

Article

Running a Sustainable Social Media Business: The Use of Deep Learning Methods in Online-Comment Short Texts

Weibin Lin ^{1,2}, Qian Zhang ¹, Yenchun Jim Wu ^{3,4,*}  and Tsung-Chun Chen ⁵ 

¹ Business School, Huaqiao University, Quanzhou 362000, China

² TSL Business School, Quanzhou Normal University, Quanzhou 362000, China

³ MBA Program in Southeast Asia, National Taipei University of Education, Taipei 106, Taiwan

⁴ Graduate Institute of Global Business and Strategy, National Taiwan Normal University, Taipei 106, Taiwan

⁵ Department of Business, Dongguan City College, Dongguan 523419, China

* Correspondence: wuyenchun@gmail.com

Abstract: With the prevalence of the Internet in society, social media has considerably altered the ways in which consumers conduct their daily lives and has gradually become an important channel for online communication and sharing activities. At the same time, whoever can rapidly and accurately disseminate online data among different companies affects their sales and competitiveness; therefore, it is urgent to obtain consumer public opinions online via an online platform. However, problems, such as sparse features and semantic losses in short-text online reviews, exist in the industry; therefore, this article uses several deep learning techniques and related neural network models to analyze Weibo online-review short texts to perform a sentiment analysis. The results show that, compared with the vector representation generated by Word2Vec's CBOW model, BERT's word vectors can obtain better sentiment analysis results. Compared with CNN, BiLSTM, and BiGRU models, the improved BiGRU-Att model can effectively improve the accuracy of the sentiment analysis. Therefore, deep learning neural network systems can improve the quality of the sentiment analysis of short-text online reviews, overcome the problems of the presence of too many unfamiliar words and low feature density in short texts, and provide an efficient and convenient computational method for improving the ability to perform sentiment analysis of short-text online reviews. Enterprises can use online data to analyze and immediately grasp the intentions of existing or potential consumers towards the company or product through deep learning methods and develop new services or sales plans that are more closely related to consumers to increase competitiveness. When consumers experience the use of new services or products again, they may provide feedback online. In this situation, companies can use deep learning sentiment analysis models to perform additional analyses, forming a dynamic cycle to ensure the sustainable operation of their enterprises.

Keywords: online-comment short texts; sentiment analysis; deep learning; BERT



Citation: Lin, W.; Zhang, Q.; Wu, Y.J.; Chen, T.-C. Running a Sustainable Social Media Business: The Use of Deep Learning Methods in Online-Comment Short Texts. *Sustainability* **2023**, *15*, 9093. <https://doi.org/10.3390/su15119093>

Academic Editors: Alessia Passanisi, Stefano Ruggieri and Harris Wu

Received: 27 February 2023

Revised: 6 May 2023

Accepted: 1 June 2023

Published: 5 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, social media has become a major part of people's lives and has significantly impacted the way businesses communicate with their customers. However, with the increasing amount of online content and user-generated comments, it has become increasingly challenging for businesses to monitor and manage the sentiments and opinions of their customers. With the popularity of the Internet at present, the presence of social media is changing almost every aspect of people's daily lives and is gradually becoming an important channel for Internet users to communicate and share their ideas [1]. Because of their characteristics of convenience and rapidity, Microblog, WeChat, QQ, and Zhihu influence people's daily lives and help them to be aware of the latest current affairs and exchange their thoughts [2]. Microblog is a leading social media platform for people to create, share, and discover information online. It not only provides users with more freedom

and a faster way to communicate, express their opinions, and record their feelings, but it also has the advantages of providing rich information, being simple to operate, providing rapid dissemination, and having a wide-ranging audience. As a result, increasingly more people have begun to use it [3,4]. According to the financial report data for Q3 on Microblog in 2018, the number of monthly active Microblog users increased to 446 million, continuing to maintain a year-on-year net growth of 70 million, and the number of daily active users also increased to 195 million [5].

Numerous short texts are created in the process of using Microblogs to communicate, share, and disseminate information, which includes people's emotions, such as happiness, sadness, and joy, and certain sentiments, such as like, dislike, and hate [6]. People desire to express their emotions while communicating through Microblogs, such as their views on news events, opinions on social hot topics, movie comments, and item evaluations, so that a high quantity of online data are generated every day. An effective sentiment analysis of these data on Microblog has many practical uses. For example, for consumers, it can provide a reference for purchasing goods. For e-commerce platforms, goods are marketed on the Microblog. It can also be used to help relevant government departments make decisions and improve network security and public-opinion monitoring systems [7,8].

However, online-comment short texts sent on Microblogs are different from traditional texts, such as those sent on forums and the news. Their specific characteristics are as follows: First, the number of words in the texts should be limited to 140. In addition to the general text and punctuation, the former contains a variety of emoticons, special symbols, stop words, and pictures [9,10]. Second, the language used is not standardized because the users' knowledge and language use are different; for example, network terms such as "skr" and "lemon essence" exist [11]. As time passes, online-comment short texts on Microblogs have the characteristics of a large amount of data, more neologisms, serious colloquialisms, and rich content, which pose a challenge to the technical requirements of mining users' sentiment orientations from large online-comment short texts presented on Microblogs [12]. In recent years, deep learning techniques have emerged as a powerful tool for analyzing and mining text data. Wang et al. (2020) proposed a deep learning approach to Chinese Microblog topic modeling [3]. Jiang and Yang (2019) developed a hybrid approach for a product quality assessment based on the user-generated content presented on social media [4]. Huang et al. (2019) used a hierarchical deep learning model to perform fine-grained emotion analysis on Chinese online reviews [5]. Nishida et al. (2019) proposed an active learning approach for social media text classification via query generation based on kernel density estimation [6]. Chen et al. (2020) developed a novel deep learning-based approach for detecting fake news on social media [7]. Furthermore, deep learning techniques have also been applied to text mining and sentiment analysis in social media (Li et al., 2019; Guo and Zhang, 2019) [8,9]. Convolutional and recurrent neural networks have also been compared in the literature for sentiment analysis (Fang et al., 2019) [10], and event extraction and rumor detection activities in social media have been tackled by novel deep learning approaches (Li et al., 2020; Fu et al., 2020) [13,14]. Therefore, a series of models are proposed in the present study for the construction and optimization of the sentiment analysis of online-comment short texts through deep learning technology and the related neural network model, aiming to rapidly analyze the sentiment analysis of online-comment short texts and facilitate the development of related e-commerce marketing, public opinion monitoring, and other businesses. Moreover, precision marketing and services have been used in the field to enhance enterprise competitiveness and seek opportunities for sustainable operations. If enterprises can use online data and conduct sentiment analyses through deep learning techniques, they can rapidly grasp the reactions and apprehend the intentions of existing and potential consumers concerning their products. Based on this information, enterprises can further modify, develop, or produce sales, services, and new products that are more in line with consumer needs at present. When consumers receive new services or products, they provide feedback online, which forms new online data. At

this time, enterprises can use deep learning sentiment analysis models to perform further analyses, creating a dynamic cycle and achieving sustainable business operations.

