Fuzzy Similarity Clustering for Consumer-Centric QoS-aware Selection of Web Services

Wei-Li Lin1, Chi-Chun Lo1, Kuo-Ming Chao2,3, Nick Godwin3

1 Institute of Information Management, National Chiao-Tung University, Hsin-Chu, 300, Taiwan
weilin.iim92@nctu.edu.tw, cclo@faculty.nctu.edu.tw
tel: +886-3-5712121 ext.57401
fax: +996-3-5723792
2 Software School, Fudan University, China
3 DSM Research Group, Department of Computing and the Digital Environment, Coventry University,
Coventry, CV1 5FB, UK
k.chao@coventry.ac.uk, a.n.godwin@coventry.ac.uk

ABSTRACT

Appropriate use of group consensus on service consumers’ QoS opinions can improve web service discovery. Web service participants with different backgrounds or preferences may not be easy to reach consensus on their Web service QoS opinions, so they should be treated as multi-groups. Also, to prevent to possible useful opinions from being omitted unintentionally, fuzzy clustering criteria for the “multi-groups” framework should be adopted. In this paper, we have proposed a FMGSAM (Fuzzy Multi-Groups based SAM) to improve the accuracy in group opinion similarity analysis and the efficiency in generating group consensus.

1. Introduction

QoS (Quality of Service) has been considered as a significant criterion in the selection of web services recently [1][2][3][4][5][6]. 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 (defined by W3C [1][2][7]). However, the challenges to the single group based opinion similarity and preference analysis for web service selection include:

1. If the fuzzy QoS opinions were collected from web service participants with very different backgrounds which could lead to diverse perceptions, the obtained consensus may not be effective.

2. For some outliers identified via our previous proposed approach which groups all users into one single group can be re-classified into other appropriate groups if multi-groups is adopted, as these opinions may have meaningful correlation. The omission of those outliers without further examination may not be appropriate.

In this paper, an attempt to use Multi-groups based consensus for web service selection to address the above challenges is reported. If the groups contain highly similar fuzzy QoS opinions, the analysis on the preference order among QoS attributes will be more credible.

However, multi-attributes based clustering is much complicated than single-attribute based clustering. Haixun Wang et al [8] proposed pCluster model in 2003 for pattern similarity clustering in large data sets. The pCluster model can be used for clustering similar objects which were connected by shifting or scaling their relationships in multi-dimensional space. Their work considers that all attributes are equally important, so the weight distribution over the attributes was not discussed. In our applications, the attributes cannot be treated as “equal weight” due to different perceptions.

In 2005, M Fazeli et al [9] proposed a parallel algorithm to tackle multi-features data clustering problems in multi-computer with star topology. The raw data is depicted with a feature vector v, as a set of measurements (v1, v2, ..., vM) that map to properties of a collection of data into a Euclidean space of dimension M. The squared-error algorithm is taken for the multi-features data clustering and divides N multi-features data into K clusters. However, for some disqualified data which is very close to the “boundary” (lower than the required similarity threshold but still can be considered as “certain degree of similarity” in fuzzy concept) could be omitted and this leads to possible distorted classifications.

Therefore, the challenges for building multi-groups based QoS-aware selection of web service including:

1. Associated weight on each QoS attribute should be introduced due to different preference orders given by the users. Therefore, a weighted multi-attributes QoS similarity should be defined.

2. To prevent from not intentionally removing possible meaningful data which just falls outside of the pre-defined group boundaries, the multi-attributes based clustering criteria should be formulated with fuzzy evaluation.
FMG-SAM, an extension of SAM used in QCMA framework, attempts to provide an effective scheme for multi-groups based QoS clustering for web service selection. All the incoming multi-QoSs opinions will be fuzzily clustered into different QoS group. FMG-SAM is the proposed mechanism to improve the insufficiency of SAM in the problem domains associated with multi-groups clustering with multi-attributes.

The following section in the paper including: Section 2 briefly describes the SAM in QCMA from our previous work [1]. Section 3 presents the proposed FMG-SAM. Section 4 reports on experimental results with a case study of multi-QoS attributes clustering for a number of hotel booking web services. Finally, Section 5 concludes this work and remarks on the future research direction.

