Available online at www.sciencedirect.com
ScienceDirect
Information Sciences 178 (2008) 1087–1097
INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL
www.elsevier.com/locate/ins

A clustering method to identify representative financial ratios

Yu-Jie Wang [a,*], Hsuan-Shih Lee [b]

[a] Department of Shipping and Transportation Management, National Penghu University, Penghu 880, Taiwan, ROC
[b] Department of Shipping and Transportation Management, National Taiwan Ocean University, Keelung 202, Taiwan, ROC

Received 4 October 2004; received in revised form 2 September 2007; accepted 3 September 2007

Abstract

When companies evaluate their performance, it is impractical to take all of their financial ratios into consideration. To evaluate the financial performance of a company, only a fraction of the available financial ratios are considered and selected as evaluation criteria. In general, financial ratios presented as sequences (or called financial ratio sequences), are first clustered and then a representative indicator is chosen from each cluster to serve as an evaluation criterion. To cluster financial ratios, we propose a clustering method in which the financial ratios of different companies with similar variations are partitioned into the same cluster. In other words, a fuzzy relation is proposed to represent the similarity between the financial ratios, and a cluster validation index is also provided to determine the number of clusters. Once the financial ratios are clustered, the representative indicator for each cluster will be identified.

Keywords: Clustering; Financial ratios; Fuzzy relation; Representative indicator

1. Introduction

In order to evaluate the financial performance of a company, financial ratios [22] are commonly used as evaluation criteria. Since some financial ratios have quite similar patterns, it would be inefficient to take all financial ratios into consideration for evaluation. To avoid repeating the evaluation process on similar financial ratios, we should partition the financial ratios into clusters. After partitioning, the characteristics of the financial ratios in the same cluster will be more similar, whereas the inter-cluster characteristics will be less similar. Therefore, we can select a representative indicator in each cluster as the evaluation criterion. For this reason, the evaluation criteria depend highly on the clustering method adopted.

Numerous clustering methods [2–5,9,11,13,17,19,20] have been proposed so far. These methods include cluster analysis, discriminant analysis, factor analysis, principal component analysis [10], grey relation analysis [7] and K-means [1,12,18,21,23]. Cluster analysis, discriminant analysis, factor analysis and principal component analysis are often applied in classic statistical problems, such as for a large sample or long-term data. To

* Corresponding author.
E-mail address: knight@axpl.stm.ntou.edu.tw (Y.-J. Wang).

0020-0255/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2007.09.016

deal with a small sample or short-term data, grey relation analyses and K-means are preferred. Because most financial ratios are short-term data, classic statistical methods are not suitable means to cluster financial ratios. Although K-means, proposed by MacQueen [18], can partition items into clusters because the clustering number is known, this method is sometimes combined with other approaches, such as a self-organizing map (SOM) or neural network [14–16], so that the clustering number is determined automatically. However, we prefer that the financial ratios presented by sequences be clustered in cases where the clustering number is unknown, hence the original K-means approach is not suitable for the clustering problem.

In this paper, we suggest a new clustering method based on a fuzzy relation between financial ratio sequences. The fuzzy relation is constructed on a compatible relation which represents the variation between two financial ratio sequences. A cluster validation index is provided as well in order to objectively determine the number of clusters. Since clusters based on this fuzzy relation may overlap, additional mechanisms are introduced to resolve ambiguities. After the disjoint clusters of financial ratios are identified, a representative indicator will be drawn from each cluster through comparison of the financial ratios within a given cluster.

For the sake of clarity, mathematical preliminaries are presented in Section 2. Financial ratios are stated in Section 3. In Section 4, we introduce our clustering method and two numerical examples. Finally, an empirical study and a comparison of our method and the K-means is provided in Section 5.

2. Mathematical preliminaries

To select the representative indicators of financial ratios, relevant mathematical theories are stated as follows. First, we review the compatible and equivalence relations [6,17].

A fuzzy binary relation R on $X \times Y$ is defined as the set of ordered pairs:

$$R = \{((x,y), r(x,y)) | (x,y) \in X \times Y\},$$

where r is a function that maps $X \times Y \rightarrow [0,1]$.

In particular, the relation R is called a fuzzy binary relation on X when $X = Y$.

Let R be a fuzzy binary relation on S. The following conditions may hold for R:

1. R is reflexive, if $r(x,x) = 1\ \forall x \in S$.
2. R is symmetric, if $r(x,y) = r(y,x)\ \forall x,y \in S$.
3. R is transitive, if $r(x,y) \geqslant \max_{y \in S} \min\{r(x,y), r(y,z)\}\ \forall x,y,z \in S$.

If R is reflexive and symmetric, R is said to be a fuzzy compatible relation on S. If R is reflexive, symmetric and transitive, R is said to be a fuzzy equivalence relation on S [6,17].

