

Deep Sentiment Learning for Measuring Similarity Recommendations in Twitter Data

S. Manikandan^{1,*}, P. Dhanalakshmi², K. C. Rajeswari³ and A. Delphin Carolina Rani⁴

¹Department of Information Technology, E.G.S. Pillay Engineering College, Nagapattinam, Tamil Nadu, India

²Department of CSSE, Sree Vidyanikethan Engineering College, Tirupati, Andhra Pradesh, India

³Department of Computer Science and Engineering, Sona College of Technology, Salem, Tamil Nadu, India

⁴Department of Computer Science and Engineering, K. Ramakrishnan College of Technology, Samayapuram, Tiruchirappalli, Tamil Nadu, India

*Corresponding Author: S. Manikandan. Email: manikandan@egspec.org

Received: 09 September 2021; Accepted: 29 December 2021

Abstract: The similarity recommendation of twitter data is evaluated by using sentiment analysis method. In this paper, the deep learning processes such as classification, clustering and prediction are used to measure the data. Convolutional neural network is applied for analyzing multimedia contents which is received from various sources. Recurrent neural network is used for handling the natural language data. The content based recommendation system is proposed for selecting similarity index in twitter data using deep sentiment learning method. In this paper, sentiment analysis technique is used for finding similar images, contents, texts, etc. The content is selected based on repetitive comments and trending information. Hash tag is also considered for data collection and prediction. The number tweets are accountable and each character is taken for evaluation. Deep belief network is generated using $512 \times 512 \times 3$ layers system and 1056 trained data, 512 test data that are taken for convolution process. The deep belief network is generated using TensorFlow. TensorFlow is used to simulate the deep learning environments. Semantic analysis is applied for handling Twitter Data. The deep learning processes are classified into clustering, regression and prediction that are evaluated by step by setup approach. The experiments are carried out using similarity index calculation and measuring of accuracy. The results of similarity recommendation are compared with existing method and the results are recorded. Our proposed system gives better results comparing with existing experiments.

Keywords: Deep sentiment learning; similarity recommendations; sentiment analysis; TensorFlow; prediction

1 Introduction

Nowadays, Social media plays a major role for sharing information, prediction and content forwarding. Twitter is one of the best social media platforms to share our views, opinion, thoughts and information at different modes. The amount of data used daily can be increased and huge volume of data are processed



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

everyday [1]. As a user we express our valuable views and thoughts by using opinions polls such as political news, weather forecasting, cultural programs, movies, daily serials, online shopping, advertisement, etc. Users can visit their pages and order the products or poll their opinions. Hence, calculating the similarity index is tedious and need of data processing and analytics techniques [2].

Sentiment analysis is a method to find or classify the results as satisfied, neutral or dissatisfied.

The users can rate their opinions by using the like, dislike icons also they can comment and share. Rating too is considered for evaluation of accuracy factor [3]. According to, Industry 4.0 standards online platform usages are increased and all the common people are using social media for rating, delivering, commenting and ordering. The sentiment analysis method is applied to calculate machine learning inputs, natural language data and contents. In this method, the extraction of data [4], classification of the data [5], labelling [6] and deep learning process are taken into account.

Deep learning method is applied for finding similarity and predicting the results. The feature can be extracted from each contents and images. Similarity index can be measured by using sentiment analysis method [7]. The major challenges are processing the automated extraction, capturing contents and keeping large volume of trained data. The above three factors are mainly considered for evaluation process. Supervised learning techniques are used for measuring trained and labelled data for classification and regression techniques. Unsupervised learning techniques are applied for measuring unlabeled data for clustering and prediction techniques [8].

Plenty of researchers are investigating on handling huge volume of data for the process of prediction and regression analysis [9]. The existing methods are used only for handling any one of the learning techniques either supervised or unsupervised. Currently the different form of data can be used such as multimedia data and natural language data, etc,. So, both learning system are required for calculating sentiment and finding accuracy factor. So, the deep learning is proposed for calculating accuracy and similarity index. In this paper, the twitter data is taken for calculating accuracy and similarity using deep learning techniques.