2. Similarity Aggregation Method in QCMA

QCMA is employed to obtain and moderate group consensus on QoS in web services [1]. SAM was developed for resolving conflicts that arise from different consensus on QoS in web services [1]. In SAM the different fuzzy opinions will be aggregated into consensus classes so that they can be measured by their similarities to each other. The procedure to perform SAM is organized into 7 steps described as below [12].

1. Each participant represents his/her subjective fuzzy preference on each specific criterion with a positive trapezoidal fuzzy number shown as Figure 1.

![Figure 1: A Trapezoidal Fuzzy Number](image)

Each fuzzy QoS opinion donated as $w_{s_k}^a$, which is a user’s opinion $k$ in the set of all opinions $K$ on QoS attribute $a$, which follows the definition in Figure 1. $W_S$ is represented as a collection of $w_{s_k}^a$, which is formally defined as follows.

$$w_{s_k}^a = \{(x_1)_{s_k}^a, (x_2)_{s_k}^a, (x_3)_{s_k}^a, (x_4)_{s_k}^a\}$$

where $0 \leq (x_1)_{s_k}^a \leq (x_2)_{s_k}^a \leq (x_3)_{s_k}^a \leq (x_4)_{s_k}^a \leq 10$

2. The opinion similarity between $w_{s_k}^a$ and $w_{s_k}^b$, which is denoted as $Sim_{ab}$, can be obtained via the following equation:

$$Sim_{ab} = \frac{\int \min(\tilde{\mu}(w_{s_k}^a), \tilde{\mu}(w_{s_k}^b)) dx}{\int \max(\tilde{\mu}(w_{s_k}^a), \tilde{\mu}(w_{s_k}^b)) dx}$$

(2)

3. To build an Agreement Matrix (AM), which can be represented as equation (3):

$$AM_{wsa} = \begin{bmatrix}
1 & Sim_{12} & \cdots & Sim_{1n} & \cdots & Sim_{1n} \\
Sim_{12} & 1 & \cdots & \cdots & \cdots & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \cdots & 1 & Sim_{n} & \cdots \\
Sim_{1n} & Sim_{2n} & \cdots & Sim_{nn} & 1 & \cdots \\
\vdots & \vdots & \cdots & \vdots & \vdots & 1
\end{bmatrix}_{n \times n}$$

(3)

4. To calculates an average agreement degree (AAD), denoted as $A(w_{s_k}^a)$ is from equation (4):

$$A(w_{s_k}^a) = \frac{1}{n-1} \sum_{k=1}^{n} Sim_{kk}$$

(4)

5. To obtain the RAD (Relative Agreement Degree) for each individual opinion uses the following formula.

$$RAD(w_{s_k}^a) = \frac{A(w_{s_k}^a)}{\sum_{j=1}^{n} A(w_{s_j}^a)}$$

(5)

6. Assign a weighting variable, $w_k$, to each opinion.

7. Obtain the CDC (Consensus Degree Coefficient) for each participant:

$$CDC(w_{s_k}^a) = \beta \times w_k + (1 - \beta) \times RAD(w_{s_k}^a)$$

(6)
where $\beta$ is a control variable to indicate the relation between CDC and RAD. To simplify the operation of CDC, we set $\beta$ as zero so that CDC is equal to RAD.

4. The Proposed Approach - FMGSAM

The proposed FMGSAM, derived from SAM, is designed for similarity analysis under multi-groups based QoS-aware selection of web service. The challenges on clustering multi-attributes based QoS are much higher than single dimensional clustering schemes due to the difficulty of identifying appropriate the number of groups and their sizes. In addition, the weight distribution among all QoS attributes must take into account. In order to increase grouping precision and to eliminate unnecessary data loss (outliers), a fuzzy group boundary is introduced.

