Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2023 SCPE. Volume 24, Issues 3, pp. 287–298, DOI 10.12694/scpe.v24i3.2371

DESIGN OF SENTIMENT ANALYSIS FRAMEWORK OF DIGITAL MEDIA SHORT TEXT BASED ON MULTI-PATTERN SENTIMENT LEXICON

SHUQIN LIN*

Abstract. Along the continuous advancement of the network and the rise of digital media, the amount of data produced by the exponential explosion. And how to use these data to provide personalized services for users is one of the current research focuses. To address the issue of insufficient coverage in the current sentiment lexicon and the difficulty of constructing sentiment lexicon in specific fields, this study proposes a multi-modal emotional thesaurus. Semi-supervised learning is used to solve the problem of insufficient coverage of emotional thesaurus, and a semi-supervised classification algorithm is realized by using a large number of unlabeled sample data combined with a small number of labeled sample data. Optimized learning is used to solve the problem of difficult construction of emotional thesaurus in specific fields, the corresponding specific emotional thesaurus is constructed by adaptive adjustment of emotional word score, and finally the improved emotional thesaurus is used to build a digital media short text sentiment analysis framework. For testing, the NLPCC dataset was used in this study, Experiments show that the framework constructed in this study requires 87 iterations, a Recall value of 0.912, a F1 value of 0.753, and an average accuracy of 83.39%, all of which are better than the sentiment analysis framework without the use of multi-pattern sentiment lexicon. In the simulation experiment, the recognition accuracy reached 85.88%, which was 16.85%, 11.57% and 6.72% higher than the test scenarios using a single emotion thesaurus selected in this study. The above results show that the digital media short-text sentiment analysis framework built in this research based on multi-pattern sentiment lexicon can carry out short-text sentiment analysis more accurately and efficiently, so as to accurately analyze users' needs and provide customized services precisely.

Key words: Sentiment lexicon; Sentiment analysis; Semi-supervised learning; Optimization learning

1. Introduction. As the growth of the Internet, it has changed our life with more social media entering our daily life and more people beginning to evaluate the current hot events on digital media as a daily activity [1]. The increase of network users leads to the proliferation of data, which can provide reference information, but because the user's expression is not formal, the data will show confusion. Therefore, how to extract effective information from these data is one of the key types of research [4]. Text sentiment analysis (SA) technology in natural language processing technology can effectively process data to judge users' emotions, attitudes and expression of views [16]. The most basic semantic unit of SA is emotion words, and the collection of emotion words and their emotion scores and labels is sentiment lexicon (SL), through which the emotion polarity or emotion intensity of text can be analyzed and calculated [15]. Text sentiment analysis technology has achieved outstanding results in theoretical research and practical application, but with the development of the Internet, the coverage of emotional words in some specific fields such as business, economics, film and television is insufficient, and the problem of difficulty in building emotional thesaurus in new fields has become increasingly prominent [12]. To solve these problems, this study puts forward a multi-patternSL. There are two innovative points in this study. First, semi-supervised learning (SSL) is adopted to solve the insufficient coverage of SL. The second is to adopt optimal learning to solve the difficult construction of SL in a specific field. The main structure of this study contains four parts. The first part is a summary of the research status in related fields; The second part is to construct the digital media short text analysis framework by combining various models; The third part is the effect verification analysis of the framework constructed in this study; The fourth part is the summary of this research.

2. Related works. As an important part of text SA, SL assigns polarity scores of text words through marking emotion scores of words, so as to judge the emotion polarity of text, which has extremely important significance for text SA. Zhou et al. introduced emotion information into human-computer dialogue and proposed a dynamic SL combining the dual replication mechanism, which was superior to other alternative methods in

*School of Arts & Communication, Xiamen Institute of Technology, Xiamen, 361021, China (lin2023126@126.com)

many aspects [20]. Wei et al. proposed a dynamic acoustic SL constructed from the acoustic lexical levels of different emotional categories to improve the ability of speech emotion recognition [17]. Navarrete et al. redesigned and constructed a SL, focusing on raising the SA in the text, enriching the emotional intensity of words, and helping to solve the problems of cyberbullying and violence in the digital society [13]. Shoujian et al. integrated existing emotion, degree, negative and network words to achieve effective SA on Weibo, and proposed a new method to construct SL, which improved the accuracy and recall rate of the method [14]. Mahadzir et al. solved the lack of emotion words in SA research in the context of Malaysian language by using a new polarity score allocation technique to construct a new Malay-English bilingual SL, ultimately improving the comprehensive performance of mixed language SL [11]. Zha et al. designed and constructed a large SL of Chinese ultra-short comments to solve the lack of large-scale and high-precision SL in Chinese book reviews. This construction method solved the issues caused by immature segmentation techniques and imperfect language models [19]. Lijo et al. put forward a rapid polarity detection method based on multiple lexical features, and built a SL with high expansivity and high polarity detection efficiency [9]. Garg et al. proposed an emotion classification method based on centrality, and created a SL named HindiEmotionNet for Hindi, achieving satisfactory results in emotion classification and recognition [7].

