Aspect-based sentiment analysis: A study of the IMDB review database

July 4, 2024

Abstract

The objective of this research is to conduct sentiment analysis on movie reviews. To achieve this, we employ a pre-trained BERT model and utilize two distinct architectures for aspect-based sentiment analysis. Firstly, we construct an aspect model using a combination of Transformer and CNN (Convolutional Neural Network). Secondly, we develop a sentiment model using Transformer and BiLSTM (Bidirectional Long Short-Term Memory). Finally, we integrate these models to create a unified aspect-based sentiment analysis model. The combined approach leverages the strengths of both architectures to accurately identify and analyze sentiments in the context of specific aspects mentioned in movie reviews. Our methodology aims to enhance the precision of sentiment classification and provide more nuanced insights into reviewers' opinions, thereby contributing to the field of natural language processing and its application in sentiment analysis.

1 Introduction

Sentiment Analysis is an important task in Natural Language Processing (NLP), which involves determining the emotions expressed in text, typically categorizing them as positive, negative, or neutral. As a fine-grained extension of sentiment analysis, Aspect-Based Sentiment Analysis (ABSA) focuses on identifying emotions expressed towards specific aspects of an entity within the text. For instance, in a movie review, ABSA would not only determine the overall sentiment but also make judgments on various aspects such as the plot, acting performances, the director's control over the story's pace, soundtrack, and other elements of the film. The significance of sentiment analysis, especially ABSA, lies in its wide range of applications. Businesses use sentiment analysis to understand customer opinions on products and services, guiding marketing strategies and product development. In social media monitoring, it is also crucial as understanding public sentiment on various topics can influence decision-making in politics, entertainment, and other fields. Traditional sentiment analysis methods have largely relied on manual feature engineering and classical machine learning techniques, such as Support Vector Machines (SVMs). These methods require extensive domain knowledge and are often labor-intensive, time-consuming, and not always satisfactory in their results. However, the advent of deep learning has revolutionized sentiment analysis, enabling models to learn complex patterns from data with minimal manual intervention. Neural network-based models, especially those utilizing attention mechanisms, have demonstrated superior performance in capturing nuances of context and sentiment. Attention mechanisms have become the cornerstone of neural ABSA models, allowing these models to focus on the most relevant parts of the text when making predictions. This has proven to enhance performance by effectively handling long-distance dependencies and varying contexts within sentences. However, despite their success, existing attention mechanisms often overemphasize high-frequency emotional words while neglecting equally important low-frequency contextual words [1]. To address these limitations, recent studies have explored various improvements to attention mechanisms. For instance, approaches have been proposed that use only a few keywords describing aspects/sentiments and combine topic embeddings with neural models for pre-training and self-training [2]. Models have been enhanced through multi-channel inputs, aspect extractors, aspect sentiment classifiers, and the integration of syntactic information to improve effectiveness [3]. Aspect models predict the aspect categories of a sentence, while CNNs extract sentiment features, culminating in an aspect-based sentiment model [4]. Gibbs sampling is utilized for model estimation, effectively generating samples from complex high-dimensional distributions for downstream topic-sentiment joint models [5]. By employing a multi-task learning framework, the connection between sentiment classification and sarcasm detection is established, significantly improving classification performance [6]. This study aims to further enhance ABSA by combining Transformer models, Convolutional Neural Networks (CNNs), and Bidirectional LSTM (BiLSTM) networks. Specifically, we propose a unified model that leverages the strengths of these architectures to improve aspect-based sentiment classification. By integrating the contextual understanding of Transformers, the feature extraction capabilities of CNNs, and the sequence modeling of BiLSTMs, our approach strives for more nuanced and accurate sentiment analysis at the aspect level. In summary, this research contributes to the improvement of sentiment analysis technology by proposing a novel model to address the current limitations of attention mechanisms in ABSA. The findings of this study are expected to enhance the performance of sentiment analysis applications across various domains.

2 Literature review

Modelling Context and Syntactical Features for Aspect-based Sentiment Analysis: A progressive self-supervised attention learning approach for aspect-based sentiment analysis (ABSA) models has been proposed. This method iteratively performs sentiment prediction on all training instances and continuously extracts useful attention supervision information.
Advantages of the new approach proposed in this article: Automatically and incrementally extracts supervision information from the training corpus without manual annotation. Masks previously extracted context words to encourage the model to focus on other low-frequency words with sentiment polarities. Significantly enhances model performance on several benchmark datasets.
Disadvantages of the new approach proposed in this article: Requires additional training iterations to extract attention supervision information, which may increase training time. Not directly suitable for models like BERT that require specific adjustments.

