

UNHIDE: UNSUPERVISED HUMAN-INTERPRETABLE DIALOGUE EXPLORATION

Anonymous ACL submission

Abstract

Organizations increasingly rely on human or virtual agents for customer service, generating vast amounts of dialogue data. Interpreting this data is essential for improving communication and customer satisfaction but remains challenging due to its volume, complexity, and nuances. We present UNHIDE, an unsupervised framework that supports human interpretation of dialogue data by discovering dialogue flows and computing interpretable metrics. These reveal insights into common paths, complexity, sentiment progression, and response speed. The potential of UNHIDE is showcased in dialogues generated with the previous variables, which can be observed in the produced analysis, demonstrating UNHIDE's potential for uncovering meaningful dialogue structures and improving conversational strategies.

1 Introduction

The development of traditional chatbots often relies on dialogue flows, as in tools like Google DialogFlow (Sabharwal et al., 2020) or Rasa (Kong et al., 2021). Such flows define possible conversation paths using a directed graph, where nodes represent states, and edges are transitions triggered by predefined actions. But dialogue flows can also be discovered from any collection of conversations, whether between humans (Sastre Martinez and Nugent, 2022; Carvalho et al., 2024) or humans and generative chatbots (Agrawal et al., 2024), without relying on predefined structures. Analyzing flows discovered in large collections of dialogues highlights communication trends, helps identifying issues or undesired scenarios, and informs actions to improve user experience. We present **Unsupervised Human-Interpretable Dialogue Exploration**, a framework that leverages unsupervised methods to support human interpretation of large dialogue collections. Since no single feature can capture all underlying nuances, UN-

HIDE goes beyond visualizing discovered flows, by providing a broad range of interpretable metrics. Extracted insights aid in analyzing trends, validating predefined protocols, and pinpointing issues, like negative sentiment or slower response speed, supporting informed decisions to enhance customer service and dialogue systems.

UNHIDE was validated on a synthetic task-oriented dialogue, generated with controlled utterance count, sentiment variation, and response speed, all reflected in the discovered flows and metrics. Our main contributions are:

- A framework for human-interpretable dialogue analysis, encompassing flow discovery, and a range of dialogue and flow metrics;
- A dataset of $\approx 10k$ task-oriented dialogues with varying sizes, sentiment variation, and response speed, valuable for future research;
- A demonstration of the framework on the new dataset, highlighting key dialogue variables and their impact.

The remainder of the paper is structured as follows: Section 2 reviews previous work on flow discovery and dialogue-collection analysis. Section 3 introduces UNHIDE, detailing dialogue flow discovery and the adopted metrics. Section 4 describes the experimental setup, including dialogue data generation, and showcases UNHIDE in action, discussing the results. Section 5 concludes the paper and outlines directions for future work.

2 Related Work

Dialogue flows have been mostly discovered to support dialogue systems development (Bouraoui and Lemaire, 2017; Sastre Martinez and Nugent, 2022; Xi et al., 2022; Sohn et al., 2023). They are typically represented as graphs, where nodes result from clustering semantically similar utterances (Bouraoui and Lemaire, 2017; Xi et al., 2022; Ferreira et al., 2024b; Agrawal et al., 2024;

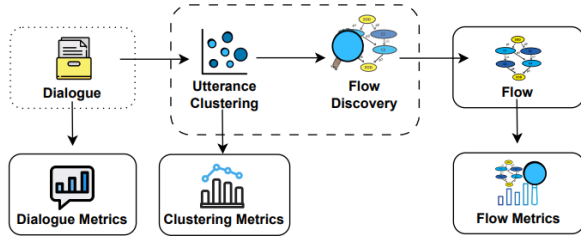


Figure 1: UNHIDE framework.

Kim et al., 2024), and transitions between clusters are derived from utterance sequences. To enhance interpretability, discovered clusters can be assigned natural language labels, often using LLMs (Agrawal et al., 2024; Ressurreição et al., 2024), which have also been used for annotating dialogues (King and Flanagan, 2024). Moreover, dialogue flows were used for constraining LLMs to specific domains (Agrawal et al., 2024), and for generating aligned question-answering dialogues (Kim et al., 2024). Relying on LLMs for producing synthetic dialogues with specific features is now common (Li et al., 2022; Terragni et al., 2023). A few studies (Holstrup et al., 2020; Carvalho et al., 2024; Ferreira et al., 2024a) have explored dialogue flow discovery for analysing trends and extracting insights. UNHIDE extends existing flow discovery and enhances it with metrics that analyze key features such as complexity, sentiment progression, and response speed.

3 UNHIDE

UNHIDE, illustrated in Figure 1, centers on flow discovery and computes metrics from three different representations: raw dialogues, clustered utterances, and flow graph.

3.1 Dialogue Flow

Given a collection of text-based dialogues between two participants, e.g., agent/system and customer/user, the flow is discovered in two steps, along the lines of Ferreira et al. (2024b):

A. Utterance Clustering groups each participant’s utterances, represented by their embeddings, according to their semantic similarity. Clusters define flow states. For better interpretability, human-readable labels are assigned to clusters. Various methods can be used, including an LLM, which can generate labels when prompted and given a sample of clustered utterances (Agrawal et al., 2024; Ressurreição et al., 2024).

B. Flow Discovery identifies paths between states and computes their probability, resulting in

a transition graph $G(C, T)$. Each node $c \in C$ corresponds to a state, while edges $t(c_1, c_2, p_{12}) \in T$ represent transitions between states c_1 and c_2 , weighted by their probability p_{12} (i.e., the proportion of utterances in c_1 that are followed by utterances in c_2). All flows begin and end with special states, SOD (Start of Dialogue) and EOD (End of Dialogue). A threshold (θ) can be applied to filter out low-probability transitions.

3.2 Metrics

To enhance human interpretability, the discovered flow is enriched with metrics offering diverse data perspectives. They are categorized by input: (i) dialogue metrics from raw dialogues; (ii) clustering metrics from discovered clusters; and (iii) flow metrics from the overall structure. Table 1 details each metric, its description and computation method.

4 Experimentation

UNHIDE was implemented and tested on synthetic dialogues that match predefined variables. All experiments were run on an NVIDIA RTX A6000 with 48GB of VRAM.

4.1 Setup

From a set of written dialogues, utterances are embedded in 384-dimension vectors, obtained from the sentence transformer *all-MiniLM-L6-v2*¹, and clustered with the scikit-learn implementation of K-means². The number of clusters is selected for maximizing the Silhouette score (Rousseeuw, 1987), according to Optuna³, which tests configurations within a range of 3-10 clusters for each participant. θ is 0.1 in all experiments.

Natural-language labels are generated for the clusters by a quantized Llama3-8b model, running in Ollama⁴. For each cluster, the model is prompted to provide a label that captures the main actions in a random sample of 30 utterances. Llama3 is also prompted for classifying the sentiment of each utterance, as positive, negative, or neutral. The temperature is always set to 0.1, and used prompts are in Appendix A.

4.2 Dataset

A task-oriented dialogue dataset was generated automatically, with subsets of dialogues designed to target specific variables expected to be reflected

¹[hf.co/sentence-transformers/all-MiniLM-L6-v2](https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2)

²<https://tinyurl.com/4ymet8ff>

³<https://optuna.org>

⁴<https://ollama.com/>

	Metrics	Description	Notation
Dataset	Utterances in Dataset	For each dataset, total number of utterances	$ U $
	Utterances in Dialogue	For each dialogue, average number of utterances	$AUD = \frac{ U }{ D }$
	Dialogue Duration	For each dialogue, average time since the previous utterance	$ADD = \frac{\sum_{d \in D} \text{time_start}_d^{\text{last}}}{ D }$
	Sentiment Distribution	For each dialogue, proportion of positive and negative sentiments	$SD^+ = \frac{ U^+ }{ U }, SD^- = \frac{ U^- }{ U }$
Clustering	Number of States	Total number of discovered dialogue states	$ C $
	Silhouette Score	An indicator of cluster quality, considering cohesion and separation	$SS = \max \left(\frac{b(s) - a(s)}{\max(a(s), b(s))} \right)$
	Utterances in State	For each state, number of clustered utterances	$ U_c $
	Utterances/State	Average number of utterances per state	$AU_c = \frac{\sum_{c \in C} U_c }{ C }$
	Tokens/utterance	For each state, average number of tokens per utterance (NLTK tokenizer)	$TU = \frac{\sum_{u \in U} W_u }{ U }$
	TimeSince(SOD)	For each state, average time since SOD	$TS(SOD) = \frac{\sum_{u \in U_c} \text{time_start}_u}{ U_c }$
	TimeSince(Prev)	For each state, average time since the previous utterance	$TS(\text{Prev})_c = \frac{\sum_{i=1}^{ U_c -1} (\text{time}_{i+1} - \text{time}_i)}{ U_c -1}$
	Sentiment in State	For each state, average sentiment of utterances	$S_c = \frac{\sum_{u \in U_c} S_u}{ U_c }$
	Sentiment Cohesion	For each state, average standard deviation of sentiment	$FSC = \frac{\sum_{c=1}^{ C } \sigma(S_c)}{ C }$
	Number of Transitions	Total number of transitions between dialogue states	$ T $
Dialogue	Initial Sentiment	Average sentiment of states connected to SOD	$IS = \frac{\sum_{c \in C_{SOD}} S_c}{ C_{SOD} }$
	Final Sentiment	Average sentiment of states connected to EOD	$FS = \frac{\sum_{c \in C_{EOD}} S_c}{ C_{EOD} }$
	$\Delta Sentiment$	Sentiment variation between SOD and EOD	$\Delta Sentiment = S_{EOD} - S_{SOD}$
	EOD_-	Proportion of utterances reaching EOD with negative sentiment	$EOD_- = \frac{ U_{EOD}^- }{ U_{EOD} }$
	Flow Density	Graph density, an indicator of complexity	$FD = \frac{2 T }{ C (C -1)}$
	Flow Accuracy	Proportion of transitions covered in unseen (test) dialogues	$Acc = \frac{ \{t \in T_{test} \cap t \in T_{Flow}\} }{ T_{test} }$

Table 1: Metrics computed in raw dialogues, clustering, and flow, where utterances ($u \in U$) contain tokens ($w \in W$) with sentiment (+, −, none) and timestamps. States link through transitions ($t \in T$).

in UNHIDE’s analysis, namely: (i) sentiment progression, (ii) number of utterances, and (iii) response speed. Rather than generating dialogues from scratch, the dataset content was aligned with the widely-used MultiWOZ 2.2 (Zang et al., 2020). For each MultiWOZ dialogue, Llama3 was prompted (temperature=0.7) to generate a new dialogue based on the original, incorporating a predefined sentiment progression and a target number of utterances (see prompt template in Appendix A).

Target utterance counts were set based on MultiWOZ dialogues, using the first quartile (*at most 10 utterances*) for shorter dialogues and the third (*at least 18*) for longer ones. Shorter and longer MultiWOZ dialogues served as references for the few- and many-utterance categories, maintaining the original train-test split. The combination of four sentiment progressions, two utterance lengths, and two response speeds resulted in 16 ($4 \times 2 \times 2$) dialogue subsets (see Table 2)⁵.

After generation, timestamps were assigned to each utterance as having fast or slow response speed. Time gaps were randomly drawn from a normal distribution: fast responses had a mean of 10 ± 5 s, and slow had a mean of 120 ± 60 s. Examples of generated dialogues are in Appendix B.

⁵To balance subset sizes, 12 additional dialogues were added to training and 8 to testing, as MultiWOZ’s 8,436 training and 1,000 test dialogues are not evenly divisible by 16.

4.3 Analysis

To validate how well target variables are reflected in the discovered flow, we define key hypotheses: (i) the number of utterances should be reflected in $|U|$ and, together with response time, in ADD ; (ii) the number of utterances should also influence complexity, i.e., shorter dialogues should produce simpler flows, with fewer states and transitions, and vice-versa; (iii) sentiment progression should be reflected in the color of the transitions (green, red) and in metrics like IS , FS , EOD_- , $\Delta Sent$; (iv) response speed could also impact sentiment dynamics, but, since they were generated after the dialogues, this correlation cannot be analyzed.

Table 2 summarizes the target variables for each subset and reports selected dialogue and clustering metrics. Additional state-specific cluster metrics and discovered flows are in Appendix C.

For illustrative purposes, we describe the flow for S12, in Figure 2. As expected, the sentiment evolves from initially negative ($IS = 0.39$) to positive ($FS = 0.98$), highlighted by transition colors shifting from red to green. Dialogues begin with *Emotional Venting* in 35% of the cases, where the user expresses negative sentiments. In 48% of instances, the system responds with *Restaurant Recommendations*, attempting to uplift sentiment. Towards the end, 38% of dialogues follow a predominantly positive trajectory, moving through

Subset	Generation Variables				Dialogue Metrics					Flow Metrics					
	Sentiment	Responses	Utters		U	AUc	AUD	ADD	SD ⁺ /SD ⁻	IS	FS	$\Delta Sent$	EOD ₋	FD	Acc
S1	Pos	Fast	Few		4,909	272.72 $\pm$ 70.23	9.32 $\pm$ 1.17	79.21 $\pm$ 17.21	2,658/14	0.75	0.75	0.25	0.00	0.14	0.70
S2	Neg	Fast	Few		5,145	285.83 $\pm$ 113.28	9.81 $\pm$ 0.59	83.70 $\pm$ 13.64	682/1,560	0.41	0.43	0.12	0.42	0.18	0.80
S3	Pos to Neg	Fast	Few		5,076	282.00 $\pm$ 103.62	9.67 $\pm$ 0.73	83.37 $\pm$ 14.28	2,481/1,631	0.84	0.42	-0.28	0.78	0.14	0.69
S4	Neg to Pos	Fast	Few		5,000	333.33 $\pm$ 144.97	9.72 $\pm$ 0.69	82.63 $\pm$ 17.47	2,212/1,710	0.29	0.86	-0.31	0.00	0.19	0.83
S5	Pos	Slow	Few		4,998	294.00 $\pm$ 183.31	9.47 $\pm$ 1.02	977.70 $\pm$ 191.63	2,581/16	0.76	0.76	0.009	0.00	0.18	0.74
S6	Neg	Slow	Few		5,085	635.00 $\pm$ 280.99	9.76 $\pm$ 0.70	1,005.15 $\pm$ 185.63	658/1,461	0.43	0.38	-0.12	0.15	0.30	0.93
S7	Pos to Neg	Slow	Few		5,022	456.55 $\pm$ 199.26	9.68 $\pm$ 0.70	986.05 $\pm$ 191.31	2,424/1,680	0.82	0.40	-0.89	0.58	0.24	0.85
S8	Neg to Pos	Slow	Few		4,977	355.52 $\pm$ 189.74	9.70 $\pm$ 0.74	977.19 $\pm$ 220.57	2,172/1,679	0.12	0.78	0.60	0.00	0.17	0.82
S9	Pos	Fast	Many		11,727	781.80 $\pm$ 414.27	22.83 $\pm$ 5.79	206.14 $\pm$ 48.40	6,197/40	0.77	0.96	0.003	0.00	0.16	0.75
S10	Neg	Fast	Many		12,875	858.33 $\pm$ 459.29	25.65 $\pm$ 7.13	226.88 $\pm$ 49.58	2,340/3,058	0.42	0.62	-0.009	0.68	0.19	0.74
S11	Pos to Neg	Fast	Many		13,976	735.58 $\pm$ 269.71	27.70 $\pm$ 7.39	247.69 $\pm$ 54.23	6,085/2,350	0.79	0.40	-0.81	0.37	0.12	0.65
S12	Neg to Pos	Fast	Many		11,138	795.57 $\pm$ 242.17	22.92 $\pm$ 6.78	194.95 $\pm$ 42.75	4,486/2,328	0.39	0.98	0.006	0.00	0.18	0.81
S13	Pos	Slow	Many		11,699	835.64 $\pm$ 315.17	23.39 $\pm$ 6.07	2,458.81 $\pm$ 534.14	6,241/40	0.78	0.81	-0.72	0.00	0.21	0.83
S14	Neg	Slow	Many		13,408	957.71 $\pm$ 370.77	26.92 $\pm$ 7.61	2,830.63 $\pm$ 688.10	2,414/3,606	0.43	0.68	-0.25	0.00	0.21	0.85
S15	Pos to Neg	Slow	Many		13,711	761.72 $\pm$ 248.99	27.66 $\pm$ 7.97	2,918.13 $\pm$ 621.87	5,716/2,687	0.76	0.40	-0.74	0.44	0.13	0.64
S16	Neg to Pos	Slow	Many		11,416	761.07 $\pm$ 243.02	23.35 $\pm$ 6.47	2,377.37 $\pm$ 467.82	4,681/1,875	0.42	0.99	0.28	0.00	0.18	0.81

Table 2: Subsets of the synthetic dialogue dataset, target variables, and corresponding dialogue and flow metrics.

