International Conference on e-Business 2007 (INCEB 2007)
22 - 23 November 2007
P19

A fuzzy product clustering with flexible linguistic quantifier

Bunthit Watanapa
School of Information Technology,
King Mongkut's University of Technology Thonburi
Bangkok, Thailand

Email: bunthit@sit.kmutt.ac.th

Abstract

This paper extends the fuzzy-approach clustering method for hierarchically grouping products based on personal preference on selected attributes. The linguistic quantifier is defined as the parameter-adjustable non-linear function which yields flexible discriminating power of the clustering. Given a personalized preference on attributes, all available products on the database can be clustered into hierarchical levels of product clusters. Numerical examples are also given to illustrate the computation involved, to provide comparative results to the existing method, and give some technical insights.

Key Words: Fuzzy, Clustering, Classification, Linguistic quantifier, Personalization.

1. Introduction

E-Commerce has spread out worldwide and become a norm nowadays. On the one hand, this enables market penetration and selling opportunity; on the other hand, the overwhelmingly substitutable products deteriorate the effective product selection of customers. A solution to this effect is to classify the products of interest into different levels according to individual customer's preference. This can be viewed as a personalization approach that supports customers or merchandisers to systematically filter and organize prospective products into hierarchical level. One can expect an improvement on sales, throughput and customer satisfaction from the implementation of product e-catalogue according to the hierarchically prioritized level.

The hierarchical level of preference will serve as a decision-making aid to the customer in a way that he/she can realize the comparative satisfaction level of the chosen product to the other available ones in the system. Having a reference to the best satisfaction (top level of hierarchy), customer can also perceive how much he/she has to compromise if having the chosen product.

One of the main difficulties for this is the ambiguity when measuring the preference in customer's mind. Fuzzy logic is a concept that has been proven to be an effective tool for quantifying such the fuzziness into computable numbers. Once, having a proper set of product's attributes, we can determine fuzzy value for each attribute that will collectively cluster the products. There have been many attempts to address this e-shopping problems, however, most of them propose a static solution of generic product hierarchy. Recently, Mohanty and Bhasker [1] proposed the idea of product classification based on Fuzzy logic with a piece-wise linearly fixed linguistic quantifier to articulate the attribute values. This gives into a non-constraint hierarchical product echelon of customer preference.

This paper proposes an improved method for fuzzy-based product clustering by introducing a nonlinear function of linguistic quantifier to dynamically control the discriminating power of the system. This should yield higher dynamicity to customers/merchandiser in reclassifying the product on the fly, e.g. when there are too many products in a particular class (very similar in terms of overall attribute values).

In Section 2, a brief overview for existing works is provided. Section 3 explains how to quantify the attribute's value based on fuzzy logic and the modeling of the proposed methods. Section 4 illustrates numerical examples to give insights into comparative results of the proposed model to the base method. The conclusion of paper and some suggestions for future research are given in Section 5.

2. Review of the existing works

There are some available web sites that support customers making internet shopping, for example, www.dealtime.com, www.Pricescan.com, and www.activebuyersguide.com.

PriceScan.com offers itself as an internet agent that can search for a low price on a specified product or for a specified supplier. There is also pricewatch.com that works in a similar fashion with PriceScan.com in which the price is the only main attribute to classify products of interest. Dealtime.com also acts as an internet agent which basically collect not only price but also some other desirable product features from the internet for a specified product.

ActiveBuyersGuide.com can be viewed as the next level of agents that can assist customers in selecting a suitable product based on the attributes provided by the buyers. However, customer cannot easily identify his/her desire level of satisfactions of the product features. Only the importance of the attributes or the range of minimum and maximum levels of attributes can be determined by the users

The more natural approach of fuzzy logic is proposed by Mohanty and Bhasker [1] which can overcome the limitation of the previous approach. The main idea is to allow products with slight deviation from the domain but offer a better overall desirability to the customers to be on the candidate list. This reflects the customer's compromising attitude implicitly presented in most of the customer's mind.

Prior to Mohanty and Bhasker [1], Ryu [3] and Lee and Widmeyer [5] proposed similar idea of classifying the products based on attributes specified by customer. If the desired product according to the prescribed attributes is available on the internet, the customer obtains the product. If not available, customer is offered the next product which is closest to the targeted attribute values. This behavior of the system is defined as flexibility value which is chosen arbitrarily by customers [3]. In [5], the e-shopping program concept is introduced. The program can flexibly suggest alternative products that are closest to the requested product in the taxonomy hierarchy. However, the approach is limited to the search which is performed in a single generic hierarchy.

