Automatic Dialog Flow Extraction and Guidance

Patrícia Ferreira
CISUC, Univ. Coimbra
DEI, Univ. Coimbra
patriciaf@dei.uc.pt

Abstract

Today, human assistants are often replaced by chatbots, designed to communicate via natural language, however, some disadvantages are notorious with this replacement. This PhD thesis project consists of researching, implementing, and testing a solution for guiding the action of a human in a contact center. It will start with the discovery and creation of datasets in Portuguese. Next, it will go through three main components: Extraction for processing dialogs and using the information to describe interactions; Representation for discovering the most frequent dialog flows represented by graphs; Guidance for helping the agent during a new dialog. These will be integrated in a single framework. In order to avoid service degradation resulting from the adoption of chatbots, this work aims to explore technologies in order to increase the efficiency of the human's job without losing human contact.

1 Introduction

In the past, a consumer's only option for customer service was to speak directly with a service employee. Now, many customer interactions are handled by automated systems powered by artificial intelligence called chatbots (Tran et al., 2021).

During the last few years, there has been a growing interest in text-based chatbots. However, despite the market's enthusiastic predictions, chatting with this type of agent raises some technological limitations, directly involving the human side of the interaction (Rapp et al., 2021).

That said, this thesis project proposes to prevent call-center service degradation and customer dissatisfaction through the use of chatbots, taking advantage of technologies that can make a human's job more efficient without losing human contact.

This work consists of researching, implementing and testing a solution to aid communication between participants, suggesting appropriate responses, thus anticipating their interventions. This guidance can be supported by the history of interactions, where information is extracted from and frequent dialog flows are discovered, which may then be used for guiding humans engaging in new dialogs of the same kind. The approaches will be applied to task-oriented dialog transcriptions (e.g. call center), providing a more efficient and facilitated service.

It begins by identifying, collecting and annotating dialogs written in Portuguese to be used in the experimentation and make available to the community. We plan to tackle the problem with a three-component pipeline: Extraction, for processing dialogs, extracting information from them and classifying interactions; Representation of the most frequent dialog flows, by graphs of interaction classes; Guidance, for assisting a human agent during a new dialog. All components are often tackled in the scope of Dialog Modelling (DM) (Budzianowski et al., 2018) to allow the reproduction of aspects of a natural conversation. Research in the area of dialogs is currently booming, with interest in chatbots, but most systems are developed for English. Instead, this work has innovative potential in the area because it is targeted for Portuguese.

In the next section, important concepts for understanding this research are introduced and an overview of related work is given. It includes an introduction to Portuguese datasets, research work on chatbots and its limitations, with the remaining subsections divided according to the three components of the project. In section 3, the research proposal and the intended methodologies are presented. Finally, section 4 presents some preliminary experiments, using NLP tools and related extraction tasks.

2 Background and Related Work

This section starts with the presentation of a Portuguese dataset, then it is checked how chatbots and human-based customer service can be so different,

Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop, pages 112–122

and finally it is divided into the three components of the project where concepts and related work for each are found.

2.1 Datasets

One of the first objectives of this work is the identification or creation of datasets in Portuguese.

There are several dialog datasets, of different natures, mainly for English (Oliveira et al., 2022), however, this work will focus on Portuguese, where dialog datasets are scarce and thus it is possible to explore approaches for low resource scenarios.

Existing resources for Portuguese are composed of audios containing only read and prepared speeches, and there is a lack of datasets that include spontaneous speeches, essential in different applications. An exception is a new dataset in Portuguese designated as CORAA (Junior et al., 2021) that is composed of five different corpora of European and Brazilian Portuguese conversations. They tried to bridge the gap of lack of spontaneity and formal speech by having only real conversations.

2.2 Human agents and chatbots

Chatbots are the result of advances in Artificial intelligence (AI), in order to interact and respond with suggestions appropriate to certain needs (Shum et al., 2018).

Human-chatbot communication has notable differences in content and quality compared to human-human. The crucial difference is empathy, as chatbots are less capable of conversational understanding than humans. However, chatbots are gradually becoming more aware of their interlocutor's feelings (Adamopoulou and Moussiades, 2020).

The first known chatbot, developed in 1966, was Eliza [1] (Weizenbaum, 1966). Its purpose was to behave like a psychologist. It used simple patterns and user sentences returned in the form of a question. Its conversational ability was not very good, but it was enough to start the development of other chatbot systems (Bradeško and Mladenić, 2012).

Most chatbots tend to respond with the same message, have a very limited vocabulary, and often provide wrong information. To demonstrate the lack of language capabilities of chatbots, a comparison was made between chatbots responses and human responses (Feine et al., 2020), by analyzing an existing human chat dialog analyzed from the Conversational Intelligence Challenge 2 (ConvAI2[2]), where the lexical diversity was analyzed of all chatbot and human messages, through Part of Speech (PoS) and counted the adjectives, adverbs and verbs that are relevant for expressing emotion which is an inherently human ability. The results indicate that human users used 75% more adjectives, 65% more adverbs, and 76% more verbs than the ConvAI2 chatbots. Therefore, this work reveals that human language use is far from superior in terms of lexical and emotional diversity.

