Selecting Skyline Web Services for Multiple Users Preferences

Karim Benouaret, Djamal Benslimane
University of Lyon, LIRIS/CNRS
Villeurbanne, France
{*karim.benouaret, djamal.benslimane*}@*liris.cnrs.fr*

Allel Hadjali
Rennes 1 University, IRISA/ENSSAT
Lannion, France
allel.hadjali@enssat.fr

***Abstract*—In this paper, we introduce a novel concept called collective skyline to deal with the problem of multiple users preferences. We then conduct a set of experiments that demonstrate the effectiveness of the introduced concept.**

***Keywords*-Service selection; skyline analysis; preferences; optimization;**

I. Introduction

User preferences are important for helping users to identify desirable Web services [1], [2]. Moreover, in many practical situations, multiple users with different preferences need to make a collective decision.

Table I: Example of Services

Service	price	year	consumption	power
s_1	$[7000, 11000]$	$[2007, 2012]$	$[3.5, 5.5]$	$[60, 110]$
s_2	$[5000, 10000]$	$[2005, 2011]$	$[4, 6]$	$[70, 115]$
s_3	$[6000, 12000]$	$[2000, 2010]$	$[4, 6]$	$[70, 110]$
s_4	$[8000, 12000]$	$[2002, 2012]$	$[3.5, 5]$	$[75, 130]$
s_5	$[9000, 15000]$	$[2009, 2012]$	$[4, 7]$	$[90, 130]$

Example. Consider a set of car trading services in Table I. Each service has its constraints on its proposed cars. For instance, the cars proposed by s_1 have price in $[7000, 11000]$, and year in $[2007, 2012]$. Now, assume that three users, say u_1, u_2, u_3, want to buy a car. u_1, prefers cars having price in $[7000, 10000]$ and year in $[2005, 2010]$; u_2, prefers cars having consumption in $[3, 5]$; u_3, prefers cars having year in $[2008, 2012]$ and power in $[80, 100]$. As services proliferate, it is crucial to provide the users with the most relevant ones.

In this paper, we introduce a new selection criterion to address the problem of multiple users preferences.

II. Our Proposal

A. Finding Relevant Services

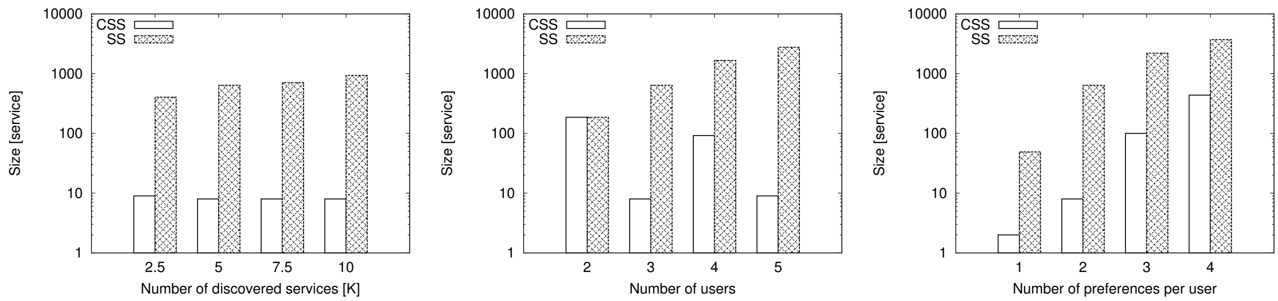
To determine the relevance of each service, we compute the matching degrees between users preferences and services' constraints using the Jaccard coefficient: $J(I_1, I_2) = \frac{|I_1 \cap I_2|}{|I_1 \cup I_2|}$. Table II shows the matching degrees of services in Table I w.r.t. each user preference attribute. For instance, 0.80 and 0.67 represent the matching degree of service s_1 w.r.t. preference attributes price and year of user u_1, respectively.

Table II: The Matching Degrees of Services

Service	u_1: (price, year)	u_2: consumption	u_3: (year, power)
s_1	$(0.80, 0.67)$	0.62	$(0.83, 0.41)$
s_2	$(0.67, 0.86)$	0.35	$(0.57, 0.47)$
s_3	$(0.57, 0.54)$	0.35	$(0.23, 0.51)$
s_4	$(0.50, 0.54)$	0.76	$(0.45, 0.38)$
s_5	$(0.57, 0.25)$	0.27	$(0.80, 0.27)$