Let R^λ be a binary relation on S defined as

$$R^\lambda = \{(x,y) | r(x,y) \geqslant \lambda\ \forall x,y \in S\},$$

where $0 \leqslant \lambda \leqslant 1$.

Lemma 2.1. *Let R be a fuzzy equivalence relation on S. Then R^λ is an equivalence relation on S.*

Proof. R^λ is an equivalence relation on S since it satisfies the following three conditions:

1. Reflexive: $r(x,x) = 1 \geqslant \lambda\ \forall x \in S$.
2. Symmetric: If $(x,y) \in R^\lambda$, then (y,x) is also in R^λ for $r(x,y) = r(y,x) \geqslant \lambda\ \forall x,y \in S$.
3. Transitive: Suppose both (x,y) and (y,z) are in R^λ, then (x,z) is also in R^λ for $r(x,z) = \min\{r(x,y), r(y,z)\} \geqslant \lambda\ \forall x,y,z \in S$. $\square$

If R is a fuzzy compatible relation on S, then R^λ is a compatible relation on S. Generally speaking, a partition of items may root in a compatible relation or an equivalence relation.

A partition or classification of S is a family of disjoint subsets, say $\{S_1, S_2, \ldots, S_i, \ldots\}$. The partition must satisfy the following conditions. That is

$$S_1 \cup S_2 \cup \cdots \cup S_i \cup \cdots = S$$

and

$$S_i \cap S_j = \varnothing \quad \forall i \neq j.$$

Given an equivalence relation R defined on S, we can use the equivalence relation to form a partition or classification of S. A partition of S is induced by R^{λ}. However, it is difficult to determine the threshold value λ without performing a lot of experiments on different λ's.

3. Financial ratios

In accounting, financial ratios on a balance sheet or income statement can be initially classified into four categories: solvency, profitability, asset and debt turnover, and return on investment.

The financial ratios that fall within these four categories are shown in Table 1.

We chose to partition these financial ratios based on these four categories. Financial ratios in different categories are considered to be unrelated, so that ratios in the same category are clustered together. The representative indicator for each cluster will be identified by the method presented in Section 4.

4. The clustering method

In this section, we propose a clustering method based on the variations of financial ratio sequences of different companies. This method partitions financial ratio sequences into clusters and then selects the representative indicators of the financial ratios from these clusters. The details are as stated below:

Let D be the matrix consisting of n financial ratios of m companies:

$$D = [y_{ij}]_{n \times m}, \tag{1}$$

where y_{ij} is the value of financial ratio i of company j.

Table 1
The four categories of financial ratios

Source	Category	Ratio	Formula
Balance sheet	Solvency	Current ratio	Current assets/current liabilities
		Fixed ratio	Stockholder's equity/fixed assets
		Equity ratio	Stockholder's equity/total assets
		Fixed/long-term ratio	Fixed assets/long-term liabilities
		Debt ratio	Total liabilities/ total assets
		Equity/debt ratio	Total liabilities/ stockholder's equity
Income statement	Profitability	Operation cost ratio	Operation cost/operation revenue
		Gross profit ratio	(Operation revenue – operation cost)/ operation revenue
		Operation profit ratio	Operation income(loss)/ operation revenue
		Income before tax ratio	Income(loss) before tax/operation revenue
		Net income ratio	Net income(loss)/operation revenue
Balance sheet and income statement	Return on investment	Return on current assets	Net income(loss)/current assets
		Return on fixed assets	Net income(loss)/fixed assets
		Return on total assets	Net income(loss)/total assets
		Return on stockholder's equity	Net income(loss)/stockholder's equity
		Return on operation to capital	Operation income(loss)/average capital
		Return on income before tax to capital	Net income(loss)/average capital
	Asset and debt turnover	Current assets turnover	Operation revenue/current assets
		Fixed assets turnover	Operation revenue/fixed assets
		Total assets turnover	Operation revenue/total assets
		Stockholder's equity turnover	Operation revenue/ stockholder's equity
		Current liabilities turnover	Operation revenue/current liabilities
		Long-term liabilities turnover	Operation revenue/long-term liabilities
		Total liabilities turnover	Operation revenue/total liabilities

Assume Y is the set composed of all financial ratios. Let $Y_i = (y_{i1}, y_{i2}, \ldots, y_{ik}, \ldots, y_{im}) \in Y$ denote the sequence of financial ratio i consisting of m entries. Define $M_i(k, k+1)$ to indicate the variation of company k to $k+1$ for Y_i, where

$$M_i(k,k+1) = \frac{y_{i,k+1} - y_{ik}}{\sqrt{\sum_{t=1}^{m} (y_{it})^2}}. \tag{2}$$

