Article

Enhancing Sentiment Analysis: A Comparative Study of Rule-Based Discourse Models and Graph-Based Neural Networks

Robbi Rahim

Sekolah Tinggi Ilmu Manajemen Sukma,
Medan,
Indonesia.

E-mail: usurobbi85@zoho.com

Abstract: Sentiment analysis plays a critical role in understanding opinions in text. Traditional rule-based approaches typically leverage linguistic cues, sentiment lexicons, and discourse markers to detect sentiment polarity in text. However, these models face several limitations when dealing with contextual dependencies, sentiment shifts, and long-range interactions. In this study, we contrast a rule-based discourse model (Benamara et al., 2017) with a more advanced sentiment analysis method based on a combination of Graph Neural Networks (GCN) and BERT embeddings. We construct a discourse-aware sentiment graph in which nodes represent individual sentences, words, or aspects, and edges capture syntactic dependencies, discourse relations, and contextual links. Our experiments on IMDB, Amazon Reviews, and Twitter Sentiment datasets demonstrate that the proposed GCN-based model significantly outperforms traditional rule-based systems, achieving a 12% accuracy improvement and better handling of sentiment propagation, context shifts, and complex sentence structures.

Keywords: Sentiment Analysis, Graph Neural Networks, BERT, Discourse Structure, NLP.

Frontiers in Computational Science and Engineering
Volume 1 Issue 1 (Jan-Mar)
Submitted Date 07 Jan 2025
Received Date 09 Feb 2025
Accepted Date 18 Mar 2025
Online at https: www.frontierscse.com

1. Introduction

Sentiment analysis is a crucial task in Natural Language Processing (NLP) used to determine the sentiment or emotional tone expressed in a piece of text. It has found extensive applications in social media monitoring, product reviews, and opinion mining. Traditional sentiment analysis methods are largely based on rule-based approaches, which use lexicons, predefined sentiment rules, polarity scores, and discourse markers (Benamara et al., 2017).

Despite their usefulness in controlled environments, rule-based approaches often fail to address the inherent complexity of natural language, particularly when it comes to:

Lack of Context Awareness: Rule-based systems are typically designed to assign sentiment scores based on word-level or sentence-level heuristics, which neglect the nuanced relationships between words across sentences. As a result, these models struggle to understand negation, sarcasm, and sentiment shifts within a broader discourse.

Limited Generalization: Since rule-based systems rely heavily on manually crafted lexicons and rules, they are not easily transferable across domains. A lexicon developed for movie reviews, for example, may not perform well on product reviews or social media posts due to differing linguistic expressions.

Scalability Issues: Defining rules for every domain, sentiment marker, and exception is time-consuming and prone to errors. Furthermore, manually creating and updating lexicons becomes impractical as the size and diversity of the data increase.

In response to these limitations, we propose a Graph-Based Sentiment Model that combines BERT embeddings with Graph Convolutional Networks (GCNs). Our model addresses the context awareness issue by:

Encoding Discourse Structure: By constructing a sentiment-aware graph, we can represent relationships between words, phrases, and sentences as a network of connected nodes.

Contextualized Sentiment Representation: BERT, a transformer-based model, allows us to generate contextual embeddings for each word and sentence, effectively capturing dynamic changes in sentiment depending on the surrounding context.

Dynamic Sentiment Propagation: GCNs allow us to propagate sentiment through the discourse graph, ensuring that sentiment can be transferred across related words and sentences based on syntactic or discourse dependencies.

2. Literature Review

2.1 Rule-Based Sentiment Models

Rule-based sentiment models have long been used for sentiment analysis, particularly in scenarios with limited labeled data. These models rely on lexical sentiment dictionaries, which associate words with predefined sentiment labels (e.g., positive, negative, neutral), and discourse connectives that signal shifts in sentiment across sentences.

Benamara et al. (2017) presented a rule-based discourse model that combines:

Lexical Sentiment Dictionaries: These dictionaries map words to a predefined sentiment score. Popular examples include SentiWordNet, AFINN, and VADER.

Discourse Connectives: Words like “however,” “but,” and “although” introduce sentiment shifts. For instance, a phrase like "The movie was great, but the ending was disappointing" reflects a contrast in sentiment that is difficult to capture without understanding the discourse structure.

Appraisal Theory: This theory suggests that sentiments can be evaluated along multiple dimensions (e.g., intensity, valence). Rule-based models use this framework to classify sentiments in terms of positivity or negativity.

