논문 2025-2-14 http://dx.doi.org/10.29056/jsav.2025.06.14

A Topic Refinement Framework Using GPT and Coherence Filtering to Enhance Topic Coherence in Social Media Data

Ika Widiastuti*, Hwan-Seung Yong*†

Abstract

Identifying coherent topics from social media data is challenging due to its informal style, short length, and high noise levels. Additionally, manually processing such data to extract useful information is both time-consuming and resource-intensive given the massive volume of user-generated content online. To address these challenges, this study evaluates the effectiveness of a previously proposed topic refinement method applied to social media data, including YouTube comments and Twitter posts. Prior to analysis, the datasets underwent preprocessing steps involving the removal of spam comments, offensive content, and short or non-informative comments. The refinement method was then applied post-topic extraction to identify and replace misaligned words using GPT and coherence filtering. The experimental results indicate that the method consistently improves coherence scores across all models and both domains, demonstrating its effectiveness and generalizability in enhancing topic quality in real-world social media discourse

keywords : Topic Modeling, Topic Refinement, Topic Coherence, GPT, Coherence Filtering, Social Media Data

1. Introduction

Topic modeling is a fundamental technique in natural language processing (NLP) used to uncover latent semantic structures in large text corpora. Its cross-domain applicability makes it valuable for processing extensive textual data in fields like social media analysis, market research, and healthcare[1]. With the rise of user-generated content, platforms like YouTube and Twitter have become rich sources for analyzing public discourse. Comments on these platforms reflect sentiment, public opinion, emerging trends, and engagement[2], while also offering insights into behaviors, cultural discourse, and learning practices. However, these comments are often informal, brief, and noisy, posing challenges for generating coherent and interpretable topics[3]. Issues such as grammatical errors and trivial content

* Division of Artificial Intelligence and Software, Computer Science and Engineering Department, Ewha Womans University
† Corresponding Author : Hwan-Seung Yong (email: hsyong@ewha.ac.kr)

Submitted: 2025.05.07. Accepted: 2025.05.13. Confirmed: 2025.06.20.

further complicate meaningful analysis. Coherence improvement is thus crucial, particularly for applications in sentiment tracking, healthcare, trend monitoring, and academic research.

Manually analyzing such vast comment streams is impractical, making automated NLP techniques essential. Latent Dirichlet Allocation (LDA)[4] is widely used for this purpose, modeling documents as mixtures of topics and topics as distributions over words[5]. This approach allows users to summarize large corpora and focus on relevant themes—for instance, a topic related to computing may include words like 'computer', 'model', 'algorithm', and 'data'[6].

Improving coherence and interpretability is therefore essential in the context of social media data analysis, particularly when such analyses are used to inform decision-making processes in areas like public sentiment tracking, healthcare, technology trend monitoring, or academic research.

Beyond LDA, methods like Non-negative Matrix Factorization (NMF)[7] and neural models such as G-BAT[8] and BERTopic[9] have been developed to improve topic interpretability and coherence. While G-BAT models topics as Gaussian distributions in embedding space, BERTopic leverages BERT for capturing nuanced semantics.

Despite these advancements, topic models still face challenges, including low coherence[10], inconsistent topic generation[11], and the presence of irrelevant or misaligned words[12]. Such issues can degrade topic quality and hinder meaningful interpretation.

To address these limitations, we apply our previously proposed post-extracted topic refinement method [13], which enhances coherence by identifying and replacing misaligned words using a combination of GPT and WordNet. This study applies the method to real-world social media data, focusing on YouTube comments and Twitter posts.

The remainder of this paper is organized as follows: Section 2 reviews related work, Section 3 details the methodology, Section 4 presents results and discussion, and Section 5 concludes the paper and suggests future directions.

2. Related Word

2.1 Topic Modeling for Analyzing Social Media Data

Topic modeling helps extract meaningful patterns from large volumes of unstructured data, such as YouTube comments. Studies have applied Latent Dirichlet Allocation (LDA) to uncover latent topics in these comments, providing insights into user behavior and sentiment. For instance, study[14] used LDA to identify 16 product-related themes from customer reviews (e.g., battery, price, camera) and paired it with BLSTM for sentiment classification. This approach helps businesses identify user preferences and concerns.

