Fuzzy-based Clustering of Web Services' Quality of Service: A Review

Mohd Hilmi Hasan, Jafreezal Jaafar, and Mohd Fadzil Hassan
Universiti Teknologi PETRONAS, Tronoh 31750 Perak, Malaysia
Email: mhilmi_hasan@petronas.com.my; jafreez@petronas.com.my; mfadzil_hassan@petronas.com.my

Abstract**—**Clustering has been proposed in numerous researches on web services due to its ability in enhancing computational efficiency. The objective of this paper is to present a review on web services clustering and to propose fuzzy clustering of web services' based on quality of service (QoS). A total of 49 manuscripts, collected from ISI, SCOPUS and Google Scholar indexing databases, were critically reviewed. The review revealed that there is no report has been published before regarding fuzzy-based clustering of web services' QoS. It also revealed the importance of clustering in ensuring the efficiency of web services delivery particularly in their discovery, selection and recommendation processes. The results of the proposed fuzzy clustering can be used as reference to guide requestors to subscribe suitable services. They could also be implemented as the inference components for fuzzy-based web services applications. For future work, the generated clustering results will be implemented in web services QoS monitoring application.

Index Terms**—**Clustering, fuzzy clustering, Fuzzy C-Means, QoS clustering, web services clustering.

I. Introduction

The growing number of available web services has made the procedures such as the discovery, recommendation and selection, require higher computational resources. Numerous researches have been carried out to attend this problem, which include solutions through web services categorization [1]. Categorization of web services reduces computational time and complexity in a way that the execution of procedure is performed upon the matched group only instead of the whole available services [2]. The categorization of web services can be carried out in two ways, namely clustering and classification. The former is an unsupervised method where the categorization is performed using unlabelled data and without the predefined desired output. On the other hand, the latter is a supervised method which categorizes web services using labeled data and with proper descriptions on the expected output [3], [4]. It is apparent in literature that numerous works have been proposed to solve web services categorization problems using clustering [5]-[7] and classification [8]-[10] methods. However, most of the previous works focused on categorizing web services based on functional criteria. We acknowledge that the categorization of web services based on their non-functional criteria, namely quality of service (QoS) is also equally important. As mentioned earlier, this non-functional-based web services categorization also reduces the computational time and complexity for web services processes. Moreover, it can also be used as a reference during services subscription as the requestors may identify which services are considered as good, moderate or poor, for instance. This is significantly important as in general, not all requestors can differentiate between the realistic and unrealistic QoS values when finding for suitable services [11].

We are also interested in categorizing web services' QoS fuzzily. This is due to the fact that results of fuzzy-based categorization can eventually be used as the inference components of fuzzy-based web services applications. It is apparent in literature that there is a growing number in the development of fuzzy-based web services applications in the areas of service selection [11], [12], service reputation [13] and overhead forecasting [14]. The fuzzy inference based on categorization can be a better option than deriving from expert knowledge. The main reasons are inference derivation based on expert knowledge may result in loss of accuracy [15] and normally consumes a lot of time. Moreover, in some areas, the expert knowledge may not always available [16]. Thus, inference from data has become the preferred alternative for the development of fuzzy-based applications. There are many available algorithms that can be used to perform the categorization task, but it is evident that none of them can produce an optimal solution for all kinds of data sets [17]. Hence, in this paper, we propose that the web services' QoS categorization is conducted based on clustering method using Fuzzy C-Means algorithm.

In this paper, the previous findings on web services clustering were critically reviewed. The review was conducted through a structured approach in which the manuscripts were gathered from the three well-known indexing databases namely ISI, SCOPUS and Google Scholar. As a result, 49 manuscripts were collected for review. Based on the review, it was evident that the proposed fuzzy clustering of web services' QoS that employs FCM algorithm has not been published before. Furthermore, the review also disclosed the importance of web services clustering. In this paper, we present the clustering of the actual web services' QoS data gathered

Manuscript received August 5, 2013; revised October 15, 2013.
Corresponding author email: hilmihasan@gmail.com.
doi:10.12720/jcm.9.1.81-90

from the network, which comprised three parameters namely response time, availability and latency. For each parameter, a data set of 1500 values was used in the experiments. We also present the validation results which were conducted to ensure that the clustering was optimally executed. These were conducted using Xie and Beni clustering validity index.

In a nutshell, the objectives of this paper are to present a review on the previous works that are related to web services clustering and to propose web services' QoS fuzzy clustering. The process was performed upon the actual web services data using FCM algorithm. Section 2 presents the conducted literature review. Section 3 presents the QoS data sets used in the experiments, and the clustering and validation results. Finally, section 4 contains conclusion that summarizes the outcomes of the work and outlines some potential future directions.

II. Literature Review

The review involved the findings that were taken from the three well-known indexing databases namely ISI, SCOPUS and Google Scholar. The searches from the three databases were conducted using two keywords, namely "web service AND clustering" and "web services AND clustering". The search result produced by ISI yielded 16 manuscripts, while SCOPUS and Google Scholar yielded 54 and 50 manuscripts respectively. All manuscripts were then filtered in order to remove the unrelated as well as the redundant ones. As a result, a total of 49 manuscripts were selected to be reviewed and presented in this paper.

In general, those selected papers can be categorized into three areas of study namely improvement of web services clustering algorithm, as well as web services clustering based on functional and non-functional criteria. The subsections that follow present the review on these manuscripts.

A. *Improvement of Web Services Clustering*

Ref. [18] presents an improved web services clustering method which uses Peano Space filling curve. The work argues that an effective clustering technique is needed to support web services selection and retrieval processes. The technique must have the capabilities to consider all the possible parameter matches and include multiple criteria in the processes of parameter matching. The technique proposed by the work is able to meet these requirements, and is compared with the previous work that employs Hilbert space filling curve. The results show that the proposed technique is better than the previous work in terms of fairness, scalability and irregularity.

Ref. [19] claims that cloud web services have to face performance problems and congestions due to the encoding of a huge amount of SOAP messages. Hence, the work proposes the fractal clustering model, which is able to group the SOAP messages into clusters. The clustering is performed based on similarity of the messages. The aggregation of the messages into clusters has helped to reduce the performance problems and congestions in cloud web services.

