Study on cluster analysis of prediction about Hail

Ma li, Wang Heling

Applied Mathematics, Xinjiang University of Finance and Economics, Urumqi 830012, China
*corresponding author's email: 1352690560@qq.com

Abstract— **Hail is one of the frequent meteorological disasters all over the world. In order to reduce its economic losses, the prediction of hail cloud has been studied deeply in various countries. Based on the overseas and home experiences, an effective hail cloud recognition model was proposed, which is mainly calculated from the morphological characteristics of the hail and the basic reflectivity factor. First, the effective areas were extracted of hail from the training samples, then, the cluster analysis and discrimination were carried out. By clustering, summarizing and summarizing the morphological characteristics of Hail Cloud monomer and the general law of reflection intensity research, the discriminant function was established. According to this model, calculated reflection intensity probability which more than 45dBZ, obtaining a parameter value from it, when the parameters of the probability is greater than the predetermined value, it could be hail weather. Finally, the accuracy of the model is verified by simulation experiments. The results show that the model has a good discrimination effect on Hail Cloud and can be applied to practical life.**

Keyword—radar reflectivity, k-means clustering, Mahalanobis distance, analog simulation

I INTRODUCTION

Hail is a kind of solid precipitation, formed in the strong cumulonimbus cloud. Generally, the diameter is about 5 to 50mm.it is a tremendous meteorological disaster caused by severe convective weather. China is a country with frequent hail disasters, different levels of hail disaster occur in individual areas every year. It causes huge economic loss to our country, these losses involving agriculture, transportation, communications, electric power and other aspects. It even threatens the safety of people's life and property. Therefore, it is urgent to understand the weather changes of hail disasters, and make predictions scientifically, so as to reduce the loss from hail disaster. According to the present situations, there are two main types of hail suppression measures, one is explosion method, another is catalytic method. These two methods are based on the ground meteorological data and sounding data, refer to the data and model obtained from image processing, then assess the recent weather combined with the local situation to achieve artificial intervention effect. Thus, the image processing is particularly important in this aspect of the data mining classification. The Cloud Radar image processing more precise and the data are more scientific, the hail prediction is more accurate.

II RADAR REFLECTIVITY IMAGE

The electromagnetic wave of Doppler weather radar will scattering with particle in the atmosphere, resulting in a part of the energy scattered back to the radar, the radar reflectivity is the space in the unit volume of atmospheric substances on the microwave radar scattering cross section. Among them, the density distribution and scale size of precipitation particles are measured by the reflectivity factor. The unit is expressed by dBZ, which is used to estimate the intensity of precipitation and the severe convective weather such as gale and hail. When the reflectivity reaches a certain value, weather warning will appear, this part is called monomer. In general, the reflectivity value is more greater more easy to appear strong convective weather. So that is why we are looking at Radar reflectivity image. The radar reflectivity image is based on the legend of the basic reflectivity factor map, The principle of color in RGB color space to read the reflection map. According to the reflection intensity of different images show different colors. For observing, the reflectance intensity is divided into 17 grades, from -35dBZ to 94dBZ, which are represented by 17 different colors. Black indicates an invalid echo, and the higher the grade, the greater the likelihood of extreme weather. As shown in Figure1.

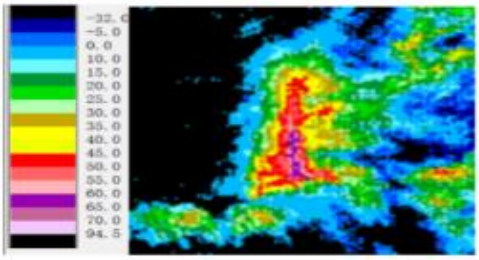


Figure 1 Radar reflectivity image

For heavy rain and hail, their particle for water polo and hockey, water polo in more than a certain limit will rupture, ice hockey is not, so the size of ice particles will be greater than the water polo, which shows the reflectivity of hail monomer was greater than heavy rain. According to the existing data show that when the reflection intensity is exceed or equal to 40dbz, is likely to appear thunderstorms, when its value at 45dbz or above 45dbz, the relatively large possibility of strong convective weather, such as rain, strong winds, hail and other extreme weather. Therefore, when study the hail weather will select areas whose reflection intensity is greater than 45dbz.

