Opinion Analysis of Implicit and Explicit Aspects of Product Reviews using Deep Neuro Fuzzy Network and Deep Maxout Network

T. Priyanka, **A. Mary Sowjanya**
Department of Computer Science and Systems Engineering,
College of **Engineering,** Andhra University**,** Visakhapatnam, AP, India

Article Info

Page Number: 10543 - 10556

Publication Issue:

Vol 71 No. 4 (2022)

Article History

Article Received: *15 September 2022*

Revised: *25 October 2022*

Accepted: *14 November 2022*

Publication: *21 December 2022*

Abstract

Online product shopping has dramatically become attributable to the e-commerce industry's tremendous boom. Because of the extensive product choices, effortless shopping, and attractive deals offered by all these platforms, customers or even producers have become familiar with them. By expressing their views and opinions about an item, service, theme, etc. online in the form of reviews, comments on blogging, discussion boards, social platforms, etc., users are making more user-generated content through all these e-commerce websites.Understanding implicit or explicit opinions conveyed in e-commerce posts is beneficial for many stake holders. The feature based sentiment analysis helps the customer to take an informed decision. In this paper, devised a new Hybrid Deep Learning Network to do the task effectively. The Aspect Term Extraction stage retrieves the BERT tokens and utilizes these to extract relevant aspects from the provided data. The aspect phrases are indeed taken into consideration and processed through the phase of feature extraction for predicting the sentiment rating. Senti-word Net, Word length, TF-IDF, Elongated Words, and BoW are all a handful of the aspects obtained during the process of feature extraction and then sent to the Hybrid Deep Learning network. The DNFN and DMN are utilized to develop the hybrid deep learning network, and the results from both models are integrated using weight correlation to provide the final sentiment score. The developed Hybrid DL network demonstrated impressive performance with the highest precision, recall, and F1-score compared to existing methods.

Keywords: Entity wise Sentiment prediction, Aspect Extraction, BERT, Deep Learning, Neuro Fuzzy Network, Maxout Network.

1.Introduction

Digitalization has researchers gifted with extraordinary prospects for reaping the potential of investigating the wealth of textual information for understanding the current trends. With the advancement in the field of Artificial Intelligence (AI), NLP has seen major advances [1]. Moreover, the development in speech recognition, sentiment analysis, pattern matching, and machine translation has transformed both business practices and led to progressive changes in the day-to-day lives of individuals [2]. Further, the expansion of social media and its usage over the internet across the globe has offered a cost-effective way of communication. Organizations started scanning the discussions, posts, tweets, reviews, blogs, etc. for determining the perception of individuals. These perceptions, opinions, feelings, views, and attitudes comprise a significant factor in assessing the individual's conduct and are referred to as sentiment. The process of evaluating the sentiments of individuals towards an article or an object is called sentimental analysis or opinion mining [3]. In the context of e-commerce, it can be defined as the process of obtaining the opinions of customers towards a service or a product. Sentiment reviews are automatically extracted and categorized using computational methods, text analysis and NLP in sentiment analysis. Evaluation of the opinions as well as sentiments is utilized in many areas, like business, marketing, and consumer information [4]. The task can make use of embeddings to generalize the task well even with less training data and helps to improve the performance in this NLP task by understanding the semantics easily. As they are trained on numerous datasets, pre-trained embeddings [5]are capable of conveying the syntactical and semantic sense of a word. In the context of NLP, Transfer Learning[6] manifests as word embeddings, which are extensively used to represent words of a document as multi-dimensional numeric vectors in place of traditional word representations like one hot vectors, TF-IDF vectors, etc., The examples are GloVe[7] and Word2Vec[8]. They capture the semantics of a word and place the words with similar meaning in close proximity. However, they are context independent and hence cannot handle polysemy. The later development in the form of Context Vectors dynamically learns contextualized word embeddings from a large collection of text using biLSTM[9] in support of machine translation. However, it represents the words based only on the final layer of the model, leaving out the information from the lower layers. The word embeddings provided by ELMo are considered as Universal Embeddings[10], since the generic Language Model once pre-trained, can be used for a wide variety of NLP tasks with appropriate fine-tuning.