Ref. [20] proposes an intelligent web services clustering method through combination of human knowledge and artificial intelligence (AI) technique to generate a taxonomy structure. In the process, web services are transferred into standard vector format through WSDL document. In addition, the self organizing neural network based learning algorithm is also employed in order to enable the intelligence behavior of the method.

Ref. [21] claims that the filtering method of web services is time consuming due to the complexity in reasoning as well as processing of the descriptions. Hence, they propose an architecture that has the capabilities to filter and cluster the semantic web services. The architecture comprises a number of components that are grouped together to deal with one aspect of the task. The groups are then chained together that finally represents a step-by-step process of filtering.

Ref. [22] argues that web services matching processes are slow as a result of complex semantic calculations. Thus, they propose a clustering method that is based on user preferences and ontology. The proposed method is capable of removing irrelevant web services in its process, hence producing semantic calculation results with less time.

Ref. [23] suggests semi-supervised clustering algorithm for web services composition. The algorithm performs the clustering based on tags and constraints, and uses a huge number of unlabeled data in order to support the supervised learning procedure.

Ref. [24] presents a web services clustering method called WTCluster, which employs both WSDL documents and tags annotation in its operation. The work argues that considering information that is gathered only from WSDL for web services clustering can actually limit its accuracy. Hence, the WTCluster combines the WSDL information with the web services' functional descriptions, and contextual and semantic information. This additional information is sought from the tags that are provided by web services search engines.

Ref. [25] investigates the ranking and clustering of web services search results that are related to the notion of dominance. The proposed methods perform the matching processes by employing multiple criteria and do not aggregate the matching scores of each service parameter. In relation to clustering, they evaluate two types of algorithm namely Approximate Skyline Clustering (ASC) and Heuristic Skyline Clustering (HSC). The experiments are conducted in order to find the most representable services for clustering. Consequently, the clusters are able to reveal the trade-offs that exist between the matched parameters.

Ref. [26] proposes a statistical clustering technique which can improve the current distributed vector space search engine for web services clustering. Based on the experiments, their technique is able to handle a very large repository of web services.

Apparently, all of the manuscripts only focus on producing methods that can improve web services clustering. None of them emphasizes on the clustering of web services based on their QoS data as proposed by this paper. Moreover, this paper also focuses on web services fuzzy clustering which is not covered in any of the reviewed manuscripts.

B. Functional-based Web Services Clustering

Ref. [27] investigates the clustering of web services' users based on their interests. This information is analyzed and compared with web services' characteristics to discover the matching services. Once the matching services are discovered, the QoS is evaluated in order to confirm whether it meets the users' expectation or not.

Ref. [28] proposes the clustering of web services for discovery purposes. The clustering method proposed by the work employs web services ontology. That means, web services matching is calculated based on their functions and processes using an accurate semantic similarity concept of the domain ontology.

Ref. [29] argues that web services discovery which is based only on registries is not adequate in guaranteeing good process delivery. In addition, the increasing number of web services has also caused bad performance and low efficiency of discovery process. Hence, they propose the discovery method that is based on semantics and clustering. In this work, the web services' descriptions similarity is used for clustering purpose.

Ref. [30] considers the application of clustering method in the process of providing an interface adapter for web services. The interface adapter converts the interface of the requested service into that of the available service. The process requires the interface matching between the source and the target interfaces through a procedure known as methods matching. Moreover, the work also proposes the clustering of methods based on their parameter similarity, in order to precisely perform the matching process.

Ref. [31] argues that clustering of web services is important especially in the current trend of the growing number of web services' application development. Hence, this work presents a similarity model to evaluate the similarity among web services for clustering purpose. The model employs a preprocessing approach that takes into account the programming style and naming rules of services information. This approach also employs the Structural Clustering Algorithm for Networks (SCAN) that can further improve the similarity calculation. The model focuses on evaluating the functional similarity of web services.

Ref. [32] proposes a fuzzy semantic clustering algorithm that can perform the calculation of semantic similarity among web services. The algorithm is based on ontology concept that enables the grouping of web services into semantically meaningful clusters. The algorithm utilizes the functionality and interface information that are contained in WSDL document. They include name and text descriptions, annotations of ontology files or elements, operation descriptions, and input and output descriptions. Besides, Ref. [5] and Ref. [33] present another semantic web services clustering approach using ant ontology concept. The semantic similarity is calculated using a matching method and a metrics set. The metrics measure the matching degree of two services based on their ontology descriptions. In general, the proposed clustering approach is based on functional similarity of web services.

Ref. [34] presents an algorithm for web services clustering that utilizes graph theory and a corresponding algorithm of web services discovery. The proposed algorithms are designed for semantic web services, hence, ontology is used to describe the input and output parameters of the services. In addition, Ref. [35] proposes yet another semantic clustering of web services that can resolve the problem of semantic web services discovery in the server center. The clustering method is performed based on semantic descriptions of web services' functionalities. Moreover, Ref. [36] also investigates the issue of semantic-based web services clustering to resolve the discovery problems. The work proposes the use of Semantic Web Services Clustering (SWSC) method and utilizes the information provided by UDDI's components that includes OWL-S file, WSDL and services description. The method grants different weight to each of the components according to their degree of significance in services discovery.

Furthermore, Ref. [37], [38] present the clustering algorithm for semantic web services retrieval engine. They argue that the proposed algorithm can produce more effective results than the traditional ranked-list approach. In its processes, the algorithm utilizes the general descriptions of requirements or web services' input and output parameters or the combination of both.

Ref. [39] proposes a clustering approach to improve the search process provided by UDDI registries. They argue that the UDDI's search functionality has limitations in terms of handling a huge number of services and it does not employ semantic capability in its operation. Therefore, they produce a clustering semantic algorithm, which has the capability to eliminate the return of irrelevant services from users' queries. They also employ a machine learning method namely Probabilistic Latent Semantic Analysis (PLSA). The PLSA is able to capture the semantics that are hidden in the queries' words and the services' descriptions.

