Proceedings of the 2nd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/73/20240399

Sentiment classification for movie reviews based on machine learning

Enzheng Chen

Wuhan University, No.299 Bayi Road, Luojiashan Street, Wuchang District, Wuhan City, Hubei Province

victorchenez@foxmail.com

Abstract. The research of this article aims to achieve the classification of positive and negative emotions in film review texts. The dataset used is the IMDB dataset from the Kaggle website, which contains approximately 50,000 movie reviews that are classified as positive or negative emotions. This is a dataset for binary-sensitive classification containing substantively more data than previous benchmark datasets. This research will use two different models, SVM and CNN, to classify the dataset and then compare and analyze them. After completing the experiment, the SVM model achieved an accuracy of nearly 90%, while the CNN model only had an accuracy of about 82%. This research found that for binary classification problems, as long as the input data is processed well, SVM can also achieve results that are no less than or even better than CNN. SVM is simple and efficient but difficult to implement for large-scale training samples; CNN is suitable for solving problems with complex internal mechanisms, but its learning speed is relatively slow. Finally, we proposed our prospects for future research.

Keywords: sentiment, machine learning, SVM, CNN.

1. Introduction

Emotion is a psychological activity mediated by individual desires and needs, which is described as sudden reactions to important internal or external events. Positive emotions can reduce stress and improve physical health; Positive emotions can enhance motivation and self-confidence as well as creativity and productivity; Positive emotions can promote interpersonal relationships. On the contrary, negative emotions can lead to psychological health problems such as depression, anxiety, and emotional disorders, as well as affect physical health, including decreased resistance and increased risk of chronic diseases. Negative emotions may lead to individuals feeling hopeless and helpless in their lives, thereby affecting their quality of life and happiness. Therefore, the recognition of emotions is a significant topic. However, due to the extremely complex physiological and psychological representations of emotions, accurate and rapid recognition of human emotions is only limited to theory, and its application in practice is still a long-term goal in the field of natural language.

In recent years, emotional analysis of online texts has become a research hot spot in psychology. The research results of psychology on emotion have essential reference value for mining and analyzing network information. More and more scholars are realizing this, adding some psychological elements to traditional emotional analysis tools and constructing multiple new research tools based on the emotional structure theory of psychology, injecting psychological ideas into the emotional analysis of online texts.

© 2024 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/73/20240399

The emotional classification of online text may predict stock prices and carry out crisis management. Classifying the emotions of user-generated content can also be used for personalized recommendations, user modeling, customized advertising, and more. Therefore, emotion classification has become a research focus in data mining and natural language processing.

This paper will investigate the effectiveness of SVM and CNN classification methods for sentiment classification on IMDB datasets.

2. Literature Review

2.1. Emotions of natural language

Emotion is a complex concept. There is no specific definition until now because psychologists have yet to reach a consensus. However, a widely accepted statement is that emotional structure can be divided into mainly two approaches: categorical approach and dimensional approach[1][2].

2.1.1. Categorical approach

This theory focuses on various aspects of emotions, such as the physiological mechanisms, external manifestations, etc. It suggests that human beings are born with basic emotions that can be distinguished by facial expressions and physiological processes and emotions are composed of basic emotions. In the early years, Ekman's theory suggested that there are six basic emotions (happy, sad, anger, disgust, fear, and surprise), and other emotions are considered as the combinations of these basic emotions[3]. Later, Plutchick used a wheel model to show a new concept of eight basic emotions (joy, trust, surprise, fear, sad, disgust, anger, and anticipation)[4], and later Izard added a ten basic emotions model (fear, anger, shame, disgust, contempt, guilt, distress, interest, joy, and surprise)[5]. To be more specific, researchers at the University of California, Berkeley have identified 27 types of emotions including admiration, amusement, anger and so on. In conclusion, many researchers believe that these basic emotions are a special class of discrete emotions, and their key features include separateness, finiteness, and singularity, which means these emotions that have a certain quantity are different from others and can not be duplicated by simply adding up some features of others[6].Based on this theory, it is not difficult to know that basic emotions are likely to become the cornerstone of other more complex emotions and more complex emotional states. Although there are still many disagreements in this theoretical field and complete consistency has not been achieved, their basic viewpoints in this regard generally have many similarities.

