



## Multi-group QoS consensus for web services

Wei-Li Lin<sup>a,\*</sup>, Chi-Chun Lo<sup>a</sup>, Kuo-Ming Chao<sup>b,c</sup>, Nick Godwin<sup>c</sup>

<sup>a</sup> Institute of Information Management, National Chiao-Tung University, Hsin-Chu, 300, Taiwan

<sup>b</sup> Software School, Fudan University, China

<sup>c</sup> DSM Research Group, Department of Computer Science, Coventry University, Coventry, CV1 5FB, UK

### ARTICLE INFO

#### Article history:

Received 10 June 2009

Received in revised form 13 September 2009

Available online 29 January 2010

#### Keywords:

Web service QoS

Multi-attributes clustering

Similarity analysis

### ABSTRACT

QoS has been considered as a significant factor for web service marketing and selection. The interpretation of QoS value from web service consumers and providers would be very different. However, a large group of web service participants with different backgrounds may have difficulties in reaching consensus on the values of multi-dimensional web service QoS, so they may have to be clustered in multi-groups in order to improve effectiveness and efficiency. The similarity of clustered fuzzy QoS dispositions as well as their preference order over these attributes should be analyzed to form a multi-groups consensus framework. A soft multi-groups clustering approach could be adopted to prevent opinions from being excluded unintentionally. The group boundaries and similarity thresholds which are used for clustering and analyzing fuzzy QoS opinions can be moderated dynamically according to the feedback from the internal learning mechanism and the web service consumers. As a result, a model for marketing web services based on multi-group consumers' QoS consensus, the FMG-QCMA (Fuzzy Multi-Groups based QoS Consensus Moderation Approach), is proposed to meet the above requirements. The proposed FMG-QCMA is also evaluated through a case study to demonstrate its effectiveness and efficiency in relation to an existing framework, QCMA (QoS Consensus Moderation Approach).

© 2010 Elsevier Inc. All rights reserved.

## 1. Introduction

Web services have increasingly become a popular technology for design of e-business systems. The number of e-business systems such as e-auction, e-mail, hotel booking service and travel ticket booking (including airline, train, and amusement park) implemented through web services is increasing [1–4]. As a result, web service selection and web service marketing are expected to become key issues in the development of service-oriented computing.

QoS (Quality of Service) has been considered as a significant criterion in the selection of web services [1,5–9] and marketing. In our previous work [1], we have designed a model of consumer-centric QoS-aware selection, QCMA (QoS Consensus Moderation Approach), to analyze the group consensus based on their fuzzy opinion similarity and QoS preference with a number of QoS attributes (defined by W3C [1,5,10]) for depicting how the group of consumers selecting a web service. QCMA utilized SAM (Similarity Aggregation Method) and RMGDP (Resolution Method for Group Decision Problems) [10–14] to perform group consensus in similarity and preference analysis over the QoS attributes. However, there are still some challenges that the single group based QCMA with opinion similarity and preference analysis for web service selection has not addressed. These include the following:

\* Corresponding author.

E-mail address: weililin.iim92@nctu.edu.tw (W.-L. Lin).

1. If the fuzzy QoS opinions were collected from web service participants with very different backgrounds and potentially diverse perceptions, the obtained group consensus may not be effective. The resulting group opinion could attempt to represent too diverse a population, so one set of values for the QoS criteria may not be effective for web service selection or marketing.
2. Even though the single group consensus in QCMA is built upon opinion similarity and QoS preference order [1], the criterion for combining QoS similarity and preference orders requires careful review in the context of multi-group consensus analysis.
3. Some outliers identified by our previous proposed approach should be re-analyzed. Some of them could be re-classified into other appropriate groups if a multi-groups approach is adopted, as these outliers' opinions may have meaningful correlation with others. The omission of those outliers without further examination can be inappropriate. Furthermore, due to the multi-attributes structure, some identified outliers based on a specific attribute may not necessarily be outliers when preference ordered multiple attributes are included. Multi-attributes based outlier identification, however, is more difficult than is the case with single attribute based outlier identification.
4. The operational complexity of SAM in QCMA is relatively high when the number of inputs is large. This is due to the required exhaustive calculation on the similarity analysis for each pair of QoS opinions. With  $n$  QoS opinions the complexity will be  $O(n^2)$ . This growth in the number of operations as the number of inputs increases will hinder the system scalability.

The proposed approach is formulated in order to address the above issues: improvement of opinion classifications; reduction of computational complexity, and, integrating the approaches of fuzzy clustering and multi-group classification with multiple attributes. The next section reports studies on a number of existing multi-attributes based information clustering approaches and the analysis of their advantages and disadvantages. The remainder of the paper is organized as follows. Section 3 presents the proposed FMG-QCMA which includes the system behavior and similarity analysis. Some descriptions on QCMA will be included in this section. Section 4 reports on experimental results with a case study of hotel booking web services. Finally, Section 5 presents the conclusions.

## 2. Multi-attributes based information clustering – existing solutions

The approaches for multi-attributes based information clustering have been studied for some time. The amount of the literatures reporting the theoretical developments and their applications are vast [2,15–20,22,23]. However, this area of research can be approximately classified into three main categories: shifting or scaling based clustering, parallel clustering and fuzzy clustering [18–20]. In this section, we only briefly describe and analyze three important sources that are related to this research. A comprehensive literature review on this area can be found in [15,20].

Haixun Wang et al. [18] proposed a pCluster model for multi-dimensional pattern similarity clustering in large data sets. In the research, the similar patterns of data are identified by a “shifting relationship” or by a “scaling relationship” rather than by traditional distance based similarity such as Euclidean distance, Manhattan distance, cosine distance, etc. Therefore, the “shifting patterns” or “scaling patterns” in the pCluster model is very effective for clustering large data sets.

However, pCluster does not address the issue regarding weight distribution to the attributes which is the main characteristics in the application of multi-attributes QoS-aware web service selection or marketing. For instance, the “4 type patterns” in pCluster was denoted as  $(a, b, c, d)$ ,  $0 \leq a, b, c, d \leq 10$ . In the study,  $(1, 2, 3, 6)$ ,  $(2, 3, 4, 7)$  were identified as “similar” by using shift comparison. However, each type in the “4 type patterns” should not be assigned with equal weight to multi-attributes based applications. In other words, each dimension (attribute) should be given weight in order to reflect the importance of these dimensions. That is, if the weight distribution for  $(1, 2, 3, 6)$  is  $(20\%, 30\%, 15\%, 35\%)$  and the weight distribution for  $(2, 3, 4, 7)$  is  $(25\%, 40\%, 30\%, 5\%)$ , then  $(1, 2, 3, 6)$  and  $(2, 3, 4, 7)$  should not be classified to have similar pattern.

M. Fazeli et al. [19] proposed a parallel algorithm to tackle multi-features data clustering for multi-computers with star topology. The proposed parallel algorithm completes a clustering problem of  $N$  data patterns with  $M$  features per pattern and  $K$  clusters in complexity of  $O(K + S^2 - T^2)$ , where  $N \cdot M = S!$ ,  $K \cdot M = T!$ . In the study, the data is depicted with a feature vector  $v$  which is a set of measurements  $(v_1, v_2, \dots, v_M)$  that map to the properties of a collection of data into a Euclidean space of dimension  $M$ . It divides the  $N$  multi-feature data patterns into  $K$  clusters via a set of clustering criteria and the  $K$  clusters can be represented as  $(S_1, S_2, \dots, S_K)$  which is shown below:

$$S_k = \{i \mid C[i] = k, 0 \leq k \leq K - 1\} \quad (1)$$

A popular clustering technique, *squared-error algorithm*, is adopted for the multi-features data clustering with the square distance  $d2$  between pattern  $i$  and cluster  $k$  as shown below:

$$d2[i, k] = \sum (F[i, j] - \text{centre}[k, j])^2 \quad (2)$$

where the cluster centre is obtained by mean of feature matrix  $F[i, j]$ , which indicates  $i$ th data with  $j$ th feature and is represented as a  $(1 \times M)$  vector. With the  $|S_k|$ , the centre of cluster  $k$  can be defined as shown below:

$$\text{centre}[k, j] = \frac{1}{|S_k|} \sum_{i \in S_k} F[i, j], \quad 0 \leq j \leq M \quad (3)$$

The *squared-error algorithm* is used to compute the distance  $d2[i, k]$  of each pattern  $i$  from each cluster  $k$ , and to choose the minimum distance to all cluster centers. Therefore, all patterns can be efficiently clustered into the right cluster according to the minimum distance from the corresponding cluster centre. Even though the multi-features data can be clustered via the parallel algorithm, the possible weight distribution over these  $M$  dimensions was not discussed.

Rui Xu [20] has conducted a significant review on fuzzy clustering related research. In [21], all selected objects can be clustered into the most appropriate groups with a certain degree of membership. The concept of fuzzy clustering and fuzzy boundary was considered in FCM [22]. FCM attempts to find a partition ( $c$  fuzzy clusters) for a set of data points  $x_j \in \mathfrak{R}^d$ ,  $j = 1, \dots, N$ , while minimizing the cost function as shown below:

$$J(U, M) = \sum_{i=1}^c \sum_{j=1}^N (u_{i,j})^m D_{ij} \quad (4)$$

where  $U$  indicates the fuzzy partition matrix  $[u_{i,j}]_{c \times N}$ ,  $u_{i,j}$  indicates the membership coefficient of  $j$ th object in the  $i$ th cluster and  $u_{i,j} \in [0, 1]$ .  $M$  indicates the cluster prototype (mean or cluster) matrix,  $M = [m_1, m_2, \dots, m_c]$  and  $m \in [1, \infty)$  [23].  $D_{ij}$  indicates the distance measure between  $x_j$  and  $m_i$ , and  $D_{ij} = D(x_j, m_i)$ .

The standard FCM, in which the Euclidean or  $L_2$  norm distance function is used, is summarized as follows:

1. Select appropriate values for  $m$ ,  $c$ , and a small positive number  $\varepsilon$ . Thereafter, initialize the prototype matrix  $M$  randomly and set step variable  $t = 0$ .
2. Calculate (at  $t = 0$ ) or update ( $t > 0$ ) the membership matrix  $U$  by:

$$u_{ij}^{(t+1)} = \frac{1}{\left(\sum_{l=1}^c \left(\frac{D_{lj}}{D_{ij}}\right)^{\frac{1}{1-m}}\right)} \quad \text{for } i = 1, \dots, c \text{ and } j = 1, \dots, N. \quad (5)$$

3. Update the prototype matrix  $M$  by

$$m_i^{(t+1)} = \frac{\sum_{j=1}^N (u_{ij}^{t+1})^m x_j}{\left(\sum_{j=1}^N (u_{ij}^{t+1})^m\right)} \quad \text{for } i = 1, \dots, c. \quad (6)$$

4. Repeat steps 2–3 until  $|M^{(t+1)} - M^t| < \varepsilon$ .

In this approach, assigning appropriate initial values to  $m$ ,  $c$ , and  $\varepsilon$  could be an issue. A number of unexpected outliers could appear if  $m$ ,  $c$ , and  $\varepsilon$  are set with improper values.

The aforementioned research papers have made significant contributions to the area of multi-attributes information clustering. However, the challenges for building a multi-groups-based information model that can be used for effective decisions making, such as multi-groups based QoS-aware selection of web service or marketing, still remain and these are summarized as follows:

1. Associated weightings on each QoS attribute should be considered, due to different preference orders given by the web service consumers. Therefore, a weighted multi-attributes QoS similarity should be defined.
2. The method of identifying outliers can be improved, as some members/opinions can be re-analyzed and re-classified. To prevent removing possibly meaningful data which falls just outside of the pre-defined group boundaries, the multi-attributes based clustering criteria should be formulated with fuzzy evaluation.

The proposed FMQ-QCMA compliments the existing research work mentioned above by incorporating fuzzy clustering, the SAM analysis approach, and RMGDP QoS preference order analysis to reach multi-group consensus and to measure the quality of grouping. It is an iterated assessment mechanism, as the system takes into account feedback from previous rounds to carry out further refinements on grouping. The quality of grouping is expected to improve as the process progresses. The efficiency of handling multi-groups based QoS-aware selection of web service or marketing, compared with the single group approach, should be better, as the computational complexity is expected to reduce. Members in each sub-group will have greater similarity with each other than with those in other groups. Also, the spread of similarity measurement within each sub-group will be lower than the spread in one single group. As a result, the service selection within multi-group consensus is expected to be more precise than one group consensus. The FMQ-QCMA inherits some of the features from QCMA such as SAM for group preference similarity measurement and RMGDP for preference order analysis. So, the next sub-section briefly describes QCMA architecture and its overall procedure.

### 3. The proposed approach – FMG-QCMA

#### 3.1. Previous work: QCMA

QCMA (QoS Consensus Moderation Approach) is employed to obtain and moderate group consensus on QoS in selecting web services [1]. QCMA enhances the moderation process by introducing a modified method of reaching group opinion similarities and preferences on QoS attributes.

