Computing and Informatics, Vol. 30, 2011, 357–380

CLUSTERING OF STEEL STRIP SECTIONAL PROFILES BASED ON ROBUST ADAPTIVE FUZZY CLUSTERING ALGORITHM

Chenglong Tang

School of Mechanical and Dynamical Engineering
Shanghai Jiao Tong University
No. 800 Dong Chuan Road, Minhang District
Shanghai 200240, P. R. China
&
Research Institute of BaoShan Iron and Steel Co., Ltd.
No. 655 of Fujin Road, Bao Shan District
201900, Shanghai, P. R. China
e-mail: tangcl@sjtu.org

Shigang Wang

School of Mechanical and Dynamical Engineering
Shanghai Jiao Tong University
No. 800 Dong Chuan Road, Minhang District
Shanghai 200240, P. R. China

Yunpeng Chen

BaoShan Iron and Steel Co., Ltd.
No. 885 of Fujin Road, Bao Shan District
201900, Shanghai, P. R. China

Manuscript received 7 September 2009; revised 25 May 2010
Communicated by János Fodor

Abstract. In this paper, the intelligent techniques are applied 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 cross-sectional profiles is the core of the system. The fuzzy clustering algorithm is used to establish the classifier. Secondly, a novel fuzzy clustering algorithm is proposed and used in the real application. The results, under the comparisons with the results obtained by the conventional fuzzy clustering algorithm, show the new algorithm is robust and efficient and it can not only get better clustering prototypes, which are used as the classifier, but also easily and effectively detect the outliers; it does great help in improving the performances of the new system. Finally, it is pointed out that the new algorithm's efficiency is mainly due to the introduction of a set of adaptive operators which allow for treating the different influences of data objects on the clustering operations; and in nature, the new fuzzy algorithm is the generalized version of the existing fuzzy clustering algorithm.

Keywords: Fuzzy clustering, robust, adaptive operators, outlier mining, steel strip profile

Mathematics Subject Classification 2000: 68Txx, 68Uxx

1 INTRODUCTION

As well known, steel strip cold rolling process is a typically complex process which is characterized by the terms of multi-variables, nonlinear and strong co-couple. A newly built cold tandem rolling mill in BaoSteel aims at producing high-grade steel strips specially supplied to automobile manufactures. While thickness of the strips is reduced to the specified value by the tandem 5 stands, flatness and transversal thickness deviation of the strip are strictly limited. There are a lot of factors to influence the quality of the final cold rolled products, the cross-sectional profile of the incoming strip being an important one [1, 2], affecting the rolling process considerably. Currently, in most of the existing cold rolling mills, normally the influence of the cross-sectional profile of the incoming strip is neglected. The experienced and popular approach is that the profiles of the incoming strip are treated as a fixed constant. A term of crown is normally given as a basic parameter to describe the cross-sectional profiles of the whole coil. There are two reasons: one, the existing quality control system can meet the common requirements of the customers. There was not the impetus to develop more complex systems. The other reason is that the high precision measuring device was not accessed over the past two decades; but now things change, producing thinner and stronger strips becomes the main direction in the cold rolling technologies. The existing system cannot meet the high quality requirement; and with the quick development of computing and measuring technologies, it becomes possible to develop the new control system on the basis of the advanced measuring devices.

The authors are currently devoted to developing a new automatic flatness and transversal thickness co-control system. The goal is that large amounts of products can be fabricated with the closer tolerances. In the new system, the incoming profile is considered to be an important factor and is treated in an online way. In order to realize this, a customized multi-sensor thickness measuring device is installed to measure the profiles. The control processes can be simply explained as follows: The thickness measuring device gets the profile signals in the stream way, the signals are treated by a kind of feature abstracting technique. The hydraulic actuators will respond according to the functional relationship between the outputs of the actuators and the signals. The functional relationship is a complex mathematical model and the calculation is generally an iterating process and even takes several minutes. The normal treating measure is to simplify some parameters to shorten the calculating time. As a result it degrades the control precision. In our research, as an alternative, the intelligence methods are introduced to ease these difficulties to a great extent. The steaming profiles signals are recognized to belong to one of the standard patterns, the hydraulic actuators will compensate the output according to the standard patterns. The online work will now classify the signals.

It is pointed out that the emphasis of this paper is not put on how the new proposed system works, this will be another topic. The emphasis of this paper is put on a kind of new developed clustering algorithm and its successful application in the new control system.

The paper is organized as follows: Section 2 puts forward a background of the industrial application. Section 3 addresses the mathematical models, the existing fuzzy clustering algorithm is introduced and the related work is reviewed, and then a new kind of adaptive fuzzy clustering algorithms is proposed and the detailed reasoning process is given as an appendix. In Section 4, the new algorithm is used to cluster the real-world data set, and the results are compared with the results obtained by the existing clustering algorithm. Section 5 summarizes the research results of this paper.

2 BACKGROUND OF THE APPLICATION

Figure 1 plots the overview of the new control scheme. A customized multi-sensor thickness measuring device (TTDG) is installed in the entry section of the tandem cold mill. It can simultaneously output the profile signals in the stream way. The features of the online signals are abstracted and the signals are classified by the classifier. The classifier will be established in advance. Clustering technique is used to establish the classifier. Because of the existence of abnormal profiles, the classifier should be robust. In the online classification phase, each real-time measured profile is regarded to belong to one of the standard profile patterns P_i $(i = 1, \dots, c)$ and the corresponding relationship between the standard patterns and the compensations of the hydraulic actuators ΔF_{ij} can be built up and updated in the offline manner. Additionally the classifier will be updated in a sliding way.

The equipped hydraulic actuators, back-up roll leveling, working roll bending, and immediate rolls bending and shifting, will compensate their outputs according

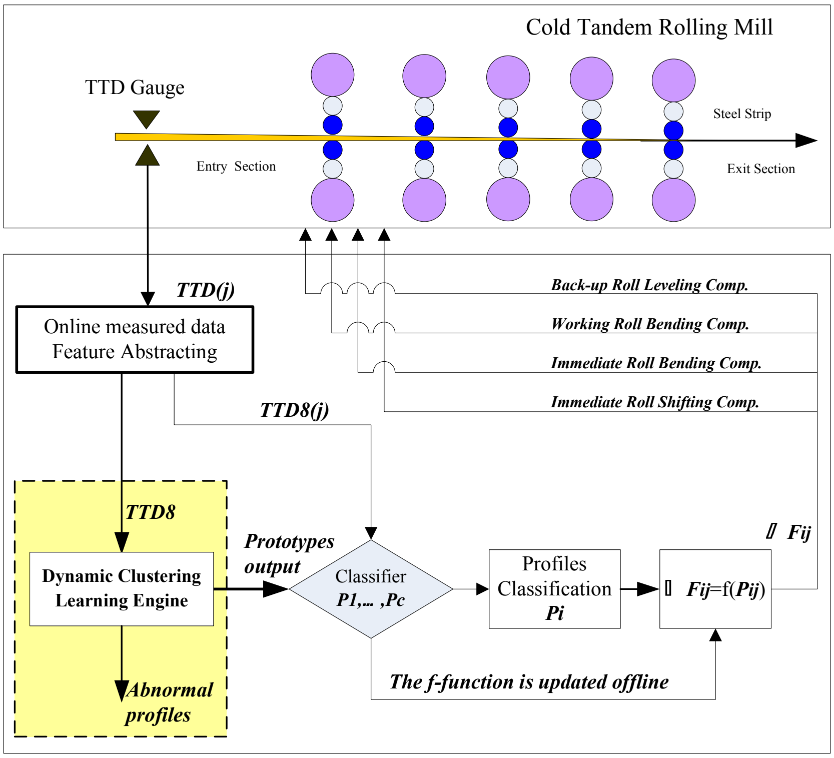

Fig. 1. Overview of the proposed intelligent control scheme

to the profile pattern on the basis of the relationship between the profile patterns and the outputs of the hydraulic actuators.

