INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

CauSE: Causal Search Engine for Understanding Contact-Center Conversations

Anup Pattnaik[1], *Tanay Narshana*[1], *Aashraya Sachdeva, Cijo George, Jithendra Vepa*

Observe.AI, India

{anup.pattnaik, tanay.narshana, aashraya.sachdeva, cijo.george, jithendra}@observe.ai

Abstract

Contact centers sit on multitude of conversational data that contains helpful information which can assist businesses to deliver better outcomes like improving customer experience. However, finding such information manually is hard. Towards this end, we propose *CauSE, a causal search engine for understanding contact center conversations* that assist in finding relevant answers to a question. Using topic modelling, the engine identifies themes within conversational contexts to help reason for the given question. To address the challenge of multiple topics in a single context, we divide the context into Elementary Discourse Units (EDUs) and perform topic modelling on EDUs to better identify coherent themes as topics. Subsequently, we employ a novel contrastive ranking algorithm to surface meaningful topics, and LLM-prompting to obtain descriptions for the topics. Our evaluations of the resultant topics and proof of value exercises demonstrate the strength of the proposed engine.

Index Terms: conversational language, contact centre, topic modelling, topic ranking, mining transcripts, topic description, search engine

1. Introduction

Organizations manage customer interactions across various channels, such as phone calls, chats, and emails, via contact centers. Contact centers possess a multitude of data in the form of conversational transcripts, which encapsulate valuable insights into the behaviour of customers and agents during interactions. Such information can help contact centers deliver better business outcomes, such as increased sales and customer satisfaction. Moreover, it can also enable contact centers to identify flaws in their internal processes. Such varied business goals require probing possible reasons for different questions, such as why customers are unsatisfied, why customers want to escalate to a manager, and why sales are not happening in certain conversations. Here, the challenge lies in the fact that parsing through data at such a scale and identifying relevant instances of the conversations for each business-driven question demands extensive manual effort and is costly.

Motivated by this, we propose *CauSE*, *a search engine for understanding contact center conversations*. The engine analyses multiple conversational contexts together. A conversational context is a localized region within a conversation of interest. For example, to understand the cause of customer escalations, a conversational context would encompass the section of the conversation where the escalation took place. Such conversational contexts are obtained via key phrases, called a *query*, provided as input to the engine.

[1]Equal contribution

Broadly, our search engine attempts to identify themes within conversational contexts that can help reason for the question at hand. We tackle the challenge of identifying themes using topic modelling via clustering document embeddings. However, multiple themes may exist in a single conversational context, leading to small or non-coherent clusters. For instance, in the snippet "my name is Brenda, registered phone number is ******, I am calling in to cancel the subscription", a customer is calling in, verifying their phone number, and asking for a cancellation. To address this, we divide the context into *Elementary Discourse Units* (EDUs), which are atomic semantic units within discourses. Each EDU often consists of a single theme, allowing us to assume the overall semantics of a discourse when considering all its EDUs together. Moreover, clustering EDUs reduces the perturbations in document embeddings by reducing the expected number of ASR (speech transcription) errors per document. So, we apply topic modelling on EDUs of conversational contexts to better identify coherent themes as topics.

When clustering EDUs, we see an explosion in the number of clusters obtained, making them difficult to consume for a human. Most of the clusters obtained contain uneventful parts of the conversations. Thus, it becomes important to identify relevant/important clusters for the given business use case. For this, we attempt to solve the problem of topic ranking that aims to recommend relevant topics. This is challenging since there is no direct way of inferring relevant topics for the questions at hand. Therefore, we develop an innovative ranking algorithm that attempts to identify relevant clusters obtained from conversational contexts of interest by contrasting them with clusters obtained from contexts that are not of interest. Finally, we employ a fine-tuned LLM to generate descriptions for the topics.

We provide details and empirical results of our approach in the following sections.

2. System Overview

As shown in Figure 1, the search engine comprises three modules described in detail in this section.

2.1. Discourse Segmentation

The input query fetches conversational contexts of interest that contain about 300 tokens on average per context. Since these are parts of a conversation, each context often consists of multiple themes under discussion. However, we observe that clustering these conversational contexts result in either small or incoherent clusters. To obtain coherent themes as topics, we employ a neural segmentation model [1] to divide the longer text into smaller elementary discourse units (EDUs). On average, the segmentation module produces 50 EDUs for each call section. Subsequently, these EDUs are independently passed into

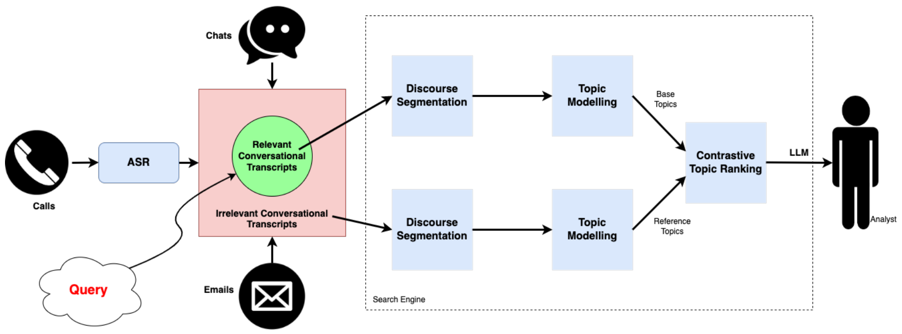

Figure 1: *Architecture Diagram of CauSE*

the topic model system. We manually validate that when EDUs are clustered, the topics obtained are more coherent and diverse.

