

Plagiarism - Report

Originality Assessment

21%



Overall Similarity

Date: Mar 27, 2024

Matches: 528 / 2564 words

Sources: 16

Remarks: Moderate similarity detected, you better improve the document (if needed).

Verify Report:

Application of 2 Convolutional Neural Network in Sentiment Analysis

Hritika Kadam

MSc Computer Science

SVKM's Mithibai College (Autonomous)

Mumbai, India

hritika.kadam04@svkmmumbai.onmicrosoft.com

Jayasree Ravi

Assistant Professor

Dept. of Computer Science

Mumbai, India

Abstract

The current rate development in the Artificial Intelligence models have put Sentiment Analysis to a forefront. With many AI models being developed and integrated into applications of various fields, the requirement for the appropriate assessment of the emotions is necessary to understand and process the data for the end user. This paper focuses 2 on a Deep Learning approach, wherein we understand the applications of Convolutional Neural Network in Sentiment Analysis. Here we try to understand the implementation of CNN models paired with Word Embedding techniques to achieve accurate results. The paper discusses the different variations of the CNN models and their rate of accuracy on an extensive range of emotions. The paper also explains the experimentation of a proposed CNN model on the “GoEmotions” dataset.

Keywords: CNN Models, GoEmotions, Sentiment Analysis

I. Introduction

With the introduction of LLM like ChatGPT, BARD and Copilot, the interest in research in MLP has peaked highly. The end of goal such models is to understand the request of the user, to understand the tonality of it and provide an accurate response which is akin to a human-like response. This entire

process can be broken down into different categories, such as the coherence of the response, the sensitivity of the response to the user's tone, how natural the response sounds. There are various techniques that **1 are used to** develop such models, such as the LSTM (Long short-term memory), BERT (Bidirectional Encoder Representations from Transformers), RNN (Recurrent Neural Networks) which are some of the frameworks of NLP.

In this paper we discuss the implementation of CNN for the purpose **2 of Sentiment Analysis**. This implementation model is further divided into two parts-Word Embedding, which is required in data pre-processing, and the CNN model itself.

Convolutional Neural Network

CNNs are famously used for tasks such as image recognition and video recognition, medical image analysis and computer vision. The CNN models are designed and trained so that they are able to recognize high level spatial features. The architecture consists of three main components – Convolution Layers, to extract the features of the data and create feature maps for the characteristics, Pooling Layers, to maintain a low complexity while reducing the parameters and **4 a fully connected** network, that maps the extracted features in a final output.

CNNs **15 are also used** for text classification, but they have limitation when it comes to recognizing the semantic relations between texts. So, for the task of text classification, while the architecture remains similar an additional component is added to the model – Word Embeddings. Word Embeddings help represent the textual data in a continuous vector space, making it feasible for the model to recognize the features and map them.

1 Word Embeddings

Word embeddings are dense numerical representations of words in a continuous vector space, where each word is mapped to a high-dimensional vector of real numbers. These embeddings are designed to capture semantic relationships between words and contextual information within a given corpus of text. By representing words in a continuous space, embeddings enable algorithms to understand language more effectively by leveraging similarities between word vectors.

Various techniques exist for generating word embeddings, such as Word2Vec, GloVe, and

transformer-based models like BERT and GPT. Word2Vec learns embeddings by predicting surrounding words given a target word or vice versa, while GloVe combines global word co-occurrence statistics with matrix factorization methods. Transformer-based models like BERT and GPT generate contextualized word embeddings by considering bidirectional context or predicting the next word in a sequence.

Sentiment Analysis

Sentiment Analysis is a common opinion mining technique used for feedbacks and reviews, providing a range of three categories – Positive, Negative and Neutral. It extracts the subjective information from the text to determine the emotion. It's used across a varied range of domains to gain insights on the user's views. This paper attempts to understand 2 its applications in setting with wider range of emotions.

II. Related Research

The authors of this paper [1] propose a new model, Character to Sentence 3 Convolutional Neural Network (CharSCNN) aiming to resolve the issue revolving around the size of the text. The short texts whose length are at most one sentence long, such as posts on X (Twitter), do not provide enough context of the information. The study focuses on how to accurately perform sentiment analysis on such short texts using the SST dataset.

Xi Ouyang [2] works on the CNN + Word2Vec framework, the overall goal for his research seems to be enhancing the understanding of context with respect to the user content 4 on social media. The paper proposes a model that labels the sentences rather than the common approach of labelling the words.

