2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Universal Fuzzy Clustering Model

Mika Sato-Ilic

***Abstract*— A universal fuzzy clustering model is proposed in order to adapt to the variability of a similarity data structure. For the purpose of the consideration of the variability of a similarity data, a nonlinear fuzzy clustering model has been proposed. In the nonlinear fuzzy clustering model, the similarity is represented by common degree of membership of a pair of objects to "each" fuzzy cluster and an ordinary aggregation operator is used for adjusting variety of the common degree of membership of a pair of objects to "each" fuzzy cluster. However, the ordinary aggregation operator is a binary operator and can only adjust for variety of common degree of membership of a pair of objects to "each" fuzzy cluster, it cannot adapt the variety of common degree of membership of a pair of objects to "all" fuzzy clusters. That is, this model cannot satisfactorily adjust to the variability of the obtained similarity data structure. Therefore, we define a new aggregation operator called a generalized aggregation operator in a linear product space spanned by "all" fuzzy clusters and propose a universal fuzzy clustering model based on this generalized aggregation operator in order to adjust to the variability of the obtained similarity data structure.**

I. Introduction

Recently, clustering techniques for analyzing the increasing number of large and noisy data have received tremendous attention from many researchers due to the necessity of cleaning, summarizing, and reducing the size of complex data. We have developed a nonlinear fuzzy clustering model [10] in order to deal with the noisy data which is an extended model of an additive fuzzy clustering model [8], [9]. This model is one example of model-based clustering which is a category of clustering techniques whose essential feature is the assumption of a structure into data. Through this feature, the mathematical properties of the obtained result tend to be clearer when compared with non-model-based clustering. In this model, we assume that all objects have some common properties and each common property is defined as a fuzzy cluster. The similarity between a pair of objects is defined by using some fuzzy clusters shown as some common properties. That is, the similarity between a pair of objects is assumed to consist of some shared common properties of the objects. The shared common property is defined as common degree of memberships of a pair of objects to a fuzzy cluster. Therefore, we do not need to define the metric to represent the similarity of objects. Since the difference of the definition of the metric causes the different clustering results, avoiding any definitions of metric has a benefit for the clustering. Exploiting this property has been discussed in certain areas such as genetics which have some merit for utilizing this model [18]. Also many algorithms related with this model have been developed through several areas [6], [7], [17]. Since this model has been discussed in a framework of "additive" clustering models, the common degree of memberships of a pair of objects to a fuzzy cluster independently contributes to the similarity between the objects, so the interaction of a pair of objects for "different" fuzzy clusters cannot be considered. Therefore, we have taken the perspective of a nonlinear relationship among fuzzy clusters and have proposed a nonlinear fuzzy clustering model by extending the additive fuzzy clustering model for explaining the complexity of the noisy data.

However, in the nonlinear fuzzy clustering model, although the variety of the common degree of memberships of a pair of objects to "each" fuzzy cluster is adjusted by using an ordinary aggregation operator, the variety of the obtained similarity cannot be explained due to the fact that the aggregation operator is a binary operator and so it cannot explain the common degree of memberships of a pair of objects to "all" fuzzy clusters. Since the obtained similarity has various structures, we must consider the common degree of membership of a pair of objects over all fuzzy clusters. Therefore in this paper, we propose a general-purpose clustering model, the universal model which is inclusive of the adaptable variety defined by the common degree of memberships of a pair of objects to "all" fuzzy clusters. In order to represent the common degree of memberships of a pair of objects to "all" fuzzy clusters, we introduce a generalized aggregation operator [11] which is defined as a function on a product space of linear spaces spanned by all fuzzy clusters and have similar conditions of aggregation operators.

Such a function has been discussed mathematically as a metric on the product space considering a probabilistic space [15], [16]. However, our proposed aggregation operator dose not have the restriction of metric and so we can exploit the merit in which we do not need to consider the difference of the kinds of metric used in the fuzzy clustering model. In addition, several definitions of multidimensional aggregation operators [1], [2], [5] have been proposed. However, mathematical features of these aggregation operators are unclear. Therefore, we propose generalized aggregation operator which satisfies conditions similar to those of the aggregation operator which has suitable conditions for the clustering model in which it takes advantage of the property of degree of memberships.

This paper consists of seven sections. The following section describes the additive clustering model. Then in section 3, we state an additive fuzzy clustering model. In section 4, a nonlinear fuzzy clustering model is described. Section

Mika Sato-Ilic is with the Faculty of Systems and Information Engineering, University of Tsukuba, Tsukuba, Japan (email: mika@risk.tsukuba.ac.jp).

5 proposes a universal fuzzy clustering model. Section 6 shows several numerical examples and section 7 contains the concluding comments.

II. Additive Clustering Model

The additive clustering model [14] is defined as follows:

$$s_{ij} = \sum_{k=1}^{K} w_k p_{ik} p_{jk} + \varepsilon_{ij}, \tag{1}$$

where s_{ij} $(i, j = 1, 2, \cdots, n)$ is a similarity data between objects i and j, K is the number of clusters, and w_k is a weight representing the salience of the property corresponding to the cluster k. n is the number of objects and ε_{ij} is an error. p_{ik} shows the status of belongingness of an object i to a cluster k. If an object i has the property of a cluster k, then $p_{ik} = 1$, otherwise it is 0. Therefore, p_{ik} satisfies the following condition:

$$p_{ik} \in \{0.1\}, \tag{2}$$

and the product $p_{ik}p_{jk}$ is unity only if both objects i and j belong to the cluster k. The cluster is defined, in this model, as a subset of all objects in which the objects included in the cluster share a common property. When a pair of objects has some common properties, this model assumes that these common properties "additively" contribute to the similarity between the pair of objects. That is, the degree of contribution of each common property to the similarity is mutually independent. For example, if a pair of objects i and j together belong to clusters l_1, l_2, $\cdots$, l_m, then the similarity s_{ij} is represented by the sum of weights of these clusters as follows:

$$s_{ij} = w_{l_1} + w_{l_2} + \cdots + w_{l_m} + \varepsilon_{ij}. \tag{3}$$

Therefore, the similarity is represented by the degree of shared common properties.

