

Article

Not peer-reviewed version

Aspect-Targeted Opinion Word Extraction with Aspect Subword Segmentation

Jacob Frank^{*}, [Rodolfo Patel](#), Cleva Terry

Posted Date: 18 December 2023

doi: 10.20944/preprints202312.1280.v1

Keywords: Sentiment Analysis; Opinion Words Extraction; Syntax Feature



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Aspect-Targeted Opinion Word Extraction with Aspect Subword Segmentation

Jacob Frank *, Rodolfo Patel and Cleve Terry

Briar Cliff University

* Correspondence: frank@briarcliff.edu

Abstract: Contemporary advanced models for aspect-targeted opinion word extraction (ATOWE), which predominantly utilize BERT-based encoders at a word level, have shown limited advancements when integrated with graph convolutional networks (GCNs) for syntactic tree assimilation. Recognizing the prowess of BERT subwords in encapsulating rare or context-poor words, this study pivots from syntactic trees to BERT subwords, omitting GCNs from the structural framework. Our approach, named Aspect-Enhanced Wordpiece Extraction Model (AEWEM), focuses on refining aspect representation during encoding. We propose an input format of paired sentence-aspect, diverging from traditional single-sentence inputs. AEWEM demonstrates superior performance on benchmark datasets, establishing a robust foundation for future explorations in this domain.

Keywords: sentiment analysis; opinion words extraction; syntax feature

1. Introduction

Aspect-Targeted Opinion Word Identification (ATOWE; Fan et al. [1]) is a nuanced sub-domain within Aspect-Based Sentiment Analysis (ABSA; Pontiki et al. [2]). It zeroes in on pinpointing words that articulate opinions about specified targets within a sentence. For example, in “The *delightful* pastry was a treat.”, ATOWE aims to identify “*delightful*” as the opinion word about the aspect **pastry**. This explicit pairing of aspect and opinion aids in enhancing downstream tasks such as opinion summarization [3] and information retrieval [2,4–8].

Contemporary ATOWE methodologies [9,10] often employ BERT [11] for sentence encoding. BERT’s contextual awareness significantly bolsters ATOWE capabilities. Yet, many existing approaches, despite their intricacy, yield marginal improvements. They typically incorporate syntax trees via a graph convolutional network (GCN) [14]. For example, Veyseh et al. [9] integrates an ordered-neuron LSTM [15] with a GCN, while Jiang et al. [20] employs an attention-focused relational GCN. Mensah et al. [10] combines BERT embeddings with a BiLSTM [21,22] and a syntax-aware GCN.

Despite these advancements, the integration of syntactic information via GCNs in ATOWE often results in limited improvement [10]. Additionally, the incorporation of subword tokens, such as BERT wordpieces, into GCN models presents challenges due to syntax trees being comprised of whole words. Subword-based models, balancing between character- and word-level encoding, excel in representing rare or contextually sparse words. Refer to Table ?? for an illustration where “##cake” in “cheese##cake” and “cup##cake” provides valuable contextual linkages, enhancing understanding of both aspects.

In our study, we introduce the Aspect-Enhanced Wordpiece Extraction Model (AEWEM), significantly simplifying syntax-aware ATOWE models [9,10,20,23,24] by substituting syntax trees with subword data, while maintaining robust predictive performance. AEWEM addresses the challenge of aspect representation loss during encoding, a factor that impairs ATOWE efficacy, by introducing a novel input structure: a sentence-aspect pair instead of a single sentence. Our experimental results corroborate AEWEM’s effectiveness, with it outperforming models like Mensah et al. [10] without necessitating a GCN component.

Our primary contributions are:

1. Introducing AEWEM, which innovatively utilizes BERT wordpieces to enhance the representation of rare and out-of-vocabulary words in ATOWE.

2. Proposing the use of sentence-aspect pairs as input, replacing the conventional single-sentence format, to mitigate aspect information loss during encoding.
3. Conducting extensive experiments, our results validate AEWEM's state-of-the-art performance in the ATOWE domain.

2. Related Works

The field of Aspect-Based Sentiment Analysis (ABSA), a vital area of natural language processing, encompasses the specialized task of target-oriented opinion word extraction, a concept initially brought to light by [1]. This emergent sub-discipline has primarily been explored through the lens of neural network architectures. Here, we provide a comprehensive review of both foundational and recent advances in this field, setting the stage for the introduction of our Aspect-Enhanced Wordpiece Extraction Model (AEWEM).

The pioneering work by Fan et al. [1] marked a significant milestone in ABSA. They introduced the Inward-Outward LSTM, a novel approach that seamlessly integrates the target information within the sentence context. This integration, combined with a global contextualized representation formed through a BiLSTM, laid the groundwork for effective sequence labeling in TOEWE. This methodology not only set a new standard in the field but also opened avenues for subsequent research focusing on target-context interplay.