Let

$$R = \{((Y_i, Y_j), r(Y_i, Y_j)) | Y_i, Y_j \in Y\} \tag{3}$$

be a fuzzy relation defined on Y, where

$$r(Y_i, Y_j) = \frac{1}{m-1} \sum_{k=1}^{m-1} \left(1 - \frac{|M_i(k,k+1) - M_j(k,k+1)|}{\|M\|}\right)$$

and

$$\|M\| = \max_{i,k}\{M_i(k,k+1)\} - \min_{i,k}\{M_i(k,k+1)\}.$$

Then $r(Y_i, Y_j)$ measures the similarity between the two financial ratio sequences Y_i and Y_j, where $0 \leqslant r(Y_i, Y_j) \leqslant 1$. The financial ratio Y_i is closer to financial ratio Y_j, as $r(Y_i, Y_j)$ approaches 1. On the other hand, the financial ratio Y_i is farther from financial ratio Y_j, as $r(Y_i, Y_j)$ approaches 0.

Lemma 4.1. *Based on the two sequences Y_i and Y_j, we know that the fuzzy relation $r(Y_i, Y_j) = 1$ iff $y_{ik} = ty_{jk}$, $t > 0$ $\forall k$, where $y_{ik} \in Y_i$ and $y_{jk} \in Y_j$.*

Proof. By (3),

$$r(Y_i, Y_j) = \frac{1}{m-1} \sum_{k=1}^{m-1} \left(1 - \frac{|M_i(k,k+1) - M_j(k,k+1)|}{\|M\|}\right) = 1.$$

To satisfy the equation, $1 - \frac{|M_i(k,k+1) - M_j(k,k+1)|}{\|M\|} = 1\ \forall k$.

That is, $M_i(k,k+1) - M_j(k,k+1) = 0\ \forall k$.

$$\frac{y_{i,k+1} - y_{ik}}{\sqrt{\sum_{t=1}^{m} (y_{it})^2}} = \frac{y_{j,k+1} - y_{jk}}{\sqrt{\sum_{t=1}^{m} (y_{jt})^2}} \quad \forall k,$$

so $y_{ik} = ty_{jk}$, $t > 0$ $\forall k$.

That is to say, if $r(Y_i, Y_j) = 1$, then $y_{ik} = ty_{jk}$, $t > 0$ $\forall k$.

On the other hand, we can also prove $r(Y_i, Y_j) = 1$ if $y_{ik} = ty_{jk}$, $t > 0$ $\forall k$. To prove this statement, the previous procedure should simply be reversed. Therefore, this proof has been omitted. □

Lemma 4.2. *The fuzzy relation R defined in (3) is a fuzzy compatible relation on Y.*

Proof

1. Reflexive: $r(Y_i, Y_i) = 1$ $\forall Y_i \in Y$.
2. Symmetric: $r(Y_i, Y_j) = r(Y_j, Y_i)$ $\forall Y_i, Y_j \in Y$.

Since R satisfies both the reflexive and symmetric law, R is a fuzzy compatible relation on Y. □

Based on R of Lemma 4.2, we can construct R^λ with a given threshold value λ. In Section 2, $R^\lambda = \{(Y_i, Y_j) | r(Y_i, Y_j) \geqslant \lambda\ \forall Y_i, Y_j \in Y\}$ and $0 \leqslant \lambda \leqslant 1$. Since R is a fuzzy compatible relation, R^λ is also a compatible relation. In the case where $\lambda = 0$, R^λ partitions n financial ratios into one cluster. In case that $\lambda = 1$, R^λ commonly partitions n financial ratios into n clusters. In what follows, we present a method to determine λ objectively.

Let $C_n(\lambda)$ be the number of clusters partitioned by R^λ. $C_n(\lambda)$ is closer to n as λ approaches 1, whereas $C_n(\lambda)$ is closer to 1 as λ approaches 0. Once λ is determined, $C_n(\lambda)$ is decided as well. To determine a suitable value of λ, a validation index is defined as follows:

$$\lambda - C_n(\lambda)/n. \tag{4}$$

The rationale for the validation index is presented as follows. A good partition should have a low inter-cluster relation value and a high intra-cluster relation value. If there is only one cluster, there is no inter-cluster relation. That is, the number of clusters increases as the value of the inter-cluster relation rises. Hence, we use $C_n(\lambda)/n$ to represent the inter-cluster relation, where $C_n(\lambda)/n \leqslant 1$. On the other hand, the intra-cluster relation is represented by λ. The relation between the financial ratios within a cluster would be high if λ is large. Thus the validation index $\lambda - C_n(\lambda)/n$ would be large, as λ is large and $C_n(\lambda)/n$ is small, which proves to be a good partition.

If R^λ is an equivalence relation, we can use the equivalence relation to form a partition or classification of Y. The partition or classification of Y refers to a family of disjoint subsets, say, $\{SP_1, SP_2, \ldots, SP_i, \ldots\}$. The partition must satisfy the following conditions, i.e.

$$SP_1 \cup SP_2 \cup \cdots \cup SP_i \cup \cdots = Y \tag{5}$$

and

$$SP_i \cap SP_j = \varnothing \quad \forall i \neq j, \tag{6}$$

where SP_i indicates the set composed of the financial ratios in cluster i.