III. Additive Fuzzy Clustering Model

From equation (3), it can be seen that since the similarity s_{ij} is observed as continuous values, in order to obtain the better fitness in which ε_{ij} is substantially small, the number of clusters tends to increase to explain observed similarity. In order to solve this problem, the additive fuzzy clustering model has been proposed. The additive fuzzy clustering model is defined as follows:

$$s_{ij} = \varphi(\rho_{ij}) + \varepsilon_{ij}, \tag{4}$$

where,

$$\rho_{ij} = (\rho(u_{i1}, u_{j1}), \cdots, \rho(u_{iK}, u_{jK})) \in R^K. \tag{5}$$

Suppose that there exist K fuzzy clusters on a set of n objects, that is, the partition matrix $U = (u_{ik})$ is assumed to exist under the following conditions:

$$\sum_{k=1}^{K} u_{ik} = 1, \quad i = 1, \cdots, n, \tag{6}$$

$$u_{ik} \in [0, 1], \quad i = 1, \cdots, n, \ \ k = 1, \cdots, K, \tag{7}$$

where u_{ik} shows a degree of membership of an object i to a cluster k. Let $\rho(u_{ik}, u_{jk})$ be a common degree of membership of a pair of objects i and j to a cluster k, namely, a degree of shared common property. To state simply, we assume that if all of $\rho(u_{ik}u_{jk})$ are multiplied by α, then the similarity is also multiplied by α. Therefore, the function φ itself must satisfy the condition "positively homogeneous of degree 1 in the ρ", that is,

$$\alpha\varphi(\rho_{ij}) = \varphi(\alpha\rho_{ij}), \quad \alpha > 0. \tag{8}$$

We consider the following function as a typical function of φ:

$$s_{ij} = \varphi(\rho_{ij}) + \varepsilon_{ij} = \{\sum_{k=1}^{K} \rho^r(u_{ik}, u_{jk})\}^{\frac{1}{r}} + \varepsilon_{ij}, \quad 0 < r < +\infty. \tag{9}$$

We will deal with (9) $(r = 1)$ hereafter, that is,

$$s_{ij} = \sum_{k=1}^{K} \rho(u_{ik}, u_{jk}) + \varepsilon_{ij}. \tag{10}$$

The degree ρ is the aggregation operator satisfied the following conditions defined in definition 1.

Definition 1 An aggregation operator (AO) is a binary operator ρ on the unit interval $[0, 1]$, that is a function ρ: $[0, 1] \times [0, 1] \to [0, 1]$, such that $\forall a, b, c, d \in [0, 1]$, $\ a, b, c, d \in R$, the following conditions are satisfied:

$$\rho(a, 0) = \rho(0, a) = 0, \ \ \rho(a, 1) = \rho(1, a) = a.$$

$$\rho(a, c) \le \rho(b, d), \ \ \text{whenever } a \le b, \ c \le d.$$

$$\rho(a, b) = \rho(b, a).$$

Where $[0, 1] \times [0, 1]$ shows a product space.

The first condition denotes the boundary condition which means that if one object belongs to a cluster completely, then the common degree of membership to the cluster equals the degree of the other object to the cluster, and if one object does not belong to the cluster, then it is 0. The second condition shows the condition of monotonicity that the greater the degree of membership of objects to a cluster, the greater the common degree of membership of the objects. The third condition means the condition of symmetry which is that the common degree of membership of objects i and j is equivalent to the common degree of objects j and i. T-norm [4], [12] is a typical example which satisfies the conditions in definition 1.

Algebraic product is an example of the t-norm, so if we assume as

$$\rho(u_{ik}, u_{jk}) = u_{ik}u_{jk}, \tag{11}$$

then the model (10) is represented as follows:

$$s_{ij} = \sum_{k=1}^{K} u_{ik}u_{jk} + \varepsilon_{ij}. \tag{12}$$

In this model, if we put

$$u_{ik} = \sqrt{w_k} p_{ik}, \tag{13}$$

then the additive fuzzy clustering model shown in equation (12) is reduced to be the additive clustering model shown in equation (1). Therefore, the additive clustering model is a special case of the additive fuzzy clustering model which in turn is an extended model of the additive clustering model. Moreover, if we assume equation (13) which shows the additive clustering model, from equation (2), u_{ik} will have only two values for all i as follows:

$$u_{ik} \in \{0, \sqrt{w_k}\}, \ \ \forall i. \tag{14}$$

This means that the flexibility of the representation of the fuzzy clustering result shown in equation (7) is substantially reduced to the two values shown in equation (14) when we use the additive clustering model. Therefore, the additive fuzzy clustering model can obtain a more flexible result by using fewer numbers of clusters when compared with the additive clustering model. This is caused by the change of the condition from equation (2) to equation (7), which shows the change from the hard clustering model to the fuzzy clustering model. Therefore, by introducing the concept of fuzzy logic to the additive clustering model, we can obtain a more flexible result.

IV. Nonlinear Fuzzy Clustering Model

Since the additive fuzzy clustering model shown in equation (10) assumes the mutual independence among shared common properties, the interaction of different fuzzy clusters, which is a degree of shared common properties of objects to "different" fuzzy clusters, can not be reflected in this model. Therefore, given the noisy data and the models lack of power to produce an explanation, the result tends to be imprecise.

