UDK 519.62

D.S. Nazarenko[1], I.V. Afanasieva[2], N.V. Golian[3]

[1]Student, Software engineering,NURE, Kharkiv, Ukraine, dmytro.nazarenko@nure.ua

[2]PhD, assistant professor, Software engineeringNURE, Kharkiv, Ukraine, iryna.afanasieva@nure.ua

[3]PhD, assistant professor, Software engineeringNURE, Kharkiv, Ukraine, nataliia.golian@nure.ua

NEURAL NETWORK APPROACH FOR EMOTIONAL RECOGNITION IN TEXT

The article is devoted to one of the most popular trends in the field of IT today – natural language processing, in particular, the extraction of emotions from the text using the neural network approach. The main task was to solve the problem of the high costs of time and human resources for companies to receive feedback from users and process emotional reactions of the second one. That to decide the task it was necessary to make modelling and learn neural network using own architecture based on the backpropagation algorithm that to recognize the emotional component in the text.The emotional component of reviews was used as a metric for evaluating user reactions. It was decided to work with five types of emotions that will help to provide better results. The neural network architecture consists of interconnected layers: embedding, bidirectional LSTM, pooling, dropout layers and two dense layers.

For the neural network learning was selected an open dataset consisted of 47,288-tagged posts from Twitter. As a result, the F-measure on the test dataset was 0.62 and which is a worthy indicator in comparison with large business solutionsю.

CLASSIFICATION, DATA, EMOTIONS, NEURAL NETWORKS, RESPONSES, TEXT, TONALITY

Introduction

In the modern world, automation has affected all spheres of human life. It is impossible to imagine life without computer technology: everything around us is connected with the IT industry. Humanity is evolving and the IT industry is developing too.

People generates huge number of data every day. It is can not be unnoticed that finally led to the emergence of a science about big data [1].

It is important to understand that the data cannot be benefit by themselves. In addition, it is important to use them in real problems. For example, all posts and photos in social networks, queries in search engine, and the benefits of choosing videos and music, and this is not an empty data set, but an important digital resource. It was the big data pushed to a revolution in the field of computer sciences – the development of artificial intelligence and machine learning [2].

A new round of software evolution is just beginning, but today you can feel its impact, which lead to the development of products built based on relevant data usage approaches will soon.

1. Analysis domain

Information technology is changing the nature of modern business. Companies are trying to automate internal processes to reduce the cost of maintaining a large staff and speed up work. Today, the number of spheres that non-automated are decreasing.

Huge amount of text data forces business representatives to hire staff for data processing. Companies that are not capable of holding a large staff who will make market analysis cannot cope with the problem of text data analysis.

The problem is actual for both small and large businesses, because nobody is interested in additional costs.

However, not all business processes are able to automation, especially in areas of data processing and analysis where human skills are needed. These areas include face recognitions in photos and real-time video [3], text processing [4].

For such tasks, decisions should use actual approaches in the world of information technologies – neural networks.

Neural networks are built on the principle of the human brain that allows them to be trained on the provided data [5].

So feedback from users to business is needed.

2. Problem Statement

The main task is to solve the problem of the high costs of time and human resources for companies to receive feedback from users and process emotional reactions of the second one using the neural network approach.

It is necessary to make modelling and learn neural network using own architecture based on the backpropagation algorithm [6] that to recognize the emotional component in the text.

That to prepare training data it is necessary to make data collection. First, to choose the text data that will be used for learning neural network, as well as its validation.

Also it is important to improve the quality of the learning process to make comprehensive data pre-processing. The dataset should be divided into a training one, which will be used during the training process and a test one, it is necessary to make a conclusion about the ability of the neural network generalization [7].

The key criterion of the neural network operation is the accuracy of emotion classification [8]. To do this, it is important to choose a metric that will help to evaluate the classification accuracy. The F1 score metric will allow to get a more realistic measure of the accuracy classifier.

Potential customers of this application will be companies of medium and large sizes, which are engaged in sales and provision of services. An important advantage for companies will be to reduce staffing in order to solve these types of problems and in particular to analyze the reaction of consumers of a product or company's services on the Internet.

3. Decision making

It is necessary to use various methods for text analysis that to receive user reviews on the certain product. It is required to understand how products satisfy the users and what emotions they feel.

The emotional component of reviews is used as a metric for evaluating user reactions. This metric allows you to determine the overall customer attitude to the product. Usually reviews are divided into 3 categories: positive, negative and neutral [9,10-13].

