Reimagining Intent Prediction: Insights from Graph-Based Dialogue Modeling and Sentence Encoders

By:
Ledneva Daria, ledneva.dr@mipt.ru
Denis Kuznetsov, kuznetsov.den.p@phystech.edu

Graph Models: The Future of Dialogue Systems

Revealing the Power of Graph-Based Models in Dialogue Systems

Within our study:

➔ Dive into the future of Conversational AI with our groundbreaking research 🌟

➔ Explore scenario dialogue graphs: the solution for context comprehension 🔍

➔ Unlock the secrets behind accurate intent prediction in closed-domain dialogue systems 💡

➔ Elevate your dialogue systems to new heights with insights from our study! 💬

Dialogue Data Characteristics

Understanding the Dynamics of Dialogue Data

- **Features of dialogues:**
  - Dialogues have a regular structure
  - Participants play different roles
  - Contextual dependencies
- **Intention** (dialogue state) – the goal/purpose of a dialogue participant in a dialogue utterance
- **Intent prediction** in a dialogue system is the determination of the intention of the next utterance in a dialogue based on the context

Multipartite Scenario Dialogue Graph

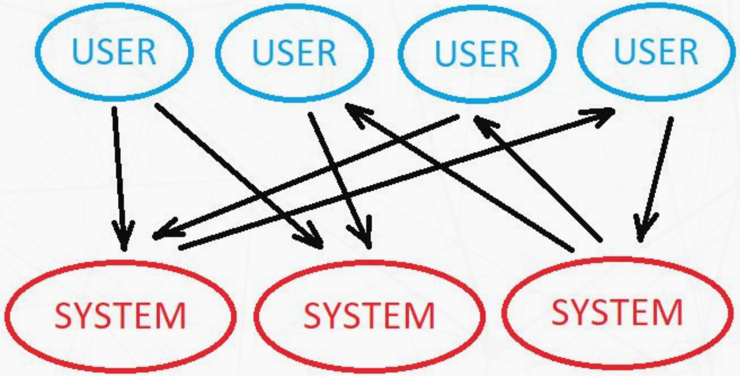
Visualizing Dialogue Systems: Understanding Multipartite Scenario Dialogue Graphs

- Two types of dialogue systems:
  - With an **open** domain
  - With a **closed** domain
- Dialogues in the dialogue systems with a **closed** domain:
  - Restricted to a narrow subject area
  - Can be modeled as a **chain of intents** with transitions between them
- A **multipartite graph** – an interpretable representation of a dialogue system
- Each **partite** of the graph represents one of the **roles** of the dialogue participants
- The **role** defines the function or position of each participant in the dialogue

Multipartite Scenario Dialogue Graph

Visualizing Dialogue Systems: Understanding Multipartite Scenario Dialogue Graphs