Text SA refers to the analysis of text information with emotion, so as to extract and summarize the emotional tendency in it. Short text SA is mainly aimed at the subjective short text on network digital media, such as Weibo, e-commerce platforms, etc., for emotion analysis and classification. Alwehaibi et al. used three deep learning topological models and proposed an emotion classification method for short texts in Arabic dialects based on deep learning to solve the difficulties caused by the excessively complex morphology and grammar of Arabic [2]. Kota et al. combined convolutional neural networks (CNNs), bidirectional long- and short-term memory and attention mechanisms to apply deep learning to SA, achieving remarkable results and excellent performance [8]. Barnes et al. proposed a multi-tasking method to incorporate negative information into SA and optimize the priority polarity of short texts [3]. Feng et al. put forward a multi-channel CNNSA model based on multi-attention mechanism in view of the relatively limited text features of short texts, which has higher classification accuracy and lower training time cost [6]. To solve the bullying and hate speech in the Internet, Chen et al. put forward a text classification model based on CNN, which can well handle short text tasks [5]. Luo et al. proposed a text SA method integrated with neural networks for the instability of single emotion classification model in classification, which significantly improved the accuracy of text SA and effectively predicted the emotional polarity of text [10]. To solve the problem that time series in emotion is not taken into account in traditional text emotion analysis, Zhao et al. proposed the method of fusion of window word vector and classifier in the decision-making level of short text emotion analysis, which had better performance [19].

From the above content that text SA, as one of the most active research fields in natural language processing, has an extremely broad prospect in information retrieval, data mining, intelligent recommendation and other aspects. With the development of the Internet, text SA plays a more prominent role. Text SA can reflect users' subjective views through text analysis in digital media. As the basic part of it, SL plays a crucial role. The initial SL is obtained by manual summary through experience. Although it will cost labor and time, its accuracy rate is high. However, there is often a problem of insufficient coverage of emotion words in common SL, and it is difficult to construct SL in some specific fields. To solve the above problems, this study proposed to use SSL and optimized learning mechanisms to construct multi-pattern SL and build short text SA model in digital media, which has strong positive significance in emotion analysis.

3. Digital Media Short Text SA Framework Based on Multi-pattern SL. As the speed growth of the Internet, digital media has quickly changed the way people live and work. As the continuous growth of the Internet, new words appear constantly, which makes the coverage of the traditional emotional thesaurus insufficient, and also makes the construction of SL in a specific field difficult. Therefore, this study adopts the idea of SSL and optimal learning to improve the SL, and then applies it to the short text emotion analysis.

3.1. A Digital Media Short Text SA Framework Based on Semi-supervised Learning SI. DUTIR is the largest Chinese SL available, among which, How Net SL and Dalian University of Technology Chinese SL ontology library (DUTIR) are common Chinese SL. However, with the development of the Internet, many new words have emerged. This has resulted in poor coverage of DUTIR sentiment words. Therefore, this

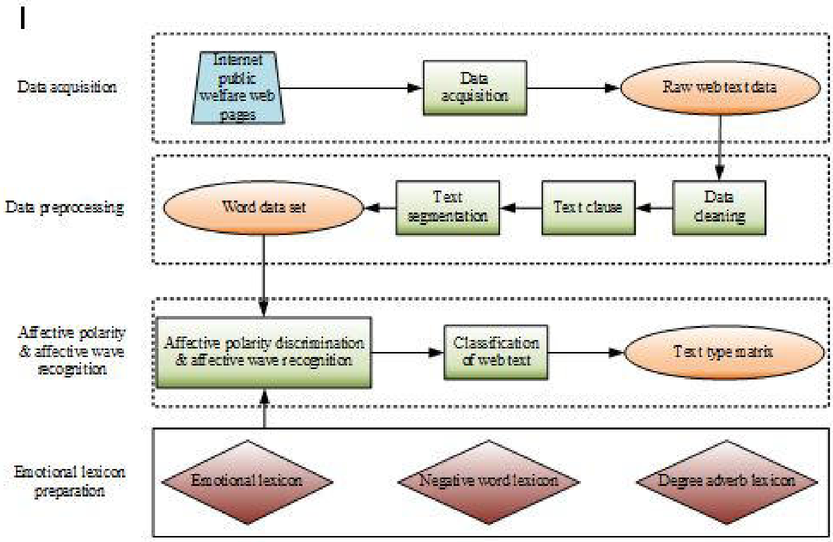


Fig. 3.1: Text SA Framework Based on SL

research applies the idea of SSL, extracts potential emotion words from the existing corpus of text data, identifies them, and expands DUTIR. The SA framework for short texts of digital media based on SL is shown in Figure 3.1.

