

American Journal of Basic and Applied Sciences (ISSN:2637-6857)



A mathematical model for grading clouds to study solar flares based on fuzzy c-means algorithm

Xiaoxi Men¹, Guozhu Chen^{2*}, Xiaoxiong Zuo³, Haijun Guo²

1.College of Electrical Engineering & New Energy, China Three Gorges University, Yichang, 443002, China. 2.College of Hydraulic & Environmental Engineering, China Three Gorges University, Yichang, 443002, China. 3.College of Science, China Three Gorges University, Yichang, 443002, China.

ABSTRACT

In this paper, a cloud-level grading comprehensive evaluation model based on fuzzy c-means algorithm is established. The image is binarized by the maximum inter-class variance method to obtain the dim curve. According to the dim curve of the four quadrants, the distribution and thickness of the cloud layer are discriminated. Three attributes of the effective area of the cloud layer, the total amount of clouds, and the thickness of the cloud amount are selected as the characteristic data of the comprehensive evaluation model.

Keywords: Dim curve; Fuzzy c-means algorithm; Cloud grading.

*Correspondence to Author:

Guozhu Chen

College of Hydraulic & Environmental Engineering, China Three Gorges University, Yichang, 443002, China.

How to cite this article:

Xiaoxi Men, Guozhu Chen, Xiaoxiong Zuo, Haijun Guo. A mathematical model for grading clouds to study solar flares based on fuzzy c-means algorithm. American Journal of Basic and Applied Sciences, 2019, 2:10

 eSciPub
eSciPub LLC, Houston, TX USA.
Website: <http://escipub.com/>

1. Introduction

The solar flare is a violent explosion that occurs inside the solar atmosphere. It releases a large amount of energy quickly in a short period of time, causing local areas to be instantly heated, emitting various electromagnetic radiation outwards, and the surrounding particles are enhanced. It is generally believed that the flare is produced by the rapid release of the magnetic field energy of the active area near the sunspot. Flare occurs when the magnetic field of the sunspot produces a magnetic field of another small-scale special structure.

Long-term observations indicate that most of the flare phenomenon occurs between the low corona zone above the sunspot group and the top atmosphere of the sun chromosphere. And the more complex the sunspot group structure, the higher the probability of occurrence of large flares. A normally developing sunspot will produce a flare on average for a few hours.

In recent years, most researchers have favored the study of flare phenomena. But the photographs of the flares obtained must be processed before they can be used for research because the observations are often occluded by clouds. According to above, a mathematical model is established to classify the amount of clouds to study solar flares.

2. Model Preparation

2.1 Extraction of Data Features

After analyzing the data characteristics of the solar flare, three attribute characteristics of the cloud amount are obtained, which are the effective area of the cloud layer, the total amount of the cloud, and the difference in the thickness distribution of the cloud amount. And they are defined as follows.

Effective area of the cloud: We select the cloud amount of the baseline as a reference. The cloud amount is the difference between the amount of cloud before the filtering and the amount of cloud after the guided filtering. We compare the cloud amount of each pixel of the sample to be tested with the reference cloud of the corresponding pixel, and the area of the pixel above the reference cloud is considered valid.

Total amount of clouds: The sum of the clouds at each pixel of the sun's full-day map.

Difference in thickness distribution of cloud

amount: Since the cloud amount of each pixel on the full-day map is regionally distributed, in order to show the difference of regional distribution, the variance of each region is selected to show the difference of cloud amount.

2.2 Data Standardization

Since the dimensions of these three attributes are different, the data needs to be normalized first. Standardization refers to removing the unit limit of the data and converting it into a dimensionless, pure value so that indicators of different units or magnitudes can be compared and weighted. The basic idea of using standardized processing is to standardize data based on the quotient division of the maximum and minimum values of the original data. The expression is as follows.

$$x' = \frac{x - \min x}{\max x - \min x}$$

In the above formula, a represents sample data before unnormalized processing, and b is data after normalization processing.

2.3 Solving the Dim Curve of the Edge

Dimness on the edge means that the brightness of the sun's round surface gradually darkens from the center of the sun to the edge of the sun. We can observe the solar surface cloud distribution and the general situation or specific location of the flare distribution by observing the dark curve of the edge, thus facilitating the subsequent cloud processing.

1) Binary Image

The key to image binarization is to find the appropriate threshold to better segment the foreground and background of the image. Considering the overall gray distribution of the image, the maximum inter-class variance method is selected to automatically threshold the image. The same area often has gray-scale similar characteristics, while different areas show significant gray-scale differences. When the gray level difference between the two regions separated by the threshold is large, the difference between the average gray level of the two regions and the average gray level of the entire image is also large, and the variance between the regions is an effective parameter for describing the difference. The maximum inter-class variance method is determined by maximizing the variance between classes to determine the threshold of binarization.

