A Hybrid Capsule Network with Attention and BiLSTM for Opinion Mining in Text

Tapas Guha*
Department of CSE, Presidency University, Bengaluru, India.
E-mail: tapasguha@presidencyuniversity.in

Dr.K.G. Mohan
Professor, Department of CSE, GITAM University, Bengaluru, India.

Received March 12, 2021; Accepted June 28, 2021
ISSN: 1735-188X
DOI: 10.14704/WEB/V18SI04/WEB18138

Abstract

Examining the product review or review on web service facilitate to increase the product quality or web service. By means, comments from online shopping sites such as Amazon, Flipkart, EBay etc will not only assist the users to purchase the product however besides be able to guide the producer/supplier to identify the advantages and disadvantages of the goods. Mining online shopping websites with their information will becomes a major important task. Web Mining plays a major important role to mine the details of online websites efficiently. Web mining is the process of data mining that learns without human intervention and mine information obtained from the documents in web and also services. Sentiment categorization and web mining has become a truly significant task recently, with profound business and research impact. Machine learning algorithms and soon after deep learning methods have been the market leaders in sentiment analysis. The advent of capsule networks has been a landmark event in deep learning. It has been truly proficient in image processing. In case of text classification, standalone capsule networks are not optimally suitable. Here, a hybrid BiLSTM-Capsule framework is introduced for sentiment analysis of web texts of reviews from various datasets. In the model beginning, there is a bidirectional LSTM layer after which is an attention layer and a final capsule layer. This review analysis will helps to improve the products from amazon, increase the movie quality. The analysis of outcome depending on MR, IMDB, SST and Amazon datasets indicated the introduced framework performs better than some benchmark deep learning models. Significantly, the BiLSTM-Capsule can put its words in sentimental trend showing the capsules’ attributes without utilizing the linguistic knowledge.
Keywords

Web Mining, Product Reviews, Amazon, Sentiment Analysis, Deep Learning, Capsule Networks, BiLSTM, Word2vec, Dynamic Routing.

Introduction

Online shopping websites will give more information to seller and buyer to identify the importance of their product. Once an individual decides to buy an item, her/his immediate step will be to seek the item in web. Web provides the individual with a variety of options based on their trademark, cost, model type, properties, shade, class, ranking, offer, etc. Everyday there are many items with new style, new type of model, new make, new company, new tools, new services, new promotion policies. All these make the consumer puzzled and mystified to decide on the option. While selecting from the option becomes really tough, feedback is obtained from the individuals who purchased and utilized these items. The consumers record their feedback on online purchase portal. Among the ten top online portals, Amazon place the first. This site provides the whole star rating of every item depending on the star rating obtained from the consumers. However based on the whole star rating an individual cannot decide on item’s quality. The producer of the item might need to enhance their item’s quality. By obtaining the item’s feedback the producer can enhance the item’s quality and also their marketing of item. Final result for all these issues is to examine the text reviews to identify the specified feature of the item that is lacking in consumer fulfillment. Web mining is the efficient solution for knowing the product reviews.

One main application of data mining is web mining where the contents in the web are mined. Based on the mined data type, web mining can be categorized into three types. Initially, mining the contents of web pages that contain videos, text and images are called as web content mining. Next, mining the data that are structured to obtain the information is called web structure mining. Lastly, mining the behavior of web users through their navigation and click streams are called as web usage mining. Furthermore, the process web mining contains four steps namely, retrieval of information, extraction of information, generalization, investigation. This work majorly focus on the web content mining to mine the contents of the text in the online websites. The products related reviews are mined using the web content mining, then sentiment study is performed to classify the content as positive or negative reviews.

The supervision of views, sentiment and the text that are subjective are referred as sentimental analysis (Yeole et al., 2015). It also provides knowledgeable information that
are based on public opinion by examining various reviews and tweets. This acts as a established tool to judge various important events like performance of films in box office, common elections, etc (Heredia et al., 2016). The sentiments include different valued features such as bi-grams, tri-grams based on combinations and polarities. Hence they are being reviewed for positive and negative aspects via various SVM, utilizing training methods. The label belonging is computed using neural network sin sentimental analysis. The uncertain dependencies between various nodes and edges on the acyclic graph are operated through Bayesian networks are utilized for context level data extraction. In social media, the accuracy of learning and data are obtained by the optimization of sentences and words. Tokenization of data is employed at the root level of a word to obtain positive and negative aspects. In order to acquire increased precision of social media data by reducing the sentimental analysis errors, many approaches are utilized (Singh et al., 2016).

