International Journal on Emerging Technologies **11**(3): 435-441(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Multi Co-Trained Gated Recurrent Neural Network for efficient Document Indexing: MCTGRNN

***Y. Krishna Bhargavi*[1]*, Y.S.S.R. Murthy*[2] *and O. Srinivasa Rao*[3]**
[1]Assistant Professor, Department of Computer Science and Engineering,
GRIET, Hyderabad, (Telangana), India.
[2]Professor, Department of Information Technology,
SRKR Engineering College, Bhimavaram, (Andhra Pradesh), India.
[3]Professor, Department of Computer Science and Engineering,
University College of Engineering, JNTUK Kakinada (Andhra Pradesh), India.

(Corresponding author: Y. Krishna Bhargavi)
(Received 26 February 2020, Revised 23 April 2020, Accepted 25 April 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: The tremendous increase in the generation of documents has given a significant pavement towards research in Document Indexing with Big Data. Different researchers have earlier proposed many techniques for document indexing but they all yield low performance in the classification owing to chosen feature selection techniques. The existing methods are analyzed and identified that the accuracy of classification is degraded due to the feature selection techniques and the amount of training data. In the current research work, the Multi Co-Trained Gated Recurrent Neural Network (MCT-GRNN) method is used in the document indexing to increase the performance of the classification. In this method, Term-Frequency – Inverse Document Frequency and Latent Dirichlet Allocation collectively select the features and provide the relevant and significant features to GRNN for classification. The GRNN technique adaptively analyzes the semantic information in the sentence and further analyzes the relationship in the document representation. With the Reuters dataset, the effectiveness of the proposed MCT-GRNN method is analyzed. The experimental result shows that the proposed MCT-GRNN has performed with an overall accuracy of 94.77% when compared with the methods like Naïve Bayes, LSTM.

Keywords: Document indexing, Latent Dirichlet Allocation, Multi Co-Trained Gated Recurrent Neural Network, Term-Frequency – Inverse Document Frequency.

Abbreviations: MCTGRNN, Multi Co-Trained Gated Recurrent Neural Network.

I. INTRODUCTION

In Big Data, automated text classification is the method of categorizing the documents with the use of machine learning techniques. The machine learning techniques may include supervised, semi-supervised or unsupervised based on the type of data set and classification applied [1]. With the enormous volume and diversity in digitized documents, Data mining techniques have become essential to obtain the knowledge discovery. The sources and applications in text mining are also diverse. A wide range of applications in Automatic Document Classification include design of digital libraries, spam filters, search on web [2]. Automated Information Retrieval helps in understanding, reading, indexing and tracking the textual data. With this necessity, the research is progressing rapidly in the fields of document retrieval, computational linguistics and text mining [3, 4].

The keyword matching techniques considers the out shell text information but deficits in identifying the semantic in documents [5]. The existing research shows that document similarity measures are efficient with LDA than with Bag-of Words alongside TF-IDF [6]. LDA derives the document representation using the global semantic topical structures in the document. LDA uses posterior maximization with Gibbs Sampling to obtain the final outcomes of topic-to-word and document-to-topic distribution [7]. To build effective conventional and deep learning approaches, large volumes of labeled data is required in each target data set to train and confirm a learning model [8].

These techniques utilize the neural network to find semantic relationships between all words in a document, but they are ineffective in modeling the inherent relations between sentences [9].

In this research, the MCT-GRNN has been proposed to enhance the efficiency of document indexing. The features are extracted from the document using TF-IDF and LDA. "These features are applied to the LSTM and GRNN for indexing of the document. The experimental result analyzes the effectiveness of the MCT-GRNN and compared with existing methods".

The paper is formulated as literature survey and problem definition are given in the section II, explanation about the TF-IDF, LDA and GRNN is given in the section III, analysis of experimental results is given in the section IV, and conclusion is given in the section V.

II. LITERATURE SURVEY

As the data increases in the internet, this is difficult to process and retrieve the relevant information. The document indexing technique is applied to easily retrieve the relevant document. The recent research related to the document indexing techniques is surveyed in this section and stated the problem.

