Wasserstein Metric Based Adaptive Fuzzy Clustering Methods for Symbolic Data

Hong Li
School of Management, Fuzhou University, Fujian 350108, P.R.China, Tel: 18950281892
e-mail: 16484492@qq.com

Abstract

Given the current limitations in fuzzy clustering metric, the aim of this paper is to present new wasserstein metric based adaptive fuzzy clustering methods for partitioning symbolic interval data. Wasserstein metric shows adavantages in digging distribution information in symbolic interval data. Besides, the proposed fuzzy clustering methods also emphasize correlation structure between indices. Based on it, fuzzy partitions and prototypes for clusters are determined by optimizing adequacy criteria. Finally, the applicability and effectiveness of the proposed methods are validated through experiments with synthetic data sets.

Keywords: *adaptive fuzzy clustering, symbolic interval data, wasserstein metric, optimization model*

1. Introduction

SDA (Symbolic Data Analysis) is a new research field in knowledge discovery and data management, and is closely related with multi-dimensional data analysis, pattern recognition and artificial intelligence. It aims to use appropriate methods to analyze, dig implicit information in different symbolic data. Fuzzy clustering is an important branch of fuzzy pattern recognition, it is an unsupervised pattern recognition method, and was widely used in many fields. At present, fuzzy clustering method is generally divided into 5 categories in the international academic research: clustering method based on similarity relation, clustering method based on objective function, based on the transitive closure of fuzzy relation, clustering neural network and clustering method based on advanced algorithm. With the development of computer and the actual problem, clustering method based on objective function has become the mainstream of fuzzy clustering method. Many scholars have made a useful research. Souza, De Carvalho and Diday [1-3] respectively use city-block distance, Hausdorff distance and the Euclidean distance to study the fuzzy clustering of interval data. The three distance formula above are often used in clustering algorithm. But the distance above give too much emphasis to endpoints of interval data, and neglect concentration and discrete distribution of data, the results may be easy to lose information in data distribution. In this paper, we will be the first study of fuzzy clustering using wasserstein measure into interval data, and through the CR index contrasted with other methods, we get the superiority of the method. Research also shows that, if we pay equal attention to indices of interval data, we may ignore the index itself and the inherent correlation structure between indices. In this paper, the first and the second part introduce constraint relationship between indices, and give fuzzy clustering theory model of single index and double index of interval data based on wasserstein measure. The third part identify the relevant methods through simulation experiment and get advantages compared with prior classic methods. Finally gives the conclusion and the improvements discussed.

2. The Proposed Method

2.1. Wasserstein Metric

If random variables X and Y have distribution functions $\Psi(X)$ and $\Phi(Y)$ respectively, then the wasserstein L_2 metric is defined as follows:

Received June 24, 2013; Revised August 24, 2013; Accepted September 17, 2013

$$d(\Psi(X),\Phi(Y)) = \int_0^1 \left|\Psi^{-1}(X) - \Phi^{-1}(Y)\right| dt \qquad (1)$$

Where Ψ^{-1} and Φ^{-1} are the inverse functions of the two distributions.

In 1999, the distance is extended to wasserstein metric by Barrio [4].

$$d(\Psi(X),\Phi(Y)) = \int_0^1 (\Psi^{-1}(X) - \Phi^{-1}(Y))^2 dt^{\frac{1}{2}} \qquad (2)$$

In 2007, based on the first moment, and two moments of distribution functions, the wasserstein metric is decomposed by Irpino and Romanoas [5] follows:

$$d_w^2(\Psi_X,\Phi_Y) = \underbrace{(\mu_X - \mu_Y)^2}_{Location} + \underbrace{(\sigma_X - \sigma_Y)^2}_{Size} + \underbrace{2\sigma_X\sigma_Y(1-\rho_{QQ}(\Psi_X,\Phi_Y))}_{Shape} \qquad (3)$$

$$\rho_{QQ}(X,Y) = \frac{\int_0^1 (\Psi^{-1}(t) - \mu_X)(\Phi^{-1}(t) - \mu_Y)dt}{\sigma_X\sigma_Y} = \frac{\int_0^1 \Psi^{-1}(t)\Phi^{-1}(t)dt - \mu_X\mu_Y}{\sigma_X\sigma_Y} \qquad (4)$$