One of the multidisciplinary field is sentimental analysis as it contains various fields like linguistics in computation, retrieval of information, semantics, NLP, ML and AI (Aydoğan & Akcayol 2016). Sentimental analysis classification can be performed at three extraction stages namely, aspect/element, text and sentence (Singh et al., 2016).

Initially, deep learning was introduced by G.E. Hinton (2006), a division of machine learning that is termed as Deep Neural Network (Day & Lee 2016). The neural network is based on the brain of a human that has many neurons which forms a network. DL networks can assist both categories of learning like unsupervised and supervised (Vateekul & Koomsubha 2016).

The DL contains various types of networks like Convolutional Neural Networks, Recurrent Neural Networks, Deep Belief Networks, Recursive Neural Networks and so on. The neural network can be used in generation of text, representation of vector, estimation of word representation, classification of sentences, modeling the sentence and presentation of features (Zhang et al., 2016).

The Deep learning is prominent in supervised and unsupervised methods of learning, various researchers employ deep learning for performing sentiment analysis. The DL has many efficient and prominent techniques that are employed for solving various issues (Ouyang et al., 2015). A well-known (Socher et al., 2011) example RNN is employed for rottentomatoes.com website to review the films.

In spite of many achievements of neural network frameworks, there exists some issues. Initially, the current framework concentrate on and depends on the instance representation
quality. The instance may be a complete document, any paragraph or a single sentence. Representing sentiment through vectors are limited because of difficult and delicate reviews. Increased capacity is provided to model sentiments through capsule structure of framework.

Next, linguistic knowledge like sentiment glossary, negative words (like nothing, never, not, no), and powered words (like strongly, extremely, very), are to be integrated to the framework to comprehend their terms for high accuracy of prediction. On the other hand, developing linguistic knowledge needs considerable efforts. Additionally, some domain definite datasets cannot utilize the provided sentiment lexicon.

The capsule network is used to overcome the above mentioned point and also provides a high instinctive way of grouping information in hierarchical level for resolving highly complicated issues. The network vector is replaced for single neuron node of traditional neural network for capsule network and dynamic routing method is employed for training process. The dynamic routing method enhance the vital features’ weight and also discover many hidden features that enhance the network’s performance.

Here, this research article introduced hybrid BiLSTM-Capsule framework that performs sentiment analysis of texts for various datasets. In the beginning of framework, a bidirectional LSTM layer is followed by attention and capsule layer. The experimental evaluation is performed for SST, Amazon, IMDB, MR databases that indicated the introduced method performs better than other existing ML and DL techniques.

The remaining paper is arranged in five main portions. In section 2, all related sentiment classification approaches are reviewed. Section 3 explains the components of introduced framework and its detailed architecture. In section 4, the datasets used, model hyperparameters and optimization techniques applied are discussed in depth. Section 5 explains the model performance. Finally, section 6 concludes this paper and discusses some possible enhancements.

**Related Works**

Most of early techniques of sentimental analysis rely on manually formed rules. By advancement in deep learning methods, neural network related techniques forms the majority. Based on this, most of the researchers applied linguistic knowledge to acquire good sentimental analysis performance.
In conventional sentiment classification, various types of feature representations like representation of bag-of-words, co-occurences of word, syntactic contexts (Graves 2013) have been employed. Authors in (Collobert et al., 2011) came up with a straight forward and efficient way of learning distributed word and phrases representations. Models that are based on Neural network showed a huge achievement for various NLP (natural language processing) problems.

RNNs, Recursive neural networks, LSTMs is explored by numerous researchers because of their capability to deal with sequential data. One prominent approach proposed a new hybrid architecture named RNN Encoder-Decoder by merging auto-encoders and recurrent neural nets which produced good performance in capturing all relevant representation in statistical language transformation (Cho et al., 2014).

In line with the already mentioned approaches, a plethora of scientists have applied deep learning techniques to analyze synthetic emotion recently. One prominent work trained labeled data to classify sentiment using transfer learning (You et al., 2025). A thorough survey of sentiment classification using deep neural networks is offered in (Dai & Le 2015).