Ref. [40] argues that the traditional web services' semantic clustering approaches, which are based on the measurements of semantic similarity distance and threshold, suffer from poor accuracy. Hence, they propose the use of taxonomic clustering algorithm that groups web services based on their functional similarity.

Ref. [41] introduces WordNet lexical database to work with the Vector Space Model (VSM) feature vectors for web services' patterns representation. Besides, they employ Self Organizing Map (SOM) neural network approach in clustering method. The good representation of web services' patterns produced by the lexical database has helped to improve the clustering process. This work however only considers functional descriptions in performing the clustering.

Ref. [6] investigates web services' functionality analysis which utilizes clustering method in its process. Through this clustering method, service taxonomy's hierarchy is created. Any web services will be labeled accordingly after they went through the clustering process. Any new web services will be compared with the clustered ones and then assigned with the matched labels.

Ref. [42] proposes a context-based web services clustering to replace the link- or link-content-based clustering. In its clustering process, the proposed method considers web services context, which is a comprehensive description of services' activities.

Ref. [43] presents a web services discovery method using keyword clustering and concept expansion. These are performed based on the content of web services, requests' reasoning and service matching. The keyword clustering is performed by employing the Pareto Principal word filtering method.

Ref. [44] proposes another web services discovery technique which uses clustering method and string matching. The technique also leverages the web services' semantic description which is available in WSDL. For clustering, they investigate three methods namely non-hierarchical methods such as K-Means, hierarchical methods like complete linkage and single linkage, and neural network-based method using SOM. As a result, they choose the unsupervised artificial neural network to be implemented in the proposed web services discovery technique.

Ref. [45] proposes the employment of biologically-inspired method in semantic web services clustering in order to produce a more efficient discovery process. The method evaluates the similarity of two services by comparing the degree of match of their semantic descriptions. In the other work [46], Particle Swarm Optimization is proposed as a method for web services clustering. The method applies the similar concept to cluster services, which is by evaluating the similarity of services' semantic descriptions. Moreover, the work also defines a set of metrics which is able to perform this similarity evaluation.

Ref. [47] presents a clustering-based semantic matching method to support web services discovery. The method applies the Automated Knowledge Acquisition approach, which has the capability to extend the ontology in the case when the user's query word does not contain in the ontology.

Ref. [48], [49] propose the automatic clustering of WSDL documents which can support services' search engine and reduce the search space in discovery process. The proposed technique is based on functional similarity that is able to collect, discover and cluster the web services into groups.

Ref. [50] argues that web services orchestration only takes into consideration the descriptions of what the services are and how they are provided. It ignores the other important considerations that describe when and where the services are served. Consequently, web services search will become inefficient and produce too many results due to a very broad search filtering queries. Hence, the work proposes a clustering algorithm which is based on divide-and-conquer rule. In its operation, the algorithm uses spatial and temporal attributes of web services.

Ref. [51] argues that web services must also be able to support a variety of mobile devices in mobile ad-hoc network (MANET), the environment in which mobile devices move independently. Therefore, they propose a method for services discovery in MANET. The proposed method performs the clustering based on the devices' mobility. The services discovery is carried out based on the clusters' characteristics.

In addition, a few other works also present the web services functional-based clustering, which include the clustering based on trust [52], registries [53], WSDL documents [54], task objects [55], geotaggedphotos [56] and clustering to solve web services discovery on MANET [57], [58].

Generally, all of the reviewed findings focus only on clustering approaches that are based on the functional aspects of web services. In contrast, this paper emphasizes on the non-functional-based clustering of web services. Moreover, none of the reviewed manuscripts discusses fuzzy-based QoS clustering as proposed by this paper.

C. Non-Functional-based Web Services Clustering

Ref. [59] proposes a web services' recommendation system which ranks web services based on their predicted QoS. The QoS prediction is performed based on its patterns. Due to the dynamic nature, each pattern does not have an exact cluster. Hence, the work argues that fuzzy clustering can produce more accurate prediction results than hard clustering since the former does not assign each of the QoS patterns with an exact cluster. Instead, each pattern has more than one cluster with different degrees of membership. The work however, focuses on improving fuzzy clustering method to produce a better QoS prediction for web services' recommendation system. There is no description on categorizing web services' QoS into fuzzy clusters as proposed by this paper.

Ref. [60] also develops a clustering-based QoS prediction solution for web services' recommendation system. The work argues that the real-time QoS values are needed to ensure the accuracy of the prediction. This is due to the fact that web services are dynamic in nature. Hence, they propose the use of a number of computers, known as landmarks, to perform the gathering of the real-time QoS data. These landmarks are distributed over the Internet. Furthermore, these landmarks are clustered based on their QoS, for instance latency time. Using this clustering of landmarks, users may perform the evaluation and select the cluster that meets their needs. Then, the QoS prediction is performed using information gathered from the landmarks of the particular cluster. Overall, this work gives more focus on developing a more accurate QoS prediction rather than QoS clustering. Moreover, the clustering method presented in their work applies the hard technique, which is different from fuzzy technique as proposed by this paper.

Ref. [7] focuses on improving the efficiency of web services discovery. They propose services clustering as a solution for the issue. In the solution, the web services similarity is calculated based on service name, capability, interface, and QoS similarities. The QoS similarity calculation is however not presented in the paper. In general, it sets a similarity threshold value. Every service that has similarity values more than the threshold value will be grouped into clusters. Thus, it can be concluded that the work does not emphasize on fuzzy clustering of web services' QoS as proposed by this paper.

Ref. [61] proposes an algorithm that uses clustering technique to cluster a huge number of web services into a number of groups according to QoS properties. The algorithm is able to reduce computational time and produce near-optimal web services selection process. The clustering is performed by employing the OPTICS algorithm and does not consider fuzzy clustering as proposed by this paper.

Ref. [62] performs web services clustering based on QoS properties using genetic algorithm (GA). Through clustering, it is believed that the overhead can be reduced and the discovery efficiency can be improved. Though their work emphasizes on QoS properties, the aim is set to discover the hard clustering instead of the fuzzy clustering as proposed by this paper. Furthermore, their work also employs different algorithm than FCM that is proposed by this paper. Moreover, they also do not present any fuzzy quality model that clusters web services based on QoS properties.

