Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT

Fuzzy Sets and Systems 135 (2003) 241–257

FUZZY sets and systems

www.elsevier.com/locate/fss

A fuzzy c-means variant for the generation of fuzzy term sets

T.W. Liao[a,*], Aivars K. Celmins[b], Robert J. Hammell II[b]

[a] *Army Research Laboratory, Attn: AMSRL-CI-CT, Building 321, Aberdeen Proving Ground, MD 21005-5067, USA*

[b] *Formerly Intelligent Systems Branch, Army Research Lab, Aberdeen Proving Ground, MD 21005, USA*

Received 5 January 2001; received in revised form 6 November 2001; accepted 15 November 2001

Abstract

A fuzzy c-means (FCM) variant is proposed for the generation of fuzzy term sets with $\frac{1}{2}$ overlap. The proposed variant differs from the original mainly in two areas. The first modification ensures that two end terms take the maximum and minimum domain values as their centers. The second modification prevents the generation of non-convex fuzzy terms that often occurs with the original algorithm. The optimal number of terms and the optimal shape of the membership function associated with each term are determined based on the mean squared error criterion. The exponential weight, m, used in the algorithm is found to greatly affect the shape of the membership function. The effect of data size used for the generation of fuzzy terms is also discussed. A generalized π-shaped function with a tunable parameter along with its complement is developed to fit all term sets generated by the FCM variant using various m values.

Keywords: Fuzzy clustering; Membership functions; Fuzzy term sets; Linguistic modeling

1. Introduction

Fuzzy set theory was proposed by Zadeh [28], in his words, "to provide a natural way of dealing with problems in which the source of imprecision is the absence of sharply defined criteria of class membership rather than the presence of random variables". In his classic paper, Prof. Zadeh stated that, "A fuzzy set A in X is characterized by a membership function $f_A(x)$ which associates with each point in X a real number in the interval [0,1], with the value of $f_A(x)$ at x representing the 'grade of membership' of x in A". The fuzzy set concept intends to capture the vagueness of meaning in words used in most of our daily communications to describe concepts, objects, events,

* Corresponding author. Department of Industrial and Manufacturing Systems Engineering, Louisiana State University, Baton Rouge, LA 70803, USA. Tel.: 1-410-2225-578-5365; fax: +1-410-2225-578-5109.

E-mail address: ieliao@lsu.edu (T.W. Liao).

PII: S0165-0114(02)00136-7

phenomena, or statements themselves. A linguistic variable is a variable that takes words used in our natural language as its values, with each word interpreted as a fuzzy set. Zimmermann [30] defined a linguistic variable as a quintuple $(y, T(y), X, G, M)$ in which y is the variable name; $T(y)$ denotes the term set of y, that is, the set of names of *linguistic values* of y with each value being a fuzzy set Y defined over a universe of discourse X associated with the basic variable x; G is a syntactic rule (which usually has the form of a grammar) for generating the name, Y, of variables of x; and M is a semantic rule for associating with each Y its meaning $M_Y(x)$, which is a fuzzy subset of X.

As pointed out by Dubois and Prade [6], there is no uniformity in the interpretation of what a membership grade means. In decision-making problems, $f_A(x)$ represents the degree of preference in favor of object x. In Zadeh's theory of approximate reasoning, $f_A(x)$ is the degree of possibility that a parameter y has value x, given that all that is known about it is that "y is A". In classification and data analysis applications, a membership grade $f_A(x)$ is interpreted as the degree of proximity of x to prototype elements of the fuzzy set A. This study follows the last interpretation because radiographic data are used to classify whether an object is a weld or not.

The issue of membership function generation is vital to the application of fuzzy set theory because the success of an application depends on the membership function used. Earlier works focused mostly on the determination of membership functions that reflect subjective perceptions about vague or imprecise concepts such as human height [17,19,8] and summarized by Turksen [24] under the framework of measurement theory. Recently, Medasani [18] provided a general overview of several methods for generating membership functions from domain data for fuzzy pattern recognition applications. They classified these methods into various categories including histogram-based methods, transformation of probability distributions to possibility distributions, fuzzy nearest neighbor techniques, neural-network-based methods and clustering methods. Subsequent studies appeared in the open literature include Chen and Wang [4], Runkler and Bezdek [22]. Both of them can be classified as clustering methods.

In their review, Medasani et al. [18] stated the pros and cons of using fuzzy c-means (FCM) as the clustering method. The pros are: (i) it can be used as an unsupervised algorithm; (2) it can be used to generate multi-dimensional membership functions; and (iii) the shape of the membership functions can be controlled by using a different distance measure. The cons are: (i) the number of classes must be provided to run the algorithm; and (ii) the algorithm is sensitive to outliers. However, they did not show any application. Chen and Wang [4] presented an enhanced FCM algorithm for initializing a fuzzy model, which incorporates a fuzzy validity function to find the optimal number of clusters, c, and a heuristic method to calibrate the fuzzy exponent, m. Runkler and Bezdek [22] developed an approach to obtain the LHS membership functions of a smooth first-order Takagi Sugeno system, given the points and the slopes of the RHS functions.

This study develops a modified FCM algorithm and applies it to generate fuzzy terms (or membership functions) for a three-feature data set obtained from an industrial application, viz., radiography of welds. The proposed variant differs from the original mainly in two areas. The first modification ensures that two end terms take the maximum and minimum domain values as their centers. The second modification prevents the generation of non-convex fuzzy terms that often occurs with the original algorithm. The fuzzy terms so defined are thus more reasonable. The optimal number of terms and the optimal shape of the membership function associated with each term are determined based on the mean squared error criterion. The effect of data size used for the generation of fuzzy terms is also investigated. In addition, a generalized π-shaped function with a tunable parameter along with its complement is developed to fit all term sets generated by the FCM variant using various m values.

The FCM variant is presented in the next section. The data used in this study are described in Section 3. The term set and membership functions generated are presented in Section 4. Section 5 gives more testing results. Section 6 presents a generalized s-shaped function along with its complement developed for fitting fuzzy terms generated by the modified FCM algorithm. Possible applications of the proposed methodology are discussed in Section 7, followed by the conclusion.