Kim, *et al*. presented the Multi-co-training (MCT) to increase the efficiency in the document classification. The three feature selection techniques namely Term-frequency-Inverse Document Frequency (TF-IDF), Document to Vector (DoC2vec) and Latent Dirichlet Allocation (LDA) are used. The performance of this technique is evaluated using Random Forest and Naïve Bayes classifiers. The document classification is good but it is not efficiently handling the data imbalance in the documents and computation complexity is high [10].

D. Srikanth, and S. Sakthivel, introduced a notion of vantage point, referred as Vantage Point Latency Semantic Indexing (VP-LSI). The user query is explored in the Rayleigh cluster objects index. With this method, the multimedia web documents are retrieved faster and also achieved improvement in accuracy. As this technique doesn't follow the reduction of predominant words, the execution time is more [11].

Chen, *et al*., proposed two feature selection techniques such as Multi-class Odds Ratio (MOR) and Class Discriminating Measure (CDM) on Naive Bayes classifiers. The two datasets-Reuters and the Chinese text classification are used to evaluate the performance and identified to be high compared to the existing techniques. The effectiveness of the text classification is low due to the feature selection criteria [12].

A. Mourao and Joao Magalhaes, introduced Balanced K-means Singular Value Decomposition (B-KSVD). It is suitable for searching over-complete patterns. It proved to be working good even when the search limit is low and achieved good distribution over small number of partitions (512 to 1024). However, the smaller reconstruction coefficients of hash dimensions decrease the performance gains of the method [13].

Jiang, *et al*., developed the text classification method based on hybrid technique of Deep Belief Network (DBN) and softmax regression. The sparse high-dimensional computation problem is resolved with deep belief network. The DBN technique is used for feature extraction and the softmax regression is used for text classification in the feature space. The DBN and softmax are trained and the parameter settings are optimized during the fine-tuning. To evaluate this method, Reuters – 21578 and 20- Newsgroups data sets are used. This method resulted in better performance due to parameter optimization. Selection of significant features may still increase its efficiency [14].

Liu & Guo developed the Attention-based Bidirectional Long Short-term Memory with convolution layer as AC-BiLSTM to address several issues such as sparsity and high dimensionality of text data. Both local features of phrases and global sentence semantics are captured by AC-BiLSTM. To validate the effectiveness of this method, the experiments are carried out on six sentiment classification datasets. With increase in stride size, more semantic information is lost in short sentences than the longer ones [15].

Sun *et al.,* solved the feature sparse problem by developing the Deep Neural Network (DNN) for classifying the sentiment of Chinese micro-blog. The posts and their related comments are combined and transformed into micro-blog conversation by using content extension framework for extracting the features. The experimental results stated that the DNN model performed better results for short-length document classification than long-sentence. During training, when the number of DNN layers increases, the time consumption for system also increases, which lead poor performance in classification [16].

Wang *et al.,* developed an eight layer stacked residual LSTM to predict the sentiment intensity for a given text. The optimization is easily performed by using this deeper network. The prediction accuracy is improved and task of classification is successfully done by using more number of LSTM layers. The experiments are conducted to test the efficiency of eight layers LSTM against Pearson correlation coefficient and Mean Absolute Error. But, LSTM model provides very poor training efficiency for time-sensitive language processing tasks and it is still highly competitive for the model with a GPU support [17].

Fu *et al.,* considered the semantic information and addressed the complex structure of neural network, a simple method called Bag-of Meta-Words (BoMW) is proposed. In meta-words vector, each meta-word represents the semantic information. The complementary models namely feature combination meta-words (FCM) and naive interval meta-words (NIM) extracted meta-words from pre-trained word embeddings. Two datasets such as Pang's and IMDB dataset are used in the experiments for validating the performance of BoMW against traditional VSM models. In a sentence-level classification, the BoMW ignores the word order and provides lower performance [18].