In order to overcome this problem, one simple idea is to consider

$$\rho(u_{ik}, u_{jl}), \ k \neq l$$

as a common degree of memberships of a pair of objects i and j to clusters k and l and put this to the model shown in equation (4). Then equation (4) may be rewritten as follows:

$$s_{ij} = \varphi(\tilde{\rho}_{ij}) + \varepsilon_{ij}, \tag{15}$$

where,

$$\tilde{\rho}_{ij} = (\rho(u_{i1}, u_{j1}), \cdots, \rho(u_{i1}, u_{jK}), \cdots, \rho(u_{iK}, u_{j1}), \\ \cdots, \rho(u_{iK}, u_{jK})) \in R^{K^2}. \tag{16}$$

From equations (5) and (16), it can be seen that

$$R^K \subset R^{K^2}.$$

From the condition shown in equation (8), as a typical function of φ shown in equation (15), the following function may be considered.

$$s_{ij} = \varphi(\tilde{\rho}_{ij}) = \{\sum_{k=1}^{K}\sum_{l=1}^{K} \rho^r(u_{ik}, u_{jl})\}^{\frac{1}{r}} + \varepsilon_{ij}, \quad 0 < r < +\infty. \tag{17}$$

When $r = 1$, equation (17) is as follows:

$$s_{ij} = \sum_{k=1}^{K}\sum_{l=1}^{K} \rho(u_{ik}, u_{jl}) + \varepsilon_{ij}. \tag{18}$$

If ρ shows algebraic product, then model (18) is written as follows:

$$s_{ij} = \sum_{k=1}^{K}\sum_{l=1}^{K} u_{ik}u_{jl} + \varepsilon_{ij}. \tag{19}$$

Now, we consider two terms,

$$u_{is}u_{jt}, \ \ u_{it}u_{js}, \ \ \exists s, t \in \{1, \cdots, K\}, \ \ s \neq t$$

in model (19). Both $u_{is}u_{jt}$ and $u_{it}u_{js}$ are the product which shows the common degree of memberships of objects i and j to clusters s and t, however, the values of these two products are not always the same. In spite of the above fact, $u_{is}u_{jt}$ and $u_{it}u_{js}$ are mutually independent due to the additivity over the clusters in model (19). That is, these two terms act independently with different values to the model, even if they are the same incidence. Therefore, model (19) is not adaptable for the inclusion of the interaction of different fuzzy clusters.

In order to solve this problem, we must consider the interaction of elements of $\tilde{\rho_{ij}}$ shown in equation (16) to the model. If we assume the interaction between $u_{is}u_{jt}$ and $u_{it}u_{js}$ as follows:

$$(u_{is}u_{jt}) \times (u_{it}u_{js}), \tag{20}$$

then we obtain

$$\rho(u_{is}, u_{jt})\rho(u_{it}, u_{js}) = \rho(u_{is}, u_{js})\rho(u_{it}, u_{jt}). \tag{21}$$

From this equation, we can see that we just need to consider the interaction of the elements in model (10). Moreover, from the condition of u_{ik} shown in equations (6) and (7),

$$u_{is}u_{jt}, \ u_{it}u_{js} \in [0, 1].$$

Also, since the algebraic product is a typical t-norm, equation (20) satisfies the condition of monotonicity shown in definition 1. That is, even if $u_{is}u_{jt}$ is not always equal with $u_{it}u_{js}$, both the terms become larger, and then equation (20) monotonically becomes larger. Therefore, we involve the nonlinearity among elements of ρ_{ij} shown in equation (10) to the model.

By introducing the nonlinearity among elements of ρ_{ij} shown in equation (10), a nonlinear fuzzy clustering model is defined as follows:

$$s_{ij} = \phi\left(\sum_{k=1}^{K} \rho(u_{ik}, u_{jk})\right) + \varepsilon_{ij}, \ \ i, j = 1, \cdots, n. \tag{22}$$

We denote

$$g_\rho(\mathbf{u}_i, \mathbf{u}_j) \equiv \sum_{k=1}^{K} \rho(u_{ik}, u_{jk}), \tag{23}$$

$$\mathbf{u}_i = (u_{i1}, \cdots, u_{iK}), \ \mathbf{u}_j = (u_{j1}, \cdots, u_{jK}),$$

then model (22) can be rewritten as follows:

$$s_{ij} = \phi \circ g_\rho(\mathbf{u}_i, \mathbf{u}_j) + \varepsilon_{ij}, \quad i, j = 1, \cdots, n. \tag{24}$$

Considering inner product on Hilbert space, we introduce the kernel function κ from $R^K \times R^K$ to R which satisfy the following conditions [13]:

$$\kappa(\mathbf{u}_i, \mathbf{u}_j) = \kappa(\mathbf{u}_j, \mathbf{u}_i). \tag{25}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \kappa(\mathbf{u}_i, \mathbf{u}_j)\mathbf{u}_i\mathbf{u}_j \geq 0, \ \mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n \in R^K. \tag{26}$$

Then model (24) can be rewritten as follows:

$$s_{ij} = \kappa(\mathbf{u}_i, \mathbf{u}_j) + \varepsilon_{ij}, \quad i, j = 1, \cdots, n, \tag{27}$$

under the condition of

$$\phi \circ g_\rho(\mathbf{u}_i, \mathbf{u}_j) = \kappa(\mathbf{u}_i, \mathbf{u}_j), \quad i, j = 1, \cdots, n. \tag{28}$$

We call this model (27) a kernel fuzzy clustering model. Since the function κ satisfy the conditions (25) and (26), there exists a function Φ which satisfies the following:

$$\kappa(\mathbf{u}_i, \mathbf{u}_j) = \langle \mathbf{\Phi}(\mathbf{u}_i), \mathbf{\Phi}(\mathbf{u}_j) \rangle, \quad \forall \mathbf{u}_i, \mathbf{u}_j \in R^K, \tag{29}$$

where

$$\mathbf{\Phi}(\mathbf{u}_i) = (\phi_1(\mathbf{u}_i), \cdots, \phi_M(\mathbf{u}_i)),$$

and

$$\Phi : R^K \to R^M, \quad K < M, \tag{30}$$

where $\langle \cdot, \cdot \rangle$ shows inner product on Hilbert space. An example of κ which satisfies equations (25), (26), and (28) is as follows:

$$s_{ij} = \kappa(\mathbf{u}_i, \mathbf{u}_j) + \varepsilon_{ij} = \langle \mathbf{u}_i, \mathbf{u}_j \rangle^\alpha + \varepsilon_{ij} = \left(\sum_{k=1}^{K} u_{ik} u_{jk} \right)^\alpha + \varepsilon_{ij}, \tag{31}$$

$$\alpha \geq 1, \quad i, j = 1, \cdots, n.$$

Notice that the model (31) when $\alpha = 2$ can consider the interaction of elements of $\tilde{\rho}_{ij}$ shown in equation (20). When $\alpha = 1$, equation (31) is equivalent to equation (22) when $\rho(u_{ik}, u_{jk}) = \sum_{k=1}^{K} u_{ik} u_{jk}$ and ϕ is an identity mapping. That is, the additive fuzzy clustering model shown in equation (12) is a special case of equation (31). Therefore, the additive fuzzy clustering model is a special case of the kernel fuzzy clustering model and the kernel fuzzy clustering model is a special case of the nonlinear fuzzy clustering model. From equation (30), the kernel fuzzy clustering model can estimate the solution $\mathbf{u}_i$ in a higher dimensional space.

