Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

Emotional Transfer Framework of Short Texts in Chinese

Ningjing Gong[1*], Zhiyuan Gong[2]
[1]Department of Information Technology, Hubei University of Police, Wuhan, 430034 China
[2]Network Information Center, Wuhan University, Wuhan, 430073 China
*Corresponding Author.

Abstract:

The emotion analysis technology based on deep learning is widely studied and applied. However, its dependence on large data set makes this technology unable to play a role in small domains with only light data sets. To break the application barrier of this technology, we combine the knowledge in deep learning and cross-domain transfer and put forward the emotional transfer framework of short texts in Chinese. The framework takes the task domain as the target domain and selects the one that has sufficient text data and semantic feature intersection as the source domain. It extracts semantic features from the source domain and transfers them to the target domain to make up for the lack of sufficient data in the target domain. It helps the text emotion analysis task work effectively when there are less data in the target domain. This paper introduces the principle of the framework from four modules: the data preprocessing module, the semantic feature transfer module, the training module for the words order model, and the monitoring module for model innovation. Finally, the experimental results show that The framework is feasible, the model effect is good, and the accuracy of emotion classification is 89.0%. It is better than other methods in recent years.

Keywords*: Deep learning, Emotion transfer, Short texts, Text emotion analysis.*

I. INTRODUCTION

Text emotion analysis is used to extract emotional information such as users' views, emotions, evaluations, attitudes from text data. It is the most effective technology, which can support product recommendation, customer management, and evaluation analysis. It can help businesses to respond quickly and grasp business opportunities in time. When the amount of text data reaches a great scale, deep learning technology can be used to build a model for the text emotion analysis task, and complete the task rapidly and effectively with an algorithm instead of manuals. However, this kind of modeling by deep learning is highly dependent on large data sets and is always used in mature domains that have sufficient data accumulation. What can we do, when facing the demands of text emotion analysis in a new or immature domain? A new or immature domain doesn't have sufficient data accumulation. We all know the model could not get correct and real semantic features of the text data from a special domain when modeling with poor and lite data sets. I try to solve this problem by using the idea of transfer learning by crossing

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

domains. I call it an emotion transfer method. Here comes out another problem. There are many differences in speaking and writing representation between Chinese and English. So the emotion transfer method with Chinese could not copy the one with English directly. Furthermore, there are many differences in semantic features distribution in different domains under Chinese text data. So there was no general formula for the emotion transfer method with Chinese text. In many small areas (such as a new user feedback system launched recently), users have little enthusiasm for feedback, the amount of text data from feedback is small, and one feedback text always has only a few words or a sentence. It is a great challenge and also a good chance to find a feasible emotion transfer method dealing with this kind of lite and short Chinese text data. So I select a small data set that comes from feedback data in a catering takeout system to simulate feedback text data from a new platform for order and delivery for meals. (In recent years, more and more online platforms like it appeared in China. With these platforms, customers just need to open a special app installed on mobile at home, choose the restaurant they liked and order from the menu and pay online. About twenty minutes later, they can enjoin delicious foods at home.) In this scenario, I studied and proposed the emotional transfer framework of short texts in Chinese. Lately, I proved its effect through experiments later.

The remainder of this paper is organized as follows: in section II, several existing related works on emotion transfer methods are introduced. In section III, the emotional transfer framework of short texts in Chinese is described, which includes the theories it used, its structure, and the modeling process. Section IV gives experiment details, results, and analysis of results.

II. RELATED WORKS

At present, there are many emotion transfer methods at home and abroad. From the perspective of the target domain, there are direct emotional transfer, inductive emotional transfer, single-source cross-domain emotional classification, multi-source cross-domain emotional classification, and so on. From the perspective of transfer strategy, there are instance transfer, feature transfer, and model transfer. Later, new transfer strategies have evolved, such as transfer methods based on graph models, dictionaries, and joint emotional topics [1]. The representative achievements of the above methods include the text emotion analysis method realized by Remus R. [2] by selecting samples similar to the target domain from the source domain training set. Bolegala [3] established training tasks from three aspects: shared features, marked data in the source domain, and original data in the target domain to solve the training of cross-domain models. This method is discussed and studied in the context of English. Xia [4] realizes cross-domain text emotion analysis through the idea of feature integration and sample integration. Blitzer [5] proposed a text emotion analysis method of structured related learning.

III. EMOTION TRANSFER FRAMEWORK

3.1 Problem

There are many problems in using deep learning for emotional transfer dealing with Chinese short text. First, Chinese and English languages and text representation are different. In English, a word is a word.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

Words in a sentence are separated by spaces. But all words in a Chinese sentence are close together without any separation. And words in Chinese can be composed of different numbers of Chinese characters, sometimes one Chinese character can be a word. If you do not understand Chinese, you can't even tell which Chinese characters next to each other in a Chinese sentence compose a word and how many words there are in the sentence. Fig 1 shows an English sentence and a Chinese sentence with the same meaning. The English sentence has 41 characters (including space characters) and 9 words. The Chinese sentence has 12 Chinese characters. (Use Chinese Pinyin instead of Chinese characters, and separate the Pinyin of each Chinese character with additional spaces). Those characters are pieced together into eight words. Every word is marked with an underline. If no these marks, foreigners do not know which Chinese characters are next to each other in the sentence composing a word. So Chinese text analysis is more difficult than English text analysis.

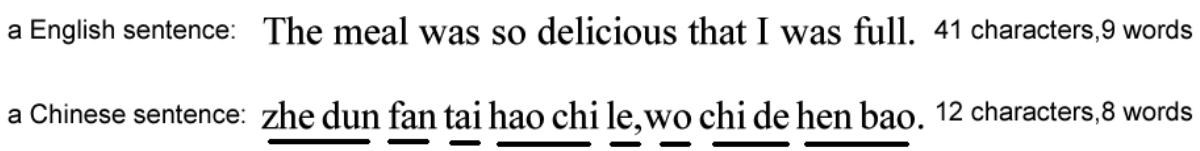

Fig 1: An English sentence and a Chinese sentence with the same meaning

Second, there are more than 90,000 Chinese characters in China, and about 70,000 are commonly used. Every Chinese character has its unique meaning. Many characters have multiple pronunciations, and each pronunciation corresponds to a different meaning. Sometimes the same pronunciation corresponds to different meanings in different words and different contexts. Therefore, there are great differences in the features distribution of Chinese text in different domains. It is difficult to ensure the applicable effectiveness of the model that is trained and has caught semantic features in one domain and transferred them to another domain. Third, to help the model catch the semantic features and judge the emotional characteristics of a text record, the longer the record length, the more information, and details be sent to the model, and the algorithm is easier to learn, the model can catch more real semantic features. When the record length is too short, (such as a record has only one sentence including a few words.) the model can catch little details for judgment from only a few words. Coupled with the small scale of the data set, it is very difficult to analyze the text emotion, and it is difficult to achieve the expected effect.

