International Journal of Future Generation Communication and Networking
Vol. 13, No. 4, (2020), pp. 3520–3526

SENTIMENT CLASSIFICATION USING PS-POS EMBEDDING WITH BILSTM-CRF AND ATTENTION

V. ROSELINE[1], Dr. G. HEREN CHELLAM[2]

[1]Research Scholar, Register Number: 18221172162009, [2]Asst. Professor in Computer Science,
[1]Rani Anna Govt. College for Women, Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli – 627 012, Tamilnadu, India.
[2]Department of Computer Science, Rani Anna Govt. College for Women, Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli – 627 012, Tamilnadu, India.
[1]rose_vasee@yahoo.com, [2]hchellam@gmail.com

***Abstract-----*Sentiment analysis uses natural language processing and machine learning to interpret and classify emotions in subjective data. Sentiment analysis is often used in business to detect sentiment in social data, gauge brand reputation, and understand customers. In this paper a novel target extraction which is performed by a combination of two methods PS-POS (Predecessor Successor-Parts of Speech) and character based BiLSTM-CRF (Bidirectional Long Short Term Memory - Conditional Random Field). Initially, a PS-POS word embedding layer is built to integrate the features of the sentimental words into the word vector during the word vector training and then BiLSTM-CRF attention layer is dedicated to extract the targets in the text as single target, multi target and no target. Finally the 2D-CNN method was introduced to classify the targeted words. The word vector is used characterize the emotional words as input, and the output represents the different stress level of the information. According to this new model, a text is classified as positive, very positive, negative, very negative and neutral based on semantic similarity. The simulation result shows that the proposed method is more effective than the existing method in considering the relation between features in the text. Compared with state-of-the-art of network methods, the target classification performance of developed model is greatly enhanced.**

Keywords---- PSPOS, BiLSTM-CRF, 2D-CNN

1. INTRODUCTION

Text Sentiment analysis is a significant field in language handling. The nostalgic highlights of text are extricated by words or succession information, among which the most widely recognized opinion of the text, incorporates both positive and negative highlights. Stress classification arranges text depending on text attributes. Sentimental examination is the field of study that dissects individuals' conclusions, slants, evaluations, perspectives, and feelings toward substances and their qualities communicated in composed content. With the fast development of web-based media on the web, for example, re-sees, gathering conversations, web journals, news, and remarks, an ever increasing number of individuals share their perspectives and feelings on the web. Thus, this entrancing issue is progressively significant in business and society. One of the principle headings of stress examination is sentence-level feeling investigation. A great part of the current examination on this theme zeroed in on distinguishing the extremity of a sentence (for example positive, very positive, negative, negative and neutral) in light of the language hints extricated from the literary substance of hints extricated from the literary substance of sentences. The objective was to propose start to finish notion extraction technique that utilize general language text embeddings: BiLSTM with an extra discretionary CRF layer and with cutting edge neural system design CNN for classification of stress. The work is generally applied on Stanford Sentiment Treebank SST2.

ISSN: 2233 - 7857 IJFGCN

2. LITERATURE REVIEW

Investigating supposition of text is a difficult issue in Natural Language Process. There is a wide range task in extricating assessment structure text.There many work done in regards to this sentimental analysis processing [3]. Past work is centered around extremity arrangement [4, 5], feeling extraction [6] and assessment source task [7, 8]. These frameworks has illuminated grouping estimation at deferent degree of granularity, structure lexical level, expressed level, sentence level, just as record level [9, 10]. Distinctive application has various necessities, so examination of grouping feeling on different levels is significant. For instance, an inquiry noting framework may require assumption of sentence level; an arrangement of item audit would generally require extremity characterization at state level or sentence level; a framework which decides articles from on line news source would require record level. This wok centers around building a model on various level and examinations worldwide assessment and nearby conclusion of record at various granularity. There are a few sorts of models utilized in investigation opinion. Organized model has been utilized for notion examination.

