Dialogue Topic Segmentation Enhancement Method Based on Collaborative Graph Convolutional Networks and Large Language Models

Yaoxin Xu*

Faculty of Computer and Mathematical Sciences, the Hong Kong Polytechnic University, 999077 Hong Kong, China

Abstract. With the rapid development of dialogue systems and Retrieval-Augmented Generation (RAG) technologies, accurate dialogue topic segmentation has become a core link in improving context understanding and response quality. However, existing methods have some limitations: semantic similarity-based methods (e.g., Sentence-BERT) only capture local correlations and easily miss indirectly related segments, while Large Language Model (LLM)-based judgments, though precise in reasoning, are computationally expensive and slow, unsuitable for large-scale initial screening. Thus, this paper proposes a GCN-LLM collaborative topic segmentation framework for addressing these issues. It constructs a graph structure using LLM's "topic continuity" judgments as the "gold standard", designs a lightweight Graph Convolutional Network (GCN) to aggregate neighbor features and generate "enhanced embeddings" to overcome locality limitations, and adopts an "assist-not-replace" strategy (GCN optimizes initial screening features, LLM determines final results). Experiments on the dialseg_711.json dataset show the framework achieves an average F1-score of 0.8393, significantly outperforming existing methods (the highest F1 of comparison methods is 0.6606), and its graph dynamic update mechanism adapts to real-time dialogues. This study provides a new precision-efficiency-balanced solution for long-dialogue topic segmentation, especially suitable for context-sensitive RAG systems and intelligent customer service scenarios.

1 Introduction

With the rapid development of dialogue systems and Retrieval-Augmented Generation (RAG) technologies, accurate dialogue topic segmentation has become a core link in improving context understanding and response quality. In long conversations, user intentions evolve dynamically through multiple turns, and efficiently filtering historical dialogue segments that are topic-continuous with the current query directly affects the model's grasp of context. However, existing solutions have some clear limitations: Semantic similarity-based methods (e.g., Sentence-BERT) only capture local correlations, easily missing indirectly related segments; while Large Language

*Corresponding author: 23108099d@connect.polyu.hk

Model (LLM)-based judgments, despite precise reasoning, are computationally expensive and slow, making them unsuitable for large-scale initial screening.

To balance efficiency and accuracy, this paper proposes a collaborative framework of GCN and large models for topic segmentation: using Graph Convolutional Networks (GCN) to enhance the global correlation capture capability of semantic embeddings, while strictly retaining the LLM's final judgment logic. Its innovations include three aspects: constructing graph structures with LLMs' "topic continuity" judgments as the "gold standard," which aligns with real dialogue patterns without manual annotations; designing lightweight GCN to aggregate neighbor features and generate "enhanced embeddings," overcoming the locality limitation of original semantic similarity; and adopting an "assist-not-replace" strategy—GCN only optimizes feature representation in initial screening, with final results still determined by LLMs to retain their strong reasoning ability.

Experiments show that the framework performs excellently on public datasets (e.g., Ubuntu Dialogue Corpus), achieving an average F1-score of 0.8393, significantly outperforming existing solutions (the highest F1 of comparison scheme is 0.6606). The dynamic update mechanism of the graph structure enables efficient adaptation to new dialogues, providing a feasible path for real-time dialogue systems. This study offers a new precision-and-efficiency-balanced solution for topic segmentation in long conversations, especially suitable for context-sensitive RAG systems and intelligent customer service scenarios.

2 Related works

Dialogue topic segmentation, a critical prerequisite for context understanding in dialogue systems and RAG, has spurred extensive research across unsupervised learning, pre-trained model integration, graph-based modeling, and domain-specific adaptation.