A certain number of the profile objects will be collected to be the learning set. The tasks, the contents of the dashed frame in Figure 1, include

1. the prototypes obtained by the Dynamical Clustering Learning Engine should be robust and efficient;
2. the outliers in the learning data set can be detected for the abnormal profiles contain abundant information and will be used in later analysis.

In order to fulfill the tasks, clustering technique is adopted. In the beginning, the widely-used conventional fuzzy clustering algorithm was used, but the clustering result was not so good, and detecting the abnormalities is difficult. Based on the conventional algorithm, a robust adaptive fuzzy clustering algorithm is developed. The improved algorithm is used to cluster the real profiles. The result comparisons between the conventional algorithm and the improved version show the improved version is more suitable to cluster the profiles, and more interesting;

detecting the abnormal profiles with the improved version becomes very easy and effective.

The developed new control frame was applied and testified in the newly building tandem cold rolling mill of BaoSteel. The learning profile signals are collected by the customized TTDG which has 40 channels and can output a row of thickness values simultaneously.

A polynomial of maximum order 8 is used to fit the measured thickness profiles, that is to say, every row of measured values is transformed to a 9-dimension coefficients vector by help of the least square fitting method. The explanation of the coefficients is easily understood in the following Equation (1). Here b_0 is the constant term and it is equal to the central thickness of the strip.

$$H = \sum_{i=1}^{8} b_i x^i + b_0 \tag{1}$$

where

- H is the reconstructing curve of the cross-sectional profile
- x is the Normalized abscissa from -1 to 1, in the width direction of the strip
- b_j are coefficients, $j = 1, \ldots, 8$.

In Equation (1), eight coefficients b_j, $j = 1, \ldots, 8$, are the research objectives and can be rewritten in the form of a vector, which is denoted as $\mathbf{B} = [b_8, b_7, b_6, b_5, b_4, b_3, b_2, b_1]$. The data set is named TTD8.

3 ROBUST ADAPTIVE FUZZY CLUSTERING ALGORITHM: RAFCM

3.1 Related Work

3.1.1 The Fuzzy Clustering Algorithms

Although the probabilistic fuzzy C-means clustering (FCM) algorithm (detailed description can be seen in the appendix and related literature) and its variants are proven to be very helpful in data mining, pattern recognition, image process etc., they also have some disadvantages, for example, the clustering results are sensitive to abnormal data objects. Many research works [3, 4, 5, 6, 7] had been done in order to improve the performances of the FCM and its variants. All these works hold a common point, cf. Equation (2) (the symbols are explained in the appendix), the objective function of the FCM can be rewritten in that way; it is clear that each data point has the same coefficient, "1". That means each data point has the same contribution to the total value of the objective function. It can be understood that the FCM treats data objects equally and neglects the possible different influences of the different data objects in the process of optimization of the objective function. Actually, in the real world data set, outliers are often contained. It is unreasonable because a bad point has the same influence on clustering just like a good point.

$$J_{CFM}(\mathbf{X},\mathbf{U},\mathbf{V}) = \sum_{j=1}^{n}\sum_{i=1}^{c} u_{ij}^m d_{ij}^2 = \underbrace{1\cdot\sum_{i=1}^{c} u_{i1}^m d_{i1}^2}_{x_1} + \underbrace{1\cdot\sum_{i=1}^{c} u_{i2}^m d_{i2}^2}_{x_2} + \ldots + \underbrace{1\cdot\sum_{i=1}^{c} u_{in}^m d_{in}^2}_{x_n} \tag{2}$$

In particular, in the aspect of the new development of the fuzzy clustering algorithms, there are a lot of references [8–14] discussing the "weighting" approaches. The weighting approaches concern mainly features [8, 9, 10, 11], prototypes [12], fuzzy membership degree [13], as well as distance [14]. Among them, the feature-weighted approaches are more widely studied and applied. It is necessary to point out that this paper also proposes a kind of adaptive approach. In nature it also can be seen as a kind of weighting approach. However, the "weight" of this paper has completely different meaning than all these existing "weights". In this paper, the concept of "weight" or "adaptive operator" means the contribution that each data point does to the objective function is self-adapted in the clustering process and its weight (adaptive operator) is updated in each iterative step.

3.1.2 The Outlier Mining Algorithms

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 many cases, such as the abnormal fluctuation of the complex industrial process, rare events are often more interesting than the normal ones. In the fields of outlier mining, the outliers themselves become the "focus", which is different from the case of the fuzzy clustering discussed above.

Outlier mining task generally can be described as a "Top-k" principle [15]: 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.

In [15] the widely used outlier mining methods are grouped into four categories, namely statistics-based, distance-based, density-based and feature deviation-based. The existing outlier mining methods use mainly 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. The other is to assign an 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 [16, 17, 18, 19] and distance-based [20, 21] methods are more studied and applied. There are also clustering techniques used to mine the outliers. However, since the main destination of the existing clustering algorithms is "cluster", they are not generally optimized for outlier detection; because of the problem of treating all data objects equally, the existing fuzzy algorithms are not suitable for detecting the outliers. They should be modified, just like the PCM [22] algorithm and the NC [23] algorithm.

Briefly, the existing outlier mining algorithms have the following 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 is sometimes difficult to explain; for example, in the case of the LOF algorithm [18], it defines the outliers in accordance with the local density in the sparse space, and it is probably 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, and thus the computation takes long time.

Generally speaking, both the existing fuzzy clustering algorithms and the outlier mining algorithms have several disadvantages. The new models need to overcome the disadvantages in these two issues in the theoretical researches and engineering applications, especially in the case when the two issues are integrated as a whole. The contribution of this paper is to propose a new model to solve the two issues as a whole.

3.1.3 The Robust Adaptive Fuzzy Clustering Algorithm with Outliers Detection

The work in this paper is aimed at solving the clustering problems by a kind of novel fuzzy clustering algorithm. A robust adaptive clustering approach is proposed on the basis of the existing fuzzy clustering algorithms. The core idea of this novel approach is that the nature of individual data point in the data set is "different". "Differently" treating the individual data points is more reasonable than treating them "equally". The proposed approach overcomes the disadvantage of the existing fuzzy clustering algorithms. When the novel approach is used to partition the data set, it can get better clustering quality than the existing methods, and it can easily detect the outliers and effectively mine much information related to the outliers.

The new robust adaptive fuzzy clustering algorithm is presented as a complete appendix attached hereto.

3.1.4 Calculating Flowchart of the RAFCM Algorithm

The calculation steps of the RAFCM algorithm are shown in Figure 2. There are two parts: the upper frame is the clustering part, and the bottom frame is the outliers mining part which is based on the clustering part. The detailed reasoning processes of the robust adaptive fuzzy clustering algorithm are given in the appendix.

3.1.5 Main features of the RAFCM algorithm

In contrast to the FCM, a new vector, adaptive operators $E(e^{t_j})$ and a new parameter s ($s \neq 0$) are introduced to the RAFCM and they jointly affect the final clustering result. In addition, note that the adaptive operator e_{t_j} works as a whole in the iterative process. According to Equation (A.18), the value of the adaptive operator e_{t_j} depends on the basic parameters, just like, u_{ij}, d_{ij}, s, n and c. There is no exponential calculation. In this paper the adaptive operator adopts the form of exponent, but it is not absolute. It is possible to adopt other forms, the only necessary condition is that the product of all operators should be one. Moreover, it can be seen from Figure 2 that the set of parameters t_j is only used to generate the initial adaptive operators. Then, these parameters are not needed again.

