INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Cross Domain Hybrid Feature Fusion based Sarcastic Opinion Recognition Over E-Commerce Reviews Using Adversarial Transfer Learning

Parvati Kadli[1]* Vidyavathi B M[2]

[1]*Department of Computer Science and Engineering, Proudhadevaraya Institute of Technology, Hosapete, Karnataka, India*
[2]*Department of Artificial Engineering and Machine learning, Ballari Institute of Technology and Management, Ballari, Karnataka, India*
*Corresponding author's Email: kadli.parvati@gmail.com

Abstract: Sarcasm detection is a challenge in review based e-commerce product recommendation systems. Failure to detect sarcasm causes assignment of directionally opposite ratings to reviews and faults the recommendation systems. This work proposes a cross domain hybrid feature fusion based sarcastic opinion recognition system. Hybrid deep feature integrating content and context is formed and feature vector volume is expanded using cross domain adversarial learning. The hybrid features are used for sarcasm detection in two combinations. First combination is hybrid features with bi-directional long short term memory (LSTM)and second is hybrid features with traditional machine learning classifiers. Hybrid feature with bi-directional LSTM is able to capture temporal relation between sentences of the review for better identification of sarcasm. Traditional machine learning classifiers trained with these hybrid features are also able to provide higher accuracy of sarcastic detection in reviews. The effectiveness of the proposed cross domain hybrid feature fusion features is tested against various experimental setups with Bi-directional LSTM and traditional machine learning classifiers. The proposed solution is able to achieve peak accuracy of 95% in classifying e-commerce reviews.

Keywords: Sarcasm, Hybrid features, Cross domain adversarial learning, Temporal correlation.

1. Introduction

E-commerce portals have become the most preferred shopping destinations. Recent surveys on E-commerce acceptance in United States [1] found that E-commerce accounts for 20% of total retail sales and every year it increases at rate of 14.2%. Reviews play an important role in influencing the product purchase decisions of consumers in E-commerce systems. Almost 97.7 % consumers refer to online reviews before making an online purchase [2]. Product recommendation systems recommend products based on translating reviews to rating and collaborative filtering on those ratings [3]. Most of collaborative recommendation system obtains ratings from reviews through quantization of sentiments expressed in reviews through natural language processing. Presence of sarcasm can make this process of sentiment quantization erroneous and move the ratings to directionally opposite end. This makes the product recommendation faulty and presents irrelevant or uninteresting products to consumers. To prevent these problems and learn the sentiment accurately, sarcastic opinions must be classified accurately. Though many works have been proposed for Sarcasm detection, they lack accurate detection due to two important problems: (i) lack of context and domain cues in classification models, (ii) lack of labelled sarcastic dataset for e-commerce domain. Insufficient context knowledge and specific topic information makes it difficult to detect sarcastic utterances accurately [4]. The context features must be augmented with context and domain cues to learn more insight necessary for accurate detection of content polarity [5]. Lack of sufficient labelled sarcastic data for e-commerce reviews is also a

Received: November 7, 2022. Revised: January 3, 2023.

problem in building machine learning based sarcasm recognition classifiers. Addressing these two problems, this work proposes a hybrid deep feature integrating content and context cues. To compensate for lack of sufficient labelled sarcastic e-commerce dataset, cross domain learning from different domains is facilitated using adversarial learning. The hybrid features are used in combination with Bi-directional LSTM and traditional machine learning classifiers for sarcasm detection. Following are the contributions of this work.

(i) A novel hybrid deep learning feature combining content and context cues for sarcasm recognition
(ii) Hybrid feature augmentation with cross domain learning to compensate for lack of sufficient labelled sarcastic data in e-commerce domain.
(iii)Exploit the temporal correlation between reviews of sentences using a Bi-directional LSTM with hybrid features for sarcasm detection
(iv)Combining hybrid feature with traditional machine learning models for sarcasm detection

In the rest of paper, section II presents the survey of existing works on sarcasm detection. Section III presents the proposed hybrid features for sarcasm detection. Section IV presents the results of hybrid features in combination with Bi-directional LSTM and traditional machine learning algorithms. Section V presents the conclusion and the scope of future work.