2.1.2. Dimensional approach

This theory considers emotion as a fuzzy synthesis, and when researching, we should focus on different dimensions of emotions, since they are related among various kinds of human emotions. In 1974, Russel and Mehrabian proposed two-dimensional and three-dimensional models one after another. They proposed the pleasure-arousal-dominance(PAD) model for emotions, which means we can totally divide emotions into 8 types using the three dimensions' arrangement and combination[7][8]. However, some scholars believe that positive and negative emotions are two independent dimensions rather than two poles of one dimension, such as the positive and negative affect model proposed by Watson and Tellegen, which believes that positive affect and negative affect are two relatively independent and basic dimensions[9]. More generally speaking, the transformation of emotions is achieved through an increase in positive emotions and a decrease in negative emotions (or vice versa), rather than a transformation from positive emotions to negative emotions (or vice versa).

2.2. Classification methods

Emotional analysis, also known as opinion mining, is a type of text classification where the main method is to extract the representation features of the text and classify them based on these features. Emotional analysis refers to various types of methods, such as natural language processing, text mining, and

DOI: 10.54254/2755-2721/73/20240399

computer linguistics, to identify and extract subjective emotional tendencies from the original text material.

This article mainly focuses on reviews about movies in the datasets IMDB, which are marked as positive reviews and negative reviews. In the sentiment classification field, the main research methods can be divided into two categories nowadays: one is based on a sentiment dictionary, and the other one is based on machine learning and deep learning.

There are two main methods for sentiment analysis based on dictionaries: rule-based methods and statistical methods. These two approaches are simple but have a limitation of scalability. The rule-based approach is to conduct emotional analysis based on certain rules and logic, such as calculating the number of positive and negative emotional words appearing, and judging the emotional tendency of the text based on their proportion; Statistical methods use statistical models such as naive Bayes and support vector machines for sentiment analysis.

Such methods can be traced back to the SentiWordNet. It is the earliest sentiment dictionary among countries other than China. It combines words with consistent meanings based on WordNet and assigns positive or negative polarity scores to the words. The sentiment polarity score can truthfully reflect the user's emotional attitude. As technology developed, other dictionaries like General Inquirer, Opinion Lexicon, and MPQA also appeared. Domestic dictionaries include the National Taiwan University Sentiment Dictionary, How Net, and Chinese Emobank. Most emotional lexicon above categorizes vocabulary based on a simple dichotomy of positive and negative emotions, neglecting the richness and diversity of human emotions. However, some of them, like constructed emotional lexicon based on emotional structure theory, subdivide into more diverse types or dimensions of emotions. For example, WordNet-Affect expands the core word range based on Wordnet, and a total of 4787 emotional words are obtained, which are divided into four basic emotions: happiness, sadness, anger, and fear[10]. They are asked to rate the pleasure, arousal, and dominance of each emotional word based on the PAD model of emotional structure rather than simply dividing it into positive and negative ones. Similarly, the Beijing University of Aeronautics and Astronautics has developed the first Weibo sentiment analysis system in China, Moodlens, which divides Weibo emotions into four types based on emoticons: happiness, sadness, anger, and disgust. They have studied mood fluctuation patterns and established a real-time monitoring system[11].

There are two main methods for emotion analysis based on machine learning: supervised learning and unsupervised learning. Supervised learning requires the use of training data annotated with emotional polarity for model training in order to obtain a classifier that can automatically recognize emotions. Commonly used supervised learning methods include support vector machines, decision trees, random forests, etc. Unsupervised learning does not require annotated data and uses clustering and other techniques to identify emotions in text.

In terms of deep learning methods, commonly used deep learning models include recurrent neural networks (RNN), long short-term memory networks (LSTM), and convolutional neural networks (CNN). RNN and LSTM can process sequential data, which is suitable for processing long text data. CNN can extract local features from the text. So, this article will take one method in each aspect, and they will be mentioned below, especially SVM and CNN.

2.3. Machine learning methods: SVM and CNN

2.3.1. SVM

The SVM model was originally built for binary classification, so it is chosen in this article. Pang et al. investigated machine learning methods based on training data and utilized techniques such as Naive Bayes, Maximum Entropy, and SVM. In the result, the text is distinguished as positive or negative and shows that an SVM classifier with unigram presence features outperforms other competitors. They found that Naive Bayes tends to do the worst and SVM tends to do the best[12]. Wang and Manning proposed a different SVM and used another feature to classify sentiment polarity. They showed that SVM

DOI: 10.54254/2755-2721/73/20240399

performed at a higher accuracy about long reviews; on the contrary, Naive Bayes was better at short reviews[13].