Choi et al. [7, 8] use CRFs to become familiar with a worldwide arrangement model to order and appoint sources to suppositions. Mao and Lebanon [9] has estimated conclusion on sentence level and built assessment stream of creators in news surveys by utilizing a CRFs relapsing model. Fell models for estimation examination were concentrated by Pang and Lee. In their work an underlying model ordered each sentence as being abstract or target utilizing a worldwide min-cut induction calculation that considered neighborhood marking textures. The top abstract sentences are contribution to a standard report level extremity classifier with improved outcomes. The current work varies from that in Pang is that we present worldwide supposition and nearby feeling dependent on semantic examination and develop a numerous level model which decides assumption of record through neighborhood notion.

Zhang et al. [11] proposed a content supposition characterization calculation dependent on CNN and bidirectional LSTM combination, which accomplished the improvement of calculation execution by melding various highlights of various systems. Li et al. [12] proposed an assumption characterization calculation for Weibo text dependent on broadened highlight grid and twofold layer convolutional neural system so as to take care of the issues of information inadequacy in Chinese Weibo short content grouping. Li et al. [13] proposed a repetitive neural system language model dependent on long haul and transient memory, which can multi-order text. Ren et al. [14], utilizing LSTM to consolidate highlights with Twitter short content for extraction, the trial results have additionally been improved, and LSTM based content conclusion characterization is one of the most widely recognized techniques. The other works regarding this is the innovation of PS-POS embedding layer [2]. In light of the structure of convolutional neural system and repetitive neural system, this paper proposes an improved supposition order calculation dependent on BiLSTM-CRF[1]. In this calculation, the connection between the multi-dimensional highlights of the content is thought of, and the element extraction impact is superior to that of the customary characterization calculation, so this model can improve expectation results than the conventional model. The above calculations have great outcomes on text order issues, yet the above calculations overlook the connection between changed content highlights. In light of this reality, this paper proposes a BiLSTM-CRF-2D-CNN model.

3. SENTIMENT CLASSIFICATION USING PS-POS

This is the primary endeavor to assess target characterization model (Fig 1) utilizing PS-POS word embeddings stretched out with character embeddings. In the following segments a concise depiction of BiLSTM based arrangement labeling models with potential CRF layer is given. CNN has accomplished brilliant execution in PC vision handling, discourse acknowledgment, and regular language preparing [15] because of the capacity to extricate neighborhood relationships of spatial or transient structures.