2. Related work

Govindan et al [6] used hyperbole features for sarcasm detection. Tweets were processed to extract features like intersection, intensive etc. These features are classified to sarcastic class using traditional machine classifiers like random forest, support vector machine etc. The classifiers were able to achieve maximum sarcasm classification accuracy of 78.74%. The classifier was trained with only two labels of sarcasm and racism, but in realistic environment classification need to done for multiple class and in this case the accuracy will be even low. Guo et al]7] exploited commonality across multiple sarcasm datasets using transfer learning. Adversarial neural transfer is used to make the shared feature distributions of the source domain and the target domain as similar as possible, while simultaneously optimizing for domain-specific performance. But there is a problem of domain-specific classifiers ignores the shared features and no meaningful transfer occurs between domains. To solve this problem, authors proposed a Latent optimized Adversarial neural transfer. The classification accuracy is increased by joint use of multiple datasets. Joint utilization increased the semantic understanding necessary for sarcasm detection. The performance gain can be still improved by removing the datasets contributing to domain losses. Bedi et al [8] proposed an attention based multi modality classification model for sarcasm classification in Hindi-English code mixed conversational dialog. Textual and acoustic features are extracted from the conversations. Word vector feature is the textural feature. MFCCs (Mel-frequency cepstral coefficients) are the acoustic features. For classification of sarcasm and humor, authors developed a multi modal dataset for Hindi-English code mixed conversations. Utterance level hierarchical model was used to learn enriched textual representations. To learn the context of the dialog, two LSTM was used – one for textual and another for acoustic features. The solution was able to achieve 80 % accuracy in sarcasm classification. The dataset is limited considering the text and acoustic vocabulary of English and Hindi language combined. Onan et al [9] proposed a sarcasm detection framework combining neutral language models and deep neural networks for classifying sarcastic text documents. A novel weighted word embedding model is proposed which provides more importance to word ordering information. The features were classified to sarcastic or not sarcastic classes using a three layer stacked bi-directional LSTM. The model was able to provide 95.30 % accuracy. The performance of the model can be improved using linguistic feature sets like conventional lexical, pragmatic, implicit incongruity and explicit incongruity based features. Zhang et al [10] integrated quantum theory and fuzzy logic to the problem of sarcasm detection in texts. Integration of fuzzy logic is done to address the uncertainty and vagueness in sarcasm detection. The text is converted to complex valued vector using the quantum theory integrated fuzzy system. The proposed solution is able to achieve 68% accuracy. Eke et al [11] proposed a sarcasm classification model addressing the problem of failing to capture the contextual information in the sarcastic expression and ignoring sentiment polarity of the words. Authors proposed a context based feature technique with three classification models of deep learning, BERT and conventional machine learning. In deep learning model, Global Vector representation (GloVe) was used for word embeeding and features were classified with bidirectional LSTM. Bidirectional encoder representation and transformer (BERT) is the second model. Hybrid feature fusion of BERT, sentiment, syntactic and GloVe with traditional machine learning classifier is the third model. All the three models were able to achieve more than 95% accuracy against datasets. But their performance against real data is not explored. Jamil et al [12] proposed a hybrid approach for sarcasm detection for multi domain datasets. Convolutional neural network (CNN) is used for feature extraction. LSTM is used for sarcasm classification. From text document, word embedding feature vector is created using CNN. LSTM is trained with this feature vector to classify sarcasm. Author experimented with three different word embedding algorithms - TF-IDF, Bag of words and Global vector for representation. The method was able to achieve 92.2% accuracy. Sonawane et al [13] proposed term co-occurrence feature for sarcasm detection. The correlation between features and the sarcasm label (positive and negative) is found using ANOVA standard t test. Through ANOVA test, the words with higher correlation to Sarcasm are detected and feature vector is formed with these words. Machine learning classifiers are trained to classify sarcasm using these features. The method was able to achieve 94% accuracy. But this method highly dependent on the dataset and in case of new arrivals with significant deviation from training patterns, the accuracy will be very low in this approach. Eke et al [14] proposed a hybrid feature fusion technique for sarcasm detection. The solution has two stages. In the first stage, lexical features are extracted using bag of words technique. In the second stage, the features are classified using traditional machine learning classifiers like decision tree, support vector machine, k-nearest neighbour and random forest etc. The results of each of the classifier are fused with average ensembling to provide final prediction. The context features are lexical, length of microblog, hashtag, discourse markers, emoticons, syntactic, pragmatic, semantic (GloVe embedding), and sentiment related features. Traditional machine learning features were used to classify the fused features. The solution achieved 93% accuracy. The solution works only if contradicting sentiment is expressed in same sentence. It cannot detect sarcasm spread across multiple sentences. Pramanick et al [15] proposed a multi modal approach called Multimodal learning using Optical transport for sarcasm detection from video and image text pairs. The proposed solution utilizes self-attention to exploit intra-modal correspondence and optimal transport for cross-modal correspondence. Finally, the modalities are combined with multimodal attention fusion to capture the inter-dependencies across modalities. The method was able to achieve 71% accuracy. Aliwy et al [27] used following features of emojii, emoticon, text conflict and hashtag to detect sarcasm. Conflicts in text were identified by checking for presence of antonyms. But this work does not consider sentence contradiction and sentiment contradiction spread across multiple sentences in same review. In addition, the work does not consider cross reference learning from multiple datasets to enhance the features. Wen et al [28] proposed a enhanced attention neural network model for sarcasm detection. Fine grained portrayal of word and context information such as title are used as feature. But sentiment contradiction across sentences in texts were not considered. Alhariri et al [29] extracted lexical features from texts and classified it using traditional machine learning classifiers. Authors did not consider sentiment context and contradiction for detection of sarcasm. Govindan et al [30] extracted five hyperbole features of interjection, intensifier, capital letter, punctuation mark and elongated word and classified it using traditional machine learning classifiers to detect sarcasm. Sentiment and sentence contradiction were not considered for sarcasm detection.

The summary of the survey is presented below

Author	**Method**	**Problem**
Govindan et al [6]	Content feature(hyperbole) with conventional machine learning classifiers	Accuracy is less than 79%
Guo et al]7]	Transfer learning from multiple datasets	Learning ignored domain specific shared features, so cannot be used for cross domain datasets.
Bedi et al [8]	Multi modal features – text + audio for sarcasm detection.	Tested with limited utterances and most reviews are not multimodal data.
Onan et al [9]	weighted word embedding model + LSTM	Considered content feature alone, cannot detect sarcasm spread across multiple

		sentences
Zhang et al [10]	Content features based on quantum theory along with fuzz logic reasoning	Considered content feature alone , cannot detect sarcasm spread across multiple sentences
Eke et al [11]	Hybrid feature fusion of BERT, sentiment, syntactic and GloVe	Overall sentiment across whole texts were not considered and sentiment contradiction was not explored
Jamil et al [12]	Word embedding content feature along with CNN and LSTM for classification	Considered content feature alone , cannot detect sarcasm spread across multiple sentences
Sonawane et al [13]	Term co-occurrence feature with traditional machine learning classifiers	Dataset dependent and does not consider sentiment contradiction
Eke et al [14]	Content+ context feature along with traditional machine learning classifiers	Cannot detect sarcasm spread across multiple sentences
Pramanick et al [15]	Multi modal approach combining video, text and audio	Most of the review data are texts and the approach is not suitable for large application context
Aliwy et al [27]	emojii, emoticon, text conflict and hashtag features to detect sarcasm.	Does not consider sentence contradiction and sentiment contradiction spread across multiple
		sentences in same review
Wen et al [28]	Word representation and context information as feature	Sentiment contradiction across the sentences are not considered.
Alhariri et al [29]	Lexical features with traditional machine learning classifiers	Sentiment spread across sentences were not considered for sarcasm detection
Govindan et al [30]	five hyperbole features with traditional machine learning classifiers	Sentiment and sentence contradiction were not considered for sarcasm detection

From the survey, most existing works for sarcasm detection from texts used only content features like Glove embedding model, TF-IDF. Very few works considered limited context information mostly sentiment. But they could not identify the semantic sentiment contradictions across the sentences in the review text. Also cross reference learning from the knowledge spread across multiple datasets was not considered in existing works. This limited the classification accuracy. This work attempts to solve this problem by considering more context features and employing cross reference learning to improve the classification accuracy.

3. Hybrid feature fusion

The hybrid feature proposed in this work fuses content based and context related features. The fused features are then used for sarcasm detection. The overall architecture of the proposed solution is given in Fig. 1. As seen from Fig. 1, pre-processing, feature extraction and training/classification are the three stages of the proposed solution. In the pre-processing stage, the reviews are pre-processed through multiple process involving filtering, tokenization, stop word removal and stemming. The pre-processed reviews are passed to the next stage of feature extraction. Three features of content and context are extracted from the pre-processed reviews. The fused features

Received: November 7, 2022. Revised: January 3, 2023.

Table 1. Sentiment emoticon mapping

<table>
<tr><th>Sentiment</th><th>Emoticon patterns</th></tr>
<tr><td>Positive</td><td>:-(, :(, :-|, ;-(,
;-<,|- {</td></tr>
<tr><td>Negative</td><td>-), :), :o, :-}, , ;-}, :->, ;-)</td></tr>
<tr><td>Sarcastic</td><td>(, [:,
;], -?[), p, P]</td></tr>
</table>

are used to train Bi-directional LSTM and traditional machine learning classifier. The performances of the trained classifiers are measured in terms of various metrics accuracy, precision, recall and F1-measure. The different stages of the proposed solution are detailed in below sub sections.

