Research and Analysis in Fine-grained Sentiment of Film Reviews Based on Deep Learning

Jin Zheng¹, Limin Zheng¹,²* and Lu Yang¹,²

¹College of Information and Electrical Engineering, China Agricultural University, Beijing, 100083, China
²Beijing Laboratory of Food Quality and Safety, China Agricultural University, Beijing, 100083, China
*Corresponding author’s e-mail: zhenglimin@cau.edu.cn

Abstract. Films are one of the main entertainment ways for people. By exploring, analyzing and summarizing the reviews that people published on the internet, audiences can make better viewing choices, while investors can get a more convenient way to understand the audience's feedback. The text proposes a method of deep learning to perform fine-grained sentiment analysis on the film reviews, and restores the user's real emotion as much as possible. The method first preprocesses the data, and converts the words into vectors using the Word2Vec model, then inputs the word sequence in the Long Short-Term Memory network (LSTM) to learn the semantic dependence. Finally, the logical regression classifier is used to classify. Compared with Naive Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Back Propagation Neural Network (BPNN), and Convolutional Neural Network (CNN) models, the method achieves 83.6% accuracy in the binary classification, and 76.1% and 51.2% accuracy in the positive and negative fine-grained emotional classification, which has achieved the best results proved by experiments.

1. Introduction

Film reviews emotional analysis refers to judging its emotional tendency by analyzing and digging subjective information in film reviews, identifying audience's praise and derogation evaluation of films from reviews, conducting fine-grained emotional analysis, restoring audience's true emotions as far as possible, and providing help to the relevant personnel of film. But in fact, fine-grained emotional analysis of film reviews is relatively difficult. The audience's evaluation of films is actually the evaluation of people, which is versatile. Therefore, the accuracy of emotional analysis of film reviews by the same model is usually lower than that of other categories of data. There are three main methods for emotional classification: emotional word tagging, traditional machine learning, and deep learning. Emotional word tagging is not extensive. Because new vocabulary continues to be produced, the emotional dictionary needs to be constantly updated, which will consume a lot of manpower and material resources, so it is generally based on traditional machine learning methods and deep learning based methods. Pang et al. applied traditional machine learning methods to emotionally classify texts to achieve 80% accuracy[1]. Kim et al. used CNN in the deep learning method to classify sentences and set up different parameter models to study the effect of correct classification of sentence emotions[2]. Zhang et al. studied the classification results of character-level distributed vectors trained by CNN model in LSTM model[3]. This paper proposes to use the most advanced deep learning method to conduct fine-grained sentiment analysis on the short evaluation information in Douban film top250, and compare it
with different classification models to verify the correct rate of our proposed model. Traditional text categorization methods usually include two stages: design feature engineering and selecting classifier. The design feature engineering stage requires text preprocessing, text representation and feature extraction, and focuses on designing hand-made features\textsuperscript{[4]}. The selection of the classifier stage is to select the statistical classification method, which is classified by the polynomial linear logistic regression classifier. The performance of the method depends largely on the representation of the data. This method often uses the term frequency/inverse document frequency algorithm (TF-IDF) to extract features, construct a text vector space model, and implements text classification using NB, SVM, KNN, and BPNN algorithms. We use these models to perform fine-grained sentiment analysis on the film reviews, and compare their correctness rates on the Douban film top250 short reviews data set, and select the optimal classification model.

With the development of deep learning, some classical deep learning models are also used for text categorization tasks in natural language processing, because linear classifiers cannot share parameters between features and classes, their sparse and discrete features are not conducive to considering the relationship between sentences. These factors may limit the generalization ability of classifiers. To solve this problem, the deep learning method is used for text classification, such as CNN\textsuperscript{[5]} and LSTM\textsuperscript{[6]} to learn text representation and achieve good results. The deep learning-based approach first uses the Word2Vec model\textsuperscript{[5]} to represent the input document as a series of words, each word corresponding to a vector, multiplied by a weight matrix to form a dense sequence of real-valued vectors, forming a continuous vector space, and then the real-valued vector sequence is fed into the neural network to generate prediction probabilities that maximize the classification accuracy on the training set. CNN can capture the local features of sentences well and has a structure of local and weight sharing. The spatial structure relationship can greatly reduce the number of parameters that the neural network needs to learn, and reduce the complexity of the network model while preventing over-fitting. However, since the acceptance domain of each convolutional layer is very small, many layers are needed to capture the long-term dependence in the sentence\textsuperscript{[8]}, while LSTM only needs one layer to capture the semantic dependencies between words in long sentence sequences, and each hidden state is calculated based on the entire input sequence. In this paper, CNN and LSTM classify the Douban film top250 short reviews data set and compared with the traditional machine learning method to obtain the best fine-grained sentiment analysis model suitable for the data set.

