Trend Relational Analysis and Grey-Fuzzy Clustering Method*

Zhijie Chen[1], Weizhen Chen[2,3], Qile Chen[4] and Mian-Yun Chen[2]

Abstract
In this paper, the latest researches on trend relational analysis and grey-fuzzy clustering method are presented. Grey systems and fuzzy systems are found everywhere. The researched results can open new prospects for the development and application of systems methodology to data mining.

Keywords Trend relational analysis, grey-fuzzy clustering method, grey systems, fuzzy systems.

1 Introduction

It is well known that the notion of a system is rather broad, and can be traced to antiquity. So to speak, any an object investigated, such as the motion of a macroscopic particle, or some socioeconomic phenomenon, may be qualified as a system. Systems such as social, economic, agricultural, industrial, ecological, and biological systems are usually those of great complexity. These complicated objects apparently have the following characteristic features:

- There is no physical prototype;
- The operation mechanism is not clear;
- The relationships between the inputs and the outputs are not obvious;
- The indeterminateness is very strong;
- Oft times, only a few of discrete data observed can be obtained.

Thus, in the studies of such objects, how to build a system model, how to forecast, how to make decision and how to control rely to a great extent on the use of information from objective reality. From a practical point of view, however, it is very difficult, or impossible, to get the adequate or complete information from investigated object in many situations. The phenomenon with incomplete information, therefore, is usually encountered. "Incomplete information" is the fundamental meaning of being "grey". The name of "grey system" is chosen based on the amount of known information. Consider a "black box" stands for an object such that its internal structure is totally unknown to the investigator. Here, the word "black" represents unknown information, "white" for completely known information, and "grey" for those information which are partially known and partially unknown. Accordingly, systems with completely known information are called as white systems. Systems with completely unknown information as black systems, and the systems with partially known and partially unknown information as grey systems, respectively [1,2].

Since 1982, there has been a quick development in grey systems theory[3,4] in China, and it is also very successful in the application of the theory to many real projects, such as agriculture, society, economics, engineering, IT, data mining, management, biological

* This project was supported by the National Natural Science Foundation of China(69874018, 79970025), GF-PR-528.
1 School of Economics and Management, Wuhan University of Science and Engineering, Wuhan 430073, China. Email:chen_zj@tom.com
2 Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China. Email:mychen@mail.hust.edu.cn. Email:chenwz_2005@126.com
3 Department of Information and Electrical Engineering, Wuhan Polytechnic University, Wuhan 430023, China. Email:chenwzh@whpu.edu.cn
4 Department of Computer Science, Wuhan University of Science and Engineering, Wuhan 430073, China. Email:cql@wuse.edu.cn

protection, ecology, environmental studies, etc.. After over 20 years of rapid development, the theory of grey systems consists of the following main blocks of concepts and results:

- Foundation, consisting of grey numbers, grey elements and grey relations;
- Grey systems analysis, including grey incidence analysis, grey statistics, grey clustering, etc.;
- Grey systems modeling, through the use of generation of grey numbers or functions so that hidden patterns can be found;
- Grey prediction;
- Grey decision making;
- Grey control;
- Grey process;
- Integral generating transform;
- Trend relational analysis; and
- Systems clouds.

The reader who takes an interest in this subject should refer to Kybernetes, Vol.33, No.2, 2004.(《Grey Systems Theory and Applications》,Guest Editors: Mian-Yun Chen, Sifeng Liu, and Yi Lin).

In this paper, the latest researches on trend relational analysis and grey-fuzzy clustering method are presented. The researched results here can open new prospects for the development and application of systems methodology to data mining.

2 On Grey Process

An excessively complex or complicated object, which generally shows a lack of completed model information, may be looked upon as a grey system. That is, with the aid of the grey systems approach we are able to solve the problems of the analysis and design of complicated systems, or excessively complex systems, including the data systems. Such a system might be looked upon as a data organizing framework according to which some data are considered to be relevant, others not. From a grey system's point of view, all the indeterminate or random concepts can be regarded to be grey. In order to describe an investigated object with incomplete information and to get a reasonably stable picture which can be communicated, some grey concepts are defined as follows:

- The most basic ingredient, which exists in a grey system, with incomplete information is called the grey element, denoted by $\otimes$;
- An indeterminate variable whose amplitude varies over a suitable range is called a grey variable, denoted by $X(\otimes)$;
- A function defined on the Cartesian product space $\otimes_{\Sigma} \times T$, denoted by $X(\otimes, t)$, is called a grey process, where $\otimes \in \otimes_{\Sigma}$ is a grey element, and $t \in T$ represents time;
- The observable output of an investigated object, which is a function of t, denoted by $X^{(0)}(t)$, is called a whitening function of grey process, $X^{(0)}(t) \in X(\otimes, t)$.

More often than not, only a few of the discrete data observed from the investigated object can be obtained, such as

$$X^{(0)} = \left\{ X^{(0)}(k) \mid X^{(0)}(k) \geq 0, k = 1,2,\ldots,n \right\}$$

which is called the original time series corresponding to $X^{(0)}(t)$. We consider that, for an investigated object, all behavioral information is contained in $X(\otimes, t)$, and all relevant information through observation is contained in $X^{(0)}(t)$ or $X^{(0)}(k)$. Thus, $X^{(0)}(t)$ or $X^{(0)}(k)$ has provided the basis for constructing systems model[5].