Recently, Mohanty and Passy [4] proposed to further manipulate the "most" quantifier as a tool for ranking products available on the internet market. The customer's own preference on product features and the information of other customers' opinion extracted from search engines are combined to determine the preference ranking of all products on the e-business site.

3. Method to cluster the products

To cluster the available products into different hierarchical levels, three main computations are required in this method. The first is to quantify the perceived value of an attribute in customers' mind into fuzzy membership value [2]. The second is to measure the extent to which an attribute value is satisfied by all available products. Finally, the third is to compute the level of satisfaction achieved by a particular product compared with the overall as obtained in the second computation. These are described in the following three Subsections.

3.1 Quantifying attribute values

This step is to transform the imprecisely perceived value of each attribute perceived by a customer into a quantifiable value according to the traditional fuzzy concept [6, 7]. For example, one may consider that a car with an average yearly maintenance cost at 5% of its price is the most satisfactory. Cars with higher or lower maintenance costs than this one are lower in satisfaction level may because it may be too expensive (unreasonable and unjustifiable cost) or too cheap (risk of improper maintenance). Given the membership profile below, the satisfaction level is fullest, or $\mu_{maintenance} = 1$, when the maintenance cost is at 5% of its original price. Zero satisfaction incurs when it is at 1% (or lower) or at 13% (or higher): $\mu_{Maintenance} \le 1\%$ or $\mu_{Maintenance} \ge 13\%$. See Figure 1 below.

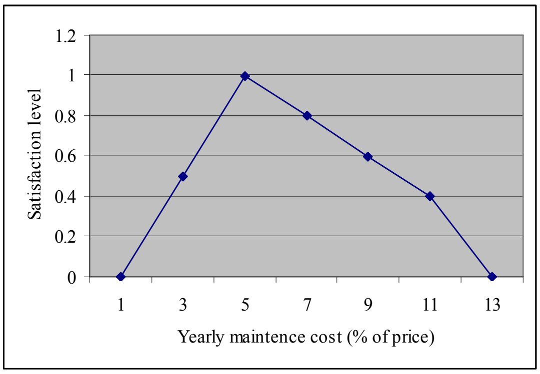

Fig. 1 Satisfaction level as member value for yearly maintenance cost.

Usually, the attribute values are conflicting with each other. At times, they are imprecise (e.g. how comfortable is a car) and non-commensurable (e.g. how to compare value of price to value of the comfort level). Fuzzy can help to address such problems and offer a way to compromise these conflicts too.

Provided there are N products $\{P_1, P_2, ..., P_N\}$ available in the database for e-merchandizing, each has M attributes, then we may measure the average level of satisfaction to customer for attribute m by all the products, or S_m, as follows [5]:

$$S_m = \frac{1}{N}\sum_{n=1}^{N}\mu_{mn} \qquad (1)$$

where μ_{mn} is the membership value of the fuzzy set mapped with the customer's perceived value on the n^{th} product for the m^{th} attribute.

3.2 Evaluate the satisfaction level

Instead of addressing the fully simultaneous satisfaction to a set of selected attributes as performed in the traditional decision science, this approach yields a more realistically flexible and adaptive solution. Users may choose a quantifier representing their desire, e.g. "most" to stands for the requirement that most of the attributes are satisfied. This natural representation of requirement enables us to articulate how far the customers have to sacrifice his/her absolute satisfaction per each attributes to keep flexibility of choices.

While the linguistic quantifier "most", as defined in the works of Kacprzyk and Yager [2], Mohanty and Bhasker [1], Ryu [3] and Hohanty and Passi [4], is a crisply piece-wise linear function, we propose to adopt a nonlinear "most" function which renders a finer model of value perception in customers' mind. Generally, this can be presumed as any continuous function, e.g. exponential, lognormal, etc. In this paper, we adopt the logistic-alike curve to represent the S-shape distribution of the perception. The perception of "most" value is acceleratedly improved in the beginning and then deceleratedly improved at the same scale when approaching the fully satisfaction level.