V. Universal Fuzzy Clustering Model

In the nonlinear fuzzy clustering model shown in equation (22), since the aggregation operator ρ is defined as a binary operator on $[0, 1]$ shown in definition 1, the aggregation operator can only adjust for variety of common degree of membership of a pair of objects to "each" fuzzy cluster, that is,

$$\rho(u_{ik}, u_{jk}), \quad i, j = 1, \cdots, n, \quad k = 1, \cdots, K,$$

and it cannot adapt the variety of common degree of membership of a pair of objects to "all" fuzzy clusters.

So, in the nonlinear fuzzy clustering model, although we can consider the variability of the common degree of memberships to "each" fuzzy cluster, we cannot consider it to "all" fuzzy clusters. In order to consider the common degree of a pair of objects over all fuzzy clusters, we must consider a space consisted of all the fuzzy clusters. Therefore, we assume a linear space spanned by all the fuzzy clusters. Then generalized aggregation operator is defined as a function on a product space of the linear spaces as follows:

$$\tilde{\rho}(\mathbf{u}_i, \mathbf{u}_j), \quad \mathbf{u}_i = (u_{i1}, \cdots, u_{iK}), \ \mathbf{u}_j = (u_{j1}, \cdots, u_{jK}),$$
$$i, j = 1, \cdots, n.$$

Definition 2 A generalized aggregation operator (GAO) is a function $\tilde{\rho}$: $X \times X \to [0, 1]$, such that $\forall \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}, \mathbf{0}, \mathbf{1} \in X$, where $\boldsymbol{a} = (a_1, \cdots, a_K)$, $\boldsymbol{b} = (b_1, \cdots, b_K)$, $\boldsymbol{c} = (c_1, \cdots, c_K)$, $\boldsymbol{d} = (d_1, \cdots, d_K)$, $\mathbf{0} = (0, \cdots, 0)$, $\mathbf{1} = (1, \cdots, 1)$, $a_k, b_k, c_k, d_k \in [0, 1]$, $k = 1, \cdots, K$, the following conditions are satisfied:

$$\tilde{\rho}(\boldsymbol{a}, \mathbf{0}) = \tilde{\rho}(\mathbf{0}, \boldsymbol{a}) = 0, \quad \tilde{\rho}(\boldsymbol{a}, \mathbf{1}) = \tilde{\rho}(\mathbf{1}, \boldsymbol{a}) = \alpha, \ \alpha \in [0, 1]. \tag{32}$$

$$\tilde{\rho}(\boldsymbol{a}, \boldsymbol{c}) \leq \tilde{\rho}(\boldsymbol{b}, \boldsymbol{d}), \quad \text{whenever } \boldsymbol{a} \leq \boldsymbol{b}, \quad \boldsymbol{c} \leq \boldsymbol{d}. \tag{33}$$

Where the following equivalence relation is assumed:

$$\boldsymbol{a} \leq \boldsymbol{b} \quad \leftrightarrow \quad a_k \leq b_k, \ k = 1, \cdots K.$$

$$\tilde{\rho}(\boldsymbol{a}, \boldsymbol{b}) = \tilde{\rho}(\boldsymbol{b}, \boldsymbol{a}). \tag{34}$$

Based on the definition 2, the following theorem is proven.

Theorem 1

The operators shown in equations (36)-(38) are GAO if the following condition shown in equation (35) is satisfied.

$$\sum_{k=1}^{K} a_k = \sum_{k=1}^{K} b_k = \sum_{k=1}^{K} c_k = \sum_{k=1}^{K} d_k = 1, \quad 2 \leq K \leq n \tag{35}$$

- Generalized Algebraic Product:

$$\tilde{\rho}(\boldsymbol{a}, \boldsymbol{b}) = \boldsymbol{a}\boldsymbol{b}^t. \tag{36}$$

- Generalized Hamacher Product

$$\tilde{\rho}(\boldsymbol{a}, \boldsymbol{b}) = \frac{\boldsymbol{a}\boldsymbol{b}^t}{\boldsymbol{a}\mathbf{1}^t + \boldsymbol{b}\mathbf{1}^t - \boldsymbol{a}\boldsymbol{b}^t}. \tag{37}$$

- Generalized Einstein Product

$$\begin{aligned} \tilde{\rho}(\boldsymbol{a}, \boldsymbol{b}) &= \frac{\boldsymbol{a}\boldsymbol{b}^t}{\mathbf{2}\mathbf{1}^t - (\boldsymbol{a}\mathbf{1}^t + \boldsymbol{b}\mathbf{1}^t - \boldsymbol{a}\boldsymbol{b}^t)} \\ &= \frac{\boldsymbol{a}\boldsymbol{b}^t}{2K - (\boldsymbol{a}\mathbf{1}^t + \boldsymbol{b}\mathbf{1}^t - \boldsymbol{a}\boldsymbol{b}^t)}. \end{aligned} \tag{38}$$

Proof

1) From condition (35), $\sum_{k=1}^{K} a_k = 1$, $a_k, b_k \in [0, 1]$. Therefore, $0 \leq a_k b_k \leq a_k$. Then $0 \leq \sum_{k=1}^{K} a_k b_k \leq \sum_{k=1}^{K} a_k = 1$. That is, $0 \leq \boldsymbol{a}\boldsymbol{b}^t \leq 1$. In equation (36), $\boldsymbol{a}\mathbf{0}^t = \mathbf{0}\boldsymbol{a}^t = 0$. From condition (35), $\boldsymbol{a}\mathbf{1}^t =$

$\mathbf{1}\boldsymbol{a}^t = \sum_{k=1}^K a_k = 1$. Hence, equation (36) satisfies conditions shown in equation (32).

