IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

Emotional analysis System of book review based on Neural network

To cite this article: Yanxiong Sun *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **563** 042012

Emotional analysis System of book review based on Neural network

Yanxiong Sun[1,a], Yeli Li[1,b], Qingtao Zeng[1,c], Yuning Bian[1,d]

[1]Beijing Institute Of Graphic Communication, China

[a]tongxin_yxsun@163.com, [b]liyl@bigc.edu. com, [c]954276545@qq. com, [d]774200483@qq.com

Abstract: Contemporary society, is an era of the Information explosion. The explosive growth of book publishing makes readers choose a good book, and publishers choose to publish a book with a market, which has become a very important topic. Aiming at the non-effective analysis of the current book publishers to obtain the real feelings of the users on the book, this paper puts forward that the neural network as the main algorithm is enough to classify the user's comments emotionally. In the collection of data, through the award-winning books and the two-star score under the book review, take different weights of the calculation, the emotional words of the eigenvectors to plump. And after the book evaluation is pre-processed, the user's credibility parameters are added to the input word vector. As a result, the accuracy of classification and judgment of the whole system is improved, the consumption of human, material and financial resources is reduced, and the resource can be invested in how to improve the level of publications, which will add to the cause of book publishing in China.

1. Introduction

Along with the advent of the era of big data, our lives have changed dramatically. Of course, the advent of the Big Data era has also affected the development of the book publishing industry and services. In the past, it was difficult for many publishers to get the books they published and actually evaluate them in the minds of their users. Publishers ' conclusions are based only on second-hand information returned by some book critics or agents, so they can not really effectively modify the publishing direction of the publishing house and the publication plan according to the needs of readers. With the rapid development of cloud computing, artificial intelligence and neural networks, natural language Processing (NLP) technology has ushered in a era of great development. In particular, the emotion analysis based on text is the hotspot and focus of domestic law research in recent years. Based on the neural network, this paper optimizes the content of the dataset and introduces the trust score of the commentator, which makes the system have outstanding performance in the test set of processing the data set (Edbrup).

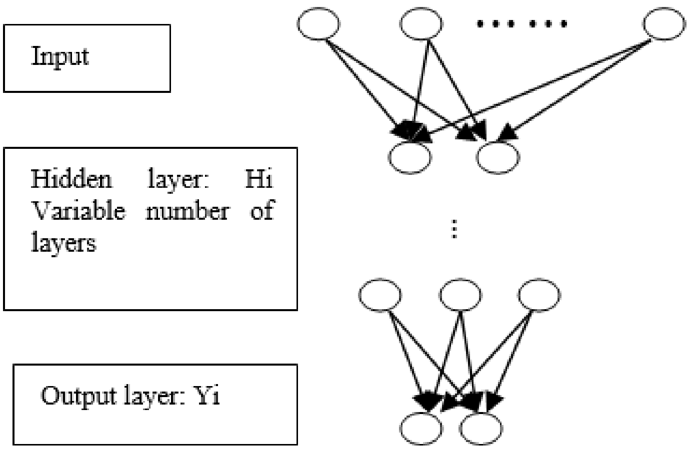
2.The main algorithms of UA-W2V-BP model

2.1 Introduction of neural network algorithm

As shown in the following figure 1 Shown: The leftmost is the input layer of the neural network and the middle layer (removing the beginning and end two layers) is the hidden layer and the far right is the output layer. The circle represents a neural element, and the bottom of each layer is the bias point of each layer, which is the same as the y=k*x+b of B in the unary equation. Xi represents input, the number of neurons per layer is not limited, according to the performance of the model can be deleted or deletions. Neurons between layers are linked by solid lines, and the output of each layer is calculated using function uo=f (wt*x+b). F is called an activation function, there are a wide variety of activation functions, commonly used are tangent functions, sigmoid functions, Softmax functions and so on. In this paper, the Softmax function is used as the activation function (Formula 2-1).

$$soft\max(x)_j = \frac{e^{x_j}}{\sum_{k=1}^{K} e^{z_k}} \qquad (2\text{-}1)$$

The Value field (-1,1) of the Softmax function, which defines the domain（-∞， ＋∞） so that the data is mapped to (0,1), can be interpreted as probability, which facilitates subsequent interpretation studies.



