

BiDAF MODEL IN SENTIMENT ANALYSIS TASK

Luong Thi Minh Hue

University of Information and Communication Technology,
Thai Nguyen University, Vietnam

ARTICLE INFORMATION

Journal: Vinh University
Journal of Sciences
ISSN: 1859-2228

Volume: 52

Issue: 2A

* **Correspondence:**
lmhue@ictu.edu.vn

Received: 10 February 2023

Accepted: 21 March 2023

Published: 25 May 2023

Citation:
Luong Thi Minh Hue (2023).
BiDAF model in sentiment
analysis task. *Vinh Uni. J. Sci.*
Vol. 52 (2A), pp. 18-24
doi: 10.56824/vujs.2023a021

OPEN ACCESS

Copyright © 2023. This is an
Open Access article distributed
under the terms of the [Creative
Commons Attribution License](#)
(CC BY NC), which permits
non-commercially to share
(copy and redistribute the
material in any medium) or
adapt (remix, transform, and
build upon the material),
provided the original work is
properly cited.

ABSTRACT

Sentiment analysis is a critical job in natural language processing. Controlling and evaluating customer feedback on their goods is a task that companies are especially interested in. For reading comprehension problems including attention processes, the BiDAF model is developed. Attention processes have recently been expanded and effectively used for natural language processing problems. In this study, we use the BiDAF model to perform sentiment analysis on Amazon product evaluations at the sentence level. The BiDAF model is a multi-layered processing model that reflects context at multiple levels and uses the BiLSTM model. Furthermore, we investigate the sentence's attention weight distribution using the attention mechanism. With a recall measure, the model achieves an accuracy of up to 99.9%. We discovered that the attention weights of important phrases are equivalent to, if not higher than, the attention weights of sentiment words in the sentence.

Keywords: Natural language processing; LSTM; deep learning; machine learning; BiDAF.

1. Introduction

E-commerce is quickly expanding these days. The fact that more and more people are shifting their shopping habits to online platforms via e-commerce is a prominent feature of the Internet era. Shopping online is extremely efficient, easy and thus is popular among individuals of various ages. However, when clients purchase online, they can only see virtual items rather than genuine goods. Customers frequently examine product reviews as an essential reference resource before making a purchase to confirm that the product fits the description. As a result, product reviews are an important indication for appraising items. As a result, sentiment analysis is a problem that may be used to evaluate products [1]. Such a quantitative analysis can benefit both producers and buyers.

Currently, sentiment analysis is divided into numerous classification strategies, including aspect-level sentiment classification, sentence-level sentiment classification, and text-level sentiment classification. Furthermore, sentiment categorization is categorized according to a variety of domains and approaches. [2]

The document is the primary unit for text-level sentiment analysis. The paper is regarded as a completed component on a certain subject. Onan [3], for example, suggested a CNN-LSTM architecture with five layers for text-level sentiment analysis. Fauzi [4] investigated the SVM model in conjunction with the TF-IDF.

Sentiment categorization at the sentence level entails extracting fundamental units and recognizing individual terms such as nouns and adjectives. These adjectives and nouns reflect a certain point of view. For example, “amazing” indicates a good attitude, but “degrade” expresses a bad viewpoint. However, if the word “not” is introduced, the perspective shifts. Smetanin [5] used CNN and pre-trained Word2Vec to do sentiment analysis at the phrase level.

The basic purpose of aspect-level sentiment analysis is to find views and aspects in a text. For example, “Excellent keyboard, long battery life, largest hard drive, and Windows 7”. The aspect phrases in these statements are “keyboard”, “battery life”, “hard drive”, and “Windows 7.” The emotion words recovered were “good”, “long”, and “biggest” [6]. Aspect-level sentiment categorization is also commonly used.

Many research studies are now focusing on the three primary categories of sentiment analysis difficulties stated above. These models' interpretability, however, remains limited. In this study, we use the BiDAF model to solve the problem at the phrase level. Attention processes are used in this model to make it more interpretable.

The (2) Relevant Work, (3) Suggested Model, (4) Results and Discussion, and (5) Conclusion will all be covered in the next portion of our work.

2. Related work

In general, the proposed methodologies are lexicon-based, machine learning-based, and deep learning-based. Many academics have improved traditional machine learning and deep learning algorithms to better handle sentiment analysis tasks. The key adjustments are based on sentiment lexicons and produce better outcomes [8].

The sentiment lexicon's structure is the most important modification. It determines how emotionally charged specific words or phrases in a text are. Each word in this technique must be assigned a sentiment value and be placed in a lexicon of positive and negative terms. A text passage is fundamentally represented as a collection of words in the lexicon-based method. In this bag of words, all positive and negative words or phrases in the text have emotional values provided from the dictionary.

