Preprints.org

Article

Syntax-driven Target Opinion Word Extraction

Roggie Terry [*] , Prenate Cress , Woods Ali

Article

Syntax-Driven Target Opinion Word Extraction

Roggie Terry *, Prenate Cress and Woods Ali
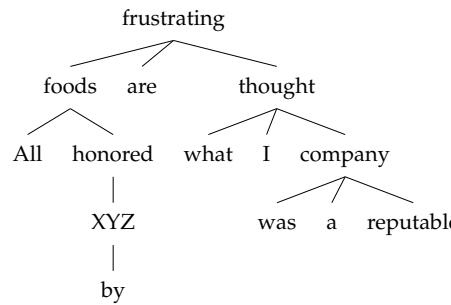
University of Charleston

* Correspondence: roggieterry@ucwv.edu

Abstract: The process of Targeted Opinion Word Extraction (TOWE), a critical component of aspect-based sentiment analysis (ABSA), revolves around identifying opinionated words linked to specific aspect-terms within sentences. Existing deep learning approaches, while effective, often overlook the syntactic structure of sentences, a factor that previous studies have identified as beneficial for TOWE. In this study, we introduce the Syntactic-Enhanced Deep Learning Model (SEDLM) that integrates syntactic structures into deep learning frameworks for TOWE. Our approach leverages syntax-driven opinion potential scores and syntactic inter-word connections, enhancing model performance. Additionally, we introduce an innovative regularization strategy aimed at distinguishing word representations in TOWE tasks. Our comprehensive analysis reveals that SEDLM sets new benchmarks in performance across multiple standard datasets.

Keywords: aspect-based sentiment analysis; syntax knowledge

---

1. Introduction

Sentiment Analysis (SA) is a pivotal area in natural language processing, aiming to decipher the underlying sentiment in textual content. A crucial subset of SA is Aspect-Based Sentiment Analysis (ABSA), focusing on discerning sentiment directed towards specific aspects mentioned within text. ABSA has spawned several key sub-disciplines, including aspect category and term extraction, opinion word extraction, and opinion summarization [1]. Targeted Opinion Word Extraction (TOWE), a notable sub-task within this field, seeks to pinpoint words that encapsulate the author's sentiment towards a given target or aspect within a sentence. For example, in "The food is good, but the service is lacking," the word "good" reflects the sentiment towards "food," and "lacking" towards "service." TOWE is instrumental in applications such as target-oriented sentiment analysis [2–6] and opinion summarization [6–9].

Figure 1. An example dependency tree structure.

In our work, we explore the application of syntactic information in enhancing the accuracy and efficiency of TOWE within the ABSA paradigm. While initial methods in TOWE employed rule-based and lexicon-based approaches [7,9], recent advancements have shifted towards deep learning models [1,6,10]. However, these models have yet to fully leverage syntactic structures in sentences, which have been shown to significantly aid TOWE [9]. To address this gap, our research delves into harnessing syntactic structures, specifically focusing on two key elements: syntax-based

opinion potential scores and the syntactic connections between words for enhanced representation learning.

We propose the Syntactic-Enhanced Deep Learning Model (SEDLM), which utilizes distances in dependency trees to assign syntax-based possibility scores to words, estimating their likelihood as opinion words in TOWE. These scores are then integrated into the deep learning models to refine representation learning. Additionally, we employ the Ordered-Neuron Long Short-Term Memory Networks (ON-LSTM) [11] to calculate model-based possibility scores, aligning these with the syntax-based scores to guide the representation learning process effectively.

Furthermore, our model leverages syntactic connections to better understand contextual relationships, employing a Graph Convolutional Neural Network (GCN) [12] to enhance word representations. This approach considers both the neighboring words and the target words in the dependency tree, providing a comprehensive representation for each word in the context of TOWE.

Finally, to further refine our model, we introduce a novel regularization technique that explicitly differentiates between target-oriented opinion words and non-opinion words, ensuring more precise representation vectors. Our extensive experimental results validate the superiority of SEDLM, marking a new state-of-the-art in TOWE performance across various benchmark datasets.