Figu Figure 1: Plain Neural network overview

2.2 UA-W2V-BP model.

2.2.1Model flowchart

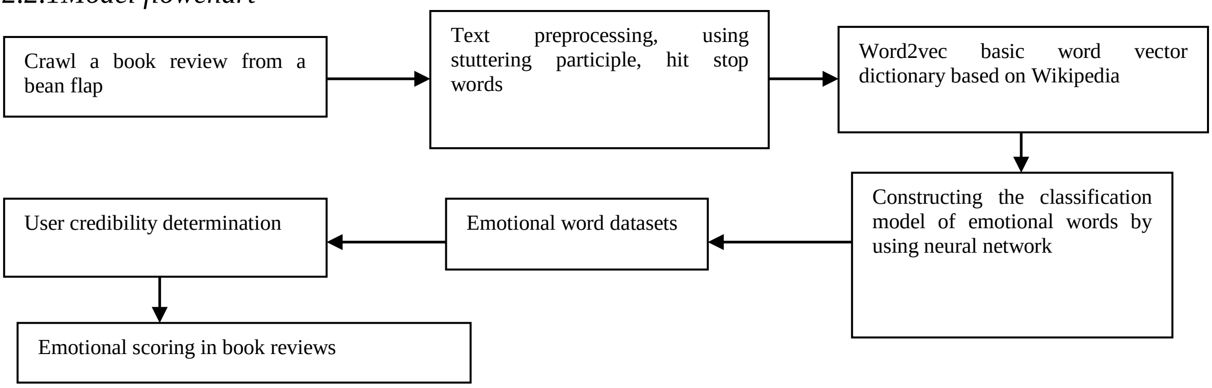


Figure 2: Flowchart of this system

2.2.2 Input of book text vectors.
General word vectors, such as one-hot word vector encoding, generally can not contain the semantic information of words, so in order to obtain the semantic information of each word and context, this paper uses word2vec Wikipedia word vector as the reference word vector dictionary (dictionary content is not limited to this). ei: A word vector that represents wi. Input of the xi:UA-W2V-BP model.

$$x_i = e_i \otimes score(w_i) \tag{2-2}$$

2.2.3 Model Training.
The final output of this article is output by the Softmax function. Suppose the training set contains a V Training sample:$\{(x^{(1)},y^{(1)}),(x^{(2)},y^{(2)}),\cdots,(x^{v},y^{v})\}$, Where x^{i}represents the input eigenvector, the class marks $y^{i} \in \{0,1\}$. Using x as input, through the processing of the model, the emotional classification label of the sentence is obtained. Therefore, the probability of classifying x as J is as shown in the formula (2-3):

$$p(y^{(i)} = j \mid x^{(i)};\theta) = \frac{e^{\theta_j^T x^{(i)}}}{\sum_{i=1}^{K} e^{\theta_j^T X^{(I)}}} \tag{2-3}$$

Where, k represents the number of labels, because the text focuses only on both positive and negative types of comments, the value of K is 2. Then, the training parameters θ are followed by our model to minimize the cost function, such as formula（2-4）:

$$loss = -\frac{1}{m}[\sum_{i=1}^{m}\sum_{j=0}^{1} 1\{y^{(i)} = j\}\log p(y^{(i)} = j \mid x^{(i)};\theta)] \tag{2-4}$$

The next step is to need more of our loss function to derive and find the minimum value point. Formula (2-5):

$$\nabla_{\theta_j} loss = -\frac{1}{m}\sum_{i=1}^{m}[x(i)(1\{y^{(i)} = j\} - p(y^{(i)} = j \mid x^{(i)};\theta))] \tag{2-5}$$

The optimal model is obtained by using the gradient descent method to train the parameters by means of the deflection guidance operation of the upper formula.

2.2.4 The management of user and data centralization in this model.
DataSet: Two types of book reviews from Douban. One is the review of the award-winning work of the contradictory literary prize over the years, which is used as our current example dataset, and the second is the review of literary books selected from the Douban between 0 and 2, as a negative example dataset of our dataset.

$$Freq(w_i) = a_1 * Neg(w_i) + b_1 * pos(w_i) \tag{2-6}$$

$$score(w_i) = 10 * sig\,\mathrm{mod}\,e[Freq(w_i)] \tag{2-7}$$

$$score(w_i^{'}) = 5*[b_1 * \frac{pos(w_i^{'})}{pos(w_i^{'}) + pos(w_i)} - a_1 * \frac{Neg(w_i^{'})}{Neg(w_i^{'}) + Neg(w_i)}] \tag{2-8}$$

$pos(w_i)$ indicates the frequency with which the first i emotional word w_i appears in the comments of a well-known book; $Neg(w_i)$ indicates the frequency of the comments of the i emotional word w_i in the 0 to 2 star Book; $Freq(w_i)$ indicates the frequency of the first i emotional word w_i in the dataset (Edbrup); $pos(w_i')$ indicates the frequency with which the first non-emotional word w_i' appears in the comments of well-known books; $Neg(w_i')$ indicates the number of frequencies in the comments of the first non-emotional word w_i' in books of 0 to 2 stars; $score(w_i)$ represents the score of the first emotional word $score(w_i')$ indicates that the first is a score in the dataset (Edbrup) that does not have an emotional predisposition word.