Emotion polarity recognition is the most important part and its function is to calculate the emotion tendency of text. Text emotion recognition calculates the emotion score of each sentence in the text and then adds all the emotion score together to get the emotion score of the whole text. If the emotion score is greater than 0, it is regarded as positive emotion; if the emotion score is less than 0, it is regarded as negative emotion; if it is equal to 0, it is regarded as neutral emotion. Its calculation is shown in Equation 3.1.

$$V = \sum_{i=1}^{n} v_i \tag{3.1}$$

In Equation 3.1, V denotes the emotion score of the whole text;n indicates the number of sentences in the text;v_i denotes the emotion score value of each clause. By matching the emotion lexis with the text words after word segmentation, the initial score of the emotion words is finally obtained. Then, the inversion of negative words to emotion in the sentence is calculated, which is mainly obtained by statistical negative words in the sentence, as shown in Equation 3.2.

$$v_i = (-1)^k \cdot \sum_{j=1}^{m} w_j \cdot w_d \tag{3.2}$$

In Equation 3.2, v_i denotes the emotion score of each clause;K indicates the number of negative words in each sentence; W_j expresses the corresponding score of emotion words; W_d refers to the weight of degree adverb; m stands for the amount of emotion words in each sentence. The next step is to calculate the emotion wave of the text. The effect of emotion wave calculation is to analyze the degree of emotion change between each sentence. The variance is calculated by the emotion score of each sentence, and then the emotion score fluctuation is obtained. The calculation is shown in Equation 3.3.

$$S = \frac{1}{n} \cdot \sum_{i}^{n} (v_i - M)^2 \tag{3.3}$$

In Equation 3.3,S represents the variance;M means the mean value of the emotion score of each text. The greater the variance, the more obvious the emotion fluctuation. IG (Information Gain), as a feature selection method with obvious effect, is usually applied to measure the importance of words in the text. It is calculated for each word in the pre-processed corpus, and then a threshold is determined. According to the level of IG, words higher than the threshold are added to the candidate emotion lexical database. The calculation is shown in Equation 3.4.

$$IG\left(t_i\right) = Entropy\left(S\right) - ExpectedEntropy\left(S_{t_i}\right) \tag{3.4}$$

In Equation 3.4, $Entropy(S)$ means the actual entropy and $ExpectedEntropy(\S_t)$ denotes the desired entropy. The specific calculation of is shown $Entropy(S)$ in Equation 3.5.

$$Entropy\left(S\right) = -\sum_{j=1}^{T} P\left(C_j\right) \times \log P\left(C_j\right) \tag{3.5}$$

In Equation 3.5, $P(c_j)$ represents the probability of C_j class documents appearing and T denotes the total number of texts. $ExpectedEntropy(S_{t_i})$ is calculated by Equation 3.6.

$$\begin{aligned} \text{Expected\_Entropy}\left(S_{t_i}\right) =& P\left(t_i\right) \times \left[-\sum_{j=1}^{T} P\left(C_j \left| t_i\right.\right) \times \log P\left(C_j \left| t_i\right.\right)\right] \\ &+ P(t_i) \times \left[-\sum_{j=1}^{T} P(C_j|\overline{t_i}) \times \log P(C_j|\overline{t_i})\right] \end{aligned} \tag{3.6}$$

In Equation 3.6, $P(C_j)$ represents the probability that C_j class documents appear; $P(t_i)$represents the probability that word documents are included; $P(C_i|t_i)$ represents the probability that they belong to the type C_j if they contain words t_i ; $P(\overline{t}_i)$ refers to the probability that documents do not contain words t_i; $P(C_j|\overline{t}_i)$ represents the probability that they belong to the type C_j if they do not contain words t_i; T represents the total number of texts. The SSL adopted in this study is essentially the combination of supervised learning and unsupervised learning. It realizes the semi-supervised classification algorithm by combining a large number of unlabeled sample data with a small number of labeled sample data. In this study, the labelled samples P and unlabeled samples U will be defined to identify the words belonging to them from U through the SSL method. The main steps are shown in Figure 3.2.

In SSLto enhance the performance of classification and further improve it, this study uses multi-layer perceptron for classification. The input layer is the representation of the word vector of a single word. The word vector is obtained by Word2Vec training, and its vector dimension is 200. Finally, the probability of a single word belonging to class +1 is calculated by Sigmod activation function. The structure of the multi-layer perceptron is shown in Figure 3.3.

3.2. The Framework of Digital Media Short Text SA Based on the Optimized Learning SL. Digital media has changed people's way of working and living. Users can express their attitudes and views on current events, products and life at any time. The amount of data generated by digital media every day shows a blowout trend, and SA can extract users' views and opinions from it. Since there are different scenarios for SA, it means that some emotional intensity and even emotional polarity will change in different fields. Therefore, the problem of the current SL is that the universality is poor, and the general emotion word database fails to contain the domain specific emotion words. The construction of the usual domain SL is shown in Figure 3.4.