In order to clarify the differences of the FCM and the RAFCM, the objective function of the RAFCM algorithm, Equation (A.4) is rewritten in the form of the following Equation (3).

$$\begin{aligned} J_{RAFCM}(\mathbf{X},\mathbf{U},\mathbf{V},\mathbf{E}) &= \sum_{j=1}^{n} e^{st_j} \sum_{i=1}^{c} u_{ij}^m d_{ij}^2 = e^{st_1} \cdot \underbrace{1 \cdot \sum_{i=1}^{c} u_{i1}^m d_{i1}^2}_{x_1} \\ &\quad + e^{st_2} \cdot \underbrace{1 \cdot \sum_{i=1}^{c} u_{i2}^m d_{i2}^2}_{x_2} + \ldots + e^{st_n} \cdot \underbrace{1 \cdot \sum_{i=1}^{c} u_{in}^m d_{in}^2}_{x_n} \end{aligned} \quad (3)$$

According to Equation (2), in the case of the FCM, each data point has the same contributing weight of "1" to the optimal value of the J_{FCM} while according to Equation (3), in the case of the RAFCM, the contribution of each data point is different, the weights, or the adaptive e^{t_j} are different. It shows that the RAFCM has more flexibility than the FCM. Considering a special case, letting all adaptive e^{t_j} be fixed to be 1, then the product of all adaptive e^{t_j} is still 1; that means, the constraint of the adaptive operators (Equation (A.4)) is still valid. Then, the RAFCM becomes the FCM right now. Thus, the RAFCM is a new kind of generalized FCM.

In the development of the clustering techniques, it was a big skip for the clustering methods were transferred from the *crisp* model to the *fuzzy* model. In this transfer, two kinds of parameters were introduced: the fuzzy membership degree u_{ij} and the fuzzy exponent m ($m \neq 0$). The fuzzy C-means clustering algorithm is essentially an extension of the crisp C-means algorithm where the notion of the degree of fuzziness, taking values not less than one, is introduced. Moreover, a constraint of the fuzzy membership degree u_{ij} is given and a constant m is introduced to adjust the influence of the fuzzy membership degree u_{ij}.

It is a new transit from the basic Fuzzy C-means clustering to the proposed robust adaptive fuzzy C-means clustering in the development of the clustering technologies. In the RAFCM, the adaptive operators, which are used to control the different influences of different data points on the clustering result, are similar to the roles that the fuzzy membership degrees play when the clustering model changes from the "crisp" time to "fuzzy" time. The new parameter s under the RAFCM is very similar to the role that the exponent m plays in the FCM. Even in the RAFCM, fuzzy membership degree u_{ij} and the fuzzy exponent m still exist and have the same

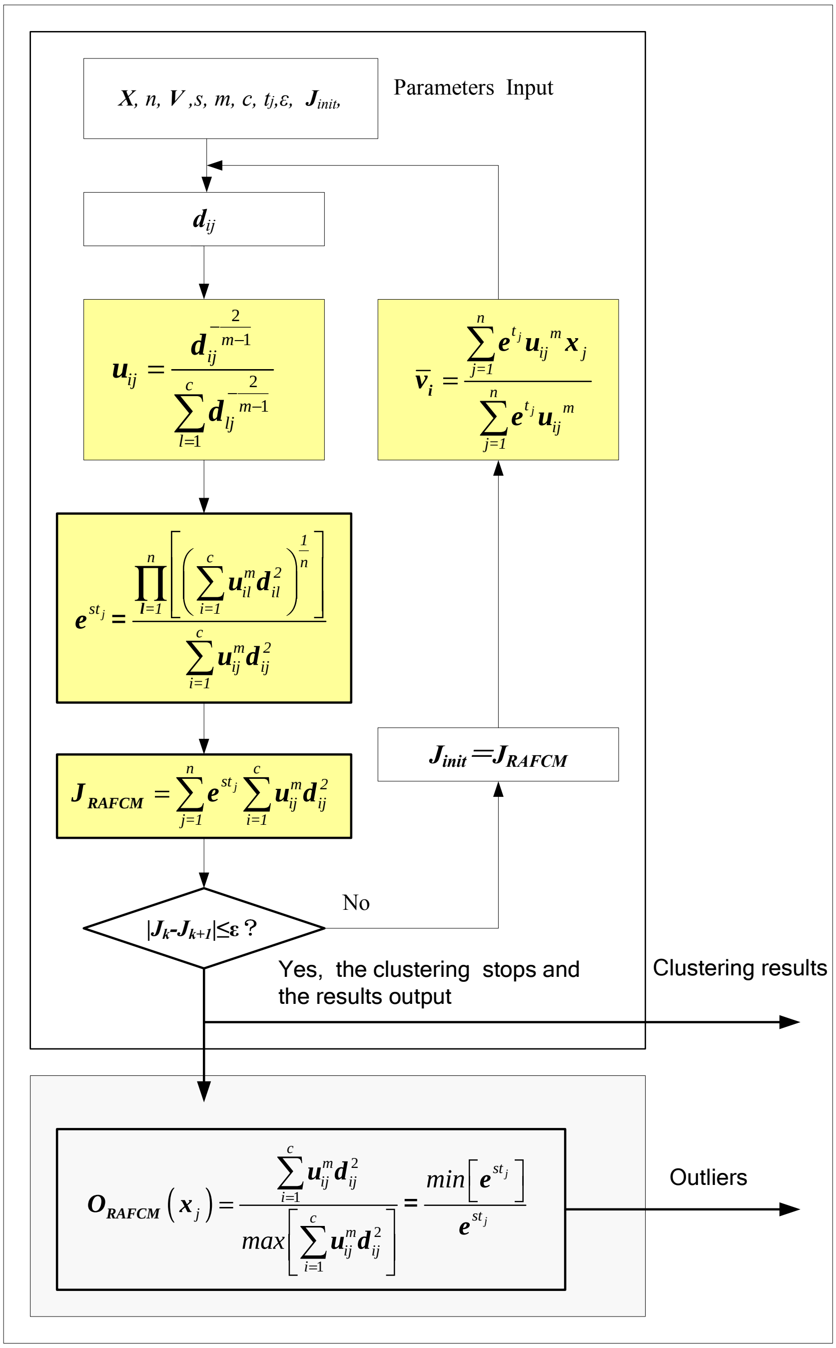

Fig. 2. Clustering and outlier mining flowchart of the RAFCM algorithm

definitions. Just like the fact that the basic fuzzy clustering model has been proven in the literature that it can get better clustering results than the crisp model, the RAFCM can get better clustering performances than the FCM. It is possible and reasonable.

As to time complexity of the robust adaptive fuzzy C-means clustering, it should be paid attention to, especially when the RAFCM is used in real time applications. When evaluating time complexity of one algorithm, total time consumed is a direct index. For the objective function-based clustering algorithms, the final results are generally got in an iterative manner; the time one algorithm needs to concern two aspects, the time needs of each iterative step and the total iterative times when the objective function is converged to its optimum.

For the RAFCM, because of the existence of a set of the adaptive operators, the time that each iterative step needs is more than that the FCM needs, but it is possible that the total iterative times of the RAFCM are reduced. The reason is the total iterative times are probably reduced, or, say, the objective function is converged to its optimum quicker. As a result, comparing with the FCM, the total time complexity of the RAFCM is probably the same or even lower. Subsection 4.3 gives the experimental results of time complexity of the RAFCM on the real data set.

As announced above, the RAFCM can easily and effectively detect the outliers in a data set. The detailed ways can be seen in Figure 2 and in the appendix.

