Clustering Web Services based on Multi-Criteria Service Dominance Relationship using Peano Space Filling Curve

Sarang Sukumar A, Jayakumar Loganathan and T Geetha
Department of Computer Science,
Pondicherry University,
Pondicherry, India
sarangsukumar@gmail.com, jkayitboy@gmail.com, tgeetha01@gmail.com

Abstract - **As the use of the web is greatly increased so as the need for effective web service selection and retrieval techniques. So many good algorithms have been developed for web service selection problem. The need for effective web service clustering technique is also came forward, which will help the users to search within their needed clusters, that will reflect different tradeoffs such as price , quality etc. Recent works on web service selection and retrieval techniques has found out that for getting greater accuracy and stability, all the algorithms for web service selection and clustering must satisfy the following two criteria. First, it should consider all the parameter matches. Second, multiple criteria should be taken into consideration while performing the parameter matching. In this paper we are proposing an efficient web service clustering technique based on multi criteria service dominance relationship, which satisfies the above criteria. Recent work on the same area proposed the use of Hilbert space filling curve, and then applying a simple algorithm for forming the clusters. Space filling curve is a way of mapping multi-dimensional space into the one dimensional space. So the effectiveness of the formed clusters are having a direct relationship with how effective the used space filling curve is. In this paper we propose the use of Peano Space filling curve for the multidimensional reduction, which tends to show less irregularity, more fairness and more scalability than Hilbert space filling curve.**

Keywords- Web services matchmaking; Clustering; Peano Space Filling Curve.

I. INTRODUCTION

WEB services are software entities that have a well-defined interface and perform a specific task. Examples of web services include services returning information to the user, such as news or weather forecast services, or services altering the world state, such as on-line shopping or booking services. A Web service is formally described in a standardized language (WSDL). The service description may include the names and types of input and output parameters, preconditions and effects, as well as Quality of Service (QoS) attributes, such as price, execution time, availability, and reputation.

As the use of the web is greatly increased, there will be a large number of candidate web services providing the same kind of services. So we need effective web service selection and retrieval techniques for identifying the most appropriate ones. In a typical scenario, a user provides a complete definition of the requested service. There will be large web service repository which contains a number of advertised service descriptions and the repository make use of an efficient matchmaking algorithm to identify the appropriate candidate web service relevant to the given request.

Since we are left with a large number of candidate services, the match making process will be cumbersome. So we are going to organize the candidate services into different clusters so that services within a cluster provide similar matches with respect to the request. And thus we can filter out the irreverent clusters and focus only on the single cluster which will make the matchmaking process more efficient. Since several parameters are involved in the matchmaking process, finding a service that provides a high degree of match for all parameters is difficult; instead, it is often needed to decide between different tradeoffs. Clustering the search results allows the user to identify an interesting advertisement and then browse similar results, i.e., those found within the same cluster.

As a first step the matchmaking algorithm we have to assign to each considered parameter, a parameter degree of match (PDM) with respect to the requested service. Then, an overall degree of match (ODM) can be computed as an aggregate of the individual PDMs. Several approaches for combining PDMs exist. One direction is to assign weights to individual scores [1], but it requires prior knowledge about the user preference. In the lack of such information, ODM calculation is based on the lower degree of match among parameters, but it may lead to significant information losses. Recent work [2] Web service selection suggested that Web service matchmaking algorithm should explicitly take into account all PDMs (R1) for avoiding information losses. Similarly there is no single criteria that can optimally determine the match between parameters. Keyword-based matchmaking [1] fails to properly identify extract semantics. On the other hand, semantic-based matchmaking [3] is hindered by the lack of available ontologies, Therefore multiple criteria should be taken into consideration while estimating the PDMs (R2). In this work we are using the concept of multi criteria service dominance relationship proposed in a recent work [2] on web service selection. Accordingly, candidates service Si dominates another Sj, with respect to a given request, if Si has higher PDMs compared to Sj in all parameters and according to all criteria. We are using the notion of uncertain dominance [4] since dominating services are rare to find.

This paper focuses mainly on the clustering of web services using multi criteria service dominance relationship. We propose the use of Peano space filling curve which is a space filling curve, for forming the clusters of web services from the multi criteria service dominance relationship, which is represented as multidimensional space. Peano space filling curve tends to show less irregularity, more fairness and more scalability which makes it more appropriable for multi-dimensional reduction and hence more effective the formed clusters are.