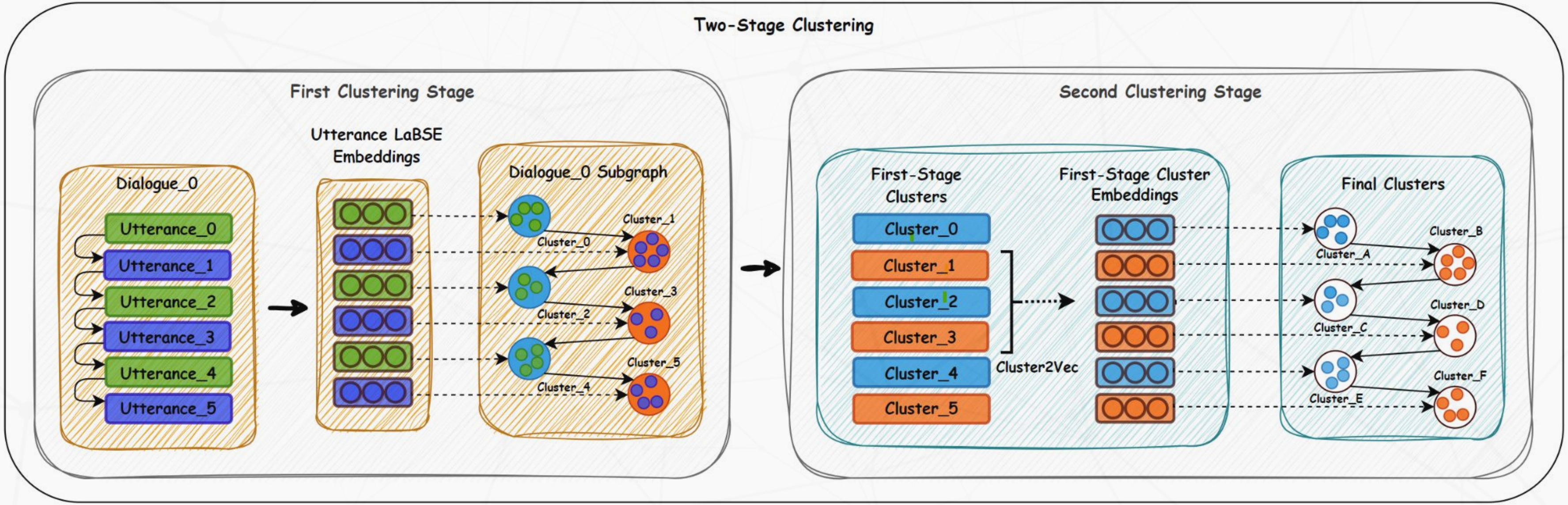
- Each **node** of the graph defines a unique **intention** in the dialogue
- The **edges** in the graph are **transitions** between states of the dialogue
- **Closed** domain datasets: 2 roles (user, system) and a bipartite graph
- **Open** domain datasets: 1 role (dialogue participant) and a unipartite graph



Two-stage Clustering Algorithm

Clustering Algorithm: Two-Stage Approach for Constructing Nodes in a Dialogue Graph

- ➔ The first stage: the semantics of utterances
- ➔ The second stage: contextual dependencies
- ➔ **Cluster2Vec**: the clusters play the role of "words"



Examples of Dialogue Graph Nodes

Dialogue Graph Nodes: Utterances with Similar Semantic and Contextual Occurrence

Samples from the graph nodes, two-stage clustering method			
User cluster #1	**User cluster #2**	**Dialogue system cluster #1**	**Dialogue system cluster #2**
Can I please have the phone number and address for that place?	Yes, please book a table for 4 people at 12:15 on Tuesday.	Thank you for contacting us and have a nice day.	I'm sorry. There is still no availability. Would you like to try a different hotel then?
Could you tell me the price, address and phone number?	Book it for the same number of people at 14:30 on the same day.	Thank you for using Cambridge Town Info centre, have a great day!	I'm sorry, there were no rooms available. Perhaps you'd like to find another hotel?
How about Jesus Green Outdoor pool. Could I have their address and phone number?	I don't have a preference for food type. I do need reservations for 8 at 12:00 on Thursday.	You're very welcome, enjoy your time in Cambridge!	I'm sorry, there are no rooms available for that length of stay. Could you shorten your stay or book a different day possibly?
Yes, please. Can I get the address and phone number for the one you recommend?	Can you see if there's anything at 20:00?	Great! I'm happy to help. Goodbye!	The booking for the Acorn Guest House was unsuccessful. Would you like me to look for another hotel for you?
Do you have there phone number?	La Mimosa sounds good. Can your reserve me a table for 1 on Saturday at 11:15?	I'm glad I was able to help. Please call back if you have any more questions!	I am sorry, but the Leverton House was not available for your party on Tuesday. Would you like me to look for another hotel?