Word embeddings[11] have limitations in natural language comprehension as it is unnatural to understand text, word-by-word rather than larger chunks of texts, such as, phrases, sentences and paragraphs. Sentence embeddings, if learned accurately, are more semantically robust than word embeddings as they capture a holistic essence of the sentence. However, representation of text using sentence embeddings is not as successful as that of word embeddings for the present. Recently, [12] explored Seven different architectures for directly extracting sentence embeddings from SNLI data set and declared that biLSTM with maxpooling network could achieve better word embeddings through supervised learning on SNLI dataset. The effectiveness or accuracy of an embedding can be established if it could be transferred to a downstream task of NLP [13] to surpass the performance levels that could be reached by traditional machine learning algorithms.

Sentiment analysis acts as a tool for converting the information available into "actionable knowledge" for making decisions for enhancing the performance of an organization. The data gathered from customer reviews can be effectively utilized for developing new strategies or taking suitable decisions regarding service/product requirements for commercial planning, strategic analysis as well as customer satisfaction [14]. The term "aspect-based sentiment analysis" (ABSA) was coined to capture aspect views stated toward particular review features because researchers observed that the generic opinions retrieved from the textual data were unsatisfactory. To improve the shopping experience for their customers, all e-commerce platforms allow individuals to just provide feedback for every product they order[15]. Although ratings are accessible, they only reflect an average score and do not offer information on specific product attributes.As a result, reading reviews posted online by other users to create an overall opinion or decide whether to buy an item, watch a film or not, schedule a trip, get an idea of resorts, tourist hotspots, etc., seems to be quite prevalent[16]. User opinions regarding a product, event, or topic can be found in reviews. Both commercial organizations and end customers can benefit from these reviews in terms of optimizing their products and making prudent judgements[17]. If a customer reads only a select few reviews or the most notable evaluations, they start experiencing a skewed opinion of the item. So the Aspect-wise Sentiment score helps the user to get an overall view of all opinions at a glance to take a decision regarding the product[18]. As sentiment analysis is performed with the aim of analyzing the opinion of people towards any entity systematically, it is necessary to determine whether the aspect, sentence, or opinion conveys neutral, positive, or negative sentiment [19]. Moreover, the extraction of sentiment comprises a large number of operations, which requires categorizing a text as negative or positive sentiment and can be thought of as a text classification issue. Further, the text available online comprises abbreviations, idiomatic expressions, or orthographic mistakes, and also the text is not often in a structured format as in newspapers and books [20].

2. Literature review

Kane, B., *et al*. [21] introduced a CLC model comprising of CNN, Bidirectional LSTM (BiLSTM) and Conditional Random Field (CRF) for performing Aspect-Based Sentiment Analysis (ABSA). Here, word-level depiction was obtained by encoding the word-level details with the help of CNN and from each word, the context details were modeled by the BiLSTM. Later, the sentiments as well as the aspect sequence were obtained using the sequential CRF. This scheme offered high adaptability; although it did not employ an attention mechanism to enhance the efficiency. An attention-based method was proposed in [22], where Trueman, T.E. and Cambria, E. developed a convolutional stacked BiLSTM for extracting sentiments as well as aspects. Here, the higher-level sentiment features as well as aspects were extracted by using a Convolutional layer and the information flow was controlled by the BiLSTM in the two directions. Moreover, the context vectors were found by using the multiplicative attention scheme. This approach produced high accuracy of prediction; but it did not utilize large-scale sentiments for improving the efficiency. Large scale classes were utilized in [23], where Song, W, *et al*. proposed a Semantics perception and refinement network (SPRN) for carrying out ABSA. The SPRN scheme was extremely effective in handling the noise associated with the input data; but no feature driven technique was considered to improve the performance. The drawback in [23] was overcome in [24], where Abdelgwad, M.M. *et al.* devised two models, namely interactive attention network based on bidirectional GRU (BGRU) (IAN-BGRU) for extracting the aspects and BGRU-CNN-CRF for determining the sentiment polarity.

A CLC model was devised in [14] for detecting sentiments, which simultaneously determined the sentiments as well as aspects. This method, however, did not consider elimination of the reliance on annotations by using an unsupervised model. Moreover, no attention schemes were considered to enhance the effectiveness of ABSA scheme. The drawback listed in [14] was overcome in [20], where the SPRN technique was developed for analyzing the sentiments. But this method suffered from a main issue, where it failed to utilize ensemble models for integrating the knowledge from meaningful patterns, common sense knowledge, cultural awareness, and social norms for improving the aspect representations.