ISSN: 2233 - 7857 IJFGCN

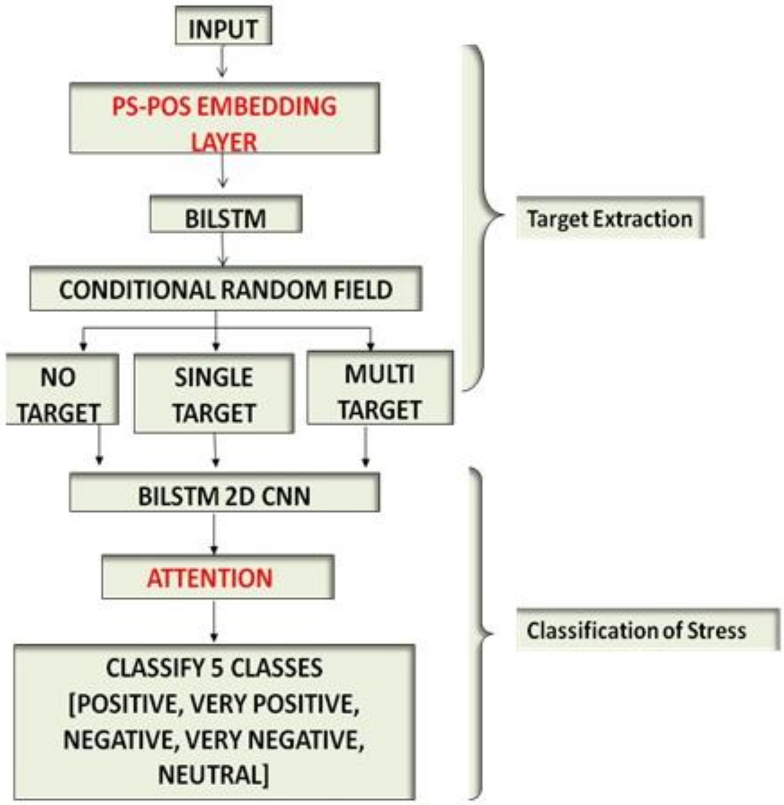


Fig. 1 Flow Chart of Proposed Methodology

3.1 PS-POS EMBEDDING LAYER

This layer utilizes likeness between words as an invariant for example they expect that words which happen in comparable settings are more looks like to share same significance. GloVe is a word depiction of Global Vectors in which the overall corpus bits of knowledge are gotten truly by the model for the coursed portrayal of the words. This model is a solo learning estimation for getting vector depiction of words. This is cultivated by planning words into a critical space where the detachment between words is related to semantic likeness. The POS provides a lot of data about a word and its neighbors, syntactic classes of words (noun, verb, adjective, adverb, and so on) and likenesses and dissimilarities among them. The produced POS tag is converted to a fixed vector and merged with PS-POS embedding vectors. Consequently, PS-POS embedding vectors will have synthetic data of words.

Mathematical Analysis:

The arbitrary event of a word being POS is expressed as

$$P(w_n = w_0) = \sum_{j=1}^{L(w)-1} \sigma\big[n(w, j+1) = ch\big(n(w, j)\big)\big] * Vec_n'(w, j)^T * h \quad (1)$$

Where $Vec_n'(w, j)$is vector form of word n(w, j) h is output of each word;

$$h = Vec_{w1}^T \quad (2)$$

ch(n) is the derivative of nth word

ISSN: 2233 - 7857 IJFGCN

The arbitrary event of moving towards left of each POS is

$$\sigma(\mathrm{Vec}_n^{\prime T} * h) \tag{3}$$

The arbitrary event of moving towards right of each POS is

$$1- \sigma(\mathrm{Vec}_n^{\prime T} * h) = \sigma(-\mathrm{Vec}_n^{\prime T} * h) \tag{4}$$

The arbitrary event of a word being output word after POS processing is expressed as

$$P(w_m = w_0)$$

$$= P(n(w_m, 1), \mathrm{left})\ P(n(w_m, 2), \mathrm{left}) *, \ldots\ldots, P(n(w_m, \max), \mathrm{right}) \tag{5}$$

$$=\sigma\left(V_n'(w_m, 1)^T * h\right) * \sigma\left(V_n'(w_m, 2)^T * h\right) *, \ldots. ,\sigma\left(V_n'(w_m, \max)^T * h\right) \tag{6}$$

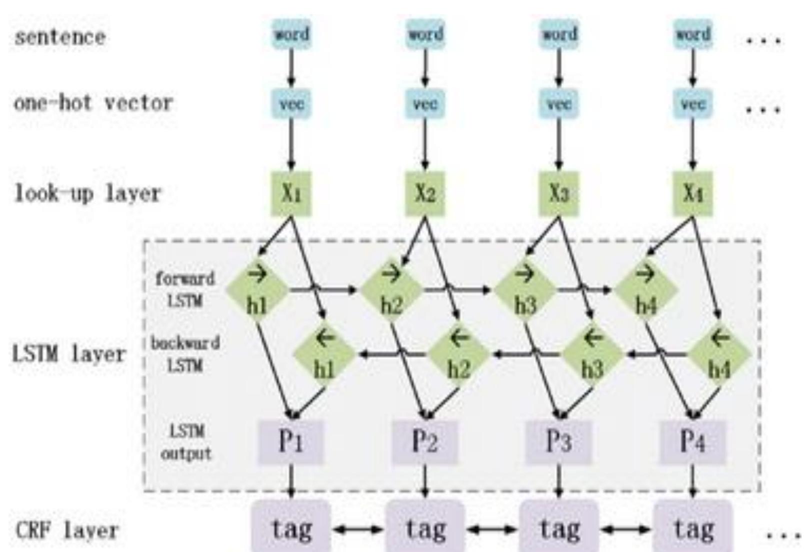
The final vector representation is constructed for the output word.