There are several solutions on the market that allow the development of chatbots such as DialogFlow (Sabharwal and Agrawal, 2020), Amazon Lex (Sreeharsha et al., 2022), Rasa (Sharma and Joshi, 2020), etc. Using one of these solutions, it is possible to develop a chatbot, through dialog flows for selecting responses or actions based on the identification of expressions that the agent should recognize, also called intents. However, these are limited to maintaining a dialog based on flows that are created manually. The limitations of chatbots motivate the need for human involvement.

2.3 Knowledge Extraction from Dialog

The extraction component should start by processing transcripts of real dialogs between humans, using an NLP pipeline (Tenney et al., 2019). This pipeline starts by segmenting the text into tokens and using a set of parsing processes such as Information Extraction (IE) (Grishman, 2019) or Semantic Parsing (Berant et al., 2013), the latter being the task of deriving a representation of meaning from the language sufficient for a given task, since IE of the text can be characterized as representing a certain level of semantic parsing.

It is also necessary to clean up the transcriptions used from what is not relevant to the creation of the dialog flow and segment it into individual utterances, from the two interlocutors, linking them together to create an objective dialog line. This process can be implemented using contextual models, based on Sentence Embeddings (Reimers and Gurevych, 2019), created from Deep Learning approaches (Li et al., 2018), and also from approaches for Intent Classification (Chen et al., 2019), which allow transforming words into vector representations and, thus, mapping the words used into known concepts or intentions.

It is also fundamental to identify entities, through Named Entity Recognition (NER) (Mo-

[1] http://psych.fullerton.edu/mbirnbaum/psych101/eliza.htm

[2] https://paperswithcode.com/dataset/convai2

hit, 2014), an IE task that consists in identifying and classifying only some types of information elements, called Named Entity (NE). DM (Bai et al., 2021) is a subarea of NLP that covers tasks aimed at learning how humans use this language to interact with each other, and exploiting it in computational applications. This typically includes intent recognition (Sukthankar et al., 2014), which maps utterances with responses or actions that the system has to perform, and can be used for dialog summarization (Goo and Chen, 2018; Liu et al., 2019), human assistance (De et al., 2021) in communication, prediction of the next interactions (Ritter et al., 2011) of a human user with a dialog system, among others.

Dialogues are sequences of utterances, commonly classified according to: intents, which represent the end-user's intents; or DAs, which represent the action performed by the speaker (Austin, 1962) and can be seen as more generic intents.

DAs function as action labels for the utterances in a given conversation (e.g., ask, explain, speak, request, etc.), thus helping to characterize intents and enabling a better understanding of conversations (Hoxha et al., 2016). On the other hand, an intent categorizes an end-user's intent for one conversation turn (Truong et al., 2004) and is usually more specific, depending on the given scenario. Thus, DAs recognition can be accomplished by identifying the function-related DAs of a single utterance or segment, unrelated to a specific domain or task. This is relevant during an ongoing conversation, as it allows for interpretation or knowledge extraction taking into account the intent and simplifies the identification of related segments in the dialog history. A dialog representation is a sequence of DAs that are useful for their interpretation by humans, conversational systems, or computational methods and for summarizing the conversation or predicting future utterances (Hoxha et al., 2016).

Dialog Act Classification (DAC) is useful for identifying patterns and extracting common flows in dialogs. Several approaches have been developed for automatic DAC. Most adopt a supervised approach, with models trained on dialog corpus where DAs are manually annotated (Bangalore et al., 2008). Some use traditional methods of classification considering the context of the dialog, using previous interactions, and capturing hierarchical relationships between tasks. In Sordoni et al. (2015) they formulate a neural network architecture for data-driven response generation trained from social conversations, in which the generation of responses is constrained by dialog utterances that provide contextual information.

Others methods adopt learning by considering utterances in isolation. However, since there may be a dependency between the current interaction and previous ones, that is, between consecutive utterances (e.g., usually after a question comes an answer) DAC should be approached as a sequential classification problem and not as a simple classification problem.

Hidden Markov Models (HMMs) (Stolcke et al., 2000) present time intervals, where the process evolves from one state to another, depending only on its last state. The hidden states of the model are the DA labels that generate the sequence of words. Another widely used alternative path to HMMs to address DAC as a sequence labeling problem is the use of neural network models associated with a Conditional Random Field (CRF) (Zimmermann, 2009; Kim et al., 2010) as the last layer. The CRF implements dialog state management to keep track of conversation history and current state in order to decide on the next conversation step and models the conditional probability of the DA label sequence given the input sequence. Long Short Term Memory (LSTM) (Yu et al., 2019; Barahona et al., 2016) is an artificial replay of the neural network (ANN) (Diehl et al., 2016) that can process not only single data points (such as images) but also entire sequences of data.

Statements provide knowledge that can be extracted in pairs, such as questions and their corresponding answers. Thus, in order to learn through dialogues, an agent must be able to identify what these statements are, and thus DAs must be identified by automatically recognizing the generic DAs conveyed by each segment (Searle, 1969). For this, it will be useful to recognize the communicative functions defined by ISO 24617-2 for the annotation of DAs (Bunt et al., 2012, 2017), which are hierarchically organized and feature a specific branch for knowledge-providing functions.

As the dialog progresses, some systems maintain a state representation in a process called Dialog State Tracking (DST) (Henderson et al., 2014), thus representing the user's intentions, which involves filling in predefined slot values.

