Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 163–168

DOI: 10.15439/2015F174
ACSIS, Vol. 5

New similarity index based on the aggregation of membership functions through OWA operator

Amine AÏT YOUNES, Frédéric BLANCHARD, Michel HERBIN
Université de Reims Champagne Ardenne, France
CReSTIC,
Email: {amine.ait-younes, frederic.blanchard, michel.herbin}@univ-reims.fr

***Abstract*—In the field of data analysis, the use of metrics is a classical way to assess pairwise similarity. Unfortunately the popular distances are often inoperative because of the noise, the multidimensionality and the heterogeneous nature of data. These drawbacks lead us to propose a similarity index based on fuzzy set theory. Each object of the dataset is described with the vector of its fuzzy attributes. Thanks to aggregation operators, the object is fuzzified by using the fuzzy attributes. Thus each object becomes a fuzzy subset within the dataset. The similarity of a reference object compared to another one is assessed through the membership function of the fuzzified reference object and an aggregation method using OWA operator.**

I. Introduction

ASSESSING the similarity between samples is a key of success in data analysis process. Many methods rely on similarity indices. Most of clustering ones uses pairwise comparisons when aggregating or separating samples [11], [12]. In the framework of case-based reasoning, solving problem needs for searching similar cases and assessing their similarities [10]. Recommender systems also deal with similarity between objects [18]. Thus the search for pairwise similarity indices remains an active field of research [2], [14]. The choice of similarity measures depends on the representation of objects we compare [5] [6]. In this paper, we restrict the scope of this study to the comparison of vector data.

When data is described with multidimensional vectors, the use of metrics remains the classical way to assess pairwise similarity [19], [6]. Unfortunately, database objects could have qualitative features making difficult to obtain standardized quantitative vectors of attributes from the objects. Thus the popular distances (Euclidean distance, Mahalanobis distance, Minkowski metric, Cosine distance, Correlation distance,...) become often inoperative. Moreover noise or vagueness can corrupt data and the curse of dimensionality is also an obstacle in processing queries in high-dimensional space [3]. These drawbacks lead us to propose a similarity index which is not based on a distance function or a metric in the data space.

To overcome these difficulties, the fuzzy set theory gives a framework to design similarity indices [17], [20]. In this paper, the dataset forms the context for our pairwise comparisons between objects.

Each object X_i of the dataset is fuzzified. Let $\tilde{X}_i$ be the fuzzy set obtained from the crisp object X_i, $\tilde{X}_i$ is the fuzzy version of X_i. The membership degree of the crisp object X_j to the fuzzy set $\tilde{X}_i$ is considered as the similarity value from X_j to the reference object X_i. Therefore the similarity indices we propose are only fuzzy membership functions. Note that such similarity indices based on membership functions do not necessarily define symmetric relations.

The challenge using our approach becomes to obtain a fuzzification of an object X_i within the dataset [13]. To achieve this goal, the attributes of the object X_i are considered as fuzzy numbers or fuzzy quantities. Then each crisp object X_j is described with a vector of membership degrees relative to the fuzzy attributes of X_i. Thanks to fuzzy logic operators, we aggregate these membership degrees to obtain the aggregated membership value of X_j to the fuzzy set $\tilde{X}_i$. The critical issue of this approach is the aggregation method we use. This communication proposes to adapt OWA operators [16] to define our aggregation method.

The paper is organized as follows.

In The sections 2 and 3 we present our approach for the fuzzification of the attributes. The section 4 exposes the methodology we use to evaluate the sensibility and specificity of each attributes. After the fuzzification of the data, we present in the section 5 the aggregation procedure used to build a new similirity index using an Ordered Weighted Aggregation operator. Before concluding, we present in the section 6 a comparison of our new pairwise similarity indice with the popular metrics.

II. Domain of fuzzification

Let E be a set of n objects defined by:

$$E = \{X_i \ / \ 1 \leq i \leq n\} \tag{1}$$

where X_i are the n objects of E.

Each object is described by a vector of p attributes. The object X is represented by the p-tuple (x_{ik}) with $1 \leq k \leq p$ where x_{ik} is the value of the k-th attribute of the object X_i. These p attributes are either quantitative or qualitative. If the k-th attribute is quantitative, then its values lie within an interval $[a_k, b_k]$ of $\Re$. If the k-th attribute is qualitative, then its values are within a set $\{v_1, v_2, v_3, ...v_l\}$ of l values. In both cases, we call D_k the set we use to define the k-th attribute.

