# GCNDA: Graph Convolutional Networks with Dual Attention Mechanisms for Aspect Based Sentiment Analysis

Junjie Chen[1,2], Hongxu Hou[1(✉)], Jing Gao[2], Yatu Ji[1], Tiangang Bai[1], and Yi Jing[3]

[1] College of Computer Science, Inner Mongolia University, Hohhot, China
chenjj@imau.edu.cn, cshhx@imu.edu.cn, jiyatu0@126.com
[2] College of Computer Science and Information Engineering, Inner Mongolia Agricultural University, Hohhot, China
gaojing@imau.edu.cn
[3] Faculty of Science, The University of Sydney, Camperdown, Australia
yjin5439@uni.sydney.edu.au

**Abstract.** As the amount of user-generated content on the web continues to increase, a great interest has been shown in aspect-level sentiment analysis, which provides more detailed information than general sentiment analysis. In recent years, neural-based models have achieved success in this task because of their powerful representation learning capabilities. However, they ignore that the sentiment polarity of the target is related to the entire text structure. In this paper, we present a method based on graph convolutional neural networks named GCNDA, in which the given text is considered as a graph and the target is the specific region of the graph. Dual graph-based attention models are used to concentrate on the relation between words and certain regions of the graph. We conduct comprehensive experiments on publicly accessible datasets, and results demonstrate that our model outperforms the state-of-the-art baselines.

**Keywords:** Aspect Based Sentiment Analysis · Graph Convolutional Networks · Attention mechanisms

## 1 Introduction

Sentiment analysis [19], also known as opinion mining [13,20], is a vital task in text mining. For example, consumers want to know sentiment of existing users about products, meanwhile business want to obtain public opinions for their decision making. Due to its great value in practical applications, it has attracted widespread attention from both the industry and academic communities.

Supported by Inner Mongolia Natural Science Foundation of China (2018MS06005, 2015MS0628) and Inner Mongolia Autonomous Region Key Laboratory of Big Data Research and Application of Agriculture and Animal Husbandry.


T. Gedeon et al. (Eds.): ICONIP 2019, CCIS 1142, pp. 189–197, 2019.
https://doi.org/10.1007/978-3-030-36808-1_21

Typically, users write both positive and negative aspects in the same review, although the general sentiment may be positive or negative. Given the review *"the food is so good and so popular that waiting can really be a nightmare"*. It expresses negative sentiment towards "waiting" while holding positive sentiment towards "food". Aspect Based Sentiment Analysis (ABSA) is a fine-grained task in the field of sentiment classification [13,22]. The goal of this subtask is to predict the sentiment polarity of aspects that appear in a given text.

The existing deep learning models mainly rely on Recurrent Neural Networks (RNNs) [12,16,23,26,30], Memory Networks [4,6,14,24] and Convolutional Neural Networks (CNNs) [28]. However, these architectures run on a grid or sequential structure without using the entire graphical text structure. Therefore, they are difficult to obtain a structural relationship between words [29].

Text graph has been widely used in NLP tasks [1,7,18,27], which can grasp the entire text structure. Moreover, the Graph Convolutional Networks (GCNs) [9] have demonstrated the powerful ability to obtain hidden representation of nodes in a graphical structure. Inspired by this, we develop a deep learning framework based on graph convolutional networks to obtain structural information. To emphasis the relation between context and aspect words, we design dual graph-based attention models for the ABSA task. One is Graph Attention Mechanism (GAM), which learns attention weights for words to different neighbor words in the context. The other is an Aspect-based Structural Attention Mechanism (ABSAM), which focus on the part of aspect terms in the text graph.

## 2 Graph Convolutional Networks

Graph convolutional networks [3] have been proposed for learning over graphs. The majority of these methods do not scale to large graphs or are designed for whole-graph classification. Kipf and Welling [9] proposed a localized first-order approximation of spectral graph convolutions, which is very effective for the semi-supervised nodes classification. Since the GCNs succeeded in nodes classification, they have been introduced into NLP tasks such as semantic role labeling [17], machine translation [2].