A standard recurrent neural network which predicts depending only on the past word information falls short in some cases where tagging both the past and future word of a sentence is required (Hochreiter et al., 2001). Here comes the significance of bidirectional recurrent neural network and LSTM. One landmark approach proposed a hybrid deep learning model by combining CNN and bidirectional RNN for efficient character-level text categorization (Xiao & Cho, 2016).

In CNN, many vital information are lost in max-pooling as only the active neuron are mainly selected for moving to the following layer. Due to which important spatial information are missing among the layers. In order to overcome the problem with CNN, (Hinton et al., 2011) came up with the concept of "capsule" for the first time. CNNs are only translation equivariant. He proposed a method called “routing-by-agreement” which makes a capsule network viewpoint equivariant. A system is said to be viewpoint equivariant if at all levels, a change in input viewpoint results in a similar viewpoint variation in the output.

With reference to Hinton work, (Sabour et al., 2017) introduced a capsule network, that employed vector output capsule as a substitute for conventional convolution neural network scalar output, employing dynamic routing algorithm instead of pooling. Some
scholars (Katarya & Arora 2019) reviewed the research work based on capsule networks, indicating that there are less applications in capsule networks so far, with wide development hope.

Getting inspired by enhanced performance of capsule network in MNIST, authors (Zhao et al., 2018) utilized capsule network for text based problems and analyzed their performance with various dataset, the starting work performed by capsule network in actual modeling task. Soon after work (Wang et al., 2018) united capsule network with attention mechanism for sentiment analysis related to text and identified the routing calculation complexity. This framework not only concentrate more on features while training, it also efficiently regulate various features’ parameters.

Capsule network address all the issues mentioned and introduce a highly instinctive method of grouping information at the hierarchical level for solving highly complicated issues. In this network, single neuron of conventional network is replaced by neuron vector and dynamic routing method is employed for training process. The method dynamic routing enhance the important feature’s weight and identify many unknown features that enhance the network’s performance.

Proposed Architecture

This article introduced a capsule network framework integrated with BiLSTM named as BiLSTM-caps. The framework contains two sections, namely a BiLSTM module followed by an attention layer and a capsule combination module. In this work, initially the product reviews of the Amazon website is collected through web. The reviews of each product are collected via web content mining, and then a sentence characterized by a word embedding is provided as input for the framework. After extraction of initial features followed by iterations of dynamic routing, obtained output vector is utilized to identify the sentiment polarity. The producer/supplier of the is also benefited by the feedbacks. The supplier/producer can enhance their item by identifying the features that lack consumer fulfillment. This work analyze the text reviews of every product to create statistical details based on the feedback of dissatisfied customers of every feature.

BiLSTM-Capsule

The design of introduced BiLSTM-based capsule framework presented in Figure 2. Here, many sentiment classes defines quantity of capsules needed. Three low level sentiment categories: ‘positive’, ‘neutral’, ‘negative’ are modelled by three capsules. Basically, each capsule is attributed to one category.
The initial BiLSTM layer computes the input and feeds the same representation into the capsules. The BiLSTM encodes any text instance that are indicated as a dense vector by generating hidden vectors as output. These hidden vectors act as inputs to all the capsules.

The capsules having the maximum probability in that state shall be ‘enabled’ and the remaining ‘disabled’. While model training, one target is to increase the capsules’ probability along the lines of the real sentiment class and decreasing the probabilities of the remaining. The similarity between the reproduced capsule representation aligned with the actual class and the instance representation is maximized and the same is minimized for the other capsules. During testing a capsule having the highest probability instance is expected to be enabled and all the remaining capsules disabled. This capsule’s output is considered as the emotion class of that instance.

**BiLSTM**

Recursive neural network undergo vanishing and exploding gradient issues. Also RNN cannot capture long term sentiment variance. To resolve these issues, LSTM was designed to take care of long term temporal dependencies.

LSTM is built on RNN. It adds some gates in the cell circuitry to solve the gradient problems. LSTM has gates like Input, Forget, output gates and also has a memory line in contrast to standard RNN. The vector that is generated by the LSTM layer mainly taking into account the output from the preceding cell and the input in the current cell. In the end, sentiment is predicted depending on the output of last hidden layer.

The embedded sequence is transformed to a series of real-valued dense vectors in the introduced model which acts as input to the BiLSTM layer. Then two opposite series vectors are generated.

