Computer Fraud and Security
ISSN (online): 1873-7056

Research on Emoji Sentiment Recognition Based on Convolutional Neural Networks

Yan Li
Xiamen Ocean Vocational College, Xiamen, China

Abstract:
Emoji usage has become increasingly prevalent in social interactions, making emojis essential tools for communication and marketing. This study proposes an emoji recognition method based on convolutional neural networks (CNN). The approach involves grayscale processing of emoji images, feature extraction using convolutional kernels, dimensionality reduction through pooling methods, and finally, emoji classification using a Softmax classifier. The CNN-based recognition method achieved an accuracy rate of 83.54% in emoji recognition tasks.

Keywords: Emoji; Convolutional Neural Network; Image Recognition.

INTRODUCTION

With the rapid development of computer and internet technologies, social platforms have profoundly transformed the way people communicate. Platforms like WeChat, Weibo, and TikTok have not only become primary venues for daily communication but also enabled diverse forms of information expression. Against this backdrop, emojis, as simple and intuitive communication tools, have gained worldwide popularity, seamlessly integrating into daily social interactions and significantly influencing communication habits and expression styles. Emojis not only serve as a supplement to textual communication but often surpass text in conveying richer and more nuanced emotional information. This unique characteristic has elevated emojis from mere communication tools to popular commercial elements and powerful marketing tools. For instance, many platforms, such as Facebook, Twitter, Taobao, Xiaohongshu, and Dianping, provide extensive emoji libraries for users. Notably, Xiaohongshu has even developed an emoji matching feature to further enhance user experience.

However, the use of emojis can exhibit ambiguity and variability in emotional expression across different contexts and cultural backgrounds. For example, the same emoji may be interpreted to express entirely different or even opposing emotions depending on the audience or context. This uncertainty in emotional expression poses challenges for online sentiment analysis, making the study of emoji-based sentiment analysis a key area of research. Exploring the mechanisms of emoji-based emotional expression and their roles in various contexts is thus of significant importance. Such research not only provides theoretical support for understanding emotional transmission in digital communication but also offers practical insights for commercial decision-making, sentiment monitoring, and user behavior analysis.

Currently, mainstream approaches to emoji sentiment recognition primarily rely on sentiment dictionaries, which categorize existing emojis by constructing emoji dictionaries. For instance, Wang Hao combined text and emojis to study gender prediction by analyzing the sentiments expressed through emojis in Weibo posts [1]. Li Nan analyzed the distribution and emotional divergence of emojis in Weibo comments, uncovering dynamic patterns of emoji usage and studying their emotional characteristics [2]. Guo Xianwei and colleagues classified frequently used, emotionally charged emojis into seven categories based on the Dalian University of Technology Sentiment Lexicon and constructed an emoji sentiment dictionary to improve the precision of sentiment classification in Weibo comments. The core principle of sentiment dictionary-based methods is to assign inherent sentiment values to emojis, typically determined through usage frequency statistics and textual sentiment integration. However, this approach has certain limitations, as it overlooks the feature variations of emojis during their usage, leading to gaps in the analysis of dynamic emotional changes.

Scholars have also explored emoji sentiment issues based on machine learning. For instance, Sanjaya constructed the first emoji library for machine recognition, enabling machines to associate emojis with their specific contextual meanings [4]. Zhao Xiaofang proposed a novel multidimensional sentiment classification method that integrates emojis and short texts, using deep learning models to analyze the sentiments of both text and emojis [5]. Zhang

Qian and colleagues investigated positive and negative sentiment issues based on text and emojis using an LDA model [6]. Chen Yaru and others demonstrated that combining emojis with a self-attention mechanism and BiGRU deep learning network models can effectively improve sentiment classification accuracy [7].

Due to the limited information and lack of semantic features in online short texts, traditional vector space model methods are often unsuitable, which introduces certain limitations to machine learning-based approaches. In response, this paper proposes an emoji sentiment recognition method based on convolutional neural networks (CNNs). Unlike traditional approaches relying on text and corpus construction, this method processes emoji images and extracts features using CNNs to determine the sentiment of emojis.

2. MODEL CONSTRUCTION

This paper constructs a CNN-based neural network for emoji sentiment recognition. As a feedforward neural network, CNN is capable of extracting features from two-dimensional images and optimizing model parameters using the backpropagation algorithm. The CNN model consists of a total of eight layers: one input layer (96×96 emoji), three convolutional layers (C1, C2, and C3), two pooling layers (S1 and S2), one fully connected layer, and one Softmax layer.

