1000-9825/2002/13(10)1905-10 ©2002 Journal of Software 软 件 学 报 Vol.13, No.10

# An Integrated Fuzzy Clustering Algorithm GFC for Switching Regressions*

WANG Shi-tong [1,2], JIANG Hai-feng [2], LU Hong-jun [2]

[1](*School of Information Technology, Southern Yangtse University, Wuxi* 214036, *China*);

[2](*Department of Computer Science, Hong Kong University of Science and Technology, Hong Kong, China*)

E-mail: wxwangst@yahoo.com.cn

Received March 29, 2001; accepted August 31, 2001

**Abstract:** In order to solve switching regression problems, many approaches have been investigated. In this paper, an integrated fuzzy clustering algorithm GFC that combines gravity-based clustering algorithm GC with fuzzy clustering is presented. GC, as a new hard clustering algorithm presented here, is based on the well-known Newton's Gravity Law. The theoretic analysis shows that GFC can converge to a local minimum of the object function. Experimental results show that GFC for switching regression problems has better performance than standard fuzzy clustering algorithms, especially in terms of convergence speed.

**Key words:** switching regression; fuzzy clustering; gravity-based clustering

Switching regression models have been extensively used in economics[1~3] and data mining in databases. Many scholars[2~9] discussed switching regression models in varying details. Let $S=\{(x_1,y_1),\ldots,(x_N,y_N)\}$ be a set of data where each independent observation $x_k \in R^t$ has a corresponding dependent observation $y_k \in R$. In switching regression models, we assume the data to be drawn from $C$ models:

$$RE_i : y = f_i(\vec{x},\vec{\beta}_i)+\varepsilon_i, \quad 1\le i\le C \tag{1}$$

where $f_i(\vec{x},\vec{\beta}_i)$ is a polynomial function about $\vec{x}$, each $\vec{\beta}_i \in \Omega \subset R^{k_i}$, $k_i \le n$, and $\varepsilon_i$ is a random vector with mean vector $u_i$=0 and covariance is $\delta_i$. When $C$=1, switching regression models become single regression model problem, in which we assume that a single functional relationship between x and $y$ holds for all the data in $S$. In a single regression model problem, the vector $\vec{\beta}$ can be well estimated using classical statistical method. However, when $C$>1, this problem becomes very subtle, that is, for a given datum $(\vec{x}_k, y_k)$, it is unknown which regression model from (1) applies.

At present, there exist three approaches for switching regression model problems. One is based on mathematical statistics, such as EM algorithm[10,11]. The second approach presented by Hathaway and Bezdek[6] is based on fuzzy clustering and the third approach is based on hard partition algorithms[1].

The work initiated by our questioning shows that the best way to solve switching regression problems may not be by using either fuzzy clustering or hard clustering only. The following example gives a very good illustration for our suspicion. In this example, we assume that all the data are generated from either of the two regression models

---

* **WANG Shi-tong** was born in 1964. He is a professor and doctorial supervisor. His research interests include AI, fuzzy systems, neural networks, pattern recognition and knowledge discovering. **JIANG Hai-feng** is a Ph.D. candidate. His research interests include pattern recognition and knowledge discovering. **LU Hong-jun** is a professor and doctorial supervisor at HKUST. His research interests include database/data warehousing, knowledge discovering and data mining, pattern recognition.

(of course, we do not know the optimal regression models for real world data):

$$y_1 = f_1(x,\vec{\beta}_1) = \beta_{11}x + \beta_{12} + \varepsilon_1,$$

$$y_2 = f_2(x,\vec{\beta}_2) = \beta_{21}x^2 + \beta_{22}x + \beta_{23} + \varepsilon_2.$$

The data points and the optimal switching regression models are shown in Fig.1.

