International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Convolutional Neural Network for Customer's Opinion on Amazon Products

T.Kumaravel, B.Bizu

Abstract: ***Opinion mining and sentiment analysis are valuable to extract the useful subjective information out of text documents. Predicting the customer's opinion on amazon products has several benefits like reducing customer churn, agent monitoring, handling multiple customers, tracking overall customer satisfaction, quick escalations, and upselling opportunities. However, performing sentiment analysis is a challenging task for the researchers in order to find the users sentiments from the large datasets, because of its unstructured nature, slangs, misspells and abbreviations. To address this problem, a new proposed system is developed in this research study. Here, the proposed system comprises of four major phases; data collection, pre-processing, key word extraction, and classification. Initially, the input data were collected from the dataset: amazon customer review. After collecting the data, pre-processing was carried-out for enhancing the quality of collected data. The pre-processing phase comprises of three systems; lemmatization, review spam detection, and removal of stop-words and URLs. Then, an effective topic modelling approach Latent Dirichlet Allocation (LDA) along with modified Possibilistic Fuzzy C-Means (PFCM) was applied to extract the keywords and also helps in identifying the concerned topics. The extracted key-words were classified into three forms (positive, negative and neutral) by applying an effective machine learning classifier: Convolutional Neural Network (CNN). The experimental outcome showed that the proposed system enhanced the accuracy in sentiment analysis up to 6-20% related to the existing systems.***

Index-values:Convolutional neural network; latent dirichlet allocation; lemmatization; modified possibilistic fuzzy c-means; Adam optimization algorithm; sentiment analysis.

I. INTRODUCTION

Presently, online shopping is growing effectively in the fields of e-commerce in which the customer experiences are achieved after the purchase over interest. The non-objective components assume a significant job in the improvement and action of social developments through online media, which have the principle sway on viral spreading. The multi-investigation approach reveals new insight into genuine issues of the useful structure in the online interpersonal organizations that relies upon the measure of feeling go with the messages [1-2]. The data is effectively open and intellectual exertion diminishes the heuristics, which assumes an imperative job in the online buy choice to have an appropriate deals and sheds understanding on the significance of evaluations and feelings over a successive basic leadership process [3-4].

Manuscript published on 30 September 2019
* Correspondence Author
T.Kumaravel*, Asst. Prof, Dept of CSE, Kongu Engineering College, Perundurai, Tamilnadu, India
B.Bizu, Asst. Prof, Dept of CSE, Kongu Engineering College, Perundurai, Tamilnadu, India

The assumption investigation for the social web is profoundly helpful for revealing insight into the job of feeling in both on the web and disconnected [5-10].

The finding of mindfulness impact for online client audits is astounding as online surveys under the examination are presented on a similar site and are not expected to expand item mindfulness [11-14]. The differential impact of consumer reviews across product category suggests online marketing on product and consumer characteristics. The social web texts identify the spurious sentiment patterns that related to the topics rather than sentimental phenomena,which is vigorous enough to be applied to a wide range of social web settings [15-19].

Nowadays, sharing data becomes a trend among business partnerships, as it is supposed to be a mutually beneficial way of increasing productivity. In order to predict the customer's opinion on Amazon products, a new machine learning classification approach is implemented with an appropriate clustering approach. In this research, sentiment analysis was performed on the reputed dataset i.e. Amazon Customer Review dataset. After the collection of input data, pre-processing was carried-out by applying lemmatization, review spam detection, and removal of stop-words and URLs from the input data. The lemmatization converts the words of a sentence into dictionary form in order to extract the proper lemma. In addition, review spam detection finds the customers' untruthful opinions like positive spam reviews and negative spam reviews. Then, the pre-processed data were used to extract the key words by applying an effective topic modelling approach LDA. In addition, PFCM was used to cluster the extracted key words on the basis of amazon products. Here, quantum inspired methodology was used for obtaining the correct cluster number (three). The quantum inspired methodologyoperates on the smallest information representation named as quantum bit (qubits). The output of modified-PFCM was given as the input for CNN classifier in order to classify the opinions of the customers for amazon products: neutral, positive and negative. Here, a new optimization algorithm: Adam optimization algorithm is applied for optimizing the moments of feature values in CNN- Long Short Term Memory Networks (LSTM) classifier that helps to find the condition of best accuracy, and also to minimize the estimated error. Finally, the proposed system performance was compared with other existing systems by means of precision, recall, f-measure, classification accuracy, and Area under Curve (AUC).

This research paper is arranged as follows. Several recent papers on sentiment analysis are reviewed in section 2. Problem statement about the existing methods are illustrated in section 3. Detailed explanation of the proposed system is given in section 4.

Section 5 illustrates about the quantitative analysis and comparative analysis of proposed system. The conclusion is done in section 6.

II. LITERATURE REVIEW

Researchers developed numerous methodologies on dissimilar stages of sentiment analysis. In this literature section, a brief review of some important contributions to the existing literatures is presented in the table 1.

