Computer & Robotics

Perturbed Masking for aspect-based sentiment analysis

Akram Karimi Zarandi[a], Sayeh Mirzaei[a,*]

[a] School of Engineering Science, College of Engineering, University of Tehran, Iran
Received 11 October 2024, Accepted 16 December 2024

Abstract

The aspect-based sentiment analysis (ABSA) model has benefited from advancements in pre-trained models, but significant challenges remain in extracting syntactic and semantic information. In this study, we introduce an innovative model that employs the Perturbed Masking method to simultaneously leverage both syntactic and semantic features. Experiments conducted on the SemEval 2014 and Twitter datasets demonstrate meaningful improvements in accuracy and F1 metrics compared to the best previous models. Overall, this model has achieved approximately enhanced F1 scores on the restaurant and laptop datasets, as well as the Twitter dataset, compared to previous models. These results indicate the model's strong capability in accurately discerning sentiments and its flexibility in handling diverse datasets.

***Keywords**:* Aspect-based sentiment analysis, Deep Learning, Transformer, Perturbed Masking, Bert.

1.INTRODUCTION

With the rapid expansion of the internet, users can complete a wide range of activities online, from shopping to social networking, which has led to an exponential growth in unstructured textual data. This development has accelerated the need for efficient sentiment analysis within natural language processing (NLP), as understanding public sentiment offers valuable insights across commercial, political, and academic domains[1, 2]. Sentiment analysis enables companies to assess opinions on products and services, empowering businesses to make data-driven improvements and consumers to make informed decisions. Beyond commercial benefits, sentiment analysis has applications in political and social research, providing insights into public opinion and societal trends.

Motivation:

Sentiment analysis, first introduced in 2004, aims to identify emotional tones within text, traditionally at three levels—document, sentence, and aspect-based sentiment analysis (ABSA) [3, 4]. While document- and sentence-level analyses classify general sentiment as positive, negative, or neutral, ABSA offers a finer-grained approach by targeting specific aspects within text (e.g., assessing the quality of food or service in a restaurant review). This detailed approach is particularly valuable for businesses and organizations aiming to refine specific aspects of their offerings[1, 3]. Table 1 presents examples of ABSA, showing how sentiment is linked to various aspects in user reviews.

Table 1. examples of aspect-based sentiment analysis.

User Review	Aspect	Polarity
I've heard wonderful things about the unique dishes at that restaurant.	Food	Positive
This laptop performs admirably.	Speed	Positive
The food is great, but the service is not good.	Food/Service	Positive/Negative

Challenges in Previous Studies:

Despite significant advancements in sentiment analysis, several challenges remain. Handling diverse writing styles, informal language, and subtleties like sarcasm or irony often obscures the true sentiment, requiring more sophisticated models for accurate analysis. Traditional approaches, including lexicon-based and basic machine learning models, have struggled with these complexities. However, recent breakthroughs with pre-trained models (PTMs) like BERT[5, 6], have significantly enhanced the performance of sentiment analysis, particularly for ABSA tasks, by capturing both syntactic and semantic nuances of language. Nonetheless, integrating syntactic structures into PTMs remains an area with room for improvement.

*Corresponding Author. Email: s.mirzaei@ut.ac.ir

Research Contributions:

This study introduces a novel approach that combines PTMs with fine-tuning strategies to enhance ABSA performance. By leveraging the Perturbed Masking method, we extract syntactic information, allowing our model to harness both syntactic and semantic features of text. Our approach provides a comprehensive analysis of sentiment by improving the model's understanding of word relationships and sentence structure, leading to more accurate sentiment predictions. Our model is evaluated on benchmark datasets, including SemEval 2014 restaurant and laptop reviews, and a Twitter dataset, demonstrating its robustness and applicability across diverse domains.

Key Contributions of This Work:

- Enhanced Sentiment Analysis Using BERT: Leveraging pre-trained language models to improve sentiment analysis accuracy.
- Integration of Syntactic Information via Perturbed Masking: Allowing simultaneous utilization of syntactic and semantic data for a deeper analysis.
- Robustness Across Datasets: Extensive evaluation on diverse datasets, showcasing the model's adaptability and effectiveness in various contexts.

In summary, this work contributes to the field of sentiment analysis by advancing ABSA through the innovative use of pre-trained models and syntactic analysis techniques. The Perturbed Masking method exemplifies a meaningful step towards improved sentiment classification, with potential implications across multiple industries. The following sections elaborate on the methodology, experimental setup, and results, highlighting this study's contribution to the advancement of sentiment analysis.