To perform ABSA, [25] utilized the rationality database available and developed a graph model using the CNN called SenticGCN, helpful to grasp the reliance among the contextual and feature words along with the sentiment reflected words. Aspect extraction phase which retrieves all the crucial features from the dataset devised through Spider-Taylor ChOA for performing the fine-grained sentiment analysis [26] unable to retrieve the words considering the words case, having integer values etc., identified that can improve by making use of embeddings along with leveraging the CNN and RNN techniques.

[27] Proposed a method to find the polarity among the user-generated reviews aspect wise, including multiple aspects in a single sentence, C^3DA. The method helpful to generate samples of aspects using adversarial training, along with the pre-trained language model performed the fine-tuning step. After the generator network capable of finding the polarity of those aspects leveraging the benefits from data augmentation found lacking in applying to understanding tasks in a natural language. [28] proposed a masked attention model with varying methods for generating the mask. The largest of all weights is determined as the threshold and maintains only those having scores above this, then those top entities are selected assuming as less prominent and ignored others, but fails when the given sentence is large.

3. Method Developed

Hybrid Deep Learning for Sentiment Rating Prediction

A different technique was developed for sentiment rating Hybrid deep learning. Initially, the input review data is acquired from datasets[29] subjected to Tokenization using Bidirectional Encoder Representations from Transformers (BERT) [30], wherein the input sentence is divided into individual words or tokens. The tokens are then forwarded to the Aspect Term Extraction (ATE) [31] phase for extracting the specific aspects from the input data. For predicting the sentiment rating, the aspect terms alone are considered, and they go through the feature extraction stage. In the feature extraction phase, features like Senti-word Net, Word length, Term

Frequency -Inverse Document Frequency (TFIDF), Elongated Words, and Bag of Words are extracted, which are then provided to the Hybrid Deep learning network The hybrid deep learning network is developed by using the DNFN [32] and DMN [33], and the outputs obtained from both the networks are fused using weight correlation to yield the output sentiment rating value. Figure 1 depicts the schematic view of the developed Hybrid deep learning for sentiment rating prediction.

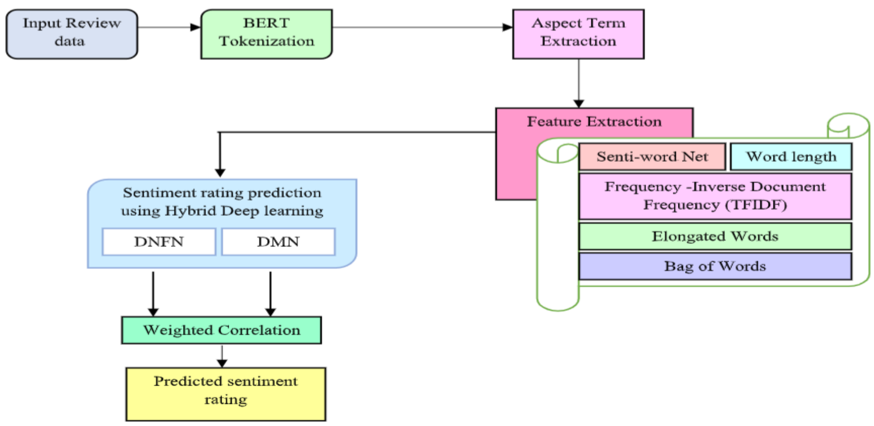

Figure 1. Schematic representation of developed Hybrid deep learning for sentiment rating prediction

3.1. The Input Text Document Acquisition

The process of abstractive summarization and sentiment rating prediction is carried out by considering the input review data accumulated from the database in such a way that the meaning of the text is not altered. Consider a database Z with a total of δ text documents, which is represented as,

$$Z = \{Z_1, Z_2, ..., Z_w, ..., Z_\delta\} \quad (1)$$

Here, δ represents the overall count of text documents and Z_a represents the a^{th} input text that is forwarded to the BERT tokenization process.

3.2. BERT Tokenization

The input text document Z_a is fed subjected to the BERT tokenization[30], where the input document is divided into tokens. A single or a pair of sentences can be combined together to form a sequence, which is fed as the input to the BERT. A total of 30,000 token vocabularies in the WordPiece embedding is utilized by the BERT. Here, the initial token in the sequence is considered to be the special classification token (CLS) and the classification process is represented by the last hidden state of the CLS. A single sequence is generated by combining the sentence pairs. The sentences can be differentiated using two approaches, wherein they are separated by a special token (SEP) in the first approach and in the latter type, every token is provided with a learned embedding to indicate the sentence to which it belongs. The position embeddings, segment, and the equivalent token are summed up to obtain the input depiction for an input token. The output of the BERT tokenization is represented by $Y_a$and is forwarded to the ATE process.