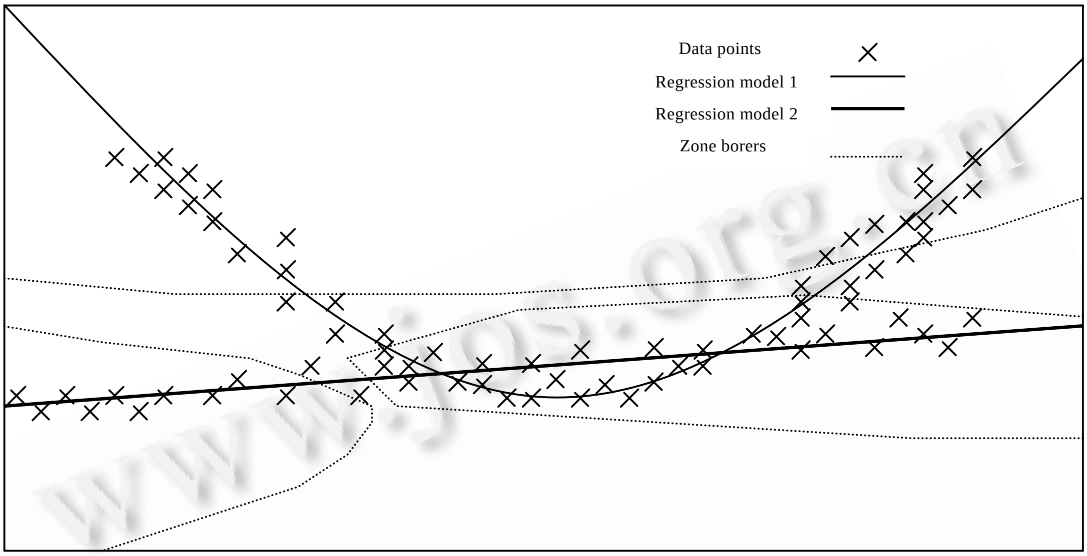



Fig.1 An example of a switching regression problem

According to Fig.1, it is obvious that data points in zones 1 and 3 should preferably be hard clustered while data points in zone 2 is better to be fuzzy clustered in order to reach the optimal state more quickly. Similar cases exist in most switching regression problems. However, if we use fuzzy clustering, then the data points in zones 1 and 3 will be unnecessarily assigned membership $(\mu)$ to the regression models that they indeed do not belong to. On the other hand, if only the hard clustering is applied, then data points in zone 2 will be hard clustered either to regression model 1 or 2 when indeed they are better off fuzzily clustered to both regression models. Therefore, in order to effectively solve switching regression problems, we should integrate hard/fuzzy clustering approaches.

The purpose of this paper is to present a new integrated approach for switching regression problems, based on gravity-based clustering and fuzzy clustering.

Our approach here keeps the advantages of Hathaway's approach[6], that is, this approach will produce estimates of $\{\vec{\beta}_1,\vec{\beta}_2,...,\vec{\beta}_c\}$ and at the same time assign a fuzzy label vector to each datum in $S$. Besides these, our approach has the following advantages over other fuzzy clustering approaches for switching regression problems:

• Gravity-Based clustering approach (GC) here is a new hard clustering one, based on well-known Newton's gravity law. It is very suitable for curve/shell clustering.

• The integrated clustering algorithm (GFC) can converge and minimize the objective function simultaneously. To best of our knowledge, to date, no one gives a clustering algorithm that combines both hard clustering and fuzzy clustering, although fuzzy clustering comes from hard clustering and their numerous variants have been presented.

• Our experiments show that GFC require fewer iterations than other fuzzy clustering approaches. It is well known that other approaches, such as EM clustering algorithm, need even more iterations than fuzzy clustering. So, our approach is superior in terms of convergence speed.

## 1 A Gravity-Based Clustering Algorithm (GC)

Consider a regression model:

$$y = f_i(\vec{x}, \vec{\beta}_i) + \varepsilon_i, i \in \{1,2,\ldots,C\}. \tag{2}$$

In general, we take *y* as a line or ellipse or other curves, as shown in Fig.2. We regard curve 1 as object 1 and curve 2 as object 2. These two objects will produce gravity force for data point A. As the result of gravity force acting on A, data point A will obviously be clustered into curve 1.

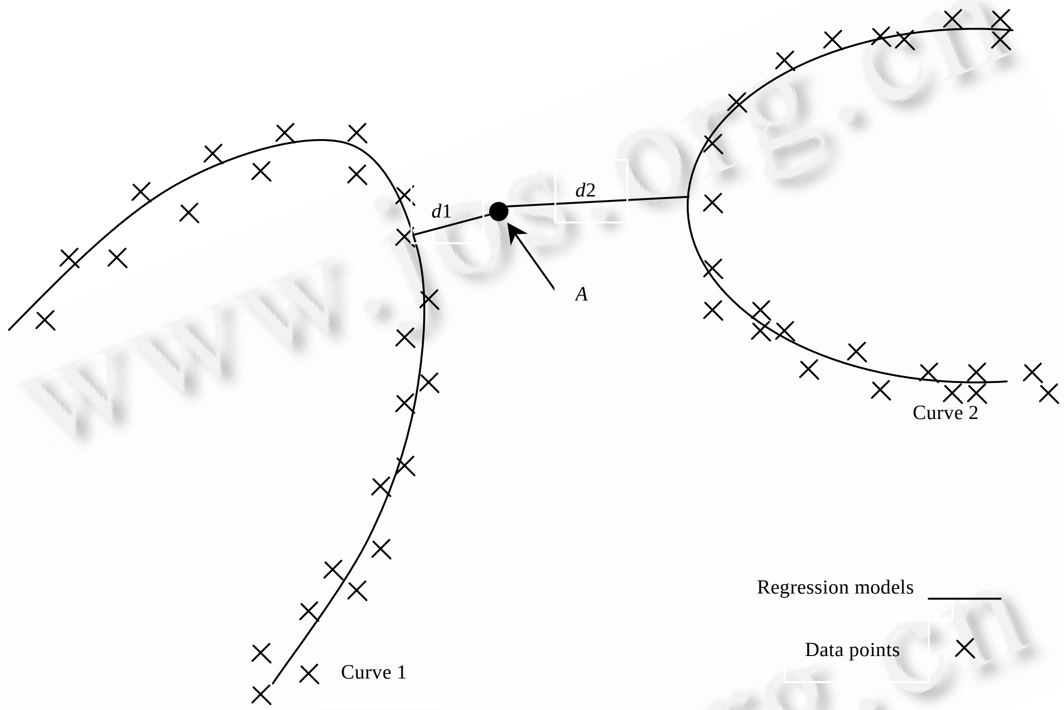



Fig.2 An example of GC clustering

It is well-known that Newton's gravity law can be formulated as:

$$F = \frac{k \cdot m_1 \cdot m_2}{d^2} \tag{3}$$

where $F$ denotes the gravity force between object 1 with mass $m_1$ and object 2 with mass $m_2$, $d$ denotes the distance between object 1 and object 2, $k$ is a coefficient and $k$=2. In order to effectively apply Newton's gravity law in our gravity-based clustering problem, we make the following assumptions:

- The quality of each data point is 1.
- $m_i^{(t)}$ data points have been clustered into curve $i(i=1,2)$ at time $t$, as shown in Fig.3.
- Each data point belonging to a cluster has the same potential. Based on this, when calculating the gravity force, we assume all data points flow into a point and its mass is the number of all current data points in this curve, as shown in Fig.3.

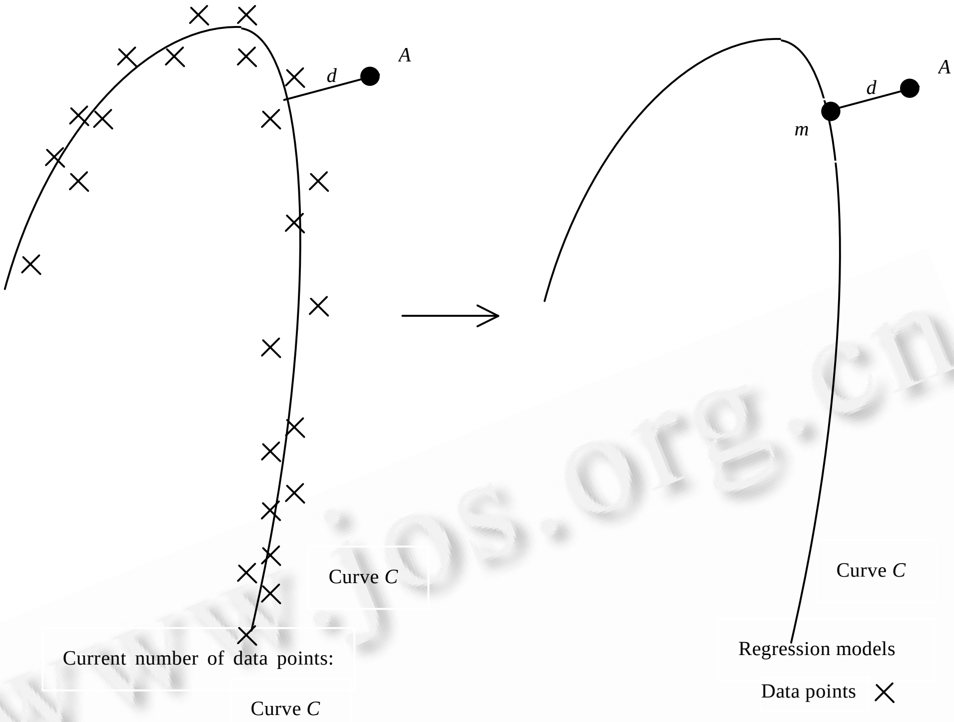



Fig.3 A cluster with $m$ points is regarded as a cluster with one point with mass $m$

Based on the above assumptions, the gravity force $F^{(t)}$ between point $A=(x_k,y_k)$ and a curve, say $c=f(\vec{x},y)=0$, can be calculated as:

$$F^{(t)}=k\frac{1\times m_c^{(t)}}{d^2}=\frac{2m_c^{(t)}}{d^2}=\frac{2m_c^{(t)}}{(y_k-f(\vec{x}_k,\vec{\beta}_c))^2}. \tag{4}$$

Let us define the objective function $J_{GC}$ as:

$$J_{GC}=\sum_{i=1}^{C}\sum_{(\vec{x}_k,y_k)\in S_i}d^2\left[(\vec{x}_k,y_k),RE_i\right] \tag{5}$$

where $S_i$ denotes the set of data points clustered to $i$th regression model $RE_i$, $d(\vec{x}_k,RE_i)$ is the square distance between $(\vec{x}_k,y_k)$ and the regression model $RE_i$. In terms of (1), we have

$$J_{GC}=\sum_{i=1}^{C}\sum_{(\vec{x}_k,y_k)\in S_i}(y_k-f_i(\vec{x}_k,\vec{\beta}_i))^2. \tag{6}$$

Now we present the following gravity-based clustering algorithm GC for finding the approximation minimum of $J_{GC}$.