Bhushan and Danti, developed an efficient similarity measure to compute the similarity among two document sets. In real time scenarios, these methods have a capability of measuring the level of similarity between text documents which consists of either the presence or absence of features. In addition to these models, a compression modeling for measuring the similarity of documents is also developed. The method used four various datasets for evaluating the performance in terms of precision, recall and F-Score. The technique provides poor performance due to insufficient information on text summarization [19].

A. Problem Definition

Many existing methods have involved in the document indexing and these methods show low efficiency in the classification. The major problems in the existing methods are provided as:

– The classification accuracy of the existing methods is low due to the feature selection techniques. Irrelevant features are also included in these techniques.

– Existing methods have low performance when the dataset has less number of data.

– Efficient feature selection techniques classify the data in categories even if their availability isles.

– The performance of the existing methods in document indexing is degraded because of low amount of training data. Some applications may have less testing data and still a technique may be needed to handle this.

To prevail over the limitations of the methods on hand, the MCT-GRNN method is proposed in document indexing. The detailed description of MCT-GRNN method is provided in the next section.

III. PROPOSED METHOD

Text data has been increasing in various fields and this requires an effective technique to handle. Document indexing is the process of categorization of the text from the dataset. Numerous document indexing methods have been applied in the last few decades. The efficiency of the document indexing still needs to be improved using the feature selection technique. In this research, the MCT-GRNN has been proposed to enhance the efficiency of document indexing. The features are extracted from the document using TF-IDF and LDA. These features are applied to the LSTM and GRNN for indexing of the document. The experimental result analyzes the effectiveness of the MCT-GRNN and compared with existing methods. The architecture of the proposed MCT-GRNN method is portrayed in Fig. 1.

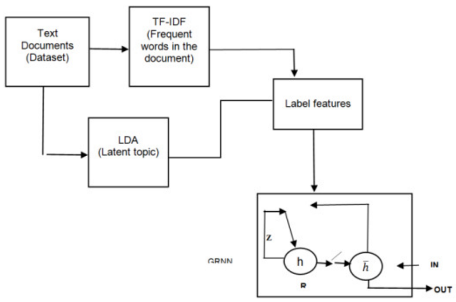

Fig. 1. Architecture of MCT-GRNN Method.

A. Term Frequency-Inverse Document Frequency

The intuition behind TF-IDF is that a query team that occurs in several documents is not a good qualifier rather it should be assigned with less weight than the term that occurs in fewer documents. Term weight is calculated using Eqn. (1) [20].

$$w_{m,n} = t\,f_{m,n} \times \log\left(\frac{N}{d\,f_m}\right) \qquad (1)$$

where,

$w_{m,n}$ is the weight for term m in document n

N is the number of documents in the collection

$t\,f_{m,n}$ is the term frequency of term m in document n

$d\,f_m$ is the document frequency of term m in the collection.

Depending on the importance of a term in a document, it can be categorized into two classes. It measures the relevance of a term in a given document. There are some issues with TF-IDF in text representation. It is not derived from mathematical model; instead it is taken from information theory by Shannon. Another issue is with the dimensionality; it involves in huge computation on the entire size of vocabulary over the data set.

B. Latent Dirichlet Allocation

Latent Dirichlet allocation (LDA) [21] is a hierarchical Bayesian model. It maps a text document into a latent low dimensional space extended by a set of automatically learned topical bases. Eqn. (2) provides the generative probability of all documents with an assumption that each document contains K topics.

$$p(v, z, \theta|\alpha, \pi) = p(\theta|\alpha) \prod_{m=1}^{M} p(Z_m|\theta)\, p(v_m|Z_m\pi) \quad (2)$$

Where

θ ~ Dirichlet $(\theta)_{0_t}$ is a K -dimensional vector that represents the topic proportion $\sum_{k=1}^{k} \theta_k = 1 - z_m$ is a K-dimensional Indicator vector referring to the mth word. (All elements are zeroes and my one element is 1).

π is a matrix with K topics for K rows and each row corresponds to a multi-nominal distribution over words in the specified vocabulary.

This model is used to Identify the topics he. Latent features of a document.