3 Trend Relational Analysis for Grey Systems

Let us consider dynamic relationships between h factors that are present in an investigated object. Naturally, we can get

$$X_i^{(0)}(k), k = 1,2,\ldots,n, \qquad i = 1,2,\ldots,h.$$

Definition 3.1 *We call* $X_r^{(0)}$ *the reference factor (choose freely),* $X_c^{(0)}$ *the compared factor,* $r, c \in \{1,2,\ldots,h\}$. *Correspondingly,* $\{X_r^{(0)}(k)\}$ *is called the reference time series,* $\{X_c^{(0)}(k)\}$ *the compared time series.*

In order to express the approximateness and similarity between $\{X_r^{(0)}(k)\}$ and $\{X_c^{(0)}(k)\}$, we proceed to processing of data as follows:

Definition 3.2 *All of the following transformations are called to be mapping of quantity:*

$$M_r : X_r^{(0)} \times X_r^{(0)} \to \Delta X_{rr}^{(0)}$$

$$M_c : X_c^{(0)} \times X_c^{(0)} \to \Delta X_{cc}^{(0)}$$

$$M_{rc} : X_r^{(0)} \times X_c^{(0)} \to \Delta X_{rc}^{(0)}$$

where

$$X_r^{(0)} = \{X_r^{(0)}(1), X_r^{(0)}(2), \ldots, X_r^{(0)}(n)\}$$

$$X_c^{(0)} = \{X_c^{(0)}(1), X_c^{(0)}(2), \ldots, X_c^{(0)}(n)\}$$

$$\Delta X_{rr}^{(0)} = \{\Delta X_{rr}^{(0)}(2), \Delta X_{rr}^{(0)}(3), \ldots, \Delta X_{rr}^{(0)}(n)\}$$

$$\Delta X_{cc}^{(0)} = \{\Delta X_{cc}^{(0)}(2), \Delta X_{cc}^{(0)}(3), \ldots, \Delta X_{cc}^{(0)}(n)\}$$

$$\Delta X_{rc}^{(0)} = \{\Delta X_{rc}^{(0)}(1), \Delta X_{rc}^{(0)}(2), \ldots, \Delta X_{rc}^{(0)}(n)\}$$

$$\Delta X_{rr}^{(0)}(k) = X_r^{(0)}(k) - X_r^{(0)}(k-1)$$

$$\Delta X_{cc}^{(0)}(k) = X_c^{(0)}(k) - X_c^{(0)}(k-1)$$

$$\Delta X_{rc}^{(0)}(k) = X_r^{(0)}(k) - X_c^{(0)}(k).$$

Definition 3.3 *Let* $\underline{Mrc}$ *be a mapping. If*

$$\underline{Mrc} : \Delta X_{rc}^{(0)} \times \Delta X_{rr}^{(0)} \times \Delta X_{cc}^{(0)} \to \xi(k) \in [0,1],$$

then we call $\xi(k)$ *the trend relational function of both* $\{X_r^{(0)}(k)\}$ *and* $\{X_c^{(0)}(k)\}$.

Theorem 3.1 *Based on* $\{X_r^{(0)}(k)\}$ *and* $\{X_c^{(0)}(k)\}$, *if*

$$\xi_{rc}(k) = \xi_{rc}\left(\{X_r^{(0)}(k)\}, \{X_c^{(0)}(k)\}\right)$$
$$= \left[1 + \beta \,|\, \Delta X_{rc}^{(0)}(k) + \Delta X_{rc}^{(0)}(k-1) \,|\, + \gamma \,|\, \Delta X_{rr}^{(0)}(k) - \Delta X_{cc}^{(0)}(k) \,|\, \right]^{-1}$$
$$k = 2,3,\ldots,n, \beta, \gamma \in [0,1],$$

then $\xi_{rc}(k)$ *is a kind of trend relational functions.*

Proof. The minimal amount of information about $\{X_r^{(0)}(k)\}$ and $\{X_c^{(0)}(k)\}$ is involved in $\xi_{rc}(k)$. Thus, $\xi_{rc}(k)$ can express the dynamic relationship between $\{X_r^{(0)}(k)\}$ and $\{X_c^{(0)}(k)\}$ sufficiently, and meet the conditions in Definition 3.3. $|\Delta X_{rr}^{(0)}(k) - \Delta X_{cc}^{(0)}(k)| \Rightarrow$ similarity; $|\Delta X_{rc}^{(0)}(k) + \Delta X_{rc}^{(0)}(k-1)| \Rightarrow$ approximateness on certain similarity; $\{X_r^{(0)}(k)\} = \{X_c^{(0)}(k)\} \Rightarrow \xi_{rc}(k) = 1$ [6-11].

Corollary 3.1 *If* $\xi_{rc}(k)$ *=const,* $k = 2,3,\ldots,n$, *then* $\{X_c^{(0)}(k)\}$ *is completely similar to* $\{X_r^{(0)}(k)\}$.

Definition 3.4 $\xi_{rc}(k)$ *is a trend relational function between* $\{X_r^{(0)}(k)\}$ *and* $\{X_c^{(0)}(k)\}$, *then*

$$\Xi_{rc} = \sum_{k=2}^{n} \xi_{rc}(k)m(k) / \sum_{k=2}^{n} m(k)$$

is defined as the trend relational grade, where $m(k) = m[k-1, k]$ *as a measure.*

Corollary 3.2 *Taking* $m(k) = 1$ *, the trend relational grade is*

$$\Xi_{rc} = (1/(n-1)) \sum_{k=2}^{n} \xi_{rc}(k) .$$

Consider that $r = 1,2,\ldots, j, c = 1,2,\ldots,h$, we can obtain the trend relational grade matrix as follows:

$$M_{RC} = \begin{bmatrix} \Xi_{11} & \Xi_{12} & \cdots & \Xi_{1h} \\ \Xi_{21} & \Xi_{22} & \cdots & \Xi_{2h} \\ \vdots & \vdots & \cdots & \vdots \\ \Xi_{j1} & \Xi_{j2} & \cdots & \Xi_{jh} \end{bmatrix} .$$

Theorem 3.2 *The maximum element in* M_{RC} *is the optimal* Ξ_{rc} *with the system modeling criterion, denoted by*

$$\Xi_{rc.om} = \max_{r=1,2,\ldots,j} \left\{ \max_{c=1,2,\ldots,h} \{\Xi_{rc}\} \right\} .$$

Proof. Omitted.