4 APPLICATION

4.1 Preparation

The cross-sectional profile signals of 3 coils and totally 1074 profile signals are sampled by the TTD gauge. The original data set is a $40 \times 1\,074$ matrix. The polynomial fitting method is adopted to reduce the dimensionality of the original data set. The maximum order 8 coefficients are abstracted as the main features. As a result, an $8 \times 1\,074$ matrix is obtained and replaces the original data set to be analyzed. The data set is named TTD8 and it works as the learning data set. The way to abstract the features has been described in Section 2.

The FCM and the RAFCM are used to cluster the TTD8 under the same combinations of the number of clusters c and the fuzzy exponent m. Moreover, the m gets its widely used value, 2 [24]. 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 ε is set to be 0.001, and the new parameter s needed by the RAFCM is set to be 4.5. On the theoretical side, the FCM is proved to be sensitive to initialization. So, the crisp C-means clustering algorithm is used to generate a set of prototypes to initialize both the FCM and the RAFCM with respect to the number of clusters in order to keep the comparisons reasonable.

4.2 Cluster Prototypes Obtained by the FCM and the RAFCM

Considering a typical case, c is 9, the clustering results are given. Prt_RAFCM_9 and Prt_FCM_9 are the prototype matrices obtained by the RAFCM and the FCM, respectively, and here 9 means the number of clusters is 9.

The values of the Prt_RAFCM_9 and Prt_FCM_9 are as follows.

$$\text{Prt\_RAFCM\_9} =$$
$$\begin{bmatrix}
0.2151 & 0.2793 & -0.0282 & -0.3934 & -0.2174 & 0.5742 & 0.4440 & -0.6269 \\
-0.1686 & 0.3334 & 0.1024 & -0.3180 & -0.0105 & 0.2409 & -0.0652 & -0.2235 \\
0.4862 & 0.6696 & -0.2449 & -0.7690 & -0.0823 & 0.9368 & 0.4148 & -0.9861 \\
-1.1796 & -0.3804 & 1.2754 & 0.5616 & -1.2828 & -0.6483 & 0.9635 & 0.7192 \\
-0.0076 & -0.4579 & -0.1853 & 0.4533 & 0.3966 & -0.4627 & -0.5358 & 0.4663 \\
-0.2985 & -0.4964 & 0.1136 & 0.4923 & 0.1139 & -0.4956 & -0.3156 & 0.4997 \\
0.6824 & 0.9735 & -0.4691 & -1.0601 & 0.1719 & 1.1435 & 0.1507 & -1.1846 \\
-0.8340 & 0.8477 & 0.7349 & -0.7884 & -0.5704 & 0.6667 & 0.4092 & -0.6158 \\
1.8447 & -1.5431 & -1.9013 & 1.5042 & 1.8613 & -1.5386 & -1.5352 & 1.4956
\end{bmatrix}$$

$$\text{Prt\_FCM\_9} =$$
$$\begin{bmatrix}
0.4311 & 0.5903 & -0.1981 & -0.6951 & -0.1157 & 0.8690 & 0.4252 & -0.9201 \\
-0.3601 & 0.1841 & 0.2432 & -0.1469 & -0.0831 & 0.0581 & -0.0391 & -0.0279 \\
0.4506 & 0.6377 & -0.2180 & -0.7368 & -0.0965 & 0.9028 & 0.4100 & -0.9517 \\
-1.1637 & -0.4118 & 1.2555 & 0.5937 & -1.2601 & -0.6786 & 0.9440 & 0.7488 \\
-0.0760 & -0.3914 & -0.0978 & 0.3863 & 0.2964 & -0.3982 & -0.4347 & 0.4067 \\
-0.4977 & -0.4754 & 0.2889 & 0.4883 & -0.0194 & -0.5083 & -0.2582 & 0.5097 \\
0.5664 & 0.8167 & -0.3897 & -0.9332 & 0.1504 & 0.9917 & 0.1016 & -1.0404 \\
-0.3619 & 0.4132 & 0.2569 & -0.3685 & -0.1134 & 0.2636 & 0.0133 & -0.2283 \\
1.8312 & -1.6176 & -1.9045 & 1.5770 & 1.8924 & -1.6010 & -1.6072 & 1.5524
\end{bmatrix}$$

It is difficult to directly evaluate which algorithm is better only according to these two number matrices. It is necessary to have an evaluating method.

4.3 Comparisons of Clustering Performances under the RAFCM and the FCM

In the research of the new clustering algorithms, the evaluation of the clustering performances can be regarded to contain three aspects: one, the clustering validity problem: for a given data set, how many clusters the data set are divided is better. Two, the clustering quality problem: if the number of clusters has been given in advance, which kind of partition is better. Three, the computing efficiency problem: if a new algorithm is heavily time-consuming, even if it can give a good partition, it is still not a good method.

The clustering validity problem is usually discussed inside a particular algorithm. For one clustering algorithm, the new cluster validity functions are developed and then the best number of the clusters can be known by calculating the 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. For the clustering is an unsupervised learning method, the data is unlabelled and the ideal output actually does not exist in advance. So, when comparing the clustering qualities under the two algorithms, the comparisons are relative. The problem of the computing efficiency of clustering is easier to understand. It is evaluated by recording the time consumed or by the total iteration times. Especially, when treating the data set online, this problem needs more attention.

The main topics of this paper are a novel clustering approach and its actual application. The number of the clusters has been given as a known value; so it does not involve the validation problem of new clustering. The clustering quality problem and the computing time complexity problem are the two key points of this paper.

In subsection 4.2, the FCM is selected to be the counterpart algorithm, and the related parameters are given in advance, the respective cluster prototypes are obtained by the RAFCM and the FCM. The next work is to compare which result is better and which algorithm is more suitable for this actual application.

4.3.1 Clustering Quality Evaluation Function

In the development of new clustering algorithms, public data sets, such as the UCI database, are commonly used to check the performances of the new algorithms. Because the prototype matrix and the data labels are known in advance, the clustering results can be compared with the known values. However, there are a lot of different cases, for example, in the practical profiles classification, the prototype matrix and the data labels of the objective data set are unknown in advance. It is necessary to propose the new evaluating ways to compare and assess the clustering performances of two different algorithms when treating that kind of data set without a priori knowledge, just like the TTD8 data set.

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, see Equation (4). The compactness and separation functions are briefly written as CMP and SPT, the definitions of the CMP and SPT are given in Equations (5) and (6), respectively.

$$EVA = \frac{CMP}{SPT} \tag{4}$$

$$CMP = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{c}(x_j - v_i)^2 \tag{5}$$

$$SPT = \min_{i \neq k} \| \bar{v}_i - \bar{v}_k \|^2 \tag{6}$$

When the data set is divided 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 partial information of the clustering results. 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, the better the clustering quality.

4.3.2 The Comparison Result and Analysis

Table 1 lists the EVA values under the FCM and the RAFCM in three cases, $c = 6$, 9 and 12. All computations are repeated five times and the average values are used as the final results. The EVA values show the RAFCM is more robust than the FCM. Additionally, when discussing the time complexity of one algorithm, two time indices are generally used, the total time consumption and the total iterative cycles. Table 1 also gives these two indices for each calculation case.