1. RESEARCH BACKGROUND

In recent decades, emotion detection in text has become a critical and widely used task in the field of NLP[1]. Since the 1980s, the advent of machine learning algorithms has fundamentally transformed research focused on NLP. These algorithms initially treated text as a mere collection of words, often concentrating on word frequency and distribution rather than grammatical structures. While these models achieved notable success in tasks such as machine translation, they frequently struggled to capture the intricate meanings and contextual subtleties embedded in human language[1, 7].

From 2010 onwards, significant advancements in word representation techniques such as Word2Vec and GloVe emerged, converting words into fixed-dimensional vectors, which allowed for a more nuanced understanding of word semantics and relationships. This shift marked a critical development in NLP, enabling models to better capture semantic similarities and contextual meanings [1, 6, 7].

Following this, various neural network architectures, including recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and convolutional neural networks (CNNs), were employed to model text more effectively. These architectures facilitated improvements in tasks such as sentiment analysis, information extraction, and question-answering systems[3]. The rise of deep learning methodologies further enhanced the performance of numerous NLP tasks, revolutionizing the way machines understand and process language[6].

A pivotal moment in this progression came in 2018 with the introduction of BERT[5]. This neural network-based technique significantly improved the performance of a wide array of NLP tasks, often surpassing human-level performance. BERT leverages context-sensitive word embeddings that dynamically adjust based on surrounding words, enabling it to capture multiple meanings and interpretations of words in various contexts. This capability is especially important for tasks such as sentiment analysis, where sentiment can drastically change depending on contextual cues [5, 6].

Recent research, such as the work by May et al., underscores the importance of integrating syntactic and semantic information for ABSA. Models such as Relational Graph Attention Networks (R-GAT) and Knowledge-Aware Dependency Graph Networks (KDGN) have aimed to enhance sentiment analysis by incorporating syntactic knowledge and domain-specific information [8, 9]. Advanced models like MWGCN and AG-VSR ABSA have further refined sentiment prediction by leveraging multi-layer graph neural networks and innovative weighting methods for local and aspect-related attributes[10, 11].

Consequently, our proposed model builds upon transformer-based architectures, specifically BERT, to improve emotion and sentiment detection in text. By utilizing context-sensitive embeddings and advanced training techniques in BERT, our model aims to enhance the accuracy and effectiveness of sentiment analysis, providing deeper insights into the emotional undertones of textual data.

2. METHODOLOGY

2.1 BERT

The BERT model, first released in 2018, is built on transformer networks. Its architecture comprises multiple encoder layers, which generate a vector representation for each input word. There are two main versions of BERT: BERT Base, featuring 16 transformer layers, and BERT Large, containing 24 layers. We utilize BERT Base in our research[3, 5, 12].

BERT can be trained in two modes: masked language model and next sentence prediction, or a combination of both. In the masked language model mode, BERT learns to predict missing words in a sentence based on the surrounding context. In the next sentence prediction mode, it determines whether two given sentences are consecutive in the original text. By integrating both training modes, BERT develops a comprehensive understanding of the relationships between words and sentences, allowing it to effectively capture nuanced semantic meanings and syntactic structures[5, 13].

The tokens used in the input data serve as the foundational elements of BERT's language understanding capabilities, enabling it to process and analyze textual information with remarkable accuracy and efficiency. Throughout the training process, the input data comprises the following essential tokens[3]:

- **[CLS]**: Represents the entire text.
- **[SEP]**: Separates two segments of text.
- **[MASK]**: Used to replace tokens for prediction.

After pre-processing the input data according to the masked language model or next sentence prediction tasks, we can fine-tune BERT for specific applications.

2.2 Perturbed Masking method

The Perturbed Masking technique aims to identify syntactic information from pre-trained models in an unsupervised manner, as proposed in[14]. This method allows for the extraction of syntactic trees from the pre-trained model without introducing additional parameters. Overall, the Perturbed Masking approach is applicable to a wide range of PTMs, and we have utilized this method in our work. In this section, we will outline the core concept of this technique.

Theis technique operates by introducing controlled perturbations to the input tokens and observing the resulting changes in the model's predictions. By systematically altering the input in this manner, the technique can uncover syntactic structures encoded within the model. This unsupervised approach allows for the extraction of valuable syntactic information without the need for additional labeled data. Furthermore, the parameter-free nature of this method enhances its versatility across various pre-trained models.