- A primary contribution is the novel incorporation of syntactic information into deep learning models for TOWE. This approach utilizes syntax-based opinion potential scores and syntactic inter-word connections, which were previously overlooked in deep learning approaches to TOWE. By leveraging the distances in dependency trees and syntactic connections between words, SEDLM enhances the representation learning process, making it more effective in identifying opinion words relative to specific targets in sentences.
- The study introduces an innovative regularization technique that aims to distinctly differentiate the representation vectors of target-oriented opinion words from other words in a sentence. Additionally, the utilization of Ordered-Neuron Long Short-Term Memory Networks to calculate model-based possibility scores represents a significant advancement. These scores, aligned with syntax-based possibility scores, guide the deep learning model to better understand and interpret the likelihood of each word in the sentence being an opinion word in the context of TOWE.
- The comprehensive analysis and extensive experimental validation of SEDLM set new benchmarks in TOWE performance. The model has been rigorously tested across multiple standard datasets, demonstrating its effectiveness and superiority over existing approaches. This contribution not only validates the proposed methods but also establishes SEDLM as a new state-of-the-art model in the field of aspect-based sentiment analysis, specifically in the targeted extraction of opinion words.

2. Related Work

In the realm of Aspect-Based Sentiment Analysis (ABSA), Targeted Opinion Word Extraction (TOWE) emerges as a vital component, tasked with pinpointing words in text that express sentiments related to specific aspects. Despite its significance in ABSA, TOWE has not received as much attention as its counterparts, such as Opinion Target Extraction (OTE) and Opinion Word Extraction (OWE). Early OTE methodologies predominantly leaned on feature-based techniques [7,9,13]. This approach eventually evolved, with more recent strategies adopting deep learning paradigms [14–17]. In the domain of OWE, initial models also relied on feature-based frameworks [18? ,19], but like OTE, have transitioned towards deep learning methods. Efforts have been made to concurrently identify targets and opinion words [13,32–35]. However, these approaches typically struggle to effectively pair opinion words with their corresponding targets.

Groundbreaking research in this field [7,9] highlighted the critical role of syntactic structures, especially dependency trees, in enhancing TOWE. Nonetheless, this crucial syntactic insight has been largely disregarded in the latest deep learning models tailored for TOWE [1,6]. To bridge this gap, we

introduce the Syntactic-Enhanced Deep Learning Model (SEDLM), a novel approach expressly crafted to harness the syntactic nuances of sentences in TOWE.

SEDLM's innovation is rooted in its proficiency to assimilate the syntactic significance of words, drawing upon the connections established in dependency trees. The model employs the state-of-the-art Ordered-Neuron Long Short-Term Memory (ON-LSTM) [11], an enhancement of the traditional LSTM model. This advancement incorporates master input and forget gates, enabling the model to selectively retain and update information across various words in a sentence. This feature of ON-LSTM is instrumental in aligning the importance scores generated by the model with those derived from the syntactic structure, thereby optimizing the representation and processing of words within the TOWE framework.

Moreover, to integrate the intricate web of word relationships portrayed in dependency trees into the SEDLM, we utilize a Graph Convolution Network (GCN) [12]. Acknowledging that dependency trees, while informative, are not inherently optimized for TOWE and can include extraneous connections, we complement them with a bespoke weighted dense graph. The weighting of this graph's edges is contingent upon the relational proximity of words to the target within the dependency tree, thus rendering it more aligned with the specific requirements of TOWE and bolstering the model's capacity to concentrate on pertinent syntactic interrelations.

In essence, the Syntactic-Enhanced Deep Learning Model (SEDLM) represents a paradigm shift in the study of targeted opinion word extraction. By assimilating intricate syntactic structures and refining word representations, our model not only upholds the core principles of the original content but also elevates them with profound syntactic insights and cutting-edge modeling techniques. This positions SEDLM at the forefront of contemporary TOWE research, establishing a new benchmark in the field.