	FCM			RAFCM		
c	*EVA*	Seconds	Total iterative cycles	*EVA*	Seconds	Total iterative cycles
6	0.0302	0.6844	48	0.0257	0.4013	32
9	36.3010	2.3245	133	0.2026	0.5149	32
12	4.6608	1.6477	73	0.1228	0.5235	26

Table 1. The EVA values of the FCM and the RAFCM when the c are 8, 12 and 16

It can be seen that in all three cases, the quality obtained by the RAFCM is better than that obtained by the FCM. In the view of time complexity, it can be known, whatever c is, that the AFCM consumes less time and the calculation is faster. It shows that the RAFCM can get better quality meanwhile keeping the computation low time consuming by selecting an available value of s. Briefly, the RAFCM shows the same or better level of time complexity when compared to the FCM. The reason is as follows: The computing time will be longer for the introduction of the adaptive operators in each iterative step, but meanwhile for adaptive operators the total iterative times are probably reduced. As a synthesized effect, the RAFCM is not a huge time-consumption algorithm. This feature is very important in the dynamical application.

4.4 Abnormal Profiles Detected by the RAFCM

Figure 3 plots the distribution of the outliers in the TTD8. The bold line is the dividing line and the y-coordinate of the line depends on the rule "k points are above the line, they are outliers. The remaining points are normal points". According to Figure 3, it is a binary decision of one data point, an object is an outlier or not. Some cluster-based outlier mining models detect the outliers in this way, the information about the outliers is not effectively mined. For the RAFCM, it can meanwhile give the cluster properties of all data points, including the outliers (cf. Figure 4). Figure 4 covers all information that Figure 3 can supply, and moreover, the cluster-property of each data is clear, seeing "Cl-i", i is from 1 to 9. This is the advantage and interesting aspect of the RAFCM, it can do the clustering operation, meanwhile it can detect outliers. It is exactly our need to fulfill the tasks.

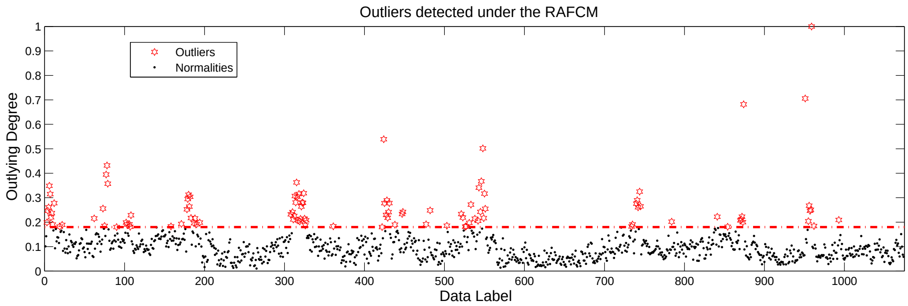

Fig. 3. Distribution of the outliers in the data set detected by the RAFCM

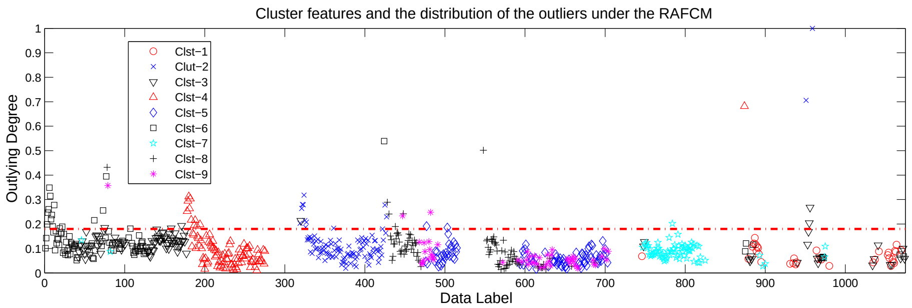

Fig. 4. Distribution of the clusters and outliers detected by the RAFCM

5 CONCLUSION

In this paper, intelligent technologies are introduced to steel strip rolling. A control scheme based the intelligent techniques is proposed. A novel robust adaptive fuzzy C-means clustering algorithm is reported and proved to be more suitable for mining the steel strip cross-sectional profile signals. Real applications on the learning profile data set TTD8 show the RAFCM has the following advantages:

1. The prototypes obtained by the RAFCM are more effective than those obtained by the FCM, the RAFCM is proved to be able to get better clustering quality. The new algorithm's efficiency is mainly due to the introduction of a set of adaptive operators and an adjusting parameter s which jointly allow treating the different influences of data objects on the clustering operation. It can

be regarded that the RAFCM supplies a new channel to refine the clustering operations. So, it is reasonable that the RAFCM can get better performances.

2. The RAFCM is more robust than the FCM. It can reduce or alleviate the bad influences of the outliers by the functions of both the adaptive operators and the available selected parameter s.
3. The abnormal profiles in the TTD8 data set can be easily detected by the RAFCM. The value of adaptive operator of each data point reflects a kind of the distance relationship between the individual data and the whole data set and it can be used as a kind of measure to detect the abnormal extent. Thus, the RAFCM can be used as a tool to mine the outliers.
4. The RAFCM shows the same or better level of time complexity when compared to the FCM.
5. It shows that the FCM is a special case of the RAFCM, in other words, the RAFCM in nature is a generalized version of the FCM.

Acknowledgement

This work is supported by National Natural Science Foundation of China under Grant No. 50875169.

Appendix: The Robust Adaptive FCM Algorithm (RAFCM)

Cluster analysis is one of the several important tools in modern data analysis. The general philosophy of cluster analysis is to divide the data set into several homogenous groups based on the similarity or dissimilarity. In such cases, the objects in the same group tend to being as similar as possible while the objects in different groups tend to being as dissimilar as possible.

There are crisp and fuzzy clustering algorithms. Fuzzy clustering is meant to provide a richer means for representing data structure. The fuzzy clustering methods are enabled to find more realistic models, since, in fact, boundaries between many groups are very badly delineated. In the field of fuzzy clustering, two types of fuzzy clustering algorithms have been developed, namely the probabilistic fuzzy clustering and the possibilistic fuzzy clustering. They differ in the constraints they place on the membership degrees and how the membership degrees should be interpreted. This paper only discusses the most widely used type, the objective functions-based probabilistic fuzzy clustering methods, more especially, the class of methods based on the fuzzy C-means (FCM) algorithm. Notice that in literature they are sometimes just called fuzzy clustering (omitting the word probabilistic). This paper did it too.

The clustering task can then be formulated as a function optimization problem. The objective function depends on both the cluster prototypes and the memberships of data points in the clusters. It cannot be optimized directly and therefore the Alternative Optimization (AO) scheme is usually used that optimizes one group of parameters meanwhile holding the other group fixed and vice versa. This iterative updating scheme is repeated in the hope to approach the optimum of the objective function.

In the following sections, a new kind of fuzzy clustering algorithm, the robust adaptive fuzzy clustering algorithm, is proposed on the basis of the existing fuzzy clustering algorithms. Firstly, the objective function-based FCM algorithm is reviewed. Second, the new fuzzy clustering algorithm is given in details. Finally the relationship between the two types of fuzzy clustering algorithms is investigated.

1 THE FCM ALGORITHM

The FCM has been successfully employed in a wide variety of fields and demonstrated a high degree of adapting to different data sets. The FCM is essentially an extension of the crisp C-means algorithm where the notion of the degree of fuzziness, taking values not less than one, is introduced.

The objective function J_{FCM} has the form as given in Equation (A.1) and a constraint related to the fuzzy membership degrees is given too.

$$J_{FCM}(\mathbf{X}, \mathbf{U}, \mathbf{V}) = \sum_{j=1}^{n} \sum_{i=1}^{c} u_{ij}^{m} d_{ij}^{2} \tag{A.1}$$

$$s.t. \qquad \sum_{j=1}^{n} u_{ij} > 0, \quad \sum_{i=1}^{c} u_{ij} = 1.$$

Here, $\mathbf{X}$ is the data set, it contains n data objects, U is the fuzzy partition matrix, and u_{ij}? $[0, 1]$ is the membership degree of data x_j in cluster v_i. $\mathbf{V}$ is the prototype matrix. c is the number of clusters, m is the fuzzy exponent and d_{ij} is the distance between x_j and v_i.