Currently, most NLP tasks use Transformer neural network-based models which is an encoder-

decoder architecture that allows the model to focus on the relevant parts of input sequences, especially long sequences such as sentences and paragraphs. Improvements can be achieved if utterances are encoded by a transformer network-based (BERT) (Devlin et al., 2018).

Since manually creating the dialog flows used by conversational agents is complex and time-consuming, there are academic works focused on automatic extraction of dialog flows. One of the identified works presents structure extraction in task-oriented dialogs by representing the dialog flow with probabilistic transitions between different states of the flow, based on HMMs (Stolcke et al., 2000). A still preliminary work (Negi et al., 2009) presents the identification of dialog flows for use in chatbots, using clustering of similar expressions and their sequencing. These works are only some parts of the process we intend to develop, and most of them focus on specific domains, with narrow scope and scale, so they are not applicable to dialogs in a generic way, and therefore their use is not feasible in a real environment. Thus, there is a need to study and develop a suitable framework to guide humans in a dialog, and represent knowledge extracted from past interactions.

A supervised approach (Bangalore et al., 2008) was exploited based on a dataset of annotated dialogs, exploited the id and the speaker's word trigrams of the current utterance. In a first attempt to incorporate context, for DAC only, the previous statements were considered. DAs were discovered from open domain Twitter conversations (Ritter et al., 2010). Each post in a conversation is represented by a bag of words. DAs will correspond to clusters of statements, representing their sequential behavior, which is captured by an HMM. In addition, to separate between content words and dialog indicators, the HMM is combined with a Latent Dirichlet Allocation (LDA) topic model. The clusters have to be inspected manually. Negi et al. (Negi et al., 2009) were based on initiated conversations by replacing named entities with their type. Clustering was applied to similar utterances based on frequent words. When these clusters are discovered, calls are represented by sequences of clusters and subtasks are discovered based on sequences of frequent utterances.

2.4 Representation of Dialog

In order to create a single representation that integrates all dialog flows and their variations, we must first study the best approach to aggregate the expressions that represent the same intent, information, or action. In this way, we can apply approaches to dialogs, such as Topic Modeling (Vayansky and Kumar, 2020) and Automatic Summarization (Gupta et al., 2009), to reveal high-level topics covered in dialogs and compress their content, or use Text Clustering (Aggarwal and Zhai, 2012), which allows the clustering of similar utterances. Since DAs are less tied to the scenario or domain than intentions, this is also a better representation for recognizing common patterns in dialogs. DA Identification (Omuya et al., 2013) may help in a more abstract representation of the flow, by classifying the various interactions into different types.

DAs and transition graphs allow the discovery of different types of interactions and the most common dialog flows that will be useful for classifying the current dialog and recommending the next interactions. Automatic response generation techniques are based on sequence-to-sequence models (Yuan and Yu, 2019) learned from large collections of dialogs. One of the problems with such models is that they are not able to model the context and history of the dialog. To solve this, the model can be extended with a latent representation of the dialog history or encapsulated in a hierarchical dialog model (Sordoni et al., 2015; Lowe et al., 2017).

If annotated data is unavailable, clustering of utterances can be done, in the expectation of grouping them according to DAs or intents. Human intervention is required for interpretation, which involves looking at the utterances in each of the clusters.

The flow can be represented by a graph where the nodes represent an expression of the dialog and the arcs, directed between nodes, represent the different expressions that can follow. A single tree can encapsulate the task structure (domain and precedence relations between tasks), the DA structure (sequences of DAs), and the linguistic structure of utterances. We can also represent the probability associated with each of the possible transitions between expressions (Ritter et al., 2010), as well as the conditions, based on the extracted context, that make each transition possible or impossible.

There are annotation schemes designed for open domain human-machine conversations, such as Midas (Yu and Yu, 2019). This has a hierarchical

structure, including a semantic and functional ordering tree, and supports multi-label annotations. Since dialogs in a large collection are represented by sequences of tasks (Bangalore et al., 2008) or DAs, hierarchical relationships between the latter can be discovered from common patterns and represented by trees or graphs that are friendly for human analysis, including transition probabilities.

Young et al. (2010) described a dialog manager *Hidden Information State System* (HIS) where each utterance is a DA and is designed for information retrieval tasks. However, compared to simple slot-filling systems, it supports a richer set of user goal representations based on tree-like structures built from classes that represent related values and sub types that are specific variants of a class.

To be used as the input of most of the previous approaches, textual utterances need to be represented in a vector space where semantically similar words or utterances are closer to each other. This is typically done at preprocessing and may resort to models of vector semantics. For instance, when performing intent classification, Hashemi et al. (Hashemi et al., 2016) used pretrained models of word embeddings, such as word2vec (Mikolov et al., 2013), for representing utterances. Park et al. (Park et al., 2022) obtains various intent clustering results with different embeddings, namely the Sentence Transformer's MiniLM-L6 and MiniLM-L12 models. More recent efforts obtain sentence embeddings from available transformers fine-tuned in sentence similarity tasks (e.g., (Vulić et al., 2022; Park et al., 2022)). An alternative to using pretrained embeddings is to learn embeddings and part of the training process (e.g., first layer of the neural network) (Firdaus et al., 2021).

