

DEEP-LEARNED CROSS-DOMAIN SENTIMENT CLASSIFICATION USING INTEGRATED POLARITY SCORE PATTERN EMBEDDING ON TRI MODEL ATTENTION NETWORK

Parvati Kadli

Associate Professor, Computer Science Engineering,
Visvesvaraya Technological University, Jnana Sangama, Belagavi
Belagavi, Karnataka, 590018/North, India

kadli.parvati@gmail.com

<http://www.pdit.ac.in>

Dr. Vidyavathi B M

Professor, Artificial Intelligence and Machine Learning, BITM, Jnana Gangotri Campus, Near Allipur, Ballari
Ballari, Karnataka, 583104/North, India

Vidyahm1@gmail.com

<http://www.bitm.edu.in>

Abstract

Classification of sentiments is a challenging task when it is done across multiple domains. The dissimilarity between the domains, reflected in their sentiments makes the classification tougher. This work proposes a deep learned novel architecture named Integrated Polarity Score based Pattern Embedding on Tri Model Attention (IPSPE_TMA) Network which is a model based on Bidirectional Long-short Term Memory (Bi-LSTM), Bi-Gated Recurrent Unit (GRU) and Convolutional Neural Network (CNN) with three stage novel embedding architecture and attention network. The information is extracted by incorporating attention in word and sentence level. In order to generate robust vectors, a novel polarity score based embedding is proposed and used along with Glove and Fast Text in this paper. Integration of three models with three kinds of embeddings and attention mechanism strengthens this model. This network implementation for sentiment analysis on cross domain dataset gives better performance than many of the previous works.

Keywords: Cross-Domain, Word Embedding, Polarity Score Pattern.

1. Introduction

Sentiment analysis paves a path in today's ever-changing world to understand and get modified themselves to the feedbacks of others. By detecting and analysing the overall opinion about the brand which reflects the reason for drop or rise, the decisions on their marketing strategies. In social media, this strategy is used for getting the public opinion and in teaching-learning system, useful information can be gathered about teaching methodology, course curriculum, students' learning curve etc. Sentiment classification classifies orientation of sentiments into appropriate labels like positive or negative feel. Sentiment classification [Pang *et al.*, (2002); Turney, (2002)] Pang and Lee, (2008)] may be based on textual document, emoji, face image or video. Though binary sentiment classification is common, analysis of documents based on three-class classification (positive, neutral and negative) and multiple-class sentiment classification problem (angry, sad, happiness etc) are also used by researchers. Sentiment analysis based on text data can be done lexicon-based or machine learning-based methods. Lexicon which shows emotions are used to gather the keywords which shows the emotions. Mass crowdsourcing [Xing *et al.*, (2019)] is a method used to predict the feel of tweets. This is done based on the features which are present in the dataset. The works use lexicons [Pang *et al.*, (2002)] to get trained and classified concurrently. The wrongly predicted sentences are identified and used as the learning for further testing. The usage of machine learning algorithms improved the effectiveness of sentiment analysis system. Naive Bayes (NB), Maximum Entropy (ME)

and support vector machine (SVM) (Pang *et al.*, 2002) are used for identifying the sentiment classes and its analysis in some works. Desai *et al.* analysed sentiments using different machine learning techniques for twitter data. Machine learning models learn the system based on the labels of data. Accuracy will get increased, if more labelled data are used for training. But usage of more data makes the machine learning system complex since manual labelling of data is expensive and time consuming. This problem was solved by proposing a technique called transfer learning [Meng *et al.*, (2015); Pan and Yang, (2010)]. In this method, source and target domains are analysed for finding the likeness and the information about the labelled data will be transferred from source to the target, if they both are similar. This further enhances the learning by inherently transferring the knowledge. New problem which is based on the previous problem can be solved by transferring the learning from the old problem and by modifying it.

In text sentiment analysis [Tahmoresnezhad and Hashemi, (2017); Wu and Tan, (2011)], many applications uses this learning technique. Spectral feature alignment [Pan *et al.*, (2010)] was an algorithm used to find the alignment of words which belongs to specific domain and cluster those words for future use. Structural correspondence learning (SCL) algorithm [Blitzer *et al.*, (2007)] was used for identifying the likeness between the target and source domain data. Deep learning based architectures [Ganin *et al.*, (2015)] are also proposed in transfer learning where feed forward network was used for transfer learning.

The increased research in sentiment classification further improved their usage in wider applications like product recommendations, business, movie reviews etc. Evolution of deep learning enhanced better sentiment representations and its classifications. This helped in classifying sentiments using cross-domains. It uses two domains, source and target. Sentiments using cross-domains learns the knowledge by using source domain data to classify the sentiments of target domain. Inter-domain transferring of knowledge will be done in cross-domain sentiment classification which is a challenging task because of the difference in representing sentiments for varying domains. If the reviews are taken for two products: Food and Furniture, first set may register opinions as “spicy” or “yummy” and second set may have “strong” and “attractive”. Lesser similarity in the emotional tendencies of different domains, challenges the classification techniques in bridging those domains to diverge into a class.

This paper proposes a novel deep learning framework which does the sentiment classification in cross domain with the hybridization of three kinds of network models which are integrated with the state of art embeddings such as Glove and Fast Text with proposed IPSPE embedding scheme. Word Embedding glove and Fast text give more generalized vector representation of data and don't provide clear information on the polarity or emotional information. Hence to obtain the semantic context in the more discriminative way an embedding scheme IPSPE is proposed based on the polarity score.

2. Related Work

Cross-domain sentiment classification developed as an important research area because of its wider applications and effectiveness. Domain-based sentiment words are used for finding the correlation between the words in pivot-based methods [Blitzer *et al.*, (2007); Jianfei *et al.*, (2016); Stephan *et al.*, (2012); Yftah Ziser and Roi Reichart, (2018)]. These methods find the alignment of non-pivot or domain specific opinion words along with pivot words or domain shared opinion words. Prediction algorithms based on multiple pivot words are used in these techniques for classification. Many researches have been made in domain adaptation challenge [Bollegala, *et al.*, (2011); yulan *et al.*, (2011); He *et al.*, (2011)] in the sentiment classification where sentiment classification performance is very poor for a classifier which has trained from another domain. Selection of pivot words for implementing these methods is very difficult and inaccurate because the emotional tendency of different domains is different. Denoising auto encoder [Yi Yang and Jacob E, (2014); Minmin *et al.*, 2012; Xavier, (2011)] was proposed for domain-based sentiment classification by reconstructing the input noise for adapting to the domain. Though these techniques give better accuracy, discrete representation of this is the problem with these methods. Latent feature representations [Werner *et al.*, (2017); Baochen *et al.*, (2016)] are used to address this challenge by minimizing the differences between the features of both domains. Differences between the features of the domains were reduced by proposing different adversarial networks [Yaroslav G *et al.*, (2016); Zheng *et al.*, (2017).] The reduction in feature difference was made possible by fooling the discriminator in the domain. A domain classifier was exploited in the domain adversarial network [Yaroslav G *et al.*, (2016)] for minimizing the difference between the two domains. Though the fooling of discriminator can be achieved by these adversarial methods, category-specific decision boundaries are ignored here and alignment of the distribution in margins are not finite. By considering samples as unlabelled, semi-supervised learning algorithms [(Olivier *et al.*, (2009); Yucen *et al.*, (2006)] were proposed for this classification problem.

Transfer learning based methods are also used to implement sentiment classification. The ways of categorizing sentiment classification strategies with respect to transfer learning are based on feature, relationship, model and sample. Feature-based methods for transfer learning, works with the assumption of absence of both domain features in common space, which further makes the system transformed to similar space. A feature-based learning method was proposed where some unique features are mapped from one space to all the other spaces. This learning

is based on the structural correspondence learning. Some works [Yucen *et al.*, (2016), Long *et al.*, (2016), (2017)] are proposed by combining with neural network models. Relationships between both domains are used in the relationship-based transfer learning method. It mainly focuses on the relation between the target and source domain samples. Markov logic network [Davis and Domingos, (2009)] is an example for this method. The model-based transfer learning methods are usually implemented using neural network models [(Xavier *et al.*, (2011)], which helps in transferring the shared parameters. These kinds of methods perform better than traditional and sample based methods because of the reduction in the dependence of pivot words and labelled data of source domain. Samples are transferred through weight reuse in sample-based transfer learning method. Adaboost [Dai *et al.*, (2007)] was proposed as the sample-based transfer learning method for improving sample weights which are beneficial for target classification task. Transitive transfer learning [Tan *et al.*, (2015)] and distant domain transfer learning [Tan *et al.*, (2017)] are two other examples for sample-based methods.