2.2.1 Core Concept of Perturbed Masking

In general, for a given sentence $X = \{x_{1,\dots,}x_T\}$, PTMs generate a textual representation $H_\theta(X)_i$, as shown in Figure 1. When these representations are obtained, a network (such as a multi-layer perceptron) is typically used to predict whether there is a dependency relationship between pairs of words in the given sentence (where one label indicates the existence of a dependency and another indicates its absence). Training such a network requires learning many parameters[14].

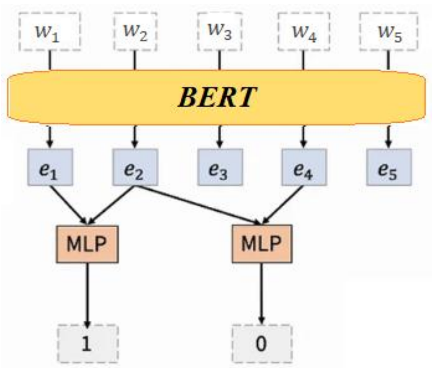

Figure 1: Detection of dependency relationships between pairs of words using BERT.

The Perturbed Masking technique aims to achieve similar results with fewer parameters. In the first step, the value of $f(x_i,x_j)$ is obtained, which indicates the effect that token x_j has on another token x_i. To extract this value, the word "[MASK]" is used to replace token x_i, which returns a representation $H_\theta(X\backslash\{x_i\})_i$ for the masked x_i. The second token is then masked, returning a representation $H_\theta(X\backslash\{x_i,x_j\})_i$ in which both x_i and x_j are masked. For example, **Figure 2** demonstrates how the impact of token x_2 on the ratio of token x_1 is determined by this method.

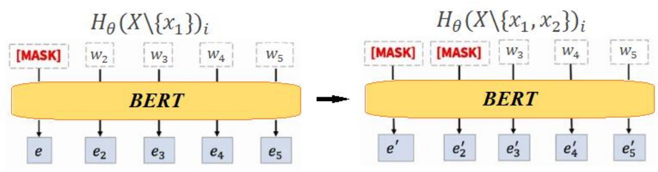

Figure 2: An example of how Perturbed Masking works.

Finally, the value of $f(x_i,x_j)$ is calculated using the Euclidean distance as follows[14]:

$$f(x_i,x_j) = \left\| H_\theta(X\backslash\{x_i\})_i - H_\theta(X\backslash\{x_i,x_j\})_i \right\|_2 \qquad (1)$$

By repeating this process for both tokens in the sentence, an Impact Matrix in the form of $M \in R^{T*T}$ is obtained, as shown in Figure 3, where $M_{i,j} = f(x_i,x_j)$.

	w_1	w_2	w_3	w_4	w_5
w_1	-	$f(x_1,x_2)$	...	...	
w_2	$f(x_2,x_1)$	-			
w_3	...		-		
w_4				-	
w_5					-

Figure 3: Impact Matrix.

2.2.2 Tree Decoding and Application

Once the impact matrix is obtained, a tree decoding algorithm such as Eisner's algorithm is used to extract the dependency tree from the Impact matrix, as shown in Figure 4.

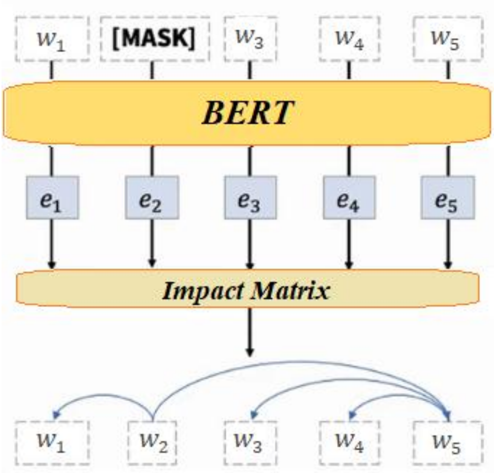

Figure 4: Induced tree resulting from Perturbed Masking.

This technique can be applied to any layer of PTMs, such as BERT or RoBERTa. Experiments in [14] have demonstrated that syntax trees recovered from BERT using this method significantly outperform those generated by traditional methods.

2.2.3 Summary of Key Benefits

- **Parameter Efficiency**: The method requires fewer parameters compared to traditional approaches.
- **Unsupervised**: No labeled data is needed to uncover syntactic structures.
- **Versatility**: Applicable across multiple PTMs like BERT and RoBERTa.
- **Improved Performance**: Experiments indicate significant improvement in syntactic tree accuracy when using this technique on models like BERT.