Unsupervised methods dominate scenarios lacking manual annotations, with innovations focusing on efficiency and domain robustness. Park et al. proposed HyperSeg [1], an unsupervised approach leveraging hyperdimensional computing (HDC) to generate turn embeddings via permutation binding and majority voting. It auto-sets segmentation thresholds using similarity statistics, outperforming SOTA methods like GraphSeg on 4 out of 5 datasets while running 10x faster on CPU—addressing parameter sensitivity but lacking global correlation modeling [1]. Maraj complemented this with unsupervised strategies such as Coherence Graph (storing contextual keywords in graphs) and GTUTS (combining LLM-extracted topics with BERT embeddings) [2], which outperformed traditional techniques on Choi and WikiSection datasets, especially for short segments. These works validate unsupervised feasibility but struggle with indirect semantic associations, a gap addressed by the GCN's global feature aggregation.

Pre-trained models and coherence modeling have become pivotal for enhancing segmentation precision. Ghinassi et al. systematically compared 13 neural sentence encoders (NSEs) across domains, finding Transformer-based NSEs (e.g., TopSeg) outperformed traditional baselines like TextTiling, though text-similarity-finetuned models (e.g., SimCSE) failed to improve segmentation—highlighting that topic segmentation demands distinct semantic modeling from typical similarity tasks, aligning with the use of all-MiniLM for base encoding and GCN for enhanced representation [3]. Yu et al. tackled coherence deficiencies in long-document segmentation by introducing two auxiliary tasks: Topic-Aware Sentence Structure Prediction (TSSP) and Contrastive Semantic Similarity Learning (CSSL) [4]. Their Longformer-based model achieved 77.16 F1 on WIKI-727K, yet it focused on long documents rather than dialogue turn correlations. Maraj further advanced supervised pre-trained models with BERT-based Tri-encoders (integrating Bi-LSTM and attention) and LDA-BERT (fusing topic features), which outperformed Badjatiya et al.'s model on Clinical and Fiction datasets but remained computationally heavy for large-scale screening [2].

Graph-based models excel at capturing non-local dialogue associations. Lee and Choi proposed TUCORE-GCN, a heterogeneous graph model for dialogue relation extraction (RE) and emotion recognition (ERC), constructing graphs with turn, speaker, and argument nodes [5]. While it demonstrated strong contextual sensitivity via turn attention, it targeted RE/ERC rather than topic segmentation [5]. Xia et al. addressed inter-segment dependency gaps with PEN-NS, a parallel extraction network combining role-specific BiLSTMs and neighbor smoothing (NS) for boundary noise reduction [6]. It outperformed Transformer BERT by 3.2 F1 points on CSTS but relied on supervised label smoothing, unlike the LLM-guided unsupervised graph construction.

Domain-specific and downstream-task-aligned methods highlight segmentation's practical value. Liu et al. developed a joint segmentation-categorization model for clinical spoken dialogues (diabetes follow-ups) [7], building a domain dataset with 16 topics and using a Transformer-LSTM framework to achieve 0.7853 F1 in segmentation. This validates domain adaptation but is limited to clinical scenarios [7]. Ren et al. linked segmentation to dialogue summarization via NTSA, which assigns topics to turns while evaluating their importance—boosting BART's ROUGE scores on CSDS and SAMSUM datasets [8].

Previous works have verified the efficiency of unsupervised methods, the semantic modeling potential of pre-trained models, the non-local association capture ability of graph models, and the value of downstream adaptation, yet they still have common and specific limitations: unsupervised methodsgenerally lack global semantic association modeling; pre-trained models either suffer from performance bias due to focusing on similarity rather than topic segmentation or have high computational costs due to complex structures; downstream adaptation schemes have scenario limitationsor poor cross-task generalization. To address these, the proposed GCN-LLM framework in this study integrates GCN's global feature aggregation and LLM's zero-shot adaptation capability, maintaining high efficiency while improving topic modeling accuracy to achieve multi-scenario compatibility.