2) If $a_k \le b_k,\ \ c_k \le d_k,\ \ k = 1, \cdots, K,\ a_k, b_k, c_k, d_k \in [0,1]$, then $\sum_{k=1}^K a_k c_k \le \sum_{k=1}^K b_k d_k$. Hence, equation (36) satisfies a condition shown in equation (33).

3) $\sum_{k=1}^K a_k b_k = \sum_{k=1}^K b_k a_k$. Hence, equation (36) satisfies a condition shown in equation (34).

From the above 1-3, we can prove that a generalized algebraic product shown in equation (36) is a GAO.

4) From condition (35), $0 \le a_k b_k \le \sqrt{a_k b_k} \le \frac{a_k + b_k}{2}$ $(a_k, b_k \in [0,1])$. Therefore, $0 \le \sum_{k=1}^K a_k b_k \le \sum_{k=1}^K a_k + \sum_{k=1}^K b_k - \sum_{k=1}^K a_k b_k$. Then $0 \le \frac{\boldsymbol{a}\boldsymbol{b}^t}{\boldsymbol{a}\mathbf{1}^t + \boldsymbol{b}\mathbf{1}^t - \boldsymbol{a}\boldsymbol{b}^t} \le 1$. From $\boldsymbol{a}\mathbf{0}^t = \mathbf{0}\boldsymbol{a}^t = 0$, equation (37) satisfies $\tilde{\rho}(\boldsymbol{a}, \mathbf{0}) = \tilde{\rho}(\mathbf{0}, \boldsymbol{a}) = 0$. From condition (35), $\tilde{\rho}(\boldsymbol{a}, \mathbf{1}) = \frac{\boldsymbol{a}\mathbf{1}^t}{\boldsymbol{a}\mathbf{1}^t + \mathbf{1}\mathbf{1}^t - \boldsymbol{a}\mathbf{1}^t} = \bar{\boldsymbol{a}} = \frac{\displaystyle\sum_{k=1}^K a_k}{K} = \frac{1}{K}$. In a similar way, we can show that $\tilde{\rho}(\mathbf{1}, \boldsymbol{a}) = \frac{1}{K}$. Hence, equation (37) satisfies conditions shown in equation (32).

5) If $\boldsymbol{a} \le \boldsymbol{b},\ \ \boldsymbol{c} \le \boldsymbol{d}$ is satisfied, then from equation (35), $0 \le \sum_{k=1}^K a_k c_k \le 1,\ \ 0 \le \sum_{k=1}^K b_k d_k \le 1$. Therefore, $1 \le (2 - \sum_{k=1}^K a_k c_k)(2 - \sum_{k=1}^K b_k d_k) \le 4$ and $\sum_{k=1}^K a_k c_k - \sum_{k=1}^K b_k d_k < 0$. Then $\tilde{\rho}(\boldsymbol{a}, \boldsymbol{c}) - \tilde{\rho}(\boldsymbol{b}, \boldsymbol{d}) = \frac{2(\displaystyle\sum_{k=1}^K a_k c_k - \sum_{k=1}^K b_k d_k)}{(2 - \displaystyle\sum_{k=1}^K a_k c_k)(2 - \sum_{k=1}^K b_k d_k)} < 0$. Hence, equation (37) satisfies a condition shown in equation (33).

6) $\frac{\displaystyle\sum_{k=1}^K a_k b_k}{\displaystyle\sum_{k=1}^K a_k + \sum_{k=1}^K b_k - \sum_{k=1}^K a_k b_k} = \frac{\displaystyle\sum_{k=1}^K b_k a_k}{\displaystyle\sum_{k=1}^K b_k + \sum_{k=1}^K a_k - \sum_{k=1}^K b_k a_k}$. Hence, equation (37) satisfies a condition shown in equation (34).

From the above 4-6, we can prove that a generalized hamacher product shown in equation (37) is a GAO.

7) $\sum_{k=1}^K a_k + \sum_{k=1}^K b_k - \sum_{k=1}^K a_k b_k \equiv A$. From $0 \le A \le 2,\ \ 2 \le K \le n,\ 2K - A \ge A$. Then from equation (35) and the above proof shown in the 4th item, the following inequality is obtained. $0 \le \frac{\displaystyle\sum_{k=1}^K a_k b_k}{2K - (\displaystyle\sum_{k=1}^K a_k + \sum_{k=1}^K b_k - \sum_{k=1}^K a_k b_k)} \le \frac{\displaystyle\sum_{k=1}^K a_k b_k}{\displaystyle\sum_{k=1}^K a_k + \sum_{k=1}^K b_k - \sum_{k=1}^K a_k b_k} \le 1$. Therefore, in equation (38), $0 \le \tilde{\rho}(\boldsymbol{a}, \boldsymbol{b}) \le 1$. Also, it is trivial that $\tilde{\rho}(\boldsymbol{a}, \mathbf{0}) = \tilde{\rho}(\mathbf{0}, \boldsymbol{a}) = 0$ from $\boldsymbol{a}\mathbf{0}^t = \mathbf{0}\boldsymbol{a}^t = 0$. $\tilde{\rho}(\boldsymbol{a}, \mathbf{1}) = \tilde{\rho}(\mathbf{1}, \boldsymbol{a}) = \bar{\boldsymbol{a}} = \frac{\displaystyle\sum_{k=1}^K a_k}{K} = \frac{1}{K}$. Hence, equation (38) satisfies conditions shown in equation (32).