The wasserstein metric takes comprehensive consideration of three factors: first, the center position (Location): two distribution functions may differ in position, wasserstein metric uses the mean difference description. Secondly, wasserstein metric use s ρ_{QQ} tandard deviations and correlation coefficent to describe fluctuation difference between size and shape. Notablly, the correlation coefficent ρ_{QQ} is different from traditional Pearson correlation coefficient. ρ_{QQ} measures the difference of density function shape. ρ_{QQ} =1, if and only if the two standardized distribution function is the same. Compared with the traditional city-block distance, Hausdorff distance and Euclidean distance, wasserstein metric is no longer focused on the endpoint, but to capture the data distribution information, consider the center of distribution and fluctuation difference, so can fully use the information provided by distribution function.

In this paper, we assume a uniform distribution of X and Y in a range of interval, then the wasserstein metric of interval $X = [a,b]$ and interval $Y = [u,v]$ is defined as follows:

$$d_W(U(a,b),\ U(u,v)) = \sqrt{(\mu_X - \mu_Y)^2 + (\sigma_X - \sigma_Y)^2} \qquad (5)$$

Where $\mu_X = \frac{1}{2}(a+b)$， $\mu_Y = \frac{1}{2}(u+v)$， $\sigma_X = \sqrt{\frac{(b-a)^2}{12}}$， $\sigma_Y = \sqrt{\frac{(v-u)^2}{12}}$

For the p -dimension interval variables, the above formula can be extended to:

$$d_w^2(X,Y) = \sum_{j=1}^{p}[\left|\frac{a_j+b_j}{2} - \frac{u_j+v_j}{2}\right|^2 + \frac{1}{3}\left|\frac{b_j-a_j}{2} - \frac{v_j-u_j}{2}\right|^2] \qquad (6)$$

2.2. Adaptive Fuzzy Clustering Theory and Model for Single-index

Let $\Omega = \{1,\cdots,n\}$ be a set of n patterns $k = 1,2,\cdots,n$. Each pattern is described by p symbolic interval variables. A symbolic interval variable X is a correspondence, which is defined from Ω to R so that for each $k \in \Omega$, $X(k) = [a,b] \in \Lambda$, where $\Lambda = \{[a,b] | a,b \in R, a \le b\}$ is the set of closed intervals defined in the real number set R. Each pattern k is represented as a vector of intervals $x_k = (x_{k1}, x_{k2},..., x_{kp})$, where

$x_{kj}=[a_{kj},b_{kj}]\in\Lambda$. In this paper, an interval datum matrix $(x_{kj})_{n\times p}$ is made up of n rows representing the n patterns to be clustered, and p columns representing p symbolic interval variables. Each entry of this matrix is an interval $x_{kj}=[a_{kj},b_{kj}]\in\Lambda$ ($k=1,2,\cdots,n$; $j=1,2,\cdots,p$). ($k=1,2,\cdots,n$; $j=1,2,\cdots,p$).

Our goal is to divide n patterns into c categories. So let the prototype g_i of each cluster P_i ($i=1,2,\cdots,c$) be represented as a vector of intervals $g_i=(g_{i1},g_{i2},...,g_{ip})$, where $g_{ij}=[\alpha_{ij},\beta_{ij}]\in\Lambda$ ($j=1,2,\cdots,p$).