Given S_m in (1) that evaluates the level of satisfaction for attribute m over the entire products on the database by a customer, the most function, says $\mu_{most}(S_m)$, is defined as:

$$\mu_{most}(S_m) = \frac{1}{1+\alpha\cdot e^{-\beta(2S_m-1)}} \qquad (2)$$

where α is the control parameter for the acceleration of increasing μ_{most} value reaching the fullest satisfaction or decreasing value reaching the zero satisfaction. When $\alpha > 1$, the decelerating rate is higher than the accelerating rate and when $\alpha < 1$, the accelerating rate is higher than the decelerating rate. Finally, when $\alpha = 1$, the accelerating rate equals the decelerating rate. β is used for controlling the asymptoticity of the curve. Namely, the higher the beta is, the closer to the value of 1 and 0 the μ_{most} of 1 and 0 are, respectively. Figure 2a-d shows the effects of varying α and β as described above.

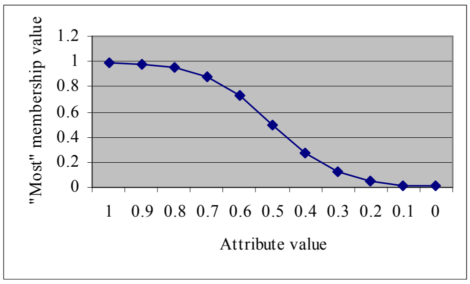

Fig 2.a μ_{most} with $\alpha = 1$, $\beta = 5$

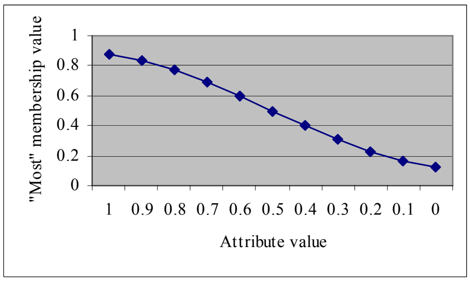

Fig 2.b μ_{most} with $\alpha = 1$, $\beta = 2$

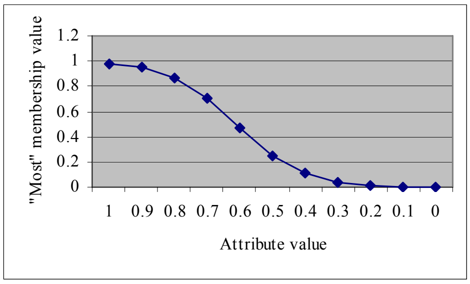

Fig 2.c μ_{most} with $\alpha = 3$, $\beta = 5$

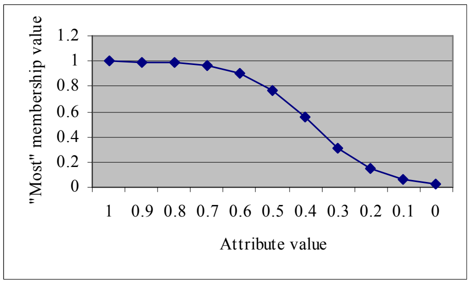

Fig 2.d μ_{most} with $\alpha = 0.3,\ \beta = 5$

3.3 Flexibility range

Based on the degree of satisfaction of attribute m over most of the available product as stated in (2), we can determine the extent to which a customer has to compromise the desired level (d_m) for the m^{th} attribute, denoted as f_m:

$$f_m = \begin{cases} 0 & d_m \le \mu_{most}(S_m) \\ d_m - \mu_{most}(S_m) & d_m > \mu_{most}(S_m) \end{cases} \tag{3}$$

If a customer requires full satisfaction on attribute m, or d_{most} = 1, then the flexibility range, f_m, equals 1-$\mu_{most}(S_{.m})$. Similarly, f_m is zero if the desired level of customer is lower than the satisfaction level of most of the available products. The flexibility range will enable us to hierarchically cluster products by the level of sacrificing or compromising the satisfaction level as described in Section 3.4.

Applying (3) to the decision system, the flexibility value of each product as per the m^{th} attribute is as follows:

$$f_{mn} = \begin{cases} 0 & ,\ \mu_{mn} \le \mu_{most}(S_m) \\ \mu_{mn} - \mu_{most}(S_m), & \mu_{mn} > \mu_{most}(S_m) \end{cases} \tag{4}$$

Where μ_{mn} is the membership value of
Based on the operational rule (4), we obtain discriminating procedure for clustering all available products into hierarchical groups as described in the next section.

3.4 Clustering method

Given a set of products $P = \{P_1, P_2, ..., P_N\}$ and a set of corresponding attributes $m = \{m_1, m_2, ..., m_M\}$ which is ranked by customer's perceived importance level, the process of clustering can be explained as follows:

Step 1: Considering the most important attribute m_1, obtain fuzzy membership value for each product(μ_{mn}).