Ref. [63] reports a clustering of users' QoS opinions. Due to the diverse backgrounds and preference among users, the web service's QoS opinions are considered as multi-groups. In consequent, the process of clustering becomes more complicated than the single-group case. To handle this issue, they propose the use of associated weight on each QoS parameter. In another perspective, they also propose the use of fuzzy evaluation method to prevent unintentional removal of meaningful data that do not fall within the defined groups' boundaries. Hence, they employ Fuzzy Multi-Groups based SAM (FMGSAM) technique in their clustering method. The work however only considers similarity matching of users QoS opinions. In contrast, this paper proposes the clustering of web services based on the real QoS data collected from the network. Furthermore, the difference also exists in terms of the proposed fuzzy technique, in which their work suggests the use of FMGSAM instead of FCM.

D. *Significance of the Work*

Table I shows the summary of all the reviewed manuscripts. On the whole, the presented review on the previous works reveals that the report on web services' QoS fuzzy-based clustering using real data has never been published before. Moreover, the review also proves that clustering is highly important to provide efficient web services delivery especially in the discovery, selection and recommendation processes.

The significance of the work proposed by this paper can be viewed in three perspectives. Firstly, it provides an exploration on how the real web services' QoS data are clustered fuzzily using unsupervised method of FCM. Secondly, the generated fuzzy clustering results can be used as reference during contract negotiation and services selection. This is especially useful for the requestors who possess little technical knowledge on web services. By applying the same processes and technique as proposed by this paper, the fuzzy clustering results can always be updated based on new and latest QoS data. Finally, the presented results of the QoS clustering are also important for the development of fuzzy-based web services' applications. This is due to fact that the clustered QoS as proposed by this paper can become a better choice for generating fuzzy inference as compared to the expert knowledge. The inference based on expert knowledge suffers from the possibility of being less accurate and unavailability of proficient people to provide the input. Moreover, seeking knowledge from experts normally involves a time-consuming process.

III. Results

A. *QoS Data Sets*

In this work, the clustering was conducted upon the actual QoS data collected from 365 web services using Web Service Crawler Engine (WSCE). It was provided by Al-Masri and Mahmoud [64], [65]. For experimental purposes, we chose three QoS parameters from the particular data set. They were response time, latency and availability. These data were separately contained in three different data sets where each of the data sets consisted of 1500 data points.

B. *Clustering Validation*

Finding the optimal number of clusters that fits the underlying data set has become one of the main issues in data clustering [66]. This is due to the fact that data clustering, which is unsupervised in nature, does not provide the optimal number of clusters a priori [67], [68]. As apparent in literature, there are numerous researches that propose validation indices to evaluate the number of clusters [69]. They are known as clustering validity indices.

TABLE I: Web SERVICES CLUSTERING

Author (s) (pub. Year)	Ref. No.	Clustering of web services' QoS	Fuzzy clustering of web services
Improvement of web services clustering technique			
Sarang Sukumar et al. (2012)	[18]	No	No
Al-Shammary et al. (2011)	[19]	No	No
Gao et al. (2009)	[20]	No	No
Abramowicz et al. (2007)	[21]	No	No
Li and Yang (2012)	[22]	No	No
Zheng and Wang (2012)	[23]	No	No
Chen et al. (2011)	[24]	No	No
Skoutas et al. (2010)	[25]	No	No
Platzer et al. (2009)	[26]	No	No
Functional-based web services clustering			
Zheng et al (2012)	[27]	No	No
Xie et al. (2011)	[28]	No	No
Wen et al. (2011)	[29]	No	No
Kim et al. (2011)	[30]	No	No
Zhu et al. (2010)	[31]	No	No
Gholamzadeh and Taghiyareh (2010)	[32]	No	Yes
Chifu et al. (2010)	[5]	No	No
Pop et al. (2010)	[33]	No	No
Li (2010)	[34]	No	No
Wang et al. (2009)	[35]	No	No
Nayak and Lee (2007)	[36]	No	No
Dong and Chi (2008, 2008)	[37, 38]	No	No
Ma et al. (2008)	[39]	No	No
Dasgupta et al. (2010)	[40]	No	No
Chen et al. (2010)	[41]	No	No
Liang et al. (2009)	[6]	No	No
Zhang et al. (2009)	[42]	No	No
Zhou et al. (2008)	[43]	No	No
Ram et al. (2006)	[44]	No	No
Pop et al. (2012)	[45]	No	No
Nagy et al. (2011)	[46]	No	No
Liu et al. (2009)	[47]	No	No
Liu and Wong (2008, 2009)	[48, 49]	No	No
Yang et al. (2007)	[50]	No	No
Kim et al. (2011)	[51]	No	No
Yan and Zhang (2010)	[52]	No	No
Sellami et al. (2010)	[53]	No	No
Elgazzar et al. (2010)	[54]	No	No
Zhang et al. (2009)	[55]	No	No
Lo (2009)	[56]	No	No
Shim et al. (2009)	[57]	No	No
Zhang et al. (2009)	[58]	No	No
Non-functional-based web services clustering			
Zhang et al. (2012)	[59]	No	Yes
Zhu et al. (2012)	[60]	Yes	No
Liu et al. (2011)	[7]	Yes	No
Xia et al. (2011)	[61]	Yes	No
Zhou and Li (2009)	[62]	Yes	No
Lin et al. (2009)	[63]	No	Yes

The optimal clustering is found when it produces the high degrees of compactness and separation [70], [71]. The compactness is measured based on the closeness among cluster members. A better compactness means the members of each cluster are closer to each other. In contrast, separation measures the distance among clusters. A good separation means the clusters are widely spaced. In calculating their results, some validity indices only consider the membership values, while some others consider both the membership values as well as the data set itself. It is evident that the latter can produce more representing results [67], [71]. Therefore, we propose the use of Xie and Beni (XB) validity index as it posses this characteristic. Furthermore, Ref. [72] also claim that XB index performs well when $2 \leq$ *number of clusters* ≤ 10 and $1.01 \leq$ *fuzzy weighting component* ≤ 7.