2.3.2. CNN

CNN is a type of deep learning (neural networks) or representation learning method. In 2008, Collobert and Wetson proposed a new CNN architecture for modeling sentences[14]. Thanks to the inspiration of their work, Kim used a four-layer CNN model for sentence classification, and he achieved good results in over half of his experiments[15]. Based on these tests, his team found that when it came to some pre-trained words, CNN worked clearly superior to others. Kalchbrenner proposed a dynamic convolutional neural network based on the k-max pooling mechanism and did several experiments to validate the model. Their model can not only recognize single n-grams but also learn to identify n-grams with different syntactic and semantic[16].

3. Research Methodology

The IMDB dataset from Kaggle contains 50,000 highly polarized comments from the Internet Movie Database (IMDB). The dataset is divided into 25000 comments for training and 25000 comments for testing, both of which contain 50% positive comments and 50% negative comments. Two classification models named SVM and CNN are utilized in the existing framework for classifying input data into two different classes. The performance analysis of both approaches is performed on the basis of accuracy.

3.1. SVM Classifier for Sentiment Analysis

In 1963, Vapnik proposed the support vector method when solving pattern recognition problems, with the decisive sample being the support vector. In 1995, they officially proposed statistical learning theory, and support vector machines were also developed based on this theory[17]. Support vector machines are originally built for binary classification, whose basic model is a linear classifier defined in the feature space with the largest interval, which distinguishes it from perceptual machines. Support vector machines also include kernel techniques, making it essentially a nonlinear classifier. Therefore, since the purpose of this paper is to analyze the positive and negative emotions of movie reviews, support vector machines are a good choice.

For ordinary binary classification problems, support vector machines can find the largest hyperplane in the feature space, maximizing the distance from all samples to that plane, thus achieving the best classification effect. The problem of finding the maximum interval hyperplane can actually be transformed into a mathematical optimization problem, but there is a prerequisite here that this hyperplane can correctly classify samples, and thus the following expression can be obtained.

$$\max \frac{1}{\|\omega\|}, s.t.\, y_i\left(\omega^T x_i + \beta\right) \geq 1, i = 1, \ldots n \tag{1}$$

However, $\|\omega\|$ is a value with a root sign. For ease of operation, we equate $\|\omega\|$ to the square of $\|\omega\|$ and multiply it by a coefficient of 1/2 for easy differentiation. Therefore, the final optimization problem expression is as follows.

$$\min_{\omega,\beta} \frac{1}{2}\|\omega\|, s.t.\, y_i\left(\omega^T x_i + \beta\right) \geq 1 \tag{2}$$

Now, it's actually about finding the optimal solution under constraint conditions. At this point, we can use the Lagrange multiplier method to solve and obtain the final hyperplane for classification.

DOI: 10.54254/2755-2721/73/20240399

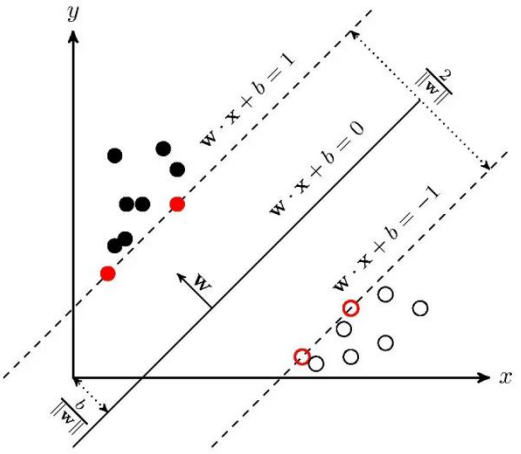

Figure 1. Finding the optimal hyperplane.

We also studied several different multiple classification methods based on binary classifiers. For example, one of them is to construct K "one-versus-rest" classifiers and select the class that classifies the test data with the maximum margin. Another method is to construct $\frac{K(K-1)}{2}$ "one-versus-one" classifiers and select the class with the highest probability, that is, the most classifier selection. However, in the classification of positive and negative emotions, multiple classifiers do not have much practical application value and will not be discussed too much.