The domain of definition D of E is defined by:

$$D = \prod_{1 \leq k \leq p} D_k \tag{2}$$

Then we have: $E \subset D$ with: $\#E = n$. In the following each object becomes a fuzzy subset of E relatively to its attributes. Thus E is called domain of fuzzification.

III. Fuzzification of the attributes

This section is devoted to the fuzzification of an object X_i within the dataset E. Although the fuzzification of an attribute is itself beyond the scope of this document, we firstly describe the way we used to fuzzify each attribute value of the object X_i. Thus we obtain k fuzzy attributes for X_i. Then we merge these fuzzy attributes to build the fuzzy object $\tilde{X}_i$ defined in E.

Let X_i be an arbitrary reference object of the data set E. Let x_{ik} be the value of the k-th attribute of X_i. The values of attributes are often imprecise and the meaning could be vague. Therefore it is convenient to represent such imprecise or vague values by fuzzy sets. Thus x_{ik} is represented by a fuzzy subset of D_k. The membership function m_k^i of this fuzzy subset is defined by:

$$m_k^i : \begin{array}{ccc} D_k & \longrightarrow & [0,1] \\ x & \longmapsto & m_k^i(x) \end{array} \tag{3}$$

In this paper, these fuzzy sets are normalized with $m_k^i(x_{ik}) \leq 1$.

In this paper, we propose a simple and empirical approach of the data fuzzification. Each numeric value is represented by a conventional trapezoidal membership function defined by (a, b, c, d) with (cf. fig. 1):

$$m_k^i(x) = \begin{cases} 0 & \text{if } x < a_i \\ \frac{x - a_i}{b_i - a_i} & \text{if } a_i \leq x < b_i \\ 1 & \text{if } b_i \leq x < c_i \\ \frac{d_i - x}{d_i - c_i} & \text{if } c_i \leq x < d_i \\ 0 & \text{if } d_i \leq x \end{cases} \tag{4}$$

Let $\overline{x}_k$ and σ_k be respectively the mean and the standard deviation of the k-th attribute within D_k. If dev_k^i is the deviation between x_{ik} and $\overline{x}_k$ (i.e. $dev_k^i = |x_{ik} - \overline{x}_k|$), an empirical study leads us to propose:

$$\begin{cases} a_i = x_{ik} - \sigma_k - 0.5\, dev_k^i \\ b_i = x_{ik} - 0.5\, \sigma_k - 0.1\, dev_k^i \\ c_i = x_{ik} + 0.5\, \sigma_k + 0.1\, dev_k^i \\ d_i = x_{ik} + \sigma_k + 0.5\, dev_k^i \end{cases} \tag{5}$$

If the k-th attribute is qualitative, x_{ik} is fuzzified using a degree of membership for each possible value of the attribute. Then m_k^i is defined by l values $(m_k^i(v_1), m_k^i(v_2), ...m_k^i(v_l))$ within D_k.

This paper proposes to use m_k^i to fuzzify the object X_i within E in respect with its k-th attribute. The membership function of the object X_i is defined by:

$$\mu_k^i : \begin{array}{ccc} E & \longrightarrow & [0,1] \\ X_j & \longmapsto & \mu_k^i(X_j) = m_k^i(x_{jk}) = \mu_{jk}^i \end{array} \tag{6}$$

with $1 \leq j \leq n$ and $1 \leq k \leq p$.

If the value of x_{ik} is not set, then we propose to define μ_k^i by simply $\mu_k^i(X_j) = \frac{1}{2}$ in order to ensure the robustness of the proposed approach.

At this stage of the communication, several points should be noted. Each object X_i gives rise to p fuzzy subsets of E.

Each subset is associated with an attribute. These fuzzy subsets are defined with reference to the object X_i. They are normalized because $\mu_{jk}^i \leq 1$.

We propose to consider the membership degrees μ_{jk}^i (with $X_j \in E$) as similarity values from X_j to the reference X_i with respect to the k-th attribute. If $\mu_{jk}^i = 1$, then X_j and X_i are considered as similar with respect to the k-th attribute. In contrast, if $\mu_{jk}^i = 0$, then X_j and X_i are considered as dissimilar with respect to the attribute. The more μ_{jk}^i is close to 1, the larger the similarity from X_j to X_i for the k-th attribute. Thus the membership function μ_{ik} (with $1 \leq k \leq p$) is considered as a similarity index to X_i with respect to its attribute x_{ik}.