C. Gated Recurrent Neural Network

The approach is analogous to recently emerged LSTM and gated neural network is developed from LSTM. The semantics of sentences and their inherent relations in document representations are adaptively encoded using GRNN. These are used to identify the label for each category of documents.

D. Long Short Term Memory

The topics of documents are predicted using both the latest and the previous data. The RNN model can deal with long term dependence issues using the self-feedback technique at hidden layer. It has certain difficulties in sensible applications. In LSTM, the input, forget and output gates are used to store and update the information in a memory cell [22, 23].

At time ^{t,} the input is given to the LSTM cell using x_t, the output at previous moment of the LSTM cell is given with $h_{t-1,}$ c_t provides the value of cell, output of the cell is emitted through h_t the computation at the LSTM unit is as follows:

— Determine the value of candidate memory cell $\widetilde{C}_t$

$$\widetilde{C_t} = \tanh(W_c.[h_{t-1}, x_t] + b_c) \qquad (3)$$

W_c. : Weight matrix

b_c : Bias

— Compute the value of the input i_t gate. It controls the current input data up dating to the memory cell

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \qquad (4)$$

σ = Sigmold Function

W_i = Weight Matrix

b_i = Bias

— Calculate the value of the forget gate. It controls the update of the historical data to the state value of the memory cell.

$$f_t = \sigma\left(W_i.[h_{t-1}, x_t] + b_f\right) \qquad (5)$$

W_f = Weight Matrix

b_f = Bias

— Determine the value of the current moment at memory cell c_t

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \qquad (6)$$

* represents the dot product. C_{t-1} Is the state value of the last LSTM unit the state values of the candidate and previous cells are used to update z the memory cell. It is controlled by input and forget gate.

— Compute ^{0}t which is the value at output gate. It controls the memory cell.

$$0_t = \sigma(W_0.[h_{t-1}, x_t] + b_0) \qquad (7)$$

W_0 = Weight Matrix

b_0 = Bias

— At last compute the output of the LSTM unit $^h{}_t$ as shown in Eqn. (8)

$$h_t = 0_t * \tanh(C_t) \qquad (8)$$

LSTM can read, store, reset and update long term information easily with the help of these three gates. The dimensions of the weight matrix are used to control the output. A long time delay is established by LSTM between input and the feedback. As a continuous error flow is maintained in the architecture, the gradient will neither detonate nor decrease because of the internal state of the memory cell.

A gated recurrent unit (GRU) adaptively captures the dependencies of different time scales at each recurrent unit [28]. Similar to the LSTM unit, the GRU has gating units that modulate the flow of information inside the unit, however, without having a separate memory cells.

Eqn. (9) is a linear interpolation between the previous activation h_j^{t-1} and the candidate activation $\tilde{h}_t^j$ at time, represented by the activation of GRU at h_t^j.

$$h_t^j = (1 - z_t^j)h_{t-1}^j + z_t^j \tilde{h}_t^j \qquad (9)$$

An update gate z_t^j identifies the amount of unit that updates its activation. The Eq. (10) computes this update gate.

$$z_t^j = \sigma(W_z x_t + U_z h_{t-1})^j \qquad (10)$$

As the LSTM unit, the GRU also follows this process of considering a linear sum between the existing state and the newly computed state. But the GRU renders the whole state every time as it does not have any method to control the degree of exposure of a state.

As in Eqn. (11), the candidate activation $\tilde{h}_t^j$ is calculated as the traditional recurrent unit (as in Eqn. (9)).

$$\tilde{h}_t^j = \tanh\left(W x_t + U(r_t \odot h_{t-1})\right)^j \qquad (11)$$

r_t^j Is an element-wise multiplication and is a set of reset gates. When off (r_t^j close to 0), the reset gate successfully makes the unit proceed as if it is evaluating the first symbol of input sequence, and sets aside the previously computed state.