In QCMA, an initial set of web services and web service consumers' opinions have to be established in order to build a preliminary group consensus. The consumers and providers have to make a judgment on the quality of the participating web services by expressing and defining their subjective opinions such as good reliability, bad performance and high availability etc., on all pre-determined 13 QoS attributes as well as giving their preference ordering over these attributes. The QCMA, including a set of reasoning approaches, is able to analyze and compute the opinions and their preferences to determine group QoS consensus on these services. So, the QoS of each service can be advertised in UDDI for service discovery and selection according to the reached consensus. QCMA also provides a moderation mechanism to accommodate the new opinions from new consumers and new services as well as to reflect the changes from the consumers and users in the dynamic environment. One of the characteristics of QCMA is its flexibility that allows the consumers to express fuzzy opinions. So, the fuzzy QoS opinions from these consumers were analyzed through two phases: group similarity analysis via SAM and QoS preference order analysis via RMGDP.

SAM was developed for resolving conflicts that arise from different opinions [11,12]. In SAM the different fuzzy opinions will be aggregated into opinion consensus classes so that they can be measured by their similarities to each other. The procedure to perform SAM is organized into 8 steps as stated below [24,25].

1. First, each participant  $k$  represents his/her subjective fuzzy QoS opinion on each specific QoS attribute  $a_i$ , which is denoted as  $wsa_{a_i}^k$ , with a positive trapezoidal fuzzy number shown in Eq. (7).

$$wsa_{a_i}^k = ((x_1)_{a_i}^k, (x_2)_{a_i}^k, (x_3)_{a_i}^k, (x_4)_{a_i}^k), \quad 0 \leq (x_1)_{a_i}^k \leq (x_2)_{a_i}^k \leq (x_3)_{a_i}^k \leq (x_4)_{a_i}^k \leq 10 \quad (7)$$

The structure of such a fuzzy number can be also illustrated as in Fig. 1:

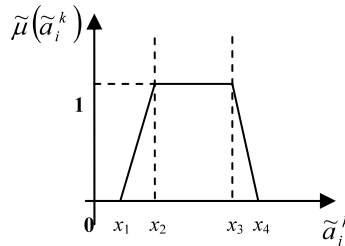


Fig. 1. A trapezoidal fuzzy number.

2. To obtain opinion similarity between any two fuzzy QoS opinions of each QoS attribute  $a_i$ ,  $wsa_{a_i}^k$  and  $wsa_{a_i}^j$ . The similarity between  $wsa_{a_i}^k$  and  $wsa_{a_i}^j$ , which is denoted as  $Sim_{a_i}^{jk}$ , can be obtained via the following equation:

$$Sim_{a_i}^{jk} = \frac{\int (\min\{\tilde{\mu}(wsa_{a_i}^j), \tilde{\mu}(wsa_{a_i}^k)\}) dx}{\int (\max\{\tilde{\mu}(wsa_{a_i}^j), \tilde{\mu}(wsa_{a_i}^k)\}) dx} \quad (8)$$

where  $\int (\min\{\tilde{\mu}(wsa_{a_i}^j), \tilde{\mu}(wsa_{a_i}^k)\}) dx$  indicates the consistent area between  $wsa_{a_i}^j$  and  $wsa_{a_i}^k$  which can be depicted as Fig. 2, and  $\int (\max\{\tilde{\mu}(wsa_{a_i}^j), \tilde{\mu}(wsa_{a_i}^k)\}) dx$  indicates the total area including  $wsa_{a_i}^j$  and  $wsa_{a_i}^k$  which can be depicted as

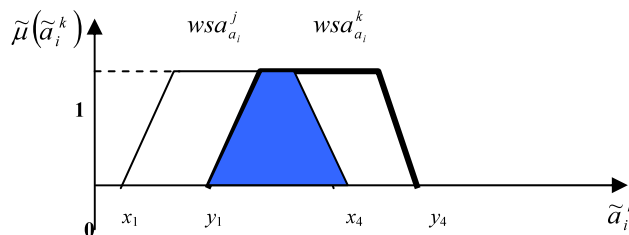


Fig. 2. The consistent area between two opinions:  $wsa_{a_i}^j$  and  $wsa_{a_i}^k$ .

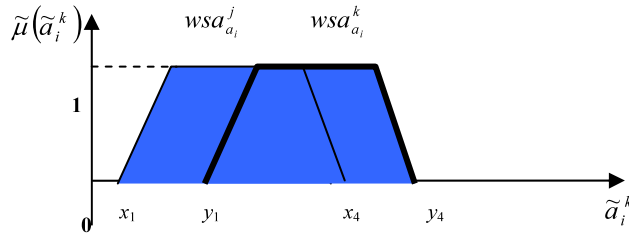


Fig. 3. The total area including two opinions:  $wsa_{a_i}^j$  and  $wsa_{a_i}^k$ .

Fig. 3. Although Eq. (8) is the definition of similarity used in the original formulation of SAM, it is possible to change this step and use alternative measures of similarity. Such a change does not require alterations to the other steps in the method.

- To build an AM (Agreement Matrix), which can be represented as Eq. (9), showing each similarity between pairs of participants in the group

$$AM_{a_i} = \begin{bmatrix} 1 & Sim_{a_i}^{12} & \dots & Sim_{a_i}^{1j} & \dots & Sim_{a_i}^{1n} \\ Sim_{a_i}^{21} & 1 & \vdots & \vdots & \vdots & \vdots \\ \vdots & \dots & 1 & \vdots & \vdots & \vdots \\ Sim_{a_i}^{j1} & \dots & \dots & 1 & \vdots & Sim_{a_i}^{jn} \\ \vdots & \dots & \dots & \dots & 1 & \vdots \\ Sim_{a_i}^{n1} & Sim_{a_i}^{n2} & \dots & Sim_{a_i}^{nj} & \dots & 1 \end{bmatrix}_{n \times n} \quad (9)$$

- To calculate an AAD (Average Agreement Degree), denoted as  $A(wsa_{a_i}^k)$ , for each opinion  $wsa_{a_i}^k$  in the group. The value of  $A(wsa_{a_i}^k)$  can be obtained from Eq. (10):

$$A(wsa_{a_i}^k) = \frac{1}{n-1} \sum_{\substack{j=1 \\ k \neq j}}^n Sim_{kj} \quad (10)$$

- To obtains an RAD (Relative Agreement Degree) for each individual opinion using the following formula.

$$RAD(wsa_{a_i}^k) = \frac{A(wsa_{a_i}^k)}{\sum_{j=1}^n A(wsa_{a_i}^j)} \quad (11)$$

- This step involves the assignment of a weighting variable,  $w_k$ , to each opinion.
- This step obtains the CDC (Consensus Degree Coefficient) for each participant:

$$CDC(wsa_{a_i}^k) = \beta \times w_k + (1 - \beta) \times RAD(wsa_{a_i}^k) \quad (12)$$

where  $\beta$  is a control variable to indicate the relation between the experts and the unmoderated opinions of the users. All the  $RAD(wsa_{a_i}^k)$  can be obtained through similarity analysis. However, the variation between the consensus using the  $RAD(wsa_{a_i}^k)$  and the consensus using  $CDC(wsa_{a_i}^k)$  can be quite small for a large population of users (it can be verified in the FMG-QCMA Validation and Evaluation). Therefore, in some cases it is possible to simplify the use of  $CDC(wsa_{a_i}^k)$  by setting  $\beta$  in  $CDC(wsa_{a_i}^k)$  as zero so that  $CDC(wsa_{a_i}^k)$  is equal to  $RAD(wsa_{a_i}^k)$ . Nevertheless, in cases considered later a non-zero value for  $\beta$  is used in developing FMG-QCMA.

- Aggregate the fuzzy opinions by the CDC in (12) as the formula as below:

$$\tilde{R}_{a_i} = \sum_{k=1}^n CDC(wsa_{a_i}^k) \bullet wsa_{a_i}^k \quad (13)$$

where  $\tilde{R}_{a_i}$  indicates an “overall” fuzzy number of combining all opinions on QoS attribute  $a_i$ . (The “multiplication” used in (13) represents the weighted combination of the 4-vectors representing the trapezoidal numbers.)

Opinion similarity enables the service consumers to reach a consensus on the interpretation of a QoS attribute for web services. These may have different preferences on the attributes. Therefore, RMGDP is proposed to resolve their differences on preferences via three phases: the transformation phase, the aggregation phase and the exploitation phase [1]. The group preference over QoS attributes of each sub-group will be obtained via the RMGDP process.

### 3.2. System behavior of FMG-QCMA

FMG-QCMA (Fuzzy Multi-Groups based QCMA), an extension of QCMA framework by incorporating a fuzzy clustering mechanism, attempts to provide an effective architecture/mechanism for fuzzy multi-groups based web service selection and service provider's market segmentation. Differing from QCMA which analyzes the fuzzy opinions and preferences given by the service consumers and providers on a collection of pre-determined web services QoS in attempt to reach a single consensus on the chosen subjective terms and their preference orders for web service selection or marketing, FMG-QCMA is capable of clustering service consumers (fuzzy opinions) into a number of sub-groups according to consumers' similar dispositions on pre-determined web services QoS attributes and focuses on the assessment of a specific collection of recommended web services for each clustered sub-group.

In FMG-QCMA, we assume that the service consumers' dispositions on QoS for service selection are static over a period of time. Once the consumers' dispositions in QoS are obtained, the service providers supporting various levels of QoS can promote the right quality level of services to the right group of service consumers. When a service request is issued by a service consumer, the service providers will look up the service consumer's profile and provide close match services according to the consumer's past selection patterns. Each service consumer needs to express his/her dispositions on all 13 QoS attributes [5] with a selection from a set of pre-defined scales and their associated trapezoidal fuzzy number as well as his/her preference order over these QoS attributes. The selected disposition for each service consumer will be treated as his/her fuzzy QoS opinion (or perception) that will be used in selecting a web service.

For reaching the objective above, each of the thirteen QoS attributes is possible to find a numerical measure of quality in the context of the type of service required. The values of this measure can then be scaled to correspond to numbers in the range [0, 10]. For each service consumer there will be a range of values that will be considered appropriate for the service they require. At the lower end there will be a cut off value and services with lower values will not be considered in any circumstance. At the higher end there will be a value above which improvement in quality will not be relevant to their needs and services above the threshold will only be considered if they do not cost any extra. So for each attribute a service consumer must choose four points in the range of values.

1. Below this level a service cannot to be considered in any circumstances.
2. This is the lower end of the normal expected quality for a service.
3. This is the upper end of the normal expected quality for a service.
4. Getting above this level could not be used to justify extra investment.

Given the choice of these values for each attribute by a service consumer a corresponding set of trapezoidal numbers over the standardized scale of [0, 10] can be defined.

For instance, for the attribute, performance, the natural measure is a response time in seconds. The upper limit of quality is immediate response, 0 seconds (standardized quality value = 10). The lower limit is context dependent but assuming a straight single retrieval requirement, 10 seconds is taken as the lower limit (standardized quality value = 0.0). The standardization scaling can most conveniently be presented as a table showing measures corresponding to the eleven scaled values [0.0, 0.5, 1.0, . . . , 9.0, 9.5, 10]. This is shown in Table 1. With reliability the natural quality measure is the percentage of transactions that will be completely successful. The scaling is shown in Table 2.

There are similar tables for each of the thirteen attributes which can be presented to service consumers for their choice of the four key levels.

FMG-QCMA, then, can collect these fuzzy QoS opinions to proceed the following four phases of FMG-QCMA operations which is depicted in Fig. 4.

**Table 1**  
The natural measure for performance quality.

Performance quality rating	Response time in seconds	Performance quality rating	Response time in seconds
0.0	10.00	5.5	1.75
0.5	8.00	6.0	1.50
1.0	7.00	6.5	1.25
1.5	6.00	7.0	1.00
2.0	5.00	7.5	0.75
2.5	4.00	8.0	0.50
3.0	3.00	8.5	0.25
3.5	2.75	9.0	0.05
4.0	2.50	9.5	0.02
4.5	2.25	10	0.00
5.0	2.00		

**Table 2**  
The natural measure for reliability quality.

Reliability quality rating	Percentage transaction success	Reliability quality rating	Percentage transaction success
0.0	50.0%	5.5	91.5%
0.5	60.0%	6.0	92.5%
1.0	70.0%	6.5	94.0%
1.5	72.5%	7.0	95.0%
2.0	75.0%	7.5	97.0%
2.5	77.5%	8.0	99.0%
3.0	80.0%	8.5	99.3%
3.5	82.5%	9.0	99.5%
4.0	85.0%	9.5	99.8%
4.5	87.5%	10	100%
5.0	90.0%		

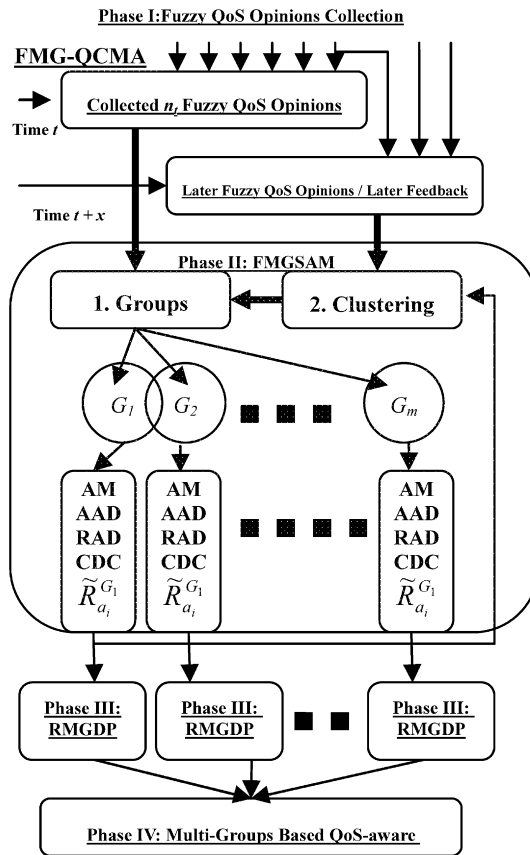


Fig. 4. FMG-QCMA system behavior.