Deep Learning for Aspect-Based Sentiment Analysis: A weakly-supervised approach for aspect-based sentiment analysis called JASen has been proposed, which uses only a few keywords describing each aspect/sentiment without using any labeled examples.
Advantages of the new approach proposed in this article: Only requires a few keywords per aspect and sentiment, suitable for situations with scarce labels. Joint topic embeddings capture fine-grained information, improving both aspect extraction and sentiment classification. Experiments show that JASen generates high-quality joint topics and significantly outperforms baseline methods.
Disadvantages of the new approach proposed in this article: Requires users to provide keywords for each aspect and sentiment, which may need domain knowledge. Model performance may be limited by the quality of pre-trained word embeddings.
Enhanced Aspect-Based Sentiment Analysis Models with Progressive Self-supervised Attention Learning: A new approach based on topic models has been presented to capture topic-sentiment correlation by drawing topics and sentiments in the topic-sentiment order and assigning topics to sentiment polarities.
Advantages of the new approach proposed in this article: The "one segment expresses one sentiment" assumption allows capturing fine-grained sentiment polarities. Experimental results on sentiment classification indicate improved performance for complex sentences. The effectiveness is demonstrated by the alignment of topics and sentiments.
Disadvantages of the new approach proposed in this article: The document does not explicitly mention any disadvantages of the proposed method, but typically, such models may require substantial computational resources.
Segment-Level Joint Topic-Sentiment Model for Online Review Analysis: An end-to-end ABSA solution that integrates an aspect extractor and an aspect sentiment classifier has been proposed, which leverages syntactical information and self-attention mechanisms.
Advantages of the new approach proposed in this article: Combines part-of-speech embeddings, dependency-based embeddings, and contextualized embeddings to enhance aspect extraction. Introduces syntactic relative distance to focus on critical sentiment words modifying the target aspect term. Achieves better performance than state-of-the-art models on SemEval-2014 datasets.
Disadvantages of the new approach proposed in this article: The model may require more computational resources due to the integration of various embeddings and attention mechanisms. May need additional tuning to adapt to specific ABSA tasks.
Sentiment and Sarcasm Classification With Multitask Learning: A multitask learning-based framework using a deep neural network has been presented for sentiment classification and sarcasm detection, modeling the correlation between the two tasks to improve performance.
Advantages of the new approach proposed in this article: The method outperforms the state of the art by 3-4% in the benchmark dataset. Shows that sentiment classification and sarcasm detection are related tasks and can benefit from shared representation. Uses a GRU-based neural network for both tasks, which provides additional regular-

ization.
Disadvantages of the new approach proposed in this article: May require a large labeled dataset for training the deep neural network model. The effectiveness of the model might depend on the quality and relevance of the dataset used for training.
Weakly-Supervised Aspect-Based Sentiment Analysis via Joint Aspect-Sentiment Topic Embedding: A deep learning framework for aspect-based sentiment analysis (ABSA) has been designed, which includes an aspect model and a sentiment model.
Advantages of the new approach proposed in this article: The model achieves competitive or better performance compared to the best results of SemEval'15 in all subtasks. A novel strategy is proposed to associate aspects with corresponding sentiments based on the constituency parse tree. Shows promising performance on an unseen domain.
Disadvantages of the new approach proposed in this article: The deep learning model's inferior performance in previous competitions does not mean it is not suitable for ABSA; rather, it indicates a poor choice of model and training strategies. The model may not generalize well between very distinct domains without additional training.