2. Method

2.1. The BERT Model

BERT (bidirectional encoder representations from transformers) is a representative model of bidirectional encoders based on Transformer [15]. The BERT model was proposed by Google in 2018, and it uses the Transformer encoder as the main structure and is stacked by the encoder structures of multiple Transformer models. The structure is presented in Figure 1 [16].

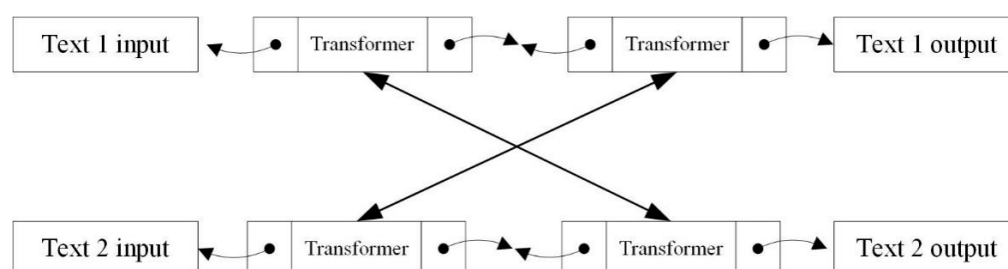


Figure 1. Structure of the BERT model.

Figure 1 shows that the BERT model uses MLM (masked language model) and next-sentence prediction methods to capture sentiment representations at word and sentence levels, respectively [17]. The main structure of the BERT model is stacked by Transformer, and the text encoding of the Transformer is based on the attention mechanism [18]. Therefore, different weights can be assigned to each word according to the relationships evident between the words and sentences [19].

2.2. CNN (Convolutional Neural Network)

In the 1960s, Hubel and Wiesel proposed the CNN, whose concept comes from the neocognitron model, proposed by Japanese scholar Fukushima, based on the concept of the perceptual field [20]. The core of a CNN is to capture local correlations. The CNN is mainly used in the field of image recognition, and it has achieved great success in the recognition of different types of images, such as plants, animals, and food [21]. Yi used the CNN to identify different types of birds [22]. The convolution layer is the core of the whole CNN. Unlike the convolution layer in the image field, the convolution kernel no longer uses the square convolution kernel commonly used in the image field but proposes a rectangular convolution kernel, which presents a good effect on the short-range semantic acquisition of a text [23,24]. In a study, Sun proposed a model for text analysis using the CNN. The word-cut text obtained following the pre-processing stage is transformed into a word vector by the word vector tool as the input of the CNN, and the labeled texts are used to train the model. The trained model is used to extract the features of the text analysis [25]. The main features of the CNN are local perception, parameter sharing, multi-core convolutions, and downsampling. A CNN is good at extracting short-range semantic features in a text analysis [26].

2.3. Recurrent Neural Network

The RNN (recurrent neural network) is a kind of neural network model used for processing sequence data. In comparison to the traditional neural network model, the nodes in the hidden layer of the network model are connected, which makes the structure and parameters of the RNN shared at each time point [27]. The greatest feature of sequence data-processing behavior is that the output of the current moment contains the output of the previous moment, which continues to preserve the information of the previous moment

through the cyclic preservation process. The feature provides the RNN with “memory” [28]. Figure 2 presents the structure of the RNN.

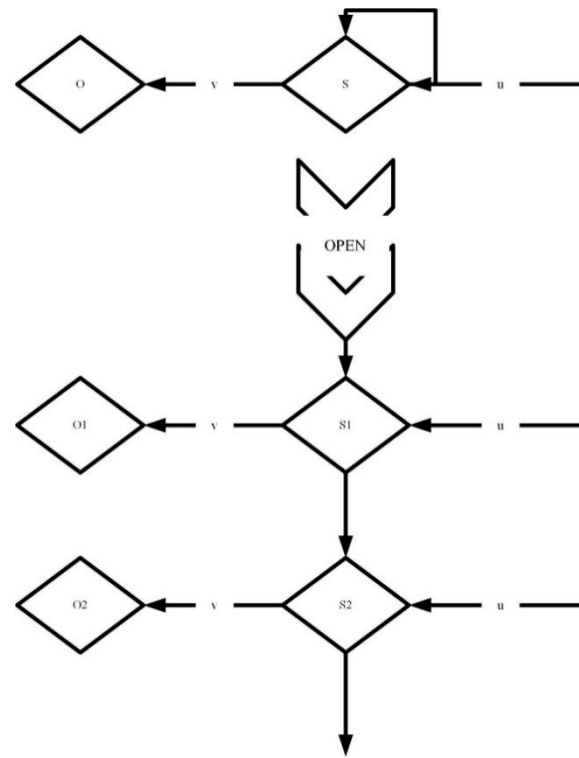


Figure 2. Structure of RNN.

When the RNN is expanded according to time, the calculation method of the cyclic network can be obtained, as shown in Equations (1) and (2) [29]:

$$s_t = h(Ux_t + ws_{t-1}) \quad (1)$$

$$o_t = f(Vs_t) \quad (2)$$

Here, t and $t + 1$ are the time series; x is the input sample; s_t is the memory of the sample at time t ; U is the weight from the input layer to the hidden layer; w is the weight of the sample in the hidden layer; V is the weight of the output sample (w , U , and V are shared at all times); and f and g are activation functions— f can be tanh, relu, sigmoid, and other activation functions, and g can be SoftMax and other functions. In the initial state, s_t is usually set to 0, and the network hidden-layer output $o_t = \text{softmax}(Vs_t)$ at time t . The memory of the RNN is limited. With the increase in sequence length, the amount of long-distance information that can be memorized reduces and the information-processing time increases, resulting in the problem of gradient disappearance and explosion. Given this situation, researchers also proposed an LSTM (long short-term memory) model, which controls the input and output values of the current time information through the structure of the “gate”, including the forgetting, input, and output gates. In order to simplify the structure of the model, the GRU (gated recurrent unit), as a variant of the LSTM, can also solve the problem of gradient disappearance and explosion in the long-term dependence of recurrent neural networks, such as the LSTM model [30].