In Fig. 4, there are four phases for handling all incoming fuzzy QoS opinions:

Phase I: To collect consumers’ fuzzy QoS opinions which reflect consumers’ disposition in QoS, their preferences order over QoS attributes, and initializing parameters for grouping such as similarity thresholds for any pair of fuzzy QoS opinions and sub-groups’ fuzzy boundaries. These values of system parameters will be evaluated by the system so they can be changed or adjusted at later stages, if they are inappropriate.

Phase II: To cluster all the collected fuzzy QoS opinions into sub-groups via *Groups Clustering*. Each allocation will be evaluated via *Clustering Verification*.

The operation *Groups Clustering* realized by *Algorithm Fuzzy\_Clustering* (see Appendix A) is to populate sub-groups with the collected fuzzy QoS opinions according to measurements of fuzzy QoS opinion similarity. Each fuzzy QoS opinion can be allocated into one or two sub-groups, as it depends on the degree of similarity to the related (close) sub-groups and the pre-set fuzzy boundaries. Each sub-group will be assessed by using Agreement Matrix (AM)/Average Agreement Degree (AAD) and Relative Agreement Degree (RAD)/Consensus Degree Coefficient (CDC) in order to find the

group similarity on fuzzy QoS opinions. In other words, it examines the degree of group consensus over the concept of disposition on the pre-defined QoS attributes.

The operation *Clustering Verification* is performed by *Algorithm Clustering Verification* (see Appendix C) and used for performing an analysis on new fuzzy QoS opinions from new web service consumers or misallocated existing opinions. In *Clustering Verification*, two scenarios will possibly occur:

1. There are two categories of similarities defined in the system: full and partial membership. Each opinion sub-group has fuzzy boundaries. Two neighboring sub-groups are likely to have overlapping areas in which members belong to both groups. When a member has full membership to a group, it means that the opinion has been assigned to the right group. The process for allocating this opinion will stop. However, if an opinion has been evaluated as partial membership to a group, it will be evaluated against adjacent group in order to identify its degree of membership. These opinions can be preliminarily clustered into arbitrary number of groups. So, producing good quality in grouping in the first instance is not expected. However, the system can evaluate the quality by measuring group similarity co-efficiency by using FMGSAM. If it does not reach desired level, the system boundaries or number of groups will be changed accordingly. This process will be iterated until the satisfactory results produced or it could be terminated after a number of tries. All the new fuzzy QoS opinions will have to be explored and analyzed to ensure that they are classified appropriately.
2. The purpose of grouping and identifying consensus on the QoS attributes is to recommend the right services to the consumers. If the consumers are often not satisfied with the services recommended by the system, this could be derived from inappropriate settings for the group boundaries or changes of consumer's pattern on service usage. We assume that the consumer will inform the system of the changes. Another set of processes will be activated to resolve the issue which will not be discussed here. For the other cases, the system records the events and accumulates these incidents. When the unsatisfactory number reaches or grows beyond the pre-set threshold, then the fuzzy boundaries for the sub-groups will be adjusted in order to improve the accuracy of recommending the appropriate services to the consumer to select. When this occurs, all fuzzy QoS opinions will be re-clustered into new opinion sub-groups.

Phase III: Once the quality of grouping presents a satisfactory result, the preference order for each sub-group can be calculated and obtained via RMGDP.

Phase IV: Through FMGSAM in Phase II and RMGDP in Phase III, the system is ready for use. Since the service consumer group consensus on QoS profiles and their preference orders can be obtained, the service providers can advertise and provide their services according to their target groups. The service consumer issuing the request to the system will receive a list of recommended web services which QoS can satisfy the required fuzzy opinions. Another filtering process based on individual QoS preference ordering will be applied in order to reduce unqualified services. After these processes, the consumers can select the desired services.

The proposed multi-attributes and multi-groups service selection is expected to produce better result than the single opinion group approach for service selection. The members in a sub-group will be more closely correlated than the single group. The system should be able to recommend a close match of services to the requests issued by the service consumer. Since the single group has been divided into a number of sub-groups and the size of each sub-group is smaller than or equal to the single group, the computational complexity can be reduced and the system efficiency can be improved. The following gives more detailed descriptions of the key steps in FMG-QCMA.

### 3.3. FMGSAM & RMGDP analysis

The proposed FMGSAM, derived from SAM in QCMA, is designed for similarity analysis under multi-groups framework. Following the system behavior of FMG-QCMA in the previous section, the FMGSAM can be organized with seven steps.

1. Represent All Fuzzy QoS Opinions: Based on the  $wsa_{a_i}^k$  represented in (7), the multi-attributes based fuzzy QoS opinion from web service participant  $k$ ,  $wsa_{S_Q}^k$ , is represented for all QoS attributes defined in  $S_Q$ , the set of QoS terms in W3C [5], as shown in (14).

$$wsa_{S_Q}^k = (wsa_{a_1}^k, wsa_{a_2}^k, \dots, wsa_{a_{13}}^k), \quad S_Q = \{a_1, a_2, a_3, \dots, a_{13}\} \quad (14)$$

The set of all the collected fuzzy QoS opinions  $wsa_{S_Q}^k$ , which is donated as  $WSA_{S_Q}$ , can be defined as follows:

$$WSA_{S_Q} = \{wsa_{S_Q}^k \mid k \in K, a_i \in S_Q\} \quad (15)$$

2. There are two conditions to use operations: *Groups Clustering* or *Clustering Verification*, in this step.

#### Condition to use "Groups Clustering"

The operation *Groups Clustering* is activated by either pre-set system time (as time  $t$ ) or the event of "re-clustering" from the operation *Clustering Verification*. When the operation *Groups Clustering* commences, all collected  $wsa_{S_Q}^k$  in  $WSA_{S_Q}$  will



be clustered into appropriate groups  $(G_1, G_2, \dots, G_m)$  through *Algorithm Fuzzy\_Clustering* and *Algorithm SimVerifier* (see Appendix B) based on the similarity threshold  $\tilde{d}_{S_Q}$  and the multi-attribute based similarity  $Sim_{S_Q}^{jk}$  between selected  $wsa_{S_Q}^j$  and  $wsa_{S_Q}^k$ .

$Sim_{S_Q}^{jk}$  can be obtained by the following equation:

$$Sim_{S_Q}^{jk} = (so_{a_1}^{jk} \times Sim_{a_1}^{jk}, so_{a_2}^{jk} \times Sim_{a_2}^{jk}, \dots, so_{a_{13}}^{jk} \times Sim_{a_{13}}^{jk}) \tag{16}$$

where  $Sim_{a_i}^{jk}$  indicates the similarity between  $wsa_{a_i}^j$  and  $wsa_{a_i}^k$  on QoS attribute  $a_i$  and can be obtained as follows:

$$Sim_{a_i}^{jk} = \frac{\min\{((x_1)_{a_i}^j + \int \tilde{\mu}(wsa_{a_i}^j) dx), ((x_1)_{a_i}^k + \int \tilde{\mu}(wsa_{a_i}^k) dx)\}}{\max\{((x_1)_{a_i}^j + \int \tilde{\mu}(wsa_{a_i}^j) dx), ((x_1)_{a_i}^k + \int \tilde{\mu}(wsa_{a_i}^k) dx)\}} \tag{17}$$

It can be noted that this measure of the similarity of two trapezoidal numbers is not the same as (8). This chosen formula is easier to calculate and gives comparable results. The element  $so_{a_i}^{jk}$  indicates the similarity of preference order between  $o_{a_i}^j$  and  $o_{a_i}^k$  and it can be obtained for the  $q$  QoS attributes by the following equation (for definition in W3C [5],  $q = 13$ ):

$$so_{a_i}^{jk} = \frac{q - |o_{a_i}^j - o_{a_i}^k|}{q} \tag{18}$$

$\tilde{d}_{S_Q}$  is a set of pairs of similarity thresholds given by an expert to emphasis the more extreme similarities given by the components of (16) in the selection and rejection of consumers for clusters. (See steps 3 and 4 in Appendix B.)

$$\tilde{d}_{S_Q} = (d_{S_Q}^l, d_{S_Q}^u), \quad 0 \leq d_{S_Q}^l < d_{S_Q}^u \leq 1 \tag{19}$$

Each  $Sim_{S_Q}^{jk}$  will be compared with the similarity threshold,  $\tilde{d}_{S_Q}$ , through the operators such as  $\succsim, \succ, \preccurlyeq, \prec$  and  $\cong$  that are defined in *Algorithm SimVerifier*. The pairs of values for  $\tilde{d}_{S_Q}$  determine the ways in which the individual similarities can influence the overall similarity. The clustering process requires the contributions to be added together and the total will determine the inclusion, semi-inclusion or exclusion of consumer from a cluster. The thresholds applied to the totals are the values  $f_{c-S_Q}$ . (See Appendix B steps 9, 11, 13, 15 and 17.)

**Condition to use “Clustering Verification”**

The operation *Clustering Verification* is launched by the addition of fuzzy QoS opinions contributed from the new web service consumers or by new feedback (mismatch) on unsatisfactory web services recommended by the system. Each new set of fuzzy QoS opinions will be assessed and assigned to appropriate opinion sub-groups if it is either similar to (with full membership) or nearly similar to (with partial membership). (*E\_Fail\_CDC*, *E\_Fuz\_Sim* or *E\_Abs\_Sim* in Appendix C.)

If the threshold of “re-clustering all fuzzy QoS opinions” is reached due to too many mismatch cases or the sub-group opinion consensus coefficient is too low, then a “re-clustering” event will be triggered to activate the operation *Groups Clustering* and this will moderate the threshold (boundaries) of subgroups in order to re-cluster the opinions.

- Determine Agreement Matrixes  $(AM_{p-a_i})_{n_p \times n_p}$  for each clustered opinion sub-group  $G_p$ . In the construction of the clusters all the necessary similarities (16) that are need to form the agreement matrixes shown in step 3 of SAM have been calculated.

$$(AM_{1-a_1}) = \begin{bmatrix} 1 & Sim_{a_1}^{12} & \dots & Sim_{a_1}^{1j} & \dots & Sim_{a_1}^{1n_1} \\ Sim_{a_1}^{21} & 1 & \vdots & \vdots & \vdots & \vdots \\ \vdots & \dots & 1 & \vdots & \vdots & \vdots \\ Sim_{a_1}^{j1} & \dots & \dots & 1 & \vdots & Sim_{a_1}^{jn_1} \\ \vdots & \dots & \dots & \dots & 1 & \vdots \\ Sim_{a_1}^{n_1 1} & Sim_{a_1}^{n_1 2} & \dots & Sim_{a_1}^{n_1 j} & \dots & 1 \end{bmatrix}_{n_1 \times n_1}, \dots,$$

$$(AM_{m-a_{13}}) = \begin{bmatrix} 1 & Sim_{a_{13}}^{12} & \cdots & Sim_{a_{13}}^{1j} & \cdots & Sim_{a_{13}}^{1n_m} \\ Sim_{a_{13}}^{21} & 1 & \vdots & \vdots & \vdots & \vdots \\ \vdots & \cdots & 1 & \vdots & \vdots & \vdots \\ Sim_{S_Q}^{j1} & \cdots & \cdots & 1 & \vdots & Sim_{a_{13}}^{jn_m} \\ \vdots & \cdots & \cdots & \cdots & 1 & \vdots \\ Sim_{a_{13}}^{n_m 1} & Sim_{a_{13}}^{n_m 2} & \cdots & Sim_{a_{13}}^{n_m j} & \cdots & 1 \end{bmatrix}_{n_m \times n_m} \quad (20)$$

4. Determine the Average Agreement Degrees: As in step 4 of SAM (definition (10)) it is possible to find the average agreement degree for each clustered opinion sub-group.
5. Determine the Relative Agreement Degrees: The RAD values within the clusters for each of the customers can be found using step 5 of the SAM process (definition (11)).
6. Determine the Consensus Degree Coefficients: As shown in step 6 of the SAM process it is possible to moderate the purely customer defined RAD values using weightings  $w_{k(G_p)}$  for each  $wsa_{S_Q}^{k(G_p)}$  in opinion sub-group  $G_p$ . With  $w_{k(G_p)}$  and assigned  $\beta$  the CDC for  $wsa_{S_Q}^{k(G_p)}$  can be obtained using definition (12).  
If the value of CDC is less than the pre-defined threshold, the group boundaries will be adjusted in order to increase group consensus coefficients. For other cases, the system progresses to the next step. This criterion is for the self-assessment mechanism to improve the quality of grouping.
7. If it is necessary definition (13) of the SAM process can be used to provide a consensus trapezoidal numbers for the clusters.