This paper proposes to use the LSTM model to perform fine-grained sentiment analysis on the Douban film top250 short reviews data set. The main contributions are as follows.

• Establish Douban film top250 short reviews and sentiment analysis standard, and the positive and negative polarity labeling of the data set is carried out by star rating. Based on the 7 kinds of emotions of the emotional vocabulary ontology DUTIR\textsuperscript{[7]} of Dalian University of Technology, the fine-grained emotion analysis is further carried out.

• Pre-processing and filtering the established data set, deleting content that is not related to the film review or repeating a lot, changing the traditional characters, typos, etc. in the film reviews, adding the network new words and emojis to the sentiment dictionary, and using the dictionary for word segmentation and word removal.

• The traditional machine learning methods (NB, SVM, KNN, and BPNN) are compared with the deep learning methods (CNN and LSTM) to verify the best classification model for the data set and to achieve a fine-grained sentiment analysis of the Douban film top250 short reviews.

2. Proposed method
In this paper, we propose to obtain the Douban film top250 short reviews data set, preprocess and filter the data set, then use LSTM to perform fine-grained sentiment analysis on the data set. The CBOW model in the Word2Vec network is used to learn the word vector representation. LSTM is used for sequence prediction between words in sentences, automatically extracts features, captures semantic dependencies, and finally realizes text classification using logistic regression classifiers.
2.1. Acquisition and annotation of data sets
This paper studies the fine-grained sentiment analysis of the film reviews and selects the Douban Film Website with high audience recognition and attention as the information source of the data set and obtains the Douban Film Top250 short reviews as the data set. 5 stars means strong recommendation, 4 stars means recommendation, 3 stars means okay, 2 stars means poor, 1 star means very bad, According to the star rating, these short reviews are divided into positive and negative two polarities, 1, 2 stars are negative emotions, 4, 5 stars are positive emotions, and 3 stars are neutral emotions. The polarity of the label is analyzed using the SnowNLP library and the results are shown in figure 1. Since the short comment of 3 stars is generally not polar, it is discarded. Our correctness for polarity partitioning has been verified in the distribution of the above figure.

![Figure 1. Distribution results of various stars on SnowNLP.](image)

Based on a Chinese ontology resource DUTIR compiled and marked by the Information Retrieval Research Laboratory of the Dalian University of Technology, the above positive and negative polarities are further fine-grained. The results are shown in table 1.

According to the above criteria, the obtained Douban film top250 short reviews data set is first classified into two levels, and then fine-grained emotional labeling is performed on the basis of the two-level labeling.

| Binary classification | Fine-grained classification | Emotional words | Case |
|-----------------------|-----------------------------|-----------------|------|
| happy                 | happy                       | peace of mind   | joy, laughter |
| respect               |                             | practical, reassuring, conscience | respectful, awe-inspiring |
| positive              | praise                      | handsome, excellent, reasonable, realistic | |
| good                  | believe                     | trust, reliability, unquestionable | |
|                      | favorite                    | admiration, baby, love | |
|                      | wish                        | longing, blessing, long life, longevity | |

Table 1. Partition of positive and negative data sets.
2.2. Data preprocessing

Use the Jieba word segmentation tool in python to segment the data. The data set contains a large number of network terms and emoji. Therefore, the dictionary function is used to assist the word segmentation. When identifying new words in the network, the word frequency, mutual information and information entropy are used to determine the word structure. A threshold is set to extract network new words from the data set, and the identified network new words are added to the dictionary, ensure the correctness of the word segmentation as much as possible, and facilitate the extraction of the subsequent text feature[10]. Then stop the words and remove the words that are used to maintain the grammatical structure but have no practical meaning and are meaningful but cannot express the meaning of the sentence.