3. Key algorithms(experiments) in UA-W2V-BP models.

3.1 Collection of datasets

At present, most of the data packets in the book review set are missing user portraits, so the analysis of book reviews lacks some objectivity, based on the author in the process of collating and collecting data sets, the number of articles, continuity, books of concern, the age of the reviewers, and the preference emotion of comments are crawled separately. Comments on the use of comments by commentators to rate the credibility of the comments, input into the final scoring model. The selected positive books are all book reviews that have won the Mao Dun Literary Prize, and the negative examples are all literature books with a frequency score of less than two stars. and the number of comments is not less than 10,000. The number of books is not less than 100. Experiments randomly extracted 80% of the data as a training set, the remaining 20% data as a test set.

3.2 Optimization of data dictionaries

As we all know, the construction of a good model lies first in having a good dataset, after which the neural network algorithm can be used to process the data and train the model. Therefore, some sources of the feature words in the Emotional Dictionary of this paper are manually screened in the bean Flap Review, and the other part is selected from the hownet Emotional Dictionary of the Zhong Wenjing, negative emotional word set. It contains a variety of non-textual forms of pictures, dynamic diagrams, expressions, such as more than 200 kinds.

3.3 Word vector dictionary and deactivated Word dictionary

In the processing of book reviews, this paper uses Gensim's Word2vec word vector to train in Wikipedia. It also contains more than 200 dimension vector representations of 575,821 words. For words that can no longer be found in the Wikipedia thesaurus in the dataset, we use the word vector corresponding to the "0" element in the Datum word vector dictionary to represent the lexical vector. In addition, deactivation words will also exist in our comments, so we use hit's deactivated thesaurus, a total of 1893 deactivation words, but also contains useless symbols, such as ",", "" This "and so on.

3.4 User Portrait analysis

In order to analyze the portraits of different users, the user is scored. Select more than 100 gold users from the bean flap, with a small number of portraits of the remaining users of each level, with scores ranging from low to high, as shown in Figure 3.

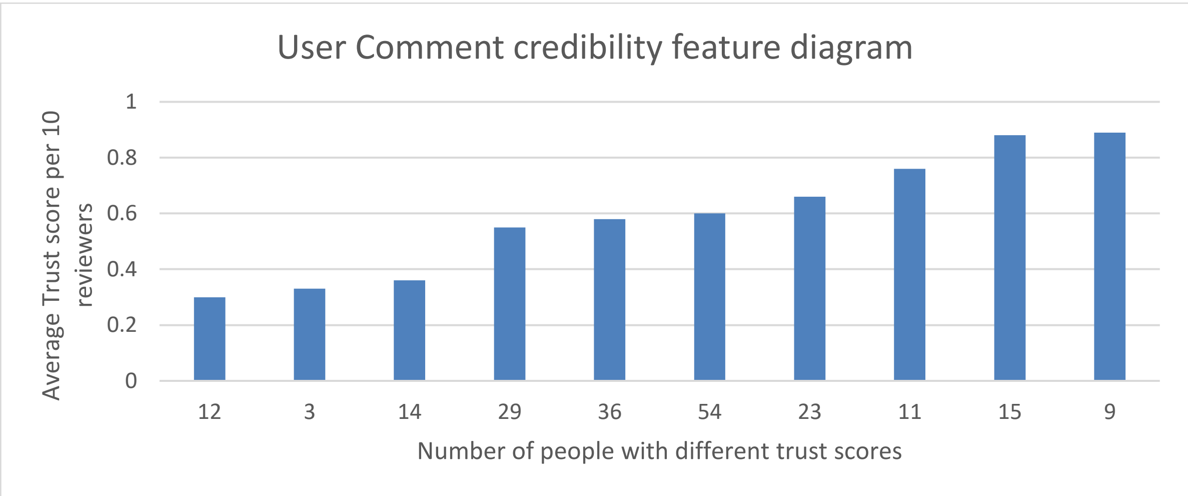


Figure 3: User Comment credibility

3.5 Analysis of experimental results.
In this paper, the precision rate (precision), recall rate (recall) and f1-measure, which are often used in classical NLP (Natural language Processing), are taken as the important performance parameters to evaluate this model.

$$precisim = \frac{\sum_{c_j \subset c} \text{Correct judgment of } \mathrm{c_j} \text{ Books}}{\sum_{c_j \subset c} \text{Test focus correctly Judge } \mathrm{c_j} \text{ Books}} \qquad (3\text{-}1)$$

$$recall = \frac{\sum_{c_j \subset c} \text{was correctly judged to be } \mathrm{c_j} \text{ books.}}{\sum_{c_j \subset c} \text{The test set actually belongs to the } \mathrm{c_j} \text{ class book}} \qquad (3\text{-}2)$$

$$F1 = \frac{2 * pre * rec}{pre + rec} \qquad (3\text{-}3)$$

Table 1 is the results of two models on the evaluation of the emotional dataset Edbrup (emotional dataset of book reviews and user portraits) with user portraits.

