A kind of generalized fuzzy C-means clustering model and its applications in mining steel strip flatness signal

TANG CHENG-LONG, WANG SHI-GANG, XU WEI

Mechanical-Electrical Design and Knowledge-Discovery Institute
Shanghai Jiao Tong University
No.800 Dong Chuan Road, Min Hang District, Shanghai, 200240
P. R. CHINA
Tangchenglong@sjtu.edu.cn

Abstract: - In this paper, the intelligent techniques are utilized to enhance the quality control precision in the steel strip cold rolling production. Firstly, a new control scheme is proposed, establishing the classifier of the steel strip flatness signal is the basis of the scheme. The fuzzy clustering method is used to establish the classifier. Getting the high quality clustering prototypes is one of the key tasks. Secondly, a kind of new fuzzy clustering model, generalized fuzzy *C*-means clustering (GeFCM) model, is proposed and used as the mining tools in the real applications. The results, under the comparisons with the results obtained by the basic fuzzy clustering model, show the GeFCM is robust and efficient and it not only can get much better clustering prototypes, which are used as the classifier, but also can easily and effectively mine the outliers. It is very helpful in the steel strip flatness quality control system in one real cold rolling line. Finally, it is pointed out that the new model's efficiency is mainly due to the introduction of a set of adaptive degrees $\boldsymbol{w}_j$ (j=1…n, and n is the number the data objects) and an adaptive exponent $\boldsymbol{p}$ which jointly affect the clustering operations. In nature, the proposed GeFCM model is the generalized version of the existing fuzzy clustering models.

Keywords: - Steel strip flatness signal; Generalized fuzzy clustering; Outliers mining; Adaptive degrees; Adaptive exponent

1 Introduction

The technologies that cold rolling the steel strip to the desired quality have attracted considerable attention in the past decade due to the potential operational benefits obtainable through advanced mills and other related techniques [1, 2]. Especially, among the quality indices, the increasing flatness and transversal thickness profile requirements have to meet a never-ending challenge particularly to the performances of the automatic control systems.

A newly building cold tandem rolling mill in BAOSTEEL aims at producing the high-grade steel strip which is specially supplied to the automobile manufactures. The thickness of the strip is reduced to the specified value by the tandem 5 stands and meanwhile the flatness and transversal thickness deviations of the final products are strictly limited. As it is well known, cold rolling thin strip is a typically complex process which is characterized with the terms of "multi variables, nonlinear and strong co-couple". The commonly-equipped back forward control system encounters difficulties to further improve the quality control precision. The authors took a task to develop a new automatic flatness and transversal thickness co-control system for this mill, the goal is that large amounts of products can be fabricated with the closer tolerances.

For many existing tandem cold rolling mills, the flatness quality control systems are normally equipped. A flatness measurement roll with multi-sensor is installed right behind the final stand and outputs the actual flatness signal periodically. These traditional control systems have been proven to have difficulties to meet the strict quality requirements for the following disadvantages: One, the control accuracy is beyond expectations. The flatness signal measured are normally treated by the common polynomial fitting method, most often, four orders polynomial is used, that means, a row of flatness signal is transferred to a vector of 4 elements (the constant term is neglected). 4-order polynomial fitting is a bit simple and it loses much available information of the real signal and as a result the control accuracy becomes weak. Two, the computation is heavily time consumed. In the simulation phase, it is found that the output results of the hydraulic actuators were got by the iterative way and it even takes several seconds. This also explains why only the simple polynomial methods can be used, it will take long time for computation if the raw signal is treated with more complex ways. It is not permitted in the real production. Three, the abnormal flatness signal cannot be detected by the current system. Abnormal signal reflect rare aspects of the production process. It is helpful in analyzing the producing process.

Comparatively, as an alternative, Artificial Intelligence (AI) technologies in some extent can ease the difficulties [3, 4]. In this research, the AI scheme as a new approach replaces the traditional control scheme and is introduced to the new control system. Fig. 1 plots the overview of the artificial intelligent control system. The upper part of the Fig. 1 plots the cold rolling environment, the tandem cold rolling mill includes five stands and each stands have six rolls, two back-up rolls with the biggest diameter, two work rolls which directly contact the strip and two immediate rolls which are installed between the back-up roll and the work roll respectively. The hydraulic actuators are equipped to change the rolling process. The actions include leveling of the back-up rolls, bending of work rolls and immediate rolls, shifting of immediate rolls, that means, the immediate rolls can be axially shifted to change the contact lengths both between the immediate roll and work roll and between the immediate roll and back-up roll. The lower part of the Fig. 1 shows the thought of the new control approach based on the AI scheme. The measured flatness signal are pre-treated (features abstracting) and sent to the computing engine in a stream manner, the subsequent tasks include, establishing the classifiers and classifying the real time flatness stream signal based on the classifiers. Classifiers are updated with a certain time interval for the signal appears in the stream manner. In establishing the classifier, a certain number of the signal are sampled and worked as in the learning data set. Clustering, as one of the most important and frequently used data analysis techniques, is used to establish the classifiers. The basic thought of clustering is taken, that means, the flatness stream signal can be grouped into several groups, each group has its own prototype and the signal inside one cluster are treated to be the respective prototype, then all prototypes can be used as a classifier. Importantly, a certain number of the abnormal signal are unavoidably contained in the data set and they normally contain much interesting information about the strip rolling process itself, the abnormal flatness needs to be detected and recorded for the analysis of the production process. In the classifying phase, the relationship between the patterns classified and the compensating outputs of the hydraulic actuators can be established in an offline manner. The main advantage is the long time computation can be avoided and done in the offline manner. The online work becomes to recognize the patterns of the flatness signal, and then the hydraulic actuator can do the standard patterns. The function between the standard patterns and the outputs of the hydraulic actuators can be set up with the more precise models. And moreover, the AI approach, often as the tool of data analysis, can discriminate the outliers in the data set.

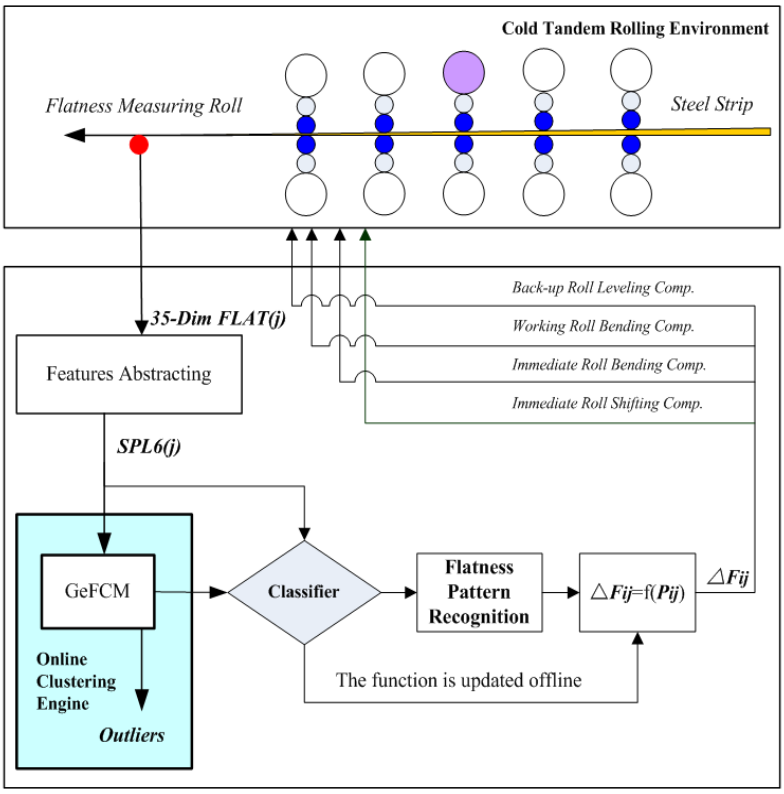

Fig. 1: Overview of the artificial intelligent control system of the cold rolling mill