This makes Perturbed Masking an efficient and powerful approach for syntactic information extraction from pre-trained models.

2.3 Our Model

In aspect-based sentiment analysis (ABSA), most contemporary models use dependency trees to map relationships between aspects and sentiment-related words, integrating these connections within neural network architectures. Although traditional dependency trees are common, induced trees—especially those derived from pre-trained models such as BERT—have led to notable performance gains in ABSA tasks. Pre-trained models like BERT inherently capture dependency structures, adding crucial contextual insights valuable for sentiment analysis.

Recent research consistently demonstrates that transformer architectures, especially BERT, deliver impressive results across various sentiment analysis tasks. BERT's success in numerous natural language processing (NLP) applications makes it an ideal foundation for our model. The architecture of our proposed model is illustrated in Figure 5.

Studies such as [15] highlight the superior performance of tree structures induced by pre-trained transformers (PTMs) over those produced by conventional dependency parsers. This induced tree, extracted from PTMs, serves as a key component in our approach. The Perturbed Masking technique, as discussed earlier, allows us to extract essential syntactic information from the BERT model, which is then leveraged in our proposed model.

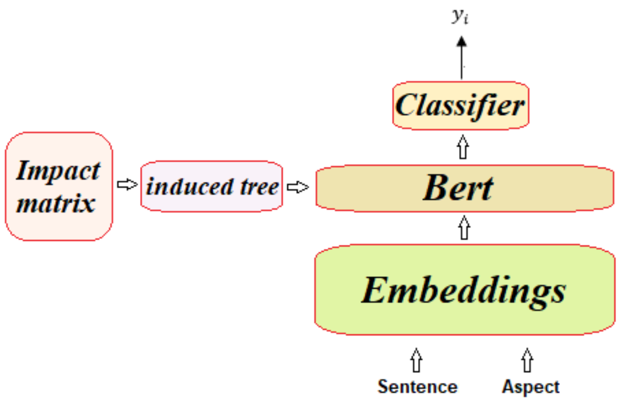

Figure 5: Architecture of the proposed model.

2.3.1 Model Architecture

In our model, we first derive the BERT-induced tree using the Perturbed Masking method. This induced tree, along with the sentence and its corresponding aspect, is then fed into a fine-tuned BERT model. By using both the induced tree and the sentence-aspect pair, we enhance the model's comprehension of syntactic relationships, which improves its ability to understand sentiment connections in the text.

This integration of syntactic structures derived from the induced tree, combined with BERT's rich contextual knowledge, enables our model to capture intricate relationships between various aspects and their associated sentiments. As a result, our model not only improves sentiment classification accuracy but also exhibits robustness across diverse ABSA tasks.

Furthermore, our approach integrates the BERT-induced tree directly into the neural framework, eliminating the need for complex external parsers and reducing model overhead. This streamlined architecture maintains the depth of syntactic understanding necessary for high-performance sentiment analysis

3. Experiment

3.1 Datasets

We conduct our experiments on three widely recognized English datasets: the SemEval 2014 restaurant and laptop datasets[16], along with a Twitter dataset[17]. The SemEval 2014 datasets comprise user comments from the restaurant and laptop domains, providing a rich source of insights for sentiment analysis in these specific contexts. This dataset is particularly valuable as it reflects the opinions and sentiments of users regarding their experiences in dining and product usage, enabling the examination of how sentiment manifests across different contexts.

The Twitter dataset, on the other hand, includes tweets that capture a diverse array of sentiments related to various aspects of life. This breadth makes the Twitter dataset especially useful for understanding the dynamics of sentiment in informal and real-time communications, where linguistic nuances can significantly impact sentiment interpretation.

Each entry in these datasets is annotated with positive, negative, or neutral polarity, facilitating a comprehensive analysis of sentiment across multiple domains. The distribution of these annotations is summarized in Table 2, which outlines the number of examples corresponding to each category.