The reset gate r_t^j is calculated as the update gate, as represented in Eqn. (12).

$$r_t^j = \sigma(W_r x_t + U_r h_{t-1})^j \qquad (12)$$

(i) **Pseudo code:** MCT-GRNN	
Obtain the input document	
Measure TF_IDF	\\TF-IDF based on Eqn. (1)
Measure LDA	\\LDA based on Eqn. (2)
$x_t = TF\_IDF + LDA$	\\Both features are given as input to GRNN
$\tilde{c}_t = (W_c, x_t, b_c)$	\\Candidate memory cell with weight and bias value
$i_t = \sigma(W_c, x_t, b_c)$	\\Calculate the input gate
$r_t^j = \sigma(W_c, x_t)$	\\The reset and update the gate based on Eqn. (12)
$o_t = \sigma(W_0, x_t, b_0)$	\\Calculate the output of GRNN
9. $l_i = o_t(x_t)$	\\Classify the document

The pseudo code of the MCT-GRNN is shown below:

Once, the proposed MCT-GRNN classifies the data into the different categories, the metrics are used to analyze the performance. The next section gives the detailed explanation about the performance of the proposed MCT-GRNN method

IV. EXPERIMENTAL RESULTS

The document indexing techniques have been applied to classify the document based on the categories. Many techniques like machine learning and deep learning were applied to improve the performance of document indexing method. The existing method shows low performance in the document indexing due to the feature selection technique. In this research, the GRNN is proposed for the document indexing technique. Since GRNN method analyzes every input and store the information in the memory cell. The GRNN method stores the information and filters the irrelevant features that solve the vanishing gradient problem.

Datasets: Reuters' dataset is collected by Carnegie group from the Reuters news agency. The Reuters-21578 benchmark dataset consists of 21578 documents distributed in the 135 classes. It provides the train and test data sets with the seven most frequent classes in its split version.

Experimental setup: The experiment is conducted on the system consists of the 16 GB RAM and 2.5 GHz Intel i7 processor. The proposed GRNN is implemented in the Python to evaluate the performance.

Evaluation Metrics: F1 score, recall, precision and accuracy are the metrics used to analyze the proposed GRNN. The formula for calculating these metrics are as given in the Eqn. (13-16).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (13)$$

$$Precision = \frac{TP}{TP+FP} \qquad (14)$$

$$Recall = \frac{TP}{TP+FN} \qquad (15)$$

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (16)$$

Where

T_P =True Positive

F_P=False Positive

T_N= True Negative

F_N=False Negative

Parameter settings: The GRNN layer is set as five including dense layer. The dropout value of GRNN is set as 0.9. The cross validation is set as 80/20 and the results are evaluated with the various cross validation.

A. Performance Evaluation

The MCT-GRNN method is proposed in the document indexing for classifying the documents into different categories. The Reuters dataset is worked on to examine the effectiveness of the adopted method in the document indexing. The metrics are evaluated from the results of proposed method and weighed against with the deep learning techniques of LSTM and DNN.

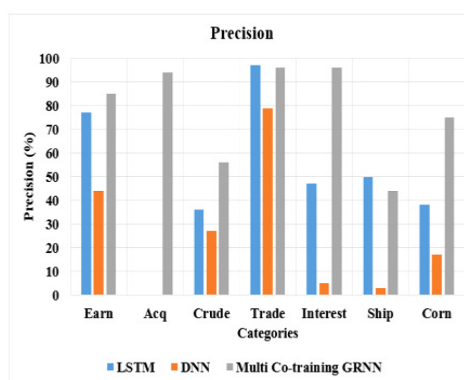

Fig. 2. Comparison of PRECISION on Dataset.

The proposed MCT-GRNN method with multi co-training is measured with precision for the different categories in the Reuters dataset and shown in the Fig. (2). It exhibits that the method is efficient in most of the categories. It has higher precision value in the document indexing except Trade categories. In trade categories, the LSTM has the higher performance. In the Acq categories, there are only 98 instances and the existing method doesn't category the data in the Acq due to less number of data and irrelevant features. The proposed method analyzes the data and provides the classification in Acq categories.

The recall value is measured for the different number of categories in the Reuters dataset and compared with the existing method such as LSTM, and DNN. The proposed MCT-GRNN method has higher recall value in most of the categories. As low number of instance present in the Acq category, the existing method doesn't classify the data in the Acq category. The LSTM has the higher performance in the crude category. The proposed MCT-GRNN has the higher precision value than other two existing methods. It indicates that the method suitable to apply for efficient document indexing.