The steel strip that the line can produce are divided into several different groups according to the dimensions and the grades of the products, a classifier corresponds to a group. The establishment of classifiers is the basis of the AI control approach. This paper only discusses the online clustering engine, seeing the bold frame in the Fig. 1. In the beginning of our solution, the basic objective function-based fuzzy clustering model, FCM, was utilized. It was found that the FCM can cluster the flatness signal with a certain precision, but this model had two deficiencies, one, the classifiers established by the FCM is not robust enough, it is sensitive to the abnormal signal in the data set. Two, the basic FCM model almost cannot directly detect the existence of the abnormal flatness signal. Even some modified FCM models can do the detection of the outliers, but these models generally treat the abnormal signal as the by-products and it cannot meet the requirements of effectively mining the rich information of the abnormal flatness signal.

Based on the above description, the clustering prototypes are used to be the classifiers, so getting the high quality of the clustering prototypes becomes one of the key tasks in the actual application. This paper proposes a new kind of FCM model; it is named as the Generalized FCM, shortly GeFCM, and the GeFCM is utilized in the actual application. The effects are verified to be good. The GeFCM can obtain much better clustering quality than the basic FCMs do because of the robustness of the GeFCM, needless to say, that the GeFCM can fulfill the common fuzzy clustering operations just like the FCMs do, meanwhile the GeFCM is verified to be able to easily detect the abnormal signal and effectively mine the useful information about the abnormal signal, because the cluster properties of the abnormal signal are also obtained in the clustering process. This paper summarizes these findings.

The remainder of the paper is organized as follows: Section 2 provides the background information on the cold rolling process and steel strip flatness signal. Second 3 addresses the theoretical part of this paper, firstly, the basic FCM model is introduced and the related work on the clustering and outliers mining techniques is reviewed, and then the new GeFCM and how the GeFCM works and detects the outlier are proposed. Section 4 is about the application, the GeFCM is used to mine the steel strip flatness signal, and the results are compared with the results obtained by the basic FCM model. Section 5 summarizes the research results.

2 Background of application: Steel strip flatness signal

The developed new control frame was applied and testified in the newly building tandem cold rolling mill of BAOSTEEL. The learning flatness signal is collected by the flatness measurement roll. The roll is a type of multi-sensor device which has maximum 35 channels corresponding to the maximum width of the strip. The device can output a row of flatness values simultaneously. About the flatness signal, it is necessary to introduce a bit more as follows, the measured values are the absolute values, sees the top part of Fig. 2, and in the actual production, a standard curve will be given in advance and normally the curve is plotted with the form of bars, sees the immediate part of the Fig. 2. The objective signal in actual production are relative values, that is to say, the difference of the absolute values and the standard values, sees the bottom part of the Fig. 2. The ideal phenomenon is the relative values are zeros, but it is generally impossible, that is why it needs to be controlled.

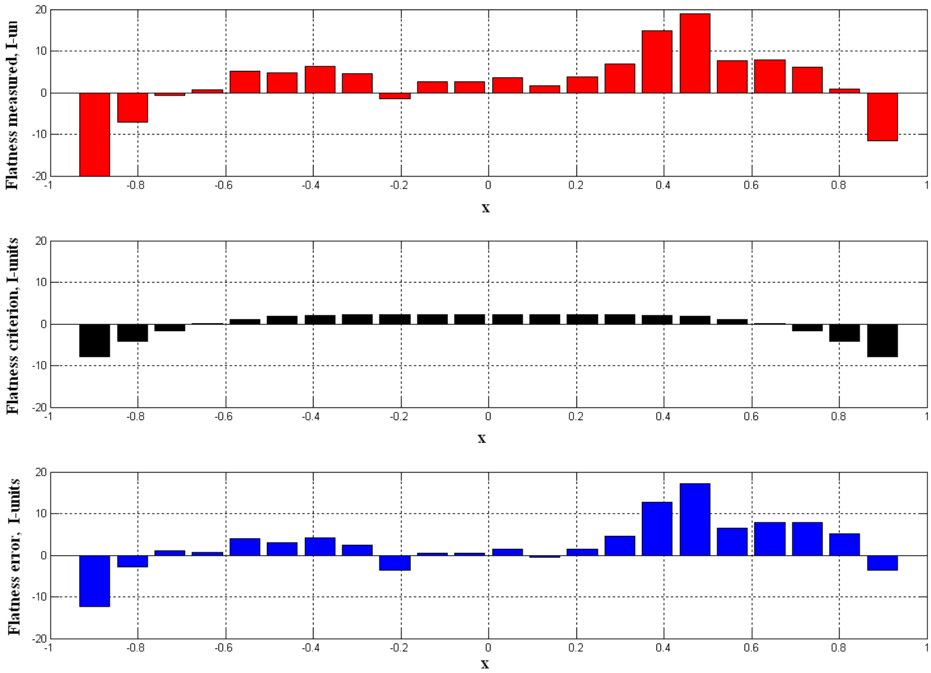

Fig. 2: Flatness bar display, including signal measured, standard and relative

The 6-order Legendre polynomial is used to fit the measured flatness signal, 35-dimension FLAT, seeing the Fig. 1. Every row of the 35-dimension flatness signal is transformed to a 6-dimension coefficients vector with the Legendre polynomial fitting method. The Legendre polynomial methods can be easily found in the mathematical literatures, here the following Eqns. (1, 2) are given directly.

$$H=\sum_{i=0}^{n} L_i \phi_i(x) \tag{1}$$

$$\phi_i(x)=\frac{1}{2^i i!}\frac{d^i}{dx^i}\left[\left(x^2-1\right)^i\right], i=0,1,2,\ldots \tag{2}$$

H: The reconstructing curve of the flatness signal.

$\boldsymbol{x}$: Normalized abscissa from -1 to 1, in the width direction of strip.

L_i: Legendre Coefficients, i=0, 1, 2…

$\boldsymbol{\phi}_i(x)$: The basic functions, i=0, 1, 2… , any two basic functions are orthogonal.

In the Eqn. (1), six coefficients L_j, j =1…6, are the research objectives and they can be rewritten in the form of a vector, which is denoted as $L=[L_1, L_2, L_3, L_4, L_5, L_6]$.

3 Mathematical Model: Generalized fuzzy C-means clustering model

3.1 The Fuzzy C-means clustering model

Clustering is the one of several important tools in modern data analysis. The general philosophy of clustering is to divide a data set into several homogenous groups based on the similarity or the dissimilarity. In such cases, the objects in the same group are tend to be as similar as possible to each other while the objects in different groups are as dissimilar as possible. There are crisp and fuzzy clustering models, among them, the fuzzy C-means clustering model, briefly FCM [5], has been successfully employed in a wide variety of fields. The FCM model is essentially an extension of the crisp c-means model where the notion of the degree of fuzziness, taking values not less than one, is introduced.