In this paper, R^λ merely satisfies the reflexive and symmetric laws, but may not satisfy the transitive law. It may be that $SP_i \cap SP_j \neq \varnothing, \exists i \neq j$, that is, a financial ratio may belong to two different clusters. In such cases, a transitive closure [8,17] is commonly employed to construct an equivalence relation from the compatible relation. Here, we deal with the situation using a different approach. To solve the problem, two additional definitions are introduced as follows.

Definition 4.1. If $r(Y_i, Y_k) \in R^\lambda$, $r(Y_i, Y_j) \in R^\lambda$ and $r(Y_k, Y_j) \in R^\lambda$, then Y_i, Y_k and Y_j belong to the same cluster.

Definition 4.2. If $Y_t \in SP_i \cap SP_j$ and $RSP_i(Y_t) \geqslant RSP_j(Y_t)$, then $SP'_j = SP_j - \{Y_t\}$ and SP'_j substitutes for SP_j until $SP_i \cap SP'_j = \varnothing$, for $i \neq j$, where $RSP_i(Y_t) = \min_{Y_k \in SP_i}\{r(Y_t, Y_k)\}$ and $RSP_j(Y_t) = \min_{Y_l \in SP_j}\{r(Y_t, Y_l)\}$.

A simple example is presented to demonstrate that the method is rationale. Assume there are four financial ratio sequences as shown in the following:

$Y_1 = (1, 2, 3, 4)$,
$Y_2 = (100, 200, 300, 400)$,
$Y_3 = (40, 30, 20, 10)$
and
$Y_4 = (400, 300, 200, 100)$.

Observing the four patterns, we find that Y_1 and Y_2 are similar in variation, as are Y_3 and Y_4, although Y_2 and Y_3 are more similar than Y_1 and Y_2. Given the pattern variation, Y_1 and Y_2 should be in the same cluster, and Y_3 and Y_4 should be in another cluster. Applying our clustering method to partition Y_1, Y_2, Y_3 and Y_4, we have

$$M_i(k, k+1) = \begin{cases} 0.183 & \text{if } k = 1,2,3 \text{ and } i = 1,2; \\ -0.183 & \text{if } k = 1,2,3 \text{ and } i = 3,4. \end{cases}$$

Then, the fuzzy relation defined on Y_1, Y_2, Y_3 and Y_4 is presented as follows:

$$r(Y_i, Y_j) = \begin{cases} 1 & \text{if } 1 \leqslant i,j \leqslant 2 \text{ or } 3 \leqslant i,j \leqslant 4; \\ 0 & \text{otherwise.} \end{cases}$$

Based on the above fuzzy relation, we can construct a 4×4 matrix corresponding to the relation:

$$\begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ 0 & 0 & 1 & \\ 0 & 0 & 1 & 1 \end{bmatrix},$$

where $r(Y_i, Y_j) = r(Y_j, Y_i)$, so $r(Y_j, Y_i)$ is omitted when $j < i$.

The partitions for different λ's are shown in Table 2.

According to the validation index, the sequences are partitioned into two clusters, (Y_1, Y_2) and (Y_3, Y_4), which coincide with our intuition.

Another example is also presented to demonstrate the effectiveness of the method. Assume that there are six financial ratio sequences, $f_1, f_2, \ldots, f_6$ of four companies A_1, A_2, A_3, A_4, shown in Table 3.

According to Table 3, the variation for each pair of different companies is presented in Table 4.

Based on Table 4, the matrix of the fuzzy relation is

$$\begin{bmatrix} 1 & & & & & \\ 0.56 & 1 & & & & \\ 0.93 & 0.62 & 1 & & & \\ 0.98 & 0.57 & 0.94 & 1 & & \\ 0.64 & 0.92 & 0.67 & 0.63 & 1 & \\ 0.57 & 0.97 & 0.62 & 0.58 & 0.93 & 1 \end{bmatrix}.$$

In the matrix, the entry $r(Y_i, Y_j) = r(Y_j, Y_i)$, $i,j = 1, 2, \ldots, 6$, so $r(Y_j, Y_i)$ is omitted when $j < i$.

The partitions of λ's with their validation indices are shown in Table 5, where we find that the maximum value of the validation index is 0.587 with $\lambda = 0.92$. By taking $\lambda = 0.92$, we partition the financial ratios into two clusters.