3 Method

3.1 Overview of the framework

The proposed GCN-LLM collaborative framework shown in Fig. 1 aims to enhance the efficiency of topic segmentation by improving dialogue segment representations while retaining the high-precision judgment capability of large models. The framework consists of four core layers as shown in Fig. 1: 1) Dialogue Data Layer provides raw dialogue text and basic semantic features (text embeddings, similarity scores). 2) Graph Structure Layer constructs a dialogue association graph based on LLM's topic continuity judgments. 3) GCN Model Layer generate enhanced embeddings by aggregating global associated features through graph convolution. 4) Integrated Application Layer performs initial screening using enhanced embeddings, with final segmentation results output by the LLM.

The core logic is: GCN only optimizes feature representation to improve initial screening accuracy without participating in final decision-making. This reuses the LLM's reasoning ability while addressing the locality limitation of original semantic features.

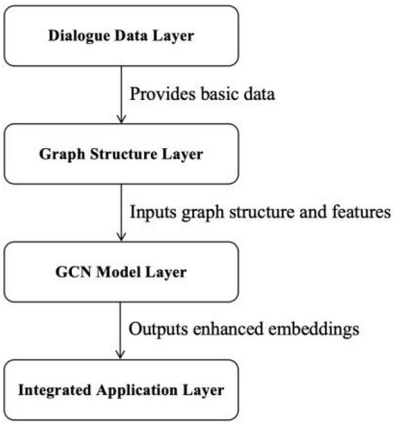

Fig. 1. Overall Structure of the GCN and Large Language Model Collaborative Framework (Picture credit: Original).

3.2 Key technical components

3.2.1 Dialogue data layer

This layer provides foundational data for subsequent graph construction and feature learning. Let the dialogue dataset be $D = \{(M_K, q_k)\}_{k=1}^{N}$, where $M_k = [m_1, m_2, \ldots, m_T]$ denotes the historical dialogue list of the k-th data (m_t is a single − turn dialogue segment), and q_k is the current query.

Text Embedding Generation: The pre-trained all-MiniLM-L12-v2 is used as the base semantic encoder to convert each dialogue segment m_t into a 384-dimensional vector:

$$\boldsymbol{h_t^{raw}} = \text{SimilarityModel}(m_t) \in \mathbb{R}^{384} \tag{1}$$

where $\boldsymbol{h_i^{raw}}$ is the raw text embedding, reflecting the segment's local semantic features.

LLM Judgment and Similarity Calculation: For any dialogue pair (m_i, m_j), topic continuity is determined via LLM inference, and the cosine similarity of raw embeddings is computed:

- Cosine similarity: Measures the local correlation strength of raw embeddings:

$$s_{ij} = \frac{\boldsymbol{h_i^{raw}} \cdot \boldsymbol{h_j^{raw}}}{\left\|\boldsymbol{h_i^{raw}}\right\| \left\|\boldsymbol{h_j^{raw}}\right\|} \tag{2}$$

- Topic continuity judgment: $\text{LLM}(\text{m}_\text{i}, \text{m}_\text{j}) \in \{0,1\}$, where $y_{ij} = 1$ indicates that the two dialogues are topic-continuous (LLM judges "yes").

3.2.2 Graph structure layer

The graph is constructed using LLM judgments as the "gold standard" to ensure alignment with real dialogue patterns.

Node Definition: Each single-turn dialogue segment m_t corresponds to a node v_t in the graph, with initial features as raw embeddings h_t^{raw}. Let the node set be $\mathrm{V} = \{\mathrm{v}_1, \mathrm{v}_2, \ldots, \mathrm{v}_\mathrm{N}\}$, and the feature matrix be $H^{raw} = [\boldsymbol{h_1^{raw}}, \boldsymbol{h_2^{raw}}, \ldots, \boldsymbol{h_N^{raw}}]^{\boldsymbol{T}} \in \mathbb{R}^{N\times 384}$

Edge Definition: Undirected edges are only established between dialogue pairs judged as topic-continuous by the:

$$\text{LLM: } (v_i, v_j) \in E \Leftrightarrow y_{ij} \tag{3}$$

Edge weights use raw cosine similarity s_{ij}, so the adjacency matrix $A \in \mathbb{R}^{N\times N}$ is define as:

$$A[i][j] = \begin{cases} s_{ij} \text{ if } y_{ij} = 1 \text{ and } i \neq j \\ 0 \text{ otherwise} \end{cases} \tag{4}$$