Table 2. Statistics of Datasets

Dataset		#Pos.	#Neu.	#Neg.
Rest 14	***Train***	***2164***	***637***	***807***
	Test	***728***	***196***	***196***
Lap14	***Train***	***994***	***464***	***870***
	Test	***341***	***169***	***128***
Twitter	***Train***	***1561***	***3127***	***1560***
	Test	***173***	***346***	***173***

Figure 6 illustrates an example from the dataset where syntactic dependencies have not yet been extracted. This particular example highlights a sentence with multiple words; however, the relationships and connections between the words remain unrepresented, which limits the depth of sentiment analysis.


```
▼<sentences>
  ▼<sentence id="32897564#894393#2">
     <text>The bread is top notch as well.</text>
    ▼<aspectTerms>
       <aspectTerm term="bread" polarity="positive" from="4" to="9"/>
     </aspectTerms>
    ▼<aspectCategories>
       <aspectCategory category="food" polarity="positive"/>
     </aspectCategories>
```


Figure 6. Example of the Dataset

In contrast, Table 3 provides an example of the dataset after preprocessing, which includes syntactic information ready for embedding. The preprocessing steps are crucial for enhancing the dataset's usability, allowing for a more nuanced understanding of sentiment through the identification of syntactic structures.

Table 3. Example of a Dataset After Preprocessing

token	**["The " ," bread " ," is " ," top " ," notch " ," as " ," well "," ."]**
pos	["DET"," NOUN"," AUX"," ADJ"," NOUN"," ADV"," ADV"," PUNCT"]
head	[2,4,4,0,4,7,4,4]
deprel	["dep"," nsubj"," cop"," root"," dep"," advmod"," dep"," punct"]
aspects	[{"term": ["bread "],"from": 1,"to": 2,"polarity": " positive"]

The tokenization, part-of-speech tagging, dependency parsing, and aspect extraction outlined above are fundamental processes in preparing the data for effective sentiment analysis. By transforming raw data into a structured format, we can leverage advanced techniques in NLP to enhance sentiment classification accuracy.

3.2 Experiment Setup

In this study, 300-dimensional word embedding vectors were initialized using pre-trained GloVe embeddings, and the BERT model was fine-tuned on three specific datasets: Rest14, Lap14, and Twitter, to ensure domain-specific contextual understanding. To prevent overfitting, a dropout rate of 0.1 was applied, and the model was trained for 40 epochs with a batch size of 32. The learning rate was set to 2e-4, and the AdamW optimizer was used with default parameters.

The hyperparameter configurations were selected based on common practices for similar models to ensure convergence and optimal performance. The model's structure also followed the standard BERT setup, with primary parameters set to commonly accepted defaults.

tion.

3.3 Experiment Results and Model Comparison

In this section, we present the outcomes of our proposed model alongside a comparative analysis against several baseline models, as shown in Table 4. All models in the comparison utilized the same dataset, ensuring a fair evaluation across the same experimental conditions.

Our proposed model's performance, especially when benchmarked against baseline models, demonstrates substantial improvements in both accuracy and F1 scores across all datasets. These enhancements can be attributed to the model's integration of syntactic structures through the induction tree, a feature introduced in earlier sections. Notably, when evaluated without the induction tree, the model's scores were noticeably lower. This result aligns with the theory that syntactic information enhances the model's contextual understanding, supporting more nuanced sentiment classification.

Table ٤: Results and Comparison with Other Models

Models	Rest14		Laptop14		Twitter	
	Acc	F1	Acc	F1	Acc	F1
IAN [18]	78.60	73.12	72.10	73.64	-	-
R-GAT [8]	83.30	76.02	77.42	73.76	75.57	73.82
MGAN [19]	81.25	71.49	75.39	72.47	72.54	70.81
AF-LSTM [20]	77.13	72.85	72.32	68.21	66.60	60.82
MWGCN [10]	82.56	74.58	76.36	72.28	72.86	70.73
BERT-SPC [5]	84.11	76.68	77.59	73.28	75.18	74.01
AG-VSR [11]	86.34	80.88	79.22	75.85	76.45	75.04

KDGN [9]	87.01	81.94	81.32	77.59	77.64	75.55
Our proposed model						
FT_Bert	**87.66**	**79.99**	81.23	76.20	**78.66**	**76.38**
Our proposed model + FT_Bert	**88.02**	**81.93**	**81.73**	**78.99**	**78.93**	**77.96**

The results in Table 4 show that our model, especially when fine-tuned with BERT and the induction tree, outperformed baseline models across all datasets. This performance improvement is grounded in the model's capability to capture both syntactic and semantic features, as described in Section 3.2 on the Perturbed Masking method. This method induces a syntactic structure within the model's contextual embeddings, thereby reinforcing the relationship between sentiment-bearing terms and their syntactic roles, which enhances aspect-specific sentiment prediction.