**Algorithm 1**. Gravity-Based clustering algorithm (GC).

1. Given data set $S=\{(\vec{x}_1,y_1),(\vec{x}_2,y_2),\ldots,(\vec{x}_N,y_N)\}$
2. Fix $C(2\le C\le N)$ and set sensitivity parameter $\gamma$, which must be determined by experts/users
3. Initialize $\vec{\beta}^{(0)}=(\vec{\beta}_1^{(0)},\vec{\beta}_2^{(0)},\ldots,\vec{\beta}_C^{(0)})^T, t=0$
4. Initialize $S_i=\{(x_i^*,y_i^*)\,|\,d^2((\vec{x}_i^*,y_i^*),RE_i)=\min_k d^2((\vec{x}_k,y_k),RE_i)\}, i=1,2,\ldots,C$
5. for $j$=1 to $N$ do
6.    for $k$=1 to $C$ do
7.       if $d^2[(\vec{x}_j,y_j),RE_k]<\gamma$ then

8. $\quad\quad S_k = S_k \cup \{(\vec{x}_j, y_j)\}$
9. $\quad$ end if
10. end for
11. Find $i(1 \le i \le C)$ that satisfies: $\dfrac{2m_i^{(t)}}{d^2[(\vec{x}_j, y_j), RE_i]} = \max_k \left\{ \dfrac{2m_k^{(t)}}{d^2[(\vec{x}_j, y_j), RE_k]} \right\}$, where $m_k^t = \|S_k\|$, i.e. $m_k^t$ is the number of elements in $S_k$.
12. $S_i = S_i \cup \{(\vec{x}_j, y_j)\}$
13. end for
14. Compute $\vec{\beta}^{(t+1)} = (\vec{\beta}_1^{(t+1)}, \vec{\beta}_2^{(t+1)}, \ldots, \vec{\beta}_C^{(t+1)})^T$, by solving the following linear systems:

$$\frac{\partial J_{GC}}{\partial \vec{\beta}_i} = 2 \sum_{(\vec{x}_k, y_k) \in S_i} (f_i(\vec{x}_k, \vec{\beta}_i) - y_k) \frac{\partial f_i(\vec{x}_k, \vec{\beta}_i)}{\partial \vec{\beta}_i} = 0 \tag{7}$$

15. if $\left\|\vec{\beta}^{(t+1)} - \vec{\beta}^{(t)}\right\| < \varepsilon$ then
16. $\quad$ Stop.
17. else
18. $\quad t \leftarrow t+1$, goto Line 4.
19. end if

The GC Algorithm shown above is quite reasonable from the following intuitive viewpoints:

• It uses Newton's Gravity Law to cluster data points while this famous law seems to be suitable for such clustering according to the discussion above.

• In order to minimize $J_{GC}$, all $\dfrac{\partial J_{GC}}{\partial \vec{\beta}_j}$ should be zero. Once $S_i$ is determined, (7) can be solved, i.e. $\vec{\beta}^{(t+1)}$ can be calculated.

• If $RE_i$ and $RE_j$ are intersected, the sensitivity parameter $\gamma$, which is very small, can make sure that a data point very close to $RE_i$ and $RE_j$ can be simultaneously classified into $RE_i$ and $RE_j$. The circled data points in Fig.4 would be simultaneously "hard" clustered to both the clusters.

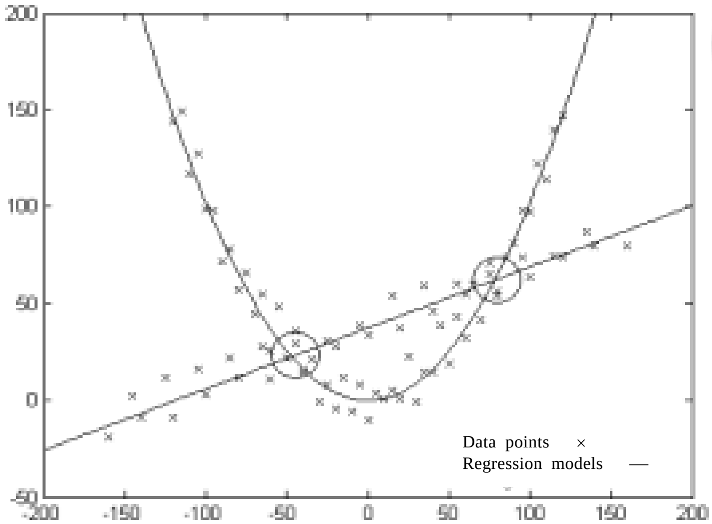



Fig.4 Points in a circle will be clustered to both clusters due to the sensitivity parameter $\gamma$

## 2 Integrated Fuzzy Clustering Algorithm GFC

Since Bezdek presented fuzzy *c*-means clustering algorithm[6], many of its variants have been proposed to improve its performance to satisfy different requirements. However, to date, no one has integrated it with the hard clustering algorithm. In this section, we will present the new algorithm GFC that integrates fuzzy *c*-means clustering with hard clustering. As shown in the previous analysis, GFC algorithm seems to be very rational and efficient for switching regression problems.