2.4. The BiGRU Network Model

The BiGRU (bi-directional gated recurrent unit) network model combines two common unidirectional GRU networks into a two-way GRU network structure to obtain the sequence

characteristics of the text. The structure of the network model of BiGRU is presented in Figure 3 [31].

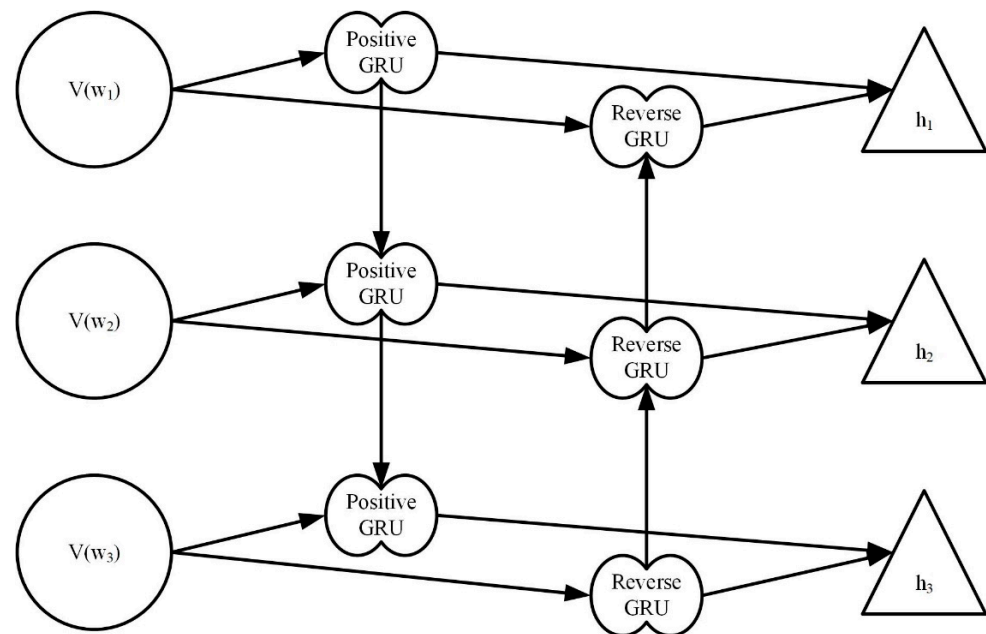


Figure 3. Structure of the BiGRU network model.

Figure 3 shows that, since a single GRU neural network lacks consideration of the following information in a text analysis, a two-way GRU network can be used to obtain the semantics in the context [17]. Among them, $V(W_1)$, $V(W_2)$, and $V(W_3)$ represent the word vectors transformed by the vector representation model, and h_1 , h_2 , and h_3 are the final vectors obtained by splicing following the feature selection of forwarding and reverse GRU networks [32].

2.5. The Attention Mechanism

As shown in Figure 4, $V(W_1)$, $V(W_2)$, and $V(W_3)$ are the word vectors transformed by the model using vector representation, and h_1 , h_2 , and h_3 are the final vectors obtained by splicing the forward and reverse GRU feature selections and input into the attention mechanism. y is the final result achieved following the analysis [33]. The specific calculations are presented in Equations (3)–(6):

$$e_t = f(u, h_t) \quad (3)$$

$$f(u, h_t) = u^t h_t \quad (4)$$

$$a_t = \frac{e_t}{\sum_{t=1}^n e_t} \quad (5)$$

$$y = \sum_{t=1}^n a_t h_t \quad (6)$$

Here, $t = 1, 2, 3 \dots$, u is the parameter of the attention mechanism, h is the output vector value, and a_t is the normalized attention score, which is obtained by Equation (5). y is the summation of the product of output vector value h in the BiGRU layer and the normalized attention score a_t , as presented in Equation (6).

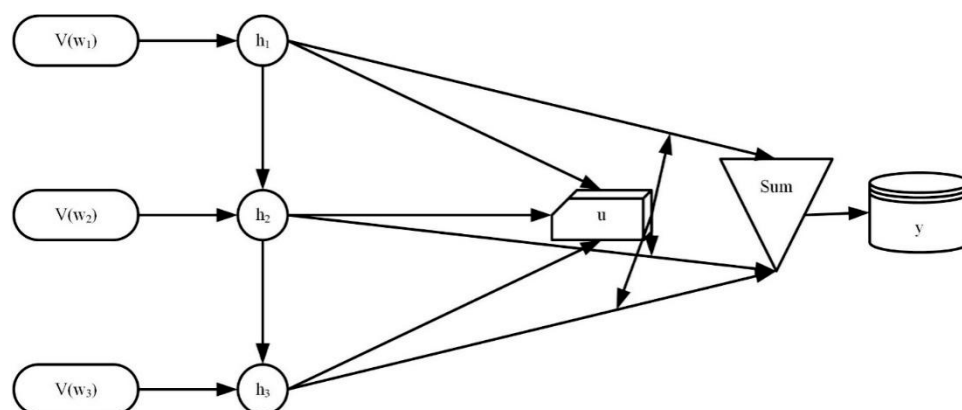


Figure 4. The attention model.

2.6. The Vector Representation Model of Online-Comment Short Texts

In the research conducted on sentiment analysis methods based on deep learning networks, the vectorization of texts is an important step, and it is necessary to use vector representation technology to represent the text as a vector and then as the input of the sentiment analysis model based on deep learning methods.

2.6.1. Construction of the Vector Representation Model

The vector representation model is used to represent texts as vectors, thereby providing complete data for the sentiment analysis model based on deep learning methods. Prior to the text vector representation stage, the texts must be pre-processed, and then vector representation models, such as Word2Vec (Word To Vector), are used to transform texts into vectors. The vector representation process of online-comment short texts on Microblog is presented in Figure 5.

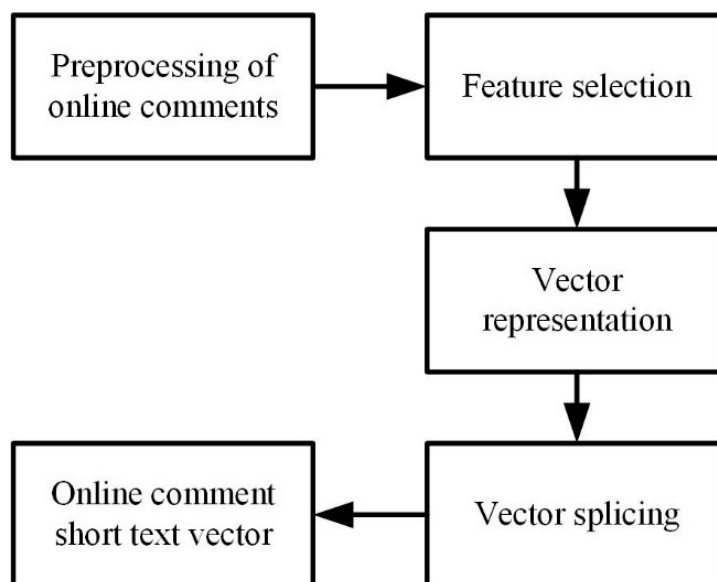


Figure 5. The vector representation process of online-comment short texts on Microblog.