2.5 Call Guidance

The orientation component will take advantage of past dialogs, represented according to the previously defined, to identify the recommendations it should provide to the agent during a new interaction. Dialogs represented as a sequence or a graph of DAs can be exploited in live conversations, either to guide dialog systems that may include automatic response generation, or to guide human agents in a call. The system can be useful for classifying utterances according to specific goals as quickly as possible. The call can be redirected to a different agent that has access to different knowledge bases and/or different streams. For example, Gunrock (Chen et al., 2018), a social bot, maps users' intentions to a topic, selects the most appropriate module for the topic, and advances the user's request to this module. In addition to the topic or goal, the current DA can be classified in real time using approaches already described, allowing the anticipation of the next dialog with different probabilities, which can be used to narrow down the automatically generated responses.

It is important to use approaches such as Semantic Textual Similarity (Cer et al., 2017), using techniques that consider the words used and their relevance, such as TF-IDF or more comprehensive models based on embeddings.

Therefore, it is necessary to look at approaches such as Recommender Systems (RS) (Resnick and Varian, 1997) that help users find items of interest and can be based on past behavior.

The design of flows is especially relevant for task-oriented dialogue systems and can steer the conversation in specific directions, avoiding purely reactive responses to what the user says (Grassi et al., 2022). It encompasses the definition of task-specific intents and training phrases, among other decisions, and generally ends up being created manually, often with the help of tools like Google's DialogFlow[3], Microsoft Luis[4], or the open source platform Rasa[5].

As the dialogue progresses, the recommendation system accumulates the user's information and builds his profile. Thus, it can provide a recommendation based on user preferences reflected in the conversation. A recommendation system can be based on conventional collaborative filtering algorithms (Resnick et al., 1994; Sarwar et al., 2001) or based on neural networks (Wang et al., 2018; He et al., 2017; Ying et al., 2018).The generated graphs can be used in a recommendation or guidance system.

3 Research Proposal

The main goal of this PhD thesis is the research and development of approaches to help communication between interlocutors in a dialog, in Portuguese, guiding the operator's action that can be supported by previous interactions. Information is to be automatically extracted and frequent dialog flows are identified, allowing their representation to guide the

[3] https://cloud.google.com/dialogflow/
[4] https://www.luis.ai/
[5] https://rasa.com/

human in how to respond. To this extent, we define five specific objectives to be achieved throughout the development of the research work:

1. Collection and creation of a corpus of dialogs in Portuguese that can be used in the project.
2. Studying, developing and experimenting with approaches for extracting structured dialog information from the various interactions.
3. Studying, developing and experimenting with approaches for representing interactions and flows extracted from those interactions.
4. Studying, developing and experimenting with approaches for guiding the human by exploiting the knowledge extracted from dialogs, interactions, and common flows.
5. Evaluation on data collected and created, using automated and manual metrics.

To achieve the five defined objectives presented above, the following tasks were defined:

1. Deepen the study of the state of the art to understand important concepts for research;
2. Collection or creation of the data to be used;
3. Approaches for IE;
4. Approaches for representing dialog flows;
5. Approaches for dialog guidance;
6. Framework with approaches explored;
7. Tests and final evaluation;
8. Writing of the thesis and scientific articles.

We intend to explore generalized approaches applied to different types of task-oriented dialogues, where one contribution will be to increase the efficiency of call centers.

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

Several alternatives will be explored to obtain the data: Following the WOZ paradigm (Green et al., 2004), where a conversation is held between two interlocutors in which one is assigned a certain task and to accomplish this task, this user must interact, using natural language, with another who will have access to more information about the domain (for example, a database or a service such as Booking[6]) and will be able to provide appropriate answers. Interaction can be done through any chat application, such as Slack or Microsoft Teams; Transcripts of existing dialogs in Portuguese, such as CORAA (Junior et al., 2021); Customer support services on social networks, such as conversations with telecom operators on Twitter[7]; Movie subtitles (Lison and Tiedemann, 2016); Translation of English datasets (e.g. DailyDialogue (Li et al., 2017) or MultiWOz (Budzianowski et al., 2018)) into Portuguese, from which it will be possible to import existing annotations.

The data will be used in the development of a framework consisting of three components:

- Extraction - Process transcripts of dialogs and extract information;
- Representation - Discovery of the most frequent dialog flows, represented by graphs;
- Guidance - will take advantage of the flow representation to guide the human.

The first component processes real dialog transcripts and extract useful information from them to represent the interactions, such as keywords, entities or actions. The extraction of some of these items may resort to an NLP pipeline (Tenney et al., 2019), but some additional development may be required, considering the language (Portuguese) and the type of text (dialog).

The extracted information can be used to better describe utterances, classifying intentions and filling slots. However, performing these tasks is usually based on supervised learning, which involves data annotation, something to consider during data definition. It can also be used to group similar utterances, using clustering. This process can make use of sentence embedding techniques (Reimers and Gurevych, 2019) to represent utterances. During the extraction process, it is necessary to remove from the text private information about the client in order to ensure the confidentiality of the data.

In Figure 1, we show an example of a dialog in which the customer requests the cancellation of an order he previously placed.