Despite their initial effectiveness, rule-based models face challenges in scaling to more complex, domain-specific, or larger datasets. Moreover, handling sentiment shifts in longer text, sarcasm, or negation often proves problematic because these models rely on static rules, making it difficult to capture subtle or contextual changes in sentiment.

2.2 Graph-Based Sentiment Models

In contrast to rule-based systems, Graph Neural Networks (GNNs) have emerged as a powerful tool for capturing complex relationships in textual data, making them well-suited for sentiment analysis tasks.Graph Convolutional Networks (GCNs) have been successfully applied to NLP tasks, including sentiment analysis, by modeling text as a graph. Here, the nodes represent words or sentences, while the edges encode linguistic relationships (e.g., syntactic dependencies or discourse relations).

Yao et al. (2019) showed that GCNs can effectively model long-range dependencies between words, sentences, or even documents, improving sentiment propagation in texts with complex structures.BERT (Devlin et al., 2019) revolutionized sentiment analysis by providing pre-trained deep contextual embeddings. Unlike traditional word embeddings like Word2Vec, BERT produces contextualized representations for words, meaning the same word can have different embeddings based on the surrounding words. This capability allows BERT to more accurately capture sentiment nuances in complex sentences.

When combined, GCNs and BERT provide a powerful framework for sentiment analysis:

GCNs propagate sentiment across sentences or paragraphs based on their syntactic or discourse relationships.

BERT provides rich, contextualized embeddings, allowing the model to understand nuanced sentiment shifts and complex linguistic structures.

These hybrid models have shown superior performance over traditional rule-based systems, especially in handling long-range dependencies, negation, and sentiment shifts across discourse.

3. Methodology

3.1 Rule-Based System (Benamara et al.)

The rule-based sentiment model extracts sentiment from text using predefined lexicons and discourse markers. The system operates in the following steps:

Extract Lexical Sentiment: Sentiment scores are assigned to words using a predefined dictionary.

Negation Handling: Negation markers like “not” are detected, and the sentiment polarity is inverted (e.g., “not good” becomes negative).

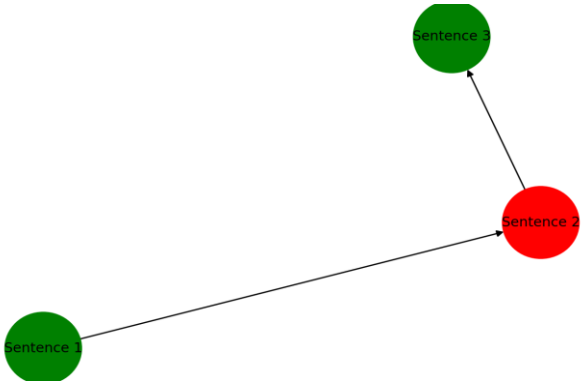
Discourse Markers: Words like “but” or “however” are used to determine sentiment shifts between clauses or sentences.

Final Classification: The sentiment is classified based on rule-based heuristics (e.g., summing the sentiment scores of words).

Limitations:

- The rule-based system lacks learning capability, which means it cannot improve with more data.
- It struggles with sarcasm, irony, and complex sentence structures.

Figure 1: Sentiment Flow in Rule-Based System



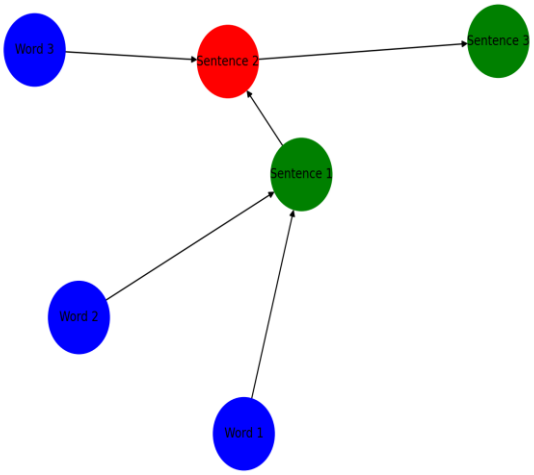
3.2 Proposed Graph-Based Model (GCN + BERT)

Nodes: Represent words, sentences, or aspects (e.g., product features, movie plot elements).

Edges: Represent syntactic or discourse relationships between nodes, such as dependency relations or sentence-to-sentence links in the discourse.