3.2 BI-LSTM-CRF LAYER

The target extraction and sentence type classification is described with BiLSTM-CRF.Target extraction is similar to the classic problem of named entity recognition (NER), which views a sentence as a sequence of tokens usually labeled with IOB format (short for Inside, Outside, Beginning).

BiLSTM uses two LSTMs to learn each token of the sequence based on both the past and the future context of the token. One LSTM processes the sequence from left to right, the other one from right to left. The input tokens X1,X2… are passed through the forward state and through the reverse state to the LSTM. At last, the two yields got from the forward and in reverse pass could be connected into one vector speaking to each protest in the arrangement. BiLSTM empowers to prepare neural system in a brisk manner; it interprets a portrayal of a word in the unique situation.

As another widely used sequence model, conditional random fields (CRF) is a type of discriminative undirected probabilistic graphical model, which represents a single log-linear distributions over structured outputs as a function of a particular observation input sequence. dropout technique is used after the input layer of BiLSTM-CRF to reduce overfitting on the training data. Fig 2 demonstrates the clear flow of the BiLSTM-CRF. After target extraction by BiLSTM-CRF, all opinionated sentences are classified into non-target sentences, one-target sentences and multi-target sentences, according to the number of targets extracted from them.



ISSN: 2233 - 7857 IJFGCN

Fig. 2 BiLSTM-CRF word embedding architecture

3.3 CNN CLASSIFIER

Sentiment characterization should be performed by supervised and unsupervised manner. In regulated learning, a preparation model will be made by investigating the highlights of preparing information. In view of this preparation model, the new information will be arranged. CNN is a calculation for arrangement in regulated learning technique. CNN is a sort of feed-forward ANN which propelled by the neurons and cerebrum. It is a variety of multilayer perceptron which needs just less pre-preparing procedures and it is simpler to prepare. Fig. 3 shows the three layers of CNN.