Table 1. Review of Existing literatures

Reference	Methodology	Advantage	Limitation
M. Giatsoglou, et al. [20]	Hybrid machine learning approaches	The developed system adopts a machine learning approach with textual documents for training a polarity classification model. Here, many document vector representation approaches were applied; lexicon-based, word embedding based and hybrid vectorizations for improving the sentiment classification task.	A number of features were considered for data feature extraction, which may decrease the accuracy of data classification.
T. Alsinet, et al. [21]	Argumentative approach	An important aspect of the developed system was the notion of an uncertainty threshold that characterizes how much uncertainty in probability values were willing to reject criticism and support relationships.	The major drawback of this research work was the developed approach did not concentrate on the pre-processing of twitter data.
A. Balahur, and J.M. Perea-Ortega, [22]	Multilingual sentiment analysis system	In this research paper, hybrid features, multilingual, machine-translated data attained better relevant features for sentiment classification and also increase the precision of sentiment analysis systems.	In some cases, the developed system inefficient for large scale databases like amazon customer review, twitter-sanders-apple 2 datasets that leads to two major problems; computational complexity and poor data classification.
M. Bouazizi, and T. Ohtsuki, [23]	A Pattern based methodology for multi-class conclusion examination in twitter information	This approach is scalable and classifies texts into more classes. Here, a new tool: SENTA was used to select out a wide variety of features that fit the most for the application, to run the classification, through an easy-to-use graphical user interface.	Mostly, the binary classification supports only structural data. It does not support unstructured data, which was considered as one of the major drawbacks in this research work.
D. Yu, et al, [24]	Hierarchical topic modeling	This system automatically mines the hierarchical dimension of tweets' topics that helps to improve the classification accuracy. The experimental results demonstrate that the developed approach outperformed other current topic models in mining and constructing the hierarchical dimension of tweeter topics.	As a drawback, LDA (topic modeling approach) has limitation to support latent sub-topics within a determined label or any global topics.
S. Bharathi, et al, [25]	Hybrid features with sentence level sentiment score algorithm	This system achieved high precision in stock market prediction by combining the Sensex points with really simple syndication news feeds and tweets.	In this research paper, performing classification was very difficult, when the dimensions of the collected data values were very high.
H. Saif, et al, [26]	SentiCircles, (lexicon based approach)	This methodology offers a fixed and static earlier slant polarities of words paying little heed to the unique situation. SentiCircles considers the co-event examples of words in various settings in tweets to catch the semantics and update the pre-doled out quality and extremity in notion vocabularies as needs be.	The lexicon based approaches need more repeated words for updating the pre-assigned strength and polarity of the words.
Y. Ren, et al, [27]	LDA and support vector machine	Here, LDA automatically extracts the key-words from the collected data and then the support vector machine classified the twitter sentiments. The experiment outcome showed that the topic enhanced word embedding was very effective in twitter sentiment classification.	SVM is a binary classifier, which supports only two-class classification.

III. PROBLEM DEFINITION AND SOLUTION

This section describes a problem statement in sentiment analysis and also detailed about how the proposed system gives the solution to the problem statement.

3.1 Expert knowledge is required to select an appropriate classifier

After extracting the key-words from the pre-processed data, classification is carried out to classify the opinions of the customers for amazon products. In sentiment analysis, binary classifiers like Support Vector Machine are a well-known classifier that is designed for the two-class problem. The success of binary classifier depends on the decision boundary that delivers good generalization performance [28, 29, and 30]. The major two problems accomplished in binary classifiers are ineffective in high dimensional data and only applicable for two-class classification. To address these issues, multiclass classification approaches are developed. **Solution:** In this research work, a new classifier: CNN is implemented for multiclass classification. The CNN classifier effectively diminishes the size of resulting dual issue by developing a relaxed classification error bound. In addition, the undertaken classification approach quickly speeds up the training process by maintaining a competitive classification accuracy.

Retrieval Number: C5670098319/2019©BEIESP
DOI:10.35940/ijrte.C5670.098319
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

IV. PROPOSED SYSTEM

In recent periods, sentiment analysis gained much attention among the researchers. It plays a major role in several applications like HealthCare, marketing, retail industry, education, etc. This research study tackles the issue of sentiment polarity categorization that is one of the major issues of sentiment analysis. The input data utilized in this research are online product reviews collected from amazon.com. The proposed system majorly consists of four phases: data collection, pre-processing, key word extraction and classification. Figure 1 shows the work flow of the proposed system. The detailed explanation about the proposed system is described below.

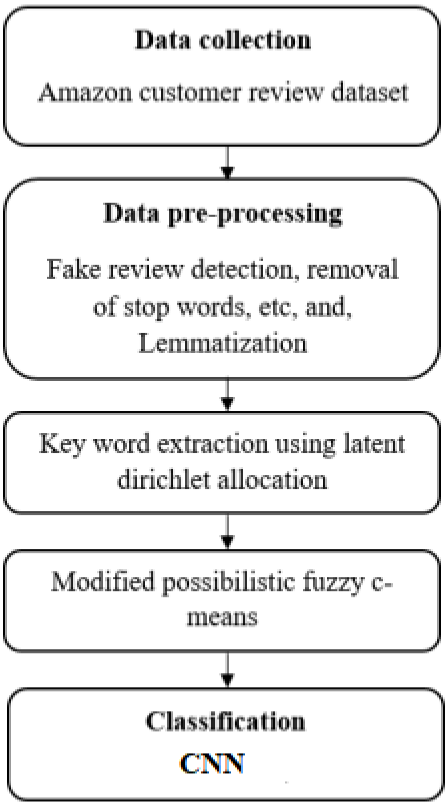


Figure 1. Work flow of proposed system

4.1 Data collection

At first, the input data are collected from the dataset: amazon customer review dataset. It is comprised of customer reviews from amazon website. The time span of the amazon customer review dataset is eighteen years that include approximately thirty-five million reviews up to March 2013. The reviews comprise of product ratings, user information, product information and a plain text review. This dataset contains potential duplicates, due to products whose reviews amazon merges. Table 2 describes the data characteristics of the amazon customer review dataset.

Table 2. Data statistics

Dataset statistics	
Number of users	6,643,669
Number of reviews	34,686,770
Number of products	2,441,053
Time span	June 1995-March 2013
Median no. of words per review	82
Users with>50 reviews	56,772

4.2 Data pre-processing

After the collection of data from amazon customer review dataset, data pre-processing is carried out in order to enhance the quality of collected data. Generally, the raw data contains more noise in terms of stop-words, and URLs, which are all removed effectively from the collected data. In addition, lemmatization technique and review spam detection are applied to further enhance the quality of data.

- **Review spam detection:** The main task of review spam detection is to identify the customers' untruthful opinions like positive spam reviews and negative spam reviews.
- **Lemmatization:** It transforms the words of a sentence into dictionary form. In order to extract the proper lemma, it is essential to analyse each morphological word. An example of lemmatization is denoted in table 3.