3.2. data for SVM

In most cases, we usually cannot maintain consistency in the amount of data for both positive and negative samples. However, in this paper, we will use a consistent number of positive and negative samples for problem discussion, which will help us conduct theoretical analysis of various situations.

In this paper, we selected 10000 pieces of data from the IMDB dataset, with the same number of positive and negative samples as mentioned above, namely 5000 pieces each. Then we use the "train_test_split" function to randomly divide these data into training and testing sets, with the training set accounting for two-thirds and the testing set accounting for one-third.

3.3. CNN Classifier for Sentiment Analysis

In 1962, Hubel and Wiesel proposed a new concept of "receptive field" through their research on the visual cortex of the cat brain[18], which had important implications for the development of artificial neural networks in the future. In 1980, Fukushima proposed a neural cognitive machine and a convolutional neural layer with weight sharing based on the receptive field theory of biological neuroscience, which is regarded as the prototype of convolutional neural networks[19]. In 1989, LeCun invented a convolutional neural network by combining a backpropagation algorithm with a weight-sharing convolutional neural layer, and successfully applied the convolutional neural network to the handwritten character recognition system of the United States Post Office for the first time.

A convolutional Neural Network (CNN) is a feed-forward neural network composed of several convolutional and pooling layers, particularly in image processing. The basic structure of CNN consists of an input layer, convolutional layer, pooling layer, fully connected layer, and output layer. The convolutional layer and pooling layer are generally taken as several layers, and they are alternately set up, that is, one convolutional layer is connected to a pooling layer, and then another convolutional layer is connected after the pooling layer, and so on. Due to the local connection between each neuron in the output feature map in the convolutional layer and its input, and the weighted sum of the corresponding connection weights and local inputs, coupled with bias values, to obtain the input value of the neuron, this process is equivalent to the convolutional process, hence the name CNN. Convolutional neural

DOI: 10.54254/2755-2721/73/20240399

networks are often used to process image features, mainly due to their superior feature extraction and classification capabilities. Therefore, we can try to treat the text as a one-dimensional image and use CNN to classify the text.

3.4. data for CNN

In this paper, we used approximately 50,000 comments from the IMDB dataset. The data are then divided into two lists, one for reviews and the other for emotions. Then values that corresponding to the comments' emotions are set in order to determine whether our reviews are positive or negative in the future. After separating and labeling the dataset, 90% of it is used as the training set and the remaining 10% is used as the test set. The reason why not all datasets are included is that SVM is not ideal for such large amounts of reviews.

4. Research findings

4.1. SVM

For an SVM classifier, the expected input is not a direct review text, but a number with features. Therefore, we need to convert the text file into a number vector. A bag of words(BOW) is applied in this paper because we are concerned with the frequency of words in text comments rather than the order of words in text comments. English documents have natural space separators, so word segmentation is not required. We use the "TfidfVectorize" function to directly convert the original document set into a TF-IDF feature matrix, thereby achieving text vectorization. Each element in the vector represents the importance of a word in the text, which is evaluated using the TF-IDF method by multiplying the term frequency of the word by the inverse document frequency. Simply put, the higher the frequency of a word appearing, the more important it becomes, but when the word appears in all texts, it becomes less important in distinguishing emotions. We will use vectorized reviews as our input.

In the dataset, our input (reviews) and output (emotions) are already identified. Therefore, this is a supervised machine learning problem. We use the kernel function "linear" for our classification and then use "classification_ report" function to evaluate the classification results and obtain output shown in table 1.

Table 1. The result of SVM classification.

	Precision	Recall	F1-score	support
Positive	0.87	0.90	0.88	1632
Negative	0.89	0.87	0.88	1668
Accuracy			0.88	3300
Macro avg	0.88	0.88	0.88	3300
Weighted avg	0.88	0.88	0.88	3300

After setting up the SVM model and training it, we obtained the model we needed. In this test, it has an accuracy of 88.30%. From the results of multiple experiments, the average accuracy of the model is approximately 88%, so we chose this time as our representative.

4.2. CNN

We use "Tokenizer" function for text cleaning and standardization, which involves removing unnecessary characters from the text as well as unifying capitalization, removing punctuation marks, and so on. At the same time, this process converts the text into integers, where each integer is an index marked in the dictionary that includes all the words.