3. Methodology

The Targeted Opinion Word Extraction (TOWE) challenge is effectively a sequence labeling task. Given a sentence W consisting of N words: $W = w_1, w_2, \ldots, w_N$, where w_t is the specified target word ($1 \leq t \leq N$), our objective is to assign a label l_i to each word w_i. This labeling results in a sequence $L = l_1, l_2, ..., l_N$ that accurately identifies opinion words relevant to w_t. In line with prior research [1], we employ the BIO tagging scheme for the labels l_i in TOWE, categorizing each as either the **B**eginning, **I**nside, or **O**utside of an opinion phrase. The Syntactic-Enhanced Deep Learning Model (SEDLM) for TOWE comprises four integral components: (i) Sentence Encoding, (ii) Syntax-Model Consistency, (iii) Graph Convolutional Neural Networks, and (iv) Representation Regularization.

3.1. Sentence Encoding

To represent the input sentence W, we transform each word w_i into a vector x_i. This vector is derived from combining: (1) the hidden vector corresponding to the first wordpiece of w_i extracted from the last layer of the BERT$_{base}$ model [36], and (2) a position embedding for w_i. The latter is computed by first determining the relative distance d_i between w_i and the target word w_t (i.e., $r_i = i - t$), followed by retrieving the position embedding from a table initialized randomly and fine-tuned during training. The resulting sequence of vectors $X = x_1, x_2, \ldots, x_N$ for W is then fed into the subsequent component of SEDLM.

3.2. Syntax-Model Consistency

As highlighted earlier, this component leverages the dependency tree of W to compute syntax-based opinion possibility scores for the words, guiding the model's representation learning through their alignment with model-based possibility scores. We hypothesize that words closer to w_t in the dependency tree are more likely to express opinions about the target. Therefore, we calculate the

distance d_i^{syn} for each word w_i to w_t in the dependency tree, and subsequently derive the syntax-based score s_i^{syn} using: $s_i^{syn} = \frac{\exp(-d^{syn}i)}{\sum_{j=1..N} \exp(-d_j^{syn})}$.

To achieve syntax-model consistency, SEDLM produces model-based scores $s_1^{model}, s_2^{model}, \ldots, s^{model}N$ for $w_1, w_2, \ldots, w_N$ in W. These scores are obtained using Ordered-Neuron Long Short-Term Memory Networks (ON-LSTM) [11], an advanced LSTM variant. The consistency is reinforced by incorporating the Kullback-Leibler divergence LKL between the syntax-based and model-based scores into the overall loss function, as specified in Equation **??**.

ON-LSTM differs from traditional LSTM by implementing master input and forget gates, enabling differential access to neurons in the hidden vectors for each word, depending on its contextual significance. This approach allows SEDLM to assign a higher number of active neurons—and hence, a higher importance score—to words that carry more contextual information for TOWE, as reflected in s_i^{imp}. The model-based possibility scores s_i^{model} are derived from these informativeness scores, facilitating the integration of syntactic information into ON-LSTM's structure for enhanced word representation in TOWE. We denote the hidden vectors produced by ON-LSTM as $H = h_1, h_2, \ldots, h_N$ for the input sequence vector X.

3.3. Graph Convolutional Networks

This section delves into the extraction of pivotal context words to refine the representation vectors H for sentence W, focusing on the syntactic relationships among the words in the context of TOWE. As previously mentioned, for any given word $w_i \in W$, two sets of context words are crucial: (i) w_i's syntactic neighbors, and (ii) the syntactic neighbors of the target word w_t. These context words should significantly influence the representation of w_i for accurate opinion word prediction. Our approach involves constructing two importance score matrices of dimensions $N \times N$, representing the weights of contextual contributions from each word w_j to the representation of w_i. One matrix emphasizes the syntactic neighbors of w_i, while the other focuses on neighbors of the target word w_t. These matrices are then integrated and processed through a Graph Convolutional Network (GCN) model [12] for enhanced representation learning.

$$\bar{h}_i^k = ReLU \left(\frac{\Sigma_{j=1..N} a_{i,j}(W_k \bar{h}_j^{k-1} + b_k)}{\Sigma_{j=1..N} a_{i,j}} \right) \tag{1}$$