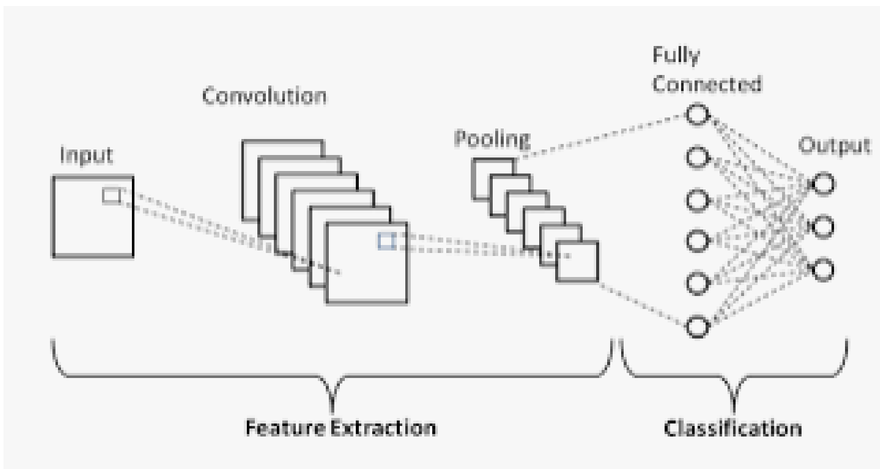


Fig. 3 CNN Architecture

Training Phase: In preparing stage, the contribution of the CNN calculation will be the pre-handled preparing information. As the info goes iteratively through each layer (convolutional, pooling, and completely associated), it will recognize the highlights in the pictures. The layers in the calculation will locate the best highlights which are required for characterization utilizing highlight maps. There must be at any rate six layers in the calculation to get a model. As the quantity of layers expanding, the preparation model will be increasingly proficient. Around nine convolutional layers, eight pooling layers and two completely connected layers are the mix of layers we might want to use in the current work. This mix will cause the calculation to work more precisely and effectively. We can undoubtedly arrange the picture with these quantities of layers.

3.4 ATTENTION LAYER

We add an attention layer for finding the contribution of each word to the whole sequence. The attention mechanism assigns a weight wi to each word feature hi with a focus on results. The hidden states are finally calculated to produce a hidden sentence feature vector r by a weighted sum function. Formally:

$$e_i = \tanh(W_h h_i + b_h),\ e_i \in [-1, 1] \tag{7}$$

$$w_i \frac{exp\ (e_i)}{\sum_{t=1}^{N} \exp\ (e_t)}, \sum_{i=1}^{N} w_i = 1 \tag{8}$$

$$r = \sum_{i=1}^{N} w_i h_i, r \in R^{2L} \tag{9}$$

ISSN: 2233 - 7857 IJFGCN

where W_h and b_h are the weight and bias from the attention layer.

4. RESULTS AND DISCUSSION

In this paper, a BiLSTM-CRF attention layer is combined with CNN model to classify the attributes of words in order to understand the feelings. To think about the presentation of this model on notion grouping errands, we utilize some conventional AI strategies for preparing in a similar informational index. What's more, to encourage the correlation of trial results, the attributes of the conventional model are additionally settled. The accompanying table I shows the preparation results for each model. The proposed method gives a better performance than the existing method 93.93% accuracy for SST2 (Fig 3) dataset.

Table II. Classification performance with SST2 dataset

S. No.	Method Used	Accuracy (%)
1.	BiLSTM-CRF + 1d-CNN	88.3
2.	BiLSTM + CNN	92.91
3.	BiLSTM-CRF + 2DCNN + Attention	93.93

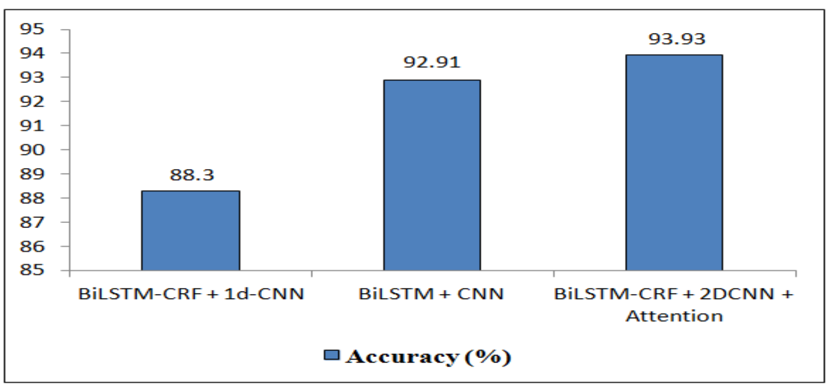


Fig 3: Chart for classification performance with SST2 dataset

5. CONCLUSION

In this paper, a convolutional neural network model is combined with BiLSTM and CRF attention layer to classify the emotions present in social media in order to characterize the sentiment analysis. PS-POS embedding layer with BiLSTM-CRF is the sentimental analysis tool to recognize the sentence of social media as single target, multi target and no target. Further the classification of the targeted words is handled with the 2D-CNN method with attention layer. The result gives an accuracy of 93.93% for SST2 dataset which is higher than the previous methods. From the result, it is affirmed that the proposed method shows improved performance than the existing methods. Still there are rooms for improvement, the data set is to be collected for certain age group say teenagers to help them recover out of stress and also the amount of

ISSN: 2233 - 7857 IJFGCN

training data is small which may cause model training be insufficient. The future research is to focus on this issue.

