

## ABSTRACT

### *Aspect-Based Sentiment Analysis with Syntactic and Semantic Aware Heterogeneous Graph Convolutional Network (SSHGCN)*

William Hilmy Susatyo

21/472585/PA/20308

In recent days, the immense growth of user-generated content online has led to an increased demand for sentiment analysis techniques that can understand the sentiment expressed towards specific aspects of a product, service, or topic, outlining the importance of the Aspect-Based Sentiment Analysis (ABSA) task. The existing approaches for this task include the utilization of the variation of the Graph Convolutional Network and the conventional text embedding method, such as LSTM and Word2Vec.

However, there are still several disadvantages regarding this approach, including a lack of the ability to discern subtle variations in sentiment related to specific aspects. In this research, the author proposes the method called Syntactic and Semantic Aware Heterogeneous Graph Convolutional Network (SSHGCN). The difference between this approach and the existing approach is that the utilization of the BERT layer is in the hidden layer, apart from only the GCN layer. This approach utilized the dataset of SemEval 2014, SemEval 2015, and SemEval 2016, which is considered to have imbalanced labels in both the training and testing datasets. Additionally, this approach combined the syntactic and semantic weight module in the feature extraction step by using the adjacency matrix and pre-trained BERT embeddings.

This proposed method obtained the best accuracy of 0.814 in the dataset of SemEval 2016, while the best F1-Score of 0.678 in the dataset of SemEval 2014 (Laptop). These results highlight the robustness of the proposed model, despite several better results obtained in the baseline. Furthermore, the utilization of the aspect embedding results in the decrease of the corresponding performance metrics. Despite that condition, this research serves as a basis, and hopefully the future research would give a better outcome by also paying attention to the process of obtaining the aspect instead of only using the aspect as a helper to obtain the sentiment polarity of the whole sentences. Therefore, it is also suggested to explore the related task, such as aspect term extraction.

**Keywords:** Natural Language Processing, Sentiment Prediction, Benchmark Datasets, Multi-modal Graph Representation, Feature Learning

## ABSTRAK

### *Analisis Sentimen Berbasis Aspek dengan Syntactic and Semantic Aware Heterogeneous Graph Convolutional Network (SSHGCN)*

William Hilmy Susatyo

21/472585/PA/20308

Belakangan ini, pertumbuhan konten yang dibuat pengguna secara daring telah menyebabkan peningkatan permintaan akan teknik analisis sentimen yang dapat memahami sentimen yang diungkapkan terhadap aspek-aspek tertentu dari suatu produk, layanan, atau topik, yang menunjukkan pentingnya Analisis Sentimen Berbasis Aspek (ABSA). Pendekatan yang ada untuk tugas ini mencakup pemanfaatan variasi Graph Convolutional Network dan metode ekstraksi fitur konvensional, seperti LSTM dan Word2Vec.

Kendati demikian, masih ada beberapa kelemahan mengenai pendekatan ini, termasuk kurangnya kemampuan untuk membedakan variasi dalam sentimen yang terkait dengan aspek-aspek tertentu. Dalam penelitian ini, penulis mengusulkan metode yang disebut Syntactic and Semantic Aware Heterogeneous Graph Convolutional Network (SSHGCN). Perbedaan antara pendekatan ini dan pendekatan yang sudah ada adalah bahwa penggunaan *layer* BERT pada *hidden layer*. Pendekatan ini memanfaatkan dataset SemEval 2014, SemEval 2015, dan SemEval 2016, yang memiliki label yang tidak seimbang baik dalam dataset pelatihan maupun pengujian. Selain itu, pendekatan ini menggabungkan *syntactic and semantic weighted module* pada tahapan ekstraksi fitur dengan menggunakan *adjacency matrix* dan *embedding* BERT yang telah dilatih sebelumnya.

Metode yang diusulkan ini memperoleh akurasi terbaik sebesar 0,814 dalam dataset SemEval 2016, sedangkan Skor F1 terbaik sebesar 0,678 dalam dataset SemEval 2014 (Laptop). Hasil ini menunjukkan bahwa model yang diusulkan sudah cukup baik, meskipun pada *baseline* diperoleh beberapa hasil yang lebih baik. Lebih lanjut, ditemukan bahwa penggunaan *aspect embedding* mengakibatkan penurunan *performance metrics*. Dengan demikian, diharapkan penelitian selanjutnya dapat memberikan hasil yang lebih baik dengan tidak hanya menggunakan aspek tersebut sebagai alat bantu untuk memperoleh polaritas sentimen dari keseluruhan kalimat, tetapi juga memperhatikan proses perolehan aspek tersebut. Oleh karena itu, disarankan pula untuk mengeksplorasi *task* terkait, seperti *aspect term extraction*.

**Keywords:** Natural Language Processing, Sentiment Prediction, Benchmark Datasets, Multi-modal Graph Representation, Feature Learning