Next, we will standardize all different comments once again. This standardization requires changing all digitized comments to the same length to facilitate our input. Long comments will be truncated, while

DOI: 10.54254/2755-2721/73/20240399

short comments will be filled with meaningless characters.In this method, meaningless characters are zeroes.

This method use the Sequential model provided by Keras. The model consists of several Dense and Dropout layers, and after the experiment swish is chose as activation function. The embedding layer is used for dimension reduction, and the principle of dimension reduction is matrix multiplication. The Flatten layer is used to "flatten" the input, that is, to one-dimensional the multidimensional input. Dropout layers can avoid overfitting in our model. The whole model and its relevant parameters are shown as table 2.

Table 2. The whole model of CNN classification.

Layer	Output Shape	Param
Embedding	(None,120,128)	192000
Flatten	(None,15360)	0
Dense	(None,128)	1966208
Dropout	(None,128)	0
Dense	(None,64)	8256
Dropout	(None,64)	0
Dense	(None,32)	2080
Dropout	(None,32)	0
Dense	(None,16)	528
Dropout	(None,16)	0
Dense	(None,8)	136
Dropout	(None,8)	0
Dense	(None,1)	9

For the CNN model, after model training, it achieved an accuracy of 81.58%.

Relu and tanh are common choices among activation functions, but we have also tried other activation functions in an attempt to find the one with the best result. Since mixing activation functions rarely occurs in practical situations, we do not explore them too much, even though they may still have good results.The test_accuracy of different activation functions and their combinations are shown in table 3.

Table 3. Different activation functions and their test_accuracy.

activation	test_accuracy
sigmoid	49.40%
tanh	80.83%
relu	79.36%
swish	81.58%
swish+relu	80.54%
tanh+relu	81.56%
tanh+sigmoid	80.91%

Different optimizers are also tried to determine which was the best option for our model. We can find out that the adam and RMSprop models have similar high accuracy. Although the SGD model is fast, its accuracy is very low.Table 4 shows the different test_accuracy of these three optimizers

DOI: 10.54254/2755-2721/73/20240399

Table 4. Different optimizers and their test_accuracy.

optimizers	test_accuracy
Adam	80.47%
RMSprop	81.56%
SGD	49.40%

Due to the fact that sentiment classification for comments is a binary classification problem, we directly used the "binary_crossentropy" function without much exploration, as this function theoretically meets our needs very well.

And the our final result is shown in table 5.

Table 5. The Training Process and result of CNN Model.

Epoch	Accuracy	Loss	Val_accuracy	Val_loss
1	0.7572	0.6448	0.8376	0.5925
2	0.8292	0.5657	0.8208	0.4563
3	0.8738	0.6065	0.8380	0.9880
4	0.9142	0.5688	0.8226	1.5463
5	0.9371	0.5655	0.8208	1.7687
6	0.9463	0.5665	0.8206	1.9171
7	0.9534	0.5417	0.8146	2.4655
8	0.9556	0.5396	0.8158	2.5011

5. Discussion and analysis of findings

For SVM model, we can see a table representing the results. Precision and recall are two different parameters and f1-score is calculated based on them.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

$$\text{F1} - \text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

And support means the number of occurrences of each category for the true value. Macro avg calculates the arithmetic mean of precision of each category. Weighted avg adds the weight of each class, which is the size of its support.

To find out if there are better parameters, we used "GridSearchCV" function to loop through all candidate parameters to find the best performing one. This is because we use SVM model with fewer parameters and also have limitations on the number of datasets. Through the final running results, it can be found that,SVC(C=1, kernel='linear') is the best one. But this method has a drawback that it is very time-consuming.

DOI: 10.54254/2755-2721/73/20240399

For CNN model, we have found out the best model. The Swish activation has a simple form. It is almost a patchwork of sigmoid and relu, with most of the advantages of both. And we determined that the combination of RMSprop and Binary Cross Entropy was the best option.

For the binary classification problem, we can find that the SVM model has higher accuracy and also has a faster calculation rate in the actual operation process. However, the CNN model has better scalability, which means that in the future, when we will conduct more detailed multi classification of emotions, SVM will significantly decrease its effectiveness in handling such problems. However, CNN can optimize and identify more detailed emotions, such as anger, sadness, happiness, surprise, and so on.