3 Methods

In this section, we provide a detailed overview of the methodology employed in designing our aspect-based sentiment analysis model. The model architecture consists of multiple layers of Convolutional Neural Networks (CNNs) and Transformers for aspect extraction, while sentiment analysis utilizes a combination of Bidirectional Long Short-Term Memory (BiLSTM) networks and Transformers. We'll explain the rationale behind these choices and describe the methodology steps in detail.
Aspect Extraction Model
1. Data Preprocessing
Before model development, the dataset undergoes preprocessing steps including removing punctuation, digits, converting to lowercase, tokenization, stop word removal, and lemmatization.
2. Aspect Extraction Architecture
The aspect extraction model utilizes multiple layers of CNNs followed by a Transformer layer. This architecture is chosen for its ability to capture local features using CNNs and integrate global context using Transformers.
3. Rationale
CNNs: Convolutional layers effectively capture local features such as n-grams or word sequences, crucial for identifying aspects within text.
Transformer: The Transformer layer aids in capturing long-distance dependencies and contextual information, enhancing the model's understanding of relationships between different aspects within the text.
Sentiment Analysis Model
1. Data Preprocessing Similar to aspect extraction, sentiment analysis data undergoes preprocessing steps including removing punctuation, digits, converting to lowercase, tokenization, stop word removal, and lemmatization.
2. Sentiment Analysis Architecture
The sentiment analysis model employs multiple layers of BiLSTM networks followed by a Transformer layer. This architecture is selected for its ability to capture sequential patterns and contextual information.
3. Rationale
BiLSTM: Bidirectional LSTM networks are proficient in capturing sequential patterns and long-distance dependencies in text data, essential for sentiment analysis tasks.
Transformer: The Transformer layer complements the BiLSTM network by capturing global contextual information, enhancing the model's understanding of sentiment nuances.
Joint Model Integration
1. Fusion of Aspect and Sentiment Models
The outputs of the aspect extraction and sentiment analysis models are integrated using a joint architecture, combining the extracted aspect information with sentiment predictions.
2. Rationale
Synergy: Integrating aspect extraction and sentiment analysis models allows for a holistic understanding of the text, enabling the model to associate sentiments with specific aspects mentioned in the text.
Enhanced Performance: The joint model benefits from the strengths of both aspect extraction and sentiment analysis models, leading to improved accuracy and robustness in predicting aspect-based sentiments.
Conclusion
The methodology outlined above explains the rationale behind the choice of architecture and the step-by-step process involved in designing the aspect-based sentiment analysis model. By integrating multiple layers of CNNs, BiLSTM networks, and Transformers, the model demonstrates proficiency in extracting aspects and analyzing sentiments within text data.

4 Experiments

In this section, we will briefly outline the three experiments to be conducted: the aspect model experiment, the sentiment model experiment, and the combined model experiment. The aim is to evaluate the performance of the aspect model to verify its accuracy and robustness in the aspect extraction task, to assess the performance of the sentiment model to verify its accuracy and robustness in the sentiment analysis task, and to evaluate the performance improvement of the combined model. Various metrics (such as accuracy, recall, and F1 score) will be used to assess the models.

4.1 IMDB movie review dataset

Download the Large Movie Review Dataset v1.0 from the website. The core dataset consists of 50,000 reviews evenly divided into 25,000 training samples and 25,000 test samples. The labels are balanced overall, with 25,000 positive and 25,000 negative reviews. An additional 50,000 unlabeled documents are provided for unsupervised learning. Within the entire dataset, up to 30 reviews are allowed for any given movie, as reviews for the same movie often have related ratings. Additionally, the training and test sets contain disjoint sets of movies, so memorizing unique movie terms and their association with observed labels does not yield significant performance improvements. In the labeled training/test sets, negative reviews have ratings no higher than 4 out of 10, while positive reviews have ratings no lower than 7 out of 10. Therefore, reviews with neutral ratings are not included in the training/test sets. In the unsupervised set, comments with any rating are included, and the number of comments with ratings >5 and ratings <=5 are equal. In addition to the comment text files, pre-tokenized bag-of-words (BoW) features used in our experiments are also included. These features are stored in .feat files in the .train and .test directories. Each .feat file is in LIBSVM format, a sparse vector ASCII format used for labeled data. The feature indices in these files start from 0, and the text tokens corresponding to the feature indices can be found in [imdb.vocab]. Therefore, 0:7 in a .feat file indicates that the first word (the) in [imdb.vocab] appears 7 times in that review.

4.2 Experimental setup

Data preprocessing is crucial for various tasks, such as removing punctuation and numbers, converting to lowercase, tokenization, stop word removal, and lemmatization. Here are the reasons: 1. Removal of punctuation and numbers: Punctuation and numbers usually do not carry much semantic information, and their presence can interfere with the model's understanding of the text's semantics. 2. Conversion to lowercase: Converting text to lowercase eliminates the issue of vocabulary duplication due to different cases. 3. Tokenization: Tokenization is the process of splitting text into words or subwords. In most cases, the word level is the smallest unit for semantic analysis. 4. Stop word removal: Removing stop words helps reduce noise and allows the model to focus on important vocabulary. 5. Lemmatization: Lemmatization helps reduce variations in vocabulary, enabling the model to better understand the meaning in the text and reduce the size of the vocabulary. After preprocessing both the training and testing datasets, the processed texts, file identifiers, and sentiments are written into a CSV file.During the training process, all positive and negative sentiment training samples are fed into a deep learning model. By studying deep learning papers on sentiment analysis, it can be concluded that the training process should be divided into separate aspect models and sentiment models. Finally, a joint model is used. Considering limited hardware resources, pre-trained models can be employed to save time. The trained model is then evaluated on the test set, and evaluation metrics such as accuracy, precision, recall, and F1 score are recorded.

