

A Multi-Label Classification for Anime Image based on Deep Neural Network

Deng Pengfei, Ren Jinkai, Lv Shengbo, Feng Jiadong, Kang Hongyuan
Informatics College of Xiamen University

Abstract

Although deep convolution neural network (CNN) has achieved great success in single-label image classification, the images in ACG usually contain multiple tags, which can correspond to different objects, scenes, actions, and attributes in the image. Accurate identification of these tags can significantly optimize the accuracy of user search and help the ACG image sharing website to serve users. Picture recommendation better. Traditional multi-label image classification methods learn independent classifiers for each category and sort or threshold segment the classification results. Although these techniques work well, they can't make use of the label dependency in images explicitly. We use RNN combined with CNN to learn a joint image tag embedding method to represent the semantic dependence and relevance of image tags and can train the two information end-to-end into a unified framework.

1. Introduction

With the progress and development of science and technology, as an important medium of information dissemination, image has been widely studied in many fields, such as communication, unmanned driving, medical image analysis, aerospace, remote sensing and so on, and plays a more important role in the national society and economic life. People pay more and more attention to the research of image, which makes the field of computer vision usher in the golden age of vigorous development.

As a basic task in the field of computer vision, image classification is an important support for object detection and semantic segmentation. Its goal is to divide

different images into different categories and achieve the minimum classification error. After nearly 30 years of research, image classification has been successfully applied to all aspects of social life. Nowadays, it can be seen everywhere in our life - automatic classification of photo albums of smartphones, product defect identification, unmanned driving, and so on. According to the different objectives of the classification task, the image classification task can be divided into two parts: single-label image classification and multi-label image classification.

The ACG is an abbreviation of "Anime, Comic, and Games," used in some subcultures of Greater China. Because a strong economic and cultural connection exists between anime, manga, and games in the Japanese market, ACG is used to describe this phenomenon in relative fields. For some ACG image sharing websites (such as pivix, etc.), multi-label classification of animation images can automatically add corresponding tags to the uploaded images, greatly optimizing the accuracy of user search. At the same time, the website can also use image tags combined with user portraits to recommend animation pictures that users may like. Based on this, our group used a neural network model combined with CNN and RNN to write CNN-RNN to classify animation images with multiple tags.

2. Related Works

The traditional solution of the multi-label problem includes machine learning methods and deep learning methods.

Machine learning methods are generally as follows:

1. Problem migration, which transforms the multi-label

classification problem into a single label classification problem, such as converting tags into vectors, training multiple classifiers, etc.;

2. According to the characteristics of multiple tags, a new adaptive algorithm is proposed.

ML-KNN [Zhang and Zhou \[2007\]](#): the nearest k neighbor samples are obtained by the KNN algorithm, and then the number of adjacent samples belonging to a certain label is counted according to the labels of K adjacent samples. Finally, the label set contained in the test sample is determined by the maximum a posteriori probability principle (map).

Rank SVM [Li et al. \[2004\]](#): Based on SVM, this model adds ranking loss function and corresponding marginal function as constraints and extends the objective function to propose a multi-label learning algorithm. Firstly, the function $s(x)$ is defined as the size of label set of sample x , and then R is defined $r_k(x) = wk^T x + Bk$ if R is obtained r_k . In the maximum $s(x)$ elements $r_k(x)$ value is $r_1(x), r_2(x), \dots, r_Q(x)$, the label K is selected for the sample x , otherwise it is not selected. In the process of solving, a new sort function R is defined $r_k(x) - r_l(x)$. In this paper, a new sorting function based on the new sorting function is proposed to minimize the ranking loss, and the objective function and constraints suitable for multi-label classification are derived.

Multi-label decision tree [Vens et al. \[2008\]](#): a decision tree is constructed recursively based on the information gain criterion of multi-label entropy. Firstly, the information gain of each feature is calculated, and the feature with the largest gain is selected to divide the sample into left and right subsets, and then recursion is carried out until the stop condition is satisfied. For the new test sample, a path is traversed along with the root node to the leaf node, and the probability of each label is 0, and 1 in the leaf node sample subset is calculated. If the probability exceeds 0.5, the label is contained. After traversing all the paths to different leaf nodes, we can judge all the label information covered.

