A Comparative Analysis of Long Short-Term Memory Recurrent Neural Network Classifiers for Text Classification

Saeed Ullah*, Jamal Uddin†

Abstract

The classification of news text is hard due to tremendous variations in interpretations of text and events such as various groups and vast volume of text contents. Text recognition and classification are challenging research areas in artificial intelligence and machine learning. Deep learning is a type of machine learning that is widely used to build the recognition method for human behavior and text classification systems. A popular deep learning technique Recurrent Neural Network is used to identify the points of interest and remove the attributes from the text for the identification and classification of the human language. The Recurrent Neural Network is a form of neural network in which the output from the previous stage is fed to the current phase data. All inputs and outputs are independent of each other in standard neural networks. It is important to determine next term in a sentence in some situations; the prior words are needed to recall the previous words. Hence, Recurrent Neural Network came in with the aid of a hidden layer, which solved this problem. The hidden state is the primary and most significant function of Recurrent Neural Network. The Recurrent Neural Network is based on classifiers that use combination of layers. Long Short-Term Memory, Bi-Directional Long Short-Term Memory, Deeper Long Short-Term Memory and Deep Bi Directional Long Short-Term Memory are all Recurrent Neural Network layered base classifiers. They store the previous data block as well as the new data block and merge it into one row as output. In this research, these classifiers are analyzed for a better, generalize and efficient Long Short-Term Memory Recurrent Neural Network text classification classifier. For this purpose, the BBC news platform is selected for text classification. Experiments are conducted on mentioned classifiers and the simulation was done through MATLAB. The comparative findings are presented in form of tables and are analyzed for exploring a better classifier. It is also observed that different evaluation parameters had an effect on the process of text analysis. Deep Bi Directional Long Short-Term Memory has proven its efficiency and better performance comparatively for text classification on performance evaluation measures like Accuracy (74%), True Positive rate (90%), and False Positive rate (6%), Recall (90%), Precision (75%) and F-measure (74%).

---

*Planning and Development Department, Government of KP & saeedullah845@gmail.com

†Qurtuba University of Science & Information Technology, Peshawar & jamuddin1983@gmail.com

Keywords: machine learning, deep learning, recurrent neural network, convolution neural network, long short-term memory.

Introduction

Text classification is the process of splitting text into sequential groups. To flush out irrelevant document from a large-scale corpus, text classification can be considered as a preprocessing technology. The role of classifying texts is to assign documents to predefined subjects, such as politics and sport. Text classifiers are used to analyze the text automatically and then assigned label to content of the text. The origin of the grouping of text dates back to the early '60s. Since the 1980s, text classification (or text categorization) has been widely studied by many researchers. From last 20 years, printable data was generated and which was used in word preparation programs. Some Text classification applications are automatic handwriting generation, language translations, health care and Natural Language Processing (NLP). Machine learning has been mostly used in NLP applications, because it is the way for understanding the changing structure of a sentence and the text structure of different arguments. Machine learning is an informatics field which gives computers the ability to learn without explicit programming. Training of the program is conducted with the aid of data derived information.

Machine Learning methods were successfully applied to text classification in the late 90s. In 1999, Vector Machines are introduced to text classification. In recent years, multi-label classification of multi-topic text has been explored. Investigating data for identification of specific data models in order to assist for correct estimates and examination is done through classification. Many classification approaches like Decision tree, Naive Bayes, and Support Vector Machine (SVM) have been in existence through the struggle of researchers involving data classification procedures for specific text characteristics. Neural Networks field emerged from understanding multiple numbers of outlets, yet simulating the human brain despite larger questions over cloning individual qualities such as speech. Neural networks are categorized into internal connected node layers, in which each node produces a non-linear influence from its inputs and outputs. Deep Learning is concerned with algorithms inspired by the structure and functions of the brain called artificial Neural Network. It is a subfield of machine learning that represents data and is extremely good at patterns of learning which means that it works on sequential and labeled data. Deep learning techniques have demonstrated much better results for analysis of communication in form of text. In addition, these techniques are capable of showing a deeper grading of arrangements in neural networks to represent the data. In various areas such as, applications, speech recognition, configuration recognition, and data arrangement, complex structure can learn more easily depending on way.

For example, Deep Learning Model is used for sentiment analysis in conjunction with the SVM Classifier for text related data (Y. Chen & Zhang, 2018). Further text, in text processing system is categorized using text classification problem of machine learning. Deep Learning is also used to capture hard-to-model linguistic concepts such as denials and mixed emotions, the current state of the art for sentiment analysis (Yin et al., 2017). Deep learning algorithms attempt to learn (multiple levels of) representation by using a multilayer hierarchy. Another machine learning algorithm Naive Bayes is used for text classification themselves. It is achieved by removing key-phrases or ad-hoc sentences leaving the sentences unchanged. While abstraction involves paraphrasing the context-conscious sentences after a language understanding. A text categorization system is developed for Arabic language (Goudjil et al., 2013). First of all, preprocessing was performed on the Arabic language data for the steps of text classification. Then extracted functions are classified through Support which discusses the applications such as text analysis and computer vision (Buzic & Dobsa, 2018)

Two classes of Deep Learning are Convolutional Neural Network (CNN) and Recurrent Neural Networks (RNN) (Young et al., 2018). The CNN is mostly designed for images and videos. RNN models work on sequences rather than features, sequences examples are text and signal data. In case to classify features using RNN, first the feature should be converted to sequences. A special kind of RNN architecture is Long Short Term Memory (LSTM), having capacity to learn long-term dependencies (Lipton et al., 2015). It is the influence of a given input on the hidden layer, and also on the network output. Though, several deep learning techniques particularly LSTM based were successfully applied for text classification. However, using simple and low performance classifier or similar method can also lower the ability and precision of the high-scale text classification. To overcome this issue, the outcomes from different deep learning-based LSTM classifiers for text classification need to be further analyzed for a real news data set in terms of traditional evaluation parameters of classification process. Moreover, the effect of increasing instances on the performance of LSTM based classifiers is also required.