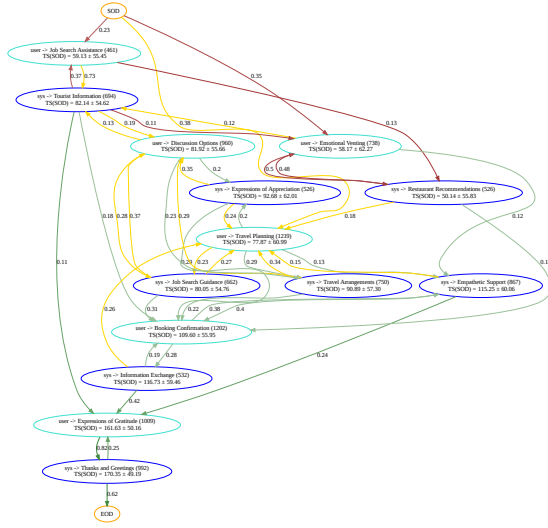


Figure 2: Dialogue flow discovered for Subset 12.

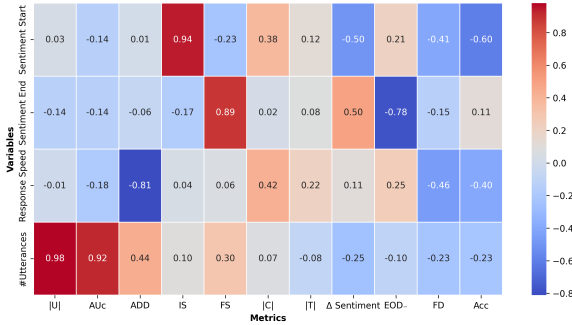


Figure 3: Correlation of variables vs. metrics.

Empathetic Support by the system, followed by *Expressions of Gratitude* from the user, and end with *Thanks and Greetings* from the system.

To validate the computed metrics, correlations between subset variables and metrics were analyzed (Figure 3). Sentiment progression was split into two variables. Sentiment Start correlates strongly with *IS*, while Sentiment End correlates with *FS* and *EOD-*, and both are moderately correlated with $\Delta Sentiment$. Response Speed correlates negatively with *ADD*, confirming that faster responses lead to shorter dialogues. The

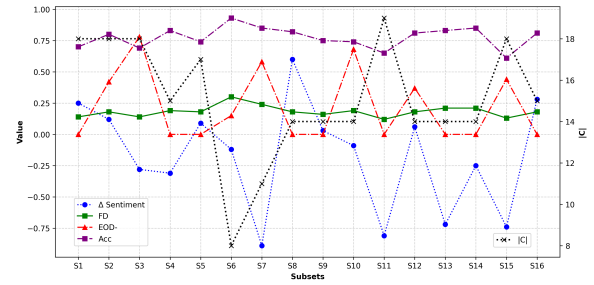


Figure 4: Variation of different metrics across subsets.

Number of Utterances is highly correlated with $|U|$, but we could not confirm its correlation with the number of clusters ($|C|$), transitions ($|T|$), nor density (*FD*), meaning that it does not affect flow complexity, only the size of each cluster (AU_c).

The variation of a selection of metrics for the 16 subsets is depicted in Figure 4. In addition to our hypotheses, we highlight the stable $\Delta Sentiment$ in shorter dialogues, and the correlation between higher density (*FD*) and representation of unseen dialogues (i.e., higher accuracy).

5 Conclusion and Future Work

Customer service dialogue analysis improves communication but is challenging due to data volume and complexity. UNHIDE is a framework for analyzing large dialogue collections with flow discovery and interpretable metrics. It was validated on a novel task-oriented dataset, varying in utterances, response speed, and sentiment, that will be released upon publication.

The produced analysis confirmed most of our hypotheses on how each variable would be observed, and further showed the lower variation of sentiment in shorter dialogues and that more dense flows capture more transitions in unseen dialogues (test).

Future work includes simulating user profiles, adjusting LLM temperature for dialogue variability, exploring additional metrics, and adapting UNHIDE for LLM interpretability.

6 Limitations

The main limitations of this work rise from the use of synthetic data. One of the reasons for this is because the real data we have tested it on is proprietary and cannot be publicly released. Generating data was shortcut to validate the platform, with the additional advantage that we could control some variables by simple prompting. This means that the dialogues used may not reflect a real-world scenario, which adds to the fact that they are inspired by dialogues from MultiWOZ, which does not exactly present a real-world scenario. On the other hand, since we know the variables that these dialogues follow, we see them as the best way to validate the framework, and thus set expectations on the produced analysis. Nevertheless, in the future, the utility of the framework will be validated in real dialogue data, and its analysis will have to be done together with experts on the used data.

Even though the options taken should not affect much the conclusions, the obtained results are limited to the adopted implementation of the framework. Utterances are represented by generic embeddings, which might not be best suited for dialogue; the ideal number of clusters was selected exclusively based on the Silhouette method; and LLMs (small, 8b), which are known problems such as inconsistency and hallucination, were used for classifying the sentiment of the utterances and labelling the clusters, in a zero-shot scenario. However, the most important was that the predominant sentiments were there and that the flows were interpretable, which, from the resulting flows and computed metrics, we believe to have been accomplished.

Nevertheless, we highlight that most of the computation of UNHIDE, i.e., clustering, flow discovery, metrics, can be run locally, without any specific computational requirements. Moreover, if necessary, the LLM-based methods for cluster labelling and sentiment classification may be replaced by less demanding methods, e.g., keyword-based and more traditional classifiers.

References

Stuti Agrawal, Pranav Pillai, Nishi Uppuluri, Revanth Gangi Reddy, Sha Li, Gokhan Tur, Dilek Hakkani-Tur, and Heng Ji. 2024. [Dialog flow induction for constrainable LLM-based chatbots](#). In *Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 66–77,

Kyoto, Japan. Association for Computational Linguistics. 313
314

Jean-Leon Bouraoui and Vincent Lemaire. 2017. Cluster-based graphs for conceiving dialog systems. In *Proceedings of ECML-PKDD 2017 Workshop on Interactions between Data Mining and Natural Language Processing*. CEUR-WS.org. 315
316
317
318
319

Isabel Carvalho, Patrícia Ferreira, Ana Alves, Catarina Silva, and Hugo Gonalo Oliveira. 2024. Analysing customer-support trends in social networks through dialogue flow discovery. In *Proceedings of Iber-SPEECH 2024*, pages 56–60, Aveiro, Portugal. ISCA. 320
321
322
323
324
325

Patrícia Ferreira, Isabel Carvalho, Ana Alves, Catarina Silva, and Hugo Gonalo Oliveira. 2024a. Sentiment-aware dialogue flow discovery for interpreting communication trends. In *Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 274–288, Kyoto, Japan. Association for Computational Linguistics. 326
327
328
329
330
331
332

Patrícia Ferreira, Daniel Martins, Ana Alves, Catarina Silva, and Hugo Gonalo Oliveira. 2024b. Unsupervised flow discovery from task-oriented dialogues. *arXiv preprint arXiv:2405.01403*. 333
334
335
336

Alexander Holstrup, Lasse Starklit, and Andrea Burattin. 2020. Analysis of information-seeking conversations with process mining. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE. 337
338
339
340
341

Yongil Kim, Yerin Hwang, Hyunkyung Bae, Taegwan Kang, and Kyomin Jung. 2024. Flowlogue: A novel framework for synthetic dialogue generation with structured flow from text passages. *IEEE Access*. 342
343
344
345

Brendan King and Jeffrey Flanigan. 2024. [Unsupervised end-to-end task-oriented dialogue with LLMs: The power of the noisy channel](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 8283–8300, Miami, Florida, USA. Association for Computational Linguistics. 346
347
348
349
350
351
352

Xiaoquan Kong, Guan Wang, and Alan Nichol. 2021. *Conversational AI with Rasa: Build, test, and deploy AI-powered, enterprise-grade virtual assistants and chatbots*. Packt Publishing Ltd. 353
354
355
356

Zekun Li, Wenhui Chen, Shiyang Li, Hong Wang, Jing Qian, and Xifeng Yan. 2022. [Controllable dialogue simulation with in-context learning](#). In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 4330–4347, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics. 357
358
359
360
361
362
363

David Ressurreiao, Patrícia Ferreira, Ana Alves, and Hugo Gonalo Oliveira. 2024. Optimising dialogue flow discovery. In *EPIA Conference on Artificial Intelligence*, pages 163–175. Springer. 364
365
366
367

Peter J Rousseeuw. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

Navin Sabharwal, Amit Agrawal, Navin Sabharwal, and Amit Agrawal. 2020. Introduction to google dialogflow. *Cognitive virtual assistants using google dialogflow: develop complex cognitive bots using the google dialogflow platform*, pages 13–54.

Javier Miguel Sastre Martinez and Aisling Nugent. 2022. Inferring ranked dialog flows from human-to-human conversations. In *Proceedings 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 312–324, Edinburgh, UK. ACL.

Sungryull Sohn, Yiwei Lyu, Anthony Liu, Lajanugen Logeswaran, Dong-Ki Kim, Dongsub Shim, and Honglak Lee. 2023. *TOD-flow: Modeling the structure of task-oriented dialogues*. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 3355–3371, Singapore. Association for Computational Linguistics.

Silvia Terragni, Modestas Filipavicius, Nghia Khau, Bruna Guedes, André Manso, and Roland Mathis. 2023. In-context learning user simulators for task-oriented dialog systems. *arXiv preprint arXiv:2306.00774*.

Xiangyu Xi, Chenxu Lv, Yuncheng Hua, Wei Ye, Chaobo Sun, Shuaipeng Liu, Fan Yang, and Guanglu Wan. 2022. A low-cost, controllable and interpretable task-oriented chatbot: With real-world after-sale services as example. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3398–3402.

Xiaoxue Zang, Abhinav Rastogi, Srinivas Sunkara, Raghav Gupta, Jianguo Zhang, and Jindong Chen. 2020. MultiWOZ 2.2 : A dialogue dataset with additional annotation corrections and state tracking baselines. In *Procs of the 2nd Workshop on Natural Language Processing for Conversational AI*, pages 109–117, Online. ACL.

A Prompts to LLM

The prompts used for (i) labelling clusters of semantically-similar utterances and (ii) classifying the sentiment of each utterance, are in Figures 5 and 6 respectively. They were both used with Llama3, with temperature 0.1.

B Generated Dialogues

Table 3 presents dialogue examples from synthetic dataset with different variables, namely: the first is mostly positive, fast responses, with few utterances (6); the second is mostly negative, slow responses,

Please provide a label that captures the main actions of the following sentences in English: [INPUT]. Answer with the label only in English, using the format label: {your_label}, and with a maximum of three words.

Figure 5: Prompt model for generating labels for dialogue states, where the INPUT variable is replaced by a random sample of 30 clustered utterances.

You are an assistant specialized in analyzing sentiment in customer service dialogues. Classify the sentiment of the utterance below according to the following definitions: Negative: the utterance expresses complaints, frustration, dissatisfaction, or something with a negative context. Neutral: the utterance is neutral, factual, or does not convey a clear positive or negative sentiment. Positive: the utterance conveys satisfaction, gratitude, or something positive. {utterance} Only respond with the corresponding word (Negative, Neutral, or Positive).

Figure 6: Prompt model for sentiment classification of each dialogue utterance, with utterance replaced by the respective utterance:

Generate a new dialogue based on the structure and style of the original dialogue provided below. The new dialogue should meet the following criteria:

- The sentiment of the dialogue {sentiment}.
- The dialogue must have {number_utterances}.

Original dialogue: {original_dialogue}. Each utterance of the dialogue must start with ‘USER:’ or ‘SYSTEM:’.

Just put the dialogue without any text before or after the dialogue.

Figure 7: Prompt template for dialogue generation, with variables: {sentiment}, with possible values *positive*, *negative*, *positive to negative*, *negative to positive*; {number_utterances}, *at most 10 utterances*, *at least 18 utterances*; {original_dialogue}, from MultiWOZ.

with few utterances (8); the third shifts from positive to negative, fast responses, with many utterances (18); the fourth shifts from negative to positive, slow responses, with many utterances (18).

C Report

The following sections report, in detail, the results of applying UNHIDE to sixteen subsets, generated with different variables regarding sentiment, response time, and number of utterances. The report includes tables with the value of cluster metrics, as

Speaker	Dialogue	Sentiment	TS(SOD)
Dialogue 1: Positive sentiment, fast response time, and few number of utterances			
User	I'm looking for a place to grab dinner in Cambridge, do you have any recommendations?	Neutral	0
System	There's a lovely Italian restaurant called Bella Vita just off the main square. Would you like their address?	Positive	11.23
User	Yes, that sounds great! What's the address again?	Positive	19.01
System	It's 34 Market Street. They're known for their delicious pasta dishes and warm atmosphere.	Positive	33.46
User	That sounds perfect, thank you so much!	Positive	37.72
System	You're welcome! If you need any more help or have any other questions, feel free to ask. Have a great evening!	Positive	48.74
Dialogue 2: Negative sentiment, slow response time, and few number of utterances			
User	I heard terrible things about the food at Cow Pizza Kitchen and Bar.	Negative	0
System	Yes, unfortunately it's true that many customers have complained about the quality of their dishes.	Negative	180
User	What kind of complaints are they?	Neutral	360
System	Mostly about overpriced food and poor service. Some people also reported getting sick after eating there.	Negative	525.26
User	That sounds awful. Have you heard anything about their cleanliness?	Negative	705.25
System	Unfortunately, yes. There have been reports of dirty floors and utensils being reused.	Negative	735.08
User	Ugh, I'm glad I never went there.	Negative	854.94
System	Yeah, it's definitely not a place to recommend.	Negative	936.03
Dialogue 3: Positive to negative sentiment, fast response time, and many utterances			
User	I'm really looking forward to trying some high-end Japanese cuisine with my friends.	Positive	0
System	That sounds like a great idea! There are several excellent sushi restaurants in the area that fit your description. Would you like me to recommend one?	Positive	9.15
User	Yeah, that would be awesome! We're willing to spend up to \$50 per person.	Positive	21.18
System	I think Sushi-Genki is an excellent choice. It's known for its high-quality fish and unique presentation.	Positive	24.25
User	That sounds perfect. What's the address again?	Positive	32.61
System	The address is 123 Main St, but you can also get there by taking the subway to the nearest stop.	Neutral	35.84
User	Got it, thanks! So what's the atmosphere like at Sushi-Genki? Is it romantic or more lively?	Neutral	41.72
System	It's actually a great spot for both. They have private rooms available if you want to make a special night out of it.	Neutral	56.09
User	That sounds perfect. We'll definitely consider that option.	Positive	56.55
System	And don't forget to try their signature roll, the "Dragon Roll"! It's a unique combination of flavors and textures.	Positive	69.73
User	Mmm, that does sound delicious. Thanks for the recommendation!	Positive	84.72
System	No problem at all! I hope you have a great time at Sushi-Genki.	Positive	96.40
User	Me too! But wait, what's this? A bad review just popped up on Yelp saying the service is terrible.	Negative	107.16
System	Oh no, I'm so sorry to hear that. Unfortunately, it looks like Sushi-Geni has had some issues with their staff in the past.	Negative	115.99
User	What?! That's ridiculous! How can a high-end restaurant have such poor service?	Negative	130.97
System	I know it's disappointing. Maybe you could try contacting them directly and asking about their current situation?	Negative	140.01
User	Yeah... maybe... but this is really putting a damper on our plans. We were so excited to go there.	Negative	149.21
System	I'm sorry to hear that. Sometimes even the best places can have off days. Would you like me to suggest an alternative?	Negative	153.80
Dialogue 4: Negative to positive sentiment, slow response time, and many utterances			
User	I'm having a terrible day, can you help me book a taxi to go from Jinling Noodle Bar over to the train station?	Negative	0
System	Sorry to hear that, but of course! When would you like the taxi to arrive at the train station?	Negative	68.73
User	Honestly, I just want it there by 18:00. That's the only thing I can think about right now.	Negative	210.50
System	Booking completed. Please look out for a silver Toyota! Contact number is 07654321987. Is there anything else I can help with?	Neutral	390.51
User	No...I don't think so. Just need to get home...	Negative	511.78
System	Okay, I hope things start looking up for you soon. Would you like me to remind you about the pickup time?	Neutral	653.93
User	Yeah, that would be great. Thanks.	Positive	821.57
System	Sure thing! I'll send a reminder at 17:45. Is there anything else on your mind that's bothering you?	Neutral	893.41
User	Just this stupid cold I've got. It's been going around and I think I picked it up from someone in the office.	Negative	1003.46
System	Ah, sorry to hear that! Maybe some hot tea or chicken soup will help clear things up?	Neutral	1104.96
User	Yeah... I hope so. Thanks for listening, actually.	Positive	1205.91
System	No problem at all! Sometimes talking about what's bothering us can really help put things into perspective. Do you have any plans for the weekend?	Neutral	1380.58
User	Not really. I was thinking of just taking it easy and watching some movies or something.	Neutral	1560.57
System	Sounds like a great plan to me! Maybe you could even treat yourself to a nice dinner or a favorite movie?	Positive	1661.98
User	That actually sounds pretty good right now. Thanks for chatting with me, I feel a bit better already!	Positive	1841.99
System	You're welcome! It was my pleasure. If you need anything else at all, don't hesitate to reach out.	Positive	1941.15
User	No, that's it. Thanks again!	Positive	2053.01
System	Enjoy the rest of your evening and get well soon! Goodbye!	Positive	2183.69

Table 3: Dialogue examples with varying sentiment, response speed, and utterance count.

well as the corresponding dialogue flows discovered.