The evolution of deep learning algorithms influenced and helped to achieve better results in the sentiment classification [Kunpeng *et al.*, (2011)] too. Different categories of deep learning are used for sentiment classification. Convolutional Neural Network (CNN) [(Nal Kalchbrenner *et al.*, (2014) ; Yoon Kim, (2014)] is the first category algorithm used for classification, which learns the patterns across the sentiment to classify it. Recurrent Neural Network (RNN) [Zichao *et al.*, (2016)]; Peng *et al.*, (2016); Preethi *et al.*, (2017) is the next category of algorithm which works based on the temporal data. Recursive Neural Network (RecNN) (Richard *et al.*, 2013) is another category which learns recursively in the tree structure by assigning same weights. (Zhang *et al.*, 2018) reviewed the various sentiment classifications and found that deep-learning methods perform well. Sentiment analysis for Chinese text [Lai *et al.*, (2017)] was done by CNN and word vector. End-to-end adversarial memory network (AMN) [Li *et al.*, (2017)] was proposed where two-parameters are used. CNN and AMN can be combined together to reduce the classification error. Cellular neural network (Fortuna *et al.*, (2001) is another network proposed for sentiment classification. In this, an embedding layer will be used for mapping word to dense vector. These dense vectors are then fed to different network architecture models to generate classification result.

3. Proposed Polarity Score based Embedding Pattern

3.1. Embedding Schemes

Bag-of-words model was used for representing text for sentiment classification since it is easy to construct. But the dimensions of features are very huge because of its data are very huge. This will lead to sparse representation of word vector matrix which may lead to overfitting. This motivated Mikolov *et al.*, (2013) to introduce a model called word embedding which helps to learn vector representation of words. Representation of sentences can be done by mapping the words into low dimensional vectors which is known as word embedding [Bengio *et al.*, (2003)]. Here the vectors are represented in a continuous space and also the words will be mapped onto the neighbouring points. An architecture for natural language processing was proposed by Collobert *et al.*, (2008, 2011) which focuses on same embedding for different tasks. Another work [(Dragoni and Petrucci, (2017)] was proposed for predicting the sentiments by using neural word embedding.

Predictive methods like Probabilistic Language Models (NNLM) (Bengio *et al.*, 2003), N-grams and Word2Vec (Mikolov *et al.*, 2013) are used for embedding. Word2Vec uses a feedforward network. It has input, projection and output layers. The two models used for prediction are Continuous Skip-gram Model (Skip-gram) and Continuous Bag-of-Words Model (CBOW). Skip gram considers the words in the sentence for maximising the classification of words. Yitao and Wan, (2019) proposed a multidomain neural sentiment classification model which is shared completely for understanding and learning the domain-aware word embeddings. This is done by sharing word embeddings of various domains for achieving domain-specific information. A new word representation named GloVe (Pennington *et al.*, 2014) was proposed for NLP tasks. It uses the advantages of local context window methods and global matrix factorization. Further two word embedding models named FastText [Bojanowski *et al.*, (2017)] and ELMo [Peters *et al.*, (2018)] are also proposed by researchers.

3.1.1. Proposed Polarity Score Binary Pattern Vector (PSBPV)

This paper incorporate the strength of rule based sentiment lexicon especially VADER (Valence Aware Dictionary for sEntiment Reasoning) to generate a bit binary pattern for word embedding. VADER [Gilbert and Eric, (2014)] is especially attuned to microblog-like contexts. It maps the lexical features of a given text to emotion intensities called sentiment scores. This tool finds the intensity of each word in the text and sums it to calculate the sentiment score of the text. It provides score for positive, negative, neutral and compound form. It provides score in the range of -1 to 1 that is from most negative to most positive. In our scheme, the embedding bit pattern is generated with the help of compound score. After the pre-processing of the sentence, such as stop word removal, stemming and etc, the sentence is split into words. For each word the polarity score is estimated with the help of VADER.

In this model compound score is estimated for each word and the pattern is generated based on this value belongs to a particular range. The generation of Polarity Score based Binary Pattern is explained in the following Pseudocode of table 1.

```

for each word w in sentence:
    [positivescore, negativescore, neuralscore, compoundscore] = Vader Polarity Score(w)
    if compoundscore >= 0.6
        PBPw = [1, 0, 0, 0, 0]
    else if compoundscore >= 0.3
        PBPw = [0, 1, 0, 0, 0]
    else if compoundscore >= 0
        PBPw = [0, 0, 1, 0, 0]
    else if compoundscore >= -0.5
        PBPw = [0, 0, 0, 1, 0]
    else
        PBPw = [0, 0, 0, 0, 0]
    end
end

```

Table 1. Pseudocode

For Example:

Let us consider the input sentence is like this

I	would	Highly	Recommend	this	drive	for	performance	and	ease	of	use.
---	-------	--------	-----------	------	-------	-----	-------------	-----	------	----	------

After removing stop words it will be

Highly	Recommend	Drive	performance	ease	Use
--------	-----------	-------	-------------	------	-----

The scores obtained using VADER for each word as follows

	High	Recommend	Drive	Performance	Ease	Use
Positive Score	0.0	0.1	0.0	0.0	1.0	0.0
Negative Score	0.0	0.0	0.0	0.0	0.0	0.0
Neutral Score	1.0	0.0	1.0	1.0	0.0	1.0
Compound Score	0.0	0.3612	0.0	0.0	0.3612	0.0
Polarity Score based Binary Pattern Vectors	0 0 1 0 0	0 1 0 0 0	0 0 1 0 0	0 0 1 0 0	0 1 0 0 0	0 0 1 0 0

Apart from the state of art embedding such as GloVe, FastText and Word2vec, the proposed embedding allocates coefficients to each word to distinguish its sentiment lexicon. By this sentiment score based word level embedding it helps the model to predict the sentiment in cross domain. Each word has its unique score and unique meaning, but it represents different sentiment based on the domain and neighbouring words. Hence this embedding along with BiLSTM model is incorporated in this proposed Tri model network to improve the performance of the sentiment analysis in cross domain.

3.1.2 GloVe Model

The number of occurrences of words which is the primary source of available information is used in this model by generating the meaning between these. By using this information, a new model was created for representing the words called as GloVe (Global Vectors). Consider i , j and k are three words. The number of co-occurrences for these words are N_{ij} and N_{jk} . N_{ij} is the co-occurrence of i and j and N_{jk} is the co-occurrence of j and k . By using this information along with its probability, word embedding is made possible in GloVe.

3.1.3 FastText Model

FastText is proposed by Facebook AI in the year 2016. It is a NLP library used for word embedding. FastText uses Continuous Bag of Words (CBOW) model for representing the words with a hierarchical classifier. CBOW can predict the word by predicting a category. In this hierarchical softmax replaces softmax layer.

Sentiment classification yields more output when the focus is more on words and phrases. Attention mechanism [Dzmitry,(2014); Minh et al.,(2015); Ashish et al.,(2017)] was used by the researchers for achieving this goal by capturing the distance dependency. Hierarchical attention networks [Zichao et al.,(2016)] were proposed both in sentence and word level for enhancing the importance of word and sentence. Another attention model [Zhigang et al.,(2018)] was proposed for multi-tasking which shares the feature extractor. Shareable sentence representation learning [Renjie et al.,(2018)] uses attention mechanism along with task-independent query vectors for generating various representations. Cai and Wan, (2019) proposed a multidomain neural sentiment classification model which made use of domain-aware attention mechanism. Zhang et al.,(2019) proposed an Interactive attention transfer network and analyzed the effect of aspect network attention and sentence network attention.

4. Proposed Architecture – Tri Model with IPSBPE (IPSPE_TMA)

Long short-term memory (LSTM) is a Recurrent Neural Network (RNNs) which can learn the dependence of order in predicting the sequence. Text classification using LSTM showed considerable output. Sparsity in the text data leads to high complexity when LSTM is used. Because of the complexity of this algorithm when it is used for text classification, bidirectional LSTM (BiLSTM) was proposed. This work proposes a unified architecture which contains a BiLSTM [Yitao and Wan, (2019)]. BiGRU [Dzmitry et al.,(2014)] and CNN attention mechanism along with new word embedding to strengthen the pre-trained word embedding. The proposed model architecture is shown in the figure 1.