It is easy to extend standard fuzzy *c*-means clustering to make it suitable for switching regression problems. According to standard fuzzy *c*-means algorithm, the objective function $J_{CM}$, for a switching regression problem, is defined as:

$$J_{CM}=\sum_{i=1}^{C}\sum_{j=1}^{N}(\mu_{ij})^m d^2[(\vec{x}_j,y_j),RE_i] \tag{8}$$

where

$$\mu_{ij}=1/T,T=\sum_{k=1}^{C}\left(\frac{d[(\vec{x}_j,y_j),RE_i]}{d[(\vec{x}_j,y_j),RE_k]}\right)^{\frac{2}{m-1}} \tag{9}$$

and $m\in(1,\infty),\sum_{i=1}^{C}\mu_{ij}=1$.

In GFC, we define the new objective function:

$$J_{GFC}=J_{GC}\cdot J_{CM}\ . \tag{10}$$

Based on (10), we present the new integrated fuzzy clustering algorithm GFC as Algorithm 2.

**Algorithm 2**. An integrated clustering algorithm GFC.

1. Given data set $S=\{(\vec{x}_1,y_1),(\vec{x}_2,y_2),\ldots,(\vec{x}_N,y_N)\}$
2. Fix $C(2\le C\le N)$ and set sensitivity parameter $\gamma$
3. Initialize $\vec{\beta}^{(0)}=(\vec{\beta}_1^{(0)},\vec{\beta}_2^{(0)},\ldots,\vec{\beta}_C^{(0)})^T,t=0$
4. Initialize $S_i=\{(x_i^*,y_i^*)\,|\,d^2((\vec{x}_i^*,y_i^*),RE_i)=\min_k d^2((\vec{x}_k,y_k),RE_i)\},i=1,2,\ldots,C$
5. for *j*=1 to *N* do
6.     for *k*=1 to *C* do
7.         if $d^2[(\vec{x}_j,y_j),RE_k]<\gamma$ then
8.             $S_k=S_k\cup\{(\vec{x}_j,y_j)\}$
9.         end if
10. end for
11. Find $i(1\le i\le C)$ that satisfies: $\dfrac{2m_i^{(t)}}{d^2[(\vec{x}_j,y_j),RE_i]}=\max_k\left\{\dfrac{2m_k^{(t)}}{d^2[(\vec{x}_j,y_j),RE_k]}\right\}$
12. $S_i=S_i\cup\{(\vec{x}_j,y_j)\}$
13. end for
14. Compute $J_{GC}$ using (5)
15. for *i*=1 to *C* do
16.     for *j*=1 to *N* do
17.         Compute $\mu_{ij}$ using (9)
18.     end for
19. end for
20. Compute $J_{CM}$ using (8)
21. Compute $\vec{\beta}^{(t+1)}$ by solving the following linear systems:

$$\frac{\partial J}{\partial\vec{\beta}_i}=J_{CM}\cdot\frac{\partial J_{GC}}{\partial\vec{\beta}_i}+J_{GC}\cdot\frac{\partial J_{CM}}{\partial\vec{\beta}_i}=0$$

i.e. $$J_{CM}\sum(f_i(\vec{x}_k,\vec{\beta}_i)-y_k)\cdot\frac{\partial f_i(\vec{x}_k,\vec{\beta}_i)}{\partial\vec{\beta}_i}+J_{GC}\sum\mu_{ij}^m(f_i(\vec{x}_j,\vec{\beta}_i)-y_j)\cdot\frac{\partial f_i(\vec{x}_j,\vec{\beta}_i)}{\partial\vec{\beta}_i}=0 \tag{11}$$

22. if $\left\|\vec{\beta}^{(t+1)}-\vec{\beta}^{(t)}\right\|<\varepsilon$ then
23. Stop.
24. else
25. $t\leftarrow t+1$, goto Line 4.
26. end if

In GFC, the definition of the objective function $J$ is rational since a clustering algorithm should minimize $J_{GC}$ and $J_{CM}$ simultaneously in order to solve the switching regression problem efficiently. Now we investigate the convergence properties of this new integrated clustering algorithm.

**Theorem 1**. In GFC, $\mu_{ij}(i=1,2,...,C,\ j=1,2,...,N)$ *and* $\vec{\beta}=(\vec{\beta}_1,\vec{\beta}_2,...,\vec{\beta}_C)^T$ is a local minimum for $J_{GFC}$ only if

$$\mu_{ij}=\frac{1}{\sum_{k=1}^{C}\left(\frac{d[(\vec{x}_j,y_j),RE_i]}{d[(\vec{x}_j,y_j),RE_k]}\right)^{\frac{2}{m-1}}} \tag{12}$$

subject to $\sum_{i=1}^{C}\mu_{ij}=1$ and $\vec{\beta}$ is a solution of (11).