Table 3. A sample example of lemmatization

Form	Morphological information	Lemma
Studyin g	Gerund of the verb study	study
Studies	Singular number, third person, present tense of the verb study	study

4.3 Key word extraction using latent dirichlet allocation

After pre-processing the collected data, keyword extraction is carried out by utilizing LDA approach. By and large, the LDA is a probabilistic point model, while each record is indicated as an irregular blend of idle subjects. Every single inactive point is named as a dissemination over a fixed arrangement of words in LDA, which is used to distinguish the basic inert subject structure based on watched information. More often than not, the words are produced in a two-stage process for each archive. In the primary stage, a circulation over themes is haphazardly chosen for each report. In LDA, a word is an unmistakable information from a jargon index{1,..,V}, a progression of N words are.

The LDA is watched for the three-layered portrayal, π and μ parameters are analysed during the age of a corpus. For every single archive, the report level subject factors are researched. Individually, the word level factors are analysed for each expression of the report in LDA. A joint conveyance over arbitrary variable is spoken to as the generative procedure of LDA. The likelihood thickness capacity of the k-dimensional dirichlet irregular variable is dictated by utilizing the Eq. (1). Progressively, the joint appropriation of a point blend and the likelihood of a corpus are assessed by utilizing the Eq. (2) and (3).

$$p(\aleph|\pi) = \frac{\Gamma\left(\sum_{i=1}^{k}\pi_i\right)}{\prod_{i=1}^{k}\Gamma(\pi_i)} \aleph_1^{\pi_1-1} \dots\dots \aleph_k^{\pi_k-1} \quad (1)$$

$$p(\aleph, x, y|\pi, \mu) = p(\aleph|\pi) \prod_{n=1}^{N} p(x_n|\aleph)\, p(y_n|x_n, \beta) \quad (2)$$

$$p(D|\pi, \mu) = \prod_{d=1}^{M} \int p(\aleph_d|\pi) \times \left(\prod_{n=1}^{N_d} \sum_{x_{dn}} p(x_{dn}|\aleph_d) p(y_{dn}|x_{dn}, \mu)\right) d\aleph_d \quad (3)$$

Where $\aleph$ is represented as the document-level topic variables, π is indicated as the dirichlet parameter, M is represented as the document, N is characterised as the number of words, μ is denoted as the topics, x is indicated as the per-word topic assignment, and y is indicated as the observed word.

In a report, the estimation of the back circulation of the concealed variable is a significant inferential assignment of the LDA model. The accurate deduction of the back dissemination of the concealed variable is an immovable issue. The blend of LDA with estimate calculations like Gibbs examining, Markov chain, Laplace, and variational guess are widely used for watchword extraction. The negative and positive catchphrases are removed with an individual weight esteem and these watchwords are put away in lexicon. So as to achieve negative and positive weight esteem, the testing assessment information are facilitated with the word reference in the testing stage. In the wake of acquiring the negative and positive weight esteems, the bunching procedure is done by utilizing the adjusted PFCM calculation.

4.4 Modified possibilistic fuzzy c-means

Clustering is a task that identifies the hidden groups in a set of objects accurately. Also, it is an unsupervised approach, so clustering does not require previous knowledge of both outputs and inputs. In this research study, the PFCM approach is utilized for clustering the membership grade. Generally, the PFCM clustering considers every object as a member of each cluster with a variable degree of "membership function". The similarity between the objects is defined by a quantum inspired method that plays an essential role in obtaining correct clusters, here, the number of clusters is three. The quantum inspired methodology operates on the smallest information representation named as a quantum bit (qubits). The classical bits are represented as "1" and "0" that stores the information at a time. Where a single qubit has the capability to a store number of information at a time with the help of a probability feature. Qubit represents the linear superposition of '1' and '0' bits probabilistically, which is mathematically stated in the Eq. (4).

$$Q = \alpha|0> + \beta|1> \tag{4}$$

Where, α and β are represented as the complex numbers, which appears in two states, state "0" and state "1". Thus, α^2 and β^2 are denoted as the probabilities of a qubit in state "0", and state "1", which is described in the Eq. (5).

$$\alpha^2 + \beta^2 = 1; 0 \le \alpha \le 1, 0 \le \beta \le 1 \tag{5}$$

As mentioned in the Eq. (5), the qubit is denoted as the linear super-position of two states; state "0", and state "1". For instance, one and two qubit systems perform the operation on two and four values. Thus, n-qubit perform the operation on $2n$ values. So, quantum bit individual contains a string of q quantum bits. Let us consider the example of two quantum bits that are represented in the Eq. (6) and (7).

$$Q = \left\langle \begin{matrix} 1/\sqrt{2}|1/\sqrt{2} \\ 1/\sqrt{2}|1/\sqrt{2} \end{matrix} \right\rangle \tag{6}$$

$$Q = \left(\frac{1}{\sqrt{2}} \times \frac{1}{\sqrt{2}}\langle 00\rangle + \frac{1}{\sqrt{2}} \times \frac{1}{\sqrt{2}}\langle 01\rangle + \frac{1}{\sqrt{2}} \times \frac{1}{\sqrt{2}}\langle 10\rangle + \frac{1}{\sqrt{2}} \times \frac{1}{\sqrt{2}}\langle 11\rangle\right) \tag{7}$$

The concept of quantum bits' representation is used for achieving the global optimization in PFCM. Here, PFCM depends on the reduction of the objective function that is mathematically represented in the Eq. (8), (9) and (10).

$$J_{PFCM}(U,T,V) = \sum_{i=1}^{C}\sum_{j=1}^{n}(u_{ij}^{m} + t^{n})d^2(x_j, v_i) \tag{8}$$

Where,

$$\sum_{i=1}^{c}\mu_{ij} = 1, \forall j \in \{1, \dots, n\} \tag{9}$$

$$\sum_{j=1}^{n} t_{ij} = 1, \forall i \in \{1, \dots, c\} \tag{10}$$

Where, $\boldsymbol{V}$ represents the vector of cluster centres, $\boldsymbol{T}$ is characterised as the typicality matrix, $\boldsymbol{U}$ is denoted as the partition matrix, and $\boldsymbol{J_{PFCM}}$ is indicated as an objective function. In this research work, the objective function is obtained by using the degree of membership and the number of cluster centres. Where the degree of membership and the cluster centres are represented in the Eq. (11), (12) and (13).

$$\mu_{ij} = [\sum_{k=1}^{c}(\frac{dx_j,v_i}{dx_j,v_k})^{\frac{2}{m-1}}]^{-1}, 1 \le i \le c, 1 \le j \le n \tag{11}$$

$$t_{ij} = [\sum_{k=1}^{n}(\frac{dx_j,v_i}{dx_j,v_k})^{\frac{2}{n-1}}]^{-1}, 1 \le i \le c, 1 \le j \le n \tag{12}$$

$$v_i = \frac{\sum_{k=1}^{n}(u_{ik}^{m} + t_{ik}^{n})x_k}{\sum_{k=1}^{n}(u_{ik}^{m} + t_{ik}^{n})}, 1 \le i \le c \tag{13}$$

Where, c is indicated as the quantity of group focuses, n is connoted as the quantity of information focuses that are depicted by the co-ordinates(x_j,v_i) and it is used for figuring the separation between bunch focuses and informational collections.

