2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)

Research on Fishery Trajectory Analysis and Fishing Ground Discrimination Based on CNN

Wenjin Xu[1,2]
[1]College of Information Science and Engineering,
Ocean University of China
[2] College of Information Science and Technology,
Qingdao University of Science & Technology
Qingdao, China
e-mail: wenjin@qust.edu.cn

Qin Xie[2]
[2]College of Information Science and Technology,
Qingdao University of Science & Technology
Qingdao, China
e-mail: xieqin1013@163.com

Yue Ma[2]
[2]College of Information Science and Technology,
Qingdao University of Science & Technology
Qingdao, China
e-mail:1375345029@qq.com

Abstract---With the continuous progress of science and technology, people's production and lifestyle which has lasted for thousands of years are also changing quietly. More and more things, people, data, and the Internet are connected together, which also makes the power of big data calculation and mining increase exponentially. The navigation data of Zhejiang coastal fishermen are recorded by satellite. Using big data calculation can combine the ocean climate information, fishing vessel location information, and the data formed by fishing grounds into a big fishery data platform. After many captains use the data provided by this platform for fishing guidance, they have achieved a very good catch. They use scientific and technological innovation to change fishing methods so that fishermen can no longer rely on heaven for food. The risk is difficult to predict. The research in this paper is to find the common characteristics of fishing grounds by analyzing the navigation data of fishermen in the Zhejiang coastal area recorded by satellite and at the same time, expand the search range to estimate the new fishing ground location. This work can be used to mine data by using convolutional neural network.

Keywords: data mining; marine fishing; CNN; deep learning

I. INTRODUCTION

As early as 1988, the Portuguese government at that time had presided over the development of the earliest fishing vessel monitoring system (VMS), which was used to detect fishery activities. In order to better manage and conserve limited fishery resources, all countries with territorial waters have basically formulated relevant laws and regulations and started to develop their own fishery regulatory systems. In the process of actual supervision, European, American, Japanese, and Russian fishing powers have gradually developed a more perfect VMS system. In order to obtain the information of their own fishing vessels, many developing countries and regions have also established a fishing vessel positioning system for the management of Marine Fisheries and fishing vessels. Generally speaking, VMS has become an important means for coastal countries to protect and manage fishery resources. This technology

2473-3636/20/$31.00 ©2020 IEEE
DOI 10.1109/DCABES50732.2020.00030

also in the constant exchange of countries has obtained considerable development.

China's fishing vessel management mainly depends on the ship borne Beidou satellite navigation system, which is also a kind of VMS. Beidou satellite system integrates communication and navigation functions and can capture and record high-precision navigation data of fishing vessels in real-time. The system has been tested and put into operation and has been widely used in the marine fishery supervision of China's coastal provinces.

The behavior characteristics and rules embodied in the fishing data of fishing vessels are an important basis for the management and protection of fishery resources. These data will be affected by various factors, for example, the fishing habits of fishermen in different countries and regions are different. In this paper, different fishing targets (fishing boat, shrimp boat, crab boat, etc.) are analyzed. Through deep mining of the data returned by the fishing boat equipped with the Beidou navigation system, not only the operation type of the fishing boat can be identified, but also the navigation condition and fishing state of the fishing boat can be judged. The fishing behavior characteristics of the fishing boat can be analyzed, traced and obtained, so as to provide a more accurate reference for the fishery department.

II. ALGORITHM

K-means (K-means): hard clustering algorithm, membership degree is 0 or 1, and the sum of squares within the class is minimized.

Fuzzy c-means clustering (FCM): fuzzy clustering algorithm, the membership degree is taken as [0,1], the sum of squares of weighted errors within the class is minimized.

A. K-means

Firstly, K objects are selected from all the samples, and these objects are regarded as cluster centers, and the cluster centers generated in this step are the initial points. Then, the distance between each object and each seed cluster center is calculated. According to the calculated distance, all objects are corresponding to the nearest cluster center. After finishing the above work, the centers of all the clusters formed will be recalculated according to the current situation due to the continuous addition of objects. Random selection - forming cluster centers - classifying according to distance - recalculating cluster centers, and this process will continue to circulate. Guide the program to determine that the predetermined termination condition is reached, which can be any of the following three conditions:

1) No (or minimum) subjects are reassigned to different clusters.

2) No (or minimum) experimental cluster centers changed again.

3) The sum of squares of errors is a local minimum.

The similarity between data is measured by Euclidean distance.

B. *Fuzzy c-means clustering algorithm*

In practice, many objects in a dataset can't be assigned to distinct sets, but it's difficult to manually transfer an object to another set, and the final result will have a large deviation. In order to solve this problem, we can first assign a weight to each object and each set. Obviously, the calculation of probability is the most effective way to get a fit degree. However, in practice, it is difficult to determine the statistical model suitable for experimental objects. However, the non-probabilistic characteristics of fuzzy C-clusters provide great universality for researchers. This method has been widely used.

1) Initial fit matrixU(0)If there are n samples and the number of classes is k, then the fit matrix should be n * k matrix;
2) According to formula

$$m_j = \frac{\sum_{i=1}^{n}[\mu_j(x_i)]^b x_i}{\sum_{i=1}^{n}[\mu_j(x_i)]^b} \quad .$$

And recalculate the cluster center $m_j, j = 1,2,3,\dots,k$;

3) Update the membership matrix $U(t), U(t+1)$;

4) If the termination conditions are met

$$max_{ij}\left\{\left|\left|U(t+1)_{ij} - U(t)_{ij}\right|\right|\right\} < \varepsilon,$$

stop iteration, otherwise return to step 2.