## 3 Our Model

We assume that sentiment polarity of aspect terms is not only related to target words, but also to the whole text. Since it is necessary to take fully account of the relation between words in a given text, our model is based on the graph where the text is regarded as a graph and the aspect terms are considered part of the graph.

Our framework (GCNDA) can be divided into two parts, one is a text graph representation and the other is an aspect-based structural attention. We employ multiple GCN layers with GAM to get the text hidden state, and the aspect-based structural attention model to obtain the specific regions representation.

Then, two hidden states are fed forward to the fully connected (FC) layer. Finally, the representation vectors are put into the softmax layer to get the class label. In the following, we will explain each part of the framework in detail.

### 3.1 Graph Construction

In our approach, operations are performed on the text graph, so the structure of the text is important. To illustrate the effectiveness of our model, we construct undirected text graph in two ways, one based on co-occurrence information and the other on syntactic dependencies, which are widely used in the literature.

For the given text, each vertex corresponds to a word. For the co-occurrence graph, if two nodes $v_i$ and $v_j$ have a co-occurrence relation, the edge $(v_i, v_j)$ is established. Where co-occurrence relation is defined as two nodes co-occur within the specific window size. The edge weight of $(v_i, v_j)$ is the number of co-occurrences. For the syntactic dependency graph, establish connections for the dependencies where two nodes belong to a specific part of speech set. The adjacency matrix $A$ is obtained by the undirected graph structure.

### 3.2 Text Graph Representation

After the graph construction, the given text is converted to a graph. The text graph representation is obtained by input nodes and adjacency matrix. Each node corresponds to the word $w_{si}$ and is represented by the vector $x_i$ of the dimension D after the embedding layer. Usually, the length N of words set in the corpus is larger than the length M of the set in the text. Since it is not necessary to build the adjacency matrix A$\in \mathbb{R}^{N\times N}$ for each given text, we covert nodes sequence to the current text words sequence, then the adjacency matrix A becomes $\mathbb{R}^{M\times M}$, and L represents the Laplacian matrix of A.

After embedding layer, the initial representation of nodes is defined as $H^{(0)} = X \in \mathbb{R}^{M\times D}$. The GAM produces the importance of node $N_j$ to node $N_i$, and the attention coefficient is computed by

$$e_{ij} = score(H_i^{(0)}, H_j^{(0)}) \tag{1}$$

Where $score(\cdot)$ is the attention function. As mentioned in [15], the score function can be divided into "dot", "general" and "concat". In our model the "general" is used, the attention score is computed by following formula.

$$e_{ij} = H_i^{(0)}(W_{att}H_j^{(0)})^T \tag{2}$$

Where $W_{att}$ is trainable parameters. The equation indicates the importance of the node to each node in the graph without any structural information. We perform masked attention similar to [25], injecting the graph structure into the mechanism as Eq. 6.

$$Att_{ij} = \begin{cases} \frac{\exp(e_{ij})}{\sum_{k\in N(i)}\exp(e_{ik})} & if\ j \in N(i) \\ 0 & others \end{cases} \tag{3}$$

Where $N(i)$ is the neighbor set of node i.

The output of the $l-th$ GCN hidden layer can be obtained by the following equation.

$$H^{(l+1)} = Relu(\beta \times LH^{(l)}W^{(l)} + (1-\beta) \times AttH^{(l)}W^{(l)}) \tag{4}$$

Where $\beta$ is a hyperparameter between 0 and 1, $W^{(l)} \in \mathbb{R}^{F \times F^{'}}$ is a weight matrix in $l-th$ GCN layer and, and $Att \in \mathbb{R}^{M \times M}$ is an attention matrix obtained from Eq. (6). $F$ and $F^{'}$ are the input and output feature sizes, respectively. Compared with Eq. (3), the GCN layers in our model incorporate the graph attention mechanism. Unlike the attention method in the graph attention network [25], the attention weights in our model are shared among all GCN layers which is called GAM. Equation (6) describes the GCN hidden representation consists of two parts: one is calculated by the attention matrix and the other is the adjacency matrix. We assume the attention matrix represents the knowledge of the relationship between the nodes acquired by training, while the adjacent matrix is the relationship between nodes in the current context, both of which are proportionally combined output vectors. The output of the GCNs is forwarded to the pooling operations layers, which is the element operation on hidden vectors to get representation of the text graph.