8) If $\boldsymbol{a} \le \boldsymbol{b},\ \ \boldsymbol{c} \le \boldsymbol{d}$ is satisfied, then from equation (35), $0 \le \sum_{k=1}^K a_k c_k \le 1,\ \ 0 \le \sum_{k=1}^K b_k d_k \le 1$. Therefore, $1 \le (2(K-1) - \sum_{k=1}^K a_k c_k)(2(K-1) - \sum_{k=1}^K b_k d_k)$ and $\sum_{k=1}^K a_k c_k - \sum_{k=1}^K b_k d_k < 0$. Then $\tilde{\rho}(\boldsymbol{a}, \boldsymbol{c}) - \tilde{\rho}(\boldsymbol{b}, \boldsymbol{d}) = \frac{2(K-1)(\displaystyle\sum_{k=1}^K a_k c_k - \sum_{k=1}^K b_k d_k)}{(2(K-1) - \displaystyle\sum_{k=1}^K a_k c_k)(2(K-1) - \sum_{k=1}^K b_k d_k)} < 0$. Hence, equation (38) satisfies a condition shown in equation (33).

9) $\frac{\displaystyle\sum_{k=1}^K a_k b_k}{2K - (\displaystyle\sum_{k=1}^K a_k + \sum_{k=1}^K b_k - \sum_{k=1}^K a_k b_k)} = \frac{\displaystyle\sum_{k=1}^K b_k a_k}{2K - (\displaystyle\sum_{k=1}^K b_k + \sum_{k=1}^K a_k - \sum_{k=1}^K b_k a_k)}$. Hence, equation (38) satisfies a condition shown in equation (34).

From the above 7-9, we can prove that a generalized einstein product shown in equation (38) is a GAO.

Then the universal fuzzy clustering model is defined as follows:

$$s_{ij} = \phi \circ \tilde{\rho}(\mathbf{u}_i, \mathbf{u}_j) + \varepsilon_{ij},\ \ i, j = 1, \cdots, n. \tag{39}$$

When ϕ is an identity function and $\tilde{\rho}$ is a generalized algebraic product shown in equation (36), equation (39) is the additive fuzzy clustering model shown in equation (12). Also, when ϕ is an exponential function such as $\phi(x) = x^\alpha$, and $\tilde{\rho}$ is a generalized algebraic product shown in equation (36), the universal fuzzy clustering model shown in equation (39) is the kernel fuzzy clustering model shown in equation (31). That is, both the additive fuzzy clustering model and the kernel fuzzy clustering model are special cases of the universal fuzzy clustering model.

When $\tilde{\rho}$ is a generalized algebraic product, the following models are examples of the universal fuzzy clustering model.

$$s_{ij} = \langle \mathbf{u}_i, \mathbf{u}_j \rangle^\alpha + \varepsilon_{ij}.$$

$$s_{ij} = \frac{\langle \mathbf{u}_i, \mathbf{u}_j \rangle^\alpha}{\langle \mathbf{u}_i, \mathbf{1} \rangle^\alpha + \langle \mathbf{u}_j, \mathbf{1} \rangle^\alpha - \langle \mathbf{u}_i, \mathbf{u}_j \rangle^\alpha} + \varepsilon_{ij}. \quad (40)$$

$$s_{ij} = \frac{\langle \mathbf{u}_i, \mathbf{u}_j \rangle^\alpha}{2K - (\langle \mathbf{u}_i, \mathbf{1} \rangle^\alpha + \langle \mathbf{u}_j, \mathbf{1} \rangle^\alpha - \langle \mathbf{u}_i, \mathbf{u}_j \rangle^\alpha)} + \varepsilon_{ij}.$$

VI. Numerical Examples

In order to investigate the performance of the universal fuzzy clustering model, we use several artificially created data. Since we intend to evaluate the performance of the universal fuzzy clustering model for noise of data and the kernel fuzzy clustering model is known as a robust model for noise [10], we use the kernel fuzzy clustering model as the contrastive model.

Figures 1-5 show values of random numbers from two different normal distributions, $N(\boldsymbol{\mu}_1, \sigma_1)$, $N(\boldsymbol{\mu}_2, \sigma_2)$. The variance is 1 for all distributions, however, means are changed. We generate 50 objects with respect to bivariate for each distribution. In figures 1-5, abscissa shows the values of variable 1 and ordinate shows the values of variable 2.

Figure 6 shows a result of the universal fuzzy clustering model shown in equation (40) when each data shown in figures 1-5 are applied to this model, respectively. We obtain similarity s_{ij} as $s_{ij} = 1 - d_{ij}/\max(d_{ij})$, where d_{ij} is Euclidean distance between objects i and j and $\max(d_{ij})$ shows the maximum value of d_{ij}, $i,j = 1, \cdots, n$. The number of clusters is assumed to be 2. In this figure, abscissa shows the objects and the ordinate shows the values of degree of membership of objects to cluster 1. From the conditions shown in equations (6) and (7), it is enough to show only the degree of membership to cluster 1. From this figure, we see an unclear classification for the data shown in figures 4 and 5, but we can see the data shown in figures 1-3 are classified into the two clusters.

Figure 7 shows the result of the kernel fuzzy clustering model ($\alpha = 2$) shown in equation (31) when each data shown in figures 1-5 are applied to this model, respectively. From this figure, we can see the case where the data shown in figure 1 is classified into two clusters, but this model failed for the classification for the other data.

From the comparison between the results shown in figures 6 and 7, we can see that the results of the universal fuzzy clustering model show more adaptable results even if the data show the noise situation which is an overlapping situation of the two groups. Although the kernel fuzzy clustering model is a special case of the universal fuzzy clustering model, by considering the variety of obtained similarity data structure with the use of the different generalized aggregation operator, this comparison shows the capability to handle a variety of obtained similarity data by the use of the universal fuzzy clustering model.

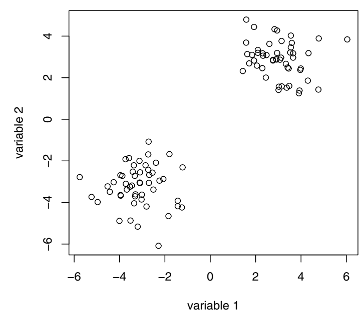

Fig. 1. Data ($\boldsymbol{\mu}_1 = (3,3), \boldsymbol{\mu}_2 = (-3,-3)$)

Figures 8 and 9 show the results of clustering for Breast Cancer data. The data used in this paper was from the Wisconsin Diagnostic Breast Cancer (WDBC) in UCI Machine Learning Repository [19]. The data represents the features that are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass with a total number of instances: of 569. Out of 30 real-valued input features 5 significant features were selected. The class distribution of instances were 357 diagnosed benign and 212 diagnosed as malignant. We selected 100 instances for each class.