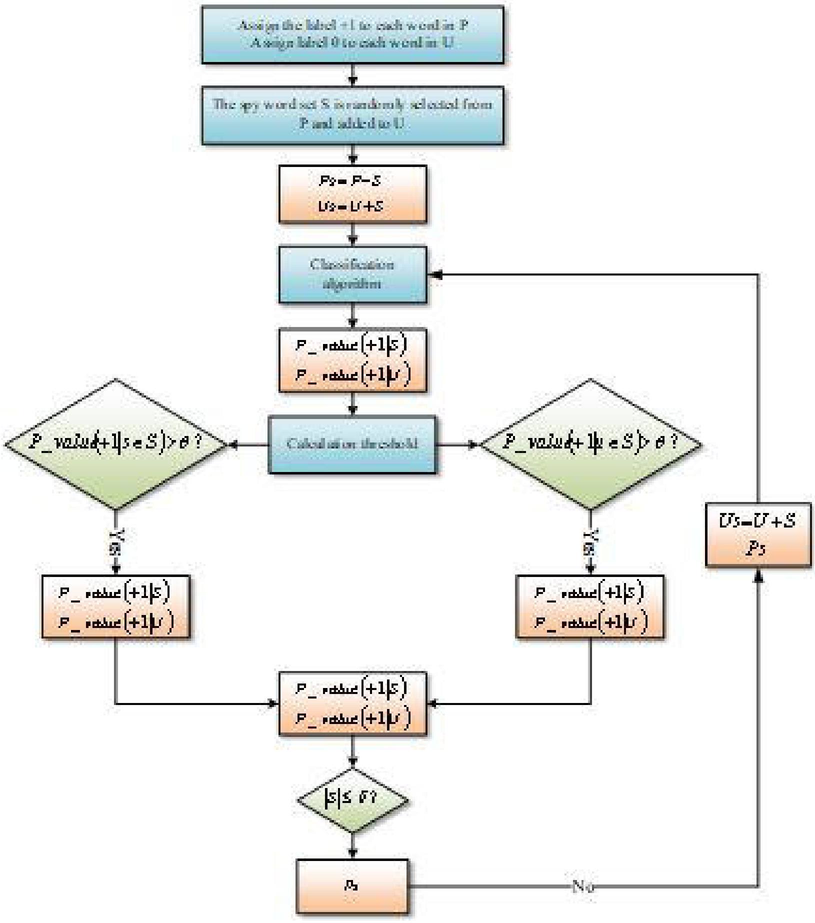


Fig. 3.2: Emotion Words Recognition Based on Semi-Supervised Learning

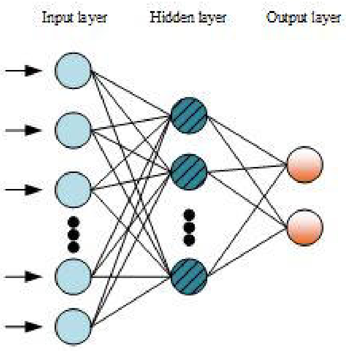


Fig. 3.3: Multilayer Augmented Perceptron

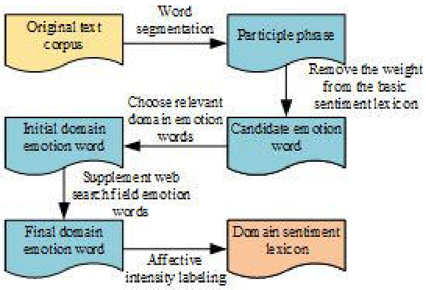


Fig. 3.4: Construction of Domain SL

Therefore, to solve the above shortcomings, aiming at the construction of SL in a specific field, this study puts forward an optimization learning method, which adaptively adjusts the score of emotion words to build corresponding specific SL. Firstly, the emotion polarity in the text is calculated, the emotion words in the corpus are matched, and the emotion score of each emotion word is extracted in turn. After adding the emotion score of all emotion words in the text, the emotion polarity of the text is judged according to the total score. The calculation for the emotion polarity of the text is shown in Equation 3.7.

$$P = \sum_{w=1}^{n} V_w \tag{3.7}$$

In Equation 3.7,P means the emotional score of the text;V_w denotes the emotional score of the $W_t h$ emotional word,n and represents the total amount of emotional words in the text. If $p > 0$, it indicates that the text is a positive text; If $P < 0$, it means the text is negative. Then, it is necessary to determine the accuracy rate of emotion classification. By comparing the artificial label with the calculated emotional polarity of the text, the accuracy rate of emotion classification of the text by emotion lexicon can be calculated, as shown in Equation 3.8.

$$Accuracy = \frac{C}{N} \tag{3.8}$$

In Equation 3.8, C means the number of texts whose emotional labels are consistent with those manually marked in the emotional lexicon. N is the amount of all texts in the corpus. Finally, the score of emotion words in the SL is adjusted in a random way to improve the accuracy of emotion classification, and the effectiveness of the adjustment scheme is verified by verifying whether the accuracy of classification is increased. The optimization learning flow chart in this study is shown in Figure 3.5.