Theorem 3.3 *Let*

$$\xi_{rc.\min}(k) = \min\{\xi_{rc}(k)\}$$

$$\xi_{rc.\max}(k) = \max\{\xi_{rc}(k)\} .$$

If $\xi_{rc}(k)$ *is satisfied with the criterion below*

$$J_{opt} = \min\{\xi_{rc.\max}(k) - \xi_{rc.\min}(k)\} ,$$

then it is called the optimal similar $\xi_{rc}(k)$*, denoted by* $\xi_{rc.os}$ *, where* $c = 1,2,\ldots,h$ *,* $k = 2,3,\ldots,n$ *. Correspondingly,* $\Xi_{rc.os}$ *is the optimal similar* Ξ_{rc} *.*

Proof. Omitted.

$\Xi_{rc.om}$ and $\Xi_{rc.os}$ are the efficient tools of grey dynamic modeling.

Consider that $X_r^{(0)}$ and $X_c^{(0)}$ stand for knowledge sets, where $X_r^{(0)}$ is known about model, and $X_c^{(0)}$ unknown about model. We have

Theorem 3.4 *The minimum element in* M_{RC} *is the optimal* Ξ_{rc} *with the knowledge discovery criterion, denoted by*

$$\Xi_{rc.kd} = \min_{r=1,2,\ldots,j} \left\{ \min_{c=1,2,\ldots,h} \{\Xi_{rc}\} \right\} .$$

Proof. Omitted.

$\Xi_{rc.kd}$ is an efficient tool for discovering new knowledge.

4 Application of Ξ to Grey Dynamic Modeling

Let us first consider a grey system with a single factor, and we can obtain a realization through observation, say, a single time series $X^{(0)}(k) \in X(\otimes, k)$, $k = 1,2,\cdots,n$. Our task here is to construct the system forecasting model, which is rather satisfactory, based on $\{X^{(0)}(k)\}$.

At first, we may transform $\{X^{(0)}(k)\}$ into $\{\overline{X}^{(1)}(k)\}$ through the Integral Generating Transform (IGT)

$$\left\{X^{(0)}(k)\right\} \xrightarrow{IGT} \left\{\overline{X}^{(1)}(k)\right\},$$

where

$$\overline{X}^{(1)}(k) = \sum_{m=2}^{k} \overline{X}^{(0)}(m),$$

$$\overline{X}^{(0)}(k) = 0.5(X^{(0)}(k-1) + X^{(0)}(k)),$$

$$k = 2,3,\cdots,n.$$

The determinacy of $\left\{\overline{X}^{(1)}(k)\right\}$ is stronger than $\left\{X^{(0)}(k)\right\}$.

Assume that there is a set of known function, denoted by $S(f)$,

$$S(f) = \left\{f_j(k)\right\},\ k = 1,2,\cdots,n,\ j = 1,2,\cdots,m.$$

Through the trend relational analysis between $\left\{\overline{X}^{(1)}(k)\right\}$ and $\left\{f_1(k), f_2(k),\cdots, f_m(k)\right\}$, respectively, we can get their trend relational grades as follows:

$$\Xi_{xf_1}, \Xi_{xf_2}, \cdots, \Xi_{xf_m}.$$

In order to find a satisfactorily comparing function, we have to analyze the trend relational functions and the trend relational grades. Thus answer is explicit. Suppose that

$$\Xi_{xf_{sc}} = \max\left\{\Xi_{xf_j}\right\},\ j = 1,2,\cdots,m,$$

then f_{sc}, which is sought out, may stand for a $\left\{\overline{X}^{(1)}(k)\right\}$'s latent law approximately.

f_{sc} is a determinate function defined on time domain, for example,

$$f_{sc}(k) = be^{a(k-1)} - c,$$

where $a,b,c \in R$, $k = 1,2,\cdots,n$. We may recognize that $f_{sc}(k)$ implies $\overline{X}^{(1)}(k)$ in the sense of trend relation satisfactorily, i.e. $f_{sc}(k)$ is a latency of $\overline{X}^{(1)}(k)$. Hence, we may use data of $\left\{\overline{X}^{(1)}(k)\right\}$ to fit $f_{sc}(k)$ like a glove, and find the system forecasting model as follows:

$$\hat{X}^{(0)}(k) = \hat{a}\hat{b}e^{\hat{a}(k-1)},$$

where

$$\hat{a} = \ln\left\{[\sum_{k=3}^{n}\overline{X}^{(0)}(k-1)\overline{X}^{(0)}(k)]/[\sum_{k=3}^{n}(\overline{X}^{(0)}(k-1))^2]\right\},$$

$$\hat{b} = \frac{(n-1)\sum_{k=2}^{n}e^{\hat{a}(k-1)}\overline{X}^{(1)}(k) - \left(\sum_{k=2}^{n}e^{\hat{a}(k-1)}\right)\left(\sum_{k=2}^{n}\overline{X}^{(1)}(k)\right)}{(n-1)\sum_{k=2}^{n}e^{2\hat{a}(k-1)} - \left(\sum_{k=2}^{n}e^{\hat{a}(k-1)}\right)^2},$$

This is a system forecasting model by name $SCGM(1,1)$ ($SCGM(1,1)_b$)[12-14,20].