To solve these problems, I propose the emotional transfer framework of short texts in Chinese which is described in Fig 2. The framework includes four parts: the data preprocessing module, the semantic feature transfer module, the training module for the words order model, and the monitoring module for model innovation.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

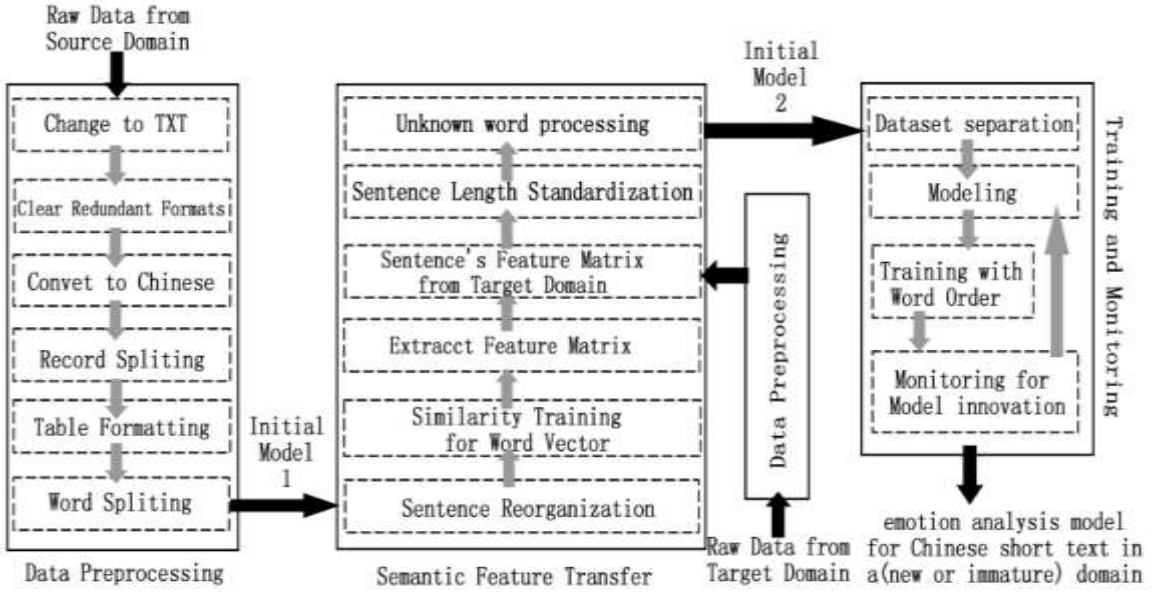

Fig 2: The emotional transfer framework of short texts in Chinese

3.2 Source Domain Selecting and Data Preprocessing

The transfer on emotion transfer methods is catching emotional semantic features from the source domain and applying them to the target domain. So an important thing is to select a proper source domain for the target domain that your emotion analysis task is in. A suitable source domain must meet the following conditions. The first is that the source domain must have sufficient data accumulation. I mentioned earlier that the target domain is a new or immature domain with poor or lite data. It can't meet the need to build a good model. The key point is that model can't catch real semantic features without enough training using sufficient text data. So I need a large source domain. It is easy to get semantic features of the source domain supporting with sufficient data. The second is that the source domain's semantic feature must be applied to the target domain. The source domain's semantic feature must be able to contain most parts of the semantic features of the target domain. Although the semantic feature distributions of the two domains are different, the transfer model's parameters will retain and carry forward the same semantic features in the two domains and punish those different ones from the target domain in the later model training.

In the data preprocessing module, firstly, it is required to extract the original data packets that come from the source domain and put the data into TXT files. The format of characters in those files must be changed to UTF-8. After that, clean data in those files and ensure that all Chinese characters' coding formats in those files are the same. For example, we can convert those coding formats to GB2312 for all characters.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

Secondly, in the step of data cleaning, other symbols or formats irrelevant to data records in the text need to be cleaned out. To split the whole data by every record easily in the next step, the space character and necessary tab characters in the text need to be retained. Different original data packets will have different structures. Therefore, we need to preview some of the original data and use the specific detailed operation to complete the above steps. Since some data would come from a database that is in Taiwan or Hong Kong, we need to check whether the characters in those files are traditional Chinese or simplified Chinese. It is necessary to convert the overall Chinese characters to simplified Chinese coding. Then we can get the pure Chinese text data.

Thirdly, to ensure that the pure Chinese text data will be organized into the training dataset form that can be understood by the algorithm, data will be split by every record using the separation characters that are retained in the upper steps. Furthermore, Words in Chinese sentences are close together without any separation. The algorithm cannot run successfully dealing with original Chinese sentences. Every record will be split by every Chinese word. The technology of Chinese word segmentation will be used in word splitting. At present, there is no unified word segmentation standard. It always needs to formulate different word segmentation standards according to different needs. Ambiguity recognition and new word recognition are two major problems that have not been completely broken through in the process of Chinese word segmentation. However, these have little impact on the function of the emotional transfer framework of short texts in Chinese, if the data in the source domain and the target domain adopt the same operations in word segmentation.

The data that comes from the target domain also need data preprocessing and be cleaned to pure data.

3.3 Semantic Feature Transfer

The target domain is the domain that the task of text emotion analysis is in. So the training text data used to model for the task is also in the target domain. The target domain is new and immature. The training data is lite and insufficient, is not enough to train an effective model unless the model already gets the semantic feature of the target domain, it can hold the relation and correspond between words in training text data well. How do we get the semantic feature of the target domain? We cannot get the semantic feature from the target domain directly. It has insufficient data. So the framework chooses a large source domain for the target domain. It has sufficient data and its semantic feature can contain most parts of the semantic features of the target domain. I call the data from the source domain corpus. The framework will extract the semantic feature from the corpus and transfer it to the target domain. It can help text data in the target domain to train an effective model.

3.3.1 Semantic feature extracted from the source domain

The framework chooses a One-Hot encoding representation to identify each word uniquely in the corpus. A dictionary must be established first. The statistical method is used to count the divided words in the corpus, and all non-repeated words are extracted and made into a dictionary. Then every word in the

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

corpus can be found in the dictionary. Words' One-Hot codes related to the dictionary are described in Fig 3. Assuming that the number of words in the established dictionary is n and the index number of "apple" in the dictionary is i, the One-Hot code of "apple" is an n-dimensional vector. Except that the ith component is 1, all other components are 0. Similarly, the One-Hot vector of any word in the corpus can be described. Given a word is x and its index number is i, the number of words in the dictionary is n, then the One-Hot vector of x has n components, which can be expressed as x_1 to x_n respectively. Except that the component x_i corresponding to the index number of x in the dictionary is 1, all other components are 0.