REFERENCES

[1] V. Roseline, Dr. Heren Chellam G. (2020). PS-POS Embedding Target Extraction Using CRF and BiLSTM. International Journal of Advanced Science and Technology, Vol. 29, no. 3, Mar. 2020, pp. 10984 -95.
[2] V. Roseline, Dr. Heren Chellam G. (2020). Performance Analysis of Target Dependent Sentiment Analysis using POS-R Embedding with LSTM and CRF. TEST Engineering and Management, Volume 83 Page Number: 11808 - 11816 Publication Issue: May-June 2020.
[3] V. Roseline, Dr. Heren Chellam G. (2019). ROLE OF SENTIMENT ANALYSIS USING DEEP LEARNING, International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.6, Issue 2, page no.116-122, February-2019.
[4] Turney (2002). "Thumbs up or thumbs down? Sentiment orientation applied to unsupervised classification of reviews". In EMNLP.
[5] Pang and L. Lee (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proc. ACL.
[6] Pang, L. Lee, and S. Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In EMNLP.
[7] Choi, C. Cardie, E. Riloff, and S. Patwardhan (2005). Identifying sources of opinions with conditional random fields and extraction patterns. In Proc. HLT/EMNLP.
[8] Choi, E. Breck, and C. Cardie (2006). Joint extraction of entities and relations for opinion recognition. In Proc. EMNLP.
[9] Mao and G. Lebanon (2006). Isotonic conditional random fields and local sentiment flow. In Proc. NIPS.
[10] Thomas, B. Pang, and L. Lee (2006). Get out the vote: Determining support or opposition from congressional floor debate transcripts. In Proc. EMNLP.
[11] C. Zhang and M. J. Zhou, "A text sentiment classification method based on the fusion of CNN and two-way LSTM," Computer Times, ed. Yunhe Pan, Zhejiang: Department of Science and Technology of Zhejiang Province, vol. 12, pp. 38-41, 2019.
[12] W. J. Li, J. Yi, "Weibo text sentiment classification based on extended feature matrix and double-layer convolution neural network,"Computer Applications and Software, ed. Sanyuan Zhu, Shanghai: Shanghai Computer Society, vol. 36, pp. 150-155, 2019.
[13] D.Wang and H. Du, "Chinese Film Comment Sentiment Classification Baesed on Deep Neural Network," Computer and Information Technology, ed. Bin Chen, Hunan: Hunan Institute of Electronics, vol. 27, pp. 19-20, 2019.
[14] Y. B. Jia, N. Li and Y. Ya, "Dynamic convolutional neural network Extreme Learning Machine for Text Sentiment Classification," Journal of Beijing University of Technology, ed. Tieyu Zuo, Beijing: Beijing University of Technology, vol. 43, pp. 29-35, 2017.
[15] Z. W. Chen, W. Feng, J. He, "Text sentiment classification based on 1D convolutional hybrid neural network," Journal of Computer Applications, JingZhong Zhang Ed. SiChuan: Sichuan Province Computer Federation, vo1. 7, pp.1937-1941, 2019.
[16] Q. Sun and Z. Y. Guo, "Vehicle following model based on LSTM neural network," Journal of Jilin University, Jie zhang, DongLiang Liu and QiaoLiang Liu Eds. JiLin: Jilin University, vol. 41, 2019.
[17] Y. Bengio, P. Simard, and P. Frasconi. Learning long-term dependencies with gradient descent is difficult. *Trans. Neur. Netw.*, 5(2):157–166, March 1994.
[18] Sepp Hochreiter and J¨urgen Schmidhuber. Long short-term memory *Neural computation*, 9(8):1735–1780, 1997.

ISSN: 2233 - 7857 IJFGCN