3.3. ATE

The tokenized data Y_a is fed to the ATE [31] phase, where the aspects are extracted automatically by learning the aspect features from the text, thereby reducing time and effort required. The aspects are extracted and clustered concurrently based on the seed words supplied by the users for different aspect classes, wherein each class contains the identical aspects. ATE can also be performed by exploring co-occurrence allocation of words by using word embedding and then applying an attention approach for abating the inappropriate words for improving the aspect consistency. ATE can be considered as the process of labeling the sequence and preparing the input with respect to the Inside, Outside, and Beginning (IOB) labels, which are designated as I_{as}, O, B_{as} corresponding to the aspect terms. The ATE task can be performed by considering the input, “The price is reasonable although the service is poor”, which is represented as $X = \{Y_1, Y_2, \ldots, Y_b\}$, where Y represents the tokenized text and the overall token count is given by $b = 10$. The above text can be represented as, $G = \{O, B_{as}, O, O, O, O, B_{as}, O, O\}$.

ATE is implemented using four steps like

i. *BERT-SPC*: Performs sentence pair categorization,
ii. E***stimating BERTg:*** The shared layer of the BERT fixed in the Global Context Feature Generator (GCFG) denoted byBERTg.
iii. ***Aspect polarity classifier:*** In this phase, the hidden states are extracted by considering the equivalent position of the primary token in the input sequence by using head-pooling. Further, the sentiment polarity is predicted by using the softmax function.
iv. ***Aspect term extractor:*** Token-level labeling is performed for every token in this step.

3.4 Feature Extraction

The aspect terms are utilized in the process of feature extraction, which is performed to extract the salient information in the text. Various significant features, like Senti-word Net, Word length, TFIDF, Elongated Words, and Bag of Words are extracted for preserving the most vital information and also to reduce the dimensionality of the data. The following subsections contain the short description of the features extracted.

f1: SentiWordNet [34] clusters terms sharing the similar meaning under a single type called synsets, wherein each term is related to two numerical values in the range [0,1]. The positive as well as negative bias of the synsets is indicated by the scores, which mirror the concurrence

between the classifier for producing a negative or a positive label for a word. By using SentiWordNet, positive and negative scores of every word can have a non-zero value.
f2: Word length can be computed by eliminating the sentences, like datelines or author names that are too short, as these sentences are not considered for producing summary.
f3: TF-IDF [35] is a numerical estimate which is used to assess the significance of a word inside a document and it acts as a weighting parameter for retrieving information.
f4: The elongated word indicates that a word is reiterated more than twice in a review. The values 0 and 1 correspond to the presence as well as absence of elongated words. Elongated word feature is computed with a dimension of [10000 * 1].
f5: Bag of words is used for collecting the words that are accessible in the text document, and the following expression gives the Bag of words with words located according to their occurrence in the document.
The features obtained above are concatenated to obtain a feature vector, which is represented as,

$$f = \{f_1, f_2, f_3, f_4, f_5\} \tag{2}$$

Wherein, f_1 represents the senti-word net, f_2 indicate the word length, f_3 signifies the TFIDF, f_4 refer to the elongated words, and f_5 denotes the bag of words features. The feature vector is then forwarded to the hybrid deep learning network for predicting sentiment rating.

3.5. Hybrid Deep Learning for Predicting Sentiment Rating

The hybrid deep learning network consists of the DNFN and the DMN to which the feature vector is applied. The output obtained from both the deep learning networks are combined using weight correlation to get the predicted rating.

3.5.1. DNFN

The DNFN [32] is a hybrid network which is created by combining the fuzzy logic with Deep Neural Network (DNN). Initially, the DNN is applied followed by the application of the fuzzy-logic for estimating the system objectives. The DNFN comprises several layers, such as input, output, as well as multiple hidden layers. The input layer uses the system's fuzzification values as well as multiple input parameters, whereas the defuzzification layer forms the output layer. On the other hand, defuzzification, normalization, and rule layers form the hidden layers. The premises and consequences of the system denote the major trivial system parameters, wherein the premises and consequences are related to the fuzzification as well as defuzzification tasks.This network utilizes a Fuzzy Inference System (FIS) for performing rule generation.