II. RELATED WORKS

In this section we discuss related work regarding Web service match making and web service clustering.

A. Web service matchmaking

Current industry standards for Web service description and web service match making focus mainly on keyword-based matching. Integrating multiple external matching services to a UDDI registry is proposed in [5]. Semantic-based approaches that use ontologies to enhance the service descriptions and address the matchmaking as a logic inference task was mentioned in [3]. Web services match making based on schema matching has been proposed in [6]. Similarity-based Web Service Matchmaking has been proposed in [7]. In a recent paper [8] proposes content-based matching and focuses on data-intensive Web services. A hybrid matchmaker OWLS-MX was proposed in [9].

B. Web service clustering

Only a very few works are there related to clustering of web services so as to provide an efficient tradeoff among different parameters. In a recent paper [2] Dimitrios Skoutas suggested the use of Hilbert Space Filling Curve for the multidimensional reduction of input graph into single dimensional graph, followed by simple heuristic algorithm for extracting cluster representatives. Multidimensional reduction using Hilbert SFC curve causes irregularity in the mapping, which will affect the quality of the selected of cluster representatives. Also Hilbert SFC also shows less scalability and less fairness which will have a direct impact on the quality of the cluster formation as the number of dimension increases.

III. PROBLEM DESCRIPTION

A. Multi criteria service dominance relationship

According to multi criteria service dominance relationship, a candidate service Si dominates another Sj, with respect to a given request, if Si has higher PDMs compared to Sj in all parameters and according to all criteria.

Table 1 illustrates the above relationship. For simplicity, we assume all services have one input Pin and one output Pout parameter, and that there exist four advertisements (candidates) A, B, C, D. Furthermore, for the given request, three different matching filters (e.g., different string similarity measures), m1, m2, and m3, are applied, resulting in the PDMs shown in Table 1. We can find that under any criterion, service A constitutes the best match with respect to both parameters. Hence, A dominates B, C, and D. However, there is no clear winner among the other three. For instance, according to m1, B is definitely better than D. On the other hand, m2 suggests that B has a lower match degree for the input parameter but a higher for the output.

TABLE 1

Services	Parameter	m1	m2	m3
A	Pin	0.96	1.00	0.92
	Pout	0.92	0.96	1.00
B	Pin	0.80	0.60	0.64
	Pout	0.80	0.88	0.72
C	Pin	0.84	0.88	0.72
	Pout	0.84	0.64	0.60
D	Pin	0.76	0.68	0.56
	Pout	0.76	0.64	0.68

Clearly dominating services are rare to find because of the trade-offs offered by the advertisements. Rather, most services will only be having good matches in some parameters while bad in others. Further, even for a particular parameter, different criteria may provide conflicting degrees of match. In this case we can adopt the notion of uncertain dominance from [4]. Briefly, a service dominates another with a probability that depends on multiple criteria PDMs.

B. Web service clustering

The purpose of web service clustering is to organize the matched services (using the previously described multi criteria service dominance relationship) into *l* groups that represent the different tradeoffs among all the considered parameters. Overall, the benefit of the clustering is that it diversifies the search results, allowing the users to focus and drill down on that subset of the results that better meets their requirements.

Table 2 gives a simple illustration of web service clustering with two matching filters m_1 and m_2, with one input parameter Pin and one output parameter Pout, and a set of matched services, A through I. Each service comprises two match instances, represented as points in the Pin×Pout plane shown in Fig.1; for example, service A consists of instances a_1 and a_2.