Fig. (4) Identifies that the adopted MCT-GRNN method has higher F1-Score value compared to the existing methods. The F1-Score is higher for the Acq category.

The proposed MCT-GRNN method and the existing method is tested on the different cross-validation such as 80:20, 70:30 and 60:40 training and testing data, as shown in Fig. (5). The Reuters dataset is used analyze the effectiveness of the proposed and existing technique in the different cross-validation.

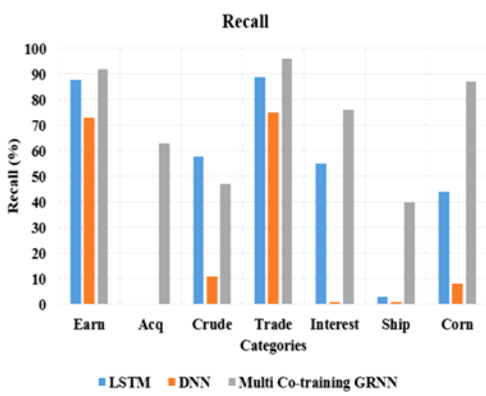

Fig. 3. Comparison of RECALL on Dataset.

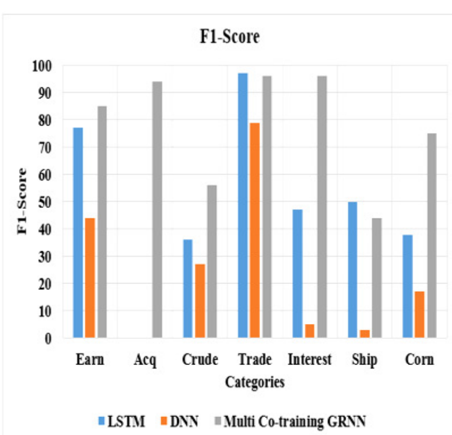

Fig. 4. F1 Score of MCT-GRNN against state-of-art Methods.

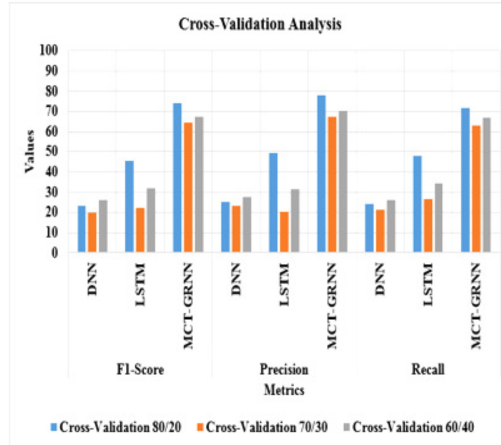

Fig. 5. Performance Analysis of Proposed Method at Different Cross-Validations.

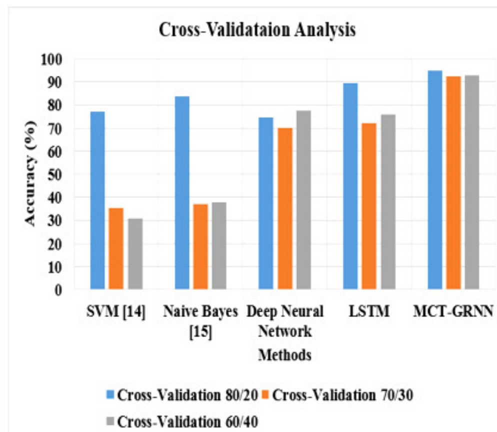

Fig. 6. Comparative Analysis at Various Cross Validations.

