RAPTM: RETRIEVAL-AUGMENTED PROMPTING FOR SHORT-TEXT TOPIC MODELING

Rui Wang[1], *Ziang Li*[1], *Shiyu Sun*[1], *Guozi Sun*[1], *Haiping Huang*[1,*], *Jialin Yu*[2], *Yuxiang Zhou*[3], *Philip Torr*[2]

[1]School of Computer Science, Nanjing University of Posts and Telecommunications
[2]Department of Engineering Science, University of Oxford
[3]School of Electronic Engineering and Computer Science, Queen Mary University of London

ABSTRACT

Advances in short-text topic modeling often utilize large language models, either to expand the data or to summarize topics, to mitigate data sparsity. However, expansion can hallucinate irrelevant content, while summarization often introduces out-of-vocabulary (OOV) terms that do not appear in the corpus. In this paper, we propose a novel **R**etrieval-**A**ugmented **P**rompting **T**opic **M**odel (RAPTM), which addresses these limitations through retrieval-augmented prompting and a contextualized topic-word mining procedure based on dynamic word representations from a language model. Extensive experiments on two news headline datasets demonstrate that our approach outperforms state-of-the-art models in both topic coherence and diversity metrics. Moreover, its superior performance in clustering shows that RAPTM can produce informative and well-separated topic representations.

***Index Terms*—** Short Text Modeling, Retrieval Augmented Generation, Large Language Models, Prompting

1. INTRODUCTION

Short-text data such as news headlines, tweets, and user reviews are widely available and valuable for discovering emerging topics, trends, and opinions. However, short texts are inherently sparse and lack sufficient context, making topic modeling on such data particularly challenging [1]. Traditional topic models such as Latent Dirichlet Allocation (LDA) [2] and more recent neural variants like the Embedding Clustering Regularized Topic Model (ECRTM) [3], often fail to capture meaningful topics in this setting due to limited word co-occurrence information.

To address this limitation, a range of approaches have been proposed. Yan et al. relax the topic mixture assumption of LDA by modeling the generation of word-co-occurrence patterns directly and propose the probabilistic Biterm Topic Model (BTM) [4]. In contrast, Nguyen et al. integrate global clustering context information into the modeling process and propose the Global Clustering Contexts Topic Model (GloCOM) [5]. Researchers have also explored leveraging contextualized neural representations, for example the Disentangled Contextualized-neural Topic Model (DisCTM) [6]. More recently, with the advent of large language models (LLMs),

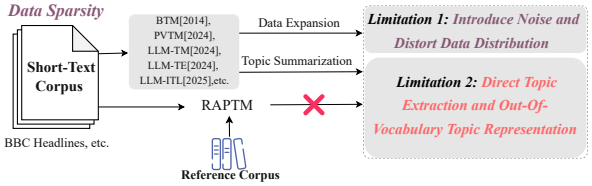


Fig. 1. Key limitations of different strategies for combating data sparsity in advanced short-text topic models.

two complementary strategies have emerged. The first is *expansion-based approaches*, where LLMs are employed to extend short texts into longer sequences; for example, Akash et al. introduce the Prefix-tuned Variational Topic Model (PVTM) [7], based on an autoencoding framework [8]. The second is *direct topic extraction approaches*, which rely on prompt engineering to guide LLMs in generating topic words or summaries from the raw corpus; representative examples include the LLM-driven Topic Extraction (LLM-TE) [9] model and the LLM-based Topic Model (LLM-TM) [10].

However, these approaches still face the following limitations: 1). Expansion-based approaches fail to ensure semantic consistency between the original short texts and their extended versions, often introducing noisy information and distorting the original corpus distribution. 2). Direct topic extraction approaches typically operate on sparse original inputs and frequently generate out-of-vocabulary (OOV) words, which reduces the coherence and interpretability of the resulting topics.

Thus, to address the above limitations, we propose the novel **R**etrieval-**A**ugmented **P**rompting **T**opic **M**odel (RAPTM) for short text topic modeling. Specifically, RAPTM first produces semantically relevant demonstrations by retrieving documents from an external reference corpus to enrich the local context and generates the corresponding topical summaries for each document using the retrieval-augmented summarization prompt. *This step mitigates the noise and semantic drift commonly introduced by expansion-based approaches*. It then analyzes the topical summaries—which contain only the relevant topical information, with extraneous words from the original text filtered out—and iteratively infers a set of topic names using a batch-wise topic-name generation prompt. Subsequently, it aligns each document with the inferred topic name that shares the highest semantic relevance using a topic alignment prompt and constructs

* Corresponding author.