2. FCM and FCM variant

FCM is an unsupervised classification method, belonging to the partitional clustering category. It was derived from the hard (or crisp) c-means algorithm. The hard c-means and its variants [1] are based on the minimization of the sum of squared Euclidean distance between data (x_k, $k=1,\ldots,n$) and cluster centers (v_i, $i=1,\ldots,c$), which indirectly minimizes the variance as follows:

$$Min\, J_1(U,V)=\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^2\|x_k-v_i\|^2. \tag{1}$$

In the above equation, $\mathbf{U}=[u_{ik}]$ denotes the matrix of a hard c-partition and $\mathbf{V}=\{v_i\}$ denotes the vector of all cluster centers. The partition constraints in c-means are (1) $u_{ik}\in\{0,1\},\forall i,k$, (2) $\sum_{i=1,c} u_{ik}=1,\forall k$, and (3) $0<\sum_{k=1,n} u_{ik}<n$, $\forall i$. In other words, each x_k belongs to one and only one cluster.

Dunn first extended the hard c-means algorithm to allow for fuzzy partition with the objective function as given in Eq. (2) below [7]:

$$Min\, J_2(U,V)=\sum_{i=1}^{c}\sum_{k=1}^{n}(\mu_{ik})^2\|x_k-v_i\|^2. \tag{2}$$

Note that $\mathbf{U}=[\mu_{ik}]$ in this and following equations denotes the matrix of a fuzzy c-partition. The fuzzy c-partition constraints are (1) $\mu_{ik}\in[0,1]$, $\forall i,k$, (2) $\sum_{i=1,c}\mu_{ik}=1,\forall k$, and (3) $0<\sum_{k=1,n}\mu_{ik}<n$, $\forall i$. In other words, each x_k could belong to more than one cluster to a fractional degree between 0 and 1. Bezdek [2] generalized $J_2(\boldsymbol{U},\boldsymbol{V})$ to an infinite number of objective functions, i.e., $J_m(\boldsymbol{U},\boldsymbol{V})$, where $1\leqslant m\leqslant\propto$. The new objective function subject to the same fuzzy c-partition constraints is

$$Min\, J_m(U,V)=\sum_{i=1}^{c}\sum_{k=1}^{n}(\mu_{ik})^m\|x_k-v_i\|^2. \tag{3}$$

Note that both hard c-means and FCM algorithms try to minimize the variance of those data within each cluster. To solve the above model, an iterative procedure is required. Please refer to the original paper for the solution procedure.

In applying FCM, the number of clusters, c, must first be specified. Various validity measures have been proposed to determine the optimal number of clusters in order to address this inherent drawback of FCM [20,21,26,27,29]. Sugeno and Yasukawa [23] proposed another measure in their use of FCM to qualitative modeling. In this study, the optimal number of terms is defined as the one that has the lowest mean squared error (MSE). The least MSE measure is also used to identify the most appropriate form of membership functions, which cannot be done by using any previously proposed validity measures.

Another drawback of FCM as pointed out by Medasani et al. [18] is its sensitiveness to outliers. Our past uses of FCM in real world data such as weld identification [15], welding flaw detection [14], and part families/machine cells formation [11] repeatedly confirm that there are always some errors in the clustering results generated by FCM. Therefore, one cannot blindly trust the results given by FCM in any application. For a set of well-defined fuzzy terms, one would expect, in general, a domain value farther away from a term center to have a lower membership value than another one that is closer to the term center. In other words, the resultant membership function for each term should be convex. Nevertheless, our preliminary results indicate otherwise. In addition, one would expect that for the left end term a larger domain value always has a smaller membership value and that the converse is true for the right end term. Since the original FCM finds the center of the end clusters (or terms), the membership function of each end term obtained by FCM is counter-intuitive. To correct the above two problems, the original FCM was modified as follows:

A. Set the lowest domain value to be the center of the left end term and the highest domain value to be the center of the right end term.
B. Redistribute the concave part of a membership function to the other two more appropriate terms by a normalization operation.

In summary, the solution procedure of the modified FCM has the following steps:

(1) Choose c $(2 \leqslant c \leqslant n)$, m $(1 < m < \infty)$, and initialize the membership matrix.
(2) Read in the data set and find the maximum and minimum values.
(3) Calculate cluster centers using the equation in the original algorithm but force the two clusters with the largest and smallest values to take the maximum and minimum domain values.
(4) Update the membership matrix according to the equation in the original algorithm.
(5) Compute the change of each value in the membership matrix and determine whether the maximum change is smaller than the threshold value chosen to stop the iterative process (set at 0.01 throughout this study). If not, return to Step 3.
(6) Redistribute erroneous membership values to the other two more appropriate terms proportional to their current membership values.

The algorithm guarantees that each domain value has a summation of membership value of one. Therefore, a complete membership to a certain fuzzy term simultaneously exhibits a complete exclusion from all the remaining terms. Fuzzy terms so defined are normal and convex. The original and modified FCM algorithms were implemented in C language for this study. It should be noted that the FCM variant determines only the term centers. Because the term sets most likely are not uniformly spaced, the fuzzy terms are thus asymmetrical. The shape of each fuzzy term is not explicitly defined by the FCM variant. To determine the shape, the data points of (d, m) pairs, with d and m stand for domain value and membership value respectively, are fitted to some more commonly used functions such as triangular, s-shaped, and Gaussian. Based on the MSE value of each function, the best shape is then determined to be the one having the lowest value.