This proposed network consists of three kinds of model. In the First model, the Glove embedding is used to generate the embedding sequences, which is then fed into BiLSTM layer with 256 nodes and followed by an Average pooling layer with pool size of 2 and Context Attention layer with dropout of 0.5. Finally 100 dense features are extracted from this first model. The second model is constructed with the help of BiGRU with 128 nodes on Fast Text embedding. Similar to first model the subsequent layers are applied and finally extract 100 dense features. The third integrated model consists of three kinds of convolutional layers with different number filters and kernel size. The first convolutional layer with 100 filters with kernel size as 3 and relu activation on glove embedding, whereas the second and third convolutional layers are constructed with number of filters as 6 and 300 and its corresponding kernel size as 3 and 5. Relu and softplus activations are applied after the second and third convolutional layer respectively on proposed polarity score based embedding and FastText embedding. Each convolutional layer fed into average pool layer with pool size as 2. Average pool from second convolutional layer on PSBP is fused with the average pool of the first and third convolutional layer using concatenation layer, which forms an Integrated Polarity Score based Binary Pattern Embedding (IPSBPE) with Glove and Fasttext CNN model.

The output from those two concatenation layer is further effectively processed through Attention layer followed by Dropout layer and dense layer with 100 units. All those features from the four kinds of dense layers are concatenated together to form an effective deep features to represent the sentiment polarity of a sentence with the help of Softmax layer.

In this proposed model, the proposed PSBPV is a 5 bit word embedding which extracts the global and local information of a sentence efficiently. BiLSTM accesses both the preceding and succeeding context representations. Attention mechanism is employed to give different focus on the information outputted from the hidden layers of BiLSTM, BiGRU and Integrated model of CNN. Finally, context information is processed by the softmax classifier appended to the dense layer to classify the processed. This Architecture is able to capture both the global sentence semantics and as well as the local feature of phrases.

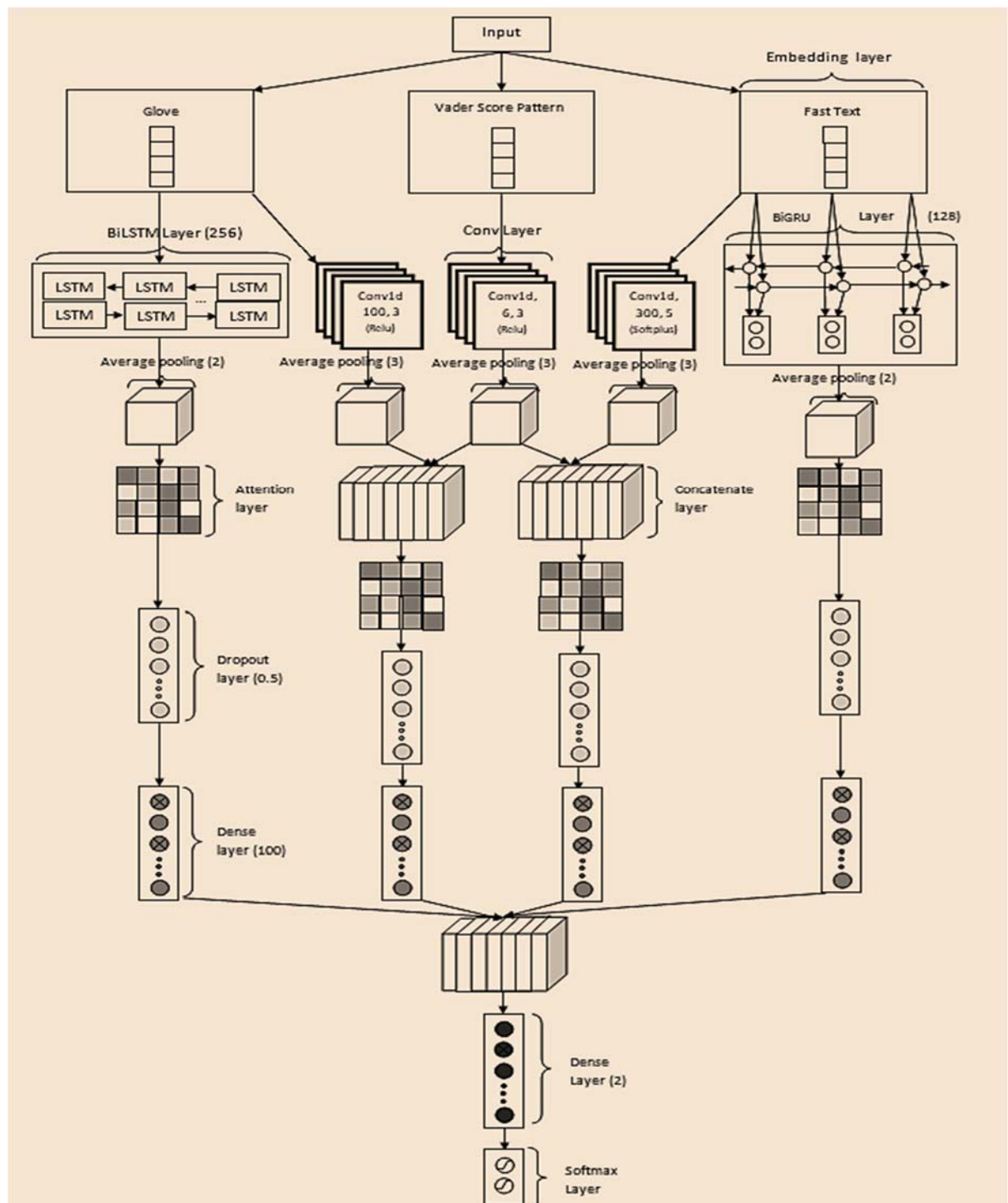


Fig.1. Architecture of IPSPE_TMA

4.1. BiLSTM

The proposed method has applied BiLSTM model for classification. BiLSTM represents a sequence processing form which contains two LSTMs. First LSTM collects the input in a forward direction. Second LSTM works in a backward direction. Word embedding is generated with the help of Glove and those extracted features are fed into BiLSTM layer to extract sentence level features to represent the each sentiment polarity of the sentence. Domain exact attention system is utilized for choosing important features along with the BiLSTM layer to improve the performance of the approach.

4.2. Bi-GRU

To get annotations of words, a bidirectional GRU is used. Here summarizing of information takes place from both directions for words, and thus incorporates the contextual information in the annotation.

The bidirectional GRU consist of the forward GRU \vec{f} and the backward GRU \tilde{f} . The forward GRU \vec{f} examines the sentence s_i from w_{i1} to w_{iK} as shown in Eq. (2). The backward GRU \tilde{f} examines the sentence s_i from w_{iT} to w_{i1} as shown in Eq. (3).

$$x_{ik} = W_e w_{ik}, k \in [1, K] \quad (1)$$

$$\vec{h}_{ik} = \overrightarrow{GRU}(x_{ik}), k \in [1, K] \quad (2)$$

$$\tilde{h}_{it} = \overleftarrow{GRU}(x_{ik}), k \in [K, 1] \quad (3)$$

An annotation for a given word w_{ik} is obtained by concatenating the forward hidden state \vec{h}_{ik} and backward hidden state \tilde{h}_{ik} , i.e., $h_{ik} = [\vec{h}_{ik}, \tilde{h}_{ik}]$. This model also summarizes the data of the entire sentence w_{ik} .

Gated Recurrent Unit (GRU) is a specific type of Recurrent Neural Network (RNN). GRU [Dzmitry *et al.*, (2014)] combines the forget gate and the input gate into a single update gate. GRU even combines the cell state and the hidden state at the same time. The unit structure is shown below in the Figure 2.