Subset 1 (S1) has a predominantly positive sentiment, a fast response time, and a low number of utterances. Table 4 shows the cluster metrics for the user and the system. Figure 8 presents a dialogue flow discovered for this subset, with $|T| = 54$ transitions, $|C| = 9$ states for both the user and the system. For the user, the SS is 0.06

and the FSC is 0.22; for the system, the SS is 0.06 and the FSC is 0.23.

Subset 2 (S2) is characterized by a higher prevalence of negative sentiment, fast interactions, and a low number of utterances. The clustering metrics for the user and system are detailed in Table 5. For the user, there are $|C| = 9$ states, with a SS of 0.04 and a FSC of 0.29. Similarly, for the system, $|C| = 9$ states, with a SS of 0.05 and a FSC of 0.31.

State		U_c	TU	TS(SOD)	TS(Prev)	S_c
User						
C0: Restaurant Recommendations	365	10.43 ± 4.34	31.54 ± 21.46	8.66 ± 4.81	0.70 ± 0.25	
C1: Cambridge Recommendations	276	13.14 ± 3.69	22.78 ± 27.56	4.73 ± 5.64	0.82 ± 0.24	
C2: Positive Confirmations	406	13.53 ± 4.27	17.45 ± 24.45	4.25 ± 5.50	0.73 ± 0.25	
C3: Hotel Booking Requests	198	9.02 ± 2.81	48.79 ± 23.02	9.60 ± 4.31	0.99 ± 0.07	
C4: Booking Conversations	254	13.48 ± 3.98	26.73 ± 27.46	6.24 ± 5.92	0.70 ± 0.25	
C5: Expressive Agreement	356	13.56 ± 4.18	46.97 ± 22.55	9.20 ± 4.05	0.77 ± 0.25	
C6: Closing Thanks	88	4.32 ± 1.29	40.59 ± 22.06	9.66 ± 3.89	0.80 ± 0.25	
C7: Hotel Discussion	245	8.44 ± 2.70	62.53 ± 18.22	9.62 ± 4.04	0.88 ± 0.21	
C8: Information Requests	277	13.01 ± 4.51	39.07 ± 25.40	8.40 ± 5.03	0.74 ± 0.25	
System						
C0: Museum Recommendations	220	22.43 ± 6.86	31.34 ± 23.35	9.77 ± 4.18	0.79 ± 0.25	
C1: Travel Assistance	271	19.79 ± 7.87	44.96 ± 27.80	9.56 ± 4.36	0.62 ± 0.21	
C2: Park Recommendations	321	22.52 ± 7.15	35.40 ± 25.07	9.26 ± 4.32	0.77 ± 0.26	
C3: Vacation Recommendations	203	17.38 ± 5.25	27.00 ± 23.32	9.42 ± 4.41	0.71 ± 0.25	
C4: Closing Remarks	278	11.69 ± 5.05	72.11 ± 18.57	9.94 ± 4.16	0.97 ± 0.11	
C5: Confirmation Process	279	22.06 ± 6.45	54.75 ± 25.55	9.67 ± 4.04	0.74 ± 0.26	
C6: Confirmation Transactions	299	20.38 ± 6.81	56.58 ± 22.25	9.40 ± 4.22	0.67 ± 0.24	
C7: Hotel Recommendations	244	23.29 ± 7.73	40.46 ± 25.12	9.57 ± 4.53	0.75 ± 0.25	
C8: Restaurant Options	329	22.34 ± 6.96	35.38 ± 26.13	9.44 ± 4.19	0.80 ± 0.25	

Table 4: Clustering metrics for Subset 1.

Figure 9 illustrates a dialogue flow discovered for this subset, with $|T| = 67$ transitions.

Subset 3 (S3) is characterized by an initial positive sentiment that gradually shifts towards a more negative sentiment, with fast response times and a low number of utterances. The clustering metrics for the user and system are detailed in Tables 6. The discovered dialogue flow consists of $|C| = 9$ states, with a SS of 0.05 and a FSC of 0.37 for the user, and $|C| = 9$ states, with a SS of 0.05 and a FSC of 0.40 for the system. Figure 10 illustrates this flow, with a total of $|T| = 52$ transitions and $AU_c = 282 \pm 103.62$.

Subset 4 (S4) contains dialogues with an initially negative sentiment that gradually shifts throughout the conversation into a positive sentiment, with fast response times and a low number of utterances. The clustering metrics for the user and the system are detailed in Table 7. The discovered dialogue flow has $|C| = 5$ states, with an SS of 0.06 and an FSC of 0.38 for the user, and $|C| = 10$ states, with an SS of 0.07 and an FSC of 0.33 for the system. Figure 11 illustrates this flow for this subset, with a total of $|T| = 51$ transitions.

Subset 5 (S5) contains dialogues with a positive sentiment, with slow response times and few utterances. The clustering metrics for the user and the system are detailed in Tables 8. The discovered dialogue flow has $|C| = 8$ states, with an SS of 0.06 and an FSC of 0.24 for the user, and $|C| = 9$ states, with an SS of 0.06 and an FSC of 0.23 for

the system. Figure 12 illustrates this flow for this subset, with a total of $|T| = 62$ transitions.

Subset 6 (S6) contains dialogues with a negative sentiment, with slow response times and few utterances. The clustering metrics for the user and the system are detailed in Table 9. The discovered dialogue flow has $|C| = 4$ states, with an SS of 0.05 and an FSC of 0.28 for the user, and $|C| = 4$ states, with an SS of 0.06 and an FSC of 0.32 for the system. Figure 13 illustrates this flow for this subset, with a total of $|T| = 27$ transitions.

Subset 7 (S7) includes dialogues that start with a positive sentiment and end with a negative one, characterized by slow response times and few utterances. The clustering metrics for the user and the system are detailed in Table 10. The discovered dialogue flow has $|C| = 5$ states, with an SS of 0.05 and an FSC of 0.38 for the user, and $|C| = 6$ states, with an SS of 0.05 and an FSC of 0.40 for the system. Figure 14 illustrates this flow for this subset, with a total of $|T| = 37$ transitions and $AU_c = 456.55 \pm 199.26$.

Subset 8 (S8) includes dialogues that start with a negative sentiment and evolve into a positive one, characterized by slow response times and few utterances. The clustering metrics for the user and the system are detailed in Table 11. The discovered dialogue flow, illustrated in Figure 15, has $|C| = 9$ states, with an SS of 0.06 and an FSC of 0.34 for the user, and $|C| = 5$ states, with an SS of 0.07 and an FSC of 0.32 for the system, totaling $|T| = 41$

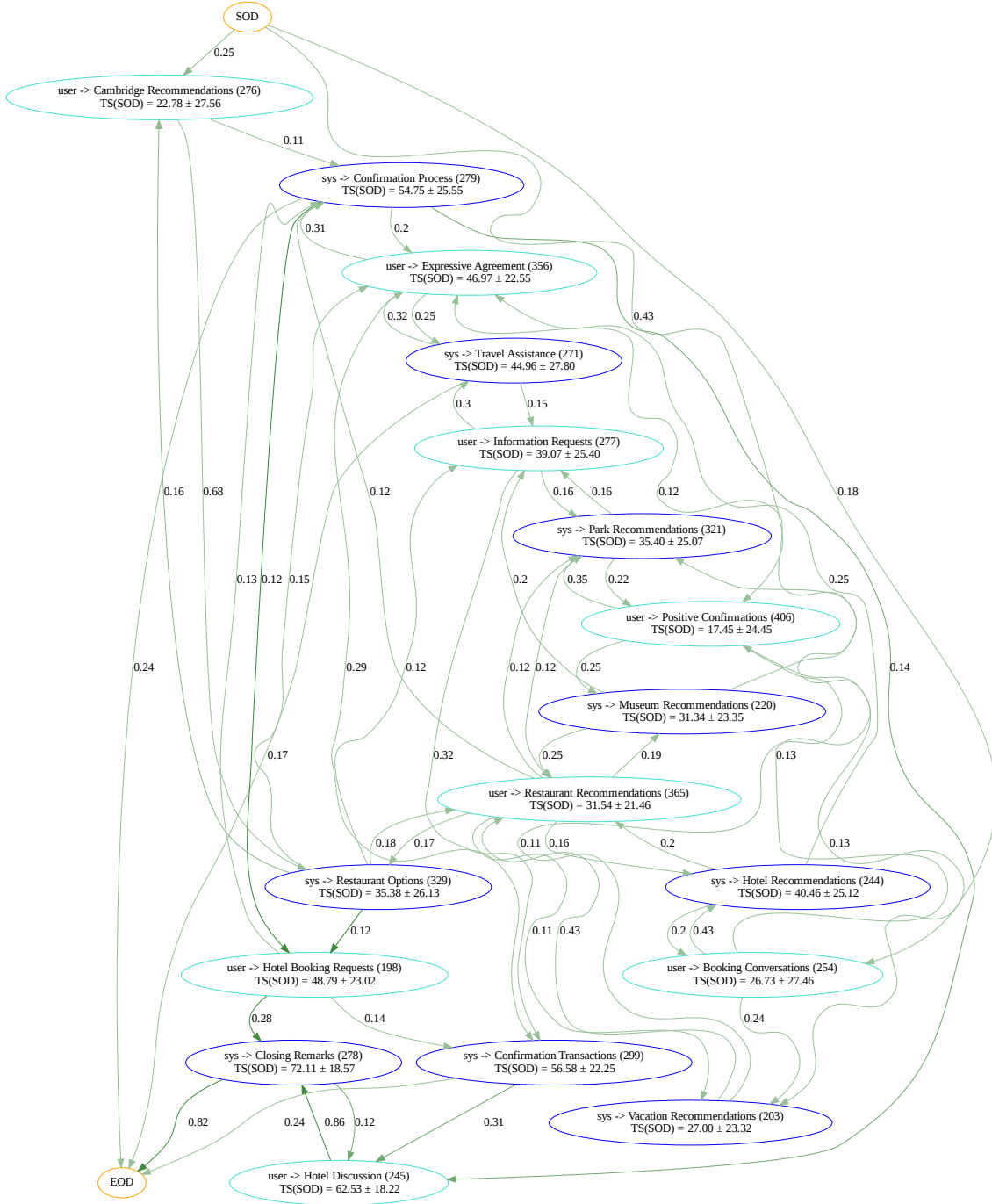


Figure 8: Dialogue flow discovered for Subset 1.

transitions.

Subset 9 (S9) consists of dialogues that maintain a consistently positive sentiment throughout. These dialogues are characterized by fast response times and many utterances. The clustering metrics for both the user and the system are detailed in Table 12. The discovered dialogue flow, illustrated in Figure 16, has $|C| = 9$ states, with an SS of

0.06 and an FSC of 0.25 for the user, and $|C| = 5$ states, with an SS of 0.07 and an FSC of 0.23 for the system, with a total of $|T| = 39$ transitions.

Subset 10 (S10) consists of dialogues with negative sentiment. These dialogues are characterized by fast response times and many utterances. The clustering metrics for both the user and the system are detailed in Table 13. The discovered dialogue

State	U_c	TU	TS(SOD)	TS(Prev)	S_c
User					
C0: Job Search Struggles	193	14.93 ± 4.36	22.94 ± 25.97	5.07 ± 5.45	0.47 ± 0.24
C1: Disgruntled Reactions	265	9.59 ± 4.05	65.36 ± 19.37	9.91 ± 4.09	0.49 ± 0.34
C2: Hotel Search	404	13.33 ± 4.15	25.31 ± 26.13	5.94 ± 5.78	0.46 ± 0.28
C3: Customer Service Frustrations	351	12.96 ± 4.47	35.79 ± 27.65	7.45 ± 5.44	0.37 ± 0.30
C4: Job Search Frustration	364	11.89 ± 3.73	37.85 ± 21.70	8.87 ± 4.66	0.46 ± 0.29
C5: Venting Frustrations	475	11.83 ± 4.89	43.69 ± 25.11	9.16 ± 4.82	0.41 ± 0.31
C6: Frustrated Conversations	287	10.78 ± 4.94	50.40 ± 21.56	9.48 ± 4.03	0.50 ± 0.25
C7: Restaurant Search	133	12.93 ± 3.68	18.36 ± 25.06	4.05 ± 5.35	0.58 ± 0.35
C8: Job Frustration	109	12.57 ± 5.26	7.44 ± 17.65	2.16 ± 4.43	0.16 ± 0.25
System					
C0: Restaurant Options	158	18.89 ± 5.49	38.72 ± 26.58	10.13 ± 4.06	0.40 ± 0.34
C1: Unhelpful Farewells	183	12.25 ± 5.26	78.47 ± 17.69	9.55 ± 4.40	0.68 ± 0.44
C2: Career Counseling	286	19.57 ± 7.10	55.15 ± 25.34	9.51 ± 4.24	0.42 ± 0.32
C3: Customer Service Conversations	439	18.13 ± 6.36	49.39 ± 26.73	9.30 ± 4.36	0.30 ± 0.29
C4: Hotel Discussions	444	20.27 ± 6.55	40.97 ± 26.84	9.76 ± 4.08	0.41 ± 0.29
C5: Job Search Support	343	20.24 ± 6.48	41.09 ± 26.46	9.73 ± 4.00	0.37 ± 0.30
C6: Sorry hear that	119	14.22 ± 4.72	20.13 ± 24.91	9.96 ± 3.92	0.04 ± 0.16
C7: Price Discussion	312	18.38 ± 6.37	42.84 ± 27.18	8.97 ± 4.28	0.40 ± 0.31
C8: Negative Reviews	280	15.07 ± 5.79	50.89 ± 24.93	9.65 ± 4.16	0.44 ± 0.30

Table 5: Clustering metrics for Subset 2.

