nlptuducd at SemEval-2025 Task 10: Narrative Classification as a Retrieval Task through Story Embeddings

Arjumand Younus
School of Information
and Communication Studies
University College Dublin
Dublin / Ireland
arjumand.younus@ucd.ie

M. Atif Qureshi
School of Business Technology,
Retail and Supply Chain
Technological University Dublin
Dublin, Ireland
atif.qureshi@tudublin.ie

Abstract

Media framing is a widely used technique in misinformation campaigns, where narratives are constructed through specific angles to align with political agendas. Such narrative twisting involves complex dynamics of rhetoric, topic selection, and pattern repetition, yet typically maintains coherence aligned with the intended media agenda. The SemEval-2025 Task 10 (Subtask 2) challenges participants to classify online news articles according to a pre-defined, two-level narrative taxonomy. In this paper, we present a retrieval-based approach to narrative classification using the E5 variant of the Mistral 7B language model. Rather than relying on supervised training, our method frames narrative detection as a semantic similarity task via story embeddings. This design exploits the inherent coherence across similarly framed news stories. Our system achieves a sample-level F1 score of 0.226, outperforming the official baseline. We highlight both the strengths and challenges of using retrieval-based embeddings for narrative understanding and propose future directions for improving label precision and generalizability.

1 Introduction

Media framing, an interdisciplinary area of study (Otmakhova et al., 2024), plays a central role in shaping online disinformation. SemEval-2025 Task 10 (Piskorski et al., 2025) advances this field by introducing a richly annotated multilingual dataset focused on controversial topics such as climate change and the Russia–Ukraine conflict. This paper presents our submission for Subtask 2: narrative classification of English news articles using a predefined two-level taxonomy (Stefanovitch et al., 2025). The task is framed as a multi-label, multi-class classification problem where each article can be associated with multiple narrative labels.

Prior work on narratives has primarily adopted an event-centric lens (Piper et al., 2021), which often overlooks propagandistic intent and disinformation framing. In contrast, Task 10 brings attention to narrative construction in the context of agenda-driven reporting, introducing new challenges in narrative modeling and classification.

To address these challenges, we frame narrative classification as a retrieval problem. This approach is motivated by several observations from the disinformation literature:

- News articles advancing specific media frames often rely on repetitive rhetorical patterns (Entman, 2003).
- These frames typically focus on a limited set of recurring topics (DiMaggio et al., 2013).
- Articles promoting similar disinformation narratives—particularly in crisis events—tend to exhibit semantic and thematic coherence (Baden and Stalpouskaya, 2015; Van der Meer et al., 2014).

Motivated by these insights, we propose a narrative classification system grounded in topical coherence and semantic similarity. Our system leverages story embeddings to retrieve semantically aligned training articles and transfers their narrative labels to test samples. Figure 1 provides a high-level overview of this pipeline, where narrative detection is framed as a retrieval problem based on embedding similarity rather than supervised classification. We adopt the approach proposed by Hatzel and Biemann (2024), which utilizes contrastive learning and adapter-finetuned large language models (E5 variant) (Wang et al., 2024) for robust narrative representation.

This paper details our system pipeline, experimental results, observed limitations, and future research directions.

2 System Description

To address the task of narrative classification, we adopt a retrieval-based approach inspired by story

Proceedings of the The 19th International Workshop on Semantic Evaluation (SemEval-2025), pages 1742–1746
July 31 - August 1, 2025 ©2025 Association for Computational Linguistics

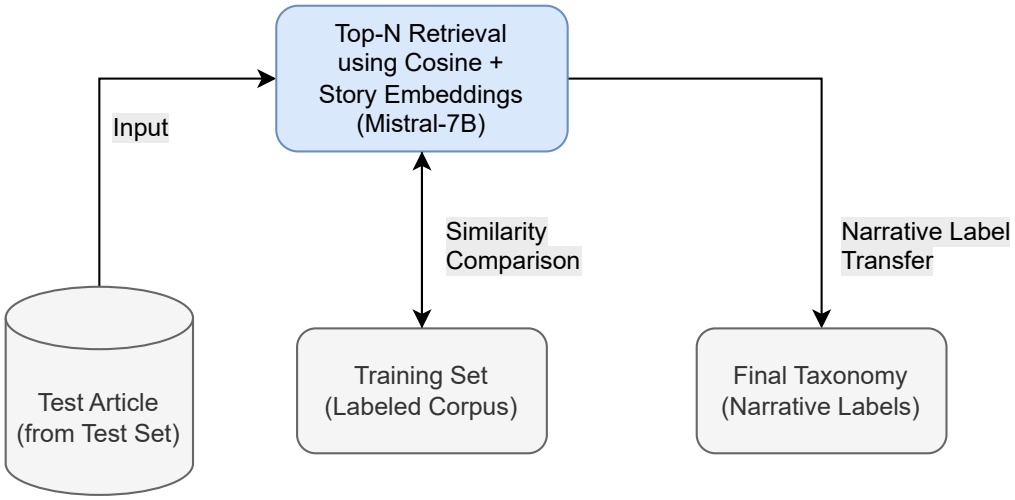

Figure 1: High-level Overview of Our Subjectivity Analysis Pipeline

embeddings proposed by Hatzel and Biemann (2024). Their method models narrative coherence through contrastive learning applied to narrative chains. We hypothesize that narrative labels in news articles can be transferred effectively based on semantic proximity in embedding space. Therefore, rather than training a supervised classifier, we treat the problem as one of story retrieval and label transfer.