In the DNFN, every input as well as output parameters are matched to distinct nodes or entities in each layer. Further, assignment of degree of 0 or 1 is performed for every input depending on the criteria of the system. The entities in the initial layers are trained by an output value. Consider a consequent c, and two premises d_1 and d_2, which are expressed as,

$$R_{1j,} = \vartheta Q_j(d_1) \text{ or } R_{1,j} = \vartheta P_{j-2}(d_2), \quad \forall\, j = 1,2,3,4. \tag{3}$$

Here, $R_{1j,}$ represents the membership degree, ϑQ_j and ϑP_{j-2} signifies the antecedent membership functions and d_1 and d_2 refer to the inputs applied to every j_{th} node. The membership functions are indicated by using bell shaped function, which have a minimum as well as maximum values of 0 and 1 respectively, which is represented as,

$$\vartheta Q_j(d_1) = \frac{1}{1 + \left| \frac{d_1 - e_j}{g_j} \right| 2h_j} \tag{4}$$

Here, e_j, g_j and h_j represent the membership functions associated with the premise parameters, that have been optimized with respect to the training task.

The rule base layer forms the layer and is utilized for defining the set of rules. Each node in the rule base layer multiplies the linguistic variable, in order to fulfill the membership degree. The product of the membership variables reveals a rule's firing strength, which is given by the following expression.

$$R_{2,j} = \phi_j = \vartheta Q_j(d_1)\vartheta P_{j-2}(d_2)\ , \forall\ j = 1,2 \tag{5}$$

In layer 3, normalization is performed for computing the firing strength ratio of the j^{th} rule to the sum of the firing strength of all rules for every entity. Here, ϕ_j signifies the generic network parameter. A rule's firing strength normalizes the output of each rule and is represented by the following equation.

$$R_{3,j} = \bar{\phi}_j = \frac{\phi_j}{\phi_1 + \phi_2}\ , \forall j = 1,2 \tag{6}$$

The defuzzification layer forms the fourth layer, wherein the consequences of every rule are estimated for representing the total impact of the output, and is given by,

$$R_{4,j} = \bar{\phi}_j k_j = \bar{\phi}_j \left(p_j d_1 + q_j d_2 + r_j \right)\ ,\ \forall j = 1,2 \tag{7}$$

Here, p, q, and r refer to the consequent parameter set. The summation layer denotes the last layer, wherein all the previous outputs are added to attain the final result, which is represented as,

$$R_{5,j} = \sum_j \bar{\phi}_j k_j = \frac{\sum_j \phi_j k_j}{\sum_j \phi_j} \tag{8}$$

All the parameters are allocated an initial arbitrary value and they are utilized for computing the final output. Let us assume the output of the DNFN be denoted by M_1.

3.5.2. DMN

DMN is a multi-layer structure that consists of multiple maxout layers connected successively. The DMN offers the benefit of improved performance in a heavily resource-limited environment and can be effectively utilized in scenarios with limited vocabulary. The maxout layer in the DMN comprises maxout functions which are utilized in the generation of the hidden activations.

Further, the layers comprises of hidden units with non-overlapping groups. The activation functions generated by the DMN can be trained and the expressions given below represent the activation functions.

$$O_{t,u}^{1} = \max_{u\in[1,m_1]} f^{T} L_{..tu} + J_{tu} \tag{9}$$

$$\vdots$$

$$O_{t,u}^{n} = \max_{u\in[1,m_n]} O_{t,u}^{n-1^{T}} L_{..tu} + J_{tu} \tag{10}$$

$$\vdots$$

$$O_{t,u}^{l} = \max_{u\in[1,m_l]} O_{t,u}^{l-1^{T}} L_{..tu} + J_{tu} \tag{11}$$

$$O_{t} = \max_{u\in[1,m_l]} O_{t,u}^{l\ \ T} \tag{12}$$

Here, m_n indicates the hidden unit count present in the n^{th} layer having weight $L_{..tu}$ and bias J_{tu}, l refers to the overall layer count of the DMN. The DMN has the ability to estimate any arbitrary function by considering various values of m, and while $m > 2$ any non-linear activation function can be estimated. The output of the DMN O_t gives the value of the sentiment rating and is then forwarded to the weight correlation phase.