State	U_c	TU	TS(SOD)	TS(Prev)	S_c
User					
C0: Requesting Contact Info	82	13.12 ± 3.31	44.87 ± 19.56	9.40 ± 3.90	0.72 ± 0.39
C1: Hotel Inquiry	405	15.16 ± 4.83	28.80 ± 27.00	6.34 ± 5.67	0.82 ± 0.34
C2: Disappointment Unfolds	354	12.88 ± 5.30	55.90 ± 23.99	9.41 ± 4.38	0.35 ± 0.42
C3: Job Search Conversations	186	14.59 ± 5.08	25.92 ± 25.84	5.98 ± 5.54	0.82 ± 0.33
C4: Price Inquiries	294	13.08 ± 3.87	39.87 ± 25.76	8.74 ± 4.89	0.68 ± 0.40
C5: Travel Arrangements	322	15.11 ± 4.76	46.42 ± 24.21	9.10 ± 4.70	0.64 ± 0.41
C6: Information Seeking	338	10.63 ± 5.10	54.00 ± 22.15	9.53 ± 4.29	0.53 ± 0.46
C7: Need Recommendations	213	13.95 ± 4.02	22.92 ± 27.22	5.40 ± 5.73	0.86 ± 0.33
C8: Travel Planning	373	14.09 ± 4.34	19.28 ± 23.81	5.06 ± 5.78	0.89 ± 0.29
System					
C0: Restaurant Recommendations	253	22.59 ± 7.79	39.96 ± 29.24	10.11 ± 4.15	0.61 ± 0.46
C1: Problems and Discussion	320	19.82 ± 7.80	69.74 ± 23.46	9.90 ± 4.38	0.15 ± 0.35
C2: Travel Recommendations	527	22.47 ± 6.70	38.40 ± 27.27	9.58 ± 4.08	0.62 ± 0.43
C3: Bad Recommendations	413	20.84 ± 7.37	53.06 ± 26.35	9.41 ± 4.17	0.36 ± 0.41
C4: Job Discussion	172	22.87 ± 7.57	37.89 ± 25.88	9.78 ± 4.08	0.61 ± 0.42
C5: Product Recommendations	201	19.67 ± 6.28	35.29 ± 26.31	10.08 ± 3.83	0.65 ± 0.42
C6: Price Discussion	178	21.93 ± 7.43	52.85 ± 24.29	9.74 ± 4.24	0.33 ± 0.36
C7: Providing Information	178	20.28 ± 6.99	51.84 ± 23.62	9.91 ± 3.95	0.33 ± 0.34
C8: Travel Planning	267	22.38 ± 7.32	39.20 ± 25.15	9.77 ± 4.02	0.61 ± 0.42

Table 6: Clustering metrics for Subset 3.

flow has $|C| = 9$ states, with an SS of 0.05 and an FSC of 0.29 for the user, and $|C| = 5$ states, with an SS of 0.06 and an FSC of 0.30 for the system. Figure 17 illustrates this flow for this subset, with a total of $|T| = 45$ transitions.

Subset 11 (S11) has dialogues with sentiments that start positive and end negative. These dialogues are characterized by fast response times and many utterances. The clustering metrics for both the user and the system are detailed in Table 14. The discovered dialogue flow, shown in Figure 18, consists of $|T| = 49$ transitions and $|C| = 10$ states, with an SS of 0.05 and an FSC of 0.31 for the user, and $|C| = 9$ states, with an SS of 0.05

and an FSC of 0.34 for the system.

Subset 12 (S12) has dialogues with sentiments that start negative and end positive. These dialogues are characterized by fast response times and many utterances. The clustering metrics for both the user and the system are detailed in Table 15. The discovered dialogue flow, shown in Figure 2, consists of $|T| = 44$ transitions and $|C| = 6$ states, with an SS of 0.06 and an FSC of 0.34 for the user, and $|C| = 8$ states, with an SS of 0.08 and an FSC of 0.30 for the system.

Subset 13 (S13) consists of dialogues with positive sentiment. These dialogues are characterized by slow response times and many utterances. The

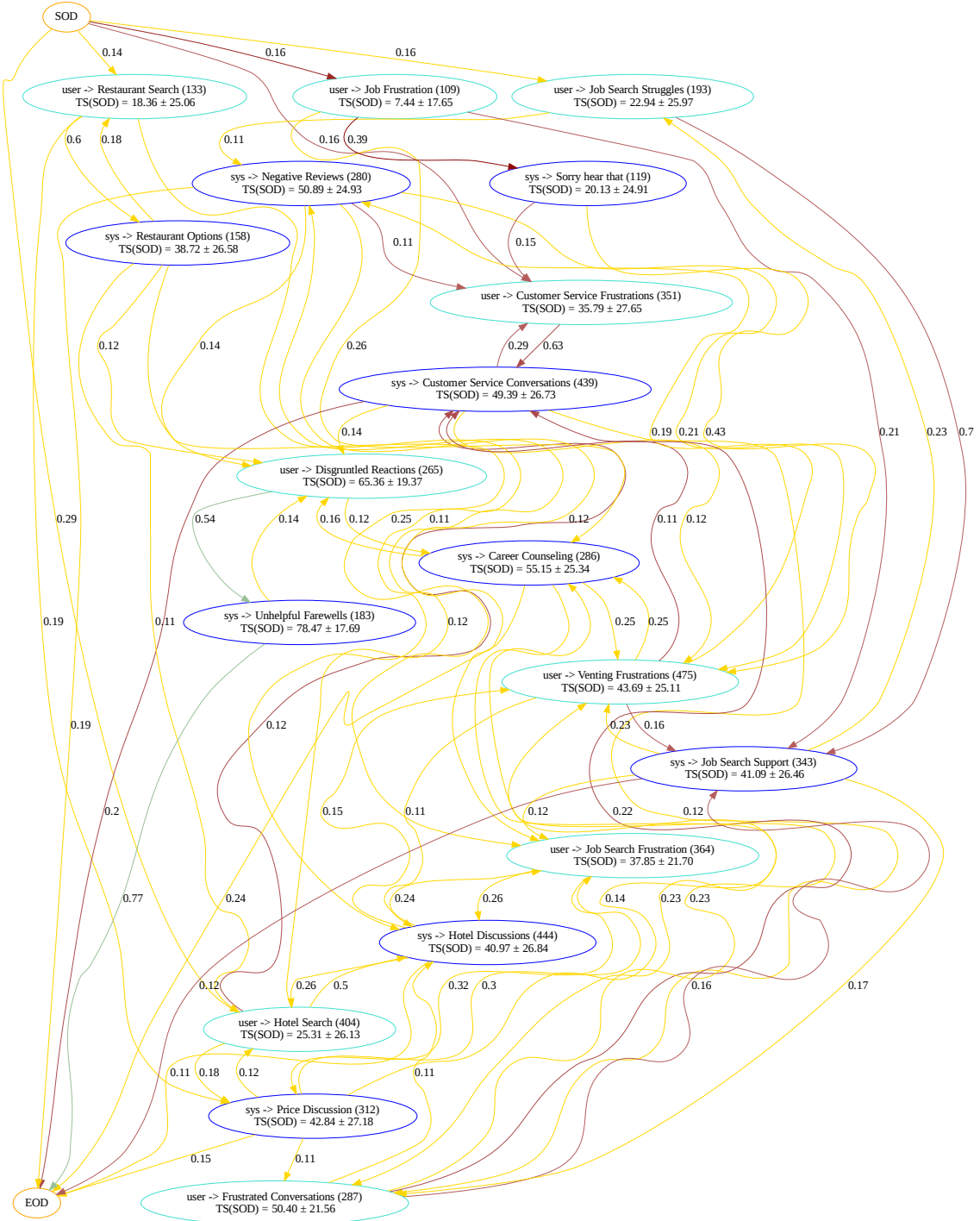


Figure 9: Dialogue flow discovered for Subset 2.

clustering metrics for both the user and the system are detailed in Table 16. The discovered dialogue flow has $|C| = 6$ states, with an SS of 0.05 and an FSC of 0.24 for the user, and $|C| = 8$ states, with an SS of 0.05 and an FSC of 0.22 for the system. Figure 19 illustrates the discovered dialogue flow for this subset, with a total of $|T| = 50$ transitions.

Subset 14 (S14) consists of dialogues with negative sentiment. These dialogues are characterized by slow response times and many utterances. The clustering metrics for both the user and the system are detailed in Table 17. The discovered dialogue flow has $|C| = 6$ states, with an SS of 0.05 and an FSC of 0.30 for the user, and $|C| = 8$ states, with an SS of 0.06 and an FSC of 0.32 for the system.

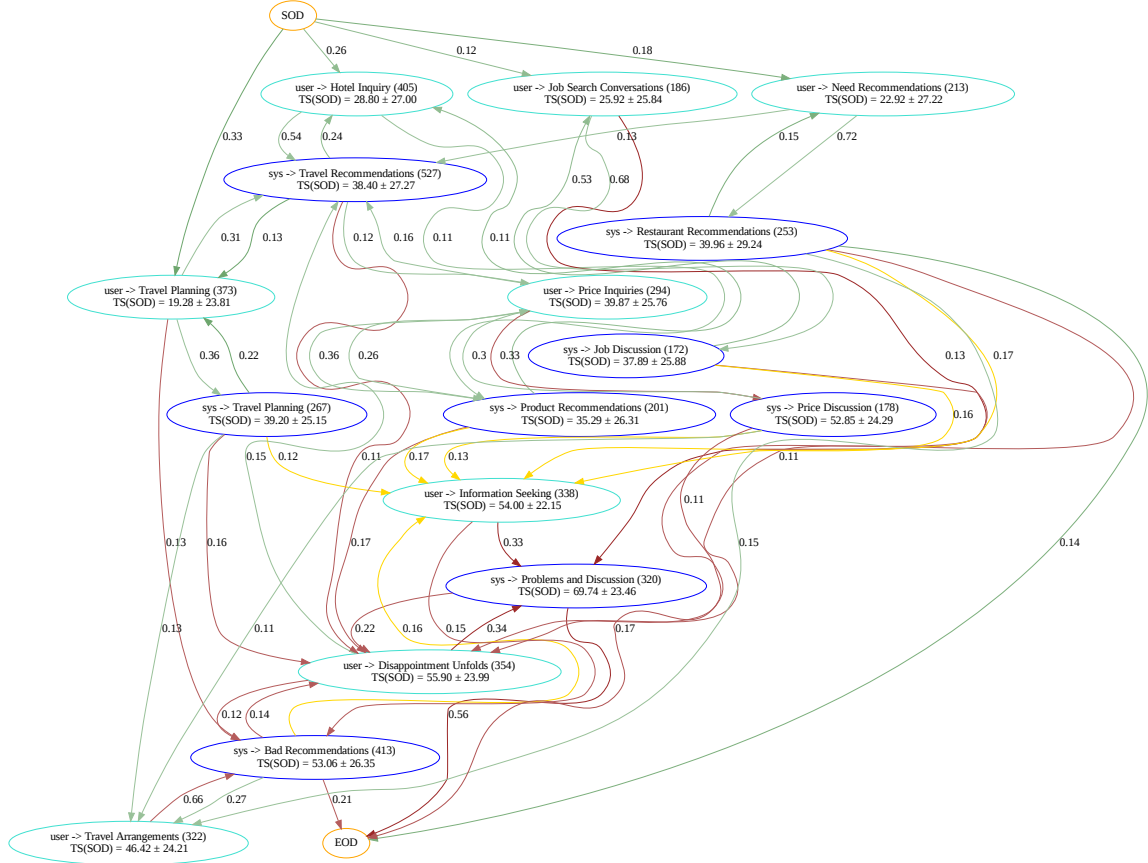


Figure 10: Dialogue flow discovered for Subset 3.

State	U_c	TU	TS(SOD)	TS(Prev)	S_c
User					
C0: Travel Arrangements	575	14.53 $\pm$ 4.56	28.42 $\pm$ 27.81	6.57 $\pm$ 5.79	0.37 $\pm$ 0.45
C1: Job Search Conversations	287	14.54 $\pm$ 4.33	20.05 $\pm$ 24.18	4.91 $\pm$ 5.62	0.23 $\pm$ 0.36
C2: Emotional Support	473	15.00 $\pm$ 4.57	25.77 $\pm$ 27.76	5.93 $\pm$ 5.68	0.25 $\pm$ 0.39
C3: Booking Confirmation	639	10.96 $\pm$ 4.41	60.00 $\pm$ 19.96	9.81 $\pm$ 4.09	0.85 $\pm$ 0.27
C4: Travel Planning	553	12.42 $\pm$ 4.47	39.81 $\pm$ 22.33	9.11 $\pm$ 4.35	0.48 $\pm$ 0.42
System					
C0: Express Gratitude	260	18.12 $\pm$ 6.72	79.47 $\pm$ 17.40	10.15 $\pm$ 4.13	1.00 $\pm$ 0.06
C1: Restaurant Recommendations	209	22.05 $\pm$ 7.31	37.98 $\pm$ 25.53	9.81 $\pm$ 4.06	0.59 $\pm$ 0.44
C2: Booking Confirmations	286	21.44 $\pm$ 6.78	61.76 $\pm$ 24.26	9.87 $\pm$ 4.29	0.78 $\pm$ 0.33
C3: Options Discussion	221	19.91 $\pm$ 7.05	37.47 $\pm$ 26.84	9.79 $\pm$ 3.92	0.52 $\pm$ 0.43
C4: Hotel Options	238	23.79 $\pm$ 8.22	42.24 $\pm$ 25.26	9.44 $\pm$ 4.05	0.58 $\pm$ 0.40
C5: Sorry to Hear	155	12.68 $\pm$ 3.33	11.15 $\pm$ 5.32	10.61 $\pm$ 3.53	0.01 $\pm$ 0.06
C6: Empathetic Guidance	226	21.41 $\pm$ 6.80	44.47 $\pm$ 24.57	9.90 $\pm$ 4.12	0.57 $\pm$ 0.40
C7: Travel Recommendations	306	21.50 $\pm$ 7.11	43.72 $\pm$ 25.08	9.68 $\pm$ 4.24	0.64 $\pm$ 0.40
C8: Customer Support	205	19.69 $\pm$ 6.47	55.63 $\pm$ 27.38	9.54 $\pm$ 4.03	0.73 $\pm$ 0.37
C9: Job Search Support	367	22.63 $\pm$ 7.46	39.87 $\pm$ 25.19	9.83 $\pm$ 3.99	0.55 $\pm$ 0.41

Table 7: Clustering metrics for Subset 4.

Figure 20 illustrates the discovered dialogue flow for this subset, with a total of $|T| = 51$ transitions.

Subset 15 (S15) has dialogues with sentiments that start positive and end negative. These dialogues are characterized by slow response times and many utterances. The clustering metrics for both the user and the system are detailed in Ta-

ble 18. The discovered dialogue flow has $|C| = 10$ states, with an SS of 0.05 and an FSC of 0.33 for the user, and $|C| = 8$ states, with an SS of 0.05 and an FSC of 0.35 for the system. Figure 21 illustrates this flow for this subset, with a total of $|T| = 51$ transitions.