Mathematically, a triangular membership function, $TMF(x)$, is defined as

$$TMF(x) = \begin{cases} (x-\alpha)/(\gamma-\alpha), & \alpha \leqslant x \leqslant \gamma, \\ (x-\delta)/(\gamma-\delta), & \gamma \leqslant x \leqslant \delta, \end{cases} \tag{4}$$

where $\alpha \leqslant \gamma \leqslant \delta$. Two adjacent fuzzy terms with triangular membership functions that meet the above-mentioned assumptions should always have $\frac{1}{2}$ overlap at the midpoint between the two term centers. A Gaussian membership function, $GMF(x)$, is defined as

$$GMF(x) = \mathrm{e}^{-1/2(x-\mu/\sigma)^2}, \tag{5}$$

where μ and σ are the mean and the standard deviation, respectively. Two adjacent fuzzy terms defined as Gaussian membership functions that have the same standard deviation should always have 0.223 overlap at the midpoint between the two term centers. Therefore, the assumption that each domain value has a total membership value of one is violated. Because of this reason, one does not expect that Gaussian will fit well with the (d, m) pairs generated by the proposed algorithm. An s-shaped membership function, $SMF(x)$, is defined as

$$SMF(x) = \begin{cases} 0, & x \leqslant \alpha, \\ 2\left(\dfrac{x-\alpha}{\gamma-\alpha}\right)^2, & \alpha \leqslant x \leqslant \beta, \\ 1-2\left(\dfrac{x-\gamma}{\gamma-\alpha}\right)^2, & \beta \leqslant x \leqslant \gamma, \\ 1, & x \geqslant \gamma, \end{cases} \tag{6}$$

where $\alpha \leqslant \beta \leqslant \gamma$. The complement of a SMF is an z-shaped membership function (ZMF). An s-shaped fuzzy term and the complement of the adjacent s-shaped fuzzy term that meet the above-mentioned assumptions form a π function. An SMF and its complement always have 0.5 overlap at their crossover or the midpoint between the two term centers, i.e., $\beta = (\alpha+\gamma)/2$.

3. Data description

The data used in this study are part of those used in our previous studies [15,16]. To show the term set and membership functions generated, 600 tuples of such data were employed with each tuple comprising three features. The first feature is the width of the object measured along the x-direction in terms of number of pixels, an integer value. The second feature is the MSE error derived when the intensity profile of the object is compared with a Gaussian form. The third feature is the gray level (or intensity) of the object. Please refer to the above two references for a detailed description of the feature extraction process. The range of each feature values forms the universe of discourse for the set of fuzzy terms to be defined over it.

In addition, four more 600-tuple data sets were generated from a larger file with 2170 tuples using the random sampling method. The modified FCM procedure is then applied to each of these randomly generated sets to determine the best term set and the corresponding membership functions.

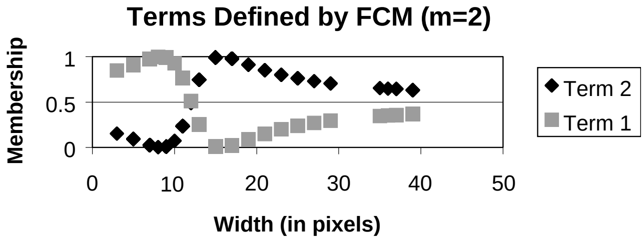

Fig. 1. Set of 2-term for Feature 1 defined by the original FCM.

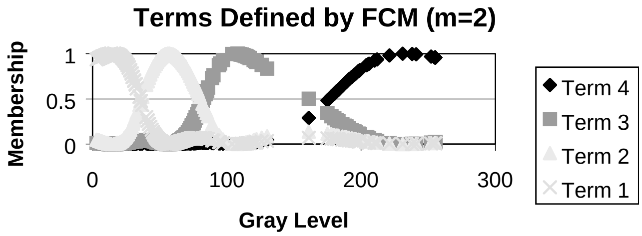

Fig. 2. Set of 4-term for Feature 3 defined by the original FCM.

Let F_{in}, F_{out}, t_i and n be the input file, output file, ith tuple, and number of randomly generated samples, respectively. Algorithm RS shows an implementation of simple random sampling.

Algorithm RS(F_{in}, n)

begin
  $F_{out} := \emptyset$;
  while $|F_{out}| \leqslant n$ **do**
    $i := random(1, |F_{in}|)$;
    $F_{out} := F_{out} \cup \{t_i\}$;
  enddo
  return (F_{out});
end;

Note that in this algorithm the sampling is done with replacement. That is, each tuple has the same chance at each draw regardless of whether it has already been sampled or not.

4. Fuzzy term set and associated membership functions

The original FCM with $m = 2$ was first applied to the data set feature by feature. Figs. 1 and 2 show the results when two and four fuzzy terms are specified for Features 1 and 3, respectively. Note that many terms generated are non-convex and all end terms do not give the membership value of one for the extreme domain values. Instead of giving semantic meaning, all terms are numbered

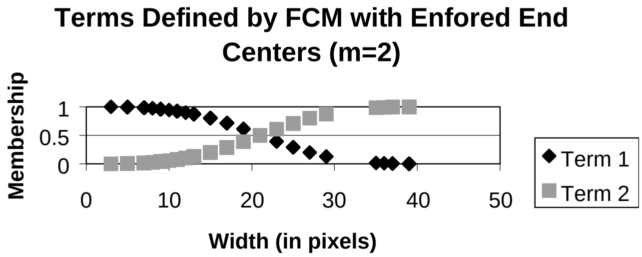

Fig. 3. Set of 2-term for Feature 1 by FCM with enforced and centers.

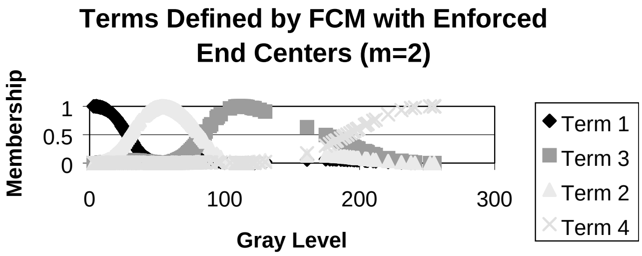

Fig. 4. Set of 4-term for Feature 3 by FCM with enforced and centers.