The objective function of the FCM model is written as J_{FCM} and its form is as the Eqn. (3) and a constraint condition is given, the Eqn. (4).

$$\boldsymbol{J}_{FCM}(\boldsymbol{X},\boldsymbol{U},\boldsymbol{V})=\sum_{j=1}^{n}\sum_{i=1}^{c}\boldsymbol{u}_{ij}^{q}\boldsymbol{d}_{ij}^{2} \tag{3}$$

$$s.t. \qquad \sum_{j=1}^{n}\boldsymbol{u}_{ij}>0, \qquad \sum_{i=1}^{c}\boldsymbol{u}_{ij}=1 \tag{4}$$

$\boldsymbol{X}$ is the data set, $\boldsymbol{U}$ is the fuzzy partition matrix and $\boldsymbol{V}$ is the prototype matrix, n is the number of the objects and c is the number of clusters, $\boldsymbol{d}_{ij}$ is the distance between the object $\boldsymbol{x}_j$ and the cluster $\boldsymbol{v}_i$, $\boldsymbol{u}_{ij}$ □ [0, 1] is the membership degree of the object $\boldsymbol{x}_j$ to the cluster v_i . The parameter q is the fuzzy exponent.

When solving the optimal value of the J_{FCM}, two update equations are needed, the Eqn. (5) is to update the fuzzy partition degrees and the Eqn. (6) is to update the prototypes.

$$\boldsymbol{u}_{ij}=\frac{\boldsymbol{d}_{ij}^{-\frac{2}{q-1}}}{\sum_{i=1}^{c}\boldsymbol{d}_{ij}^{-\frac{2}{q-1}}} \tag{5}$$

$$\boldsymbol{v}_i=\frac{\sum_{j=1}^{n}\boldsymbol{u}_{ij}^{q}\boldsymbol{x}_j}{\sum_{j=1}^{n}\boldsymbol{u}_{ij}^{q}} \tag{6}$$

For the FCM, the distance measure is basically Euclidean distance, there are also several FCM variants with the some others distance measures.

3.2 Related work

This section reviews the relate work in fuzzy clustering and outlier mining fields.

3.2.1 Fuzzy clustering models

Although the FCM and its variants are proven to be very helpful in the pattern recognition, data mining, image process etc. They have also been proven to have some disadvantages. The clustering results are badly influenced by the outliers in the data set and the cluster quality sometimes is beyond expectation. Much research work has been done in order to improve the performances of the FCM models. In particular, in the aspect of the new development of the fuzzy clustering algorithms, there are a lot of literatures which discussed the “weighting” methods. The typical weighting approaches are mainly around prototypes [6], features [7, 8, 9, 10 and 11], fuzzy membership degree [12], as well as distance [13]. Among them, the feature weighted approach is more widely studied and applied in order to obtain better clustering quality. The reason that we specially mention the weighting approaches is , in this paper, a kind of generalize fuzzy clustering model is proposed by introducing a set of parameters named as the adaptive degrees and the set of the adaptive degrees has the similar role as the weight. It is pointed out that the concept of the adaptive degrees means the contribution that each data point does to the objective function is adapted and the set of adaptive degrees is updated in each iterative step. So it is clear that the adaptive degree of this paper holds completely different role and meaning from all “weights” listed in these literatures [6-13].

3.2.2 Outlier mining models

Detecting outliers in a large set of data objects is a major data mining task aiming at finding different mechanisms responsible for different groups of objects in a data set. In many cases, such as abnormal fluctuation of the complex industrial process, rare events often are more interesting than the normal events. In the field of outlier mining, the outliers themselves become the “focus”, which is different from the case of the fuzzy clustering discussed in the above where the outliers probably are treated as the by-products.

Outlier mining task generally can be described as a "Top-k" principle [14]: given a data set which contains n data objects and given the expected number of outliers, k, one or several methods are chosen to find the first k points which are the most significant anomalies or inconsistencies. Outlier mining issue usually begins from the basic problem of what kind of points is defined as outliers.

Literature [14] groups the widely used outlier mining methods into four categories, statistics-based, distance-based, density-based, feature deviation-based and the existing outlier mining methods mainly use two strategies: One is a binary decision of whether or not a data point is an outlier, such as the statistics-based, distance-based methods. Two is to assign a outlying degree to each data point, for example, “outlier factor” is a value characterizing each data point in “how much” this data point is an outlier, such as the density-based methods; Among the four major methods, the density-based [15, 16 , 17] and distance-based [18, 19] methods are more studied and applied. About the density-based methods, the typical representative is the LOF method proposed by Breuning [15]. The LOF calculates an outlying degree for each data and then determines k outliers in accordance with the “Top-k” principle. About the distance-based methods, Amol and Srinivasan et al [18] proposed a distance-based fast outliers detecting method suitable for the high-dimensional space. Weng and Shen et al [19] proposed a method to mine the outliers in the time series data set, they computed the distance using the expanded Frobenius norm then to determine the existence of outliers.

There are also clustering techniques used to mine the outliers, for example, PCM [5] and NC [20] algorithms. However, since the main objective of the existing clustering method is “cluster”, they are not optimized for outlier detection. The outliers detected by the clustering models are by-products. Furthermore, in many cases, the outlier detection criteria are implicit and cannot easily be inferred by the clustering procedure.

Briefly, the existing outlier mining models have these disadvantages: First of all, the information that they mine is probably not enough, for example, if an outlier is only detected to be an “outlier” by a method, the abundant information that the outlier contains apparently cannot be mined. Second, the physical meaning of the outlier sometimes is difficult to be explained, for example, in the case of the LOF, it defines the outliers in accordance with the local density in the sparse space, and it is sometimes unreasonable. Third, the computational efficiency is a problem. For example, the most widely used density-based methods have to scan the entire data set when calculating the parameters concerning each data point, the computation takes long time.

Generally speaking, both the existing fuzzy clustering models and the outlier mining models have several disadvantages. The new models are in demands to overcome the disadvantages in these two issues in the theoretical researches and engineering applications, especially, in the case of the two issues are integrated as a whole, which is termed as "clustering meanwhile mining outliers". The contribution of this paper is to propose a new model to solve the two issues as a whole and the applications show this model works well.

3.3 Generalized fuzzy C-means clustering model: GeFCM

3.3.1 The objective function of the GeFCM

The main improvement of the new model presented in this paper is that it can treat each data object discriminatively, not equally any more. The new model is named as Generalized FCM model (GeFCM) because the new fuzzy clustering model is proven to be the generalized form of the basic fuzzy clustering models. This subsection gives the detailed reasoning process of the GeFCM.

The new objective function $\boldsymbol{J_{GeFCM}}$ is proposed as the Eqn. (7) with two constraints, seeing the Eqn. (8).

$$\boldsymbol{J_{GeFCM}}\left(\boldsymbol{X,U,V,W}\right)=\sum_{j=1}^{n}\boldsymbol{w}_j^p\sum_{i=1}^{c}\boldsymbol{u}_{ij}^q\boldsymbol{d}_{ij}^2 \tag{7}$$

$$\begin{aligned} s.t. \quad & \sum_{j=1}^{n}\boldsymbol{u}_{ij}>0, \qquad \sum_{i=1}^{c}\boldsymbol{u}_{ij}=1 \\ & \prod_{j=1}^{n}\boldsymbol{w}_j=1 \end{aligned} \tag{8}$$

The $\boldsymbol{J_{GeFCM}}$ is developed from the J_{FCM}, so the same symbols appeared in the Eqn. (2) have the same definitions and the constrain about the fuzzy membership degree $\boldsymbol{u}_{ij}$ is still valid, and besides, there are two kind of new introduced parameters and a new constrain, $\boldsymbol{w}_j$ is named as the adaptive degree of object $\boldsymbol{x}_j$ and it reflects the influence of object $\boldsymbol{x}_j$ on the objective function. The parameter $\boldsymbol{p}$ is name as the adaptive exponent and it is a preselected real-valued parameter and used to adjust the influence of the adaptive degree. The additional constraint shows the product of all adaptive degrees is subject to be zero.