We can see in "fig. 1" that x_{j_3k} is considered as similar to x_{ik} but x_{j_1k} is not comparable to x_{ik}. This similarity value is asymmetric. In "fig. 2" x_{j_4k} is not comparable to x_{ik} : $\mu_k^i(x_{j_4k}) = 0$ but $\mu_k^{j_4}(x_{ik}) > 0$.

The membership functions μ_{ik} give p indices of similarity to X_i within the set E. Let us define two characteristics of these indices that we call the sensibility and the specificity to the similarity with X.

Let $sens_{ik}$ be the mean of μ_{jk}^i when $X_j \in E$:

$$sens_{ik} = \frac{1}{n} \sum_{X_j \in E} \mu_{jk}^i \tag{7}$$

The value $sens_{ik}$ lies between 0 and 1. It assesses an average similarity between the reference object X_i and the whole dataset E in respect with the k-th attribute.

If $sens_{ik}$ is close to 1, then the n similarity values μ_{jk}^i are also rather close to 1. Then the n objects X_j of E are rather similar to X_i. In this case, the values μ_{jk}^i are sensitive indicators of the similarity to X_i. Since these values are rather equal to 1 or close to 1, then the value μ_{jk}^i becomes highly symptomatic of a dissimilarity (non-similarity) with X_i when the indicator of similarity μ_{jk}^i is close to 0. Thus the k-th attribute is considered as an attribute sensitive to the similarity with X_i.

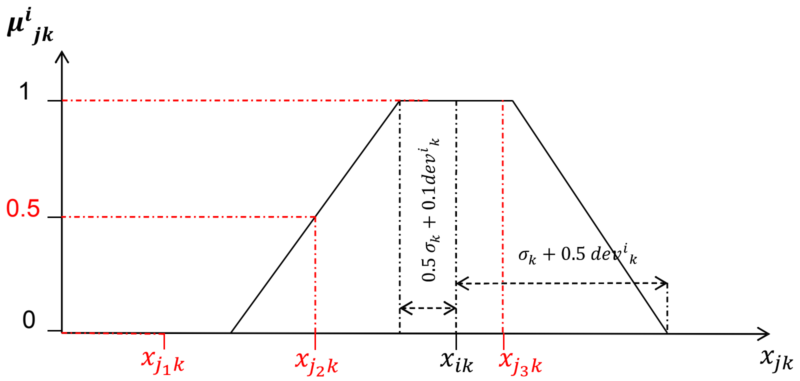

Fig. 1. Fuzzy representation.

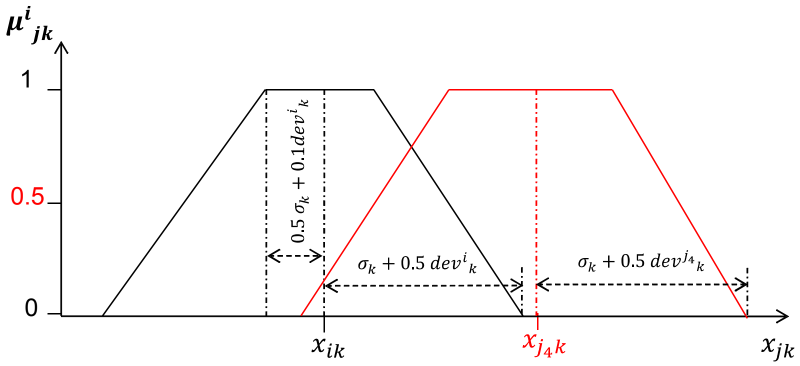

Fig. 2. Asymmetric similarity values.

In contrast, if $sens_{ik}$ is close to 0, the n similarity values μ^i_{jk} are also rather close to 0. Then the n objects X_j are rather dissimilar (non-similar) to X_i. In this case, the values μ^i_{jk} are specific indicators of the similarity to X_i. Since these values are rather equal to 0 or close to 0, then the value μ^i_{jk} becomes highly symptomatic of a similarity to X_i when the indicator of similarity μ^i_{jk} is close to 1. Thus the k-th attribute is considered as an attribute specific of the similarity with X_i.