Dynamic Update Mechanism: When a new dialogue m_{new} is added, its raw embeding $\boldsymbol{h_{new}^{raw}}$ is gengerated, and $y_{new,t}$ and $s_{new,t}$ are computed with all existing nodes v_t. if $y_{new,t} = 1$, a new node v_{new} and edge (v_{new}, v_t)(with weight $s_{new,t}$) are added without reconstructing the global graph.

3.2.3 GCN model layer

GCN aggregates neighbor features (i.e., LLM-verified continuous dialogues) to generate enhanced embeddings fused with global correlations.

Inputs: Node feature matrix H^{raw} and adjacency matrix A. To avoid ignoring self-features, the adjacency matrix is superimpose with an identity matrix: $\widehat{\mathrm{A}} = \mathrm{A} + \mathrm{I}$ (I is the identity matrix).

Graph Convolution Operation: 1-2 GCN layers are used for feature aggregation, with each layer updating node features via weighted summation of neighbors. For layer l, the update formula for node v_i is:

$$\boldsymbol{h_i^{(l+1)}} = \mathrm{ReLU}\left(W^{(l)} \cdot \left(\boldsymbol{h_i^{(l)}} + \sum_{j\in N(i)} \widehat{\mathrm{A}}\,[\mathrm{i}][\mathrm{j}] \cdot \boldsymbol{h_j^{(l)}}\right) + \boldsymbol{b^{(l)}}\right) \tag{5}$$

where:

- $N(i)$ is the neighbor set of node i(including itself).
- $W^{(l)} \in \mathbb{R}^{384\times 384}$and $b^{(l)} \in \mathbb{R}^{384}$are learnable parameters of layer l.
- ReLU is used as the activation function to avoid gradient vanishing.

Output Features: After L layers ($L = 1\ or\ 2$), the enhanced embedding $\boldsymbol{h_i^{enhanced}} = \boldsymbol{h_i^{(L)}} \in \mathbb{R}^{384}$ is output, which integrates global features of all LLM-verified related dialogues.

3.2.4 Integrated application layer

The integration follows an "assist-not-replace" principle: the GCN only optimizes feature representation in the initial screening stage to narrow down the candidate range, while the final topic continuity judgment fully relies on the semantic reasoning capability of the large language model (LLM), ensuring high precision of core decisions.

GCN-Assisted Initial Screening: For the current query q, its raw embedding $\boldsymbol{h_q^{raw}}$,which integrates global correlation features. To filter out historically relevant dialogues potentially related to the query, the cosine similarity between $\boldsymbol{h_q^{enhanced}}$ and enhanced embedding $\boldsymbol{h_i^{enhanced}}$ for all historical dialogue segment is computed:

$$\widehat{s_q^i} = \frac{\boldsymbol{h_i^{enhanced}} \cdot \boldsymbol{h_q^{enhanced}}}{\left\|\boldsymbol{h_i^{enhanced}}\right\|\left\|\boldsymbol{h_q^{enhanced}}\right\|} \tag{6}$$

A similarity threshold $\theta = 0.4$, (consistent with the base screening logic) is set. Historical dialogue segments with $\widehat{s_q^i} \geq \theta$ are retained to form the initial candidate set C.

LLM Final Judgment: For each historical dialogue segment m_i in the candidate set C, the pair (m_i, q)(composed of m_i and the current query q) is the input to the LLM for topic continuity reasoning. The binary judgment result $y_{iq} \in \{0,1\}$ output by the LLM serves as the final decision basis: if $y_{iq} = 1$, m_i is determined to be topic-continuous with q and retained as valid historical context; if $y_{iq} = 0$, the segment is discarded.