Table 2
Partitions of different λ's

λ	$C_n(\lambda)$	Clustering arrangement	Validation index
1	2	$(Y_1, Y_2), (Y_3, Y_4)$	0.5
0	1	(Y_1, Y_2, Y_3, Y_4)	−0.25

Table 3
The financial ratio sequences of four companies

Ratio	A_1	A_2	A_3	A_4
f_1	0.62	0.84	0.86	0.60
f_2	0.55	0.12	0.16	0.60
f_3	0.19	0.26	0.30	0.29
f_4	0.50	0.70	0.75	0.55
f_5	0.66	0.25	0.26	0.71
f_6	0.88	0.20	0.25	0.85

Table 4
The pair-wise variation for two different companies

Ratio	$M_i(A_1, A_2)$	$M_i(A_2, A_3)$	$M_i(A_3, A_4)$
f_1	0.149	0.014	−0.176
f_2	−0.513	0.048	0.525
f_3	0.133	0.076	−0.019
f_4	0.158	0.039	−0.158
f_5	−0.396	0.010	0.435
f_6	−0.538	0.040	0.474

Table 5
Partitions of different λ's

λ	$C_n(\lambda)$	Clustering arrangement	Validation index
1	6	f_1,f_2,f_3,f_4,f_5,f_6	0
0.98	5	$(f_1,f_4),f_2,f_3,f_5,f_6$	0.147
0.97	4	$(f_1,f_4),(f_2,f_6),f_3,f_5$	0.303
0.93	3	$(f_1,f_3,f_4),(f_2,f_6),f_5$	0.430
0.92	2	$(f_1,f_3,f_4),(f_2,f_5,f_6)$	0.587
0.56	1	$(f_1,f_2,f_3,f_4,f_5,f_6)$	0.393

Finally, we can select a representative indicator from each cluster according to Definitions 4.3 and 4.4. The two definitions are described below.

Definition 4.3. Let cluster SP_i be a set composed of several financial ratios. Then Y_t is a candidate representative indicator of SP_i as $Y_t \in SP_i$ and $RSP_i(Y_t) \geqslant RSP_i(Y_l)$, $\forall Y_l \in SP_i$.

Then, we can select a representative indicator by the following definition.

Definition 4.4. Let CSP_i be the set consisting of candidate representative indicators in the cluster SP_i. Then Y_t is a representative indicator of SP_i, if $Y_t \in CSP_i$ and $ERSP_i(Y_t) \leqslant ERSP_i(Y_j)$, $\forall$ $Y_j \in CSP_i$, where $ERSP_i(Y_t) = \max_{Y_k \notin SP_i}\{r(Y_t, Y_k)\}$ and $ERSP_i(Y_j) = \max_{Y_k \notin SP_i}\{r(Y_j, Y_k)\}$. Once the number of the representative indicators in SP_i is greater than one, we may choose any of them to serve as the representative.

By the above definitions, we can apply the clustering method to partition the financial ratio sequences and select the representative indicators for the empirical study as stated below.

5. Empirical study

An empirical study is described in this section. The data come from four shipping companies, A_1, A_2, A_3 and A_4, in Taiwan. Their financial ratios are shown in Table 6.

Table 6
The financial ratios of four shipping companies

Ratio	A_1	A_2	A_3	A_4
Current ratio (F_1)	1.002	1.573	1.421	1.025
Fixed ratio (F_2)	2.742	2.409	1.621	1.428
Equity ratio (F_3)	0.578	0.532	0.527	0.665
Fixed/long-term ratio (F_4)	1.044	1.834	1.311	4.399
Debt ratio (F_5)	0.422	0.468	0.473	0.335
Equity/debt ratio (F_6)	0.729	0.879	0.898	0.503
Operation cost ratio (F_7)	0.844	0.983	0.979	0.933
Gross profit ratio (F_8)	0.156	0.017	0.021	0.067
Operation profit ratio (F_9)	−0.054	−0.014	0.003	0.032
Income before tax ratio (F_{10})	−0.069	0.049	0.019	0.123
Net income ratio (F_{11})	0.059	−0.005	0.036	0.079
Return on current assets (F_{12})	0.071	−0.013	0.143	0.377
Return on fixed assets (F_{13})	0.072	−0.030	0.114	0.189
Return on total assets (F_{14})	0.015	−0.007	0.037	0.088
Return on stockholder's equity (F_{15})	0.026	−0.012	0.071	0.133
Return on income before tax to capital (F_{16})	0.045	−0.019	0.091	0.169
Current assets turnover (F_{17})	1.205	2.586	3.929	4.784
Fixed assets turnover (F_{18})	1.212	6.042	3.151	2.401
Total assets turnover (F_{19})	0.256	1.335	1.024	1.119
Stockholder's equity turnover (F_{20})	0.442	2.509	1.943	1.682
Current liabilities turnover (F_{21})	1.207	4.067	5.583	4.901
Long-term liabilities turnover (F_{22})	1.265	11.083	4.132	10.562
Total liabilities turnover (F_{23})	0.606	2.855	2.164	3.345

In Table 6, the debt ratio, equity/debt ratio and operation cost ratio are expressed in the form of a reciprocal.