3.6. Fusion using weight correlation coefficient

The output of the DNFN M and DMN O_t are then subjected to the weight correlation phase, where both the outputs are fused for predicting the final sentiment rating. Here, the sentimental rating is performed by using the weight correlation coefficient. The final predicted sentiment rating is given by the following expression.

$$D = \frac{\sum \omega_i M O_t - \sum \omega_i M \sum \omega_i O_t}{\sqrt{\sum \omega_i M^2 - \left(\sum \omega_i M\right)^2} \sqrt{\sum \omega_i O_t^2 - \left(\sum \omega_i O_t\right)^2}} \tag{13}$$

Here, ω_i indicates the weighted attribute, M denotes the output of the DNFN and O_t signifies the output of the DMN. The final sentiment rating predicted is denoted by D.

4. Results and discussion

This section details the experimental results of the devised Hybrid deep learning for sentiment rating prediction. The devised schemes are examined for their effectiveness based on parameters, such as precision, F1-score, as well as recall.

ISSN: 2094-0343
2326-9865

The evaluation of the techniques is carried out with Amazon Reviews [29]. The dataset comprises reviews regarding fine foods available on Amazon. A total of 568,454 food reviews were collected till October 2012 over a span exceeding 10 years. The dataset contains ten attributes, such as Id, ProductId, UserId, ProfileName, Time, summary, text, score, Helpfulness Denomination and Helpfulness Numerator. Further, the reviews contain plain text review, ratings, user as well as product information, the developed Hybrid deep learning for sentiment rating prediction are implemented on a system with Windows 10 OS, Intel i3 core processor and 8GB RAM using Python.

4.1. Evaluation measures

The proposed Hybrid deep learning for sentiment rating prediction is evaluated for its effectiveness using parameters, like Precision, Recall, as well as, F1-score.

4.2. Performance Assessment

The developed Hybrid deep learning based sentiment rating prediction is evaluated for its effectiveness using metrics, like recall, precision, as well as, F-measure by varying the percentage of reviews.

Assessment of Hybrid deep learning based sentiment rating prediction

Figure 7 illustrates the evaluation of the devised Hybrid deep learning based sentiment rating prediction based on the percentage of reviews by considering various numbers of epochs.

The below assessment graph depicts the efficiency of Sentiment Rating in terms of Precision, Recall and F-measure vary as a function for number of reviews ranging from 60 – 90%. When compared with the existing methods like C3DA, Sentic GCN and Masked Attention, the proposed Hybtid Deep Learning Model has shown improvement in the performance corresponding to epochs varying from 10 to 50.

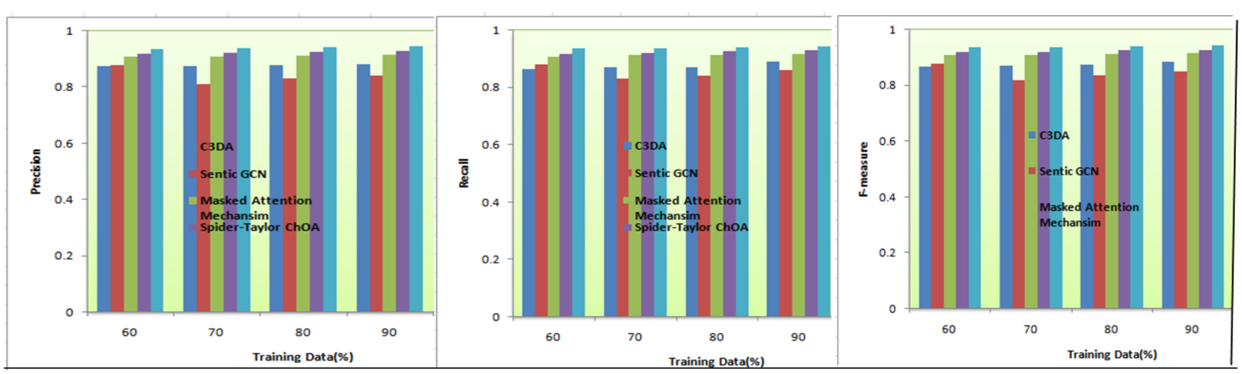

Figure 7. Assessment of Hybrid Deep Learning Network based sentiment rating prediction using a) Precision b) recall c) F-measure