3.1 Preprocessing

The raw review texts are pre-processed before passing to feature extraction stage. The review comments with very few words are filtered out. The remaining reviews are tokenized to words and space characters are removed. From the tokenized words, the most common stop words in English are removed. The remaining words are then stemmed to standard form using Porter stemmer algorithm.

3.2 Feature extraction

Feature are extracted in two categories: content and context space.

Content features

Emoticon features and word embedding features are the content features used in this work. Emoticons visually represent the facial expression. Emoticons have high correlation to uttered sarcastic statements [16]. A smiley emoticon with negative situation word is an example of sarcastic utterances. The emoticons patterns considered in this work are given Table 1.

The reviews are processed to look up for the emoticon patterns and the frequency of the emotion patterns. The feature vector is constructed for emoticon patterns in form of

$$E = < f_p, f_n, f_s > \qquad (1)$$

Where $\mathrm{f_p}$ is the frequency of positive emoticons, $\mathrm{f_p}$ is the frequency of negative emoticons and $\mathrm{f_s}$ is the frequency of sarcastic emoticons.

Glove word embedding features [17] are extracted from the pre-processed reviews after removing all the emoticons. Glove is a powerful word embedding algorithm. It learns the words vector representation by doing dimensionality reduction on the co-occurrence matrix. Glove is an unsupervised technique to obtain meaningful vector for each individual word in the corpus. The vector for the word is constructed in such a way that similar words cluster together and different words repel against each other. Glove embedding is selected in this work due to its ability to capture both local statistics and global statistics in forming word vectors compared to other word embedding models like word2vec.

Amazon product review [23] is the most important e-commerce dataset. Though the dataset has about 142.8 million reviews, the dataset does not have any explicit label to classify sarcastic reviews. Transfer learning is adopted to compensate for lack of large labeled e-commerce datasets for sarcastic detection. This work compensates for lack of sarcastic label in Amazon product review with Reddit sarcastic dataset [24] and Twitter sarcastic dataset [25]. The architecture of the transfer learning is given in Figure 2. The architecture learns shared feature space across two sources Redditand Twitter and target Amazon and a domain-private feature space for Reddit and Twitter using convolutional neural network (CNN). The transfer learning used in this work bridge cross domain discrepancy of all domains to predict the sarcastic labels (sarcastic/normal) for the Amazon reviews. Both private and shared feature spaces are exploited by transfer learning. Private feature space is specific to Reddit and Twitter data. Shared feature space is across all the three datasets. The feature space learning is facilitated by the adversarial loss function to split private and shared representations. Two private encoders are trained with Reddit and Twitter data with Glove embedding of Reddit and Twitter labelled dataset. The shared feature space stores the common knowledge learned from three datasets of Reddit, Twitter and Amazon. This improves the sarcastic classification performance with Amazon dataset. The transfer learning must be done in such a way that shared feature space has more sharable features so that prediction on sharable features does not discriminate the three datasets. To achieve the above goal, adversarial loss function is used. Let $\mathrm{F(f_c; \theta_d)}$ be the domain classifier parameterized by $\mathrm{\theta_d}$ mapping the shared feature space $\mathrm{f_c}$ to the output label (sarcastic or not). The adversarial loss function (L) optimizes $\mathrm{\theta_d}$ to make correct classification. It is defined as

$$\mathrm{L = min_{\theta_c}(max_{\theta_d}\left(\sum_{j=1}^{J}\sum_{i=1}^{N_j} d_i^j \log d_i^j + \left(1 - d_i^j\right)\log\left(1 - d_i^j\right)\right))} \qquad (2)$$

Where $\mathrm{d_i^j}$ is the ground truth label (sarcastic or not) for instance i. The adversarial loss function,