2.3. Word embedding

The embedding layer implements a distributed representation of words, and each word is represented as a low-dimensional, continuous real-valued vector. The word2vec neural network framework used in this paper includes the input layer, the projection layer, the hidden layer, and the output layer. The CBOW model is used to predict the current word in the output layer by inputting n-1 words around the current word to the projection layer. The structure is shown in figure 2.

Figure 2. Word2Vec network structure and CBOW model.

$W$ and $U$ represent weight matrix, and $p$ and $q$ represent representation bias vectors. The input corpus is $C$, the length of the word vector is $m$, $w$ is the word in the corpus $C$, let $n$ be the context length of $w$, and $Context(w)$ take n-1 words in front of the current word, $(Context(w), w)$ forms a training sample of
binary pairs, the vector $x_w$ of the projection layer joins the first bits of the $n$-1 word vectors of the input layer to form a long vector of length $(n-1)m$, and the samples are trained using the equation in the Word2Vec network (1) (2) (3) for calculation.

$$z_w = \tanh(Wx_w + p)$$
$$y_w = Uz_w + q$$
$$p(w|\text{Context}(w)) = \text{softmax}(y_w) = \frac{\exp(y_{w}))}{\sum_{w=1}^{n} \exp(y_{w}))}$$

2.4. Long short-term memory network layer

The Recurrent Neural Network (RNN) is capable of processing sequences of any length by using mechanisms of backpropagation and memory\[9\][19]. However, if the input sequence is long, the RNN training will suffer from the problem of gradient disappearance or explosion\[10\], and it is difficult to model long-distance correlation in the sequence. LSTM can avoid these problems. Compared with RNN, LSTM has a complex internal structure in the recurrent layer and uses four different layers to control information interaction. LSTM implements input, output, forgotten gates, and candidate memory cells by designing a "gate" structure, and calculates\[11\] by equation (4) - (9).

$$f_t = \sigma(Wf_x + Uf_h + b_f)$$
$$i_t = \sigma(Wi_x + Ui_h + b_i)$$
$$g_t = \tanh(Wg_x + Ug_h + b_g)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot g_t$$
$$o_t = \sigma(Wo_x + Uo_h + b_o)$$
$$h_t = o_t \odot \tanh(c_t)$$

The LSTM maintains the existing memory cell $c_t \in \mathbb{R}^n$ at time $t$, the input at each time step $t$ is $x_t$, $h_{t-1}$, $c_{t-1}$, the output is $h_t$ and $c_t$, and $\sigma$ represents sigmoid function, $\odot$ denotes the vector product, $h_0$ and $c_0$ can be initialized to 0, $W$ and $U$ represent weights, and $b$ represents the bias. The corresponding network structure inside the LSTM is shown in figure 3.

2.5. Classification layer

The classification layer is essentially a logistic regression classifier. The prediction probability of all categories is calculated by the softmax function\[12\], and the probability prediction of the classification is performed using equation (10).

$$p(y = k|X) = \frac{\exp(W^T_x y + b_k)}{\sum_{k=1}^{K} \exp(W^T_x y + b_k)}$$

$K$ denotes the number of categories, $W$ denotes the weight, and $b$ denotes the bias value.
3. Experiments

3.1. Datasets
Short reviews of 250 Douban films from the Douban Website are used as the datasets for this article. Each film has 5 stars, each of which receives 100 pieces of data for each star, totaling 125,000 pieces of data. After removing 3 stars short reviews, the remaining 100,000 data are obtained. According to star classification, 50,000 data of 1,2-star reviews are negative emotions, and 50,000 of 4,5-star reviews are positive emotions. 5,000 data are randomly extracted from positive and negative emotions datasets as experimental data and labeled. The training set and test set are divided into 7:3 scales and the results are shown in table 2.

| category | number of training sets | number of test sets |
|----------|-------------------------|---------------------|
| positive | 3500                    | 1500                |
| negative | 3500                    | 1500                |

3.2. Experimental device
Word2Vec model is used to obtain the word vector, and the word embedding dimension is set to 200. The stochastic gradient descent method is used to train the model[^13], which makes the model converge faster. The dimension of the LSTM hidden layer is set to 128, the dropout parameter is used to set it to 0.5. Finally, the cross-entropy loss function is used to train the model and early stopping strategy is used to prevent over-fitting. The hyperparameters of the model compared with LSTM are extracted from the relevant literature, and the optimal parameters are selected through training. The parameters and feature vectors of the whole neural network are updated during training. The correct rate is used as a measure of the classification performance of the model.