Despite their utility, social media data poses challenges due to informal language, slang,

topic shifts, and noisy or low-quality content[15], making it difficult to produce coherent and interpretable topics.

2.2 Topic Modeling Refinement for Improving Topic Coherence

Short-text data such as tweets and comments often suffer from sparsity and lack of word co-occurrence, degrading topic coherence and model stability. Text quality issues—like noise, misspellings, and abbreviations—further hinder interpretability. To address this, [16] proposed a Social Media Data Cleansing Model (SMDCM) that improves coherence by preprocessing short text (e.g., removing URLs, normalizing slang).

Traditional topic models may also generate inconsistent results across runs. This instability is more severe with short texts that lack sufficient context, making topic assignments unreliable.

To overcome these limitations, several refinement methods have been proposed. These include embedding-based word replacement, keyphrase enhancement, and large language model (LLM)-driven refinements[17-19]. For example, [17] introduced an interactive visual framework that lets users adjust topics based on semantic relationships. In [18], keyphrases were extracted with an RNN and used to refine LDA-generated topics by adjusting document-topic mappings.

Most notably, [19] presented a model-agnostic refinement technique for short texts using LLMs. This method identifies semantically misaligned intruder words via prompt-based evaluation and replaces them with coherent alternatives.

Building on this, [13] introduced a post-extraction refinement approach that improves topic quality without modifying the initial topic modeling step. It identifies misaligned words and replaces them using WordNet and GPT-generated candidates, selecting replacements based on coherence improvement.

3. Methodology

This section outlines the methodological framework employed in this study, which consist of three main components: (1) Data Collection and Dataset Construction, (2) Data Preprocessing, and (3) Topic extraction and refinement.

3.1 Data Collection and Dataset Construction

We used two types of social media data—YouTube comments and Twitter posts. For YouTube, two domain-specific datasets were constructed, each containing comments from 15 videos. The Technology Dataset was compiled from trending videos in the Technology category, selected based on relevance to technological trends, consumer electronics, and innovation, with preference given to videos with high comment volume and user engagement. The Academic-Science Dataset

was built using keyword-based searches (e.g., “scientific paper,” “peer-review,” “AI research paper”, etc.), with 15 videos collected per keyword and their comments merged into a single dataset. Comments from both datasets were extracted and stored in CSV format.

For Twitter data, we used a subset of the TweetPap dataset[20], which contains tweets referencing scientific papers published on arXiv. This subset captures social media discourse related to academic research. All YouTube comments were collected using the official YouTube Data API v3, which allows programmatic access to video metadata and user-generated content.

3.2 Data Preprocessing

YouTube comments and tweets often contain spam, offensive language, and low-value content that hinder meaningful analysis. Prior to topic modeling, these irrelevant texts must be removed.

Spam comments include repetitive, irrelevant, or promotional content such as “Subscribe to my channel!” or “Click here for free giveaways.” They are often generated by bots or used for self-promotion and do not contribute to genuine discussion. Offensive content, including profanity, hate speech, insults, or sexually explicit language, also degrades discourse quality. Comments like “Only dumb person would believe this garbage”exemplify such harmful language.

To address these issues, a multi-step preprocessing pipeline was developed. First, the dataset is loaded, and empty or null entries are removed. Spam detection is then performed by matching comments against a list of predefined spam-related keywords (e.g., subscribe, click here, bit.ly), and flagged entries are excluded.

Next, offensive language is filtered using a curated list of abusive terms. This list is enhanced with the help of GPT-4, which identifies frequently used but potentially offensive words in context. Comments containing any flagged terms are removed. The pipeline also filters out short, low-information comments using a minimum word count threshold.

Finally, standard text normalization procedures (e.g., lowercasing, removing punctuation) are applied to ensure consistency. The cleaned dataset is then stored for topic modeling.

Table 1. Preprocessing Result of Technology Dataset

Step	Preprocessing	
	Before	After
Total Comments	11,562	8,685
Spam Comments Removed	159	0
Offensive Comments Removed	168	0
Short, Non-Informative Comments Removed	2550	0

Table 1 summarizes the preprocessing results for the Technology dataset. From 11,562 comments in the Technology dataset, 8,685 remained after removing spam, offensive language, and short texts—reflecting a 24.87% reduction.