3.3.1 PERFORMANCE TRENDS OVER TIME

To further understand performance trends, we conducted more than 40 training sessions across the "Restaurant," "Laptop," and "Twitter" datasets, focusing on both accuracy and F1 scores. Below is a breakdown of the performance patterns observed across each dataset, supported by analysis based on training dynamics and stability.

Restaurant Dataset:
The Restaurant dataset demonstrated a consistent upward trend, with accuracy and F1 scores steadily increasing from an initial accuracy of 84.28% and F1 score of 78.28% to final values of 88.02% and 81.93%, respectively. This stability supports the theoretical advantage of incorporating syntactic structures, as the model's semantic understanding is reinforced over time, particularly for contextually rich datasets like Restaurant.

Laptop Dataset:
The Laptop dataset, characterized by more diverse linguistic styles, showed fluctuations in accuracy and F1 scores during training. Beginning with 76.01% accuracy and 73.56% F1, the model eventually achieved 81.73% accuracy and 78.99% F1. The variability observed aligns with the model's adaptation process in dealing with broader vocabularies and styles, highlighting its capability to generalize over time.

Twitter Dataset:
The Twitter dataset initially presented moderate accuracy and F1 values, starting at 74.52% accuracy and 73.52% F1, and reached 78.93% accuracy and 77.96% F1 by the end of training. While growth was less stable, the model's ability to improve despite the informal language typical of Twitter highlights its robustness. This dataset illustrates the model's flexibility, adapting to complex expressions such as sarcasm, which were challenging for baseline models.

The patterns across these datasets validate the model's theoretical framework introduced in prior sections, where the inclusion of both syntactic and semantic data is designed to yield robust results across domains. This approach, grounded in syntactic reinforcement, ultimately enhances our model's ability to handle diverse linguistic challenges more effectively than traditional sentiment analysis methods.

3.3.2 VISUAL REPRESENTATION OF PERFORMANCE

To better illustrate these findings, Figures ٧ and ٨ show the changes in accuracy and F1 scores over time for all three datasets. These graphs offer a clear visual representation of the models' evolution during the training process.

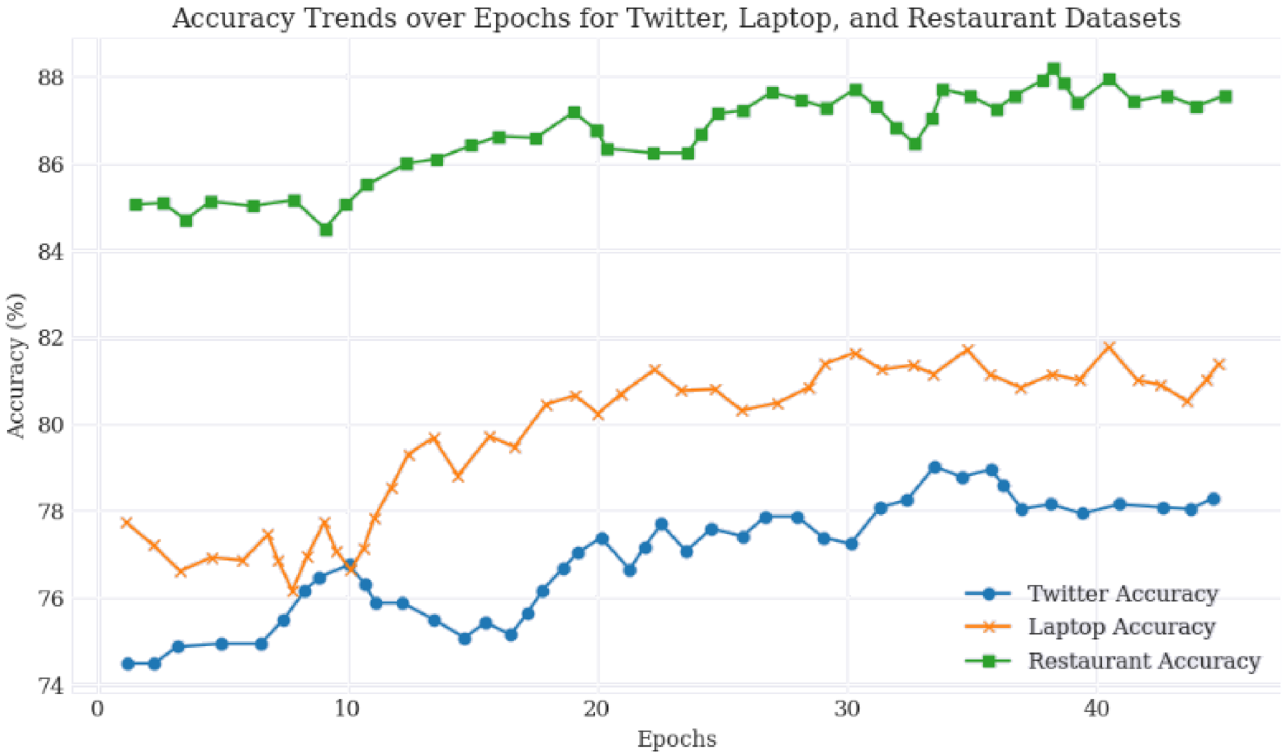

Figure ۷: Changes in Accuracy Over Time for Different Datasets.