When solving the optimal value of the J_{FCM} by the AO scheme, two update equations are needed. Equation (A.2) is to update the fuzzy partition matrix $\mathbf{U}$ (u_{ij}) and Equation (A.3) is to update the prototypes matrix $\mathbf{V}$ (v_i).

$$u_{ij} = \frac{d_{ij}^{-\frac{2}{m-1}}}{\sum_{l=1}^{c} d_{lj}^{-\frac{2}{m-1}}} \tag{A.2}$$

$$v_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} x_j}{\sum_{j=1}^{n} u_{ij}^{m}} \tag{A.3}$$

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

2 THE RAFCM ALGORITHM

This section first presents the new objective function of the RAFCM algorithm. In this algorithm, three sets of parameters should be updated in each iterative step to solve the optimal value of the new objective function.

2.1 The New Objective Function of the RAFCM Algorithm

The new objective function is proposed with two additional constraints, see Equation (A.4).

$$J_{RAFCM}\left(\mathbf{X},\mathbf{U},\mathbf{V},\mathbf{E}\right)=\sum_{j=1}^{n}\left(e_j^t\right)^s\sum_{i=1}^{c}u_{ij}^m d_{ij}^2 \tag{A.4}$$

$$s.t. \qquad \sum_{j=1}^{N}u_{ij}>0, \sum_{i=1}^{c}u_{ij}=1, \quad \prod_{j=1}^{n}e^{t_j}=1$$

The reason why the new algorithm is named robust adaptive FCM algorithm, shortly RAFCM, is that a set of adaptive operators is introduced into the new objective function. An adaptive operator corresponds to a data object. All adaptive operators make up a vector. This vector is named the adaptive operator vector, shortly $\mathbf{E}(e^{t_j})$.

Comparing the *JFCM* and the *JRAFCM*, the same symbols in Equations (A.1)–(A.3) have the same definitions as those in Equation (A.4) and the constraint about the fuzzy membership degree holds unchanged. Besides, there is a set of new parameters e^{t_j} and an additional constraint in the RAFCM. The e^{t_j} is named the adaptive operator of the data x_j and reflects the influence of the data x_j on the *JRAFCM*; t_j is named the data factor of the data x_j. s $(s \neq 0)$ is a preselected real-valued parameter and is used to adjust the influence of the adaptive operator. It is named the adaptive exponent. The additional constraint shows that the product of all adaptive operators is subject to be one.

Just like the basic FCM, the optimal value of the *JRAFCM* is also solved by the AO scheme. The adaptive operator vector is also updated. It is necessary to give the expression of e^{t_j}.

The minimization of the objective function represents a nonlinear optimization problem that can usually be solved by means of the Lagrange Multiplying Method. While taking two kinds of constraints into account, the optimization expression can be written as follows (cf. Equation (A.4)):

$$J_{RAFCM,\lambda_1,\lambda_2}\left(\mathbf{X},\mathbf{U},\mathbf{V},\mathbf{E}\right)=\sum_{j=1}^{n}e^{st_j}\sum_{i=1}^{c}u_{ij}^m d_{ij}^2+\lambda_1\left(\sum_{i=1}^{c}u_{ij}-1\right)+\lambda_2\left(\prod_{j=1}^{n}e^{t_j}-1\right) \tag{A.5}$$

where λ_1 and λ_2 are two Lagrange multiplying operators corresponding to the two constraints, fuzzy membership degrees and adaptive operators.

In Equation (A.5), the third term of the right side is the product of all $e^{t_j} = 1$, it is inconvenient in use. A new constraint (A.6) says that the summary of all $t_j = 0$ $(j = 1, \dots, n)$ is given. It is obvious that the constraint in Equation (A.6) is completely equivalent to that in Equation (A.4).

$$\sum_{j1}^{n} t_j = 0 \tag{A.6}$$

Therefore, Equation (A.5) can be rewritten as follows:

$$J_{RAFCM,\lambda_1,\lambda_2}(\mathbf{X},\mathbf{U},\mathbf{V},\mathbf{E}) = \sum_{j=1}^{n} e^{st_j} \sum_{i=1}^{c} u_{ij}^m d_{ij}^2 + \lambda_1 \left(\sum_{i=1}^{c} u_{ij} - 1 \right) + \lambda_2 \left(\sum_{j=1}^{n} t_j - 0 \right) \tag{A.7}$$

The next steps are to solve the expression of the fuzzy membership degree u_{ij} and the adaptive operators e^{t_j} when the new objective function **JRAFCM** reaches its optimal value.

The u_{ij} and e^{t_j} (replaced by t_j) are partially differentiated:

$$\begin{cases} \frac{\partial J_{RAFCM,\lambda_1,\lambda_2}(\mathbf{X},\mathbf{U},\mathbf{V},\mathbf{E})}{\partial u_{ij}} = m e^{st_j} u_{ij}^{m-1} d_{ij}^2 + \lambda_1 \\ \frac{\partial J_{RAFCM,\lambda_1,\lambda_2}(\mathbf{X},\mathbf{U},\mathbf{V},\mathbf{E})}{\partial t_j} = s e^{st_j} \sum_{i=1}^{c} u_{ij}^m d_{ij}^2 + \lambda_2 \end{cases}. \tag{A.8}$$

Let two partial derivatives be 0; the expressions of the two multiplying operators, λ_1 and λ_2, can be obtained (cf. Equations (A.9) and (A.10).

$$\lambda_1 = -m e^{st_j} u_{ij}^{m-1} d_{ij}^2 \tag{A.9}$$

$$\lambda_2 = -s e^{st_j} \sum_{i=1}^{c} u_{ij}^m d_{ij}^2 \tag{A.10}$$

There are three groups of parameters which will be updated in the process of solving the optimal value of the robust adaptive fuzzy clustering algorithm. They are fuzzy partition matrix **U**, adaptive operator vector **E** and cluster prototypes matrix **V**.

2.2 Fuzzy Partition Matrix U (u_{ij}) Update Equation

The expression of the fuzzy membership degree u_{ij} can be obtained by transforming Equation (A.9); Equation (A.11) is obtained subsequently:

$$u_{ij} = \left(\frac{\lambda_1}{-m e^{st_j}} \right)^{\frac{1}{m-1}} d_{ij}^{-\frac{2}{m-1}} \tag{A.11}$$

According to the constraint of the fuzzy membership degree u_{ij} (cf. Equation (A.11)), the following Equation (A.12) can be obtained.

$$\sum_{i=1}^{c} u_{ij} = \left(\frac{\lambda_1}{-me^{st_j}}\right)^{\frac{1}{m-1}} \sum_{i=1}^{c} d_{ij}^{-\frac{2}{m-1}} \tag{A.12}$$

Noting the fact that the left side of Equation (A.12) is equal to 1 and transforming Equation (A.12) further, Equation (A.13) is obtained as follows:

$$\left(\frac{\lambda_1}{-me^{st_j}}\right) = \frac{1}{\sum_{i=1}^{c} d_{ij}^{-\frac{2}{m-1}}}. \tag{A.13}$$

So, the expression of the fuzzy membership degree u_{ij} can be obtained based on Equations (A.11) and (A.13) (cf. Equation (A.14)).

$$u_{ij} = \frac{d_{ij}^{-\frac{2}{m-1}}}{\sum_{l=1}^{c} d_{lj}^{-\frac{2}{m-1}}} \tag{A.14}$$

Equation (A.14) shows that under the RAFCM the membership degree u_{ij} holds the same definition as the FCM (cf. Equation (A.2). u_{ij} is updated in each iterative step according to Equation (A.14).