Table 2. Preprocessing Result of Academic-Science Dataset

Step	Preprocessing	
	Before	After
Total Comments	16,032	13,745
Spam Comments Removed	836	0
Offensive Comments Removed	58	0
Short, Non-Informative Comments Removed	1388	0

Table 2 summarizes the preprocessing results for the Academic-Science YouTube dataset. From an initial 16,032 comments, 836 spam, 58 offensive, and 1,388 low-content comments were removed. This resulted in a refined set of 13,745 comments used for topic modeling and subsequent analysis.

Table 3. Preprocessing Result of TweetPap Dataset

Step	Preprocessing	
	Before	After
Total Comments	15,129	14,609
Spam Comments Removed	509	0
Offensive Comments Removed	0	0
Short, Non-Informative Comments Removed	11	0

Table 3 presents the preprocessing results for the TweetPap dataset. Of the original 15,129 comments, 14,609 were retained after filtering. The minimal removal of offensive or low-value content indicates that the dataset was generally clean and well-suited for topic modeling.

3.3 Topics Extraction and Refinement

In this stage, we apply four topic modeling techniques—LDA, NMF, BERTopic, and GBAT—to extract topics from the preprocessed YouTube and TweetPap datasets. The extracted topics are then refined using our previously proposed method[13] to improve coherence. This method involves two key functions: misaligned word detection, which identifies semantically inconsistent words based on their deviation from the topic centroid, and misaligned word replacement, which uses WordNet and GPT to suggest contextually appropriate alternatives. The detection process is illustrated in Figure 1. We utilize the bert-base-uncased model to generate contextualized BERT embedding, allowing more accurate identification of misaligned words in complex or ambiguous contexts.

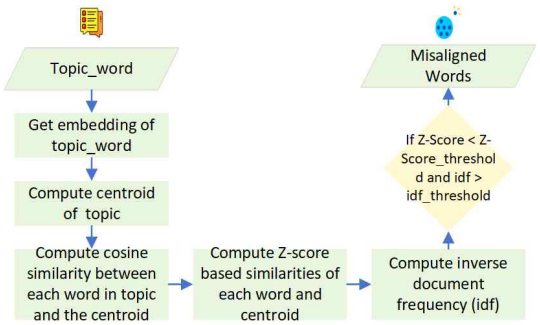

Fig. 1. Flowchart of Misaligned Word Detection

After detecting misaligned words, the next step is to replace them to improve topic coherence. Figure 2 illustrates the replacement process, which combines WordNet and GPT to generate contextually appropriate alternatives.

This dual approach first uses WordNet to identify synonyms and semantically related terms based on the topic centroid, then

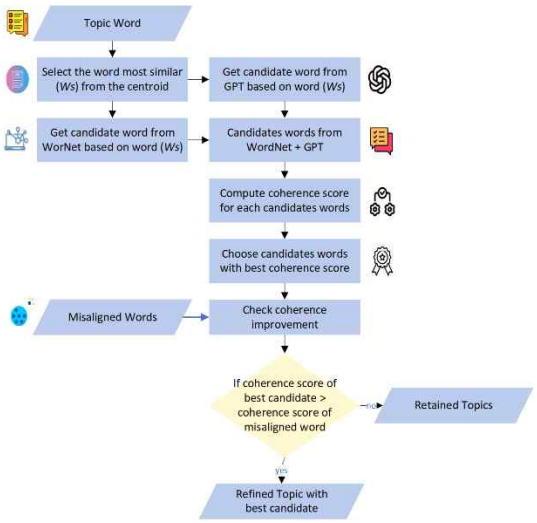

Fig. 2. Flowchart of Misaligned Word Replacement

employs GPT to produce additional candidate words. The combined set is filtered for uniqueness, and each candidate is evaluated by temporarily replacing the misaligned word and recalculating the topic's coherence score. A replacement is accepted if it improves the score —this process is known as coherence filtering.

Coherence filtering uses Gensim's Coherence Model, focusing on the C_v metric[21], which integrates multiple measures to capture topic interpretability and semantic consistency from a human-centric perspective.