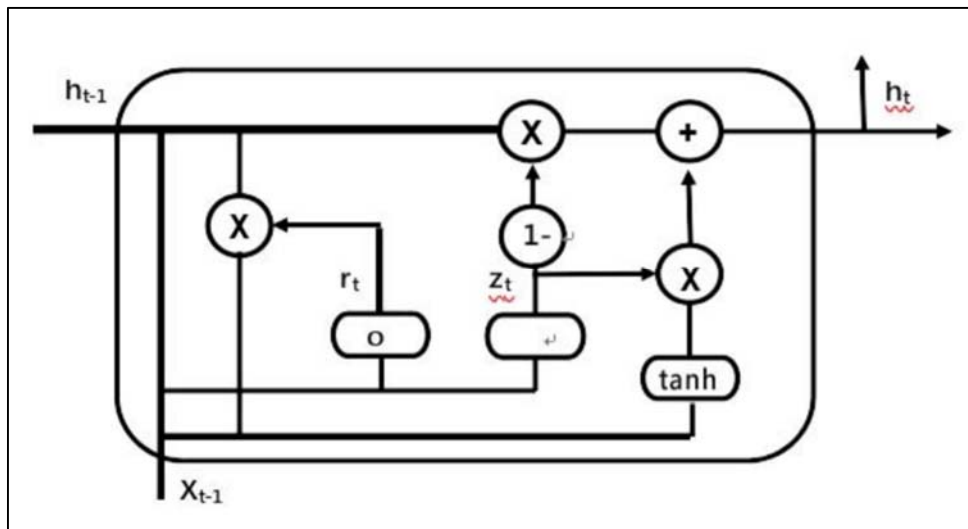


Fig.2. GRU Coding Unit [Dzmitry *et al.*, (2014)]

The specific estimation procedure is performed following Eqs (4) – Eqs (7).

$$Z_t = \sigma(W_i * [h_{t-1}, x_t]) \quad (4)$$

$$r_t = \sigma(W_i * [h_{t-1}, x_t]) \quad (5)$$

$$\tilde{h}_t = \tanh(W_c * [r_t \cdot h_{t-1}, x_t]) \quad (6)$$

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \quad (7)$$

Where σ is represented as the sigmoid function.

\cdot is represented as the dot product.

x_t is represented as the input vector at time t .

h_t is represented as the hidden state at time t .

z_t is represented as an update gate.

r_t is represented as a reset gate.

This update gate controls the information flows into the next process. Hence, the control information is lost. The reset gate determines the output of the hidden state.

4.3. Convolution layer

Convolutional layer is formed from a set of convolutional kernels where every neuron acts as a kernel. Convolution operation becomes a correlation operation [Ian *et al.*, (2017); Bouvrie, (2006)] when a kernel is symmetric. Convolution operation can be expressed as shown in Eq. (8).

$$f_l^k(p, q) = \sum_c \sum_{x,y} i_c(x, y) \cdot e_l^k(u, v) \quad (8)$$

where $i_c(x, y)$ is represented as an element of the input tensor I_c .

This input is element wise multiplied by value of $e_l^k(u, v)$. $e_l^k(u, v)$ is considered as the index position of the k^{th} convolutional kernel l of the 1^{st} layer. The output feature-map of the k^{th} convolutional operation is represented as $F_l^k = [f_l^k, \dots, f_l^k(p, q), \dots, f_l^k(P, Q)]$.

Normally, various sets of features are extracted using sliding kernel concept for an input. Hence, convolutional operation has ability of weight sharing system. In fully connected networks, CNN parameter provides efficient results with the same set of weights on an input, the input may be an image or an audio, text and signal. Convolution operation is also divided into various types based on the type of padding, the size of filters and the direction of convolution [(Le Cun *et al.*, (2015))]. In this work, three kinds of one dimensional Convolution layer is used to extract cross domain features which has significant strength to represent each domain properly. Two Convolutions layer fetch the embedding feature from the traditional embedding layer, and one Convolutional layer extract the features in the form of correlation mode with the polarity based embedding and traditional embedding layers.

4.4. Average pooling

Pooling method [Zhang *et al.*, (2019)] is used to decrease the spatial size of the representation and preserving the important features. It is also known as a form of non-linear down-sampling method. Thus, in this approach after getting the hidden state representations of sentences the pooling method is adopted. There are various non-linear functions to execute pooling. Among the functions, pooling works in a superior way. Finally, the sentence pooling vector (i.e., h_s^p) is estimated throughout by applying the following Eq. (9).

$$h_s^p = \sum_{i=1}^n h_a^i / n. \quad (9)$$

4.5. Context attention layer

All the words don't donate evenly to the representation of the important to the sentence meaning. Hence, we utilize the Attention mechanism technique to find out the most dominant feature vector of words to represent the sentence. The representation of those informative words is used to form a sentence vector [Zichao *et al.*, (2016)]

$$u_{it} = \tanh(W_w h_{it} + b_w) \quad (10)$$

$$\alpha_{it} = \frac{\exp(u_{it}^T u_w)}{\sum_t \exp(u_{it}^T u_w)} \quad (11)$$

$$s_i = \sum_t \alpha_{it} h_{it} \quad (12)$$

The word annotation h_{it} is fed throughout from one layer MLP to get u_{it} as a hidden representation of h_{it} . The significant of the word similarity of u_{it} is identified with a word level context vector u_w . This method gets a normalized importance weight α_{it} throughout with a softmax function. The sentence vector s_i is estimated with a weighted sum of the word annotations depending on the weights. The context vector u_w is available at a high level representation of a set query. For example, the query can be "what is the informative word". This method is mostly available in all the memory networks [Sainbayar *et al.*, (2015), (Ankit *et al.*, (2015))]. The word context vector u_w is accidentally initialized and thus during the training procedure period it is jointly learned.

4.6. Dropout layer

Dropout is applied in this proposed work to evade over fitting problem. In this model after the attention mechanism the dropout process is done with the help of dropout layer with dropout rate as 50%. This dropout layer is utilized to concentrate on the distinguished upcoming parameters as an alternative approach to avoidable parameters [Sangeetha and Prabha, (2021)].

4.7. Dense layer

A dense layer consists of neurons of network layer. Dense layer collects the input from all layers using the neurons present there. Each dense layer has components, namely a weight matrix w , a bias vector b , and an activations of previous layer a . The dense layer is represented by Eq. (13).

$$y = (a(x.w) + b) \quad (13)$$

where a is represented as the element wise argument, w is represented as the weights matrix and bias is represented as a bias vector produced by the layer [Sangeetha and Prabha, (2021)].

4.8. Softmax layer

The output layer of each Neural Networks (NN) is represented as the Softmax layer. Depending on the deep score, the nature of the input text and the functions of the hidden layer are predicted. In the Softmax layer, the sentiment polarity of review in cross domain [(Han *et al.*, (2020))] is predicted by applying the deep score. The review sentence illustration z_s is given into the Softmax classifiers to envisage the probability distribution $\hat{y}_s \in \mathbb{R}^c$ of sentiment categories as represented in Eq. (14).

$$\hat{y}_s = \text{softmax}(W_s z_s + b_s) \quad (14)$$

where c is represented as the number of sentiment categories, and assigned $c = 2$. In this proposed model, $W_s \in R^{c \times d_h}$ and $b_s \in R^c$ are represented as the weight matrix and the bias respectively.

5. Performance Evaluation

The performance metrics applied to assess the classification results are F-measure, Precision, Recall, and Accuracy. These performance metrics are calculated depending on the values of True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) for the given classes [Parvati and Vidyavathi, (2020)]

True Positive (TP): It is known as the correctly calculated positive values. TP defines that the value of real class is represented as yes and the value of calculated class is represented as also yes.

True Negatives (TN): It is known as the correctly calculated negative values. TN defines that the value of real class is represented as no and the value of calculated class is represented as also no.

False Positives (FP): FP defines that the value of real class is represented as no and the value of calculated class is represented as yes.

False Negatives (FN): FN defines that the value of real class is represented as yes and the value of calculated class is represented as no.

5.1. Precision

Precision is known as the ratio of accurately calculated positive annotations to the total calculated positive annotations. It is given by Eq. (15).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (15)$$

5.2. Recall (Sensitivity)

Recall is known as the ratio of accurately calculated positive annotations to the all annotations in real class. Recall is also called as Sensitivity. It is given by Eq. (16).

$$\text{Recall} = \frac{TP}{TP+FN} \quad (16)$$

5.3. F1 Score

F1 Score is known as the weighted average of Precision and Recall. This is a bench mark metric. Hence, this score considers both False Positives (FP) and False Negatives (FN) into relation. It is given by Eq. (17).

$$F1_Score = \frac{2 * (\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})} \quad (17)$$

5.4. Accuracy

Accuracy is known as the most perceptive performance measure. It is defined as a ratio of accurately calculated annotation to the total annotations. It is given by Eq. (18).

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (18)$$

6. Implementation Details

6.1. Datasets

The datasets used for our work are Amazon Reviews of four domains built by Blitzer *et al* (2007). The domains are Kitchen(K), DVD(D), Book(B) and Electronics(E). Each of these has positive and negative Reviews. All reviews marks of a rating between (0-5 stars), a reviewer name and location, a product name, a review title and date, and the review text. Reviews with rating > 3 are labeled positive, those with rating < 3 were labeled negative, and the rest discarded because their ambiguous polarity. After this conversion, there are 1000 positive and 1000 negative examples for each domain, a balanced composition. The model is trained with these reviews of one domain and tested on 400 reviews of another domain (200 positive and 200 negative).