Received: November 7, 2022. Revised: January 3, 2023. 157

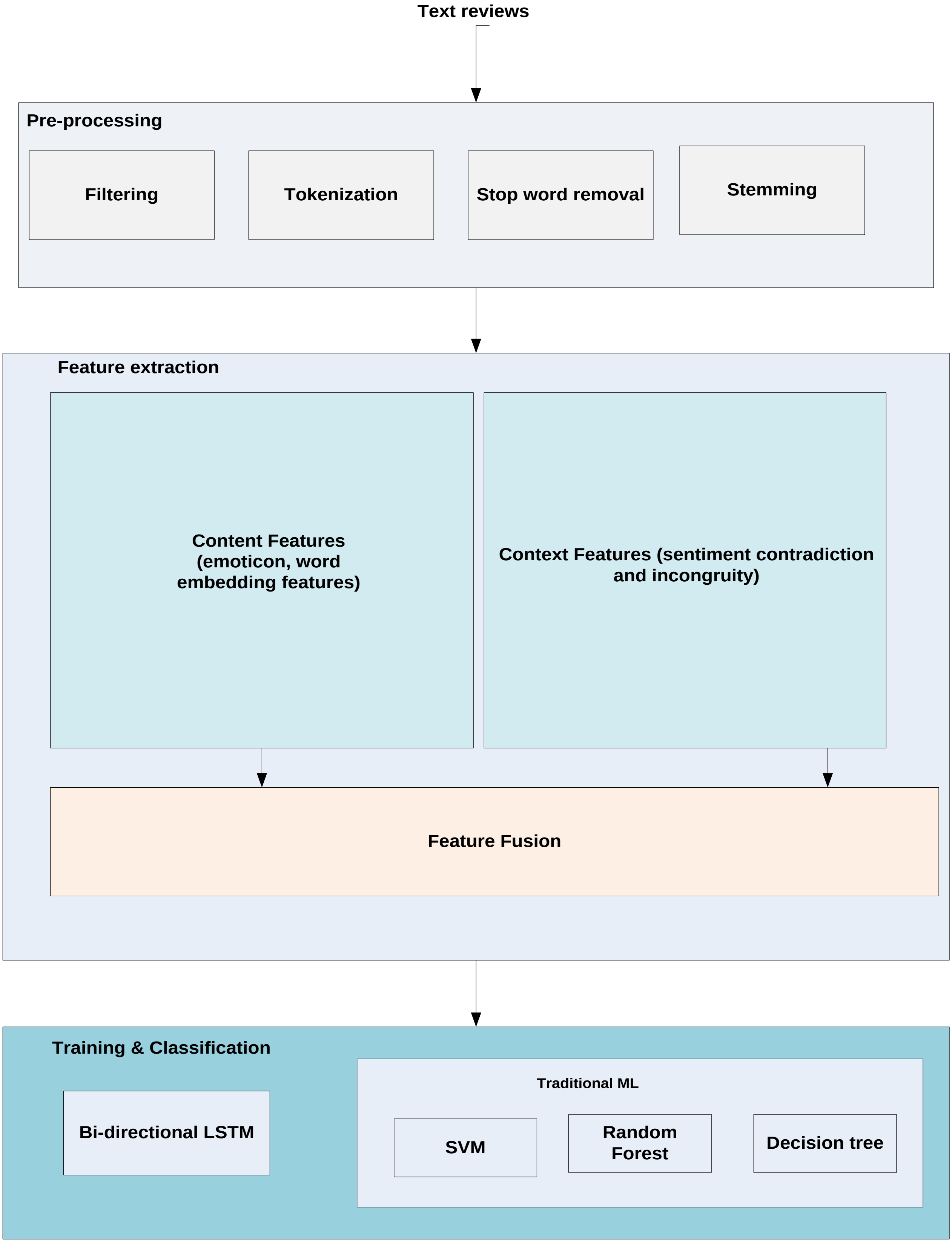

Figure. 1 Hybrid feature fusion architecture

maximize the cross entropy for domain classification with respect to θ_d and minimizing with respect to θ_c

CNN model is used for transfer learning in this work. The architecture is given in Fig. 4. Review sentences are mapped to Glove feature vectors of dimension d=100 and passed to convolution layer.

Received: November 7, 2022. Revised: January 3, 2023.