The system scenario for multi-groups based similarity analysis can be depicted as Figure 2. A number of opinion groups will be formed and these groups have overlapping boundaries.

\[
WSA_k = \{wsa_{k}^{a} | k \in K, a \in S_a\} \quad (7)
\]

(2) Perform QoS fuzzy clustering which allocates all $wsa_{k}^{a}$ into appropriate groups $(G_1, G_2, \ldots, G_m)$. For each pair of $wsa_{j}^{a}$ and $wsa_{k}^{a}$ in clustered $G_p$, $Sim_{j,k}^{a}$ can be obtained via following formulas:

\[
Sim_{j,k}^{a} = (so_{a}^{i} \times Sim_{j,k}^{a} \ldots, so_{a}^{i} \times Sim_{j,k}^{a}) \quad (8)
\]

where $Sim_{j,k}^{a}$ is similarity between $wsa_{j}^{a}$ and $wsa_{k}^{a}$, $so_{a}^{i}$ is the similarity in preference order between $wsa_{j}^{a}$ and $wsa_{k}^{a}$, and $q$ indicates the number of QoS attributes. By definition in W3C [2], $q = 13$. $Sim_{j,k}^{a}$ and $so_{a}^{i}$ can be obtained by following equations:

\[
Sim_{j,k}^{a} = \min \left\{ (x_j^i)^{\beta} + \int (\beta(wsa_j^a)dx) + (x_k^i)^{\beta} + \int (\beta(wsa_k^a)dx) \right\} \\
\max \left\{ (x_j^i)^{\beta} + \int (\beta(wsa_j^a)dx) + (x_k^i)^{\beta} + \int (\beta(wsa_k^a)dx) \right\} \quad (9)
\]

\[
so_{a}^{i} = q - \frac{|\mu_i - \mu_j|}{q} 
\]

(3) Determine Agreement Matrixes $(AM_p)_{a,s}$ for each clustered fuzzy QoS group $G_p$. The corresponding agreement matrixes can be represented as follows:

\[
(AM_1)_{a,s} = \begin{bmatrix}
1 & Sim_{j_0,k_0}^{a_1} & \ldots & Sim_{j_0,k_0}^{a_q} \\
Sim_{j_0}^{a_1} & 1 & \ldots & \vdots \\
& \ddots & \ddots & \ddots \\
Sim_{j_0}^{a_q} & \vdots & \ddots & 1 \\
Sim_{k_0}^{a_1} & Sim_{k_0}^{a_2} & \ldots & Sim_{k_0}^{a_q}
\end{bmatrix}_{a,s} 
\]

\[
(AM_2)_{a,s} = \begin{bmatrix}
1 & Sim_{j_0,k_0}^{a_1} & \ldots & Sim_{j_0,k_0}^{a_q} \\
Sim_{j_0}^{a_1} & 1 & \ldots & \vdots \\
& \ddots & \ddots & \ddots \\
Sim_{j_0}^{a_q} & \vdots & \ddots & 1 \\
Sim_{k_0}^{a_1} & Sim_{k_0}^{a_2} & \ldots & Sim_{k_0}^{a_q}
\end{bmatrix}_{a,s} 
\]

\[
\text{……………………………}
\]

Figure 2: System Scenario for Multi-Groups Similarity

Before FMGSAM, $n$ multi-attributes based fuzzy QoS collected at time $t$ will be used for a preliminary analysis form a number of opinion groups. Then, the detailed processes of the FMGSAM can be carried which are illustrated as the following steps:

(1) Represent Fuzzy QoS, $wsa_{k}^{a}$, as a multi-dimensional fuzzy trapezoidal number as $(wsa_{k}^{a_1}, \ldots, wsa_{k}^{a_q})$. $WSA_{a}$ is a set of $wsa_{k}^{a}$ as below:
Algorithm 4.1: Fuzzy_Comparison