2.3 Adaptive Operators Vector E (e^{t_j}) Update Equation

Equation (A.10) is further transformed, and then the adaptive operator e^{t_j} is written as follows:

$$e^{t_j} \left[\frac{-\lambda_2}{s\sum_{i=1}^{c} u_{ij}^m d_{ij}^2}\right]^{\frac{1}{s}}. \tag{A.15}$$

Noting that a data point corresponds to an adaptive operator, taking all adaptive operators into consideration and multiplying all adaptive operators, Equation (A.16) can be got:

$$\prod_{j=1}^{n} e^{t_j} = \prod_{j=1}^{n} \left[\frac{-\lambda_2}{s\sum_{i=1}^{c} u_{ij}^m d_{ij}^2}\right]^{\frac{1}{s}} \tag{A.16}$$

The left side of Equation (A.16) is 1 following the constraint of adaptive operators. So the expression of λ_2 can be rewritten as follows:

$$\lambda_2 = -s \prod_{j=1}^{n} \left[\left(\sum_{i=1}^{c} u_{ij}^m d_{ij}^2\right)^{\frac{1}{n}}\right]. \tag{A.17}$$

Bringing the operator λ_2 back to Equation (A.15), e^{t_j} is as follows:

$$e^{t_j} = \frac{\prod_{l=1}^{n}\left[\left(\sum_{i=1}^{c} u_{il}^m d_{il}^2\right)^{\frac{1}{s}\frac{1}{n}}\right]}{\left(\sum_{i=1}^{c} u_{ij}^m d_{ij}^2\right)^{\frac{1}{s}}}. \tag{A.18}$$

According to Equation (A.18), if a prototype sits exactly on a data object, it is sure that one membership degree will be zero, and then (A.18) cannot be computed. In order to avoid this, giving the available initial prototypes is the key. It is suggested that HCM or FCM is used to produce the initial prototypes. In addition, the extreme outliers or false data points will probably produce very high values of the adaptive operators. It is advisable to handle the data set before it is analyzed by the RAFCM, for example, to remove the extreme outliers.

The adaptive operator vector is updated in each iterative step according to Equation (A.18).

2.4 Cluster Prototypes Matrix V (v_i) Update Equation

The cluster prototype is given as follows. Here $\bar{v}_i$ is the adaptive cluster prototype.

$$\bar{v}_i = \frac{\sum_{j=1}^{n} \bar{u}_{ij}^m x_j}{\sum_{j=1}^{n} \bar{u}_{ij}^m} \tag{A.19}$$

where $\bar{u}_{ij}$ is the adaptive fuzzy membership degree and its calculating equation is as follows (Equation (A.20) has nothing to do with the parameter s):

$$\bar{u}_{ij}^m e^{t_j} u_{ij}^m. \tag{A.20}$$

Based on Equations (A.19) and (A.20), the new cluster prototype equation can be rewritten as follows:

$$\bar{v}_i = \frac{\sum_{j=1}^{n} e^{t_j} u_{ij}^m x_j}{\sum_{j=1}^{n} e^{t_j} u_{ij}^m}. \tag{A.21}$$

It should be pointed out that adaptive exponent s is not included in (A.20) or (A.21). The reason is as follows. In order to make the explanation clear, (A.18) is rewritten as (A.22). Note that the right side of (A.22) has nothing to do with the parameter s and the left side e^{st_j} appears as a whole and joins the objective function, see (A.4). So, if e^{st_j} and not e^{t_j} is introduced to (27) or (28), the parameter s will not play its role to adjust the adaptive degrees.

$$e^{st_j} = \frac{\prod_{l=1}^{n}\left[\left(\sum_{i=1}^{c} u_{il}^m d_{il}^2\right)^{\frac{1}{n}}\right]}{\sum_{i=1}^{c} u_{ij}^m d_{ij}^2} \tag{A.22}$$

Although s is not included in (A.20) and (A.21), it cannot be regarded that the calculation of the cluster prototypes has nothing to do with s. In contrast, the prototypes are influenced by s, the influencing relationship of both is indirect. The reason is that e^{st_j} works as a whole in the calculation of the objective function, but different s will output different e^{t_j}. So s indirectly influences the cluster prototypes, see (A.21); the influence is realized by iteration. Their relationship between the s and prototypes is implicit.

3 MINING OUTLIERS WITH THE RAFCM ALGORITHM

The RAFCM algorithm attaches importance to the different contributions of different data points. On the one hand, it can obtain better clustering quality; on the other hand, it can be used as a tool to detect the outliers in the data set. In the field of mining outlier, the definition of the outlier is a basic issue which should be solved in the beginning. When the RAFCM is used to mine the outliers, the definition of outliers will be given in advance too. In this paper, the fuzzy distance between the data point and the adaptive prototype matrix is used to define the outlier.

$\sum_{i=1}^{c} u_{ij}^m d_{ij}^2$ is the fuzzy square distance of the data x_j. The term expresses a kind of fuzzy distance relationship between the data x_j and the prototype matrix. The greater the fuzzy square distance, the bigger the possibility that data x_j is an outlier. When the objective function J_{RAFCM} is converged to its optimal value, each data point gets its own fuzzy square distance. All n the fuzzy square distance values are rearranged with the order from the maximum to the minimum. k data points corresponding to the largest k values are defined to be the outliers.

Transforming Equation (A.18) further, Equation (A.23) is obtained as follows:

$$e^{st_j} \sum_{i=1}^{c} u_{ij}^m d_{ij}^2 \prod_{l=1}^{n} \left[\left(\sum_{i=1}^{c} u_{il}^m d_{il}^2 \right)^{\frac{1}{n}} \right] \tag{A.23}$$

When the J_{RAFCM} is converged to its optimal value, the right side of Equation (A.23) is surely a confirmed constant, it is even unnecessary to know its exact value. It can be known that, for any data x_j, the product of e^{st_j} and $\sum_{i=1}^{c} u_{ij}^m d_{ij}^2$ is equal to each other. The largest k fuzzy square distance values correspond to the k smallest e^{st_j}. Therefore, under the RAFCM, the outliers can be mined in accordance with the values of e^{st_j}. Here a term of the outlying degree of data is defined as follows:

$$O_{RAFCM}(x_j) = \frac{\sum_{i=1}^{c} u_{ij}^m d_{ij}^2}{\max\left[\sum_{i=1}^{c} u_{ij}^m d_{ij}^2\right]} \frac{\min\left[e^{st_j}\right]}{e^{st_j}}. \tag{A.24}$$

$O_{RAFCM}(x_j)$ is the outlying degree of the data x_j and it is a normalized value $(0, 1]$. The bigger the $O_{RAFCM}(x_j)$, the more likely the data x_j is an outlier. Obviously, the definition of the outlying degree of the data x_j expresses a kind of global relationship between the data x_j and the whole data set.

4 THE RAFCM IS A KIND OF GENERALIZED FCM

This section discusses the relationship between the FCM and the RAFCM. It can be seen that the fuzzy membership degree u_{ij} holds the same definition under both the FCM and the RAFCM by comparing Equations (A.2) and (A.14) .This shows that the RAFCM algorithm does not change the definition of fuzzy membership degree. Comparing Equations (A.3) and (A.20), the prototype equations are almost the same except that the fuzzy membership degrees u_{ij} have to be replaced by the adaptive fuzzy membership degrees $\bar{u}_{ij}$.