Several RNN based classifiers like Long Short-Term Memory, Bi-Directional Long Short-Term Memory, Deeper Long Short-Term Memory and Deep Bi Directional Long Short-Term Memory are developed and proved successful individually. Accordingly, in this research we will compare them altogether against same data set. This empirical comparison provides the opportunity to adopt a more general and better LSTM classifier. The different variants of Recurrent Neural Network (RNN) Classifiers for Text Classification are comparatively analyzed. This research determines empirically which one among these is more efficient in term of accuracy, precision, recall and other performance measures. For this purpose, specific dataset was developed for the classification of text obtained from BBC Website (www.bbc.com/news) having no missing data and data extraction. The text data is preprocessed using stop word removal, stemming and Bag of Words (BOW) to get strong structures from written documents. This text classification system automatically and efficiently recognizes whether a text is related to the describe classes of politics, religion, social media, entertainment and sports. The present research also explores the effect of increased data on performance improvement of text classification system through utilizing and manipulating the RNN's multi layered deep features. This significant ability of deep learning approach that the increase in data instances improves overall performance is due to utilization of large number of epochs, bitch size and other parameters.

Literature Review

Text Classification

Text data is also a sequential data in which features contents can be easily understood. RNNs are used to interpret these sequences but it forgets the inputs earlier to form a long-term sequence. Such problem is called vanishing gradient problem, which is solved using special RNN known as LSTM. LSTM learns long term dependencies through structure with input, output, and forget gates that handle the long-term sequence recognition. For classification of text data, LSTM based RNN was trained which was used to predict/classify text data into its corresponding classes. Convolutional Neural Networks (CNNs) using deep feed forward networks is used to study local characteristics from arguments or phrases, while Recurrent Neural Networks (RNNs) are capable of learning sequential dependencies in sorted data(labeled) (Wang et al., 2018). Specifically, RNN are continuously used for analyzing the text related data, displacing networks for feed forward (Hochreiter & Schmidhuber, 1997). For obtaining large amount of beneficial features, the concept of spread representation of arguments is used in which each input is denoted by several kinds and each kind is related in several possible inputs.

In the text processing system, text can be categorized in order of the classification questions in machine learning. Naive Bayes is classification algorithm successfully utilized in applications such as document summing, computer vision (Maheswari & Priya, 2018). They predict a singer's success based solely on lyrics using the Naive Bayes classification algorithm. The classification of medical text (bio medicine) is done using three up to date classificatory for machine learning, i.e. KNN, the SVM and Naïve Bayes (Silalahi et al., 2018). In terms of precision, Naïve Bayes classifier gets better accuracy. The SVM and other Multi perceptron Neural Network classifier is also used for the analysis of consumers and dependent on consumer purchases groups. Different collection strategies of apps have been applied and tested on six separate test text databases in order to enhance the efficacy and consistency of the text recognition system (Asim et al., 2018). Comprehensive computational analysis is conducted, and the efficiency of the classifier is measured using the output parameter F1-score.

A structure is used to identify more than two groups using SVM and allow it comfortably to use a binary or more classifier to solve multiple text class problems (Lin & Wang, 2014). A benchmark dataset was used to train the classifier and obtain output through the source of internet for a more reliable categorization of consumers and divided them into 6 grades. It can help for the categorization and identification of vector machine and KNN classifiers.

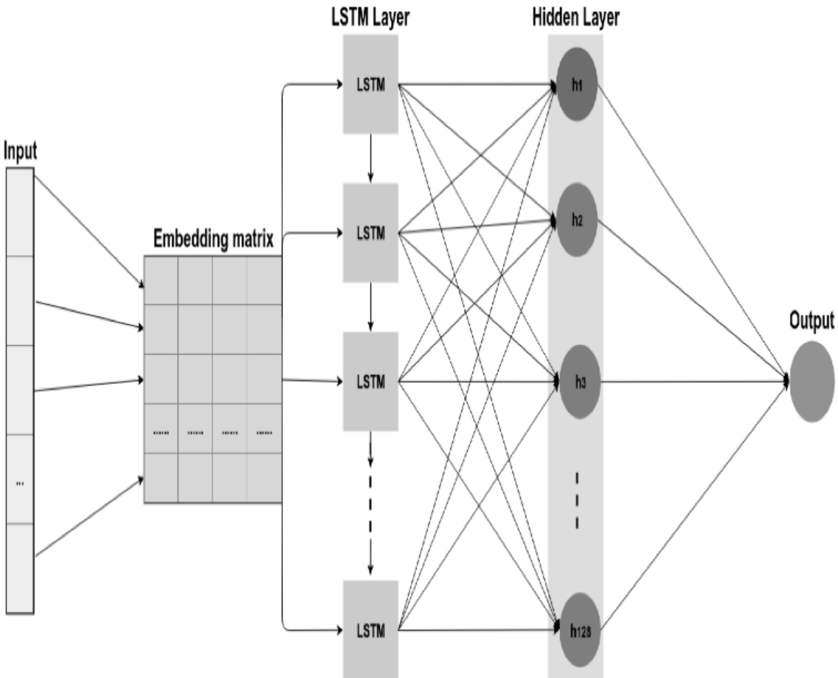
The data related to text are processed using Hidden Markova model by incorporating features section method (D. Chen & Liu, 2010). This classification approach uses two or more classifiers for solving multiple class problems for solution of multiclass problems. Experimental tests reveal that using Naïve Bayes and SVM classifier, the HMM based SVM approach for text categorization achieves a higher precision than independently. The categorization of Arabic languages framework is being developed (Goudjil et al., 2013). Prior to other steps, the data of language were pre-processed for stop-word removal, trailing, and tokenization. The derived functions are listed through the use of SVM classifier. The proposed method achieved better precision and the findings contrasted were in the same domain with certain updated research work.

A text recognition framework is developed using CNN model combined with Support Vector Classifier created to conduct an overview of emotions focused on data from email (Rhanoui & Mikram, 2019). The data extraction tool used for text shaping applications is Word2vec which needs little human input, like just only word meaning functionality. The suggested system's experimental findings gain better precision than traditional CNN's. An automatic Chinese-language sentiment recognition system is being developed using various stop word removal methods (J. Zhou et al., 2019). They pursued for describing the successful stop-word-removal strategy for classifying the file. The experiment is based on the sentimentality remarks that were grabbed on the Site, features removal is done through SVM classifier.