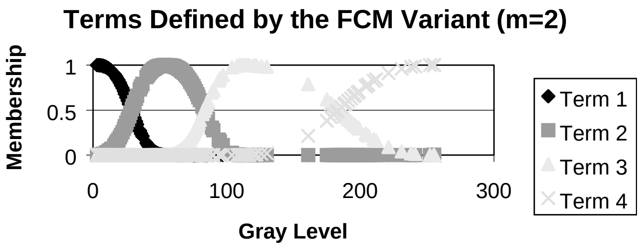

Fig. 5. Set of 4-term for Feature 3 defined by the modified FCM.

from left to right. Unfortunately, the algorithm does not necessarily number the cluster in the same manner. This is reflected in the figures shown throughout this paper.

The FCM with the first modification alone, i.e. setting the left and right term centers to be the minimum and maximum of the feature values, was applied next. The results are shown in Figs. 3 and 4. From these figures, it can be seen that: (1) the end terms now are more appropriate; (2) some fuzzy terms are still not convex. One example is the second term of Feature 3, in which some domain value farther from the center is assigned a higher membership value than another value that is closer to the center. This concave behavior is counter-intuitive.

The FCM with both modifications as described in Section 2 was then applied to Feature 3 data. Fig. 5 shows the results. Note that non-convex fuzzy terms no longer exist. Feature 1 data was not applied in this step because both terms are already convex after the first modification.

The following discussion considers the goodness of fit when all the (d,m) pairs generated by the FCM variant were fitted with three different functions: triangular, Gaussian, and π-shaped. To

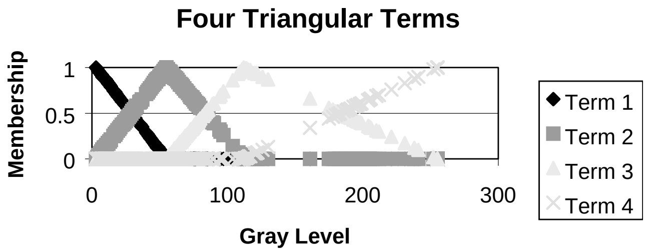

Fig. 6. Set of four triangular terms for Feature 3.

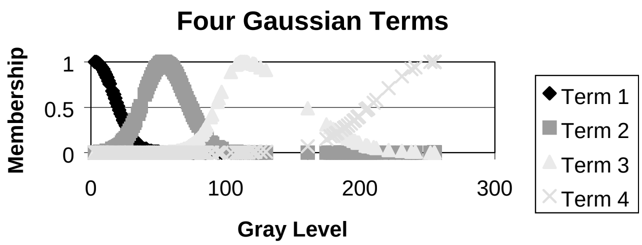

Fig. 7. Set of four asymmetric Gaussian terms for Feature 3.

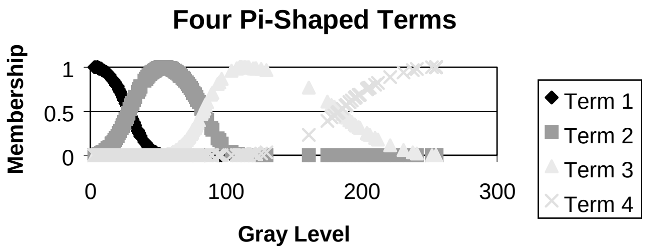

Fig. 8. Set of four π-shaped terms for Feature 3.

save space, only Feature 3 data is shown. Figs. 6–8 show the fitted triangular, Gaussian, and π-shaped functions, respectively. Note that most terms are asymmetric and the overlap of two adjacent Gaussian terms is not $\frac{1}{2}$, as mentioned in Section 3.

Using MSE as the criterion, it was determined that π-shaped functions consistently give smaller values than the other two functions. Triangular functions give either equivalent or better fit than Gaussian functions. The poor performance of Gaussian functions is expected because of the lower crossover points between two adjacent terms. π-shaped functions fit better than triangular functions because of their non-linearity, which happens to fit the subject data better when $m=2$ is used to run the modified FCM algorithm. From Figs. 9–11, one can tell that for the data set used here the best sets of terms are: two terms of π-shaped function for width, two terms of π-shaped function for MSE of intensity profile, and four terms of π-shaped function for gray level. This result will not hold when a different m value is used. More details are given in the next section.

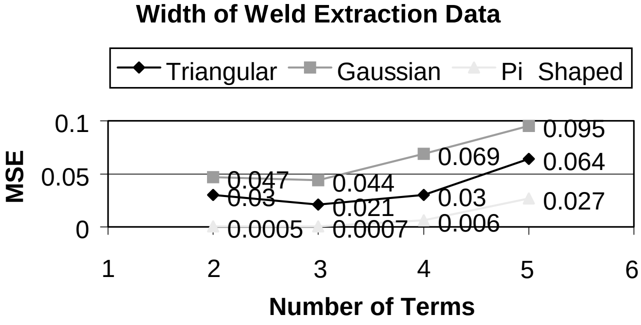

Fig. 9. MSE vs. number of terms for three different functions for Feature 1.

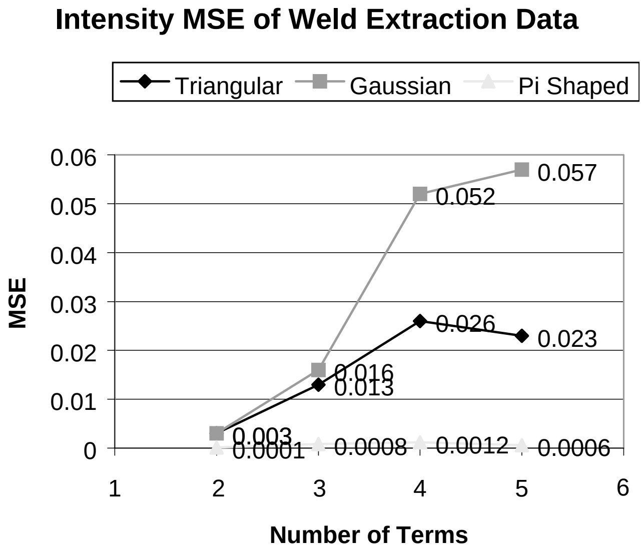

Fig. 10. MSE vs. number of terms for three different functions for Feature 2.