In this work, we assumed that the value of fuzzy weighting exponent, m is 2. This was based on Pal and Bezdek [72] which claim that $m=2$ is normally used for the type of algorithm like FCM. We also assumed that the maximum number of clusters to be evaluated was $c_{max}=5$. This assumption of the reasonable c_{max} value was based on a report in [73]. Hence, the validated numbers of clusters were $c_i \in \{2, 3, 4, 5\}$. The optimal number of clusters was represented by the minimum value of the results produced by XB index.

Table II shows the results of XB validation upon the three data sets. The availability data set contained data points measured in percentage (%). On the other hand, the other two data sets comprised data points measured in milliseconds (ms). Based on Table II, it is shown that $c=4$ is the optimal number of clusters for response time, while $c=3$ is for the other two parameters.

TABLE II: XB VALIDITY INDEX SCORES

QoS Parameter	Number of Clusters			
	2	3	4	5
Response time	2476.1000	263.2772	148.4363	2362.9000
Availability	18.8142	15.8146	59.2782	16.0355
Latency	6535.2000	537.7606	958.0795	26506.0000

C. Clustering with FCM

FCM categorizes a collection of data into a number of different clusters where it assigns each data point to a specific cluster with a membership degree [74]. FCM performs the clustering of data points through an iterative process that minimizes the objective function. In every repetition of this process, the membership functions of the data points, the cluster centers, and the objective function are computed. These computed values will replace the values generated from the previous repetition. The process stops when it reaches the threshold value which is the minimum value of the objective function [75].

Based on Table II, we named the response time clusters as *Good*, *Moderate High*, *Moderate Low* and *Poor*. Moreover, the clusters of availability and latency comprised *Good*, *Moderate* and *Poor*. The clustering results are shown in Fig. 1, Fig. 2 and Fig. 3. Each curve in the figures represents a cluster, while its peak corresponds to the cluster center.

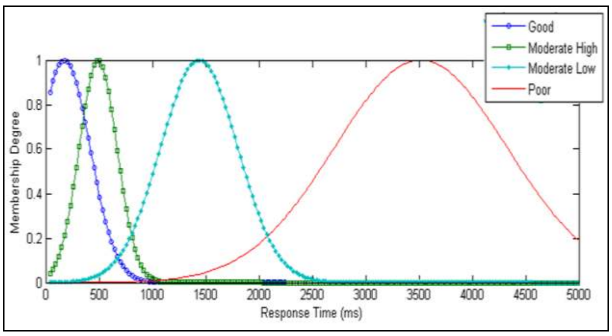

Fig. 1. Fuzzy quality model of web services' response time.

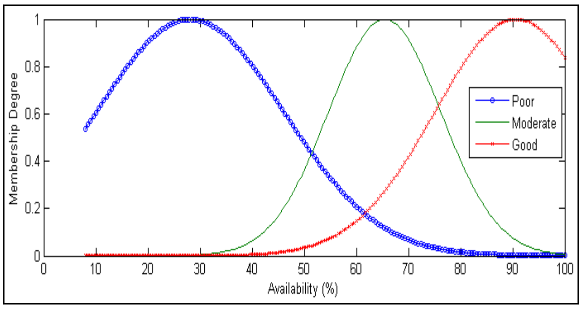

Fig. 2. Fuzzy quality model of web services' availability.

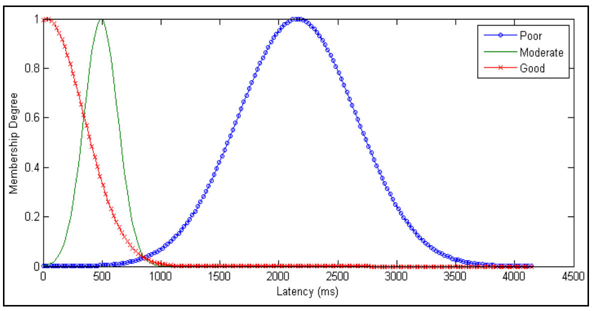

Fig. 3. Fuzzy quality model of web services' latency.

As discussed earlier, these clustering results can be used as the reference to guide requestors in choosing suitable web services. This is due to the fact that they can distinguish the good, moderate and poor web services' QoS by looking at the generated clusters. Furthermore, the clustering results can also be utilized in the inference components of fuzzy-based web services applications. For example, Fig. 4 shows that the generated clusters for response time are implemented in a fuzzy inference system on Matlab.

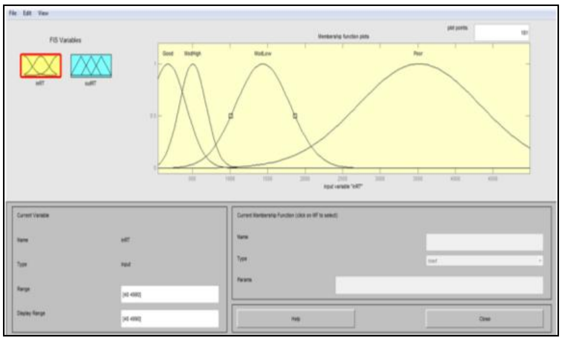

Fig. 4. Employment of response time's clustered web services' QoS in fuzzy inference system.

IV. Conclusion

The conducted review showed that clustering is significantly important to produce efficient web services' delivery, particularly in their discovery, selection and recommendation processes. The review also discovered that the clustering of web services based on their QoS has never been published before. The proposed fuzzy clustering of web services' QoS is believed could overcome two main issues. Firstly, it helps the requestors who have limited technical knowledge about the web services to understand what the realistic QoS is so that they can subscribe services that give value for their money. The outcomes of the clustering can also be used as reference for the requestors in the process of requirements negotiation and specification. Secondly, they also contribute towards the development of fuzzy-based web services. As fuzzy inference using knowledge from the experts requires more time, and suffers from less accuracy and unavailability, the automatic generation of the inference components like the proposed QoS clustering can be considered as the better option.

In this work, the fuzzy clustering of web services' QoS was performed using FCM algorithm. It was performed based on the data gathered from the real web services' which included three QoS parameters namely response time, availability and latency. Furthermore, this work also used XB validity index to ensure that the clustering was conducted optimally.