The convolutional and pooling layers contain several feature maps, with each feature map connected locally to the feature maps of the previous layer. The convolutional layers C1, C2, and C3 employ 32, 64, and 128 convolution kernels, respectively, with kernel sizes of 5×5. The pooling layers S1 and S2 use a sampling window size of 2×2. The fully connected layer comprises 300 neurons and is fully connected to the S2 pooling layer. The Softmax layer contains three neurons that classify the output features of the fully connected layer into three categories: positive sentiment, negative sentiment, and neutral sentiment.

2.1 Convolutional Layer

The convolutional layer serves as the feature extraction layer, where each neuron connects to a subregion of the image from the previous layer to extract local features. The first convolutional layer, C1, uses a 5×5 convolutional kernel to perform convolution on the input image of size 96×96 pixels. Each neuron is assigned a 5×5 local receptive field, resulting in a feature map size of 92×92 after the convolution operation. By applying 32 different convolutional kernels, 32 feature maps are generated, capturing 32 distinct local emoji features. Within the same feature map, all neurons share the same weights (using the same convolutional kernel) but receive inputs from different local receptive fields. The second convolutional layer, C2, applies 64 convolutional kernels of size 5×5 to the feature maps output by layer C1, producing 64 feature maps, each with a size of 88×88. The third convolutional layer, C3, employs 128 convolutional kernels of size 5×5 to perform convolution on the feature maps output by pooling layer S1, generating 128 feature maps, each with a size of 40×40.To prevent the vanishing gradient problem during backpropagation, the ReLU activation function is used as the activation function in all convolutional layers. This ensures efficient training and better feature representation.

2.2 Pooling Layer

Pooling layer S1 performs downsampling on the feature maps output by convolutional layer C2 using a 2×2 window, resulting in feature maps with a size of 44×44. The number of feature maps remains unchanged after downsampling, so there are still 64 feature maps. Similarly, pooling layer S2 applies a 2×2 window to downsample the feature maps output by convolutional layer C3, producing 128 feature maps, each with a size of 20×20.

2.3 Fully Connected Layer

The input to the fully connected layer must be a one-dimensional array. Since the feature maps output by the previous pooling layer, S2, are two-dimensional arrays, each feature map is first converted into a one-dimensional array. Subsequently, the 128 one-dimensional arrays are concatenated into a single 51,200-dimensional feature vector, which serves as the input for each neuron in the fully connected layer. The ReLU activation function is then applied to activate the neurons.

2.4 Softmax Layer

The final layer of the CNN is a Softmax classifier, which is a multi-output competitive classifier. When a sample is input, each neuron outputs a value between 0 and 1, representing the probability that the input sample belongs to a particular category. The category corresponding to the neuron with the highest output value is selected as the classification result.

EXPERIMENTAL ANALYSIS AND RESULTS

The experiment was conducted using the deep learning library MXNet based on the Python programming language. The dataset consisted of 3,308 emoji images collected from platforms such as WeChat, Weibo, Xiaohongshu, and other online sources. Among these, there were 935 positive emojis, 1,044 negative emojis, and 1,329 neutral emojis. Since the emoji images were sourced from different platforms, some were in color. To facilitate the experiment, all emojis were first cropped and normalized into 96×96 grayscale images, as shown in Figure 1, 2, 3.



Figure 1. Positive Emojis   Figure 2. Negative Emojis   Figure 3. Neutral Emojis

To improve the reliability of recognition results, the experiment employed a 3-fold cross-validation method. The 3,308 emoji images were evenly divided into four subsets, each containing three categories of emojis. For each experiment, three subsets were used as the training sample set, and the remaining subset was used as the test sample set. This recognition experiment was repeated four times, and the average recognition accuracy across the four iterations was taken as the final performance metric. To verify the effectiveness of CNN in emoji recognition, the experiment compared recognition algorithms based on Support Vector Machine (SVM), Multilayer Perceptron (MLP), and CNN. The results are presented in Table 1.