Long Short Term Memory (LSTM)

LSTM was developed by Jurgen Schmid Huber and Sepp Hochreiter. They made the findings on this subject in their thesis of 1991, but LSTM started in 1997. Their recurrent network worked very well compared to previous RNNs. The architecture of RNN LSTM classifier is presented in Figure 1.

Figure 1: The architecture of LSTM classifier



Long Short-Term Memory (LSTM) is a common approach to enhance RNNs ability to store progressive information over a longer period. Widening and adding layers can increase the capacity of an

LSTM network. RNN provides base for LSTM networks since the 1980s. Their architecture permits information to be accumulated during operation, and for remembrance of feedback in network call of previous state. There are a variety of drawbacks to the classical recursive networks that restrains their ability to solve more complex issues. With the advent of LSTM the world again started to pay them special attention as this technique performs much better than previous recursive networks. Each time we use this technique we alert the same cell, changing the state of its internal structure. The most important elements of LSTM cells include;

1. Cell state: the cell state transferred sequentially after the next steps
2. Forget gate: the gate which determines which information should be omitted
3. Input gate: a gate which determines what should be forwarded to the next activation

In a traditional RNN during the gradient back-propagation phase, the gradient signal can end up being multiplied a large number of times (as many as the number of times steps) by the weight matrix associated with the connections between the neurons of the recurrent hidden layer. This means that the magnitude of weights in the transition matrix can have a strong impact on the learning process. LSTM is debated in social media for the grouping of short texts with distributed representation. It uses word injecting model for terms including short texts as vectors, based on Word2Vec.

LSTM is recycled to show long-term, small-level dependences across word sequences. The last outcome from the last point in time is used as the expected result. Production matches the study in LSTM with Naïve Bayes (NB) technique and Extreme Learning System (ELM) in these tests on several social datasets for classification of emotions. The experimental analysis indicates, the proposed approach can produce greater results with vast volumes of training data than traditional probabilistic model and artificial Neural Networks (ANN). It indicates the potential of utilizing deep learning approaches for an interpretation of emotion. To check the utility of the proposed solution at a larger scale, further analysis is needed. ANN is a construction of a network supported by human brain neurons. Nodes are rendered into layers, and in head-to - head layers, edges link nodes. A feed forward process is used to calculate and errors are recorded back to earlier layers for weight management of the compliant sides.

Extreme learning machines (ELMs) are a given type of ANN in which the weights are not acquainted with back propagation (Huang et al., 2006). The secret nodes are allocated randomly, and not modified. So the weights are taught often in a single stage, and also operate even quicker. Deep learning techniques are allowed for more complicated relationships which use several secret layers. Mostly it requires some analytical effort for larger network organizations. RNNs are a kind of neural networks in which submerged layer inputs depend on previous outputs of the secret layer in the present point in time (Pal et al., 2018). This enables consecutive relationships like speech recognition to be dealt with in a time sequence. RNNs are more operational in sentiment analysis than CNNs according to earlier reasonable training of RNN and CNN in NLP. In the training of traditional RNNs, Strong LSTM was designed to gain strong term dependency over a longer period epoch in order to cope with the issue of the disappearing gradient (Hochreiter & Frasconi, 2001). Gates in LSTM also included, ignored windows including input and output windows. They are also used for estimation of time series, and for identification of handwriting.

The usage of LSTM in short texts of information sentiment is discussed and analyzed and explored that the fragmented associations of word existences in texts are useful for NLP (Silva & Ribeiro, 2007). The easy method of doing one hot encoding is to view the presence of each term as a binary vector in identifications the term embedding models are recycled in semantic delivery to map from a single-hot vector space to a continuous vector. In short texts the Word2Vec word embedding model is used here to represent words. Then, the LSTM classifiers are taught to capture the long-term relation between words in short texts. Every text's mood may be marked as positive or negative. The LSTMs work better on sequence-based activities with long term dependences, although a wide range of traditional LSTM modifications have been suggested in recent times, a large-scale study of the LSTM model revealed that none of the variants can substantially recover from the regular LSTM architecture.

The regular LSTM is adopted as part of the network structure (Hefron et al., 2017). Here, the secret layer is the distinction part between traditional LSTM architecture and RNN architecture. The secret layer of LSTM is also known as LSTM node. Like RNNs, an LSTM cell often consists of layer input xt, and layer output ht, for any iteration t. LSTM can contract with long-term dependencies because of the gated structure to allow valuable information to pass through the LSTM network. A LSTM cell includes three walls, including an input gate, a forget gate, and an output gate. The forget gate especially in the gated construction of LSTM is used for operative and accessible model for many knowledge issues related to successive data.

Stacked/Deeper LSTM (DLSTM)

Deeper (multilayer) LSTM constructions with multiple hidden layers is used to build up gradually further stages of illustrations for order data, and hence, achieve efficient results (S. Zhang et al., 2018). The deeper LSTM constructions are multiple networks with stacked LSTM hidden layers, where value of the hidden layer LSTM is used as the entry to the higher stage of the hidden layer LSTM. The problem of text classification is approached differently from current document classification methods that view the problem as multi-class classification (Kowsari et al., 2017). Instead they perform hierarchical classification using an approach, we call Hierarchical Deep Learning (HDL) for Text classification. HDL works stacks of deep learning constructions of specialized kind at all level of the document hierarchy.

A novel neural method is suggested for generation of paraphrases is discussed (Prakash et al., 2016). Conventional methods for generating paraphrases either exploit handwritten rules and alignments based on lexicons, or use the principles of statistical machine learning. A stacked residual LSTM network became the key target, where residual links between LSTM layers were introduced. This helps deep LSTMs to be trained effectively. The algorithm was trained on three separate datasets, along with other updated deep learning models: like PPDB, Wiki Answers, and MSCOCO; Appraisal. Results indicate that the model outperforms set, focus, and bidirectional. It has been showed that LSTM RNNs are better performer than deep neural networks and conventional RNNs in term of text classification.

Bi Directional Long Short Term Memory (BDLSTM)

The definition of BDLSTM stems from bidirectional (BD) RNN where methods are implemented with two separate concealed layers on sequence data in both feeds forward and reverse orders. BDLSTM uses dual hidden layers to the adjacent output plate where the efficiency of BDLSTM over simple uni-directional is discussed (Iung, 2013). BD networks have been found to be considerably safer in several respects than one directional systems, such as phoneme of text related classification and speech recognition which is consist of multiple character of language(Graves et al., 2013).