Now considering a special case (cf. Equation (A.4)), let all adaptive operators e^{t_j} be fixed to be one, or equivalently, all data factors t_j $(j = 1, \ldots, n)$ be fixed to be zero. The constraint in Equation (A.4) or Equation (A.6) is still valid, the RAFCM becomes the basic FCM right now; it tells the fact that the basic FCM is a special case of the RAFCM. In other words, the RAFCM is a kind of generalized FCM.

Compared with the FCM, a new adaptive operator vector $\mathbf{E}(e^{t_j})$ and a new parameter s $(s \neq 0)$ are introduced in the RAFCM and they jointly affect the final clustering results. In addition, it is known that the adaptive operator e^{t_j} works as a whole in the iterative process according to Equation (A.18). In this paper the adaptive operators adopt the form of exponent, but it is not absolute. It is possible to use other forms, the only condition is that the product of all operators should be one. Actually the set of parameters t_j is only used to generate the initial adaptive operators. Then these parameters are not needed again.

This paper proposes a new kind of generalized fuzzy clustering model. The FCM to RAFCM transit is a new feature in the development of clustering technologies. In the RAFCM, adaptive operators, which are used to control different influences of different data points on the clustering results, are similar to the roles that the fuzzy membership degrees play when the clustering methods change from "crisp" time to "fuzzy". The role of the new parameter s under the RAFCM is very similar to that of the exponent m in the FCM. Even in the RAFCM, fuzzy membership degree u_{ij} and the fuzzy exponent m still exist and have the same roles. Just like the fuzzy clustering algorithm and its variants has been proven by numerous literature that it can get better clustering results than the crisp model does, the RAFCM and its variants can get better clustering performances than the FCM. It is possible and reasonable.

REFERENCES

[1] Auzinger, D.—Djumlija, G. et al.: Application of Advanced Technology Packages for Improved Strip Profile and Flatness in Hot Strip Mills. AISE Steel Technology 2003, Vol. 30, pp. 17–21.

[2] Jalali, M.—Wolf, A.: Advanced Flatness in EKOs Tandem Cold Mill. La Revue de Metallurgic-CIT, Vol. 10, 2003, pp. 987–992.

[3] Gao, X.-B.—Jiao, L.-C.: A New Feature Weighted Fuzzy Clustering Algorithm. In: Slezak et al. (Eds.) Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Springer 2005, pp. 412–420.

[4] Borgelt, C.: Feature Weighting and Feature Selection in Fuzzy Clustering. [C]. Fuzzy Systems, FUZZ-IEEE 2008, Hong Kong, 2008, pp. 838–844.

[5] Chuang, K.—Tzeng, H.—Chen, S.: Fuzzy *C*-Means Clustering With Spatial Information for Image Segmentation. Computerized Medical Imaging and Graphics, Vol. 30, 2006, pp. 9–15.

[6] Yin, Z. H.—Tang, Y. G.—Sun, F. C.: Fuzzy Clustering With Novel Separable Criterion. Tsinghua Science and Technology, Vol. 11, 2006, pp. 50–53.

[7] Celikyilmaz, A.—Turksen, I. B.: Enhanced Fuzzy System Algorithms With Improved Fuzzy Clustering Algorithm. IEEE Transactions on Fuzzy Systems, Vol. 16, 2008, pp. 779–794.

[8] Wang, X.-Z.—Wang, Y. D.—Wang, L. J.: Improving Fuzzy *C*-Means Clustering Based on Feature-Weight Learning. Pattern Recognition Letters, 2004, pp. 1123–1132.

[9] Xue, Z. A.—Cen, F.—Wei, L. P.: A Weighted Fuzzy Clustering Algorithm Based on Euclidean Distance. The 15^{th} international conference on fuzzy systems and knowledge discovery, 2008, pp. 172–175.

[10] Chieh, Y. T.—Chuang, C. C.: Developing a Feature Weight Self-Adjustment Mechanism for a *K*-Means Clustering Algorithm. Computational statistics and data analysis, 2008, pp. 4658–4672.

[11] Li, J.—Gao, X.-B.—Jiao, L.-C.: A New Feature Weighted Fuzzy Clustering Algorithm. Electronic Journal, Vol. 34, 2006, pp. 9–13.

[12] Leski, J. M.: Generalized Weighted Conditional Fuzzy Clustering. Transactions on Fuzzy systems, 2003, pp. 709–715.

[13] Chan, E. Y.—Wai, K. C.—Ng, M. K. et al.: An Optimization Algorithm for Clustering Using Weighted Dissimilarity Measures. Pattern Recognition, 2004, pp. 943–952.

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

[15] Han, J.—Kamber, M.: Data Mining: Concepts and Techniques. 2^{nd} edition, Elsevier 2006.

[16] Breunig, M.—Kriegel, H.—Ng, R. T. et al.: LOF: Identifying Density-Based Local Outliers. Proceedings of ACM SIGMOD International Conference on Management of Data, Dallas, Texas, ACM Press 2000, pp. 93–104.

[17] Huang, T. Q.—Qin, X.-L.—Ye, F.-Y.: New Approach Based on Square Neighborhood to Detect Outliers. Control and Decision, Vol. 21, 2006, No. 5, pp. 145–545.

[18] Breunig, M.—Kriegel, H.—Ng, R. T. et al.: LOF: Identifying Density-Based Local Outliers. Proceedings of ACM SIGMOD International Conference on Management of Data, Dallas, Texas, ACM Press 2000, pp. 93–104.

[19] Cao, G.-Q.—Si, W.-Z.—Zhu, Y. et al.: Enhancing Effectiveness of Density-Based Outlier Mining. International Symposium on Information Processing 2008.

[20] Ghoting, A.—Parthasarathy, S.—Otey, M. E.: Fast Mining of Distance-Based Outliers in High-Dimensional Dataset. Data Mining Knowledge Discovery, Vol. 16, 2008, pp. 349–364.

[21] Weng, X.-Q.—Shen, J.-Y.: Detecting Outlier Samples in Multivariate Time Series Dataset. Knowledge-based systems, Vol. 21, 2008, pp. 807–812.

[22] Doring, C.—Lesot, M.-J.—Kruse, R.: Data Analysis With Fuzzy Clustering Methods. Computational statistical and data analysis, Vol. 51, 2006, pp. 192–214.

[23] Davè, R. N.: Characterization and Detection of Noise in Clustering. Pattern Recognition Letters, Vol. 12, 1991, pp. 657–664.

[24] Pal, N. R.—Bezdek, J. C.: On Cluster Validity for the Fuzzy *C*-Means Algorithm. IEEE Transactions on Fuzzy Systems, Vol. 3, 1995, pp. 370–378.

Chenglong Tang received the Master degree in 2004 and the Ph. D. degree in 2010 from the School of Mechanical and Dynamical Engineering of Shanghai Jiao Tong University. Now he is a Senior Engineer at Research Institute of Bao-Steel in Shanghai, China. His research interests include artificial intelligence, data mining and knowledge discovery, as well as theories and applications of steel strip production.

Shigang Wang received the Ph. D. degree from HuaZhong University of Science and Technology (HUST), Wuhan, China in 1993, as a postdoctoral fellow in Department of Mechanical Engineering at HUST from 1993 to 1995. Working in the School of Mechanical Engineering, SJTU since 1995. He is currently a Chief Professor of the Institute of Mechatronics and Automation Technology and also the Director of Research Center for Microelectronic Equipment at SJTU. His research interests include machine vision and pattern recognition, robotics, design and control of complex mech-electrical systems, etc.

Yunpeng Chen received the Bachelor degree from Chongqing University, Chongqing, China in 1994 and the Master degree from Fudan University, Shanghai, China in 2006. He is now a Senior Engineer and the Director of Cold Rolled Thin-Gauge Plate Plant of Baosteel Co., Ltd. His research interests include steel plate rolling and coating technologies.