Sometimes, the information found in knowledge bases is not organized in a way that facilitate its use

[6]http://www.booking.com
[7]http://www. twitter.com

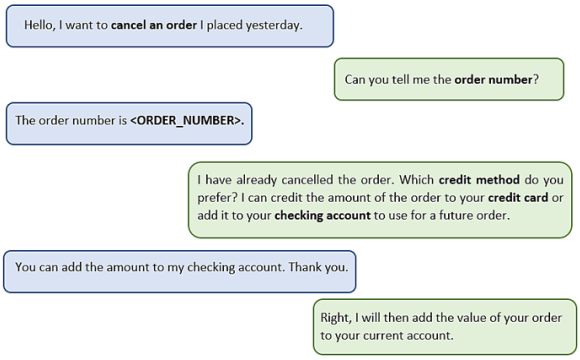

Figure 1: Example of a dialog flow created by the extraction component

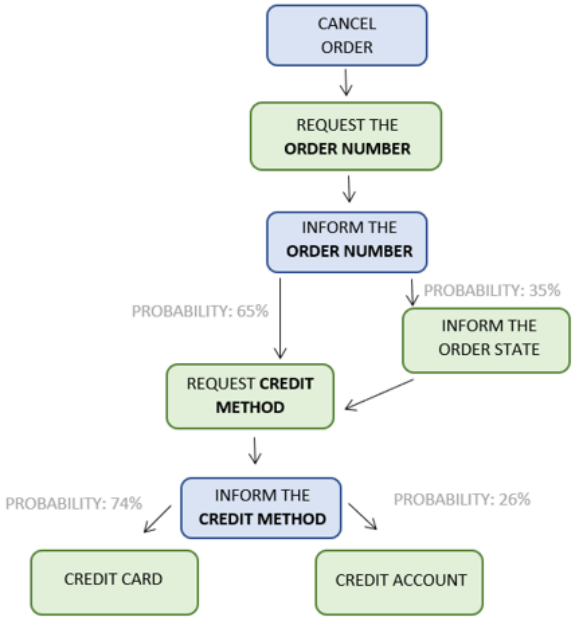

Figure 2: Example of a dialog graph created by the representation component

by the agent in a conversation (Taylor et al., 2002). Therefore, the project to be developed will answer the question of how to improve the organization and representation of information that supports the agent's assistance during a conversation.

Interactions with customers are dependent on need and context. Even if the information is up to date, we need to make sure that we effectively map the need expressed by the possible solutions and use the context to ensure that we choose the solution that best fits that specific case. To this end, the project will be able to answer the question of how to find the best solution for customer's need and how to ensure that this solution fits its context.

The second component will aim to discover the most frequent dialog flows, represented by graphs, where the vertices represent speech classes or groupings, and the arcs represent transitions between them, with probabilities. In this component one can apply the classification of interactions into more generic classes (DAs) or, if there is a lack of data to make the system less domain-dependent, perform a grouping that approximates these acts. To facilitate human interpretation, it will be important to have a way to describe the groupings/classes through relevant n-grams or verb phrases. Figure 2 shows an example of a dialog graph, generated from the previous dialog flow.

Finally, the guidance component will take advantage of the representation of flows to guide the human.

The ability to take into account the previous statements is key to building dialog systems that can keep conversations active and engaging (Sordoni et al., 2015). Past interactions are an important source of information about customers and how their needs are met by agents, however, due to the complexity of working with past interactions, they are generally ignored. We aim to find the best approaches to extract from past interactions the knowledge needed to guide agents on how to respond to customer needs.

In each interaction, previous interactions will be considered to suggest responses, while anticipating the next interactions. It will function as a RS (Resnick and Varian, 1997) in the sense that we want to recommend speech and/or actions. Figure 3 shows an example of a user interface with expressions used in dialog and the recommendations provided by the guidance component.

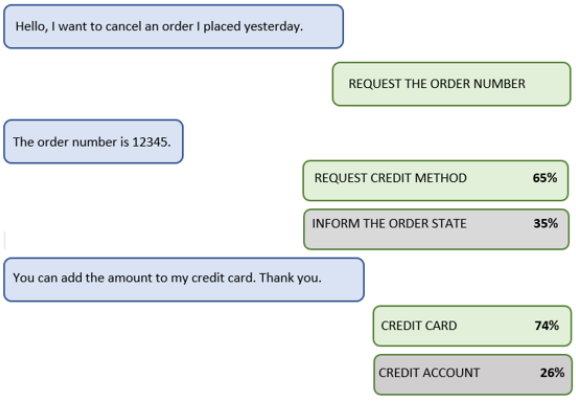

Figure 3: Example of user interface - recommendations are provided to the agent by the guidance component

The approaches resulting from tasks 3, 4, and 5 will be evaluated independently but a final eval-

uation of their integration into the framework is required. In this way, the approaches will be evaluated on the data collected and created using metrics for classification, used when annotations are produced, or metrics for clustering when this is not the case. Due to the subjectivity of the quality of the flows identified and the guidance produced, a subjective evaluation based on human opinions using the resulting framework is imperative, as well as an objective evaluation where a control group that uses the node-generated graphs and a group that does not is selected and the average success rates and response times are then compared.

4 Preliminary Experiments

The exploration of some NLP tools, through the spaCy [8] library such as PoS Tagging considering verbal syntagma, NER and coreference resolution was performed and DAC, taking into account different vector representations, was done for these tasks in order to generalize utterances.

Given that DAs can be seen as generic intents, a possible representation of dialog flows is through a graph of DAs. Transitions between DAs may further have assigned probabilities, computed from the dialogue history. This can be see as a Markov chain, and its inspection may further enable the identification of communication patterns.