2.2. Topic Modeling

We utilize BERTopic [2] to cluster EDUs. EDUs identified as outliers are discarded since they often contain EDUs with infrequent themes (including those with ASR errors). We represent each topic cluster with the 15 most diverse EDUs from within the cluster using Maximal Marginal Relevance (MMR) [3]. The topics obtained from the relevant contexts are called base topics (B). For our topic ranking algorithm (presented in Section 2.3), we also apply the same methodology on irrelevant conversational contexts and call the topics obtained to be reference topics (R).

2.3. Contrastive Topic Ranking and Topic Descriptions

We see an explosion of number of topics (> 100) when clustering EDUs which are difficult for a human to consume. Furthermore, a substantial number of topics that appeared at the top of the BERTopic output list comprised mundane or unrelated sections of conversations, such as customer detail verification or agent greetings. This is expected since conversations are typically long and mostly consist of common conversational exchanges. Thus, it is important to surface the most relevant topics given the context. Towards this end, we compare each topic in base topics (B) to identify if similar topics exist in reference topics (R) (refer to Section 2.2). We then assign a score to each topic in B based on how contrasting it is from topics in R. Based on this score, we assign ranks to each topic in B. We experimentally validate the merit of this approach and see that it improves the ranking of query-specific topics, pushing the more generic topics to the bottom. This ranking methodology ensures that topics in B that are semantically distinct from topics in R are ranked higher.

Finally, we prompt a fine-tuned LLM model based on Cerebras-GPT [4] to generate descriptions for each topic by using its representational EDUs.

3. Experimental Results and Conclusion

In this section, we present empirical evidence to demonstrate the strength of our search engine, *CauSE*. We show the quantitative performance of our ranking algorithm and the top five "important" reasons for a business use case.

To evaluate the performance of the contrastive ranking algorithm, we let five annotators mark the importance of topics obtained for two different use cases (data sets) on a scale of 1 to 3. The Fleiss Kappa score for these annotations was more than 6.5, indicating a good annotation agreement. We combine these annotations to assign each topic a final label of importance. The evaluation of the contrastive ranking algorithm shows that it surfaces at least 231% more relevant topics in the top 20 recommended topics than when each topic was equally likely to be in the top 20. Moreover, it surfaces from 100% to 450% more relevant topics than the default BERTopic ranks.

Table 1 displays the top five most relevant reasons obtained by the search engine for why customers want to escalate to managers. These reasons are very relevant and present actual concerns raised by customers for escalations.

To conclude, CauSE facilitates answering questions by identifying relevant topics for a given query. Our proof of value exercises with our customers and empirical results demonstrate the strength of this engine towards obtaining possible reasons for a question by understanding contact-centre conversations.

Table 1: *Top five relevant reasons for customer escalation for one of our clients.*

Topic Descriptions
It seems that the customers are expressing concerns and frustration about emails, such as not receiving them, needing them urgently, and requesting assistance in sending or receiving them.
It appears that the customers are complaining about being charged multiple times for the same thing, or being charged for things that they shouldn't be charged for.
The customers seem to be discussing issues related to their card information, such as providing or confirming the correct card number, billing, and usage. Some customers express concerns about the security of their card information or previous issues they have experienced.
The customers appear to be expressing frustration and dissatisfaction with the customer service they have received.
The customers seem to be discussing issues related to owing money, such as disputing the amount owed or claiming to have paid their balance in full. Some customers are also suggesting that the company owes them money or that there has been an error in the billing.

4. References

[1] Y. Wang, S. Li, and J. Yang, "Toward fast and accurate neural discourse segmentation," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 962–967.

[2] M. Grootendorst, "Bertopic: Neural topic modeling with a class-based tf-idf procedure," *arXiv preprint arXiv:2203.05794*, 2022.

[3] J. Carbonell and J. Goldstein, "The use of mmr, diversity-based reranking for reordering documents and producing summaries," in *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, 1998, pp. 335–336.

[4] N. Dey, G. Gosal, H. Khachane, W. Marshall, R. Pathria, M. Tom, J. Hestness *et al.*, "Cerebras-gpt: Open compute-optimal language models trained on the cerebras wafer-scale cluster," *arXiv preprint arXiv:2304.03208*, 2023.