### 3.3 Aspect-Based Structural Attention

For the general graph classification based on sequential methods, one challenge is to give the order of the recurrent neural networks [10], while for text mining, word sequences in the original text can be used naturally.

We extend the structural attention model [11] for ABSA task. See articles [11] for more details on the structural attention model. The sequences are produced through the aspect sequence generator. First the generator produces a set of nodes, referred to herein as aspect structural nodes, which are in the largest connected subgraph containing the aspect words. Then two agents are defined, one from the left and the other from right traversing to aspect terms. At each step, agents move along the text sequence to the word that are in aspect structural nodes. Obviously, in our approach the rank vector is the distance between the current node and aspect structural nodes. The ultimate goal of the agent is to collect enough information about the aspect terms. Finally, the generator generates a left sequence and a right sequence that are fed forward to the left and right LSTM layers, respectively. The hidden state of the aspect structural attention is obtained by concatenating the last hidden vectors of the left and right LSTM.

## 4 Experiments and Results

### 4.1 Experimental Setting

We test our model on three public datasets, two of them come from SemEval 2014 [22], and the third is a collection of twitters [5]. SemEval 2014 includes user-generated reviews of laptop and restaurant domains, following previous work [24],

**Table 1.** Statistics of aspects in different datasets

| Datasets | Positive | | Negative | | Neutral | |
|---|---|---|---|---|---|---|
| | Train | Test | Train | Test | Train | Test |
| Restaurant | 2164 | 728 | 805 | 196 | 633 | 196 |
| Laptop | 987 | 341 | 866 | 128 | 460 | 169 |
| Twitter | 1561 | 173 | 1560 | 173 | 3126 | 346 |

we removed a few examples having the "conflict" label. The statistics of the datasets are shown in Table 1.

Our models are performed on co-occurrence graph and on syntactic dependency graph, denoted as GCNDAc and GCNDAs, respectively. The window size is set to 2 for co-occurrence graph construction, and Stanford parser[1] is used for syntactic dependency graph construction.

In experiments, four-layer GCNs with RELU activation function is developed. We use Adam [8] optimizer with the learning rate 0.01, dropout 0.2, and the maximum number of epoch 50. 300-dimensional word embeddings pre-trained by GloVe [21] are utilized, which are not tuned during training time.

### 4.2 Compared Methods

We compare our model with following baseline methods:

TD-LSTM [23] is a model based on LSTM network, in which two LSTM models are used to model the preceding and following contexts surrounding the target string for sentiment classification.

MemNet [24] is a neural attention model over an external memory, which consists of multiple computational layers.

RAM [4] is a framework that adopts multiple-attention mechanism on recurrent neural network.

IAN [16] is an interactive attention networks model. It uses two attention networks to model the target and context interactively.

Cabasc [14] is based on the memory model, which can solve the semantic mismatch problem through two attention mechanisms, namely sentence-level content attention mechanism and context attention mechanism.

GCAE [28] is based on convolutional neural networks and gating mechanisms.

We note that in the different literature, different results are reported for the same model performed on the same dataset. We think that the results of the baseline methods are affected by text preprocessing and word embeddings, as mentioned in [16]. To reveal the capability of models, same word vectors used in our models are applied to all baselines.

[1] https://nlp.stanford.edu/software/lex-parser.shtml.

### 4.3 Main Results

For all methods accuracy evaluation is used as metric, and results are shown in Table 2. The best scores are highlighted in bold and the underlines indicate the second best performances. As the results show, our two models, GCNDAc and GCNDAs, consistently outperform all comparison methods on these three datasets. GCNDAs outperforms GCNDAc on Laptop and Restaurant datasets. This may be due to the fact that the syntactic dependency graph establishes a connection between two long distance words, shortening the distance between the aspect and the related words. However, the text in Twitter is irregular and short, the dependency parsing is not guaranteed to work well. Although the performance of GCNDA on the co-occurrence graph is not optimal, it is easier to construct a co-occurrence graph than to build a syntactic dependency graph, and its performance is superior to other baselines.