The deep learning method, the development of deep learning, has greatly improved the accuracy of image classification. The neural network has strong

nonlinear representation ability, which can learn more effective features in large-scale data.

CNN-RNN [Wang et al. \[2016\]](#) has a strong ability to extract information. According to this theory, firstly, CNN is used to train the input image to obtain the corresponding features, and then the corresponding features of the image are projected into the space consistent with the label, and the RNN is used to search the words in this space. The algorithm fully considers the correlation between categories and can effectively identify the tags with certain relationships in the image.

HCP(hypotheses-CNN-Pooling) [Wei et al. \[2016\]](#) where an arbitrary number of object segment hypotheses are taken as the inputs, then a shared CNN is connected with each hypothesis, and finally, the CNN output results from different hypotheses are aggregated with max-pooling to produce the ultimate multi-label predictions.

Many pieces of research on weakly-supervised segmentation based on image-level make full use of the information of a multi-label classification network. The main idea is to unify the labels into vector forms, construct a matrix label (such as $[0,0,0,1,1,0]$ for each image, and use special loss functions (Hamming loss, ranking loss, etc.) for training. This method successfully transforms the complex multi-label problem into a single label problem so that the traditional classification network can be used for training.

3. Plan

Collect and process datasets [Anonymous et al. \[2020\]](#). In the process of deep learning model, the selection of data set is very important. A high quality data set can improve the quality of model training and prediction accuracy. The selection and construction of the model is very important, and the training data is also very important to the model. While changing the model architecture to try to improve the accuracy of the model prediction, we also need to pay attention to improving the quality of the input data, and at the same time, we should consider increasing the number of input data to see whether the prediction effect of the

model can be improved.

Build model. Our model is based on CNN-RNN model. CNNVallet and Sakamoto [2015] has a strong ability to extract semantic information, while RNN can establish the association between information. According to this theory, firstly, CNN is used to train the input image to obtain the corresponding features, and then the corresponding features of the image are projected into the space consistent with the label, and the RNN is used to search the words in this space. The algorithm fully considers the correlation between categories, and can effectively identify the tags with certain relationship in the image.

Evaluation. We will design an evaluation function to evaluate the performance of the algorithm. More importantly, we will show and visualize the results and try to analyze the pros and cons of the results.

Reference

- Anonymous, Danbooru community, and Gwern Branwen. Danbooru2019: A large-scale crowdsourced and tagged anime illustration dataset. <https://www.gwern.net/Danbooru2019>, January 2020. URL <https://www.gwern.net/Danbooru2019>. Accessed: DATE. 2
- Xuchun Li, Lei Wang, and Eric Sung. Multilabel svm active learning for image classification. In 2004 International Conference on Image Processing, 2004. ICIP’04., volume 4, pages 2207–2210. IEEE, 2004. 2
- Alexis Vallet and Hiroyasu Sakamoto. A multi-label convolutional neural network for automatic image annotation. *Journal of Information Processing*, 23(6):767–775, 2015. doi: 10.2197/ipsjip.23.767. 3
- Celine Vens, Jan Struyf, Leander Schietgat, Sašo Džeroski, and Hendrik Blockeel. Decision trees for hierarchical multi-label classification. *Machine learning*, 73(2):185, 2008. 2
- Jiang Wang, Yi Yang, Junhua Mao, Zhiheng Huang, Chang Huang, and Wei Xu. Cnn-rnn: A unified framework for multi-label image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 2
- Y. Wei, W. Xia, M. Lin, J. Huang, B. Ni, J. Dong, Y. Zhao, and S. Yan. Hcp: A flexible cnn framework for multi-label image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(9):1901–1907, 2016. doi: 10.1109/TPAMI.2015.2491929. 2
- Min-Ling Zhang and Zhi-Hua Zhou. Ml-knn: A lazy learning approach to multi-label learning. *Pattern recognition*, 40(7):2038–2048, 2007. 2