International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 11, Issue 3 [May.-June. 2022] PP: 376-381

Combined Deep Learning Procedure to handle Domain Level Sentiment Analysis in Social Media

T. VEERANNA[1], M. MADHAVI[2], G. ANUSHA[3], M. BHAVANA[4], P. CHANDRA KIRAN[5]

[1]Associate Professor,Dept. of CSE, Sai Spurthi Institute of Technology,Khammam,Telangana,India
[2,3,4,5] B.Tech Student, Dept. of CSE, Sai Spurthi Institute of Technology, Khammam,Telangana,India

Abstract

Interpersonal organizations have changed the correspondence designs altogether. Data accessible from various long range informal communication destinations can be very much used for the investigation of clients assessment. Subsequently, the associations would benefit through the improvement of a stage, which can break down open opinions in the web-based media about their items and administrations to offer a benefit expansion in their business cycle. Throughout the most recent couple of years, profound learning is exceptionally famous in the space of picture grouping, discourse acknowledgment, and so forth Notwithstanding, research on the utilization of profound learning technique in feeling investigation is restricted. It has been seen that at times the existing AI strategies for opinion investigation neglect to separate some understood perspectives and probably won't be exceptionally valuable. Subsequently, we propose a profound taking in approach for angle extraction from text and investigation of clients opinion comparing to the perspective. A seven layer profound convolutional neural organization (CNN) is utilized to label every viewpoint in the stubborn sentences. We have consolidated profound learning approach with a bunch of rulebased way to deal with work on the presentation of viewpoint extraction strategy just as opinion scoring technique. We have likewise attempted to further develop the current principle based methodology of viewpoint extraction by angle classification with a predefined set of viewpoint classes utilizing bunching strategy and contrasted our proposed technique and a portion of the best in class strategies. It has been seen that the general exactness of our proposed strategy is 0.87 while that of the other best in class strategies like changed standard based technique and CNN are 0.75 and 0.80 separately. The general exactness of our proposed strategy shows an addition of 7–12% from that of the state of-the-workmanship techniques.

Keywords: Opinion mining, CNN, Sentiment analysis, Text mining, POS tagging.

Date of Submission: 02-06-2022 Date of Acceptance: 15-06-2022

---

1. INTRODUCTION

These days, different interpersonal interaction locales like Twitter, Facebook, Linkedin, and so on have become a famous stage for trading feelings, and consequently giving criticism with respect to explicit items, association, administrations, motion pictures, people, political occasions, themes and so forth, which can assist with further developing the item quality and administrations of an association. Angle level assessment mining builds up a connection between various parts of a thing and its extremity. A part of an item implies a trait or element of an item. For opinion investigation, the distinguishing proof of perspectives is a vital issue. There are two kinds of perspective, unequivocal viewpoint, and understood angle. For instance, in the sentence,"The goal of the telephone is truly overall quite the telephone is reasonable", the "goal" is a part of the telephone and there is a positive assessment for the telephone. In the above model, "goal" is unequivocally referenced in the text, however "reasonable" is a verifiable angle. Unequivocal viewpoints are all around arranged while understood angles are not effectively characterized. So it is hard to remove certain angle. The majority of the past works for perspective level opinion examination have utilized SVM based calculation contingent irregular fields or some standard based methodologies utilizing regular language andling devices. Yet, these techniques have a few limits. CRF needs colossal number of elements to work appropriately and the standard put together strategy depends with respect to the syntactic exactness of sentences and it can just recognize express viewpoints however neglects to separate understood perspectives. Other than this, occasionally wrong perspectives have been labeled by utilizing grammatical features tagger as grammatical forms tagger considers all the thing or thing phrases as perspective terms. Be that as it may all things are not generally important viewpoint terms. Here CNN is utilized which have many less associations and boundaries. Each word in the sentences isn't named in this technique and we just need to name the entire sentences. So for huge information, we can without much of a stretch train the organization. ReLU is the most generally involved initiation work in CNN. ReLU is direct (personality) for every single positive worth, and zero for generally adverse qualities. It joins quicker and arranges information all the more without any problem. In this paper, we have attempted to defeat every one of the limits of existing strategies by utilizing CNN. A creative procedure is used to recognize the viewpoints by applying POS Tagging, reliance parsing utilizing CoreNLP followed by various leveled bunching, which diminishes the quantity of erroneous viewpoints and extricated perspectives are arranged with some predefined set of angles. Then, at that point, the worked on existing strategy is contrasted and the CNN based methodology of viewpoint extraction. Yet, it has been seen that CNN classifier at some point neglects to distinguish some substantial viewpoint terms. So a standard based methodology has been presented which is joined with CNN to additionally work on the exhibition of angle extraction technique. Moreover, we have likewise further developed the feeling scoring technique by presenting seven-point scaling. The vast majority of the past works utilized three degrees of opinion arrangement. In any case, now and then assessment word is related with some solid and frail modifier which adjust the opinion score. Here, the clients feeling has been arranged into seven gatherings (Almost Positive, Positive, Very Positive, Practically Negative, Negative, Very Negative, and Neutral). In this paper, item surveys from a well known interpersonal interaction site called Twitter, SST-1for the film survey and SemEval Task 4 for Restaurant Review has been gathered. Precise angle extraction and extremity discovery strategy help for suggestion frameworks, item quality, and administration improvement. It likewise permits the client to distinguish which highlights are significant and which not based on criticism. The remainder of the paper is coordinated as follows, in Section 2, we have introduced related work, and Section 3 portrays the foundation on opinion investigation. Segment 4 depicts the subtleties of the exploratory arrangement. Segment 5 contains results and examination and Section 6 briefs the finishing up piece of the paper.