To provide a more detailed analysis of the clusters of customers it is useful to find consensus values for their preferences. The clusters identify similarities of quality expectations and the preferences will show the group's attitudes to the relative importance of these expectations. Therefore, all the QoS opinions  $wsa_{S_Q}^{G_p(k)}$  in  $G_p$  will be further analyzed via RMGDP according to associated preference order over all QoS attributes. For all clustered opinion sub-groups ( $G_1, G_2, \dots, G_m$ ) there are  $m$  RMGDP processes that will be performed respectively. In the FMGSAM, the individual consumer's preference ordering over QoS attributes was taken into consideration when the sub-groups are forming. Therefore at this stage, the members in a group should have strong consensus on the preference ordering.

### 3.4. Precision and efficiency

Calculating similarity for each pair of fuzzy QoS opinions in a group, is the dominant step in the complexity of FMG-QCMA and QCMA frameworks. The improvement on this step without compromising the precision of measurement of opinion similarities can significantly improve system efficiency. In the QCMA, the number of opinions in a single group is  $n$ , so its complexity is  $O(n^2)$  for  $AM$  generation in SAM. The number of processes and its associated complexity in FMG-QCMA can be significantly reduced, as it has multiple opinion sub-groups to fabricate  $(AM_1), (AM_2), \dots,$  and  $(AM_m)$  giving complexity of the form  $O(n_1^2 + n_2^2 + \cdots + n_m^2)$  and this will be lower than  $O(n^2)$  since  $n = n_1 + n_2 + \cdots + n_m$ .

In addition, FMG-QCMA can improve the precision in opinion similarity measurement which is illustrated as the following steps:

1. Let  $PSim_{FMQ}$  is denoted as precision (lowest similarity) for FMG-QCMA which is obtained from minimal  $Sim_{S_Q}^{G_p j}$  in generated  $(AM_1), (AM_2), \dots, (AM_m)$  defined in (20). Also, let  $PSim_Q$  is denoted as precision (lowest similarity) for QCMA which is obtained from minimal  $Sim_{S_Q}^{jk}$  in generated  $AM$  defined in (9).
2. The precision improvement by FMG-QCMA can be defined as  $Plmpr(\text{FMG-QCMA}/\text{QCMA})$ :

$$Plmpr(\text{FMG-QCMA}/\text{QCMA}) = (PSim_{FMQ}/PSim_Q) - 1 \quad (21)$$

The example below, where it is feasible to calculate the full set of similarities and the similarities used in FMG-QCMA, shows the improvement in precision introduced by the method.

## 4. Validation and evaluation of FMG-QCMA

This section presents how the proposed FMG-QCMA achieves marketing web services support via a case study, hotel booking web services. There were sixty fuzzy QoS dispositions collected from sixty consumers at time  $t$  as initial inputs to FMG-QCMA. This output from FMG-QCMA process contains a number of opinion sub-groups. Based on the framework, the preference order over 13 QoS attributes for each opinion sub-group on hotel booking web services will be obtained via RMGDP.

**Table 3**

The fuzzy QoS opinions (QoS disposition and preference order) from the first service consumer.

$wsa^i$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$	$a_9$	$a_{10}$	$a_{11}$	$a_{12}$	$a_{13}$
$o_i$	5	10	2	4	9	8	3	1	6	13	7	11	12
$x_1$	6.0	4.0	5.0	6.0	3.0	5.5	5.5	6.0	5.5	7.0	4.5	2.5	4.0
$x_2$	6.5	4.5	5.5	6.5	3.5	6.5	6.5	7.0	6.0	7.5	5.5	3.5	4.5
$x_3$	7.5	5.5	6.5	7.5	4.5	7.5	7.5	8.0	7.0	9.5	6.5	4.5	5.5
$x_4$	8.0	6.0	7.0	8.0	5.0	8.5	8.5	9.0	7.5	10.0	7.5	5.5	6.0

**Table 4**

The multi-attributes similarity analysis.

$G_1$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$	$a_9$	$a_{10}$	$a_{11}$	$a_{12}$	$a_{13}$	$sim\_result$
<b><math>Sim^{01\_01}</math></b>	–	–	–	–	–	–	–	–	–	–	–	–	–	–
<b><math>Sim^{01\_02}</math></b>	0.790	0.839	0.706	0.865	0.684	0.938	0.865	0.808	0.659	0.579	0.716	0.519	0.727	<b>0.461</b>
<b><math>Sim^{01\_03}</math></b>	0.339	0.564	0.733	0.677	0.479	1.000	0.793	0.740	0.659	0.401	0.923	0.692	0.716	<b>0.289</b>
<b><math>Sim^{01\_04}</math></b>	0.492	0.621	0.733	0.769	0.692	0.667	0.865	0.688	0.808	0.583	0.733	0.586	0.671	<b>0.286</b>
<b><math>Sim^{01\_05}</math></b>	0.721	0.423	0.667	0.862	0.495	0.564	0.923	0.538	0.769	0.567	0.643	0.755	0.725	<b>0.273</b>
<b><math>Sim^{01\_06}</math></b>	0.600	0.688	0.706	0.769	0.297	0.865	0.800	0.538	0.725	0.490	0.846	0.544	0.727	<b>0.257</b>
<b><math>Sim^{01\_07}</math></b>	0.513	0.423	0.533	0.492	0.385	0.867	0.747	0.846	0.790	0.486	0.521	1.000	0.846	<b>0.296</b>
$(Sim^{01\_08})$	0.369	0.604	0.800	1.000	0.489	0.564	1.000	0.346	0.923	0.389	0.586	0.252	0.725	<b>0.184</b>
$(Sim^{01\_09})$	0.790	0.818	0.467	0.361	0.476	0.721	0.800	0.875	0.791	0.437	0.781	0.346	0.587	<b>0.222</b>
$Sim^{01\_10}$	0.433	0.769	0.857	0.692	0.750	0.692	0.282	0.615	0.929	0.097	0.467	0.519	0.846	<b>0.145</b>
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
$Sim^{01\_58}$	0.481	0.348	0.824	0.538	0.692	0.692	0.149	0.718	0.533	0.692	0.144	0.635	0.923	<b>0.023</b>
$Sim^{01\_59}$	0.846	0.242	0.198	0.814	0.431	0.564	0.917	0.641	0.320	0.857	0.333	0.431	0.564	<b>0.007</b>
$Sim^{01\_60}$	0.433	0.423	0.875	0.369	0.577	0.867	0.149	0.556	0.852	0.256	0.333	0.564	0.781	<b>0.014</b>

Table 3 shows the first fuzzy QoS opinion of the sixty fuzzy QoS dispositions/preference ordering over these attributes from the first service consumer in the case study. Each consumer of the sixty service consumers follows the similar way to select his/her dispositions in format of (7) on each QoS attribute ( $a_1, a_2, \dots, a_{13}$ ) based on the available definitions given in Table 3. He/she also expresses a preference ordering (in row of  $o_i$ ) over these attributes which are shown on row for  $o_i$  in Table 3. “1” means the most important attribute and “13” represents the least important one. The new fuzzy QoS opinions and feedback as well as their new preference ordering, which will be used to demonstrate FMG-QCMA Moderation Process, follow the same format of the fuzzy QoS opinions defined in Table 3 to provide data.

The similarity threshold,  $\tilde{d}_{S_Q}$ , is initialized as (0.5, 0.6) and the  $f_{c\_S_Q}$  for similarity range is initialized as (0.15, 0.25). If the number of unsatisfactory feedbacks on the recommended web services is more than 3% of the whole opinion population, the resulting cluster is determined as inappropriate. In other words, if the system receives more than 3 unsatisfactory feedbacks from the users, the threshold  $\tilde{d}_{S_Q}$  needs to be moderated. Consequently it also re-clusters all fuzzy QoS opinions (sixty-eight opinions) by going through validation and evaluation process in *Clustering Verification*.

4.1. Reaching consensus

4.1.1. FMGSAM process

After the required inputs have been obtained, the FMGSAM starts to process the sixty  $wsa_{S_Q}^k$  in  $WSA_{S_Q}$ . One of tasks in Algorithm Fuzzy\_Clustering is to select an appropriate fuzzy QoS opinion (the first fuzzy opinion which has not been grouped) from the opinion pool to act as a group centre of a specific clustered group, so the other fuzzy QoS opinions (from those which have not been clustered into groups) will be evaluated against the center based on their similarity measurement. The result of the similarity analysis for the first clustered group ( $G_1$ ) is shown in Table 4.

In Table 4,  $Sim_{S_Q}^{jk}$  is represented as  $Sim^{j-k}$  which indicates the similarity between  $wsa_{S_Q}^j$  (group centre) and  $wsa_{S_Q}^k$ . The  $Sim_{S_Q}^{jk}$  represented as bold “ **$Sim^{j-k}$** ” indicates that  $Sim_{S_Q}^{jk}$  is similar to the group  $p$  and has its full membership to the group. The  $Sim_{S_Q}^{jk}$  represented as “ $(Sim^{j-k})$ ” with regular bracket indicates that  $Sim_{S_Q}^{jk}$  only has some degree similarity to the group, so it only has partial membership to the group  $p$ . Due to the analysis in Table 4 all the sixty fuzzy QoS opinions can be clustered into 13 sub-groups and represented with the index of fuzzy QoS opinion ( $k$  in  $wsa_{S_Q}^k$ ) in Table 5. The first column of each row in the table is the group ID and the rest of the columns are the group members. The values without brackets mean that these opinion sets are the core group members. The entries with bracket are the members, but they cannot be classified as core group members.

With the 13 clustered sub-groups and each having 13 QoS attributes, there are 169 agreement matrixes (AM) being generated. Some of these matrixes are shown in Fig. 5.

**Table 5**  
The clustered groups and opinions.

Group 1	1	2	3	4	5	6	7	(8)	(9)	12	(14)	(17)	(19)	22	23	(25)	(28)	(29)	31	(32)	(34)	(38)	42	46	47	(50)	55	(58)	(60)
Group 2	10	(17)	(18)	20	25	(26)	33	35	36	39	(41)	(49)																	
Group 3	11	9	(34)	(40)	(41)	(44)	(48)	50																					
Group 4	13	(8)	(19)	(21)	(26)	(28)	(29)	(30)	34	40	(44)	(56)	58																
Group 5	15	(19)	(21)	(32)	(38)	(43)	45																						
Group 6	16	(14)	(48)	(49)	53																								
Group 7	24	29	(30)	(32)	(38)																								
Group 8	27	(8)	14	57																									
Group 9	37	8	(19)	(21)	(43)	(49)																							
Group 10	51	30																											
Group 11	52	(38)	(44)	48	(60)																								
Group 12	54	43																											
Group 13	59	(21)	49																										

$$(AM_{1-a_1}) = \begin{bmatrix} 1 & 0.93_{a_1}^{1,2} & 0.88_{a_1}^{1,3} & \dots & 0.67_{a_1}^{1,58} & 1.00_{a_1}^{1,60} \\ 0.93_{a_1}^{2,1} & 1 & \vdots & \vdots & \vdots & \vdots \\ 0.88_{a_1}^{3,1} & \dots & 1 & \vdots & \vdots & \vdots \\ \vdots & \dots & \dots & 1 & \vdots & \vdots \\ 0.67_{a_1}^{58,1} & \dots & \dots & \dots & 1 & 0.67_{a_1}^{58,60} \\ 1.00_{a_1}^{60,1} & 0.93_{a_1}^{60,2} & \dots & \dots & 0.67_{a_1}^{60,58} & 1 \end{bmatrix}_{29 \times 29}, \dots, (AM_{13-a_{13}}) = \begin{bmatrix} 1 & 0.92_{a_{13}}^{21,49} & 0.92_{a_{11}}^{21,59} \\ 0.92_{a_{13}}^{49,21} & 1 & 1.00_{a_{13}}^{49,59} \\ 0.92_{a_{13}}^{59,21} & 1.00_{a_{13}}^{59,49} & 1 \end{bmatrix}_{3 \times 3}$$

Fig. 5. AMs generation for all clustered groups.