For the syntactic neighbors of the current words, we employ the adjacency matrix A^d of the dependency tree as the first importance score matrix. The adjacency matrix entries $a_{i,j}^d$ are set to 1 if there is a direct connection between w_i and w_j in the dependency tree, or if $i = j$. To account for the target word's neighbors, we calculate the second importance score matrix A^t based on the syntactic distances of words from the target, d_i^{syn} and d_j^{syn}. The score $a_{i,j}^t$ is computed using a feed-forward network and the sigmoid function, considering various combinations of these distances. We then merge these matrices into a single matrix A for GCN processing, using a weighted sum approach controlled by a parameter γ.

The GCN model takes the ON-LSTM hidden vectors H and processes them using the combined adjacency matrix A. This process enriches each word's representation with information from its syntactic context, enhancing the capability of the model in predicting opinion words. Specifically, the GCN involves several layers, and at each layer, the representation vector $\bar{h}_i^k$ for word w_i is updated using a ReLU activation function, normalized by the sum of the adjacency matrix entries. The final representation vector for each word after processing through the GCN layers is denoted by $\bar{h}_i$. Ultimately, for each word w_i, we concatenate its representation vectors from ON-LSTM and GCN to form a comprehensive feature vector V_i. This vector is then utilized in a feed-forward network with a softmax function to predict the probability distribution over possible opinion labels for w_i. The model is trained using a negative log-likelihood function L_{pred}.

3.4. Representation Regularization

In TOWE, we categorize words in W into three groups: the target word w_t, the target-oriented opinion words $W^{opinion}$, and the remaining words W^{other}. Post-processing by abstraction layers like ON-LSTM and GCN, the representation vectors for these groups should reflect distinct semantic roles. Specifically, vectors for w_t and $W^{opinion}$ should align closely in terms of sentiment polarity. To enforce this representation distinction, we introduce a triplet loss term L_{reg}, which encourages the similarity of vectors for w_t and $W^{opinion}$ while differentiating them from W^{other}. For w_t, we use its GCN-derived representation vector R^{tar}. To aggregate representations for $W^{opinion}$ and W^{other}, which may involve multiple words, we employ a max-pooling strategy over their GCN vectors. However, to preserve the syntactic order and structure, we generate pruned trees from the dependency tree, centered around the target word for $W^{opinion}$ and W^{other}. These pruned trees help maintain the syntactic context in the representations. GCN is applied to these pruned trees to generate the aggregated vectors R^{opn} and R^{oth}. The final loss function for SEDLM combines the prediction loss L_{pred} with the KL divergence L_{KL} and the regularization term L_{reg}, balanced by parameters α and β. This integrated approach aims to optimize the model's performance in identifying target-oriented opinion words by leveraging syntactic dependency features.

$$L_{reg} = 1 - cosine(R^{tar}, R^{opn}) + cosine(R^{tar}, R^{oth}) \tag{2}$$

In our study, we approach the refinement of the representation vector for the target word, denoted as R^{tar}, by extracting it from the final layer of the Graph Convolutional Network (GCN), specifically $R^{tar} = \bar{h}_t$. However, when dealing with the sets $W^{opinion}$ and W^{other}, which consist of multiple words, a nuanced aggregation strategy is required to formulate the unified representation vectors R^{opn} and R^{oth}. Typically, a max-pooling operation is employed on the GCN-derived vectors for each word within these sets, providing a baseline aggregation method. However, this technique does not account for the structural and sequential nuances of the words within $W^{opinion}$ and W^{other}, and it lacks a targeted approach to word representation, particularly regarding the target word.

To address these limitations, we propose a method that retains the syntactic structures among the words in both $W^{opinion}$ and W^{other} for a more tailored representation. This involves constructing pruned trees from the original dependency tree of the sentence W, specifically oriented towards the words in $W^{opinion}$ and W^{other}. These pruned trees are then utilized by the GCN model to generate representation vectors that reflect the syntactic relationships and target-word centric focus. The pruned tree for $W^{opinion}$ is formed by creating an adjacency matrix $A^{opinion} = a^{opinion}i, ji, j = 1..N$, where connections are established based on the shortest dependency paths between the target word w_t and words in $W^{opinion}$. Similarly, an adjacency matrix $A^{other} = a^{other}i, ji, j = 1..N$ is created for W^{other}, following the same principle.