TABLE 2

Service	Parameter	m_1	m_2
A	P_{in} P_{out}	0.55 0.93	0.53 0.91
B	P_{in} P_{out}	0.60 0.92	0.62 0.90
C	P_{in} P_{out}	0.82 0.84	0.85 0.79
D	P_{in} P_{out}	0.87 0.85	0.85 0.83
E	P_{in} P_{out}	0.76 0.84	0.81 0.80
F	P_{in} P_{out}	0.93 0.63	0.95 0.55
G	P_{in} P_{out}	0.96 0.58	0.91 0.60
H	P_{in} P_{out}	0.72 0.70	0.70 0.66
I	P_{in} P_{out}	0.74 0.68	0.72 0.65

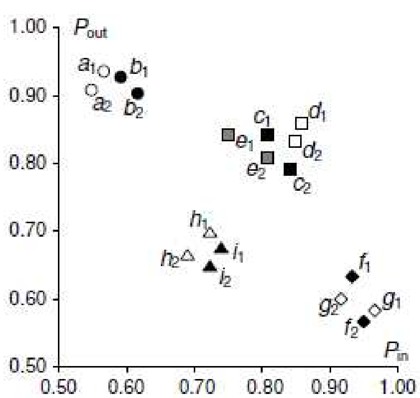

Fig.1. Clustering web services based on multi-criteria service dominance relationship.

From the Fig.1 one can easily deduce that the above matched service can be organized into four clusters namely {A,B}, {C,D}, {H,I} and {F,G} which represents the different tradeoffs. Cluster {A,B} represents the services which have high value for Pout, but low Pin values, cluster {G,H} represents the services having high Pin value, but low Pout value. Clusters {C, D}, {H, I} have average values for Pin and Pout.

C. Problem statement: Selecting l cluster representatives.

This paper addresses the problem of forming efficient web service clusters satisfying multi criteria service dominance relationship. Clustering may involve two high level steps: selecting *l* cluster representatives and forming the clusters by assigning each of the remaining services to the closest representatives. This paper deals with the problem of selecting *l* cluster representatives from the multidimensional graph representing the multi criteria service dominance relationship.

Input to our web service clustering algorithm is a multidimensional graph of points, from which the cluster representatives have to be extracted. As the number of parameters used in the match making process increases so does the dimensionality of the graph. Therefore extracting cluster centers that represent different tradeoffs will become practically infeasible when dimensionality of the input graph goes beyond four or more dimensions. In a recent work [2] on web service ranking and clustering Dimitrious Skoutas and Dimitris sacharids have suggested the use of Hilbert space filling curve for the multidimensional reduction of the input graph into a single dimensional graph and then applying a simple heuristic to extract the cluster representatives.

But this method is hampered by the following disadvantages

1) Multidimensional reduction using Hilbert SFC curve causes irregularity in the mapping, which will affect the quality of the selected of cluster representatives.

2) Hilbert SFC also shows less scalability and less fairness which will have a direct impact on the quality of the cluster formation.

3) Hilbert SFC fails to discover the exact cluster patterns, when compared with Peano SFC.

Our problem statement is associated with the dimensional reduction of multi-dimensional graph to a single dimensional graph. Space Filling Curve is a way of mapping multidimensional space into a single dimensional graph. So the effectiveness of the cluster formation is having a direct relationship with how effective the used space filling curve is. So here the problem statement is how to improve the multi-dimensional reduction technique so as the formed clusters are less hampered by irregularity and scalability. And is there any way to improve the heuristic technique for extracting cluster representatives from the two dimensional graph.

IV. PROPOSED SOLUTION

Space Filling Curve is a way of mapping multidimensional space into single dimensional space. In a recent research [10] on irregularity in multi-dimensional space filling curves with application in multimedia database, Mohammed F Mokbel and Walid G Areg performed a comparative study on the cluster forming capabilities of different space filling curves such an Hilbert SFC, Peano SFC and Grey SFC and found that Peano SFC is far better than Hilbert SFC and Grey SFC, in the sense that it is showing less irregularity and more fairness in multidimensional reduction. Peano Space Filling Curves also supports Intentional Biasing; i.e. giving importance to some dimensions while not with the others. This property makes Peano Space Filling Curves much suitable for application that has different parameters with different priority. In our case we can make use of this property for assigning appropriate weights to different QoS parameters.

So the proposed system is to use Peano SFC instead of Hilbert SFC for extracting cluster representatives. Peano space filling curve tends to show less irregularity, more fairness and more scalability which makes it more appropriable for multi-dimensional reduction and hence more effective the formed clusters are. Also Peano SFC tends to discover the exact cluster patterns as compared with Hilbert SFC. So here we may use the Peano space filling curve for the multidimensional reduction and evaluate the performance of clusters with respect to irregularity, fairness and scalability.