When we consider the k-th attribute, $sens_{ik}$ is a coefficient of the sensibility of this attribute to the similarity with X_i and 1-$sens_{ik}$ is a coefficient of the specificity of the attribute for the similarity with X_i. These two coefficients characterize the k-th attribute with reference to the object X_i within E.

Let us consider an example (see Table I) to explain these coefficients. The dataset E has six objects X_1, X_2, X_3, X_4, X_5 and X_6. Each object is described with four attributes. The reference object is X_1. The four attribute values of X_1 are fuzzified. The four membership functions μ^1_1, μ^1_2, μ^1_3 and μ^1_4 indicate the degrees of similarity to X_1. In this example, the sensitivities of the four attributes are respectively 0.767, 0.400, 0.583 and 0.680. The 1st attribute is the most sensitive one. Only X_6 has 1st attribute value dissimilar from the one of X_1. Thus the 1st attribute reveals the dissimilarity with X_1. The specificities of the four attributes are respectively 0.233, 0.600, 0.417 and 0.320. The 2nd attribute is the most specific one. Only X_2 has 2nd attribute value similar to the one of X_1. Thus the 2nd attribute reveals the similarity with X_1.

TABLE I
SENSITIVITY AND SPECIFICITY OF THE ATTRIBUTES IN RESPECT WITH THE REFERENCE OBJECT X_1: EXAMPLE OF 6 OBJECTS WITH 4 FUZZY ATTRIBUTES, μ_k ARE THE DEGREES OF MEMBERSHIP TO X_1 RELATIVE TO THE k-TH ATTRIBUTE WITH $1 \leq k \leq 4$

Objects	Fuzzy attributes			
	μ^1_1	μ^1_2	μ^1_3	μ^1_4
X_1	1	1	1	1
X_2	0.9	1	0.3	0.9
X_3	0.9	0.1	0.4	0.7
X_4	0.9	0.1	0.5	0.5
X_5	0.9	0.1	0.6	0.3
X_6	0	0.1	0.7	0.2
sensitivity	0.767	0.400	0.583	0.680
specificity	0.233	0.600	0.417	0.320

IV. AGGREGATION WITH OWA OPERATORS

Let us consider the reference object X_i in E. The fuzzy subset defined by the membership function μ^i_{jk} depends on the value x_{ik} of the k-th attribute of X_i. We propose to aggregate these p fuzzy subsets taking into account all the attributes. The goal is to fuzzify the reference object X_i within E defining a new membership function μ^i fusing the functions μ^i_k.

The aggregation operators give a classical way to merge the fuzzy subsets in E. Let $aggreg$ be an aggregation operator. The function μ^i is defined by:

$$\begin{array}{rccc} \mu^i: & E & \longrightarrow & [0,1] \\ & Y & \longmapsto & \mu^i(X_j) = \underset{1\leq k\leq p}{aggreg}(\mu^i_k(X_j)) \end{array} \tag{8}$$

The aggregation operators are well studied in literature [7], [8]. The minimum is the reference operator to obtain a conjunction and the maximum is the one for a disjunction. The operators used in this paper are a tradeoff between the conjunction (AND) and the disjunction (OR).

If the similarity index μ^i_k is very sensitive ($sens_{ik}$ close to 1), then the similarity index μ^i_k should contribute to μ^i using a conjunction operator. Indeed, a conjunction seems desirable because significant information is obtained when μ^i_k is close to 0. In contrast, if the similarity index μ^i_k is very specific ($sens_{ik}$ close to 0), a disjunction operator seems preferable because significant information is obtained when μ^i_k is close to 1.

In this paper, the tradeoff between conjunction and disjunction is defined using an Ordered Weighted Aggregation (OWA operators proposed by R. Yager [16]).

Let us describe the function μ^i obtained when using such an OWA operator.

For an object X_j in E, the membership degrees $\mu^i_k(X_j)$ are ordered by decreasing order. We obtain :

$$\mu^i_{(1)}(X_j) \geq \mu^i_{(2)}(X_j) \geq \mu^i_{(3)}(X_j) \geq ... \geq \mu^i_{(p)}(X_j)$$

The aggregation is defined by:

$$\mu^i(X_j) = \sum_{1\leq k\leq p} w_k \times \mu^i_{(k)}(X_j) \tag{9}$$

We denote W the weighting vector :

$$W = [w_1, w_2, \ldots, w_p] \tag{10}$$

with $\sum_{1\leq k\leq p}(w_k) = 1$ and $w_k \in [0,1]$

The weights are not associated with attributes but with their ordered positions. The challenge is to determine the weights.