5. Testing results and discussion

Table 1 summarizes the mean squared values of fitting the (d, m) pairs generated by the FCM variant $(m = 2)$ with three different membership functions feature by feature for each one of the four randomly sampled 600-tuple data sets. Note that these four data sets are all different from the one used in Section 4. The π-shaped functions consistently give the lowest MSE values (in italic). For the first feature (width), two terms emerge to be the best 50% of the time $(\frac{2}{4})$. One out of four times (25%), either three terms or four terms are better. Based on these results, it seems quite appropriate to conclude that two terms are the best for the first feature. This is even so if the result of Section 4 is also considered. For the second feature (MSE of intensity profile), two terms emerge to be the best 100% of the time. This is identical to the result obtained in Section 4. For the third feature (gray level), there is a tie between three and four terms, with each giving the lowest MSE value

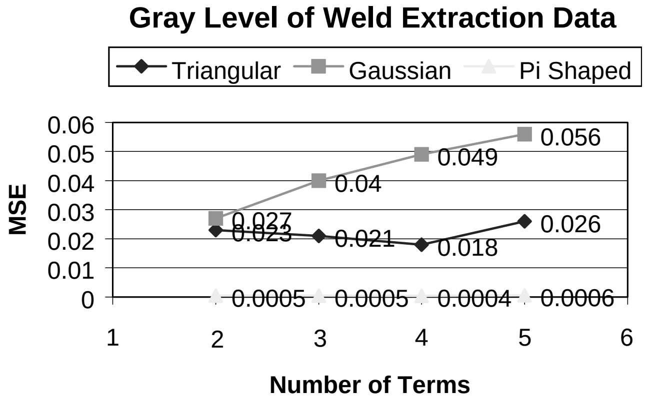

Fig. 11. MSE vs. number of terms for three functions for Feature 3.

50% of the time. However, four terms become better than three if the result obtained in Section 4 is used as the tiebreaker.

The exponential weight, m, in the FCM algorithm reduces the influence of "noise" (points further away from the centers) when computing the cluster centers (term centers here). All the results presented up to this point were obtained by setting $m=2$. To investigate the effect of m, three additional m values (3, 5 and 10) were used to rerun the data set used in Section 4. The selection of these three values intends to capture the wide range of effect. It will become clear later that the effect is actually gradual and by simply comparing these results one can see the transition. Figs. 12–14 show the set of two fuzzy terms for Feature 2 when the modified FCM algorithm was applied to the 600-tuple data set used in Section 4 with $m=3$, $m=5$, and $m=10$, respectively. Comparing these figures, it can be observed that:

(1) The tendency for membership values to be pull toward the $\frac{1}{2}$ value becomes stronger as m increases.
(2) For $m=3$, the fuzzy terms look very much like triangular.
(3) For $m<3$, the membership values of domain values above (below) the crossover point take a convex (concave) shape. The opposite is true when $m>3$.

The above observations are not unique to Feature 2 with two fuzzy terms. They are also observed in other features and different number of fuzzy terms. However, they are not shown to save space. The same phenomena were observed when a feature from another set of data was used to generate term sets using the proposed algorithm with different m values. It is appropriate to note that Liao et al. [15] reported the first observation in another study. This behavior is expected as a result of Eq. (3).

The above results imply that triangular shape will fit better if $m=3$ is used. To verify this observation, the FCM variant was applied to the 600-tuple data set used in Section 4 to define two terms for Feature 1, two terms for Feature 2, and four terms for Feature 3. The membership values were subsequently fitted with triangular, Gaussian, and π-shaped functions and the corresponding MSE values were calculated. Table 2 summarizes the MSE values of each type of membership function when different m values were used. These results confirm that triangular functions fit the best when $m=3$. Among all three membership functions tested, triangular functions also gave the best fit when $m>3$ was used. However, the MSE value increases with m, which indicates that

Table 1
Testing results of MSE

Data set	Feature	Membership function	2 terms	3 terms	4 terms	5 terms
1	1	Triangular	0.02510	0.02322	0.00910	0.08242
		Gaussian	0.02692	0.04011	0.05101	0.12863
		s-Shaped	0.00058	0.00313	***0.00039***	0.06452
	2	Triangular	0.00270	0.02153	0.02426	0.02286
		Gaussian	0.00143	0.04505	0.04386	0.04798
		s-Shaped	***0.00003***	0.00036	0.00062	0.00048
	3	Triangular	0.02104	0.02027	0.01836	0.01780
		Gaussian	0.02391	0.04155	0.03888	0.04892
		s-Shaped	0.00046	0.00055	***0.00036***	0.00062
2	1	Triangular	0.02997	0.02240	0.03017	0.08710
		Gaussian	0.04654	0.03211	0.06207	0.13884
		s-Shaped	0.00069	***0.00059***	0.00325	0.06837
	2	Triangular	0.00289	0.01823	0.02605	0.02448
		Gaussian	0.01776	0.02300	0.05072	0.05716
		s-Shaped	***0.00004***	0.00035	0.00068	0.00072
	3	Triangular	0.02540	0.02091	0.01820	0.01966
		Gaussian	0.02552	0.03673	0.04952	0.04640
		s-Shaped	0.00060	***0.00037***	0.04640	0.00040
3	1	Triangular	0.02673	0.02501	0.0481	0.07769
		Gaussian	0.02901	0.04514	0.08562	0.11051
		s-Shaped	***0.00062***	0.00381	0.0307	0.07769
	2	Triangular	0.00246	0.02264	0.02402	0.02494
		Gaussian	0.00095	0.04442	0.04919	0.06222
		s-Shaped	***0.00003***	0.00114	0.00046	0.02494
	3	Triangular	0.02251	0.02109	0.01965	0.01961
		Gaussian	0.02395	0.03769	0.05066	0.04645
		s-Shaped	0.00051	***0.00049***	0.00122	0.01961
4	1	Triangular	0.02589	0.01469	0.03531	0.04363
		Gaussian	0.02878	0.02996	0.06838	0.08502
		s-Shaped	***0.00061***	0.00071	0.08502	0.02098
	2	Triangular	0.00296	0.01651	0.02619	0.02523
		Gaussian	0.00244	0.02702	0.04466	0.05831
		s-Shaped	***0.00004***	0.00050	0.00081	0.00081
	3	Triangular	0.02400	0.02142	0.02122	0.01911
		Gaussian	0.02836	0.04257	0.03638	0.04698
		s-Shaped	0.00053	0.00058	***0.00047***	0.00073