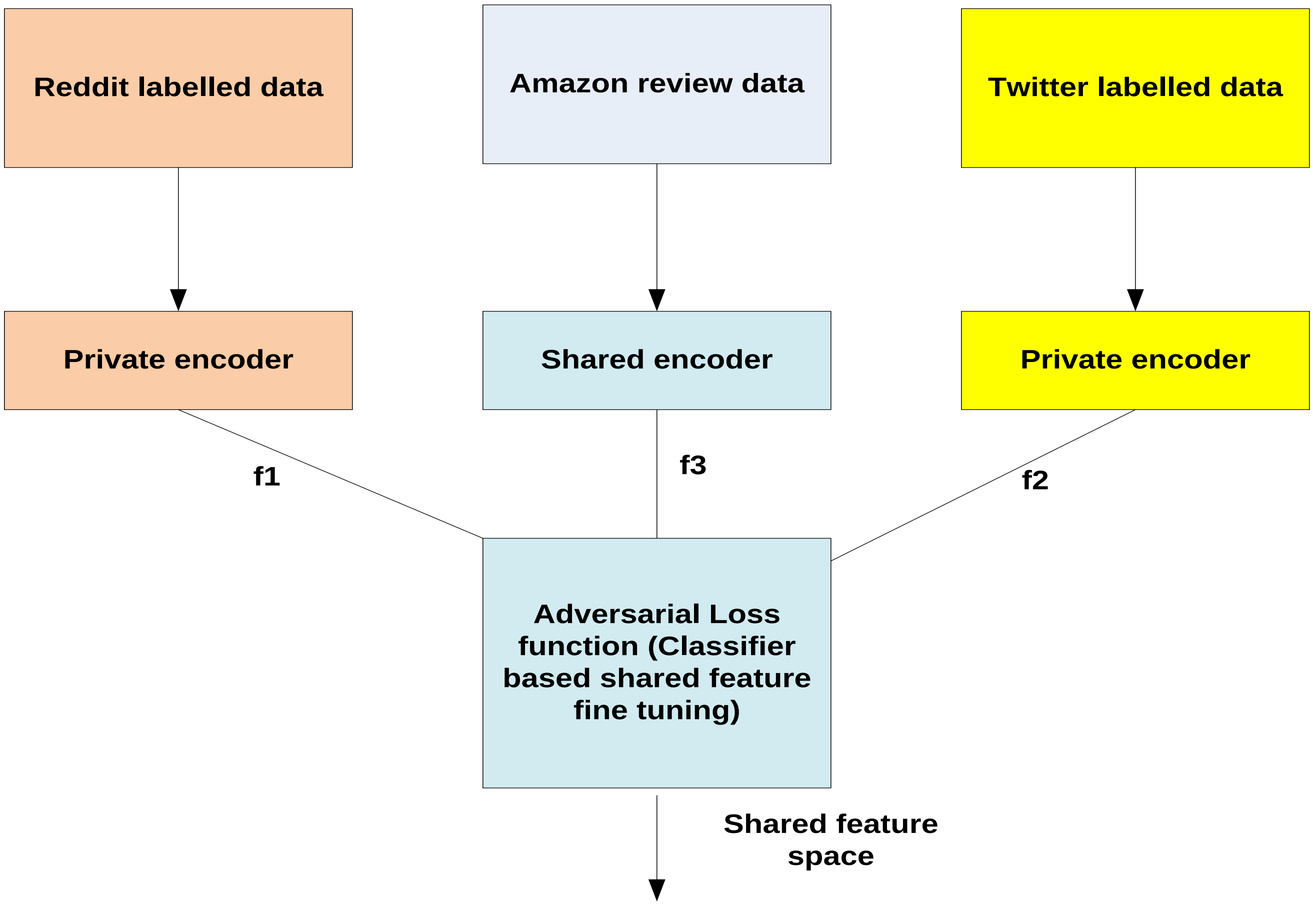

Figure. 2 Adversarial training

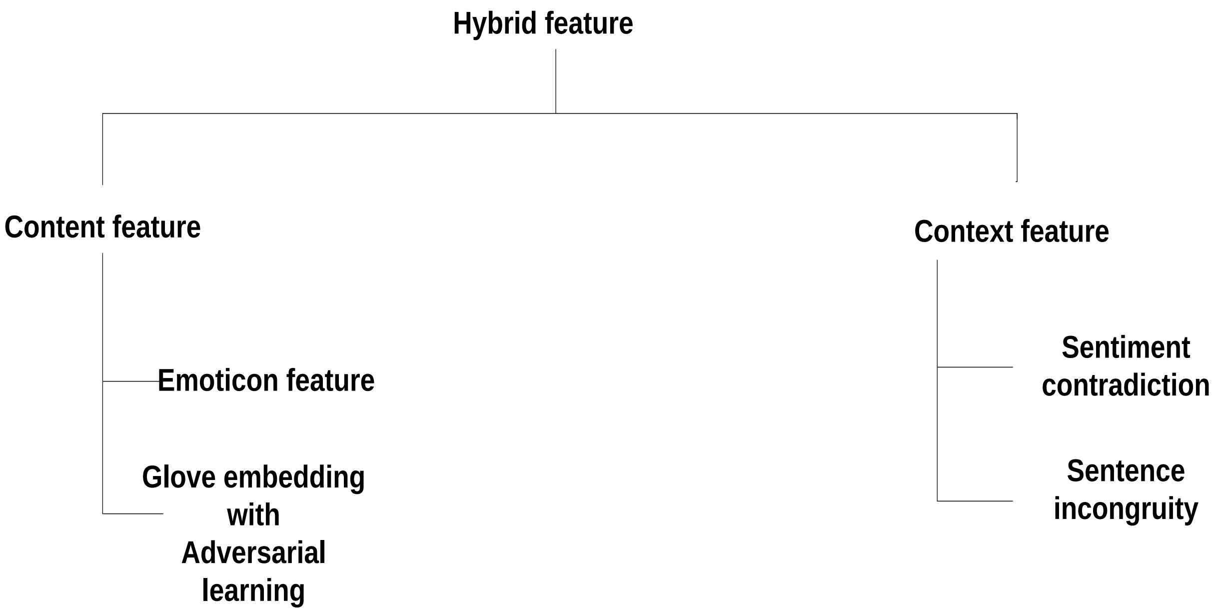

Figure. 3 Feature taxonomy

Three filters (3,4,5) are applied on feature maps of dimension 100×1 with ReLU activation function. The result is passed to Pooling layer where features are flattened to a dimension (300×1).

Context features

Sentiment contradiction and sentence incongruity are the context features extracted from the reviews. Sentiment contradiction refers to the conflicting

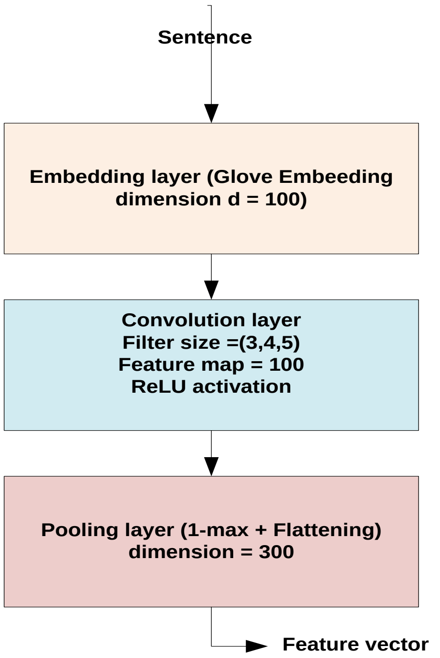

Figure. 4 CNN Model for learning features

sentiments expressed in the same review. Conflicting sentiments is an indicator of sarcasm.

Conflicting sentiments must be identified only for sentences with coherence (sentences referring to same concept). The sentences in the reviews must be grouped based on coherence and conflict must be identified in each group. The strength of sentiment contradiction is calculated for each group of sentence and provided as a feature.

Coherence sentences have similarity in terms of subjects or objects of the sentences. Say for two sentences St_1, St_2 the coherence between two sentences is defined in terms of relationship between the noun/noun variants in two sentences. POS tagging of the sentences is performed to extract tags of nouns, pronouns, noun phrases. The sentence are represented in terms of their extracted tags are

$$St_1 = \sum_{i=1}^{p} w_{1i} \quad (3)$$

$$St_2 = \sum_{i=1}^{q} w_{2i} \quad (4)$$

The conditions for existence of coherence are stated below

(i) There exists atleast any w_{1i} which is almost equal to any of w_{2i}
(ii) There exits atleast one pronoun w_{1i} which is almost identical to any of w_{2i}
(iii) There existing a name entity w_{1i} which is matching to any of w_{2i}

Once two sentences St_1 and St_2 are found as coherent, the sentiment contradiction is measured between them. The polarity of each of the words in the sentences is calculated and if the polarities are opposite between the sentences, contradiction is detected. The number of contractions divided by the total number of sentences is calculated and this score is referred as sentiment contradiction strength (SCS).

The polarity of each word in sentence is found using two lexicons SentiStrength(ss) [18] and Senticnet(sn) [19]. The sentiment score of each word (w) is calculated as below

$$p(w) = \begin{cases} ps(w), \forall\, w \in ss \text{ or } sn \\ (ps_{ss}(w) + ps_{sn}(w))/2 \\ \dfrac{1}{|C|} \sum_{c \in C} ps(c), \text{otherwise} \end{cases}, w \in ss \,\&\, sn \quad (5)$$

Where ps is the polarity score found from ss or sn. The polarity of word (p(w)) is the polarity score if the word is present in one of ss or sn. It is average of polarity score when the word is present in both ss and sn. In case the word is not present in both ss and sn, the top K concepts which the word is associated is found using Conceptnet[20] lexicon and the polarity score for those concepts is found and added up.

The polarity score for sentence is found as average of polarity of all words in the sentence

$$p(s) = \frac{\sum_{w \in s} p(w)}{|w|} \quad (6)$$

Contradiction is decided as 1, when polarity of two sentence is just opposite to each other. The sentiment contradiction in whole review is found as sum of contradictions divided by total number of sentences.

$$sc(r) = \frac{\text{total number of contradictions}}{|s|} \quad (7)$$

Sentence incongruity is the concept of polarity contrast between the positive candidate term and negative phrase or negative candidate term with positive phrase. The order of occurrence is not important. Camp [21] detailed the incongruity patterns in English language sentences and the summary is presented in Table 2.

The sentences are POS tagged and count of number of patterns as defined in Table 2 are found and given as sentence incongruity (si) feature. A hybrid feature combining emoticon patterns (E),

Table 2. Sentence incongruity rules

Candidate term(Verb positive/negative)	Positive/Negative patterns
Verb	Verb followed by Verb
Verb present participle	Verb followed by Adverb
Verb Gerund	Adverb followed by Verb
Verb past participle	Verb followed by proposition
Verb past form	Verb followed by adjective
Verb present participle third person singular	Verb followed by noun

glove word embedding improved with adversarial learning (Glal), sentiment contradiction feature(sc) and sentence incongruity feature (si) is formed. This hybrid feature vector is used as input for the next stage of sarcastic review classification.

3.3 Classification

Two modes of classification: deep learning and traditional machine learning are explored in this work. In deep learning mode, Bi-directional LSTM classifier is trained with hybrid feature vector to detect sarcastic reviews. In traditional machine learning mode, support vector machine (SVM), decision tree and random forest classifiers are trained to detect sarcastic reviews.