If to divide all reviews into positive and negative, we will lose information that is useful for companies. In real life, comments and reviews are rarely absolutely positive or completely negative. The emotional colour of a comment has a whole palette of emotions. According to the spectrum of emotions, we can study the reaction of the market in detail and make conclusions [8].

However, some hidden emotions are difficult to detect even for person. Specific and unpopular emotions do not have a big impact on the overall user reaction, because of business is interested in indicators of the consumer market as a whole, and not of the user individually. Therefore, it is necessary to choose the basic emotions that it makes sense to extract but the result of the analysis will have business value.

Selection of emotions from the text should not be limited by binary classification [9]. If we will divide text data only to positive and negative classes, then the results of the analysis will lose flexibility and will not sufficiently characterize what emotions the users had exactly.

Nevertheless, during research the number of emotions should not exceed seven [8], otherwise, there will be an information that does not have an impact on the user reaction.

Thus, several classes should be extracted for classification, since binary classification entails a lack of informativeness in the results. In addition, an information overload should be avoided by optimizing the learning process of neural network on some layers.

As the investigated emotions were chosen: 1 – happiness; 2 – sadness; 3 – anger; 4 – aggression. They have good informational content in tasks of comments and feedback processing from users. Types of emotions are common and well expressed in text messages. It is suitable characteristic for user reactions analysis. Also, we decided add fifth emotion – neutral, because of not brightly expressed emotions (unemotional colouring of the text).

It was decided to work with five types of emotions that will help to provide better results.

4. Neural network architecture

The neural network architecture was created. Our interconnected layers are consist of embedding, bidirectional LSTM, pooling, dropout layers and two dense layers, so let us describe each layer more detail.

In general, embedding is a form of words representation [14]. It helps to connect human and machine in language understanding. This type of layer provides distributed representations of text in n-dimensional space. Word embedding is a family of natural language processing methods aimed at mapping semantic meaning into geometric space. It is implemented by associating a numerical vector with each word in the dictionary, so that the distance between any two vectors encompasses part of the semantic relations between two related words. For example, "coconut" and "polar bear" are words that semantically rather different, but embedding will present them as vectors that will be located at a great distance from each other; but the words "kitchen" and "dinner" are related words, so they will be located closer to each other.

Word embedding is computed by applying dimension reduction methods to co-occurrence datasets between words in the body of the text and implemented using the GloVe approach.

The next layer is bidirectional long short-term memory (bidirectional LSTM). LSTM is a type of recurrent neural network capable of learning long-term dependencies [15]. LSTMs are designed to eliminate long-term dependency problems and the specialty is to memorize of information for long periods of time.

A bidirectional recurrent neural network (BRNN) [16] connects in opposite directions two hidden layers with the same input. The output layer of the recurrent neural network can receive information from past (previous) and future states (next) simultaneously due to the form of generative learning. BRNN has been created to increase the amount of incoming information available to the network, as shown in Figure 1.

For example, multilayer perceptron (MLP) [17] and time delay neural network (TDNN) [16] have limitations on the input data flexibility, because of requirements of the input data recording. The standard recurrent neural network (RNN) [15] also has a limit, since future input information from the current state can not be reached. On the contrary, the BRNN does not require the recording of input data. Moreover, future input information is available from the current state. It allows future data to be taken into consideration during training and improve the train ability of the model. BRNN is especially useful when the presence of the context of the input data improves the result. For example, accuracy can be enhanced by the words that placed consecutive in the sentence when we are recognizing word by context.

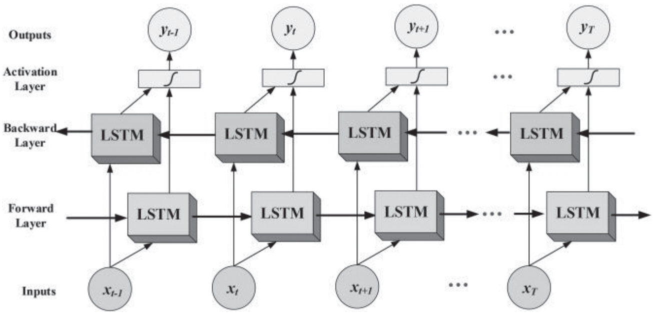

Fig. 1. BRNN scheme

The next layer is pooling [18] is a non-linear compaction of matrices with numbers. Numbers in matrices are words that pass non-linear transformation. The most used function is the max function in GlobalMaxPooling [18]. Therefore, transformations affect non-intersecting rectangles or squares, where each is compressed into a number that has a maximum value. The pooling can significantly reduce the spatial data volume and is interpreted as follows: if at the previous operation some data attributes have already been identified, then for further processing such detailed data are not use, and they are condense to less detailed. In addition, unnecessary parts filtering helps the neural network not to be retrain.