A. Multidimensional reduction using peano space filling curve

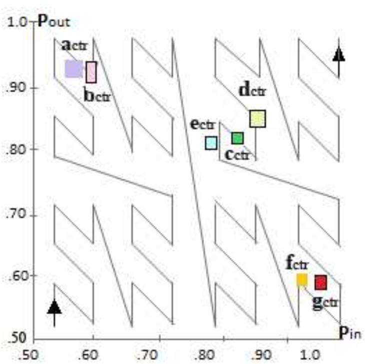

Fig.2 Dimensional reduction using Peano Space Filling curve

Fig.2 is a slight modification of Fig.1, which represented the previously considered multi criteria service dominance relationship in the section 3. The difference is that instead of showing all the match instances Fig 2 shows the corresponding centroids (in different colures) only. For example instances a1 and a2 are replaced by their centroid a_{ctr}. Fig. 2 also shows the application of Peano space filling curve for the dimensional reduction to single dimensional space. The points are extracted in to one dimensional space as they are traversed by the Peano space filling curve.

B. Lay out of heuristic algorithm for forming clusters.

Input: A graph representing the multi criteria service dominance relationship with *N* services each having *M* instances and *d* parameters.

Output: *l* clusters representing the different tradeoffs.

Heuristic Algorithm

Step 1: Select the services with skyline dominance score greater than zero. Let this set be S.

Step2: Find out the centroid s_{ctr} of all the objects $s \in S$.

Step 3: Insert s_{ctr} in a list L.

Step4: Sort L using the Peano space filling curve.

Step5: Divide L into *l* partitions of equal width.

Step6: Extract centroids of each partition as the cluster representatives.

V. EXPERIMENTAL SETUP

We make use of the publicly available service retrieval test collection OWLS-TC v2 (http://www-ags.dfki.uni-sb.de/_klusch/owls-mx/), which contains real-world Web service descriptions, retrieved mainly from public IBM UDDI registries. More specifically, our data set comprises: (a) 576 service descriptions,(b) 28 sample requests, and (c) a manually identified relevance set for each request.

For measuring the PDMs of the services, we use the OWLSMX matchmaker [9], which matches I/O parameters from the service descriptions, exploiting either purely logic-based reasoning (M0) or combined with some content-based, IR similarity measure. In particular, the following measures are provided: loss-of-information measure (M1), extended Jaccard similarity coefficient (M2), cosine similarity (M3), and Jensen-Shannon information divergence based similarity (M4).

VI. RESULTS

In this section we present the results of web service clustering based on multi-criteria service dominance relationship using the proposed method (Peano space filling curve) and also compares it with the existing clustering method (Hilbert space filling curve)

A. Webservice clustering using Hilbert Space filling curve(Existing method)

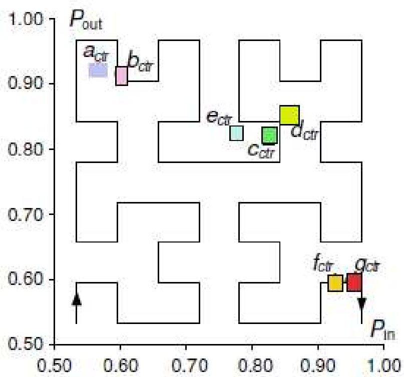

Fig.3 Dimensional reduction using Hilbert Space filling curve

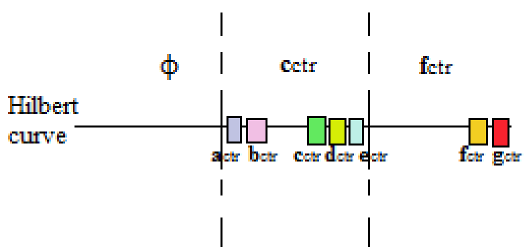

Fig.4 Output of Heuristic algorithm when Hilbert space filling curve is applied.

