Extracting Dialog Structure and Latent Beliefs from Dialog Corpus

Aishwarya Chhabra, Pratik Saini and C. Anantaram

TCS Research, Tata Consultancy Services Ltd, Gwal Pahari, Gurgaon, India {aishwarya.chhabra, pratik.saini, c.anantaram}@tcs.com

Abstract

Dialog corpus captures various real world human interactions in a particular domain. However, to build a task-based chat-bot for carrying out humanmachine interactions in that domain, it is essential to extract the dialog structure and the latent beliefs in that corpus to build an effective chatbot. We examine this problem and propose a machine learning based solution. Our method categorizes the utterances into corresponding dialog states (sub-tasks) using a domain ontology, extracts the required information using machine learning based approach, maps it with the appropriate state, and automatically builds a finite-state-machine based dialog model. Further, since each human utterance is in a context, a set of utterances consists of latent beliefs that the human uses while conversing on a topic. Our method identifies the latent beliefs in conversations and uses them to appropriately tailor the chat-bot's responses based on the extracted finite state machine. We show how our method can lead to better conversational experience with a chatbot.

1 Introduction

Customer support systems and Planning systems in domains such as Product support, Travel planning, Student-advising etc. have transcribed dialog corpus capturing human-human conversations that are largely task-oriented. In order to implement chatbot service in such domains, it is essential to extract the dialog model that captures the information regarding the states of the dialog. Most of the task-oriented systems still use significant engineering and expert knowledge to implement the backbone of the dialog manager that carries out the dialogues. Usually, dialog systems are either trained on a huge general corpus or driven through a rule base. For this reason, these tend to behave in a restricted way and fail to capture beliefs and emotional state. It is observed that most of the time chatbots behave mechanically and do not take customer beliefs into account while conversing. As shown in Figure 1, we can see the user tells the bot repetitively that dates and budget are not flexible for him, but the bot keeps on asking for the change of dates or budget instead of suggesting a new place. This behaviour of the system leads to a significant downturn in customer satisfaction.

In this work, the focus is on taking dialog corpus captured in human-human interactions, and use that to learn the underlying dialog model for conversations in a domain. In addition to the extraction of the dialog model, our method also identifies and learns the latent beliefs of the user to drive the conversation in a meaningful direction.

User: I'm looking for a trip to Gotham City leaving from Kakariko Village on Saturday, August 13, 2016. 3 adults for no more than \$2400
Bot: I have a trip available to Vancouver for these dates within your price range. Would you like to increase your budget or adjust the dates?
User: I'd like to adjust the departure city to Caprica but I cannot adjust the dates or budget.
Bot: Still no availability for this destination. Would you like to increase your budget or adjust the dates?
User: The dates cannot be changed. How about going to Theed with 2 adults, leaving from Kakariko Village, on a budget of \$2400?
Bot: Still no availability.

User: Then I will bring my business elsewhere. Thank you.

Figure 1: A sample conversation without beliefs

In a way, dialog transcript data-sets encode the domain structure information. Our framework automatically learns this domain structure information using deep learning models. We make use of domain ontology to enhance the accuracy of this learned dialog model. There have been number of attempts to build end to end dialog systems. However, such systems have not focused on extracting the latent beliefs in the conversations that is required to tailor the chatbot interaction for each user. Our framework also learns the latent beliefs of the customer from these transcripts and effectively incorporates these beliefs to tailor its dialog suitably.

This remainder of the paper is organized as follows. Section 2 discusses the related work. Section 3 describes the proposed architecture; Section 4 contains details on extracting latent beliefs. Section 5 evaluates our models qualitatively and quantitatively, and finally conclusion in Section 6.

Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2 Related Work

Automatic extraction of dialog structure and latent beliefs from a given corpus is a relatively less explored area. Previously, most of the work has been done using supervised learning [Feng et al., 2005]. [Bangalore et al., 2008] uses a classification-based approach to automatically create task structures for task-oriented dialogues. They use a dialog modeling approach that tightly couples dialog act and task/sub-task information. There has been some work done in direction of unsupervised learning for discovering the dialog model. [Zhai and Williams, 2014] proposed three models to discover the structure of the dialogue. They synthesize hidden Markov models and topic models to extract the underlying structure in dialogues. Their models achieve superior performance on held-out log likelihood evaluation and an ordering task. [Negi et al., 2009] presented a method to build a task-oriented conversational system from call transcript data in an unsupervised manner. The work in [Shi et al., 2019] focuses on task oriented dialog and use Variational Recurrent Neural Network(VRNN) to extract the dialog structure and dynamics in dialog.

Neural dialogue generation has also shown promising results recently. [Serban et al., 2015; Henderson et al., 2014] uses generative neural models to produce system responses that are autonomously generated word-by-word. [Liu et al., 2018] combined knowledge base with neural dialogue generation for generating meaningful, diverse and natural responses for both factoid-questions and knowledge grounded chit-chats. [Wu et al., 2018] has shown a method to represent conversation session into memories upon which attentionbased memory reading mechanism can be performed multiple times for generating optimal responses step-by-step. [Bordes and Weston, 2016] use the Memory Networks to build the dialog system on DSTC2 dataset [Williams et al., 2016]. Although quite a number of attempts have been made to build dialogue systems [Weston, 2016], the use of epistemic rules in driving the dialogue in a consistent way with the beliefs has not yet been tackled. Various approaches to dialog management and discovering dialog structure have been proposed. But these approaches failed to take user's beliefs into account to tailor the dialogues. [Prabhakaran *et al.*, 2018] analyses how author commitment in text reveals the underlying power relations and how to incorporate this information to detect the power direction in actual conversation. [Kawabata and Matsuka, 2018] focuses on the construction of mutual belief in spoken task oriented dialogues.

For the extraction of latent beliefs, [Chhabra *et al.*, 2018] and [Sangroya *et al.*, 2018] have shown how beliefs can be used to design a more meaningful conversation. However, no work seems to have been done regarding extraction of dialog structure with latent beliefs from dialog corpus.

3 Architecture

The problem of automatically discovering the dialog model can be viewed as extracting all the relevant sub-tasks and its valid ordering. For example, the task of hotel booking can have following sub tasks : destination city, budget, hotel rating, location

preference, number of people, dates,

amenities, confirm booking. In this example, the first seven sub-tasks are independent from each other and can be performed in any order. The task "Confirm booking" will always be the last sub-task that needs to be performed to complete the task successfully. In our work, we are focusing on finding all the valid orders of the sub tasks. This is in contrast to previous work where only a fixed ordering of sub-tasks is considered. Our approach consists of several steps. We initially split the utterances into agent utterances and user utterances and then analyze these separately.

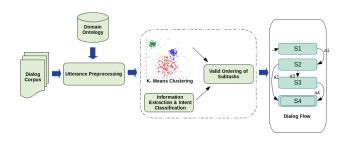


Figure 2: Architecture

3.1 Cleaning and tagging

We remove stop words and then identify the domain-specific and general-purpose tags from the agent utterances. For example, from the agent utterance 'I can also offer you 5 days at Scarlet Palms Resort, a 3.5 star rating hotel, for 1358.78 USD'. In this sentence, we identify the tags like person, location, etc. This utterance will be changed to 'I can also offer you *n_days* at *location*, a *n_rating* hotel, for *price'*. We use Stanford Core NLP [Manning *et al.*, 2014] to identify the general-purpose tags. For domain-specific tags, we make use of domain ontology and find domain-specific terms from the utterance to replace it with the domain tags corresponding to the terms. The tagged data helps us achieve clean clusters.

3.2 Clustering

We have observed that the agent utterances follow a standard sequence to help users accomplish a task. On the other hand, the user utterances have a lot of variations in their responses. Hence, for clustering purposes, we are considering the agent utterances exclusively. The idea behind clustering of agent utterances is to identify the states/sub tasks in a dialog. For example, the task of booking a hotel may consist of several sub-tasks. By using clustering, we cluster together all the agent utterances that fall into one common state. For example, 'Which place are you planning to go?' and 'Where would you like to go?' will be clustered together. We use K-means clustering, where k value is determined by elbow method, to create clusters. For example, in the hotel booking domain we create 8 clusters. For each agent utterance, we generate a feature vector, where n-grams words ($n \le 2$) are used as features. The clusters are used together with extracted information to determine dialog states.