Using approaches such as Word2Vec and FastText, word vector characteristics may be retrieved from supplied texts. Because these algorithms cannot distinguish the emotional components of words and phrases, emotion ratings must still be supplied manually. Conditional Random Fields (CRFs), a type of probabilistic model, Naive Bayes, Support Vector Machines, Maximum Entropy, Random Forests, and Conditional Maximum Entropy are among the most prominent machine learning approaches [9-10].

Deep learning algorithms have recently gained popularity due to their high performance in processing enormous bodies of data. There are more popular models than CNN, RNN, bidirectional RNN, and, most recently, attention-based networks. Li et al. [11] established a framework for studying emotions that includes multiple levels of prior knowledge into embedding. To boost geographic structure embedding even further, Yu et al. [12] used emotional vocabulary and deep learning techniques to create a model that

embeds emotion intensity ratings, considerably enhancing spatial structure embedding. RNN and its variants, particularly RNN and LSTM, are the most common deep-learning models for processing sequential data [13-14].

Using biLSTM-CRF, Chen et al. [15] described a divide-and-conquer strategy for sentence-level sentiment classification. Utilizing the attention mechanism, Yuan et al. Jiang [16] suggested a method to classify emotions into a Question & Answer system using an attention-based GRU network, combining the goals of models A and B.

3. Model

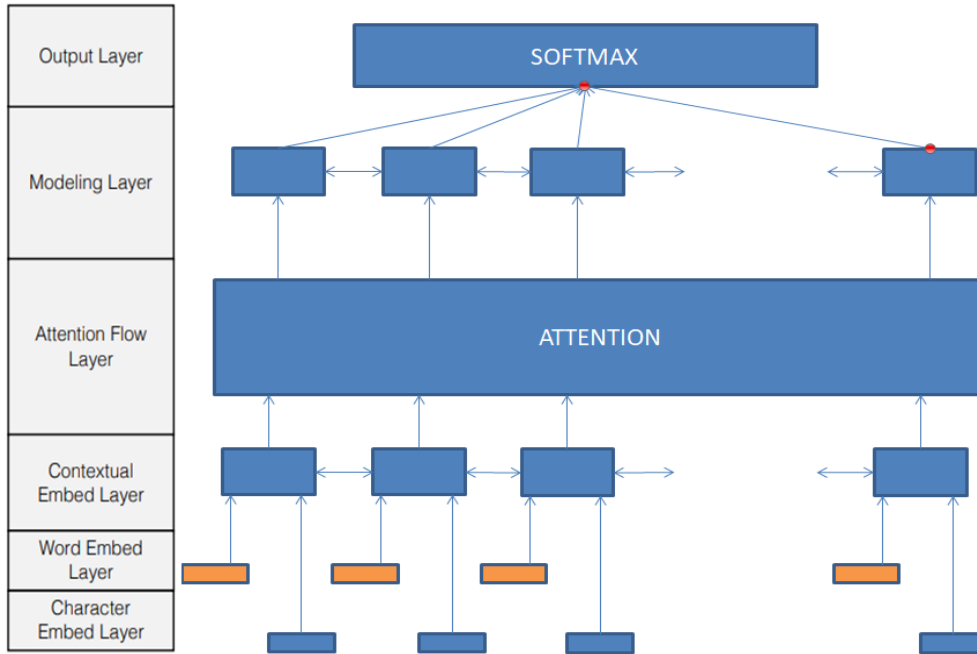


Figure 1: The BiDAF model for sentiment classification [17]

The BiDAF model [17] utilized in the emotion classification challenge is depicted in Figure 1. The BiLSTM model was initially designed for reading comprehension tests, however we use it for sentiment categorization tasks as follows:

Class **Character Embed Layer:** The vector from the character level is represented using the charCNN model.

Class **Word Embed Layer:** The Glove model is a popular method for learning word embeddings. Each word in the text is mapped to a vector space using a pre-trained embedding model.

The input of the next BiLSTM model combines the character-level vector and word embedding vector from the two layers above.