3.3 Training and inference pipline

3.3.1 Offline training (GCN model)

The GCN model is trained using unsupervised contrastive learning, with dialogue association relationships verified by the LLM to construct positive and negative samples, eliminating the need for manually annotated supervision signals.

Positive/Negative Sample Construction:

- Positive pairs: All node pairs satisfying $(v_i, v_j) \in E$ (i.e., dialogue pairs judged "topic-continuous" by the LLM), which naturally reflect the association patterns of real dialogues;
- Negative pairs: Randomly sampled from non-edge node pairs $(v_i, v_j) \notin E$ (dialogue pairs judged "topic-discontinuous" by the LLM). To avoid class imbalance, the ratio of positive to negative samples is controlled at 1:5.

Loss Function: Contrastive Loss is adopted to optimize model parameters. The core goal is to minimize the distance between enhanced embeddings of positive pairs and maximize the distance of negative pairs, enabling the GCN to learn more accurate association feature representations:

$$\mathcal{L} = -\sum_{(i,j)\in E} \log\left(\frac{\exp\left(\mathrm{sim}\left(\boldsymbol{h_i^{enhanced}}, \boldsymbol{h_j^{enhanced}}\right)/\tau\right)}{\sum_{k\notin N(i)} \exp\left(\mathrm{sim}\left(\boldsymbol{h_i^{enhanced}}, \boldsymbol{h_j^{enhanced}}\right)/\tau\right)}\right) \tag{7}$$

Here, $\mathrm{sim}(\cdot,\cdot)$ denotes cosine similarity calculation, $\tau = 0.7$ is the temperature coefficient (used to adjust the smoothness of the similarity distribution), and $N(i)$ is the neighbor set of node i(including itself).

3.3.2 Online inference (topic segmentation)

1. Input Reception: Receive the user's current query q and the corresponding historical dialogue list $M = [m_1, m_2, \ldots, m_i]$(where m_t is a single-turn historical dialogue segment).
2. Embedding Generation: Generate raw embeddings for all m_t and q via the base semantic encoder, then feed these raw embeddings into the pre-trained GCN model to obtain the corresponding enhanced embeddings $h_t^{enhanced}$ (historial dialogues) and $\boldsymbol{h_q^{enhanced}}$ (current query);

3. Initial Screen: Compute the similary $\widehat{s_q^i}$ between $\boldsymbol{h_t^{enhanced}}$ and each $\boldsymbol{h_q^{enhanced}}$,retain m_t with $\widehat{s_q^i} \geq$0.4, and form the candidate set $C = \{m_t | \widehat{s_q^i} \geq 0.4\}$;

4. Final Judgment: For each m_t in the candidate set C, input (m_t, q) to the LLM for topic continuity reasoning. Based on the y_{tq} output by the LLM, filter out segments with $y_{tq} = 1$ as the final valid historical context;

5. Dynamic Graph Update: Create a new node v_q for the dialogue segment corresponding to the current query q(with its raw embedding as the feature). If the LLM judges m_t to be continuous with q(i.e., $y_{tq} = 1$), add a new edge between the new node v_q and the historical node v_t, with the edge weight set to $\widehat{s_{tq}^i}$(no need to reconstruct the global graph structure).

4 Experimental results

4.1 Experimental dataset

This experiment uses the public dialogue topic segmentation dataset dialseg_711.json as the sole test set to ensure the reproducibility and comparability of experimental results (dialseg_711.json dataset) Each sample in the dataset is a JSON object, containing three core fields:

dial_id: An integer variable that uniquely identifies each dialogue (e.g., 0, 1, 2, etc.);

utterances: A string array that records each utterance in the dialogue in chronological order (including alternating content from users and systems, but without separate role labels);

segments: An integer array where each element represents the "number of sentences" in a topic segment. The sum of all elements is exactly equal to the length of utterances, and segment boundaries can be directly determined by accumulating the element values.