3.3 Deep Learning based Information Extraction

After clustering agent utterances, user utterances are considered. A information extraction model is trained through supervised learning to provide all the tags for given user utterances. In order to efficiently extract tags, deep neural networks based sequence tagging model is used. Our architecture consists of a bi-directional LSTM network along with a CRF (conditional random field) output layer. For a given sentence $(x_1, x_2, ..., x_n)$ containing n words, each represented as a d-dimensional vector, a bi-LSTM computes the word representation by concatenating the left and right context representation, $\mathbf{h}_t = [\mathbf{h}'_t; \mathbf{h}_t]$. We use ELMO embedding computed on top of two-layer bidirectional language models with character convolutions as a linear function of the internal network states [Peters et al., 2018]. Next, we map the extracted tags to the states/ sub tasks extracted from the clusters identified in the previous step. For example, from an utterance 'I want to visit Denver for 4 days.', "Denver" and "4 days" will be extracted and mapped to destination and number of days for travel. The annotation of this example can be seen in Figure 3.

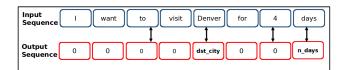


Figure 3: Example of Input Output Sequence

3.4 Find Valid Ordering of States

The next step is to find the valid order of the states for both the agent and user. For this, we are using a revised version of the Apriori algorithm. Using this algorithm, we can find the implications like the sub-task 'Confirm booking' will always come as the last state, 'Booking' is always preceded by 'dates' etc. by using the agent and user state information to find out the implication rules by determining the support in the corpus for such transitions. This provides us with a more appropriate ordering of the sub tasks and what to do when User provides a response for a particular state.

3.5 Intent Classification

Intent classification needs to be done when a new user utterance has to be processed. The intent classifier was implemented as a Bi-directional LSTM classification model trained through supervised learning. Intent classification determines the state in the Dialog model that the new user utterance starts from.

We merge all these components: states extracted, their valid order, and intent classification model to build the finite

state machine. This finite state machine is now able to tailor the dialog for the learnt task oriented dialog model.

4 Identifying Latent Beliefs

Understanding a user's opinion is extremely important to initiate and maintain a meaningful conversation. Every user can have a different sentiment, these dissimilar users need different conversational flow and a the set of dialog policy needs to be tailored for each case. It is a challenging task to build an automatic system that can understand the latent beliefs of users. It is possible to handcraft it, such an approach has several flaws. An alternative to hand-crafting belief rules is to automatically learn it from a large annotated corpus of utterances and corresponding labeled beliefs [Chhabra *et al.*, 2018; Sangroya *et al.*, 2018].

Latent belief extraction module is implemented and evaluated in two domains: Student-Advisor domain [Chulaka Gunasekara and Lasecki. 2019 and the Frames data [El Asri et al., 2017], also used to extract the dialogue model. Better performance and reduced number of turns recorded with tailoring of dialog model. In the Student-advisor domain, three belief classes were identified: curious, neutral, confused. For example, a confused student wrote <code>'I</code> have no sense of where I want my life to go and am unable to determine what classes to take. Can you help me decide what to do?', who can be significantly disoriented and may require a one-on-one counseling session with an expert. This category of students need serious attention and a specific flow of questions to help them make a precise selection. An illustrative conversation is shown in Figure 4.

Student: Hi, my class selections for next semester are under consideration. What are some suggestions that can be given by you?

Advisor: As for requirements, do you have any left? **Student:** Not to my knowledge.

Advisor:Do you have a precise preference as to course selection?

Student: I do prefer classes with a lighter work load. **Advisor:**What do you think about EECS183, Elementary Programming Concepts? The class is entry level.

