

# A Review on Text Classification Based on CNN

<sup>1</sup>Sachin Sambhaji Patil, <sup>2</sup>Anthon Rodrigues, <sup>3</sup>Rahul Telangi, <sup>4</sup>Vishwajeet Chavan

<sup>1</sup>Asst Professor, <sup>2</sup>BE Students,

<sup>1-4</sup>Department of Computer Engineering, Zeal College of Engineering and Research, Pune, Maharashtra, India

## ABSTRACT

Text can be an incredibly rich source of information, but extracting information from it can be difficult and time-consuming due to its unstructured nature. However, thanks to advances in natural language processing and machine learning, both of which are under the broad umbrella of artificial intelligence, it is getting easier and easier to organize textual data. It works by automating and structuring documents quickly and cost-effectively, so businesses can automate processes and uncover insights that help make good decisions. Instead of relying on manually generated rules, text classification using machine learning learns to perform classification based on past observations. By using pre-tagged examples as training data, machine learning algorithms can learn different associations between text fragments.

Keywords: - NLP , CNN , DL , MLP

## Article Info

Volume 9, Issue 6

Page Number : 622-624

## Publication Issue

November-December-2022

## Article History

Accepted : 05 Nov 2022

Published : 30 Dec 2022

## I. INTRODUCTION

Text classification is a common NLP task that assigns labels or classes to text. There are many practical applications of text classification widely used in production by some of the largest companies today. One of the most common forms of text classification is sentiment analysis, which assigns labels such as positive, negative, or neutral to a text string. Text classification is a machine learning technique that assigns a predefined set of categories to open text. Text classifiers can be used to organize, structure, and categorize nearly any type of text - from documents, medical studies, and files, and across entire websites. For example, new articles can be organized by topic; Support tickets can be organized by urgency; Conversations can be organized by language, brand

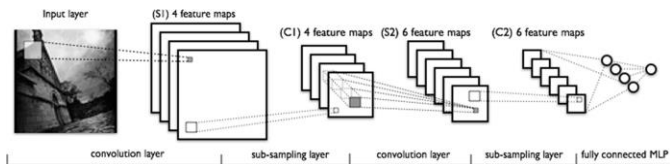
mentions can be organized by emotion, and so on. Text classification is one of the basic tasks of natural language processing.

## II. CONVOLUTIONAL NEURAL NETWORK

They were originally developed in the neural network image processing community, where they achieved breakthrough results in recognizing an object of a predefined category (e.g. cat, bicycle, etc)

A convolutional neural network usually consists of two operations, which can be thought of as feature extractors: convolution and aggregation.

The output of this sequence of operations is then usually connected to a fully connected layer, which is in principle the same as a traditional multi-layer perceptual neural network (MLP).



### III. RELATED WORK

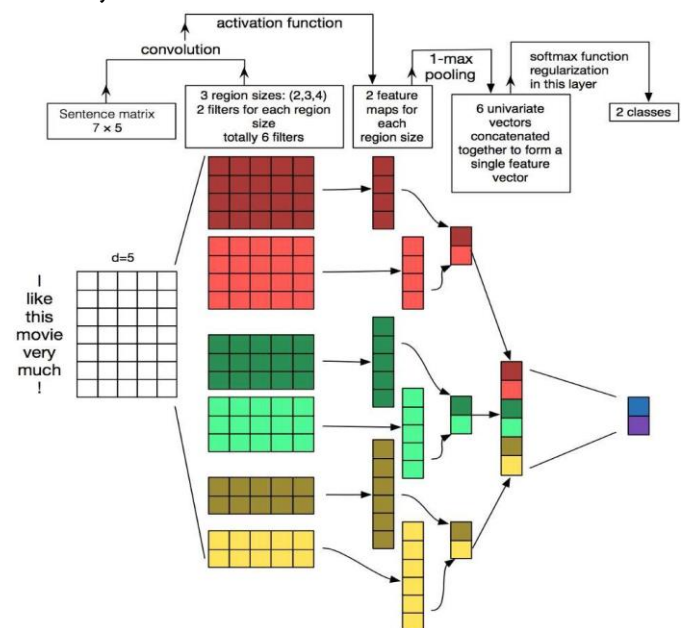
Deep learning is one of the latest trends in machine learning and artificial intelligence research, and many important breakthroughs have been made in this field around the world [4]. Similarly, in the field of natural language processing, the use of deep learning has produced an effect that easily surpasses traditional methods. Reference [5] briefly introduces the architecture and methods of deep learning and its application in natural language processing, and then discusses current state of the art in detail and gives some suggestions for future research in this area. Convolutional neural networks are widely used in the image domain. [Kim, 2014] first proposed the application of convolutional neural network in text classification task. He used a simple single-layer convolutional neural network in many classifier datasets while achieving excellent classification results and performing detailed parameter tuning [6]. It also prompts us to use deep learning methods in certain tasks without the need for complex network structures.