III K_MEANS CLUSTERING

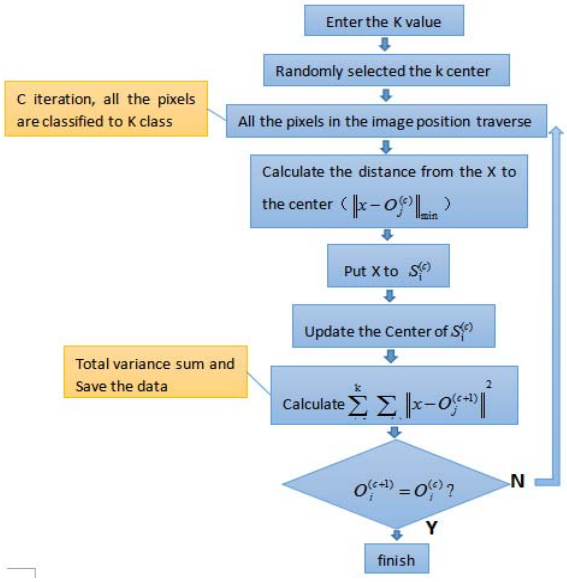
Clustering analysis divided a large number of complex data into n sub class of the same nature. in the group can find the common characteristics , and can be

978-1-5386-1401-3/17 $31.00 © 2017 IEEE
DOI 10.1109/ICSCSE.2017.58

compared between the groups in the observation, to find out the differences between the groups and to excavate the distribution of data and some inherent rules. Therefore, clustering analysis is widely used in many fields such as image processing, pattern recognition, data compression, industrial economic forecasting and so on. The K shows the specified number of categories, and the K-means algorithm needs to determine the K value in advance. The performance of the algorithm depends largely on the K. Means is the center of the cluster, and the initial cluster centers are randomly selected, and all the samples should move closer to these centers in accordance with the principle of proximity.

Through the investigation and analysis of K-means clustering algorithm, combined with the gray level processing of digital image, two value detection and edge detection technology, a clustering based image text recognition and extraction algorithm is designed.

This is algorithm flow chart:



IV DISTANCE DISCRIMINANT ANALYSIS

Distance discrimination: the basic idea is to obtain the barycentric coordinates of each category from training samples , and then find the distance of the new sample to the center of gravity of each category, and classify them into the nearest class. The most common distance is Mahalanobis distance, and sometimes use Euclidean distance.

In this paper, there are two known overall: G_1(hail cloud) and G_2 (not hail cloud), a test sample of X .We judge the test sample X belong to the general (G_1 hail cloud) or overall G_2 (not hail cloud), Mainly to see who it is near. From the comparison between the $D(X,G_1)$ and $D(X,G_2)$, if $D(X,G_1)$ < $D(X,G_2)$, then X belong to the $D(X,G_1)$; if $D(X,G_1)$ > $D(X,G_2)$, then X belong to the $D(X,G_2)$; if $D(X,G_1)$ = $D(X,G_2)$, then X is testing sample.

If the X to the i class $\bar{x}^{(i)}$ is defined as the Mahalanobis distance, i.e.:

$$D^2(X,G_i)=(X-\bar{x}^{(i)})'(\sum_i)^{-1}(X-\bar{x}^i) \quad (i=1,2) \tag{1}$$

The criterion can be written as follows:

$$W(X)=D^2(X,G_2)-D^2(X,G_1)$$

$$\begin{cases} X\in G_1, & if \quad W(X)>0 \\ X\in G_2, & if \quad W(X)<0 \\ \text{to be confirmed}, & if \quad W(X)=0 \end{cases} \tag{2}$$

According to the nearest distance principle differentiate and classify for X. First, calculate the Mahalanobis distance from sample X to 2 ensembles: $D_i^2(X), i=1,2$, then give X to the smallest one.

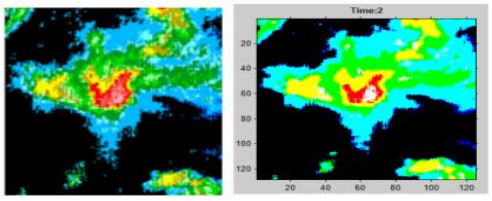
$$\text{If} \quad D_1^2=\min_{i=1,2}\{D_i^2(X)\} \tag{3}$$

Like that $X\in G_1$

V MODEL ESTABLISHMENT AND TEST

1.Establishment of Model

According to meteorological knowledge, extreme weather generally occurs in areas where the intensity of reflection is greater than 45 dBZ, The area in which the image is red or more than red. Therefore, the paper only chooses the area of 11 - 17 of the magnitude as the object to study. Firstly, radar reflectivity image preprocessing. Take K-means to cluster for image, using MATLAB program to realize this purpose (including sample figure below).



Unprocessed image clustering results

figure 1

Next, the segmentation process is done on the MATLAB software, which is based on the color difference to the clustering results. The results are shown in figure 2.

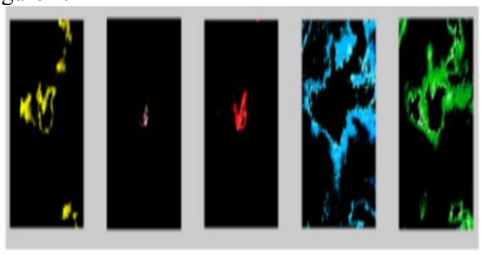


figure 2 segmentation results

In this paper, Select rainfall and hail storm images as samples, each category of the 50, and the training samples of each 30 images were extracted randomly.