Figure 4: A sample conversation from Student Advisor Domain

We identified the sentiments of the users in a hotel booking domain can be broadly classified into 5 categories: Flexible, Satisfied, Neutral, Disappointed, Inflexible. User responses are classified to one of these categories to tailor the dialog with the user. For example, 'All I have left in this life is my burgeoning bank account. So no budget, just get me something I'll like.' is categorized as disappointed and there is no budget constraint for booking, so the states of *asking hotel rating* and *amenity* can be skipped, and user should be suggested the options with high rating and luxurious amenities.

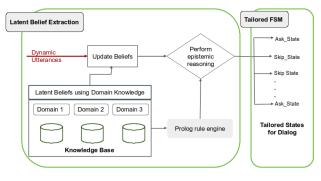


Figure 5: Extracting Latent Beliefs

If an utterance belongs to more critical categories such as disappointed, it is assigned a higher weight as compared to category such as of a flexible user. This is instinctive that the user who is disappointed would need a different response and dialog policy. We used LSTM based classification model on *five* categories. A high level architecture is illustrated in Figure 5

4.1 Epistemic Reasoning over Latent Beliefs

The extracted latent beliefs and the domain knowledge trigger the epistemic rules. For example "Belief (disappointed) and Budget(high) => Knows-Agent (user to be given luxury suite with all amenities), Knows-Agent (skip-state(ask hotel rating))"asserts facts about the current epistemic state of the agent. The epistemic logic is written using prolog in the working system. The beliefs and epistemic rules helped tailor the dialog to the customer expectations.

5 Experiments and Results

5.1 Dataset

We used Frames data-set [El Asri *et al.*, 2017] consists of conversations for finding an appropriate vacation package. The corpus has 1369 human-human dialogs with an average of 15 turns per dialog, for a total of 19986 turns in the data-set. We used human-human task oriented conversations as they include the real-world contexts and are rich in terms of user beliefs. A sample conversation from this dataset can be seen in Figure 6.

5.2 Information Extraction

We have taken 10400 utterances from around 1400 dialogs. We annotated these utterances with around 50 tags like budget, str_date , n_adults , price, etc. An accuracy of 97.23% achieved in training and 93.54% in testing phase.

User: Hi there, I am from Vitoria and I want to go on a vacation. Wizard: Where would you like to go? User: I would like to go to Santo Domingo. Wizard: Would a 7 day trip work for you? User: Yes that sounds fine, looking to leave on the 19th Wizard: Great, I have a flight departing on the 19th and returning on the 25^{th} of August. User: What is the hotel like? Wizard: It is called the Rose Sierra Hotel and it is a 3-star hotel that includes free breakfast, wifi and parking. The total cost would be 2170.90 USD. User: What type of flight is that going to be? Wizard: It is an economy class flight. User: Let's book it please. Wizard: Perfect. Have a great trip. User: Thank you.

Figure 6: A sample conversation from Frames data

5.3 Clustering

For clustering, we evaluated results by comparing the results against a manually tagged data-set for 7 clusters. We achieved an accuracy of 89%.

5.4 Latent Belief Extraction

For the Student-Advisor domain, total number of 3500 utterances including the paraphrases of those utterances, labelled across 3 categories were trained on an LSTM based classifier. The model achieved an accuracy of 84%.

For Frames dataset, the similar classifier was used as in the student-advisor domain. In this domain, accuracy of 87% was achieved over 5 classes.

6 Conclusion

This paper presents a framework to automatically extract a dialog model and latent beliefs from transcribed dialog corpus with good results at each component level. Our approach takes latent beliefs of customers into account to tailor the finite state machine to give better and more personalized experience. Our experimental evaluation demonstrates the efficacy of the proposed methods.

References

- [Bangalore *et al.*, 2008] S. Bangalore, G. Di Fabbrizio, and A. Stent. Learning the structure of task-driven human-human dialogs. *IEEE Transactions on Audio*, *Speech, and Language Processing*, 16(7):1249–1259, Sep. 2008.
- [Bordes and Weston, 2016] Antoine Bordes and Jason Weston. Learning end-to-end goal-oriented dialog. *CoRR*, abs/1605.07683, 2016.
- [Chhabra *et al.*, 2018] Aishwarya Chhabra, Pratik Saini, Amit Sangroya, and C. Anantaram. Learning latent beliefs and performing epistemic reasoning for efficient and meaningful dialog management. *CoRR*, abs/1811.10238, 2018.