As discussed earlier, each LSTM cell contains $i_t$, input gate, $o_t$ output gate, $f_t$ forget gate, $c_t$ memory unit. Formula for calculation is shown in (1)-(6).

![Fig. 1 Standard BILSTM architecture](http://www.webology.org)
\[ i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (1) \]
\[ f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (2) \]
\[ o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (3) \]
\[ u_t = \tanh (W_u x_t + U_u h_{t-1} + b_u) \quad (4) \]
\[ c_t = i_t * u_t + f_t * c_{t-1} \quad (5) \]
\[ h_t = o_t * \tanh (c_t) \quad (6) \]

Here \( x_t \) denotes the input during time \( t \), \( \sigma \) the sigmoid activation function, \( \ast \) indicates elements’ corresponding multiplication. Also, \( c_t \) is employed to decrease the issue of gradient disappearance/explosion, subsequently LSTM discover longer information dependence. \( f_t \) forget gate act as reset memory unit, \( i_t \) and \( o_t \) indicate the input, output gates that are employed to manage memory unit’s input and output.

However LSTM cannot perform information encoding back and front. The bidirectional LSTM network is introduced to process the fetures in backward and forward for a given time. Both forward and backward LSTM are integrated with BiLSTM, that provides further context information with enhanced and quicker ability of learning. BiLSTM structure is depicted in Figure 1. Here, BiLSTM is utilized to study the capsule layer’s output for enhancing the network features’ fitting effect and generalization of new data set.

In BiLSTM, two hidden layers are trained over the same input sequence, one on the unchanged input series, and the second on the inverted duplicate of the same input series. The goal is to associate both the hidden layers of inverse directions to the same desired output.

In the introduced framework, embedded sequence is transformed to a series of real-valued dense vectors, then a one dimensional convolutional layer is employed to yield an undersized sequence of feature vector, as shown in (7).

\[ FV = (f v_1, f v_2, ..., f v_T) \quad (7) \]

This generated vector is then supplied to a BiLSTM layer, producing a pair of reverse sequence vectors as in (8) and (9).

\[ H_f = (\overline{h}_1, \overline{h}_2, ..., \overline{h}_T) \quad (8) \]
\[ H_r = (\overline{h}_1, \overline{h}_2, ..., \overline{h}_T) \quad (9) \]
Where: $H_f$ and $H_r$ are the forward and reverse vectors respectively. The final hidden states of both directions are concatenated to produce a fixed-dimensional vector as shown in (10).

$$h_{fd} = [\overline{h_r}, \overline{h_1}]$$  \hspace{1cm} (10)

This BiLSTM encoding acts as instance representation to all the capsules in the next layer. Each instance is represented from the average of the hidden vectors generated by the bi-LSTM.

Capsule Structure

The actual capsule layer’s input vectors appear from the mean of hidden vectors from the BiLSTM. Three functional sections constitute each capsule. Attention mechanism is used by the visual illustration section to create the representation $c_{v,i}$ of the capsule. Then an activation function like sigmoid is used by the Likelihood section in order to predict the probability $p_{bi}$ of that instance of the active capsule. Finally, Restoration section calculates the instance restoration representation by multiplying the capsule vector representation $c_{v,i}$ and the active state probability $p_{bi}$.

As discussed, the BiLSTM hidden vectors pass through the attention layer to build internal representation of a capsule. It judges the importance of a term relevant to the prediction task in hand. We use an attention mechanism with only one parameter in capsule:

$$e_{t,i} = h_tw_{a,i}$$  \hspace{1cm} (11)

$$\nu_{c,i} = \sum_{j=1}^{N_S} a_{t,j} h_t$$  \hspace{1cm} (12)

In (10) and (11), $h_t$ indicate word representation at position $t$ (RNN hidden vector) and $w_{a,i}$ is attention layer’s parameter of $ith$ capsule. Score of attention importance for every position, $a_{t,i}$ is acquired by getting product of representations and weight matrix. Then compression function is the normalizing the probability distribution on the words that squash the input vectors’ modulus.

Hidden vectors $H = [h_1, h_2, ..., h_{N_S}]$ is the input for capsule from the BiLSTM. The active state probability $p_{bi}$ is calculated after acquiring the capsule representation vector $c_{v,i}$. Based on the previously mentioned objectives the parameters are studied, i.e., maximizing the capsule’s state probability of that re chosen by ground truth sentiment, and minimizing
other capsule(s) state probability. During testing, the state of a capsule will be active when $pb_i$ is the highest among entries capsules.