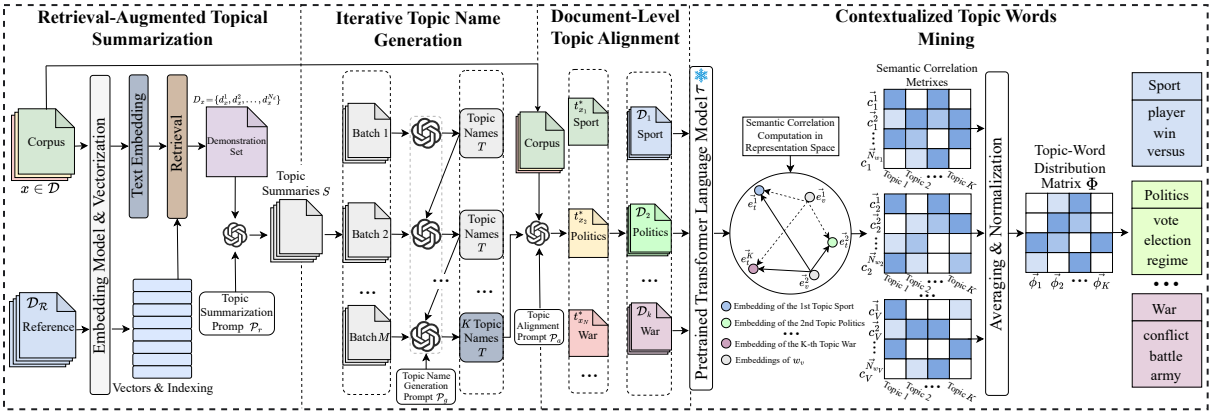


Fig. 2. The framework of the Retrieval-Augmented Prompting Topic Model (RAPTM).

the topic-specific corpus. Finally, RAPTM employs a transformer [11] to produce contextualized word representations for the corpus and designs a contextualized topic mining procedure, *which mitigates the OOV issue commonly observed in direct topic extraction approaches and improves topic quality*.

The main contributions of this work are summarized as:

- We propose the novel **R**etrieval-**A**ugmented **P**rompting **T**opic **M**odel (RAPTM), which is, to our best knowledge, the first attempt at incorporating the retrieval-augmented prompting for short text modeling.
- RAPTM enriches the local context of short texts through retrieval-augmented prompting, which enhances the quality of topical summaries and improves the accuracy of subsequent topic-name generation and topic alignment. Moreover, it incorporates contextualized word representations into the topic-word mining process, thereby improving topic interpretability.
- Experimental results on two headline datasets confirm the effectiveness of RAPTM for short-text modeling, as measured by topic coherence and diversity metrics. Moreover, its superior performance in text clustering shows that RAPTM produces more informative document–topic distributions than state-of-the-art baselines.

2. METHODOLOGY

As depicted in Figure 2, our proposed RAPTM consists of four main components: 1). *Retrieval-Augmented Topical Summarization*: For each document x in the corpus $\mathcal{D}$, this module first retrieves a set of semantically relevant demonstrations $D_x = \{d_x^1, d_x^2, ..., d_x^{N_d}\}$ [1] from the external reference corpus $\mathcal{D}_\mathcal{R}$ to enrich its local context. It then applies a retrieval-augmented prompt $\mathcal{P}_r$ to generate a concise topical summary s_x that conveys the key semantic information of the document, thereby constructing the summary set $S=\{s_{x_1}, s_{x_2}, ..., s_{x_N}\}$, N denotes the number of documents in corpus $\mathcal{D}$. 2). *Iterative Topic Name Generation*: Given the topical summaries, this module employs a topic name generation prompt $\mathcal{P}_g$ to iteratively infer and refine topic names set $T = \{t_1, t_2, ..., t_K\}$. 3). *Document-Level Topic Alignment*: This module uses a topic alignment prompt $\mathcal{P}_a$ to align each document $x \in \mathcal{D}$ with the topic name $t_x^* \in T$ that exhibits the greatest similarity, partitioning the corpus $\mathcal{D}$ into topic-specific subsets $\mathcal{D}_k$ ($k \in \{1, 2, ..., K\}$). 4). *Contextualized Topic Word Mining*: For the k-th topic, it generates the topic-word distribution $\vec{\phi}_k$ based on contextualized word representations of documents in $\mathcal{D}_k$, produced by the transformer language model $\mathcal{T}$. We will introduce the details of these components below, repsectively.