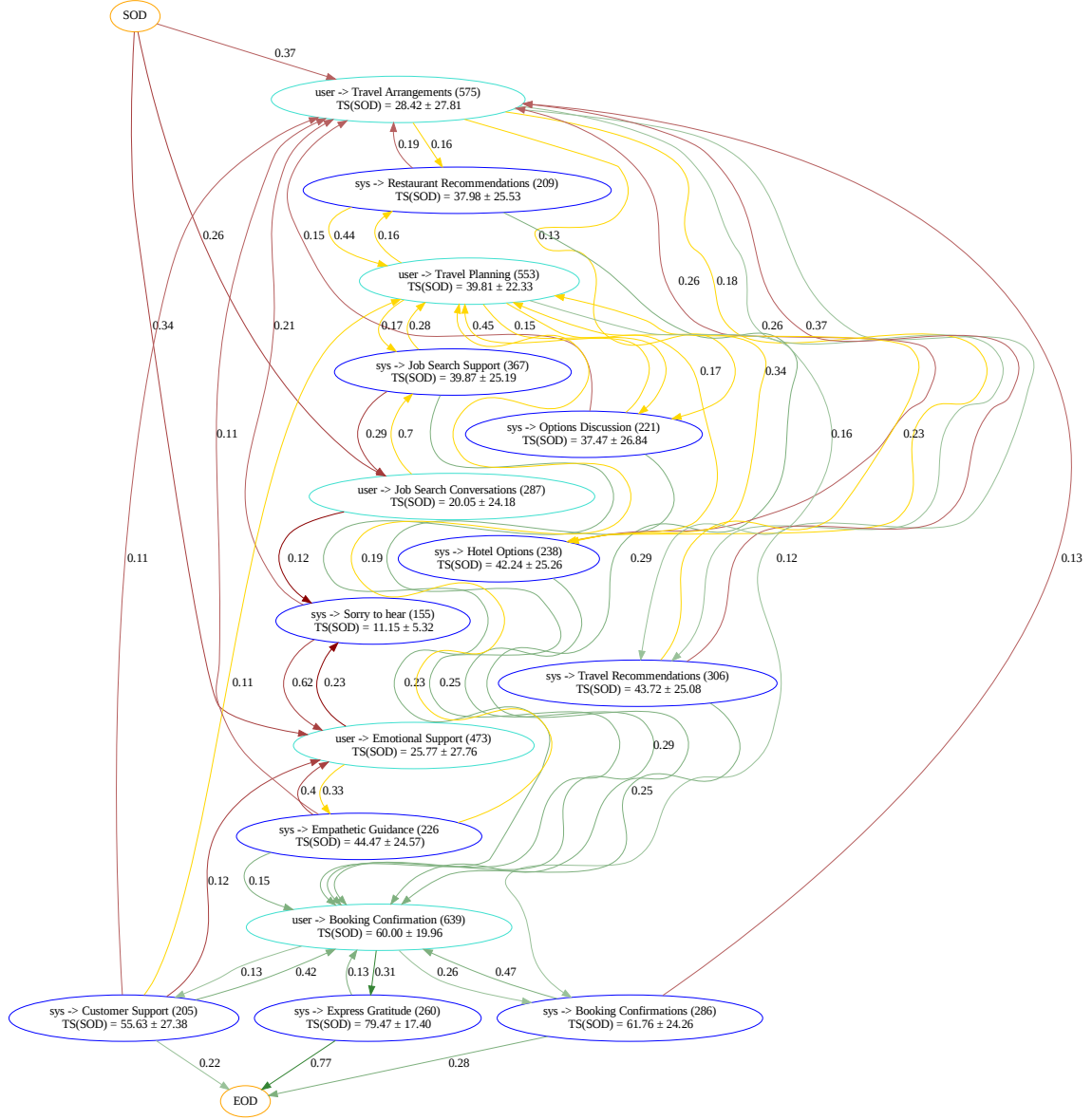


Figure 11: Dialogue flow discovered for Subset 4.

Subset 16 (S16) has dialogues with sentiments that start negative and end positive. These dialogues are characterized by slow response times and many utterances. The clustering metrics for both the user and the system are detailed in Table 19. The discovered dialogue flow has $|C| = 6$ states, with an SS of 0.05 and an FSC of 0.32 for the user, and $|C| = 9$ states, with an SS of 0.08 and an FSC of 0.27 for the system. Figure 22 illustrates the discovered dialogue flow for this subset, with a total of $|T| = 50$ transitions.

State	U_c	TU	TS(SOD)	TS(Prev)	S_c
User					
C0: Travel Planning	515	13.85 $\pm$ 4.34	264.43 $\pm$ 303.31	66.24 $\pm$ 70.17	0.74 $\pm$ 0.25
C1: Price Inquiry	143	12.45 $\pm$ 4.00	505.45 $\pm$ 247.40	114.67 $\pm$ 50.31	0.73 $\pm$ 0.26
C2: Booking Conversations	363	14.47 $\pm$ 5.05	568.89 $\pm$ 275.64	115.63 $\pm$ 52.81	0.75 $\pm$ 0.25
C3: Request Clarification	352	8.77 $\pm$ 4.20	477.32 $\pm$ 275.75	108.27 $\pm$ 51.81	0.69 $\pm$ 0.24
C4: Gift Planning	326	12.37 $\pm$ 4.63	311.28 $\pm$ 272.54	85.55 $\pm$ 65.68	0.73 $\pm$ 0.25
C5: Restaurant Recommendations	297	13.81 $\pm$ 4.18	235.86 $\pm$ 344.68	43.57 $\pm$ 62.87	0.81 $\pm$ 0.24
C6: Closing Thank You	308	8.82 $\pm$ 2.97	781.81 $\pm$ 231.62	112.71 $\pm$ 51.87	0.90 $\pm$ 0.20
C7: Information Requests	208	13.45 $\pm$ 3.72	536.84 $\pm$ 302.74	108.19 $\pm$ 55.47	0.79 $\pm$ 0.25
System					
C0: Museum Recommendations	208	24.00 $\pm$ 7.26	435.31 $\pm$ 295.29	122.91 $\pm$ 47.41	0.75 $\pm$ 0.26
C1: Restaurant Recommendations	324	22.42 $\pm$ 7.07	449.51 $\pm$ 319.74	119.40 $\pm$ 47.19	0.80 $\pm$ 0.25
C2: Recommendations	284	19.10 $\pm$ 6.58	365.64 $\pm$ 289.15	113.09 $\pm$ 45.90	0.70 $\pm$ 0.25
C3: Travel Planning	228	18.36 $\pm$ 7.52	517.15 $\pm$ 323.06	113.26 $\pm$ 47.55	0.63 $\pm$ 0.24
C4: Park Options	325	24.52 $\pm$ 7.12	450.62 $\pm$ 317.61	116.48 $\pm$ 49.78	0.79 $\pm$ 0.25
C5: Providing Information	214	21.01 $\pm$ 7.40	661.62 $\pm$ 303.37	112.03 $\pm$ 48.63	0.60 $\pm$ 0.20
C6: Hotel Recommendations	295	24.17 $\pm$ 8.34	455.44 $\pm$ 305.75	115.37 $\pm$ 50.04	0.75 $\pm$ 0.26
C7: Confirmation Process	347	22.46 $\pm$ 5.87	742.48 $\pm$ 255.78	117.05 $\pm$ 50.09	0.72 $\pm$ 0.26
C8: Closing Remarks	261	11.95 $\pm$ 4.25	906.35 $\pm$ 222.80	110.76 $\pm$ 50.63	0.98 $\pm$ 0.11

Table 8: Clustering metrics for Subset 5.

State	U_c	TU	TS(SOD)	TS(Prev)	S_c
User					
C0: Problems and Frustrations	239	13.98 $\pm$ 4.68	184.12 $\pm$ 271.08	49.89 $\pm$ 68.59	0.38 $\pm$ 0.27
C1: Frustrated Traveler Conversations	1071	11.34 $\pm$ 5.06	565.75 $\pm$ 318.13	108.01 $\pm$ 56.46	0.39 $\pm$ 0.27
C2: Travel Planning	674	13.78 $\pm$ 4.43	328.45 $\pm$ 337.97	70.50 $\pm$ 69.38	0.51 $\pm$ 0.29
C3: Seeking Options	573	11.98 $\pm$ 4.18	510.38 $\pm$ 293.35	108.17 $\pm$ 55.22	0.49 $\pm$ 0.30
System					
C0: Unfortunately and Sorry	390	16.81 $\pm$ 6.54	477.34 $\pm$ 351.03	119.91 $\pm$ 45.94	0.24 $\pm$ 0.29
C1: Customer Support	665	16.73 $\pm$ 6.39	733.07 $\pm$ 324.55	112.87 $\pm$ 50.93	0.44 $\pm$ 0.38
C2: What happened	422	18.75 $\pm$ 6.02	497.43 $\pm$ 311.83	113.40 $\pm$ 47.05	0.40 $\pm$ 0.31
C3: Travel Recommendations	1051	20.95 $\pm$ 7.08	526.01 $\pm$ 323.38	117.92 $\pm$ 49.10	0.42 $\pm$ 0.32

Table 9: Clustering metrics for Subset 6.

State	U_c	TU	TS(SOD)	TS(Prev)	S_c
User					
C0: Travel Planning	865	15.16 $\pm$ 4.58	252.65 $\pm$ 309.22	60.70 $\pm$ 68.39	0.85 $\pm$ 0.32
C1: Product Inquiry	498	13.62 $\pm$ 4.45	477.71 $\pm$ 301.23	109.54 $\pm$ 54.35	0.66 $\pm$ 0.41
C2: Job Search Conversations	171	15.22 $\pm$ 4.45	332.38 $\pm$ 335.21	76.28 $\pm$ 70.85	0.79 $\pm$ 0.36
C3: Travel Arrangements	567	14.41 $\pm$ 5.29	563.05 $\pm$ 262.40	113.62 $\pm$ 52.23	0.65 $\pm$ 0.43
C4: Conversational Chaos	446	12.85 $\pm$ 5.98	681.50 $\pm$ 287.44	115.79 $\pm$ 51.80	0.31 $\pm$ 0.39
System					
C0: Apology Conversations	370	19.88 $\pm$ 7.95	807.86 $\pm$ 285.47	115.57 $\pm$ 47.68	0.15 $\pm$ 0.32
C1: Job Discussion	182	21.75 $\pm$ 7.32	445.66 $\pm$ 298.39	118.55 $\pm$ 47.54	0.64 $\pm$ 0.41
C2: Budget Discussion	330	20.68 $\pm$ 7.12	487.59 $\pm$ 327.39	110.84 $\pm$ 52.07	0.54 $\pm$ 0.43
C3: Booking Conversations	565	22.15 $\pm$ 7.52	647.66 $\pm$ 313.77	118.49 $\pm$ 50.10	0.31 $\pm$ 0.38
C4: Restaurant Recommendations	345	22.70 $\pm$ 6.61	484.53 $\pm$ 324.78	114.57 $\pm$ 49.44	0.61 $\pm$ 0.44
C5: Travel Recommendations	683	22.88 $\pm$ 7.47	422.80 $\pm$ 312.80	112.85 $\pm$ 50.02	0.65 $\pm$ 0.43

Table 10: Clustering metrics for Subset 7.

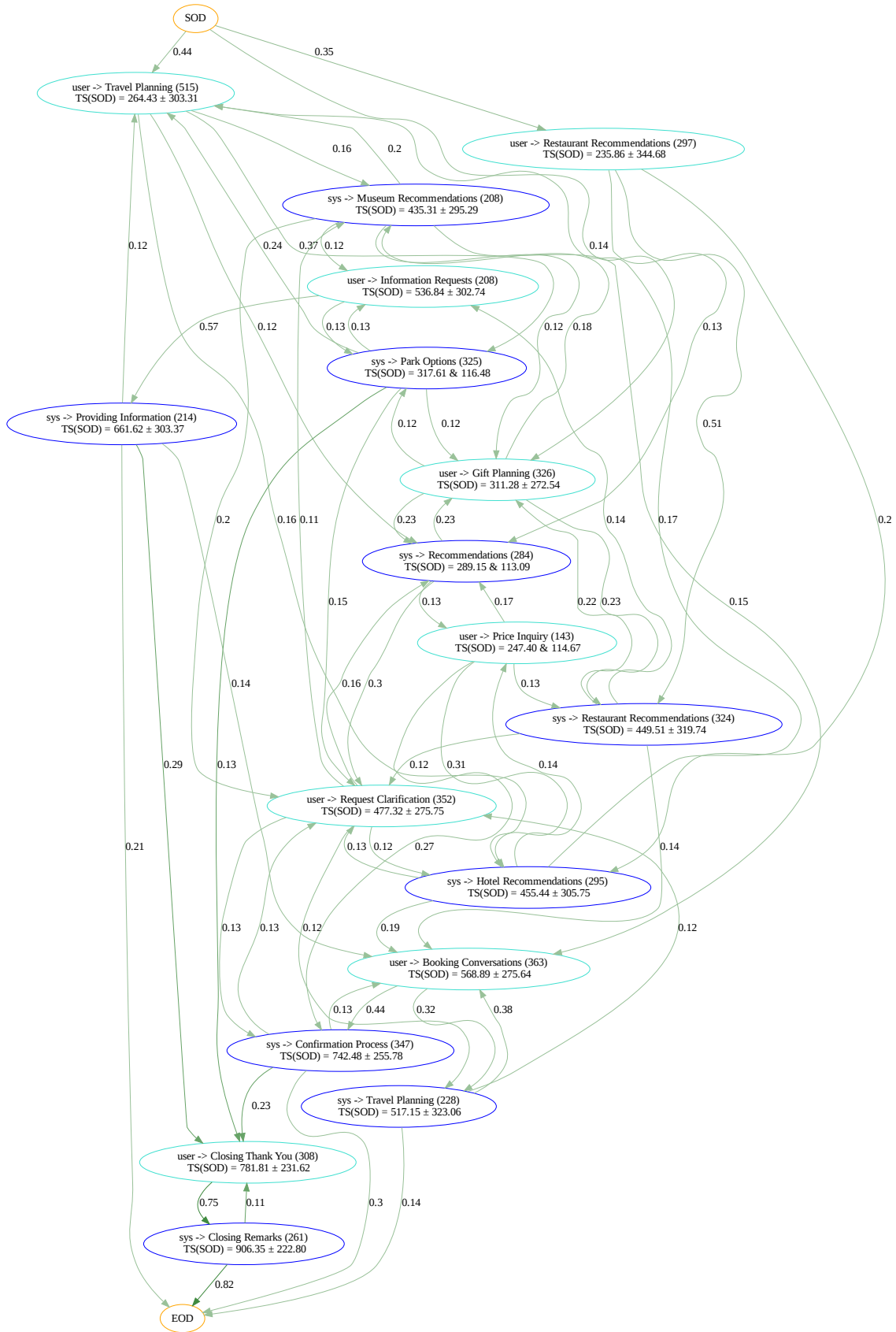


Figure 12: Dialogue flow discovered for Subset 5.

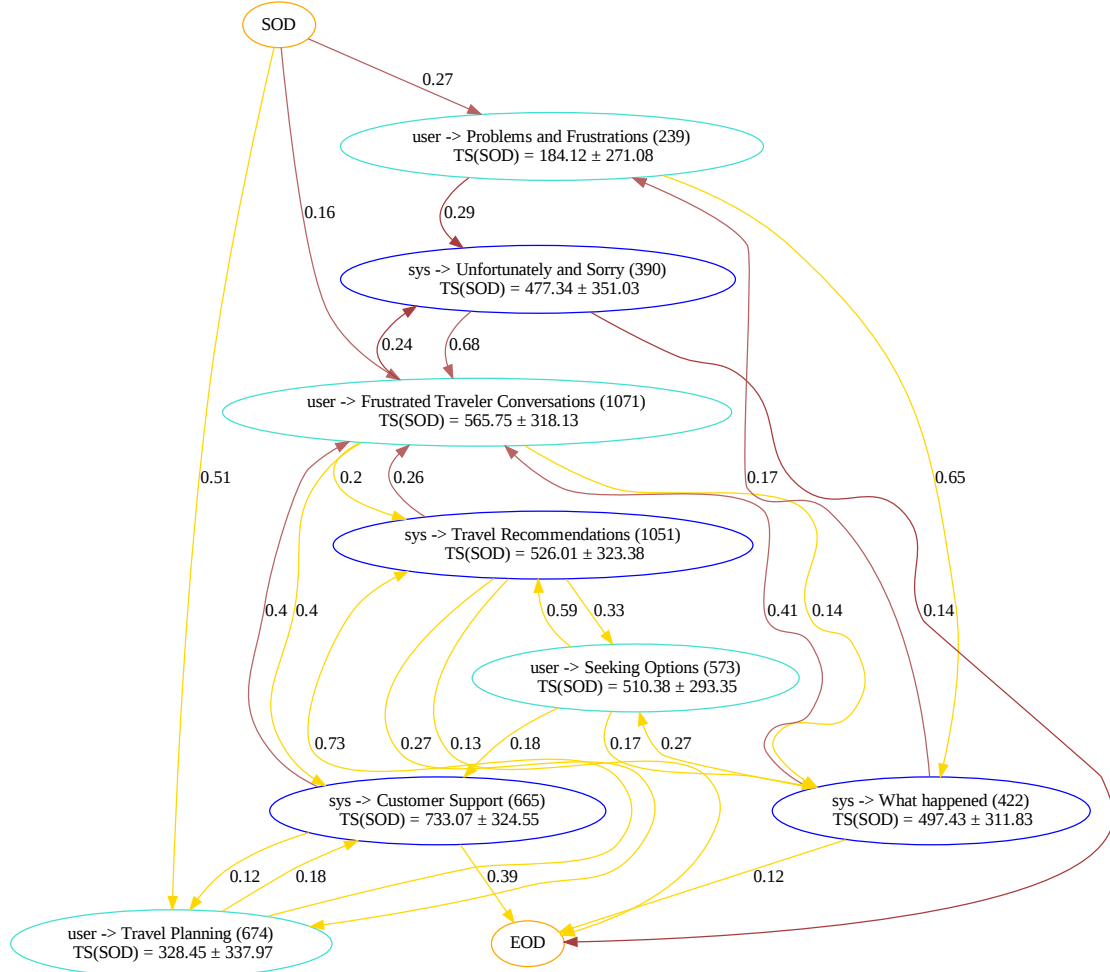


Figure 13: Dialogue flow discovered for Subset 6.