B. Webservice clustering using Peano Space filling curve(Proposed method)

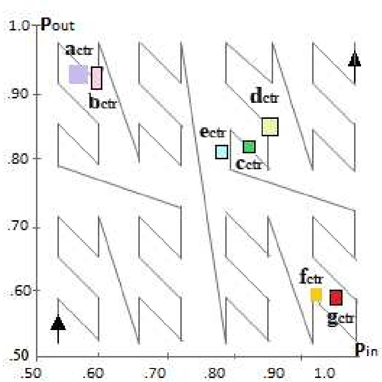

Fig.5 Dimensional reduction using Peano Space Filling curve

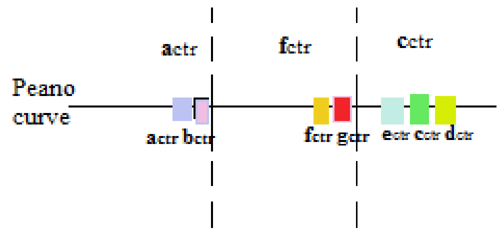

Fig.6 Output of Heuristic algorithm when Hilbert space filling curve is applied

C. Retreival Effectiveness and computational cost of the Proposed Method.

From the outputs of the Heuristic algorithm we can deduce that our proposed method, which uses the Peano Space filling curve for multi dimensional reduction finds out the exact clusters as compared to the existing one. From Fig.4 it is clear that the existing method fails identify any useful clusters in initial section, but our proposed method successfully identifies all the clusters. Moreover in a recent work [10] Mohammed F Mokbel and Walid G Areg has mathematically proved that Peano SFC is far better than Hilbert SFC , in the sense that it is showing less irregularity and more fairness in multidimensional reduction. So our proposed method does outperform the existing one.

Computational cost of the proposed heuristic algorithm can be computed as given below. Selecting all those services with a skyline score greater than zero requires comparing the instances of all the services with each other. The complexity of this process is $O(dN^2M)$. Sorting the selected objects S requires a single scan of S and partitioning it according to the order imposed by the Space filling curve. Hence the overall complexity of the Heuristic Algorithm is $O(dN^2M+|S|)$

VII. CONCLUSIONS

In this paper we have addressed the problem of clustering the web services so as to reflect the different tradeoffs among the different parameters. Proposed web service clustering is based on multicriteria service dominance relation which takes into consideration of multiple evaluating criteria without taking the aggregate of individual parameter matches. We proposed the use of Peano space filling curve for the dimensional reduction of multi-dimensional objects to a linear space for extracting the cluster representatives, which is expected to show less irregularity, more fairness and more scalability than the existing methods. Our proposed method gives more accurate results than the existing one.

REFERENCES

[1] X. Dong, A. Y. Halevy, J. Madhavan, E. Nemes, and J. Zhang, "Similarity Search for Web Services," in VLDB, 2004, pp. 372–383.

[2] Dimitrios Skoutas, Dimitris Sacharidis, Alkis Simitsis, Timos K. Sellis: "Ranking and Clustering Web Services Using Multicriteria Dominance Relationships." IEEE T. Services Computing 3(3): 163-177 (2010).

[3] M. Paolucci, T. Kawamura, T. R. Payne, and K. P. Sycara, "Semantic Matching of Web Services Capabilities." in ISWC, 2002, pp. 333–347.

[4] J. Pei, B. Jiang, X. Lin, and Y. Yuan, "Probabilistic Skylines on Uncertain Data," in VLDB, 2007, pp. 15–26.

[5] J. Colgrave, R. Akkiraju, and R. Goodwin, "External Matching in UDDI," in ICWS, 2004, p. 226.

[6] Yanan Hao and Yanchun Zhang, "Web Services Discovery Based on Schema Matching" ,in ACSC '07 Proceedings of the thirtieth Australasian conference on Computer science - Volume 62.

[7] Jian Wu and Zhaohui Wu, " Similarity-based Web service matchmaking", in 2005 IEEE International Conference on Services Computing

[8] J. Caverlee, L. Liu, and D. Rocco, "Discovering and Ranking Web Services with BASIL: A Personalized Approach with Biased Focus," in ICSOC, 2004, pp. 153–162.

[9] M. Klusch, B. Fries, and K. P. Sycara, "Automated Semantic Web service discovery with OWLS-MX," in AAMAS, 2006, pp. 915–922.

[10] Mohamed F. Mokbel, and Walid G. Aref, "Irregularity in Multidimensional Space Filling Curves with Applications in Multimedia Databases", ACM 2001.