The conjunction operator (i.e. the minimum) is obtained if :

$$W_* = [0, 0, \ldots, 1] \tag{11}$$

The disjunction operator (i.e. the maximum) is obtained if :

$$W^* = [1, 0, \ldots, 0] \tag{12}$$

The ordinary average is recovered if :

$$\bar{W} = \left[\frac{1}{p}, \frac{1}{p}, \ldots, \frac{1}{p}\right] \tag{13}$$

Pérez and Lamata [15] discuss the weights determination by means of linear functions that we use in this paper.

To emphasize the conjunction, we propose to use decreasing weights w_{min} defined by using the linear orders of Borda [4] where :

$$w_{min}(k) = \frac{2k}{p(p+1)} \tag{14}$$

where $1 \leq k \leq p$.

To emphasize the disjunction, we propose to use increasing weights w_{max} defined by increasing linear orders with:

$$w_{max}(k) = \frac{2(p+1-k)}{p(p+1)} \tag{15}$$

where $1 \leq k \leq p$.

Then we have:

$$w_{min}(k) + w_{max}(k) = \frac{2}{p} \tag{16}$$

For example, if $p = 4$ $\begin{cases} W_{min} = \left[\frac{2}{20}, \frac{4}{20}, \frac{6}{20}, \frac{8}{20}\right] \\ \\ W_{max} = \left[\frac{8}{20}, \frac{6}{20}, \frac{4}{20}, \frac{2}{20}\right] \end{cases}$ (17)

The membership function μ^i_k is a similarity index with X_i in respect with the k-th attribute. $sens_{ik}$ describes the sensitivity of the index. In this paper, $sens_{ik}$ is considered as the ANDness of the attribute and 1-$sens_{ik}$ defines the specificity i.e. the ORness of the attribute. Then the weights of OWA operator we use are defined by:

$$w_k = C\left(\ (sens_{(ik)})w_{min}(k) + (1 - sens_{(ik)})w_{max}(k)\ \right) \tag{18}$$

where C is the coefficient for obtaining $\sum_{1\leq k\leq p}(w_k) = 1$.

The membership degrees $\mu^i(X_j)$ with $X_j \in E$ are obtained through these weights. Such an OWA operator in E permits us to define a similarity index sim_i from the reference X_i to the other objects X_j of E by:

$$sim(X_i, X_j) = \mu^i(X_j) \tag{19}$$

Note that the similarity index we propose is not necessarily symmetrical (cf. fig:fuzzy2). In fact $sim(X_i, X_j)$ is not always equal to $sim(X_j, X_i)$.

V. Comparison with popular metrics

The way we describe to design pairwise similarity between multidimensional data leads us to propose a new pairwise similarity index based on fuzzy logic operators. This section is devoted to the assessment of the new similarity index we propose. First we define a criterion to compare the similarity indices. Second we apply this criterion for comparing our new pairwise similarity index with more classical indices based on the popular metrics.

A. Assessment of a similarity index

Such a pairwise similarity indices are often used in the context of data clustering [9]. In this paper, we take the problem upside down using the clusters for assessing the similarity indices. The clusters define a partition of E, we call C_{X_i} the cluster to which the object X_i belongs and we call sim a similarity index between two objects of E.

Let us consider all pairs of objects (X_i, X_j) within the sample data E. If the two objects X_i and X_j belong to the same cluster, then an optimal similarity index from X_i to X_j should be equal to one (i.e. X_i and X_j are similar). On the contrary, if the two objects X_i and X_j belong to two different clusters, then an optimal similarity index from X_i to X_j should be equal to zero (i.e. X_i and X_j are dissimilar). Thus we define the intra-cluster similarity of sim with:

$$intra(sim) = \frac{1}{n_1} \sum_{C_{X_i}=C_{X_j}} sim(X_i, X_j) \qquad (20)$$

where n_1 is the number of couples (X, Y) where X and Y belong to the same cluster. The inter-cluster similarity is defined with:

$$inter(sim) = \frac{1}{n_2} \sum_{C_{X_i} \neq C_{X_j}} sim(X_i, X_j) \qquad (21)$$

where n_2 is the number of couples (X_i, X_j) where X_i and X_j belong to two different clusters.