Building upon this, Wu et al. [27] embarked on a novel direction by harnessing latent opinion knowledge from rich review sentiment datasets. This transfer learning approach effectively mitigated the resource scarcity in TOEWE, showcasing the potential of leveraging cross-domain knowledge transfer in ABSA.

The advent of BERT embeddings marked a transformative phase in TOEWE. A line of studies [9, 10, 20, 28, 29] adopted BERT for its exceptional ability to capture contextual nuances. The integration of Graph Convolutional Networks (GCNs) further enhanced this by adding a layer of syntactic understanding. For instance, Veyseh et al. [9] innovatively merged an ordered-neuron LSTM for contextual dynamics with a GCN to incorporate syntax, creating a more holistic representation of sentences. Similarly, Jiang et al. [20] designed an attention-based relational GCN, leveraging the structural richness of syntax trees for encoding, thereby enriching the semantic interpretation of sentences in ABSA.

Mensah et al. [10] introduced a unique blend of BiLSTM with relative position embeddings and BERT embeddings. This amalgamation was designed to capture the intricate positional interrelations of words concerning the aspect, subsequently enhanced by the syntactic enrichment offered by a GCN. This model stood out for its ability to intertwine positional and syntactic information, paving the way for more nuanced aspect-opinion pair extractions.

While syntax-based methods have been predominant, recent explorations have ventured into new territories. The study by [29] marked a departure from syntax reliance, adopting self-attention mechanisms to directly target the identification of opinion words. This approach emphasized the intrinsic ability of models to discern relevant opinions without the scaffolding of syntactic structures.

Further broadening the scope, Gao et al. [32], Wu et al. [33] redefined the TOEWE task as akin to a machine reading comprehension challenge. By formulating the problem as a question-answer paradigm, they leveraged the inherent interrogative nature of language processing, opening new pathways for understanding aspect-opinion dynamics in text.

In contrast to these varied methodologies, our work introduces the Aspect-Enhanced Wordpiece Extraction Model (AEWEM). AEWEM diverges from the complex architectures of prior models by adopting a streamlined approach focused on semantic word representations. Utilizing a blend of BERT and BiLSTM networks, AEWEM addresses the intricacies of target-oriented opinion word extraction with a novel perspective. Its simplicity in design does not compromise its efficacy; instead, it facilitates easier deployment in real-world scenarios. Moreover, AEWEM's state-of-the-art performance, as

evidenced by our extensive experiments, underscores its potential as a new benchmark in the field of ABSA.

3. Methodology

3.1. Task Formalization

In this paper, we redefine the Target-Oriented Opinion Word Extraction (TOWE) task. Here, the objective is to discern an opinion word in a given sentence $S = \{w_1, \dots, w_{n_s}\}$, particularly focusing on a specified aspect $w_a \in S$. The sentence is broken down into tokens $T = \{t_1, \dots, t_{n_t}\}$, varying in granularity from subwords to full words. Each token t_a represents a fragment or entirety of the aspect w_a , with the relationship $n_s \leq n_t$. The core task involves labeling each token according to the IOB scheme [34], designating whether it falls Inside, Outside, or at the Beginning of the opinion word relative to the aspect.

3.2. Baseline Approaches to TOWE

Existing TOWE methodologies, particularly those incorporating syntax-aware strategies [9,10,20], primarily utilize BERT [11] for text encoding. These approaches are augmented with position [35] or category embeddings [20] to achieve aspect-aware whole word representations. A Graph Convolutional Network (GCN) is typically employed to fuse syntactic information into the overall model structure.

Ordered-Neuron LSTM GCN (ONG):

The ONG model, as proposed by Veyseh et al. [9], integrates an ordered-neuron LSTM (ON-LSTM; Shen et al. [15]) with a GCN. The ON-LSTM layer, an innovative variant of LSTM, processes the input sequence, including BERT and position embeddings, and models the dependencies therein. The GCN component then enriches these representations with syntactic structural insights.

BERT+BiLSTM+GCN:

In a variation of the ONG model, Mensah et al. [10] employed a BiLSTM in place of the ON-LSTM. This adaptation is aimed at more effectively capturing the immediate dependencies between aspect and opinion words, thereby refining the extraction process.

Attention-based Relational GCN (ARGCN):

The ARGCN model by Jiang et al. [20] merges BERT's contextualized embeddings with category embeddings, specifically IOB tags, to enrich aspect information. This model then utilizes a relational GCN [36], in conjunction with a BiLSTM, to simultaneously incorporate syntactic and sequential data into TOWE.