this result cannot be generalized because a better function might exist but was not tested. The generalized π-shaped function to be discussed in Section 6 could be one of such functions, given that an appropriate parameter is chosen.

It is known that FCM does not always give the number of clusters specified. For instance, FCM might give only three clusters when four is asked. This unwanted result could be caused by the lack

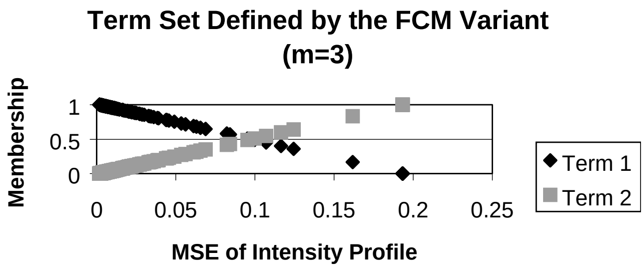

Fig. 12. Terms defined by the modified FCM when $m = 3$ was used.

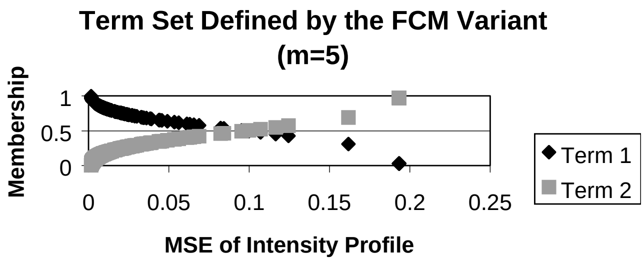

Fig. 13. Terms defined by the modified FCM when $m = 5$ was used.

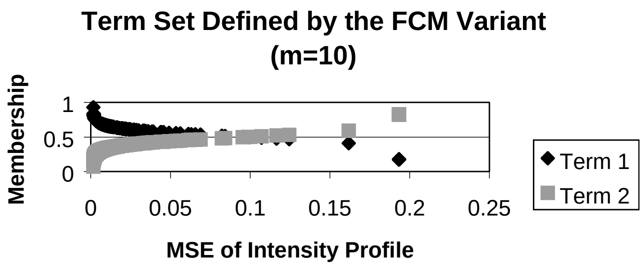

Fig. 14. Terms defined by the modified FCM when $m = 10$ was used.

of data and/or their poor distribution. Feature 3 of the 600-tuple data set used in Section 4 was chosen to investigate the effect of data size. The random sampling method was used to sample only 25 out of the 600 tuples. Fig. 15 shows that four terms could still be formed with 25 tuples only. However, the number of data points for some terms (Term 4 in particular) is very few. If the number of tuples is further reduced, four terms might not be possible. Figs. 16 and 17 show the frequency distribution of 25- and 600-tuple data sets, respectively. The number of modes in the distribution seems to be consistent with the number of terms. The frequency of data also reflects where terms will be dense or sparse. For instance, the high frequency of lower gray level value (Feature 3) leads to more cluttered terms on the left side of the universe. Understandably, the term centers move as the number of data points and the composition of data sets change. As an example, Table 3 gives the change in four term centers generated for Feature 3 when different sizes of data set were used.

Table 2
MSE values for each membership function when different m values were used

Feature	No. of terms	Membership function	$m=2$	$m=3$	$m=5$	$m=10$
Width	2	Triangular	0.03000	***0.00002***	***0.03131***	***0.13654***
		Gaussian	0.04700	0.04852	0.10656	0.19043
		s-Shaped	***0.00050***	0.02179	0.09524	0.18910
MSE of intensity profile	2	Triangular	0.00300	***0.00002***	***0.02102***	***0.14289***
		Gaussian	0.00300	0.00448	0.03577	0.16359
		s-Shaped	***0.00010***	0.00266	0.03565	0.16502
Gray level	4	Triangular	0.01800	***0.00050***	***0.02331***	***0.08634***
		Gaussian	0.04900	0.05376	0.09763	0.17936
		s-Shaped	***0.00040***	0.01322	0.06735	0.15692

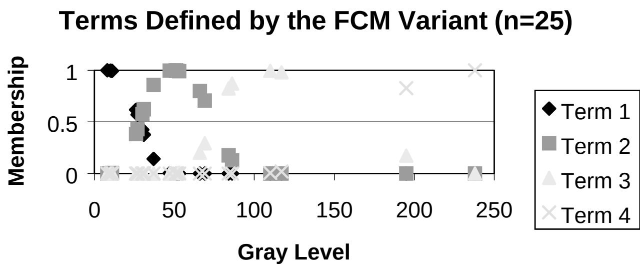

Fig. 15. Terms defined by the FCM variant for Feature 3 with 25 data points.

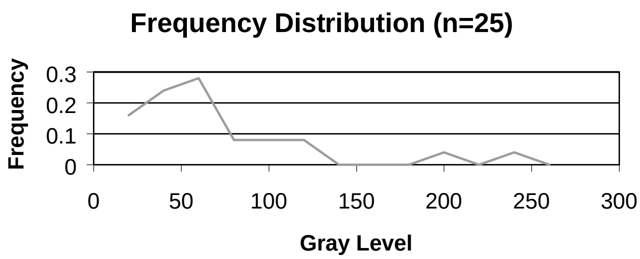

Fig. 16. Frequency distribution of 25 Feature-3 data points.