The altered PFCM develops conceivable outcomes and enrolments with ordinary models and bunch communities for each group. Picking the target capacity is a significant part of the exhibition of group strategy for achieving better bunching. Though, the grouping execution depends on the goal work, which is used for bunching. For building up a compelling target work, the accompanying arrangement of necessities is considered. Separation between the groups ought to be decreased. Separation between the information focuses, which apportioned in the groups ought to be diminished.

The attractive quality between the groups and information are displayed by the goal work. Further, the target capacity of changed PFCM is improved by utilizing driven model learning of parameter α. The learning method α is reliant exponential detachment quality between the groups and it is refreshed at each emphasis. The parameter α is spoken to in the Eq. (14).

$$\alpha = exp\,(-min_{i \neq k}\frac{\|v_i - v_k\|^2}{\beta}) \tag{14}$$

Where, β is represented as the sample variance that is represented in the Eq. (15).

$$\beta = \frac{\sum_{j=1}^{n}\|x_j - \bar{x}\|^2}{n} \tag{15}$$

Where, $\bar{x} = \frac{\sum_{j=n}^{n} x_j}{n}$

At that point, a weight parameter is acquainted with figure the normal estimation of α. Each purpose of the database comprises of weight in association with each bunch. Along these lines, the utilization of weight capacity conveys a superior order result, especially on account of boisterous information. The general condition of weight capacity is resolved in the Eq. (16).

$$w_{ji} = \exp\left(-\frac{\|x_j - v_i\|^2}{\left(\sum_{j=1}^{n}\|x_j-\bar{v}\|^2\right)\times c/n}\right) \quad (16)$$

Where, w_{ji} is denoted as the weight function of the point j with the class i. The step by step procedure of modified-PFCM is represented in the Fig. 2, and also effectively explained below,

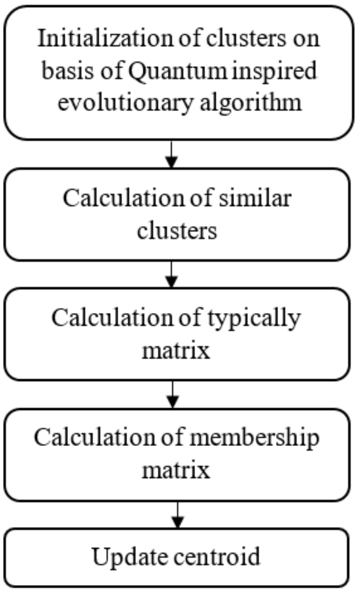


Figure 2. Step by step procedure of modified-PFCM

Instatement: Initially, the quantity of bunches is expected by the client based on quantum-propelled transformative calculation.

Count of similitude separation: After expecting the quantity of groups, assess the separation between information focuses and centroids for every single fragment.

Figuring of commonality grid: After the estimation of separation lattice, the normality networks are assessed that is gotten from the altered PFCM.

Figuring of enrollment grid: Evaluate the participation network M_ik by surveying the enrollment estimation of information point that are accumulated from the altered PFCM.

Update centroid: After creating the bunches, centroid modernization is refreshed.

The relative procedure is performed over and over till the modernized centroids of every single bunch ends up indistinguishable in progressive emphasess. Subsequent to extricating the watchwords, information characterization is done by utilizing CNN classifier.

4.5 Classification of data using CNN classifier

Generally, CNN-LSTM is a multi-layer feed forward network, which is designed to recognize the features in the sentiment data. The neurons in CNN-LSTM classifier considers a small portion of the data that is called as sub-data. Then, the respective sub-data are used for feature extraction, for instance, a feature may be a vertical line, arch, or circle. The features are captured by the respective feature maps of the network. A combination of features is utilized to classify the data. In addition, multiple different feature maps are used to make the network more robust. In this research, an optimization algorithm: Adam optimization algorithm is used to optimize the moment of features values from one layer to another layer. Due to this action, the unwanted convolutions happening in the convolutional layer are avoided. A few major advantages of Adam optimization algorithm are; works effectively even with a little tuning of hyper parameters and relatively low memory requirements. Figure 3 represents the general architecture of CNN classifier.

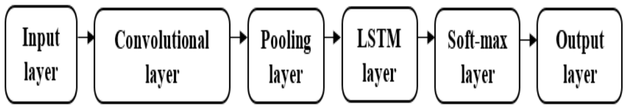


Figure 3. General architecture of CNN classifier

The convolutional layer is a primary layer in CNN classifier, which extracts the local information of the data. Moreover, convolutional operation improves the input features and reduce noise interference. The mapping operation in the convolution process is mathematically expressed in Eq. (17).

$$x_j^l = f_c\left(\sum_{i\in M_j} x_i^{l-1} \times k_{i,j}^l + \theta_j^l\right) \quad (17)$$

Where, x_j^l is specified as the j^{th} mapping set of convolutional layer l, x_i^{l-1} is denoted as thei^{th} feature set indicating in the $(l-1)$ convolutional layer, and $k_{i,j}^l$ is indicated as the convolutional kernel between the i^{th} feature set andj^{th} mapping set in the convolutional layer l. The variable θ_j^l is represented as bias and f_c is denoted as activation function. The next step is the pooling process, which reduce the possibility of over-fitting during training process. The pooling process is mathematically denoted in Eq. (18).

$$x_j^l = f_p\left(\beta_j^l down\left(x_i^{l-1}\right) + \theta_j^l\right) \quad (18)$$

Where, $down(.)$ is represented as the downsampling approach from layer $(l-1)$ to layerl^{th}, θ_j^l and β_j^l are indicated as the additive bias and multiplicative bias, and $f_p(.)$ is represented as the activation function. Generally, the pooling process is sub-divided into two types such as, average and maximum pooling. The final pooling layer (matrix features) are arranged to form a rasterization layer, which is further connected with the fully connected layer. The output of node j is mathematically stated in the Eq. (19).

$$h_j = f_h\left(\sum_{i=0}^{n-1} w_{i,j} x_i - \theta_j\right) \quad (19)$$

Where, $w_{i,j}$ is denoted as the connection weight of input vectorx_i, θ_j is stated as the node threshold, and $f_h(.)$ is represented as the activation function. The next layer is the LSTM layer that helps to capture the sequential data by considering the prior data.