3.3. Our Methodology: AEWEM

Recent findings by Mensah et al. [10] indicate that incorporating a GCN for syntax tree information has a marginal impact on TOWE performance. Additionally, the integration of subword tokens, a key feature in modern NLP, poses a challenge in GCN-based models. Addressing these challenges, AEWEM proposes a streamlined TOWE model that eschews the GCN component and instead leverages subword tokens, particularly BERT's Wordpieces [11], for input processing. This approach, while simpler, capitalizes on the rich semantic information embedded in subword representations.

3.4. Formatting BERT Input for AEWEM

To format the input for BERT in AEWEM, we first segment the sentence S into a sequence of wordpieces $T = \{t_1, t_2, \dots, t_{n_t}\}$. The structured BERT input is then represented as:

$$T^{(S)} = \{[\text{CLS}], T, [\text{SEP}]\} \quad (1)$$

where [CLS] and [SEP] are special tokens denoting the start and end of the sentence, respectively. While this format suffices for many NLP tasks, it can be suboptimal for capturing aspect nuances in sentiment classification [37]. To address this, AEWEM adopts a modified input format, incorporating a sentence-aspect pair $T^{(S,A)}$. This format combines $T^{(S)}$ with the aspect subsequence t_a , enhancing aspect representation.

$$T^{(S,A)} = \{[\text{CLS}], T, [\text{SEP}], t_a, [\text{SEP}]\} \quad (2)$$

3.4.1. Relative Position Embeddings in AEWEM

In the Aspect-Enhanced Wordpiece Extraction Model (AEWEM), we advance the concept of position embeddings within the BERT framework. The standard BERT model employs token, segment, and position embeddings to capture the absolute position of wordpiece tokens, effectively encoding word order and eschewing a bag-of-words representation. In the context of TOWE, where the task is defined as tagging each wordpiece in relation to a given aspect, capturing the relative position to the aspect becomes crucial for accurately tracking opinion word positions.

For a wordpiece token $t_i \in T^S$, its relative position p_i to the aspect t_a is computed as $p_i = i - a$. We map p_i to a d_p -dimensional relative position embedding \mathbf{p}_i , distinct from BERT's absolute position embeddings. This embedding, initially randomly initialized, is refined during training. This distinction allows AEWEM to utilize patterns in opinion word proximity to aspects [38], which has been shown to significantly enhance TOWE performance [10]. By integrating BERT's wordpiece representation \mathbf{e}_i with \mathbf{p}_i , we obtain a comprehensive token representation $\mathbf{e}_i^p = [\mathbf{e}_i, \mathbf{p}_i]$. This concatenated representation \mathbf{E}^p for the entire sequence T is then processed through a BiLSTM for further refinement.

3.4.2. BiLSTM Encoding Layer in AEWEM

In AEWEM, we encode the sentence-aspect pair $T^{(S,A)}$ using BERT, ensuring that $T^{(S)} \in T^{(S,A)}$ captures essential aspect information. Our focus is on enriching the T^S representation by employing a BiLSTM [21] that processes a combination of vectors: (1) the BERT-derived semantic representation \mathbf{e}_i and (2) the relative position embedding \mathbf{p}_i . This combination allows AEWEM to effectively track the locations of opinion words in relation to the aspect's position, going beyond the mere encoding of word order as done by BERT's original position embeddings.

Formally, for each token $t_i \in T^{(S)}$, we calculate its relative position p_i to the aspect t_a as $p_i = i - a$, and map this to a d_p -dimensional embedding \mathbf{p}_i . The resultant BiLSTM input $\mathbf{e}_i^p = [\mathbf{e}_i, \mathbf{p}_i]$ for t_i forms a rich representation that is fed into the BiLSTM, producing the output $\mathbf{H} = \{\mathbf{h}_i\}_{i=1}^{|T^{(S)}|}$.

$$\mathbf{H} = \text{BiLSTM}(\mathbf{E}^p) \quad (3)$$

Here, each BiLSTM hidden state $\mathbf{h}_i = \{h_{ij}\}_{j=1}^{d_h}$ encapsulates both semantic and relative positional information, with d_h being the dimension of the hidden state.

3.5. Classification and Optimization in AEWEM

In AEWEM, sequence labeling is anchored on the representation of the first wordpiece of a word. For instance, for the word "surfboard", segmented into "surf" and "##board", the model focuses on the "surf" representation for classification, despite "surf" receiving contextual information from "##board" via the BiLSTM. This approach effectively simplifies the classification process without losing contextual richness.