The similarity index sim is optimal for the clusters when $intra(sim) = 1$ and $inter(sim) = 0$. Therefore we define a criterion to evaluate sim with:

$$crit(sim) = intra(sim) - inter(sim) \qquad (22)$$

The value $crit(sim)$ lies always between -1 and 1.

The higher $crit(sim)$, the more optimal sim with respect to the clusters.

We propose to use this criterion to assess our new pairwise similarity index.

B. Applications

This paper proposes a new way to evaluate the similarity between multidimensional vector data. The most classical way consists in using the popular metrics when data are quantitative. In this paper we consider Euclidean distance, Manhattan distance, Chebyshev distance, Canberra distance and Mahalanobis distance (see Table II). In fact, these distances are dissimilarity indices that we transform into similarity indices with:

$$simil(X, Y) = 1 - \frac{dist(X, Y)}{\max\limits_{A,B \in E} dist(A, B)} \qquad (23)$$

where $dist$ is the distance that we use.

Then we have five similarity indices we call $Euclidean$, $Manhattan$, $Chebyshev$, $Canberra$, and $Mahalanobis$ which are based on their three respective popular metrics.

We compare these similarity indices based on distances with two indices we propose based on the aggregation operators of membership functions. The first one is called $simOWA$ that is the index which is described in this paper. It is based on OWA operators. The second one replaces the OWA operator with the arithmetic mean of the membership functions. This second one is called $Arithmetic$.

The similarity indices are computed using the databases from *Machine Learning Repository of UCI* [1].

In this paper we propose to use six numerical multivariate clustering databases that are $iris$, $wine$, $ecoli$, $glass$, $seeds$ and $haberman$. The number of attributes lies between 3 and 15. The number of objects lies between 100 and 500. The number of clusters lies between 3 and 10. $iris$ is the classical database that has 150 iris plants with 4 attributes and three clusters. The $wine$ recognition database has 178 objects with 13 attributes and three clusters. $ecoli$ is the database of sites of protein localization, it has 336 objects with 7 attributes and eight clusters. The $glass$ identification database has 214 objects with 9 attributes and seven clusters. The $seeds$ database of wheat varieties has 210 objects with 7 attributes and three clusters. $Haberman$'s survival database has 306 objects with 3 attributes and two clusters.

The results obtained are in Table III. We can see that the similarity indices proposed in this paper (SimOWA and Arithmetic) are better than the others in 5 cases and ranked second for one of them (*glass* Database). In 5 cases of 6, the similarity indice based on the OWA operator gives us better results than the one based on the arithmetic mean.

VI. Conclusion

The approach we propose has a significant advantage, it allows us to deal with imperfection that is a general case with real data. In medicine and biology, data is often imprecise mainly due to the inherent variability of biological data. In physics, data is also imperfect and it is usual to assign a value from a sensor with the accuracy of the measurement. Qualitative data is also imprecise or vague. Thus the use of the fuzzy set theory is relevant in this context of imperfect data.

In this paper we propose a simple method of fuzzification for imperfect multidimensional data. With this fuzzification we define a new similarity indice that will allow us in future works to identify the main features of the dataset and build a robust classification. We can also use this approach to compare a new object with the existing data, for example by finding the nearest objects.

References

[1] Bache, K., Lichman, M., *UCI* Machine learning repository. http://archive.ics.uci.edu/ml, University of California, Irvine, School of Information and Computer Sciences (2013)

[2] Barrena, M., Jurado, E., Marquez, P. and Pachon, C., *A flexible framework to ease nearest neighbor search in multidimensional data spaces*, Data and Knowledge Engineering, 69, pp. 116–136 (2010)

TABLE II
DEFINITION OF THE MOST POPULAR METRICS BETWEEN QUANTITATIVE DATA

Popular metrics	
Euclidean	$dist(X,Y) = \sqrt{\sum_{1 \le k \le p} (x_k - y_k)^2}$
Manhattan (city block)	$dist(X,Y) = \sum_{1 \le k \le p} \lvert x_k - y_k \rvert$
Chebyshev	$dist(X,Y) = \max_{1 \le k \le p} \lvert x_k - y_k \rvert$
Canberra	$dist(X,Y) = \sum_{1 \le k \le p} \frac{\lvert x_k - y_k \rvert}{\lvert x_k \rvert + \lvert y_k \rvert}$
Mahalanobis	$dist(X,Y) = \sqrt{(X-Y)^T C^{-1} (X-Y)}$