LSTM classifier is an improvement over standard recurrent neural network to solve the vanishing gradient problem [22]. LSTM is designed to learn long distance dependencies within the sequential data. LSTM has ability to forget long term information and upgrade to new information with gating mechanism and cell activation state. Cell activation output can also be controlled by separating the hidden state with cell activation state. Each LSTM node takes the input vector x and the previous hidden state. The LSTM node calculates the weighted sum of input and bias b and passes to the hyperbolic tangent activation function. The resulting output is given as

$$c_t = \emptyset_t(W_c x_t + U_c h_{t-1} + b_c) \tag{8}$$

c_t is the candidate cell activation. x_t is the input vector. W and U are the weight matrices.h_{t-1} is the hidden state vector at the previous time step and b_c is the bias.

The gates control how much of activation must be retained and how much to forgot. Input gate control decides how must activation to retain and forget gate decides how much cell activation to forget. The final gate is incorporated to calculate the hidden state in terms of forgot vector (f_t), input gate vector(i_t) and output gate vector(o_t) as

$$f_t = \emptyset_s(W_f x_t + U_f h_{t-1} + b_f) \tag{9}$$

$$i_t = \emptyset_s(W_i x_t + U_i h_{t-1} + b_i) \tag{10}$$

$$o_t = \emptyset_s(W_o x_t + U_o h_{t-1} + b_o) \tag{11}$$

Bi-LSTM is adaptation of LSTM that can learn compositional information in a sentence in an effective manner. It has forward and backward operation network to learn clause information in both directions. The hidden states learnt in both directions are combined to result in hidden state of Bi-LSTM. The output is computed in Bi-LSTM as

$$y_t = \sigma(\vec{h}, h^{\leftrightarrow} t) \tag{12}$$

In the above equation, σ can any function (summation, multiplication, concatenation or average) which takes two sequences and provides the output in form of vector. LSTM results are concatenated to provide the final output results. The labelled Reddit and Twitter datasets are converted to hybrid features and Bidirectional LSTM is trained with the hybrid features as input and label (sarcastic or not) as the output. The trained Bidirectional LSTM is then used for classify the Amazon review test datasets.

4. Results

The performance of the proposed solution is evaluated against Amazon product review datasets [23]. A total of 3000 reviews are taken from Amazon product reviews and manually labelled with sarcastic or not. This dataset is split in ratio of 80:20 to evaluate the performance. The performance is measured in terms of standard metrics: accuracy, precision, recall and F-measure. The performance is first compared between proposed hybrid feature with Bi-directional LSTM classifier and hybrid feature with traditional machine learning classifiers. The performance of proposed hybrid features is then compared against multi feature fusion framework proposed by Eke et al [14], CNN feature with LSTM model proposed by Jamil et al [12], hyperbole feature based sarcasm detection method proposed by Govindan et al [30].

The performance of proposed hybrid features with Bi-directional LSTM and traditional machine learning classifier is given in Table 3. The accuracy (Fig. 5) of proposed solution is atleast 1% higher compared to existing solution. The precision and recall are also higher in the proposed solution. The accuracy in Bi-directional LSTM is higher due to use of temporal correlation between the features over the window length but traditional classifiers are snapshot based. Use of temporal information across sentences in the same review has increased the recall in Bi-directional LSTM. The multi feature frame work by Eke et al, CNN feautr by Jamil et al and hyperbole feature by Govidan et al did not consider context information like sentiment spread in the texts, sentence contradiction etc in feature modelling. Also these approaches did not consider feature enhancement by learning from multiple datasets. Due to this, these features had lower discriminating ability compared to proposed solution in detection of sarcasm.

The performance of proposed hybrid features with and without adversarial learning from cross domain dataset is compared and the result is given in Table 4. The accuracy has increased by 4.9% due to cross domain transfer learning. Cross domain data has enhanced the feature space of the Glove embeddings and this has contributed to higher accuracy. Cross domain learning improvises the feature vocabulary both in content and context level. By exploiting it, the proposed solution has enhanced the features and this increased the accuracy in proposed solution.

Effectiveness of features with Bi-directional LSTM classifiers is given in Table 5. Hybrid features have atleast 5% more accuracy compared to content features and 6% more accuracy compared to context features. By combining both content and context features, the overall accuracy is improved in proposed solution. This gain is higher considering both content and context features in separation.

The performance of proposed hybrid features with Bi-directional LSTM is compared against existing works and the result is given in Table 6. The accuracy in proposed solution is atleast 1% higher compared to Eke et al. It is 11% higher compared to Jamil et al and 18% higher compared to Govindan et al.

The use of hybrid features along with transfer learning has increased the accuracy in proposed solution. Though Eke et al also used multiple feature fusion; they did not consider sentiment contradiction and sentence incongruity in their features. Due to this, their accuracy was atleast 1% lower compared to proposed solution. Jamil et al used only word embedding feature (content feature) and Govindam et al used only hyperbole features(context) and their accuracy is very low compared to the proposed hybrid features. The features designed in works of Jamil et al, Govindan et al and Eke et al cannot capture the sentiment spread across the document and sentence contradiction. But the proposed solution designed unique feature to capture sentiment spread and sentence contradiction. This has allowed the proposed solution to detect sarcasm spread across multiple sentences in the review text and this increased the accuracy in the proposed solution.

Table 3. Accuracy comparison for hybrid features

Classifiers	**Accuracy**	**Precision**	**Recall**	**F1-score**
Bi-directional LSTM	0.949	0.951	0.938	0.936
SVM	0.937	0.943	0.941	0.942
Decision tree	0.931	0.94	0.912	0.911
Random Forest	0.933	0.939	0.929	0.923

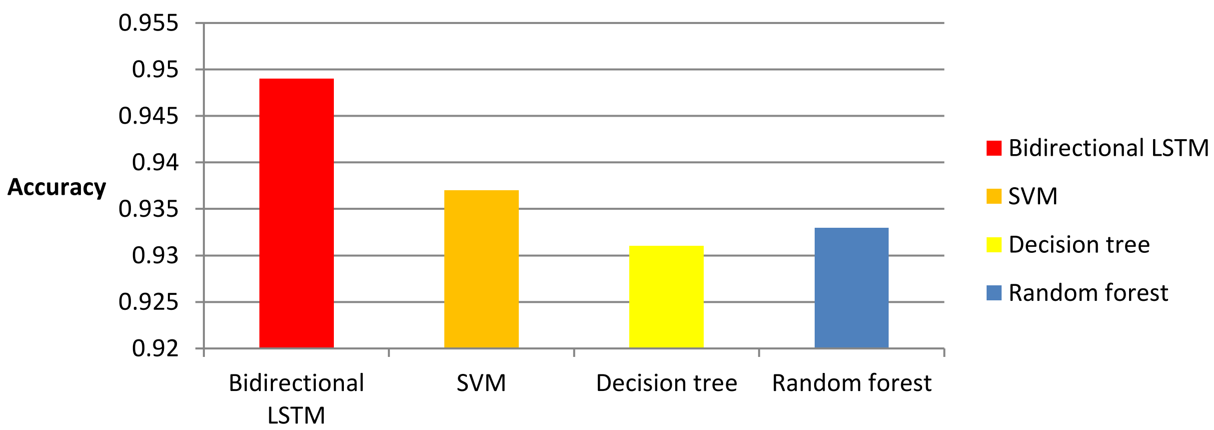

Figure. 5 Comparison of accuracy for hybrid features

Table 4. Comparison with/without transfer learning

Classifiers	Accuracy	Precision	Recall	F1-score
Bidirectional LSTM with adversarial learning	0.949	0.951	0.938	0.936
Bidirectional LSTM without adversarial learning	0.901	0.891	0.902	0.924