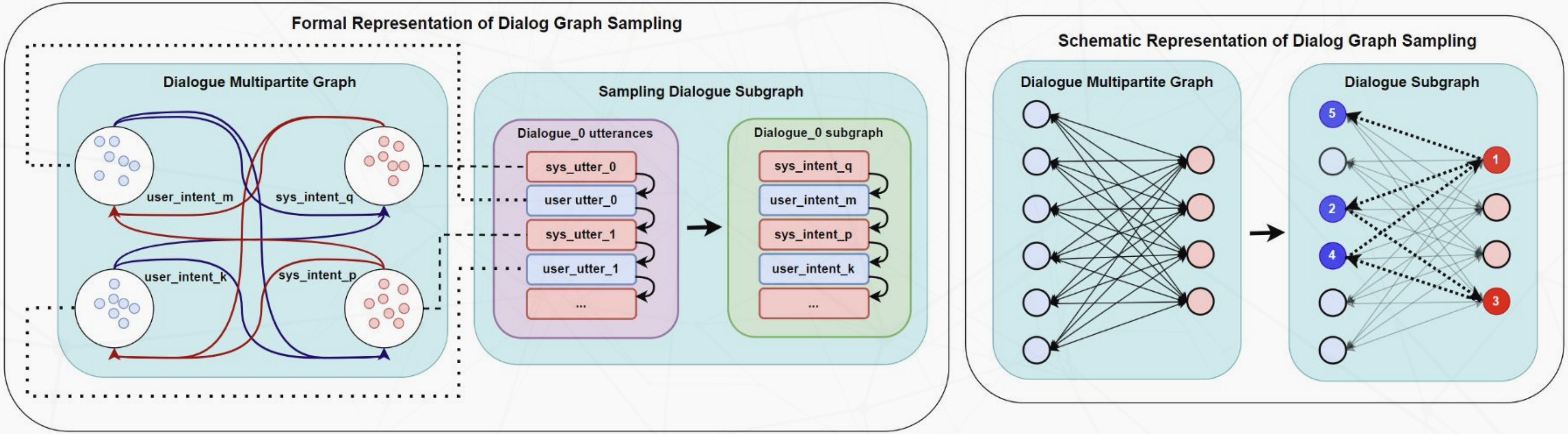
Table 5: Samples from the user and dialogue system MultiWOZ 2.2 graph nodes.

Dialogue Subgraph Sampling

Dialogue Subgraph Construction: Extracting Structure from Dialogues Using Dialogue Graphs

- ➔ Dialogue -> Dialogue Subgraph ***G = (V, E)***
- ➔ Vertex **(v_i)** contains the intention of the utterance **(u_i)**

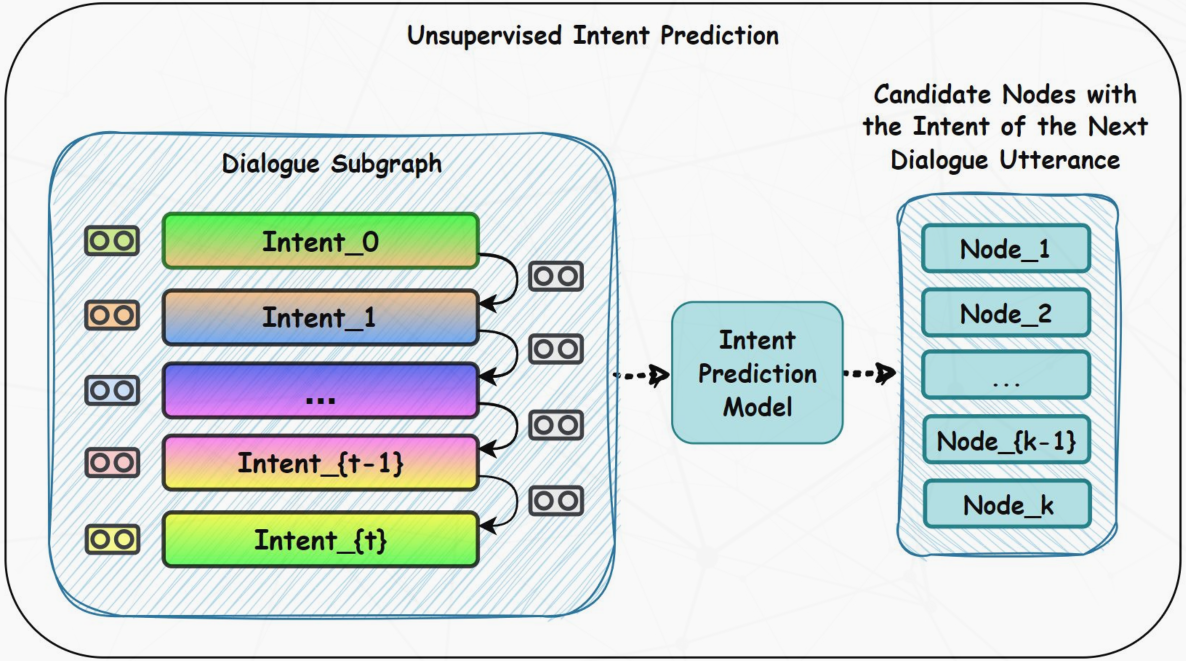
$$V = \text{unique}(\{v_1, v_2, ..., v_t\}) \qquad E = \{(v_1, v_2), (v_2, v_3), ..., (v_{t-1}, v_t)\}$$