In twitter, multimedia data, natural language data and NoSQL data are used. So the sentiment analysis is proposed by using deep learning techniques. In this paper, Section 2 describes the various related works, Section 3 explains the deep learning method for finding learning operations. Section 4 gives the deep belief network preparation and algorithm for evaluation. Section 5 explains the experimental setup, results and discussions.

2 Related Works

Sentiment analysis is used to measure the content quality based on opinion poll. R-tool, Weka tool, Ruby on Rails are used to analyze the data based on results and prediction of the further actions. Manikandan et al, proposed a semantic analysis of disease and treatment method for evaluating image annotation values. Data analysis tools are proposed for handling pre-trained data or structured query data. The table of formatted data are analyzed by using real time data analysis tools. Nowadays, non structured and unformatted data are used for various applications such as social media, online shopping, e-learning contents, etc [10,11]

Machine learning and predictive analysis methods are available to handle customer previews, rating and feedbacks. Wong et al, suggested that human vision is also plays an important role for predicting customer reviews and rating. Based on computer vision and repetitive advertisements the users can opt their services using service request mode. In recent years, commercial mobile apps available for processing the requests, reviews and ratings. For example food ordering apps, cab booking services, ticket booking services are trending today based on opinion polls [12].

The deep neural network suggested for handling online and unformatted data for analytical process. Decision trees and dimensional reduction methods are used for convolution and recurrent neural network

operations. Most of the countries are using Twitter for data sharing, commenting, sending wishes and sharing their thoughts. As per the survey by Nivedia 2019, 79% of the whole world are depending Twitter media and tweet everyday. Twitter data can be analyzed by using classification and regression methods. This supervised learning method is used for measuring current polls and decision. Rayung et al, supports that vector machine is suggested for handling images and their restoration operations [13].

Bayesian network and support vector machines are proposed for handling semantic data and it is proposed for tweet data prediction. Various research works are analyzed and tweets are evaluated by using sentiment analysis method [14]. Yung et al, proposed the sentiment analysis method for handling natural language data, lexical elements, content selections and data analysis. So, we need to compare a set of tweets and their pool responses. Based on various literature surveys, the single deep learning framework model is prepared and structured in the simulating environments [15]. TensorFlow is used for simulating the experiments and each data are to be trained. The trained data can be analyzed by using sentiment analysis method in deep learning environments.

3 Deep Learning: A Process

In this section, a dataset is prepared and structured in the simulating environment by using deep belief network. The different datasets are collected from various regions such as Asia, Pacific, Australia, Europe and America. From these data, totally 55,000 tweets, 2,65,000 words and 20,000 vocabularies are selected. Next, the dataset is applied for training process. In each stage, the redundant data, unwanted information, symbols and other irrelevant data are removed. This phase is called as emotional selection of trained dataset.

Let as considered G is the goal or sentiment index which is selected from trained data (d) and input values of x , where as $G \in \{d,x\}$. Fig. 1, from this a set of input is represented as $G(x)$ as $\{x_0, x_1, \dots, x_{n-1}\}$ and sentiment index is calculated as follows,