Now let us consider a grey system with h factors, and we can get the original time series $\left\{X_i^{(0)}(k)\right\}$, $i = 1,2,\cdots,h$, $k = 1,2,\cdots,n$. Our task is to construct the system model based on $\left\{X_i^{(0)}(k)\right\}$. As stated above, we summarize the following theorem.

Theorem 4.1 *Let* $\left\{X_i^{(0)}(k)\right\}$, $i = 1,2,\cdots,h$, $k = 1,2,\cdots,n$, *be the original time series. Correspondently, there are the mean-value time series* $\left\{\overline{X}_i^{(0)}(k)\right\}$ *and its generating mean-value time series*

$\left\{\overline{X}_i^{(1)}(k)\right\}$. *If*

$$\left\{\overline{X}_i^{(1)}(k)\right\}\xrightarrow{Mtr}\left\{f_j(k)\right\},$$

where $f_j(k)$ *, which is known, represents the nonhomogeneous exponential function with respect to discrete time,* $k=1,2,\cdots,n$ *,* $j=1,2,\cdots,m$ *, and* Mtr *stands for a satisfactorily trend relation, then the grey model of systems clouds,* $SCGM(1,h)_b$ *by name, can be constructed as*

$$\dot{\hat{X}}^{(1)}(t)=\hat{A}\hat{X}^{(1)}(t)+\hat{U}\ ,t\geq 0 \text{ continuously} \quad (4\text{-}1)$$

and its solution is

$$\hat{X}^{(1)}(t)=e^{\hat{A}t}(\hat{X}^{(1)}(0)+\hat{A}^{-1}\hat{U})-\hat{A}^{-1}\hat{U} \quad (4\text{-}2)$$

or

$$\hat{X}^{(1)}(k)=e^{\hat{A}(k-1)}(\hat{X}^{(1)}(1)+\hat{A}^{-1}\hat{U})-\hat{A}^{-1}\hat{U}\ , \quad (4\text{-}3)$$

where

$$\hat{A}=\ln[\overline{X}_K^T\overline{X}_{K-1}(\overline{X}_K^T\overline{X}_{K-1})^{-1}], \quad (4\text{-}4)$$

$$\overline{X}_{K-1}=\begin{bmatrix}\overline{X}_1^{(0)}(2) & \overline{X}_2^{(0)}(2) & \cdots & \overline{X}_h^{(0)}(2)\\ \overline{X}_1^{(0)}(3) & \overline{X}_2^{(0)}(3) & \cdots & \overline{X}_h^{(0)}(3)\\ \vdots & \vdots & & \vdots\\ \overline{X}_1^{(0)}(n-1) & \overline{X}_2^{(0)}(n-1) & \cdots & \overline{X}_h^{(0)}(n-1)\end{bmatrix}$$

$$\overline{X}_K=\begin{bmatrix}\overline{X}_1^{(0)}(3) & \overline{X}_2^{(0)}(3) & \cdots & \overline{X}_h^{(0)}(3)\\ \overline{X}_1^{(0)}(4) & \overline{X}_2^{(0)}(4) & \cdots & \overline{X}_h^{(0)}(4)\\ \vdots & \vdots & & \vdots\\ \overline{X}_1^{(0)}(n) & \overline{X}_2^{(0)}(n) & \cdots & \overline{X}_h^{(0)}(n)\end{bmatrix}$$

$$\overline{X}_i^{(0)}(k)=0.5(X_i^{(0)}(k-1)+X_i^{(0)}(k)),$$

$k=2,3,\cdots,n$, $i=1,2,\cdots,h$;

$$\hat{B}=\left[(n-1)\sum_{k=2}^{n}e^{\hat{A}^T(k-1)}e^{\hat{A}(k-1)}-\left(\sum_{k=2}^{n}e^{\hat{A}^T(k-1)}\right)\left(\sum_{k=2}^{n}e^{\hat{A}(k-1)}\right)\right]^{-1}$$

$$\cdot\left[(n-1)\sum_{k=2}^{n}e^{\hat{A}^T(k-1)}\overline{X}^{(1)}(k)-\left(\sum_{k=2}^{n}e^{\hat{A}^T(k-1)}\right)\left(\sum_{k=2}^{n}\overline{X}^{(1)}(k)\right)\right], \quad (4\text{-}5)$$

$$\hat{C}=\frac{1}{n-1}\left[\left(\sum_{k=2}^{n}e^{\hat{A}(k-1)}\right)\hat{B}-\sum_{k=2}^{n}\overline{X}^{(1)}(k)\right], \quad (4\text{-}6)$$

$\hat{U}=\hat{A}\hat{C}$;

$\overline{X}^{(1)}(1)=\hat{B}-\hat{C}$;

$X^{(1)}(k)=\left[\hat{X}_1^{(1)}(k),\hat{X}_2^{(1)}(k),\cdots,\hat{X}_h^{(1)}(k)\right]^T$;

$X^{(1)}(t)=\left[\hat{X}_1^{(1)}(t),\hat{X}_2^{(1)}(t),\cdots,\hat{X}_h^{(1)}(t)\right]^T$.

Proof. Selecting $e^{\hat{A}t}$ to be the integral factor, and multiplying both sides of Eq. (4-1) by $e^{-\hat{A}t}$, we have

$$e^{-\hat{A}t}[\dot{\hat{X}}^{(1)}(t)-\hat{A}\hat{X}^{(1)}(t)]=e^{-\hat{A}t}\hat{U}$$

$$\frac{d}{dt}\left[e^{-\hat{A}t}\hat{X}^{(1)}(t)\right]=e^{-\hat{A}t}\hat{U}$$

$$e^{-\hat{A}t}\hat{X}^{(1)}(t)\Big|_0^t=\int_0^t e^{-\hat{A}t}dt\hat{U}\ .$$

After rearranging, Eq. (4-1) becomes Eq. (4-2) or Eq.(4-3), where Eq.(4-3) is a discrete solution of Eq.(4-1).