Next Intention Prediction

Visual Representation: Predicting Next Intentions with Dialogue Subgraphs

➔ The task of predicting the next intention:



Baseline Approaches

Description of the Approaches Compared to the Graph Methodologies in the Study

- **Markov Chains**
  - Based on probabilistic transitions in a multipartite dialogue graph
- **Encoder**
  - Obtaining vector representations for utterances, predicting next dialogue utterance and their intent based on these representations
- **ConveRT**
  - Dual encoder model
  - Takes into account more than one dialogue history utterance
- **ConveRT-MAP**
  - ConveRT + Context-Response Feed-Forward Neural Network
  - Contrastive loss based on cosine distance is used as a loss
- **Gradient Boosting (CatBoost)**

Graph-Based Approaches

A Comprehensive Explanation of Methodologies Utilizing Graphs

- **Homogeneous** configuration (**Message Passing**):
  - One type of the edges and vertices in the graph
  - Graph Attention Networks (GATs)[1] was used alongside other Message Passing Neural Networks
  - GATs characterized by its attention mechanism on the graph
- **Heterogeneous** configuration (**FastGTN**[2]):
  - Various types of the edges and vertices in the graph
  - A separate weight matrix for each vertex type
  - Complex structural dependencies are taken into account in addition to neighbouring vertex representations

Datasets

An Overview of Open and Closed Domain Data Employed for Evaluation

- **Open Domain Datasets:**
  - PersonaChat: 160,000+ conversational exchanges on diverse topics
  - DailyDialog: 13,000+ dialogues spanning life events and interests
- **Closed Domain Datasets:**
  - MultiWOZ 2.2: 10,000+ dialogues across 7 domains like hotels and restaurants
  - FoCus: 14,000+ dialogues centered on geographical landmarks
  - Taskmaster: 13,000+ dialogues across 6 domains, including written and spoken interactions

Experiment Setup

Details of Experiment Design and Metric Descriptions

- ➔ **Metrics** (the accuracy of predicting the intention of the next utterance):
  - ◆ ***Recall@k:*** *k* ∈ *{1, 3, 5, 10}*
  - ◆ ***MAR (Mean Average Recall):***
    - The average value of *Recall@k* for *k* ∈ *{1, 3, 5, 10}*
  - ◆ **Separate** metrics for **different** dialogue roles
- ➔ To ensure result stability, each approach was trained on **three** different sets of clusters, and the metrics were then averaged
- ➔ Each approach was run on **three** configurations of cluster numbers:
  - ◆ [200, 30], [400, 60], [800, 120]
- ➔ The choice of the **number of clusters** depends on the **unique characteristics** of each dataset and the specific requirements of the task

One-Stage vs Two-Stage Clustering

Comparing Approaches for Constructing Dialogue Graph Nodes

➔ Employing a two-stage clustering approach outperforms single-stage clustering for next-intention prediction tasks

Models	MPNet	MPNet-one-stage
# of Parameters	109M	109M
	Encoder	
Recall@1	**23.63** $\pm$ **0.531**	19.18 $\pm$ 0.421
Recall@3	**47.87** $\pm$ **0.469**	41.31 $\pm$ 0.435
Recall@5	**58.92** $\pm$ **0.738**	53.99 $\pm$ 0.157
Recall@10	**74.19** $\pm$ **1.109**	72.21 $\pm$ 0.023
	Message Passing	
Recall@1	**46.94** $\pm$ **1.135**	37.79 $\pm$ 0.818
Recall@3	**74.40** $\pm$ **0.277**	67.12 $\pm$ 0.386
Recall@5	**83.45** $\pm$ **0.136**	80.46 $\pm$ 0.470
Recall@10	**92.74** $\pm$ **0.352**	92.61 $\pm$ 0.703
	Markov Chain	
Recall@1	**37.62** $\pm$ **0.503**	27.56 $\pm$ 1.007
Recall@3	**63.86** $\pm$ **0.282**	55.20 $\pm$ 0.993
Recall@5	**75.19** $\pm$ **0.474**	70.81 $\pm$ 1.164
Recall@10	**88.56** $\pm$ **0.728**	88.23 $\pm$ 0.483