The average value of precision, recall and F1-Score is measured for 7 categories on Reuter's datasets. The result shows that the adopted MCT-GRNN method is effective on different cross-validation. The proposed MCT-GRNN method has the higher performance in all the categories in the dataset due to the relevant feature selection. The proposed MCT-GRNN and the existing methods are tested on the different cross-validation analysis on Reuters' dataset, as shown in the Fig. (6). The accuracy is measured for the different methods on various cross- validation. The existing methods such as LSTM and Naive Bayes exhibit good performance when there is high training data. Reduction in training data for Naive Bayes and LSTM affects their performance. The deep learning techniques such as DNN and LSTM have the considerable performance in all three cross-validation analysis. The proposed MCT-GRNN method has higher performance and this shows that the proposed MCT- GRNN perform document indexing with low testing data compared to existing methods.

Therefore, the proposed MCT-GRNN method has the higher performance in document indexing method compared to the existing method. Hence, the proposed MCT-RNN is suitable for applying in the document indexing method.

B. Comparative Study

The existing methods of document indexing like Naive Bayes [12] and LSTM are compared for performance against the proposed method. The proposed MCT-GRNN has the advantages of processing the complex data and analyzing every input for the relevant feature selection. Table 1 shows the comparative study of accuracy among the existing and proposed methods. The proposed MCT-GRNN has higher accuracy of 94.77% in the Reuters dataset, compared to the existing method of SVM 77.3%.

Table 1: The comparative study of the proposed method.

Methods	Accuracy
MCT-GRNN (proposed)	94.77
LSTM	89.59
DNN	74.69
Naive Bayes [12]	83.79

V. CONCLUSION AND FUTURE WORK

Document indexing has been used to classify a number of documents to their categories. Different techniques have been proposed in the document. In this research, MCT-GRNN method is proposed to improve the effectiveness of document indexing technique. The feature selection technique such as TF-IDF and LDA were used to analyze the execution of the system. The MCT-GRNN method is estimated with the existing methods and the deep learning techniques to analyze the performance. The contribution of the MCT-GRNN method is as follows:

– Multi Co-Training method involves in analyzing the features of TF-IDF and LDA to represent the document features. These two features are increasing the classifier performance.

– GRNN analyze the sentence semantic information and the relationship between the sentences in the document. The features' relationships further increase the performance of the classifier.

– The proposed MCT-GRNN method is tested with the Reuters dataset and compared with existing method to analyze the performance. Apart from the existing methods, two deep learning techniques such as DNN, and LSTM were applied and compared with proposed MCT-GRNN method. This shows that the MCT-GRNN exhibits high performance.

– In the Reuters dataset, the existing method shows the low performance in classifying the Acq categories of data and the proposed MCT-GRNN method has the higher performance. This is due to that proposed MCT-GRNN effectively analyzes the features in the dataset.

– The experimental results are evaluated with different cross-validation techniques to test the performance of the given method with various training and testing data. The experimental result shows that the proposed method has the higher performance with less number of testing data.

– The proposed method has higher performance compared to the existing methods in the document indexing. The proposed MCT-GRNN has accuracy of 94.77%, while the existing method of Naive Bayes has the accuracy of 83.79%. The future scope of this method involves in analyzing the more features to increase the performance of the classifiers.

REFERENCES

[1]. Mujtaba, G., Shuib, L., Raj, R. G., Rajandram, R., Shaikh, K., & Al-Garadi, M. A. (2018). Classification of forensic autopsy reports through conceptual graph-based document representation model. *Journal of biomedical informatics, 82*, 88-105.

[2]. Mourão, F., Rocha, L., Viegas, F., Salles, T., Gonçalves, M., Parthasarathy, S., & Meira, W. (2018). NetClass: A network-based relational model for document classification. *Information Sciences*, *469*, 60–78.

[3]. El Barbary, O. G., & Salama, A. S. (2018). Feature selection for document classification based on topology. *Egyptian Informatics Journal*, *19*(2), 129–132.

[4]. Rajbabu, K., Srinivas, H., & Sudha, S. (2018). Industrial information extraction through multi- phase classification using ontology for unstructured documents. *Computers in Industry*, *100*, 37-147.

[5]. Wu, Z., Zhu, H., Li, G., Cui, Z., Huang, H., Li, J.,Xu, G. (2017). An efficient Wikipedia semantic matching approach to text document classification. Information Sciences, 393, 15–28.