Retrieval Number: C5670098319/2019©BEIESP
DOI:10.35940/ijrte.C5670.098319
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

This layer considers the output vectors from the pooling layer as inputs. The LSTM layer has a number of cells or units and the input of every cell is the output from the pooling layer. The final output of this layer has a similar number of units in the network.

If the LSTM layer deals with the multiclass issue, the softmax classifier is utilized in the fully connected layer. The loss function of softmax classifier is denoted in Eq. (20).

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m}\sum_{j=1}^{k} l\{y^{(i)} = j\} \log \frac{e^{\theta_j^l}}{\sum_k e^{\theta_k^l}}] \quad (20)$$

Where, $e^{\theta_j^l}$ is represented as the input of j^{th}neuron in the l layer, $\sum_k e^{\theta_k^l}$ is denoted as the input of all the neurons, $\frac{e^{\theta_j^l}}{\sum_k e^{\theta_k^l}}$is indicated as the output of j^{th} neuron, e is stated as the constant, and $l(.)$ is represented as the indictor function. If the value in the brace is true, the result of the indicator function is one. If the value in the brace is false, the result of the indicator function is zero. Then, add the rule items in $J(\theta)$ to prevent from falling into local optimum.The loss function of softmax classifier $J(\theta)$ after adding the rule items is mathematically represented in the Eq. (21).

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} l\{y^{(i)} = j\} \log \frac{e^{\theta_j^l}}{\sum_k e^{\theta_k^l}}\right] + \frac{\rho}{2}\sum_{i=1}^{k}\sum_{j=0}^{n}\theta_{ij}^2 \quad (21)$$

Where,$\frac{\rho}{2}\sum_{i=1}^{k}\sum_{j=0}^{n}\theta_{ij}^2$ is represented as the weighted term that helps to stabilize the excessive parameters in the training set. In proposed classifier, each layer contains a Reluctant Linear Unit (ReLU) activation function for activating the neurons in each layer. The ReLU activation function makes the proposed classifier (CNN) redundant, because it effectively solves the exploding and vanishing gradient problems completely. At last, the results are obtained in two forms such as, positive and negative.

V. EXPERIMENTAL RESULT AND DISCUSSION

This section detailed about the experimental result and discussion of the proposed system and also detailed about the performance metric, experimental setup, quantitative analysis and comparative analysis. The proposed system was implemented using Python with 4 GB RAM, 1 TB hard disk, 3.0 GHz Intel i5 processor. The performance of the proposed system was compared with other classification methods and existing research papers based on amazon customer review dataset in order to assess the effectiveness of proposed system. The performance of proposed system was evaluated in terms of precision, recall, classification accuracy, f-measure and AUC.

5.1 Performance measure

The presentation measure is characterized as the customary estimation of results and results that creates dependable data about the viability of the proposed framework. Likewise, the presentation measure is the way toward revealing, gathering and breaking down data about the exhibition of a gathering or person. The scientific condition of exactness, f-measure, accuracy, and review are signified in the Eq. (22), (23), (24), and (25).

$$Accuracy = \frac{TN+TP}{TP+TN+FN+FP} \times 100 \quad (22)$$

$$F-measure = \frac{2TP}{(2TP+FP+FN)} \times 100 \quad (23)$$

$$Precision = \frac{TP}{(FP+TP)} \times 100 \quad (24)$$

$$Recall = \frac{TP}{(FN+TP)} \times 100 \quad (25)$$

Where, TP is signified as true positive, $TN$is indicated as true negative, FP is specified as false positive, and FN is indicated as false negative.

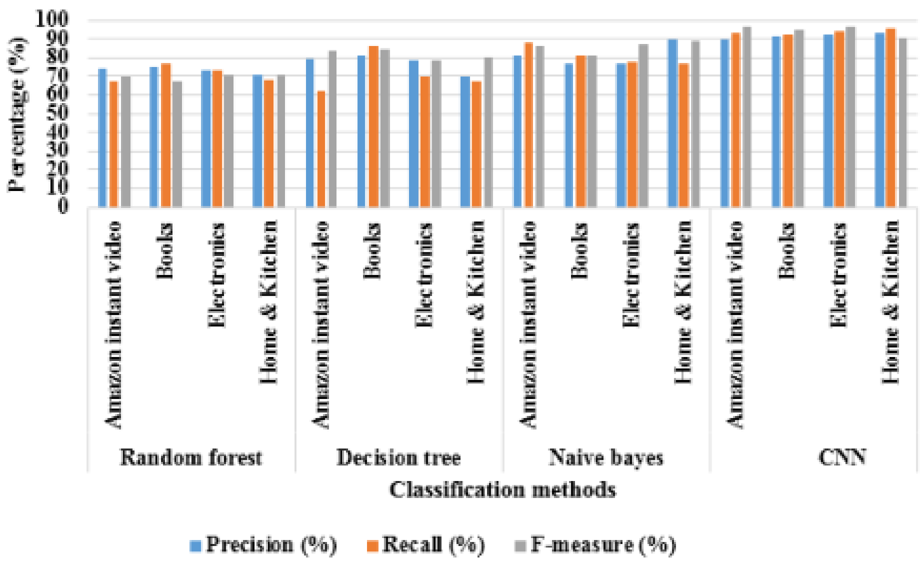
5.2 Quantitative analysis

Amazon customer review dataset is used for evaluating the performance of the proposed system and other existing classification approaches like random forest, decision tree, and Naive bayes. In this research study, the collected data is classified into three forms: positive, negative and neutral classes. In tables 4, 5, 6, and 7, the performance evaluation of the proposed system and existing classification approaches are evaluated in terms of precision, recall, accuracy, f-measure and area under a curve. Here, the performance evaluation is validated with 80% training of data and 20% testing of data. Among 2,441,053 amazon products, eight products are considered for experimental investigation such as, amazon instant video, books, electronics, home & kitchen, movie review, media, kindle, and camera.