6. Generalized π-shaped functions

This section proposes a generalized π-shaped function that includes triangular functions and π-shaped functions as special cases. The parameter of the generalized π-shaped function can be adjusted to fit the terms generated by the modified FCM algorithm when $m > 3$.

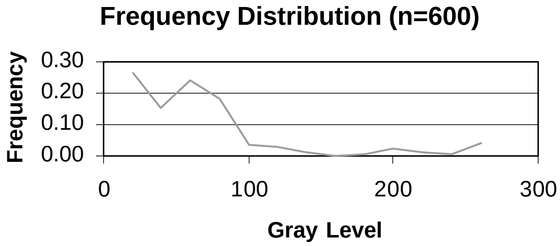

Fig. 17. Frequency distribution of 600 Feature-3 data points.

Table 3
Change in term centers for Feature 3 when different sizes of data set were used

Term center	Data size				
	25	75	150	300	600
1	8	5	3	3	3
2	48.9	44.6	46.7	47.3	54.5
3	100.3	75.5	80.7	80.3	111.9
4	238.2	255	255	255	255

The mathematical form of the generalized s-shaped membership function is given below.

$$GSMF(x) = \begin{cases} 0, & x \leqslant \alpha, \\ \left(\frac{1}{2}\right)^{1-z}\left(\frac{x-\alpha}{\gamma-\alpha}\right)^{z}, & \alpha \leqslant x \leqslant \beta, \\ 1-\left(\frac{1}{2}\right)^{1-z}\left(\frac{x-\gamma}{\gamma-\alpha}\right)^{z}, & \beta \leqslant x \leqslant \gamma, \\ 1, & x \geqslant \gamma. \end{cases} \tag{7}$$

For each middle term, Eq. (7) can represent only the left-hand side of the generalized π-shaped function. For the right-hand side, the complement of Eq. (7) or a generalized z-shaped function is needed, which takes the following mathematical form:

$$GSMF^{c}(x) = \begin{cases} 1, & x \leqslant \alpha, \\ 1-\left(\frac{1}{2}\right)^{1-z}\left(\frac{x-\alpha}{\gamma-\alpha}\right)^{z}, & \alpha \leqslant x \leqslant \beta, \\ \left(\frac{1}{2}\right)^{1-z}\left(\frac{x-\gamma}{\gamma-\alpha}\right)^{z}, & \beta \leqslant x \leqslant \gamma, \\ 0, & x \geqslant \gamma. \end{cases} \tag{8}$$

The generalized s-shaped function in Eq. (7) reduces to the left-hand side of a triangular function when $z=1$. When $z=2$, Eq. (7) essentially becomes Eq. (6), which is the s-shaped function. When $0<z<1$, the generalized π-shaped function can be used to fit terms generated by the FCM variant when $m>3$. Fig. 18 shows the shape change of the generalized s-shaped membership function for

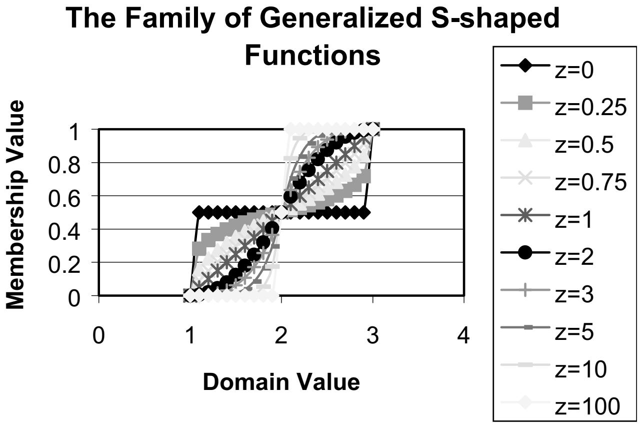

Fig. 18. Generalized s-Shaped membership functions with various z values.

a range of z values. All of the above discussions also hold for Eq. (8). For a given set of data, the optimal value of z must be found that best fits the term set generated for the data by the modified FCM algorithm.

7. Potential applications

It has been shown that the proposed algorithm can be used to generate an optimal set of fuzzy terms all at one time given a set of domain values. Different from the conventional uses of FCM, the input data set for the generation of fuzzy term set here is one-dimensional. For a multi-dimensional data set, the proposed algorithm can be repeatedly applied one dimension at a time. This was shown with the three-dimensional data set throughout this paper.

The development of the FCM variant for the generation of fuzzy term sets was part of our research effort to extend the basic framework of case-based reasoning in order to handle fuzzy attributes. Once the term set is generated, it can be used to fuzzify a numeric value into a fuzzy value with some degree of belonging. Jeng and Liang [10] have shown this idea of fuzzy retrieval in case-based systems, except they defined the term set in a heuristic way. The fuzzified data can subsequently be used to induce fuzzy classification rules. Many fuzzy rule induction algorithms take such data as their input. The algorithms developed by Castro et al. [3], Chen and Yeh [5], Hong and Chen [9], and Wang et al. [25] are a few examples. These induced rules often are later revised by some adaptive strategies to account for the new data collected in the system. This feature is particularly important to fuzzy control and classification systems.

Our method can also be used to extract fuzzy rules, but is limited to the problem domains with relatively few input features. The reason is that the antecedents of a rule extracted by our method are necessarily the Cartesian products of fuzzy terms defined for the input features. As the number of input features and/or the number of terms increase, the number of rules explodes. A more detailed discussion on the effectiveness of the rules extracted based on the proposed method is given in a follow up paper (Liao, to appear).

8. Conclusion

This paper has presented a fuzzy term set generation algorithm. The algorithm can be considered a variant of fuzzy c-means, which differs from the original fuzzy c-means with several modifications. Data sets taken from the domain of radiography-based weld inspection were used to illustrate the effectiveness of the proposed algorithm.