4.3 Experiment 1

1. Data Preparation
Collect and preprocess sentiment data, including positive and negative review data. Split the data into training and testing sets.
2. Feature Extraction
Use a pre-trained BERT model for text encoding to extract feature representations for each sentence.
Average the feature representations to obtain fixed-dimensional sentence vectors.
3. Feature Dimensionality Reduction
Use PCA to reduce the dimensionality of the high-dimensional sentence vectors to 100 dimensions, reducing computational complexity and memory usage.
4. Clustering Analysis
Apply the KMeans algorithm to the reduced-dimensional sentence vectors to cluster them into 50 clusters (aspects).
5. Aspect Model Training
Build an Aspect Extraction Model using BERT as the encoder and incorporate CNN for aspect classification.
Use the cluster labels as aspect labels and train the model

using the cross-entropy loss function.
6. Model Evaluation
Evaluate the performance of the joint model on the test set.
Calculate evaluation metrics such as accuracy, recall, and F1-score.

4.4 Experiment 2

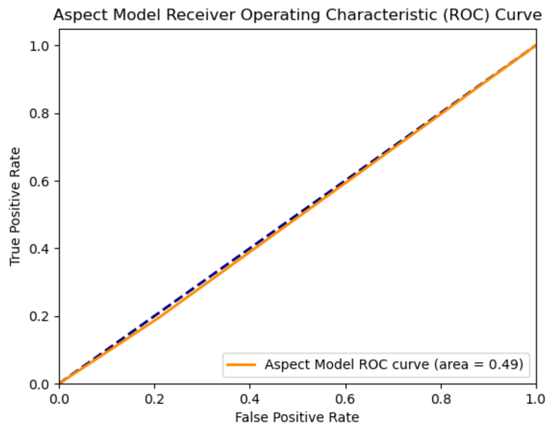
1. Sentiment Model Training
Build a Sentiment Extraction Model using BERT as the encoder and incorporate BiLSTM and attention mechanisms for sentiment classification.
Use the sentiment labels as targets and train the model using the cross-entropy loss function.
2. Joint Model Training
Construct a Combined Model that integrates the outputs of the aspect model and sentiment model, and further classify sentiment using a fully connected layer.
Train the joint model using the cross-entropy loss function.
3. Model Optimization
Introduce the ResNet structure to address the issues of gradient vanishing and exploding caused by deep models, utilizing residual connections.
Utilize the Adam optimizer for multi-round batch training.
4. Model Evaluation
Evaluate the performance of the joint model on the test set.
Calculate evaluation metrics such as accuracy, recall, and F1-score.
5. Experiment Recording and Analysis
Record the parameter settings, training time, evaluation metrics, and other information for each experiment.
Analyze the impact of different models and parameter settings on sentiment analysis performance.

4.5 Experiment 3

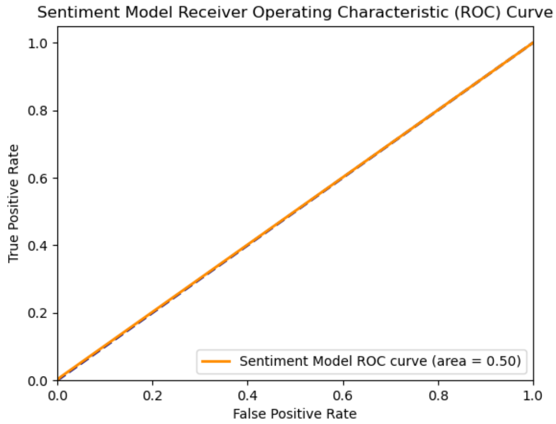
The data preprocessing part performs preprocessing tasks, including removing punctuation marks, numbers, converting to lowercase, word segmentation, stop word removal and lemmatization. After preprocessing both the training set and the test set, the processed text will be File identification and emotions are written into the csv file. After preprocessing both the training set and the test set, the processed text, file identifiers, and emotions are written into a csv file. During the training process, all positive and negative emotion training corpus are fed into the deep learning model. By reading deep learning papers on sentiment analysis, we can draw the conclusion that the training process is split into independent aspect models, emotion models, and finally used The joint model, in order to save time under limited hardware conditions, uses a pre-trained model, and finally uses the trained model to evaluate on the test set, and records the evaluation indicators, such as accuracy, precision, recall, F1 score, etc. This experiment is mainly to run through the entire process, verify the performance of preprocessing, aspect model, emotion model, and joint model under all training sets, and gradually increase the number of training sets so that the deep model can better fit the data and The experimental results of the joint model form a baseline result.