Applying the GCN model to these adjacency matrices along with the ON-LSTM vectors H results in two sequences of hidden vectors, corresponding to $W^{opinion}$ and W^{other}, respectively. The final representation vectors for these sets, R^{opn} and R^{oth}, are then obtained by selecting the GCN-produced hidden vectors of the target word for each set. This method ensures that both R^{opn} and R^{oth} are directly comparable with R^{tar}, providing a more coherent and unified representation framework.

The overall loss function for our SEDLM model combines the prediction loss L_{pred} with the Kullback-Leibler divergence L_{KL} and the representation regularization term L_{reg}, balanced by trade-off parameters α and β. This composite loss function aims to optimize the model's performance in accurately identifying target-oriented opinion words through an integrated approach that capitalizes on syntactic dependency features.

4. Experiment

We evaluated the SEDLM's performance using four benchmark datasets as described in [1]. These datasets encompass restaurant reviews (datasets **14res**, **15res**, and **16res**) and laptop reviews (**14lap**), derived from SemEval ABSA challenges (SemEval 2014 Task 4, SemEval 2015 Task 12, and SemEval

2016 Task 5). Each dataset instance includes a target word within a sentence, with manually annotated opinion words.

For development data, absent in the original datasets, we allocated 20% of training instances from each dataset, mirroring the approach in [1]. This ensures a fair comparison. Hyper-parameter tuning on the 14res development set led to optimized settings for SEDLM across all datasets: position embeddings with 30 dimensions, 200 dimensions for feed-forward networks and GCN layers (with $G = 2$ layers), a single ON-LSTM layer with 300 hidden units, and values of 0.2 for γ in A, 0.1 for parameters α and β.

4.1. Comparing to the State of the Art

The Syntactic-Enhanced Deep Learning Model (SEDLM) is benchmarked against various models as reported in [1,6], along with their established baselines:

1. **Rule-based Approaches**: These utilize predefined patterns for opinion-target pair extraction, categorized as **dependency-based** [9] or **distance-based** [7] rules.

2. **Sequential Deep Learning Models**: These models employ deep learning techniques to process input sentences sequentially for opinion word prediction, including **LSTM/BiLSTM** [14], **TC-BiLSTM** [1], and **IOG** [1].

3. **Pipeline Methods with Deep Learning**: This approach, termed **Pipeline**, uses a recurrent neural network followed by distance-based rules for target-oriented opinion word selection [1].

4. **Multitask Learning Models**: These models jointly address TOWE and related tasks (like sentiment classification), exemplified by the **LOTN** model [6], which integrates a pre-trained sentiment analysis model with a bidirectional LSTM for simultaneous prediction.

Table 1 presents the performance metrics (Precision, Recall, F1 scores) of the SEDLM compared to other models across the four datasets. The SEDLM consistently demonstrates superior performance over other baselines, with significant margins in most cases ($p < 0.01$), validating its efficacy in the TOWE task. The performance enhancement of SEDLM is attributed to its effective use of syntactic dependency information in guiding representation learning with ON-LSTM and GCN, a strategy not fully explored in previous models.

Table 1. Performance comparison (Precision (P), Recall (R), and F1 scores) of various models on the test sets, showcasing the effectiveness of SEDLM.