Table 5. Comparison between features

Classifiers	Accuracy	Precision	Recall	F1-score
Bidirectional LSTM + hybrid features	0.949	0.951	0.938	0.936
Bidirectional LSTM + content features	0.894	0.871	0.892	0.891
Bidirectional LSTM + context features	0.887	0.868	0.871	0.881

Table 6. Comparison of solutions

Classifiers	Accuracy	Precision	Recall	F1-score
Proposed	0.949	0.951	0.938	0.941
Eke et al	0.935	0.937	0.936	0.931
Jamil et al	0.832	0.835	0.832	0.832
Govindan et al	0.763	0.787	0.764	0.759

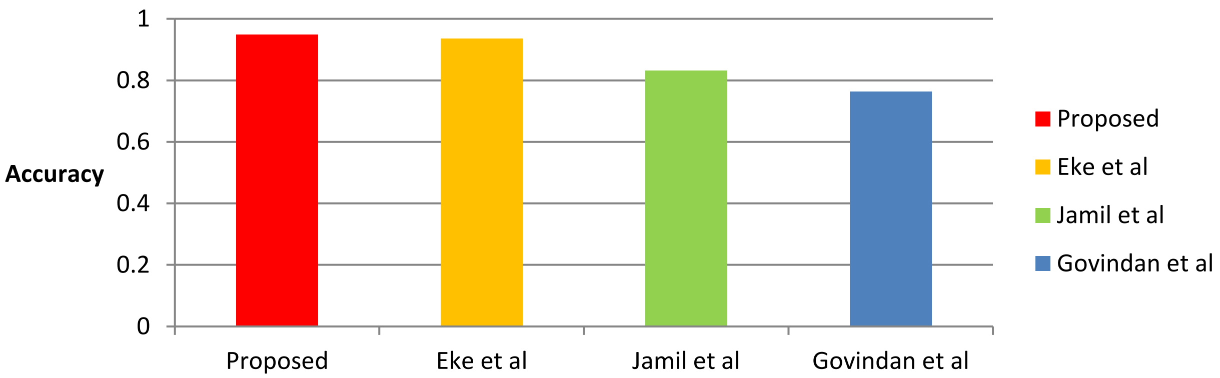

Figure. 6 Comparison of accuracy across solutions

Higher the ROC Area, better is the classifier performance in classifying the reviews. Plotting the true positive rate against various false positive rate, the ROC area in proposed solution is 0.907 which is higher compared to Eke et al (0.903), Jamil et al (0.853) and Kumar et al (0.82).

5. Discussion

Sarcasm detection is a important functionality of text classification. It is very important for various applications like depression detection, emotion detection etc. Most of the existing works considered only content information in sarcasm classification. Due to this sarcasm contradictions expressed in multiple sentences were missed out and this affects the accuracy of sarcasm detection in the text. This work proposed two solutions to this problem. In addition to content, two different context features of sentiment contradiction and sentence incongruity were extracted and used with content as a hybrid feature. Cross reference learning across multiple data sets is conducted to enrich the hybrid features, so that

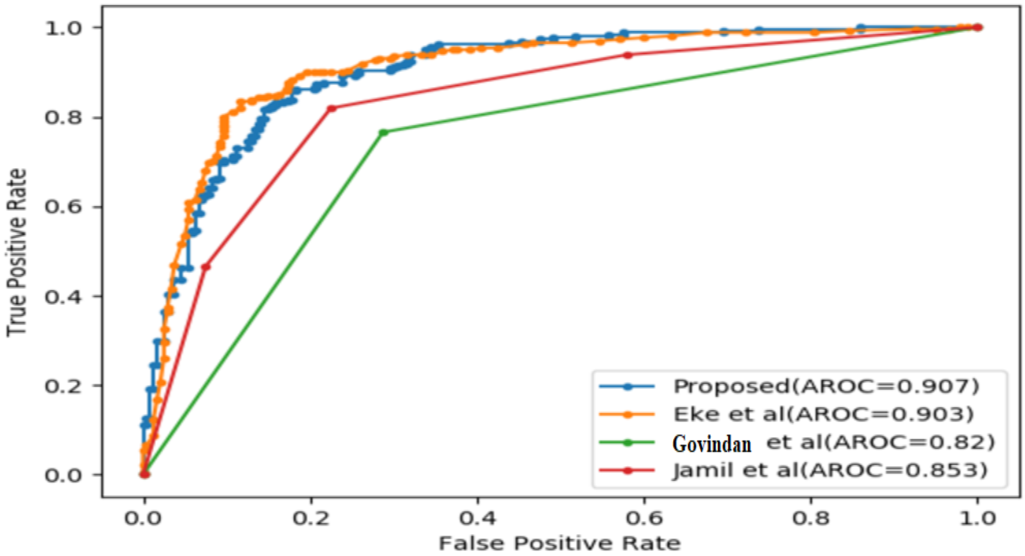

Figure. 7 ROC Curve

discriminative ability for sarcasm detection is improved. Eke et al used hybrid features for sarcasm classification. But the features were only lexical features collected using Bag of word technique. No context information was considered. Also the results did not consider temporal correlation between features. Jamil et al classified sarcasm using the deep learning features extracted from text using CNN. Though CNN features were better compared to hand crafted features, no context features were included. Also the approach lacked analysis of temporal correlation between the features. Govindan et al [30] used hyperbole features which captured the content but could not capture context in terms of sentiment contradiction and sentence contradiction. The proposed solution excelled better than existing works in context feature coverage.

6. Conclusion

A hybrid deep feature combining content, context and domain cues is proposed for sarcastic review detection in this work. The proposed solution addressed two problems of lack of context awareness and lack of sufficient labelled sarcastic dataset. The problems were addressed with hybrid features and cross domain adversarial learning. The proposed solution also considered temporal correlation between the adversarial learning enriched features with help of bi-directional LSTM. The hybrid features in combination with Bi-directional LSTM is able to achieve about 95% accuracy. The hybrid features in combination with traditional machine learning classifiers is able to achieve about 94% accuracy. The performance of the proposed solution can be further enhanced applying cross modality attention for further feature enrichment before passing to bi-directional LSTM.