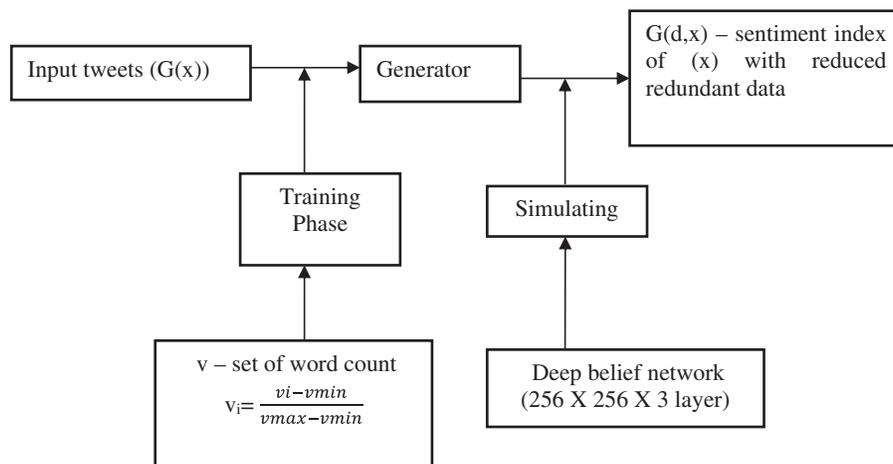


Figure 1: Sentiment index process

So the data distribution factor is calculated by eliminating noisy or redundant data.

$$S(v, g) = G(x) - (\log V(i) - \log Vmax) \tag{1}$$

The deep convolution network is generated based on training phase and data distribution factor shown in Fig. 2. The generator is used to classify the tweet data and select exact copy of the content. The complete task is monitored and it created the deep belief network with input, hidden and output layers.

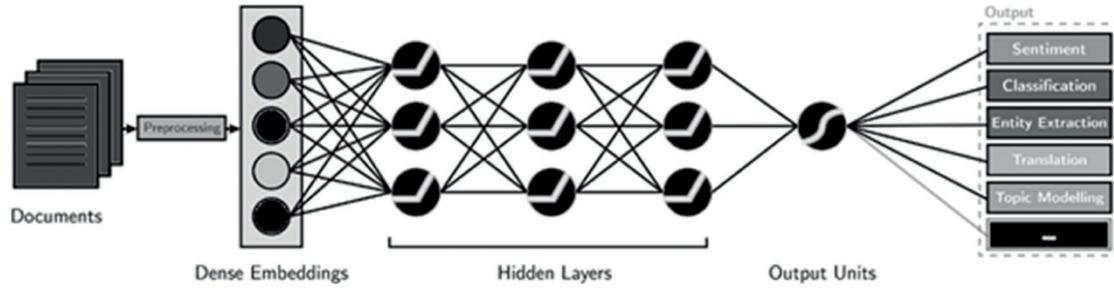


Figure 2: Twitter data pre-processing and hidden layer formation

The above architecture is fully connected model with hidden layers of trained and tested data. So, the mutual execution process is to be done for decomposing the input parameters and their representations. Here c – mutual information, λ – regularized data and D – Distribution factor so

$$V(D, G) = V(S, G) - \lambda_i(S(v, g)) \quad (2)$$

Here the upper and lower bound values are calculated so that the information can be distributed. The hidden node is tested by using the following constraints.

$$\begin{aligned} 0 & \quad i = n \\ V_{\text{hidden}}(i) &= 1 \quad i < n \\ \sum_{i=0}^{n-1} V(i) - Vi(\min D, G) \quad i \leq v(i) \leq n & \quad (3) \end{aligned}$$

From this sentence a vector is created based on tweets. Here we fixed 50 tweets as one sentence. If it exceeds, tweets are sub divided and labelled. Regions can be identified based on dependencies calculation. Here 50 tweets are devised by 25 regions of dimensional words. 3 layer regions segments are formed for processing. In this section the redundant and noisy data can be removed.

In Fig. 3, the dataset is divided into 3-layered distinct datasets. 2 for trained dataset and 1 for test dataset. The region based sentences are framed (and each 2 regions 1 for trained data, another one for test data.)?? So, 6 regions and 3-layered systems are created for deep learning process. Each 25 regions have 1056 trained data, and 512 test data are taken for evaluation.