The test results show that the exponential weight, m, has a strong effect on the shape of fuzzy terms defined by the algorithm. Among three different membership functions fitted, π-shaped functions fit better when $m=2$ was used. Triangular functions become the best fit when $m \geqslant 3$. The fitness of function is determined based on the mean squared error criterion. It was also shown that the optimal number of fuzzy terms slightly varies when a different data set was used. As the data size reduces, the term centers move and there might not be sufficient number of data points to define the desired number of terms. Gaussian functions do not fit well because of the inherent lower overlap than the (d, m) pairs generated by the proposed algorithm.

Finally, the generalized π-shaped function with an adjustable parameter, z, was proposed to fit all fuzzy terms generated by the proposed FCM variant when various exponential weight values were used. Potential applications of the proposed method were also discussed.

Acknowledgements

The authors acknowledge the support provided by the 2001 ASEE-ARL Postdoctoral Fellowship Program for this work (DAAL01-96-C-0038).

References

[1] G.H. Ball, D.J. Hall, ISODATA, an iterative method of multivariate analysis and pattern recognition, Behav. Sci. 12 (1967) 153–155.

[2] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York and London, 1987.

[3] J.L. Castro, J.J. Castro-Schez, J.M. Zurita, Learning maximal structure rules in fuzzy logic for knowledge acquisition in expert systems, Fuzzy Sets and Systems 101 (1999) 331–342.

[4] M.-S. Chen, S.-W. Wang, Fuzzy clustering analysis for optimizing membership functions, Fuzzy Sets and Systems 103 (1999) 239–254.

[5] S.-M. Chen, M.-S. Yeh, Generating fuzzy rules from relational database systems for estimating null values, Cybernet. Systems 28 (1997) 695–723.

[6] D. Dubois, H. Prade, The three semantics of fuzzy sets, Fuzzy Sets and Systems 90 (1997) 141–150.

[7] J.C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, J. Cybernet. 3 (1974) 32–57.

[8] E. Hisdal, Are grades of membership probabilities? Fuzzy Sets and Systems 25 (1988) 325–348.

[9] T.-P. Hong, J.-B. Chen, Processing individual fuzzy attributes for fuzzy rule induction, Fuzzy Sets and Systems 112 (2000) 127–140.

[10] B.C. Jeng, T.-P. Liang, Fuzzy indexing and retrieval in case-based systems, Expert Systems Appl. 8 (1) (1995) 135–142.

[11] K. Josien, T.W. Liao, Integrated use of fuzzy c-means and fuzzy KNN for GT part family and machine cell formation, Internat. J. Prod. Res. 38 (15) (2000) 3513–3536.

[12] K. Josien, G. Wang, T.W. Liao, E. Triantaphyllou, M.C. Liu, An evaluation of sampling methods to data mining with fuzzy c-means, chapter 15, in: Dan Braha (Ed.), Data Mining for Design and Manufacturing, Kluwer Academic Publishers, New York, 2001.
[13] T.W. Liao, Fuzzy reasoning based automatic inspection of radiographic welds: part I, weld recognition, J. Intell. Manuf., to appear.
[14] T.W. Liao, D.-M. Li, Y.-M. Li, Detection of welding flaws from radiographic images with fuzzy clustering methods, Fuzzy Sets and Systems 108 (2) (1999) 145–158.
[15] T.W. Liao, D.-M. Li, Y.-M. Li, Extraction of welds from radiographic images using fuzzy classifiers, Inform. Sci. 126 (2000) 21–42.
[16] T.W. Liao, K. Tang, Automated extraction of welds from digitized radiographic images based on MLP neural networks, Appl. Artif. Intell. 11 (3) (1997) 197–218.
[17] P.J. MacVicar-Whelan, Fuzzy sets, the concept of height, and the hedge VERY, IEEE Trans. Systems Man Cybernet. 8 (6) (1978) 507–511.
[18] S. Medasani, J. Kim, R. Krishnapuram, An overview of membership function generation techniques for pattern recognition, Internat. J. Approx. Reas. 19 (1998) 391–417.
[19] A.M. Norwich, I.B. Turksen, A model for the measurement of membership and the consequences of its empirical implementation, Fuzzy Sets and Systems 12 (1984) 1–25.
[20] N.R. Pal, J.C. Bezdek, On cluster validity for the fuzzy c-means model, IEEE Trans. Fuzzy Systems 3 (3) (1995) 370–379.
[21] M.R. Rezaee, B.P.F. Lelieveldt, J.H.C. Reiber, A new cluster validity index for the fuzzy c-means, Pattern Recognition Lett. 19 (1998) 237–246.
[22] T.A. Runkler, J.C. Bezdek, Function approximation with polynomial membership functions and alternating cluster estimation, Fuzzy Sets and Systems 101 (1999) 207–218.
[23] M. Sugeno, T. Yasukawa, A fuzzy logic based approach to qualitative modeling, IEEE Trans. Fuzzy Systems 1 (1) (1993) 7–31.
[24] I.B. Turksen, Measurement of membership functions and their acquisition, Fuzzy Sets and Systems 40 (1991) 5–38.
[25] X. Wang, B. Chen, G. Qian, F. Ye, On the optimization of fuzzy decision trees, Fuzzy Sets and Systems 112 (2000) 117–125.
[26] M.P. Winham, Cluster validity for the fuzzy c-means clustering algorithm, IEEE Trans. Pattern Anal. Mach. Intell. 4 (4) (1982) 357–362.
[27] X.L. Xie, G. Beni, A validity measure for fuzzy clustering, IEEE Trans. Pattern Anal. Mach. Intell. 13 (8) (1991) 841–847.
[28] L.A. Zadeh, Fuzzy sets, Inform. Control 8 (1965) 338–353.
[29] N. Zahid, M. Limouri, A. Essaid, A new cluster validity for fuzzy clustering, Pattern Recognition 32 (1999) 1089–1097.
[30] H.-J. Zimmermann, Fuzzy Set Theory and its Applications, Kluwer Academic Publishers, Boston, 1996.