Figure 5 shows that the online-comment short texts presented on Microblog must be pre-processed and feature-selected prior to the vector representation step. The pre-processing step can clean out the redundant and disorderly data present in the short texts; feature selection reduces the number of features present and the feature dimensions. Following the vector representation step, the feature vectors of multiple features must be spliced together to form the final text vector representation.

A vector representation method of online-comment short texts on Microblog based on the BERT model was proposed to transform the short texts into vector representations that could input the sentiment analysis model, aiming at the problem of sparse features and a lack of short-text semantics on Microblog. This method uses the BERT model to embed words into the pre-processed short texts and analyzes a polysemy while the texts are transformed into vectors, thus generating more accurate text representation vectors.

2.6.2. The Vector Representation Model of Short Texts Based on the BERT Model

At present, Word2Vec and Glove are mostly used in the field to train word vectors in the sentiment analysis models of online-comment short texts on Microblog based on deep learning methods. However, word vectors trained by these vector representation models belong to static vector representations; that is, the vector representation results of the same word in different contexts are the same. This leads to some deviations in the results of the sentiment analysis. For example, the word “apple” can represent either fruit or the Apple company. If the word “apple” is represented as the same vector in these different short texts, a certain deviation is evident for subsequent sentiment analyses. A text vector representation method based on BERT embedding was proposed to improve the semantic representation of the vector of the short texts since numerous polysemous words exist in the online-comment short texts on Microblog and the word vector models, such as Word2Vec, cannot express and process them well.

Each word input in the BERT model consists of three parts: the word vector that represents the current word encoding method (token embeddings), the segment vector that represents the position encoding of the current word in the sentence (segment embeddings), and the position vector that represents the position encoding of the current word in the sentence (position embeddings). The details are presented in Figure 6.

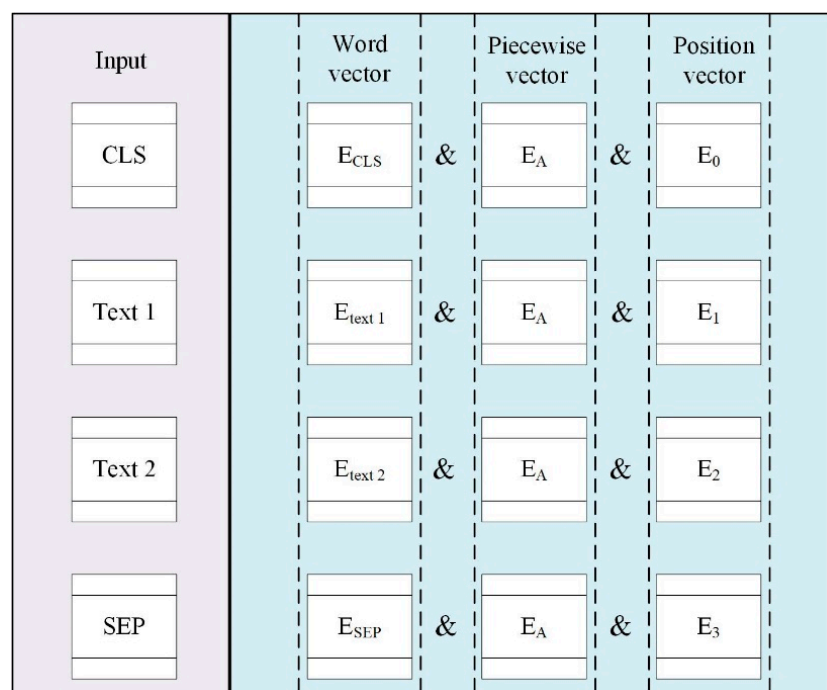


Figure 6. Input process of the BERT model.

Figure 6 shows that each text, such as Text 1 and Text 2, that composes the input texts, comprehensively considers the relationship between texts prior to the use of the BERT to train the text vector, which is reflected by the segmented and position vectors, respectively. The core idea of the BERT model for text vector representation is to calculate the relationship that exists between the texts in a sentence, which reflects the semantic correlation between

different texts. Then, the weight is allocated according to the importance of each text by using the coding mechanism of Transformer, so that the vector representation of the text contains more abundant semantics. Compared with the text vector of Word2Vec, the text vector based on the BERT model is more universal in its application.

2.6.3. Evaluation Indicators

The universal evaluation criteria were used to evaluate the effect of sentiment analyses on the experimental model. Precision, recall, and F1-scores under positive- and negative-emotion categories were used as the evaluation indicators for the effect of the model analysis. Indicators and their corresponding contents are presented in Table 1.

Table 1. Evaluation indicators and their meanings.

Indicators	Meanings
TP	Both prediction and practice are positive
FP	The prediction is positive, but the reality is negative
TN	Both prediction and reality are negative
FN	The prediction is negative, but the reality is positive

The calculation equations for precision, recall, and F1-scores under positive- and negative-emotion categories are presented in Equations (7)–(12), where pos denotes the positive-emotion label and neg denotes the negative-emotion label.

$$precision_{pos} = \frac{TP}{TP + FP} \quad (7)$$

$$precision_{neg} = \frac{TN}{TN + FN} \quad (8)$$

$$recall_{pos} = \frac{TP}{TP + FN} \quad (9)$$

$$recall_{neg} = \frac{TN}{TN + FP} \quad (10)$$

$$F1_{pos} = \frac{2 \times P_{pos} \times R_{pos}}{P_{pos} + R_{pos}} \quad (11)$$

$$F1_{neg} = \frac{2 \times P_{neg} \times R_{neg}}{P_{neg} + R_{neg}} \quad (12)$$

In the equations above, P is the precision rate, and R is the recall rate.

Accuracy represents the analytical performance of the training model for the sample, as shown in Equation (13).

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad (13)$$

2.7. The Sentiment Analysis Model Based on the BiGRU-Att Model

Increasingly more words are used to express people's opinions and emotions, which poses a challenge to sentiment analysis techniques based on deep learning methods. The problems of sentiment analysis based on deep learning methods are less-advanced data pre-processing technology, inaccurate text vector representation, and feature extraction dependent on deep neural networks. The sentiment analysis process of online-comment short texts based on deep learning methods includes data pre-processing, text vector representation, and sentiment analysis methods, as shown in Figure 7.