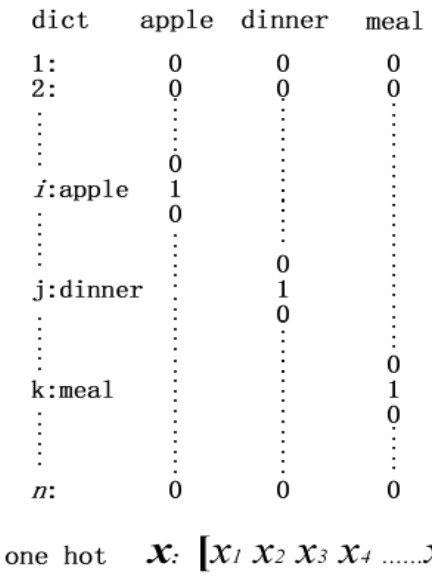

Fig 3: One-Hot encoding related to the dictionary

The semantic features extracted from the corpus are finally reflected in the association information between words. The correlation degree between two words is measured by Eqs (1).

$$\cos\theta = \frac{\vec{x}\cdot\vec{y}}{|\vec{x}|*|\vec{y}|} \tag{1}$$

Where $\vec{x}$ and $\vec{y}$ are vectors of two words. $\cos\theta$ is the cosine similarity of them. It is the inner product of two vectors divided by the modular product of two vectors. If the value of $\cos\theta$ is 1, it indicates that the two words are positively correlated; A value of 0 indicates that the two words are not related; A value of -1 indicates a negative correlation between the two words. One-Hot representation solves the identification of words in a specific corpus. But the inner product of the One-Hot vector of any two different words is 0. (One-Hot vector has only one component is 1, others are 0. And the components valued 1 in One-Hot vectors representing different words are different.) So the cosine similarity must be 0,

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

which shows that each word is isolated and there is no way to represent the related information of two words. How to represent the related information? Fig. 4 shows a way to connect two vectors of words. Where x and y are One-Hot vectors of two words. h is a vector. It has k features components, which be expressed as h_1 to h_k respectively. Components of x are fully mapped to h, and components of h are fully mapped to y too. h is connected with x and y through the network in Fig 4. So the related information between x and y is retained in parameters on this neural network, those map x to h, then to y.

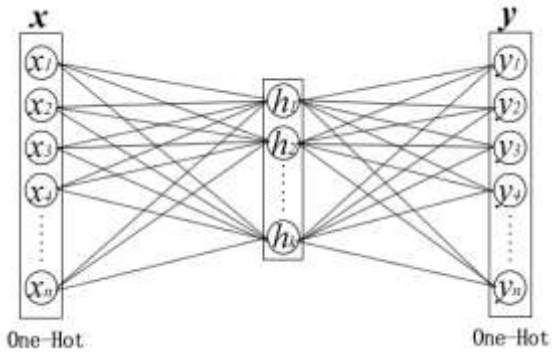

Fig 4: The neural network of the relationship between two words x and y

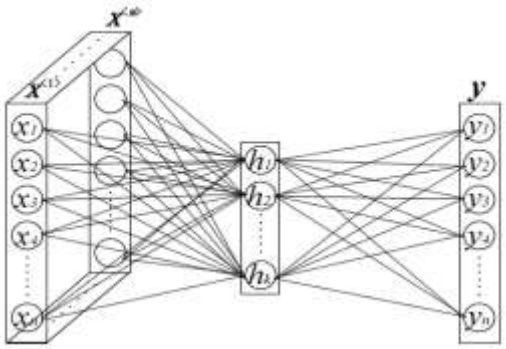

Fig 5: The neural network of the relationship between a word y and its context X

The semantic features are reflected in the association information between words in the context. So the related information between a word and its context (composed of multiple sentences) is needed to be described. Fig 5 shows the way to represent the related information of a word and its context. It is a network developed from Fig 4. X is a One-Hot matrix with a size of (m, n). It is a text record including m words, which is also the context of the word y. X includes m One-Hot vectors, which are expressed as $X^{\langle 1 \rangle}$ to $X^{\langle m \rangle}$. X is fully mapped to h by Eqs (2) and (3).

$$z = w^T X + b \tag{2}$$

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

$$h = \tanh(z) = \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}} \tag{3}$$

X is looked at as the input layer of the network, h is the hidden layer. The state before h is activated is z. w is the weight matrix with a size of (m, k) from X to z in the network and b is the bias. m is the first dimension of X, representing the number of words in the context. k is the dimension of h, representing the number of features of a word. The larger the value of k, the more inter-association information of words can be retained in the neural network. The network use $\tanh$ as the activation function to process z to h. Then h is fully mapped to y by Eqs. (4).

$$\bar{y} = w'^{T}h + b' \tag{4}$$

y is looked at as the output layer, $\bar{y}$ is the prediction of y. w' is the weight matrix with the size of (k, m) from h to $\bar{y}$ in the network and b' is the bias. The network from X to $\bar{y}$ is actually a task model for predicting a word y according to the context X. Because$\bar{y}$is the result predicted by the network, it is not a One-Hot vector. Finally, it will be processed by softmax function to recover to a One-Hot form. Then the whole neural network actually is a modified model of word2vec [7] as described in Fig. 6. The cost function of this model is Eqs. (5). $\bar{y}_j^i$ represents the value of the j_{th} component in the vector $\bar{y}^i$. And $\bar{y}^i$ is the i_{th} output corresponding to the input sample X^i. y_j^i represents the value of the j_{th} component in the vector y^i . And y^i is the i_{th} ideal value corresponding to the input sample X^i. d is the batch size in each iteration during modeling.

$$cost = \frac{1}{d}\sum_{i=1}^{d}\sum_{j=1}^{n} -y_j^i \log\bar{y}_j^i - \left(1 - y_j^i\right)\log\bar{y}_j^i \tag{5}$$

The framework uses sufficient data from the source domain to train the initial model 1 in Fig. 6 with mini-batch learning. When the cost value gradually reaches the satisfactory range, the w in this neural network model includes the semantic feature that we want to extract from the source domain. We also call it word embedding. The word embedding is a features matrix with a size of (k, m). It includes m vectors corresponding to m words in the dictionary established for the source domain. So every vector is a word vector with k feature components. This kind of word vector can identify each word uniquely in the corpus and can represent the relationship between any word and its context in the source domain.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