1. m_c creadt : 0; /* m_credit: indicator for fuzzy comparison
2. m_result : null; /* m_result: indicator for comparison results
3. for i = 1 to 13 /* Number of QoS attributes
4. if (so(\text{Sim}) \times \text{Sim}(a)_i) > \text{d}_{a_i}^u
5. m_c creadt : m_c creadt + \text{d}_{a_i}^u - \text{Sim}(a)_i
6. end if (so(\text{Sim}) \times \text{Sim}(a)_i) > \text{d}_{a_i}^u
7. if (so(\text{Sim}) \times \text{Sim}(a)_i) < \text{d}_{a_i}^l
8. m_c creadt : m_c creadt + \text{Sim}(a)_i - \text{d}_{a_i}^l
9. end if (so(\text{Sim}) \times \text{Sim}(a)_i) < \text{d}_{a_i}^l
10. end for i = 1 to 13
11. if m_c creadt > 0.1 then /* Fuzzily larger than
12. m_result : E_Large
13. elseif (-0.1 \leq m_c creadt \leq 0.1)
14. m_result : E_Equal
15. else /* Fuzzily less than
16. m_result : E_Less
17. end if m_c creadt > 1
18. if m_operator = “≤” then
19. if (m_result=E_Large) or (m_result=E_Equal) then
20. return (true)
21. else
22. return (false)
23. endif
24. else /* Fuzzily larger than
25. if m_operator = “≥” then
26. if (m_result=E_Large) then
27. return (true)
28. else
29. return (false)
30. endif
31. endif
32. End Algorithm Fuzzy_Comparison

5. Experiment

This section describes a simplified experiment using FMGSAM to conduct the similarity analysis of hotel book web services which include 5 QoS opinions on 3 different QoS attributes. The procedures are illustrated as follows:

1. Assume 5 opinions wsa_{1}^{a}, wsa_{2}^{a}, wsa_{3}^{a}, wsa_{4}^{a}, and wsa_{5}^{a} in the set WSA_{a}. Each wsa_{k}^{a} (1 \leq k \leq 5) represents three opinions on different QoS attributes (wsa_{a_1}, wsa_{a_2}, and wsa_{a_3}, and WSA_{a} = \{a_1, a_2, a_3\}). Each opinion wsa_{k}^{a} is represented as fuzzy trapezoidal numbers defined in (1). The 5 opinions collected in WSA_{a} are defined as below:

wsa_{1}^{a} = \{(4,5,6,7), (5,6,7,8), (3,4,5,6)\};
wsa_{2}^{a} = \{(6,7,8,9), (7,8,9,10), (6,7,9,10)\};
wsa_{3}^{a} = \{(3,4,5,6), (2,3,4,5), (3,4,5,6)\};
wsa_{4}^{a} = \{(2,3,4,5), (6,7,8,9), (3,4,5,6)\};
\[ \text{wsa}_0 = \{(5,6,7,8), (7,8,9,10), (6,7,8,9)\}; \]

2. The relative position of preference order denoted as \( o^a_1 \) for each \( \text{wsa}_1 \) can be defined as follows:

\[
\begin{align*}
(o^1_{w_1}, o^2_{w_1}, o^3_{w_1}) & \quad \text{for} \quad \text{wsa}_1 = (1, 3, 2) \\
(o^1_{w_1}, o^2_{w_1}, o^3_{w_1}) & \quad \text{for} \quad \text{wsa}_2 = (3, 1, 2) \\
(o^1_{w_1}, o^2_{w_1}, o^3_{w_1}) & \quad \text{for} \quad \text{wsa}_3 = (1, 2, 3) \\
(o^1_{w_1}, o^2_{w_1}, o^3_{w_1}) & \quad \text{for} \quad \text{wsa}_4 = (2, 1, 3) \\
(o^1_{w_1}, o^2_{w_1}, o^3_{w_1}) & \quad \text{for} \quad \text{wsa}_5 = (3, 2, 1) \\
\end{align*}
\]

3. For opinions clustering, the \( \bar{d}_1 \) is initialized as:

\[ \bar{d}_1 = ((0.4, 0.6), (0.4, 0.6), (0.4, 0.6)) \]

4. Let \( \text{wsa}_1 \) be selected as the first group centre. The distance of \( \text{wsa}_1 \) to the rest opinions, similarity \( \text{Sim}_{1j} \) (2 \( \leq j \leq 5 \)), will be evaluated via equation (8) which obtains the following values:

\[ \text{Sim}_{12} = (0.75, 0.26, 0.18); \quad \text{Sim}_{13} = (0.83, 0.38, 0.67); \]
\[ \text{Sim}_{14} = (0.45, 0.29, 0.67); \quad \text{Sim}_{15} = (0.28, 0.52, 0.42) \]