The gray value of the full-day sun image is between 0 and 255, which we normalize to 0 to 1. Using *MATLAB* software, the maximum inter-class variance method is used to automatically calculate the threshold value is approximately

equal to 0.3608, so the gray value of the full-day solar image is binarized. When the gray value is greater than 0.3608, it is taken as 1, and when the gray value is less than 0.3608, it is taken as 0. We take a flare image and after the above processing, the image is got (Fig. 1).

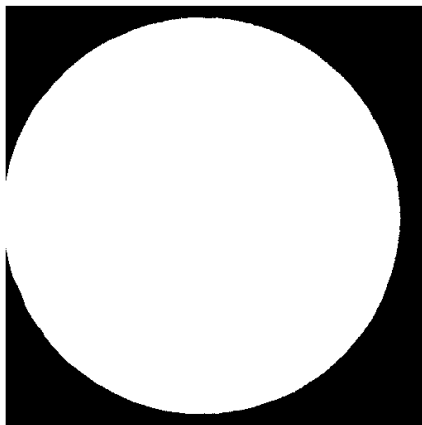


Fig. 1: The image after binarization

As can be seen from Fig. 1, the periphery of the sun is black, the sun is white, that is, the surrounding gray value is zero, and the sun gray value is one.

2)Edge Extraction

The edge extraction of binary images does not require complex operators. Here we use the sobel operator to convolve with the binary image, and the result is the edge of the image, which is shown in Fig. 2.

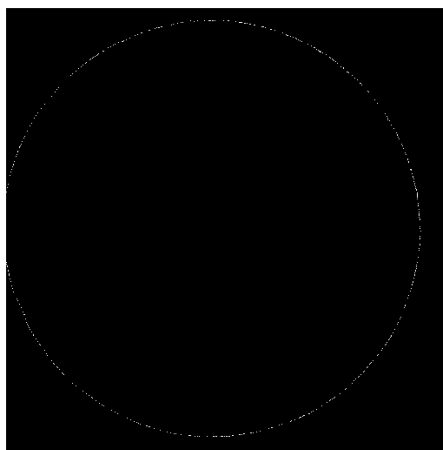


Fig. 2: Edge Extraction Diagram

3)Fitting the Image Contour to Find the Center and Radius

After obtaining the coordinates of each point on the edge of the image, in order to obtain the center and radius, a circle fitting is required. The basic idea is the least squares method.

The analytical equation of the circle is expressed as follows.

$$(x - x_c)^2 + (y - y_c)^2 = R^2$$

According to the idea of the least square's method, the following definitions are made.

$$f = \sum ((x - x_c)^2 + (y - y_c)^2 - R^2)^2$$

In the above formula, f is the sum of the squares of the distances. By letting the partial derivative

of f be zero, the variance is maximized, and the expression is expressed as follows.

$$\begin{cases} \frac{\partial f}{\partial x_c} = 0 \\ \frac{\partial f}{\partial y_c} = 0 \\ \frac{\partial f}{\partial R} = 0 \end{cases}$$

Solve the above formula, the result is as follows.

$$\begin{cases} x_c = u_c + \bar{x} \\ y_c = v_c + \bar{y} \\ R^2 = \sum ((x_i - x_c)^2 + (y_i - y_c)^2) \end{cases}$$

In the above formula, (x_c, y_c) is the center of the circle and R is the radius.

$$\begin{cases} u_c = \frac{S_{uuu}S_{uv} - S_{uuu}S_{vv} - S_{uvv}S_{vv} + S_{uv}S_{vvv}}{2(S_{uv}^2 - S_{uu}S_{vv})} \\ v_c = -\frac{-S_{uu}S_{uv} + S_{uuu}S_{uv} + S_{uv}S_{uvv} - S_{uu}S_{vvv}}{2(S_{uv}^2 - S_{uu}S_{vv})} \end{cases}$$

$$\begin{cases} S_{uuu} = \sum u_i^3 \\ S_{vvv} = \sum v_i^3 \\ S_{uu} = \sum u_i^2 \\ S_{vv} = \sum v_i^2 \\ S_{uv} = \sum u_i v_i \\ S_{uuv} = \sum u_i^2 v_i \\ S_{uvv} = \sum u_i v_i^2 \\ \begin{cases} u_i = x_i - \bar{x} \\ u_c = x_c - \bar{x} \\ v_i = y_i - \bar{y} \\ v_c = y_c - \bar{y} \end{cases} \end{cases}$$

4)Panning the Center of the Circle to the Center of the Image

In order to facilitate the subsequent operation, after acquiring the center of the circle and the radius, we remove the information outside the circle firstly, and then shift the center of the circle to the center of the image. The solar flare pattern at this time is symmetrical along the central axis, and the effect diagram is shown in Fig. 3.