**Table 6**  
AAD, RAD and CDC for all groups.

$G_1/a_1$	AAD	RAD	$\beta$	$w_i$	CDC
$wsa^1$	0.8471	0.0358	0.4000	0.0345	0.0352
$wsa^2$	0.8635	0.0365	0.4000	0.0345	0.0357
$wsa^3$	0.7794	0.0329	0.4000	0.0345	0.0335
.....					
$wsa^{58}$	0.7499	0.0317	0.4000	0.0345	0.0328
$wsa^{60}$	0.8471	0.0358	0.4000	0.0345	0.0352
.....					
$G_{13}/a_{13}$	AAD	RAD	$\beta$	$w_i$	CDC
$wsa^{21}$	0.9168	0.3236	0.4000	0.3333	0.3275
$wsa^{49}$	0.9585	0.3383	0.4000	0.3333	0.3363
$wsa^{59}$	0.9583	0.3382	0.4000	0.3333	0.3363

After the AMs have been generated, the corresponding AAD, RAD and individual CDC for each fuzzy QoS opinion by each QoS attribute can be derived. Table 6 shows their corresponding results.

In CDC,  $\beta$  is set with 0.4 and each single QoS attribute based fuzzy QoS opinion within the same opinion sub-group is set with the same weight. These parameters setting were determined by experts' opinions according to their experience. With generated CDC of each fuzzy QoS opinions, the group consensus for each opinion sub-group can be obtained and represented as a 13-attributes fuzzy trapezoidal number. Each the sub-group's consensus, which is also represented as fuzzy trapezoidal number, is shown in Table 7.

4.1.2. RMGDP process

Based on 13 clustered sub-groups obtained through FMGSAM, there are 13 groups needed to be processed by RMGDP (denoted as RMGDP<sub>1</sub>, RMGDP<sub>2</sub>, ..., RMGDP<sub>12</sub>, RMGDP<sub>13</sub>) in order to gain their 13 QoS attributes preference orderings. RMGDP starts from transformation phase to generate preference relations for all fuzzy QoS opinions in the corresponding sub-group. Each matrix of preference relations  $p_{G_p}^k$  ( $G_p, P_k$  in Table 8), which represents all  $p_{ij}^k$  defined in RMGDP [1] for  $wsa_{S_Q}^k$  in group  $p$ , can be gained via transformation phase in RMGDP shown in Table 8.

According to Table 8, the corresponding preference relations aggregation, which is given as  $p_{G_p}^C$ , can be obtained via RMGDP [1] with equal weight ( $w_i^i = 1/|G_p|$ ). With all generated aggregation of preference relations for each clustered sub-group and equations for QGNDD/QGDD [1] for each QoS attribute in selected RMGDP<sub>p</sub> with the equal weight value  $w_i$  for each  $b_i$  in FMQ Moderator ( $w_i = 0.083$ ) can be represented as in Table 9.

**Table 7**  
Multi-groups consensus by QoS attributes.

$G_1/a_1$	CDC	$x_1$	$x_2$	$x_3$	$x_4$
$wsa^1$	0.0352	0.2115	0.2291	0.2644	0.2820
$wsa^2$	0.0357	0.1783	0.2140	0.2496	0.2853
$wsa^3$	0.0335	0.2347	0.2515	0.2850	0.3018
.....					
$wsa^{58}$	0.0328	0.0984	0.1311	0.1639	0.1967
$wsa^{60}$	0.0352	0.2115	0.2291	0.2644	0.2820
Group consensus		4.8365	5.7153	6.7153	7.5941
.....					
$G_{13}/a_{13}$	CDC	$x_1$	$x_2$	$x_3$	$x_4$
$wsa^{21}$	0.3275	1.309854	1.473585	1.801049	1.964781
$wsa^{49}$	0.3363	1.345144	1.68143	2.017716	2.354001
$wsa^{59}$	0.3363	1.345003	1.681253	2.017504	2.353754
Group consensus		4.0000	4.8363	5.8363	6.6725

**Table 8**  
All matrixes of preference relation.

$G_1 P_1$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$	$a_9$	$a_{10}$	$a_{11}$	$a_{12}$	$a_{13}$
$a_1$	0.50	0.71	0.38	0.46	0.67	0.63	0.42	0.33	0.54	0.83	0.58	0.75	0.79
$a_2$	0.29	0.50	0.17	0.25	0.46	0.42	0.21	0.13	0.33	0.63	0.38	0.54	0.58
$a_3$	0.63	0.83	0.50	0.58	0.79	0.75	0.54	0.46	0.67	0.96	0.71	0.88	0.92
$a_4$	0.54	0.75	0.42	0.50	0.71	0.67	0.46	0.38	0.58	0.88	0.63	0.79	0.83
$a_5$	0.33	0.54	0.21	0.29	0.50	0.46	0.25	0.17	0.38	0.67	0.42	0.58	0.63
$a_6$	0.38	0.58	0.25	0.33	0.54	0.50	0.29	0.21	0.42	0.71	0.46	0.63	0.67
$a_7$	0.58	0.79	0.46	0.54	0.75	0.71	0.50	0.42	0.63	0.92	0.67	0.83	0.88
$a_8$	0.67	0.88	0.54	0.63	0.83	0.79	0.58	0.50	0.71	1.00	0.75	0.92	0.96
$a_9$	0.46	0.67	0.33	0.42	0.63	0.58	0.38	0.29	0.50	0.79	0.54	0.71	0.75
$a_{10}$	0.17	0.38	0.04	0.13	0.33	0.29	0.08	0.00	0.21	0.50	0.25	0.42	0.46
$a_{11}$	0.42	0.63	0.29	0.38	0.58	0.54	0.33	0.25	0.46	0.75	0.50	0.67	0.71
$a_{12}$	0.25	0.46	0.13	0.21	0.42	0.38	0.17	0.08	0.29	0.58	0.33	0.50	0.54
$a_{13}$	0.21	0.42	0.08	0.17	0.38	0.33	0.13	0.04	0.25	0.54	0.29	0.46	0.50
:													
:													
$G_{13} P_{59}$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$	$a_9$	$a_{10}$	$a_{11}$	$a_{12}$	$a_{13}$
$a_1$	0.50	0.25	0.58	0.38	0.21	0.33	0.54	0.42	0.63	0.17	0.46	0.13	0.29
$a_2$	0.75	0.50	0.83	0.63	0.46	0.58	0.79	0.67	0.88	0.42	0.71	0.38	0.54
$a_3$	0.42	0.17	0.50	0.29	0.13	0.25	0.46	0.33	0.34	0.08	0.38	0.04	0.21
$a_4$	0.63	0.38	0.71	0.50	0.33	0.46	0.67	0.54	0.75	0.29	0.58	0.25	0.42
$a_5$	0.79	0.54	0.83	0.67	0.50	0.63	0.83	0.71	0.92	0.46	0.75	0.42	0.58
$a_6$	0.67	0.42	0.75	0.54	0.38	0.50	0.71	0.58	0.79	0.33	0.63	0.29	0.46
$a_7$	0.46	0.21	0.54	0.33	0.17	0.20	0.50	0.38	0.58	0.13	0.42	0.08	0.25
$a_8$	0.58	0.33	0.67	0.46	0.29	0.42	0.63	0.50	0.71	0.25	0.54	0.21	0.38
$a_9$	0.38	0.13	0.46	0.25	0.08	0.21	0.42	0.29	0.50	0.04	0.33	0.00	0.17
$a_{10}$	0.83	0.58	0.92	0.71	0.54	0.67	0.88	0.75	0.96	0.50	0.79	0.46	0.63
$a_{11}$	0.54	0.29	0.63	0.42	0.25	0.38	0.58	0.46	0.67	0.21	0.50	0.17	0.33
$a_{12}$	0.88	0.63	0.96	0.75	0.58	0.71	0.92	0.79	1.00	0.54	0.83	0.50	0.67
$a_{13}$	0.71	0.46	0.79	0.58	0.42	0.54	0.75	0.63	0.83	0.38	0.67	0.33	0.50

Based on the result in Table 9, the preference ordering over 13 QoS attributes for each sub-group analysed by QGNDD can be represented as below.

$$\begin{aligned}
 o_{G_1}^c &= \{a_7, a_8, a_3, a_4, a_6, a_9, a_1, a_5, a_{11}, a_{10}, a_2, a_{12}, a_{13}\} \\
 o_{G_2}^c &= \{a_8 = a_{10}, a_5, a_{11}, a_6 = a_9, a_3, a_4, a_{12}, a_7, a_{13}, a_1, a_2\} \\
 &\dots\dots\dots \\
 o_{G_{13}}^c &= \{a_4 = a_{10}, a_2, a_{13}, a_8, a_5, a_{12}, a_6, a_7, a_{11}, a_3, a_9, a_1\}
 \end{aligned}
 \tag{22}$$

In (22), the preference order over 13 QoS attributes for sub-group  $G_1$ ,  $o_{G_1}^c$ , can be explicitly identified by QGNDD. For  $o_{G_{13}}^c$ , the preference order for  $a_4$  is the same as  $a_{10}$ , after they were analyzed by QGNDD. In the case of  $o_{G_2}^c$ , QGNDD can distinguish most of the attributes by ordering them, but the preference order for  $a_8$  is the same as for  $a_{10}$  and  $a_6$  is the same as  $a_9$ . Both pairs of QoS attributes  $(a_8, a_{10})$  and  $(a_6, a_9)$  were further analyzed by QGDD, then  $a_{10} > a_8$  and  $a_6 = a_9$ . Further analysis by QGDD on the preference order for each sub-group can be obtained as follows:

**Table 9**  
QoS preference order analysis via QGNDD/QGDD.

Group 1	QGNDD	UND Occurs	QGDD
$a_1$	0.923	$a_2, a_5, a_{10}, a_{11}, a_{12}, a_{13}$	0.506
$a_2$	0.788	$a_{12}, a_{13}$	0.406
$a_3$	0.992	$a_1, a_2, a_4, a_5, a_6, a_9, a_{10}, a_{11}, a_{12}, a_{13}$	0.613
$a_4$	0.966	$a_1, a_2, a_5, a_6, a_9, a_{10}, a_{11}, a_{12}, a_{13}$	0.557
$a_5$	0.912	$a_2, a_{10}, a_{11}, a_{12}, a_{13}$	0.495
$a_6$	0.957	$a_1, a_2, a_5, a_9, a_{10}, a_{11}, a_{12}, a_{13}$	0.543
$a_7$	1.000	$a_1, a_2, a_3, a_4, a_5, a_6, a_8, a_9, a_{10}, a_{11}, a_{12}, a_{13}$	0.647
$a_8$	0.998	$a_1, a_2, a_3, a_4, a_5, a_6, a_9, a_{10}, a_{11}, a_{12}, a_{13}$	0.638
$a_9$	0.956	$a_1, a_2, a_5, a_6, a_{10}, a_{11}, a_{12}, a_{13}$	0.541
$a_{10}$	0.841	$a_2, a_{12}, a_{13}$	0.440
$a_{11}$	0.864	$a_2, a_{10}, a_{12}, a_{11}$	0.457
$a_{12}$	0.697	$a_{13}$	0.351
$a_{11}$	0.633	No UND Occur	0.317
:			
:			
:			
Group 13	QGNDD	UND Occurs	QGDD
$a_1$	0.640	No UND Occur	0.321
$a_2$	0.977	$a_1, a_3, a_5, a_6, a_7, a_8, a_9, a_{11}, a_{12}, a_{13}$	0.606
$a_3$	0.747	$a_1, a_9$	0.380
$a_4$	1.000	$a_1, a_2, a_3, a_5, a_6, a_7, a_8, a_9, a_{10}, a_{11}, a_{12}, a_{13}$	0.681
$a_5$	0.924	$a_1, a_3, a_7, a_9, a_{11}, a_{12}$	0.516
$a_6$	0.894	$a_1, a_3, a_7, a_9, a_{11}$	0.486
$a_7$	0.858	$a_1, a_3, a_9, a_{11}$	0.456
$a_8$	0.935	$a_1, a_3, a_5, a_6, a_7, a_9, a_{11}, a_{12}$	0.531
$a_9$	0.726	$a_1$	0.366
$a_{10}$	1.000	$a_1, a_3, a_5, a_6, a_7, a_8, a_9, a_{11}, a_{12}, a_{13}$	0.681
$a_{11}$	0.794	$a_1, a_3, a_9$	0.411
$a_{12}$	0.908	$a_1, a_3, a_6, a_7, a_9, a_{11}$	0.501
$a_{11}$	0.963	$a_1, a_3, a_5, a_6, a_7, a_8, a_9, a_{11}, a_{12}$	0.576

$$\begin{aligned}
 o_{G_1}^c &= \{a_7, a_8, a_3, a_4, a_6, a_9, a_1, a_5, a_{11}, a_{10}, a_2, a_{12}, a_{13}\} \\
 o_{G_2}^c &= \{a_{10}, a_8, a_5, a_{11}, a_6 = a_9, a_3, a_4, a_{12}, a_7, a_{13}, a_1, a_2\} \\
 &\dots\dots\dots \\
 o_{G_{13}}^c &= \{a_4 = a_{10}, a_2, a_{13}, a_8, a_5, a_{12}, a_6, a_7, a_{11}, a_3, a_9, a_1\}
 \end{aligned} \tag{23}$$

4.1.3. Marketing web service

After the sub-groups have been identified and all the opinions have been allocated into the appropriate groups, it means that the value of CDC for each group is within an acceptable range. In addition, each sub-groups’ consensus preference order has been reached. The providers can look up these profiles to advertise their services by registering their services with UDDI. Therefore, the system is ready for recommending the services. Assume Consumer003 in sub-group  $G_1$  requires a suitable hotel booking web service based on his/her disposition on QoS. According to the result of RMGDP analysis for sub-group  $G_1$ , the preference order over QoS attributes is:  $o_{G_1}^c = \{a_7, a_8, a_3, a_4, a_6, a_9, a_1, a_5, a_{11}, a_{10}, a_2, a_{12}, a_{13}\}$ . In other words, the preference order is:

Accuracy > Integrity > Scalability > Capacity > Exception Handling > Accessibility > Performance > Robustness  
> Interoperability > Availability > Reliability > Security > Friendly GUI (Network Related QoS Requirement).