4. Result and Discussion

This section presents the results of applying the proposed topic refinement method to three datasets: (1) Technology-related YouTube comments, (2) Academic-scientific YouTube comments retrieved via keyword search, and (3) Science-related tweets. Each dataset was modeled using four algorithms—LDA, NMF, BERTopic, and G-BAT—followed by topic refinement using three candidate word generation strategies: WordNet, GPT, and their combination. Each approach offers distinct advantages in aligning topic words and enhancing interpretability. By comparing these methods, we aim to demonstrate their effectiveness in producing more coherent and meaningful topics.

4.1 Evaluating Topic Coherence Improvement Across Datasets

Figure 3 presents the C_V coherence scores for topics extracted from the Technology YouTube comments dataset. Prior to refinement, models such as LDA and NMF exhibited relatively low coherence due to the informal and fragmented nature of social media language. After applying the refinement strategies, all models showed improvement. Notably, LDA's coherence increased from 0.412 to 0.506 under the hybrid WordNet+GPT approach—the most substantial gain in this dataset. GPT-only and hybrid strategies consistently outperformed WordNet-only, highlighting the value of contextual understanding from language models in enhancing topic quality for informal text.

Figure 4 presents the C_V coherence scores for the Academic-Science dataset, which contains more structured and content-rich language. All models improved following

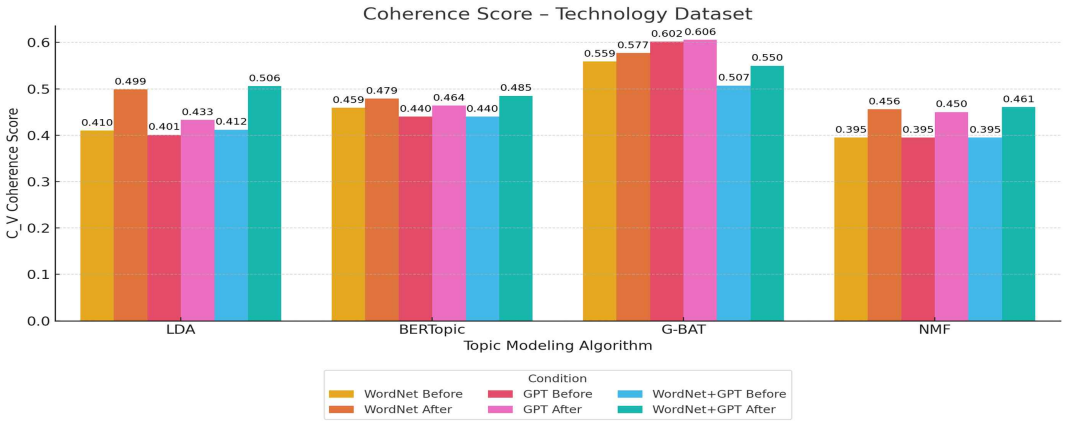

Fig 3. Coherence Score - Technology Dataset

refinement, with LDA showing the greatest gain—increasing from 0.464 to 0.575 using the hybrid WordNet+GPT approach. G-BAT exhibited smaller improvements in all replacement strategy. Compared to the Technology dataset, changes were more moderate but consistent. These results confirm that the refinement method effectively enhances topic quality even in more lexically stable domains like academic discourse.

Figure 5 presents the results for the TweetPap dataset. Despite this, substantial coherence gains were observed—particularly for LDA and NMF. For instance, LDA's coherence increased from 0.412 to 0.528 with the hybrid approach. G-BAT remained relatively stable, reflecting its strong baseline performance. These results confirm the refinement method's