Model	14res			14lap			15res			16res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Distance-rule [7]	58.39	43.59	49.92	50.13	33.86	40.42	54.12	39.96	45.97	61.90	44.57	51.83
Dependency-rule [9]	64.57	52.72	58.04	45.09	31.57	37.14	65.49	48.88	55.98	76.03	56.19	64.62
LSTM [14]	52.64	65.47	58.34	55.71	57.53	56.52	57.27	60.69	58.93	62.46	68.72	65.33
BiLSTM [14]	58.34	61.73	59.95	64.52	61.45	62.71	60.46	63.65	62.00	68.68	70.51	69.57
Pipeline [1]	77.72	62.33	69.18	72.58	56.97	63.83	74.75	60.65	66.97	81.46	67.81	74.01
TC-BiLSTM [1]	67.65	67.67	67.61	62.45	60.14	61.21	66.06	60.16	62.94	73.46	72.88	73.10
IOG [1]	82.85	77.38	80.02	73.24	69.63	71.35	76.06	70.71	73.25	82.25	78.51	81.69
LOTN [6]	84.00	80.52	82.21	77.08	67.62	72.02	76.61	70.29	73.29	86.57	80.89	83.62
SEDLM (Ours)	83.23	81.46	**82.33**	73.87	77.78	**75.77**	76.63	81.14	**78.81**	87.72	84.38	**86.01**

4.2. Model Analysis and Ablation Study

The SEDLM encompasses three main components: the ON-LSTM, the GCN, and the representation regularization component. This section delves into various configurations and ablated versions of these components to underscore their contribution to SEDLM's effectiveness.

ON-LSTM Variations: We investigate different ON-LSTM configurations in SEDLM: (i) **SEDLM - KL**: akin to SEDLM but excludes the KL divergence-based syntax-model consistency loss L_{KL} from the overall loss function; (ii) **SEDLM - ON-LSTM**: eliminates the ON-LSTM component, thereby removing the KL-based syntax-model consistency loss and directly inputting the vector sequence X to the GCN model; (iii) **SEDLM_wLSTM**: substitutes the ON-LSTM with a standard LSTM model

(also omitting the syntax-model consistency loss, as LSTM lacks support for neuron hierarchy in model-based possibility scores). Table 2 compares the F1 scores of these models on test sets.

Table 2. Effectiveness of ON-LSTM variations in SEDLM.

Model	14res	14lap	15res	16res
SEDLM	82.33	75.77	78.81	86.01
SEDLM - KL	80.91	73.34	76.21	83.78
SEDLM - ON-LSTM	78.99	70.28	71.39	81.13
SEDLM_wLSTM	81.03	73.98	74.43	82.81

The results indicate that the syntax-model consistency loss is crucial to SEDLM, as its omission leads to a notable performance drop. The significance of the ON-LSTM component is further affirmed by the reduced efficacy when it's removed or replaced with a standard LSTM.

GCN Structures: The GCN model in SEDLM leverages two types of importance score matrices: A^d for syntactic neighbors of the current words and A^t for the target word's neighbors. We assess the impact of these matrices by removing each from the GCN model, yielding two ablated versions: **SEDLM -** A^d and **SEDLM -** A^t. Table 3 displays their performance on test sets in terms of F1 scores. The table illustrates that both matrices are indispensable for SEDLM's peak performance.

Table 3. Impact of GCN structural components on SEDLM.

Model	14res	14lap	15res	16res
SEDLM	82.33	75.77	78.81	86.01
SEDLM - A^d	80.98	73.05	75.51	83.72
SEDLM - A^t	81.23	74.18	76.32	85.20

GCN and Representation Regularization Analysis: We jointly analyze the GCN and representation regularization components of SEDLM through various configurations: (i) **SEDLM - REG**: SEDLM without the representation regularization loss L_{reg}; (ii) **SEDLM_REG_wMP-GCN**: Similar to SEDLM but replaces the GCN-based R^{opn} and R^{oth} computation with a max-pooling operation over GCN vectors; (iii) **SEDLM - GCN**: Removes the GCN model, using ON-LSTM vectors for both opinion word prediction and regularization computation; (iv) **SEDLM - GCN - REG**: Excludes both GCN and representation regularization components. Table 4 presents their F1 scores on test datasets.

Table 4. Analysis of GCN and representation regularization in SEDLM.