As the standard fuzzy algorithm, the fuzzy clustering method for symbolic interval data aims to determine a fuzzy partition of a set of patterns from c clusters $\{P_1,\cdots,P_c\}$ and a corresponding set of prototypes $\{g_1,\cdots,g_c\}$ so that a criterion function W^1 measuring the fitness between clusters and their representatives (i.e., prototypes) is locally minimized. The criterion function is based on a non-adaptive distance between vectors of intervals and is defined as follows:

$$W^1=\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^2\Phi(x_k,g_i) \tag{7}$$

Where $\Phi(x_k,g_i)$ are the distances between patterns x_k and prototypes g_i, u_{ik} satisfyconditions as follows:

$$\begin{cases}\sum\limits_{i=1}^{c}u_{ik}=1 & (k=1,2,\cdots,n)\\ 0\le u_{ik}\le 1 & (i=1,2,\cdots,c;k=1,2,\cdots,n)\end{cases} \tag{8}$$

For single-index, the distance is defined according to the structure of a cluster P_i and is described by a vector of coefficients $\lambda_i=(\lambda_{i1},\cdots,\lambda_{ip})$. We define the single-index adaptive wasserstein distance between the two vectors of intervals x_k and g_i as follows:

$$d_i(x_k,g_i)=\sum_{j=1}^{p}\lambda_{ij}[(\frac{a_{kj}+b_{kj}}{2}-\frac{\alpha_{ij}+\beta_{ij}}{2})^2+\frac{1}{3}(\frac{b_{kj}-a_{kj}}{2}-\frac{\beta_{ij}-\alpha_{ij}}{2})^2] \tag{9}$$

From Equation (9), the criterion function based on the above adaptive distance between vectors of intervals is defined as follows:

$$W^1=\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^2\sum_{j=1}^{p}\lambda_{ij}[(\frac{a_{kj}+b_{kj}}{2}-\frac{\alpha_{ij}+\beta_{ij}}{2})^2+\frac{1}{3}(\frac{b_{kj}-a_{kj}}{2}-\frac{\beta_{ij}-\alpha_{ij}}{2})^2] \tag{10}$$

$$\min\{\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^2\sum_{j=1}^{p}\lambda_{ij}[(\frac{a_{kj}+b_{kj}}{2}-\frac{\alpha_{ij}+\beta_{ij}}{2})^2+\frac{1}{3}(\frac{b_{kj}-a_{kj}}{2}-\frac{\beta_{ij}-\alpha_{ij}}{2})^2]\} \tag{11}$$

$$s.t.\begin{cases}\sum\limits_{i=1}^{c}u_{ik}=1\ \ (k=1,2,\cdots,n)\\ 0\le u_{ik}\le 1\ (i=1,2,\cdots,c;k=1,2,\cdots,n)\\ \prod\limits_{j=1}^{p}\lambda_{ij}=1\ (i=1,2,\cdots,c)\\ \lambda_{ij}>0\ \ (i=1,2,\cdots,c;\ j=1,2,\cdots,p)\end{cases}$$

Then, the optimization model is constructed as follows Equation (10).

In order to solve Equation (11), the Lagrange function is constructed as follows:

$$\begin{aligned}&L(u_{ik},\alpha_{ij},\beta_{ij},\lambda_{ij},\Phi_1,\Phi_2)=\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^2\sum_{j=1}^{p}\lambda_{ij}\\&[(\frac{a_{kj}+b_{kj}}{2}-\frac{\alpha_{ij}+\beta_{ij}}{2})^2+\frac{1}{3}(\frac{b_{kj}-a_{kj}}{2}-\frac{\beta_{ij}-\alpha_{ij}}{2})^2]\\&-\Phi_1(\sum_{i=1}^{c}u_{ik}-1)-\Phi_2(\prod_{j=1}^{p}\lambda_{ij}-1)\end{aligned} \tag{12}$$

Using the Lagrange multiplier method, we get:

$$\begin{cases}\alpha_{ij}=\dfrac{\sum\limits_{k=1}^{n}(u_{ik})^2a_{kj}}{\sum\limits_{k=1}^{n}(u_{ik})^2}\\\beta_{ij}=\dfrac{\sum\limits_{k=1}^{n}(u_{ik})^2b_{kj}}{\sum\limits_{k=1}^{n}(u_{ik})^2}\end{cases} \tag{13}$$