Metrics: Sentence Encoder Selection

Comparison of Different Sentence Encoders for Dialogue Graph Node Construction

Models	MPNet	MPNet-one-stage	DistilRoBERTa	S-BERT	MiniLM	GloVe	GPT	T5
# of Parameters	109M	109M	82M	22M	33M	120M	125M	335M
				Encoder				
Recall@1	**23.63 ± 0.531**	19.18 ± 0.421	**23.92 ± 0.806**	21.22 ± 1.417	**23.15 ± 1.489**	13.35 ± 0.341	21.01 ± 1.233	**23.08 ± 0.884**
Recall@3	**47.87 ± 0.469**	41.31 ± 0.435	**47.57 ± 0.219**	43.55 ± 1.086	**47.13 ± 1.508**	32.51 ± 0.890	44.36 ± 1.241	**48.95 ± 0.719**
Recall@5	**58.92 ± 0.738**	53.99 ± 0.157	**58.81 ± 0.405**	53.67 ± 1.012	**59.50 ± 0.419**	44.07 ± 0.840	54.90 ± 1.223	**60.01 ± 0.343**
Recall@10	**74.19 ± 1.109**	72.21 ± 0.023	**73.75 ± 1.164**	68.28 ± 0.914	**74.35 ± 0.372**	61.97 ± 1.046	71.72 ± 1.541	**73.70 ± 0.271**
				Message Passing				
Recall@1	**46.94 ± 1.135**	37.79 ± 0.818	**46.55 ± 1.288**	45.82 ± 1.263	**46.33 ± 0.766**	38.77 ± 1.726	44.78 ± 0.633	**48.23 ± 0.614**
Recall@3	**74.40 ± 0.277**	67.12 ± 0.386	**74.36 ± 0.533**	71.80 ± 0.804	**72.82 ± 1.033**	64.07 ± 0.797	71.07 ± 0.212	**74.29 ± 0.687**
Recall@5	**83.45 ± 0.136**	80.46 ± 0.470	**83.63 ± 0.558**	81.62 ± 0.756	82.15 ± 0.670	76.47 ± 0.336	81.50 ± 0.211	**83.90 ± 0.532**
Recall@10	**92.74 ± 0.352**	92.61 ± 0.703	**93.17 ± 0.758**	92.27 ± 0.541	**92.35 ± 0.486**	89.99 ± 0.534	92.37 ± 0.345	**93.31 ± 0.752**
				Markov Chain				
Recall@1	**37.62 ± 0.503**	27.56 ± 1.007	**37.99 ± 0.599**	36.66 ± 1.207	**37.47 ± 0.648**	28.66 ± 1.735	36.98 ± 1.105	**36.81 ± 0.735**
Recall@3	63.86 ± 0.282	55.20 ± 0.993	**65.52 ± 0.469**	63.43 ± 0.965	**64.65 ± 0.513**	52.76 ± 1.503	61.29 ± 0.940	**65.28 ± 0.588**
Recall@5	75.19 ± 0.474	70.81 ± 1.164	**76.96 ± 0.269**	74.45 ± 0.977	**76.20 ± 0.322**	64.97 ± 1.106	72.83 ± 0.452	**76.38 ± 0.638**
Recall@10	**88.56 ± 0.728**	88.23 ± 0.483	**89.62 ± 0.564**	87.78 ± 0.730	**88.48 ± 0.223**	82.92 ± 0.151	86.71 ± 0.294	**89.37 ± 0.727**

Table 1: Evaluation of text encoders in generating vector representations for dialogue utterances in the MultiWOZ dataset and their impact on the three primary approaches: Message Passing, Encoder, and Markov Chain.