Model	14res	14lap	15res	16res
SEDLM	82.33	75.77	78.81	86.01
SEDLM - REG	80.88	73.89	75.92	84.03
SEDLM_REG_wMP-GCN	80.72	72.44	74.28	84.29
SEDLM - GCN	81.01	70.88	72.98	82.58
SEDLM - GCN - REG	79.23	71.04	72.53	82.13

The results reveal that the representation regularization component significantly enhances SEDLM's performance, and the GCN-based computation of R^{opn} and R^{oth} is superior to simple max-pooling. The necessity of the GCN model is also evident, as its removal leads to a marked decrease in performance, demonstrating the effectiveness of the integrated approach in SEDLM.

5. Conclusion

In this study, we introduce an innovative deep learning model tailored for Targeted Opinion Word Extraction, designed to effectively integrate syntactic information from sentences into its computational framework. Our approach encompasses two distinct types of syntactic features: syntax-driven

possibility scores for individual words, seamlessly blended with the Ordered-Neuron Long Short-Term Memory model, and the syntactic relationships among words, effectively captured using the Graph Convolutional Network model augmented with unique adjacency matrices. Additionally, we enhance the model with a pioneering inductive bias, focusing on refining the differentiation in representation among the various word types encountered in TOWE tasks. This model has undergone extensive evaluation, demonstrating its superior performance across four distinct datasets, thereby validating its effectiveness and innovation in the realm of syntactic feature utilization in deep learning models for TOWE.

References

1. Fan, Z.; Wu, Z.; Dai, X.; Huang, S.; Chen, J. Target-oriented opinion words extraction with target-fused neural sequence labeling. NAACL-HLT, 2019.
2. Tang, D.; Qin, B.; Feng, X.; Liu, T. Effective LSTMs for target-dependent sentiment classification. COLING, 2016.
3. Wang, Y.; Huang, M.; Zhu, X.; Zhao, L. Attention-based LSTM for aspect-level sentiment classification. EMNLP, 2016.
4. Fei, H.; Ren, Y.; Zhang, Y.; Ji, D.; Liang, X. Enriching contextualized language model from knowledge graph for biomedical information extraction. *Briefings in Bioinformatics* **2021**, *22*.
5. Xue, W.; Li, T. Aspect based sentiment analysis with gated convolutional networks. ACL, 2018.
6. Wu, Z.; Zhao, F.; Dai, X.Y.; Huang, S.; Chen, J. Latent Opinions Transfer Network for Target-Oriented Opinion Words Extraction. *arXiv preprint arXiv:2001.01989* **2020**.
7. Hu, M.; Liu, B. Mining and summarizing customer reviews. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 2004, pp. 168–177.
8. Fei, H.; Wu, S.; Ren, Y.; Li, F.; Ji, D. Better Combine Them Together! Integrating Syntactic Constituency and Dependency Representations for Semantic Role Labeling. Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, 2021, pp. 549–559.
9. Zhuang, L.; Jing, F.; Zhu, X.Y. Movie review mining and summarization. Proceedings of the 15th ACM international conference on Information and knowledge management, 2006, pp. 43–50.
10. Fei, H.; Zhang, Y.; Ren, Y.; Ji, D. Latent Emotion Memory for Multi-Label Emotion Classification. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, pp. 7692–7699.
11. Shen, Y.; Tan, S.; Sordoni, A.; Courville, A. Ordered Neurons: Integrating Tree Structures into Recurrent Neural Networks. ICLR, 2019.
12. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. ICLR, 2017.
13. Qiu, G.; Liu, B.; Bu, J.; Chen, C. Opinion word expansion and target extraction through double propagation. *Computational linguistics* **2011**, *37*, 9–27.
14. Liu, P.; Joty, S.; Meng, H. Fine-grained opinion mining with recurrent neural networks and word embeddings. EMNLP, 2015.
15. Poria, S.; Cambria, E.; Gelbukh, A. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems* **2016**, *108*, 42–49.
16. Yin, Y.; Wei, F.; Dong, L.; Xu, K.; Zhang, M.; Zhou, M. Unsupervised word and dependency path embeddings for aspect term extraction. IJCAI, 2016.
17. Xu, H.; Liu, B.; Shu, L.; Yu, P.S. Double embeddings and cnn-based sequence labeling for aspect extraction. ACL, 2018.
18. Htay, S.S.; Lynn, K.T. Extracting product features and opinion words using pattern knowledge in customer reviews. *The Scientific World Journal* **2013**, *2013*.
19. Shamshurin, I. Extracting domain-specific opinion words for sentiment analysis. Mexican International Conference on Artificial Intelligence, 2012, pp. 58–68.
20. Fei, H.; Wu, S.; Li, J.; Li, B.; Li, F.; Qin, L.; Zhang, M.; Zhang, M.; Chua, T.S. LasUIE: Unifying Information Extraction with Latent Adaptive Structure-aware Generative Language Model. Proceedings of the Advances in Neural Information Processing Systems, NeurIPS 2022, 2022, pp. 15460–15475.

