Automatic Dialogue Flow Extraction and Guidance

Patrícia Ferreira

May 31, 2022

Automatic Dialogue Flow Extraction and Guidance

Patrícia Ferreira*

CISUC, Universidade de Coimbra
patriciaf@dei.uc.pt

Abstract. Nowadays, human chat service agents are frequently replaced by conversation software agents, designed to communicate with humans by means of natural language, often based on Artificial Intelligence, namely Natural Language Processing (NLP) and Machine Learning (ML). This work will begin by identifying and annotating dialogue sets, written in Portuguese. It consists of researching and implementing a solution with the objective of aiding communication between participants, suggesting appropriate responses, thus anticipating their interventions. This guidance can be supported by the history of interactions, where information is extracted and frequent dialog flows are discovered, allowing a representation of them to guide humans. The approach will be divided into three components: Extraction to process dialogues and use the information to describe interactions; Representation for the discovery of the most frequent dialogue flows, represented by graphs of interaction; Guidance to guide the agent during a new dialogue.

Keywords: Natural Language Processing · Dialog Analysis · Dialog Information Extraction · Representation of dialog flows · Assisted guidance

1 Background and Related Work

A dialog is composed of utterances, which instantiate dialog acts (DAs), that is, abstract representations of intentions. There are several dialogue datasets, mainly for English [2], however, this work will focus on Portuguese, where public dialogue datasets are scarce.

There are several approaches for automatic classification of DAs (DAC) [1]. Most are based on supervised learning, with models trained in dialog datasets where the DAs are manually annotated [3]. Others use traditional classification, considering utterances in isolation [3][4]. However, since there may be a dependency between the current interaction and previous ones, DAs can be tackled as a sequence classification problem, with methods such as Hidden Markov Models (HMM)[5] or Conditional Random Fields (CRF) [6]. Improvements can be achieved if utterances are encoded by a transformer network-based language model (e.g., BERT), which performs tasks such as Natural Language Inference

* Supervised by Hugo Gonçalo Oliveira, Catarina Silva e Ana Alves

and automatic question answering [7]. DAs and transition graphs between them are useful representations of dialogs. When applied to large sets of dialogs, models trained will allow the discovery of different types of interactions and the most common dialog flows. These flows can even be discovered without annotating DAs, through unsupervised approaches (clustering) [8]. Furthermore, as real-time call monitoring, these flows will be useful for classifying the current dialog and recommending interactions to the human [9] [10].

2 Methodology

Overall, this work consists of researching, implementing and testing a solution that aims to improve communication between participants in a dialog by guiding their actions, which can be supported by previous interactions, where information should be automatically extracted and where frequent dialog flows.

Besides task-oriented dialogues experiments will be extended to conversations between common users, for example, in a social network, where it will be possible to analyze communication trends.

The experiments will be conducted with data in Portuguese, which will be a differentiating factor from the state of the art. They will also be limited to written text, i.e., written conversations or transcripts of oral communication.

The data can be created following the Wizard-of-Oz (WOZ) [11], where a conversation takes place between two participants with different roles. One plays the role of an ordinary user who is assigned a certain task and must interact, using natural language, with another that will have access to information about the domain (e.g., a database) and will be able to provide appropriate answers.

Data can also result from available transcriptions of existing dialogues into Portuguese, such as CORAA [12]; from customer support services, such as conversations with telecom operators on Twitter; or from movie subtitles [13]. One last possibility will be the translation of English datasets (e.g., DailyDialogue [14], MultiWOz [15]) into Portuguese, from where existing annotation can be imported.

The data will be used in the development of a framework consisting of three components. The first will process real dialog transcripts and extract useful information from them to represent the various interactions, such as keywords, entities or actions. The extraction of some of these items may resort to an NLP pipeline [16], but some additional development may be necessary, considering the language and the type of text. The extracted information can be used to better describe the utterances, by classifying intentions and filling slots. However, the performance of these tasks is usually based on supervised learning, which implies data annotation.

The extracted information can also be used to group similar utterances, using clustering. This process can also resort to Sentence Embedding techniques [17].

The second component will aim at discovering the most frequent dialog flows, represented by graphs, where the vertices represent speech classes or clusters, and the arcs represent transitions between them, with associated probabilities. In this component one can apply the classification of interactions into more generic classes or, if there is a lack of data to make the system less domain-dependent, perform a clustering that approximates these acts.

Finally, the guidance component will take advantage of the representation of flows to guide the human. In each interaction, previous interactions will be considered to suggest appropriate responses, while anticipating the next interactions.

A final evaluation of the results of integrating the three components into the framework should be done, as the approaches explored will be evaluated on the data gathered and created using metrics for classification when annotations are produced, or metrics for clustering when they are not. Manual evaluations may also be required.

3 Objectives and Expected Results

The main goal of this work is to investigate and develop approaches to improve communication between interlocutors in a dialog, in Portuguese.