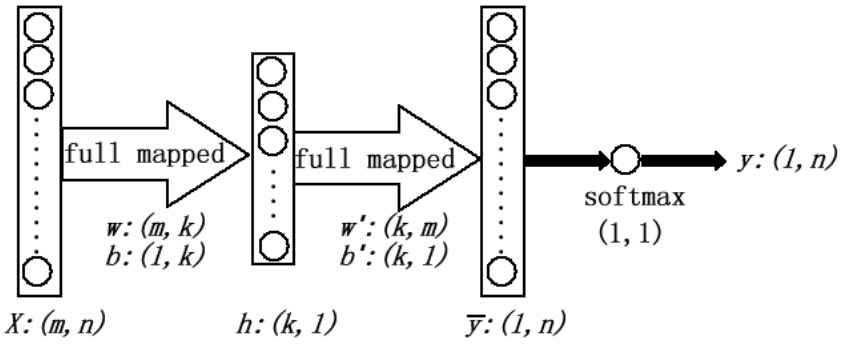

Fig 6: Initial model 1 for the task of semantic feature extracting

3.3.2 Semantic feature transfer to the target domain

Assuming the target domain is a feedback platform for a takeout catering system launched recently. Our purpose is to train a model that can judge positive and negative comments on the feedback text of customers. Although there is not much data in the new system and the feedback texts are mostly short texts, we can use the word embedding extracted from the source domain and transfer it to the target domain. The specific approach is using the word embedding extracted from the source domain to recreate each word in the target domain. Fig 7 shows the details. Assuming a record of the text data in the target domain is "apple is my favorite fruit". It is a short text with only one sentence. To convert the record to a feature matrix of it, we use the w (the word embedding). The first step is to get the record vector by it refers to the dictionary of the source domain. Find the index number for every word in the sentence referring to the dictionary, and then form the record vector according to their order in the sentence. The second step is to take out each component in the vector in turn and take the value of the component as the column mark in the w matrix to intercept the word vector in that column of the matrix. Finally, according to the order of the record vectors, these word vectors are spliced into the feature matrix of the record. For example, the first word in the record in Fig 7 is "apple", its index number in the dictionary is 35. So the first component of the record vector is 35. According to the value of 35, we intercept the word vector in the 35th column of the w and put it in the first row of the feature matrix of the record. So the first row is the word "apple" and it has sematic features brought from the source domain. Following these steps, we can convert all text records in the target domain into feature matrixes of record. These record matrices are the training data of the emotion analysis model. To process in batch, all record matrices must have the same size. The second dimension of the record matrix is the number of feature components in a word vector. All word vectors have the same number of feature components obviously. The first dimension of the record matrix is the number of words in a record. Usually, each record is different in the number of words (also called record length). To fix the first dimension of the record matrix to a proper value, the record length frequency of all records in the corpus should be counted, and the appropriate quantile should be selected as the record length parameter s according to the distribution to standardize the length of the records. Finally, the dimension of the feature matrix of records is (s, k). So that we can use insufficient text data in the target domain to model for the task of emotion analysis under semantic feature transfer.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

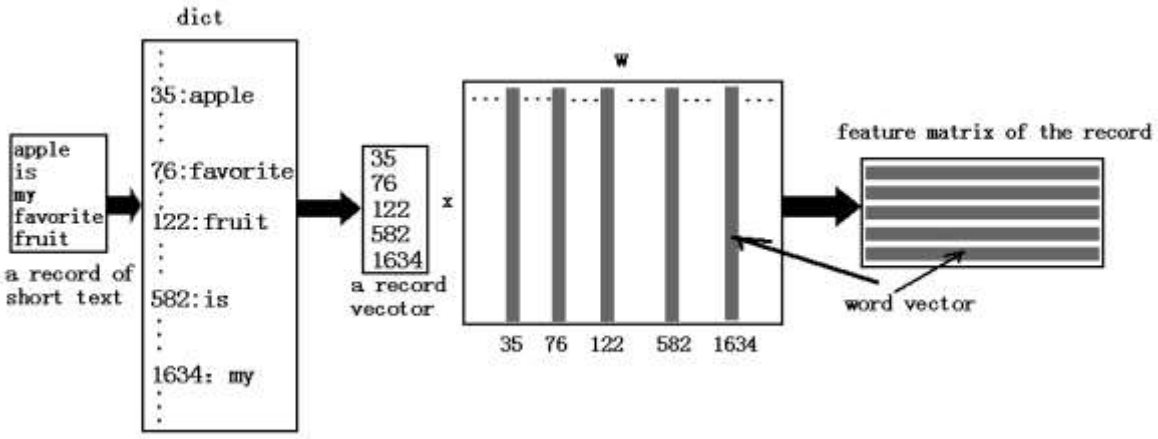

Fig 7: Convert a record to a feature matrix of the record

3.4 Modeling and Monitoring

3.4.1 Modeling with word order

Text data belongs to sequential data [6], which can be trained by the CNN model or the RNN [8] model prepared for sequential data. The framework chooses the SLTM [9] unit embedding in the RNN model to construct a neural network for the task of emotion analysis. Its advantage is to improve the long-term dependence in the RNN, and partially alleviate the gradient disappearance and gradient explosion in the RNN. Fig 8 shows the structure of the SLTM unit. The unit has three control gates. The forgetting gate determines how much of the state c_{t-1} of the t-1$_{th}$ step can be saved in the current state c_t, the updating gate determines how much of the input x_t of the t$_{th}$ step can be saved in the current state Ct, and the output gate determines how much of the current state can be used as the current output y_t. According to the parameter, record length s obtained above, an RNN model with step size s is constructed in Fig 9 for the task of emotion analysis. Each unit in the neural network is an SLTM structure. c_0 is a zero vector. x_1 to x_s are inputs and y_1 to y_s are outputs. The input of the task model is sequential data, the feature matrix of records with a size of (s, k) which is described in Fig 10. Therefore, the framework divides the matrix into s rows, corresponding to x_1 to x_s respectively. So x_t (t is from 1 to s) is a k dimensional vector. And the output is required to be a scalar. So the output only needs to get the last y_s, which is also a k dimensional vector. In order to get the scalar of good and bad prediction, we need to add a sigmoid unit after y_s for binary classification judgment and use $\bar{y}$ to replace y_s as the prediction value. Therefore, the final model for the task of emotion analysis is described in Fig 11. Where q is the number of lays that the model has.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