For illustrative purposes, we generated such a representation for the Mastodon dataset (Cerisara et al., 2018) and its annotated DAs, with the help of the NetworkX [9] package. The flow can be visualized in Figure 4, where nodes were also included for the start (SOD) and end (EOD) of dialog. Transitions with probability below 0.05 were ignored.

5 Conclusion

This research proposal aims to improve the customer/human experience when contacting a call-center, by improving the responsiveness of human agents in conversations, guided by intelligent methods and NLP about the current context and about previous interactions with customers. To achieve this goal, the project is organized into three components: extraction, representation and guidance. One of the challenges involved is that this project will be focused on Portuguese, a language for which there is little work in this area.

[8] https://spacy.io/

[9] https://networkx.org/

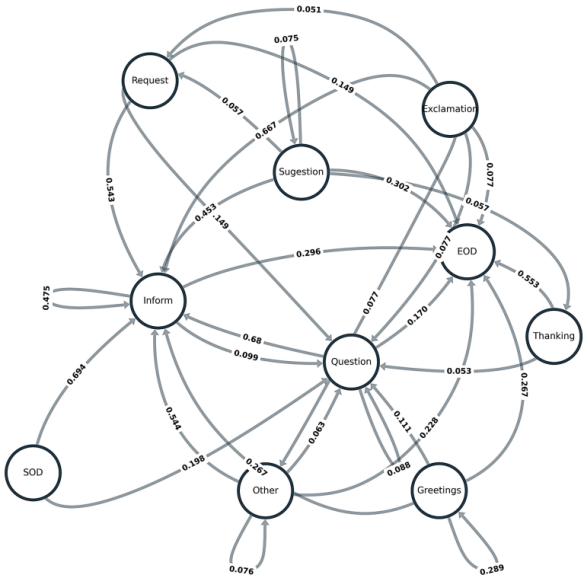

Figure 4: DAs transitions in Mastodon

The existing work for the automatic extraction of dialog flows is still underdeveloped and has been applied in a small scope and scale. Furthermore, this work is usually referenced in the context of application to chatbots, and its application is not oriented towards human agents.

The development of the proposed solution allows for the automatic extraction of dialog flows from past interactions, guiding human agents, and may represent a breakthrough in the state of the art in this area, answering the question of how to find the best solution for the customer's needs and how to ensure that this solution fits the customer's context.

Thus, the use of chatbots is increasingly present, however, we believe that human agents have a relevant role in contact centers, since they can handle situations with a level of complexity that is not yet within the reach of any chatbot and there is no distance between customers and human interlocutors. All relevant findings and results will be published in reports, articles and scientific papers, in addition to the resulting doctoral thesis.

Acknowledgements This work was financially supported by the project FLOWANCE (POCI-01-0247-FEDER-047022), cofinanced by FEDER, through PT2020, and by COMPETE 2020; and by national funds through FCT, within the scope of the project CISUC (UID/CEC/00326/2020) and by European Social Fund, through the Regional Operational Program Centro 2020. I would like to thank my supervisors Hugo Gonçalo Oliveira, Ana Alves, and Catarina Silva for all their support and to the mentor assigned by the EACL organization, Maria Jung Barrett, for her availability and help.

References

Eleni Adamopoulou and Lefteris Moussiades. 2020. Chatbots: History, technology, and applications. *Machine Learning with Applications*, 2:100006.

Charu C Aggarwal and ChengXiang Zhai. 2012. A survey of text clustering algorithms. In *Mining text data*, pages 77–128. Springer.

John Langshaw Austin. 1962. *How to do things with words: the William James lectures delivered at Harvard University in 1955*. Oxford University Press, New York.

Xuefeng Bai, Yulong Chen, Linfeng Song, and Yue Zhang. 2021. Semantic representation for dialogue modeling. *arXiv preprint arXiv:2105.10188*.

Srinivas Bangalore, Giuseppe Di Fabbrizio, and Amanda Stent. 2008. Learning the structure of task-driven human–human dialogs. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(7):1249–1259.

Lina M Rojas Barahona, Milica Gasic, Nikola Mrkšić, Pei-Hao Su, Stefan Ultes, Tsung-Hsien Wen, and Steve Young. 2016. Exploiting sentence and context representations in deep neural models for spoken language understanding. *arXiv preprint arXiv:1610.04120*.

Jonathan Berant, Andrew Chou, Roy Frostig, and Percy Liang. 2013. Semantic parsing on freebase from question-answer pairs. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1533–1544.

Luka Bradeško and Dunja Mladenić. 2012. A survey of chatbot systems through a loebner prize competition. In *Proceedings of Slovenian language technologies society eighth conference of language technologies*, pages 34–37. Institut Jožef Stefan Ljubljana, Slovenia.

Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Inigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gašić. 2018. Multiwoz–a large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling. *arXiv preprint arXiv:1810.00278*.

Harry Bunt, Volha Petukhova, David Traum, and Jan Alexandersson. 2017. Dialogue act annotation with the iso 24617-2 standard. In *Multimodal interaction with W3C standards*, pages 109–135. Springer.