For future work, this work will be extended by employing the generated clustering results into web services' QoS monitoring application. This will enable the monitoring application to perform web services' QoS monitoring based on vague expected delivery values using fuzzy linguistic variables.

References

[1] S. Saha, C. A. Murthy, and S. K. Pal, "Classification of Web services using tensor space model and rough ensemble classifier," *Foundations of Intelligent Systems Lecture Notes in Computer Science*, vol. 4994, pp. 508-513, 2008.

[2] P. Venketesh and R. Venkatesan, "A survey on applications of neural networks and evolutionary techniques in Web caching," *IETE Technical Review*, vol. 26, no. 3, pp. 171-180, 2009.

[3] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 491-502, 2005.

[4] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems," in *Proc. IEEE*, 1998, pp. 2210-2239.

[5] V. R. Chifu, C. B. Pop, I. Salomie, M. Dinsoreanu, V. Acretoaie, and T. David, "An ant-inspired approach for semantic Web service clustering," in *Proc. 9th RoEduNet IEEE International Conference*, 2010, pp. 145-150.

[6] Q. Liang, P. Li, P. C. K. Hung, and X. Wu, "Clustering Web services for automatic categorization," in *Proc. IEEE International Conference on Services Computing*, 2009, pp. 380-387.

[7] J.-X. Liu, K.-Q. He, J. Wang, and D. Ning, "A clustering method for Web service discovery," in *Proc. IEEE International Conference on Services Computing*, 2011, pp. 729-730.

[8] M. Crasso, A. Zunino, and M. Campo, "AWSC: An approach to Web service classification based on machine learning techniques," *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial*, vol. 12, no. 37, pp. 25-36, 2008.

[9] R. Ladner, E. Warner, F. Petry, K. M. Gupta, P. Moore, D. W. Aha, and K. Shaw, "Case-based classification alternatives to ontologies for automated web service discovery and integration," in *Proc. SPIE Defense & Security Symposium*, 2006, pp. 620117-1 – 620117-8.

[10] H. Wang, Y. Shi, X. Zhouy, and Q. Zhou, "Web service classification using support vector machine," in *Proc. 22nd International Conference on Tools with Artificial Intelligence*, 2010, pp. 3-6.

[11] D. Mobedpour and D. Chen, "User-centered design of a QoS-based web service selection system," *Service Oriented Computing and Applications*, vol. 7, no. 2, pp. 117–127, 2013.

[12] D. Bacciu, M. G. Buscemi, and L. Mkrtchyan, "Adaptive fuzzy-valued service selection," in *Proc. ACM Symposium on Applied Computing*, 2010, pp. 2467–2471.

[13] W. Sherchan, S. W. Loke, and S. Krishnaswamy, "A fuzzy model for reasoning about reputation in web services," in *Proc. ACM Symposium on Applied Computing*, 2006, pp. 1886–1892.

[14] M. H. Zadeh and M. A. Seyyedi, "Qos monitoring for web services by time series forecasting," in *Proc. 3rd IEEE International Conference on Computer Science and Information Technology*, 2010, pp. 659–663.

[15] S. Guillaume, "Designing fuzzy inference systems from data: An interpretability-oriented review," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 3, pp. 426-443, 2001.

[16] J.-S. R. Jang, "Self-learning fuzzy controllers based on temporal back propagation," *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 714-723, 1992.

[17] S. Vega-Pons and J. Ruiz-Shulcloper, "A survey of clustering ensemble algorithms," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 25, no. 3, pp. 337-372, 2011.

[18] A. S. Sukumar, J. Loganathan, and T. Geetha, "Clustering Web services based on multi-criteria service dominance relationship using Peano Space filling curve," in *Proc. International Conference on Data Science and Engineering*, 2012, pp. 13-18.

[19] D. Al-Shammary, I. Khalil, and L. E. George, "Clustering SOAP Web services on internet computing using fast fractals," in *Proc. 10th IEEE International Symposium on Network Computing and Applications*, 2011, pp. 366-371.

[20] H. Gao, W. Stucky, and L. Liu, "Web services classification based on intelligent clustering techniques," in *Proc. International Forum on Information Technology and Applications*, 2009, pp. 242-245.

[21] W. Abramowicz, K. Haniewicz, M. Kaczmarek, and D. Zyskowski, "Architecture for Web services filtering and clustering," in *Proc. 2nd International Conference on Internet and Web Applications and Services*, 2007.

[22] M. Li and Y. Yang, "Efficient clustering index for semantic Web service based on user preference," in *Proc. International Conference on Computer Science and Information Processing*, 2012, pp. 291-294.

[23] Q. Zheng and Y. Wang, "The application of semi-supervised clustering in web services composition," *Advances in Intelligent and Soft Computing*, vol. 169, pp. 683-688, 2012.

[24] L. Chen, L. Hu, Z. Zheng, J. Wu, J. Yin, Y. Li, and S. Deng, "WTCluster: Utilizing tags for web services clustering," *Lecture Notes in Computer Science*, vol. 7084, pp. 204-218, 2011.

[25] D. Skoutas, D. Sacharidis, A. Simitsis, and T. Sellis, "Ranking and clustering web services using multicriteria dominance relationships," *IEEE Transactions on Services Computing*, vol. 3, no. 3, pp. 163-177, 2010.

[26] C. Platzer, F. Rosenberg, and S. Dustdar, "Web service clustering using multidimensional angles as proximity measures," *ACM Transactions on Internet Technology*, vol. 9, no. 3, 2009.

[27] K. Zheng, H. Xiong, Y. Cui, J. Chen, and L. Han, "User clustering-based web service discovery," in *Proc. 6th International Conference on Internet Computing for Science and Engineering*, 2012, pp. 276-279.

[28] L.-L. Xie, F.-Z. Chen, and J.-S. Kou, "Ontology-based semantic web services clustering," in *Proc. IEEE 18th International Conference on Industrial Engineering and Engineering Management*, 2011, pp. 2075-2079.