Input: $(D, G) \equiv G$: *Set of trained and test data 512 X 512 X 3*

Output: Data distribution factor with sentiment index

Step 1: Select coordinate values with data and regions where G – group consist of regions and layers

Step 2: Calculate $V(D, G) = (1/N) \sum_{i=0}^{n-1} S(i) - \sum_{i=0}^{n-1} V(i) - V(\min x)$

Step 3: Find the distribution factor

$$Df_{\text{data}} = \frac{\sum_{i=0}^{n-1} V(i) - S(i)}{N}$$

Step 4: Update the record in layers

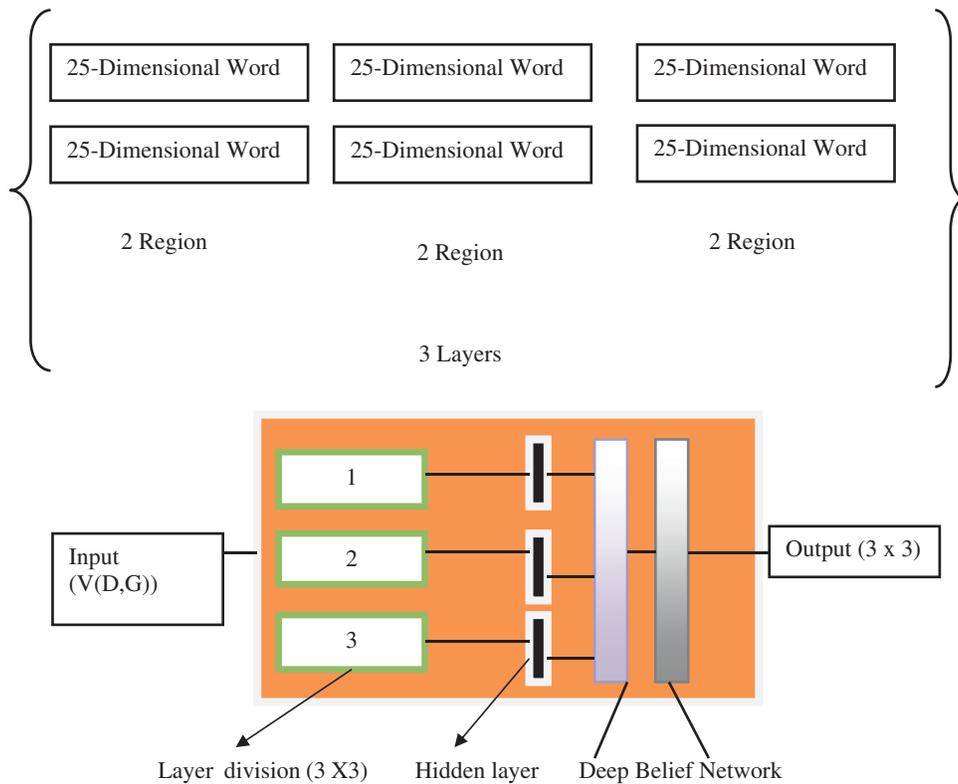


Figure 3: Deep belief network formation

The neural network configuration is done by using TensorFlow shown in Fig. 4. Supported vector machine classifier is added for finding regional and non-regional dataset. Totally 3 layers and 6 regions, 3×3 layers are framed.

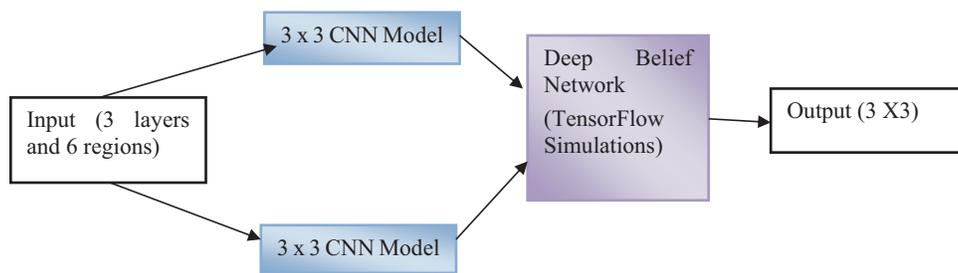


Figure 4: CNN model and deep belief network operations

The proposed deep learning model is used for two-layer operations such as deep representations of input data and trained test dataset. The below Fig. 5 shows the TensorFlow representation of our proposed recommendation systems.