Zufan Zhang [3] experimented with data of different modalities, implementing attention 1 mechanism to capture more nuanced understanding of the semantics. The authors compare three different frameworks for texts encoding- Semantic Embedding, Lexicon Embedding and Sentiment Embedding. Further integrating three different attention mechanisms in CNN – attention vector, LSTM attention and attentive pooling. The paper states that the combination of lexicon embeddings and the attention mechanisms yielded the better performance on the general

sentiment analysis out of all three.

The paper [4] introduces several approaches, integrating Lexicon Embedding with the Kim's CNN model [17], along with attention mechanism. While the study applied the attention model to each word resulting in a more robust understanding, it concludes that the applying the attention model to multiple words at a time is possible scope for the future.

Shiyang Liao [5] proposes a CNN model trained on X (Twitter) dataset to gain understanding of sentiments in real world situations. The authors also use datasets or Movie Reviews and STS Gold [15] to train the model.

The authors [6] through their study emphasize that the sizes of filter matrix as well as [3] the number of convolution layers have a significant impact on the accuracy of the model. They conclude that more convolution layers lead to decrease in accuracy, similar with [1] the size of the filter matrix, where in the matrix of larger size is not capable of capturing the details of dataset accurately.

Chen and Zhang [7] use a combination of CNN along with SVM. Implementing CNN for feature analysis and SVM [3] for text classification. The pertained word vectors are achieved using Word2Vec algorithm.

The authors [8] expand the scope dataset by utilizing the Hindi Movie Reviews. Along with implementing the CNN, they compare the results with other machine learning baseline classification models, such as, K-means, Naïve Bayes, SVM, deeming that properly trained CNNs perform better in sentiment analysis than the baseline models [4] of Machine Learning.

The paper [9] discusses the development and application of DGCN (Dialogue Graph Convolutional Neural Network) used for emotion recognition in conversations. The study aims to recognize several emotions such as, happy, sad, neutral, angry, excited, frustrated, disgust, and fear, present in the conversational data. The DGCN model has three components - [12] Sequential Context Encoder, Speaker-Level Context Encoder, and Emotion Classifier, paired along with Kim's CNN model for extracting features in combination with GloVe algorithm to represent the vectors of the data. The model performs better than [4] other baseline models on the ERC (Emotion Recognition in Conversation) datasets.

Hannah Kim [10] states research for the sentiment classification on longer texts using a CNN model.

In the study, the binary and ternary classifications are considered, where in the author claims that the network developed by them has a better accuracy than machine learning classification models such as Decision Tree, Random Forest, SVM and Naïve Bayes. This study further open up the scope for tasks such as gender classification where the proposed model might give similar results to this study.

The paper [11] has taken an multi modal approach to detect stress. It uses social media data acquired from X(Twitter) and Facebook. The methodology [4 of the paper](#) implements OCR – for image recognition, NLP, and CNN – for text content processing. The model steers into being advantageous for the health-care, helping the end user in stress detection in a non-invasive manner with quick results.

D. Christy Daniel and L. Shyamala [12] propose a method to counteract the limitation of the CNN models for sentiment analysis. As per the study, even the though CNN models extract high-level features, abstract nouns are difficult to identify. They propose a [2 hybrid Convolutional Neural Networks - Global Vectors - Complex Sentence Searching - ABstract Noun Searching \(CNN-GloVe-CSS-ABNS\)](#) model. As the emoting words tend to abstract nouns rather than adjectives or verbs, [the proposed model](#) utilizes [Complex Sentence Searching \(CSS\) algorithm](#) to extract complex sentences and [ABstract Noun Searching \(ABNS\) algorithm to identify the](#) abstract nouns [3 based on the](#) sentence structure.

The authors [13], have proposed an enhanced version of Alexandre Cunha's model[18]. Working with movie review data, they have seven-layer CNN network, which consists [1 of a word embedding](#) layer – [to represent the](#) words in vectors, two 1D-CNN layers, a global max-poling layer [to reduce the](#) dimension [of the data](#), along with a fully connected network.

Sameeksha Khare [14] focuses [2 on enhancing the accuracy of the sentiment analysis, using CNN and](#) ANN techniques. The study is performed on X(Twitter) data, wherein CNN is utilised to understand the graphical content in the user comments and ANN is utilised for the textual content.

The paper [15] focuses [3 on how to](#) characterize sentences in a comment using CNN models. The authors have used the X(Twitter) dataset, and have analysed [the performance of](#) the CNN models and their sentiment [accuracy of the](#) extracted tweets.