Let $\hat{C}=\hat{A}^{-1}\hat{U}$, and $\hat{B}=\hat{X}^{(1)}(1)+\hat{C}$.

Considering the given conditions, we define

$$\overline{X}^{(1)}(k)=\hat{X}^{(1)}(k)$$

$$=e^{\hat{A}(k-1)}\hat{B}-\hat{C}$$

$$\overline{X}^{(0)}(k)=e^{\hat{A}(k-1)}(I-e^{-\hat{A}})\hat{B}$$

$$e^{\hat{A}}\overline{X}^{(0)}(k-1)=\hat{X}^{(0)}(k).$$

When $k=2,3,\cdots,n$, $i=1,2,\cdots,h$, we can construct the matrixes: $\overline{X}_{K-1}$ and $\overline{X}_K$. If $\overline{X}_{K-1}^T\overline{X}_{K-1}$ is nonsingular and rank $\left(\overline{X}_K^T\overline{X}_{K-1}\left(\overline{X}_{K-1}^T\overline{X}_{K-1}\right)^{-1}\right)=h$, then Eq.(4-4) holds. Because $\{\overline{X}^{(1)}(k)\}$ is satisfactorily related to the fitting curves determined by Eq.(4-2) or Eq.(4-3), we can get

$$\begin{bmatrix}\hat{B}\\ \hat{C}\end{bmatrix}=\begin{bmatrix}(H_{\sum 11}-H_{\sum 12}H_{\sum 22}^{-1}H_{\sum 21})^{-1} & 0\\ -H_{\sum 22}^{-1}H_{\sum 21}(H_{\sum 11}-H_{\sum 12}H_{\sum 22}^{-1}H_{\sum 21})^{-1} & H_{\sum 22}^{-1}\end{bmatrix}\cdot\begin{bmatrix}Y_1-H_{\sum 12}H_{\sum 22}^{-1}Y_2\\ Y_2\end{bmatrix},$$

where

$$H_{\sum 11}=\sum_{k=2}^{n}e^{\hat{A}^T(k-1)}e^{\hat{A}(k-1)}$$

$$H_{\sum 12}=-\sum_{k=2}^{n}e^{\hat{A}^T(k-1)}$$

$$H_{\sum 21}=-\sum_{k=2}^{n}e^{\hat{A}(k-1)}$$

$$H_{\sum 22}=(n-1)I$$

$$Y_1=\sum_{k=2}^{n}e^{\hat{A}^T(k-1)}\overline{X}^{(1)}(k)$$

$$Y_2=-\sum_{k=2}^{n}\overline{X}^{(1)}(k).$$

where I is the identity matrix. Eqs.(4-5) and (4-6) hold. Thus, we have $\hat{X}^{(1)}(1)$, $\hat{U}$ and Eq.(4-3). Considering Eq.(4-3) $\Leftrightarrow$ Eq.(4-2), Eq.(4-1) can be obtained by differentiating both sides of Eq.(4-2) with respect t.

Theorem 4.2 *The $SCGM(1,h)_b$ forecasting model is*

$$\hat{X}^{(0)}(t)=e^{\hat{A}t}\hat{A}\hat{B}$$

or

$$\hat{X}^{(0)}(k)=e^{\hat{A}(k-1)}\hat{A}\hat{B},$$

where $\hat{A}$ and $\hat{B}$ are given by Eq.(4-4) and Eq.(4-5), respectively.

Proof. Omitted.

Corollary 4.1 *When $h=1$, Eq.(4-1) becomes $SCGM(1,1)_b$ model.*

Remarks *Parameters identification of $SCGM(1,h)_b$ on line.*

As we know, the original time series which are used to construct $SCGM(1,h)_b$ only reflect the past and present behavior on investigated object. With the lapse of time, inside and outside may be changing, correspondingly, a number of new data can be obtained. In order to make $SCGM(1,h)_b$ be able to reflect and trace the change of behavior, we should improve the original $SCGM(1,h)_b$ model, using the new data observed, so that we can keep $SCGM(1,h)_b$ all the time to be valid and precise. Considering the parameters $\hat{A}$, $\hat{B}$ and $\hat{C}$ of $SCGM(1,h)_b$ model here only, in which $e^{\hat{A}}$ is essential part, we can derive $\hat{B}$ and $\hat{C}$ from it. The detail is in the following.

Suppose that there is an original time series

$$X_i^{(0)} = \left\{X_i^{(0)}(1), X_i^{(0)}(2), \cdots, X_i^{(0)}(n)\right\}$$

which is satisfied for $SCGM(1,h)_b$ modeling. According to Theorem 4.1, we have

$$e^{\hat{A}} = \overline{X}_K^T \overline{X}_{K-1}\left(\overline{X}_{K-1}^T \overline{X}_{K-1}\right)^{-1}.$$