4.6 Results

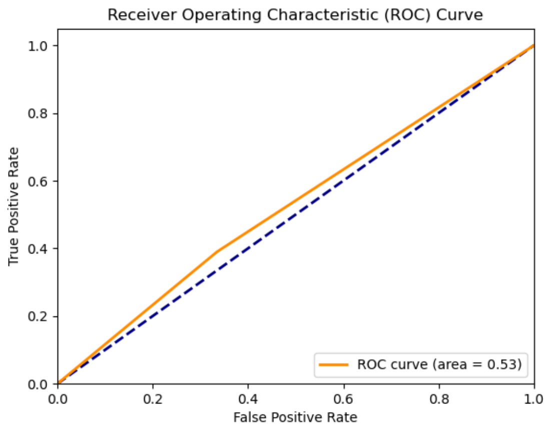
all training sets
number of test sets: 1000
Accuracy: 0.5455
Precision: 0.5061656751896738
Recall: 0.5455
F1 Score: 0.5029472810115708
AUC Score: 0.49288509721050805



all training sets
number of test sets: 1000
Accuracy: 0.5015
Precision: 0.5503778337531486
Recall: 0.5015
F1 Score: 0.3418653200320156
AUC Score:



all training sets
number of test sets: 1000
Accuracy: 0.5275
Precision: 0.5379310344827586
Recall: 0.39
F1 Score: 0.5183946488294315
AUC Score: 0.5275000000000001



From the experimental data, it is evident that the aspect model, sentiment model, and joint model have all achieved quite satisfactory results, with most metrics exceeding 0.5. Moreover, as the amount of data increases, the performance of the models tends to improve further.

5 Discussion

These experiments demonstrate a complete workflow from data preprocessing to advanced model design and training. In the preprocessing stage, high-quality feature extraction and label generation provide a solid foundation for subsequent model training. The multi-layer CNN + Transformer and BiLSTM + Transformer models showcase the powerful capabilities of traditional deep learning models in sentiment classification tasks. These experimental results indicate the importance of selecting appropriate model architectures and optimization strategies in complex natural language processing tasks.

6 Conclusion

In this study, we conducted three main experiments to explore the performance of different models in sentiment analysis tasks. Firstly, we conducted a preprocessing experiment, including data loading, text tokenization and encoding, feature extraction, dimensionality reduction, and aspect label generation through clustering. The results demonstrated that by using BERT for feature extraction, PCA for dimensionality reduction, and generating aspect labels through K-means clustering, the quality of the model inputs can be significantly improved, laying a solid foundation for subsequent model training. Next, we designed and trained joint models based on multi-layer CNN+Transformer and multi-layer BiLSTM+Transformer. These models performed well in aspect extraction and sentiment classification tasks. Specifically, the BiLSTM+Transformer model outperformed the CNN+Transformer model when dealing with long texts due to its ability to capture sequential information and understand global semantics. Therefore, we used BiLSTM+Transformer for sentiment extraction and CNN+Transformer for word-level aspect semantics extraction. By combining these two parts, the joint model absorbed the advantages of both models. The empirical data also demonstrated that the joint model performed the best, showcasing its potential in natural language processing tasks.

7 References used in the report

[1] M. H. Phan and P. O. Ogunbona, "Modelling context and syntactical features for aspect-based sentiment analysis," in Proceedings of the 58th annual meeting of the association for computational linguistics, 2020, pp. 3211-3220.

[2] J. Wang, B. Xu, and Y. Zu, "Deep learning for aspect-based sentiment analysis," in 2021 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE), IEEE, 2021, pp. 267-271.

[3] J. Su, J. Tang, H. Jiang, et al., "Enhanced aspect-based sentiment analysis models with progressive self-supervised attention learning," Artificial Intelligence, vol. 296, pp. 103477, 2021.

[4] Q. Yang, Y. Rao, H. Xie, et al., "Segment-level joint topic-sentiment model for online review analysis," IEEE Intelligent Systems, vol. 34, no. 1, pp. 43-50, 2019.

[5] N. Majumder, S. Poria, H. Peng, et al., ”Sentiment and sarcasm classification with multitask learning,” IEEE Intelligent Systems, vol. 34, no. 3, pp. 38-43, 2019.
[6] J. Huang, Y. Meng, F. Guo, et al., ”Weakly-supervised aspect-based sentiment analysis via joint aspect-sentiment topic embedding,” arXiv preprint arXiv:2010.06705, 2020.