6.2. Pre-processing

The sentences are split and stop words are removed. These are converted to word vectors using glove embeddings and fast-text embeddings and are used as input to the model

6.3. Implementation

This work focuses on a novel framework for cross-domain sentiment classification which utilises the information from one domain for classifying the data of another domain. Let us consider the source domain as Dm_s and the destination domain as Dm_d . The words and its sentiments in the source domain are used to learn which will be utilised in destination domain. Let the labelled training data in source domain be D_s^{ld} , unlabelled

data in source domain be D_s^{uld} and that of destination domain be D_d^{uld} . These data can be defined as in equations (19) to (21).

$$D_s^{ld} = (x_s^i, y_s^i)_{i=1}^{N_s^{ld}} \quad (19)$$

$$D_s^{uld} = (x_s^i)_{i=N_s^{ld}+1}^{N_s} \quad (20)$$

$$D_d^{uld} = (x_d^j)_{j=1}^{N_d} \quad (21)$$

Where N_s^{ld} the number of is labelled data in source domain, N_s is the total number of data in source domain and N_d is the total number of data in destination domain.

7. Results and Discussion

Our work is evaluated in terms of Accuracy, Precision, Recall and F1_Score, and compared with Basic ML methods [Parvati and Vidyavathi, (2020)]

Table 2: Comparison of Accuracy

		D ↓ E	D ↓ K	D ↓ B	E ↓ K	E ↓ D	E ↓ B	K ↓ D	K ↓ E	K ↓ B	B ↓ E	B ↓ D	B ↓ K	Average
Accuracy	SVM-L	73	70	69	78	72	66	72	82	67	71	75	69	71
	Bag-SVM-L	74	73	73	79	70	66	69	83	66	69	75	72	72
	IPSPE_TMA	83	81	83	83	84	84	80	85	84	83	82	83	83

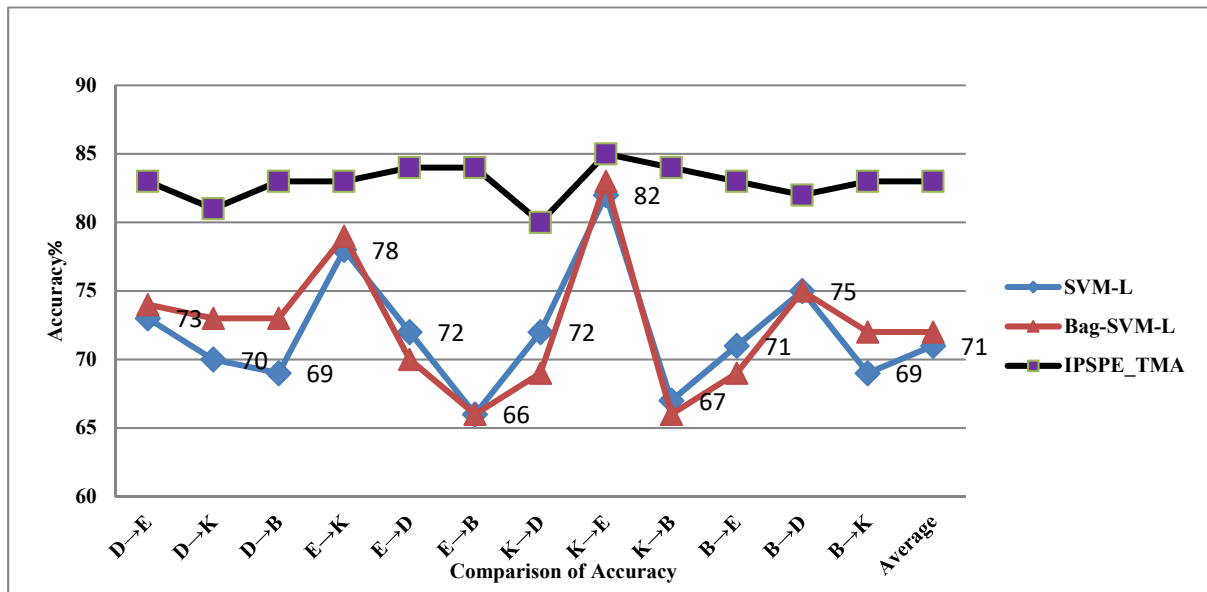


Fig.3: Comparison of accuracy between various methods [Parvati and Vidyavathi, (2020)]

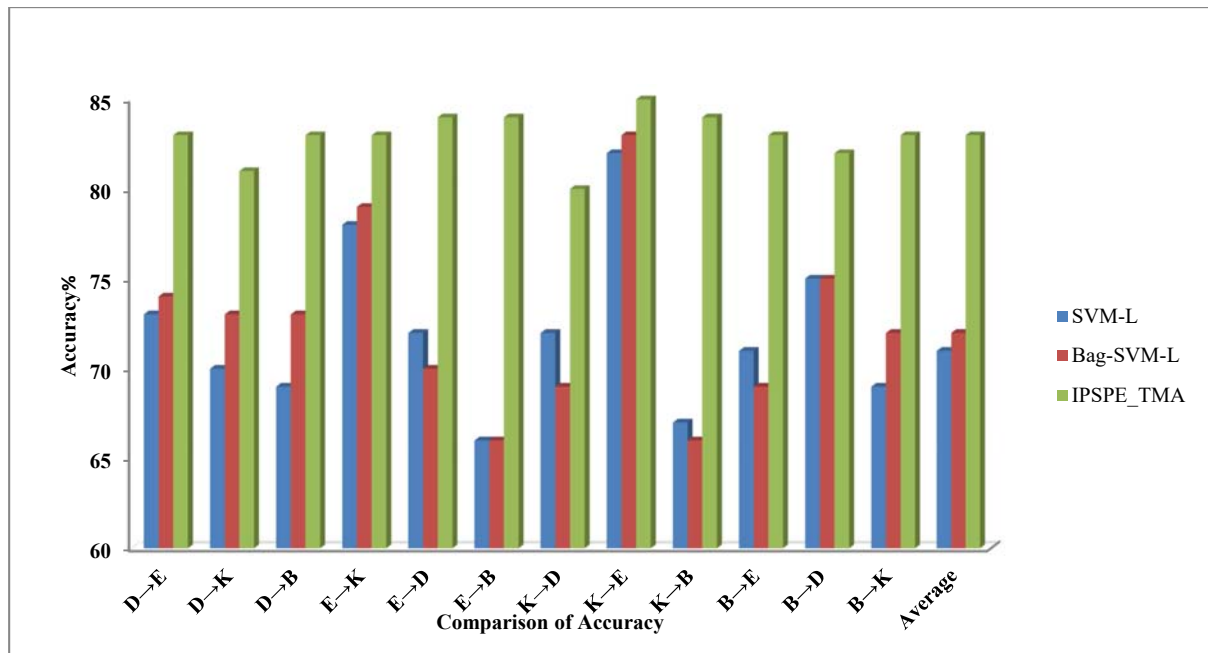


Fig.4. Comparison of accuracy between various methods [Parvati and Vidyavathi, (2020)]

The comparative result of the proposed work with the previous methods was displayed in Table 2. It is in terms of accuracy. It is clear that produce best results than any other previous works based on accuracy. Figure 4 shows the comparison of accuracy between various methods it gives 83% of accuracy which is +11 than SVM-L with Bagging, +12 than SVM-L basic ML methods.