III. LSTM FOR TRAINING

Generally speaking, using machine learning to do data mining and analysis is equivalent to cooking a piece of potato silk. The initial data is not marked with the location of the fishing ground, so it is impossible to obtain valuable information intuitively, which is equivalent to potatoes just harvested from the field. The function of LSTM is to train a model. After training, the model can predict the location of the fishing ground as long as the labeled data are input.

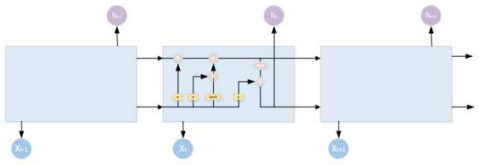

Figure 1. LSTM structure.

Figure 1 shows that the structure consists of three parts, but in fact, it represents the connection of the results of the same neuron at different times.

Four small yellow rectangles are the hidden layer structure of the general neural network. The first, second, and fourth activation functions are sigmoid, and the third activation function is tanh. The input x at time t and the output H (t-1) at time T-1 are spliced and input into the unit. It can be understood that our input x (T) is input into four small yellow rectangles. The operation of each small yellow rectangle is the same as that of the conventional neural network (matrix multiplication). The memory part is completely controlled by various gate structures (i.e. 0 and 1). At the same time, the input x (T) is input into four small yellow rectangles. The input time not only has the original data set but also increases the output result of previous data. Therefore, LSTM is similar to a common neural network, only adding some contents to the input and output. The cell can be roughly divided into two horizontal lines. The upper horizontal line is used to control long-term memory, and the lower horizontal line is used to control short-term memory.

IV. RESULT

The data used in this paper is the navigation data of Zhejiang coastal fishing vessels collected by Beidou satellite. The amount of data is extremely large, reaching nearly 5 million lines in June 2016 alone. In this paper, through the analysis of data, identify the location of the fishing ground, and then provide a certain reference for fishermen. The meanings of these data are fishing vessel ID, time stamp, longitude, latitude, heading angle and speed. As shown in Table 1.

TABLE 1. Data collected by Beidou satellite

ID	171431	270525
Time Stamp	1464710544	1464710426
Longitude	18473306	18549913
Latitude	73892856	73981132
Angle	2440	2400
Speed	89	89

How to make the clustering result obvious? According to the extracted features, the greater the distance and distance difference, the greater the angle change, the smaller the change of space and time of fishing vessels, the greater the possibility of fishing spots. Then, the samples are clustered into two categories, and the final result labeled as 1 is the location of the fishing ground. Thus, the map marked with the overall trajectories of fishing

vessels and predicted as fishing spots are drawn.

By clustering the red trajectories of fishing vessels, the purple point is the cluster center, and the blue part is the expected fishing ground location.

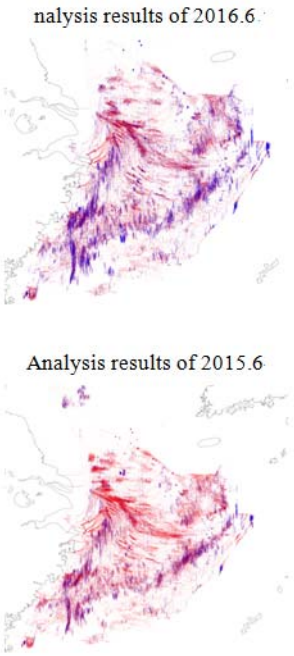

Figure 2. Comparison of different time periods

It can be seen that the predicted location is basically the same, and the suitable fishing area is a V-shaped sea area. There are a lot of fishing trajectories at the opening of the V-shaped area, but only a small number of prediction points are not piecemeal. The results show that the speed and turning speed of the maneuvering fishing vessels in this range do not conform to the fishing characteristics.

The output of this study is mainly in the form of a bit thermal diagram, which can directly see the predicted fishing ground location and the trajectories of fishing vessels in the samples. However, many calculations and results are limited by the quality of the samples. Although the amount of data that can be collected is large enough, the format is different from that of the design code. Although the sample data which can be trained has been denoised, there are still a lot of outliers in the samples, which leads to a large range of mapping.

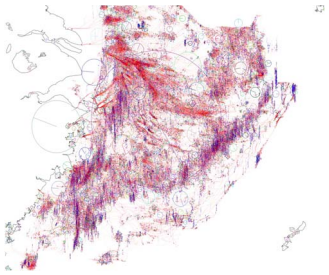

Figure 3. Feature point matching

This is a CNN algorithm to detect local features. The algorithm obtains features by finding interest points (or corner points) and their descriptors related to scale and orientation in a picture and performs image feature point matching and achieves good results.

REFERENCES

[1] Cao Shiyu, Liu Yuehu, Li Xinzhao, vehicle target detection based on fast r-cnn, China Journal of image graphics, time: May 16, 2017

[2] Yang Zhenjie, research on traffic sign recognition method based on CNN, Tianjin University of technology, time: January 17, 2017

[3] Luo Wenhui; Dong Baotian; Wang Zesheng, "short term traffic flow prediction based on cnn-svr hybrid deep learning model", transportation system engineering and information publication time: October 15, 2017

[4] Yang Liangbin, visualization analysis of research status and trend in data mining field, library and information work, published on December 15, 2015

[5] Zhang Xi, research and improvement of association rule algorithm in data mining, Beijing University of Posts and telecommunications, published on December 18, 2014

[6] Chang Kai, comparison and analysis of data mining classification algorithm based on neural network, Anhui University, published on April 1, 2014

[7] Zhang SHIMENG. Implementation and application of Hadoop based wireless network service and user mining platform [D]. Beijing University of Posts and telecommunications, 2015