Although Conditional Random Fields [39] are typically favored for sequence labeling, our preliminary experiments indicated that a softmax classifier offers comparable, if not superior, performance. Consequently, for each token t_i , the BiLSTM hidden state \mathbf{h}_i is used in a softmax classifier to generate a probability distribution \hat{y}_i over the label set \mathcal{Y} .

$$\hat{y}_i = \text{softmax}(\mathbf{h}_i \mathbf{W}_c + b_c) \quad (4)$$

Here, \mathbf{W}_c and b_c represent the classifier's weight and bias parameters, respectively. The training process involves minimizing the loss function:

$$\mathcal{L} = \sum_{(S,A)} \sum_{t_i \in T^{(S)}} \text{CrossEntropy}(\hat{y}_i, y_i) \quad (5)$$

In this function, y_i and \hat{y}_i denote the ground truth and predicted label of the wordpiece token $t_i \in T^{(S)}$, respectively.

4. Experiments

Our experimental framework involves an array of baseline models: ARGCN, BERT+BiLSTM+GCN, and ONG. We distinguish our model variants using suffixes (S) or (S,A), indicating the use of either a wordpiece sentence or a wordpiece sentence-aspect pair as input, respectively. We utilize publicly available code and optimal hyperparameters from the ARGCN¹ and BERT+BiLSTM+GCN.² The ONG model variants were implemented in-house, adhering to the authors' recommended hyperparameter configurations.³ Consistent with [1], our evaluation encompasses the Laptop (Lap14) and Restaurant (Res14, Res15, Res16) datasets [40–42]. We report the average Micro F1 scores from five runs using different random seeds for stability.

We compare AEWEM with the following existing models:

IOG [1].

IOG integrates an Inward-Outward LSTM with a BiLSTM for global context learning to pinpoint opinion words related to an opinion target.

LOTN [43].

LOTN, a multitask learning framework, facilitates the transfer of opinion knowledge from sentiment classification tasks to TOWE.

ONG [9].

ONG leverages an Ordered-Neuron LSTM with a Graph Convolution Network, infusing syntax data into BERT embeddings for TOWE.

ARGCN [20].

ARGCN employs an Attention-based Relational Graph Convolutional Network to encode syntactic information while focusing on key features for TOWE.

¹ https://github.com/samensah/encoders_towe_emnlp2021

² <https://github.com/wcwowwww/towe-eacl>

³ <https://github.com/samensah/Towe-TradeSyntax4WP>

SDRN+BERT [28]

SDRN+BERT introduces a Synchronous Double-channel Recurrent Network, extracting opinion targets and words, along with their interrelations, in tandem.

BERT-BiLSTM [10]

This model uses a BiLSTM over BERT and relative position embeddings to learn word-level features.

BERT-BiLSTM-GCN [10]

A slight improvement over BERT-BiLSTM, this model integrates syntax information into word-level features.

QD-OWSE [32]

QD-OWSE is a Question-Driven Opinion Word Span Extraction model that processes a BERT model with question and answer pairs to identify opinion words.

TMSMA [29]

TMSMA employs a Target-Specified sequence labeling with Multi-head Self-Attention, focusing on multi-head attention for opinion word identification.

Apart from IOG and LOTN, which use Glove embeddings [44], all methods are based on BERT embeddings.

4.1. Implementation Details of AEWEM

Our AEWEM model's results are the average of five independent runs to ensure consistency and reliability. For a fair comparison with other methods, we follow the experimental settings outlined in [10]. Hyperparameters were empirically determined, fine-tuned on a randomly selected 20% sample of the training set. The bert-base-uncased model served as our BERT encoder. We employed the AdamW optimizer, selecting from learning rates {1e-3, 1e-5} and a batch size of 16. Our BiLSTM encoder had a hidden size chosen from {100, 150}, with position embedding dimensions set to 100. The final hyperparameter configuration was selected based on optimal performance. Notably, our model's position embeddings differ from BERT's default, enhancing our approach's efficacy in TOWE tasks. We will make our code publicly available upon paper acceptance.

4.2. Comparative Analysis of Syntax-Aware Approaches

In our experiments, we scrutinized the performance of syntax-aware models like ONG, ARGCN, and BERT+BiLSTM+GCN, along with their modified versions. These modifications involved either removing the GCN component or incorporating wordpiece granularity in their input processing. This evaluation, detailed in Table 1, aimed to discern the impact of these alterations on their overall efficacy.

Table 1. Performance comparison of various syntax-aware models and their adaptations in terms of F1 score. The table presents results across four datasets with a highlighted average score. The ‘‘Granularity’’ column indicates whether the model processes input tokens at the word or wordpiece level.