An RNN is trained to illustrate words by switching text tokens into vectors and matrix form (P. Zhou et al., 2016). They discuss the use of 2D max pooling method to achieve a fixed-length representation of the text in order to combine the functionalities on such dimensions of the matrix. It focuses on 2D convolution, in order to train more matrix data. Experiments are carried out on 6 tasks of text classification, including emotion analysis; the models proposed produce outstanding success on 4 out of 6 tasks. Adversarial training is constructed in the direction Supervised learning algorithms and automated testing of adversaries are capable of bringing supervised learning algorithms to a semi supervised stage (Miyato et al., 2017). Nevertheless, all approaches demand that different input vector entries, which are used for low, high dimensional inputs such as single word representations, be rendered to worry small. The system produces state-of-the-art performance on several semi-supervised and strictly controlled activities. A novel multi-layer RNN model called densely connected bidirectional long short-term memory (DCBi-

LSTM) is developed (Greff et al., 2015). Here, each layer is known through merging of its hidden state and totally preceding layers' hidden states, followed through repetitive passing each layer's representation to all following layers. The proposed model checked on 5 benchmark datasets of sentence classification. DC-Bi-LSTM with depth up to 20 can be successfully trained and obtain reliable improvements over the previous Bi-LSTM with the same or even less parameters.

The technique's superiority over other text detection algorithms focused on three sets: spam base data collection, farm advertisements and amazon book ratings (Nowak et al., 2017). The results of the first two datasets were compared with feed forward neural networks joint Ada Boost. The outcome is eventually contrasted with the bag-of-words algorithm. The emphasis is on classifying category wise email, as the point of concern is that it is categorized as a SPAM or qualified post. Further, the construction of an open BDLSTM layer, covering a forward LSTM layer and a backward LSTM layer is introduced.

Deeper Bi Directional LSTM (DBD LSTM)

DBDLSTM utilizes references both forward and backward. During the function learning cycle, both the three-dimensional relation of the speed in dissimilar traffic network sites and the dimensional dependencies of the speed specifications can be taken into account when feeding the multidimensional traffic network details, the BDLSTM. The BDLSTM is more helpful in this respect for the first layer of a model to gain valuable details from dimensionally time-level results. For guess future values, like speed values in this study, the final (highest) layer of a neural network only creates the future (predicted) values. During this process, the final layer requires only to consume the structures and guess future values with the directive of the outputs from previous layers. Thus, it is not mandatory to assume the BDLSTM in the last layer.

A deeper bidirectional LSTM is the combination of BDLSTM and DLSTM that uses multiple hidden layers and also back forward facility (Cui et al., 2017). A DBDLSTM neural network (NN) for system wise traffic speed prediction is proposed and its detailed framework is introduced. Speed prediction is defined as predicting for coming time periods speeds by using pre-speed details. Methods of controlled word sense disambiguation (WSD) focused on enabling vector machine or natural trees, despite intrinsic correlations to scientific word meaning (C. Zhang et al., 2019). Two deep-learning - based models are proposed for supervised WSD that demonstrates the neural network layout for the Bi LSTM with an appropriate upper layer structure. It performs even better than the existing state-of-the-art models on the MSH WSD dataset, which is three to four times faster than the Bi LSTM model. Additionally, the trained "universal" models are designed to disassemble all ambiguous words.

A novel technique focused on machine learning for the right role of Science writing of word choice is proposed (Makarenkov et al., 2019). Proper use of terms is a generalization of activities related to lexical replacement (LS) and grammatical error correction (GEC). To this function, they illustrate and test the effectiveness through applying BDLSTM tagger. Error detection-based classifiers and methods of computer translation are used to illustrate uncontrolled approach focused solely on a high-quality text corpus which needs no manual data. The key aim was to predict the most suitable target word provided an ESL writer's original word and its sentential meaning.

A technique was proposed for online 2D sequence recognition using deep bidirectional LSTM (Zhelezniakov et al., 2019). The hard job is electronic sequence identification of Handwritten Mathematical Expressions (HME) while isolation and identification of characters is a challenging activity. They suggest a deep learning approach that employs Recurrent Neural Networks (RNNs) for structure and character recognition in grouping for phrase creation with reordering and updated CYK algorithm. They also explored several types of structural and optimization improvements to the CYK algorithm which specifically boosts performance in terms of recognition speed. The research conducted for the implemented optimization techniques showed considerable improvement in recognition speed, holding the accuracy of the recognition comparable with established models.

The question of per classifier-per-one-word Word Sense Disambiguation (WSD) algorithms is highlighted by introducing a single BLSTM network (Pesaranghader et al., 2018). This is used to jointly identify senses and construct sequences that operate on all ambiguous terms. The estimation on the SensEval3 test indicates that the efficiency of the proposed model as comparable to the WSD algorithms of top output. A bidirectional attention-based long-term memory (Att-Bi-LSTM) paradigm is suggested for service robots, capable of classifying outpatient groups according to textual material (C. W. Chen et al., 2020). For the ambulatory text recognition system, consumers can speak to a service robot regarding their condition and the robot can tell them which clinic, they should be registering with. The users dialog text was obtained at the Taiwan E Hospital as the training data collection. The information in the dialog text was extracted, sorted, and converted through NLP to train the deep learning model of the LSTM. Experimental results prove skill of the robot to respond to questions separately concluded developed chance knowledge.

Proposed Research Methodology

After going through the related research work, it is realized that a lot of work on the LSTM and its classifiers for text classification is carried out in past and is still a hot research area for researchers from different backgrounds. Various standard of NLP and machine learning techniques are applied to extract the robust features from text data. Deep learning LSTM classifiers are used by the researchers to enhance the accuracy of the class text arrangement process. Our work will add further conciseness and remove the ambiguity in term of optimal LSTM classifier selection which is also a serious issue for text classification. To accomplish this task, we comparatively and empirically analyzed LSTM, DLTM, DBLSTM and DBDLSTM altogether. A real data set used for doing multi-class distribution (Entertainment, social media, sports, religion, politics) in which True positive, False positive, Accuracy, Precision, Recall and F-measure of each class were checked and analyzed.