2.1. Retrieval-Augmented Topical Summarization

To enrich the local context and distill key semantic information from documents, RAPTM retrieves semantically similar demonstrations from an external reference corpus $\mathcal{D}_\mathcal{R}$ and generates a topical summary for each document using the retrieval-augmented prompt $\mathcal{P}_r$.

Concretely, given a corpus $\mathcal{D}$ and the reference corpus $\mathcal{D}_\mathcal{R}$, we first perform vectorization and indexing for documents in $\mathcal{D}_\mathcal{R}$ with the faiss library [2]. Then, for each document x in the corpus $\mathcal{D}$, we retrieve and select top N_d documents that share highest semantic similarity with x from $\mathcal{D}_\mathcal{R}$ to form the demonstration set $D_x = \{d_x^1, d_x^2, ..., d_x^{N_d}\}$. Subsequently, we utilize the retrieval-augmented topic summarization prompt $\mathcal{P}_r$ to analyze its textual content and produce the corresponding topical summary s_x using:

$$s_x = \mathcal{P}_r(x, D_x) \tag{1}$$

where the detailed design of $\mathcal{P}_r$ is shown in the Figure 3 (a), which follows the OpenAI released API [3]. Obviously, we could construct the corpus-level summary set S, which will be further used to generate topic names.

2.2. Iterative Topic Name Generation

To alleviate the impact of non-topic words in the original text and enable more accurate topic inference, we generate topic names by analyzing topical summaries, from which noisy information has already been filtered out. Moreover, topic names are produced in a batch-wise manner to address the input-length limitations of LLMs.

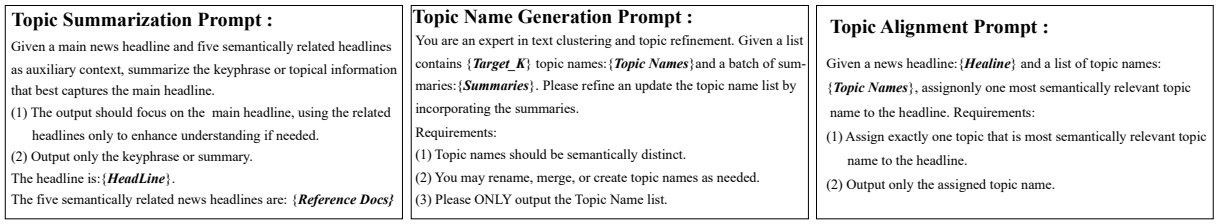
Specifically, given a batch of topical summaries $S' \subset S$, to infer or refine a set of K topic names $T = \{t_1, t_2, ..., t_K\}$ [4],

[1] N_d is the number of demonstrations for each document and is set to 5.

[2] https://pypi.org/project/faiss/

[3] https://cookbook.openai.com/

[4] T is an empty set for the first batch.



(a) Topic Summarization Prompt $\mathcal{P}_r$ **(b)** Topic Name Generation Prompt $\mathcal{P}_g$ **(c)** Topic Alignment Prompt $\mathcal{P}_a$

Fig. 3. The detailed prompting of topic summarization (a), topic name generation (b), and topic assignment (c).

we employ the topic name generation prompt $\mathcal{P}_g$ with:

$$T = \mathcal{P}_g(S', K, T) \tag{2}$$

where K denotes the topic number, the topic name set T could be updated and refined iteratively, t_k represents name of the k-th topic, details of $\mathcal{P}_g$ is shown in Figure 3 (b).

2.3. Document-Level Topic Alignment

Given the LLM-generated topic names T, this module aligns each document with the most semantically relevant topic name and constructs the topic-specific corpus $\mathcal{D}_k$ for subsequent topic-word extraction.

In detail, for each document x in the corpus $\mathcal{D}$, we emply a topic alignment prompt $\mathcal{P}_a$ to align it with the topic that shares the highest similarity using:

$$t_x^* = \mathcal{P}_a(x, T) \tag{3}$$

where t_x^* denotes the LLM-responded topic name, details of $\mathcal{P}_a$ is shown in Figure 3 (c). Thus, we could partition the corpus $\mathcal{D}$ according to topic assignments and form the topic-specific corpus $\mathcal{D}_k$ ($k \in \{1, 2, ..., K\}$).