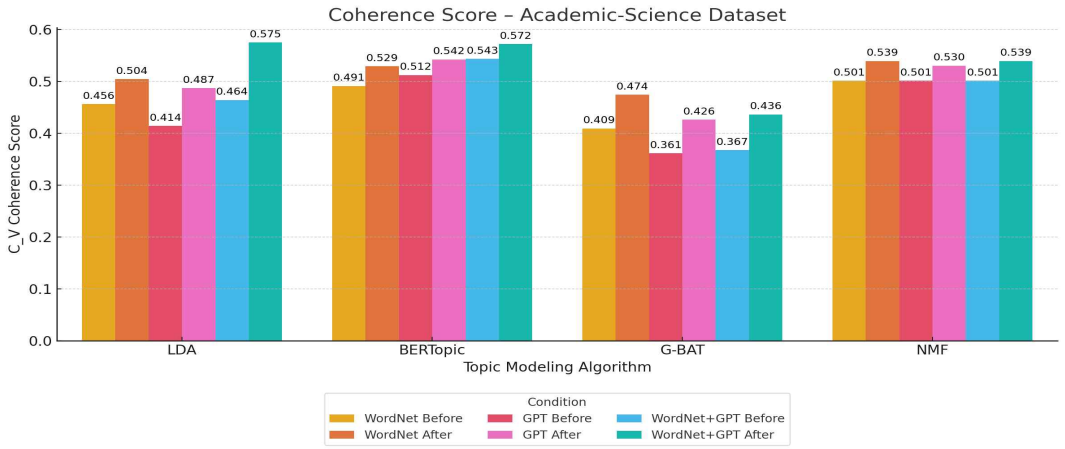

Fig 4. Coherence Score - Academic-Science Dataset

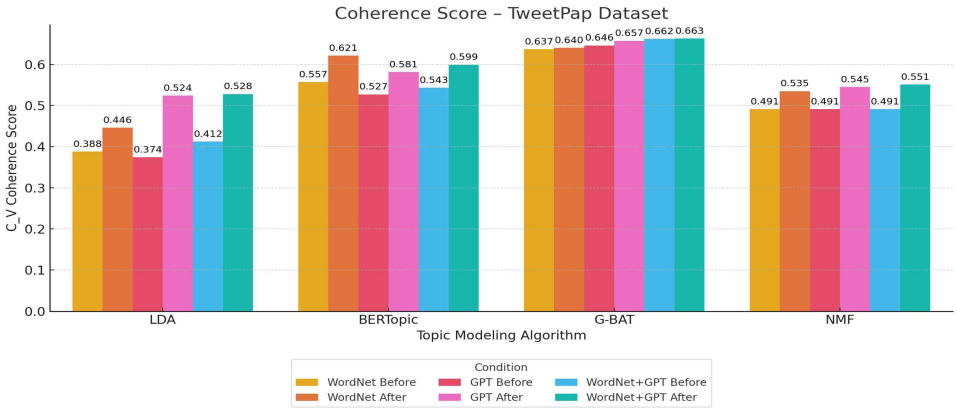

Fig 5. Coherence Score – TweetPap Dataset

effectiveness in both structured and highly compressed social media text.

Across all three datasets, the hybrid WordNet+GPT strategy consistently out performed individual methods, combining WordNet's semantic structure with GPT's contextual fluency to achieve the highest coherence scores. Figures 3 - 5 visually confirm this trend, with hybrid-refined topics showing the tallest bars.

The proposed refinement method improved topic coherence across all models and datasets. LDA showed the largest gains in the Technology dataset, while G-BAT—despite a high baseline—also improved. In the Academic-Science dataset, all models benefited, with the hybrid approach again yielding the best results. Even the short and noisy TweetPap data showed notable gains, particularly for LDA and NMF.

These results demonstrate the method's robustness and adaptability across diverse domains and modeling techniques. While traditional models like LDA gain the most, neural models such as G-BAT and BERTopic also improve, confirming that the refinement approach is both model-agnostic and domain-adaptive.

4.2 Qualitative Topic Comparison

While coherence scores provide a quantitative assessment of refinement, they may not fully reflect topic interpretability or semantic alignment. This subsection offers a qualitative comparison of selected topics before and after refinement to complement the numerical results. Table 4 presents an examples from the YouTube and Twitter datasets that illustrate how misaligned word in user-generated content are detected and replaced using hybrid WordNet+GPT—based replacement approach, resulting in more coherent and interpretable topics.