TABLE III
COMPARISON OF SIMILARITY INDICES

	Database					
	$iris$	$wine$	$ecoli$	$glass$	$seeds$	$haberman$
Number of objects	150	178	336	214	210	306
Number of attributes	4	13	7	9	7	3
Number of clusters	3	3	8	7	3	2
Index of similarity						
Euclidean	0.336	0.175	0.230	0.098	0.275	0.020
Manhattan	0.331	0.176	0.210	0.097	0.272	0.026
Chebyshev	0.344	0.175	0.211	0.085	0.258	0.017
Canberra	0.422	0.222	0.142	**0.166**	0.238	0.048
Mahalanobis	0.113	0.047	0.078	0.064	0.080	0.027
Arithmetic	0.506	0.264	0.245	0.142	0.447	**0.052**
SimOWA	**0.510**	**0.270**	**0.289**	0.151	**0.451**	0.044

[3] Bohm, C., Berchtold, S. and Keim, D.A., *Searching in high-dimensional spaces: index structures for improving the performance of multi-media databases*, ACM Computing Surveys, 33(3), pp. 322–373 (2001)

[4] de Borda, M., *Memoire sur les elections au scrutin*, Academie Royale des Sciences, Paris (1784)

[5] Cha, S.H., *Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions*, International Journal of Mathematical Models and Methods in Applied Sciences, 1(4), pp. 300–307 (2007)

[6] Cunningham, P., *A Taxonomy of Similarity Mechanisms for Case-Based Reasoning*, IEEE Trans. Knowledge and Data Engineering, Vol. 21 (11), pp. 1532–1543, (2009)

[7] Detyniecki, M., *Mathematical aggregation operators and their application to video querying*, Research Report, LIP6, Paris (2001)

[8] Dubois, D. and Prade, H., *On the use of aggregation operations in information fusion processes*, Fuzzy Sets and Systems, 142, pp. 143–161 (2004)

[9] Fred, A.N.L. and Jain, A.K., *Learning Pairwise Similarity for Data Clustering*, 18th International Conference on Pattern Recognition (ICPR 2006), vol. 1, pp. 925–928 (2006)

[10] De Mantaras, R.L., McSherry, D., Bridge, D., Leake, D., Smyth, B., Craw, S., et al., *Retrieval, reuse, revision, and retention in case-based reasoning*, Knowledge Engineering Review, 20(3), pp. 215–240 (2005)

[11] Jain, A K., Murty, M.N. and Flynn, P.J., *Data clustering: a review*,ACM Computing Surveys, 31(3), pp. 264–323 (1999)

[12] Jain, A., *Data clustering: 50 years beyond k-means*, Pattern Recognition Letters, 31(8), pp. 651–666 (2010)

[13] Nourizadeh, A., Blanchard, F., Aït Younes, A., Delemer, B. and Herbin, M., *Data Analysis of Insulin Therapy in the Elderly Type 2 Diabetic Patients*, Studia Informatica Universalis, 11(3), pp. 32–49 (2013)

[14] Novak, D., Batko, M. and Zezula, P., *Large-scale similarity data management with distributed metric index*, Information Processing and Management, 48(5), pp. 855–872 (2012)

[15] Perez, E.C. and Lamata, M.T., *OWA weights determination by means of linear functions*, Mathware and Soft Computing, 16, pp. 107–122 (2009)

[16] Yager, R., *On ordered weighted averaging aggregation operators in multicriteria decision making*, IEEE Trans. Systems, Man and Cybern., 18(1), pp. 183–190 (1988)

[17] Yager, R., *Fuzzy logic methods in recommender systems*, Fuzzy Sets and Systems, 136, pp. 133–149 (2003)

[18] Zenebe, A. and Norcio, A.F., *Representation, similarity measures and aggregation methods using fuzzy sets for content-based recommender systems*, Fuzzy Sets and Systems, 160, pp. 76–94 (2009)

[19] Zerzucha, P. and Walczak, B., *Concept of (dis)similarity in data analysis*, Trends in Analytical Chemistry, 38, pp. 116–128 (2012)

[20] Zwick, R., Caristein, E. and Badescu, D.V., *Measures of similarity among fuzzy concepts: A comparative analysis*, International Journal of Approximate Reasoning, 1, pp. 221–242 (1987)