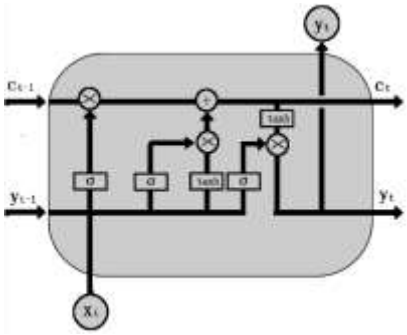

Fig 8: The structure of the SLTM unit

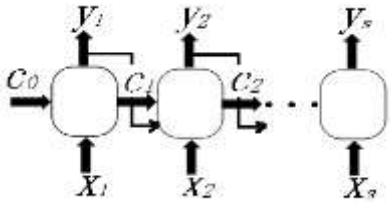

Fig 9: The RNN model

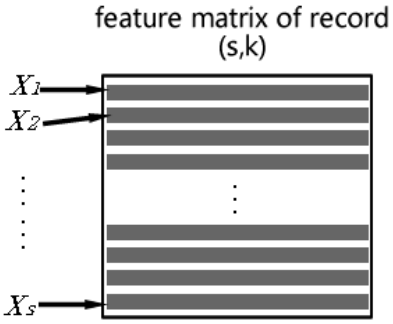

Fig 10: The input of the Initial model 2

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

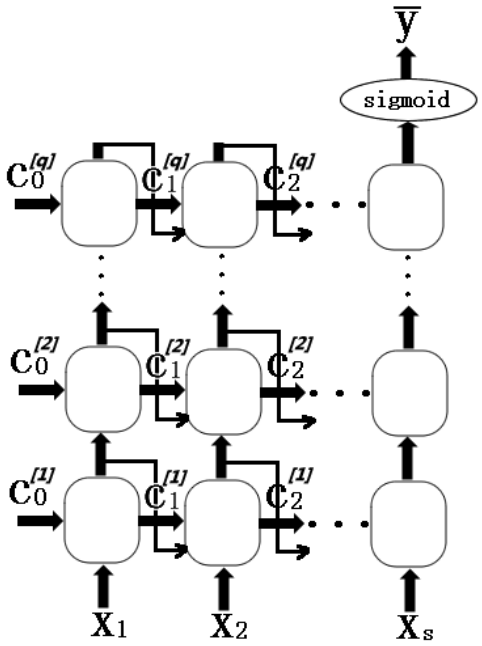

Fig 11: Initial model 2 for the task of emotion analysis

3.4.2 Monitoring for model innovation

How to train a satisfactory model with a neural network has given many good experiences to the predecessors of deep learning. Fine-tuning the hyper parameters and using cross-validation can screen the best among multiple candidate models. However, it is a time-consuming and laborious task to re-iterate and re-train the model whenever the hyper parameter is modified. No matter how well a model converges on the training set, you can only run it on the development set to know whether it is good or not. Why not intersperse the development set verification during parameter training? We put forward the idea of model innovation monitoring. Use the Jupyter-Notebook to build the neural network model and carry out visual training for the deep learning task.

The main idea of model innovation monitoring is to improve the real-time visualization of model training by inserting the verification of the development set into the process of parameter training. The convergence of the model on the training set can be monitored in real-time, and the prediction effect and convergence degree of the model on the development set under a certain iteration interval can also be monitored. Most importantly, by observing the difference of prediction error and prediction accuracy between the training set and the development set in the same state, we can timely understand the development trend of the actual deviation and variance of the model with the increase of the number of iterations in the training process. Once it is found that there is over fitting, oscillation non convergence, or under fitting caused by other reasons, the training can be ended as soon as possible, the causes can be investigated or the parameters can be adjusted rapidly. Under normal convergence, we can also find which parameters can reduce variance under different hyper parametric conditions with less iteration.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

VI. EXPERIMENTS AND RESULTS

4.1 Data Preparation

In order to verify the effectiveness of the emotional transfer framework of short texts in Chinese, the experiment uses the data set (2.87GB) of Chinese Wikipedia as the source domain corpus, which includes about 400,000 Chinese text encyclopedias involving 10 categories. There are more than 400,000 entries in the lexicon, and about 1 million records (super long text) will be retained after cleaning. And an ultra-lightweight data set (922KB) of feedback for the catering takeout system is selected as the target domain corpus. There are about 10,000 dietary evaluations (short text). Positive or negative labels have been established for all evaluations.

TABLE I. Data set description based on statistics

DATA SET	RECORD NUMBER	WORD NUMBER	DICTIONARY SIZE
Source Domain	399,511	240,000,000	1,089,805
Target Domain	11,987	164,657	11,380

The table I shows that there are only about 160 thousand words in the target domain corpus, but about 11 thousand different words. The two numbers indicate that each word in the target domain corpus appears only about 15 times in different records. If the text emotion analysis and modeling are directly carried out with the target domain data alone, the generated model can only learn the semantic feature between the texts on the existing small sample set, which seriously lacks the generalization ability and is difficult to perform well on the newly generated data in this domain. The source domain has sufficient data, and the text comes from many different encyclopedia categories. The source domain intersect with the target domain. Its data scale meets the demand of modeling for semantic feature extraction and the semantic feature transfer to the target domain.

4.2 Parameter Setting

In the data preprocessing stage, regular expressions are mainly used to remove unnecessary characters and formats in the text. The data processing of Chinese text usually includes the step of removing stop words. This operation can eliminate all kinds of auxiliary words that lack practical meaning in the text, making the semantic feature of the text more obvious. But this does not apply to this experiment. Although many auxiliary words do not express practical meaning, they play a great role in the expression of emotions, opinions, attitudes, and so on. In the text emotion analysis task, if the stop words containing a large number of auxiliary words are removed, the emotional expression of the original data will be distorted. Therefore, this experiment keeps these words in the text.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

When using initial model 1 to model and extract semantic features of the source domain, we should weigh the number of word features and the complexity of the neural network. The more features are given, the more real correlation between words can be described. But at the same time, it should increase the complexity of the model and the pressure of computation. In this experiment, the number of word features is set to 100. According to the vocabulary with 1,089,805 words in the source domain corpus, the word embedding with dimension (1089805, 100) is obtained from the model.

When the semantic features are transferred to the target domain corpus through word embedding, it should be considered that some words in the target domain may not be in the dictionary of the source domain corpus. If this happens, we regard these words as rare words or unknown words and assign them the same maximum value. It can be understood that these rare words have little correlation with other words.

When the record length of the target domain is standardized, we should first understand the record length-frequency distribution in the whole target corpus. By drawing the frequency histogram (Fig 12) and the box chart of distribution statistics (Fig 13) of the record lengths in the target corpus, it can be seen that 75% of the record length is below 42. The distribution of record lengths greater than 150 is very sparse. Finally, 90% of the quantile 71 is selected as the standard record length to standardize the record length of the records in the target domain corpus.

