Research on big data processing based on deep learning to extract social network image information

Xiao Wang

School of Engineering and Technology, Xi'an Fanyi University, Xi'an, 710105, China

Keywords: Deep learning; social network images; big data processing

Abstract: With the advent of the era of big data, a series of deep learning network structures have shown great advantages in the field of image processing. The social network represented by microblog contains a lot of valuable public opinion information about national security incidents, which exists in the text, image and network user data of social network. The emergence of massive cross media information challenges the traditional search methods. In recent years, the emerging deep neural network method provides a new solution for extracting and understanding cross-media semantic information. Data obtained from various complex systems are integrated into a network through the correlation between data, so the scientific problem faced by big data is essentially a network scientific problem. How to mine the relationship between data and analyze it quickly is a research hotspot at present.

1. Introduction

Social network analysis aims to establish a visual and measurable model for enterprises by analyzing personal social network relationships, and to unearth the business intelligence hidden in the process of information flow and value exchange among people, so that enterprises can obtain business opportunities or avoid potential. Risk [1]. With the rapid development of information technology, the era of big data where data is king has arrived, and strategic needs have also undergone major changes. Data is a strategic resource as important as natural resources and human resources, which implies huge economic value [2]. In recent years, social networks have gradually become an indispensable part of daily life. Taking Sina Weibo as an example, its monthly active users have reached 380 million, while daily active users have reached 180 million [3]. Usually, users share real-time information according to the mechanism of mutual attention. Daily user activities generate a lot of data and information, which has rich content and research value, but also brings challenges to the search of social networks. Traditional information search technology has been difficult to meet the user's multi angle and multi-level search needs [4]. The information of major emergencies affecting national security spreads rapidly and widely in social networks. If we can obtain the real and credible information of event location, time, people, social attributes and so on through the characteristics of time and space, it is of great significance to understand the cause, development process and impact of the event, make decision and disposal in time, and ensure national security [5]. Because there are a lot of irrelevant information in social networks, it is particularly important to identify the irrelevant information in Weibo and remove it to improve the search quality. The development of deep learning provides many advanced methods for this task. In recent years, big data has attracted great attention from the scientific and industrial circles. It is another technological revolution after the Internet of Things and cloud computing [6]. Through the analysis and mining of big data, we can get some users' behavior habits and other related data. The analysis and utilization of these data can further serve enterprises and give corresponding adjustment strategies for enterprise product decision-making. This undoubtedly brings huge benefits to enterprises, so the processing of big data is very important.

2. Deep learning and social network image information big data processing

In computer science, graph theory provides a simple and systematic modeling method for many

problems. Since graphs can be used to represent the association relationship between data, many problems can be represented by graphs, and the method of graph search is used to mine the association relationship between data. According to the connection mode between the nodes in the graph, the graph can be divided into two basic forms: directed graph and undirected graph. Neural networks were proposed in the 1950s. However, due to the lack of theory of network training algorithms, insufficient training samples and poor computer computing capabilities, the development of neural networks encountered bottlenecks.With the advent of the era of cloud computing and big data, the great improvement of computing power can alleviate the low efficiency of training, and the great increase of training data can reduce the risk of over fitting. Combined with the unsupervised layer by layer training strategy and the BP algorithm, it is possible to train the deep neural network. Therefore, the concept of deep learning has been widely concerned. A typical deep learning model is a deep neural network. Multiple hidden layers stack, each layer processes the output of the previous layer, so that the initial input and the output target are not closely related, which is transformed into a closer representation, making it possible to complete the task that is difficult to complete only based on the output mapping of the last layer. This learning mode can also be understood as automatic "feature learning". Image recognition is the earliest field of deep learning, in which convolutional neural network plays a huge role. Social network relationships (see Figure 1) can be simply characterized by users and their relationships, while users are often represented by vertices and user relationships are represented by edges. The research on points and edges can be extended to traditional graph theory and network topology. Both directed graph and undirected graph can be defined as G=(V, e). In which: v is a non-empty set called vertex set.



Figure 1 Social network diagram

E is a set, which includes an ordered two-tuple composed of all the elements in V, that is, a combination of edges, called an edge set. The research of image recognition is generally divided into two categories, one is the research of feature extraction, and the other is the research of feature learning. An image is composed of a matrix of pixels. When the system processes an image, if a feature learning algorithm is used for each pixel, the cost of doing so is huge, and the information contained in each pixel is limited. So we must use some feature extraction algorithms to preprocess the image, and finally use a series of features to represent the image. After this processing, the efficiency will be greatly improved by using machine learning algorithm for classification and recognition.