Here, tables 4 and 5 shows the performance investigation of proposed and existing classification methods for four amazon products; amazon instant video, books, electronics, home & kitchen. The average classification accuracy of proposed classifier is 92.85% and the existing classification approaches (random forest, decision, and Naive Bayes) achieved 75.44%, 73.85, and 85.76% of classification accuracy. Correspondingly, the average precision, recall, f-measure and area under curve of the proposed classifier is better than the existing classification approaches, because the proposed system effectively calculates the linear and non-linear properties of collected data and also significantly preserves the quantitative relationship between the high and low level features. The graphical comparison of proposed and existing classification methods for the amazon products; amazon instant video, books, electronics, home & kitchen is represented in Fig. 4, and 5.

Table 4. Performance analysis of proposed and existing classifiers in light of precision, recall, and f-measure

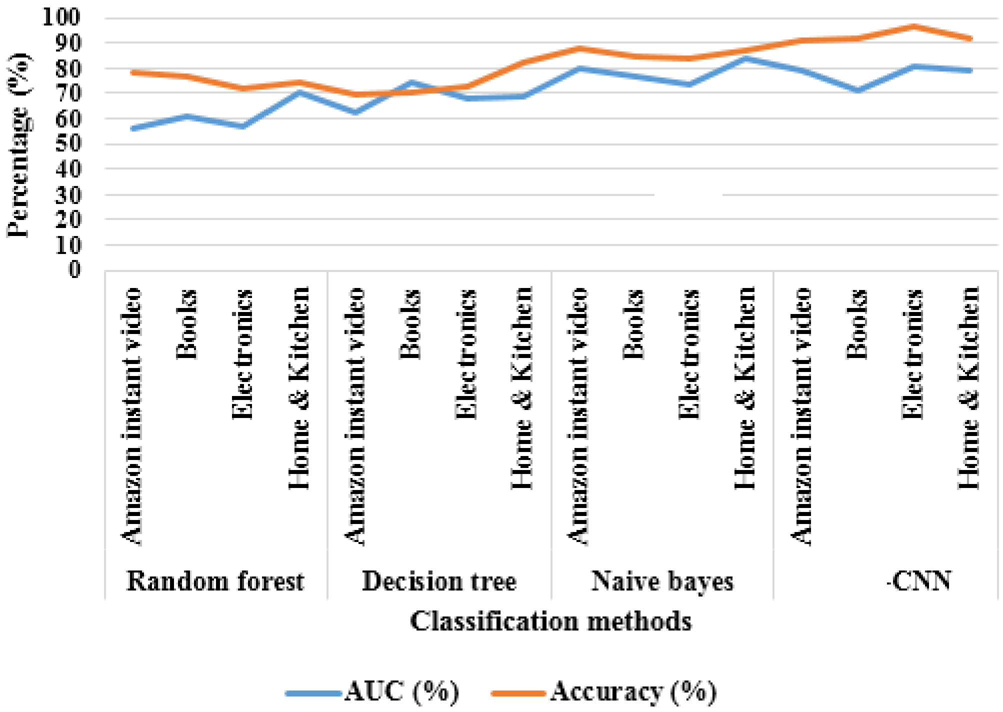
Classifiers	Class	Precision (%)	Recall (%)	F-measure (%)
Random forest	Amazon instant video	74	67.43	69.65
	Books	75.09	76.91	67.38
	Electronics	73.34	73.42	71.15
	Home & Kitchen	70.98	68.34	70.78
Decision tree	Amazon instant video	79.74	62.10	83.47
	Books	81.56	86.32	84.68
	Electronics	78.72	70.02	79.26
	Home & Kitchen	70	67	79.97
Naive bayes	Amazon instant video	81.23	88	86.45
	Books	76.45	81.24	80.87
	Electronics	77	77.76	86.74
	Home & Kitchen	89.98	77.02	89
CNN	Amazon instant video	90	93.44	96.89
	Books	91.11	92.49	94.84
	Electronics	92.67	94	96.69
	Home & Kitchen	93.33	95.55	90.67



Figures 4. Graphical comparison of proposed and existing classifiers by means of precision, recall, and f-measure

Table 5. Performance analysis of proposed and existing classifiers in light of AUC and accuracy

Classifiers	Class	AUC (%)	Accuracy (%)
Random forest	Amazon instant video	56.02	78.09
	Books	60.98	77
	Electronics	56.8	72.24
	Home & Kitchen	70.07	74.44
Decision tree	Amazon instant video	62.34	69.78
	Books	74.61	70.48
	Electronics	68	73
	Home & Kitchen	69.023	82.17
Naive bayes	Amazon instant video	79.87	88
	Books	76.95	84.31
	Electronics	73.6	83.9
	Home & Kitchen	83.65	86.86
CNN	Amazon instant video	78.75	90.65
	Books	71.11	92
	Electronics	80.64	96.87
	Home & Kitchen	79.26	91.94



Figures 5. Graphical comparison of proposed and existing classifiers by means of AUC and accuracy