Than the training data are combined and transferred to a dense layer. Moreover, further layers do not have spatial structure, but have a relatively small dimension to return the final result of the classification.

To improve the learning quality of the model, the dropout layer was included in front of the output layer [19]. The dropout is a method for regularization of artificial neural networks, and the goal is to prevent overfitting. The essence of the method is that in the process of learning from the neural network subnet is randomly allocated and training for subnet is provided. The training subnet comes from excluding of neurons from the full original network (dropping out) with probability p, thus the probability that the neuron will stay in the network is $q = 1–p$. In the process of learning the excluded neurons do not contribute to any stages of chosen backpropagation algorithm, therefore excluding at least one of the neurons is equivalent to learning new neural network.

It was decided to make the output layer as a fully layer with the five neurons equal to the number of classified emotions. The Softmax function was chosen as the neuron activation function [20]. The Softmax function is a generalization of the logistic function for the multidimensional case. The function converts a vector z of dimension K into a vector σ of the same dimension, where each coordinate σ_i of the obtained vector is represented by a real number in the interval [0,1] and the sum of the coordinates equal 1. Softmax is given by the following formula:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{k=1}^{N} e^{z_k}},$$

where z_i is the value at the output of the i-th neuron before activation, but N is the total number of neurons in the layer. The Softmax function is used in machine learning for classification problems when the number of possible classes are more than two. The coordinates σ_i of the resulting vector are interpreted as the probabilities that the text belongs to class i.

As a loss function for this model, it was decided to choose categorical cross entropy [21]. Categorical cross entropy is used to categorize one label and it means that only one category for each data point is applicable. In other words, a data sample can belong to just one class. Categorical cross entropy compares the distribution of predictions, that is, activations in the output layer, one for each class, with the true distribution, where the probability of the true class is set to 1 and 0 for other classes. In other words, the true class as a one-hot encoded vector is represented, and the closer the model outputs to this vector, the less the loss. Categorical cross entropy is used in conjunction with the Softmax activation function.

5. The neural network training

For the neural network learning, an open dataset from the github repository was selected [22]. The dataset consist of 47,288-tagged posts from Twitter in English. The data set was chosen due to the large number of textual data samples with already labelled classes of emotions. The mark-up was made in five classes: joy, sadness, anger, aggression and a neutral class of emotions to designate text samples where the emotional component is not brightly expressed. The number of examples of each class are neutral – 9643, happy – 16297, sad – 15938, hate – 4301 and anger – 1109.

The dataset was pre-processed where we used normalization and lemmatization [23]. During normalization- the texts were cleaned of punctuation marks, all letters were switched to lowercase. Then lemmatization was performed – the words were reduced to their normal form, and the stop words were deleted.

The next step of data preparation for training was to bring data samples to a form that would be convenient for using them as input parameters of a neural network. So using the tools of the Keras library [24] text was vectorized. Each word in the text was associated with a numeric index in the dictionary. As a result, each data sample represented by a vector of numbers. Each vector was supplemented with zeros to a constant length that the length of the text does not affect the final ability of the neural network to generalize.

The vectorized data representation then was split into two subsets – train and test datasets, in a ratio of 3 to 1. The train dataset was used at the entire training stage, and the test dataset was used to evaluate the quality of the model prediction on data that in the training were not involved.

So implementing the model, the Keras library was used [24] and own neural network architecture was created and used.

The vector of tokens of constant length was fed to the input of the model. The sequence of tokens was transferred to the Embedding layer. As a layer of embedding, it was decided to use the pre-trained word vectors GloVe [25]. GloVe is an unsupervised learning algorithm for generating vector representations for words. The training is performed on aggregated global word-match statistics from the corpus, and the resulting representations demonstrate linear substructures of the vector word space. The GloVe model is trained on non-zero elements of the global word match matrix and shows how often the words occur with each other in a given corpus. This matrix requires one pass through the entire corpus to collect statistics. Subsequent training iterations are much faster, because of number of non-zero elements of the matrix is usually much smaller than the total number of words in the body. However, in this model have already used filled matrix with ready-made word vectors, since the pre-trained GloVe model was used.