Taking the sample with dial_id=0 as an example: the dialogue contains 24 utterances, and its segments field is [4, 6, 6, 4, 4]. The sum of these elements equals 24, so the dialogue can be divided into 5 topic segments, with corresponding boundaries: "Sentences 1–4, 5–10, 11–16, 17–20, 21–24". The segment annotation logic of this dataset aligns with the topic flow rules of real dialogues, providing reliable ground truth for evaluating the segmentation accuracy of models.

4.2 Comparative methods and evaluation metrics

4.2.1 Comparative methods

To verify the superiority of the proposed GCNLLM collaborative framework, four mainstream dialogue topic segmentation methods in recent years are selected as baselines, with details as follows:

1) DialSTART Framework [9]: An unsupervised framework that initializes the topic encoder with SimCSE and uses NSPBERT as the coherence encoder. It optimizes representation learning through Neighbor Utterance Matching (NUM) and Relevance Modeling (RM) tasks, and realizes segmentation by combining TextTiling.

2) UtterancePair Coherence Scoring Method [10]: An unsupervised method that constructs positive and negative samples based on the assumption that "adjacent utterances are more coherent". It finetunes a coherence scoring model with NSPBERT and realizes segmentation using the depth score threshold of TextTiling, supporting multilingual scenarios;

3) Discourse Dependency Injection Method [11]: A supervised model that adopts a hierarchical attention BiLSTM as the basic architecture. It injects intersentence discourse dependency graphs generated by a retrained SentFirst parser and aggregates neighbor information through GAT to improve the ability to capture topic consistency;

4) CLDA Model [12]: An improved method based on LDA that introduces Contextual Wordpair Topic Distribution (CWTD) and ContextAware Topic Distribution (CTD). It optimizes segment merging by combining the RTM algorithm to balance topic coherence and saliency.

4.2.2 Evaluation metrics

Classic evaluation metrics for dialogue topic segmentation are used in the experiment, including Precision, Recall, and F1score:

Precision: Measures the proportion of results predicted by the model as "topiccontinuous segments" that are actually real segments;

Recall: Measures the proportion of real topic segments that are successfully predicted as "continuous" by the model;

F1score: The harmonic mean of Precision and Recall, comprehensively reflecting the overall performance of the model and avoiding bias from a single metric.

4.3 Experimental results and analysis

4.3.1 The performance comparison of various models

The evaluation results of all methods on the dialseg_711.json dataset are shown in Table 1.

Table 1. Performance Comparison of Different Methods on dialseg_711.json Dataset.

Method	Precision	Recall	F1-score
Proposed GCN-LLM	0.8207	0.8589	0.8393
DialSTART	0.5907	0.6128	0.5999
Utterance-Pair Coherence Scoring	0.6727	0.6613	0.6606
Discourse Dependency Injection	0.3359	0.2084	0.2572
C-LDA	0.1586	0.9939	0.2712

4.3.2 Results analysis

As can be seen from Table 1, the proposed GCNLLM collaborative framework significantly outperforms the four comparative methods in all evaluation metrics, with core advantages reflected in three aspects:

1. Leading Comprehensive Performance (F1score): The F1score of the proposed method reaches 0.8393, which is about 27% higher than the current bestperforming comparative method (UtterancePair Coherence Scoring, F1=0.6606). This proves that the global correlation capture of GCN and the accurate reasoning of LLM form effective collaboration—GCN solves the "local correlation limitation" of traditional semantic similarity methods, while LLM avoids the insufficient reasoning accuracy of pure graph models or unsupervised models [9, 10].

2. Balanced Precision and Recall: Both the Precision (0.8207) and Recall (0.8589) of the proposed method are at a high level with a small gap, indicating that the model can not only effectively screen real topiccontinuous segments (low misjudgment) but also reduce the missed

detection of valid segments (low omission). In contrast, although the CLDA model has a high Recall of 0.9939, its Precision is only 0.1586, suffering from serious "excessive segment merging" problems; the Discourse Dependency Injection method has both Precision and Recall below 0.4, reflecting the poor adaptability of supervised models to unannotated dialogue data [11, 12].