*Proof*. First, we assume that $\vec{\beta}$ is fixed. Then the problem is to minimize $J_{GFC}$ with respect to $\mu_{ij}$ under the constraint $\sum_{i=1}^{C}\mu_{ij}=1$. Using Lagrange multiplier method, we find that the problem is equivalent to minimizing

$$L(\vec{\beta},\lambda)=J_{CM}\cdot J_{GC}-\sum_{j=1}^{N}\lambda_j\left(\sum_{i=1}^{C}\mu_{ij}-1\right) \tag{13}$$

without constraints. The necessary condition of this problem is

$$\frac{\partial L(\vec{\beta},\lambda)}{\partial\mu_{ij}}=J_{GC}\cdot m(\mu_{ij})^{m-1}d^2((\vec{x}_j,y_j),RE_i)-\lambda_j=0 \tag{14}$$

$$\frac{\partial L(\vec{\beta},\lambda)}{\partial\lambda_j}=\sum_{i=1}^{C}\mu_{ij}-1=0 \tag{15}$$

From (14), we have

$$\mu_{ij}=\left[\frac{\lambda_j}{m\cdot J_{GC}\cdot d^2[(\vec{x}_j,y_j),RE_i]}\right]^{\frac{1}{m-1}}. \tag{16}$$

Substituting (16) into (15), we have

$$\left(\frac{\lambda_j}{m}\right)^{\frac{1}{m-1}}=\frac{1}{\sum_{k=1}^{C}\left(\frac{1}{J_{GC}d[(\vec{x}_i,y_i),RE_k]^2}\right)^{\frac{1}{m-1}}}. \tag{17}$$

Substituting (17) into (16), we get (12).

In order to show that $\vec{\beta}$ must be a solution of (11), we assume that $\mu_{ij}$ is fixed. Thus this is an unconstrained minimizing problem, and the necessary condition is

$$\frac{\partial J_{GFC}}{\partial\vec{\beta}}=0, \tag{18}$$

i.e.
$$\frac{\partial(J_{GC}\cdot J_{CM})}{\partial\vec{\beta}}=\frac{\partial J}{\partial\vec{\beta}}=J_{CM}\cdot\frac{\partial J_{GC}}{\partial\vec{\beta}}+J_{GC}\cdot\frac{\partial J_{CM}}{\partial\vec{\beta}}=0 \tag{19}$$

from which we get (11). The theorem is proved. □

**Theorem 2.** Let $\Phi(U)=J_{GC}\cdot J_{CM}$, where $U=[\mu_{ij}]_{C\times N}$, $\vec{\beta}$ is fixed, and $d[(\vec{x}_j,y_j),RE_i]\neq 0$, for all $1\le i\le C$, $1\le j\le N$, then $U$ is a local minimum of $\Phi(U)$ if and only if $\vec{\beta}$ is computed via (11).

*Proof.* The only-if part has been proved in Theorem 1. To show the sufficiency, we examine $H(\Phi)$, the $CN\times CN$ Hessian of the Lagrangian of $\Phi(U)$ evaluated at the $U$ given by (12). From (13), we have

$$h_{st,ij}(U)=\frac{\partial}{\partial\mu_{st}}\left[\frac{\partial\Phi(U)}{\partial\mu_{ij}}\right]$$

$$=\begin{cases}m(m-1)(\mu_{st})^{m-2}d^2[(\vec{x}_t,y_t),RE_s]\times J_{GC}, s=i,t=k\\ 0,\text{otherwise}\end{cases} \tag{20}$$

where $u_{st}$ is computed from (12). Thus, $H(U)=[h_{st,ij}(U)]$ is a diagonal matrix. Since $m>1$, and $d((\vec{x}_t,y_t),RE_s)$ for all $1\le t\le N$ and $1\le s\le C$, and $J_{GC}\ge 0$. We know from the above formula that Hessian $H(U)$ is positive definite and consequently, (12) is also a sufficient condition for minimizing $\Phi(U)$. □

**Theorem 3**. Let $\Phi(\vec{\beta})=J_{GC}\times J_{CM}=J_{GC}(\vec{\beta})J_{CM}(U,\vec{\beta})$, where $U=[\mu_{ij}]_{C\times N}$ is fixed, $d[(\vec{x}_j,y_j),RE_i]\neq 0$ for $1\le i\le C$ and $1\le j\le N$, and $m>1$. Then $\vec{\beta}$ is a local minimum of $\Phi(\vec{\beta})$ if and only if $\vec{\beta}$ is computed via (11).

*Proof.* The necessity was proven in Theorem 2. To show the sufficiency, we have, from (18) that

$$\frac{\partial}{\partial\vec{\beta}_i}\left[\frac{\partial\Phi(\vec{\beta})}{\partial\vec{\beta}_j}\right]=J_{GC}\frac{\partial^2 J_{CM}}{\partial\vec{\beta}_i\cdot\partial\vec{\beta}_j}+J_{CM}\frac{\partial^2 J_{GC}}{\partial\vec{\beta}_i\cdot\partial\vec{\beta}_j}=0\,. \tag{21}$$

Since $f_i(\vec{x}_k,\vec{\beta}_i)$ is a polynomial function about $\vec{x}_k$, it is very easy to prove:

$$J_{GC}\frac{\partial^2 J_{CM}}{\partial\vec{\beta}_i\cdot\partial\vec{\beta}_j}=\begin{cases}>0, & i=j\\ 0, & \text{otherwise}\end{cases}. \tag{22}$$

$$J_{CM}\frac{\partial^2 J_{GC}}{\partial\vec{\beta}_i\cdot\partial\vec{\beta}_j}=\begin{cases}>0, & i=j\\ 0, & \text{otherwise}\end{cases}. \tag{23}$$

i.e. the Hessian is positive definite and consequently (11) is a sufficient condition for minimizing $\Phi(\vec{\beta})$.