Assume there are 831 hotels booking web services available, so these are satisfied with functional requirements. These web services will be further analysed by the 13 QoS attributes according to the disposition of sub-group  $G_1$  on each QoS attribute from the most preferable QoS attribute “Accuracy” to the least preferable QoS attribute “Friendly GUI”. The inappropriate web services in these 831 will be filtered out according to the order of QoS preference. Table 10 illustrates the filtering process. In the end of this process, the system only recommends those services that meet the QoS conditions given by Consumer003. In this case, only 7 web services that are satisfied with the consumer’s functional and non-functional requirements can be recommended for selection to form a composite service.

4.2. Process of FMG-QCMA moderation

The eight new fuzzy QoS opinions  $wsa_{S_Q}^{61}, wsa_{S_Q}^{62}, \dots, wsa_{S_Q}^{68}$  and three feedback messages with the value  $E\_Not\_Sim$  from web service consumers,  $wsa_{S_Q}^{25(G_1)}, wsa_{S_Q}^{50(G_1)}$  and  $wsa_{S_Q}^{58(G_1)}$ , are processed by FMG-QCMA. Through similarity analysis

**Table 10**  
The sample scenario about recommending web services.

Preferred QoS attribute	Group consensus on QoS	Fuzzy expression	No. of services via filtering
Accuracy ( $a_7$ )	(5.6, 6.4, 7.4, 8.2)	93% ~ 98%	831 → 470
Integrity ( $a_8$ )	(5.4, 6.1, 7.1, 7.9)	Rank (1 ~ 10): 6.1 ~ 7.1	470 → 198
Scalability ( $a_3$ )	(5.7, 6.5, 7.5, 8.2)	Rank (1 ~ 10): 6.5 ~ 7.5	198 → 87
Capacity ( $a_4$ )	(5.6, 6.2, 7.2, 7.9)	Rank (1 ~ 10): 6.2 ~ 7.2	87 → 38
Exception Handling ( $a_6$ )	(5.4, 6.1, 7.1, 7.8)	71% ~ 83%	38 → 24
Accessibility ( $a_9$ )	(4.8, 5.5, 6.5, 7.3)	Rank (1 ~ 10): 5.5 ~ 6.5	24 → 19
Performance ( $a_1$ )	(4.8, 5.7, 6.7, 7.6)	0.7 sec ~ 2.1 sec	19 → 14
Robustness ( $a_5$ )	(5.3, 6.0, 7.0, 7.8)	Rank (1 ~ 10): 6.0 ~ 7.0	14 → 13
Interoperability ( $a_{11}$ )	(4.6, 5.4, 6.4, 7.2)	Rank (1 ~ 10): 5.4 ~ 6.4	13 → 12
Availability ( $a_{10}$ )	(5.0, 5.7, 6.7, 7.5)	Rank (1 ~ 10): 5.7 ~ 6.7	12 → 11
Reliability ( $a_2$ )	(4.9, 5.7, 6.7, 7.4)	89% ~ 97%	11 → 10
Security ( $a_{12}$ )	(3.8, 4.5, 5.5, 6.3)	Transaction fault rate 0.089% ~ 0.038%	10 → 8
Friendly GUI ( $a_{13}$ )	(3.9, 4.6, 5.6, 6.3)	Rank (1 ~ 10): 4.6 ~ 5.6	8 → 7
Conclusion → 7 web services will be recommended			

**Table 11**  
The re-clustering with later fuzzy QoS opinions via *Clustering Verification*.

		$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$	$a_9$	$a_{10}$	$a_{11}$	$a_{12}$	$a_{13}$	sim_result	
$wsa^{61}$	Sim <sup>01_61</sup>	0.721	0.376	0.429	0.938	0.476	0.577	0.846	0.865	0.396	0.607	0.813	0.593	0.395	0.761	To $G_1$
$wsa^{62}$	Sim <sup>13_62</sup>	0.205	0.510	0.846	0.621	0.813	0.577	0.938	0.426	0.857	0.451	0.462	0.592	0.568	0.618	To $G_4$
$wsa^{63}$	Sim <sup>11_62</sup>	0.462	0.769	0.529	0.375	0.635	0.793	0.923	0.214	0.800	0.636	0.317	0.705	0.846	0.676	To $G_3$
$wsa^{64}$	Sim <sup>13_64</sup>	0.615	0.923	0.574	0.862	0.692	0.326	0.649	0.651	0.308	1.100	0.282	0.456	0.846	0.910	To $G_4$
$wsa^{65}$	Sim <sup>27_63</sup>	0.781	0.857	0.271	0.933	0.646	0.692	0.361	0.800	0.862	0.519	0.544	0.839	0.643	1.286	To $G_2$
$wsa^{66}$	Sim <sup>13_66</sup>	0.527	0.777	0.933	0.554	0.586	0.593	0.869	0.714	0.346	0.538	0.396	0.533	0.497	0.633	To $G_4$
$wsa^{67}$	Sim <sup>11_67</sup>	0.323	0.271	1.000	0.500	0.793	0.577	0.706	0.791	0.692	0.538	0.346	0.543	0.308	0.731	To $G_3$
$wsa^{68}$	Sim <sup>01_62</sup>	0.846	0.582	0.325	0.564	0.692	0.492	0.615	0.889	0.725	0.757	0.568	0.846	0.423	0.912	To $G_1$

with (16) in operation *Clustering Verification*, the eight new fuzzy QoS opinions were processed in sequence with the thirteen groups (the first fuzzy QoS opinion of each sub-group shown in Table 5, such as  $wsa_{S_Q}^{1(G_1)}$ ,  $wsa_{S_Q}^{10(G_2)}$ ,  $wsa_{S_Q}^{11(G_3)}$ , ...,  $wsa_{S_Q}^{59(G_{13})}$ ) and the sub-group(s) to which each new fuzzy QoS opinion is allocated are illustrated in Table 11.

That is,  $wsa_{S_Q}^{61}$  and  $wsa_{S_Q}^{68}$  are allocated to sub-group  $G_1$ ;  $wsa_{S_Q}^{63}$  and  $wsa_{S_Q}^{67}$  become a member of sub-group  $G_3$ ;  $wsa_{S_Q}^{62}$ ,  $wsa_{S_Q}^{64}$  and  $wsa_{S_Q}^{65}$  are assigned to sub-group  $G_4$ ; and  $wsa_{S_Q}^{65}$  belongs to sub-group  $G_8$ .

For the three feedback messages which are associated with  $wsa_{S_Q}^{25(G_1)}$ ,  $wsa_{S_Q}^{50(G_1)}$  and  $wsa_{S_Q}^{58(G_1)}$  about inappropriate service recommendation, an event “re-clustering” is triggered to activate the operation *Groups Clustering* because the  $m\_threshold\_distortion$  flag is true. As a result, the similarity threshold,  $\tilde{d}_{S_Q}$ , was moderated from (0.50, 0.60) to (0.52, 0.62), which can be denoted as  $\tilde{d}'_{S_Q}$  by the operation *Clustering Verification*. With the moderated similarity threshold  $\tilde{d}'_{S_Q}$ , all the sixty-eight fuzzy QoS opinions are re-clustered and the new results for AM, AAD, RAD, CDC, Group Consensus and Group Preference order over QoS attributes are obtained through FMGSAM and RMGDP accordingly.

#### 4.3. Precision and efficiency – FMG-QCMA vs. QCMA

In the following experiments, we attempt to analyze the differences between FMG-QCMA and QCMA in terms of precision in similarity analysis and efficiency in operation. The estimated approaches to generate Agreement Matrix from both methods will be evaluated, as they are the most critical processes in the frameworks.

The Agreement Matrix Generation in QCMA QCMA which adopts a single group analysis approach can be expressed as in Fig. 6.

FMG-QCMA adopts multiple sub-groups analysis approach to generate multi-group agreement matrix tables. So, the differences of these two approaches are summarised in Table 12.

According to Table 12, FMGSAM produces better similarity than the results that QCMA.

Regarding the efficiency, the number of computational operations for generating AM in both SAM and FMGSAM can be summarised as in Table 13.

According to Table 13, FMG-QCMA also has better operation efficiency than QCMA. In this case, it reduces the computational complexity by 60.8%. The effort, however, in forming the clusters is not taken account.

#### 4.4. Scalability of FMG-QCMA

The scalability of FMG-QCMA is tested with three cases with different numbers of fuzzy QoS opinions sets (i.e. 60, 40, and 20 consumers) on the same number of QoS attributes (13). Fig. 7 shows the numbers of the required steps to compute

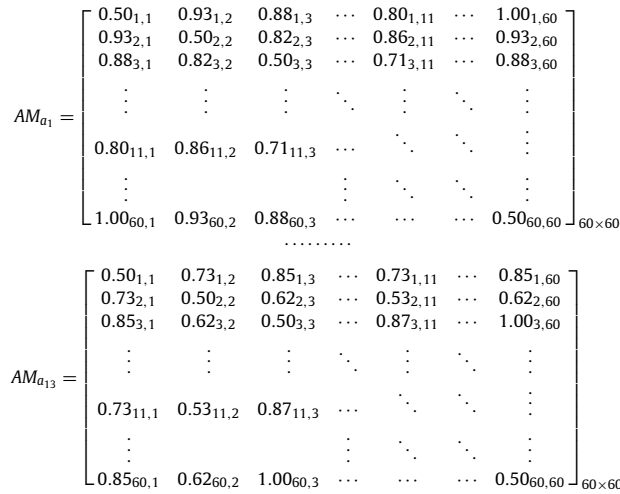


Fig. 6. AMs generation via QCMA.

Table 12  
Similarity comparison between FMGSAM and SAM.

Lowest similarity	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$	$a_9$	$a_{10}$	$a_{11}$	$a_{12}$	$a_{13}$
$(AM_1)_{FMGSAM}$	0.47	0.53	0.56	0.53	0.44	0.61	0.77	0.71	0.56	0.42	0.56	0.44	0.47
$(AM)_{SAM}$	0.39	0.47	0.50	0.50	0.44	0.44	0.56	0.50	0.44	0.37	0.50	0.41	0.47
Improvement	21%	12%	11%	6%	0%	38%	38%	41%	29%	14%	13%	6%	0%

Lowest similarity	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$	$a_9$	$a_{10}$	$a_{11}$	$a_{12}$	$a_{13}$
$(AM_{13})_{FMGSAM}$	1.00	0.93	0.64	0.88	0.73	0.69	0.79	0.75	0.69	0.78	0.93	0.67	0.92
$(AM)_{SAM}$	0.39	0.47	0.50	0.50	0.44	0.44	0.56	0.50	0.44	0.37	0.50	0.41	0.47
Improvement	157%	96%	29%	76%	65%	56%	41%	50%	58%	111%	86%	62%	96%

Table 13  
Efficiency comparison between FMGSAM and SAM.