4 Experimental Setup – Genetic Algorithm

The above Fig. 5 shows the experimental setup of TensorFlow results. The input coordinators are shown below in Tab. 1,

Based on the above representation the sentiment index is calculated using accuracy, precision, recall and measure factors.

$$\text{Accuracy} = \frac{V_{max} + V_{hidden}}{N} \tag{4}$$

$$\text{Precision} = \frac{V_{max}}{N - \text{Accuracy}} \tag{5}$$

$$\text{Recall} = 1 - \frac{\text{Accuracy} - S}{N} \tag{6}$$

$$\text{Measure} = \text{Accuracy} + \sum (\text{Precision} * \text{Recall}) - N \tag{7}$$

From this sentiment index is calculated as

$$\text{SI} = \frac{(\text{Accuracy} + \text{Precision}) - N}{\text{Recall} * \text{Measure}} \tag{8}$$

The above formulas are taken for calculating the index, [Tab. 2](#) represents the trained and test data values with iterations.

[Tab. 3](#) shows the sentiment index values for 1056 trained data, 512 test data with 100,50,25 node values of 512 X 512 deep belief network.

Also the result compared with existing method. [Tab. 4](#) shows our proposed sentiment analysis method which gives accurate percentage.

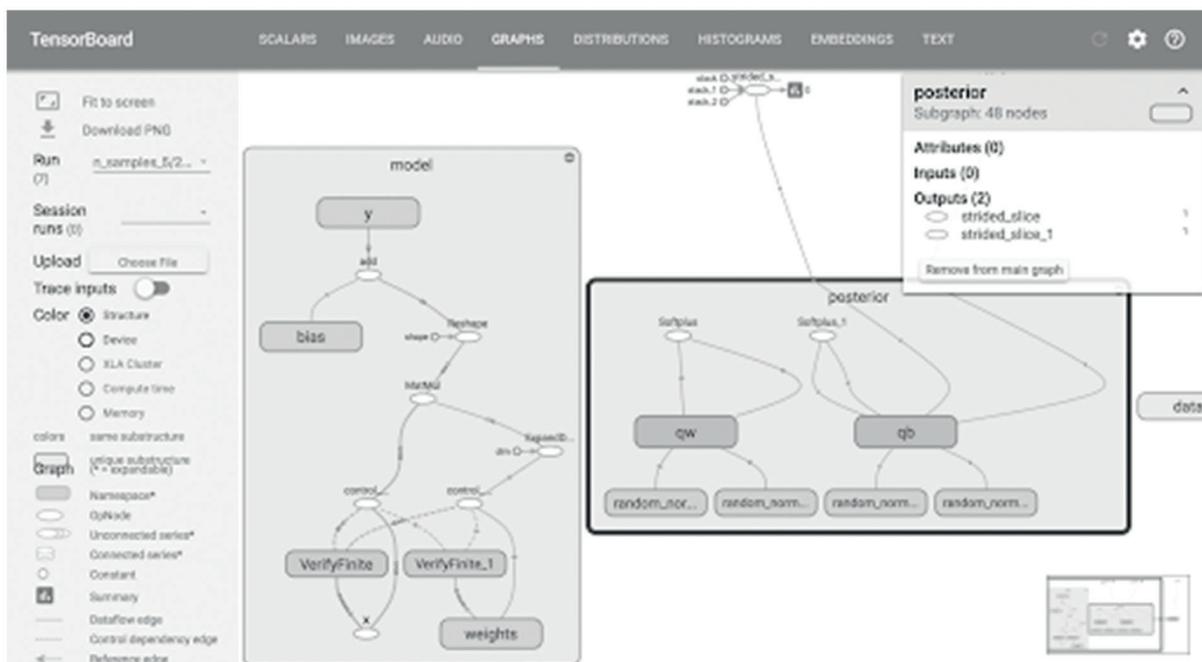


Figure 5: (Continued)