- [Chulaka Gunasekara and Lasecki, 2019] Lazaros Polymenakos Chulaka Gunasekara, Jonathan K. Kummerfeld and Walter S. Lasecki. Dstc7 task 1: Noetic end-to-end response selection. In 7th Edition of the Dialog System Technology Challenges at AAAI 2019, January 2019.
- [El Asri et al., 2017] Layla El Asri, Hannes Schulz, Shikhar Sharma, Jeremie Zumer, Justin Harris, Emery Fine, Rahul Mehrotra, and Kaheer Suleman. Frames: a corpus for adding memory to goal-oriented dialogue systems. In Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue, pages 207–219. Association for Computational Linguistics, 2017.
- [Feng et al., 2005] Junlan Feng, Patrick Haffner, and Mazin Gilbert. A learning approach to discovering web page semantic structures. In Proceedings of the Eighth International Conference on Document Analysis and Recognition, ICDAR '05, pages 1055–1059, Washington, DC, USA, 2005. IEEE Computer Society.
- [Henderson *et al.*, 2014] Matthew Henderson, Blaise Thomson, and Steve Young. Word-based dialog state tracking with recurrent neural networks. In *Proceedings of the* 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), pages 292–299. Association for Computational Linguistics, 2014.
- [Kawabata and Matsuka, 2018] Y. Kawabata and T. Matsuka. How do people construct mutual beliefs in taskoriented dialogues? In 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), pages 1299–1304, Nov 2018.
- [Liu et al., 2018] Shuman Liu, Hongshen Chen, Zhaochun Ren, Yang Feng, Qun Liu, and Dawei Yin. Knowledge diffusion for neural dialogue generation. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1489– 1498. Association for Computational Linguistics, 2018.
- [Manning et al., 2014] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In Association for Computational Linguistics (ACL) System Demonstrations, pages 55–60, 2014.
- [Negi et al., 2009] S. Negi, S. Joshi, A. K. Chalamalla, and L. V. Subramaniam. Automatically extracting dialog models from conversation transcripts. In 2009 Ninth IEEE International Conference on Data Mining, pages 890–895, Dec 2009.
- [Peters et al., 2018] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. In Proc. of NAACL, 2018.
- [Prabhakaran et al., 2018] Vinodkumar Prabhakaran, Premkumar Ganeshkumar, and Owen Rambow. Author commitment and social power: Automatic belief tagging to infer the social context of interactions. CoRR, abs/1805.06016, 2018.

- [Sangroya et al., 2018] Amit Sangroya, C. Anantaram, Pratik Saini, and Mrinal Rawat. Extracting latent beliefs and using epistemic reasoning to tailor a chatbot. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18, pages 5853–5855. International Joint Conferences on Artificial Intelligence Organization, 7 2018.
- [Serban et al., 2015] Iulian Vlad Serban, Alessandro Sordoni, Yoshua Bengio, Aaron C. Courville, and Joelle Pineau. Hierarchical neural network generative models for movie dialogues. *CoRR*, abs/1507.04808, 2015.
- [Shi *et al.*, 2019] Weiyan Shi, Tiancheng Zhao, and Zhou Yu. Unsupervised dialog structure learning. *CoRR*, abs/1904.03736, 2019.
- [Weston, 2016] Jason Weston. Dialog-based language learning. In Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS'16, pages 829–837, USA, 2016. Curran Associates Inc.
- [Williams *et al.*, 2016] Jason D. Williams, Antoine Raux, and Matthew Henderson. The dialog state tracking challenge series: A review. *D&D*, 7(3):4–33, 2016.
- [Wu et al., 2018] Xianchao Wu, Ander Martinez, and Momo Klyen. Dialog generation using multi-turn reasoning neural networks. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pages 2049–2059. Association for Computational Linguistics, 2018.
- [Zhai and Williams, 2014] Ke Zhai and Jason D Williams. Discovering latent structure in task-oriented dialogues. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 36–46, Baltimore, Maryland, June 2014. Association for Computational Linguistics.