Models	Feature	Lap14	Res14	Res15	Res16	Avg.
ONG	word	75.77	82.33	78.81	86.01	80.73
ONG w/o GCN	word	74.17	84.10	78.33	84.87	80.37
ONG(S) w/o GCN	wordpiece	79.79	86.63	80.72	88.30	83.86
ONG(S,A) w/o GCN	wordpiece	81.70	88.70	82.55	91.18	86.03
ARGCN	word	76.36	85.42	78.24	86.69	81.68
ARGCN w/o R-GCN	word	76.38	84.36	78.41	84.61	80.94
ARGCN(S) w/o R-GCN	wordpiece	80.08	85.92	81.36	89.72	84.27
ARGCN(S,A) w/o R-GCN	wordpiece	81.37	88.18	82.49	90.82	85.72
BERT+BiLSTM+GCN	word	78.82	85.74	80.54	87.35	83.11
BERT+BiLSTM	word	78.25	85.60	80.41	86.94	82.80
BERT+BiLSTM(S)	wordpiece	80.45	86.27	80.89	89.80	84.35
BERT+BiLSTM(S,A)	wordpiece	82.59	88.60	82.37	91.25	86.20

The removal of the GCN component yielded a negligible decrease in performance, suggesting that GCN’s role in these models might be less critical than previously assumed. Notably, when models were adapted to process inputs at the wordpiece level (indicated by the (S) suffix), we observed substantial improvements in their performance. This indicates that BERT’s inherent syntax capturing capabilities [45] may be sufficient, rendering the explicit syntax tree processing via GCNs less impactful. The models employing a sentence-aspect pair (marked with the (S,A) suffix) further improved performance, highlighting the importance of explicit aspect information during encoding.

Among these methods, the Aspect-Enhanced Wordpiece Extraction Model (AEWEM) configured as BERT+BiLSTM(S,A) exhibited the highest average F1 score of 86.2, underscoring the effectiveness of the sentence-aspect pair approach in TOWE tasks.

4.3. Investigating the Role of Aspect Representation

To further understand the influence of aspect representation in encoding, we evaluated the BERT+BiLSTM(S) model under various conditions, including aspect masking and the omission of position embeddings. The results, as shown in Table 2, highlight the impact of these factors on model performance.

Table 2. Impact of aspect masking and position embedding removal on BERT-BiLSTM(S) model’s F1 scores. The table reveals the model’s performance across different datasets with an aggregated average score.

Model	Lap14	Res14	Res15	Res16	Avg
BERT-BiLSTM(S)	80.45	86.27	80.89	89.80	84.35
-Mask Aspect	80.01	86.11	80.42	88.59	83.78
-Position Embedding	68.51	63.38	64.68	72.82	67.35

Masking the aspect resulted in a marginal decrease in performance across all datasets, suggesting that while aspect information contributes to model accuracy, its role is not dominant. However, the removal of position embeddings led to a significant drop in performance, emphasizing their critical

role in the model’s ability to contextualize and localize opinion words relative to the aspect. These findings support the enhancement of aspect representation in models like AEWEM to optimize TOWE task performance.

4.4. Qualitative Analysis with AEWEM

We conducted a qualitative analysis of the Aspect-Enhanced Wordpiece Extraction Model (AEWEM) and its variants on specific sentences, as summarized in Table 3. The analysis revealed that AEWEM and AEWEM(S) models sometimes missed distant opinion words from the aspect, notably in the first and second examples, where “beautiful” and “not the best” were overlooked. This gap was attributed to the models’ inability to fully contextualize co-referential terms like “it.” However, the AEWEM(S,A) variant demonstrated superior performance, successfully capturing these distant opinion words by enhancing aspect representation in the encoding process. In the third example, the advantage of using wordpieces was evident, particularly with the word “minimally,” which was not in the training set but was effectively identified due to its subword component.

Table 3. Comparative case study on opinion word extraction using different configurations of the AEWEM model. The table illustrates the model’s capability to identify opinion words relative to aspects in various sentences.

Instance	AEWEM	AEWEM(S)	AEWEM(S,A)
The OS is <i>fast</i> and <i>fluid</i> , everything is organized and it’s just <i>beautiful</i> .	<i>fast, fluid</i>	<i>fast, fluid</i>	<i>fast, fluid, beautiful</i>
Certainly <i>not the best sushi</i> in new york, however, it is always <i>fresh</i> , and the place is very clean, sterile.	<i>fresh</i>	<i>not the best</i>	<i>not the best, fresh</i>
Although somewhat loud, the noise was <i>minimally intrusive</i>	<i>loud, intrusive</i>	<i>loud, minimally intrusive</i>	<i>loud, minimally intrusive.</i>

4.5. Impact of BPE Subword Representations

Building upon the wordpiece analysis, we explored the utilization of Byte Pair Encoding (BPE) subword representations, inspired by data compression techniques. While BPE is distinct from BERT’s wordpiece approach, its application in RoBERTa models provides a suitable ground for comparison. We adapted the AEWEM model to use RoBERTa’s BPE subword representations, resulting in variants such as RoBERTa-AEWEM(S,A). The performance of these models, as shown in Table 4, underscores the effectiveness of BPE representations in TOWE tasks.