Our approach is influenced by common-sense reasoning tasks such as the Story Cloze Test (Mostafazadeh et al., 2017), where systems must choose the more coherent ending from two candidates given a four-sentence story. Similarly, we embed a test article and compare it to candidate articles from the training set to identify the most semantically similar ones and inherit their labels. The overall system pipeline, previously introduced in Figure 1, proceeds through the following stages:

Given a test article, we compute cosine similarity between its embedding and those of all training articles using the E5 variant of Mistral-7B (Wang et al., 2024). The process comprises the following steps:

- **Similarity Ranking:** We first embed all articles using the story embedding model and rank training samples by their cosine similarity to the test article.
- **Candidate Selection:** The top-n most similar training articles are selected. These represent the most semantically coherent stories with respect to the test article.
- **Anchor-Based Prediction:** We treat the test article as the anchor and embed it alongside each of the top-n stories. We then compute similarity in the story embedding space to identify the closest narrative match.
- **Label Transfer:** The narrative labels from the most similar training article are directly transferred to the test article. This unsupervised strategy assumes that narrative proximity implies label relevance. However, as illustrated in Table 1, this approach may result in the transfer of excessive or irrelevant labels in cases where the semantic similarity is partial or ambiguous.

In implementation, we use the open-source story embedding model released by Hatzel and Biemann (2024) on Hugging Face.[1] No additional fine-tuning was performed on the model. The query prompt used to guide retrieval was: *"Retrieve stories with a similar narrative to the given story:"* This framing implicitly guides the model to surface coherent narrative alignments.

While this pipeline provides a simple and generalizable baseline, we recognize that direct label transfer introduces potential noise, especially in multi-label cases. We analyze this limitation and its impact on performance in the following section.

[1] https://huggingface.co/uhhlt/story-emb

Article Snippet	Excess Label Transferred
• Title: **OBEY THE GREEN: Blue states depriving rural counties of right to reject green energy community takeovers** Snippet: Each of these so-called "blue" states – Michigan is arguably a "purple" state since it appears as though Donald Trump may have won the state, minus the fraud, in the 2020 election – is controlled by far-left politicians who are in bed with the green energy industry.	CC: Hidden plots by secret schemes of powerful groups
• Title: **European Parliament members clash over support for Ukraine** Snippet:Members of the European Parliament (EP) engaged in mutual insults during debates on the need for further support to Ukraine. Several European Parliamentarians accused advocates of continuing military assistance of madness and called for an end to arms shipments during debates in the EP's plenary session in Strasbourg.	URW: Russia is the Victim

Table 1: Examples of Excess Labels' Transfer in Development Set

3 Experiments and Evaluation

3.1 Experimental Setup

We implemented our system using Google Colab with a T4 GPU. The story embedding model from Hatzel and Biemann (2024) was accessed via Hugging Face[2] without any additional fine-tuning. The dataset used was the official English subset provided by the Task 10 organizers. After computing predictions, results were exported in the required submission format for leaderboard evaluation.

3.2 Evaluation and Label Transfer Analysis

Since gold labels for the final test set are unavailable, we conducted a detailed analysis using the development set. One key insight is that the article most similar to a test sample based on cosine similarity often differs from the one closest in the story embedding space. This similarity re-ranking effect occurred in roughly 50% of the development samples, indicating that embedding-based coherence plays a critical role.

However, we also observed a limitation: direct transfer of narrative labels from the nearest training article sometimes led to excessive or irrelevant labels. This issue is particularly evident in multi-label scenarios, where overlapping but distinct narratives may coexist. Table 1 (introduced in Section 2) provides concrete examples of this phenomenon. These findings highlight the need for more precise or supervised methods for narrative label assignment.

To assess the sensitivity of our system to the number of retrieved candidates, we varied the value of n used for label transfer. Table 2 summarizes the results. We observe that increasing n beyond 2 consistently degrades performance. This drop in F1 score is likely due to added semantic noise from additional training articles, which dilutes the coherence of the narrative signal.

[2] https://huggingface.co/uhhlt/story-emb

Number of Retrieved Articles (n)	F1 Score (Sample-Level)
2	0.2260
3	0.1830
4	0.1780

Table 2: Effect of Number of Retrieved Articles on F1 Score

These results reinforce our design choice to limit label transfer to the top-$n = 2$ most similar training stories. Selecting a larger set introduces topic drift or partial matches, increasing the risk of incorrect narrative assignments.