Figure 8 shows the result of the universal fuzzy clustering model shown in equation (40). Figure 9 shows the result of the conventional fuzzy clustering method named fuzzy c-means method [3]. The number of clusters is assumed to be 2. Each line shows each cluster. The abscissa shows each instance where 1-100 belong to the class of benign and 101-200 are the malignant instances. The ordinate shows the degree of memberships of instances to clusters. From these figures, it can be seen that the universal fuzzy clustering model can obtain a clearer result when compared with the result of the fuzzy c-means method. Tables 1 and 2 show classification rate of both methods. For the calculation of the ratio, the instances are assigned to one cluster in which the instances have larger values of degree of memberships. From these tables, it can be seen that for the benign instances, both the methods can obtain the correct results, however, for malignant instances, the universal fuzzy clustering model can obtain the better result.

TABLE I

Classification Rate of Universal Fuzzy Clustering Model for Breast Cancer Data

	Classification Rate	Misclassification Rate
Benign	100 %	0 %
Malignant	75 %	25 %

VII. Conclusions

As a model-based clustering, a universal fuzzy clustering model is proposed for implementing a general-purpose

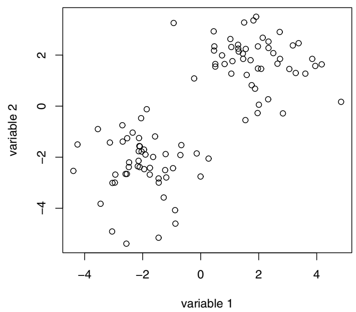

Fig. 2. Data ($\boldsymbol{\mu}_1 = (2, 2), \boldsymbol{\mu}_2 = (-2, -2)$)

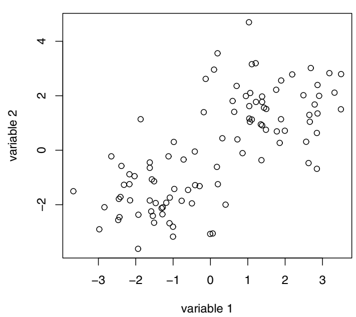

Fig. 3. Data ($\boldsymbol{\mu}_1 = (1.5, 1.5), \boldsymbol{\mu}_2 = (-1.5, -1.5)$)

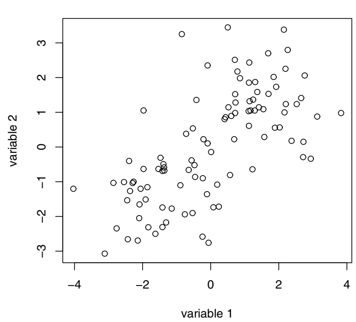

Fig. 4. Data ($\boldsymbol{\mu}_1 = (1.3, 1.3), \boldsymbol{\mu}_2 = (-1.3, -1.3)$)

TABLE II
CLASSIFICATION RATE OF FUZZY c-MEANS FOR BREAST CANCER DATA

	Classification Rate	Misclassification Rate
Benign	100 %	0 %
Malignant	54 %	46 %

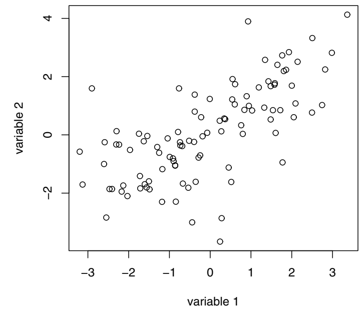

Fig. 5. Data ($\boldsymbol{\mu}_1 = (1.2, 1.2), \boldsymbol{\mu}_2 = (-1.2, -1.2)$)

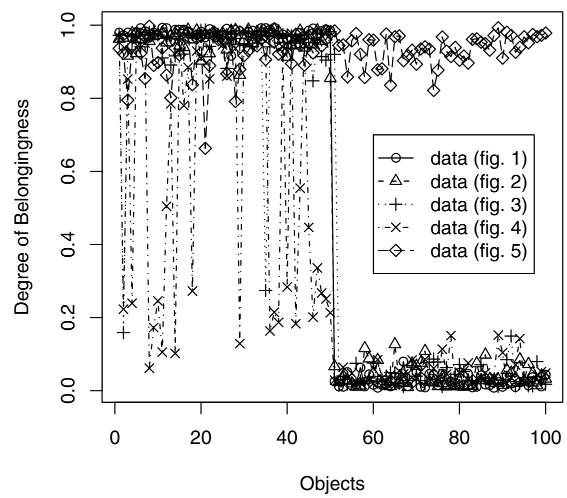

Fig. 6. Result of Universal Fuzzy Clustering Model

model. The target data of this model is similarity among objects. The universal fuzzy clustering model is defined to implement its versatility in order to adapt to the variability of the similarity data structure using the generalized aggregation operator. The generalized aggregation operator is defined as a function on a product space of linear spaces spanned by all fuzzy clusters, so we can use this operator for representing the variability of the similarity structure in the fuzzy clustering model. Several numerical examples show higher capability for dealing with the noise of data and the benefits of the use of the universal fuzzy clustering model when compared with conventional clustering methods.

REFERENCES

[1] Ai-Ping, L., J. Yan, and W. Quan-Yuan, Harmonic Triangular Norm Aggregation Operators in Multicriteria Decision Systems, Journal of Convergence Information Technology, vol. 2, no. 1, pp. 83-92, 2007.

[2] Beliakov, G, How to Build Aggregation Operators from Data, International Journal of Intelligent Systems, vol. 18, pp. 903-923, 2003.