The output data of the embedding layer is a matrix of 30 by 200, where each word is associated with its vector representation and, the data passes through the layer of a bidirectional recurrent neural network. The input sequence is served in the usual order of time for one network and in the reverse order of time for another. The outputs of the two networks at each time step are combined. This structure allows the network to have both reverse and direct information about the sequence at each time step. At the result the output, we have less matrix of 30 to 64.

The GlobalMaxPooling1D approach [18] for time data takes the maximum vector for measuring steps.

Thus, a multidimensional table with a form [10, 4,10] becomes a global multidimensional table with a form [10, 10] after merging.

Suppose we have a simple sentence with 3 words, and some vector representation of these words. In the case of GlobalMaxPooling1D, the maximum vector of this sentence is taken.

The pooling layer helps get rid of data redundancy, which allows the neural network to take up less memory and learn faster. Than output data passes through two fully connected Dense layers, between which we disable 5% of random neurons and it helps the model to avoid overfitting. At the same time, the last Dense layer in the neural network contains 5 neurons that equal to the number of classes of emotions. Each neuron has a Softmax activation function. Figure 2 shows the final architecture of the designed neural network.

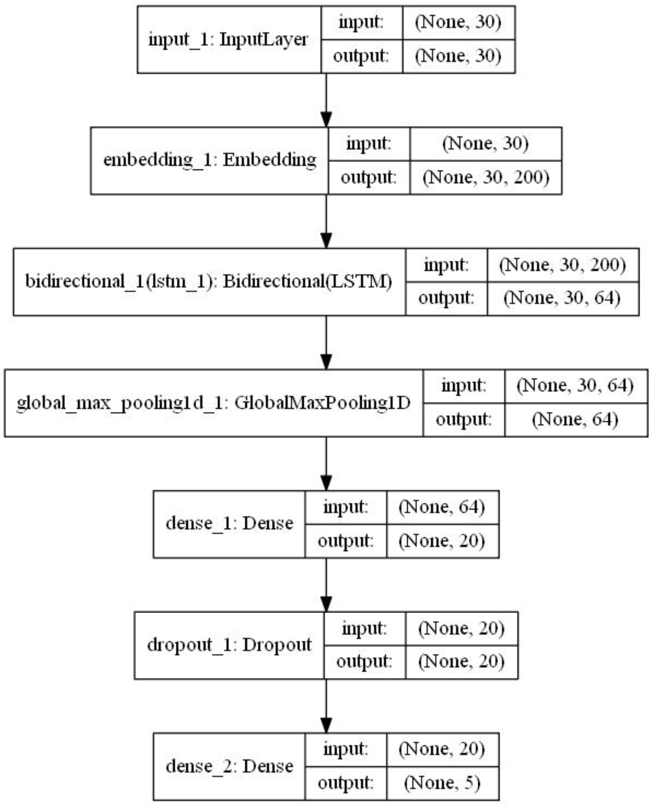

Fig. 2. Neural network architecture

As a result, the neural network is trained using the backpropagation algorithm [6]. Training by this method involves two passes through all layers of the network: direct and reverse. The input vector is fed to the input layer of the neural network with a direct pass. Than input vector propagates through the network layer by layer. As a result, a set of output signals are generated and gives the actual response of the neural network to this input image. During the direct pass, all synaptic weights of the network are fixed. During the back pass, all synaptic weights are adjusted in accordance with the error correction rule: the actual output of the neural network is subtracted from the desired and as a result is an error signal. This signal subsequently propagates through the network in the direction opposite to the direction of synaptic connections, hence the name – algorithm of backpropagation. Synaptic weights are adjusted to maximize the output signal of the network to the desired.

The backpropagation algorithm is as follows:

1. Initialize synaptic weights with small random values.

2. Choose the next training pair from the training set; submit the input vector to the network input.

3. Calculate network output.

4. Calculate the difference between the network output and the required output (the target vector of the training pair).

5. Adjust network weights to minimize errors.

6. Repeat steps 2-5 for each vector of the training set until the error on the entire set reaches an acceptable level.

In the process of training to assess the accuracy of the neural network the accuracy metric was used. It is interprets as the proportion of correct answers.