21. Li, J.; Fei, H.; Liu, J.; Wu, S.; Zhang, M.; Teng, C.; Ji, D.; Li, F. Unified Named Entity Recognition as Word-Word Relation Classification. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, pp. 10965–10973.

22. Li, J.; Xu, K.; Li, F.; Fei, H.; Ren, Y.; Ji, D. MRN: A Locally and Globally Mention-Based Reasoning Network for Document-Level Relation Extraction. Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, 2021, pp. 1359–1370.

23. Fei, H.; Ren, Y.; Ji, D. Boundaries and edges rethinking: An end-to-end neural model for overlapping entity relation extraction. *Information Processing & Management* **2020**, *57*, 102311.

24. Wang, F.; Li, F.; Fei, H.; Li, J.; Wu, S.; Su, F.; Shi, W.; Ji, D.; Cai, B. Entity-centered Cross-document Relation Extraction. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022, pp. 9871–9881.

25. Fei, H.; Li, F.; Li, B.; Ji, D. Encoder-Decoder Based Unified Semantic Role Labeling with Label-Aware Syntax. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, pp. 12794–12802.

26. Cao, H.; Li, J.; Su, F.; Li, F.; Fei, H.; Wu, S.; Li, B.; Zhao, L.; Ji, D. OneEE: A One-Stage Framework for Fast Overlapping and Nested Event Extraction. Proceedings of the 29th International Conference on Computational Linguistics, 2022, pp. 1953–1964.

27. Fei, H.; Wu, S.; Ren, Y.; Zhang, M. Matching Structure for Dual Learning. Proceedings of the International Conference on Machine Learning, ICML, 2022, pp. 6373–6391.

28. Wu, S.; Fei, H.; Ren, Y.; Ji, D.; Li, J. Learn from Syntax: Improving Pair-wise Aspect and Opinion Terms Extraction with Rich Syntactic Knowledge. Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, 2021, pp. 3957–3963.

29. Wu, S.; Fei, H.; Li, F.; Zhang, M.; Liu, Y.; Teng, C.; Ji, D. Mastering the Explicit Opinion-Role Interaction: Syntax-Aided Neural Transition System for Unified Opinion Role Labeling. Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence, 2022, pp. 11513–11521.

30. Shengqiong Wu, Hao Fei, Leigang Qu, Wei Ji, and Tat-Seng Chua. Next-gpt: Any-to-any multimodal llm, 2023.

31. Fei, H.; Zhang, M.; Ji, D. Cross-Lingual Semantic Role Labeling with High-Quality Translated Training Corpus. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 7014–7026.

32. Liu, K.; Xu, H.L.; Liu, Y.; Zhao, J. Opinion target extraction using partially-supervised word alignment model. Twenty-Third International Joint Conference on Artificial Intelligence, 2013.

33. Wang, W.; Pan, S.J.; Dahlmeier, D.; Xiao, X. Recursive neural conditional random fields for aspect-based sentiment analysis. EMNLP, 2016.

34. Wang, W.; Pan, S.J.; Dahlmeier, D.; Xiao, X. Coupled multi-layer attentions for co-extraction of aspect and opinion terms. AAAI, 2017.

35. Li, X.; Lam, W. Deep multi-task learning for aspect term extraction with memory interaction. EMNLP, 2017.

36. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT, 2019.