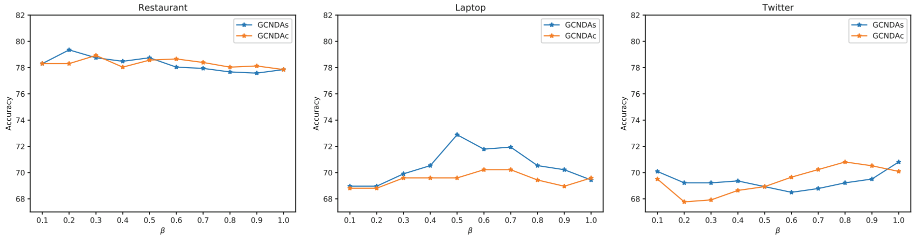
**Table 2.** Results of our model against baselines.

| Methods | Restaurant | Laptop | Twitter |
|---|---|---|---|
| TD-LSTM | 75.17 | 66.94 | 67.72 |
| MemNet | 76.88 | 68.18 | 69.63 |
| IAN | 76.96 | 67.86 | 68.63 |
| RAM | 76.87 | 67.24 | <u>69.88</u> |
| Cabasc | 77.05 | 68.65 | 67.33 |
| GCAE | 76.12 | 68.65 | 69.79 |
| GCNDAs | **79.35** | **72.88** | **70.81** |
| GCNDAc | <u>78.93</u> | <u>70.21</u> | **70.81** |

In LSTM-based models, TD-LSTM and IAN, IAN has better results than TD-LSTM because IAN uses context and target attention mechanisms, which make better use of important parts of a sentence for aspect words. MemNet is based on memory network, containing multiple attention layers, superior to LSTM-based models on Laptop and Twitter. RAM achieves the best performances on Twitter among baselines, which adopts not only the multi-hop attention mechanism but also deep bidirectional LSTM. Compared with RAM and MemNet, Cabasc enhances the ability to capture important information about a given aspect from a global perspective by sentence-level content attention mechanism and context attention mechanism, thus has a best performance in all baselines on Laptop and Restaurant. GCAE utilizes convolutional neural network with gating mechanisms, obtaining the best result as Cabasc on Laptop.

### 4.4 Effect of Hyperparameter $\beta$

The hyperparameter $\beta$ is used in GCN layers, which represents the ratio obtained from the adjacency matrix in GCN output vectors. The effect of $\beta$ on performance is shown in Fig. 1.




(a) (b) (c)

**Fig. 1.** Effect of $\beta$ on GCNDA

## 5 Conclusion

In this paper, we present a novel method based on graph convolutional networks and two attention mechanisms for the aspect-based sentiment analysis task. Compared with baselines on public datasets, the experimental results show that our model outperforms the state-of-the-art baselines.

We performed our model on co-occurrence graph structure and syntactic graph structure, and the results demonstrate that although co-occurrence graph is simple in construction, it can achieve better performance datasets.

## References


1. Angelova, R., Weikum, G.: Graph-based text classification: learn from your neighbors. In: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 485–492. ACM (2006)
2. Bastings, J., Titov, I., Aziz, W., Marcheggiani, D., Simaan, K.: Graph convolutional encoders for syntax-aware neural machine translation. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 1947–1957. Association for Computational Linguistics, Copenhagen, September 2017. https://www.aclweb.org/anthology/D17-1209
3. Bruna, J., Zaremba, W., Szlam, A., Lecun, Y.: Spectral networks and locally connected networks on graphs. In: International Conference on Learning Representations (ICLR 2014), CBLS, April 2014
4. Chen, P., Sun, Z., Bing, L., Yang, W.: Recurrent attention network on memory for aspect sentiment analysis. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 452–461 (2017)
5. Dong, L., Wei, F., Tan, C., Tang, D., Zhou, M., Xu, K.: Adaptive recursive neural network for target-dependent twitter sentiment classification. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), vol. 2, pp. 49–54 (2014)
6. Fan, C., Gao, Q., Du, J., Gui, L., Xu, R., Wong, K.F.: Convolution-based memory network for aspect-based sentiment analysis. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 1161–1164. ACM (2018)