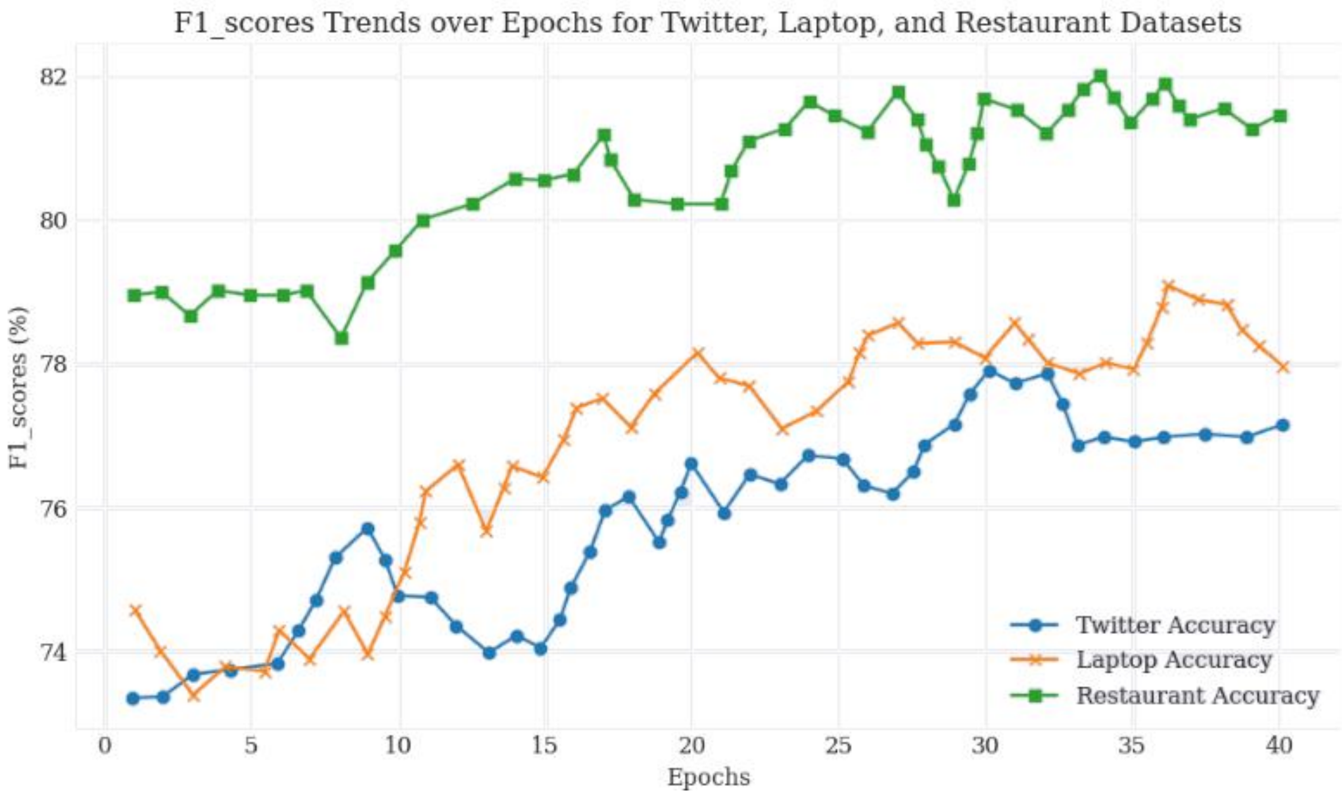

Figure ۸: Changes in F1 Scores Over Time for Different Datasets.

These figures provide a clear visual representation of the improvement in both accuracy and F1 scores across the "Restaurant," "Laptop," and "Twitter" datasets throughout the training process. The steady upward trends highlight the effectiveness of our model, particularly with the integration of the induction tree, in achieving consistent and enhanced performance over time.