The fitness value is designed according to the accuracy of text classification by SL. For all the texts in the corpus, if the SL correctly classifies a text, a reward value 1 is added to the fitness value of the SL. If no text is correctly classified, a penalty value is subtracted from that. Its specific fitness calculation expression is shown in Equation 3.9.

$$Fitness\,(k, D) = \sum_{i=1}^{n} R\,(k, D, T_i) \tag{3.9}$$

Equation 3.9 shows the fitness calculation of individual K in training set D and represents $R(K, D, T)$ the accurate measurement value of classification when individual K is used to classify the ith text T_i in training set D. Whether the text has been correctly classified can be determined by Equation 3.7. The specific classification accuracy measurement value calculation is shown in Equation 3.10.

$$R(k, D, T) = \begin{cases} 1, \text{if Accurately Predicted} \\ -\frac{|P|}{\omega}, \text{if not Accurately Predicted} \end{cases} \tag{3.10}$$

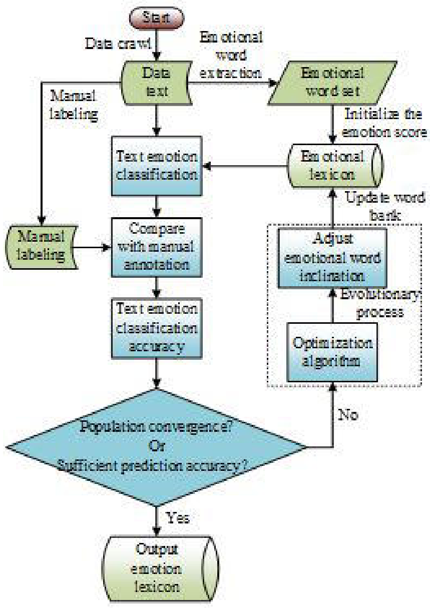


Fig. 3.5: Optimization Learning

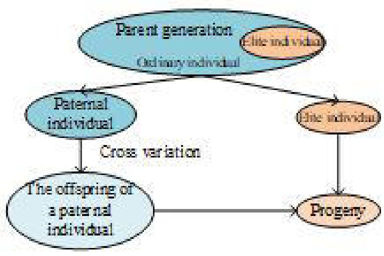


Fig. 3.6: Optimal Strategy

In Equation 3.10 shows the classification by individual, where K denotes the training set, represents the th article text, and means the penalty parameter. In traditional genetic algorithms, there are usually problems of poor learning efficiency and poor training effect. Therefore, in order to further improve the process of emotional thesaurus construction, the optimization strategy is introduced to improve the efficiency of optimal learning. Two methods of combining optimal strategies are adopted. The first method is to directly copy the excellent individuals of the previous generation into the next generation; the second method is to select some of the better individuals as the elite set and then use the individuals and ordinary individuals to generate new individuals through cross mutation. The optimal preservation strategy designed in this study is shown in Figure 3.6.

In the text emotion analysis, the emotion polarity of the text is closely related to the polarity of the emotion words in the text. By analyzing the use of emotion words, the emotion expressed by the text can be judged. Generally speaking, if a certain emotion word often appears in positive text with a high probability, then the emotion word has a high probability of being positive emotion word; Otherwise, it is a negative emotion word.

Therefore, to better guide the evolution of the score of emotion words and make the emotional polarity of text more closely related to that of emotion words, a new variation strategy is designed based on Sigmoid function [18]. The Sigmoid function is shown in Equation 3.11.

$$S\left(x\right)=\frac{1}{1+e^{-x}} \tag{3.11}$$

Sigmoid function is often used in the output of neurons in the hidden layer, and its value range is $(0,1)$.It can map a real number to the interval $(0,1)$, so as to carry out binary classification. It has the advantages of smooth and easy derivation. Equation 3.12 can be obtained by solving the inverse function of Equation 3.10.

$$f\left(x\right)=-ln\left(\frac{1}{x}-1\right) \tag{3.12}$$

Let the emotion word be, the data set be, and the probability of the emotion word appearing in the positive text of the data set be, plus the range of the emotion score in this paper be. Therefore, based on Equation 3.12, the variation guidance function is constructed, as shown in Equation 3.13.

$$f\left(P_w\right)=-\frac{5}{3}\times ln\left(\frac{1}{P_w}-1\right),P_w\in[0,1] \tag{3.13}$$

Finally, combined with Equation 3.12, a new probability variation strategy is constructed, as shown in Equation 3.14.

$$V_w=R+f\left(P_w\right) \tag{3.14}$$

In Equation 3.14, R represents the random number, V_w indicates the emotion score value of the emotion word after variation. Equation 3.15 is the restriction of Equation 3.14.

$$\begin{cases}R\in[-10,10]\\V_w\in[-10,10]\end{cases} \tag{3.15}$$

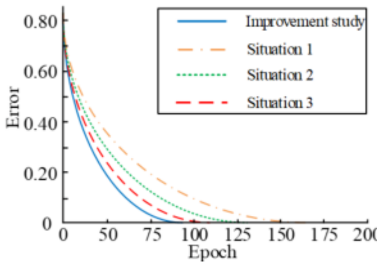
Equation 3.15 limits R value range to $[-10,10]$. In this formula, if the value of V_w exceeds $[-10,10]$, then correspondingly, it takes the co responding boundary value -10 or 10.

4. Evaluation of Short Text SA Framework of Digital Media Based on Multi-Pattern SL. To verify the validity of the framework constructed in this study, eight widely used Chinese SA data sets were selected for experiments. The four data sets were microblog SA data sets released by NLPCC, which were typical short text data sets of digital media. The other four data sets were Sina News RSS subscription channel data set (THUC News), National news data (Sogou CA), Sohu news data (Sougou CS) and a food delivery platform data (Waimai_10k). The details of the number of positive and negative text in the experimental data set are shown in Table 4.1.