3. Application of Deep Learning in Social Network Image Field

At the beginning of the popularity of social networks, some scholars have studied the network topology of Cyberspace, MySpace and Orkut. Through snowball sampling, the average path length, aggregation coefficient, degree distribution and degree correlation of Cyberspace, MySpace and Orkut are calculated. In 2009, Kwak and others grabbed 41.7 million twitter users, 1.47 billion user relationships, 4,262 topics and 106 million tweets. Based on these data, they made a series of

topological analysis and reached the following conclusions.

a) The distribution of the number of fans and followers does not conform to the power law distribution.

b) Low interactivity, that is, the number of users interacting with friends \div (the number of user followers + the number of users' fans) = 22.1%.

c) The degree of separation is 4.2 (and most social networks conform to the 6 degree separation theory).

d) Group: users with less than 1000 fans are geographically closer to their friends.

e) The number of fans is not directly proportional to the influence of users.

f) 85% of twitter content is headline or instant news.

In 2009, Jake Hofman grabbed 25 million twitter user profiles and 8 million user relationships, and used Hadoop for clustering calculation. He calculated the probability that users with the same fans/concerns are mutual friends, and concluded that many users have no fans, some users have many fans, and many users only have a few fans. The calculation results are shown in Figure 2.



Clustering coefficient

Figure 2 Jake Hofman's research results

The industry has always had a very popular view: under the condition of big data, simple machine learning models will be more effective than complex models. In many big data applications, the simplest linear model is widely used. Linear models such as the maximum entropy method and support vector machine have achieved great success in web search ranking, spam filtering systems, and content recommendation-based systems. The recent breakthroughs in deep learning have resulted in amazing results in the field of speech recognition and image processing. Deep learning technology was first applied to the direction of image recognition and achieved remarkable results.

For image features, firstly, the images that may not have practical meaning are removed according to the aspect ratio. Then the images are normalized to the same size. The residual neural network is used to extract the features. The feature of the residual neural network is that there is a shortcut of gradient conduction between convolution layer and convolution layer, so the gradient disappearance caused by the depth of neural network can be reduced. The resnet50 neural network structure is adopted, and the neural network is initialized by the weights pre-trained in ImageNet image classification task, and then the weights of the network are fine-tuned by using Weibo image data. It can be seen that the classification probability of each class is obtained by multiplying the input vector by the weight of neurons and adding deviation in the last fully connected layer of the model, so the vector input by the last fully connected layer can be used as the feature representation

of the whole image. As shown in Table 1.

	1	2		1023	1024
Image 1	0.311468	0.796759		0.330088	0.103066
Image 2	0.094519	0.389111		0.312401	0.242882
Image 3	0.360025	0.079587		0.01695	0.886343
Image 4	0.132107	0.588127		0.14223	0.027658
Image 5	0.470794	0.165772		0.062962	0.321917
1	1 1 4 1	. 1 T	.1	C' 1	1 . 1

Table 1 Image feature vector

The alexnet network proposed by Alex et al. Is the first deep convolution neural network for image recognition. Subsequently, a series of deep learning development in image recognition are based on this. Compared with the traditional CNN structure, the alexnet network is deeper and wider, with five convolution layers and three full connection layers. AlexNet network established the dominant position of deep learning in image recognition, and defined the general discipline framework of deep learning model in image processing field. Feedforward convolutional neural network: Convolutional layer and pooling layer are superimposed on each other as features. Extraction Then, connect multiple fully connected levels into classifiers to fix and unify the direction of information flow.

4. Conclusions

The 21st century is an era of information and data. For example, social relations are so important data that we must dig and learn them in depth in many aspects in order to make full use of its value. Deep learning mainly focuses on convolution neural network in image field, but convolution operation is destined to have a large amount of computation in the whole network, which leads to a long time for network training. Changing the convolution operation form to simplify the computational complexity will also become a major development direction. At present, information technology is showing the characteristics of exponential growth, and we have entered the era of big data. Data is a strategic resource as important as self-heating resources and human resources, and it implies huge economic value. How to effectively organize and process big data will play a huge role in social and economic development. Graph search and deep learning algorithm play an increasingly important role in the process of big data processing because of their strong network analysis ability and feature recognition and classification ability. The research and study of social network is endless. Scholars are also studying human beings when they study their own characteristics.

References

[1] Yin Qilin, Wang Jinwei. Overview of the application of deep learning in image processing. Journal of Higher Education, vol. 81, no. 9, pp. 78-80, 2018.

[2] Li Xiangyu, Sun Xiaoqing. Research on image classification analysis based on deep learning. China Information Technology, no. 4, pp. 48-50, 2019.

[3] Yang Zhenzhen, Kuang Nan, Fan Lu, et al. Overview of image classification algorithms based on convolutional neural networks. Signal Processing, vol. 34, no. 12, pp. 1474-1489, 2018.

[4] Ran Xiaobin, Liu Yuewen, Jiang Jinhu. Big Data Analysis of Social Network Active Behavior: The Perspective of Network Externality. Management Science, vol. 30, no. 5, pp. 77-86, 2017.

[5] Li Xiaolong. Personal information security status and protective measures of social networks in the context of big data. Exploratory Science, no. 4, pp. 62, 61, 2016.

[6] Wang Shibo. On the social network big data analysis framework and its key technologies. Science and Information Technology, no. 3, pp. 39-40, 2018.