Dr. Bhaludra R Nadh Singh and his team [16] present a study that [2 focuses on the](#) significance of

the multimodal data in analysing the emotions of the user. They propose a framework that uses CNN models to analyse the sentiment across different modalities – text, images and videos, aiming 1 to enhance the accuracy in the context of social media analytics.

III. Methodology

GloVe

8 GloVe (global vectors) is an unsupervised learning algorithm. First introduced by Stanford, it works on generating word embeddings that maps a matrix based on aggregated global word-word co-occurrence from the corpus. This results in a linear substructure of the representation 3 of the words.

Linear Substructure of the words

The model is trained to learn the word embeddings that preserve the ratios of co-occurrence probabilities between words. It's defined by 1 the objective function:

$$J = \sum_{i,j=1}^V f(X_{ij})(w_i^T w_j + b_i + b_j - \log(X_{ij}))^2$$

As a result, the dense the vector representations produced for the vocabulary are capable of capturing the semantic and syntactic relationships between the words. For this paper we'll be utilizing the 300d data of the pre-trained model provided 3 in the model.

CNN Model

Kim's CNN model

Here, we have used a variation of the Kim's CNN model [17]. The Kim's CNN model is widely used for text classification, which is applied here to understand her to categorize the semantics 2 of the sentences. We use the D convolution layer along side dropout to prevent overfitting of the data. 3 The pooling layers utilize the global max pooling function to reduce the dimensionality of the mapped features. The final classification consists of a dense layer of 28 units and the sigmoid activation function.

GoEmotions

GoEmotions: A Dataset of Fine-Grained Emotions[19], is a corpus consisting of 58,000 comments

extracted from popular English language Reddit threads, with human annotations to over 27 emotions categories. ⁵ The emotion categories are: admiration, amusement, anger, annoyance, approval, caring, confusion, curiosity, desire, disappointment, disapproval, disgust, embarrassment, excitement, fear, gratitude, grief, joy, love, nervousness, optimism, pride, realization, relief, remorse, sadness, surprise.

Range of the emotions in the dataset [19]

Categorization of the sentiment [19]

Results

After the model trained for a humble 12 rounds of learning, it got really good at figuring out ³ patterns in the data, reaching an F1 score of almost 42%. Then, we did a little experiment with different cutoff points ⁴ and found that setting the threshold at 0.25 gave us the best overall score. With this threshold in place, we looked at how well the model did for each emotion separately. It turned out that some emotions, like amusement, gratitude, and love, were pretty easy for ³ the model to predict accurately. But others, like disappointment, realization, and relief, were a bit trickier. When we plotted these scores, we could see these differences clearly.

The figure gives the F1 score as per the emotions.

iv. Conclusion

The paper focused on the use of CNN model for the categorization of 27 emotions ¹¹ for sentiment analysis. The experiment provides a ⁴ F1 score of 41%, which is commendable when compared with the F1 score of BERT at 42%. While this experimentation doesn't provide the optimal results that we

are looking for, it gives a way to **1 the development of** this model further to enhance **its ability to** better recognize the sentiments. **The development of** the CNN **model can be** then used for the enhancement of the combination models of CNN and NLP algorithms such as LSTM and BERT.

References

- [1] Cicero **10 Nogueira dos Santos, Maira Gatti**, “Deep **Convolutional Neural Networks for Sentiment Analysis of Short** Texts”, COLING (2014).
- [2] Xi Ouyang, Pan Zhou, Cheng Hua Li, Lijun Liu, “Sentiment Analysis Using Convolutional Neural Network”, IEEE (2015).
- [3] Zufan Zhang, Yang Zou, Chenquan Gan, “Textual sentiment analysis via three different attention convo15lutional **3 neural networks and** cross-modality consistent regression”, Neurocomputing (2017).
- [4] Bonggun Shin, Timothy Lee, Jinho D. Choi, “Lexicon **13 Integrated CNN Models with Attention for Sentiment** Analysis”, arXiv (2017)
- [5] Shiyang Liao, Junbo Wang, Ruiyun Yu, Koichi Sato, Zixue Cheng, “CNN for situations understanding based on sentiment analysis of twitter data”, Elsevier (2017)
- [6] **11 Moch. Ari Nasichuddin**, Teguh Bharata Adji, Widyawan³, “Performance **Improvement Using CNN for Sentiment** Analysis”, IJITEE (2018)
- [7] Yuling Chen, Zhi Zhang, “Research on text sentiment analysis based on CNNs and SVM”, IEEE (2018)
- [8] Sujata Rani, **7 Parteek Kumar**, “Deep Learning Based Sentiment Analysis Using Convolution **Neural Network**”, Arabian Journal for **Science and Engineering** (2018)
- [9] Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya and Alexander Gelbukh, “DialogueGCN: **4 A Graph Convolutional Neural Network for Emotion** Recogre3w350nition in Conversation”, COLING (2019)
- [10] Hannah Kim, Young-Seob Jeong, “Sentiment Classification Using Convolutional Neural Networks”, MDPI (2019)
- [11] T Radhika, Ch Sandhya, Punna Mahesh, K Manohar Reddy, “PRINCIPLE **9 OF SENTIMENT**