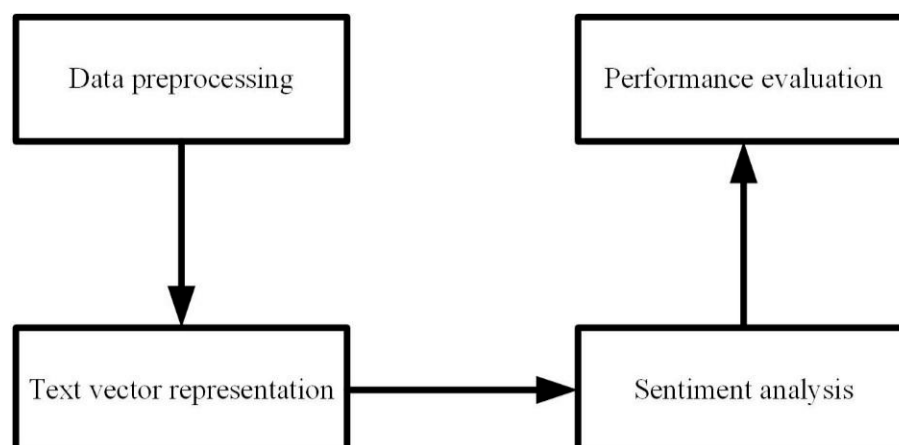


Figure 7. Sentiment analysis process based on deep learning methods.

Figure 7 shows that the data pre-processing method mainly includes data cleaning, text segmentation, and the removal of stop words. The data cleaning process removes content that is not relevant to subsequent sentiment analysis tasks, such as various symbols. Text segmentation is the use of word segmentation tools to perform text segmentation, especially knot segmentation. Text vector representation is mainly used to represent the pre-processed text as a vector using Word2Vec, Glove, and other neural networks, and then as the input of the deep learning model. Sentiment analysis refers to the use of machine learning, deep learning, and other technologies to construct models to perform the sentiment analysis of texts. Sentiment analysis based on deep learning methods should first train the model with the labeled data set, then use the trained optimal parameter model to extract the text feature, and finally use the analyzer to analyze the sentiment analysis, obtaining the final sentiment analysis results. Performance evaluation uses common evaluation criteria to evaluate the performance of different emotion analysis models.

2.7.1. The Construction of the Sentiment Analysis Model Based on the BiGRU-Att Model

Since the current sentiment analysis method of online-comment short texts on Microblog based on the deep neural network does not focus on the importance of sentiment words or phrases in the extraction of text sentiment features, the attention mechanism is introduced, and a BiGRU-Att (bi-directional gated recurrent unit attention) model based on the attention mechanism was proposed by combining the BiGRU deep neural network. The attention mechanism can assign different weights to the feature words with different sentiment contributions and add more weight to the words with more sentiment contributions. The BiGRU model can effectively obtain the semantic information of online-comment short texts on the Microblog. The model structure is presented in Figure 8.

Figure 8 shows that the entire BiGRU-Att model is mainly divided into four layers: the input, BiGRU, attention, and output layers of the sentiment analysis. The contents of each layer are as follows:

The input layer: The vector matrix of the text is input. Given text $d = \{s_1, s_2, s_n\}$, s_n denotes the sentence that constitutes the text. Sentences can be regarded as the sequence information of words. Given sentences $s = \{W_1, W_2, W_n\}$, W_n denotes words that compose the sentences. Word2Vec and other word vector representation tools are used to convert W into the form of the word vector $V(W_n)$, and the dimension of the word vector is m . Then, the input text is represented in the form of the $n \times m$ vector matrix, providing complete data for the BiGRU layer.

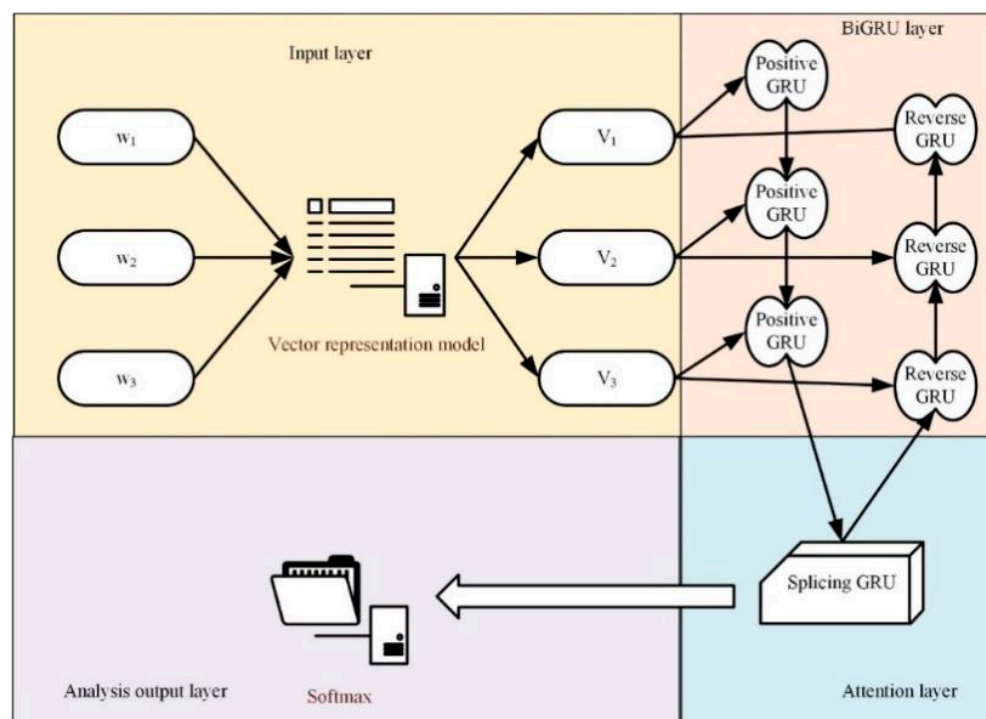


Figure 8. Structure of the sentiment analysis model based on the BiGRU-Att model.

The BiGRU layer: Used to extract the semantic information features of online-comment short texts on the Microblog. A bidirectional GRU network is used. One GRU direction is in the positive-order direction of the input sequence, and the other is in the inverse-order direction of the input sequence. In the feature extraction of the input sequence, there is no shared state between GRUs in the two directions. The state of GRUs in the positive direction is only transferred along the positive direction, and the reverse direction is only transferred along the reverse direction. However, at the same time, the output results of the GRU in two directions are spliced as the output values of the entire BiGRU layer. This process considers the semantic information presented above as well as the semantic information below.