FMGQCMA vs. QCMA	The calculation of operational computation	Total counts
AM (FMGSAM)	$((29 \times 29)_{G_1} + (12 \times 12)_{G_2} + (8 \times 8)_{G_3} + (13 \times 13)_{G_4} + (7 \times 7)_{G_5} + (12 \times 12)_{G_6} + (5 \times 5)_{G_7} + (4 \times 4)_{G_8} + (6 \times 6)_{G_9} + (2 \times 2)_{G_{10}} + (5 \times 5)_{G_{11}} + (2 \times 2)_{G_{12}} + (3 \times 3)_{G_{13}}) \times 13$	18,343
AM (SAM)	$(60 \times 60) \times 13$	46,800
Improvement of operational computation		60.8%

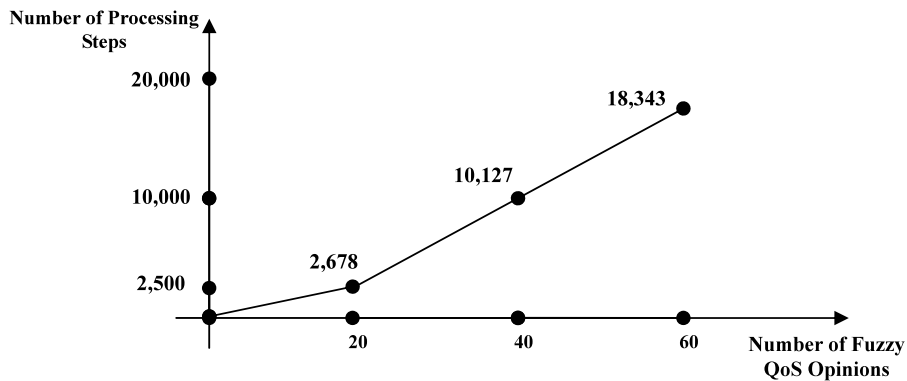


Fig. 7. The scalability test of FMG-QCMA.

the clusters using FMG-QCMA. The opinion size 20 requires 2678 processing steps in order to reach the result. The opinion collection size 40 needs 10 127 steps to complete the whole process. When the number of opinions is 60, it increases to 18 343 steps. This simple test suggests that in practice the clustering process will have a computation time growth rate that is better than  $O(n^2)$  and hence will be scalable.



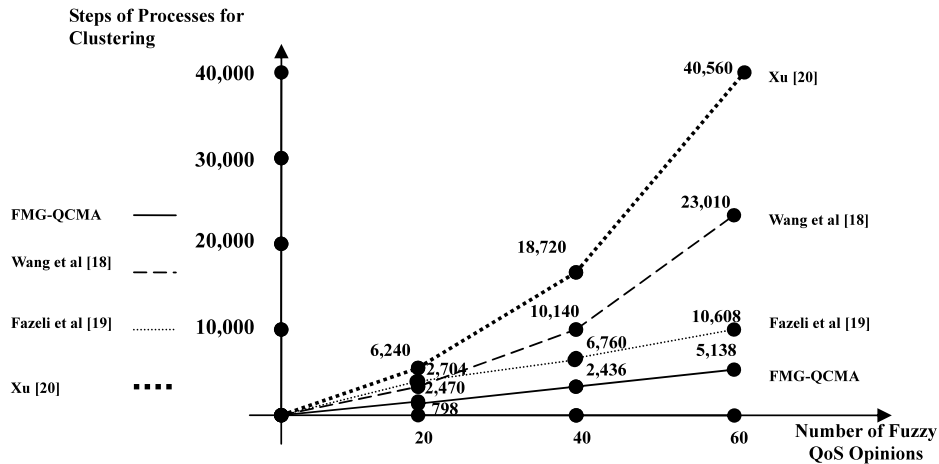


Fig. 8. The comparative analysis among FMG-QCMA, Wang et al. [18], Fazeli et al. [19] and Xu [20].

#### 4.5. Comparative analysis between FMG-QCMA and the existing solutions

In this section, we have made a rough comparison of the proposed approach with the existing ones (Wang et al. [18], Fazeli et al. [19], and Xu [20]) in terms of system complexity, by carrying out a number of experiments. It has been necessary to alter some parameters in FMG-QCMA before the experiments, due to the following reasons:

1. The existing classification approaches do not deal with fuzzy opinions, so the fuzzy representation cannot be processed by these approaches.
2. Introducing the weight distribution to the attributes in FMG-QCMA has significant differences from the existing solutions.

Nevertheless, a comparative analysis between FMG-QCMA and the existing solutions (Wang et al. [18], Fazile et al. [19], Xu [20]) in terms of system complexity (operational computation) for object clustering has been performed with the modifications of opinion representation and the removal of weighting distribution on the attributes in FMG-QCMA. Three cases with different sizes of opinions (i.e. 20, 40 and 60) were used to test the complexity of clustering. Fig. 8 shows that the number of processes that is required to cluster four different sizes of opinion sets. The processing steps required in the approaches proposed by Xu and Wang increase roughly quadratically, when the number of opinions increases linearly. Fazile's approach requires less computational steps in comparison with the other two, but it is still higher than FMG-QCMA. Therefore, the proposed approach has outperformed the others in complexity analysis.

## 5. Conclusion

FMG-QCMA is a marketing web service mechanism based on multi-groups fuzzy QoS disposition consensus of participants. The different weightings over QoS attributes and the relationship among these attributes have been taken into account in order to facilitate the consumers to reach a consensus. The service providers can utilize this result to design and market their services. The approach is a two-layers learning mechanism. In the first layer, the agreement co-efficiency index was used to evaluate the quality of grouping. The initial parameters for arbitrary group boundaries can be adjusted according to the feedback from the group agreement co-efficient. The second learning layer is based on the feedback from the users in order to adjust the number of groups. When the system received too many unsatisfactory recommendations, this implies the grouping is not appropriate and a change of boundaries cannot resolve this issue. So, the number of groups likely needs to increase.

The paper also reports its improvements on QCMA in terms of similarity measurement and system efficiency. The FMGSAM achieve higher similarity, as it adopts an effective multi-groups opinions clustering according to service consumers' QoS disposition. It also achieves higher efficiency, as its improvement in efficiency is evident as shown in Table 13.

The future work of this research will aim at exploring more aspects of perception handling for web service selection. In addition, the clusters of customers based on the multi-groups consensus on opinions and preference order can be introduced to form consumer coalitions.

## Acknowledgment

We are very grateful for anonymous reviewer's valuable comments on the earlier version of this paper. This work is partially funded by the EU FP7-ICT under Agreement no. 224609.

## Appendix A. Algorithm Fuzzy Clustering

### Algorithm Fuzzy\_Clustering( $WSA_{S_Q}$ ).

[The algorithm assumes the definitions:  $K$ , the set of consumers;  $wsa_{S(Q)}^k$ , the set of trapezoidal opinions for consumer,  $k$ , over the set of attributes,  $S(Q)$ ;  $WSA_{S(Q)}$  the collection of all the  $wsa_{S(Q)}^k$ , and,  $G_p$  is a subset of  $K$  containing the consumers in cluster,  $p$ .  $G_p.Abs\_Sim$  is the set of trapezoidal number sets  $\{wsa_{S(Q)}^k \mid k \in G_p; k \text{ is in the part of cluster } p \text{ not in any other } G_q.Abs\_Sim \ q < > p\}$ .  $G_p.Fuz\_Sim$  is those sets of sets for which the corresponding consumers belong to  $G_p$  and belong to another cluster.]

/\* Input: (1) the set of all trapezoidal numbers representing opinions of all consumers on all attributes,  $WSA_{S(Q)}$ ;  
(2) the preference order for each consumer for all attributes;  
(3) the threshold pair  $(d^u, d^l)$ ;

Output: A set of groups of consumers, for each group there is: a group representative consumer; an inner set of closely linked consumers; an outer set of consumers more weakly associated with the group representative. \*/

```

1.  $WSA\_temp_{S_Q} \leftarrow WSA_{S_Q}$ ; /* Copy all incoming opinions into a temporary set for clustering.
2.  $p \leftarrow 0$ ; /*  $p$  is set as subgroup ID and initialized as 0
3. while  $WSA\_temp_{S_Q}$  is not empty /* Clustering Loop for a created group.
    /* Find a consumer 'j' not already in the inner group for an existing cluster */
4.  $j \leftarrow \min\{k \mid k \in K, wsa_{S_Q}^k \in WSA\_temp_{S_Q} \text{ and } wsa_{S_Q}^k \notin G_{p'}.Abs\_Sim, G_{p'} \in \text{all created } G_p\}$ .
5.  $p \leftarrow p + 1$ ; /* Set Subgroup ID.
6.  $wsa_{S_Q}^{G_p} \leftarrow wsa_{S_Q}^j$ ; /* Set group centre for  $G_p$  with the vector for the selected consumer.
7.  $WSA\_temp_{S_Q} \leftarrow WSA\_temp_{S_Q} - \{wsa_{S_Q}^j\}$ ;
    /* Remove the consumer's opinion vector  $wsa_{S_Q}^j$  from the list of potential consumers. */
8.  $cluster\_temp_{S_Q} \leftarrow WSA\_temp_{S_Q}$ ; /* Initialise the pool of potential consumers to select new cluster. */
9.  $G_p.Abs\_Sim \leftarrow \{wsa_{S_Q}^{G_p}\}$ ; /* Insert group centre to "Similar Area" in  $G_p$ .
10.  $n_t^{G_p} \leftarrow 1$ ; /* Initialize  $n_t^{G_p}$ : no. of  $wsa_{S_Q}^j$  in  $G_p$ .
11. while  $cluster\_temp_{S_Q}$  is not empty /* Cluster all evaluated opinions in set for comparison.
12.  $j \leftarrow \min\{k \mid k \in K, wsa_{S_Q}^k \in cluster\_temp_{S_Q}\}$ ;
13. select  $wsa_{S_Q}^j$  in  $cluster\_temp_{S_Q}$ ;
    /* Construct vector  $Sim_{S_Q}^{G_p j}$  using vector  $wsa_{S_Q}^{G_p}$  and preference values for the two consumers. */
14. if  $SimVerify(Sim_{S_Q}^{G_p j}, " \geq ", \tilde{d}_{S_Q}) > 0$  then /*  $Sim_{S_Q}^{G_p j} \gtrsim \tilde{d}_{S_Q}$ .
15.  $n_t^{G_p} \leftarrow n_t^{G_p} + 1$ ;
16. if  $SimVerify(Sim_{S_Q}^{G_p j}, " < ", \tilde{d}_{S_Q}) > 0$  then /*  $Sim_{S_Q}^{G_p j} \gtrsim \tilde{d}_{S_Q}$ ,  $wsa_{S_Q}^j$  should be clustered.
17.  $G_p.Abs\_Sim \leftarrow G_p.Abs\_Sim + \{wsa_{S_Q}^j\}$ ; /* Insert evaluated opinion into "Similar Area" in  $G_p$ .
18.  $WSA\_temp_{S_Q} \leftarrow WSA\_temp_{S_Q} - \{wsa_{S_Q}^j\}$ ; /* Remove the evaluated opinion due to step 17.
19. else
20.  $G_p.Fuz\_Sim \leftarrow G_p.Fuz\_Sim + \{wsa_{S_Q}^j\}$ ; /* Insert evaluated opinion into "like Similar Area" but the evaluated
    opinion will be kept for next round.
21. endif /* if  $(Sim_{S_Q}^{G_p j} \gtrsim \tilde{d}_{S_Q})$ .
22. endif /* if  $(Sim_{S_Q}^{G_p j} \gtrsim \tilde{d}_{S_Q})$ .
23.  $cluster\_temp_{S_Q} \leftarrow cluster\_temp_{S_Q} - \{wsa_{S_Q}^j\}$ ; /* Remove  $wsa_{S_Q}^j$  from the evaluation for comparison.
24. end while  $cluster\_temp_{S_Q}$  is not empty /* Go evaluation for next opinion.
25. end while  $WSA\_temp_{S_Q}$  is not empty /* Go to next clustered group.
26. end Algorithm Fuzzy_Clustering( $WSA_{S_Q}$ );