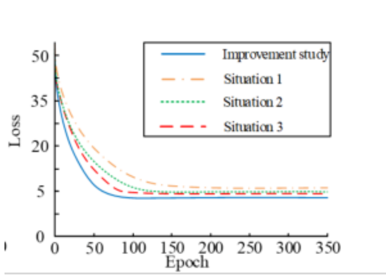
In the experiment, 80% of the data set was randomly selected to reconstruct the positive and negative sample sets, and the remaining 20 samples were utilized as test samples. To compare the effectiveness of the improved multi-pattern SL in this study, the following three scenarios were selected to compare with the framework constructed in this study: In scenario 1, the DUTIR was selected, where the sum of scores of emotion words was used as the classification criteria; In scenario 2, DUTIR and network word database (NWD) were selected, and the sum of scores of emotion words was used as the classification criteria; In scenario 3, DUTIR, NWD, and emoticon database (ED) were chosen, and the sum of values of emotion words are used as the classification standard. Firstly, data sets were used for training, and SA frameworks using different lexicon were compared. The results are shown in Figure 4.1. Where, the improved SA framework based on multi-pattern SL in this study could reach the best state after 87 training sessions, which was 65, 39 and 13 times less than scenarios 1, 2 and 3, respectively. This indicated that the convergence of the framework established in this study was better than that of other frameworks using the unimproved lexicon.

Table 4.1: Number of Positive and Negative Comments in the NLPCC Data Set

	NLPCC2013	**NLPCC2014**	**NLPCC2018**	**NLPCC2020**
Positive comments	2886	2460	1049	1214
Negative comments	2047	2728	851	1158
Total comments	4960	5188	1900	2372
	THUCNews	**SogouCA**	**SougouCS**	**Waimai_10k**
Positive comments	6715	8977	9457	4688
Negative comments	8946	7655	7546	8154
Total comments	15661	16632	17003	12842



(a) Error



(b) Load

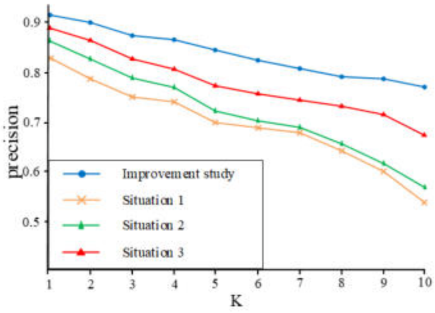
Fig. 4.1: Convergence of Four Situations

Using the selected data set, the four frameworks were tested and the accuracy rate and recall rate were obtained, as shown in Figure 4.2. Where, under the condition that the value of parameter K ranged from 1 to 10, with the increase of parameter K, the recall rate of the SA framework established in this study based on multi-pattern sentiment lexical database increased significantly compared with other frameworks, and the value of recall reached 0.912. It was 0.102, 0.154 and 0.218 higher than scenarios 1, 2 and 3, respectively, while the accuracy rate maintained a slower fading speed compared with other models.

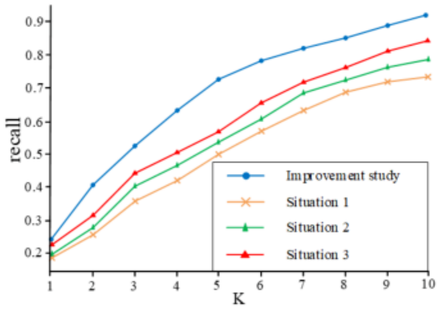
After each frame was trained to the best state, the test data was input for testing. Each frame as tested for 10 times. To avoid errors, the average value was taken as the final result, and the F1 value and AUC value of each model were compared. As shown in Figure 4.3, the value of the digital media short text SA framework established in this study based on mutil-pattern SL reached 0.753, which was higher than that of scenarios 1,2 and 3 by 0.091, 0.052 and 0.032 respectively. Its AUC value also had the best performance. The results showed that in digital media short text SA, the improved lexicon proposed in this study had the best performance.

In NLPCC and the other four data sets, the recognition accuracy of the framework established in this study and scenarios 1, 2 and 3 was compared, and the average accuracy results were shown in Figure 4.4. Where, the SA framework established in this study based on the mutil-pattern SL had an outstanding performance in the direction of accuracy. And its average accuracy in NLPCC2013,NLPCC2014,NLPCC2018 and NLPCC2020 was 85.88%, 77.28%, 80.26% and 90.15%, respectively. Its overall average accuracy was 83.39%, which was 16.85%, 11.57%, and 6.72% higher than Scenarios 1, 2, and 3, respectively. Table 4.2 shows the average accuracy, positive text accuracy and negative text accuracy of NLPCC data set in this experiment.