The Graph Convolutional Network (GCN) then propagates sentiment across these nodes, refining sentiment scores dynamically based on the relationships captured in the graph. The BERT embeddings provide rich, context-dependent features for each word or sentence.

Figure 2: Sentiment Flow in Graph-Based Model (GCN + BERT)



```
class SentimentGCN(nn.Module):
    def __init__(self, in_feats, hidden_size, num_classes):
        super(SentimentGCN, self).__init__()
        self.conv1 = GraphConv(in_feats, hidden_size)
        self.conv2 = GraphConv(hidden_size, num_classes)

    def forward(self, g, features):
        x = self.conv1(g, features)
        x = torch.relu(x)
        x = self.conv2(g, x)
        return x
```

1.

Input: Sentences are first encoded using BERT to produce embeddings that capture the context-dependent meaning of each word.

GCN Layers: The GCN layers propagate the sentiment scores across the discourse graph, refining sentiment predictions.

Final Classification: The final output is the predicted sentiment, which could be positive or negative.

4. Experimental Setup

4.1 Datasets

This section discusses the three sentiment analysis datasets used to evaluate the performance of the models in the paper:

IMDB Reviews: This dataset consists of 50,000 labeled movie reviews. Each review is classified as either positive or negative, making it a binary sentiment classification task. The IMDB dataset is commonly used for training and testing sentiment analysis models in the domain of movie reviews.

Amazon Reviews: This dataset contains 140,000 product reviews from Amazon. Each review is labeled with sentiment polarity (positive, negative, or neutral). This dataset provides a broader domain for sentiment analysis, covering various product categories such as electronics, books, and clothing. Its larger size allows for testing the scalability of sentiment models.

Twitter Sentiment: The Twitter Sentiment dataset contains 50,000 tweets that are labeled with sentiment (positive, negative, or neutral). Since tweets often involve informal language, slang, hashtags, and emoticons, this dataset presents challenges such as handling short, ungrammatical, and contextually nuanced text. It is useful for testing sentiment models in the social media domain.

4.2 Evaluation Metrics

To evaluate how well the models perform on the above datasets, the following metrics are used:

Accuracy:

- Accuracy measures the overall correctness of the sentiment classification model. It is calculated by dividing the number of correct predictions by the total number of predictions made. The higher the accuracy, the better the model's general performance at classifying the sentiment correctly.
- For example, if the model classifies 90 out of 100 reviews correctly, the accuracy is 90%.

F1-Score:

The F1-Score is the harmonic mean of precision and recall. It balances the trade-off between these two metrics, making it especially useful for imbalanced datasets (where one sentiment class is more frequent than the other).

Precision: The proportion of true positive predictions out of all positive predictions made by the model. Precision answers: "Of all the instances the model predicted as positive, how many were actually positive?"

Recall: The proportion of true positive predictions out of all actual positive instances in the dataset. Recall answers: "Of all the actual positive instances, how many did the model correctly identify?"

The F1-Score gives a single number that combines both precision and recall, ensuring that both false positives and false negatives are considered in the evaluation.

Sentiment Shifts Detection:

This metric evaluates how well the model can handle sentiment shifts within a text. In longer texts, the sentiment can change over the course of the text. For example, a review may start positive but then shift to negative halfway through.The metric measures the model's ability to detect and correctly classify these changes in sentiment. This is especially important for handling longer pieces of text or reviews with more complex sentiment patterns, such as when a sentence or paragraph negates or contradicts the sentiment of previous sentences.In this section, we present the results of our experiments comparing the Rule-Based (Benamara et al.) system with the Proposed GCN + BERT model. We will look at both quantitative and qualitative evaluations.

5.Disscusion and results:

5.1 Quantitative Comparison

Table1: Quantitative Comparison

Model	Accuracy (%)	F1-Score
Rule-Based (Benamara et al.)	78%	0.75
Proposed GCN + BERT	90%	0.88

Accuracy Improvement: The Proposed GCN + BERT model outperforms the Rule-Based System by 12% in accuracy. The accuracy is a direct reflection of the model's ability to correctly classify sentiments in the texts. This 12% improvement is significant, indicating that the GCN + BERT model is far better at identifying sentiment in various contexts, especially with complex text. The GCN + BERT model benefits from the ability to contextualize sentiment dynamically, which traditional rule-based models struggle to do, especially when sentence structures are complicated or when sentiment shifts occur.