```

## Appendix B. Algorithm SimVerifier

### Algorithm Sim Verifier( $Sim_{S_Q}^{jk}$ , $sim\_operator$ , $\tilde{d}_{S_Q}$ ).

/\*  $Sim_{a(i)}^{jk}$  and  $so_{a(i)}^{jk}$  are defined in 3.3 and (16) introduces a weighted similarity vector representing the similarity of a pair of consumers across all the attributes. The following algorithm uses the components of the vector (16) to determine if the overall similarity satisfies a particular cluster inclusion condition. \*/

```

/* Inputs: (1) similarity vector,  $Sim_{S(Q)}^{jk}$ ; (2) cluster condition,  $\succcurlyeq$ ,  $\succ$ , or  $\cong$ ; (3) threshold pair,  $(d^u, d^l)$ ,
Returns an indication of satisfaction of condition */
/* Normally applied when one of the consumers is a cluster centre. */
/*  $sim\_result$ : an indicator for the similarity verification by comparison between  $Sim_{S(Q)}^{jk}$  and  $\tilde{d}_{S(Q)}$ .
1.  $sim\_result \leftarrow 0$ ; /* Initialize  $sim\_result$  as 0.
/* Do similarity comparison over 13 QoS attributes and convert to  $sim\_result$  for further analysis.
2. for  $i = 1$  to 13
/* Add to  $sim\_result$  for weighted similarity components above upper threshold and subtract from  $sim\_result$  for the
weighted components below the lower threshold */
3. if  $(so_{a_i}^{jk} \times Sim_{a_i}^{jk}) > d_{S(Q)}^u$  then  $sim\_result \leftarrow sim\_result + |so_{a_i}^{jk} \times Sim_{a_i}^{jk} - d_{S(Q)}^u|$ ;
4. if  $(so_{a_i}^{jk} \times Sim_{a_i}^{jk}) < d_{S(Q)}^l$  then  $sim\_result \leftarrow sim\_result - |so_{a_i}^{jk} \times Sim_{a_i}^{jk} - d_{S(Q)}^l|$ ;
5. end for  $i = 1$  to 13
/* Theoretically  $(-13 * d_{S(Q)}^l) \leq sim\_result \leq (13 * (1 - d_{S(Q)}^u))$  and for significant similarity would expect a positive
total value. Aug is a variable to augment a value to become distinguishable. In this case 3 is sufficient.
6.  $Aug = 3$ ;  $sim\_result \leftarrow Aug \times (sim\_result/13)$ ;
7. Case  $sim\_operator$  of
8. " $\succcurlyeq$ ": /*  $(Sim_{S(Q)}^{jk} \succcurlyeq \tilde{d}_{S(Q)})$  is recognized.
9. if  $(f_{c-S(Q)}^l \leq sim\_result \leq 1)$  then return  $(sim\_result)$  else return  $(-1)$ ;
10. " $\succ$ ": /*  $(Sim_{S(Q)}^{jk} \succ \tilde{d}_{S(Q)})$  is recognized.
11. if  $(f_{c-S(Q)}^u \leq sim\_result \leq 1)$  then return  $(sim\_result)$  else return  $(-1)$ ;
12. " $\cong$ ": /*  $(Sim_{S(Q)}^{jk} \cong \tilde{d}_{S(Q)})$  is recognized.
13. if  $(f_{c-S(Q)}^l \leq sim\_result < f_{c-S(Q)}^u)$  then return  $(sim\_result)$  else return  $(-1)$ ;
14. " $\succcurlyeq$ ": /*  $(Sim_{S(Q)}^{jk} \succcurlyeq \tilde{d}_{S(Q)})$  is recognized.
15. if  $(0 \leq sim\_result < f_{c-S(Q)}^u)$  then return  $(sim\_result)$  else return  $(-1)$ ;
16. " $\succ$ ": /*  $(Sim_{S(Q)}^{jk} \succ \tilde{d}_{S(Q)})$  is recognized.
17. if  $(0 \leq sim\_result < f_{c-S(Q)}^l)$  then return  $(sim\_result)$  else return  $(-1)$ ;
18. end Case; /*  $sim\_operator$ 
19. End Algo. SimVerifier( $Sim_{S(Q)}^{jk}$ ,  $sim\_operator$ ,  $\tilde{d}_{S(Q)}$ );

```

### Appendix C. Algorithm Clustering Verification

#### Algorithm Clustering\_Verification( $wsa_{S(Q)}^j$ , $s\_feedback$ , $group\_ID$ ).

```

/* Identify if  $wsa_{S(Q)}^j$  was on "Similar Area" or "Like Similar Area", from the first group it was allocated. This algorithm must
be evoked after the initialized clustering process via Algorithm Fuzzy_Clustering, which is described in Appendix A, having
been completed. */
/* Input: a later incoming fuzzy QoS opinion from consumer  $j$  with later feedback  $s\_feedback$  and appointed group ID. */
1.  $p\_Sim\_Type \leftarrow GetSimType(wsa_{S(Q)}^j, group\_ID)$ ; /* Return if  $wsa_{S(Q)}^j$  is  $E\_Fail\_CDC$ ,  $E\_Fuz\_Sim$  or  $E\_Abs\_Sim$ .
2. if  $Validation(group\_ID)$  is true then /*  $group\_ID$  is valid.
/* Verify the cases of  $s\_feedback$ : Fail CDC (detecting by CDC threshold) or later mismatched feedback.
3. Case  $s\_feedback$  of
/* Verify the conditions if the CDC for  $wsa_{S(Q)}^j$  is less than the CDC threshold of evaluated clustered group.
4.  $E\_Fail\_CDC$ :
5.  $m\_count\_fdistance\_too\_long \leftarrow m\_count\_fdistance\_too\_long + 1$ ;
6. if  $m\_count\_fdistance\_too\_long \geq m\_threshold\_distortion$  then
7. if  $d_{S(Q)}^l \geq 0.02$  /* Moderate  $d_{S(Q)}^l$ .
8.  $d_{S(Q)}^l \leftarrow d_{S(Q)}^l - 0.02$ ;
9.  $d_{S(Q)}^u \leftarrow d_{S(Q)}^u - 0.02$ ;
10. Fuzzy_Clustering( $WSA_{S(Q)}$ )
11. endif /*  $d_{S(Q)}^l \geq 0.02$ .
12. endif /* if  $m\_count\_fdistance\_too\_long \geq m\_threshold\_distortion$ .
/* Verify the conditions if  $wsa_{S(Q)}^j$  was allocated into mismatched area.
13. Otherwise:
14. Case  $p\_Sim\_Type$  of
15.  $E\_Fuz\_Sim$ :

```

```

16.   if s_feedback = E_Not_Sim then
17.       m_count_fdistance_too_long ← m_count_fdistance_too_long + 1;
18.   if s_feedback = E_Abs_Sim then
19.       m_count_fdistance_too_short ← m_count_fdistance_too_short + 1;
20.   E_Abs_Sim:
21.   if (s_feedback = E_Not_Sim) or (s_feedback = E_Fuz_Sim) then
22.       m_count_fdistance_too_long ← m_count_fdistance_too_long + 1;
23.   endif /* if (s_feedback = E_Not_Sim) or (s_feedback = E_Fuz_Sim)
24.   Otherwise: /* Allocate this wsaSQi into appropriate group.
25.   for p = 1 to max_p
26.   if SimVerify(SimSQGpj, "≥", d̃SQ) > 0 then /* SimSQGpj ≥ d̃SQ.
27.     ntGp ← ntGp + 1;
28.     if SimVerify(SimSQGpj, ">", d̃SQ) > 0 then /* SimSQGpj > d̃SQ, wsaSQi should be clustered.
29.       Gp.Abs_Sim ← Gp.Abs_Sim + {wsaSQi}; /* Insert opinion into "Similar Area" in Gp.
30.       WSA_tempSQ ← WSA_tempSQ - {wsaSQi}; /* Remove the opinion due to step 17.
31.       break; /*Terminate Algorithm Clustering_Verification when just allocate wsaSQi.
32.     else
33.       Gp.Fuz_Sim ← Gp.Fuz_Sim + {wsaSQi}; /* Insert opinion into "like Similar Area" but the evaluated opinion
34.         will be kept for next round.
35.     endif /* if (SimSQGpj > d̃SQ)
36.   endif /* if (SimSQGpj ≥ d̃SQ)
37.   end for p = 1 to max_p
38.   end Case; /* p_Sim_Type
39.   /* Determine if re-clustering by moderated threshold for similarity should be enabled or not.
40.   if m_count_fdistance_too_long ≥ m_threshold_distortion then
41.     if dSQl ≥ 0.02 /* Moderate dSQl.
42.       dSQl ← dSQl - 0.02;
43.       dSQu ← dSQu - 0.02;
44.       Fuzzy_Clustering(WSASQ);
45.     endif /* dSQl ≥ 0.02
46.   endif /* if m_count_fdistance_too_long ≥ m_threshold_distortion.
47.   if m_count_fdistance_too_short ≥ m_threshold_distortion then
48.     if dSQu ≤ 0.98 /* Moderate dSQu.
49.       dSQu ← dSQu + 0.02;
50.       dSQl ← dSQl + 0.02;
51.       Fuzzy_Clustering(WSASQ);
52.     endif /* dSQu ≤ 0.98.
53.   endif /* if m_count_fdistance_too_short ≥ m_threshold_distortion.
54.   end Case; /* s_feedback.
55.   endif /* if Validation(group_ID) is true.
56.   end Algorithm Clustering_Verification(wsaSQi, s_feedback, group_ID);

```

## References

- [1] Wei-Li Lin, Chi-Chun Lo, Kuo-Ming Chao, Muhammad Younas, Consumer-centric QoS-aware selection of web service, J. Comput. System Sci. 74 (2008) 211–231.
- [2] Doo-Ywan Kim, Tae-Ywan Kim, Evaluation of classified information on web agent using fuzzy theory, Int. J. Fuzzy Syst. 5 (3) (2005) 216–221.
- [3] Media Mall, US online ad spending continues rapid growth, Media Mall 27 November 2006, [http://www.mediamall.com/blogger/2006\\_11\\_01\\_archive.html](http://www.mediamall.com/blogger/2006_11_01_archive.html).
- [4] eMarketer, An emerging consensus paints a positive picture, eMarketer 27 November 2006, <http://www.emarketer.com/Article.aspx?R=1004293>.
- [5] W3C, QoS for web service: requirements and possible approaches, W3C Working Group Note 25 November, [http://www.w3c.or.kr/kr-office/TR/2003\\_2003](http://www.w3c.or.kr/kr-office/TR/2003_2003).
- [6] Kuo-Ming Chao, Muhammad Younas, Chi-Chun Lo, Tao-Hsin Tan, Fuzzy matchmaking for web service, in: Proceedings of 19 IEEE Conference on Advanced Network and Information Application, IEEE CS, 2005, pp. 721–726.
- [7] D.A. Menasce, Composing web services: A QoS view, IEEE Internet Comput. 8 (6) (2004) 88–90.
- [8] Michael C. Jaeger, Gregor Rojec-Goldmann, Gero Mühl, QoS aggregation for service composition using workflow patterns, in: Proceedings of the 8th International Enterprise Distributed Object Computing Conference (EDOC 2004), Monterey, California, USA, IEEE CS Press, 2004, pp. 149–159.

- [9] Yutu Liu, Anne H.H. Ngu, Liangzhao Zeng, QoS computation and policing in dynamic web service selection, ACM WWW Conference, New York, USA, May 2004, pp. 66–73.
- [10] E. Herrera-Viedma, F. Herrera, F. Chiclanam, A consensus model for multiperson decision making with different preference structures, *IEEE Trans. Syst. Man, Cybern.* 32 (2002) 394–402.
- [11] F. Chiclana, F. Herrera, E. Herrera-Viedma, Integrating multiplicative preference relations in a multipurpose decision-making model based on fuzzy preference relations, *Fuzzy Sets and Systems* 122 (2001) 277–291.
- [12] F. Chiclana, F. Herrera, E. Herrera-Viedma, A classification method of alternatives for multiple preference ordering criteria based on fuzzy majority, *J. Fuzzy Math.* 34 (1996) 224–229.
- [13] F. Chiclana, F. Herrera, E. Herrera-Viedma, Integrating three representation models in fuzzy multipurpose decision making based on fuzzy preference relations, *Fuzzy Sets and Systems* 97 (1998) 33–48.
- [14] F. Herrera, et al., A rational consensus model in group decision making using linguistic assessments, *Fuzzy Sets and Systems* 88 (1997) 31–49.
- [15] A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review, *ACM Comput. Surveys* 31 (3) (1999) 264–323.
- [16] Ji-Rong Wen, Jian-Yun Nie, Hong-Jiang Zhang, Query clustering using user logs, *ACM Trans. Inform. Syst.* 20 (1) (2002) 59–81.
- [17] Ravi Kannan, Santosh Vampala, Adrian Vetta, On clustering: good, bad and spectral, *J. ACM* 51 (3) (2004) 497–515.
- [18] Haixun Wang, Wei Wang, Joing Yang, Philip S. Yu, Clustering by pattern similarity in large data sets, in: *Data Mining, Third IEEE International Conference, 2003*, pp. 187–194.
- [19] M. Fazeli, H. Sarbazi-Azad, R. Farivar, Parallel clustering on the star graph, in: *Lecture Notes in Comput. Sci.*, vol. 3719/2005, 2005, pp. 287–292.
- [20] Rui Xu, Survey of clustering algorithms, *IEEE Trans. Neural Networks* 16 (3) (2005).
- [21] L. Sadeh, Fuzzy set, *Inf. Control* 8 (1965) 338–353.
- [22] F. Höppner, F. Klawonn, R. Kruse, *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis, and Image Recognition*, Wiley, New York, 1999.
- [23] R. Hathaway, J. Bezdek, Fuzzy c-means clustering of incomplete data, *IEEE Trans. Syst., Man, Cybern.* 31 (5) (2001) 735–744.
- [24] C.-L. Huang, K.-M. Chao, C.-C. Lo, A moderated fuzzy matchmaking for web services, in: *Proceedings of The 5th International Conference on Computer and Information Technology*, Shanghai, China, IEEE CS, 2005, pp. 1116–1122.
- [25] Hsi-Mei Hsu, Chen-Tung Chen, Aggregation of fuzzy opinions under group decision making, *Fuzzy Sets and Systems* 79 (1996) 279–285.