Table 1: Evaluation results of the two models under 3 indicators (p accuracy rate, R recall rate, F1)

Model	Index	Positive	Negative	Overall
SVM	P	0.75	0.79	0.75
	R	0.63	0.83	0.75
	F1	0.68	0.81	0.75
UA-W2V-BP	P	0.87	0.83	0.85
	R	0.75	0.91	0.85
	F1	0.81	0.87	0.85

The implementation results show that in the evaluation of the DataSet (Edbrup), the fraction of the traditional method represented by SVM is significantly lower than that of this model. The key lies in the consideration of user portraits in this model, the use of neural network algorithm to construct the classification model of emotional words, the deletion of Chinese word segmentation, named entity recognition and deactivation words in the preprocessing of data, the improvement of data purity, and the weighting of comments by user portraits on the credibility of comments of different users. To sum up, the performance of affective classification of this model is improved.

4.Conclusion

This paper presents an emotional analysis system based on neural network for book evaluation, which is based on deep neural network to realize the semantic information of book review text and the emotional information of words. The representation results are then combined with the user's user evaluation credibility information and are a key point in the book score. Then the emotional polarity classification of the evaluation text is realized by Softmax layer. In this paper, a data set for the study of emotional classification of book reviews is constructed, which will provide data resources for researchers of text emotion classification. This paper takes the neural network as the main structure, combines with the Word2vec word vector method and adds the emotional analysis model system of UA-W2V-BP, which is composed of the comment user portrait and the combination of the three.

Acknowlegments

Beijing Institute of graphic communication 2018 R&D Project(Ec201802).Beijing science and technology innovation service capability constructionproject(PXM2016_014223_000025)

IOP Conf. Series: Materials Science and Engineering **563** (2019) 042012 doi:10.1088/1757-899X/563/4/042012

Reference

[1] R.Collobert,J.Weston,L.Bottou,M.Karlen,K.Kavukcuoglu,andP.Kuksa,Natural language processing (almost) from scratch,J.Mach.Learn.Res.,vol.12 :2493–2537, Aug. 2011.
[2] Wu Jie, based on the user's emotional tendency perception of Weibo emotional analysis method [J]. Journal of Shandong University (Rigaku Edition), 2019, Volume 54th, issue 3rd.
[3] PANG B,LEE L.Opinion mining and sentiment analysis[J].Foundations and Trends in Information Retrieval,2008,2(1/2):1-135.
[4] TABOADAM,BROOTEJ,TOFILOStlM,etal.Lexicon-based methods for sentiment analysis[J].Computation-al Linguistics,2011,37(2):267-307.
[5] X.Chen,L.Xu,Z.Liu,M.Sun,andH.-B.Luan,"joint learing of characters,"in Proc.51st Annu.Meetingn Assoc.Comput.Linguistics,vol.1,2013:125-134.
[6] Erhan D,Bengio Y,Couville A,Manzagol P A,VincentP,Samy B.Why does unsupervised pre-training help deep learing? Journal of Machine Learing Research,2010,11:625-660.
[7] Sun Zhi-Jun,Xue Lei,Xu Yang-Ming,Wang Zheng.Overview of deep learing.Application Reasearch of Computers,2012,29(8):2806-2810.
[8] Peng Hao.Research on Emotional Analysis Algorithms Oriented to Public Opinion of Multi-Source Social Networks[J].Information Technology.2019:43-48.
[9] text classification based on cyclic and convolution neural networks [J]. Liu Ascendas, in duality, Zhang Hongtao, Yinhong. Software. 2018 (01).
[10] Learning to Forget: Continual Prediction with LSTM[J] . Felix A. Gers,Jü,rgen Schmidhuber,Fred Cummins. Neural Computation . 2000 (10)
[11] Recurrent convolutional neural networks for text clas-sification. Siwei Lai,Liheng Xu,Kang Liu,Jun Zhao. Twenty-Ninth AAAI Conference on Arti-ficial Intelligence . 2015
[12] Learning to Forget: Continual Prediction with LSTM[J] . Felix A. Gers,Jü,rgen Schmidhuber,Fred Cummins. Neural Computation . 2000 (10)