A real data set was developed and used for text classification. First of all, data was collected from BBC news. The dataset was composed of English Text Related to Politics, Religion, Social

Media, Entertainment and Sports. For Simulation we used MATLAB which is a multi-model arithmetical calculating programming language and fourth-generation programming language. In our algorithm development we used machine learning and Deep Learinig toolbox.

Before the phase of classification, preprocessing, steps were soiled to the data. The extraction of features is an essential role in the learning of devices that removes any valuable knowledge from the text data. The classifiers were trained on the extracted features of LSTM based RNN which was used to predict/classify text data into its corresponding classes. In the last step the results were displayed and compared. The different phases of text classification process are illustrated in Figure 2 whereas the proposed research framework is presented in Figure 2.

Figure 2: Phases of Text Classification Process

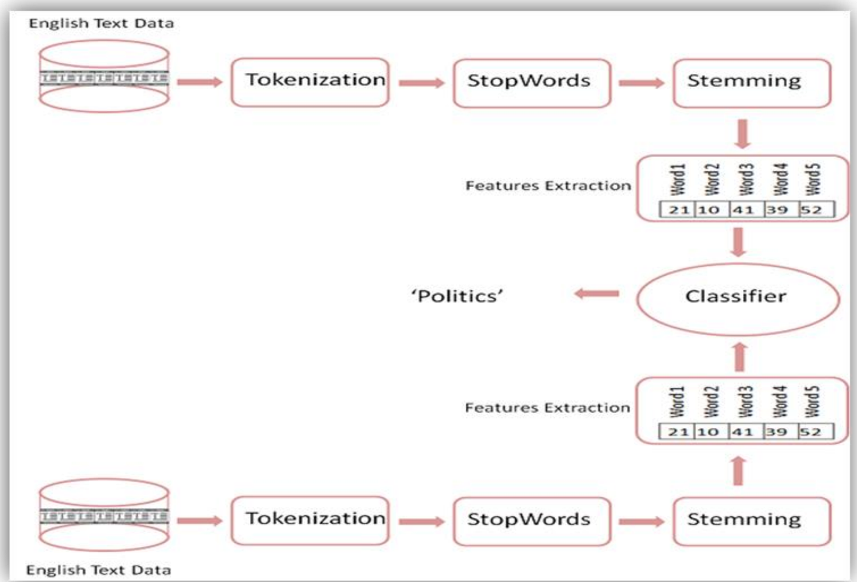
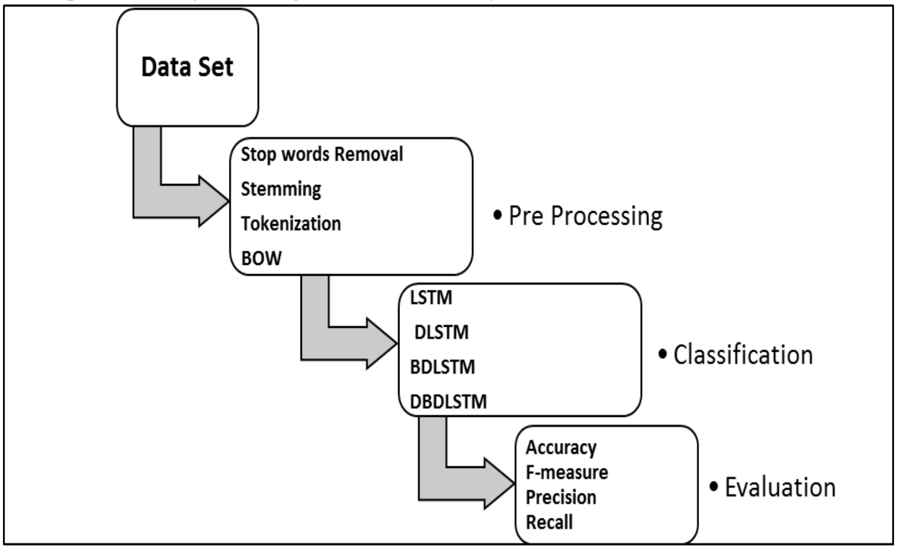


Figure 3: Proposed Frame work



Experimental Results and Discussion

The dataset comprises of phrases, terms, mathematical values and dates. This text dataset is divided into categories of training instances and for testing instances collection. MATLAB program is utilized which is an up-to-date modeling recreation platform used for text interpretation and data processing. MATLAB consists of a vast variety of machine learning and data science techniques applied in which consumers who lack a clear experience in computer science and programming are able to do so. The Experimental research was conducted on core i7 workstation, 3.7 GHZ with 24 GB RAM and 8 GB GPU at DIP Lab Islamia college University Peshawar. For comparison the performance in terms of accuracy, precision, recall etc. are noted. The results are obtained after simulation in two cases (case1=1000 instances and case2 =2000 instances). In case 1 all the described classes were checked empirically on 1000 instances for comparative analysis. First of all, training of simple LSTM was performed. The details of LSTM classifier on 1000 Instances in which correctly and incorrectly classified instances of related classes are presented. After successful training on 1000 instances next we train same classifiers one by one on data set size of 2000 instances.

First of all, LSTM was further trained on details of 2000 Instances in order to increase its accuracy but result obtained just increase by 10.67% which means that single layered LSTM is not achieving the peak with increasing data for each class.

At the initial stage of the research one of the goals of this analysis was to equate LSTM, DLSTM, BDLSTM and DBDLSTM classifiers. Hence the tests on the suggested classifiers were carried out. The model was built using BBC news text data collection for each trial. Classification comparison was carried out using Confusion Matrix, and we also calculated the classifier's true positive rate, false positive rate, precision recall and accuracy. In case 1, all the described classes were checked empirically on 1000 instances for comparative analysis. The training of simple LSTM was performed. According to Table 1, the result of classes of LSTM is presented. Here, the class entertainment has highest TP rate of 0.8018 FP rate, precision, Recall, F-Measure and accuracy of LSTM obtained.