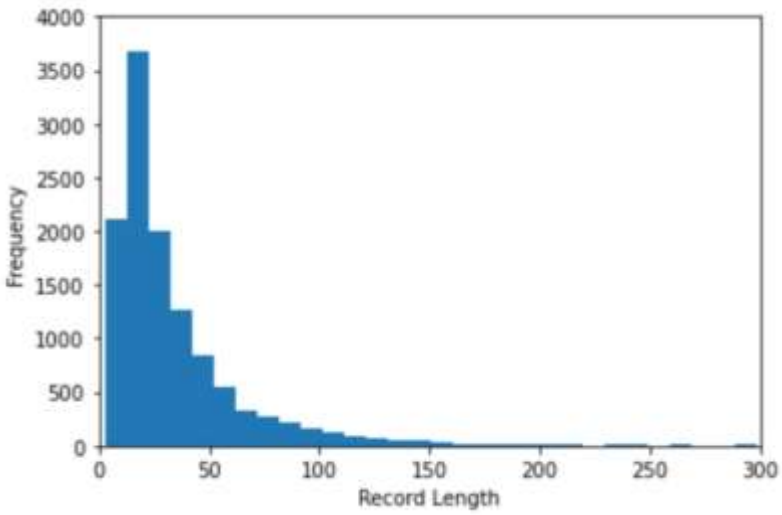

Fig 12: The frequency of record length in the target domain

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

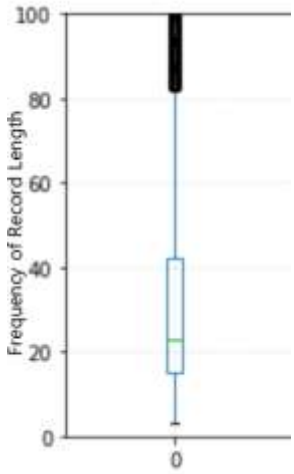

Fig 13: The distribution statistics of record length in the target domain.

4.3 Result and Discussion

In order to evaluate the effect of the emotional transfer framework of short text in Chinese, the data set in the target domain is randomly divided into three data sets: train set, develop set, and test set according to the proportion of 90%, 5%, and 5%. The training of the model (the model parameters will be updated) only contacts the train set. Develop set is used to validate and help the innovation of the model being trained (the model parameters are fixed). Finally, the selected model is used for text emotion analysis on the test set (the model parameters are fixed), then results and analysis will be obtained.

4.3.1 Evaluating indicator

The text emotion analysis model belongs to the classification task model, and the performance evaluation index of the model generally comes from the confusion matrix shown in Fig 14. In this experiment, there are two categories (0: positive and 1 negative), the dimension of the two classification confusion matrix is (2, 2), and there are four regions in the matrix. In the x coordinate, 0 indicates that the comment is predicted to be positive, and 1 indicates that the comment is predicted to be negative. In the y coordinate, 0 indicates that the comment is actually positive, and 1 indicates that the comment is actually negative. Thus, four basic indexes are obtained: TP (true positive), FP (false positive), FN (false negative), and TN (true negative).

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

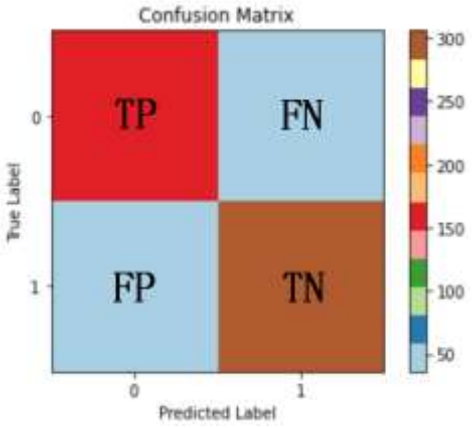

Fig 14: Binary confusion matrix

By counting these four conditions, the evaluation indexes of this experiment are got. They are accuracy, precision, and recall. The change trend of accuracy and recall rate with the increase in the number of the samples is a negative correlation. In order to consider balance, take the harmonic mean F1 score of the two as a more reasonable evaluation index. These indexes are defined as follows:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

$$precision = \frac{TP}{TP+FP} \tag{7}$$

$$recall = \frac{TP}{TP+FN} \tag{8}$$

$$F1 = \frac{2TP}{2TP+FP+FN} \tag{9}$$

4.3.2 Effect evaluation

When training the model for emotional analysis of short text, there are some hyper parameters needed to be assigned initial values. Such as the batch size of train data in every iteration step, the learning rate of updating for weights and bias, etc. different values have different effects on modeling. At the same time, different operations will also affect modeling. The impacts of different combinations on these choices are known only through experiments.

First, the data set has been divided into three parts: 90% is the train set, 5% is the developing set and 5% is the test set. But is each part taken from the data set in front to backorder, or randomly? It is uncertain whether the data set has been randomly processed beforehand, so we have carried out model training in both cases.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

Second, what is the size of the batch for the train set in each iteration step? When the size of the batch increases, the model will converge faster, require fewer iterations steps, and the model's training time is short relatively. However, at the same time, the variance may become larger and over fitting may occur. When the size of the batch decreases, the number of iterations required to scan the complete training set increases, and there may be large shocks, slow convergence speed, more iteration steps of model training, and longer training time during the error reduction. However, a small batch will highlight the individual feature of the sample and enhance the generalization ability of the model. Theoretically, it should reduce the variance and effect better on the test set.

Third, the front part of the neural network module is RNN, followed by a softmax activation unit. Should these two parts be regularized? We can set the dropout parameter for the front RNN part, and the latter part can use L1 or L2 regularization.

Fourth, what is the proper value for the learning rate? If it is too large, the step size of the parameters update will also increase, which may affect the convergence. If it is too small, the step size of the parameters update will decrease, which may slow down the convergence speed. Generally, the learning rate is higher at the initial stage of training, which makes the model approach the steady-state quickly. The learning rate decreases gradually in the later stage, so that the model can converge smoothly. We use the Adam optimization mechanism to control automatically.