Conflicts of interest (Mandatory)

The authors declare no conflict of interest.

Author contributions

Conceptualization and methodology are designed by both authors; Implementation, software validation by Parvati Kadli; Result analysisand Formal analysisis done by Parvati Kadli and Dr. Vidyavathi B M; Supervision and Project administration, investigationdone by Dr. Vidyavathi B M; Resources, data curation. writing—original draft preparation by Parvati kadli; review done by Dr. Vidyavathi B M; editing, visualization is by Parvati Kadli; supervision and project administration by Dr VidyavathiB M.

References

[1] https://www.digitalcommerce360.com/.
[2] J. Zhang, W. Zheng, and S. Wang, "The study of the effect of online review on purchase behavior: Comparing the two research methods", *International Journal of Crowd Science*, Vol. 4 No. 1, pp. 73-86, 2020.
[3] H. Cui and M. Zhu, "Collaboration filtering recommendation optimization with user implicit feedback", *Journal of Computational Information Systems*, Vol. 10, pp. 5855-5862, 2014.

[4] S. Muresan, R. Gonzalez, D. Ghosh, and N. Wacholder, "Identification of nonliteral language in social media: A case study on sarcasm", *Journal of the Association for Information Science and Technology*, Vol. 67, No. 11, pp. 581-586, 2016.

[5] P. Carvalho, L. Sarmento, M. Silva, and E. Oliveira, "Clues for detecting irony in user-generated contents", In: *Proc of 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion, ACM*, pp. 29-36, 2009.

[6] V. Govindan and V. Balakrishnan, "A machine learning approach in analysing the effect of hyperboles using negative sentiment tweets for sarcasm detection", *Journal of King Saud University – Computer and Information Sciences*, Vol. 34, No. 8, pp. 5110-5120, 2022.

[7] X. Guo, B. Li, H. Yu, and C. Miao, "Latent-optimized adversarial neural transfer for sarcasm detection", In: *Proc of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 5394-5407, 2021.

[8] M. Bedi, S. Kumar, S. Akhtar, and T. Chakraborty, "Multi-modal Sarcasm Detection and Humor Classification in Code-mixed Conversations", *arXiv:2105.09984*, 2021.

[9] A. Onan and M. Tocoglu, "A Term Weighted Neural Language Model and Stacked Bidirectional LSTM Based Framework for Sarcasm Identification", *IEEE Access*, Vol. 9, pp. 7701-7722, 2021.

[10] Y. Zhang, Y. Liu, Q. Li, P. Tiwari, B. Wang, Y Li, H. Pandey, P. Zhang, and D. Song, "CFN:A Complex-Valued Fuzzy Network for Sarcasm Detection in Conversations", *IEEE Transactions on Fuzzy Systems*, Vol. 29, No. 12, pp. 3696–3710, 2021.

[11] C. Eke, A. Norman, and L. Shuib, "Context-Based Feature Technique for Sarcasm Identification in Benchmark Datasets Using Deep Learning and BERT Model", *IEEE Access*, Vol. 9, pp. 48501-48518, 2021.

[12] R. Jamil, I. Ashraf, F. Rustam, E. Saad, A. Mehmood, and G. Choi, "Detecting sarcasm in multi-domain datasets using convolutional neural networks and long short term memory network model", *PeerJ Computer Science*, Vol. 25, No. 7, p. e645, 2021.

[13] S. Sonawane and R. Kolhe, "TCSD: Term Co-occurrence Based Sarcasm Detection from Twitter Trends", *Procedia Computer Science*, Vol. 167, pp. 830-839, 2020.

[14] C. Eke, A. Norman, and L. Shuib, "Multifeature fusion framework for sarcasm identification on twitter data: A machine learning based approach", *PLoS ONE*, Vol. 16, No. 6, p. e0252918, 2021.

[15] S. Pramanick, A. Roy, and V. Patel, "Multimodal Learning using Optimal Transport for Sarcasm and Humor Detection", In: *Proc of IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 546-556, 2022.

[16] T. Jain, N. Agrawal, G. Goyal, and N. Aggrawal, "Sarcasm detection of tweets: A comparative study", In: *Proc of Tenth International Conference on Contemporary Computing (IC3), IEEE*, pp. 1-6, 2017.

[17] A. George, B. Ganesh, A. Kumar, and K. Soman, "Significance of Global Vectors Representation in Protein Sequences Analysis", *Computer Aided Intervention and Diagnosticsin Clinical and Medical Images, Springer*, pp. 261–269, 2019.

[18] https://web.archive.org/web/20200211135658/http://sentistrength.wlv.ac.uk/

[19] https://sentic.net/

[20] https://conceptnet.io/

[21] E. Camp, "Sarcasm, pretense, and the semantics/pragmatics distinction", *Noûs*, Vol. 46, No. 4, pp. 587-634, 2012.

[22] H. Salehinejad, J. Baarb´e, S. Sankar, J. Barfett, E. Colak, and S. Valaee, "Recent advances in recurrent neural networks", arXiv:1801.01078, 2017.

[23] https://jmcauley.ucsd.edu/data/amazon/

[24] https://www.kaggle.com/datasets/danofer/sarcasm

[25] https://www.kaggle.com/competitions/gse002/

[26] A. Kumar and G. Garg, "Empirical study of shallow and deep learning models for sarcasm detection using context in benchmark datasets", *Journal of Ambient Intelligence and Humanized Computing*, Vol. 1, No. 16, 2019.

[27] A. Aliwy, A. Abbas, and M. Hadi, "Key Challenges and Proposed Solutions to Design Sentiment Analysis System", *International Journal of Intelligent Engineering and Systems*, Vol. 15, No. 4, 2022, doi: 10.22266/ijies2022.0831.24.

[28] Z. Wen, L. Gui, Q. Wang, M. Guo, X. Yu, J. Du, and R. Xu, "Sememe knowledge and auxiliary information enhanced approach for sarcasm detection", *Information Processing and Management*, Vol. 59, No. 3, p. 102883, 2022.

[29] A. Amer and T. Siddiqu, "A novel algorithm for sarcasm detection using supervised machine learning approach", *AIMS Electronics and Electrical Engineering*, Vol. 6, No. 4, pp. 345-

369, 2022.

[30] V. Govindan and V. Balakrishnan, "A machine learning approach in analysing the effect of hyperboles using negative sentiment tweets for sarcasm detection", *Journal of King Saud University - Computer and Information Sciences*, Vol. 34, No. 8, pp. 5110-5120, 2022.