2. RELATED WORK

A prologue to the field of feeling examination was found at Pang and Lee's article in 2008. Different strategies and techniques with both pragmatic and hypothetical contemplations ACI were covered by their article. These methods were utilized to investigate audits for motion pictures and items. In 2004, Kim and Hovy and all the more as of late, Bhayani and Huang (2009) performed opinion examination on Twitter furthermore Tweets were arranged as far as negative or positive feeling. Hu and Liu first presented the idea of perspective extraction from feelings and later this technique has been changed by Popescu and Etzioni and by Blair-Goldensohn et al. A language model was presented by Popescu and Etzioni. They accepted that item class is known in advance and acquainted a calculation with recognize a thing or thing phrase as an item include. They estimated point-wise common data between the thing expression and the item class. Scaffidi et al. utilized a language model to distinguish item includes. It was expected to be in their strategy that item audits contain a bigger number of quantities of item includes than in a general normal language text. However, it was observed that their strategy has low accuracy esteem and removed viewpoints are impacted by commotion. Perspective extraction technique for an item was

improved by presenting semi-administered models by Wang et al. They proposed a model to extricate viewpoints dependent on the utilization of cultivating perspectives. In this strategy, they utilized seed words to recognize subjects of explicit interest to a client and concentrate viewpoints from the audit. Later approach dependent on CNN has additionally accomplished critical improvement in execution over best in class techniques in numerous conventional NLP errands. It has been utilized in various NLP regions like data recovery and related characterization. A straightforward organization counting one-layer convolution and with a maximum pooling layer has been proposed by Kim et al. which performed feeling order effectively. Johnson et al. acquainted a pack of words model with address text rather than low-layered word vectors, which are incredibly successful for text order. Here more than 150,000 microblog postings were dissected containing marking remarks, feelings, and suppositions. They examined the in general design of these microblog postings, the kinds of articulations, and the development in sure or negative feeling. They looked at mechanized strategies for grouping feeling in these microblogs with manual coding. Collobert et al. utilized partof-discourse labeling, lumping, and named element acknowledgment assignments utilizing a perform multiple tasks grouping labeler. Xiaodong Liu et al. proposed a perform various tasks profound neural organization to group question furthermore search site positioning. Qiongxia Huang et al. in 2017 planned a model with existing CNN, LSTM, CNN-LSTM (Implement of one-layer LSTM straightforwardly stacked on one-layer CNN) what's more SVM.

Here, we have presented a combinational strategy for CNN and Rule-Based methodology which recognize viewpoint terms more precisely than the other best in class techniques.

3. SENTIMENT ANALYSIS METHODOLOGY BACKGROUND

3.1 Different Levels of Sentiment Analysis

There are three distinct degrees of feeling examination have been proposed.