Step 2: For each product, evaluate f_{mn} value as defined in (4).

Step 3.1: Select products from the set P into set C_{im}:

$$\{C_{im} \subseteq P \,|\, f_m + f_{mn} > f_m\};\ \text{for } n=1,\ldots,N$$

Step 3.2: If step 3.1 gives $C_{im} = \varnothing$, set $C_{im} = \{P\}$ due to its indiscrimination.

Step 4: Repeat Steps 1 to 3 for the rest of attributes (2,…,M) in the order of importance level so that the set of the most preferred products, C_{iM} is obtained.

Step 5: After getting the 1st cluster of products, C_{1M}, repeat Steps 1 to 4 for the remaining products , $P = P - C_{1M}$, so that other clusters, C_{iM}, can be obtained.

4. Numerical examples

To gain insights into the effect of the proposed method in comparison to that of Mohanty and Bhasker [1], the following examples are based on the same set of data (requirement for a car) as shown in Table 1 below.

Table 1: Sample data for numerical examples

Car Typ.	Cost in US$	μ_{cost}	Maint. cost in US$	μ_{maint}	Miles/ gal.	μ_{mile}
P_1	30,000	0.6	100	0.63	19	0.8
P_2	40,000	0.4	50	0.4	25	0.72
P_3	20,000	1.0	300	0.8	17	0.73
P_4	50,000	0.1	100	0.63	22	0.8
P_5	50,000	0.1	150	0.65	25	0.72
P_6	40,000	0.4	200	1.0	22	0.8
P_7	15,000	0.8	500	0.2	12	0.4
P_8	25,000	0.8	300	0.8	20	1.0

The obtained fuzzy membership values are according to the predefined fuzzy sets shown in Fig. 3below:

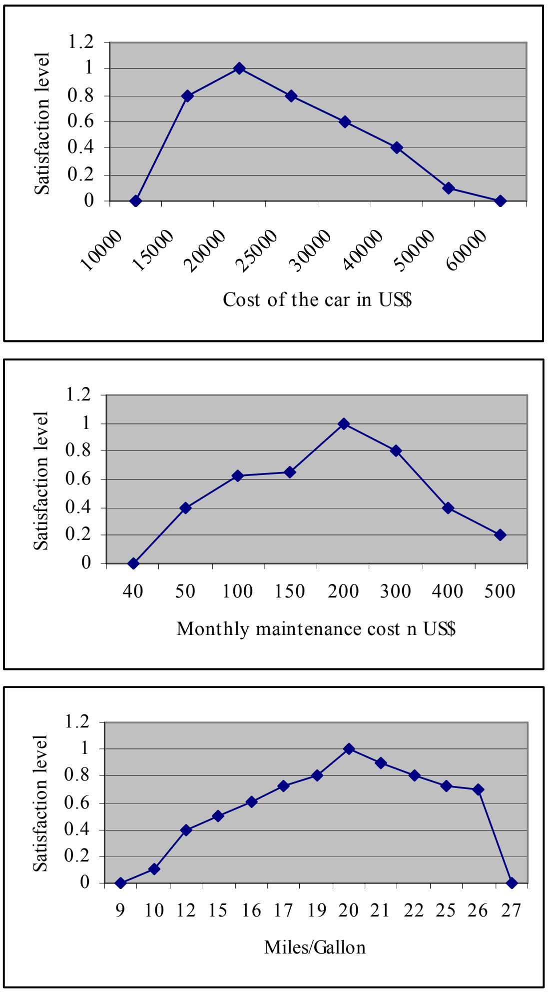

Fig. 3. Fuzzy sets of the satisfaction level on the three attributes

Applying Equations (1), (2) and (3), we obtain the average satisfaction values for each attribute (S_m), the most value of S_m with the preset value of α=1 and β=5, and the flexibility range (f_m) as shown in Table 2.