Let $a_n = e^{\hat{A}n}$ and $P_n = \left(\overline{X}_{K-1}^T \overline{X}_{K-1}\right)^{-1}$, then

$$a_n = \overline{X}_K^T \overline{X}_{K-1} P_n.$$

When a new data vector observed is obtained, we have

$$P_{n+1} = \left(\begin{bmatrix} \overline{X}_{K-1} \\ --- \\ \overline{X}^{(0)T}(n) \end{bmatrix}^T \begin{bmatrix} \overline{X}_{K-1} \\ --- \\ \overline{X}^{(0)T}(n) \end{bmatrix}\right)^{-1}$$

$$= [\overline{X}_{K-1}^T \overline{X}_{K-1} + \overline{X}^{(0)}(n)\overline{X}^{(0)T}(n)]^{-1}$$

$$= \left[\overline{X}_{K-1}^T \overline{X}_{K-1}\right]^{-1} - \left[\overline{X}_{K-1}^T \overline{X}_{K-1}\right]^{-1} \overline{X}^{(0)}(n)$$

$$\cdot \left(1 + \overline{X}^{(0)T}(n)\left[\overline{X}_{K-1}^T \overline{X}_{K-1}\right]^{-1} \overline{X}^{(0)}(n)\right)^{-1}$$

$$\cdot \overline{X}^{(0)T}(n)\left[\overline{X}_{K-1}^T \overline{X}_{K-1}\right]^{-1}$$

$$= P_n - P_n \overline{X}^{(0)}(n)(1 + \overline{X}^{(0)T}(n) P_n \overline{X}^{(0)}(n))^{-1}$$

$$\cdot X^{(0)T}(n) P_n. \qquad (4\text{-}7)$$

Let $a_{n+1} = e^{An+1}$, then

$$a_{n+1} = \left(\begin{bmatrix} \overline{X}_K \\ --- \\ \overline{X}^{(0)T}(n+1) \end{bmatrix}^T \begin{bmatrix} \overline{X}_{K-1} \\ --- \\ \overline{X}^{(0)T}(n) \end{bmatrix}\right) P_{n+1}$$

$$= a_n + \left[\overline{X}^{(0)}(n+1) - a_n \overline{X}^{(0)}(n)\right]$$

$$\cdot \left(1 + \overline{X}^{(0)T}(n) P_n \overline{X}^{(0)}(n)\right)^{-1} \overline{X}^{(0)T}(n)$$

$$\cdot P_n, \qquad (4\text{-}8)$$

where

$$\overline{X}^{(0)T}(n) = \left[\overline{X}_1^{(0)}(n), \overline{X}_2^{(0)}(n), \cdots, \overline{X}_h^{(0)}(n)\right]$$

$$\overline{X}^{(0)T}(n+1) = \left[\overline{X}_1^{(0)}(n+1), \overline{X}_2^{(0)}(n+1), \cdots, \overline{X}_h^{(0)}(n+1)\right].$$

Thenceforth, we can get

$$\hat{B}_{n+1} = \left[n\sum_{k=2}^{n+1} e^{\hat{A}_{n+1}^T(k-1)} e^{\hat{A}_{n+1}(k-1)} - \left(\sum_{k=2}^{n+1} e^{\hat{A}_{n+1}^T(k-1)}\right)\right.$$

$$\left.\cdot\left(\sum_{k=2}^{n+1} e^{\hat{A}_{n+1}(k-1)}\right)\right]^{-1} \cdot \left[n\sum_{k=2}^{n+1} e^{\hat{A}_{n+1}^T(k-1)} \overline{X}^{(1)}(k) -\right.$$

$$\left.\left(\sum_{k=2}^{n+1} e^{\hat{A}_{n+1}^T(k-1)}\right)\left(\sum_{k=2}^{n+1} \overline{X}^{(1)}(k)\right)\right].$$

$$\hat{C}_{n+1} = \frac{1}{n}\left[\left(\sum_{k=2}^{n+1} e^{\hat{A}_{n+1}(k-1)}\right)\hat{B}_{n+1} - \sum_{k=2}^{n+1} \overline{X}^{(1)}(k)\right].$$

Eqs.(4-7) and (4-8), which are two recurrence formulas, are sought out.

$\left\{X_i^{(0)}(k)\right\}$ may be arbitrary time series in the presence of uncertainty. In many situations, an investigated object may be regarded as a generalized energy system, and it is suitable to represent such system behavior by exponential function. Thus, the principle of grey dynamic modeling and $SCGM(1,h)$ model are able to provide the effective tools for modeling the intricacies of the real world.

5 Grey-Fuzzy Clustering Method

As is well known that the concept of a fuzzy set first arose from the study of problems related to pattern classification, since the recognition of patterns is an important aspect of human perception, which is a fuzzy process in nature. Fuzzy classification methods may be classified into the following four categories[15]:

- Methods based on fuzzy relation;
- Methods based on fuzzy pattern matching procedures;
- Methods based on fuzzy clustering procedures; and
- Other methods.

On clustering methods, the primary objective of clustering techniques is to partition a given data set into so-called homogeneous clusters. The term “homogeneous” means that all points in the same group are close to each other and are not close to points in other groups. Clustering algorithms may be used to build pattern classes or to reduce the size of a set of data while retaining relevant information[16]. From a practical point of view, grey systems and fuzzy systems are found everywhere, and the representation of clusters by grey-fuzzy clustering method below may seem more appropriate in certain situations.

Suppose $X=\{X_1,X_2,\ldots,X_h\}$ is a sample set, in which each element is called a sample. In terms of practical observation, every sample has n indexes $X_i=\{X_i(1),X_i(2),\ldots,X_i(n)\}$, $i=1,2,\ldots,h$. Thus, based on Theorem 3.1 and Corollary 3.2, we are able to construct the trend relational grade matrix, denoted by $M_{RC}(X_r,X_c),\ r,c\in\{1,2,\ldots,h\}$,

$$M_{RC}(X_r,X_c)=\begin{bmatrix}\Xi_{11} & \Xi_{12} & \cdots & \Xi_{1h}\\ \Xi_{21} & \Xi_{22} & \cdots & \Xi_{2h}\\ \vdots & \vdots & \cdots & \vdots\\ \Xi_{h1} & \Xi_{h2} & \cdots & \Xi_{hh}\end{bmatrix}$$

which describes the approximateness and similarity among samples on sample set X . Obviously, the approximateness and similarity are fuzzy concepts, and the trend relational grade is a fuzzy quantity, $0\le\Xi_{rc}\le1$, and the trend relational grade matrix $M_{RC}(X_r,X_c)$ is a fuzzy relation on set X . Because $M_{RC}(X_r,X_c)$ satisfies

$$M(X_r,X_c)=1,\qquad X_r,X_c\in X\ ,r=c$$

$$M(X_r,X_c)=M(X_c,X_r),\qquad X_r,X_c\in X$$

$$M^2(X_r,X_c)\le M(X_r,X_c),\qquad X_r,X_c\in X$$

that is, it is reflexive, symmetric and transitive, therefore, $M_{RC}(X_r,X_c)$ is a fuzzy equivalent relation on X .