The financial ratios in Table 1 were initially divided into four categories. We employ the fuzzy relation in Section 4 to construct relation matrices for the four categories. Let T_1, T_2, T_3 and T_4 represent relation matrices for solvency ($F_1 \sim F_6$), profitability ($F_7 \sim F_{11}$), return on investment ($F_{12} \sim F_{16}$), and asset and debt turnover ($F_{17} \sim F_{23}$), respectively.

$$T_1 = \begin{bmatrix} 1 & & & & & \\ 0.78 & 1 & & & & \\ 0.75 & 0.84 & 1 & & & \\ 0.63 & 0.59 & 0.67 & 1 & & \\ 0.72 & 0.82 & 0.97 & 0.69 & 1 & \\ 0.66 & 0.77 & 0.90 & 0.73 & 0.93 & 1 \end{bmatrix},$$

$$T_2 = \begin{bmatrix} 1 & & & & \\ 0.79 & 1 & & & \\ 0.71 & 0.61 & 1 & & \\ 0.64 & 0.53 & 0.82 & 1 & \\ 0.73 & 0.85 & 0.71 & 0.53 & 1 \end{bmatrix},$$

$$T_3 = \begin{bmatrix} 1 & & & & \\ 0.77 & 1 & & & \\ 0.96 & 0.82 & 1 & & \\ 0.88 & 0.89 & 0.93 & 1 & \\ 0.85 & 0.92 & 0.90 & 0.97 & 1 \end{bmatrix}$$

and

$$T_4 = \begin{bmatrix} 1 & & & & & & \\ 0.61 & 1 & & & & & \\ 0.77 & 0.84 & 1 & & & & \\ 0.72 & 0.89 & 0.95 & 1 & & & \\ 0.89 & 0.72 & 0.80 & 0.82 & 1 & & \\ 0.60 & 0.82 & 0.78 & 0.76 & 0.58 & 1 & \\ 0.79 & 0.76 & 0.92 & 0.87 & 0.77 & 0.81 & 1 \end{bmatrix}.$$

In the relation matrices, the entry $r(Y_i, Y_j) = r(Y_j, Y_i)$, so $r(Y_j, Y_i)$ is omitted when $j < i$.

Since financial ratios are initially divided into four categories, $C_n(\lambda)$ can be calculated respectively. The result is presented in Tables 7–10.

In Table 7, we find that the maximum value of the validation index is 0.423 at $\lambda = 0.59$. By taking $\lambda = 0.59$, we partition the financial ratios into one cluster.

Table 7
The clustering arrangements for different λ's in Category 1

λ	$C_n(\lambda)$	Clustering arrangement	Validation index
1	6	$F_1, F_2, F_3, F_4, F_5, F_6$	0
0.97	5	$(F_3, F_5), F_1, F_2, F_4, F_6$	0.137
0.90	4	$(F_3, F_5, F_6), F_1, F_2, F_4$	0.233
0.78	3	$(F_1, F_2), (F_3, F_5, F_6), F_4$	0.280
0.67	2	$(F_1, F_2), (F_3, F_4, F_5, F_6)$	0.337
0.59	1	$(F_1, F_2, F_3, F_4, F_5, F_6)$	0.423

Table 8
The clustering arrangements for different λ's in Category 2

λ	$C_n(\lambda)$	Clustering arrangement	Validation index
1	5	$F_7, F_8, F_9, F_{10}, F_{11}$	0
0.85	4	$(F_8, F_{11}), F_7, F_9, F_{10}$	0.050
0.82	3	$(F_8, F_{11}), (F_9, F_{10}), F_7$	0.220
0.73	2	$(F_7, F_8, F_{11}), (F_9, F_{10})$	0.330
0.53	1	$(F_7, F_8, F_9, F_{10}, F_{11})$	0.330

Table 9
The clustering arrangements for different λ's in Category 3

λ	$C_n(\lambda)$	Clustering arrangement	Validation index
1	5	$F_{12}, F_{13}, F_{14}, F_{15}, F_{16}$	0
0.97	4	$(F_{15}, F_{16}), F_{12}, F_{13}, F_{14}$	0.170
0.96	3	$(F_{12}, F_{14}), (F_{15}, F_{16}), F_{13}$	0.360
0.89	2	$(F_{12}, F_{14}), (F_{13}, F_{15}, F_{16})$	0.490
0.77	1	$(F_{12}, F_{13}, F_{14}, F_{15}, F_{16})$	0.570