ANALYSIS USING CNN AND ONLINE SOCIAL NETWORK DATA, A TECHNOLOGICAL
METHOD TO STRESS DETECTION”, IJCRT (2019)

- [12] D. Christy Daniel, L. Shyamala, “Hybrid CNN Classification **6** for Sentiment Analysis under Deep Learning”, IJITEE (2020)
- [13] Kifayat Ullah, Anwar Rashad , Muzammil Khan , Yazeed Ghadi , Hanan Aljuaid and Zubair Nawaz, “A Deep Neural Network-Based Approach **2** for Sentiment Analysis of Movie Reviews”, WILEY (2022)
- [14] Sameeksha Khare, “Accuracy Enhancement During Sentiment Analysis in Twitter Using CNN”, IEEE (2022)
- [15] Poornima A, Nataraj N , Nithya R , Nirmala D , Divya P, “SENTIMENT ANALYSIS OF TWEETS IN TWITTER USING CNN”, IEEE (2022)
- [16] **16** Dr. Bhaludra R Nadh Singh, Ms. SpurthiTippani, Ms. SravaniReddyaGaddam, Ms. SainniUdaya Sri, Ms. VuggeUshasree, and Ms. JampulaVandana, “Intelligent **2** Multimodal Sentiment Analysis Using a CNN-Based Framework”, IEEE (2024)
- [17] Yoon Kim, “Convolutional Neural Networks for Sentence Classification”, arXiv (2014)
- [18] A. A. L. Cunha, M. C. Costa, and M. A. C. Pacheco, “Sentiment **6** analysis of youtube video comments using deep neural networks,” International Conference on Artificial Intelligence and Soft Computing, China, 2019.
- [19] **14** Dana Alon and Jeongwoo Ko, Software Engineers, “GoEmotions: A Dataset for Fine-Grained Emotion Classification”, Google Research, 2021

Sources

1	https://www.ibm.com/topics/word-embeddings#:~:text=The primary goal of word embeddings is to,reflect the semantic similarities and relationships between words. INTERNET 5%
2	https://www.academia.edu/91429970/A_Hybrid_Deep_Learning_Architecture_for_Sentiment_Analysis INTERNET 4%
3	https://realpython.com/python-keras-text-classification/ INTERNET 3%
4	https://www.researchgate.net/publication/336999715_DialogueGCN_A_Graph_Convolutional_Neural_Network_for_Emotion_Recognition_in_Conversation INTERNET 2%
5	https://paperswithcode.com/dataset/goemotions INTERNET 1%
6	https://dl.acm.org/doi/10.1155/2022/5217491 INTERNET 1%
7	https://parteekbhatia.com/journal-publications/ INTERNET 1%
8	https://nlp.stanford.edu/projects/glove/ INTERNET 1%
9	https://ijcrt.org/papers/IJCRTV020073.pdf INTERNET 1%
10	https://dblp.org/rec/conf/coling/SantosG14 INTERNET 1%
11	https://consensus.app/papers/performance-improvement-using-sentiment-analysis-nasichuddin/fbc4fc730e3b5c8bb22b75c0c8d9ed61/ INTERNET <1%
12	https://ar5iv.labs.arxiv.org/html/1908.11540 INTERNET <1%
13	https://aclanthology.org/W17-5220.pdf INTERNET <1%
14	https://blog.research.google/2021/10/goemotions-dataset-for-fine-grained.html INTERNET <1%

15

<https://medium.com/toloka/choosing-the-best-architecture-for-your-text-classification-task-aee30ecc7870>
INTERNET
<1%

16

<https://ijsrcseit.com/paper/CSEIT1846106.pdf>
INTERNET
<1%

EXCLUDE CUSTOM MATCHES OFF

EXCLUDE QUOTES OFF

EXCLUDE BIBLIOGRAPHY OFF