In practice, we usually use $|X_{rr}^{(0)}(k)-X_{cc}^{(0)}(k)|$, that is,

$$\xi_{rc}(k)=\left[1+\gamma\,|\,\Delta X_{rr}^{(0)}(k)-\Delta X_{cc}^{(0)}(k)\,|\right]^{-1},$$

then $M_{RC}(X_r,X_c)$ is a “similarity” relation on X . In this case, there exists

$$\hat{M}_{RC}(X_r,X_c)=\bigcup_{p=1}^{\infty}M_{RC}^{p}(X_r,X_c),$$

and $\hat{M}_{RC}(X_r,X_c)$ is a fuzzy equivalent relation on X .

Consider that the α -cut $M_{RC}(X_r,X_c)_\alpha$ of $M_{RC}(X_r,X_c)$ is an ordinary relation, $\alpha\in[0,1]$. Thus, we can easily prove that $M_{RC}(X_r,X_c)_\alpha$ is a equivalent relation if $M_{RC}(X_r,X_c)$ is a equivalent relation. Based $M_{RC}(X_r,X_c)_\alpha,\alpha\in[0,1]$, we are able to partition a given data set into so-called homogeneous clusters[17-19].

6 Demonstrating Examples

6.1 Forecasting the floodwater in Huaihe River

Huaihe River is located in the south of Fuyang

District of Anhui Province, China. This river was a place famous for its more flood in history. In 1990, using the $SCGM(1,1)$ model, J.Sun et al [20] successfully forecasted the catastrophic flood happened in Huaihe River in 1991 and other years. As a example, if the discharge of river is over 7000 M^3/s, the heavy flood will happen in Huaihe River. Take abnormal value ξ as 7000 M^3/s, and write down the particular years as follows:

$$X = \{50,54,56,60,68,75,82,83\},$$

in which the discharge of the Huaihe River is more than 7000 M^3/s. We may transform X into $X^{(0)}$,

$$X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), \cdots, X^{(0)}(8)\}$$

$$= \{1,5,7,11,19,26,33,34\}.$$

Based on $X^{(0)}$, we can construct the $SCGM(1,1)_b$ forecasting model

$$\hat{X}^{(0)}(k) = \hat{a}\hat{b}e^{\hat{a}(k-1)}$$

$$= 6.94373 65 e^{0.24019(k-1)}$$

where

$$\hat{a} = 0.24019991,\ \hat{b} = 28.90936.$$

Taking $k = 9$, we can obtain the forecasting result of the time when the floodwater will take place next time.

$\hat{X}^{(0)}(9) = 47.4$ that will happen in 1996.4.

On the basis of forecasting result above, the flood control strategy was taken in advance, so the heavy losses were reduced greatly in Fuyang District.

6.2 Dynamic grey-fuzzy clustering

Suppose $X = \{X_1, X_2, X_3, X_4, X_5\}$ is the set of five people, and a relation on X is "familiar". According to grey-fuzzy clustering method mentioned above, we have the trend relational grade matrix

$$M_{RC}(X_r, X_c) = \begin{bmatrix} 1 & 0.1 & 0.8 & 0.5 & 0.3 \\ 0.1 & 1 & 0.1 & 0.2 & 0.4 \\ 0.8 & 0.1 & 1 & 0.3 & 0.1 \\ 0.5 & 0.2 & 0.3 & 1 & 0.6 \\ 0.3 & 0.4 & 0.1 & 0.6 & 1 \end{bmatrix}.$$

It is clear that $M(X_r, X_c) = 1$, when $r = c$, and $M(X_r, X_c) = M(X_c, X_r)$. Such a matrix is a fuzzy "similarity" relation on X. Thus,

$$\hat{M}_{RC}(X_r, X_c) = \bigcup_{p=1}^{\infty} M_{RC}^{p}(X_r, X_c)$$

$$= M_{RC}^{4}(X_r, X_c)$$

$$= \begin{bmatrix} 1 & 0.4 & 0.8 & 0.5 & 0.5 \\ 0.4 & 1 & 0.4 & 0.4 & 0.4 \\ 0.8 & 0.4 & 1 & 0.5 & 0.5 \\ 0.5 & 0.4 & 0.5 & 1 & 0.6 \\ 0.5 & 0.4 & 0.5 & 0.6 & 1 \end{bmatrix}$$

which is a fuzzy equivalent relation.