6. Results

As a result, at the testing stage, the resulting model was evaluated on a test dataset. It is consisted of 9457 text records from the Twitter microblogging service. As a metric, the F-measure was used – average harmonic of precision and recall. Precision is the proportion of correctly predicted instances among all found, and recall – the proportion of correctly predicted instances relative to the total number of relevants. The F-measure on the test dataset was 0.62 and which is a worthy indicator in comparison with large business solutions and other solutions of the problem.

References

[1] *Майер-Шенбергер В.et al.* Big Data. A Revolution That Will Transform How We Live, Work, and Think / пер. сангл. Инны Гайдюк.–М.: Манн, Иванов, Фербер. –2014. – 240 с.

[2] *Флах П.* Машинное обучение. – М.: ДМК Пресс.– 2015.– 400 с.

[3] *Друки А.* Система поиска, выделения и распознавания лиц на изображениях// Известия Томского политехнического университета [Известия ТПУ]. – 2011. – Т. 318, № 5. – С. 64 -70.

[4] *Zhang L., Wang S., Liu B.* Deep Learning for Sentiment Analysis: A Survey// Wiley Online Library.– 2018. –34 p.

[5] *Lawrence, Jeanette.*Introduction to neural networks: design, theory and applications // Nevada City: California Scientific Software.–1994. – 324 p.

[6] *Goodfellow I., Bengio Y., Courville A.* Deep Learning. // MIT Press. –2016. – 196 p.

[7] Bousquet O., Elisseef A. Stability and Generalization // Journal of Machine Learning Research.–1992. – P. 499-526.

[8] *Назаренко Д. Афанасьєва І.* Аналіз сервісів щодо розпізнавання емоційної складової відгуків користувачів. Актуальные вызовы современной науки// Сб. научных трудов – Переяслав-Хмельницкий.– 2019. – Вип.4(36), ч.1 – С. 85-90.

[9] *Read J. et al.* Classifier Chains for Multi-label Classification// Machine Learning Journal. Springer. –2011.–Vol. 85(3). –27 p.

[10] *ChoiY., Lee H.* Data properties and the performance of sentiment classification for electronic commerce applications//Information Systems Frontiers. Springer. – 2017. – Vol. 19, Issue 5. – P. 993–1012

[11] *Dashtipour K. et al.* Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques// Cognitive Computation. Springer. – 2016. –Volume 8, Issue 4. –P. 757–771

[12] *Shuhaida Mohamed Shuhidan et al.*Sentiment Analysis for Financial News Headlines using Machine Learning Algorithm//International Conference on Kansei Engineering & Emotion Research. KEER. – 2018.–P. 64-72.

[13] *Souma W.et al.* Enhanced news sentiment analysis using deep learning methods//Journal of Computational Social Science. –2019. – Volume 2, Issue 1. P. 33–46.

[14] *Goldberg Y., Levy O.et al.*Word2vec Explained: Deriving Mikolov et al.'s Negative-Sampling Word-Embedding Method. – 2014. – 5 p.

[15] *Hochreiter S., Schmidhuber J.* Long Short-Term Memory// Neural Computation.–1997. –№9 (8). – P. 1735–1780.

[16] *Graves A. et al.* Bidirectional LSTM networks for improved phoneme classification and recognition. Artificial Neural Networks: Formal Models and Their Applications// ICANN. Springer Berlin Heidelberg. –2005. –P.799-804.

[17] *Rosenblatt F.* Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms// Spartan Books, Washington DC. – 1961.–

[18] *Graham B.* Fractional Max-Pooling.– 2015. –10p.

[19] *Hinton G. et al.* Improving neural networks by preventing co-adaptation of feature detectors– 2012. –18p.

[20] *Gao Bolin, Pavel Lacra.* On the Properties of the Softmax Function with Application in Game Theory and Reinforcement Learning. –2017–10p.

[21] *Murphy K.* Machine Learning: A Probabilistic Perspective// MIT Press. – 2012. –P. 57-571

[22] *Liu T.*Text emotion classification dataset. –2018.

[23] *Airio Eija.* Word Normalization and Decompounding in Mono- and Bilingual IR//Information Retrieval. – 2006.– Vol. 9, Issue 3. P. 249–271.

[24] *Ketkar N.* Introduction to Keras. In: Deep Learning with Python//Apress, Berkeley, CA – 2017. – P. 97-111

[25] *Pennington J. et al.* GloVe: Global Vectors for Word Representation// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) – 2014. – P.1532-1543.

The article was delivered to your editory stuff on the 23.05.2019