[29] T. Wen, G. Sheng, Y. Li, and Q. Guo, "Research on Web service discovery with semantics and clustering," in *Proc. 6th IEEE Joint International Information Technology and Artificial Intelligence Conference*, 2011, pp. 62-67.

[30] B. Kim, N. Ho, D. Lee, and S. J. Hyun, "A clustering based schema matching scheme for improving matching correctness of web service interfaces," in *Proc. IEEE International Conference on Services Computing*, 2011, pp. 488-495.

[31] Z. Zhu, H. Yuan, J. Song, J. Bi, and G. Liu, "WS-SCAN: A effective approach for Web services clustering," in *Proc. International Conference on Computer Application and System Modeling*, 2010, pp. V5-618-V5-622.

[32] N. Gholamzadeh and F. Taghiyareh, "Ontology-based fuzzy Web services clustering," in *Proc. 5th International Symposium on Telecommunications*, 2010, pp. 721-725.

[33] C. B. Pop, V. R. Chifu, I. Salomie, M. Dinsoreanu, T. David, and V. Acretoaie, "Semantic web service clustering for efficient discovery using an ant-based method " *Studies in Computational Intelligence*, vol. 315, pp. 23-33, 2010.

[34] Y. Li, "Algorithm for semantic web services clustering and discovery," in *Proc. WRI International Conference on Communications and Mobile Computing,* 2010, pp. 532-536.

[35] L. Wang, J. Si, and X.-B. Hu, "Research on the clustering and composition of P2P-based web services," in *Proc. 2nd International Conference on Biomedical Engineering and Informatics*, 2009, pp. 1-5.

[36] R. Nayak and B. Lee, "Web service discovery with additional semantics and clustering," in *Proc. IEEE/WIC/ACM International Conference on Web Intelligence*, 2007, pp. 555-558.

[37] S. Dong and C. Chi-Hung, "A clustering-based approach for assisting semantic web service retrieval," in *Proc. IEEE International Conference on Web Services*, 2008, pp. 838-839.

[38] S. Dong and C.-H. Chi, "Effective web service retrieval based on clustering," in *Proc. 4th International Conference on Semantics, Knowledge, and Grid*, 2008, pp. 469-472.

[39] J. Ma, Y. Zhang, and J. He, "Efficiently finding web services using a clustering semantic approach," in *Proc. International Workshop on Context Enabled Source and Service Selection, Integration and Adaptation*, 2008.

[40] S. Dasgupta, S. Bhat, and Y. Lee, "Taxonomic clustering of web service for efficient discovery," in *Proc. 19th ACM International Conference on Information and Knowledge Management*, 2010, pp. 1617-1620.

[41] L. Chen, G. Yang, Y. Zhang, and Z. Chen, "Web services clustering using SOM based on kernel cosine similarity measure," in *Proc. 2nd International Conference on Information Science and Engineering*, 2010, pp. 846-850.

[42] R. Zhang, K. Zettsu, T. Nakanishi, Y. Kidawara, and Y. Kiyoki, "Context-based web service clustering," in *Proc. 5th International Conference on Semantics, Knowledge, and Grid*, 2009, pp. 192-199.

[43] J. Zhou, T. Zhang, H. Meng, L. Xiao, G. Chen, and D. Li, "Web service discovery based on keyword clustering and ontology," in *Proc. IEEE International Conference on Granular Computing*, 2008, pp. 844-848.

[44] S. Ram, Y. Hwang, and H. Zhao, "A clustering based approach for facilitating semantic web service discovery," in *Proc. 15th Annual Workshop on Information Technolgies & Systems*, 2006.

[45] C. B. Pop, V. R. Chifu, I. Salomie, M. Dinsoreanu, T. David, *et al.*, "Biologically-inspired clustering of semantic Web services. Birds or ants intelligence?" *Concurrency Computation Practice and Experience*, vol. 24, no. 6, pp. 619-633, 2012.

[46] A. Nagy, C. Oprişa, I. Salomie, C. B. Pop, V. R. Chifu, and M. Dinşoreanu, "Particle swarm optimization for clustering semantic web services," in *Proc. 10th International Symposium on Parallel and Distributed Computing*, 2011, pp. 170-177.

[47] P. Liu, J. Zhang, and X. Yu, "Clustering-based semantic web service matchmaking with automated knowledge acquisition," *Lecture Notes in Computer Science*, vol. 5854, pp. 261-270, 2009.

[48] W. Liu and W. Wong, "Discovering homogenous service communities through web service clustering," in *Proc. AAMAS International Workshop on Service-Oriented Computing: Agents, Semantics, and Engineering*, 2008, pp. 69-82.

[49] W. Liu and W. Wong, "Web service clustering using text mining techniques," *International Journal of Agent-Oriented Software Engineering*, vol. 3, no. 1, pp. 6-26, 2009.

[50] H. Yang, J. Chen, X. Meng, and B. Qiu, "Dynamically traveling Web service clustering based on spatial and temporal aspects," *Lecture Notes in Computer Science*, vol. 4802, pp. 348-357, 2007.

[51] Y.-S. Kim, Y.-S. Shim, and K.-H. Lee, *Mobile Information Systems*, vol. 7, no. 4, pp. 299-315, 2011.

[52] C. Yan and S. Zhang, "Method to dynamic classification of trust for web service based on intuitionistic fuzzy clustering," in *Proc. International Conference on Computational Intelligence and Software Engineering*, 2010.

[53] M. Sellami, W. Gaaloul, and S. Tata, "Functionality-driven clustering of Web service registries," in *Proc. IEEE 7th International Conference on Services Computing*, 2010, pp. 631-634.

[54] K. Elgazzar, A. E. Hassan, and P. Martin, "Clustering WSDL documents to bootstrap the discovery of Web services," in *Proc. IEEE 8th International Conference on Web Services*, 2010, pp. 147-154.

[55] X. Zhang, Y. Yin, M. Zhang, and B. Zhang, "Web service community discovery based on spectrum clustering " in *Proc. International Conference on Computational Intelligence and Security*, 2009, pp. 187-191.