$$\lambda_{ij}=\frac{\left\{\prod\limits_{h=1}^{p}[\sum\limits_{k=1}^{n}u_{ik}^2[(\frac{a_{kh}+b_{kh}}{2}-\frac{\alpha_{ih}+\beta_{ih}}{2})^2+\frac{1}{3}(\frac{b_{kh}-a_{kh}}{2}-\frac{\beta_{ih}-\alpha_{ih}}{2})^2]\right\}^{\frac{1}{p}}}{\sum\limits_{k=1}^{n}u_{ik}^2[(\frac{a_{kj}+b_{kj}}{2}-\frac{\alpha_{ij}+\beta_{ij}}{2})^2+\frac{1}{3}(\frac{b_{kj}-a_{kj}}{2}-\frac{\beta_{ij}-\alpha_{ij}}{2})^2]}$$

$$u_{ik}=[\sum_{h=1}^{c}\frac{\sum\limits_{j=1}^{p}\lambda_{ij}[(\frac{a_{kj}+b_{kj}}{2}-\frac{\alpha_{ij}+\beta_{ij}}{2})^2+\frac{1}{3}(\frac{b_{kj}-a_{kj}}{2}-\frac{\beta_{ij}-\alpha_{ij}}{2})^2]}{\sum\limits_{j=1}^{p}\lambda_{hj}[(\frac{a_{kj}+b_{kj}}{2}-\frac{\alpha_{hj}+\beta_{hj}}{2})^2+\frac{1}{3}(\frac{b_{kj}-a_{kj}}{2}-\frac{\beta_{hj}-\alpha_{hj}}{2})^2]}]^{-1} \tag{14}$$

2.3. Algorithm

The fuzzy clustering algorithm is summarized as follows:

(1) Initialization. Choose fixed c $(2\le c<n)$ and $\varepsilon>0$. u_{ik} $(k=1,\cdots,n$ and $i=1\cdots,c$) of pattern k belonging to cluster P_i are chosen so that $u_{ik}\ge 0$ and $\sum_{i=1}^{c}u_{ik}=1$. Let $t=1$

(2) The membership degrees u_{ik} of patterns k belonging to clusters P_i are fixed, Compute the prototypes g_i of classes P_i $(i=1,\cdots,c)$ using Equation (13).

(3) the membership degrees u_{ik} of pattern k belonging to cluster P_i and the prototypes g_i of classes P_i are fixed,compute the vector of weights λ_i $(i=1,\cdots,c)$ using Equation (14).

(4) Update the fuzzy membership degrees u_{ik} of patterns k belonging to clusters P_i ($i=1,\cdots,c$) using Equation (15).

(5) Stopping criterion. If $\left|W_{t+1}^{1}-W_{t}^{1}\right|\le\varepsilon$, then Stop, Else let $t=t+1$ and go to step 2.

2.4. Adaptive Fuzzy Clustering Theory and Model for Double-index

As described by the wasserstein metric, the advantage is reflected in its consideration of the central tendency and fluctuation of interval variables, so we consider giving different index to central and fluctuate part. And we get:

$$W^1 = \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^2\,\Phi(x_k, g_i)$$

$$= \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^2 \sum_{j=1}^{p}[\lambda_{ij}^{m}(\frac{a_{kj}+b_{kj}}{2}-\frac{\alpha_{ij}+\beta_{ij}}{2})^2+\frac{1}{3}\lambda_{ij}^{v}(\frac{b_{kj}-a_{kj}}{2}-\frac{\beta_{ij}-\alpha_{ij}}{2})^2]$$

Where,

$$\begin{cases}\sum_{i=1}^{c}u_{ik}=1\ (k=1,2,\cdots,n)\\ \prod_{j=1}^{p}\lambda_{ij}^{m}=1\ (i=1,2,\cdots,c)\\ \prod_{j=1}^{p}\lambda_{ij}^{v}=1\ (i=1,2,\cdots,c)\\ 0\le u_{ik}\le 1, \lambda_{ij}^{m}>0, \lambda_{ij}^{v}>0\ (i=1,2,\cdots,c; k=1,2,\cdots,n; j=1,2,\cdots,p)\end{cases} \tag{15}$$

And λ_{ij}^{m} is the adaptive index of central part , λ_{ij}^{v} is the adaptive index of fluctuation.