[6]. Pavlinek, M., & Podgorelec, V. (2017). Text classification method based on self-training and LDA topic models. Expert Systems with Applications, 80, 83–93.

[7]. Chen, Y., Zhang, H., Liu, R., Ye, Z., & Lin, J. (2019). Experimental explorations on short text topic mining between LDA and NMF based Schemes. *Knowledge-Based Systems, 163*, 1–13.

[8]. Kastrati, Z., Imran, A. S., & Yayilgan, S. Y. (2019). The impact of deep learning on document classification using semantically rich representations. I*nformation Processing and Management, 56*(5), 1618–1632.

[9]. Rao, G., Huang, W., Feng, Z., & Cong, Q. (2018). LSTM with sentence representations for document-level sentiment classification. Neurocomputing, 308, 49–57.

[10]. Kim, D., Seo, D., Cho, S., & Kang, P. (2019). Multi-co-training for document classification using various document representations: TF– IDF, LDA, and Doc2Vec. *Information Sciences, 477*, 15-29.

[11]. Srikanth, D., Sakthivel, S. (2019). Vantage Point Latent Semantic Indexing for multimedia web document search. *Cluster Comput., 22*, 10587–10594.

[12] Chen, J., Huang, H., Tian, S., & Qu, Y. (2009). Feature selection for text classification with Naïve Bayes. *Expert Systems with Applications, 36*(3), 5432-5435.

[13]. Mourão, A., Magalhães, J. (2018). Balancing search space partitions by sparse coding for distributed redundant media indexing and retrieval. *Int J Multimed Info Retr.*, *7*, 57–70.

[14]. Jiang, M., Liang, Y., Feng, X., Fan, X., Pei, Z., Xue, Y., & Guan, R. (2016). Text classification based on deep belief network and softmax regression. *Neural Computing and Applications, 29*, 61-70.

[15]. Liu, G., & Guo, J. (2019). Bidirectional LSTM with attention mechanism and convolutional layer for text classification. *Neurocomputing*, *337*, 325–338.

[16]. Sun, X., Li, C., & Ren, F. (2016). Sentiment analysis for Chinese microblog based on deep neural networks with convolutional extension features. *Neurocomputing, 210*, 227-236.

[17]. Wang, J., Peng, B., & Zhang, X. (2018). Using a stacked residual LSTM model for sentiment intensity prediction. *Neurocomputing, 322,* 93-101.

[18]. Fu, M., Qu, H., Huang, L. & Lu, L. (2018). Bag of meta-words: A novel method to represent document for the sentiment classification. *Expert Systems with*

Applications, 113, 33-43.
[19]. Bhushan, S. B., & Danti, A. (2018). Classification of compressed and uncompressed text documents. *Future Generation Computer Systems, 88,* 614- 623.
[20]. Arroyo-Fernández, I., Méndez-Cruz, C. F., Sierra, G., Torres-Moreno, J. M., & Sidorov, G. (2019). Unsupervised sentence representations as word information series: Revisiting TF–IDF. *Computer Speech & Language, 56*, 107-129.
[21]. Zheng, L., Caiming, Z., & Caixian, C. (2018). MMDF-LDA: an improved multi-modal latent Dirichlet allocation model for social image annotation. *Expert Systems with Applications*, 168-184.
[22]. Yildirim, O., Baloglu, U. B., Tan, R. S., Ciaccio, E. J., & Acharya, U. R. (2019). A new approach for arrhythmia classification using deep coded features and LSTM networks. *Computer methods and programs in biomedicine,* 121-133.
[23]. Liu, Y., Zhang, X., Huang, F., Tang, X., & Li, Z. (2019). Visual question answering via Attention-based syntactic structure tree-LSTM. *Applied Soft Computing, 82,* 105584.

How to cite this article: Bhargavi, Y. K., Murthy, Y. S. S. R. and Rao, O. S. (2020). Multi Co-Trained Gated Recurrent Neural Network for efficient Document Indexing: MCTGRNN. *International Journal on Emerging Technologies*, *11*(3): 435–441.