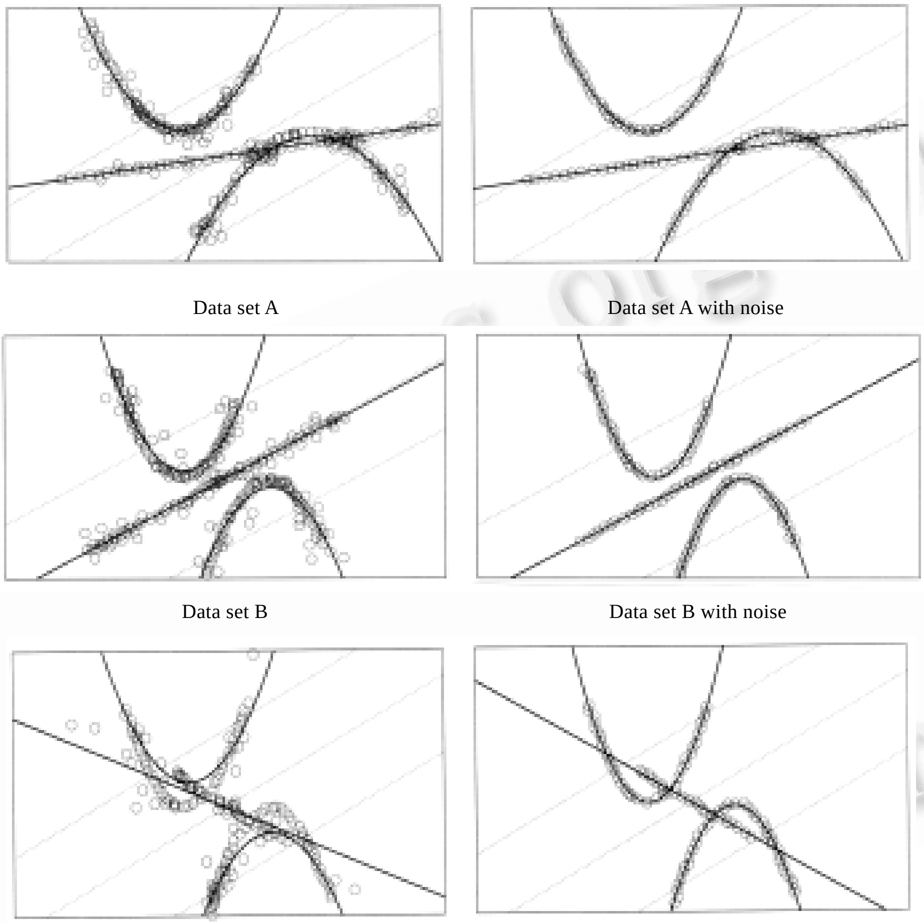
With Theorems 2 and 3, we can prove that

$$J_{GFC}(U^{t+1},\vec{\beta}^{(t+1)})=J_{GC}(\vec{\beta}^{(t+1)})\cdot J_{CM}(U^{t+1},\vec{\beta}^{(t+1)})\le J(U^t,\vec{\beta}^{(t)})$$
$$=J_{GC}(\vec{\beta}^{(t)})\cdot J_{CM}(U^t,\vec{\beta}^{(t)})\,. \tag{24}$$

In other words, $J_{GFC}$ is a decreasing function with $t$. So, the GFC algorithm will finally converge. Since the proof of (24) is similar to Bezdek's proof in Ref.[6], it is omitted here. □

## 3 Simulations

In this section, we use the numerical simulation results illustrate the effectiveness of algorithm GFC. This experiment deals with structure mining involving the mixture of curve and line under three different noise free data sets and the same data sets with noise. The data sets named as A, B, and C are shown in Fig.5. The dashed lines in Fig.5 are the initialization models. We run both the GFC and FCRM algorithm (fuzzy *c*-means clustering) with the same data and initial models. The two algorithms are nearly equally powerful in terms of finding the terminal regression models while our GFC algorithm always needs less iterations. The experimental result is shown in Fig.5 and the number of iterations needed by each algorithm for each sub-experiment (totally 6 cases) is illustrated in Fig.6. Moreover, as we can see, our GFC algorithm converges to an acceptable result models even in noisy situations.