Table 4. Evaluation of AEWEM models using RoBERTa with BPE subword representations, showcasing their performance in F1 score across different datasets.

Model	Lap14	Res14	Res15	Res16	Avg
RoBERTa-AEWEM(S,A)	82.77	88.27	83.84	91.06	86.49
RoBERTa-AEWEM(S)	81.10	86.95	82.21	88.70	84.74
RoBERTa-AEWEM	75.87	81.38	75.94	84.70	79.47
RoBERTa-AEWEM+GCN	77.57	82.09	77.85	85.37	80.72

The results revealed a substantial improvement with BPE representations, with RoBERTa-AEWEM(S) outperforming its word-level counterpart. This reinforces the advantage of subword representations in handling rare and out-of-vocabulary words. Moreover, the RoBERTa-AEWEM(S,A) model, leveraging BPE-based sentence-aspect pairs, achieved the highest average F1 score, indicating that enhanced aspect representation further refines model performance.

4.6. Comparing with Existing Models

In our evaluation, we compared the Aspect-Enhanced Wordpiece Extraction Model (AEWEM) in its sentence-aspect pair configuration (AEWEM(S,A)) with contemporary state-of-the-art models. These models include IOG [1], LOTN [43], SDRN+BERT [28], BERT+BiLSTM+GCN [10], QD-OWSE [32], and TSMSA [29]. The results, displayed in Table 5, underscore the AEWEM(S,A) model's exceptional performance.

Table 5. Comparative analysis of TOWE models' performance. The table includes recent state-of-the-art methods, highlighting the superior performance of AEWEM(S,A).

Model	Lap14	Res14	Res15	Res16	Avg
IOG	71.35	80.02	73.25	81.69	76.58
LOTN	72.02	82.21	73.29	83.62	77.79
SDRN+BERT*	73.69	83.10	76.38	85.40	79.64
ONG	75.77	82.33	78.81	86.01	80.73
ARGCN	76.36	85.42	78.24	86.69	81.68
BERT+BiLSTM+GCN	78.82	85.74	80.54	87.35	83.11
QD-OWSE	80.35	87.23	80.71	88.14	84.11
TSMSA	82.18	86.37	81.64	89.20	84.85
AEWEM(S,A)	82.59	88.60	82.37	91.25	86.20

The BERT+BiLSTM+GCN model, although integrating GCN for syntax awareness, did not demonstrate significant improvements over the BERT+BiLSTM model. This suggests that the incorporation of GCN does not necessarily enhance performance. In contrast, the AEWEM(S) model, which utilizes wordpieces, showed considerable improvement. Further enhancement was observed in the AEWEM(S,A) model, where aspect representation is emphasized.

Among the competing methods, QD-OWSE and TSMSA, both employing BERT as their base, demonstrated competitive performance. QD-OWSE inputs a generated question-answer pair to BERT, while TSMSA utilizes multi-head attention for opinion word identification. These models confirm BERT's ability to capture sufficient syntax information for TOWE tasks, even without explicit syntax trees. However, our AEWEM(S,A) model, with its simplistic yet effective architecture, achieved the highest F1 scores across all datasets, setting a new benchmark in the TOWE domain.

5. Conclusion and Future Directions

In this study, we have successfully showcased that the Aspect-Enhanced Wordpiece Extraction Model (AEWEM) sets new benchmarks in the field of Target-Oriented Opinion Word Extraction (TOWE) by innovatively substituting Graph Convolutional Networks (GCNs) with BERT wordpieces. This transformation, coupled with a strategic emphasis on aspect representation, has propelled AEWEM to achieve state-of-the-art (SOTA) results. The enhanced aspect representation within AEWEM acts as a pivotal prompt, guiding the model's focus and significantly boosting its performance. The findings of our research underline a crucial trade-off in TOWE methodologies: the balance between harnessing syntax information through traditional means and leveraging the advanced capabilities of BERT wordpieces. The substantial improvement in performance, brought about by this shift in approach, underscores the effectiveness of BERT wordpieces in capturing intricate language nuances, especially when combined with an improved aspect representation strategy. Looking forward, we aim to delve deeper into the realm of prompt-based learning [46]. This exploration seeks to refine our understanding of how different types of prompts can further enhance the aspect representation in TOWE tasks. By integrating prompt-based learning strategies, we anticipate uncovering new pathways to optimize aspect enhancement, thus elevating the efficacy of the AEWEM model even further. This

future research direction is poised to contribute significantly to the advancement of natural language processing techniques, particularly in the nuanced domain of opinion word extraction.