2.4. Contextualized Topic Words Mining

Using the topic-specific corpus $\mathcal{D}_k$ ($k \in \{1, 2, ..., K\}$), we inject the contextualized knowledge from the transformer into the topic mining procedure to improve topic interpretability.

Concretely, for each topic t_k, we first generate its topic embedding $\vec{e}_t^k$ with the formulation:

$$\vec{e}_t^k = \frac{1}{N_k} \sum\nolimits_{n=1}^{N_k} \mathcal{T}(x_n^k) \tag{4}$$

where x_n^k denotes the n-th document in $\mathcal{D}_k$, $\mathcal{T}(x_n^k)$ is the contextualized representation of x_n^k, which is computed by averaging the contextualized word representations of all the words in the document. N_k denotes the number of documents in $\mathcal{D}_k$.

Then, for the v-th ($v \in \{1, 2, ..., V\}$) word w_v in the vocabulary, we first generate contextualized word representations of all its appearances in $\mathcal{D}_k$ using a transformer $\mathcal{T}$ [11] and collect them in the word representation matrix $E_v \in \mathbb{R}^{H \times N_{w_v}}$. Here, H is the dimension of transformer representation, N_{w_v} denotes the number of times that word w_v appears in $\mathcal{D}_k$. Thus, the semantic correlation $c_{v,k}$ between word w_v and the k-th topic t_k could be calculated with:

$$c_{v,k} = \frac{1}{N_{w_v}} \sum\nolimits_{n=1}^{N_{w_v}} c_{v,k}^n = \frac{1}{N_{w_v}} \sum\nolimits_{n=1}^{N_{w_v}} \cos(\vec{e}_v^n, \vec{e}_t^k) \tag{5}$$

where $\vec{e}_v^n$ denotes the contextualized word representation of word w_v for its n-th appearence, $c_{v,k}^n$ is the local semantic correlation between the k-th topic and the n-th appearance of w_v. Similarly, we could compute the semantic correlations $\vec{c}_k$ between the k-th topic and words in the vocabulary. Thus, the topic-word distribution $\vec{\phi}_k$ of the k-th topic is computed with:

$$\vec{\phi}_k = \text{softmax}(\vec{c}_k) \tag{6}$$

where softmax$(\cdot)$ is the softmax normalization, $\vec{\phi}_k$ is the k-th volumn of the topic-word distribution matrix $\Phi \in \mathbb{R}^{V \times K}$.

3. EXPERIMENTS

3.1. Experimental Setups

Datasets: We use two headline datasets, which are the TagMyNews [5] and HuffPostNews [6], to evaluate the RAPTM's performance. For the reference corpus, we utilize the GoogleNews [7], which contains nearly 600k newspaper headlines. The details of the preprocessed datasets are listed in Table 1:

Table 1. The statistics of the processed datasets.

Dataset	# Train	# Test	Vocabs	Avg. Length	# Labels
TagMyNews	11,003	3,000	7,804	4.66	7
HuffPostNews	14,589	3,000	10,144	5.61	42

Baselines: We choose the following state-of-the-art topic modeling approaches as baselines: LDA [2], BTM [4], ECRTM [12], CWTM [13], GloCOM [5], DisCTM [6], HiCOT [14], LLM-ITL [15], LLM-TE [9] and LLM-TM [10].
Evaluation Metrics: We follow [16] and choose four widely used topic coherence metrics (C_P, C_A, $NPMI$ and UCI), calculated using the released Palmetto [8] library. Besides, we follow [17] and utilize the Unique Term Rate (UTR) metric to assess topic diversity. Additionally, for clustering performance evaluation, we follow [18, 19] and employ Purity and Normalized Mutual Information (NMI), using TopMost [9].
Implementation Details: We utilze the GPT-3.5-Turbo[10] for topical summarization prompt $\mathcal{P}_r$, topic name generation prompt $\mathcal{P}_g$ and the topic alignment prompt $\mathcal{P}_a$. For the embedding model employed in the summarization stage, we utilize the BGE-M3 [11]. In the topic mining phase, we utilize the MPNet [20] [12] to generate contextualized representations.