Just like the basic FCM, the optimal value of the $\boldsymbol{J_{GeFCM}}$ is also solved by the Alternative Optimization scheme. The AO scheme repeats the iteration process until the objective function is converged to the optimal value. For the GeFCM, each adaptive degree is also updated. It is necessary to reason the expression of $\boldsymbol{w}_j$。

The minimization of the objective function represents a nonlinear optimization problem that is usually solved by means of Lagrange Multipliers Method. When the Lagrange Multiplier Method is used to solve the optimal $\boldsymbol{J_{GeFCM}}$, two kinds of constraints are taken into account. The optimization function can be written as following Eqn. (9) based on the Eqns. (7) and (8):

$$\boldsymbol{J}_{\boldsymbol{GeFCM},\boldsymbol{\phi}_1,\boldsymbol{\phi}_2}\left(\boldsymbol{X,U,V,W}\right)=\sum_{j=1}^{n}\boldsymbol{w}_j^p\sum_{i=1}^{c}\boldsymbol{u}_{ij}^q\boldsymbol{d}_{ij}^2+\boldsymbol{\phi}_1\left(\prod_{j=1}^{n}\boldsymbol{w}_j-1\right)+\boldsymbol{\phi}_2\left(\sum_{i=1}^{c}\boldsymbol{u}_{ij}-1\right) \tag{9}$$

Where, φ_1 and φ_2 are two Lagrange operators corresponding to two constraints, the adaptive degree w_j and the fuzzy membership degree u_{ij}.

The w_j and u_{ij} are partially differentiated respectively, there are,

$$\begin{cases} \dfrac{\partial J_{GeFCM,\phi_1,\phi_2}}{\partial w_j} = p\, w_j^{p-1} \sum_{i=1}^{c} u_{ij}^q d_{ij}^2 + \phi_1 \left(\prod_{\substack{k=1 \\ k \neq j}}^{n} w_k \right) \\ \dfrac{\partial J_{GeFCM,\phi_1,\phi_2}}{\partial u_{ij}} = q\, w_j^p u_{ij}^{q-1} d_{ij}^2 + \phi_2 \end{cases} \tag{10}$$

Let two partial derivatives be 0, the expressions of the two Lagrange operators, φ_1 and φ_2, can be obtained, sees the Eqn. (11) and (12) respectively.

$$\phi_1 = \frac{-p\, w_j^{p-1} \sum_{i=1}^{c} u_{ij}^q d_{ij}^2}{\prod_{\substack{k=1 \\ k \neq j}}^{n} w_k} \tag{11}$$

$$\phi_2 = -q\, w_j^p u_{ij}^{q-1} d_{ij}^2 \tag{12}$$

3.3.2 Adaptive degree update equation

The Eqn. (11) is transformed furthermore and both sides multiplied by w_j, the Eqn. (13) is obtained. Noticing the constraint of all adaptive degrees, the operator φ_1 can be got as the Eqn. (14). The expression of the adaptive degree w_j is the Eqn. (15).

$$\phi_1 \left(\prod_{\substack{k=1 \\ k \neq j}}^{n} w_k \right) w_j = \left(-p\, w_j^{p-1} \sum_{i=1}^{c} u_{ij}^q d_{ij}^2 \right) w_j \tag{13}$$

$$\phi_1 = -p\, w_j^p \sum_{i=1}^{c} u_{ij}^q d_{ij}^2 \tag{14}$$

$$w_j = \left[\frac{-\phi_1}{p \sum_{i=1}^{c} u_{ij}^q d_{ij}^2} \right]^{\frac{1}{p}} \tag{15}$$

On the basis of the Eqn. (14), considering all data points, for all data objects, multiplying both sides respectively, then the Eqn. (16) is obtained. Transforming the Eqn. (16) furthermore and noting the constraint of the adaptive degrees again, the Eqns. (17-19) can be reasoned. On the basis of the Eqn. (15) and the Eqn. (19), the expression of the adaptive degree w_j without the Lagrange operator φ_1 is given as the following Eqn. (20).

$$\phi_1^n = \prod_{j=1}^{n} \left[-p\, w_j^p \sum_{i=1}^{c} u_{ij}^q d_{ij}^2 \right] \tag{16}$$

$$\phi_1^n = (-p)^n \left(\prod_{j=1}^{n} w_j \right)^p \prod_{j=1}^{n} \left[\sum_{i=1}^{c} u_{ij}^q d_{ij}^2 \right] \tag{17}$$

$$\phi_1^n = (-p)^n \prod_{j=1}^{n} \left[\sum_{i=1}^{c} u_{ij}^q d_{ij}^2 \right] \tag{18}$$

$$\phi_1 = -p \left[\prod_{j=1}^{n} \left(\sum_{i=1}^{c} u_{ij}^q d_{ij}^2 \right) \right]^{\frac{1}{n}} \tag{19}$$

$$w_j = \left[\frac{\left[\prod_{j=1}^{n} \left(\sum_{i=1}^{c} u_{ij}^q d_{ij}^2 \right) \right]^{\frac{1}{n}}}{\sum_{i=1}^{c} u_{ij}^q d_{ij}^2} \right]^{\frac{1}{p}} \tag{20}$$

The Eqn. (20) is used to update the adaptive degree of each data object in each iterative step.

3.3.3 Fuzzy membership degree update equation

The expression of the fuzzy membership degree $\boldsymbol{u}_{ij}$ can be obtained by transforming the Eqn. (12), subsequently; the Eqn. (21) is obtained,

$$\boldsymbol{u}_{ij} = \left(\frac{\phi_2}{-q\boldsymbol{w}_j^p} \right)^{\frac{1}{q-1}} \boldsymbol{d}_{ij}^{-\frac{2}{q-1}} \tag{21}$$

Considering the constraint of the fuzzy membership degrees, the following Eqn. (22) can be obtained on the basis of the Eqn. (21),

$$\sum_{i=1}^{c} \boldsymbol{u}_{ij} = \left(\frac{\phi_2}{-q\boldsymbol{w}_j^p} \right)^{\frac{1}{q-1}} \sum_{i=1}^{c} \boldsymbol{d}_{ij}^{-\frac{2}{q-1}} \tag{22}$$

Transforming Eqn. (22) furthermore and noting the fact that the left side of the Eqn. (22) is equal to 1, the Eqn. **(**23**)** is obtained as follow.

$$\left(\frac{\phi_2}{-q\boldsymbol{w}_j^p} \right)^{\frac{1}{q-1}} = \frac{1}{\sum_{i=1}^{c} \boldsymbol{d}_{ij}^{-\frac{2}{q-1}}} \tag{23}$$

The expression of the fuzzy membership degree $\boldsymbol{u}_{ij}$ without the Lagrange operator φ_2 can be obtained based on the Eqns. (21) and (23), seeing the Eqn. (24).

$$\boldsymbol{u}_{ij} = \frac{\boldsymbol{d}_{ij}^{-\frac{2}{q-1}}}{\sum_{i=1}^{c} \boldsymbol{d}_{ij}^{-\frac{2}{q-1}}} \tag{24}$$

It can be seen that the fuzzy membership degree $\boldsymbol{u}_{ij}$ holds the same expression under both the FCM and the GeFCM by comparing the Eqn. (24) and (3) .This shows that the definition of fuzzy membership degree is not changed in the GeFCM. It only changes the "different" contribution of individual data point to the objective function with the form of the adaptive degree.