Metrics: Closed Domain Datasets

Results of Evaluation of Approaches on Closed Domain Datasets

Approach	# Parameters	Relative Training Time	Dataset # Clusters First Stage	Dataset # Clusters Second Stage	MultiWOZ User	MultiWOZ Dialog System	MultiWOZ All	FoCus User	FoCus Dialog System	FoCus All	Taskmaster User	Taskmaster Dialog System	Taskmaster All
Markov Chain	10K	0.13	200	30	59.47 ± 0.77	75.57 ± 0.59	67.52 ± 0.48	52.55 ± 1.30	52.15 ± 2.06	52.35 ± 0.98	57.79 ± 0.45	59.63 ± 0.67	58.77 ± 0.51
			400	60	47.05 ± 1.88	66.19 ± 1.50	56.61 ± 1.60	46.67 ± 0.70	44.46 ± 0.71	45.57 ± 0.56	49.84 ± 0.86	49.06 ± 0.29	49.52 ± 0.52
			800	120	30.90 ± 1.26	48.33 ± 1.47	39.62 ± 0.43	39.67 ± 1.91	39.86 ± 0.76	39.77 ± 0.81	42.60 ± 0.44	43.57 ± 0.24	43.14 ± 0.18
Message Passing	82M + 3.7M	0.47	200	30	$\mathbf{65.24 \pm 1.09}$	$\mathbf{83.62 \pm 0.64}$	$\mathbf{74.43 \pm 0.78}$	$\mathbf{66.34 \pm 2.31}$	$\mathbf{68.80 \pm 0.70}$	$\mathbf{67.57 \pm 1.46}$	$\mathbf{72.04 \pm 0.70}$	$\mathbf{78.69 \pm 0.60}$	$\mathbf{75.41 \pm 0.45}$
			400	60	$\mathbf{52.66 \pm 0.44}$	$\mathbf{75.88 \pm 0.78}$	$\mathbf{64.27 \pm 0.33}$	$\mathbf{59.56 \pm 1.67}$	$\mathbf{63.36 \pm 0.72}$	$\mathbf{61.46 \pm 0.71}$	64.73 ± 0.53	$\mathbf{69.98 \pm 0.47}$	$\mathbf{67.40 \pm 0.33}$
			800	120	$\mathbf{35.93 \pm 0.72}$	$\mathbf{58.35 \pm 0.92}$	$\mathbf{47.14 \pm 0.67}$	$\mathbf{54.64 \pm 1.05}$	$\mathbf{56.07 \pm 0.90}$	$\mathbf{55.35 \pm 0.61}$	$\mathbf{57.56 \pm 0.41}$	$\mathbf{64.00 \pm 0.37}$	$\mathbf{60.83 \pm 0.32}$
CatBoost	82M + 2.2M	1.00	200	30	$\mathbf{65.88 \pm 0.54}$	$\mathbf{83.09 \pm 0.56}$	$\mathbf{74.48 \pm 0.45}$	$\mathbf{65.71 \pm 0.37}$	$\mathbf{69.09 \pm 0.31}$	$\mathbf{67.41 \pm 0.20}$	71.57 ± 0.30	$\mathbf{78.23 \pm 0.52}$	$\mathbf{74.94 \pm 0.24}$
			400	60	51.07 ± 1.07	73.09 ± 0.81	62.08 ± 0.83	$\mathbf{59.61 \pm 1.47}$	60.91 ± 0.46	$\mathbf{60.26 \pm 0.77}$	$\mathbf{65.03 \pm 0.34}$	68.93 ± 0.33	67.01 ± 0.24