Table 1

LSTM Class Wise detail on 1000 Instances

Class wise representation in %

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	Accuracy
Entertainment	0.8018	0.1125	0.5028	0.8018	0.4789	0.4573
Politics	0.7728	0.7130	0.3448	0.7728	0.2091	0.1500
Religion	0.1545	0.8138	0.8500	0.1545	0.8138	0.0850
Social Media	0.7998	0.0000	0.0000	0.7998	0.0000	0.0000
Sports	0.4424	0.6683	0.2489	0.4424	0.3886	0.8850
Weighted AVG	0.5942	0.4615	0.3893	0.5942	0.3780	0.3154

Deeper LSTM just overlap simple LSTM where the details of described classes. According to details illustrated in Table 2, the TP, FP rate, precision, Recall, highest F-Measure class is religion with 0.5976 and highest accuracy obtained by sports with 0.6816. The BDLSTM results of Entertainment, Politics, Religion, Social

Media and Sports classes is presented in Table 3. According to this table, the TP rate social media is highest with 0.8138, and lowest FP rate is of class politics with 0.0701 other like precision, Recall, F-Measure and accuracy. The BDLSTM was found better than earlier simple LSTM.

Table 2

Deeper LSTM Class Wise detail on 1000 Instances

	Class wise representation in %					
Class	TP Rate	FP Rate	Precision	Recall	F-Measure	Accuracy
Entertainment	0.8174	0.0920	0.5061	0.8174	0.4573	0.4171
Politics	0.8208	0.9510	0.5607	0.8208	0.5201	0.4850
Religion	0.8288	0.1227	0.5644	0.8288	0.5976	0.6350
Social Media	0.7988	0.0951	0.9049	0.7988	0.4274	0.3750
Sports	0.7858	0.1880	0.4774	0.7858	0.5615	0.6816
Weighted AVG	0.8102	0.2897	0.6027	0.8102	0.5127	0.5190

Table 3

BDLSTM Class Wise Detail on 1000 Instances

	Class wise representation in %					
Class	TP Rate	FP Rate	Precision	Recall	F-Measure	Accuracy
Entertainment	0.5966	0.4234	0.2890	0.5966	0.4047	0.6749
Politics	0.7708	0.0701	0.3253	0.7708	0.1908	0.1350
Religion	0.8128	0.0163	0.6667	0.8128	0.2176	0.1300

Social Media	0.8138	0.0200	0.6522	0.8138	0.2439	0.1500
Sports	0.7037	0.2868	0.3629	0.7037	0.4695	0.6650
Weighted AVG	0.7397	0.8166	0.4592	0.7397	0.3053	0.3509

Table 4 presents of Deeper Bi-Directional LSTM trained on 1000. The obtained accuracy is 52.22% in which class Entertainment have 66.33% accuracy and Sports have 64% accuracy was highest on 1000 instances with respect to all others further details are in.

Table 4

DBDLSTM Class Wise Detail on 1000 Instances

	Class wise representation in %					
Class	TP Rate	FP Rate	Precision	Recall	F-Measure	Accuracy
Entertainment	0.8609	0.0900	0.6471	0.8609	0.6551	0.6633
Politics	0.8098	0.0463	0.5595	0.8098	0.3310	0.2350
Religion	0.8408	0.1164	0.5903	0.8408	0.6276	0.6700
Social Media	0.7548	0.1615	0.3944	0.7548	0.4088	0.4200
Sports	0.7848	0.1790	0.4273	0.7848	0.5435	0.6400
Weighted AVG	0.8102	0.1186	0.5237	0.8102	0.5132	0.5226

After successful training on 1000 instances next, we train same classifiers one by one on data set size of 2000 instances. Table 5 shows the results by LSTM detail of classes according to TP rate, FP rate, where highest precision of class entertainment with 0.7820, and lowest F-Measure is of social media with 0.1235 and lowest accuracy 0.0750 obtained. Table 6 shows the results of DLSTM on classes. According to which, the TP rate Sports was on top with

0.9240 it also has lowest FP rate of 0.0288, precision, Recall, F-Measure and accuracy.

Table 5

LSTM Class Wise Detail on 2000 Instances

	Class wise representation in %					
Class	TP Rate	FP Rate	Precision	Recall	F-Measure	Accuracy
Entertainment	0.8750	0.0363	0.7820	0.8750	0.6256	0.5213
Politics	0.7219	0.1914	0.3289	0.7219	0.3505	0.3750
Religion	0.7539	0.1476	0.3789	0.7539	0.3692	0.3600
Social Media	0.7869	0.0350	0.3488	0.7869	0.1235	0.0750
Sports	0.7024	0.3146	0.3798	0.7024	0.5087	0.7700
Weighted AVG	0.7680	0.1449	0.4436	0.7680	0.3955	0.4202

Table 6

DLSTM Class Wise Detail on 2000 Instances

	Class wise representation in %					
Class	TP Rate	FP Rate	Precision	Recall	F-Measure	Accuracy
Entertainment	0.8869	0.0700	0.7179	0.8869	0.7161	0.7143
Politics	0.7236	0.0632	0.7710	0.7236	0.4920	0.3613
Religion	0.8664	0.0544	0.7166	0.8664	0.6223	0.5500
Social Media	0.8769	0.1126	0.6498	0.8769	0.7309	0.8350
Sports	0.9240	0.0288	0.8647	0.9240	0.7946	0.7350

Weighted AVG	0.8555	0.0658	0.7440	0.8555	0.6711	0.7361

The Bi-Directional LSTM was trained on 2000 instances. The resultant detail of classes according to TP, FP rate, precision, Recall, F-Measure and accuracy obtained is noted in Table 7. Similarly, the Table 8 presents the results on DBLSTM Entertainment, Politics, Religion, Social Media and Sports classes according to highest TP rate social media have 0.9950, FP rate, precision, Recall, F-Measure of class sports is 1.0000 highest of all and also same accuracy obtained. In which DBDLSTM was trained on 2000 instances obtaining best of all accuracy and other performance measures.