[56] S.-H. Lo, "Design and implementation the incremental clustering algorithm for geotagged photos on a map-enabled photo web service," in *Proc. 10th International Conference on Mobile Data Management: Systems, Services and Middleware*, 2009, pp. 666-671.

[57] Y.-S. Shim, Y.-S. Kim, and K.-H. Lee, "A mobility-based clustering and discovery of web services in mobile Ad-hoc networks," in *Proc. IEEE International Conference on Web Services*, 2009, pp. 374-380.
[58] C. Zhang, Z.-Y. Chen, Q.-S. Zhu, and H.-J. Liu, "A novel clustering framework for web service in MANET," in *Proc. 6th Web Information Systems and Applications Conference*, 2009, pp. 109-115.
[59] M. Zhang, X. Liu, R. Zhang, and H. Sun, "A Web service recommendation approach based on QoS prediction using fuzzy clustering," in *Proc. IEEE 9th International Conference on Services Computing*, 2012, pp. 138-145.
[60] J. Zhu, Y. Kang, Z. Zheng, and M. R. Lyu, "A clustering-based QoS prediction approach for Web service recommendation," in *Proc. 15th IEEE International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing Workshops*, 2012, pp. 93-98.
[61] Y. Xia, P. Chen, L. Bao, M. Wang, and J. Yang, "A QoS-aware Web service selection algorithm based on clustering," in *Proc. IEEE 9th International Conference on Web Services*, 2011, pp. 428-435.
[62] J. Zhou and S. Li, "Semantic Web service discovery approach using service clustering," in *Proc. International Conference on Information Engineering and Computer Science*, 2009, pp. 1-5.
[63] W.-L. Lin, C.-C. Lo, K.-M. Chao, and N. Godwin, "Fuzzy similarity clustering for consumer-centric QOS-aware selection of web services," in *Proc. International Conference on Complex, Intelligent and Software Intensive Systems*, 2009, pp. 904-909.
[64] E. Al-Masri and Q. H. Mahmoud, "Discovering the best web service," in *Proc. 16th International Conference on World Wide Web*, 2007, pp. 1257-1258.
[65] E. Al-Masri and Q. H. Mahmoud, "QoS-based discovery and ranking of Web services," in *Proc. IEEE 16th International Conference on Computer Communications and Networks*, 2007, pp. 529-534.
[66] C.-F. Chen and J.-M. Lee, "The validity measurement of fuzzy c-means classifier for remotely sensed images," in *Proc. 22nd Asian Conference on Remote Sensing*, 2001.
[67] M. R. Rezaee, B. P. F. Lelieveldt, and J. H. C. Reiber, "A new cluster validity index for the fuzzy c-mean," *Pattern Recognition Letters*, vol. 19, no. 3-4, pp. 237–246, 1998.
[68] K.-L. Wu and M.-S. Yang, "A cluster validity index for fuzzy clustering," *Pattern Recognition Letters*, vol. 26, no. 9, pp. 1275–1291, 2005.
[69] W. Wang and Y. Zhang, "On fuzzy cluster validity indices," *Fuzzy Sets and Systems*, vol. 158, no. 19, pp. 2095 – 2117, 2007.
[70] M. J. A. Berry and G. Linoff, *Data Mining Techniques For Marketing, Sales and Customer Support*, John Wiley & Sons, Inc., 1996.
[71] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," *Journal of Intelligent Information Systems*, vol. 17, no. 2-3, pp. 107–145, 2001.
[72] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy c-means model," *IEEE Transactions on Fuzzy Systems*, vol. 3, no. 3, pp. 370-379, 1995.
[73] X. Xie and G. Beni, "A Validity measure for fuzzy clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 3, no. 8, pp. 841–847, 1991.
[74] L. Wang and J. Wang, "Feature Weighting fuzzy clustering integrating rough sets and shadowed sets," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 26, no. 4, 2012.
[75] H. Guldemır and A. Sengur, "Comparison of clustering algorithms for analog modulation classification," *Expert Systems with Applications 30*, vol. 30, no. 4, pp. 642-649, 2006.

Mohd Hilmi Hasan is a senior lecturer at Universiti Teknologi PETRONAS, Malaysia. He graduated with Bachelor of Technology (Hons.) Information Technology degree from Universiti Teknologi PETRONAS in 2002. He obtained his Master of Information Technology (eScience) in 2004 from The Australian National University, Canberra, Australia. He has been teaching and being an active researcher at the Universiti Teknologi PETRONAS since 2004, in which his research interest covers areas of artificial intelligence, networking and communication systems. He has also served as programme committee for a number of conferences in the area. He is currently a candidate of PhD of Information Technology at Universiti Teknologi PETRONAS.

Jafreezal Jaafar is the Head of Computer and Information Sciences Department at Universiti Teknologi PETRONAS, Malaysia. He is also a senior lecturer and active researcher at the university, in which his research areas include artificial intelligence, software design, virtual environment and human- computer interaction. He holds a Degree in Computer Science from Universiti Teknologi Malaysia (graduated in 1998) and Master of Applied Science in Information Technology from RMIT University, Australia (graduated in 2002). He obtained his PhD (Artificial Intelligence) degree from University of Edinburgh, United Kingdom in 2009. His PhD research addressed the issue of navigating autonomous agent using Fuzzy logic. By extending his research interest in Fuzzy logic and Artificial Intelligence (AI) areas, he managed to secure a number of research grants from the government body of Malaysia as well as the university's foundation body.

Mohd Fadzil Hassan is the Dean of Graduate Studies and an associate professor at Universiti Teknologi PETRONAS, Malaysia. He obtained his BSc (Computer Information Systems) from Colorado State University, USA in 1998, and Masters of Science (Artificial Intelligence), from University of Edinburgh, UK in 2001. He gained his PhD (Informatics) from University of Edinburgh, UK in 2007, in which his research explored the use of constraints relaxation method during interactions of agents that participate in a problem solving task. His research interest covers areas of Artificial Intelligence (AI), Multi-Agent Systems (MAS) and distributed interaction protocol for Multi-Agent Systems Application of AI and MAS in fields like E-Learning & Grid-Computing. He has secured a number of national grants awarded by the government body of Malaysia in those areas of research.