### IV. SYSTEM ARCHITECTURE

1. Input
2. Word Embedding
3. Concatenate
4. Flatten
5. Dropout
6. Output

We will be using TensorFlow for our backend. When using Naive Bayes and KNN, we usually represent our text as a vector and run algorithms on that vector, but we need to consider the similarity of words across

different journals, because it will help us to look at the review as a whole and instead of focusing about the effect of each word. We use built-in predefined words available in the library. Generally, if data is not integrated, there are many open source integrations such as Glove and Word2Vec. When we do the dot product of the text representing the vectors, they can result in zero even if they belong to the same class, but if you do the dot product of these embedded word vectors to find the similarity between them, you can find the mutual relationship of words for a particular class. We then drag the filter/kernel through these embeddings to find convolutional textures and these are further reduced in size to reduce the complexity and computation of the Max Pooling layer. In the end we have the layers fully connected and the function activates on the outputs which will give the values for each layer.



#### Word Embedding

Word embedding represents the density of word vectors, unlike what we did with the vector counting tool. This is another way to preprocess data. This integration can map semantically similar words. It does not treat the text as a human language but maps the structure of the word sets used in the corpus. They aim to map words in a geometric space known as the integration space.

## V. APPLICATIONS

Mark up content or products with categories to improve navigation or identify relevant content on your site. A faster emergency response system can be created by categorizing panic conversations on social media. Marketers can track and rank users based on how they talk about a product or brand online. Scholars, law practitioners, social scientists, governments, and non-profit organizations can also use text classification technology.

## VI. CONCLUSION

It is estimated that about 80% of all information is unstructured, with text being one of the most common types of unstructured data.

A large collection of documents can provide useful information for everyone. But it is also a challenge to find information from a large collection of documents. The impact of specific adjustments is also not well studied outside of the original domain of application. For these reasons, the design of text classification systems is more of an art than an exact science.

## VII. REFERENCES

- [1]. "Convolutional Neural Network Based Intelligent Handwritten Document Recognition", Sagheer Abbas , Yousef Alhwaiti
- [2]. "Impact of convolutional neural network and FastText embedding on text classification" , Carlo Medaglia,Gyu Sang Choi
- [3]. "The accuracy comparison among word2vec, glove, and fast text towards convolution neural network (cnn) text classification", Eddy muntina dharma , ford lumban gaol , harco leslie hendric spits warnars , benfano soewito
- [4]. "Textconvonet: A convolutional neural network based architecture for text classification", Sanskar soni, satyendra singh chouhan, santosh singh rathore
- [5]. "Medical Text Classification Based on Convolutional Neural Network: A Review", Hazha Saeed Yahia, Adnan Mohsin Abdulazeez
- [6]. "Deep Learning Based Text Classification: A Comprehensive Review", Shervin Minaee, Nal Kalchbrenner, Erik Cambria, Nanyang Narjes Nikzad, Meysam Chenaghlu, Jianfeng Gao
- [7]. "A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance"
- [8]. "Fusion of heterogeneous attention mechanisms in multi-view convolutional neural network for text classification"
- [9]. "Deep Learning Techniques on Text Classification Using Natural Language Processing (NLP) In Social Healthcare Network"
- [10]. "Text Classification Based on Convolutional Neural Networks and Word Embedding for Low-Resource Languages"

### Cite this article as :

Sachin Sambhaji Patil, Anthon Rodrigues, Rahul Telangi, Vishwajeet Chavan, "A Review on Text Classification Based on CNN", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 9 Issue 6, pp. 622-624, November-December 2022. Available at doi : <https://doi.org/10.32628/IJSRST229677>  
Journal URL : <https://ijsrst.com/IJSRST229677>