7. Florescu, C., Caragea, C.: PositionRank: an unsupervised approach to keyphrase extraction from scholarly documents. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol. 1, pp. 1105–1115 (2017)
8. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: International Conference on Learning Representations (ICLR) (2015)
9. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
10. de Lara, N., Pineau, E.: A simple baseline algorithm for graph classification. arXiv preprint arXiv:1810.09155 (2018)
11. Lee, J.B., Rossi, R., Kong, X.: Graph classification using structural attention. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1666–1674. ACM (2018)
12. Li, L., Liu, Y., Zhou, A.: Hierarchical attention based position-aware network for aspect-level sentiment analysis. In: Proceedings of the 22nd Conference on Computational Natural Language Learning, pp. 181–189 (2018)
13. Liu, B.: Sentiment analysis and opinion mining. Synth. Lect. Hum. Lang. Technol. **5**(1), 1–167 (2012)
14. Liu, Q., Zhang, H., Zeng, Y., Huang, Z., Wu, Z.: Content attention model for aspect based sentiment analysis. In: Proceedings of the 2018 World Wide Web Conference on World Wide Web, pp. 1023–1032. International World Wide Web Conferences Steering Committee (2018)
15. Luong, M.T., Pham, H., Manning, C.D.: Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025 (2015)
16. Ma, D., Li, S., Zhang, X., Wang, H.: Interactive attention networks for aspect-level sentiment classification. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence, pp. 4068–4074. AAAI Press (2017)
17. Marcheggiani, D., Titov, I.: Encoding sentences with graph convolutional networks for semantic role labeling. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 1507–1516. Association for Computational Linguistics, Copenhagen, September 2017. https://www.aclweb.org/anthology/D17-1159
18. Mihalcea, R., Tarau, P.: TextRank: Bringing Order into Texts. Association for Computational Linguistics, Stroudsburg (2004)
19. Nasukawa, T., Yi, J.: Sentiment analysis: capturing favorability using natural language processing. In: Proceedings of the 2nd International Conference on Knowledge Capture, pp. 70–77. ACM (2003)
20. Pang, B., Lee, L., et al.: Opinion mining and sentiment analysis. Founda. Trends® Inf. Retr. **2**(1–2), 1–135 (2008)
21. Pennington, J., Socher, R., Manning, C.: Glove: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014)
22. Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., Manandhar, S.: SemEval-2014 task 4: aspect based sentiment analysis. In: Proceedings of International Workshop on Semantic Evaluation, pp. 27–35 (2014)
23. Tang, D., Qin, B., Feng, X., Liu, T.: Target-dependent sentiment classification with long short term memory. CoRR, abs/1512.01100 (2015)
24. Tang, D., Qin, B., Liu, T.: Aspect level sentiment classification with deep memory network. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 214–224 (2016)

25. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. In: International Conference on Learning Representations (2018). https://openreview.net/forum?id=rJXMpikCZ. Accepted as poster
26. Wang, Y., Huang, M., Zhao, L., et al.: Attention-based LSTM for aspect-level sentiment classification. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 606–615 (2016)
27. Wei, F., Li, W., Lu, Q., He, Y.: A document-sensitive graph model for multi-document summarization. Knowl. Inf. Syst. **22**(2), 245–259 (2010)
28. Xue, W., Li, T.: Aspect based sentiment analysis with gated convolutional networks. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, 15–20 July 2018, Volume 1: Long Papers, pp. 2514–2523 (2018). https://aclanthology.info/papers/P18-1234/p18-1234
29. Yang, Z., et al.: GLoMo: unsupervisedly learned relational graphs as transferable representations. arXiv preprint arXiv:1806.05662 (2018)
30. Zhang, M., Zhang, Y., Vo, D.T.: Gated neural networks for targeted sentiment analysis. In: AAAI, pp. 3087–3093 (2016)