References

1. Zhifang Fan, Zhen Wu, Xinyu Dai, Shujian Huang, and Jiajun Chen. Target-oriented opinion words extraction with target-fused neural sequence labeling. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2509–2518, 2019.
2. Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. Semeval-2014 task 4: Aspect based sentiment analysis. In Preslav Nakov and Torsten Zesch, editors, *Proceedings of the 8th International Workshop on Semantic Evaluation, SemEval@COLING 2014, Dublin, Ireland, August 23-24, 2014*, pages 27–35. The Association for Computer Linguistics, 2014. doi: 10.3115/v1/s14-2004. URL <https://doi.org/10.3115/v1/s14-2004>.
3. Hyun Duk Kim, Kavita Ganesan, Parikshit Sondhi, and ChengXiang Zhai. Comprehensive review of opinion summarization. 2011.
4. Hao Fei, Yafeng Ren, and Donghong Ji. Retrofitting structure-aware transformer language model for end tasks. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 2151–2161, 2020.
5. Duyu Tang, Bing Qin, and Ting Liu. Aspect level sentiment classification with deep memory network. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 214–224, 2016.
6. Jingye Li, Kang Xu, Fei Li, Hao Fei, Yafeng Ren, and Donghong Ji. MRN: A locally and globally mention-based reasoning network for document-level relation extraction. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1359–1370, 2021.
7. Hao Fei, Shengqiong Wu, Yafeng Ren, and Meishan Zhang. Matching structure for dual learning. In *Proceedings of the International Conference on Machine Learning, ICML*, pages 6373–6391, 2022.
8. Kai Sun, Richong Zhang, Mensah Samuel, Aletras Nikolaos, Yongyi Mao, and Xudong Liu. Self-training through classifier disagreement for cross-domain opinion target extraction. In *Proceedings of the ACM Web Conference 2023*, pages 1594–1603, 2023.
9. Amir Pouran Ben Veyseh, Nasim Nouri, Franck Dernoncourt, Dejing Dou, and Thien Huu Nguyen. Introducing syntactic structures into target opinion word extraction with deep learning. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 8947–8956. Association for Computational Linguistics, 2020. doi: 10.18653/v1/2020.emnlp-main.719. URL <https://doi.org/10.18653/v1/2020.emnlp-main.719>.
10. Samuel Mensah, Kai Sun, and Nikolaos Aletras. An empirical study on leveraging position embeddings for target-oriented opinion words extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9174–9179, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. URL <https://aclanthology.org/2021.emnlp-main.722>.
11. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina N. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, 2018.
12. Hao Fei, Yafeng Ren, and Donghong Ji. Boundaries and edges rethinking: An end-to-end neural model for overlapping entity relation extraction. *Information Processing & Management*, 57(6):102311, 2020.
13. Jingye Li, Hao Fei, Jiang Liu, Shengqiong Wu, Meishan Zhang, Chong Teng, Donghong Ji, and Fei Li. Unified named entity recognition as word-word relation classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 10965–10973, 2022.
14. Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. URL <https://openreview.net/forum?id=SJU4ayYgl>.
15. Yikang Shen, Shawn Tan, Alessandro Sordoni, and Aaron C. Courville. Ordered neurons: Integrating tree structures into recurrent neural networks. In *International Conference on Learning Representations*, 2018.

