answering challenge. *Transactions of the Association for Computational Linguistics* 7:249–266.

Saeidi, M.; Bartolo, M.; Lewis, P.; Singh, S.; Rocktäschel, T.; Sheldon, M.; Bouchard, G.; and Riedel, S. 2018. Interpretation of natural language rules in conversational machine reading. *arXiv preprint arXiv:1809.01494*.