			800	120	$\mathbf{37.16 \pm 0.58}$	55.45 ± 0.74	46.30 ± 0.59	$\mathbf{54.55 \pm 0.35}$	53.94 ± 0.74	54.25 ± 0.49	56.53 ± 0.35	62.60 ± 0.29	59.61 ± 0.30
FastGTN	82M + 1.9M	0.49	200	30	$\mathbf{65.55 \pm 0.64}$	$\mathbf{83.04 \pm 0.48}$	$\mathbf{74.30 \pm 0.26}$	$\mathbf{65.12 \pm 2.73}$	$\mathbf{68.98 \pm 1.16}$	$\mathbf{67.05 \pm 1.38}$	$\mathbf{72.53 \pm 0.41}$	$\mathbf{78.30 \pm 0.51}$	$\mathbf{75.46 \pm 0.36}$
			400	60	$\mathbf{51.84 \pm 0.66}$	$\mathbf{75.94 \pm 0.95}$	$\mathbf{63.89 \pm 0.55}$	55.89 ± 1.93	61.76 ± 0.58	58.82 ± 1.04	$\mathbf{65.84 \pm 0.50}$	$\mathbf{70.11 \pm 0.36}$	$\mathbf{68.01 \pm 0.29}$
			800	120	$\mathbf{36.40 \pm 0.90}$	$\mathbf{58.38 \pm 1.29}$	$\mathbf{47.39 \pm 0.41}$	$\mathbf{54.19 \pm 1.50}$	$\mathbf{55.91 \pm 0.28}$	$\mathbf{55.05 \pm 0.77}$	$\mathbf{57.52 \pm 0.51}$	$\mathbf{64.27 \pm 0.47}$	$\mathbf{60.93 \pm 0.43}$
Encoder	82M	0.50	200	30	34.69 ± 1.20	67.33 ± 0.90	51.01 ± 0.65	39.01 ± 1.63	59.11 ± 0.80	49.06 ± 0.77	46.08 ± 0.72	49.05 ± 0.42	47.56 ± 0.19
			400	60	24.67 ± 0.44	53.40 ± 2.03	39.04 ± 0.90	32.50 ± 0.87	50.39 ± 0.73	41.45 ± 0.56	36.35 ± 0.24	40.88 ± 0.20	38.61 ± 0.19
			800	120	15.31 ± 0.33	36.35 ± 0.74	25.83 ± 0.41	28.55 ± 0.41	43.16 ± 0.43	35.86 ± 0.26	27.82 ± 0.14	31.21 ± 0.14	29.52 ± 0.11
ConveRT	46M	0.36	200	30	32.81 ± 0.78	57.94 ± 0.94	45.38 ± 0.81	38.13 ± 0.85	60.62 ± 0.32	49.38 ± 0.50	47.52 ± 0.36	59.80 ± 0.78	53.66 ± 0.34
			400	60	21.10 ± 0.23	46.25 ± 1.00	33.67 ± 0.53	33.19 ± 0.63	52.53 ± 0.87	42.86 ± 0.45	37.87 ± 0.57	45.92 ± 0.64	41.90 ± 0.44
			800	120	12.71 ± 0.56	29.38 ± 0.69	21.04 ± 0.27	28.59 ± 0.23	45.80 ± 0.85	37.20 ± 0.47	29.54 ± 0.31	38.52 ± 0.18	34.03 ± 0.23
ConveRT MAP	46M + 2M	0.78	200	30	51.75 ± 1.87	75.97 ± 1.08	63.86 ± 1.38	55.74 ± 1.33	60.11 ± 1.49	57.92 ± 0.86	63.18 ± 0.68	70.82 ± 0.90	67.00 ± 0.68
			400	60	39.39 ± 1.33	61.44 ± 1.31	50.41 ± 1.32	44.31 ± 1.38	47.52 ± 1.40	45.92 ± 1.25	54.54 ± 0.61	58.59 ± 0.88	56.56 ± 0.53
			800	120	22.20 ± 1.21	39.75 ± 0.36	31.35 ± 0.58	37.62 ± 0.42	36.99 ± 1.43	37.29 ± 0.61	43.61 ± 1.09	49.61 ± 0.90	46.61 ± 0.99