NLP techniques will be explored, focusing on dialog modeling [18], in order to, based on the history of interactions, identify the most common ones in each application domain, discover and interpret flows, and take advantage of the latter to guide interlocutors, who may thus anticipate their interventions.

We aim at approaches, applicable to written dialogues, e.g. between users of a social network, or to transcriptions of task-oriented dialogues (e.g., call-center) to assist call-center operators in providing more efficient service.

We believe that the work will result in innovative approaches, and highlight the fact that, regardless of the possible adaptation to other languages, it will be focused on Portuguese. The work will be divided into four specific objectives, namely: identify and create sets of dialogues, in Portuguese; study, develop and experiment with approaches for extracting structured dialogue information from the various interactions; study, develop and experiment with approaches for representing interactions and dialog flows extracted from those interactions; study, develop and experiment with approaches for guiding the human by exploiting the knowledge extracted from dialogues, dialogue type and interactions, and common flows.

To achieve the defined goals, the following tasks were established: 1. To deepen the study of the state of the art; 2. Definition of the data to be used; 3. Exploration of approaches for information extraction, for representing dialog flows and for dialogue guidance; 6. Proposed framework encompassing the approaches explored; 7. Testing and final evaluation; 8. Writing of the thesis and dissemination papers. The approaches resulting from task 3 will be evaluated independently during their respective tasks, but a final evaluation of the results of their integration into the framework will be required. This integration will encompass the approaches explored for each component. The experiences are regularly described in writing in the doctoral thesis. We further believe that from tasks 2 to 7 will result contributions relevant to write scientific papers.

References

1. Ruizhe Li, Chenghua Lin, Matthew Collinson, Xiao Li, and Guanyi Chen. A dual attention hierarchical recurrent neural network for dialogue act classification. arXiv preprint arXiv:1810.09154, 2018
2. Hugo Gonçalo Oliveira, Patrícia Ferreira, Daniel Martins, Catarina Silva, and Ana Alves. A brief survey on textual dialogue corpora. In Proceedings of the 13th International Conference on Language Resources and Evaluation (LREC 2022), page In press, 2022.
3. Srinivas Bangalore, Giuseppe Di Fabbrizio, and Amanda Stent. Learning the structure of task driven human–human dialogs. IEEE Transactions on Audio, Speech, and Language Processing, 16(7):1249–1259, 2008.
4. Konstantina Christakopoulou, Filip Radlinski, and Katja Hofmann. Towards conversational recommender systems. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pages 815–824, 2016.
5. [5] R Eddy Sean. What is a hidden markov model. Nat Biotechnol, 22(10):1315–1316, 2004.
6.] Harshit Kumar, Arvind Agarwal, Riddhiman Dasgupta, and Sachindra Joshi. Dialogue act sequence labeling using hierarchical encoder with crf. In Proceedings of the aaai conference on artificial intelligence, volume 32, 2018.
7. Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer. In Proceedings of the 28th ACM international conference on information and knowledge management, pages 1441–1
8. Ronald M Kaplan, Jill Burstein, Mary Harper, and Gerald Penn. Human language technologies: the 2010 annual conference of the north american chapter of the association for computational linguistics. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2010.
9. Su Nam Kim, Lawrence Cavedon, and Timothy Baldwin. Classifying dialogue acts in one-on-one live chats. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pages 862–871, 2010.
10. Michael McTear. The role of spoken dialogue in user–environment interaction. Human-Centric Interfaces for Ambient Intelligence, pages 225–254, 2010.
11. Anders Green, Helge Huttenrauch, and K Severinson Eklundh. Applying the wizard-of-oz framework to cooperative service discovery and configuration. In RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759), pages 575–580. IEEE, 2004.
12. Arnaldo Candido Junior, Edresson Casanova, Anderson Soares, Frederico Santos de Oliveira, Lucas Oliveira, Ricardo Corso Fernandes Junior, Daniel Peixoto Pinto da Silva, Fernando Gorgulho Fayet, Bruno Baldissera Carlotto, Lucas Rafael Stefanel Gris, et al. Coraa: a large corpus of spontaneous and prepared speech manually validated for speech recognition in brazilian portuguese. arXiv preprint arXiv:2110.15731, 2021.
13. Pierre Lison and J¨org Tiedemann. Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles. European Language Resources Association, 2016.
14. Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. Dailydialog: A manually labelled multi-turn dialogue dataset. arXiv preprint arXiv:1710.03957, 2017.

15. Pawe l Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Inigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gasic. Multiwoz–a large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling. arXiv preprint arXiv:1810.00278, 2018.
16. Xavier Schmitt, Sylvain Kubler, J´er´emy Robert, Mike Papadakis, and Yves Le-Traon. A replicable comparison study of ner software: Stanfordnlp, nltk, opennlp, spacy, gate. In 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), pages 338–343. IEEE, 2019.
17. Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert networks. arXiv preprint arXiv:1908.10
18. Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. Dialogue act modeling for automatic tagging and recognition of conversational speech. Computational linguistics, 26(3):339–373, 2000.