Table 4. Qualitative examples of Topic Refinement using Hybrid-based Replacement

Dataset	Model	Extracted Topic	Refined Topic	Misaligned Word	Replacement Word
Technology dataset	LDA	**na**, **gon**, go, take, app, make, want, please, heaven, store.	**micturate**, **stimulate**, go, take, app, make, want, please, heaven, store.	na, gon	micturate, stimulate
	NMF	know, **didnt**, day, buy, app, asking, oled, girl, **yall**, man.	know, **human**, day, buy, app, asking, oled, girl, **bloke**, man.	didn't, yall	human, bloke
	BERTopic	toy, kid, memory, childhood, remember, old, **nostalgia**, **year**, **gen**, painting.	toy, kid, memory, childhood, remember, old, **reflection**, **flashback**, **storage**, painting.	Nostalgia, year, gen	Reflection, flashback, storage
	G-BAT	able, **god**, born, work, hitting, **cheaper**, eye, greedy, st, **voltage**.	able, **reach**, born, work, hitting, **contact**, eye, greedy, st, **strike**.	god, cheaper, voltage	reach, contact, strike
Academic-Science dataset	LDA	**um**, de, result, method, study, conclusion, **beijo**, section, abstract, discussion.	**give-and-take**, de, result, method, study, conclusion, **word**, section, abstract, discussion.	um, beijo	give-and-take, word
	NMF	time, got, say, right, long, thing, hard, watching, taking,**ive**.	time, got, say, right, long, thing, hard, watching, taking, **clock**.	ive	clock
	BERTopic	thesis, phd, master, dissertation, degree, **year**, university, project, page, student.	thesis, phd, master, dissertation, degree, **academic work,**university, project, page, student.	year	academic work,
	G-BAT	ability, **paper**, scientific, **website**, andy, project, money, **today**, greatly, precise.	ability, **competence**, scientific, **picture**, andy, project, money, **talent**, greatly, precise.	Paper, website, today	Competence, picture, talent
TweetPap dataset	LDA	**arxivv**, laser, effect, radiation, **physicsaccph**, note, analytical, **phys**, emission, field.	**radioactivity**, laser, effect, radiation, **radiotherapy**, note, analytical, **electromagnetic waves**, emission, field.	arXiv, physicsaccph, phys	radioactivity, radiotherapy, electromagnetic waves
	NMF	relativity, special, body, rigid, **hydrodynamics**, relabelling, **dissipation**, **arxivv**, physicsclassph, transformation.	relativity, special, body, rigid, **change**, relabelling, **conversion**, **shift**, physicsclassph, transformation.	hydrodynamics, dissipation, arxivv	change, conversion, shift
	BERTopic	characterisation, meaningful, **perturbative**, theoretical, uncertainty, continuity, epistemic, thought, empiricist, realization.	characterisation, meaningful, **hypothetical**, theoretical, uncertainty, continuity, epistemic, thought, empiricist, realization.	perturbative	hypothetical
	G-BAT	stable, adhesion, proton, conversation, break, trend, **deciphering**, transition, compactification, astrophim.	stable, adhesion, proton, conversation, break, trend, **evolution**, transition, compactification, astrophim.	deciphering	evolution

4.3 Analysis of Results and Limitations

To quantify the impact of the refinement methods, we calculated the percentage improvement in C_V coherence scores across all models, datasets, and strategies. The hybrid WordNet+GPT approach consistently achieved the highest relative improvement, confirming the complementary strengths of WordNet's semantic precision and GPT's context-aware generation. WordNet ensures topical relevance through structured lexical relations, while GPT

enhances linguistic coherence, resulting in more interpretable topic representations.

Experimental results in Section 4.1 and 4.2 validate the method's effectiveness across diverse topic models and datasets. Among the models, LDA showed the highest relative improvements—most notably in the TweetPap dataset, with a gain of over 28% using the hybrid strategy. G-BAT, despite having the highest baseline coherence, exhibited smaller relative gains due to its already strong performance.

Refinement impact varied across datasets. TweetPap, with its short, noisy scientific tweets, benefited most—highlighting the value of combining structural filtering and contextual enrichment. The Technology dataset showed moderate gains, while the Academic-Science dataset, with more structured comments, demonstrated consistent improvement.

An important limitation lies in the sensitivity of the GPT-based refinement process to prompt design. We used a standardized prompt format: "Provide alternative words for '{centroid_word}' in the context of the topic: '{topic_word}'. Please separate words with commas." While this prompt aims to elicit semantically relevant replacements, minor changes in phrasing—such as using "suggest synonyms" instead of "provide alternative words"—can lead to significantly different outputs. For example, our version returned clean replacements like system, topology, while another added noise: "Here are some option...". Moreover, the formatting of GPT's response plays an important role in downstream processing. The use of delimiters (e.g., commas vs line breaks) affects how candidates are parsed. Inconsistent formatting or inclusion of explanatory text (e.g., "Here are some option:") can lead to incorrect token extraction or noisy replacements.