3.3.4 Cluster prototypes update equation

The cluster prototype is given directly as follow by referring the prototype update equation of the basic **FCM,** seeing the Eqn. (25). Here $\overline{v}_i$ is the i-th new cluster prototype.

$$\overline{\boldsymbol{v}}_{\boldsymbol{i}} = \frac{\sum_{j=1}^{n} \overline{\boldsymbol{u}}_{ij}{}^{q} \boldsymbol{x}_{\boldsymbol{j}}}{\sum_{j=1}^{n} \overline{\boldsymbol{u}}_{ij}{}^{q}} \tag{25}$$

Where, $\overline{\boldsymbol{u}}_{ij}$ is called as the adaptive fuzzy membership degree and its calculating equation is as follow,

$$\overline{\boldsymbol{u}}_{ij}^{q} = \boldsymbol{w}_j \boldsymbol{u}_{ij}^{q} \tag{26}$$

Based on the Eqns. (25) and (26), the new cluster prototype equation under the adaptive approach can be rewritten as the Eqn. (27),

$$\overline{\boldsymbol{v}}_{\boldsymbol{i}} = \frac{\sum_{j=1}^{n} \boldsymbol{w}_j \boldsymbol{u}_{ij}^{q} \boldsymbol{x}_j}{\sum_{j=1}^{n} \boldsymbol{w}_j \boldsymbol{u}_{ij}^{q}} \tag{27}$$

It can be found that the cluster prototypes are the functions of $\boldsymbol{u}_{ij}$, m and $\boldsymbol{X}$ for the **FCM**, while for the **GeFCM**, besides $\boldsymbol{u}_{ij}$, m and $\boldsymbol{X}$, a new set of parameters $\boldsymbol{w}_j$ is introduced in the update equation, and the parameter p is used to adjust the set of $\boldsymbol{w}_j$. They jointly affect the final clustering results. So the combination of the set of $\boldsymbol{w}_j$ and p works as a new channel. So, getting better clustering performances becomes possible and reasonable.

3.4 Outliers detected by the GeFCM

The GeFCM model, just as announced before, can obtain much better clustering quality meanwhile it can easily and effectively detect the existence of the outliers in the data set. When the GeFCM is used to mine the outliers, the definition of outliers will be given in advance. In this paper, the fuzzy distance between the data point and the prototype matrix is used to define the outlier.

Seeing the Eqn. (7), $\sum_{i=1}^{c} \boldsymbol{u}_{ij}^{q} \boldsymbol{d}_{ij}^{2}$ is named as the fuzzy square distance of the data $\boldsymbol{x}_j$. The term expresses a kind of fuzzy distance relationship between the data $\boldsymbol{x}_j$ and the prototype matrix. The greater the fuzzy square distance is, the bigger possibility the data $\boldsymbol{x}_j$ is an outlier. When the $\boldsymbol{J}_{GeFCM}$ is converged to its optimal value, each data point gets its own fuzzy square distance. All $\boldsymbol{n}$ values of the fuzzy square distance are rearranged with the order from the maximum to the minimum. $\boldsymbol{k}$ data points corresponding to the largest $\boldsymbol{k}$ values are defined to be the outliers. Furthermore, transforming the Eqn. (20), clearly, the Eqn. (28) is obtained as follow.

$$\boldsymbol{w}_j^p \sum_{i=1}^{c} \boldsymbol{u}_{ij}^{q} \boldsymbol{d}_{ij}^{2} = \prod_{j=1}^{n} \left[\left(\sum_{i=1}^{c} \boldsymbol{u}_{ij}^{q} \boldsymbol{d}_{ij}^{2} \right)^{\frac{1}{n}} \right] \tag{28}$$

When the objective function is converged to its optimum, the right side of the Eqn. (28) is surely a confirmed constant, even it is unnecessary to know its exact value. It tells a fact that, for any data $\boldsymbol{x}_j$, the product of $\boldsymbol{w}_j^p$ and $\sum_{i=1}^{c} \boldsymbol{u}_{ij}^{q} \boldsymbol{d}_{ij}^{2}$ is equal to each other. The largest $\boldsymbol{k}$ fuzzy square distance values correspond to the $\boldsymbol{k}$ smallest $\boldsymbol{w}_j^p$. Therefore, the outliers can be detected in accordance with the values of $\boldsymbol{w}_j^p$. Here a term of the outlying degree of the data $\boldsymbol{x}_j$ is defined as the following Eqn. (29).

Definition:

$$\boldsymbol{O}_{GFCM}\left(\boldsymbol{x}_j\right) = \frac{\sum_{i=1}^{c} \boldsymbol{u}_{ij}^{q} \boldsymbol{d}_{ij}^{2}}{max\left[\sum_{i=1}^{c} \boldsymbol{u}_{ij}^{q} \boldsymbol{d}_{ij}^{2}\right]} = \frac{min\left[\boldsymbol{w}_j^p\right]}{\boldsymbol{w}_j^p} \tag{29}$$

$\boldsymbol{O}_{GFCM}\left(\boldsymbol{x}_j\right)$ is named as the outlying degree of the data $\boldsymbol{x}_j$ and it is a normalized value (0, 1]. The bigger the $\boldsymbol{O}_{GFCM}\left(\boldsymbol{x}_j\right)$ is, the more likely the data $\boldsymbol{x}_j$ is an outlier. Obviously, the definition of the outlying degree of the data $\boldsymbol{x}_j$ expresses the global relationship between the data $\boldsymbol{x}_j$ and the whole data set.

The calculating steps of the GeFCM with the outliers detecting can be concluded as the Fig. 2, the upper frame is "clustering" and the low frame is "mining outliers". That is we name, "clustering meanwhile mining outliers"