3.3. Baseline methods

3.3.1. Feature-based methods
The TF-IDF algorithm is used to extract the features of the data set, and the classification algorithms of NB, KNN, SVM, and BPNN are used to establish the classification prediction model. The idea of the algorithm is that if a word appears frequently in documents and rarely in other documents, it is highly representative and suitable for use as a distinction.

Naïve Bayes performs better when classifying small-scale data and can handle multi-classification tasks, but the model assumes features are independent of each other and cannot learn the relationship between features[^14]. Calculated by equation (11).

\[
p(y|x) = \frac{p(x|y)p(y)}{p(x)} = \frac{p(a_1|y)p(a_2|y_2)...p(a_m|y)p(y)}{p(x)} = \frac{p(y) \prod_{j=1}^{m} p(a_j|y)}{\prod_{j=1}^{m} p(a_j)}
\]

\[x = \{a_1, a_2 ... a_m\} \text{ is feature vector, } m \text{ is the dimension of the word vector.}\]

KNN algorithm considers that the most neighboring samples around a sample belong to the same category, so this sample also belongs to this category. This algorithm is suitable for irregular and cross-classified sample sets. The algorithm selects the K samples with the smallest distance from the predicted sample and calculates the frequency of the category. The category with the highest frequency is selected as the category of the sample \(x[^{16}]\), and the distance is calculated by the equation (12).

\[
d(x, y) = \sqrt{\sum_{k=1}^{k} (x_k - y_k)^2}
\]

\[x, y \text{ represents the sample, and } k \text{ is the dimension of the sample.}\]

SVM algorithm can analyze the case of low-dimensional linear separability. In the case of inseparability, the nonlinear function is used to map the low-dimensional input space to the high-dimensional space, and then the linear classification is continued. SVM can finally be represented as a convex optimization problem. The known mature algorithm can be used to solve the global optimal solution of the objective function, which is suitable for small and high-dimensional sample sets.
BPNN is a multilayer feedforward neural network trained according to the error back-propagation algorithm, which includes an input layer, a hidden layer, and an output layer. The network structure is shown in figure 4.

![BPNN structure diagram](image)

**Figure 4. BPNN structure diagram.**

BPNN uses ReLU function\(^{[18]}\) as the non-linear excitation function, because the function only needs a threshold to get the activation value, which saves a lot of complex operations and converges quickly.

### 3.3.2. Deep learning based method

Word2Vec neural network is used to extract the word vector automatically and fed to the convolutional neural network to extract the local features. Then, the pooling layer is iterated for many times and finally smoothed into a vector. The vector is fed to the full connection layer and then classified by logistic regression classifier. In CNN\(^{[20]}\), we use multiple filters to capture the local semantics of n-grams with different granularity\(^{[17]}\). \(tanh\) as a nonlinear activation function, and the network structure is shown in figure 5.

![CNN network structure](image)

**Figure 5. CNN network structure.**

The LSTM proposed in this paper is used for the fine-grained sentiment analysis of the Douban film top250 short reviews data set. Only one layer can capture the long-term dependence between sentences. After the input sequence, each hidden state is calculated in order, and there is no need for particularly complex debugging hyperparameters. Compare the LSTM network model with the correct rate of the proposed model on the dataset, and select the best classification model suitable for the dataset.