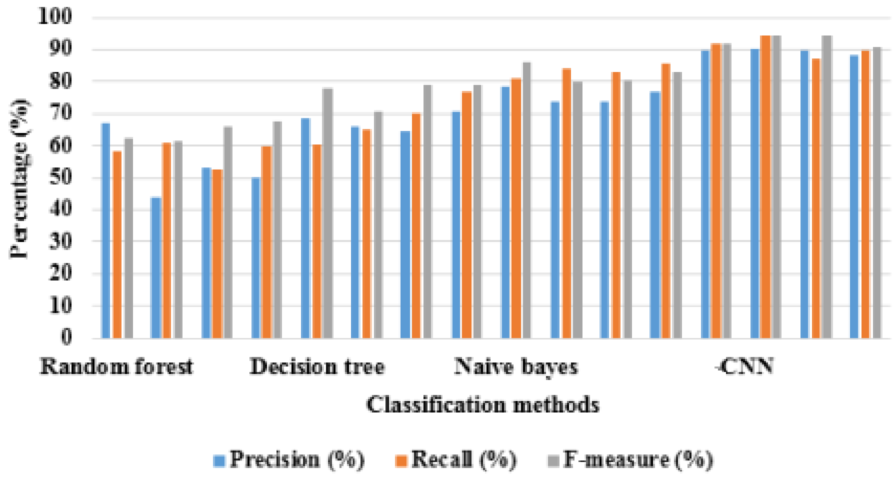
In addition, the comparative study of existing and proposed classification methods is carried-out for another four amazon product like a movie review, media, kindle, and camera. Here, the performance evaluation is validated with 80% training and 20% testing of data. Inspecting the tables 6 and 7, the proposed classification method (CNN) outperformed with the average classification accuracy of 92.8% as compared to the traditional classification methods; random forest, decision tree and naïve bayes. In addition, the existing classification methods achieved minimum precision, recall, f-measure, accuracy and AUC, compared to the proposed classification method (CNN). The graphical comparison of proposed and existing classification methods for the amazon products like a movie review, media, kindle, and a camera is represented in Fig. 6, and 7.

Table 6. Performance analysis of proposed and existing classifiers by means of precision, recall, and f-measure

Classifiers	Class	Precision (%)	Recall (%)	F-measure (%)
Random forest	Movie	67.11	58.34	62.43
	Media	44	60.78	61.34
	Kindle	53.29	52.80	66
	Camera	49.967	60	67.554
Decision tree	Movie	68.80	60.36	77.72
	Media	66.32	65.22	70.87
	Kindle	64.71	70	79
	Camera	70.94	77	79
Naive bayes	Movie	78.53	80.98	86.05
	Media	73.90	84.23	79.78
	Kindle	74	83	80.34
	Camera	77.01	85.54	83.32
CNN	Movie	89.77	92	91.84
	Media	90.41	94.27	94.50
	Kindle	90	87.23	94.19
	Camera	88	89.67	90.98

Retrieval Number: C5670098319/2019©BEIESP
DOI:10.35940/ijrte.C5670.098319
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication



Figures 6. Graphical comparison of proposed and existing classifiers by means of precision, recall, and f-measure

Table 7. Performance analysis of proposed and existing classifiers by means of AUC and accuracy

Classifiers	Class	AUC	Accuracy (%)
Random forest	Movie	54.90	58.12
	Media	46.89	60.67
	Kindle	51.67	62.98
	Camera	58.92	65.57
Decision tree	Movie	67.52	76.90
	Media	67	77
	Kindle	69.12	75.119
	Camera	73.57	78
Naive bayes	Movie	80.25	87.6
	Media	81.58	81.59
	Kindle	86.91	83.33
	Camera	84.47	89.907
CNN	Movie	89.38	92.68
	Media	88.30	91.50
	Kindle	92.20	93.55
	Camera	90.81	94

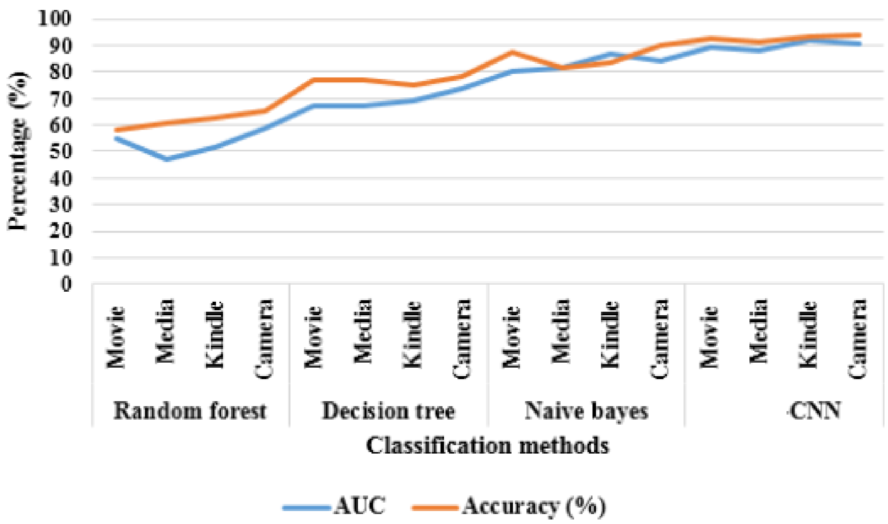


Figure 7. Graphical comparison of proposed and existing classifiers by means of AUC and accuracy

VI. CONCLUSION

In this research study, a new supervised system developed to classify the opinions of the customers for amazon selling products. The main motivation behind this experiment is to develop a proper keyword extraction method and classification approach for classifying the opinions of customers as neutral, negative, and positive forms using amazon customer review dataset. In this scenario, a key word extraction method (LDA) along with modified PFCM used for selecting the appropriate keywords. The obtained keywords classified using the classifier: CNN. The development of automated system for analysing the customer's opinion on amazon products has numerous advantages; able to handle multiple customers, effective in agent monitoring, track overall customer satisfaction, etc. Compared to the existing papers, the proposed system delivered an effective performance by means of quantitative analysis and comparative analysis. From the experimental analysis, the proposed system averagely achieved around 92.83% of classification accuracy, but the existing methodologies attained limited accuracy in amazon customer review dataset. In future work, an effective unsupervised system will be developed in order to further improve the classification accuracy of sentiment analysis.