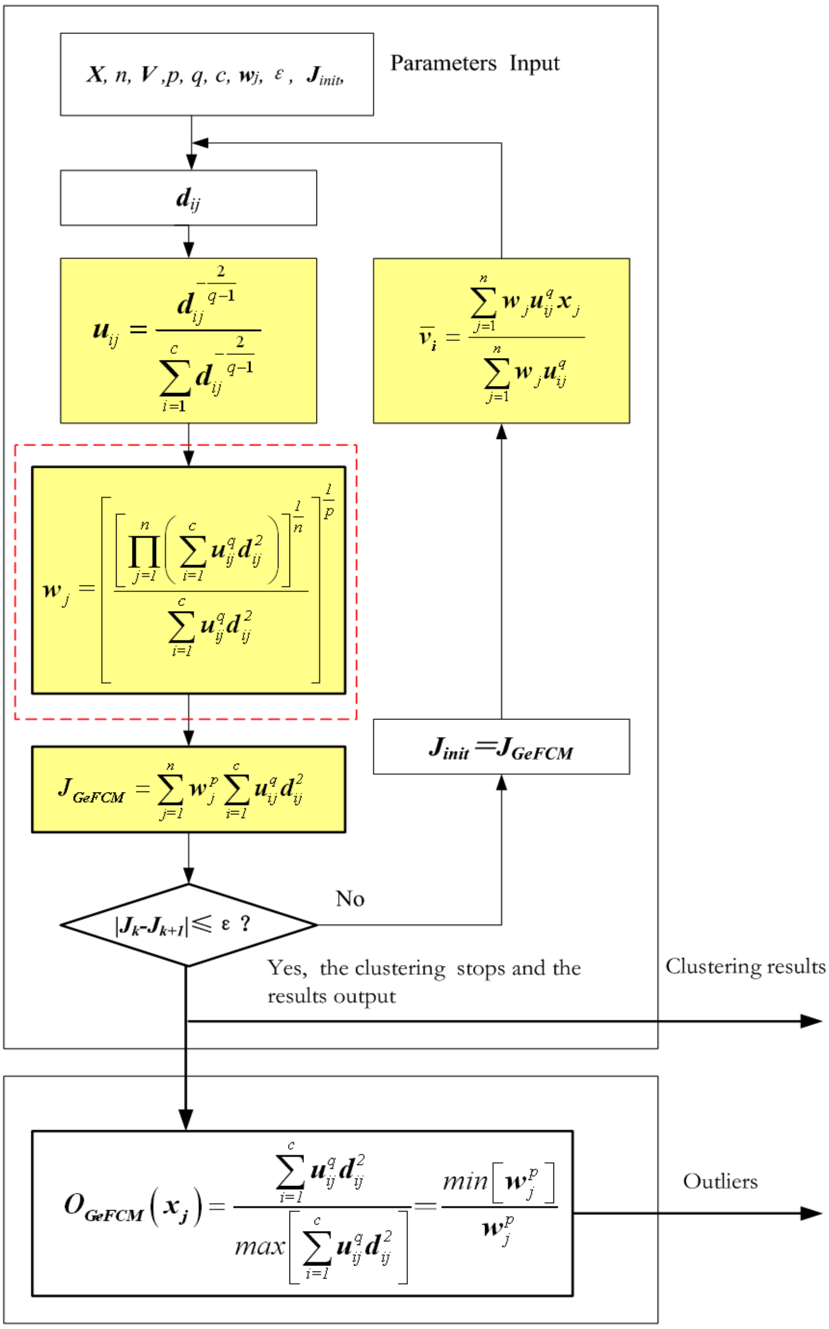

Fig. 3: The calculating flowchart of the GeFCM model with outliers mining

4 Applications: clustering the flatness signal with outliers

4.1 Data set preparation

The flatness signal of 6 coils is sampled and totally 3402 flatness signal with the fixed interval of 1 meter length are obtained. The original data set is a 35×3402 matrix. Legendre polynomial fitting method is adopted to reduce the dimensionality of the original data set. The detailed mathematical method has been described in the section 2. The order-6 Legendre coefficients are abstracted as the main features. As a result, a 6×3402 matrix is obtained and replaces the original data set to be analyzed. The data set is named as FL6L.

The FCM and the GeFCM are used respectively to cluster the FL6L under the same combinations of the number of clusters $\boldsymbol{c}$ and the fuzzy exponent $\boldsymbol{q}$. Moreover, the $\boldsymbol{q}$ gets its widely used value, it is 2 [21]. The c gets the integer values, typically 6, 9 and 12 following the need of the control system. In all cases, the convergence threshold value $\boldsymbol{\varepsilon}$ is set to be 0.001, and the adaptive exponent $\boldsymbol{p}$ needed by the GeFCM is set to be -1.25. On the theoretical side, the FCM is proved to be sensitive to initialization. So, the crisp $\boldsymbol{C}$-means clustering method is used to generate a set of prototypes to initialize both the FCM and the GeFCM with respect to different number of clusters in order to keep the comparisons are reasonable.

4.2 Cluster prototypes obtained by the FCM and the GeFCM

As a typical case, when the $\boldsymbol{c}$ is 9, the $\boldsymbol{p}$ is -1, the clustering results are given. Prt_GeFCM_9 and Prt_FCM_9 are the prototype matrices obtained by the GeFCM and FCM respectively and here 9 means the number of clusters is 9. The values of the Prt_GeFCM_9 and Prt_FCM_9 are given as follows. The clustering results when the c is 6 or 12 are omitted.

$$Prt\_GeFCM_9 = \begin{bmatrix} 1.8816 & -1.8566 & -3.3523 & 2.0102 & -2.8359 & 2.4705 \\ 2.6629 & -5.9817 & -2.1805 & 6.5885 & -5.7400 & 1.9679 \\ 2.1448 & -2.2811 & -1.4001 & 1.2487 & -4.8616 & 3.0917 \\ 2.2542 & -1.6326 & -2.6019 & -1.8914 & -2.6397 & 5.2910 \\ 1.8116 & -1.8899 & -1.9110 & 1.5286 & -2.8397 & 2.1228 \\ -0.1943 & -0.1768 & -0.5555 & 0.0068 & -1.5646 & 0.9658 \\ 2.0793 & -1.5388 & -3.1406 & 0.6964 & -2.1878 & 2.8415 \\ 4.4034 & 2.7148 & -2.6041 & -3.4510 & -5.8582 & 1.0541 \\ 2.0642 & -1.1276 & -2.6354 & -0.1749 & -2.9152 & 2.5618 \end{bmatrix}$$

$$Prt\_FCM_9 = \begin{bmatrix} 1.9276 & -1.8834 & -3.4252 & 2.1495 & -2.7603 & 2.2994 \\ 2.4647 & -5.6585 & -2.0286 & 6.2737 & -5.6111 & 1.9324 \\ 2.0804 & -2.4199 & -1.2683 & 1.4156 & -4.9540 & 3.0487 \\ 1.9381 & -1.3937 & -2.3268 & -1.4479 & -2.6407 & 4.4770 \\ 1.6802 & -1.8693 & -1.8510 & 1.6796 & -2.6479 & 1.9081 \\ -0.2877 & -0.2418 & -0.4756 & 0.1546 & -1.4822 & 0.8562 \\ 1.9625 & -1.4733 & -3.2423 & 0.7702 & -1.8580 & 2.6642 \\ 3.7351 & 1.40180 & -2.4674 & -1.8793 & -5.2342 & 1.2412 \\ 1.9579 & -1.4801 & -2.3387 & 0.6081 & -3.0916 & 2.3013 \end{bmatrix}$$

It is difficult to evaluate which model is much better only according to these two number matrices, Prt_FCM_9 and $Prt\text{-}GeFCM_9$. It is necessary to have others evaluating methods.

4.3 Comparisons of clustering performances

In the researches of the new clustering models, evaluating their performances mainly contains three aspects: One, clustering validity problem: for a given data set, how many clusters the data set are divided is much better. Two, clustering quality problem: if the number of clusters has been given in advance, which kind of partition is much better. Three, time complexity problem: if one new model is heavily time-consumed, even it can give good partition, it is still probably not a good method.

The clustering validity problem is usually discussed inside a particular clustering model, which means, for a given clustering model, the new cluster validity functions are developed and then the best number of the clusters can be known by calculating a series of values of the new validity functions under the different combinations of the parameters. The clustering quality problem mainly concerns the final clustering prototypes and the cluster-belonging of each data. Because the clustering is an unsupervised learning technique, the data set is unlabelled and the ideal output actually doesn't exist in advance. So, when comparing the clustering quality under the two models, the comparisons are relative, that is to say, one is better than the other. The problem of the time complexity of clustering is easier to understand, it is evaluated by recording the time consumed or the total iteration times. Especially, this problem needs to be paid more attention when treating the stream data set.

Before discussing the clustering quality, it is necessary pointed out that the number of clusters was predetermined. When the number of the standard patterns, see the Fig. 1, has been designed by calculating the Xie and Beni validity index [22]. It was found that the best three numbers of clusters are 6, 9 and 12. So the Classifier is designed to be one of the three cases, the number of the standard patterns is 6, 9, or 12. It can be selected according to the different control requirements.

Briefly in this application, the number of the clusters has been fixed in advance because of the design of the classifier. So the clustering quality problem and the time complexity problem are the two interesting topics.

In subsection 4.2, the FCM is selected to be the counterpart model and the number of clusters has been given in advance, the respective cluster prototypes are obtained by the GeFCM and FCM. The next work is to compare which result is much better and which model is more suitable for this actual application.