5. Conclusion

The comments or user feedback in the form of reviews are growing exponentially as the usage of e-commerce has increased drastically, which leads the user to be vague and not confident in making a decision or choice.The entity based sentiment analysis useful for the user to guide his decision making step. This research proposed a new Hybrid Deep Learning method to predict the sentiment rating score for the user generated reviews in a fine-grained approach where we incorporated the weighted correlation also. Here, BERT tokenization is utilized for splitting the input review into tokens, from which aspect terms are extracted. The aspect terms are used by the hybrid deep learning network comprising DMN and DNFN for estimating sentiment rating. Moreover, neutral sentiments can be considered for enhancing the sentiment rating prediction accuracy.By leveraging the DNFN and DMN we are able to alleviate the problems like multi-aspects/entities that may present in review sentences. The proposed Hybrid Learning Network offers an enhanced performance with highest precision, recall and highest F-measure. The developed method is inadequate in identifying the polarity of review which includes sarcasm and emotions. In future, we can boost the performance by incorporating the transfer learning techniques into the proposed Hybrid Learning model. To assess the efficacy of the developed method, more datasets can also be adapted.

References

[1] Liu, X., Lee, D. and Srinivasan, K., “Large-scale cross-category analysis of consumer review content on sales conversion leveraging deep learning”, Journal of Marketing Research, vol.56, no.6, pp.918-943, 2019.

[2] Kang, Y., Cai, Z., Tan, C.W., Huang, Q. and Liu, H., “Natural language processing (NLP) in management research: A literature review”, Journal of Management Analytics, vol.7, no.2, pp.139-172, 2020.

[3] Yadav, A. and Vishwakarma, D.K., “Sentiment analysis using deep learning architectures: a review”, Artificial Intelligence Review, vol.53, no.6, pp.4335-4385, 2020.

[4] Hussein, D.M.E.D.M., “A survey on sentiment analysis challenges”, Journal of King Saud University-Engineering Sciences, vol.30, no.4, pp.330-338, 2018.

[5] Rezaeinia, S. M., Ghodsi, A., &Rahmani, R. (2017). Improving the accuracy of pre-trained word embeddings for sentiment analysis. arXiv preprint arXiv:1711.08609.

[6] Hosna A, Merry E, Gyalmo J, Alom Z, Aung Z, Azim MA. Transfer learning: a friendly introduction. Journal of Big Data. 2022 Dec;9(1):1-9.

[7] Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).

[8] Church, Kenneth Ward. "Word2Vec." Natural Language Engineering 23.1 (2017): 155-162.

[9] Wang Y, Huang M, Zhu X, Zhao L. Attention-based LSTM for aspect-level sentiment classification. InProceedings of the 2016 conference on empirical methods in natural language processing 2016 Nov (pp. 606-615).
[10] Matthew E. Peters , Mark Neumann , Mohit Iyyer , Matt Gardner et al., Deep contextualized word representations. In Proceedings of NAACL-HLT June 2018 Association for Computational Linguistics, pages 2227–2237.
[11] Tomas Mikolov, IlyaSutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In Conference on Advances in Neural Information Processing Systems.
[12] Conneau, A., Kiela, D., Schwenk, H., Barrault, L., &Bordes, A. (2017). Supervised Learning of Universal Sentence Representations fromNatural Language Inference Data.
[13] Ruder S, Peters ME, Swayamdipta S, Wolf T. Transfer learning in natural language processing. InProceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Tutorials 2019 Jun (pp. 15-18).
[14] Choi, Y. and Lee, H., “Data properties and the performance of sentiment classification for electronic commerce applications”, Information Systems Frontiers, vol.19, no.5, pp.993-1012, 2017.
[15] Tang D, Qin B, Liu T. Aspect level sentiment classification with deep memory network. arXiv preprint arXiv:1605.08900. 2016 May 28.
[16] Zhang W, Li X, Deng Y, Bing L, Lam W. A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges. arXiv preprint arXiv:2203.01054. 2022 Mar 2.
[17] Pandey, P., &Soni, N. Sentiment analysis on customer feedback data: Amazon product reviews. In 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (pp. 320-322).
[18] Eman Saeed Alamoudi& Norah Saleh Alghamdi (2021) Sentiment classification and aspect-based sentiment analysis on yelp reviews using deep learning and word embeddings, Journal of Decision Systems, 30:2-3, 259-281.
[19] Valdivia, A., Luzón, M.V., Cambria, E. and Herrera, F., “Consensus vote models for detecting and filtering neutrality in sentiment analysis”, Information Fusion, vol.44, pp.126-135, 2018.
[20] Birjali, M., Kasri, M. and Beni-Hssane, A., “A comprehensive survey on sentiment analysis: Approaches, challenges and trends”, Knowledge-Based Systems, vol.226, pp.107134, 2021.
[21] Kane, B., Jrad, A., Essebbar, A., Guinaudeau, O., Chiesa, V., Quénel, I. and Chau, S., “CNN-LSTM-CRF for Aspect-Based Sentiment Analysis: A Joint Method Applied to French Reviews,” In ICAART, vol.1, pp. 498-505, 2021.
[22] Trueman, T.E. and Cambria, E., “A convolutional stacked bidirectional lstm with a multiplicative attention mechanism for aspect category and sentiment detection”, Cognitive Computation, vol.13, no.6, pp.1423-1432, 2021.