16. Shengqiong Wu, Hao Fei, Fei Li, Meishan Zhang, Yijiang Liu, Chong Teng, and Donghong Ji. Mastering the explicit opinion-role interaction: Syntax-aided neural transition system for unified opinion role labeling. In *Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence*, pages 11513–11521, 2022.
17. Wenxuan Shi, Fei Li, Jingye Li, Hao Fei, and Donghong Ji. Effective token graph modeling using a novel labeling strategy for structured sentiment analysis. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4232–4241, 2022.
18. Hao Fei, Yue Zhang, Yafeng Ren, and Donghong Ji. Latent emotion memory for multi-label emotion classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 7692–7699, 2020.
19. Fengqi Wang, Fei Li, Hao Fei, Jingye Li, Shengqiong Wu, Fangfang Su, Wenxuan Shi, Donghong Ji, and Bo Cai. Entity-centered cross-document relation extraction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9871–9881, 2022.
20. Junfeng Jiang, An Wang, and Akiko Aizawa. Attention-based relational graph convolutional network for target-oriented opinion words extraction. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1986–1997, 2021.
21. Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.
22. Hao Fei, Shengqiong Wu, Yafeng Ren, Fei Li, and Donghong Ji. Better combine them together! integrating syntactic constituency and dependency representations for semantic role labeling. In *Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021*, pages 549–559, 2021.
23. Shengqiong Wu, Hao Fei, Yafeng Ren, Donghong Ji, and Jingye Li. Learn from syntax: Improving pair-wise aspect and opinion terms extraction with rich syntactic knowledge. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pages 3957–3963, 2021.
24. Hao Fei, Fei Li, Bobo Li, and Donghong Ji. Encoder-decoder based unified semantic role labeling with label-aware syntax. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 12794–12802, 2021.
25. Hao Fei, Shengqiong Wu, Jingye Li, Bobo Li, Fei Li, Libo Qin, Meishan Zhang, Min Zhang, and Tat-Seng Chua. Lasuie: Unifying information extraction with latent adaptive structure-aware generative language model. In *Proceedings of the Advances in Neural Information Processing Systems, NeurIPS 2022*, pages 15460–15475, 2022.
26. Hao Fei, Yafeng Ren, Yue Zhang, Donghong Ji, and Xiaohui Liang. Enriching contextualized language model from knowledge graph for biomedical information extraction. *Briefings in Bioinformatics*, 22(3), 2021.
27. Zhen Wu, Fei Zhao, Xin-Yu Dai, Shujian Huang, and Jiajun Chen. Latent opinions transfer network for target-oriented opinion words extraction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9298–9305, 2020.
28. Shaowei Chen, Jie Liu, Yu Wang, Wenzheng Zhang, and Ziming Chi. Synchronous double-channel recurrent network for aspect-opinion pair extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6515–6524, 2020.
29. Yuhao Feng, Yanghui Rao, Yuyao Tang, Ninghua Wang, and He Liu. Target-specified sequence labeling with multi-head self-attention for target-oriented opinion words extraction. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1805–1815, 2021.
30. Hao Fei, Qian Liu, Meishan Zhang, Min Zhang, and Tat-Seng Chua. Scene graph as pivoting: Inference-time image-free unsupervised multimodal machine translation with visual scene hallucination. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5980–5994, 2023.
31. Hao Fei, Meishan Zhang, and Donghong Ji. Cross-lingual semantic role labeling with high-quality translated training corpus. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7014–7026, 2020.
32. Lei Gao, Yulong Wang, Tongcun Liu, Jingyu Wang, Lei Zhang, and Jianxin Liao. Question-driven span labeling model for aspect-opinion pair extraction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 12875–12883, 2021.
33. Shengqiong Wu, Hao Fei, Leigang Qu, Wei Ji, and Tat-Seng Chua. Next-gpt: Any-to-any multimodal llm, 2023.

34. Lance A. Ramshaw and Mitch Marcus. Text chunking using transformation-based learning. In David Yarowsky and Kenneth Church, editors, *Third Workshop on Very Large Corpora, VLC@ACL 1995, Cambridge, Massachusetts, USA, June 30, 1995*, 1995. URL <https://aclanthology.org/W95-0107/>.
35. Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao. Relation classification via convolutional deep neural network. In *Proceedings of COLING 2014, the 25th international conference on computational linguistics: technical papers*, pages 2335–2344, 2014.
36. Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In Aldo Gangemi, Roberto Navigli, Maria-Esther Vidal, Pascal Hitzler, Raphaël Troncy, Laura Hollink, Anna Tordai, and Mehwish Alam, editors, *The Semantic Web*, pages 593–607, Cham, 2018. Springer International Publishing. ISBN 978-3-319-93417-4.
37. Yuanhe Tian, Guimin Chen, and Yan Song. Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2910–2922, 2021.
38. Jie Zhou, Jimmy Xiangji Huang, Qinmin Vivian Hu, and Liang He. Is position important? deep multi-task learning for aspect-based sentiment analysis. *Applied Intelligence*, 50:3367–3378, 2020.
39. John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.
40. Maria Pontiki, Dimitrios Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, page 27–35, 2014.
41. Maria Pontiki, Dimitrios Galanis, Harris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, pages 486–495, 2015.
42. Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, et al. Semeval-2016 task 5: Aspect based sentiment analysis. In *International workshop on semantic evaluation*, pages 19–30, 2016.
43. Zhen Wu, Fei Zhao, Xin-Yu Dai, Shujian Huang, and Jiajun Chen. Latent opinions transfer network for target-oriented opinion words extraction. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 9298–9305. AAAI Press, 2020. URL <https://aaai.org/ojs/index.php/AAAI/article/view/6469>.
44. Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.
45. Kevin Clark, Urvashi Khandelwal, Omer Levy, and Christopher D. Manning. What does BERT look at? an analysis of BERT’s attention. In *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 276–286, Florence, Italy, August 2019. Association for Computational Linguistics. doi: 10.18653/v1/W19-4828. URL <https://aclanthology.org/W19-4828>.
46. Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.