From Table 4.2, the overall accuracy of scenario 1 on NLPCC2013, NLPCC2014, NLPCC2018, NLPCC2020 was 68.83%, 60.73%, 66.25% and 70.36%, respectively. After the combination of the network lexical database, the accuracy of scenario 2 in NLPCC2013, NLPCC2014, NLPCC2018 and NLPCC2020 was 75.95%, 64.80%,

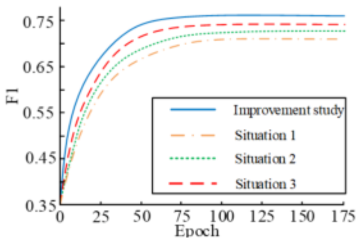


(a) Precision

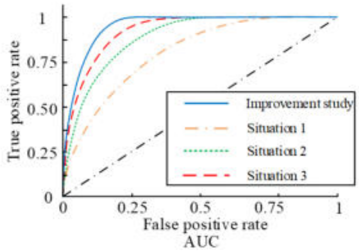


(b) Recall

Fig. 4.2: Precision and Recall of Four Situations



(a) F1

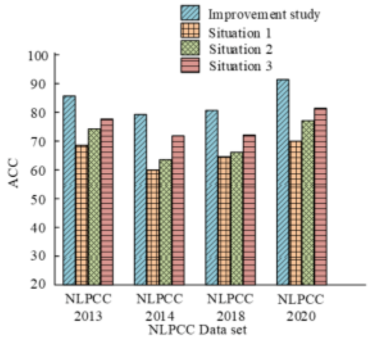


(b) AUC

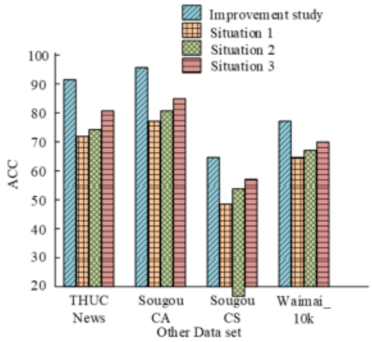
Fig. 4.3: F1 and AUC of Four Situations

68.94% and 77.59%, which was indeed improved compared with scenario 1. When the emoticon lexicon was added into scenario 3, its accuracy was 78.85%, 72.49%, 72.53% and 82.82%, which was better than scenario 1 and scenario 2. It was suggested that network lexicon and emoticons should be considered in short text SA of digital media short text. The average accuracy of the improved SA framework based on multi-pattern SL in NLPCC2013, NLPCC2014, NLPCC2018 and NLPCC2020 was 85.88%, 77.28%, 80.26% and 90.15%, respectively, which was ahead of the three scenarios. Its recognition accuracy in positive text was 87.49%, 79.08%, 81.94% and 92.76%, which were all higher than the other three scenarios. Its recognition accuracy in negative text was 84.27%, 75.48%, 78.58% and 87.54%, which was also the highest. To sum up, after comparison with all other schemes, it showed that the framework of digital media short text SA based on multi-pattern SL proposed in this study had the best performance, and it was also the best in the classification of positive and negative texts.

5. Conclusion. In the field of text emotion recognition, there were problems such as insufficient coverage of emotion words and difficult construction. Therefore, a multi-pattern SL was proposed. SSL mode was adopted to solve the insufficient coverage of emotion words, and optimization learning was combined to solve the difficulty of constructing SL. The experimental results showed that the improved framework in this study needed 87 iterations to achieve the target accuracy and the optimal Loss value, which was 65, 39 and 13 times less than scenarios 1, 2 and 3, respectively. The Recall value of the improved frame in this study reached 0.912, which was 0.102, 0.154 and 0.218 higher than the other three scenarios, respectively. Its F1 value reached 0.753, which was higher than the three scenarios by 0.091, 0.052 and 0.032, respectively. The average accuracy of



(a) NLPCC Dataset ACC



(b) Other Dataset ACC

Fig. 4.4: Accuracy Mean of Four Situations

Table 4.2: Accuracy on the NLPCC Data Set

Data Set	**Accuracy (%)**	**Average accuracy**	**Positive comments**
NLPCC2013	-	-	-
Improvement study	85.88	87.49	84.27
Situation 1	68.83	70.12	67.54
Situation 2	75.95	75.41	76.49
Situation 3	78.85	79.24	78.45
NLPCC2014	-	-	-
Improvement study	77.28	79.08	75.48
Situation 1	60.73	61.79	59.67
Situation 2	64.80	65.87	63.72
Situation 3	72.49	73.57	71.42
NLPCC2018	-	-	-
Improvement study	80.26	81.94	78.58
Situation 1	66.25	67.25	65.24
Situation 2	68.94	69.46	68.41
Situation 3	72.53	71.48	70.57
NLPCC2020	-	-	-
Improvement study	90.15	92.76	87.54
Situation 1	70.36	71.24	69.47
Situation 2	77.59	79.86	75.31
Situation 3	82.82	84.21	81.42

the four data sets was 85.88%, 77.28%, 80.26% and 90.15% respectively, and the average accuracy was 83.39%, which was 16.85%, 11.57% and 6.72% higher than the three scenarios respectively. The accuracy of positive text in several data sets was 87.49%, 79.08%, 81.94% and 92.76%, respectively, which showed the best performance. Its accuracy in negative text was 84.27%, 75.48%, 78.58% and 87.54%, respectively, which showed the best performance. The above results showed that the framework of digital media short text SA based on mutil-pattern SL established in this study could carry out digital media short text analysis accurately and efficiently, and had positive significance for the development of text SA field. However, in this study, the optimization of emotional words themselves is not directly considered, and the genetic algorithm used in this study will inevitably slow down with the increase of time and the increase of data volume, and subsequent studies can consider changing better algorithms and further optimizing emotional words.