The input instance’s reconstruction representation $rep_{s,i}$ is acquired by the product of capsule vector representation $c_{v,i}$ and the active state probability $pb_i$ as shown in (12).

$$rep_{s,i} = pb_i c_{v,i}$$  \hspace{1cm} (13)

All the modules go together with each other. The attribute and its capsule representation are matched, and the state of a capsule corresponds to the input instance. Hence, the probability module depending on capsule representation, is the highest if the sentiment of the capsule fit the input instance. Capsule representation and its state probability is used for developing the reconstruction module, hence it is able to support for the representation of input instance when it is ‘active’ state.

**Experimental Setup**

**Datasets**

To analyze the introduced methods, the experiments are conducted on three standard datasets namely, Movie Review (Pang & Lee 2005) (MR), Internet Movie Database (Pang & Lee 2007) (IMDB), Stanford Sentiment Treebank (Socher et al., 2013) (SST) and an Amazon dataset (https://www.cs.jhu.edu/~mdredze/datasets/sentiment). All the above mentioned datasets are employed for sentiment analysis. Outcome of the analysis are compared with the available algorithm results to attain the objective comparison effect.

MR: A set of English film reviews from Cornell University collected through www.rottentomatoes.com, that has 5331 optimistic reviews, 5331 pessimistic reviews. Average sentence length of 20.

IMDB: contains 50000 Internet Movie Reviews with severe polarization. It has 25000 training and 25000 testing data, that has 25000 optimistic and 25000 pessimistic reviews. average sentence length of 294.

SST: contains training, verification and testing set, is the first corpus with completely labeled parse trees, that is used for a comprehensive analysis of the compositional effects of sentiment in language (Socher et al., 2013). The corpus depends on the dataset initiated by Pang and Lee (Pang & Lee 2005). Also contains fine-grained sentiment labels for 215,154 phrases parsed by the Stanford parser. The sentiment label set of \{0,1,2,3,4\}, indicates ‘very negative’, ‘negative’, ‘neutral’, ‘positive’, and ‘very positive’,
correspondingly. As SST offers phrase-level annotations on the parse trees, few reported results are acquired depending on annotations at phrase-level. For the analysis, sentence-level annotations are only employed as our capsule model does not require expensive phrase-level annotation.

Amazon Multi-Domain Sentiment Dataset (version 2.0): The original Amazon review dataset (https://www.cs.jhu.edu/~mdredze/datasets/sentiment) has 340,000 reviews of 22 various types of items. This dataset is a collection of metadata of products and reviews on those products. Every review is tagged as positive or negative. Based on previous work, the reviews are obtained for four different domains, namely, Books (B), DVD (D), Electronics (E), and Kitchen (K), to analyze the capsule network’s performance for cross-domain sentiment classification. Every domain contains 1000 positive and 1000 negative reviews.

Model Hyper Parameters

In deep learning models, the neural network parameters and optimizations applied have considerable impact on the classifier’s performance. Additional time has been devoted in model design aspects like initial weight, number of hidden layers, activation and loss functions, optimizers used etc. and in tuning hyper parameters like number of filters in CNN, learning and dropout rate among others. For the experiment, Glove4 is used to initialize for entire word vectors. an unlabeled corpus is employed for pre-training the word embedding vectors, whose size is 840 billion.

The proposed framework is trained with incrementing batch sizes of 32, 64 and 128. A dropout layer with a parameter in the range of 0.3 to 0.5 and early stopping mechanism is introduced to counter overfitting. For repeated timings, best parameter values are obtained. Total neurons in BiLSTM layer is initialized to 128. To decrease the training loss, Adam optimizer is chosen with learning rate of 0.002. The initial number of epoch is initialized to 60.

In capsule layer, there are 32 capsules, each of size 16. For each instance representation, the dynamic routing algorithm runs for 3 iterations. For all the datasets, 0.5 is the dropout used in the capsule representation stage. The text length defines the weights’ length in the attention layer. Adam (Hochreiter et al., 2001) has been used as the optimizer. The learning rate for the embedded word vectors is $1e^{-4}$ while for other model parameters it is set to $1e^{-3}$. The Adam hyper parameters for initial decay rates, $\beta_1$ and $\beta_2$ are set to
0.899 and 0.999, respectively in order to speed up the learning phase. The capsule models are executed on TensorFlow 2.0 and the model parameters are set randomly.