Table 3: Comparison of Precision

		D ↓ E	D ↓ K	D ↓ B	E ↓ K	E ↓ D	E ↓ B	K ↓ D	K ↓ E	K ↓ B	B ↓ E	B ↓ D	B ↓ K	Average
Precision	SVM-L	62	66	66	84	80	64	66	76	78	68	81	68	72
	Bag-SVM-L	63	71	70	83	82	69	63	78	73	65	82	73	73
	IPSPE_TMA	83	74	81	83	84	84	78	85	84	83	81	83	82

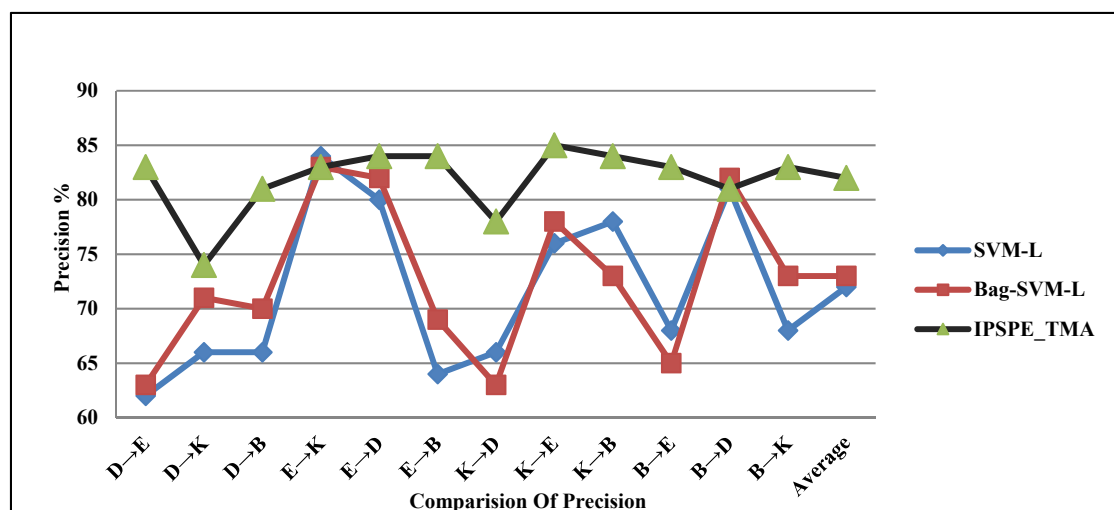


Fig.5. Comparison of precision between various methods [Parvati and Vidyavathi, (2020)]

The comparative result in terms of precision of the proposed work with the previous methods was displayed in Table 3. It is clear that produce best results than any other previous works based on precision. Figure

5 shows the comparison of precision between various methods it gives 82% of precision which is +9 than SVM-L with Bagging, +10 than SVM basic ML methods

Table 4: Comparison of F1_Score

		D ↓ E	D ↓ K	D ↓ B	E ↓ K	E ↓ D	E ↓ B	K ↓ D	K ↓ E	K ↓ B	B ↓ E	B ↓ D	B ↓ K	Average
F1 Score	SVM-L	70	69	68	79	74	65	70	81	70	70	76	68	72
	Bag-SVM-L	71	73	72	80	73	67	67	82	68	68	77	72	73
	IPSPE_TMA	83	77	82	82	83	83	78	85	83	82	81	83	82

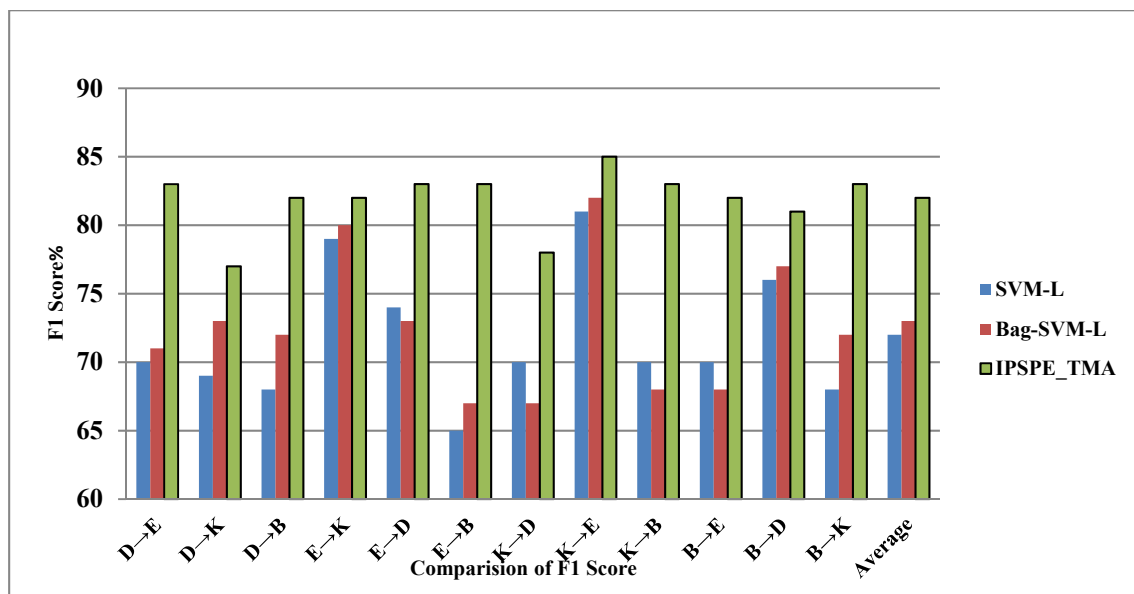


Fig.6. Comparison of F1_Score between various methods [Parvati and Vidyavathi, (2020)]

The comparative result in terms of f1_score of the proposed IPSPE_TMA with the previous methods was displayed in Table 4. It is captured that it produces best results than any other previous works based on f1score. Figure 6 shows the comparison of f1score between various methods it gives 82% of f1score which is +9 than SVM-L with Bagging and +10 than SVM-L BASIC ML methods.

Table 5: Comparison of Recall

		D ↓ E	D ↓ K	D ↓ B	E ↓ K	E ↓ D	E ↓ B	K ↓ D	K ↓ E	K ↓ B	B ↓ E	B ↓ D	B ↓ K	Average
Recall	SVM-L	79	72	70	75	69	66	75	87	64	73	73	69	73
	Bag-SVM-L	81	74	73	77	66	65	71	88	64	71	72	71	73
	IPSPE_TMA	83	81	83	83	84	84	79	85	84	83	82	83	83

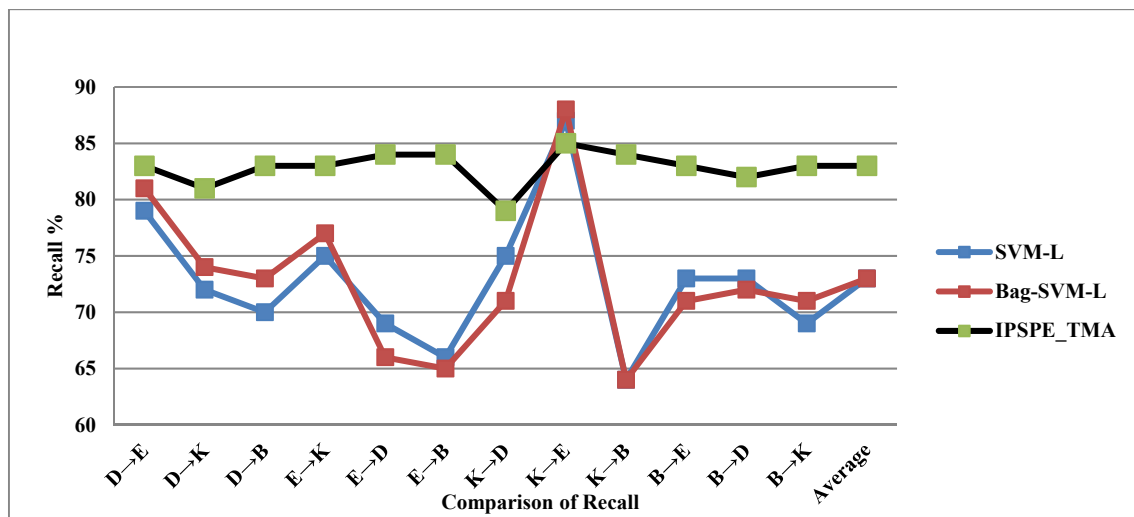


Fig.7. Comparison of Recall between various methods [Parvati and Vidyavathi, (2020)]

The comparative result in terms of recall of the proposed IPSPE_TMA with the previous methods was displayed in Table 5. It is clear that produce best results than any other previous works based on recall. Figure 7 shows the comparison of recall between various methods it gives 83% of recall which is +10 than SVM-L with Bagging, +10 than SVM BASIC ML methods.

8. Conclusion

In this work, a new embedding model is generated with the help of VADER to improve the performance of sentiment analysis especially in cross domain. A embedding model called Polarity Score based Binary Pattern is proposed to generate embedding coefficients which describe the sentiment intensity of each word effectively, which can be integrated with other embedding network model to improve its overall performance. The proposed Tri model network (IPSPE_TMA) utilizes the role of BiLSTM, BiGRU and CNN in sentiment analysis. Fusion of all network models with its attention scheme produces deep features which is robust to cross domain sentiment analysis. Accuracy and F1-Score has been improved from 10% when compared with the Machine Learning approach.