Using the Lagrange multipliers method, we get:

$$\begin{cases}\alpha_{ij}=\dfrac{\sum_{k=1}^{n}(u_{ik})^2 a_{kj}}{\sum_{k=1}^{n}(u_{ik})^2}\\ \beta_{ij}=\dfrac{\sum_{k=1}^{n}(u_{ik})^2 b_{kj}}{\sum_{k=1}^{n}(u_{ik})^2}\end{cases} \tag{16}$$

$$\lambda_{ij}^{m}=\frac{\prod_{h=1}^{p}[\sum_{k=1}^{n}u_{ik}^2(\frac{a_{kh}+b_{kh}}{2}-\frac{\alpha_{ih}+\beta_{ih}}{2})^2]^{\frac{1}{p}}}{\sum_{k=1}^{n}u_{ik}^2(\frac{a_{kj}+b_{kj}}{2}-\frac{\alpha_{ij}+\beta_{ij}}{2})^2}$$

$$\lambda_{ij}^{v}=\frac{\prod_{h=1}^{p}[\sum_{k=1}^{n}u_{ik}^2(\frac{b_{kh}-a_{kh}}{2}-\frac{\beta_{ih}-\alpha_{ih}}{2})^2]^{\frac{1}{p}}}{\sum_{k=1}^{n}u_{ik}^2(\frac{b_{kj}-a_{kj}}{2}-\frac{\beta_{ij}-\alpha_{ij}}{2})^2} \tag{17}$$

$$u_{ik}=[\sum_{h=1}^{c}\frac{\sum_{j=1}^{p}[\lambda_{ij}^{m}(\frac{a_{kj}+b_{kj}}{2}-\frac{\alpha_{ij}+\beta_{ij}}{2})^2+\frac{1}{3}\lambda_{ij}^{v}(\frac{b_{kj}-a_{kj}}{2}-\frac{\beta_{ij}-\alpha_{ij}}{2})^2]}{\sum_{j=1}^{p}[\lambda_{hj}^{m}(\frac{a_{kj}+b_{kj}}{2}-\frac{\alpha_{hj}+\beta_{hj}}{2})^2+\frac{1}{3}\lambda_{hj}^{v}(\frac{b_{kj}-a_{kj}}{2}-\frac{\beta_{hj}-\alpha_{hj}}{2})^2]}]^{-1} \tag{18}$$

3. Simulation and Results

We choose two synthetic interval data sets with different shapes and sizes to compare four fuzzy clustering algorithms considering different adaptive distances. For synthetic interval data sets, rectangles are built from three clusters of points drawn from three bi-variate normal distributions. We use the Corrected Rand (CR) index [6-8] for comparing two partitions. The CR index measures the similarity between a priori partition and a partition determined by a partitioning clustering algorithm. CR takes its values on the interval [0, 1], where 1 indicates perfect agreement between partitions, whereas values near 0 correspond to cluster agreement found by chance.

3.1. Synthetic Symbolic Interval Data Sets

In order to compare the results, we use the same data point presented in Souza and De Carvalho [1-3]. Data sets 1 and data sets 2 have 150 points respectively, Data sets 3 has 50 points.

The data points of each cluster in this data set were drawn according to the following parameters:

Class 1: $\mu_1 = 5,\ \mu_2 = 250,\ \sigma_1^2 = 5,\ and\ \sigma_2^2 = 30$

Class 2: $\mu_1 = 45,\ \mu_2 = 320,\ \sigma_1^2 = 5,\ and\ \sigma_2^2 = 30$

Class 3: $\mu_1 = 20,\ \mu_2 = 150,\ \sigma_1^2 = 5,\ and\ \sigma_2^2 = 5$

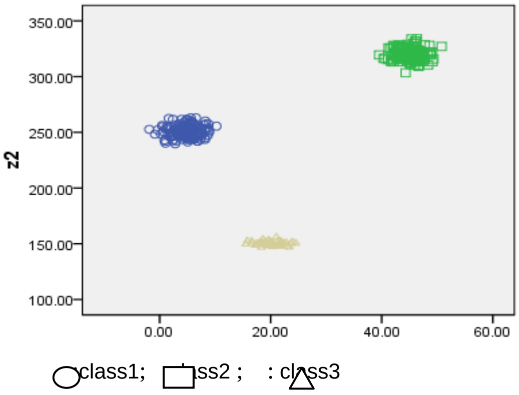

Figure 1. Data Set 1

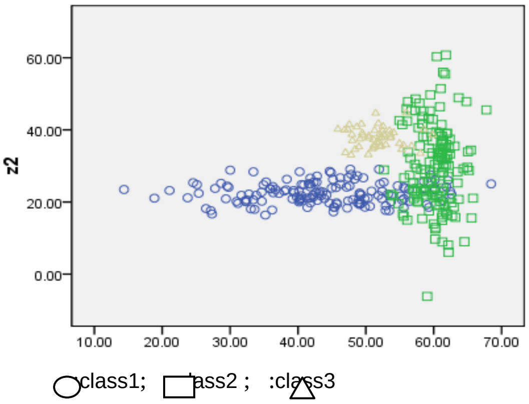

Figure 2. Data Set 2

As we can see from Figure 1, data set 1 shows well-separated clusters. Data set 2 shows overlapping clusters, depicted as in Figure 2. The data points of each cluster in this data set were drawn according to the following parameters:

In order to build interval data sets from data sets 1 and 2, each point (z1, z2) of these data sets is considered as the “seed” of a rectangle. Each rectangle is therefore a vector of two intervals expressed by $([z_1 - \gamma_1/2, z_1 + \gamma_1/2],[z_2 - \gamma_2/2, z_2 + \gamma_2/2])$. The parameters γ_1 and γ_2 are the width and the height of the rectangle. They are drawn randomly within a given range of values. In the framework of a Monte Carlo experiment, 100 replications of the previous process were carried out for parameters γ_1 and γ_2, which are drawn randomly 100 times from each of the intervals [1, 8], [1, 16], [1, 24], [1, 32] and [1, 40].

The above data sets are used to compare the following dynamic fuzzy clustering algorithms considering different adaptive distances: adaptive Hausdorff distance, one component adaptive city-block distance, the single-adaptive and double-adaptive fuzzy clustering methods proposed in this paper. For each 100 replications, the average CR index is calculated. Table 1 gives the values of the average CR index for the interval data sets 1 and 2 as well as γ_1 and γ_2 drawn from the intervals [1, 8], [1, 16], [1, 24], [1, 32] and [1, 40].

For the data configurations with well separated classes, the average CR indices of adaptive wasserstein distance are better than those of other methods. Moreover, the CR indices of the double method are better than those of the dynamic clustering algorithms: adaptive Hausdorff, city-block distances and single wasserstein distance regardless of the ranges of the predefined intervals in Table 1.

Table 1. Comparison of the Methods According to the Average CR Index

Predefined	Interval data set 1				Interval data set 2			
intervals	sin wass.	dou wass	City-block	Hausd.	sin wass.	dou wass	City-block	Hausd.
[1, 8]	0.951	0.954	0.933	0.923	0.502	0.562	0.464	0.448
[1, 16]	0.952	0.952	0.934	0.979	0.451	0.581	0.425	0.434
[1, 24]	0.979	0.987	0.987	0.957	0.423	0.632	0.399	0.418
[1, 32]	0.857	0.888	0.764	0.919	0.412	0.601	0.385	0.412
[1, 40]	0.818	0.861	0.683	0.868	0.391	0.581	0.367	0.393

Notice that, for data configurations with overlapping classes, the double wasserstein distance clustering algorithm clearly outperforms the other methods. And, while single clustering method has almost the same performance as the dynamic clustering methods based on adaptive Hausdorff and city-block distances.