State	U_c	TU	TS(SOD)	TS(Prev)	S_c
User					
C0: Job Discussion	370	13.02 ± 4.40	481.92 ± 281.28	111.90 ± 53.10	0.50 ± 0.44
C1: Venting Session	560	13.59 ± 5.66	455.12 ± 284.92	106.12 ± 55.21	0.40 ± 0.40
C2: Job Search Struggles	257	15.24 ± 4.04	131.47 ± 267.35	26.66 ± 52.08	0.14 ± 0.34
C3: Feeling Frustrated	214	15.57 ± 5.03	206.46 ± 298.71	46.64 ± 63.54	0.21 ± 0.36
C4: Budget Discussion	326	10.53 ± 3.83	743.62 ± 233.93	113.85 ± 49.28	0.90 ± 0.26
C5: Expressions of Misery	178	14.16 ± 4.62	426.05 ± 254.37	115.02 ± 53.42	0.43 ± 0.41
C6: Booking Requests	300	14.89 ± 4.87	494.87 ± 307.11	103.36 ± 58.72	0.55 ± 0.41
C7: Accommodation Struggles	116	15.16 ± 2.77	216.10 ± 165.70	0.93 ± 10.06	0.01 ± 0.09
C8: Travel Assistance	194	14.85 ± 4.58	691.90 ± 250.39	114.03 ± 50.33	0.78 ± 0.31
System					
C0: Job Search Support	388	22.98 ± 7.86	521.96 ± 337.62	118.05 ± 49.45	0.57 ± 0.41
C1: Travel Information	554	21.92 ± 7.42	688.54 ± 299.45	114.00 ± 50.53	0.77 ± 0.34
C2: Hotel Recommendations	866	22.80 ± 6.98	470.60 ± 299.77	116.54 ± 48.49	0.61 ± 0.41
C3: Condolences and Sorry	400	18.72 ± 7.37	343.34 ± 287.60	119.63 ± 48.13	0.33 ± 0.40
C4: Closing Remarks	254	18.25 ± 6.63	936.62 ± 205.91	117.07 ± 49.49	1.00 ± 0.04

Table 11: Clustering metrics for Subset 8.

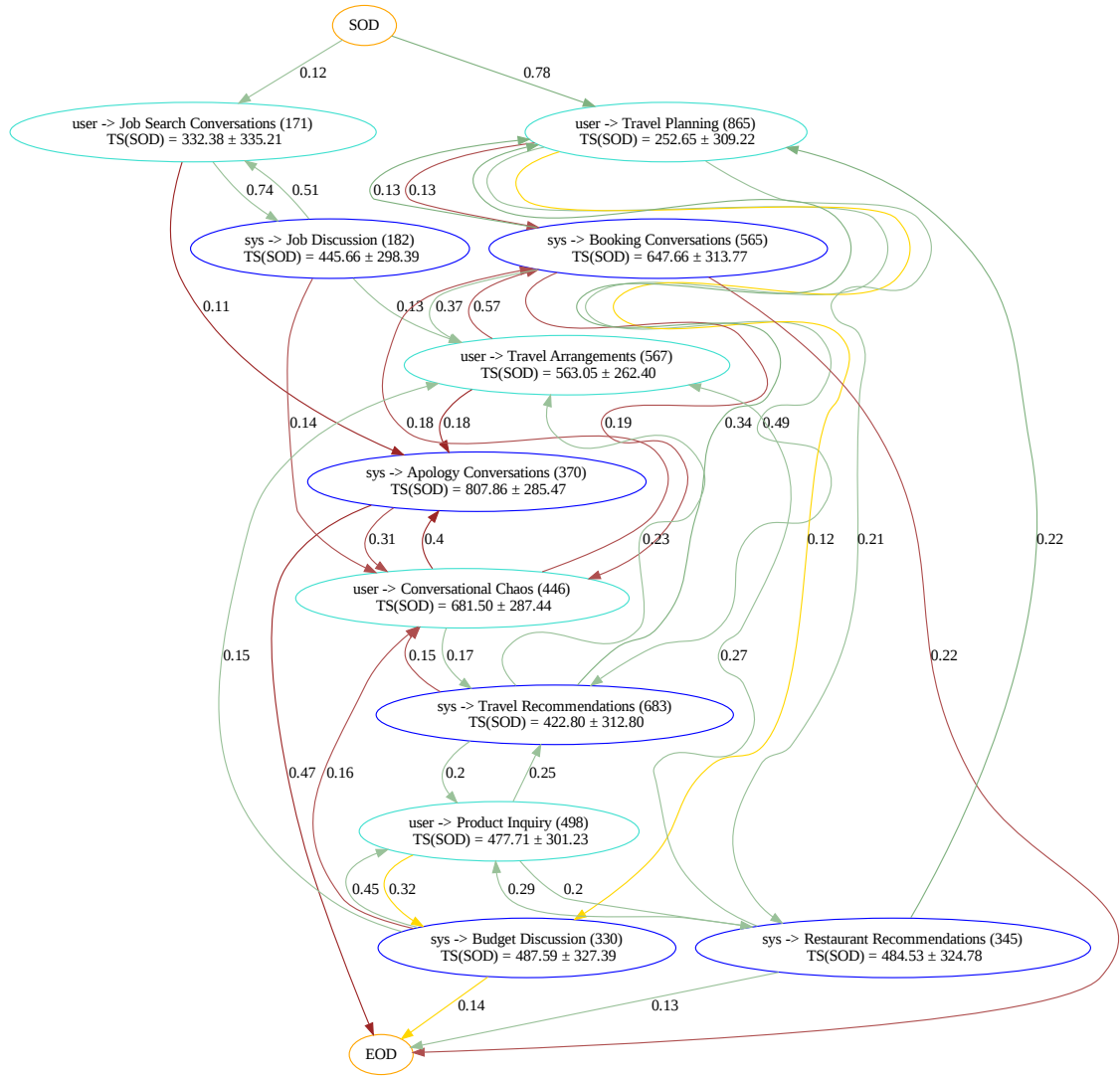


Figure 14: Dialogue flow discovered for Subset 7.

State	U_c	TU	TS(SOD)	TS(Prev)	S_c
User					
C0: Closing Remarks	213	10.28 ± 2.47	162.19 ± 57.49	9.86 ± 4.10	0.77 ± 0.25
C1: Farewell Appreciation	255	6.03 ± 3.53	170.09 ± 55.16	9.45 ± 4.17	0.68 ± 0.24
C2: Expressive Agreement	243	5.16 ± 1.98	107.03 ± 64.06	10.12 ± 4.05	0.72 ± 0.25
C3: Booking Conversations	905	13.39 ± 4.76	100.35 ± 60.08	9.68 ± 4.17	0.74 ± 0.25
C4: Travel Planning	1126	14.52 ± 4.38	71.06 ± 69.43	7.30 ± 5.53	0.78 ± 0.25
C5: Food Recommendations	558	14.30 ± 4.17	81.29 ± 66.53	7.56 ± 5.42	0.80 ± 0.25
C6: Travel Arrangements	583	13.78 ± 4.13	99.40 ± 65.20	8.95 ± 4.73	0.76 ± 0.25
C7: Price Inquiries	318	12.88 ± 3.83	90.18 ± 62.77	9.84 ± 4.05	0.72 ± 0.25
C8: Expressions of Gratitude	782	9.00 ± 3.36	158.28 ± 61.09	9.63 ± 4.23	0.88 ± 0.22
C9: Event Planning	891	12.15 ± 4.80	84.13 ± 66.84	9.03 ± 4.90	0.74 ± 0.25
System					
C0: Hotel Options	986	20.76 ± 7.39	81.50 ± 61.48	9.64 ± 4.15	0.68 ± 0.24
C1: Tourist Information	1372	22.91 ± 7.42	90.56 ± 68.91	9.79 ± 4.05	0.77 ± 0.25
C2: Booking Confirmations	1635	19.55 ± 7.13	113.88 ± 64.16	9.67 ± 4.16	0.64 ± 0.24
C3: Express Gratitude	1118	13.10 ± 6.06	177.15 ± 55.83	9.67 ± 4.20	0.96 ± 0.14
C4: Restaurant Recommendations	742	23.22 ± 7.38	93.03 ± 64.94	9.69 ± 4.04	0.77 ± 0.26

Table 12: Clustering metrics for Subset 9.

State	U_c	TU	TS(SOD)	TS(Prev)	S_c
User					
C0: Travel Arrangements	646	13.61 $\pm$ 4.96	114.24 $\pm$ 75.13	8.77 $\pm$ 4.65	0.51 $\pm$ 0.26
C1: Disheartened Goodbyes	354	7.84 $\pm$ 4.43	188.51 $\pm$ 58.06	9.73 $\pm$ 4.10	0.44 $\pm$ 0.32
C2: Closing Thank You	605	10.12 $\pm$ 3.78	176.83 $\pm$ 64.73	9.66 $\pm$ 4.07	0.60 $\pm$ 0.34
C3: Job Search Frustration	498	14.63 $\pm$ 5.09	62.38 $\pm$ 64.40	6.51 $\pm$ 5.70	0.38 $\pm$ 0.29
C4: Agreement Negotiation	620	8.70 $\pm$ 3.97	132.00 $\pm$ 70.98	9.91 $\pm$ 4.19	0.46 $\pm$ 0.21
C5: Requesting Information	358	12.67 $\pm$ 4.74	118.03 $\pm$ 72.89	9.56 $\pm$ 4.33	0.54 $\pm$ 0.27
C6: Price Discussion	397	11.96 $\pm$ 4.45	96.51 $\pm$ 63.44	9.55 $\pm$ 4.22	0.52 $\pm$ 0.23
C7: Travel Troubles	945	13.79 $\pm$ 4.36	90.25 $\pm$ 76.84	7.44 $\pm$ 5.26	0.56 $\pm$ 0.32
C8: Venting Frustrations	1280	11.95 $\pm$ 4.89	97.49 $\pm$ 71.15	8.92 $\pm$ 4.86	0.33 $\pm$ 0.31
C9: Seeking Alternatives	759	12.20 $\pm$ 4.20	108.59 $\pm$ 65.99	9.59 $\pm$ 4.12	0.47 $\pm$ 0.34
System					
C0: Closing Remarks	690	12.80 $\pm$ 5.82	194.12 $\pm$ 59.73	9.64 $\pm$ 4.12	0.87 $\pm$ 0.29
C1: Travel Assistance	1345	17.92 $\pm$ 6.69	127.38 $\pm$ 72.01	9.83 $\pm$ 4.08	0.45 $\pm$ 0.26
C2: Travel Recommendations	1992	19.86 $\pm$ 6.74	103.19 $\pm$ 69.97	9.76 $\pm$ 4.17	0.47 $\pm$ 0.31
C3: Sorry hear that	1494	15.50 $\pm$ 6.85	127.15 $\pm$ 79.58	9.70 $\pm$ 4.13	0.36 $\pm$ 0.34
C4: Job Search Guidance	892	19.83 $\pm$ 7.07	94.29 $\pm$ 66.27	9.75 $\pm$ 4.02	0.41 $\pm$ 0.30

Table 13: Clustering metrics for Subset 10.

State	U_c	TU	TS(SOD)	TS(Prev)	S_c
User					
C0: Encouraging Agreement	197	5.38 $\pm$ 2.15	128.87 $\pm$ 74.40	9.32 $\pm$ 4.04	0.52 $\pm$ 0.25
C1: Booking Conversations	1300	15.39 $\pm$ 4.55	89.31 $\pm$ 82.70	7.69 $\pm$ 5.43	0.80 $\pm$ 0.28
C2: Budget Discussion	670	14.90 $\pm$ 4.13	122.51 $\pm$ 67.46	9.45 $\pm$ 4.52	0.74 $\pm$ 0.31
C3: Job Search Conversations	539	13.32 $\pm$ 3.86	108.46 $\pm$ 70.49	9.54 $\pm$ 4.20	0.70 $\pm$ 0.30
C4: Restaurant Recommendations	364	15.16 $\pm$ 4.55	100.96 $\pm$ 83.25	8.56 $\pm$ 5.20	0.74 $\pm$ 0.30
C5: Expressions of Gratitude	504	15.02 $\pm$ 4.41	88.21 $\pm$ 73.14	7.98 $\pm$ 5.35	0.78 $\pm$ 0.32
C6: Time Management	812	12.82 $\pm$ 4.61	102.81 $\pm$ 81.24	9.18 $\pm$ 4.55	0.78 $\pm$ 0.31
C7: Frustrated Customer Service	1117	9.61 $\pm$ 3.86	177.03 $\pm$ 67.48	9.80 $\pm$ 4.16	0.58 $\pm$ 0.44
C8: Travel Planning	570	13.57 $\pm$ 4.84	108.49 $\pm$ 78.98	9.19 $\pm$ 4.66	0.67 $\pm$ 0.31
C9: Recommendation Discussion	950	11.92 $\pm$ 5.62	177.58 $\pm$ 82.82	9.51 $\pm$ 4.18	0.38 $\pm$ 0.33
System					
C0: Confirmation Process	621	19.94 $\pm$ 7.63	140.20 $\pm$ 69.88	9.50 $\pm$ 4.08	0.49 $\pm$ 0.32
C1: Price Quotes	534	19.84 $\pm$ 7.84	116.78 $\pm$ 75.78	9.82 $\pm$ 4.02	0.57 $\pm$ 0.27
C2: Scheduling Confirmed	907	20.29 $\pm$ 7.41	121.67 $\pm$ 77.12	9.60 $\pm$ 4.19	0.55 $\pm$ 0.33
C3: Restaurant Recommendations	682	22.58 $\pm$ 7.36	103.06 $\pm$ 72.07	9.89 $\pm$ 3.98	0.71 $\pm$ 0.33
C4: Closing Remarks	1005	13.95 $\pm$ 5.79	193.34 $\pm$ 64.68	9.70 $\pm$ 4.01	0.59 $\pm$ 0.48
C5: Hotel Recommendations	713	24.34 $\pm$ 8.01	98.99 $\pm$ 72.77	9.80 $\pm$ 4.08	0.65 $\pm$ 0.32
C6: Travel Recommendations	1087	22.47 $\pm$ 7.43	108.52 $\pm$ 82.04	9.72 $\pm$ 4.12	0.67 $\pm$ 0.34
C7: Apology Conversations	581	18.66 $\pm$ 9.19	211.53 $\pm$ 71.48	9.69 $\pm$ 4.06	0.21 $\pm$ 0.31
C8: Recommendations Provided	823	20.49 $\pm$ 7.82	108.07 $\pm$ 82.23	9.88 $\pm$ 4.18	0.68 $\pm$ 0.33

Table 14: Clustering metrics for Subset 11.