Table 1. Comparison of Recognition Accuracy for Different Algorithms

Algorithm	Recognition Accuracy (%)
SVM	73.03%
MLP (200-600-300)	80.26%
CNN	83.54%

To address the issue of overfitting in deep neural networks caused by insufficient training data or overtraining, dropout and data augmentation strategies were implemented to mitigate overfitting. For the dropout strategy, during recognition, the output values of neurons in the fully connected layer were set to zero with a probability of 0.5. When updating weights using the backpropagation algorithm, the weights connected to the dropped neurons were not updated. This approach is referred to as CNN+D.

For CNN+A, a data augmentation strategy was applied. Each image in the training set underwent transformations including rotation, horizontal translation, vertical translation, and horizontal flipping, with each transformation producing a new image. After augmentation, the total number of samples increased to 13,232. (Table 2)

Table 2. Comparison of Recognition Rates for CNN with Different Strategies

Strategy	Recognition Rate (%)
CNN	83.54%
CNN+D	84.97%
CNN+A	86.53%

The table data illustrates the performance of different strategies in recognition tasks, primarily comparing the recognition rates of the baseline Convolutional Neural Network (CNN) and its two extended optimization methods. Firstly, the standalone CNN model achieves a recognition rate of 83.54%, reflecting its baseline performance as a classic deep learning model in specific tasks. Secondly, the recognition rate improves to 84.97% when strategy "D" is applied to the CNN model. This strategy likely involves optimization techniques such as data augmentation, regularization methods (e.g., Dropout), or advanced feature extraction methods like deep convolution. This improvement indicates that strategy "D" effectively enhances the model's generalization capability and feature representation power.

Moreover, introducing strategy "A" into the CNN model further increases the recognition rate to 86.53%. Strategy "A" refers to the data augmentation approach. This significant improvement suggests that strategy "A" not only optimizes the model's structure but also enhances its ability to capture key features, thereby substantially improving recognition performance. Among the three methods, CNN+A achieves the highest recognition rate, demonstrating the superiority of this optimization strategy in the task.

CONCLUSION

Emoji sentiment recognition based on neural networks demonstrates high recognition accuracy, achieving a rate of 83.54% on the emoji dataset, surpassing the performance of SVM- and MLP-based methods. Additionally, neural network-based emoji sentiment recognition directly uses pixel values of emojis as input, avoiding the issue of feature loss inherent in traditional dictionary-based approaches. It also exhibits robustness to certain transformations of emojis, such as rotation and translation.

In this study, convolutional neural networks were applied to recognize emoji sentiment. The experimental results show that this method achieves high recognition accuracy and good generalization ability. Future research will focus on comprehensively considering the relationships among feature maps at each layer to further optimize the structure of convolutional neural networks.

ACKNOWLEDGMENT

Funding: 2021 Fujian Province Young and Middle-aged Teachers' Education Research Project (Science and Technology Category) - "Research and Design of P2P-based Video-on-Demand System" (Project No.: JAT210810).

REFERENCES

[1] Wang Hao, Xu Xiaoke. Gender Recognition of Social Network Users by Integrating Text and Emoji Features [J]. Complex Systems and Complexity Science, 2022, 19(04): 17-24.
[2] Li Nan, Zhang Yuhui. Research on Public Opinion Analysis by Integrating Dynamic Features of Emojis [J]. Modern Information, 2021, 41(08): 98-108.
[3] Guo Xianwei, Lai Hua, Yu Zhengtao, Gao Shengxiang, Xiang Yan. Sentiment Classification of Weibo Comments on Cases by Integrating Emotional Knowledge [J]. Journal of Computer Science, 2021, 44(03): 564-578.
[4] Sanjaya, Wijeratne, Lakshika, et al. EmojiNet: Building a Machine-Readable Sense Inventory for Emoji [J]. Lecture Notes in Computer Science, 2016.
[5] Zhao Xiaofang, Jin Zhigang. Multi-Dimensional Sentiment Classification of Weibo by Integrating Emojis and Short Texts [J]. Journal of Harbin Institute of Technology, 2020, 52(05): 113-120.

[6] Zhang Qian, Zhang Shibing, Ren Fuji, Zhang Xiaoge. Aspect-Based Sentiment Mining of Weibo Comments Using a Topic Model [J]. Journal of Chinese Information Processing, 2019, 33(06): 116-123+140.
[7] Chen Yaru, Chen Shiping. A Sentiment Analysis Model for Weibo by Integrating Self-Attention Mechanism and BiGRU Network [J]. Mini-Micro Systems, 2020, 41(08): 1590-1595.