The attention layer: A structure added between the BiGRU and output layers. It uses the splicing vectors output by all the BiGRU networks for the input of automatic learning features. According to the contribution of each word, it dynamically assigns different attention weights to each word and adds additional weights to the most relevant features of the sentiment analysis, so that the sentiment characteristics of the text are more obvious. Therefore, the output vector of the attention layer considers the context and semantic information of the text and key sentiment word features.

The analysis output layer: Selects a SoftMax analyzer to perform the final sentiment analysis. After the attention layer assigns the weights to the BiGRU layer output features, the results are input into the SoftMax analyzer. Analysts output the final results of integration in the form of arrays. Since positive and negative emotions were evident in the study, the content in the array represents the probability that the text's emotion is positive or negative.

2.7.2. The Contrast Experiment

The sentiment analysis model based on the BiGRU-Att model proposed in the study was compared with the following three commonly used sentiment analysis models to verify its effect:

- The CNN sentiment analysis model.
- The BiLSTM (bi-directional long short-term memory) emotion analysis model. The comparative experiment of the sentiment analysis model based on the BiLSTM model shows that the number of hidden neurons is 100 and the activation function is tanh.
- The BiGRU emotion analysis model. The sentiment analysis model based on BiGRU is the BiGRU-Att model, which directly connects the output results of BiGRU to the output layer of the emotion analysis following the removal of the attention layer.

3. Results and Discussion

3.1. Experimental Results and Analysis of the Short-Text Vector Representation Model Based on the BERT Model

In order to improve the quality of word vectors and the semantic representation ability of online-comment short texts on the Microblog, a vector representation method based on the BERT model was proposed to solve the problem of the sparse feature of online-comment short texts on the Microblog. The effects of two different word embedding methods were compared in the experiment. BERT and Word2Vec were used to obtain the vector of the text, and the CNN model was selected as the analysis model for the comparative experiment. The experimental results are presented in Figure 9.

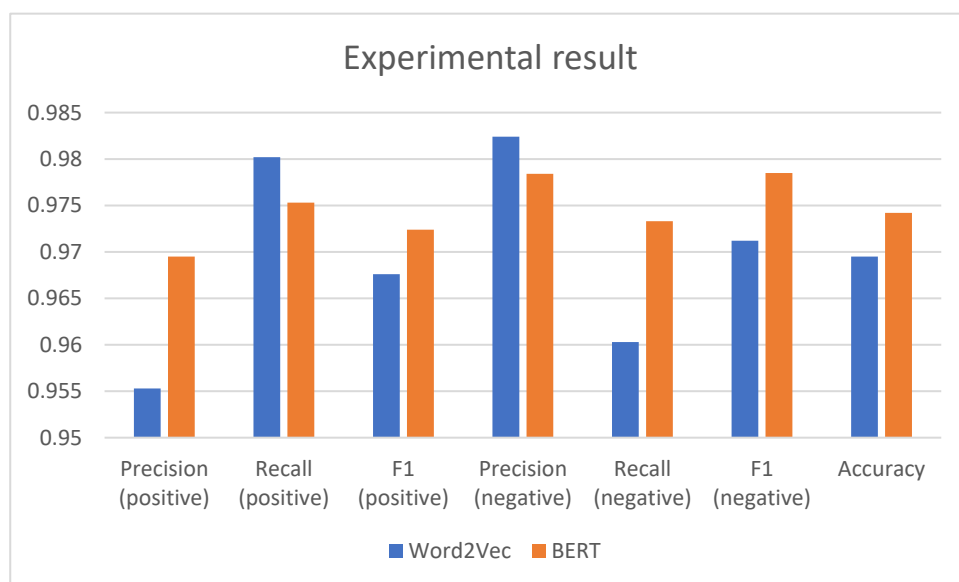


Figure 9. Experimental results for word vectors in two different models.

Figure 9 shows that the vector representation based on the BERT model is better than that based on the Word2Vec model in the sentiment analysis. The reason for this result was that the BERT transformer mechanism could obtain more semantic representations, which can be observed in the derivation of the word vector representation model in the second chapter. In summary, the effect of the pre-training word vector using the BERT model is better than that of the CBOW (continuous bag-of-words model) based on Word2Vec.

3.2. Experimental Results and Analysis of Sentiment Analysis Model Based on the BiGRU-Att Model

In order to verify the effectiveness of the BiGRU-Att model proposed in this study, the two vectors presented in Section 3.1 were used to represent the word vector trained by the model as the input of the emotion analysis model. The experimental results obtained are as follows:

The experimental results of the sentiment analysis model based on deep learning methods under the word vector based on the Word2Vec model are presented in Figure 10.

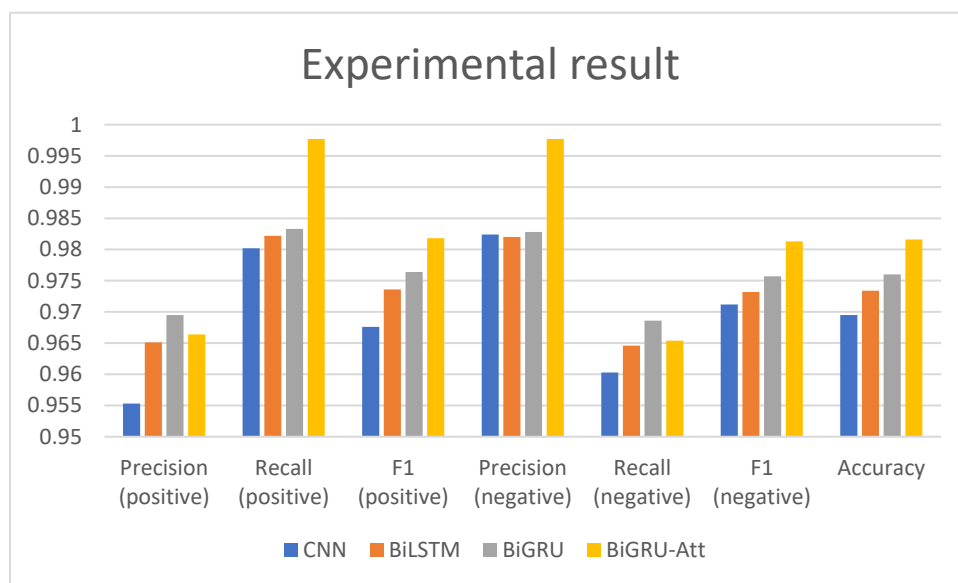


Figure 10. Experimental results based on the Word2Vec model.