According to the above four points, we trained many different models by selecting different values for batch size, regularization coefficient for L2, and deciding whether to divide data randomly. We choose the best six models in experiments and the comparison of the accuracy of each model in the training set and the development set is given in Fig 15. The data sets of groups (a), (b) and (c) in the figure were divided without random, so we mark them with "rand = 0"; The data set of groups (d), (e), and (f) were randomly divided, marked with "rand = 1". Model in (a) is without regularization, so it is marked with "lam=0", and other model use L2 to regularization with a regularization coefficient of 0.095, marked with "lam=0.095". The mark "batch" is the batch size of every model. Through comparison, we find that the feature distributions of the training set and development set of model (a) are very different without random segmentation. So the performance of model (a) in the development set is not good enough. The more convergent on the train set, the lower the accuracy on the development set. Then model (b) added regularization with a regularization coefficient of 0.095 based on model (a), the downward trend of accuracy has been partially inhibited. Model (c) increased the minimum limit of learning rate by Adam optimization, and the minimum learning rate should not be less than 0.001. Statistically, this limitation affects the convergence effect of the model on the train set. Through the performance of model (b) and model (d), it is found that under the same other conditions, the random segmentation data set can make the model converge and make the develop set have higher accuracy, which shows that the random segmentation can make the train set and develop set have more similar feature distribution. In the case of random segmentation of data set in model (e), the size of the batch is reduced from 550 to 100, and the accuracy of develop set is further improved. It shows that the smaller the batch size, the stronger the generalization ability of the model. This is also a means of deep learning regularization. However, as the

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

batch becomes smaller, the oscillation amplitude of the convergent curve of the model increases. When the batch is reduced to 78 in model (f), the model receives the best effect. Although the shock intensifies, the model does not converge on the train set prematurely, which increases the accuracy of the develop set and gradually tends to be stable on the accuracy curve of the develop set. Based on model (f), continue to reduce the batch size to 60 and 50, and we can get model (g) and model (h). When the batch size continues to decrease, the curve oscillation intensifies, the model convergence slows down, the number of iterations required for convergence on the training set increases significantly, and the time consumption of model training increases at the same time.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

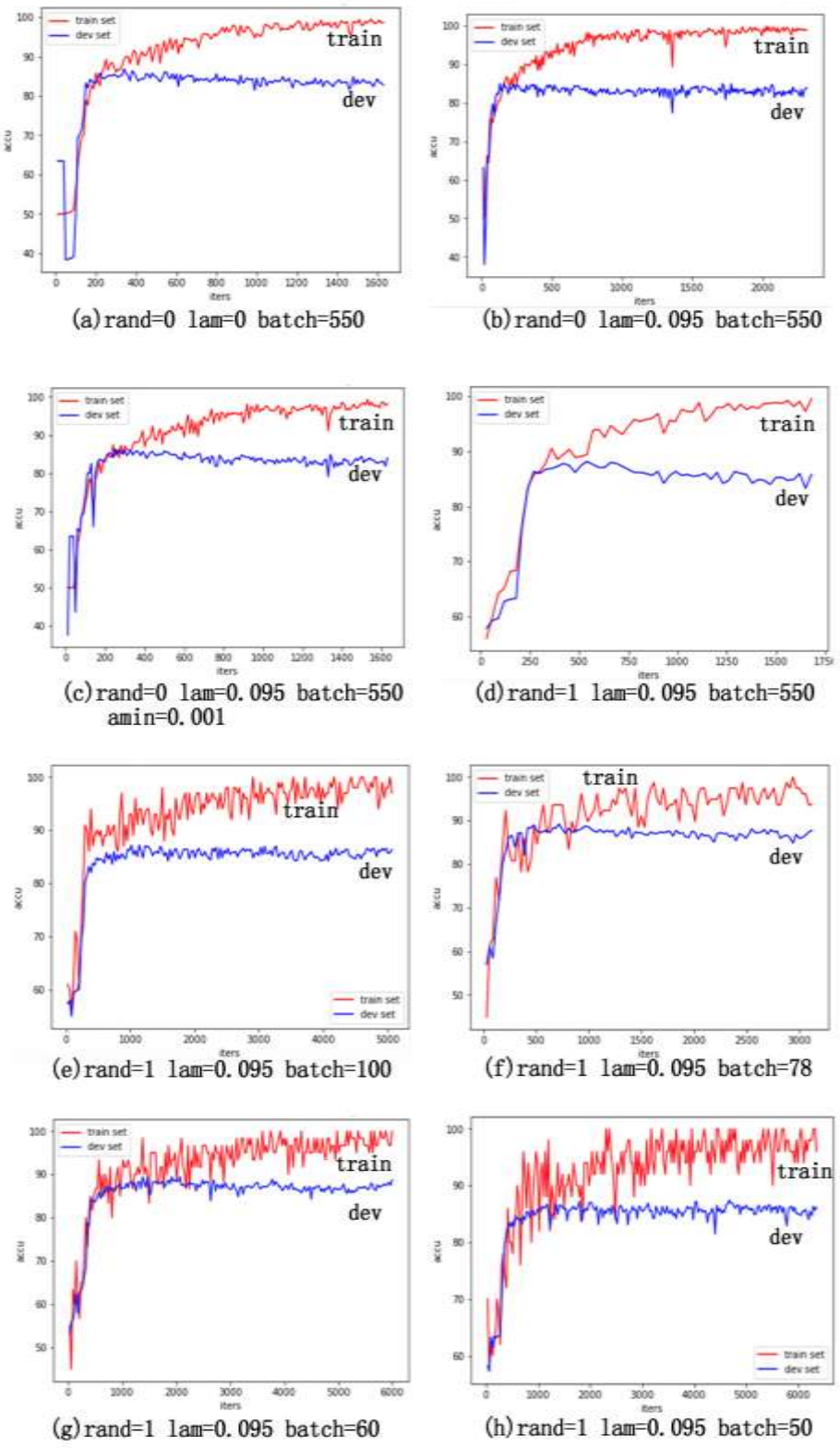

Fig 15: Comparison of accuracy in training set and development set under different parameter

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

Finally, we run the eight models selected from the experiment on the test set. Table II shows the accuracy, precision, recall, and F1 score of each model in the test set.

TABLE II. Comparison of evaluation indexes of experimental models

MODEL	ACCURACY	PRECISION	RECALL	F1
a	0.838	0.783	0.771	0.777
b	0.838	0.772	0.791	0.781
c	0.824	0.726	0.831	0.775
d	0.840	0.778	0.786	0.782
e	0.879	0.885	0.801	0.841
f	0.890	0.873	0.821	0.846
g	0.857	0.860	0.736	0.794
h	0.876	0.856	0.796	0.825

Through the data in the table, it can be found that model (b) adds regularization to the sigmoid unit of the model based on model (a), which improves its F1 score. Model (c) adds a lower limit of 0.001 for the continuous reduction of learning rate based on model (b), which affected the smooth convergence of the model, resulting in the value decline of almost all indicators on the test set. On the basis of model (c), model (d) changed the way of data set segmentation to random segmentation, which narrowed the difference of feature distribution on the train set, the develop set, and the test set, and improved the indexes on the test set a little. The following model (e), (f), (g), and (h) gradually reduce the size of the batch on the basis of model (d). The batch sizes of (e), (f), (g) and (h) are 550, 100, 78, 60 and 50 respectively. From the data of model (c), (d), and (e) in table II, we can find that the batch becomes smaller, and the indicators' values of the model are significantly improved. From the data of model (f), (g), and (h), we can find that when the batch size is too small, the oscillation amplitude of the model in the convergence process is too large. Finally, the performance on the test set is unstable. Finally, we compare the data in Fig. 15 and table II as a whole. It can be seen that the model with a good verification effect in the development set is also good in the test set generally.