HC Bunt, J Alexandersson, J Choe, C Alex, K Hasida, VV Petukhova, A Popescu-Belis, and D Traum. 2012. Iso 246170-2: A semantically-based standard for dialogue annotation. In *Proceedings of the 8th International Conference on Language Resources and Evaluation, Istanbul, Turkey*, page 8. ELRA.

Daniel Cer, Mona Diab, Eneko Agirre, Inigo Lopez-Gazpio, and Lucia Specia. 2017. Semeval-2017 task 1: Semantic textual similarity-multilingual and cross-lingual focused evaluation. *arXiv preprint arXiv:1708.00055*.

Christophe Cerisara, Somayeh Jafaritazehjani, Adedayo Oluokun, and Hoa Le. 2018. Multi-task dialog act and sentiment recognition on mastodon. *arXiv preprint arXiv:1807.05013*.

Chun-Yen Chen, Dian Yu, Weiming Wen, Yi Mang Yang, Jiaping Zhang, Mingyang Zhou, Kevin Jesse, Austin Chau, Antara Bhowmick, Shreenath Iyer, et al. 2018. Gunrock: Building a human-like social bot by leveraging large scale real user data. *Alexa Prize Proceedings*.

Qian Chen, Zhu Zhuo, and Wen Wang. 2019. Bert for joint intent classification and slot filling. *arXiv preprint arXiv:1902.10909*.

Abir De, Nastaran Okati, Ali Zarezade, and Manuel Gomez Rodriguez. 2021. Classification under human assistance. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 5905–5913.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Peter U Diehl, Guido Zarrella, Andrew Cassidy, Bruno U Pedroni, and Emre Neftci. 2016. Conversion of artificial recurrent neural networks to spiking neural networks for low-power neuromorphic hardware. In *2016 IEEE International Conference on Rebooting Computing (ICRC)*, pages 1–8. IEEE.

Jasper Feine, Stefan Morana, and Alexander Maedche. 2020. A chatbot response generation system. In *Proceedings of the Conference on Mensch und Computer*, pages 333–341.

Mauajama Firdaus, Hitesh Golchha, Asif Ekbal, and Pushpak Bhattacharyya. 2021. A deep multi-task model for dialogue act classification, intent detection and slot filling. *Cognitive Computation*, 13:626–645.

Chih-Wen Goo and Yun-Nung Chen. 2018. Abstractive dialogue summarization with sentence-gated modeling optimized by dialogue acts. In *2018 IEEE Spoken Language Technology Workshop (SLT)*, pages 735–742. IEEE.

Lucrezia Grassi, Carmine Tommaso Recchiuto, and Antonio Sgorbissa. 2022. Knowledge-grounded dialogue flow management for social robots and conversational agents. *International Journal of Social Robotics*, 14(5):1273–1293.

Anders Green, Helge Huttenrauch, and K Severinson Eklundh. 2004. 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.

Ralph Grishman. 2019. Twenty-five years of information extraction. *Natural Language Engineering*, 25(6):677–692.

Vishal Gupta, Gurpreet S Lehal, et al. 2009. A survey of text mining techniques and applications. *Journal of emerging technologies in web intelligence*, 1(1):60–76.

Homa B Hashemi, Amir Asiaee, and Reiner Kraft. 2016. Query intent detection using convolutional neural networks. In *International conference on web search and data mining, workshop on query understanding*.

Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, pages 173–182.

Matthew Henderson, Blaise Thomson, and Jason D Williams. 2014. The second dialog state tracking challenge. In *Proceedings of the 15th annual meeting of the special interest group on discourse and dialogue (SIGDIAL)*, pages 263–272.

Julia Hoxha, Praveen Chandar, Zhe He, James Cimino, David Hanauer, and Chunhua Weng. 2016. Dream: Classification scheme for dialog acts in clinical research query mediation. *Journal of biomedical informatics*, 59:89–101.

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. 2021. Coraa: a large corpus of spontaneous and prepared speech manually validated for speech recognition in brazilian portuguese. *arXiv preprint arXiv:2110.15731*.

Su Nam Kim, Lawrence Cavedon, and Timothy Baldwin. 2010. 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.

Ruizhe Li, Chenghua Lin, Matthew Collinson, Xiao Li, and Guanyi Chen. 2018. A dual-attention hierarchical recurrent neural network for dialogue act classification. *arXiv preprint arXiv:1810.09154*.

Yanran Li, 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*.

Pierre Lison and Jörg Tiedemann. 2016. Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles. *European Language Resources Association*.

Chunyi Liu, Peng Wang, Jiang Xu, Zang Li, and Jieping Ye. 2019. Automatic dialogue summary generation for customer service. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1957–1965.

Ryan Lowe, Nissan Pow, Iulian Vlad Serban, Laurent Charlin, Chia-Wei Liu, and Joelle Pineau. 2017. Training end-to-end dialogue systems with the ubuntu dialogue corpus. *Dialogue & Discourse*, 8(1):31–65.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, page 3111–3119, Red Hook, NY, USA. Curran Associates Inc.

Behrang Mohit. 2014. Named entity recognition. In *Natural language processing of semitic languages*, pages 221–245. Springer.

Sumit Negi, Sachindra Joshi, Anup K Chalamalla, and L Venkata Subramaniam. 2009. Automatically extracting dialog models from conversation transcripts. In *2009 Ninth IEEE International Conference on Data Mining*, pages 890–895. IEEE.