Figure 10 shows that, among the four sentiment analysis models based on the Word2Vec model, the BiGRU-Att sentiment analysis model proposed in this study presents the highest accuracy among the four models, reaching 98.09%. Although the LSTM and GRU models can process long sequences, the experimental results show that the accuracy of the BiGRU model is 97.58%, which is higher than that of the BiLSTM model. The experimental results also show that the accuracy increases by 0.53% compared with the unfocused BiGRU model.

The experimental results of the sentiment analysis model based on deep learning methods under the word vector based on the BERT model are presented in Figure 11.

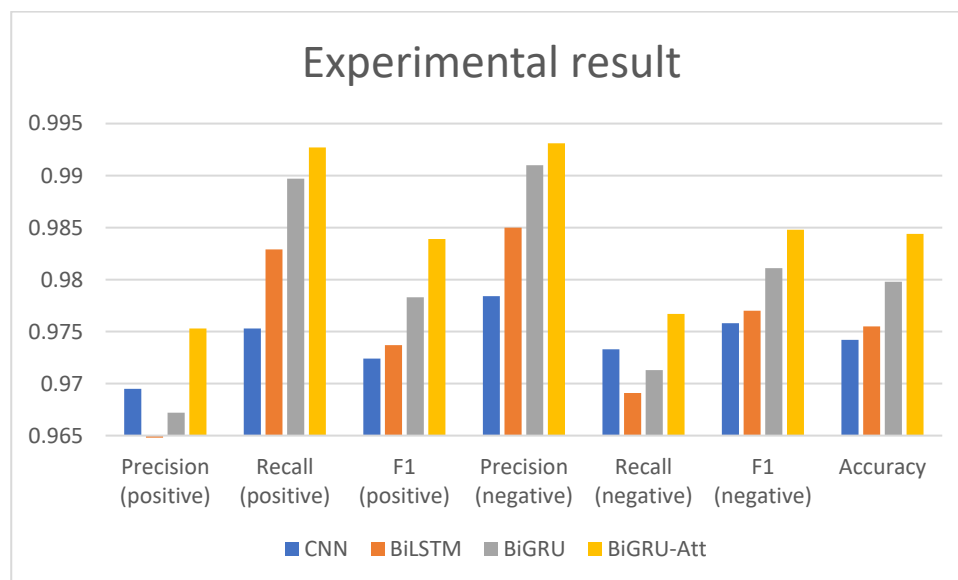


Figure 11. Experimental results of the word vector based on the BERT model.

Figure 11 shows that, among the four sentiment analysis models based on the word vector of the BERT model, the accuracy of the BiGRU-Att model is the highest, reaching 98.32%. Compared with the BiGRU model without the attention mechanism, the accuracy is increased by 0.39%, which also verifies that the BiGRU sentiment analysis model with an attention mechanism can improve the accuracy of the sentiment analysis.

The advantages of the BiGRU-Att model proposed in this study are the following: The GRU can effectively record the temporal information of the text; the BiGRU not only learns the semantic information mentioned above but also the semantic information presented below; different words are assigned different weights following the introduction of the attention mechanism, focusing on important sentiment words in the text. From the abovementioned experimental results, we observed that the BiGRU-Att sentiment analysis model proposed in this study can effectively improve the accuracy of the sentiment analysis of online-comment short texts on the Microblog.

4. Conclusions

With the popularity of social networking tools such as Microblog and WeChat, short online comments have become the disseminators of people's online dialogues and interactions. The sentiment analysis of online-comment short texts on Microblogs is, at present, a popular direction in the field of natural language processing techniques. The new word-discovery method, vector representation model, and sentiment analysis model based on the deep learning methods of online-comment short texts on Microblog were studied in this paper. The main subjects of this study were as follows: (1) Aiming at the problem of the sparse feature of short texts on the Microblog, a short-text representation model of online-comment short texts based on BERT embedding was constructed. In comparative experiments, the influence of two different word vector representation models based on BERT and Word2Vec on the accuracy of the sentiment analysis model based on deep learning methods was explored. (2) The application of deep learning methods to the sentiment analysis of short texts on Microblog was studied. Following the introduction of the attention mechanism, the BiGRU-Att sentiment analysis model of online-comment short texts on Microblog was designed. Through the comparative experiments we performed, the BiGRU-Att analysis model presented high-accuracy results for the data set compared with the CNN, BiLSTM, and BiGRU models, which proves that the correct addition of attention mechanisms can improve the accuracy of the sentiment analysis model of online-comment short texts on the Microblog. This study explored how to use online data and deep learning methods for sentiment analysis, which helps companies rapidly understand the reactions and purchase intentions of current and potential consumers towards their products. It further enables them to modify, develop, or produce sales, services, and new products that better meet consumer demands at present, creating new online data and achieving sustainable business operations through dynamic cycles.

5. Research Limitations and Future Research Directions

In this study, we present some of the shortcomings of the present research and some additional thoughts on future research directions, and the following two aspects should be emphasized and studied further: (1) Through our experiments, we observed that the text embedding method based on the BERT model had a better effect following the sentiment analysis. In the future, more auxiliary sentiment features, such as sentiment images, user portraits, and social networks, should be fused based on BERT. (2) Only the positive and negative categories were studied in our study, neglecting other emotions. Therefore, the follow-up research must focus on other emotions, such as surprise and joy.

Author Contributions: W.L., Q.Z., Y.J.W. and T.-C.C. jointly participated in the writing of the research and structure; W.L. established the index system and completed the statistical analysis; Q.Z. and T.-C.C. provided some case research paper references; Y.J.W. contributed valuable assistance during the manuscript writing and also assisted in the writing and modifying the manuscript formats. All authors have read and agreed to the published version of the manuscript.