In the future, we can combine the advantages of SVM and CNN models to further improve the accuracy and efficiency of data classification. Specifically, we can use the output of the last convolutional layer of CNN as a feature vector and then use SVM for classification. The advantage of this method is that we can utilize the feature extraction capability of CNN to convert image data into low dimensional feature vectors, thereby reducing the dimensionality of the data. This not only reduces the computational complexity of SVM, but also improves the accuracy of classification. However, CNN-SVM also has some challenges and limitations. Firstly, since CNN and SVM are two different models, their training processes are separate, requiring additional computational resources and time. Secondly, CNN-SVM may face limitations in memory and computing resources when processing large-scale datasets. In addition, the performance of CNN SVM is also influenced by factors such as the selection of CNN models and the adjustment of SVM parameters.

6. Conclusion and recommendations

This paper describes the sentiment classification of movie reviews based on SVM and CNN models, which achieve fast and high accuracy classification without the need for features provided by external resources. At the same time, we look forward to further implementing the CNN-SVM model in the future.

Acknowledgement

I would like to express my gratitude to all those who have helped me during the writing of this thesis.I gratefully acknowledge my family for their care and concern during my thesis, as well as to my thesis teacher Becky and others for their guidance.

References

[1] Yue G and Dong YH 2013 On the categorical and dimensional approaches of the theories of the basic structure of emotions *Nankai Journal(Philosophy,Literature and Social Science Edition)* 140-50

[2] Shu L, Xie J, Yang M, Li Z, Li Z, Liao D, Xu X and Yang X 2018 A Review of Emotion Recognition Using Physiological *Signals. Sensors* 18(7) 2074

[3] Ekman P 1992 An Argument for basic emotions *Cogn. Emot* 6 (3) 169-200

[4] Plutchik R 2001 The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice *Am. Sci* 89 (4) 344-50

[5] Izard CE Basic emotions 2007 Natural kinds, emotion schemas, and a new paradigm Perspect. *Psychol. Sci* 2 (3) 260-80

[6] Levenson RW 2011 Basic Emotion Questions *Emotion Review* 3(4) 379-86

[7] Russel JA 1980 A circumplex model of affect *J.Pers.Soc.Psychol* 39(6) 1161-1178

[8] Mehrabian A 1997 Comparison of the PAD and PANAS as models for describing emotions and for differentiating anxiety from depression *J. Psychopathol. Behavioural Assess* 19(4) 331-57

[9] Watson D, Tellegen A, Wiese D and Vaidya J 1999 The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence *J.*

DOI: 10.54254/2755-2721/73/20240399

Personal. Soc. Psychol 76(5) 820-838
[10] Valitutti A and Strapparava C 2010 Interfacing WORDNET-AFFECT with OCC Model of Emotions *LREC 2010 - SEVENTH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION* J16-J19
[11] Vujic A, Starner T and Jackson M 2016 *ISWC'16 - PROCEEDINGS OF THE 2016 ACM INTERNATIONAL SYMPOSIUM ON WEARABLE COMPUTERS* 36-39
[12] Pang B, Lee L and Vaithyanathan S 2002 *PROCEEDINGS OF THE 2002 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING* 79-86
[13] Wang, S and Manning C 2012 Baselines and bigrams: simple, good sentiment and topic classification *Proceedings of the 50th annual meeting of the association for computational linguistics* vol.2 90–94
[14] Collobert R and Wetson J 2008 A unified architecture for natural language processing: Deep neural networks with multitask learning *International Conference on Machine Learning*
[15] Kim Y 2014 Convolutional neural networks for sentence classification *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* 1746-51.
[16] Kalchbrenner N, Grefenstette E and Blunsom P 2014 A Convolutional Neural Network for Modelling Sentences *Proceedings of ACL 2014*
[17] Vapnik V 2006 Learning hidden information: SVM+ *2006 IEEE International Conference on Granular Computing* 22
[18] Kandel ER 2009 An introduction to the work of David Hubel and Torsten Wiesel *The Journal of Physiology* 587(12) 2733-41
[19] Tran M, Ray S and Pose R 2000 Neocognitron for rotated pattern recognition *COMPUTERS AND THEIR APPLICATIONS* 211-4