5. Select \( \text{wsa}_1 \) as an input for fuzzy comparison \( \text{Sim}_{1j} \) and use \( \bar{d}_1 = ((0.4, 0.6), (0.4, 0.6), (0.4, 0.6)) \), which is defined in algorithm Fuzzy_Comparison as threshold to determine the group membership, if \( \text{Sim}_{1j} \geq \bar{d}_1 \) is satisfied.

\[ \text{Sim}_{12} \geq \bar{d}_1 \quad \text{due to} \quad \text{m.credit} = -0.21 < -0.1 \]
\[ \text{Sim}_{13} \geq \bar{d}_1 \quad \text{due to} \quad \text{m.credit} = 0.28 > 0.1 \]
\[ \text{Sim}_{14} \geq \bar{d}_1 \quad \text{due to} \quad \text{m.credit} = -0.1 \approx \text{m.credit} \approx 0.1 \]
\[ \text{Sim}_{15} \geq \bar{d}_1 \quad \text{due to} \quad \text{m.credit} = -0.12 < -0.1 \]

\[ \Rightarrow \quad \text{In this case, } \text{wsa}_3 \text{ and } \text{wsa}_4 \text{ are clustered into the first group but } \text{wsa}_5 \text{ can also be classified into the other group due to } \text{Sim}_{15} \leq \bar{d}_1 \text{ (so that wsad}_5 \text{ is still reserved in pool of opinions that are to be clustered).} \]

\[ \Rightarrow G_1 = \{ \text{wsa}_1, \text{wsa}_3, \text{wsa}_4 \} \]

6. Similar to the generation of first group started from step 4, the \( \text{wsa}_2 \) is selected as the second group centre.

The rest opinions will be evaluated using similarity \( \text{Sim}_{2j} \) (2 \( \leq j \leq 5 \)) that is equation (8) to examine if they could belong to the second group:

\[ \text{Sim}_{24} = (0.34, 0.89, 0.38); \quad \text{Sim}_{25} = (0.88, 0.67, 0.60) \]

7. Similar to Step 5, the Fuzzy_Comparison algorithms is employed and the same threshold values \( \bar{d}_1 \) are adopted to determine their m.credit values.

\[ \text{Sim}_{24} \geq \bar{d}_1 \quad \text{due to} \quad \text{m.credit} = 0.21 > 0.1 \]
\[ \text{Sim}_{25} \geq \bar{d}_1 \quad \text{due to} \quad \text{m.credit} = 0.35 > 0.1 \]

\[ \Rightarrow \quad \text{As result, wsad}_4 \text{ and wsad}_5 \text{ are qualified for the second group.} \]

\[ \Rightarrow \quad \text{So, } G_2 = \{ \text{wsa}_2, \text{wsa}_4, \text{wsa}_5 \} \text{ is obtained.} \]

8. According to the above steps, we can get \( G_1 = \{ \text{wsa}_1, \text{wsa}_3, \text{wsa}_4 \} \) and \( G_2 = \{ \text{wsa}_2, \text{wsa}_4, \text{wsa}_5 \} \). The following figure is a diagram representation of 2 different opinion groups and their memberships.

![Figure 3: The Clustering of 5 Opinions](image)

According to the above illustration, we can conclude FMGSAM is better than SAM in terms of similarity and efficiency:

1. Regarding similarity, the lowest similarity obtained by the SAM is \( \text{Sim}_{15} \) which is \( \text{Sim}_{15} = (0.23, 0.44, 0.21) \). If we use \text{m.credit} in fuzzy comparison algorithms as criterion, \( \text{Sim}_{15} \) is -0.36. However, for FMGSAM, the lowest similarity is \( \text{Sim}_{15} \) in \( G_2 \):
\[ \text{Sim}_{15} = (0.38, 0.60, 0.21) \], and its corresponding \text{m.credit} is -0.21. The improvement of similarity analysis can be obtained via the following calculation:
The research leading to these results has received funding from the European Community’s Seventh Framework Programme FP7/2007-2013 under grant agreement n° No224609.

References