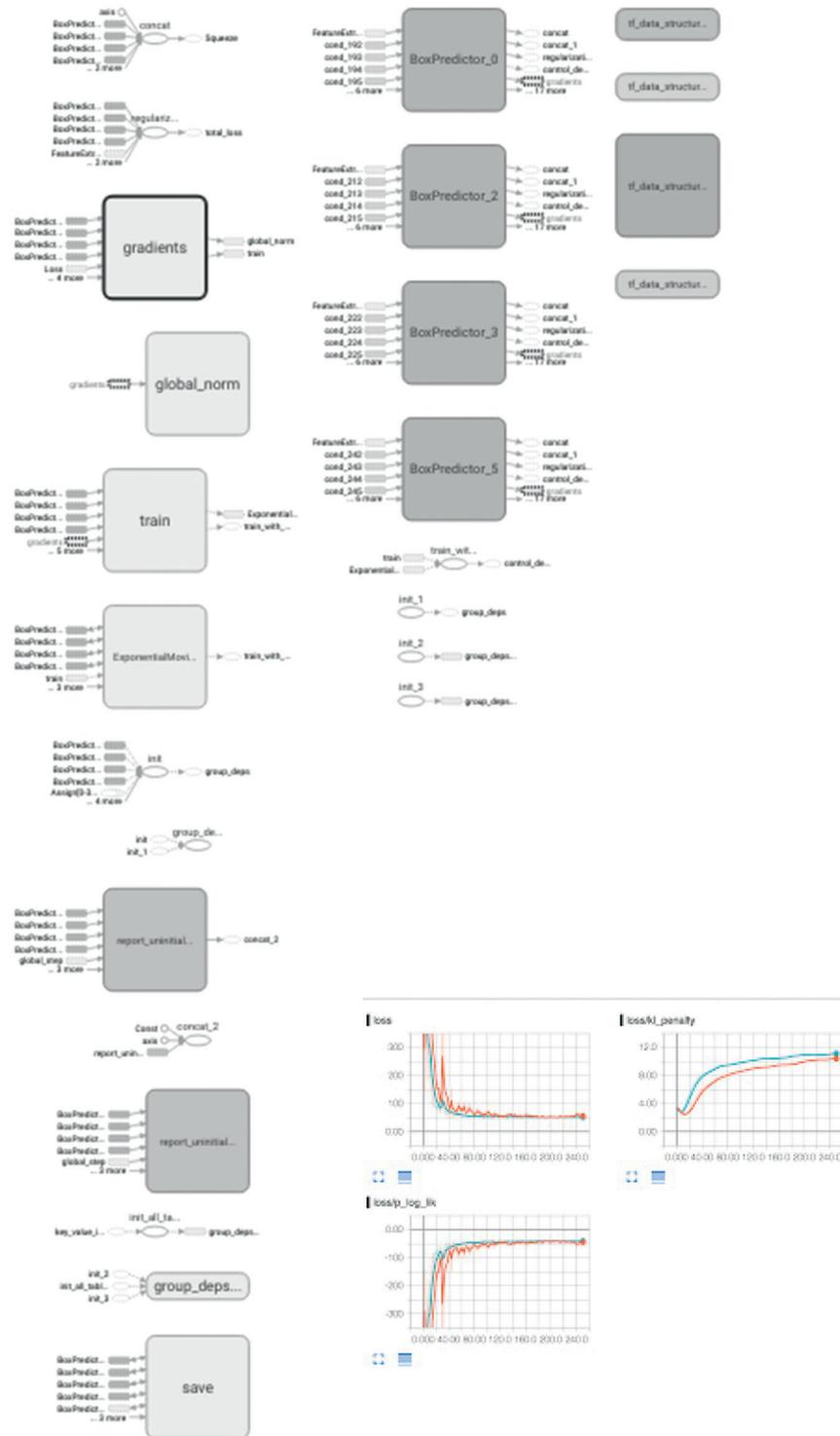


Figure 5: Tensor flow–Experimental setup

Table 1: Major considerations of input, output and hidden components

Trained data and Input values (D and X)	Generator model (G)
Input and tool	512 × 512 Convolution Model and TensorFlow
Layers	3 layers and 6 region deep belief network
Dropout point	0.5 connected layer
Hidden nodes	8 point scale model
GPU	3.75 Ghz Deep Decision tree model
Dimensionality	100 × 50 × 25
Tweets	55,000 tweet, 2,65,000 words and 20,000

Table 2: Iterations and accuracy, precision, recall, measure value of 1056 trained data, 512 test data

Iterations	Nodes	Accuracy	Precision	Recall	Measure
1	100,50,25	0.98,0.94,0.95	0.15,0.14,0.16	0.87,0.84,0.85	98,97,94
2	100,50,25	0.89,0.92,0.91	0.21,0.24,0.21	0.88,0.87,0.83	97,94,96
3	100,50,25	0.92,0.91,0.91	0.22,0.16,0.18	0.79,0.82,0.94	88,89,91
4	100,50,25	0.91,0.94,0.92	0.17,0.19,0.14	0.81,0.79,0.82	92,91,92
5	100,50,25	0.93,0.94,0.95	0.14,0.15,0.18	0.91,0.87,0.91	94,94,94
6	100,50,25	0.91,0.94,0.92	0.17,0.19,0.14	0.81,0.79,0.82	92,91,92
7	100,50,25	0.92,0.91,0.91	0.22,0.16,0.18	0.79,0.82,0.94	88,89,91
8	100,50,25	0.91,0.94,0.92	0.17,0.19,0.14	0.81,0.79,0.82	92,91,92
9	100,50,25	0.93,0.94,0.95	0.14,0.19,0.18	0.91,0.87,0.91	94,92,94