Funding: This research receives support from National Science and Technology Council, Taiwan (111-2410-H-003-072-MY3).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wu, D.; Yang, R.; Shen, C. Sentiment word co-occurrence and knowledge pair feature extraction based LDA short text clustering algorithm. *J. Intell. Inf. Syst.* **2020**, *56*, 1–23. [\[CrossRef\]](#)
2. Zhou, W.S. A Method of Short Text Representation Based on the Feature Probability Embedded Vector. *Sensors* **2019**, *19*, 3728. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Wang, X.; Zhou, C.; Wu, J. A deep learning approach to Chinese microblog topic modeling. *J. Intell. Fuzzy Syst.* **2020**, *39*, 653–663.
4. Jiang, L.; Yang, H. A hybrid approach to product quality assessment based on user-generated content in social media. *Inf. Process. Manag.* **2019**, *56*, 2015–2026.
5. Huang, X.; Zhang, L.; Ding, X. Fine-grained emotion analysis in Chinese online reviews with hierarchical deep learning. *Inf. Process. Manag.* **2019**, *56*, 874–889.
6. Nishida, T.; Hara, K.; Yoshikawa, H. Active learning for social media text classification via query generation based on kernel density estimation. *Inf. Process. Manag.* **2019**, *56*, 752–766.
7. Chen, J.; Ren, Y.; Chen, J. A novel deep learning-based approach for detecting fake news on social media. *J. Intell. Fuzzy Syst.* **2020**, *39*, 1387–1395.
8. Li, H.; Liu, T.; Li, S. A survey of deep learning for text mining. *Inf. Fusion* **2019**, *101*, 9–19.
9. Guo, Y.; Zhang, Y. Deep learning based sentiment analysis research: A survey. *Inf. Process. Manag.* **2019**, *56*, 1601–1619.
10. Fang, H.; Guo, Y.; Zhang, Y. A comparative study of convolutional neural network and recurrent neural network for sentiment analysis. *Inf. Process. Manag.* **2019**, *56*, 1841–1844.
11. Wang, H.; He, J.; Zhang, X.; Liu, S. A Short Text Classification Method Based on N-Gram and CNN. *Chin. J. Electron.* **2020**, *29*, 248–254. [\[CrossRef\]](#)
12. Zhang, T.; You, F. Research on Short Text Classification Based on TextCNN. *J. Phys. Conf. Ser.* **2021**, *1757*, 12092. [\[CrossRef\]](#)
13. Li, J.; Feng, L.; Li, C. A novel deep learning approach for event extraction from social media. *J. Intell. Fuzzy Syst.* **2020**, *39*, 3183–3193. [\[CrossRef\]](#)
14. Fu, G.; Liu, Q.; Liu, W. A deep learning approach to rumor detection in social media. *J. Intell. Fuzzy Syst.* **2020**, *38*, 7429–7437.
15. Bueno, P.; Cano, P.A.; Goelen, F.; Hertog, T.; Vercnocke, B. Echoes of Kerr-like wormholes. *Phys. Comment D* **2018**, *97*, 24040. [\[CrossRef\]](#)
16. Gao, W.; Zheng, X.; Zhao, S. Named entity recognition method of Chinese EMR based on BERT-BiLSTM-CRF. *J. Phys. Conf. Ser.* **2021**, *1848*, 12083. [\[CrossRef\]](#)
17. Liu, Y.; Lu, J.; Yang, J.; Mao, F. Sentiment analysis for e-commerce product comments by deep learning model of Bert-BiGRU-Softmax. *Math. Biosci. Eng.* **2020**, *17*, 7819–7837. [\[CrossRef\]](#)
18. Zhuang, H.; Wang, F.; Bo, S.; Huang, Y. A BERT based Chinese Named Entity Recognition method on ASEAN News. *J. Phys. Conf. Ser.* **2021**, *1848*, 12101. [\[CrossRef\]](#)
19. Huang, G.; Wang, J.; Tang, H.; Ye, X. BERT-based Contextual Semantic analysis for English Preposition Error Correction. *J. Phys. Conf. Ser.* **2020**, *1693*, 012115. [\[CrossRef\]](#)
20. Dixon, F.M.; Polson, N.G.; Sokolov, V.O. Deep learning for spatio-temporal modeling: Dynamic traffic flows and high frequency trading. *Appl. Stoch. Model. Bus. Ind.* **2019**, *35*, 788–807. [\[CrossRef\]](#)
21. Nicolas, G.; Andrei, M.; Bradley, G.; Liu, A.; Ucci, G. Deep learning from 21-cm tomography of the cosmic dawn and reionization. *Mon. Not. R. Astron. Soc.* **2019**, *1*, 282–293.
22. Yi, A.; Li, Z.; Mi, G.; Zhang, Y.; Yu, D.; Chen, W.; Ju, Y. A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system. *Neural Comput. Appl.* **2019**, *31*, 1665–1677.
23. Wang, B.; Zhang, N.; Lu, W.; Geng, J.; Huang, X. Intelligent Missing Shots' Reconstruction Using the Spatial Reciprocity of Green's Function Based on Deep Learning. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 1587–1597. [\[CrossRef\]](#)
24. Dong, X.; Ba, T.L. Rapid analysis of coal characteristics based on deep learning and visible-infrared spectroscopy. *Microchem. J.* **2020**, *157*, 104880.
25. Sun, X.; Ning, Y.; Yang, D. Research on the Application of Deep Learning in Campus Security Monitoring System. *J. Phys. Conf. Ser.* **2021**, *1744*, 42035. [\[CrossRef\]](#)
26. Yao, L.; Zhao, H. Fuzzy Weighted Entropy Attention Deep Learning Method for Expression Recognition. *J. Phys. Conf. Ser.* **2021**, *1883*, 12130. [\[CrossRef\]](#)
27. Denisenko, A.I.; Krylovetsky, A.A.; Chernikov, I.S. Integral spin images usage in deep learning algorithms for 3D model classification. *J. Phys. Conf. Ser.* **2021**, *1902*, 12114. [\[CrossRef\]](#)
28. Wang, X.; Zeng, W.; Fei, P. A Deep-learning Based Computer Framework for Automatic Anatomical Segmentation of Mouse Brain. *J. Phys. Conf. Ser.* **2021**, *1815*, 12012. [\[CrossRef\]](#)
29. Chang, Z.; Lu, Z.; Mo, F.; Hao, L.; Yu, G. Design of workpiece recognition and sorting system based on deep learning. *J. Phys. Conf. Ser.* **2021**, *1802*, 22061. [\[CrossRef\]](#)

30. Li, N.; Cheng, Z.; Woo, Y.; Ha, J.; Kim, A. Research on the adverse reactions of medicines based on deep learning models. *J. Phys. Conf. Ser.* **2020**, *1629*, 12102. [[CrossRef](#)]
31. Liu, Y.; Liu, C.; Wang, L.; Chen, Z.; Wu, Y. Chinese Event Subject Extraction in the Financial Field Integrated with BIGRU and Multi-head Attention. *J. Phys. Conf. Ser.* **2021**, *1828*, 12032. [[CrossRef](#)]
32. Rani, M.S.; Subramanian, S. Attention Mechanism with Gated Recurrent Unit Using Convolutional Neural Network for Aspect Level Opinion Mining. *Arab. J. Sci. Eng.* **2020**, *45*, 6157–6169. [[CrossRef](#)]
33. Ren, F.; Jiang, Z.; Wang, X.; Liu, J. A DGA domain names detection modeling method based on integrating an attention mechanism and deep neural network. *Cybersecurity* **2020**, *3*, 4. [[CrossRef](#)]

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.