Based on $\hat{M}_{RC}(X_r, X_c)$, X can be divided to some clusters. Choose $\hat{M}_{RC}(X_r, X_c)_\alpha$, $\alpha = 1,0.8,0.6,0.5$ and 0.4, we have

$$\hat{M}_{RC}(X_r, X_c)_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

and corresponding cluster:

$$\{\{X_1\},\{X_2\},\{X_3\},\{X_4\},\{X_5\}\};$$

$$\hat{M}_{RC}(X_r,X_c)_{0.8}=\begin{bmatrix}1&0&1&0&0\\0&1&0&0&0\\1&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{bmatrix},$$

$$\{\{X_2\},\{X_4\},\{X_5\},\{X_1,X_3\}\};$$

$$\hat{M}_{RC}(X_r,X_c)_{0.6}=\begin{bmatrix}1&0&1&0&0\\0&1&0&0&0\\1&0&1&0&0\\0&0&0&1&1\\0&0&0&1&1\end{bmatrix},$$

$$\{\{X_2\},\{X_1,X_3\},\{X_4,X_5\}\};$$

$$\hat{M}_{RC}(X_r,X_c)_{0.5}=\begin{bmatrix}1&0&1&1&1\\0&1&0&0&0\\1&0&1&1&1\\1&0&1&1&1\\1&0&1&1&1\end{bmatrix},$$

$$\{\{X_2\},\{X_1,X_3,X_4,X_5\}\};$$

$$\hat{M}_{RC}(X_r,X_c)_{0.4}=\begin{bmatrix}1&1&1&1&1\\1&1&1&1&1\\1&1&1&1&1\\1&1&1&1&1\\1&1&1&1&1\end{bmatrix},$$

$$\{X_1,X_2,X_3,X_4,X_5\}.$$

We can check that the clustering results obtained above are the same as usual fuzzy clustering analyses[15]. As seen by those mentioned above, the grey-fuzzy clustering method developed in this paper is a very useful tool of data mining.

References

[1]M.Y.Chen. *Grey system and its dynamic modeling*. Proceedings of the International AMSE Conference on "Signal, Data, Systems" Methodologies & Applications, Calcutta(India), Dec.7-9,1992,AMSE Press, Vol.2, PP.89-96.

[2]Y.Lin, M.Y.Chen and S.F.Liu. *Theory of grey systems: capturing uncertainties of grey information*. Kybernetes, Vol.33(2004), No.2, pp.196-218.

[3]J.L.Deng. *Control problems of grey systems*. Systems & Control Letters, Vol.1(1982), No.5, pp.288-294.

[4]M.Y.Chen. *Grey dynamics of the system of a boring machine*. Journal of Huazhong University of Science and Technology(in Chinese), Vol.10(1982), No.6, pp.7-11.

[5]M.Y.Chen. *System cloud and its grey model*. Proceedings of the Fourth Japanese-Sino Sapporo International Conference on Computer Applications. Hokkaido University, Sapporo 1990, pp.38-41.

[6]M.Y.Chen. *Trend relational analysis and grey dynamic modeling*. Proceedings of the Twelfth European Meeting on Cybernetics and Systems Research, University of Vienna, World Scientific, 1994, pp.19-24.

[7]M.Y.Chen, W.Z.Chen, Z.J.Chen. *Grey state model and forecasting control of complicated systems*. Proceedings of the 15th International Conference on Systems Science, Sept.7-10,2004, Wroclaw(Poland), pp.141-150.

[8]M.Y.Chen. *Uncertainty analysis and grey modeling*. Uncertainty Modeling and Analysis (ISUMA'90), IEEE Computer Society Press, 1990, pp.469-473.

[9]W.Z.Chen, Z.J.Chen, Z.H.Xie, L.Zhou and M.Y.Chen. *Multifactor process analysis and* $SCGM(1,h)_{mv}$ *forecasting model*. Journal of Wuhan University of Technology(in Chinese), Vol.28(2004), No.2, pp.291-294.

[10]M.Y.Chen, H.J.Xiong and Z.J.Chen. *An approach to modal control of grey systems*. SAMS, 1998. Vol.33, pp.47-56.

[11]M.Y.Chen. *Fuzzy control system with trend relational predictor*. The Journal of Fuzzy Mathematics, Vol.2(1994), No.4. pp.685-694.

[12]M.Y.Chen. *General systems studies and grey dynamic modeling*. General Systems Studies and Applications, Huazhong University of Science and Technology Press, 1997, pp1-9.

[13]Z.J.Chen, W.Z.Chen, Q.L.Chen and M.Y.Chen. *"Poor" signal processing and grey modeling technique*. Proceedings of Eighteenth International Conference on Systems Engineering, August 16-18, 2005, Las Vegas, Nevada(USA), pp.251-256.

[14]W.Z.Chen, Z.J.Chen, Q.L.Chen and M.Y.Chen. *Approaches to grey prediction and control of environmental systems*. Processings of Eighteenth International Conference on Systems Engineering, August 16-18, 2005, Las Vegas, Nevada(USA), pp.152-157.

[15]M.Grabisch and F.Dispot. *A comparison of some methods of fuzzy classification on real data*. Proceedings of IIZUKA'92,1992,pp.659-662.

[16]D.Dubis and H.Prade. Fuzzy Sets and Systems-Theory and Application, New York, 1980.

[17]Z.J.Chen, Q.L.Chen, W.Z.Chen and Y.N.Wang. *Grey linear programming*. Kybernetes, Vol.33(2004), No.2,pp.238-246.

[18]H.J.Xiong, M.Y.Chen and Q.Tan. *Application of trend relational grade to fuzzy clustering*. Advances in Systems Science and Applications, Special Issue 1996,pp.238-243.

[19]Y.Y.Chen. *Fuzzy Mathematics(in Chinese)*. Huazhong University of Science and Technology Press, 1986.

[20]J.Sun and J.Zhao. *The application of the model of* $SCGM(1,1)$ *about forecasting the floodwater in the Huaihe branches and the storing floodwater time in Mengwa area*. Proceedings of the International Conference on "Systems, Control, Information" Methodologies & Applications (SCI'94), Huazhong University of Science and Technology Press, 1995, pp.298-301.