10	100,50,25	0.89,0.92,0.91	0.21,0.24,0.21	0.88,0.87,0.83	97,94,96
11	100,50,25	0.91,0.94,0.92	0.17,0.19,0.14	0.81,0.79,0.82	92,91,92
12	100,50,25	0.91,0.94,0.92	0.17,0.19,0.14	0.81,0.79,0.82	92,91,92
13	100,50,25	0.91,0.94,0.92	0.17,0.19,0.14	0.81,0.79,0.82	92,91,92
14	100,50,25	0.98,0.94,0.95	0.15,0.14,0.16	0.87,0.84,0.85	98,97,94
15	100,50,25	0.91,0.94,0.92	0.17,0.19,0.14	0.81,0.79,0.82	92,91,92

Table 3: Sentiment index calculation of twitter data

Iterations	Nodes	Sentiment index	Average
1	100,50,25	94,92,94	93
2	100,50,25	92,93,92	93
3	100,50,25	88,90,91	90
4	100,50,25	92,91,89	91
5	100,50,25	94,95,96	95
6	100,50,25	94,94,95	95

(Continued)

Iterations	Nodes	Sentiment index	Average
7	100,50,25	95,89,93	93
8	100,50,25	95,92,92	93
9	100,50,25	92,92,91	92
10	100,50,25	93,94,94	93
11	100,50,25	92,89,91	91
12	100,50,25	91,92,94	93
13	100,50,25	91,90,92	91
14	100,50,25	94,93,93	93
15	100,50,25	94,92,91	93

Table 4: Proposed system vs. existing methods

Methods	Model	Index accuracy	Node
Sweeper Rep	CNN	87	100, 50, 25
Decision tree	RNN	83	100, 50, 25
SemanticNet	Machine learning	78	100, 50, 25
IndexZero	Machine learning	86	100, 50, 25
Sentiment	Deep belief network	93	100, 50, 25