Table 10
The clustering arrangements for different λ's in Category 4

λ	$C_n(\lambda)$	Clustering arrangement	Validation index
1	7	$F_{17}, F_{18}, F_{19}, F_{20}, F_{21}, F_{22}, F_{23}$	0
0.95	6	$(F_{19}, F_{20}), F_{17}, F_{18}, F_{21}, F_{22}, F_{23}$	0.093
0.89	5	$(F_{17}, F_{21}), (F_{19}, F_{20}), F_{18}, F_{22}, F_{23}$	0.176
0.87	4	$(F_{17}, F_{21}), (F_{19}, F_{20}, F_{23}), F_{18}, F_{22}$	0.299
0.82	3	$(F_{17}, F_{21}), (F_{18}, F_{22}), (F_{19}, F_{20}, F_{23})$	0.391
0.76	2	$(F_{17}, F_{21}), (F_{18}, F_{19}, F_{20}, F_{22}, F_{23})$	0.474
0.58	1	$(F_{17}, F_{18}, F_{19}, F_{20}, F_{21}, F_{22}, F_{23})$	0.437

In Table 8, we find that the maximum value of the validation index is 0.330 at $\lambda = 0.73$ or 0.53. By taking $\lambda = 0.73$, we partition the financial ratios into two clusters because $0.73 > 0.53$.

In Table 9, we find that the maximum value of the validation index is 0.570 at $\lambda = 0.77$. By taking $\lambda = 0.77$, we partition the financial ratios into one cluster.

In Table 10, we find that the maximum value of the validation index is 0.474 at $\lambda = 0.76$. By taking $\lambda = 0.76$, we partition the financial ratios into two clusters.

Clearly, the maximum values of the validation index in the four categories are 0.423, 0.330, 0.570 and 0.474, respectively. The clustering outcomes of the financial ratios in the four categories are shown in Table 11.

We then select the representative indicator of financial ratios in each cluster according to Definitions 4.3 and 4.4. The result is shown in Table 12.

Finally, we compare our clustering method with others to demonstrate the feasibility of the proposed method. Among the existing methods, K-means is commonly applied in clustering research. We intend to utilize K-means to prove whether our method is reasonable or not. A simple introduction to K-means is given below.

K-means assigns each item to the cluster with the nearest mean. The process is composed of three steps:

Step 1. Partition items into k clusters.
Step 2. Proceed through the list of items, and assign an item to the cluster whose mean is the closest; then recalculate the mean for the cluster that received a new item and for the cluster which lost the item.
Step 3. Repeat Step 2 until no more reassignments take place.

In Step 2, the distance from each item to the mean is expressed by the Euclidean distance. Thus the objective function of K-means defined with the Euclidean norm is stated as follows:

Table 11
The clustering arrangements for financial ratios in the four categories

Category	Cluster	Ratio
Category 1: Solvency	1	$F_1, F_2, F_3, F_4, F_5, F_6$
Category 2:	2	F_7, F_8, F_{11}
Profitability	3	F_9, F_{10}
Category 3: Return on investment	4	$F_{12}, F_{13}, F_{14}, F_{15}, F_{16}$
Category 4:	5	F_{17}, F_{21}
Asset and debt turnover	6	$F_{18}, F_{19}, F_{20}, F_{22}, F_{23}$

Table 12
Representative indicators for the financial ratios

Cluster	Ratio	Representative indicator
1	$F_1, F_2, F_3, F_4, F_5, F_6$	F_5
2	F_7, F_8, F_{11}	F_8
3	F_9, F_{10}	F_{10}
4	$F_{12}, F_{13}, F_{14}, F_{15}, F_{16}$	F_{15}
5	F_{17}, F_{21}	F_{17}
6	$F_{18}, F_{19}, F_{20}, F_{22}, F_{23}$	F_{19}

$$\text{Min}: \sum_{i=1}^{k} \sum_{j \in I_i} (\|x_j - z_i\|)^2, \tag{7}$$

where $z_i = \frac{\sum_{j \in I_i} x_j}{Nu(I_i)}$, I_i represents cluster i, z_i indicates the mean of cluster i and $Nu(I_i)$ denotes the number of items on I_i.

Since K-means assigns any item to the cluster with the nearest mean, we can utilize the condition to determine whether the result shown in Table 11 is reasonable or not. In Table 11, Categories 1 and 3 are partitioned into one cluster, so we merely calculate the squared distance sums in Categories 2 and 4. In addition, the financial sequences have to be normalized to meet the requirements of the K-means method. The calculation of the squared distance sums is shown as follows.

According to the K-means calculation, the minimized distance sums in Tables 13 and 14 match the clustering result in Table 11, so the proposed method yields reasonable cluster outputs. Comparing the two methods, the proposed method has advantages over the K-means method. First, the proposed method objectively determines the number of clusters. Second, our method does not reassign any items to a cluster, in contrast to the K-means method. In short, our method is more practical than the K-means method.