[5]https://github.com/cskarthik93/TagMyNews-Classification
[6]https://www.kaggle.com/datasets/rmisra/news-category-dataset
[7]https://www.kaggle.com/datasets/crxxom/daily-google-news
[8]https://github.com/dice-group/Palmetto
[9]https://github.com/BobXWu/TopMost
[10]https://openai.com/index/
[11]https://huggingface.co/BAAI/bge-m3
[12]https://huggingface.co/sentence-transformers/all-mpnet-base-v2

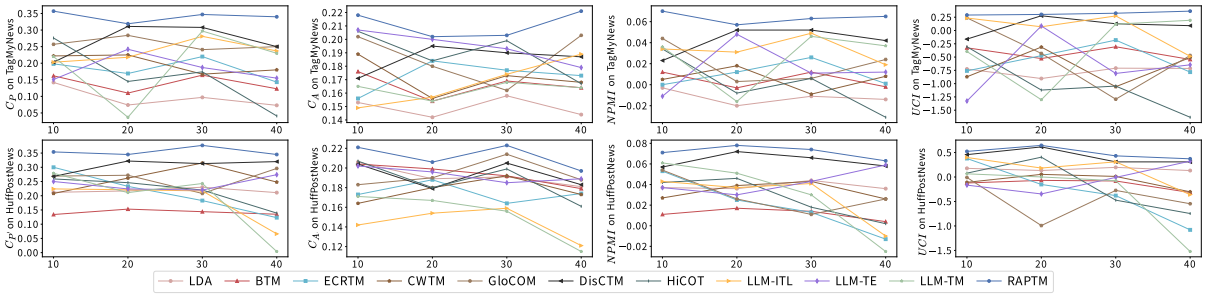


Fig. 4. The comparison of topic coherence vs. topic numbers on both datasets.

TagMyNews

Models	LDA	BTM	ECRTM	CWTM	GloCOM	DisCTM	HiCOT	LLM-ITL	LLM-TE	LLM-TM	RAPTM
C_P	0.096	0.140	0.182	0.199	0.258	0.268	0.159	0.235	0.184	0.196	**0.341**
C_A	0.149	0.166	0.172	0.172	0.187	0.186	0.189	0.167	0.195	0.163	**0.211**
$NPMI$	-0.012	0.005	0.010	0.005	0.019	0.042	0.001	0.033	0.015	0.026	**0.064**
UCI	-0.763	-0.426	-0.549	-0.688	-0.489	0.083	-1.026	0.023	-0.674	-0.347	**0.321**
UTR	0.586	0.599	0.710	0.701	0.814	0.852	0.715	0.564	0.813	0.774	**0.870**

HuffPostNews

Models	LDA	BTM	ECRTM	CWTM	GloCOM	DisCTM	HiCOT	LLM-ITL	LLM-TE	LLM-TM	RAPTM
C_P	0.217	0.142	0.210	0.258	0.261	0.306	0.216	0.182	0.242	0.184	**0.355**
C_A	0.192	0.194	0.175	0.177	0.194	0.193	0.187	0.144	0.193	0.152	**0.209**
$NPMI$	0.038	0.011	0.019	0.034	0.029	0.063	0.027	0.028	0.042	0.029	**0.071**
UCI	0.160	-0.147	-0.308	-0.086	-0.455	0.461	-0.066	0.133	-0.049	-0.383	**0.493**
UTR	0.764	0.480	0.676	0.788	0.843	0.879	0.628	0.574	0.723	0.829	**0.921**

Table 2. Average coherence and diversity values on datasets.

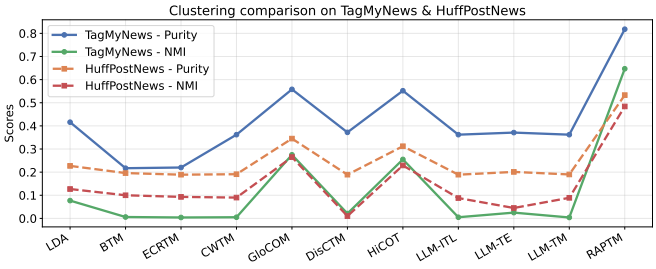


Fig. 5. Text clustering performance comparison.