4.3.3 Analysis of result

Firstly, through the performance of eight models on the test set, and related it to the accuracy curve of them on the train set and the develop set in the training process, it is concluded that the model (e) and model (f) are the most ideal models. Model (f) works best on the develop set. The effect of model (f) on the test set is also the best, with an accuracy of 89.0% and an F1 score of 0.846. The accuracy of model (e) is 87.9%, and the F1 score is 0.841, which is slightly lower than model (f). Fig 15 and table II are consistent in the performance judgment of the model, which shows that the processing of randomly

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

dividing the data set and reducing the batch size can improve the performance of the model. Of course, the randomness generated by each random sampling will slightly change the consistency of the feature distribution of the test set and the develop set, and the consistency of the performance effect reflected in the develop set and the test set will also fluctuate slightly.

Secondly, the accuracy of the best model selected in the experiment is 89.0%, and the F1 score is 0.846. The accuracy of the model is satisfactory in the task of emotional transfer for short text in Chinese. It is better than other methods in recent years. In 2011, the accuracy of the method of removing the centroid of the source domain by selecting the highly credible target domain was 74.6% [10]. In 2013, the accuracy of the method of obtaining samples in the target domain according to the heat conduction model was 71.5% [11]. In 2014, the accuracy of the dictionary-based migration method was 86.6% [12]. The accuracy of the emotion embedding method in 2018 was 81.0% [13]. The accuracy of the HANP method in 2019 is 87.76.

The performance of text emotion transfer in the test set is difficult to achieve the accuracy of general numerical task models. There are three main reasons: first, the short text data of the target domain is lite and insufficient, and there is a difference in semantic feature distribution between the source domain and target domain. Therefore, the accuracy of this kind of transfer model is lower than the model with large training data without transferring. Second, many meanings and emotional expressions in natural language are ambiguous, and different people have different understandings. It is difficult to give a well-defined judgment with appositive or negative classification. Many of these kinds of text data cannot be interpreted as positive or negative even by humans. This kind of text data in the target domain corpus will affect the training effect of the model. Third, the effect of the model is closely related to the quality of the training set. Through browsing the training set data, we find that the labels of some sample data are wrong. Too much such data will affect the training efficiency of the model. At the same time, this also reminds us that optimizing the quality of data sets is an optional idea to improve the quality of modeling.

V. CONCLUSION AND FUTURE WORK

Aiming at the problem that ordinary text emotion analysis methods cannot effectively model in a new or immature domain, we propose an emotional transfer framework of short texts in Chinese by deep learning technology. The framework can effectively obtain more comprehensive semantic features from sufficient text data in the source domain and transfer them to the target domain data, so that the target domain data set can also carry out text emotion analysis and modeling with lite data in the target domain. Experiments show that the framework can build a good model for the task of emotional analysis in a new or immature domain. The framework is feasible, the model effect is good, and the accuracy of emotion classification is 89.0%. It is better than other methods in recent years.

There are many aspects of the framework that can be improved through exploration. For example, the influence of the word's feature number in the source domain on the effect of the semantic feature transfer, the effect of the computational complexity of modeling, and the computational power

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 102-123
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

consumption brought by the specific implementation. How much does the scale of text data in the source domain affect the effect of the text emotion analysis model? All these aspects will continue to be explored in future work to improve the framework.

ACKNOWLEDGEMENTS

This work was supported by the Scientific Research Project of the Education Department of Hubei Province of China (Grant No. B2020188).

REFERENCES

[1] Zhao, C., Wang, S., Li, D., "Research progress on cross-domain text sentiment classification," Journal of Software, vol.31, no. 6, pp. 1723-1746, 2020.
[2] Remus R., "Domain adaptation using domain similarity-and domain complexity-based instance selection for cross-domain sentiment analysis," in Proc. 2012 IEEE 12th international conference on data mining workshops, 2012, pp. 717-723.
[3] Bollegala D, Mu T, Goulermas JY, "Cross-Domain sentiment classification using sentiment sensitive embedding," IEEE Trans. On Knowledge and Data Engineering, vol. 28, no. 2, pp. 398-410, Sep 2015.
[4] Xia, R., Zong, C., Hu, X. et al., "Feature ensemble plus sample selection: Domain adaption for sentiment classification," IEEE Intelligent Systems, vol. 28, no. 3, pp. 10-18, Feb 2013.
[5] Blitzer J., McDonald R., Pereira F. "Domain adaptation with structural correspondence learning," In Proc. of the 2006 Conf. on Empirical Methods in Natural Language Processing, 2006, pp. 120-128.
[6] Sutskever et al., "Sequence to sequence learning with neural networks, Advances in neural information processing systems, vol. 27, 2014.
[7] Mikolov et al., "Efficient estimation of word representations in vector space," arXiv preprint arXiv: 1301.3781, 2013.
[8] Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," arXiv preprint arXiv: 1406.1078, 2014.
[9] Hochreiter, S., & Schmidhuber, J., "Long short-term memory," Neural computation, vol. 9, no. 8, 1735-1780, 1997.
[10] Yang, W., Wang, Z., Li P., et al., "Adaptive domain sentiment classification based on centroid transfer," Computer Applications and Software, vol. 28, no.12, pp. 26-28, 2021.
[11] Wu, Q., Liu, Y., Shen, H., et al., "A unified framework for cross-domain sentiment classification," Journal of Computer Research and Development, vol. 50, no. 8, pp.1689, 2013.
[12] Mao, K., Niu, J., Wang. X., et al., "Cross-Domain sentiment analysis of product reviews by combining lexicon-based and learn-based techniques," In Proc. of the Int'l Conf. on High Performance Computing and Communications. IEEE Computer Society, pp. 351-356, 2015.
[13] Shi. B., Fu, Z., Bing, L., et al., "Learing domain-sensitive and sentiment-aware word embedding," in Proc. of the 56th Annual Meeting of th Association for Computational Linguistics, vol.1 pp. 2494-2504, 2018.