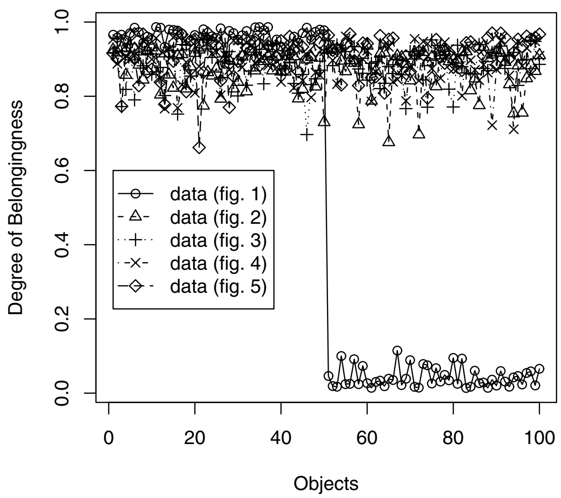

Fig. 7. Result of Kernel Fuzzy Clustering Model

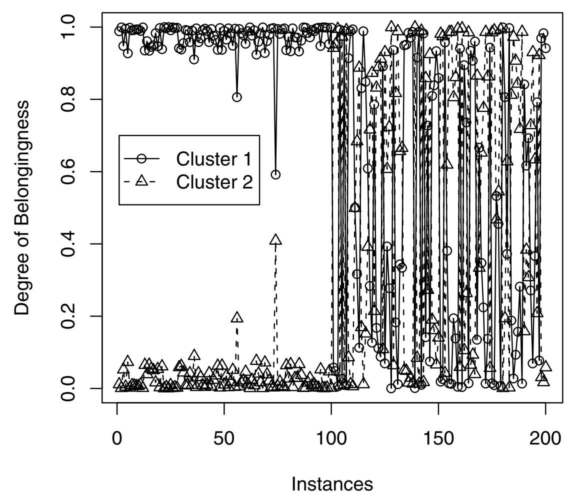

Fig. 9. Result of Fuzzy c-means for Breast Cancer Data

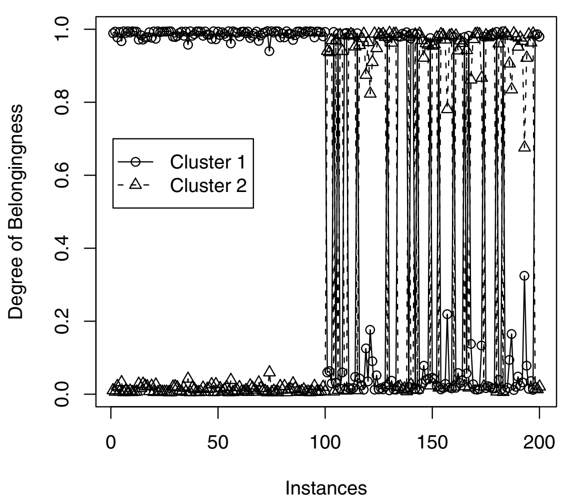

Fig. 8. Result of Universal Fuzzy Clustering Model for Breast Cancer Data

[3] Bezdek, J. C., *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, 1981.

[4] Menger, K., Statistical Metrics, Proc. Nat. Acad. Sci., USA, vol. 28, pp. 535-537, 1942.

[5] Merigo, J.M. and M. Casanovas, Fuzzy Generalized Hybrid Aggregation Operators and its Application in Fuzzy Decision Making, International Journal of Fuzzy Systems, vol. 12, no. 1, pp. 15-24, 2010.

[6] Pedrycz, W., V. Loiac, S. Senatore, Proximity based Fuzzy Clustering (P-FCM), Fuzzy Sets and Systems, vol. 148, pp. 21-41, 2004.

[7] Runkler, T. A. and F. Steinke, A New Approach to Clustering Using Eigen Decomposition, WCCI 2010 IEEE World Congress on Computational Intelligence, pp. 306-311, 2010.

[8] Sato, M. and Y. Sato, An Additive Fuzzy Clustering Model, *Japanese Journal of Fuzzy Theory and Systems*, vol. 6, pp. 185-204, 1994.

[9] Sato, M., Y. Sato, and L.C. Jain, Fuzzy Clustering Models and Applications, Springer, 1997.

[10] Sato-Ilic, M., S. Ito, and S. Takahashi, Nonlinear Kernel-Based Fuzzy Clustering Model, *Developments in Fuzzy Clustering*, D.A. Viattchenin ed., VEVER, Minsk (Belarus), pp. 56-73, 2009.

[11] Sato-Ilic, M., Generalized Aggregation Operator Based Nonlinear Fuzzy Clustering Model, Intelligent Engineering Systems through Artificial Neural Networks, vol. 20, pp. 493-500, 2010.

[12] Schweizer B., Sklar A. Probabilistic Metric Spaces, North-Holland, 1983.

[13] Shawe-Taylor, J. and N. Cristianini, Kernel Methods for Pattern Analysis, Cambridge University Press, 2004.

[14] Shepard R.N., Arabie P. Additive Clustering: Representation of Similarities as Combinations of Discrete Overlapping Properties, Psychological Review, vol. 86, no. 2, pp. 87-123, 1979.

[15] Tardiff, R.M., Topologies for Probabilistic Metric Spaces, Pacific Journal of Mathematics, vol. 65, no. 1, pp. 233-251, 1976.

[16] Tardiff, R.M, On a Functional Inequality Arising in the Construction of the Product of Several Metric Spaces, Aequationes Mathematicae, vol. 20, pp. 51-58, 1980.

[17] ter Braak, C.F.J., Y. Kourmpetis, H.A.L. Kiers, and M.C.A.M. Bink, Approximating a Similarity Matrix by a Latent Class Model: A Reappraisal of Additive Fuzzy Clustering, Computational Statistics and Data Analysis, vol. 53, pp. 3813-3193, 2009.

[18] ter Braak, C.F.J., M.P. Boer, L.R. Totir, C.R. Winkler, O.S. Smith, and M.C.A.M. Bink, Identity-by-Descent Matrix Decomposition Using Latent Ancestral Allele Models, Genetics, 2010.

[19] UCI Machine Learning Repository: http://archive.ics.uci.edu/ml/