3.2. Experimental Results

Topic Quality Evaluation: We conduct extensive experiments on TagMyNews and HuffPostNews datasets with different topic number settings [10, 20, 30, 40]. Average topic coherences (C_P, C_A, $NPMI$ and UCI) and diversity (UTR) values are listed in Table 2. Additionally, to expore if the variations of topic number settings will impact the modeling performance, we present a coherence vs. topic number comparison in Figure 4. From the statistics, we could observe that our proposed RAPTM outperforms the state-of-the-art baseline considering coherence and diversity metrics across different topic number settings on both datasets. This may be attributed to the following factors: 1). Retrieval-augmented topical summarization enriches the local context of short-length documents and helps to produce informative topic summaries. 2). Inferring topic names from informative topical summaries effectively captures the semantic concepts within the corpus, thereby enhancing the topic mining process. 3). Incorporating contextualized knowledge from a transformer helps to improve topic interpretability. Meanwhile, topics are mined from the topic-specific corpus $\mathcal{D}_k$, which addresses the OOV word limitation.

Text Clustering: To examine RAPTM's performance on downstream tasks, we conduct text clustering experiments on the TagMyNews and HuffPostNews datasets. The numbers of topics are set to 7 and 42, respectively, corresponding to the number of categories. Comprision results on Purity [21] and NMI [22] metrics are listed in Figure 5. It reveals that RAPTM produces more informative document-topic distributions than baselines and accurately captures semantic structure within the corpus.

Ablation Studies: As retrieval-augmented summarization prompt $\mathcal{P}_r$, topic name generation prompt $\mathcal{P}_g$ and transformer $\mathcal{T}$ are crucial for modeling performance, we conduct ablation studies on these components on TagMyNews with topic number set to 10, and results are presented in Table 3. Here, 'RAPTM-$\mathcal{P}_r$' denotes the exclusion of the retrieval-augmentation mechanism, 'RAPTM-$\mathcal{P}_g$' indicates that we use the K-Means clustering to substitute $\mathcal{P}_g$ for topic name generation. 'RAPTM-MiniLM' means that it uses MiniLM embedding [13] in the topic mining phase. Also, we only present DisCTM's value for clarity due to its superior performance among baselines. We observe that RAPTM and its variants sustainably outperforms the competitive DisCTM in terms of coherence and diversity metrics.

Models	C_P	C_A	$NPMI$	UCI	UTR
DisCTM	0.203	0.171	0.023	-0.162	0.980
RAPTM-$\mathcal{P}_r$	0.321	0.206	0.056	0.148	0.970
RAPTM-$\mathcal{P}_g$	0.342	0.213	0.068	0.224	**1.000**
RAPTM-MiniLM	0.343	**0.222**	0.070	0.265	**1.000**
RAPTM	**0.357**	0.218	**0.071**	**0.291**	**1.000**

Table 3. Results of ablation studies on TagMyNews.

4. CONCLUSION

In this paper, we propose the Retrieval-Augmented Prompting Topic Model (RAPTM), a novel approach that integrates retrieval-augmented prompting to address a key challenge in short-text modeling: data sparsity. To further enhance topic interpretability, it incorporates contextualized knowledge from transformers into the topic-mining process. Extensive experimental results on two datasets validate the effectiveness of RAPTM in discovering high-quality topics from headlines, as measured by topic coherence and diversity metrics. Moreover, its superior performance in text clustering demonstrates that RAPTM is able to produce more informative and well-separated document–topic representations.

[13] https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

5. ACKNOLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (Grant No. 62102192), the fellowship of China Postdoctoral Science Foundation (Grant No. 2022M710071), the Fundamental Research Funds for the Central Universities of China (Grant Nos. PA2025IISL0107 and aiia-24-01) and the Innovation and Entrepreneurship Program of Jiangsu Province (Grant No. JSSCBS20210530). JY acknowledges support from Microsoft Ltd.

6. REFERENCES

[1] Jipeng Qiang, Zhenyu Qian, Yun Li, Yunhao Yuan, and Xindong Wu, "Short text topic modeling techniques, applications, and performance: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 3, pp. 1427–1445, 2022.

[2] David M. Blei, Andrew Y. Ng, and Michael I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.

[3] Xiaobao Wu, Xinshuai Dong, Thong Thanh Nguyen, and Anh Tuan Luu, "Effective neural topic modeling with embedding clustering regularization," in *ICML, 23-29 July, Hawaii, USA*, 2023, pp. 37335–37357.

[4] Xiaohui Yan, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng, "A biterm topic model for short texts," in *WWW, Brazil, May 13-17*, 2013, pp. 1445–1456.