5 Conclusion

In this paper, the twitter data are analyzed by using sentiment index method in deep learning environments. Deep belief network is generated using $512 \times 512 \times 3$ -layered system and 1056 trained data, 512 test data. TensorFlow simulator is used for simulating recommendation system. The similarity recommendations are taken by using deep belief network results. Twitter data are analyzed and it achieved 93% accuracy of average sentiment index. Also, the results are compared with existing methods and our proposed systems that have good accuracy factors. In future the same method can be applied for social media applications.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] K. Sani and D. Goularas, "Evaluation of deep learning techniques in sentiment analysis from twitter data," in *Proc. of the 19th IEEE Conf. on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)*, China, pp. 256–268, 2019.
- [2] Betul, Z. Koyun, M. Demir and G. Aydin, "A visual similarity recommendation system using generative adversarial networks," in *Proc. of the 19th IEEE Int. Conf. on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)*, China, pp. 486–497, 2019.

- [3] S. Manikandan, P. Dhanalakshmi, S. Priya and A. Mary Odilya Teena, "Intelligent and deep learning collaborative method for E-learning educational platform using tensorflow," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 10, E-ISSN: 1309-4653, pp. 2669–2676, 2021.
- [4] Szegedy and Christian, "Rethinking the inception architecture for computer vision," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, China, vol. 13, no. 8, pp. 925–937, 2016.
- [5] J. Pennington, R. Socher and C. Manning, "Glove: Global vectors for word representation," in *Proc. 2014 Conf. on Empirical Methods Natural Language Processing*, Malaysia, vol. 15, no. 2, pp. 1532–1543, 2014.
- [6] S. Manikandan and M. Chinnadurai, "Evaluation of students' performance in educational sciences and prediction of future development using tensorflow," *International Journal of Engineering Education*, vol. 36, no. 6, pp. 1783–1790, 2020.
- [7] C. Chen and P. Xi, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 5, no. 3, pp. 58–68, 2016.
- [8] S. Manikandan, A. Delphincarolarani, C. Rajeswari, T. Suma and D. Sivabalaselvamani, "Recognition of font and tamil letter in images using deep learning," *Applied Computer Science*, vol. 17, no. 2, pp. 90–99, 2021.
- [9] M. Srivastava and S. Nitish, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2015.
- [10] S. Manikandan and M. Chinnadurai, "Intelligent and deep learning approach OT measure E-Learning content in online distance education," *Online Journal of Distance Education and e-Learning*, vol. 7, no. 3, pp. 199–204, 2019.
- [11] Y. H. Peng, D. Y. Chen, L. H. Chen, J. Y. Yu and M. J. Bao, "The machine learning based finite element analysis on road engineering of built-in carbon fiber heating wire," *Intelligent Automation and Soft Computing*, vol. 24, no. 3, pp. 531–539, 2018.
- [12] C. R. Kumarl and V. E. Jayanthi, "A novel fuzzy rough sets theory based CF recommendation system," *Computer Systems Science and Engineering*, vol. 34, no. 3, pp. 123–129, 2019.
- [13] D. J. Zeng, Y. Dai, F. Li, J. Wang and A. K. Sangaiah, "Aspect based sentiment analysis by a linguistically regularized CNN with gated mechanism," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 5, pp. 3971–3980, 2019.
- [14] Y. Yin, S. Yangqiu and M. Zhang, "NNEMBs at semeval-2017 Task 4: Neural twitter sentiment Classification: A simple ensemble method with different embeddings," in *Proc. of the 11th Int. Workshop Semantics Evaluation*, China, vol. 2, no. 1, pp. 621–625, 2017.
- [15] S. Palani Murugan, M. Chinnadurai and S. Manikandan, "Tour planning design for mobile robots using pruned adaptive resonance theory networks," *Computers, Materials & Continua*, vol. 70, no. 1, pp. 181–194, 2021.