Firstly, the dataset of Douban Film Top250 short reviews is classified into positive and negative categories. The experimental results are shown in table 3.

| classification algorithm | accuracy  |
|--------------------------|-----------|
| NB                       | 75.1%     |
| KNN                      | 73.4%     |
| SVM                      | 76.9%     |
| BPNN                     | 76.3%     |
| CNN                      | 81.7%     |
| LSTM                     | **83.6%** |
The experimental results show that LSTM has the highest classification accuracy in this data set. Therefore, LSTM is used to classify the data set into positive and negative categories, resulting in 59290 short reviews of positive emotions and 40710 short reviews of negative emotions. Then 59290 positive short reviews are classified as happy and good, and 40710 negative short reviews are classified as fine-grained emotional categories of anger, sorrow, fear, evil and shock. The ratio of the training set and the test set is still 7:3, and the corresponding training set and test set are divided as shown in table 4 below, the training results are shown in table 5.

| binary classification | fine-grained classification | number | number of training sets | number of test sets |
|-----------------------|-----------------------------|--------|-------------------------|---------------------|
| positive              | happy                       | 59290  | 41503                   | 17787               |
|                       | good                        |        |                         |                     |
| negative              | anger                       | 40710  | 28497                   | 12213               |
|                       | sorrow                      |        |                         |                     |
|                       | fear                        |        |                         |                     |
|                       | evil                        |        |                         |                     |
|                       | shock                       |        |                         |                     |

Table 5. Accuracy of fine-grained sentiment analysis.

| model | positive | negative |
|-------|----------|----------|
|       | happy, good | anger, sorrow, fear, evil, shock |
| NB    | 70.8%    | 46.3%    |
| SVM   | 71.8%    | 47.9%    |
| KNN   | 67.7%    | 45.7%    |
| BPNN  | 71.4%    | 47.6%    |
| CNN   | 75.6%    | 50.6%    |
| LSTM  | 76.1%    | 51.2%    |

4. Results and discussion

4.1. Comparison with baseline method

4.1.1. Comparison between traditional models

For the fine-grained sentiment analysis of the Douban Film top250 short reviews dataset, using traditional machine learning methods NB, KNN, SVM, and BPNN as the comparison model. Firstly, the TF-IDF algorithm is used to extract features manually, and then these classifiers are used to realize classification. These models all use the optimal parameters to classify the dataset, and the results are shown in figure 6.
Figure 6. Experimental results of traditional machine learning methods.

Experiments show that the accuracy of SVM classifier is the highest, and it is more suitable for the dataset proposed in this paper, whether it is the binary or fine-grain classification.

4.1.2. Comparison of traditional model and deep learning model

Traditional models use TF-IDF algorithm to extract features, NB, KNN, SVM and BPNN classifiers for classification, while the deep learning models use Word2Vec network model to extract features automatically, CNN and LSTM for classification. The experimental results of these models for the fine-grained sentiment analysis of the Douban film top250 short review datasets are shown in figure 7.

Figure 7. Experimental results of binary and fine-grained classification of six large models

Experiments show that the deep learning method is obviously better than the traditional classification method. It can automatically learn features, update and optimize in the training process, and achieve good classification results with fewer parameters. LSTM has better classification performance than CNN. CNN can capture the local features of sentences but lacks the learning of the semantic relationship between sentences. LSTM can capture long-term dependence in sentences. Therefore, for the Douban Film Top250 short reviews dataset, LSTM can achieve the best classification results for its binary-classification and fine-grained emotional analysis, and the longer the comments are, the better the classification effect of LSTM is.
4.2. Impact of the dataset
The fine-grained sentiment analysis of the Douban film top250 short reviews dataset is based on its binary classifications. The second classification is divided by the classifier, and there will be errors. The training set we extracted is relatively small, which will affect the accuracy of classification results. In order to reduce human error, we can use the remaining data sets to verify. In order to further study fine-grained affective analysis, more data can be obtained for experiments. Although the LSTM model proposed in this paper is more suitable for this dataset, each classification model has its own advantages. The LSTM model is only the most effective for the Douban film top250 short reviews dataset, not necessarily applicable to other datasets. It is not a specific model for all datasets.