[5] Quang Duc Nguyen, Tung Nguyen, Duc Anh Nguyen, Linh Ngo Van, Sang Dinh, and Thien Huu Nguyen, "Glocom: A short text neural topic model via global clustering context," in *NAACL-HLT, Albuquerque, New Mexico, USA, April 29 - May 4*, 2025, pp. 1109–1124.

[6] Rui Wang, Xing Liu, Yanan Wang, Shuyu Chang, Yuanzhi Yao, and Haiping Huang, "Mining topics towards chatgpt using a disentangled contextualized-neural topic model," in *WSDM, Hannover, Germany, March 10-14*, 2025, pp. 539–548.

[7] Pritom Saha Akash and Kevin Chen-Chuan Chang, "Enhancing short-text topic modeling with llm-driven context expansion and prefix-tuned vaes," in *Findings of EMNLP, Miami, USA, November 12-16*, 2024, pp. 15635–15646.

[8] Diederik P. Kingma and Max Welling, "Auto-encoding variational bayes," in *ICLR, Banff, AB, Canada, April 14-16*, 2014.

[9] Yida Mu, Chun Dong, Kalina Bontcheva, and Xingyi Song, "Large language models offer an alternative to the traditional approach of topic modelling," in *LREC/COLING 20-25 May, Torino, Italy*, 2024, pp. 10160–10171.

[10] Tomoki Doi, Masaru Isonuma, and Hitomi Yanaka, "Topic modeling for short texts with large language models," in *ACL, Bangkok, Thailand, August 11-16*, 2024, pp. 21–33.

[11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *NAACL-HLT, Minneapolis, MN, USA, June 2-7*, 2019, pp. 4171–4186.

[12] Xiaobao Wu, Chunping Li, Yan Zhu, and Yishu Miao, "Short text topic modeling with topic distribution quantization and negative sampling decoder," in *EMNLP, Online, November 16-20*, 2020, pp. 1772–1782.

[13] Zheng Fang, Yulan He, and Rob Procter, "CWTM: leveraging contextualized word embeddings from BERT for neural topic modeling," in *LREC/COLING, 20-25 May, 2024, Torino, Italy*, 2024, pp. 4273–4286.

[14] Hoang Tran Vuong, Tue Le, Tu Vu, Tung Nguyen, Linh Ngo Van, Sang Dinh, and Thien Huu Nguyen, "Hicot: Improving neural topic models via optimal transport and contrastive learning," in *Findings of ACL, Austria, July 27-August 1*, 2025, pp. 13894–13920.

[15] Xiaohao Yang, He Zhao, Weijie Xu, Yuanyuan Qi, Jueqing Lu, Dinh Phung, and Lan Du, "Neural topic modeling with large language models in the loop," in *ACL, Austria, July 27 - August 1*, 2025, pp. 1377–1401.

[16] Michael Röder, Andreas Both, and Alexander Hinneburg, "Exploring the space of topic coherence measures," in *WSDM, Shanghai, China, February 2-6*. 2015, pp. 399–408, ACM.

[17] Rui Wang, Deyu Zhou, Haiping Huang, and Yongquan Zhou, "MIT: mutual information topic model for diverse topic extraction," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 36, no. 2, pp. 2523–2537, 2025.

[18] Liang Du, Yukai Shi, Yan Chen, Peng Zhou, and Yuhua Qian, "Fast and scalable incomplete multi-view clustering with duality optimal graph filtering," in *MM, Melbourne, VIC, Australia, 28 October- 1 November*, 2024, pp. 8893–8902.

[19] Xiaobao Wu, Thong Nguyen, Delvin Zhang, William Yang Wang, and Anh Tuan Luu, "Fastopic: Pretrained transformer is a fast, adaptive, stable, and transferable topic model," in *NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15*, 2024.

[20] Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu, "Mpnet: Masked and permuted pre-training for language understanding," in *NeurIPS, December 6-12, virtual*, 2020.

[21] Bin Cao, Chen Yang, Kaibo He, Jing Fan, Honghao Gao, and Pengjiang Qian, "Internal purity: A differential entropy-based internal validation index for crisp and fuzzy clustering validation," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 10, pp. 5660–5673, 2024.

[22] Zhiyuan Dang, Cheng Deng, Xu Yang, Kun Wei, and Heng Huang, "Nearest neighbor matching for deep clustering," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, 2021, pp. 13693–13702.