Table 7

BDLSTM Class Wise Detail on 2000 Instances

	Class wise representation in %					
Class	TP Rate	FP Rate	Precision	Recall	F-Measure	Accuracy
Entertainment	0.9070	0.1106	0.6878	0.9070	0.8075	0.9774
Politics	0.8460	0.0918	0.5897	0.8460	0.5823	0.5750
Religion	0.9820	0.0100	0.9596	0.9820	0.9548	0.9500
Social Media	0.9830	0.0000	1.0000	0.9830	0.9556	0.9150
Sports	0.9780	0.0213	0.9198	0.9780	0.9466	0.9750
Weighted AVG	0.9392	0.046	0.5705	0.9392	0.8993	0.8784

Table 8

DBDLSTM Class Wise Detail on 2000 Instances

Class wise representation in %

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	Accuracy
Entertainment	0.9770	0.0288	0.8966	0.9770	0.9455	1.0000
Politics	0.9835	0.0063	0.9742	0.9835	0.9581	0.9425
Religion	0.9835	0.0000	1.0000	0.9835	0.9570	0.9175
Social Media	0.9950	0.0000	1.0000	0.9950	0.9873	0.9750
Sports	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000
Weighted AVG	0.9878	0.00702	0.9741	0.9878	0.9695	0.9670

The outcomes from listed classifiers for text classification are analyzed at the end. The classifiers are evaluated on BBC news data set using evaluation measures like TP rate, FP rate, Recall, Precision etc. This analysis show that the performance of classifiers like DLSTM, BDLSTM and DBDLSTM is found better than LSTM classifier. The comprehensive findings discussed above indicate that DBDLSTM has the higher performance results. Moreover, the results also show that increasing the instances result in better performance of all LSTM classifiers. Table 1 shows average Performance by each RNN Based Classifiers for different evaluation measures. This table shows the better performance by DBDLSTM technique as the resultant values are comparatively large.

Table 10

Average Performance by RNN LSTM Based Classifiers

Evaluation Measure	LSTM	DLSTM	BDLSTM	DBDLSTM
Accuracy	37%	64%	61%	74%
TP Rate	68%	83%	84%	90%

FP Rate	30%	47%	43%	6%
Precision	42%	67%	51%	75%
Recall	68%	83%	83%	90%
F-measure	39%	59%	60%	74%

Conclusion

The Recurrent Neural Network (RNN) is a form of Neural Network RNN. Several Deep learning techniques particularly the RNN have been used to recognize text which is a challenging task. There are different RNN layered base classifiers which are utilized successfully for example Long Short-Term Memory (LSTM), Bi Directional LSTM (BDLSTM), Deeper LSTM (DLSTM) and Deep Bi-directional LSTM (DBDLSTM). These classifiers are comparatively analyzed for text classification in this research. The news platform data set developed from BBC news website is considered which has great variations in interpretations and events. The dataset design has been hard to use in large scale research, rendering it is hard for the training and validation process.

The developed data set is further divided in five classes which are religion, entertainment, social media, sports and politics. Accordingly, four frequently used RNN LSTM classifiers were trained on the data set with 1000 instances and next on 2000 instances. The comparative results that are obtained were analyzed and presented in form of tables. The TP rate, FP rate, Recall, Precision of utilized techniques were obtained to evaluate the performance of each RNN based LSTM classifier. The overall results infer the better performance of DBDLSTM as compare to DLSTM, BDLSTM and traditional LSTM classifier. It is proved that using deep learning RNN based DBDLSTM model achieved highest efficiency. Hence, this analysis offers the better method for classifying and interpreting text using different LSTM classificatory. By doing this study, it is also suggested that a text classification system through utilizing and manipulating the RNN's multi layered deep features with increase data will improve its performance further. The reason of significant improvement in overall performance with increase in data instances is that ability of deep learning approach to utilize large epochs, bitch size and other parameters.

For abstraction and grouping of functions, specific NN (Neural network) may be tested in the future. In terms of precision RNN is known to be very reliable, multi-layered LSTM, DBDLSTM can be utilized for the improvement of the strength of our text behavior analysis method. The attributes of the English text data were categorized using various LSTM classifiers. There are a few other algorithms for grouping, such as Help Vector Machine, Victor Quantization, and K-means. Instead of LSTM, it is possible to use or check the above-mentioned algorithm to classify the functions of text files. Next analysis can be on creating further large data set and also to compare with some benchmark and real data sets. More research on rising languages may also be performed instead of just one language like Urdu written in English.

References

Asim, M. N., Wasim, M., Ali, M. S., & Rehman, A. (2018). Comparison of feature selection methods in text classification on highly skewed datasets. *1st International Conference on Latest Trends in Electrical Engineering and Computing Technologies, INTELLECT 2017*, *2018-Janua*, 1–8. https://doi.org/10.1109/INTELLECT.2017.8277634

Buzic, D., & Dobsa, J. (2018). Lyrics classification using Naive Bayes. *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2018 - Proceedings*, 1011–1015. https://doi.org/10.23919/MIPRO.2018.8400185

Chen, C. W., Tseng, S. P., Kuan, T. W., & Wang, J. F. (2020). Outpatient text classification using attention-based bidirectional LSTM for robot-assisted servicing in hospital. *Information (Switzerland)*, *11*(2). https://doi.org/10.3390/info11020106

Chen, D., & Liu, Z. (2010). A new text categorization method based on HMM and SVM. *ICCET 2010 - 2010 International Conference on Computer Engineering and Technology, Proceedings*, *7*, 383–386. https://doi.org/10.1109/ICCET.2010.5485482

Chen, Y., & Zhang, Z. (2018). Research on text sentiment analysis based on CNNs and SVM. *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 2731–2734. https://doi.org/10.1109/ICIEA.2018.8398173

Cui, Z., Ke, R., & Wang, Y. (2017). Deep Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction. *The 6th International Workshop on Urban Computing*, 22–25.