State	U_c	TU	TS(SOD)	TS(Prev)	S_c
User					
C0: Job Search Assistance	461	16.19 $\pm$ 5.00	59.13 $\pm$ 55.45	7.19 $\pm$ 5.61	0.37 $\pm$ 0.37
C1: Expressions of Gratitude	1009	9.74 $\pm$ 3.97	161.63 $\pm$ 50.16	9.90 $\pm$ 4.15	0.90 $\pm$ 0.23
C2: Booking Confirmation	1202	11.62 $\pm$ 5.35	109.60 $\pm$ 55.95	9.81 $\pm$ 4.20	0.63 $\pm$ 0.34
C3: Discussion Options	960	12.65 $\pm$ 4.27	81.92 $\pm$ 55.66	9.75 $\pm$ 4.19	0.53 $\pm$ 0.36
C4: Travel Planning	1219	15.42 $\pm$ 4.65	77.87 $\pm$ 60.99	8.03 $\pm$ 5.24	0.52 $\pm$ 0.40
C5: Emotional Venting	738	14.31 $\pm$ 4.98	58.17 $\pm$ 62.27	6.98 $\pm$ 5.38	0.28 $\pm$ 0.35
System					
C0: Empathetic Support	867	20.07 $\pm$ 8.02	115.25 $\pm$ 60.06	9.67 $\pm$ 4.02	0.64 $\pm$ 0.30
C1: Restaurant Recommendations	526	18.88 $\pm$ 7.95	50.14 $\pm$ 55.83	9.31 $\pm$ 4.24	0.22 $\pm$ 0.35
C2: Expressions of Appreciation	526	21.97 $\pm$ 7.00	92.68 $\pm$ 62.01	9.56 $\pm$ 4.11	0.63 $\pm$ 0.37
C3: Thanks and Greetings	992	16.98 $\pm$ 7.54	170.35 $\pm$ 49.19	9.66 $\pm$ 4.12	0.98 $\pm$ 0.13
C4: Job Search Guidance	662	22.11 $\pm$ 8.04	80.05 $\pm$ 54.76	9.77 $\pm$ 4.10	0.49 $\pm$ 0.31
C5: Tourist Information	694	23.36 $\pm$ 8.06	82.14 $\pm$ 54.62	9.75 $\pm$ 4.15	0.55 $\pm$ 0.35
C6: Travel Arrangements	750	22.23 $\pm$ 7.34	90.89 $\pm$ 57.30	9.68 $\pm$ 4.08	0.62 $\pm$ 0.33
C7: Information Exchange	532	19.43 $\pm$ 7.08	116.73 $\pm$ 59.46	9.68 $\pm$ 4.01	0.60 $\pm$ 0.28

Table 15: Clustering metrics for Subset 12.

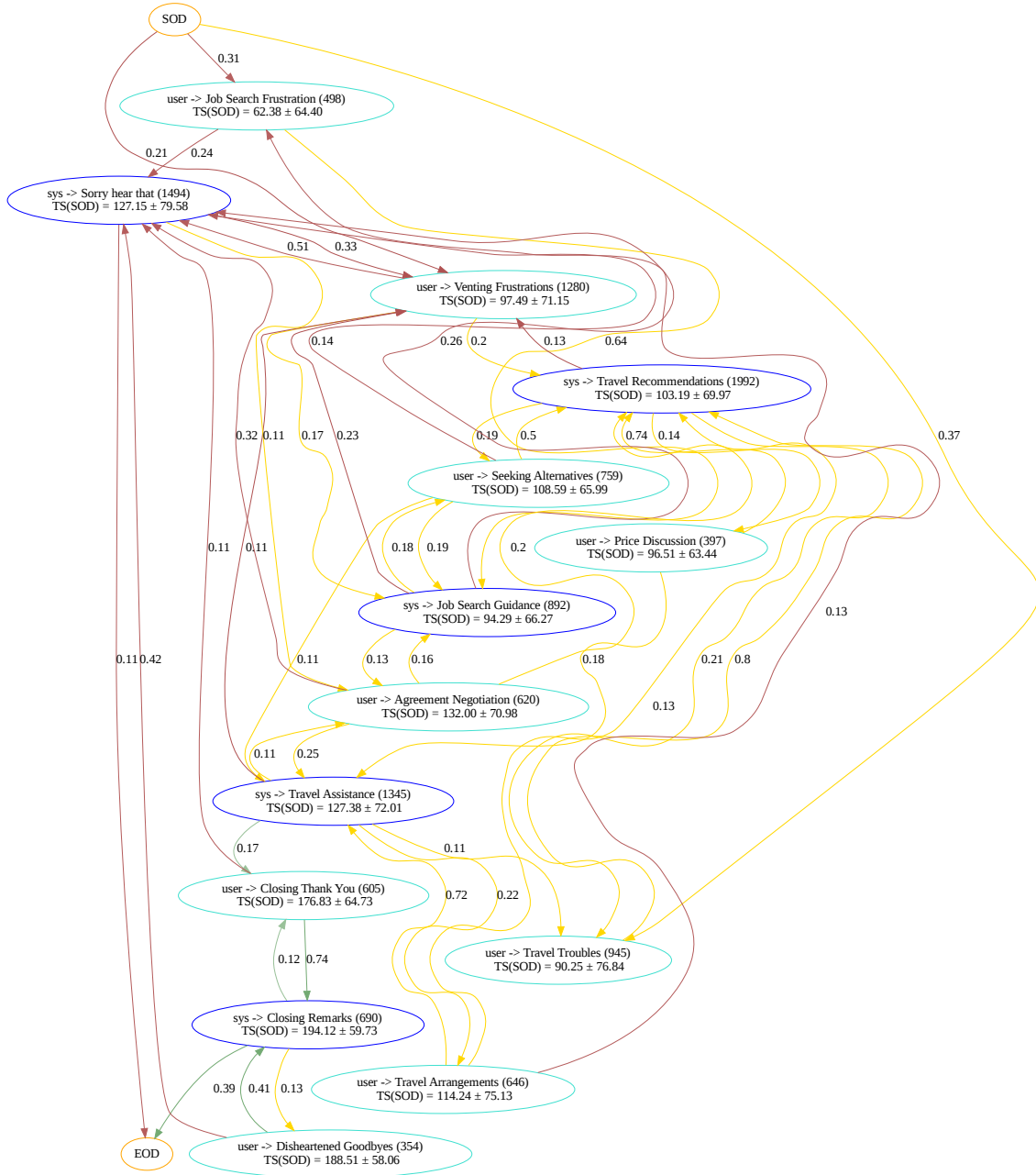


Figure 17: Dialogue flow discovered for Subset 10.

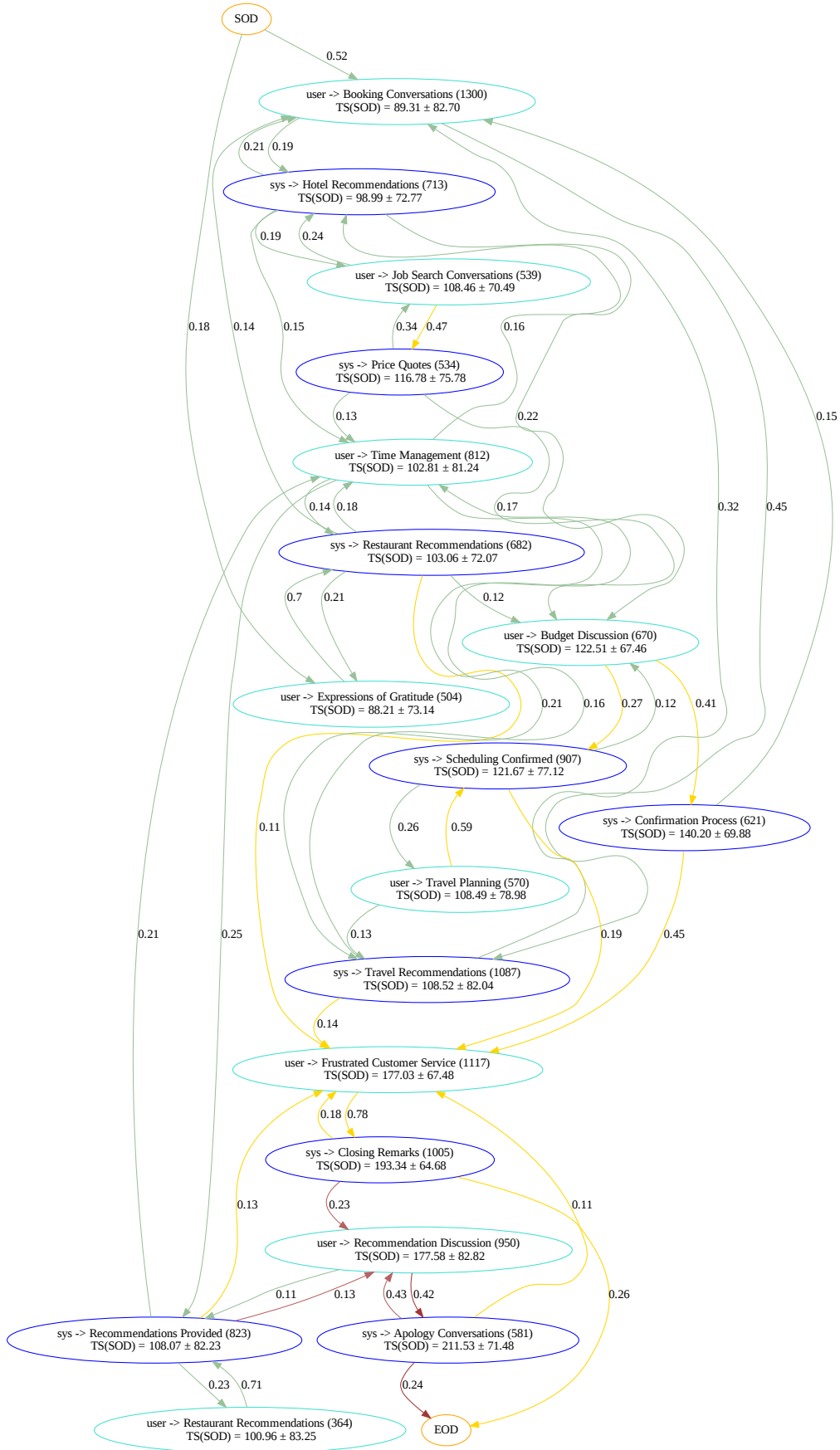


Figure 18: Dialogue flow discovered for Subset 11.

State	U_c	TU	TS(SOD)	TS(Prev)	S_c
User					
C0: Gift Shopping Conversations	1253	13.34 ± 4.68	992.11 ± 747.71	107.11 ± 55.35	0.75 ± 0.25
C1: Expressions of Gratitude	1050	9.39 ± 3.62	1934.74 ± 674.60	117.55 ± 49.73	0.85 ± 0.23
C2: Conversational Dialogue	597	7.31 ± 3.92	1333.34 ± 790.77	114.98 ± 49.58	0.61 ± 0.21
C3: Food Recommendations	530	15.10 ± 4.33	991.53 ± 735.31	95.31 ± 64.31	0.81 ± 0.24
C4: Booking Confirmation	1094	14.94 ± 4.83	1316.62 ± 721.58	115.51 ± 51.25	0.76 ± 0.25
C5: Travel Planning	1336	15.22 ± 4.66	828.58 ± 788.24	86.12 ± 66.55	0.76 ± 0.25
System					
C0: Travel Planning	710	22.54 ± 8.24	1311.36 ± 774.47	113.78 ± 50.87	0.64 ± 0.24
C1: Booking Confirmations	689	20.84 ± 7.24	1507.96 ± 740.08	116.19 ± 49.70	0.69 ± 0.25
C2: Closing Remarks	256	18.31 ± 5.86	1827.91 ± 700.79	116.65 ± 50.41	0.89 ± 0.21
C3: Price Options	441	21.85 ± 8.39	1130.15 ± 693.95	115.38 ± 49.96	0.60 ± 0.21
C4: Informational Conversations	739	25.53 ± 8.28	1133.93 ± 704.99	116.54 ± 49.73	0.79 ± 0.25
C5: Expressing Appreciation	931	23.03 ± 8.51	990.92 ± 750.26	117.60 ± 48.47	0.77 ± 0.25
C6: Food Recommendations	843	13.19 ± 5.96	2169.19 ± 614.39	115.85 ± 48.70	0.97 ± 0.11
C7: Venue Information	1230	25.20 ± 8.28	1010.14 ± 777.16	118.04 ± 50.09	0.76 ± 0.25

Table 16: Clustering metrics for Subset 13.

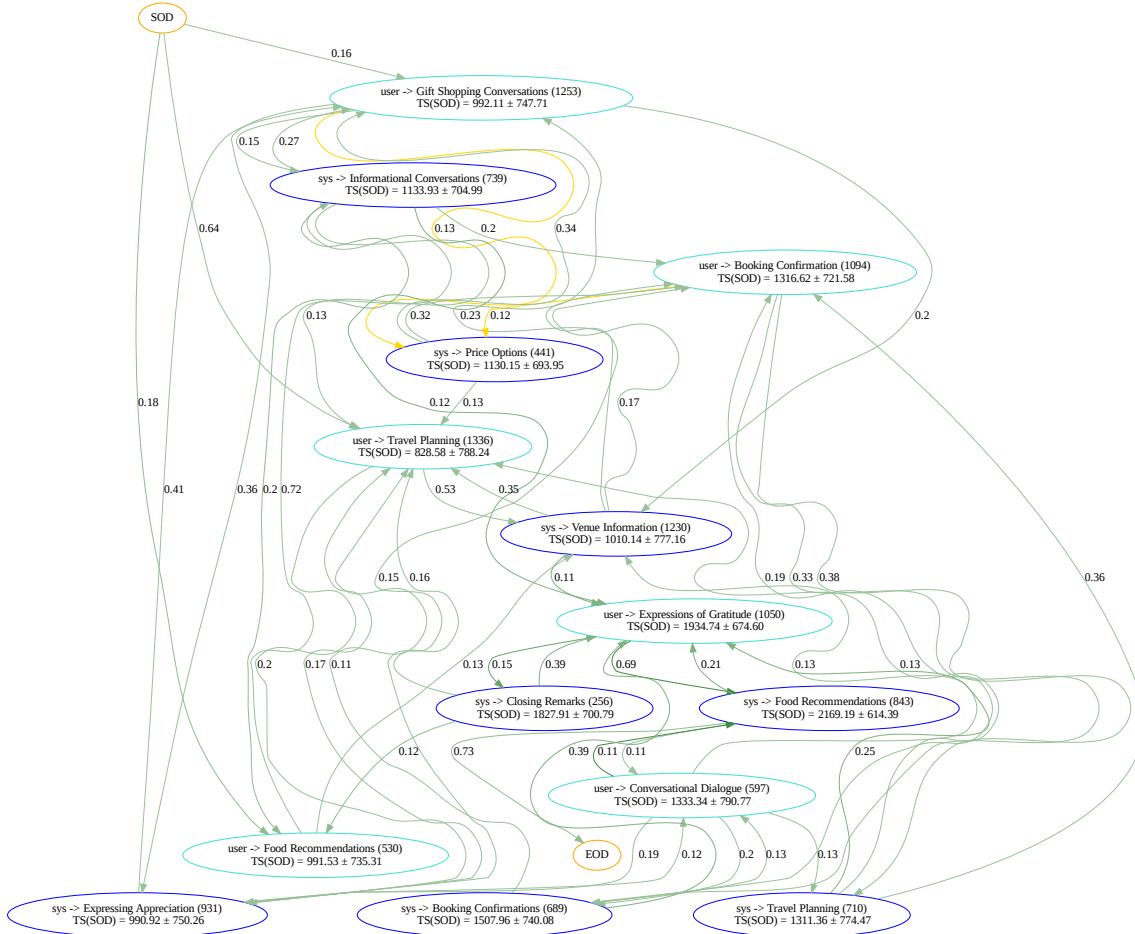


Figure 19: Dialogue flow discovered for Subset 13.