4 Conclusion and Future Work

This paper presents our system description for narrative classification in Subtask 2 of SemEval-2025 Task 10. We approached the task through a retrieval-based perspective using story embeddings derived from a Mistral-7B model variant (E5), framing narrative detection as a similarity-driven retrieval challenge. Our method leverages the narrative coherence and topical consistency typically embedded in disinformation, enabling us to transfer narrative taxonomy labels based on content proximity in the semantic embedding space. Without any additional supervised fine-tuning, our system outperformed the competition baseline and achieved an F1 score of 0.226 on the samples.

Our analysis revealed both strengths and limitations of our pipeline. Notably, the re-ranking of similar stories using embedding-based similarity rather than raw cosine similarity led to insightful improvements. However, we also observed the challenge of excessive or inaccurate label transfer during the final inference stage, highlighting the need for a more precise mechanism for label assignment.

As part of future work, we intend to address these limitations by:

- Incorporating supervised learning components to refine the label transfer step, potentially using contrastive loss or multi-label classification heads on top of retrieved embeddings.
- Exploring narrative disambiguation strategies that can handle overlapping or competing frames within articles more robustly.
- Applying zero-shot or few-shot prompt engineering techniques to leverage large language models directly for narrative prediction, reducing reliance on nearest-neighbor heuristics.
- Investigating multilingual extensions to better support narrative detection in global contexts, given the cross-cultural framing of controversial topics.

Overall, our findings suggest that framing narrative classification as a retrieval task is a promising direction, especially when paired with pre-trained embeddings that encode structural and semantic narrative features. We hope our approach sparks further innovations in retrieval-based disinformation and narrative analysis.

References

Christian Baden and Katsiaryna Stalpouskaya. 2015. Maintaining frame coherence between uncertain information and changing agendas: The evolving framing of the syrian chemical attacks in the us, british, and russian news. In *ICA Annual Conference, San Juan*.

Paul DiMaggio, Manish Nag, and David Blei. 2013. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of us government arts funding. *Poetics*, 41(6):570–606.

Robert M Entman. 2003. Cascading activation: Contesting the white house's frame after 9/11. *Political Communication,*, 20(4):415–432.

Hans Ole Hatzel and Chris Biemann. 2024. Story embeddings — narrative-focused representations of fictional stories. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 5931–5943, Miami, Florida, USA. Association for Computational Linguistics.

Nasrin Mostafazadeh, Michael Roth, Annie Louis, Nathanael Chambers, and James Allen. 2017. LSDSem 2017 shared task: The story cloze test. In *Proceedings of the 2nd Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics*, pages 46–51, Valencia, Spain. Association for Computational Linguistics.

Yulia Otmakhova, Shima Khanehzar, and Lea Frermann. 2024. Media framing: A typology and survey of computational approaches across disciplines. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15407–15428, Bangkok, Thailand. Association for Computational Linguistics.

Andrew Piper, Richard Jean So, and David Bamman. 2021. Narrative theory for computational narrative understanding. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 298–311, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Jakub Piskorski, Tarek Mahmoud, Nikolaos Nikolaidis, Ricardo Campos, Alípio Jorge, Dimitar Dimitrov, Purificação Silvano, Roman Yangarber, Shivam Sharma, Tanmoy Chakraborty, Nuno Guimarães, Elisa Sartori, Nicolas Stefanovitch, Zhuohan Xie, Preslav Nakov, and Giovanni Da San Martino. 2025. SemEval-2025 task 10: Multilingual characterization and extraction of narratives from online news. In *Proceedings of the 19th International Workshop on Semantic Evaluation*, SemEval 2025, Vienna, Austria.

Nicolas Stefanovitch, Tarek Mahmoud, Nikolaos Nikolaidis, Jorge Alípio, Ricardo Campos, Dimitar Dimitrov, Purificação Silvano, Shivam Sharma, Roman Yangarber, Nuno Guimarães, Elisa Sartori, Ana Filipa Pacheco, Cecília Ortiz, Cláudia Couto, Glória Reis de Oliveira, Ari Gonçalves, Ivan Koychev, Ivo Moravski, Nicolo Faggiani, Sopho Kharazi, Bonka Kotseva, Ion Androutsopoulos, John Pavlopoulos, Gayatri Oke, Kanupriya Pathak, Dhairya Suman, Sohini Mazumdar, Tanmoy Chakraborty, Zhuohan Xie, Denis Kvachev, Irina Gatsuk, Ksenia Semenova, Matilda Villanen, Aamos Waher, Daria Lyakhnovich, Giovanni Da San Martino, Preslav Nakov, and Jakub Piskorski. 2025. Multilingual Characterization and Extraction of Narratives from Online News: Annotation Guidelines. Technical Report JRC141322, European Commission Joint Research Centre, Ispra (Italy).

Toni GLA Van der Meer, Piet Verhoeven, Hans Beentjes, and Rens Vliegenthart. 2014. When frames align: The interplay between pr, news media, and the public in times of crisis. *Public Relations Review*, 40(5):751–761.

Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. 2024. Improving text embeddings with large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11897–11916, Bangkok, Thailand. Association for Computational Linguistics.