Table 2: Main parameters computed by Equations (1)-(3)

Parameters	Cost	Maint.	Mileage
S	0.53	0.64	0.75
μ_{most} at α=1, β=5	0.57	0.8	0.92
f	0.43	0.2	0.08

Using the five-step procedure given above, we can cluster those car products into hierarchical groups in the following manner:

Follow the steps as explained in Section 3.3 with the assumption that customer's priority on attributes is in the order of cost, maintenance, and mileage, the resulting clusters for the available product set $P=\{P_1,P_2,\ldots,P_8\}$ are as follows:

Step 1-2: Determine values of f_{1n} by (5.1-3) which are the flexibility value of each product to the most important attribute, Cost.

$$f_{1n}=\{0.03,\ 0,\ 0.43,\ 0,\ 0,\ 0.23,\ 0.23\}$$

Step 3.1: Set C_{11} which is the 1st set of qualified products according to the first attribute of Cost.

$$C_{11}=\{P_1,\ P_3,\ P_7,\ P_8\}$$

Step 4: This repeats Steps 1-3 for selecting the qualified products out of those in C_{11}. Two repetitions are required for the 2nd and 3rd attributes: maintenance and mileage, respectively. Results of the *2nd* round and the *3rd* round, C_{12} and C_{13}, are shown next.

$$C_{12}=\{P_1,\ P_3,\ P_7,\ P_8\} \text{ and}$$
$$C_{13}=\{P_8\}$$

Note that there is no product abandoned in C_{12} since all are unqualified according to the flexibility range computed in 3.1. Hence, Step 3.2 applies and it is treated evenly to all products. The final round for the cost attribute render the cluster in C_{13} which can be viewed as the best qualified group of products according to the specified fuzzy attributes and their preference level given by the customer.

Step 5: Follow the steps 1-4 for the remaing seven products which can be determined by the set property, $P=P-\{P_8\}$. The clusters obtained from Step 5 are listed below:

$$C_{23}=\{P_1,\ P_3,\ P_7\}\text{, and}$$
$$C_{33}=\{P_6\}\text{, and}$$
$$C_{43}=\{P_2,\ P_4,\ P_5\}.$$

Comparing these four clusters with those five clusters obtained in [1], this model with "most" function of logistic curve with parameter α=1, β=5 yields a lower scale of discrimination on the available products on the internet. The accelerating and decelerating speeds are at higher rate than the piece-wise linear model in [1].

If setting α=1 and β=3 which yields lower speed of acceleration and deceleration of change in "most" values, we can obtained a similar discriminating power of model to that of [1] and the resulting cluster is exactly the same, which is:

$C_{13} = \{P_8\}$,

$C_{23} = \{P_3\}$,

$C_{33} = \{P_1, P_7\}$,

$C_{43} = \{P_6\}$, and

$C_{53} = \{P_2, P_4, P_5\}$.

This model, hence, enable users to tune up the discriminating power of the clustering model according to their requirement. Users have to carefully tune the behavior of the system for optimal performance.

5. Conclusion

A fuzzy-oriented clustering methodology is extended and generalized by adopting a non-linearly linguistic quantifier of "most" function to improve the discriminating ability. This method can be applied as an engine to effectively support the e-merchandize or internet shopping. Introduction of the non-linear "most" fuzzy operation with adjustable parameters provides adaptive discriminating power to this clustering method.

The numerical examples have illustrated the effective adaptation of the clustering method. By increasing the beta value, a parameter of "most" quantifier, the discriminating power of the classifier is reduced, and vice versa. This yields flexibility in merchandizing the optimal set of products per each hierarchical level on e-catalogue.

The topics for future research may be that of the ranking procedure for those products in the same hierarchical level so that customers are supported in a finer degree. Another direction is to incorporate penalty function into the model so that we can discriminate products with very low satisfaction level out of those with "most" satisfactory and/or average satisfactory levels.

Reference

[1] Mohanty, B.K. and Bhasker, B. (2005) Product classification in the internet business- a fuzzy approach, Decision Support Systems 38: 611-619.

[2] Kacprzyk, J. and Yager, R.R. (1984) Linguistic quantifiers and belief quanlification in fuzzy multicriteria and multi-stage decision making, Control Cybern. 13:155-173.

[3] Ryu, Y.U. (1999), A hierarchical constraint satisfaction approach to product selection for electronic shopping support. IEEE Trans. Syst. Man Cybern., Part A. Syst. Humans 29: 525-532.

[4] Mohanty, B.K. and Passy, K. (2006). Web based information for product ranking in e-business: a fuzzy approach, ACM International Conference Proceeding Series, New Brunswick, Canada, 156: 558 – 563

[5] Lee, R.M. and Widmeyer, G.R. (1986) Shopping in the electronic market place, J. Manage. Inf. Syst. 21-35.

[6] Dubois, D. and Prade, H. (1978) Operation on fuzzy numbers, Int. J. Syst. Sci. 9: 613.

[7] Zadeh, L.A. (1968) Fuzzy algorithms, Information and Control, 5: 94-102.