3.1.1 Document Level: In Document Level opinion examination, it is dissected whether the record communicates a positive or negative feeling.

3.1.2 Sentence Level: In Sentence Level feeling investigation, the record is broken into a few sentences and each sentence is treated as a solitary element and investigated at a time.

3.1.3 Aspect Level: In Aspect Level, the primary errand is to remove perspective terms of the item and afterward client criticisms are examined based on the extricated angles.

3.2 Parts of Speech (POS) Tagging

Grammatical features (POS) labeling is a type of commenting on text and each word is a tag with Parts of Discourse. Tokens are set apart with their relating word by the POS Tagger. Grammatical feature labels are appointed to character strings. Each sentence can be classified into a gathering of determiners, action words, things, and so on.

3.3 Dependency Parsing

The linguistic construction of a sentence and the connections between "Fundamental" words and the word which alter those fundamental words can be acquired through a reliance parser. Here, we use reliance parser for angle extraction and seeing as their reliance connection with assessment words. For instance, "The telephone has a decent camera", here "camera" is a perspective and "great" is an assessment word. We can examine the design of the sentence "This telephone has a decent camera" in the accompanying way. Here amod: descriptive modifier, det: determiner, dobj: direct article and nsubj: ostensible subject. In the above example,"Camera" and "great" has amod connection. So from the Standford rule(rule 1) "camera" is a part of telephone and "great" is assessment word.

3.4 Cluster Analysis

Bunch examination is needed in text digging for making a gathering of items. It comprises of various strategies and calculations to bunch objects of comparable sorts into particular classes. Various leveled Clustering is utilized here. This strategy utilizes the dissimilarities (similitudes) or distances between objects while framing the bunches. Later POS labeling and reliance parsing, loads of viewpoints are gathered. To increment precision, viewpoints are sorted with the predefined set of angles utilizing various leveled grouping.

3.5 Convolutional Neural Network (CNN) for Text Classification

CNN is contained at least one convolutional layers [8,26,40], which are answerable for significant forward leaps in picture arrangement. All the more as of late, CNN is likewise applied to issues in Natural Language Processing (NLP) like data recovery and connection grouping, opinion examination [8,9,13], spam identification or subject classification. Sentences or archives that are the contribution of most NLP errands can be addressed as a network where each line addresses one token. A token might be a word or a person. The convolutional layers can be addressed as the weighted amount of the word vectors as for the common weight framework. The biggest worth is chosen in the maximum pooling layer. The conduct of the CNN is emphatically impacted by the size of the word, and all the word vectors from Google word2vec are standardized to one. The maximum pooling layer can be expanded or diminished by uniform increasing (Up or Down) of word vectors. Each CNN contains a word installing layer, a convolutional and pooling layer, and a completely associated layer.

3.5.1 Word Embedding. Word installing is a technique where words or expressions from the jargon are planned to vectors of genuine numbers. Every one of the words in the information sentence are encoded as word vector. Let the sentence length $l \in R$ and the jargon size $D \in R$. The implanting network of k-layered word vectors is $W^1 \in R^{k|D|}$ and the i^{th} word in a sentence is changed into a k-layered vector w_i by network vector item:

$$w_i = W^1 x_i \qquad (1)$$

Here x_i is the one-hot representation for the i-th word.

3.5.2 Convolution. The convolution operations are applied on the top of the vectors which are generated after encoding the input sentence to produce new features. In convolution operation, a filter u∈R^{hk} is applied to a window of h ¼ 2r+1 and a feature f_i is produced from a window of words w_{i-ri+r} by

$$f_i = g(w_{i-ri+r}.u) \qquad (2)$$

where g is a non-linear activation function (Relu) and the filter is applied to all possible windows of the input sentence to generate a feature map.

3.5.3 Pooling. In this layer, max-over the long run pooling is applied to every one of the component maps that are produced in convolutional layers. The most extreme component in each element map is taken by Max-after some time pooling and a fixedsized include vector vi ∈Rmi for the I-th task [30]. In this model, one element is removed from one channel. The model uses different channels (with changing window sizes) to get numerous highlights. In this layer, the element with the most noteworthy worth is separated for each channel. The thought is to catch the main element with the most noteworthy worth.