REFERENCES

1. M.K. Hassan, S.P. Shakthi, and R. Sasikala, "Sentimental analysis of Amazon reviews using naïve bayes on laptop products with MongoDB and R", In *IOP Conference Series: Materials Science and Engineering, IOP Publishing*, vol.263, no.4, pp.042090, 2017.
2. X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach", In *Proceedings of the 28th international conference on machine learning(ICML-11)*, pp. 513-520, 2011.
3. Z.H. Deng, K.H. Luo, and H.L. Yu, "A study of supervised term weighting scheme for sentiment analysis", *Expert Systems with Applications*, vol.41, no.7, pp.3506-3513, 2014.
4. X. Fang, and J. Zhan, "Sentiment analysis using product review data", *Journal of Big Data*, vol.2, no.1, pp.5, 2015.
5. A.C. Pandey, D.S. Rajpoot, and M. Saraswat, "Twitter sentiment analysis using hybrid cuckoo search method", *Information Processing & Management*, vol.53, no.4, pp.764-779, 2017.
6. M. Ghiassi, and S. Lee, "A domain transferable lexicon set for Twitter sentiment analysis using a supervised machine learning approach", *Expert Systems with Applications*, vol.106, pp.197-216, 2018.
7. A.S.M. Alharbi, and E. de Doncker, "Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information", *Cognitive Systems Research*, vol.54, pp.50-61, 2019.
8. M. Daniel, R.F. Neves, and N. Horta, "Company event popularity for financial markets using Twitter and sentiment analysis", *Expert Systems with Applications*, vol.71, pp.111-124, 2017.
9. F. Abid, M. Alam, M. Yasir, and C. Li, "Sentiment analysis through recurrent variants latterly on convolutional neural network of Twitter", *Future Generation Computer Systems*, vol.95, pp.292-308, 2019.
10. N. Öztürk, and S. Ayvaz, "Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis", *Telematics and Informatics*, vol.35, no.1, pp.136-147, 2018.
11. T. Singh, and M. Kumari, "Role of text pre-processing in twitter sentiment analysis", *Procedia Computer Science*, vol.89, pp.549-554, 2016.
12. K. Philander, and Y. Zhong, "Twitter sentiment analysis: capturing sentiment from integrated resort tweets", *International Journal of Hospitality Management*, vol.55, pp.16-24, 2016.
13. R.P.Schumaker, A.T. Jarmoszko, and C.S. LabedzJr, "Predicting wins and spread in the Premier League using a sentiment analysis of twitter", *Decision Support Systems*, vol.88, pp.76-84, 2016.
14. N.F. Da Silva, E.R. Hruschka, and E.R. HruschkaJr, "Tweet sentiment analysis with classifier ensembles", *Decision Support Systems*, vol.66, pp.170-179, 2014.
15. Y. Wang, L. Sun, J. Wang, Y. Zheng, and H.Y. Youn, "A novel feature-based text classification improving the accuracy of twitter sentiment analysis", In *Advances in Computer Science and Ubiquitous Computing, Springer, Singapore*, pp. 440-445, 2017.
16. B. Le, and H. Nguyen, "Twitter sentiment analysis using machine learning techniques", In *Advanced Computational Methods for Knowledge Engineering, Springer, Cham*, pp.279-289, 2015.
17. G. Jalaja, and C. Kavitha, "Sentiment Analysis for Text Extracted from Twitter", In *Integrated Intelligent Computing, Communication and Security, Springer, Singapore*, pp. 693-700, 2019.
18. M. Yang, Q. Qu, X. Chen, C. Guo, Y. Shen, and K. Lei, "Feature-enhanced attention network for target-dependent sentiment classification", *Neurocomputing*, 2018.
19. O. Araque, I. Corcuera-Platas, J.F. Sanchez-Rada, and C.A. Iglesias, "Enhancing deep learning sentiment analysis with ensemble techniques in social applications", *Expert Systems with Applications*, vol.77, pp.236-246, 2017.
20. M. Giatsoglou, M.G. Vozalis, K. Diamantaras, A. Vakali, G. Sarigiannidis, and K.C. Chatzisavvas, "Sentiment analysis leveraging emotions and word embeddings", *Expert Systems with Applications*, vol.69, pp.214-224, 2017.
21. T. Alsinet, J. Argelich, R. Béjar, C. Fernández, C. Mateu, and J. Planes, "An argumentative approach for discovering relevant opinions in Twitter with probabilistic valued relationships", *Pattern Recognition Letters*, vol.105, pp.191-199, 2018.
22. A. Balahur, and J.M. Perea-Ortega, "Sentiment analysis system adaptation for multilingual processing: The case of tweets", *Information Processing & Management*, vol.51, no.4, pp.547-556, 2015.
23. M. Bouazizi, and T. Ohtsuki, "A pattern-based approach for multi-class sentiment analysis in twitter", *IEEE Access*, vol.5, pp.20617-20639, 2017.
24. D. Yu, D. Xu, D. Wang, and Z. Ni, "Hierarchical Topic Modeling of Twitter Data for Online Analytical Processing", *IEEE Access*, 2019.
25. S. Bharathi, A. Geetha, and R. Sathiynarayanan, "Sentiment Analysis of Twitter and RSS News Feeds and Its Impact on Stock Market Prediction", *International Journal of Intelligent Engineering and Systems*, vol.10, no.6, 2017.
26. H. Saif, Y. He, M. Fernandez, and H. Alani, "Contextual semantics for sentiment analysis of Twitter", *Information Processing & Management*, vol.52, no.1, pp.5-19, 2016.
27. Y. Ren, R. Wang, and D. Ji, "A topic-enhanced word embedding for Twitter sentiment classification", *Information Sciences*, vol.369, pp.188-198, 2016.
28. P.G. Preethi, and V. Uma, "Temporal sentiment analysis and causal rules extraction from tweets for event prediction", *Procedia computer science*, vol.48, pp.84-89, 2015.
29. K.A. Kumar, N. Rajasimha, M. Reddy, A. Rajanarayana, and K. Nadgir, "Analysis of users' Sentiments from Kannada Web Documents", *Procedia Computer Science*, vol.54, pp.247-256, 2015.
30. . Amolik, N. Jivane, M. Bhandari, and M. Venkatesan, "Twitter sentiment analysis of movie reviews using machine learning techniques", *International Journal of Engineering and Technology*, vol.7, no.6, pp.1-7, 2016.