Class **Contextual Embed Layer:** The two vectors from the aforementioned layer serve as the input for the biLSTM (Bidirectional Long Short-Term Memory) model that we employ. We apply LSTMs in both directions and combine the two LSTM models' outputs. The LSTM model's formula is as follows:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + V_f c_{t-1} + b_f) \quad (1)$$

The formula for the i-th input gate is calculated as follows:

$$\begin{cases} i_t = \sigma(W_i x_t + U_i h_{t-1} + V_i c_{t-1} + b_i \\ c_{in_t} = \tanh(W_c x_t + U_c h_{t-1} + V_c c_{t-1} + b_c \\ c_t = f_t \cdot c_{t-1} + i_t \cdot c_{in_t} \end{cases} \quad (2)$$

The output gate is computed as follows:

$$\begin{cases} o_t = \sigma(W_o x_t + U_o h_{t-1} + V_o c_{t-1} + b_o) \\ h_t = o_t \cdot \tanh(c_t) \end{cases} \quad (3)$$

In this paper, use of the Bi-LSTM model is made. Both forward and reverse-directed bidirectional information can be produced. The input gate, forget gate, and output gate, respectively, are represented by the symbols i , f , and o . The matrices W , U , V , and b are the learning matrices.

Class Attention Layer: Each word in a text passage serves a certain purpose in the phrase. Different words trigger different emotions, which alters the meaning of the sentence. While certain words may have little effect on the meaning of the statement, others may have a substantial impact on the polarity of that attitude. As a result, the method necessitates a system that assigns a different weight to each word in a phrase. In this work, we used a self-attention approach to compute the attention weights of each word. The attention weight a_i for the hidden state h_i of the Bi-LSTM output is determined as follows:

$$u_i = \tan(W \cdot h_i + b) \quad (4)$$

$$a_i = \frac{e^{u_i^T \cdot u_w}}{\sum_i e_i^T \cdot u_w} \quad (5)$$

Afterward, the vector $S=(s_i)=(a_i \cdot h_i)$

In which W is the weight matrix and u_w is the global context vector.

Class Modeling Layer: The input to the modeling layer is S . This layer encodes the sequence. The output of the modeling layer captures the interaction between contextual words. This is different from the contextual embedding layer, which captures the interaction between contextual words independently. We use a bidirectional LSTM with output size d for each direction.

Class Output Layer: We aggregate the hidden layers of the modeling layer and use the softmax function to make predictions.

To train the model, we use the cross-entropy loss function:

$$L_{model} = -\frac{1}{S} \sum (y \log \hat{y} + (1 - y) \log(1 - \hat{y})) + \frac{\gamma}{2S} \|W\|_2^2 \quad (6)$$

4. Results

We utilize the Amazon video game review dataset [18] with 497,577 reviews to assess the suggested model. This dataset is divided into two halves, with 70% for training and 30% for testing. To assess the model's performance, we employed accuracy, recall, and f1 scores.

The following values are set as the model's parameters: Character-level word vectors have a dimension of 100 dimensions; embedded word vectors employ the Glove model, which has 300 dimensions; hidden layer word vectors in biLSTM models have a

dimension of 100; learning rate is set to 0.5; epochs are 17; batch size is 38; and dropout is 0.2.

Table 1: Comparison results of BiDAF model with basic models

Model	Precision	Recall	F1
SVM	48.0	50.0	49.0
Naïve Bayes	87.9	84.7	83.6
LSTM	95.2	99.9	97.5
BiDAF	95.9	99.9	97.66

According to the findings (Table 1), the BiDAF model surpasses the baseline models LSM, SVM, and Naive Bayes in terms of accuracy and F1 score, with 95.2% precision and 97.66% F1. This illustrates that the BiDAF model with attention mechanism outperforms the LSTM model in terms of prediction outcomes. This is due to the fact that the BiDAF model combines Glove word embeddings with character-level word embeddings. The use of character-level word embeddings aids in dealing with out-of-vocabulary and unusual terms in the test set.

5. Conclusion

With the growth of internet transactions, assessing product evaluations is becoming more vital and visible. The vast majority of customers will read product reviews on online shopping sites to form an accurate first impression. As a result, sentiment research is becoming increasingly important in aiding both sellers and buyers in deciding if their items meet their needs. In this research, we employed the BiDAF model to analyze sentiment. The results demonstrated that the BiDAF model had a considerable impact and outperformed earlier models. In the future, we will investigate the model's unique impact and test it on more data domains in the Amazon product review dataset.

REFERENCES

- [1] Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S. F., & Pantic, M., "A survey of multimodal sentiment analysis," *Image and Vision Computing*, 65, 3-14, 2017.
- [2] Onan, A., "Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks," *Concurrency and Computation: Practice and Experience*, 2020.
- [3] Fauzi, M. A., "Word2Vec model for sentiment analysis of product reviews in Indonesian language," *International Journal of Electrical and Computer Engineering*, 9(1), 525, 2019
- [4] Chen, X., Xue, Y., Zhao, H., Lu, X., Hu, X., & Ma, Z., "A novel feature extraction methodology for sentiment analysis of product reviews," *Neural Computing and Applications*, 31(10), 6625-6642, 2019.