This sensitivity suggests that GPT-based refinement, while powerful, relies not only the model's linguistic capacity but also on careful prompt formulation and robust response handling. Future research could explore prompt tuning, few-shot examples, or in-context learning techniques to enhance stability and cross-domain applicability.

5. Conclusion and future works

This study proposed and evaluated a topic refinement framework to improve the coherence and interpretability of topic models applied to social media content. The method combines WordNet's lexical precision with GPT's contextual generation to replace semantically misaligned or contextually inconsistent topic words.

The framework was tested on three domain-specific datasets: YouTube comments from the Technology and Academic-Science domains, and science-related tweets (TweetPap). Using four topic modeling algorithms—LDA, NMF, BERTopic, and G-BAT—the refined topics were evaluated using the C_V coherence metric. Results showed consistent improvements across all models and datasets, with the hybrid

WordNet+GPT approach yielding the most significant gains, especially for LDA and NMF. The method proved effective even in noisy, short-text environments like YouTube comments and Twitter posts.

Future work may explore alternative coherence metrics, including contextual or human-centered evaluations, to further assess topic quality. Enhancing GPT-based generation through prompt engineering or dynamic feedback could improve domain adaptability. Additionally, evaluating the impact of refined topics on downstream tasks—such as recommendation, stance detection, or misinformation tracking—could demonstrate broader practical value.

References

[1] K. Taghandiki and M. Mohammadi, “Topic Modeling: Exploring the Processes, Tools, Challenges and Applications”, Authorea Preprints, Oct.2023, doi: 10.36227/TECHRXIV.23528283.V1.

[2] T. Gokcimen and B. Das, “Exploring climate change discourse on social media and blogs using a topic modeling analysis”, Heliyon, vol. 10, no. 11, p. e32464, Jun. 2024, doi: 10.1016/j.heliyon.2024.e32464.

[3] M. Hankar, M. Kasri, and A. Beni-Hssane, “A comprehensive overview of topic modeling: Techniques, applications and challenges”, Neurocomputing, vol. 628, p. 129638, May 2025, doi: 10.1016/J.NEUCOM.2025.129638.

[4] D. M. Blei, A. Y. Ng, and J. B. Edu, “Latent Dirichlet Allocation”, Journal of Machine Learning Research, vol. 3, no. Jan, pp. 993 - 1022, 2003.

[5] M. Yousef and D. Voskergian, “TextNetTopics: Text Classification Based Word Grouping as Topics and Topics’ Scoring”, Frontier in Genetics, vol. 13, p. 893378, Jun. 2022, doi: 10.3389/FGENE.2022.893378/BIBTEX.

[6] L. Liu, L. Tang, W. Dong, S. Yao, and W. Zhou, “An overview of topic modeling and its current applications in bioinformatics”, SpringerPlus 2016 5:1, vol. 5, no. 1, pp. 1 - 22, Sep. 2016, doi: 10.1186/S40064-016-3252-8.

[7] P. Li et al., “Guided Semi-Supervised Non-Negative Matrix Factorization”, Algorithms 2022, vol. 15, no. 5, p. 136, Apr. 2022, doi: 10.3390/A15050136.

[8] R. Wang, X. Hu, D. Zhou, Y. He, Y. Xiong, “Neural Topic Modeling with Bidirectional Adversarial Training”, Proceedings of the Annual Meeting of the Association for Computational Linguistics, pp. 340 - 350, Apr. 2020, doi: 10.18653/v1/2020.acl-main.32.

[9] Z. Fang, Y. He, and R. Procter, “BERTTM: Leveraging Contextualized Word Embeddings from Pre-trained Language Models for Neural Topic Modeling”, May 2023, [Online]. Available: http://arxiv.org/abs/2305.09329

[10] C. D. P. Laureate, W. Buntine, and H. Linger, “A systematic review of the use of topic models for short text social media analysis”, Artificial Intelligence Review, vol. 56, no. 12, pp. 14223 - 14255, Dec. 2023, doi: 10.1007/S10462-023-10471-X/TABLES/5.