4. Conclusion

In this study, we explored the application of the Perturbed Masking technique within pre-trained transformer models, specifically BERT, for ABSA. Our approach leveraged syntactic information derived from induced dependency trees, resulting in substantial performance improvements. The proposed model, incorporating BERT-induced trees, achieved notable enhancements, with accuracy and F1 scores consistently surpassing baseline models across multiple datasets, including SemEval 2014 and Twitter. For instance, our model improved accuracy by up to 1.2% and F1 score by 1.3% on average compared to existing methods. These results validated the effectiveness and robustness of our approach in a variety of ABSA tasks.

For future research, further refinements could focus on optimizing the syntactic extraction process and testing the model in additional domains to enhance its applicability.

.

5. References

1. Birjali, M., M. Kasri, and A. Beni-Hssane, *A comprehensive survey on sentiment analysis: Approaches, challenges and trends.* Knowledge-Based Systems, 2021. **226**: p. 107134.
2. Chauhan, G.S., et al., *Aspect based sentiment analysis using deep learning approaches: A survey.* Computer Science Review, 2023. **49**: p. 100576.
3. Trisna, K.W. and H.J. Jie, *Deep learning approach for aspect-based sentiment classification: a comparative review.* Applied Artificial Intelligence, 2022. **36**(1): p. 2014186.
4. Hadi, M.U., et al., *A survey on large language models: Applications, challenges, limitations, and practical usage.* Authorea Preprints, 2023.
5. Devlin, J., *Bert: Pre-training of deep bidirectional transformers for language understanding.* arXiv preprint arXiv:1810.04805, 2018.
6. Yusuf, K.K., et al., *A technical review of the state-of-the-art methods in aspect-based sentiment analysis.* Journal of Computing Theories and Applications, 2024. **1**(3): p. 287-298.
7. Nath, D. and S.K. Dwivedi, *Aspect-based sentiment analysis: approaches, applications, challenges and trends.* Knowledge and Information Systems, 2024: p. 1-43.
8. Wang, K., et al., *Relational graph attention network for aspect-based sentiment analysis.* arXiv preprint arXiv:2004.12362, 2020.
9. Wu, H., C. Huang, and S. Deng, *Improving aspect-based sentiment analysis with Knowledge-aware Dependency Graph Network.* Information Fusion, 2023. **92**: p. 289-299.
10. Yu, B. and S. Zhang, *A novel weight-oriented graph convolutional network for aspect-based sentiment analysis.* The Journal of Supercomputing, 2023. **79**(1): p. 947-972.
11. Feng, S., et al., *Aspect-based sentiment analysis with attention-assisted graph and variational sentence representation.* Knowledge-Based Systems, 2022. **258**: p. 109975.
12. Pota, M., et al., *Multilingual evaluation of pre-processing for BERT-based sentiment analysis of tweets.* Expert Systems with Applications, 2021. **181**: p. 115119.
13. Zarandi, A.K. and S. Mirzaei, *A survey of aspect-based sentiment analysis classification with a focus on graph neural network methods.* Multimedia Tools and Applications, 2024. **83**(19): p. 56619-56695.
14. Wu, Z., et al., *Perturbed masking: Parameter-free probing for analyzing and interpreting BERT.* arXiv preprint arXiv:2004.14786, 2020.
15. Dai, J., et al., *Does syntax matter? a strong baseline for aspect-based sentiment analysis with roberta.* arXiv preprint arXiv:2104.04986, 2021.
16. Kirange, D., R.R. Deshmukh, and M. Kirange, *Aspect based sentiment analysis semeval-2014 task 4.* Asian Journal of Computer Science and Information Technology (AJCSIT) Vol, 2014. **4**.
17. Rosenthal, S., N. Farra, and P. Nakov, *SemEval-2017 task 4: Sentiment analysis in Twitter.* arXiv preprint arXiv:1912.00741, 2019.
18. Ma, D., et al., *Interactive attention networks for aspect-level sentiment classification.* arXiv preprint arXiv:1709.00893, 2017.
19. Fan, F., Y. Feng, and D. Zhao. *Multi-grained attention network for aspect-level sentiment classification*. in *Proceedings of the 2018 conference on empirical methods in natural language processing*. 2018.
20. Tay, Y., L.A. Tuan, and S.C. Hui. *Learning to attend via word-aspect associative fusion for aspect-based sentiment analysis*. in *Proceedings of the AAAI conference on artificial intelligence*. 2018.