**Evaluation Results and Analysis**

The proposed model has been implemented using TensorFlow 2.0 on Keras 2.3. Training the model on both CPU and GPU, it is observed that training time on CPU takes eight to ten times more than that on GPU.

Here Micro Averaged Precision, Micro Averaged Recall, Micro Averaged F1 scores are used as evaluation metrics for multi-label text classification. Scores are calculated for every class labels after which averaged for entire classes, indicated as label-based metrics. Moreover, Exact Match Ratio (ER) is also used as evaluation measure, that treats partially correct prediction as incorrect and simply considers fully correct predictions.

The introduced method caps-BiLSTM is compared with the methods like Recursive Auto Encoder (RAE) (Socher et al., 2011), BiLSTM (Cho et al., 2014), tree structured LSTM (Tree-LSTM) (Tai et al., 2015), LR- LSTM (Qian et al., 2016), DAN (Iyyer et al., 2015).
The sentiment analysis experiment is performed for above listed four datasets. Table 1 indicates the experimental results for entire datasets in comparison with these benchmarks.

The outcome of the experiment revealed that our introduced method considerably outperforms against other models for IMDB and SST. DL techniques are greater than machine learning methods represented by SVM. nevertheless, the outcome of the experiment for MRdataset cannot beat that of LR-LSTM, but still a considerable improvement over the other models. This is because BiLSTM layer optimally extracts the hidden text information, and the dynamic routing method in teh capsule network identify the capsules with superior similarity in the target value. therefore, BiLSTM can study the sequence’s context characteristics and perform classification consequently.

On the IMDB dataset, the proposed BiLSTM-Capsule model obtains highest accuracy of 88.3. between the baseline techniques, LR-Bi-LSTM performs better than other baselines in most of the datasets. However, it needs linguistic knowledge such as sentiment lexicon, intensity regularizer. Again it requires many human efforts for building linguistic knowledge. This framework doesn’t depend on linguistic context, it involves only simple linear operations above the hidden vectors acquired through the bidirectional LSTM.

**Table I Model Evaluation**

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy (%) in Datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recursive Auto Encoder (RAE) (Socher et al., 2011)</td>
<td>IMDB</td>
</tr>
<tr>
<td>BiLSTM (Cho et al., 2014)</td>
<td>79.80</td>
</tr>
<tr>
<td>Tree structured LSTM (Tree-LSTM) (Tai et al., 2015)</td>
<td>81.1</td>
</tr>
<tr>
<td>LR-LSTM (Qian et al., 2016)</td>
<td>84.6</td>
</tr>
<tr>
<td>DAN (Iyyer et al., 2015)</td>
<td>--</td>
</tr>
<tr>
<td>BiLSTM-Caps with attention</td>
<td>88.3</td>
</tr>
</tbody>
</table>

From the results, it is evident that this hybrid model achieves good result in deep sentiment representation for emotion categorization due to the ensemble of BiLSTM and CapsNet. Also this deep hybrid structure proves to be more efficient in dealing with long texts.

**Conclusion**

In this research paper, a hybrid BiLSTM based capsule network model is proposed for sentiment analysis, that enhances standalone capsule network’s performance. This research created a BiLSTM based capsule network model to show that reviews of text can be extracted through feedback for any item. Reviews of text from any Amazon can be
used in this system to increase the quality of the product. This results strongly proves that Capsule networks are not only here to replace CNNs and produce commendable improvement in feature extraction and classification of images, but they are truly efficient in multimodal sentiment classification. Capsules when preceded by bidirectional LSTM can be effectively utilized in sentiment analysis tasks. This work shows the efficiency of introduced model in sentiment classification on four landmark sentiment database. As future enhancements, other popular word embeddings like Wiki FastText embeddings may be tried to improve pre-training of review words as the input to the deep network. Other deep learning models like Gated Recurrent Units (GRU) may be merged with these Capsules may be targeted in the future.

Acknowledgements

The authors are appreciative to the Presidency University, Bengaluru for presenting all the foundations for the direct of this exploration. They want to express gratitude to the editors and all the anonymous critics for their valuable remarks and proposals in enhancing the nature of this paper.

References


https://www.cs.jhu.edu/~mdredze/datasets/sentiment

http://www.webology.org