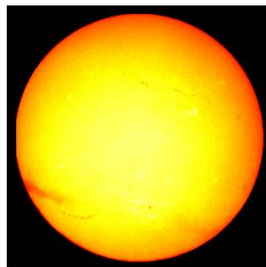


Fig. 3: Center shift effect diagram

5)Polar Coordinate Transformation

In order to obtain the dim curve of the image, it is necessary to first convert the image from a Cartesian coordinate system to a polar coordinate system. Direct conversion is not convenient, so we use the rotation sampling method to convert. First, select the gray value in one direction of the original image as the first column of the polar coordinate image, then rotate the whole image around the center by a small angle, and extract again in the same direction, and repeat this step several times. Until the entire image is turned over one week, the resulting data is a matrix and is the image in polar coordinates.

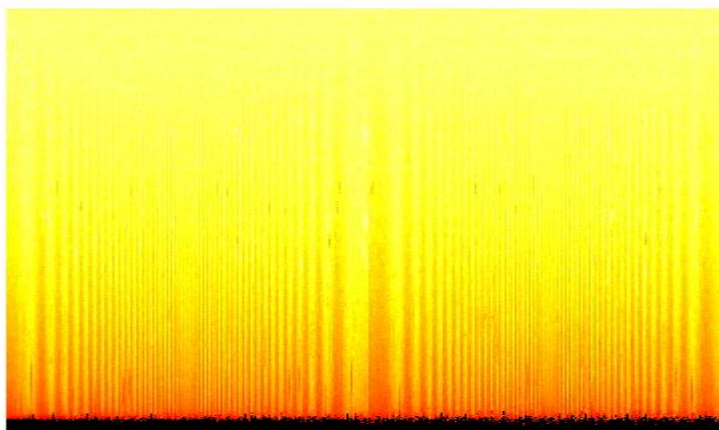


Fig. 4: Full sun image of the sun in polar coordinates

6)Getting the Dim Curve

The image in polar coordinates is divided into four quadrants, and they are averaged to obtain the average edge dim curves of the four quadrants.

The vertical coordinate pixel value of the dim curve of the edge reflects the brightness of each point different from the distance from the center of the circle. Therefore, the position of the steep drop indicates the edge of the sun, and the brightness drops sharply. The concave portion of the curve indicates a decrease in brightness, indicating that there is a cloud here. The convex portion of the curve indicates a sharp increase in brightness, indicating that there is a flare here.

3. Model Establishment

The basis of the C-means algorithm is the minimum error squared criterion. If N_i is the number of samples in the cluster Γ_i and m_i is the mean of these samples, and the expression is as follows.

$$m_i = \frac{1}{N} \sum_{y \in \Gamma_i} y$$

The sum of squared errors between each sample y in Γ_i and the mean m_i is added as follows.

$$J_e = \sum_{i=1}^c \sum_{y \in \Gamma_i} (y - m_i)^2$$

J_e is the squared error sum clustering criterion, which is a function of the sample set Y and the category set Ω . For different clusters, the error square J_e is also different, and the smallest error square J_e is the optimal result. The implementation of J_e minimization uses iterative methods to continually adjust the categories of samples. After this adjustment, the mean values of the two types are expressed as follows.

$$\tilde{m}_k = m_j + \frac{m_k - y}{N_k - 1}$$

$$\tilde{m}_j = m_j + \frac{y - m_j}{N_j + 1}$$

Correspondingly, the sum of the squared errors of the two types is also expressed as follows.

$$\tilde{J}_k = J_k - \frac{N_k (y - m_k)^2}{N_k - 1}$$

$$\tilde{J}_j = J_j + \frac{N_j (y - m_j)^2}{N_j + 1}$$

$$\frac{N_j (y - m_j)^2}{N_j + 1} < \frac{N_k (y - m_k)^2}{N_k - 1}$$

4. Cloud Grading

Based on the above model, we divide the cloud amount into five levels, and the names and characteristics are as follows.

Non-uniform cloud: Clouds are unevenly distributed and have very little cloud.

Non-uniform and thin cloud: Clouds are less unevenly distributed and have less cloud distribution.

Thin cloud in the central area: A small amount of clouds are mainly distributed in the central area.

Thick clouds in small area: Small areas have thick clouds.

Thick clouds in large area: Large areas have thick clouds.

5. Conclusion and Outlook

In this paper, the classification of the model is considered comprehensive, including not only the thickness information of the cloud layer, but also the position information and the change of

the thickness of the cloud layer with the position. At the same time, we can use the dim curve to identify whether there is a cloud. For the c-means dynamic clustering algorithm adopted, it has strong universality and is involved in the fields of classification and pattern recognition. And we use the three-dimensional c-means clustering algorithm, which can be extended to high-dimensional space to solve more complex problems, and also has higher precision.

References

1. G. Petschnigg, M. Agrawala, H. Hoppe, R. Szeliski, M. Cohen, and K. Toyama, "Digital Photography with Flash and No-Flash Image-Pairs," Proc. ACM Siggraph, 2004.
2. C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images," Proc. IEEE Int'l Computer Vision Conf., 1998.
3. Kaiming He, Jian Sun, Xiaoou Tang, Guided Image Filtering. IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 35, Issue 6, pp. 1397-1409, June 2013