[23] Song, W., Wen, Z., Xiao, Z. and Park, S.C., “Semantics perception and refinement network for aspect-based sentiment analysis”, Knowledge-Based Systems, vol.214, pp.106755, 2021.
[24] Abdelgwad, M.M., Soliman, T.H.A., Taloba, A.I. and Farghaly, M.F., “Arabic aspect based sentiment analysis using bidirectional GRU based models”, arXiv preprint arXiv:2101.10539, 2021.
[25] Liang, Bin, Hang Su, Lin Gui, Erik Cambria, and Ruifeng Xu. "Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks." Knowledge-Based Systems 235 (2022): 107643
[26] .Banbhrani, S.K.; Xu, B.; Lin, H.; Sajnani, D.K. Spider Taylor-ChOA: Optimized Deep Learning Based Sentiment Classification for Review Rating Prediction. Appl. Sci. 2022, 12, 3211.
[27] Wang B, Ding L, Zhong Q, Li X, Tao D. A contrastive cross-channel data augmentation framework for aspect-based sentiment analysis. arXiv preprint arXiv:2204.07832. 2022 Apr 16.
[28] Feng A, Zhang X, Song X. Unrestricted Attention May Not Be All You Need–Masked Attention Mechanism Focuses Better on Relevant Parts in Aspect-Based Sentiment Analysis. IEEE Access. 2022 Jan 11;10:8518-28.
[29] Amazon Reviews dataset available at, "https://www.kaggle.com/currie32/finding-sentiment-text-with-amazon-reviews/data", accessed on October 2022.
[30] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding”, arXiv preprint arXiv:1810.04805, 2018.
[31] HengYanga, BiqingZeng, JianHao Yang, Youwei Song, and RuyangXua., “A multi-task learning model for Chinese-oriented aspect polarity classification and aspect term extraction”, Neurocomputing, vol.419, pp.344-356, 2021.
[32] Javaid, S., Abdullah, M., Javaid, N., Sultana, T., Ahmed, J. and Sattar, N.A., “Towards Buildings Energy Management: Using Seasonal Schedules Under Time of Use Pricing Tariff via Deep Neuro-Fuzzy Optimizer”, In proceedings of 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC), pp. 1594-1599, 2019.
[33] Sun W, Su F, Wang L., “Improving deep neural networks with multi-layer max out networks and a novel initialization method”, Neuro computing, vol.278, pp.34-40, February 2018.
[34] Ohana, B. and Tierney, B., “Sentiment classification of reviews using SentiWordNet”, In proceedings of It& T conference, vol.13, pp.18-30, October 2009.
[35] Liu, Q., Wang, J., Zhang, D., Yang, Y. and Wang, N., “Text features extraction based on TF-IDF associating semantic”, In proceedings of IEEE 4th International Conference on Computer and Communications (ICCC), pp. 2338-2343, December 2018.

[36] Shi, Z., Hao, H., Zhao, M., Feng, Y., He, L., Wang, Y. and Suzuki, K., "A deep CNN based transfer learning method for false positive reduction", Multimedia Tools and Applications, vol.78, no.1, pp.1017-1033, 2019.
[37] Krizhevsky, A., Sutskever, I. and Hinton, G.E., "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems, vol.25, 2012.
[38] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556, 2014.
[39] Socher, R., Pennington, J., Huang, E.H., Ng, A.Y. and Manning, C.D., "Semi-supervised recursive autoencoders for predicting sentiment distributions", In Proceedings of the conference on empirical methods in natural language processing, pp. 151-161, 2011.