References

- [1] Pang, B.; Lee, L.; S. Vaithyanathan. (2002) Thumbs up? Sentiment Classification using Machine Learning Techniques. In Proceedings of EMNLP vol. 02, pp. 79-86.
- [2] Turney, P. (2002). Thumbs up or Thumbs down? Semantic Orientation Applied to Unsupervised Classification of reviews. In Proceedings of ACL (vol. 02, pp. 417-424).
- [3] Pang, B.; Lee, L. (2008). Opinion Mining and Sentiment Analysis: Foundations and Trends. Information Retrieval, (vol.2(12), pp.1-135.
- [4] Xing, F.; Pallucchini, F.; Cambria, E. (2019). Cognitive-inspired domain adaptation of sentiment lexicons. Inf. Process. Manag., vol. 56, pp. 554-564.
- [5] Bollegala, D.; D. Weir; J. Carroll. (2011). Using Multiple Sources to Construct a Sentiment Sensitive Thesaurus for Cross-Domain Sentiment Classification. In Proceedings of ACL-11, pp. 132-141.
- [6] Desai, Radhi (2018). Sentiment Analysis of Twitter Data. 114-117. 10.1109/ICCONS.2018.8662942
- [7] Desai, M.; Mehta, M.A. (April 2016). Techniques for sentiment analysis of twitter data: A comprehensive survey. In Proceedings of the International Conference on Computing, Communication and Automation, Greater Noida, India, pp. 29-30.
- [8] Meng, J.N.; Yu, Y.H.; Zhao, D.D.; Sun, S. (2015). Cross-domain sentiment analysis based on combination of feature and instance-transfer. J. Chin. Inf. Process. pp. 29, pp.74-79, 143.
- [9] Pan, S.J.; Yang, Q. (2010). A survey on transfer learning. IEEE Trans. Knowl. Data Eng., vol. 22, pp. 1345-1359.
- [10] Tahmoresnezhad, J.; Hashemi, S. (2017). Visual domain adaptation via transfer feature learning. Knowl. Inf. Syst. vol 50, pp. 585-605.
- [11] Wu, Q.; Tan, S. (2011). A two-stage framework for cross-domain sentiment classification. Expert Syst. Appl., (vol.38, pp. 14269-14275.
- [12] Pan, S.J.; Ni, X.; Sun, J.T.; Yang, Q.; Chen, Z. (2010). Cross-domain sentiment classification via spectral feature alignment. In Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, pp. 26-30.
- [13] Blitzer, J.; Dredze, M.; Pereira, F. (June 2007) Biographies, bollywood, boomboxes and blenders: Domain adaptation for sentiment classification. In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Prague, Czech Republic, pp. 25-27.
- [14] Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V. (2015) Domain-adversarial training of neural networks. J. Mach. Learn. Res. vol.17, pp. 2096-2030
- [15] Yulan, He.; Chenghua, Lin.; Harith, Alani. (2011). Automatically extracting polarity-bearing topics for cross-domain sentiment classification. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, (pp. 123-131). Association for Computational Linguistics.
- [16] Jianfei, Yu.; Jing, Jiang. (2016). Learning sentence embeddings with auxiliary tasks for cross-domain sentiment classification. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. (pp. 236-246) Association for Computational Linguistics.
- [17] He, Y.; Lin, C.; Alani, H. (2011). Automatically Extracting Polarity-Bearing Topics for Cross-Domain Sentiment Classification. In Proceeding of ACL-11, pp.123-131

- [18] Werner, Z.; Thomas G.; Edwin L.; Thomas, N.; Susanne, S. (2017). Central moment discrepancy (CMD) for domain-invariant representation learning. arxiv preprint arXiv:1702.08811.
- [19] Baochen, S.; Jiashi, F.; Kate, S. (2016). Return of frustratingly easy domain adaptation. arXiv:1511.05547 [cs.CV]
- [20] Yaroslav, G.; Evgeniya, U.; Hana, A.; Pascal, G.; Hugo, L.; Francois, L.; Mario M.; Victor, L. (2016). Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1), 2096–2030.
- [21] Zheng, L.; Yu Z.; Ying, W.; Yuxiang, W.; Qiang, Y. (2017). End-to-end adversarial memory network for cross-domain sentiment classification. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 2017)*
- [22] Kunpeng, Z.; Yu, C.; Yusheng, X.; Daniel, H.; Ankit, A.; Diana, P.; Kathy, L.; Wei-keng, L.; Alok, N. C. (2011). SES: sentiment elicitation system for social media data. In *ICDM, IEEE Computer Society*. (pp. 129–136).
- [23] Nal, K.; Edward, G.; Phil, B.; (2014). A convolutional neural network for modelling sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics Association for Computational Linguistics*. (Volume 1: Long Papers, pp. 655–665).
- [24] Yoon, K. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. (pp. 1746–1751). Association for Computational Linguist.
- [25] Zichao, Y.; Diyi, Y.; Chris, D.; Xiaodong, H.; Alex, S.; Eduard, H. (2016). Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. (pp. 1480–1489). Association for Computational Linguistics.
- [26] Peng, Z.; Zhenyu, Q.; Suncong, Z.; Jiaming, Xu; Hongyun, B.; Bo, X. (2016b). Text classification improved by integrating bidirectional lstm with two-dimensional max pooling. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, (pp. 3485–3495). The COLING 2016 Organizing Committee.
- [27] Richard, S.; Alex, P.; Jn, W.; Jason, C.; Christopher, D.; Manning, A N.; Christopher P. (2013). Recursive deep models for semantic compositionality over a sentiment tree bank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. (pp. 1631–1642). Association for Computational Linguistic
- [28] Stephan, G. GJ Van Rooyen, MIH Medialab, & YoshuaBengio. (2012). Learning structural correspondences across different linguistic domains with synchronous neural language models. In *Proc. of the xLite Workshop on Cross-Lingual Technologies, NIPS*.
- [29] Yftah, Z.; Roi, R. (2018). Pivot based language modeling for improved neural domain adaptation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, (Volume 1, Long Papers, pp. 1241–1251. Association for Computational Linguistics.
- [30] Yi, Y.; Jacob, E. (2014). Fast easy unsupervised domain adaptation with marginalized structured dropout. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers, pp. 538–544)*. Association for Computational Linguistics.
- [31] Minmin, C.; Zhixiang, X.; Kilian, W.; Fei, S. (2012). Marginalized denoising autoencoders for domain adaptation. arXiv preprint arXiv:1206.4683
- [32] Xavier, G.; Antoine, B.; Yoshua B. (2011). 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.
- [33] Olivier, C.; Bernhard, S.; Alexander, Z. (2009). Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. *IEEE Transactions on Neural Networks*, 20(3) pp.542–542.
- [34] Yucen, L.; Jun, Zhu.; Mengxi, L.; Yong, R; Bo Z. (2017). Smooth neighbors on teacher graphs for semi-supervised learning. arXiv preprint arXiv:1711.00258.
- [35] Long, M.; Wang, J.; Cao, Y.; Sun, J.; Philip, S.Y. (2016) Deep learning of transferable representation for scalable domain adaptation. volume 28, pp. 2027–2040. *IEEE Trans. Knowl. Data Eng.*,
- [36] Long, M.; Zhu, H.; Wang, J.; Jordan, M.I. (August 2017). Deep transfer learning with joint adaptation networks. (pp. 6–11). In *Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia*.
- [37] Sener, O.; Song, H.O.; Saxena, A.; Savarese, S. (2016). Learning transferrable representations for unsupervised domain adaptation. December 2016., pp.5–10. In *Proceedings of the 13th Conference on Neural Information Processing Systems, Barcelona, Spain*.
- [38] Davis, J.; Domingos, P. (2009): Deep transfer via second-order markov logic. In *Proceedings of the 26th International Conference on Machine Learning, Montreal, QC, Canada*. pp. 14–18 .
- [39] Dai, W.; Yang, Q.; Xue, Gu.; Yu, Y. (2007). Boosting for transfer learning. In *Proceedings of the International Conference on Machine Learning, Corvallis, OR, USA*, (pp. 20–24).
- [40] Tan, B.; Song, Y.; Zhong, E.; Yang, Q. (August 2015): Transitive transfer learning. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia*, (pp.10–13)
- [41] Tan, B.; Zhang, Y.; Pan, S.J.; Yang, Q. (2017): Distant domain transfer learning. In *Proceedings of the ThirtyFirst AAAI Conference on Artificial Intelligence, San Francisco, CA, USA*. (pp. 4–9)
- [42] Zhang, L.; Wang, S.; Liu, B. (2018): Deep learning for sentiment analysis: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, 8, e1253.
- [43] Lai, W.H.; Qiao, Y.P. (2018): Spam messages recognizing methods based on word embedding and convolutional neural network. *J. Comput. Appl.*, 38, 2469–2476.
- [44] Li, Z.; Zhang, Y.; Wei, Y.; Wu, Y.; Yang, Q. (August 2017): End-to-end adversarial memory network for cross-domain sentiment classification. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia*. (pp. 19–25)
- [45] Fortuna, L.; Arena, P.; Balya, D.; Zarandy, A. (2001): Cellular neural networks: A paradigm for nonlinear spatio-temporal processing. *IEEE Circuits Syst. Mag.* 1, pp. 6–21.
- [46] Preethi, G.; Venkata Krishna, P.; Mohammad S. O.; V. Saritha.; Sumanth Y. (2017): Application of Deep Learning to Sentiment Analysis for recommender system on cloud pp. 93–97 *International Conference on Computer, Information and Telecommunication Systems (ICITS)*
- [47] Yitao, C.; Xiaojun, W. (2019): Multi-Domain Sentiment Classification Based on Domain-Aware Embedding and Attention. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*.
- [48] Dragoni, M.; Petrucci, G. (2017): A neural word embeddings approach for multi-domain sentiment analysis *IEEE Trans. Affective Computing* 8 (4) pp. 457–470. <https://doi.org/10.1109/TAFFC.2017.2717879>
- [49] Collobert, R.; Weston, J. (2008): A unified architecture for natural language processing: deep neural networks with multitask learning, in: *Machine Learning*, In *Proceedings of the Twenty-Fifth International Conference (ICML 2008)*, Helsinki, Finland, June 5–9, 2008, pp. 160–167. doi:10.1145/1390156.1390177. <http://doi.acm.org/10.1145/1390156.1390177>
- [50] Collobert, R.; Weston, J.; Bottou, L.; Karlen, M.; Kavukcuoglu K.; Kuksa, P. P. (2011): Natural language processing (almost) from scratch, *Journal of Machine Learning Research* 12 pp. 2493–2537. URL <http://dl.acm.org/citation.cfm?id=2078186>
- [51] Bengio, Y.; Ducharme, R.; Vincent, P.; Jauvin, C. (2003): A neural probabilistic language model. *Journal of machine learning research* 3(Feb), 1137–1155.