- [5] Smetanin, S., & Komarov, M., "Sentiment analysis of product reviews in Russian using convolutional neural networks" In *2019 IEEE 21st Conference on Business Informatics (CBI)* vol. 1, pp. 482-486, 2019.
- [6] Yu, J., Jiang, J., & Xia, R., "Global inference for aspect and opinion terms co-extraction based on multi-task neural networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(1), 168-177, 2018.
- [7] Zucco, C., Liang, H., Di Fatta, G., & Cannataro, M., "Explainable sentiment analysis with applications in medicine" In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1740-1747), 2018.
- [8] Xin Li, Lidong Bing, Piji Li, Wai Lam, and Zhimou Yang, "Aspect term extraction with history attention and selective transformation," *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, {IJCAI-18}*, 4194-4200, 2018.
- [9] Chen, Y., Xiong, J., Xu, W., & Zuo, J., "A novel online incremental and decremental learning algorithm based on variable support vector machine," *Cluster Computing*, 22(3), 7435-7445, 2019.
- [10] Yang, L., Zhou, Y., & Zheng, Y., "Annotating the literature with Disease Ontology," *Chinese Journal of Electronics*, 26(6), 1261-1268, 2017.
- [11] Kim, Y. "Convolutional Neural Networks for Sentence Classification," 2014.[online] <https://arxiv.org/abs/1408.5882>.
- [12] Li, Y., Pan, Q., Yang, T., Wang, S., Tang, J., & Cambria, E., "Learning word representations for sentiment analysis," *Cognitive Computation*, 9(6), 843-851, 2017.
- [13] Yu, L. C., Wang, J., Lai, K. R., & Zhang, X., "Refining word embeddings using intensity scores for sentiment analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(3), 671-681, 2017.
- [14] Huang, Q., Chen, R., Zheng, X., & Dong, Z., "Deep sentiment representation based on CNN and LSTM" In *2017 International Conference on Green Informatics (ICGI)* pp. 30-33, 2017.
- [15] Hassan, A., & Mahmood, A., "Deep learning approach for sentiment analysis of short texts" In *2017 3rd international conference on control, automation and robotics (ICCAR)* pp. 705-710, 4/2017.
- [16] Chen, T., Xu, R., He, Y., & Wang, X., "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN," *Expert Systems with Applications*, 72, 221-230, 2017.
- [17] Min Joon Seo, Aniruddha Kembhavi, Ali Farhadi and Hannaneh Hajishirzi, "Bidirectional Attention Flow for Machine Comprehension," *5th International Conference on Learning Representations, {ICLR} 2017, Toulon, France, 24-26/4, 2017.*
- [18] <http://jmcauley.ucsd.edu/data/amazon/>

TÓM TẮT

MÔ HÌNH BiDAF TRONG BÀI TOÁN PHÂN TÍCH CẢM XÚC

Lương Thị Minh Huệ

Trường Đại học Công nghệ Thông tin và Truyền thông, Đại học Thái Nguyên, Việt Nam

Ngày nhận bài 10/02/2023, ngày nhận đăng 21/3/2023

Phân tích cảm xúc là một trong các bài toán quan trọng trong xử lý ngôn ngữ tự nhiên. Phân tích cảm xúc là bài toán được các doanh nghiệp rất quan tâm để kiểm soát đánh giá của khách hàng về sản phẩm của mình. Mô hình BiDAF được đề xuất sử dụng cho mô hình đọc hiểu sử dụng các cơ chế chú ý. Cơ chế chú ý gần đây được mở rộng và áp dụng thành công trong các bài toán trong xử lý ngôn ngữ tự nhiên. Trong bài báo này, chúng tôi áp dụng mô hình BiDAF vào bài toán phân tích cảm xúc ở cấp độ câu đánh giá sản phẩm từ Amazon. Mô hình BiDAF là mô hình với nhiều lớp xử lý. Mỗi lớp biểu diễn ngữ cảnh ở các mức độ khác nhau kết hợp và sử dụng mô hình BiLSTM để thu được biểu diễn ngữ cảnh. Đồng thời, chúng tôi xem xét sự phân bố trọng số chú ý của câu qua cơ chế chú ý. Mô hình có độ chính xác lên đến 99,9% với độ đo recall. Chúng tôi nhận thấy rằng các trọng số chú ý của các thuật ngữ chính bằng hoặc thậm chí cao hơn các từ ngữ chứa cảm xúc trong câu.

Từ khóa: Xử lý ngôn ngữ tự nhiên; LSTM; học sâu; học máy; BiDAF.