4.3.1 Clustering quality evaluation function

For the application of the flatness signal recognition, the clustering result, the clustering prototype matrix will be used as the classifiers, of the quality of the prototype matrix is important. Before comparing, it is necessary to propose a kind of evaluating way to compare the clustering performances of two different models when treating that kind of data set without apriority knowledge, just like the FL6L.

A kind of evaluation function, shortly *EVA*, is given in this paper. Briefly the *EVA* function is the ratio of the compactness and the separation, seeing the Eqn. (29). The compactness and separation functions are briefly written as *CMP* and *SPT*, the definitions of *CMP* and *SPT* are given in the Eqn. (30) and (31) respectively.

$$EVA = \frac{CMP}{SPT} \tag{29}$$

$$CMP = \frac{\sum_{j=1}^{n}\sum_{i=1}^{c} \boldsymbol{u}_{ij}^{q}\left(\boldsymbol{x}_j - \overline{\boldsymbol{v}}_i\right)^2}{n} \tag{30}$$

$$SPT = min_{i \neq k} \left\| \overline{\boldsymbol{v}}_i - \overline{\boldsymbol{v}}_k \right\|^2 \tag{31}$$

The *EVA* function is actually the validity function proposed by Xie and Beni [23]. In most applications, the Xie-Beni validity index is used to determine the best number of the clusters in one algorithm. Even the Xie-Beni validity index can also be utilized to evaluate the quality of the clustering for different algorithms. As mentioned at the beginning of this subsection, the number of clusters in this application has been given as the given parameters. In order to not make confused, the *EVA* function is used.

When the data set is clustered into c clusters, the *CMP* reflects the compactness, that is to say, the data points in the same cluster are as similar as possible. The *SPT* reflects the separation, that is to say, the data points in the different clusters are as dissimilar as possible. The compactness or the separation only gives the partial information of the clustering. The ratio of these two items is more reasonable to evaluate the clustering quality. In this paper, the *EVA* is used to be the final decision. That means, the smaller the *EVA* value is, the better the clustering quality is.

4.3.2 The comparison result and analysis

The experiments are conducted to compare the relative performance of the FCM and the GeFCM. Table 1 shows the EVA values obtained. Additionally, when discussing the time complexity of one model, two time indices are generally used, the time consumption and the total iterative times. Table 1 also gives these two indices for each calculating case. Five runs of each model were performed on the FL6L and the average values are used to be the final results.

Table 1: The *EVA* values of the FCM and the GeFCM when the *c* are 3, 6, and 9

	FCM			GeFCM			
c	*EVA*	Time*	Iterative Times	*EVA*	Time*	Iterative Times	*p*
6	1.6809	0.1111	12	1.8503	0.2699	16	-1.25
				1.6490↓	0.0998↓	**6↓**	-2
9	0.6947	0.4518	42	**0.6763**	0.4110	**31**	-1.25
12	1.0460	0.2816	19	**0.8983**	0.2513	**10**	-1.25

Notes: * the time is the average results of five calculations. The unit is second.

It can be seen that, when the p= -1.25, in all three cases, the c=9 and 12, the EVA values of the GeFCM are smaller than the values of the FCM. This means that the cluster prototypes obtained by GeFCM have better quality than the FCM. When the c=6, the FCM can get better clustering quality, for the FCM gets the value of the EVA 1.6809, and the FCM consumes less time and the total iterative times are less than that of the GeFCM. Here another experimental case is given, the value of the p is changed from -1.25 to -2 and the others parameters are unchanged, and then executing the GeFCM again, the time consumed now is reduced to 0.0998 seconds and the total iterative times are reduced from 16 to 6. It shows that the GeFCM can get much better quality meanwhile keeping the computation be low time consumption. In this process, selecting an available value of the p is very important. The p greatly affects the iterative speed.

As a brief, the GeFCM shows the same or better level of the time complexity by comparing with the FCM. The GeFCM is an efficient, not a heavy time-consumed model. It eliminates the suspicion that GeFCM is a complex model because more parameters are introduced to the calculation. Even in the beginning, the authors also held this suspicion.

4.4 The abnormal flatness signal detecting result

The outliers in the FL6L can be mined according the methods described in the subsection 3.4. Fig. 4 plots the distribution of he outliers. The dividing line distinguishes the outliers from normal objects. The y-coordinate of the line depends on this rule: k points are above the line, they are outliers. The remaining points are normal objects. According to the Fig. 4, it is a binary decision of one data point, an object is an outlier or not. For the GeFCM is a kind of clustering model, so it can meanwhile give the cluster properties of all data points, it can be seen in Fig. 5, The Fig. 5 covers all information that Fig. 4 can supply, and moreover, the cluster-property of each data is clear, sees “Clst.i”, i is from 1 to 9. This is the oblique advantage and interesting aspect of the GeFCM, not only it can do the clustering operation, but also, it meanwhile can detect the outliers. And it is exactly our need for the real applications.

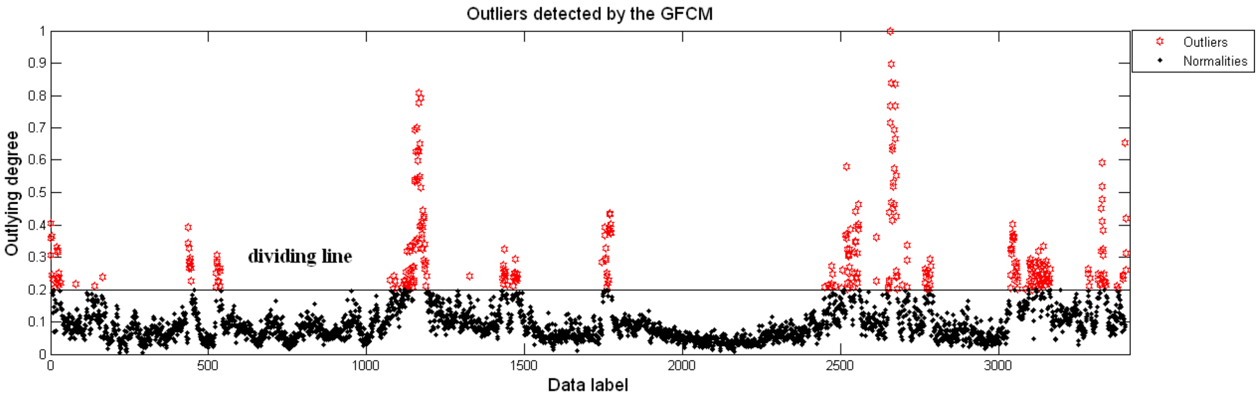

Fig. 4: The distribution of the abnormal flatness signal detected by the GeFCM

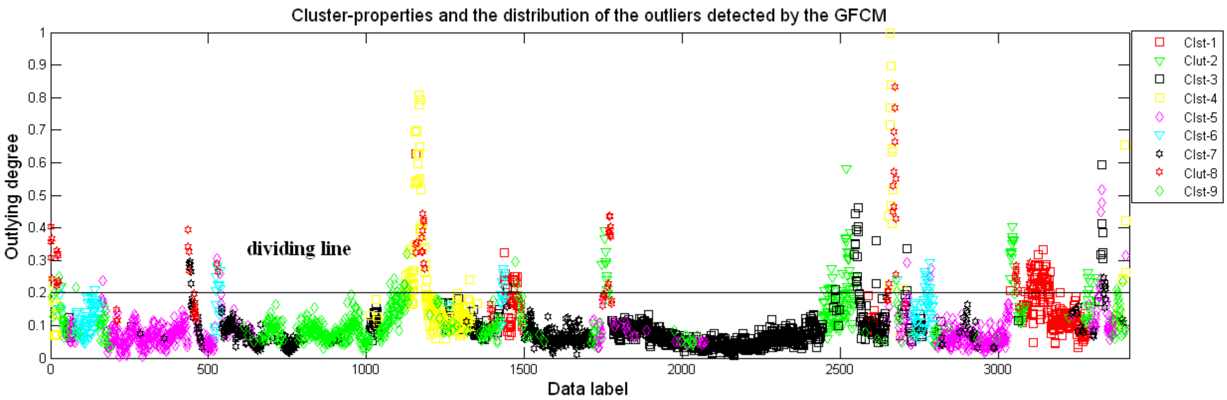

Fig. 5: The distribution of the clusters and abnormal flatness signal detected by the GeFCM