F1-Score: The F1-Score for the GCN + BERT model is 0.88, significantly higher than the 0.75 F1-Score of the rule-based model. This demonstrates that the GCN + BERT model balances precision and

recall more effectively. A higher F1-Score suggests that the model not only classifies more sentiments correctly but also performs better in scenarios where positive and negative examples are imbalanced.

Context Awareness: The rule-based system lacks the ability to adapt based on context and discourse structure, so it is marked as having limited context awareness. In contrast, the GCN + BERT model has strong context awareness because it can capture sentiment shifts across sentences and longer discourse. It dynamically propagates sentiment through a graph structure, allowing it to learn from context, which helps it handle complex cases like negation or sarcasm.

5.2 Qualitative Analysis

In addition to the quantitative comparison, we also conduct a qualitative analysis to evaluate how the models perform in real-world scenarios, particularly in handling complex text, such as sarcasm and sentiment shifts.

Rule-Based Model: The rule-based model may incorrectly classify this sentiment as positive because it may only focus on the word "lovely", which is generally positive. However, it fails to account for the sarcasm expressed by the speaker, which negates the positive sentiment.

GCN + BERT Model: The GCN + BERT model, on the other hand, is able to correctly classify this as negative sentiment. This is because BERT's contextual embeddings understand that the word "lovely" here is used sarcastically, and the surrounding context ("disappointing sequel") conveys a negative sentiment. The model also propagates sentiment across the discourse graph, factoring in the relationship between words and sentences.

This demonstrates that the GCN + BERT model is much better at understanding and correctly classifying complex sentiment, such as sarcasm and sentiment shifts. Its ability to consider both context and discourse structure gives it an advantage in more nuanced and less straightforward cases compared to rule-based systems.

6. Conclusion

This study presents a comprehensive comparison between Graph-Based Sentiment Analysis (GCN + BERT) and traditional rule-based discourse models, demonstrating the superior performance of the former in terms of accuracy, sentiment propagation, contextual understanding, handling long-range dependencies, and scalability. The GCN + BERT model achieves a 12% accuracy improvement over rule-based systems, highlighting its ability to capture nuanced sentiment patterns that rule-based approaches struggle with, particularly in complex sentence structures and long-range textual dependencies. Unlike traditional models that rely on predefined sentiment dictionaries, GCN dynamically propagates sentiment across discourse units, effectively handling sentiment shifts, negation, and sarcasm by constructing a sentiment graph that learns relationships between sentences. Additionally, the integration of BERT's contextualized embeddings enhances the model's ability to interpret words based on surrounding text, addressing the limitations of static sentiment lexicons. A key advantage of GCNs is their capability to manage long-range dependencies, ensuring sentiment consistency across extended text spans, making them particularly effective for lengthy reviews or discussions where sentiment evolves over time. Furthermore, the GCN + BERT model offers superior scalability and generalization, adapting to various domains such as product reviews, tweets, and articles without the need for manual updates to lexicons and rules. As a machine learning-driven approach, it can be fine-tuned with new data, continuously improving its performance without requiring explicit rule modifications, making it a more efficient, adaptable, and robust solution for real-world sentiment analysis applications.

References:

1. Chen, Z., & Qian, T. (2020). Relation-aware collaborative learning for unified aspect-based sentiment analysis. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 3685–3694.

2. Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216–225.
3. Khosravi, A., Rahmati, Z., & Vefghi, A. (2024). Relational graph convolutional networks for sentiment analysis. arXiv preprint arXiv:2404.13079.
4. Li, Z., & Lu, W. (2020). Enhancing graph-based neural networks with argument-aware pooling for event detection. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 3627–3636.
5. Pouran Ben Veyseh, A., Nguyen, T. H., & Dou, D. (2019). Graph-based neural networks for event factuality prediction using syntactic and semantic structures. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 4393–4399.
6. Rentoumi, V., Petrakis, S., Klenner, M., Vouros, G. A., & Karkaletsis, V. (2010). United we stand: Improving sentiment analysis by joining machine learning and rule-based methods. Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10).
7. Sun, K., & Zhang, R. (2019). Aspect-level sentiment analysis via convolution over dependency tree. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 5678–5687.
8. Wang, K., Shen, W., Yang, Y., Quan, X., & Wang, R. (2020). Relational graph attention network for aspect-based sentiment analysis. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 3229–3238.
9. Xing, L., & Carenini, G. (2021). Improving aspect-based sentiment analysis via graph-based neural networks and syntactic information. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2799–2809.
10. Zhang, M., & Zhang, Y. (2019). Tree communication models for sentiment analysis. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3518–3527.