State	U_c	TU	TS(SOD)	TS(Prev)	S_c
User					
C0: Travel Planning	1042	13.98 $\pm$ 4.43	1053.84 $\pm$ 921.97	92.74 $\pm$ 64.38	0.51 $\pm$ 0.32
C1: Travel Assistance	938	13.66 $\pm$ 4.73	1394.65 $\pm$ 905.28	108.42 $\pm$ 55.83	0.48 $\pm$ 0.29
C2: Furniture Options Discussion	1069	12.15 $\pm$ 4.53	1265.74 $\pm$ 854.80	113.46 $\pm$ 51.68	0.45 $\pm$ 0.32
C3: Complaining About Life	2013	10.84 $\pm$ 5.36	1462.80 $\pm$ 1012.45	109.77 $\pm$ 55.27	0.36 $\pm$ 0.29
C4: Closing Conversations	1233	9.13 $\pm$ 3.83	2109.26 $\pm$ 869.67	116.28 $\pm$ 49.91	0.49 $\pm$ 0.33
C5: Job Search Struggles	435	15.51 $\pm$ 5.64	709.06 $\pm$ 794.05	74.05 $\pm$ 67.86	0.36 $\pm$ 0.29
System					
C0: Travel Advice	1057	19.27 $\pm$ 6.93	1540.61 $\pm$ 870.72	117.06 $\pm$ 50.07	0.41 $\pm$ 0.27
C1: Restaurant Recommendations	1192	21.23 $\pm$ 7.73	1339.43 $\pm$ 937.20	116.51 $\pm$ 48.48	0.47 $\pm$ 0.30
C2: Job Search Support	871	14.83 $\pm$ 7.26	2101.44 $\pm$ 955.61	117.14 $\pm$ 48.96	0.52 $\pm$ 0.33
C3: Apology Closure	525	21.48 $\pm$ 7.38	1230.55 $\pm$ 870.82	118.08 $\pm$ 50.26	0.48 $\pm$ 0.35
C4: Empathetic Support	655	22.91 $\pm$ 8.47	1160.29 $\pm$ 835.88	115.38 $\pm$ 50.03	0.42 $\pm$ 0.33
C5: Travel Assistance	871	19.42 $\pm$ 7.87	1325.41 $\pm$ 914.05	117.52 $\pm$ 49.71	0.45 $\pm$ 0.30
C6: Price Inquiries	751	12.69 $\pm$ 6.59	2421.10 $\pm$ 772.24	115.80 $\pm$ 49.88	0.84 $\pm$ 0.33
C7: Supportive Conversations	756	19.35 $\pm$ 7.74	1075.53 $\pm$ 932.44	114.93 $\pm$ 50.24	0.22 $\pm$ 0.33

Table 17: Clustering metrics for Subset 14.

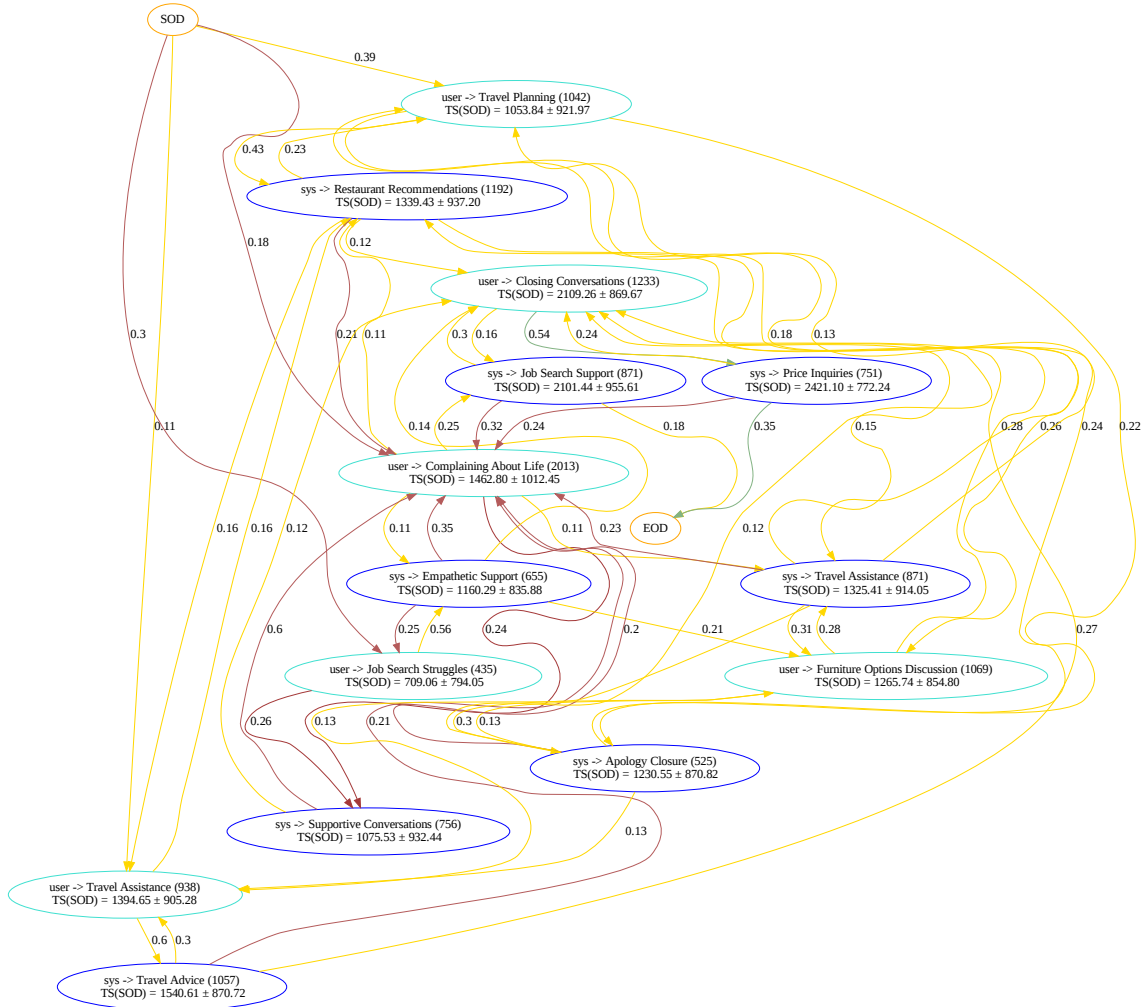


Figure 20: Dialogue flow discovered for Subset 14.

State	U_c	TU	TS(SOD)	TS(Prev)	S_c
User					
C0: Price Inquiries	453	14.68 $\pm$ 5.22	1510.22 $\pm$ 890.01	119.57 $\pm$ 49.79	0.66 $\pm$ 0.29
C1: Confirmations Given	1027	15.85 $\pm$ 4.69	1024.01 $\pm$ 911.95	95.49 $\pm$ 64.53	0.77 $\pm$ 0.31
C2: Gratitude Expressions	580	14.18 $\pm$ 4.50	1297.18 $\pm$ 858.25	116.14 $\pm$ 51.72	0.68 $\pm$ 0.30
C3: Job Interview Discussion	534	9.35 $\pm$ 4.68	1434.68 $\pm$ 866.58	114.45 $\pm$ 50.73	0.57 $\pm$ 0.38
C4: Restaurant Recommendations	809	10.49 $\pm$ 4.30	2175.05 $\pm$ 743.73	117.37 $\pm$ 49.19	0.60 $\pm$ 0.44
C5: Confusion Unfolds	628	15.88 $\pm$ 5.02	1131.15 $\pm$ 1075.83	84.94 $\pm$ 67.00	0.73 $\pm$ 0.35
C6: Booking Confirmations	850	14.03 $\pm$ 5.13	1090.46 $\pm$ 840.79	106.41 $\pm$ 57.89	0.76 $\pm$ 0.31
C7: Travel Planning	458	15.56 $\pm$ 4.57	1076.40 $\pm$ 859.44	94.65 $\pm$ 62.06	0.80 $\pm$ 0.31
C8: Attractions Inquiry	973	11.97 $\pm$ 5.96	2125.80 $\pm$ 935.31	113.43 $\pm$ 51.40	0.80 $\pm$ 0.31
C9: Accommodation Inquiry	574	15.79 $\pm$ 4.71	1537.18 $\pm$ 829.83	110.45 $\pm$ 51.02	0.70 $\pm$ 0.33
System					
C0: Price Discussions	629	21.76 $\pm$ 8.55	1447.43 $\pm$ 918.07	118.74 $\pm$ 49.82	0.53 $\pm$ 0.29
C1: Hotel Recommendations	1545	25.09 $\pm$ 8.57	1186.42 $\pm$ 929.83	117.85 $\pm$ 49.12	0.68 $\pm$ 0.35
C2: Job Discussion	828	21.71 $\pm$ 8.29	1231.56 $\pm$ 923.50	122.07 $\pm$ 47.29	0.64 $\pm$ 0.35
C3: Restaurant Recommendations	672	23.96 $\pm$ 7.67	1254.92 $\pm$ 847.78	118.46 $\pm$ 48.98	0.70 $\pm$ 0.36
C4: Customer Support	674	19.40 $\pm$ 9.55	2341.06 $\pm$ 832.24	113.99 $\pm$ 48.23	0.22 $\pm$ 0.32
C5: Booking Confirmation	841	22.04 $\pm$ 7.61	1699.50 $\pm$ 833.93	116.18 $\pm$ 48.96	0.50 $\pm$ 0.33
C6: Travel Arrangements	791	20.98 $\pm$ 8.18	1449.28 $\pm$ 930.55	118.59 $\pm$ 48.82	0.55 $\pm$ 0.34
C7: Closing Assistance	845	15.62 $\pm$ 6.35	2324.85 $\pm$ 756.35	116.45 $\pm$ 49.84	0.58 $\pm$ 0.49

Table 18: Clustering metrics for Subset 15.

State	U_c	TU	TS(SOD)	TS(Prev)	S_c
User					
C0: Conversation Options	1084	13.58 $\pm$ 4.69	1038.57 $\pm$ 657.60	111.64 $\pm$ 52.26	0.60 $\pm$ 0.35
C1: Travel Planning	1250	15.95 $\pm$ 4.85	981.55 $\pm$ 767.34	95.56 $\pm$ 64.18	0.52 $\pm$ 0.38
C2: Emotional Venting	1141	13.00 $\pm$ 6.03	913.11 $\pm$ 749.39	99.70 $\pm$ 62.03	0.40 $\pm$ 0.34
C3: Booking Confirmations	926	12.32 $\pm$ 5.49	1360.60 $\pm$ 650.15	112.39 $\pm$ 50.45	0.71 $\pm$ 0.29
C4: Job Search Struggles	409	16.17 $\pm$ 4.95	653.29 $\pm$ 660.53	84.29 $\pm$ 67.65	0.34 $\pm$ 0.35
C5: Expressions of Gratitude	917	9.99 $\pm$ 4.05	1980.01 $\pm$ 580.07	115.27 $\pm$ 50.19	0.91 $\pm$ 0.20
System					
C0: Restaurant Recommendations	554	23.64 $\pm$ 8.00	1048.69 $\pm$ 690.48	117.09 $\pm$ 51.36	0.67 $\pm$ 0.35
C1: Booking Conversations	657	21.90 $\pm$ 7.52	1421.87 $\pm$ 701.00	116.31 $\pm$ 49.50	0.61 $\pm$ 0.28
C2: Providing Information	298	15.51 $\pm$ 5.96	300.42 $\pm$ 419.90	115.38 $\pm$ 51.18	0.10 $\pm$ 0.24
C3: Tourist Information	780	19.99 $\pm$ 8.67	1269.10 $\pm$ 753.12	113.76 $\pm$ 49.69	0.66 $\pm$ 0.31
C4: Job Search Discussion	463	20.59 $\pm$ 7.60	1484.52 $\pm$ 692.35	119.22 $\pm$ 48.22	0.64 $\pm$ 0.26
C5: Hotel Options	777	23.69 $\pm$ 8.31	1132.55 $\pm$ 667.78	117.98 $\pm$ 49.57	0.64 $\pm$ 0.31
C6: Closing Remarks	588	24.32 $\pm$ 8.08	956.86 $\pm$ 620.44	112.21 $\pm$ 51.22	0.59 $\pm$ 0.32
C7: Empathetic Concerns	599	23.55 $\pm$ 8.61	957.97 $\pm$ 683.14	115.31 $\pm$ 48.81	0.52 $\pm$ 0.31
C8: Supportive Conversation	973	16.99 $\pm$ 7.42	2064.61 $\pm$ 567.83	113.52 $\pm$ 51.73	0.99 $\pm$ 0.07

Table 19: Clustering metrics for Subset 16.

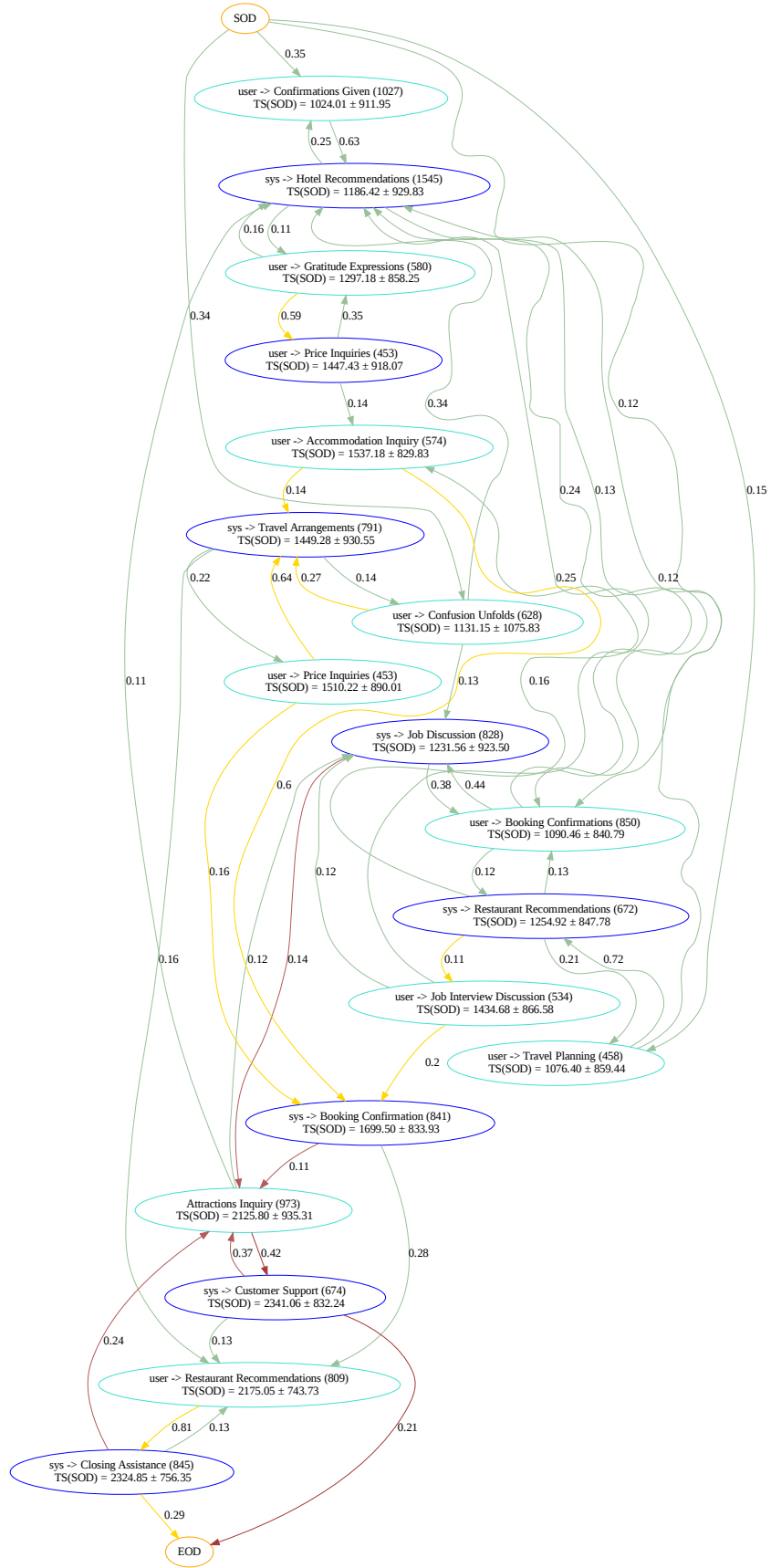


Figure 21: Dialogue flow discovered for Subset 15.