5 Conclusions

In this paper, the artificial intelligence technologies are introduced to the steel strip rolling fields. A control scheme based the AI techniques is proposed. This paper reports a novel generalized fuzzy C-means clustering model, shortly GeFCM, and proves that the GeFCM is more suitable for mining the steel strip flatness signal, “clustering meanwhile mining outliers”. The applications on the flatness signal of six real coils show the presented GeFCM holds the following advantages:

1) The GeFCM is more effective than the FCM and can get better clustering quality. The new model’s efficiency is mainly due to a set of adaptive degrees w_j which allow for treating the different influences of data objects on the clustering operation. And the adaptive exponent p can greatly affect the iterative speed. It can be understood that the GeFCM supplies a new channel to adjust the clustering operations. So, compared to the basic FCM, the GeFCM can get better performances is reasonable.

2) The GeFCM is more robust than the FCM. It can reduce or alleviate the bad influences of the abnormal data objects by the self-adaptation of the adaptive degree and the availably selected parameter

p.

3) The GeFCM can be used a tool to mine the outliers. It can easily detecting the existence of abnormal objects in the data set and the value of adaptive degree of each data point can be used as a kind of measure to detect the abnormal extent.

4) The GeFCM shows the same or better level of the time complexity when comparing with the FCM.

The core thought of the GeFCM is that the nature of individual data point in the data set is "different". The GeFCM overcomes the disadvantages that the existing fuzzy clustering models treat different data points "equally". Actually, considering a special case, let all adaptive degrees be 1, the additional constraint is still valid, but the GeFCM now becomes the basic FCM. It shows the FCM is a special case of the GeFCM and the GeFCM is the generalized form of the FCM. So, it is reasonable that the GeFCM can get much clustering performances than the FCM can do.

In the GeFCM, the role of the adaptive exponent p is very important. The future work includes the deep investigation of the adaptive exponent p and developing the methods to get the best values of the adaptive exponent.

Acknowledgement

This work is supported by National Natural Science Foundation of China under Grant 50875159.

References:

[1] D. Auzinger and G. Djumlija，et al, *Application of advanced technology packages for improved strip profile and flatness in hot strip mills*, [J], AISE Steel Technolog,Vol.30, 2003, pp.17-21.

[2] M. Jalali and A. Wolf, *Advanced flatness in EKO's tandem cold mill*, [J]. La Revue de Metallurgic-CIT, Vol.10, 2003, pp. 987-992.

[3] Minsuk Shim, Kwang Y. Lee, Dae-Sik Lee, *Intelligent controller design for the flatness control in a cold rolling process*, [C], Proceedings of the 40th IEEE conference on Decision and Control, Orland, Florida USA, December, 2001, pp.2720-2725.

[4] Luis E. Zarate, Fabricio R. Bittencout, *Neural networks and fuzzy based control for cold rolling process via sensitivity factors*, [C], The 27th Annual Conference of the IEEE Industrial Electronics Society, 2001, pp. 64-69.

[5] Christian Doring, Marie-Jeanne Lesot, Rudolf Kruse, *Data analysis with fuzzy clustering methods*, [J], Computational statistical & data analysis, Vol.51, 2006, pp.192-214.

[6] Jacek M. Leski, *Generalized weighted conditional fuzzy clustering*, [J], Transactions on Fuzzy systems, 2003, pp. 709-715.

[7] Wang Xi-Zhao, Wang Ya-Dong, Wang Li-Juan, *Improving fuzzy c-means clustering based on feature-weight learning*, [J], Pattern Recognition Letters, 2004, pp.1123-1132.

[8] Xue Zhan-Ao, Cen Feng, Wei Li-Ping. *A weighted fuzzy clustering algorithm based on Euclidean distance*, [C], The 5th international conference on fuzzy systems and knowledge discovery, 2008, pp. 172-175.

[9] Christian Borgelt. *Feature weighting and feature selection in fuzzy clustering*，[C]，Fuzzy Systems, IEEE World Congress on Computational Intelligence. IEEE International Conference on 2008.

[10] Chieh Yuan Tsai, Chuang Cheng-Chiu. *Developing a feature weight self-adjustment mechanism for a K-means clustering algorithm*, [J], Computational statistics & data analysis, 2008, pp. 4658-4672.

[11] Li jie, Gao Xin-Bo, Jiao Li-Cheng, *Fuzzy clustering algorithm based feature weighted*, [J], Electronic Journal, 2006, 34, pp. 9-13.

[12] Elaine Y. Chan, Wai Ki Ching, Micheal K. Ng, Joshua Z. Huang. *An optimization algorithm for clustering using weighted dissimilarity measures*, [J], Pattern Recognition, 2004, pp. 943-952.

[13] Sirichai Turmchokkasam , Sanya Mitaim. *Effects of Weights in Weighted Fuzzy C-Means Algorithm for Room Equalization at Multiple Locations* , [C], 2006 IEEE International Conference on Fuzzy Systems , Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada , July 16-21, 2006 , pp.1468-1475.

[14] Jiawei Han, Micheline Kamber. Data Mining: *Concepts and Techniques, 2rd edition*, [M], Publisher: Elsevier, 2006.

[15] Breunig M.Kriegel H, Ng R T et al. *LOF: Identifying density-based local outliers*, [C]. Proceedings of ACM SIGMOD International Conference on Management of Data, Dallas, Texas: ACM Press, 2000, pp. 93-104.

[16] Huang Tian-Qiang, Qin Xia-Lin,Ye Fei-Yue. *New approach based on square neighborhood to detect outliers*, [J].Control and Decision, Vol. 21, No.5,2006, pp. 145-545.

[17] Xue An-Rong, Ju Shi-Guang, He Wei-Hua. *Study on algorithms for local outlier detection*, [J], Chinese Journal of Computers, 2007, Vol.30, No.8, pp. 6541-3641.

[18] Amol Ghoting, Srinivasan Parthasarathy, Matthew Eric Otey. *Fast mining of distance-based outliers in high-dimensional dataset*, [J]. Data Mining Knowledge Discovery, Vol.15, 2008, pp. 349-364.

[19] Weng Xiao-Qing, Shen Jun-Yi. *Detecting outlier samples in multivariate time series dataset*, [J]. Knowledge-based systems, Vol.21, 2008, pp. 807-812.

[20] Frank Rehm, Frank Klawonn, Rudolf Kruse, *A novel approach to noise clustering for outlier detection*, [J], Soft Comput ,Vol. 11, 2007, pp. 489–494.

[21] N. R. Pal, J. C. Bezdek. *On cluster validity for the Fuzzy c-means Model*, [J], IEEE transactions on fuzzy systems. Vol. 3, 1995, pp. 370-378.

[22] Xie, X.L. and Beni, G. *A validity measure for fuzzy clustering*. [J], IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 13, No.8, 1991, pp. 841–847.