3. Adaptability to Unsupervised Scenarios: Compared with the unsupervised methods (DialSTART and UtterancePair Coherence Scoring), the proposed method does not require manually designing complex rules for generating positive and negative samples (e.g., pseudosegment screening for NUM tasks, adjacent utterance assumptions). It only constructs the graph structure through the "topic continuity judgment" of LLM, which not only reduces the noise interference of unsupervised learning but also simplifies the training process [9, 10].

5 Conclusion

This paper addresses the "difficulty in balancing efficiency and precision" in dialogue topic segmentation by proposing a collaborative solution of GCN and LLM. The core logic of the framework is "GCN enhances feature representation, LLM is responsible for final decision-making": it uses LLM to construct a graph structure that aligns with real dialogue patterns, and GCN makes up for the local correlation defect of semantic similarity methods while avoiding the high computational cost of LLM. On the dialseg_711.json dataset, the method achieves an F1-score of 0.8393, which is about 27% higher than the current optimal method, with balanced precision (0.8207) and recall (0.8589), and strong adaptability to unsupervised scenarios. This study successfully resolves the contradiction between the locality of traditional methods and the inefficiency of LLM; its dynamic graph update mechanism supports real-time dialogue systems, providing a practical path for dialogue topic segmentation in scenarios such as RAG and intelligent customer service. However, the method still has some limitations. LLM invocation consumes high computing power, with more obvious costs for ultra-long dialogues; small/medium teams or edge devices face high local deployment thresholds, while cloud reliance brings privacy and latency risks. Future research can explore LLM lightweighting (e.g., efficient fusion of knowledge distillation and GCN) and develop streamlined local deployment frameworks to expand applications with edge computing.

References

1. S. Park, J. Seo, J. Lee, Unsupervised dialogue topic segmentation in hyperdimensional space. arXiv preprint arXiv:2308.10464 (2023)
2. A. Maraj, Contextual topics: advancing text segmentation through pre-trained models and contextual keywords (2024)
3. I. Ghinassi, L. Wang, C. Newell, M. Purver, Comparing neural sentence encoders for topic segmentation across domains: not your typical text similarity task. PeerJ Comput. Sci. 9, e1593 (2023)
4. H. Yu, C. Deng, Q. Zhang, J. Liu, Q. Chen, W. Wang, Improving long document topic segmentation models with enhanced coherence modeling. arXiv preprint arXiv:2310.11772 (2023)
5. B. Lee, Y.S. Choi, Graph based network with contextualized representations of turns in dialogue. arXiv preprint arXiv:2109.04008 (2021)
6. J. Xia, C. Liu, J. Chen, Y. Li, F. Yang, X. Cai, H. Wang, Dialogue topic segmentation via parallel extraction network with neighbor smoothing, in Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (2022), pp. 2126–2131

7. Z. Liu, S.U.M. Salleh, H.C. Oh, P. Krishnaswamy, N. Chen, Joint dialogue topic segmentation and categorization: A case study on clinical spoken conversations, in Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track (2023), pp. 185–193
8. Z. Ren, A Novel Topic Segmentation Approach for Enhanced Dialogue Summarization (2024)
9. H. Gao, R. Wang, T.E. Lin, Y. Wu, M. Yang, F. Huang, Y. Li, Unsupervised dialogue topic segmentation with topic-aware utterance representation. arXiv preprint arXiv:2305.02747 (2023)
10. L. Xing, G. Carenini, Improving unsupervised dialogue topic segmentation with utterance-pair coherence scoring. arXiv preprint arXiv:2106.06719 (2021)
11. L. Xing, P. Huber, G. Carenini, Improving topic segmentation by injecting discourse dependencies. arXiv preprint arXiv:2209.08626 (2022)
12. W. Li, T. Matsukawa, H. Saigo, E. Suzuki, Context-aware latent dirichlet allocation for topic segmentation, in Pacific-Asia Conference on Knowledge Discovery and Data Mining (Springer, Cham, 2020), pp. 475–486