3.5.4 Dropout Regularization. Profound neural organizations experience the ill effects of overfitting due to the big number of boundaries that should be learned. So dropout regularization is added, which will haphazardly cripple a negligible portion of neurons in the layer (set to half here) to guarantee that the model doesn't overfit. This keeps neurons from co-adjusting and powers them to learn separately valuable elements.

4. RESULTS & ANALYSIS

Domain	Features	Precision	Recall	F-Score
Cellphone	WE	81.64%	72.15%	79.8%
Cellphone	WE + POS	85.24%	75.4%	82.54%
Camera	WE	72.9%	78.87%	77.30
Camera	WE + POS	76.29%	82.8%	80.5%
Laptop	WE	78.9%	83.23%	80.25%
Laptop	WE + POS	82.25%	85.45%	81.24%
Restaurant	WE	83.56%	86.8%	85.45%
Restaurant	WE + POS	85.67%	88.20%	89.34%

Table 1 Impact of the POS feature over word embedding.

Domain	Classifiers	% of Accurate Aspects	Precision	Recall	F-Score
Cellphone	CoreNLP + Rule-based	65.3%	72.4%	75.55%	74.86%
Cellphone	CNN	71.3%	75.68%	85.15%	80.56%
Cellphone	CNN + Rule-based	75%	79.24%	88.4%	82.34%
Camera	CoreNLP + Rule-based	59.8%	73.6%	79.57%	75.50%
Camera	CNN	68.7%	76.6%	88.87%	78.50%
Camera	CNN + Rule-based	72.4%	78.79%	89.9%	80.5%
Laptop	CoreNLP + Rule-based	64.8%	73.9%	81.53%	79.45%
Laptop	CNN	71.4%	76.9%	85.23%	82.35%
Laptop	CNN + Rule-based	77.4%	79.25%	88.45%	83.24%
Restaurant	CoreNLP + Rule-based	59.4%	74.46%	80.8%	79.55%
Restaurant	CNN	67.4%	77.56%	84.8%	81.45%
Restaurant	CNN + Rule-based	74.4%	79.67%	86.20%	83.34%
Movie Review	CoreNLP + Rule-based	63.7%	74.26%	78.8%	75.55%
Movie Review	CNN	69.4%	75.36%	79.8%	78.45%
Movie Review	CNN + Rule-based	75.6%	78.67%	82.20%	80.34%

Table 2. Comparison of (CoreNLP þ RuleBased), CNN and (CNN þ Rule-Based) method.

Item	Methods	Positive Score(%)			Negative Score(%)			Neutral Score(%)	Accuracy
		Almost Pos	Pos	Very Pos	Almost Neg	Neg	Very Neg		
Laptop (Dell)	CoreNlp + Rule-Based	42.10	1.05	.10	11.57	.21	.04	44.90	0.75
Laptop (Dell)	CNN	45.28	3.00	.50	4.25	.15	.3	45	0.80
Laptop (Dell)	CNN + Rule-Based	52	5.05	.68	4.57	.24	.5	65.60	0.87

Table 3. comparison of list of sentiment score and over all accuracy among existing methods and proposed method.

4.1 Significancy test using Paired T-test

Domain	Algorithm	Precision	Recall
Nikon Camera Data Set	Hu and Liu [39]	69.00%	82.00%
	Popescu and Etzioni [40]	86.00%	80.00%
	Dependency propagation [49] method	81.00%	84.00%
	Proposed Method(CNN + Rule-Based)	88.6	90.5

pı

Table 4. Comparison of proposed method with the state of the art method on Nikon Camera dataset.

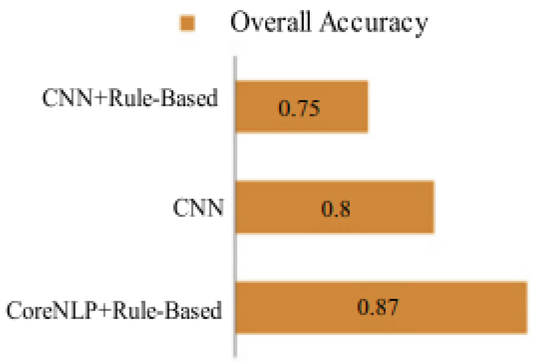

Figure 1. Comparison of the overall accuracy of CoreNLP þ Rulebased, CNN and CNN þ Rule-based approach on the dataset of laptop domain.