Table 3: Experimental results for Mean Average Recall metric: various intent prediction approaches on the closed domain datasets. The training time of the models was counted from the start of training until the Early Stopping. The all metric is the average of the user metric and the dialogue system metric. To ensure stability of results, all approaches were trained on 3 different sets of clusters and the resulting metrics were averaged.

Metrics: Open Domain Datasets

Results of Evaluation of Approaches on Open Domain Datasets

Approach	# Parameters	Relative Training Time	# Clusters		PersonaChat	DailyDialog
			First Stage	Second Stage		
Markov Chain	10K	0.13	200	30	52.50 ± 2.27	49.91 ± 0.85
			400	60	41.67 ± 2.28	40.53 ± 2.66
			800	120	32.72 ± 1.03	31.48 ± 0.91
Message Passing	82M + 3.7M	0.47	200	30	**58.86 ± 1.06**	**57.13 ± 2.28**
			400	60	**48.79 ± 0.68**	**47.15 ± 0.71**
			800	120	**42.96 ± 0.68**	**38.52 ± 0.42**
CatBoost	82M + 2.2M	1.00	200	30	**59.31 ± 1.24**	**58.67 ± 0.90**
			400	60	**50.12 ± 0.78**	**47.55 ± 1.20**
			800	120	**42.56 ± 0.63**	**39.50 ± 0.60**
FastGTN	82M + 1.9M	0.49	200	30	**60.21 ± 2.29**	55.88 ± 0.54
			400	60	**49.11 ± 0.45**	**46.35 ± 0.71**
			800	120	**41.68 ± 1.35**	**38.92 ± 0.96**
Encoder	82M	0.50	200	30	43.45 ± 2.20	48.92 ± 0.58
			400	60	30.95 ± 2.02	39.95 ± 1.61
			800	120	24.10 ± 4.06	31.16 ± 0.66
ConveRT	46M	0.36	200	30	45.39 ± 1.46	50.24 ± 2.35
			400	60	35.01 ± 2.96	40.65 ± 0.92
			800	120	27.32 ± 2.33	32.27 ± 0.57
ConveRT MAP	46M + 2M	0.78	200	30	47.08 ± 2.01	50.51 ± 2.03
			400	60	39.97 ± 1.69	38.41 ± 2.15
			800	120	20.78 ± 2.01	29.66 ± 1.82

Table 2: Experimental results for Mean Average Recall metric: various intent prediction approaches on the open domain datasets. The training time of the models was counted from the start of training until the Early Stopping. The all metric is the average of the user metric and the dialogue system metric. To ensure stability of results, all approaches were trained on 3 different sets of clusters and the resulting metrics were averaged.

Metrics: Comparative Table

Assessing Proposed Approaches: Comparative Evaluation Across Diverse Metrics and Datasets

➔ If an approach performed best within the confidence interval within a specific configuration and dataset, it was assigned a score of 1

Dataset	Markov Chain	Message Passing	CatBoost	FastGTN	Encoder	ConveRT	ConveRT-MAP	Max Score
MultiWOZ	0	**9**	4	**9**	0	0	0	**9**
FoCus	0	**9**	6	6	0	0	0	**9**
Taskmaster	0	8	3	**9**	0	0	0	**9**
DailyDialog	0	**3**	**3**	2	0	0	0	**3**
PersonaChat	0	**3**	**3**	**3**	0	0	0	**3**
Closed Domain Summary	0	**26**	13	24	0	0	0	**27**
Open Domain Summary	0	**6**	**6**	5	0	0	0	**6**

Table 4: The table shows how different intent prediction methods performed in research. Each method gets a score of 1 if it does better than others on a specific metric; otherwise, it gets a score of 0. The table summarizes all the scores for each method and dataset.

Results and Discussion

Interpreting Findings: Insights and Analysis

The following results were obtained on the proposed methods and datasets:

- **Closed Domain Datasets**
  - Graph-based approaches showed superior performance
- **Open Domain Datasets**
  - Graph-based approaches were not outperforming gradient boosting techniques
  - Open-domain datasets have a weakly regular structure
- **Asymmetry in Dialogue Roles**
  - A significant distinction between user and dialog system metrics was observed

Limitations

Study Limitations: Exploring Boundaries and Methodological Constraints

- **Language Focus**
  - Experiments primarily centered on English dialogue datasets
- **Participant Pool Size**
  - The datasets involved a relatively small number of participants
- **Traditional Dialogue Emphasis**
  - The study was focused on conventional dialogues, excluding non-standard formats like social media conversations
- **Clustering Impact**
  - The study was conducted on fixed numbers of clusters
- **Sentence Encoder Selection**
  - Dialogue encoders like DSE were not considered

Thank you for your attention!