Table 13
The squared distance sums of financial ratios in Category 2: Profitability

Cluster	Squared distance sums to group means				
	F_7	F_8	F_9	F_{10}	F_{11}
(F_7, F_8, F_{11})	0.1326	0.1218	2.5157	1.4137	0.1051
(F_9, F_{10})	1.7918	2.4983	0.1371	0.1371	1.5522

Table 14
The squared distance sums of financial ratios in Category 4: Asset and debt turnover

Cluster	Squared distance sums to group means						
	F_{17}	F_{18}	F_{19}	F_{20}	F_{21}	F_{22}	F_{23}
(F_{17}, F_{21})	0.0083	0.2883	0.0733	0.1079	0.0083	0.2055	0.0568
$(F_{18}, F_{19}, F_{20}, F_{22}, F_{23})$	0.1470	0.0636	0.0062	0.0158	0.0958	0.0483	0.0322

6. Conclusions

In this paper, we have proposed a clustering method based on the fuzzy relation derived from variations in the financial ratio sequences of different companies. As the number of clusters is unknown, the proposed clustering method objectively partitions financial ratio sequences into clusters. Therefore, the clustering method can be applied in conditions where the cluster number is not determined. This is the major difference between the proposed method and the K-means method. On the other hand, the four definitions mentioned in this paper can substitute for the transitive closure law; therefore, the representative indicators can be selected by this mechanism as well. In short, we provide a new approach to resolve tie-breaks in cluster outcomes. Additionally, the partitioned results are proven to be similar to those of the K-means method when compared in an empirical study. That is to say, the proposed method has the strengths of the K-means method without many of its weaknesses.

References

[1] S. Bandyopadhyay, U. Maulik, An evolutionary technique based on K-means algorithm for optimal clustering in R^N, Information Sciences 146 (2002) 221–237.
[2] J.S. Deogun, D. Kratsch, G. Steiner, An approximation algorithm for clustering graphs with dominating diametral path, Information Processing Letters 61 (1997) 121–127.
[3] R. Dubes, A. Jains, Algorithm that Cluster Data, Prentice-hall, Englewood Cliffs, NJ, 1988.
[4] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.
[5] K.B. Eom, Fuzzy clustering approach in supervised sea-ice classification, Neurocomputing 25 (1999) 149–166.
[6] S.S. Epp, Discrete Mathematics with Applications, Wadsworth, Canada, 1990.
[7] C.M. Feng, R.T. Wang, Performance evaluation for airlines including the consideration of financial ratios, Journal of Air Transport Management 6 (2000) 133–142.
[8] F. Guoyao, An algorithm for computing the transitive closure of a fuzzy similarity matrix, Fuzzy Sets and Systems 51 (1992) 189–194.
[9] S. Hirano, X. Sun, S. Tsumoto, Comparison of clustering methods for clinical databases, Information Sciences 159 (2004) 155–165.
[10] R.A. Johnson, D.W. Wichern, Applied Multivariate Statistical Analysis, Prentice-Hall, Englewood Cliffs, NJ, 1992.
[11] L. Kaufman, P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, Wiley, New York, 1990.
[12] S.S. Khan, A. Ahmad, Cluster center initialization algorithm for K-means clustering, Pattern Recognition Letters 25 (2004) 1293–1302.
[13] R. Krishnapuram, J.M. Keller, A possibilistic approach to clustering, IEEE Transactions on Fuzzy Systems 1 (1993) 98–110.
[14] R.J. Kuo, L.M. Ho, C.M. Hu, Integration of self-organizing feature map and K-means algorithm for market segmentation, Computers and Operations Research 29 (2002) 1475–1493.
[15] R.J. Kuo, K. Chang, S.Y. Chien, Integration of self-organizing feature maps and genetic algorithm based clustering method for market segmentation, Journal of Organizational Computing and Electronic Commerce 14 (2004) 43–60.
[16] R.J. Kuo, J.L. Liao, C. Tu, Integration of ART2 neural network and genetic K-means algorithm for analyzing Web browsing paths in electronic commerce, Decision Support Systems 40 (2005) 355–374.
[17] H.S. Lee, Automatic clustering of business processes in business systems, European Journal of Operational Research 114 (1999) 354–362.
[18] J.B. MacQueen, Some methods for classification and analysis of multivariate observations, Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, University of California Press, Berkeley, Calif, 1967, pp. 281–297.
[19] S. Miyamoto, Information clustering based on fuzzy multisets, Information Processing and Management 39 (2003) 195–213.
[20] W. Pedrycz, G. Vukovich, Logic-oriented fuzzy clustering, Pattern Recognition Letters 23 (2002) 1515–1527.
[21] H. Ralambondrainy, A conceptual version of the K-means algorithm, Pattern Recognition Letters 16 (1995) 1147–1157.
[22] B.M. Walter, F.M. Robert, Accounting: The Basis for Business Decisions, McGraw-Hill, New York, 1988.
[23] K.L. Wu, M.S. Yang, Alternative *c*-means clustering algorithms, Pattern Recognition 35 (2002) 2267–2278.