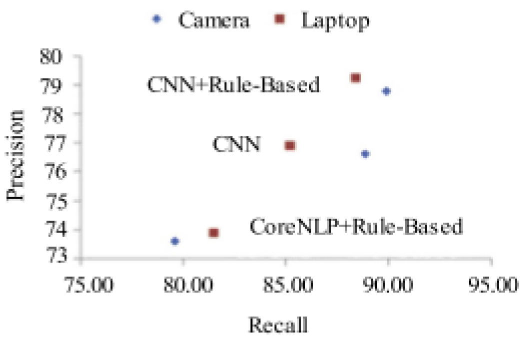

Figure 2. Comparison of the performance of CoreNLP þ Rulebased, CNN and CNN þ Rule-based approach on the dataset of camera and laptop.

	Over all Accuracy of Proposed Method	Over all Accuracy of Existing Method (CoreNLP + Rule-based)
Mean	0.87	0.78
Variance	0.0001	0.0007
Observations	3	3
Pearson Correlation	−0.188982237	
Hypothesized Mean Difference	0	
df	2	
t Stat	5.773502692	
P(T<=t) one-tail	0.014357069	
P(T<=t) two-tail	0.028714138	
t Critical two-tail	4.30265273	

Table 5. Paired T-Test for significant test.

5. CONCLUSION

In this paper, a blended methodology of profound learning technique and the standard based strategy has been presented for perspective level feeling examination by separating and estimating the viewpoint level opinions. From one viewpoint, we have utilized AI methods, POS labeling, reliance parsing, and so forth to recognize the perspectives and assessment of client identified with the viewpoint. On the other hand, a seven-layer explicit profound CNN engineering has been fostered that contains the info layer, comprising of word inserting highlights for each word in the sentence, two convolution layers, every one of them is trailed by max-pooling layer, completely associated layer, what's more the result layer. A standard based idea is likewise acquainted with work on the presentation of angle extraction. In correlation with the current strategies, the proposed strategy (CNN þ Rule-Based) brings better order precision for both the positive and negative classes.

6. REFERENCES

[1]. Samir Tartir, Ibrahim Abdul-Nabi, Semantic sentiment analysis in arabic social media, J. King Saud University – Comput. Inform. Sci. 29 (2) (April 2017) 229– 233.

[2]. H. Qiongxia, X. Zheng, Z. Dong, R. Chen, Deep Sentiment Representation Based on CNN and LSTM. 2017 International Conference on Green Informatics.

[3]. O. Kolchyna, T.P. Tharsis, P. Souza, T. Aste Treleaven, Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination, Department of Computer Science, UCL, Gower Street, London, UK, 2015.

[4]. B. Pang, L. Lee, Opinion mining and sentiment analysis, Found. Trends Inform. Retrieval 2 (2008) 1–135.

[5]. L. Xiaodong, G. Jianfeng, H. Xiaodong, Deng Li, D. Kevin, Ye Yi Wang, Representation learning using multi-task deep neural networks for semantic classification and information retrieval, Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2015, pp. 912–921.

[6]. A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, Sentiment analysis of twitter data, Proc. Assoc. Comput. Linguistics (2011) 30–38.

[7]. W. Medhat, Korashy H. HassaAn, Sentiment analysis algorithms and applications: a survey, Ain Shams Eng. J. 5 (4) (2014) 1093–1113.

[8]. B. Pang, L. Lee, A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts, in: D. Scott (Ed.), Proc. of the ACL. Morristown, 2004, pp. 271–278.

[9]. C. Scaffidi, K. Bierhoff, E. Chang, M. Felker, Red Opal: Product-feature scoring from review, in: Proceedings of the 8th ACM Conference on Electronic Commerce, ACM, 2007, pp. 182–191.

[10]. S. Dario, S. Gjorgji, M. Gjorgji, D. Ivica. Twitter Sentiment Analysis Using Deep Convolutional Neural Network published at: https://www.researchgate.net/publication/279208470June2015.

[11]. G. Qiu, B. Liu, J. Bu, C. Chen, Opinion word expansion and target extraction through double propagation, Comput. Linguist. 37 (1) (2011) 9–27.