5. Conclusions
This paper obtains the Douban Film Top250 short reviews dataset, which contains comments of different stars. LSTM model is proposed to implement fine-grained sentiment analysis of the dataset. First, the acquired data set is preprocessed, network neologisms and expression symbols are added to the emotional dictionary, which is used for words segmentation and stop words. After preprocessing, sentences are represented as word vectors by CBOW model in Word2Vec network, and then the word vectors are sent to LSTM for training and finally the logistic regression classifier is used for classification. The fine-grained sentiment analysis of the dataset is divided into two steps. The first is the binary classifications of positive and negative emotions, and then the best classification results are analyzed for fine-grained emotions. By comparing with NB, KNN, SVM, BPNN and CNN models, the LSTM model proposed in this paper is the most suitable for this dataset, whether it is binary or fine-grained classification. The experimental results show that the traditional machine learning method and deep learning method can achieve good classification effect for the binary classifications, but for the fine-grained sentiment analysis, it is not very ideal, because the complexity of Chinese leads to certain difficulties for multi-classification. The network updates rapidly, and a large number of new words are constantly appearing, which also poses challenges for classification. In order to conduct fine-grained emotional analysis of film reviews more quickly and accurately in massive texts, we can consider using more complex network models and adding emotional dictionaries to improve classification performance.

Acknowledgments
This work was supported by the National Key Research and Development Program of China (Project No. 2017YFC1601800)

References
[1] Pang, B., Lee, L., & Vaithyanathan, S. (2002) sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing, Volume 10: 79-86.
[2] Y. Kim. (2014) Convolutional neural networks for sentence classification. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing, 2014: 1746–1751.
[3] Zhang, X., Zhao, J., & LeCun, Y. (2015) Character-level convolutional networks for text classification. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing. In Advances in neural information processing systems, 2015: 649-657.
[4] Paltoglou, G., & Thelwall, M. (2010). A study of information retrieval weighting schemes for sentiment analysis. In Proceedings of the 48th annual meeting of the association for computational linguistics. Uppsala.pp. 1386-1395.
[5] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013) Efficient estimation of word representations in vector space. arXiv preprint arXiv , 2013: 1301-3781.
[6] Wang, J. H., Liu, T. W., Luo, X., & Wang, L. (2018). An LSTM Approach to Short Text Sentiment Classification with Word Embeddings. In Proceedings of the 30th Conference on Computational Linguistics and Speech Processing (ROCLING 2018). China, pp. 214-223.

[7] Chen, J. M., 2008. The construction and application of Chinese emotion word ontology. http://ir.dlut.edu.cn/news/detail/215

[8] Bengio Y, Ducharme R, Vincent P, et al. (2003) A neural probabilistic language model. Journal of machine learning research, 3(Feb): 1137-1155.

[9] Liu P, Qiu X, Huang X. (2016) Recurrent neural network for text classification with multi-task learning. arXiv preprint arXiv,1605:05101-2016.

[10] Bengio Y, Simard P, Frasconi P. (1994) Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks, 5(2): 157-166.

[11] Wang X, Jiang W, Luo Z. (2016) Combination of convolutional and recurrent neural network for sentiment analysis of short texts. the 26th International Conference on Computational Linguistics: Technical Papers, 2016: 2428-2437.

[12] Bridle J S. (1990) Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. Neurocomputing. Springer, Berlin, Heidelberg.

[13] Bottou L. (2012) Stochastic gradient descent tricks. Neural networks: Tricks of the trade. Springer, Berlin, Heidelberg.

[14] Leung K M. (2007) Naive bayesian classifier. Polytechnic University Department of Computer Science/Finance and Risk Engineering, 2007: 123-156.

[15] Tang D, Qin B, Liu T. (2015) Document modeling with gated recurrent neural network for sentiment classification. Proceedings of the 2015 conference on empirical methods in natural language processing, 2015: 1422-1432.

[16] Deng Z, Zhu X, Cheng D, et al. (2016) Efficient kNN classification algorithm for big data. Neurocomputing, 195: 143-148.

[17] Tang D, Qin B, Liu T. (2015) Document modeling with gated recurrent neural network for sentiment classification. Proceedings of the 2015 conference on empirical methods in natural language processing, 2015: 1422-1432.

[18] Glorot X, Bordes A, Bengio Y. (2011) Deep sparse rectifier neural networks. Proceedings of the fourteenth international conference on artificial intelligence and statistics, 2011: 315-323.

[19] Wang B. (2018)Disconnected Recurrent Neural Networks for Text Categorization. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Volume 1: 2311-2320.

[20] Conneau A, Schwenk H, Barrault L, et al. (2016) Very deep convolutional networks for text classification. arXiv preprint arXiv, 2016: 1606.01781.