[11] M. Bewong, J. Wondoh, S. Kwashie, J. Liu, M.Z. Islam, D. Kernot “DATM: A Novel Data Agnostic Topic Modeling Technique With Improved Effectiveness for Both Short and Long Text”, IEEE Access, vol. 11, pp. 32826 - 32841, 2023,

doi: 10.1109/ACCESS.2023.3262653.

[12] D. K. Geeganage, Y. Xu, and Y. Li, “A Semantics-enhanced Topic Modelling Technique: Semantic-LDA”, ACM Transaction on Knowledge Discovery Data, vol. 18, no. 4, Feb. 2024, doi: 10.1145/3639409/ASSET/42B6D80F-5FA4-48BE-9D82-938963894F11/ASSETS/GRAPHIC/TKDD-2022-06-0213-T04.JPG.

[13] I. Widiastuti and H. S. Yong, “TR-GPT-CF: A Topic Refinement Method Using GPT and Coherence Filtering”, Applied Sciences, vol. 15, no. 4, p. 1962, Feb. 2025, doi: 10.3390/APP15041962.

[14] Z. Wang, P. Gao, and X. Chu, “Sentiment analysis from Customer-generated online videos on product review using topic modeling and Multi-attention BLSTM”, Advanced Engineering Informatics, vol. 52, p. 101588, Apr. 2022, doi: 10.1016/J.AEI.2022.101588.

[15] S. Kinariwala and S. Deshmukh, “Short text topic modelling using local and global word-context semantic correlation”, Multimedia Tools Application, vol. 82, no. 17, pp. 26411 – 26433, Jul. 2023, doi: 10.1007/S11042-023-14352-X/TABLES/.

[16] B. A. H. Murshed, J. Abawajy, S. Mallappa, M. A. N. Saif, S. M. Al-Ghuribi, and F. A. Ghanem, “Enhancing Big Social Media Data Quality for Use in Short-Text Topic Modeling”, IEEE Access, vol. 10, pp. 105328 – 105351, 2022, doi: 10.1109/ACCESS.2022.3211396.

[17] M. El-Assady, R. Kehlbeck, C. Collins, D. Keim, and O. Deussen, “Semantic concept spaces: Guided topic model refinement using word-embedding projections”, IEEE Transaction on Visualization and Computer Graphics, vol. 26, no. 1, pp. 1001 – 1011, Jan. 2020, doi: 10.1109/TVCG.2019.2934654.

[18] K. M. H. Ur Rehman and K. Wakabayashi, “Keyphrase-based Refinement Functions for Efficient Improvement on Document-Topic Association in Human-in-the-Loop Topic Models”, Journal of Information Processing, vol. 31, pp. 353 – 364, 2023, doi: 10.2197/IPSJJIP.31.353.

[19] S. Chang, R. Wang, P. Ren, and H. Huang, “Enhanced Short Text Modeling: Leveraging Large Language Models for Topic Refinement”, ArXiv, Mar. 2024, Accessed: Nov. 30, 2024. [Online]. Available:https://arxiv.org/abs/2403.17706v1

[20] “TweetPap/LICENSE at main · lingo-iitgn/TweetPap”, Accessed: Apr. 14, 2025. [Online]. Available: https://github.com/lingo-iitgn/TweetPap/blob/main/LICENSE

[21] M. Röder, A. Both, and A. Hinneburg, “Exploring the space of topic coherence measures”, WSDM 2015 – Proceedings of the 8th ACM International Conference on Web Search and Data Mining, pp. 399 – 408, Feb. 2015, doi: 10.1145/2684822.2685324.

Authors

Ika Widiastuti

2010.8-2012.2 Master of Electrical Engineering－Multimedia Smart Networks from Sepuluh November Institute of Techmology, Indonesia
2022.8-present Ph.D student in Computer Science and Engineering, Ewha Womans University
2005.2-present Faculty member in State Polytechnic of Jember, Indonesia

<Research interests> Artificial Intelligence, Machine Learning, Data Mining, NLP

Hwan-Seung Yong

1994.2 Ph.D in Computer Engineering from Seoul National University
1985.2-1989.2 ETRI Research Member
1995.3-present Professor in Ewha Womans University

<Research interests> Databases, Data Mining, Artificial Intelligence