- [52] Mikolov T.; Chen K.; Corrado G.; Dean, J.(2013): Efficient estimation of word representations in vector space, arXiv:1301.3781.
- [53] Dzmitry, B.; Kyunghyun, C.; Yoshua, B.(2014): Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
- [54] Minh-Thang, L.; Hieu P.; Christopher, D. M. (2015): Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025.
- [55] Ashish, V.; Noam, S.; Niki, P.; Jakob U.; Llion J.; Aidan N G.; Lukasz K.; Illia P.(2017): Attention is all you need. In Advances in Neural Information Processing Systems, pp. 5998–6008
- [56] Zichao, Y.; Diyi, Y.; Chris, D.; Xiaodong, H.; Alex, S.; Eduard, H. (2016): Hierarchical attention networks for document classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics. Human Language Technologies, pp 1480–1489
- [57] Zhigang, Y.; Sixing, W.; Fangzhao, W.; Junxin, L.; Yongfeng H. (2018): Domain attention model for multi-domain sentiment classification. Knowledge-Based Systems, 155:1–10.
- [58] Renjie, Z.; Chen; Xipeng, Q.(2018): Same representation, different attentions: Shareable sentence representation learning from multiple tasks. arXiv preprint arXiv:1804.08139.
- [59] Yitao, C.; Xiaojun, W.; “Multi-Domain Sentiment Classification Based on Domain-Aware Embedding and Attention,(2019):” In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19). pp. 4904 – 4911.
- [60] Zhang, K.; Zhang, H.; Liu, Q.; Zhao, H.; Zhu, H.; Chen, E. (2019): Interactive attention transfer network for cross-domain sentiment classification, In The 33rd AAAI Conference on Artificial Intelligence (AAAI-2019) pp. 5773–5780 Honolulu, Hawaii, USA.
- [61] Pennington, J.; Socher R.; Manning, C.(2014): Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) pp. 1532–1543.
- [62] Bojanowski, P.; Grave, E.; Joulin, A.; & Mikolov, T. (2017). Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics, 5, pp.135–146.
- [63] Peters, M. E.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; Zettlemoyer, L. (2018): Deep contextualized word representations. arXiv preprint arXiv:1802.05365.
- [64] Xing Y.; Changhui, L.; Xiaodong, F. (2021): Sentiment analysis based on BiGRU information enhancement. Journal of Physics: Conference Series. 1748. 032054. 10.1088/1742-6596/1748/3/032054.
- [65] Ian Goodfellow.; Bengio, Y.; Courville, A. (2017) Deep learning Nat Methods 13:35. doi: 10.1038/nmeth.3707
- [66] Bouvrie, J. (2006): Introduction Notes on Convolution Neural Networks. doi: <http://dx.doi.org/10.1016/j.protcy.2014.09.007>
- [67] LeCun, Y.; Bengio, Y.; Hinton, G.; (2015): Deep learning. Nature 521, PP. 436–444. doi: 10.1038/nature14539
- [68] Zhang, K.; Zhang, H.; Liu, Q.; Zhao, H.; Zhu, H.; Chen, E.; (2019): Interactive Attention Transfer Network for Cross-Domain Sentiment Classification. Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), (pp. 5773–5780).
- [69] Zichao, Y.; Diyi Y.; Chris, Dyer.; Xiaodong, H., Alex, S; Eduard, H. (2016): Hierarchical Attention Networks for Document Classification. Proceedings of NAACL-HLT 2016, pages 1480–1489, San Diego, California, June 12–17, 2016. Association for Computational Linguistics.
- [70] Sainbayar, S.; Arthur, S.; Jason, W.; Rob, F. (2015): End-to-end memory networks. arXiv preprint arXiv:1503.08895.
- [71] Ankit, K.; Oza n I.; Jonathan S.; James B.; Robert E.; Brian P.; Peter O.; Ishaan G., Richard, S. (2015): Ask me anything: Dynamic memory networks for natural language processing. arXiv preprint arXiv:1506.07285.
- [72] Sangeetha, K.; Prabha, D. (2021): Sentiment analysis of student feedback using multi-head attention fusion model of word and context embedding for LSTM. J Ambient Intell Human Computing, 12, pp. 4117–4126.
- [73] Han, Y.; Liu, M.; Jing, W. (2020): Aspect-Level Drug Reviews Sentiment Analysis Based on Double BiGRU and Knowledge Transfer, in IEEE Access, vol. 8, pp. 21314–21325
- [74] Parvati, K.; Vidyavathi, B M. (2020): Cross Domain Sentiment Analysis on E-Commerce Datasets using Machine Learning and Ensemble Learning Approaches. International Journal of Advanced Science and Technology, 29(6s), 93 - 107.
- [75] John, Blitzer; Mark, D.; Fernando, P.(2007)Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification. Association of Computational Linguistics (ACL)
- [76] Gilbert; Hutto E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In Eighth International Conference on Weblogs and Social Media (ICWSM-14).

Authors Profile



Mrs. Parvati Kadli has 19 years of teaching experience. She is pursuing Ph.D from Visvesvaraya University, Belagavi, Karnataka in Computer Science and Engineering since 2017 under the guidance of Dr.Vidyavathi B.M. She has completed her masters in Computer Engineering from SJCE, Mysore in 2002, Visvesvaraya University, Belagavi, Karnataka. Her area of interest include Data Mining, Database Management Systems and Computer Networks. She is currently working as Associate Professor in Department of Electronics and Communication Engineering, PDIT, Hosapete, Karnataka since 2002. She is Life member of Indian Society for Technical Education (LMISTE). She has published 5 papers in national, international journals.



Dr. B. M. Vidyavathi has 25 years of teaching Experience. She has been awarded Doctorate in computer science and Engineering from Visvesvaraya University, Belagavi, Karnataka in the year 2010. She was working as a professor in Department of Computer science and engineering, BITM, Ballari, karnataka since 2000 and currently heading the Department of Artificial Intelligence and Machine Learning. Her area of interest includes Data mining, Pattern recognition, Software Engineering, Artificial Intelligence. She is a senior member of Computer society of India, Life member of Indian Society for Technical Education (LMISTE). Member of International Association of Engineers (IAENG), senior member of International Association of Computer Science and Information Technology (IACSIT). She has published 60 papers in national, international conferences and journals.