Goudjil, M., Koudil, M., Hammami, N., Bedda, M., & Alruily, M. (2013). Arabic text categorization using SVM active learning technique: An overview. *2013 World Congress on Computer and Information Technology, WCCIT 2013*, 7–8. https://doi.org/10.1109/WCCIT.2013.6618666

Graves, A., Jaitly, N., & Mohamed, A. (2013). Hybrid speech recognition with Deep Bidirectional LSTM. *2013 IEEE Workshop on Automatic Speech Recognition and Understanding*, *12*(4), 273–278. https://doi.org/10.1109/ASRU.2013.6707742

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2015). TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS 1 LSTM: A Search Space Odyssey. *ArXiv:1503.04069*, 1–11. https://doi.org/10.1109/TNNLS.2016.2582924

Hefron, R. G., Borghetti, B. J., Christensen, J. C., & Kabban, C. M. S. (2017). Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation. *Pattern Recognition Letters*, *94*, 96–104. https://doi.org/10.1016/j.patrec.2017.05.020

Hochreiter, S., & Frasconi, P. (2001). Gradient Flow in Recurrent Neural Nets : The Difficulty of Learning Long-Term Dependencies. A Field Guide to Dynamical Recurrent Neural Network. *A Field Guide to Dynamical Recurrent Neural Network*, 401–403. https://doi.org/10.1109/9780470544037.ch14

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, G. Bin, Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, *70*(1–3), 489–501. https://doi.org/10.1016/j.neucom.2005.12.126

Iung, B. (2013). Cœur et grossesse. *EMC - Traité de Médecine AKOS*, *8*(2), 1–4. https://doi.org/10.1016/s1634-

6939(13)59289-1

Kowsari, K., Brown, D. E., Heidarysafa, M., Jafari Meimandi, K., Gerber, M. S., & Barnes, L. E. (2017). HDLTex: Hierarchical Deep Learning for Text Classification. *Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017*, *2017-Decem*, 364–371. https://doi.org/10.1109/ICMLA.2017.0-134

Lin, Y., & Wang, J. (2014). Research on text classification based on SVM-KNN. *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS*, 842–844. https://doi.org/10.1109/ICSESS.2014.6933697

Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). *A Critical Review of Recurrent Neural Networks for Sequence Learning*. 1–38.

Maheswari, K., & Priya, P. P. A. (2018). Predicting customer behavior in online shopping using SVM classifier. *Proceedings of the 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing, INCOS 2017*, *2018-Febru*, 1–5. https://doi.org/10.1109/ITCOSP.2017.8303085

Makarenkov, V., Rokach, L., & Shapira, B. (2019). Choosing the right word: Using bidirectional LSTM tagger for writing support systems. *Engineering Applications of Artificial Intelligence*, *84*, 1–10. https://doi.org/10.1016/j.engappai.2019.05.003

Miyato, T., Dai, A. M., Goodfellow, I., & Networks, P. (2017). Adversarial Training Methods for Semi-supervised Text Classification. *ICLR 2017*, 1–11.

Nowak, J., Taspinar, A., & Scherer, R. (2017). LSTM recurrent neural networks for short text and sentiment classification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10246 LNAI*(November), 553–562. https://doi.org/10.1007/978-3-319-59060-8_50

Pal, S., Ghosh, S., & Nag, A. (2018). Sentiment Analysis in the Light of LSTM Recurrent Neural Networks. *International Journal of Synthetic Emotions*, *9*(1), 33–39. https://doi.org/10.4018/ijse.2018010103

Pesaranghader, A., Pesaranghader, A., Matwin, S., & Sokolova, M. (2018). One single deep bidirectional LSTM network for word sense disambiguation of text data. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial*

Intelligence and Lecture Notes in Bioinformatics), *10832 LNAI*, 96–107. https://doi.org/10.1007/978-3-319-89656-4_8

Prakash, A., Hasan, S. A., Lee, K., Datla, V., Qadir, A., Liu, J., & Farri, O. (2016). Neural paraphrase generation with stacked residual LSTM Networks. *COLING 2016 - 26th International Conference on Computational Linguistics, Proceedings of COLING 2016: Technical Papers*, 2923–2934.

Rhanoui, M., & Mikram, M. (2019). *A CNN-BiLSTM Model for Document-Level Sentiment Analysis*. 832–847. https://doi.org/10.3390/make1030048

Silalahi, M., Hardiyati, R., Nadhiroh, I. M., Handayani, T., Amelia, M., & Rahmaida, R. (2018). A text classification on the downstreaming potential of biomedicine publications in Indonesia. *2018 International Conference on Information and Communications Technology, ICOIACT 2018*, *2018-Janua*, 515–519. https://doi.org/10.1109/ICOIACT.2018.8350778

Silva, C., & Ribeiro, B. (2007). Scaling text classification with relevance vector machines. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, *5*, 4186–4191. https://doi.org/10.1109/ICSMC.2006.384791

Wang, J.-H., Liu, T.-W., Luo, X., & Wang, L. (2018). An LSTM Approach to Short Text Sentiment Classification with Word Embeddings. *2018 Conference on Computational Linguistics and Speech Processing ROCLING 2018*, 214–223.

Yin, W., Kann, K., Yu, M., & Schütze, H. (2017). *Comparative Study of CNN and RNN for Natural Language Processing*.

Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing [Review Article]. *IEEE Computational Intelligence Magazine*, *13*(3), 55–75. https://doi.org/10.1109/MCI.2018.2840738

Zhang, C., Biś, D., Liu, X., & He, Z. (2019). Biomedical word sense disambiguation with bidirectional long short-term memory and attention-based neural networks. *BMC Bioinformatics*, *20*(Suppl 16), 1–15. https://doi.org/10.1186/s12859-019-3079-8

Zhang, S., Lei, M., Yan, Z., & Dai, L. (2018). Deep-FSMN for Large Vocabulary Continuous Speech Recognition. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *2018-April*, 5869–5873. https://doi.org/10.1109/ICASSP.2018.8461404

Zhelezniakov, D., Zaytsev, V., & Radyvonenko, O. (2019). Acceleration of Online Recognition of 2D Sequences Using

Deep Bidirectional LSTM and Dynamic Programming. In I. Rojas, G. Joya, & A. Catala (Eds.), *Advances in Computational Intelligence, Springer Nature* (Vol. 11507). Springer International Publishing. https://doi.org/10.1007/978-3-030-20518-8

Zhou, J., Lu, Y., Dai, H. N., Wang, H., & Xiao, H. (2019). Sentiment analysis of Chinese microblog based on stacked bidirectional LSTM. *IEEE Access*, *7*, 38856–38866. https://doi.org/10.1109/ACCESS.2019.2905048

Zhou, P., Qi, Z., Zheng, S., Xu, J., Bao, H., & Xu, B. (2016). Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling. *COLING 2016 - 26th International Conference on Computational Linguistics, Proceedings of COLING 2016: Technical Papers*, *2*(1), 3485–3495.