So we get the following conclusion: no matter well seperated classes or overlapping classes, double-adaptive fuzzy clustering methods has the best performance.The reason is wasserstein metric dig the mean and variance distribution information in interval data while other metric focus on endpoints. Besides, double-adaptive fuzzy clustering methods give adaptive weights to indices, such as $\prod_{j=1}^{p}\lambda_{ij}=1$. It emphasizes inherent correlation structure between indices while others are not.

In order to compare the computing efficiency of different methods, we show objective values and interations times from resluts of program running. It is easily seen from Table 2 that computing efficiency of the adaptive wasserstein algorithm is better than those of other methods. It means methods proposed in this paper have higher computational efficiency in large-scale data operation although it need to calculate more complex parameters such as λ_{ij}^{m} and λ_{ij}^{v}.

Table 2. Comparison of Objective Functions and Average Iteration Times

Predefined intervals		Interval data set 1				Interval data set 2			
		sin wass.	dou wass.	City-block	Hausd.	Sin wass.	dou wass	City -block	Hausd.
[1, 8]	Objective values	7869.0	7235.6	19632.0	9816.0	10091.5	7523.7	27462.1	13731.2
	iteration times	10	10	14	10	23	23	26	26
[1,16]	Objective values	7869.0	7235.0	19673.0	9816.0	10091.5	7523.7	27462.2	13731.2
	iteration times	10	10	14	10	23	23	25	23
[1,24]	Objective values	6065.0	5823.0	19638.9	9815.0	7923.8	6923.6	27462.4	8891.5
	iteration times	10	10	14	10	23	23	27	23
[1,32]	Objective values	6065.0	5823.0	19632.0	6608.4	7923.8	6923.6	27462.4	8891.5
	iteration times	10	10	13	10	22	23	26	27
[1,40]	Objective values	6065.0	5823.0	19632.0	6608.48	7923.81	6923.6	27462.2	8891.5
	iteration times	10	10	13	19	22	23	22	26

4. Conclusion

The choice of a metric is an important task when a fuzzy clustering of interval data is performed.Based on the defined wasserstein metric, the single-index and double-index adaptive fuzzy clustering algorithms for symbolic interval data are introduced. Compared with other methods, wasserstein metric considers the density of points within the intervals and the mean and variance of intervals. Besides, adaptive parameters in the fuzzy clustering model was introduced which considers correlation of indices. Finally, simulation experiments are carried out with two artificial interval data sets and show the usefulness and validity of proposed clustering methods. Compared with Hausdorff and city-block distances fuzzy clustering models, the proposed methods performance better not only in clustering results but also in computing efficiency.

References

[1] Souza RMCR, De Carvalho. Clustering of interval data based on city-block distances. *Pattern Recognition Lett*. 2004; 25: 353–365.

[2] De Carvalho, Souza RMCR, Chavent M, Lechevallier Y. Adaptive Hausdorff distances and dynamic clustering of symbolic data. *Pattern Recognition Lett*. 2006: 27167–179.

[3] Billard L, Diday E. From the statistics of data to the statistics of knowledge: Symbolic data analysis*.* A*mer. Statist. Asso*. 2003; 98: 470–487.

[4] Barrio E, Matra C. Tests of goodness of fit based on the L_2 wasserstein distance. *Annals of Statistics*. 1999; 27: 1230-1279.

[5] Irpino A, Romano E. Optimal histogram representation of large data sets: Fisher vs piecewise linear approximations. *Revue des Nouvelles Technologiesde Information*. 2007; 23: 99-110.

[6] Hubert L, Arabie P. Comparing Partitions. *Journal of Classification*. 1985; l2: 193-218.

[7] Yushu Xiong. A clustering Algorithm Based on Rough Set and Genetic Algorithm. *TELKOMNIKA Indonesian Journal of Electrical Engineering.* 2013; 11(10): 5782-5788.

[8] Weilin Li, Pan Fu, Erqin Zhang. Application of Fractal Dimensions and Fuzzy Clustering to Tool Wear Monitoring. *TELKOMNIKA Indonesian Journal of Electrical Engineering.* 2013; 11(1): 187-194.