Hugo Oliveira, Patrícia Ferreira, Daniel Martins, Catarina Silva, and Ana Alves. 2022. A brief survey on textual dialogue corpora. *Proceedings of the 13th International Conference on Language Resources and Evaluation (LREC 2022)*.

Adinoyi Omuya, Vinodkumar Prabhakaran, and Owen Rambow. 2013. Improving the quality of minority class identification in dialog act tagging. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 802–807.

Jeiyoon Park, Yoonna Jang, Chanhee Lee, and Heuiseok Lim. 2022. Analysis of utterance embeddings and clustering methods related to intent induction for task-oriented dialogue. *arXiv preprint arXiv:2212.02021*.

Amon Rapp, Lorenzo Curti, and Arianna Boldi. 2021. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies*, 151:102630.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. 1994. Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pages 175–186.

Paul Resnick and Hal R Varian. 1997. Recommender systems. *Communications of the ACM*, 40(3):56–58.

Alan Ritter, Colin Cherry, and Bill Dolan. 2010. Unsupervised modeling of twitter conversations. *North American Chapter of the Association for Computational Linguistics (HLT-NAACL)*.

Alan Ritter, Colin Cherry, and Bill Dolan. 2011. Data-driven response generation in social media. In *Empirical Methods in Natural Language Processing (EMNLP)*.

Navin Sabharwal and Amit Agrawal. 2020. Introduction to google dialogflow. In *Cognitive virtual assistants using Google Dialogflow*, pages 13–54. Springer.

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295.

John R Searle. 1969. *Speech acts: An essay in the philosophy of language*, volume 626. Cambridge university press.

Rakesh Kumar Sharma and Manoj Joshi. 2020. An analytical study and review of open source chatbot framework, rasa. *International Journal of Engineering Research and*, 9(06).

Heung-Yeung Shum, Xiao-dong He, and Di Li. 2018. From eliza to xiaoice: challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering*, 19(1):10–26.

Alessandro Sordoni, Michel Galley, Michael Auli, Chris Brockett, Yangfeng Ji, Margaret Mitchell, Jian-Yun Nie, Jianfeng Gao, and Bill Dolan. 2015. A neural network approach to context-sensitive generation of conversational responses. *arXiv preprint arXiv:1506.06714*.

ASSK Sreeharsha, Sai Mohan Kesapragada, and Sai Pratheek Chalamalasetty. 2022. Building chatbot using amazon lex and integrating with a chat application. *International Journal of Scientific Research in Engineering and Management (IJSREM)*, 6(04).

Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics*, 26(3):339–373.

Gita Sukthankar, Christopher Geib, Hung Bui, David Pynadath, and Robert P Goldman. 2014. *Plan, activity, and intent recognition: Theory and practice*. Newnes.

Phil Taylor, Gareth Mulvey, Jeff Hyman, and Peter Bain. 2002. Work organization, control and the experience of work in call centres. *Work, employment and society*, 16(1):133–150.

Ian Tenney, Dipanjan Das, and Ellie Pavlick. 2019. Bert rediscovers the classical nlp pipeline. *arXiv preprint arXiv:1905.05950*.

Anh D Tran, Jason I Pallant, and Lester W Johnson. 2021. Exploring the impact of chatbots on consumer sentiment and expectations in retail. *Journal of Retailing and Consumer Services*, 63:102718.

Khai N Truong, Elaine M Huang, and Gregory D Abowd. 2004. Camp: A magnetic poetry interface for end-user programming of capture applications for the home. In *UbiComp 2004: Ubiquitous Computing: 6th International Conference, Nottingham, UK, September 7-10, 2004. Proceedings 6*, pages 143–160. Springer.

Ike Vayansky and Sathish AP Kumar. 2020. A review of topic modeling methods. *Information Systems*, 94:101582.

Ivan Vulić, Iñigo Casanueva, Georgios Spithourakis, Avishek Mondal, Tsung-Hsien Wen, and Paweł Budzianowski. 2022. Multi-label intent detection via contrastive task specialization of sentence encoders. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7544–7559, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Xiting Wang, Yiru Chen, Jie Yang, Le Wu, Zhengtao Wu, and Xing Xie. 2018. A reinforcement learning framework for explainable recommendation. In *2018 IEEE international conference on data mining (ICDM)*, pages 587–596. IEEE.

Joseph Weizenbaum. 1966. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45.

Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. 2018. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 974–983.

Steve Young, Milica Gašić, Simon Keizer, François Mairesse, Jost Schatzmann, Blaise Thomson, and Kai Yu. 2010. The hidden information state model: A practical framework for pomdp-based spoken dialogue management. *Computer Speech & Language*, 24(2):150–174.

Dian Yu and Zhou Yu. 2019. Midas: A dialog act annotation scheme for open domain human machine spoken conversations. *arXiv preprint arXiv:1908.10023*.

Yong Yu, Xiaosheng Si, Changhua Hu, and Jianxun Zhang. 2019. A review of recurrent neural networks: Lstm cells and network architectures. *Neural computation*, 31(7):1235–1270.

Lin Yuan and Zhou Yu. 2019. Abstractive dialog summarization with semantic scaffolds. *arXiv preprint arXiv:1910.00825*.

Matthias Zimmermann. 2009. Joint segmentation and classification of dialog acts using conditional random fields. In *Tenth Annual Conference of the International Speech Communication Association*.