Data set A | Data set A with noise

Data set B | Data set B with noise

Data set C | Data set C with noise

Fig.5 Initial (dashed lines) and terminal models (solid arcs) for different data sets and initializations

In most cases, algorithm GFC detects and characterize the quadratic/linear models generating these data sets correctly. Although FCRM algorithm can finally characterize the quadratic/linear models even with almost the same effectiveness as GFC does, we can obviously see that GFC prevails over FCRM in terms of the convergence speed, which can be seen in Fig.6.

## 4 Conclusions

In this paper, we present a new integrated fuzzy clustering algorithm GFC. GFC combines gravity-based clustering algorithm GC with fuzzy clustering. GC as a new hard clustering algorithm is based on the well-known Newton's Gravity Law. Our theoretic analysis shows that GFC can converge to a local minimum of the objective function. Simulations are done to show the validity and effectiveness of our GFC algorithm. We find that GFC algorithm detects and characterizes the quadratic/linear models generating the data sets used in the examples correctly in most cases. We also run the fuzzy c-means algorithm (FCRM) on the same data and find that GFC prevails considerably over FCRM in terms of converge speed.

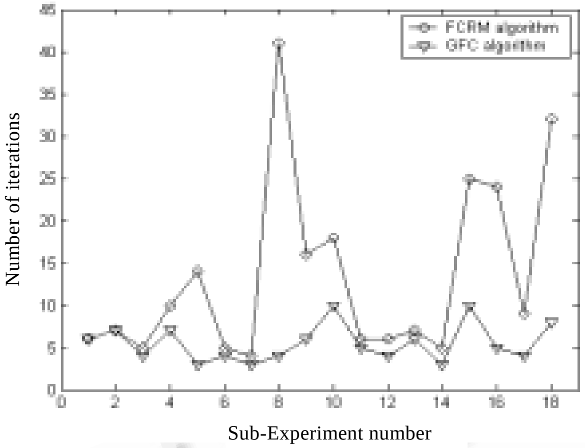



Fig.6 Convergence speed of FCRM and GFC

**References:**

[1] Hamermesh, D.S. Wage bargains, threshold effects, and the Phillips curve. Quarterly Journal of Austrian Economics, 1970,84: 501~517.

[2] Quandt, R.E. A new approach to estimating switching regressions. Journal of The American Statistical Association, 1972,67: 306~310.

[3] Quandt, R.E., Ramsey, J.B. Estimating mixtures of normal distributions and switching regressions. Journal of the American Statistical Association, 1978,73:730~752.

[4] Caudill, S.B., Acharyal, R.N. Maximum-Likelihood estimation of a mixture of normal regressions: starting values and singularities. Communications in Statistics, Simulation and Computation, 1998,27(3):667~674.

[5] Dennis, J.E. Algorithms for nonlinear fitting. NATO Advanced Research Symposium, 1981. 307~315.

[6] Hathaway, R., Bezdek, J. Switching regression models and fuzzy clustering. IEEE Transactions on Fuzzy Systems, 1993,1(3): 195~204.

[7] Hosmer, D.W. Maximum likelihood estimates of the parameters of a mixture of two regression lines. Communications in Statistics, 1974,3(10):995~1005.

[8] Kiefer, N.M. Discrete parameter variation: efficient estimation of a switching regression model. Econometrica, 1978,46(2): 427~434.

[9] Leung, Y. Ma, J.-H., Zhang, W.-X. A new method for mining regression classes in large data sets. IEEE Transactions on Pattern Analysis Machine Intelligence, 2001,23(1):5~21.

[10] Aitkin, M., Hector, G.T. Mixture models, outliers, and the algorithm. Technometrics, 1980,22(3):325~331.

[11] Redner, R.A., Walker, H.F. Mixture densities, maximum likelihood and the em algorithm. SIAM Review, 1984,26(2):195~239.

[12] Dave, R.N., Krishnapuram, R. Robust clustering methods: a unified view. IEEE Transactions on Fuzzy Systems, 1997,5(2): 270~293.

[13] Fayyad, U., Stolorz, P. Data mining and KDD: promise and challenges. Future Generation Computer Systems, 1997,13:99~115.

[14] Krishnapuram, R., Keller, J. A possibilistic approach to clustering. IEEE Transactions on Fuzzy Systems, 1993,1:98~110.

[15] Veaus, R.D. Mixtures of linear regressions. Computational Statistics and Data Analysis, 1989,8:227~245.

# 关于切换回归的集成模糊聚类算法 GFC

王士同[1,2], 江海峰[2], 陆宏钧[2]

[1](江南大学 信息学院,江苏 无锡 214036);

[2](香港科技大学 计算机系,香港)

摘要: 已经有多个方法可用于解决切换回归问题.根据所提出的基于 Newton 引力定理的引力聚类算法 GC,结合模糊聚类算法,进一步提出了新的集成模糊聚类算法 GFC.理论分析表明 GFC 能收敛到局部最小.实验结果表明 GFC 在解决切换回归问题时,比标准模糊聚类算法更有效,特别在收敛速度方面.

关键词: 切换回归;模糊聚类;引力聚类

中图法分类号: TP18 文献标识码: A