REFERENCES

[1] S. Velichety, S. Ram, *Finding a Needle in the Haystack, Recommending Online Communities on Social Media Platforms Using Network and Design Science*, Journal Of The Association Of Information Systems, 22 (2021) pp. 1285-1310

[2] A. Alwehaibi, M. Bikdash, M. Albogmi, and K. A. Roy, *Study of the performance of embedding methods for arabic short-text sentiment analysis using deep learning approaches*, Journal of King Saud University - Computer and Information Sciences, 34, p. 201.

[3] J. Barnes, E. Velldal, and L. Vrelid, *Improving sentiment analysis with multi-task learning of negation*, Natural Language Engineering, 27 (2020), pp. 249–269.

[4] K. Chan and U. M. D. media consumption: Using metrics, *patterns and dashboards to enhance data-driven decision-making*, Journal of Consumer Behaviour, 21 (2022), pp. 80–91.

[5] K. C. Chen, *Verbal aggression detection on twitter comments: convolutional neural network for short-text sentiment analysis*, Neural computing & applications, 32 (2020), pp. 10809–10818.

[6] Y. Feng and Y. Cheng, *Short text sentiment analysis based on multi-channel cnn with multi-head attention mechanism*, IEEE Access, 9 (2021), pp. 19854–19863.

[7] K. Garg and A. Lobiyal D. K. Hindi EmotionNet:, *Scalable emotion lexicon for sentiment classification of hindi text*, ACM transactions on Asian and low-resource language information processing, 19 (2020), pp. 2–35.

[8] V. R. Kota and S. D. Munisamy, *High accuracy offering attention mechanisms based deep learning approach using cnn/bi-lstm for sentiment analysis*, International Journal of Intelligent Computing and Cybernetics, 15 (2022), pp. 61–74.

[9] V. P. Lijo and H. Seetha, *Tweets sentiment analysis using multi-lexicon features and smo*, International Journal of Embedded Systems, 14 (2021), pp. 476–485.

[10] S. Luo, Y. Gu, X. Yao, and W. Fan, *Research on text sentiment analysis based on neural network and ensemble learning*, Rev. D'Intelligence Artif, 35 (2021), pp. 63–70.

[11] N. H. Mahadzir, M. F. Omar, M. Nawi, A. A. Salameh, K. C. Hussin, and S. A. MELex, *The construction of malay-english sentiment lexicon*, Computers, Materials, & Continua, 4 (2022), pp. 1789–1805.

[12] M. Naik, D. Vasumathi, and A. P. A. Kumar, *Novel approach for extraction of distinguishing emotions for semantic granularity level sentiment analysis in multilingual context*, Recent advances in computer science and communications, 15 (2022), pp. 77–87.

[13] A. S. Navarrete, C. Martinez-Araneda, C. Vidal-Castro, and C. A. Rubio-Manzano, *novel approach to the creation of a labelling lexicon for improving emotion analysis in text*, The Electronic Library, 39, p. 201.

[14] Y. U. Shoujian, B. Wang, and L. U. Ting, *Sentiment lexicon construction based on improved left-right entropy algorithm*, Journal of Donghua University, 39 (2022), pp. 65–71.

[15] J. Tang, Y. Xue, Z. Wang, S. Hu, T. Gong, Y. Chen, H. Zhao, and L. Xiao, *Bayesian estimation-based sentiment word embedding model for sentiment analysis*, CAAI Transactions on Intelligence Technology, 7 (2022), pp. 144–155.

[16] Z. Wang, P. Gao, and X. Chu, *Sentiment analysis from customer-generated online videos on product review using topic modeling and multi-attention blstm*, Advanced Engineering Informatics, 52 (2022), pp. 1588–10159.

[17] W. Wei, X. Cao, H. Li, L. Shen, and W. PA., *Improving speech emotion recognition based on acoustic words emotion dictionary*, Natural Language Engineering, 27 (2020), pp. 747–761.

[18] M. Yang, *Research on vehicle automatic driving target perception technology based on improved msrpn algorithm*, Journal of Computational and Cognitive Engineering, 1 (2022), pp. 147–151.

[19] M. Zha, C. Hu, and Y. Shi, *Sentiment lexicon construction for chinese book reviews based on ultrashort reviews*, The Electronic Library: The International Journal for Minicomputer, Microcomputer, and Software Applications in Libraries, 40 (2022), pp. 221–236.

[20] Q. Zhou, D. Ji, Y. Ren, T. H. D. copying Mechanism, and D. Emotion, *Dictionary for generating emotional responses*, Neurocomputing, 454 (2021), pp. 303–312.

Edited by: Rajni Mohana
Special Issue on: Sentiment Analysis and Affective computing in Multimedia Data on Social Network
Received: Jun 26, 2023
Accepted: Jul 24, 2023