B. Selection Criterion

Computing the skyline comes as a popular solution to select services [3], [4]. The skyline comprises the set of services that are not dominated by any other service. A service s_i dominates another service s_j if and only if s_i is better than or equal to s_j in all preference attributes, and strictly better in at least one preference attribute. For instance, service s_1 dominates service s_5. In our example, the skyline comprises services s_1, s_2, s_3 and s_4. However, a major drawback of skyline is that, when multiple users are involved, the number of services in the skyline becomes very large and no longer offer any interesting insights. The reason is that as the number of users and preferences increase, for any service s_i, it is more likely there is another service s_j where s_i and s_j are better than each other in different preference attributes. It is thus crucial to further reduce the size of the skyline. Different works have focused on this problem; e.g., [5]. However, the case of multiple users preferences is not taken into account. To this end, we propose the notion of *collective skyline*. Intuitively, it maybe reasonable to consider a service s_i better than another service s_j if s_i is better than s_j for most users. With this intuition, we propose a new extension of the dominance relationship called *collective dominance*: a service s_i *collectively dominates* another service s_j if and only if s_i dominates s_j for more than half of users. For instance, s_2 *collectively dominates* s_3. Then, the *collective skyline* comprises the set of services that are not *collectively dominated* by any other service. In our example, service s_1 *collectively dominates* services s_3, s_4 and s_5, while, services s_1 and s_2 are not *collectively dominated* by any other service. Thus, services s_1 and s_2 form the *collective skyline* – recall that the skyline comprises services s_1, s_2, s_3 and s_4. This shows that the *collective skyline* further reduces the size of the skyline.



(a) Effect of n (b) Effect of m (c) Effect of d

Figure 1: Effects of parameters on the size of the *collective skyline* and that of the skyline.

III. Experimental Evaluation

We focus on the size of the *collective skyline* (denoted as CSS). To demonstrate that the *collective skyline* further reduces the size of the traditional skyline, we also compute the size of the skyline (denoted as SS) to compare how the size of the *collective skyline* varies from that of the skyline. Table III displays the parameters under investigation and their corresponding ranges; default values are shown bold.

Table III: Parameters and Examined Values

Parameter	Symbol	Range
Number of discovered services	n	2.5K, **5K**, 7.5K, 10K
Number of users	m	2, **3**, 4, 5
Number of preferences per user	d	1, **2**, 3, 4

We computed the size of the *collective skyline* and the size of the skyline, varying the parameters. The results of this set of experiments are presented in Figure 1.

Comparing the size of the *collective skyline* and that of the skyline, we can observe that the size of the *collective skyline* is very smaller than the size of of the skyline. Thus the *collective skyline* further reduces the large size of the skyline, thereby, enabling users to make a good, quick selection.

As shown in Figure 1a, the size of the *collective skyline* is insensitive to n at all. This is because as n varies, more services may have chances not to be dominated on the one hand, and more services may have possibilities to be dominated on the other hand.

However, Figure 1b shows fluctuation in the size of the *collective skyline*. The fluctuation is related to the definition of the *collectively dominance* relationship. Indeed, with $m = 2$, the size of the *collective skyline* equals that of the skyline as a service s_i *collectively dominates* another service s_j if and only if s_i dominates s_j for the two users, i.e., $\frac{2}{2} = 100\%$ of users. Thus, the *collectively dominance* relationship is equivalent to the dominance relationship. Similarly, with $m = 3$, $m = 4$, and $m = 5$, a service s_i *collectively dominates* another service s_j if and only if s_i dominates s_j for respectively two, three, and three users, i.e., $\frac{2}{3} = 0.67\%$, $\frac{3}{4} = 0.75\%$, and $\frac{3}{5} = 0.60\%$ of users, respectively. These percentages (of "all most of users") are consistent with the fluctuation.

In contrast of n and m, the size of the *collective skyline* follows a similar trend as the size of the skyline varying d, i.e., it increases with the increase of d, as shown in Figure 1c. Since, as d increases, a service has better opportunity not to be dominated in all preference attributes w.r.t a given user.

IV. Conclusion

In this paper, we addressed the problem of preference-based Web service selection, in the context of multiple users preferences. We introduced a novel concept called *collective skyline*, which enables users to make a good, quick selection. Our experimental evaluation demonstrated the effectiveness of the introduced concept.

In the future, we plan to follow three directions for further work on this topic. The first one is to develop techniques for ranking the services returned by the *collective skyline*. The second direction is to carry out a user study to evaluate the quality of the results. Finally, we will investigate the case where the users preferences are incomplete.

References

[1] K. Benouaret, D. Benslimane, A. Hadjali, and M. Barhamgi, "Top-k web service compositions using fuzzy dominance relationship," in *IEEE SCC*, 2011, pp. 144–151.

[2] ——, "Fudocs: A web service composition system based on fuzzy dominance for preference query answering," *PVLDB*, vol. 4, no. 12, pp. 1430–1433, 2011.

[3] Q. Yu and A. Bouguettaya, "Computing service skylines over sets of services," in *ICWS*, 2010, pp. 481–488.

[4] K. Benouaret, D. Benslimane, and A. Hadjali, "On the use of fuzzy dominance for computing service skyline based on qos," in *ICWS*, 2011, pp. 540–547.

[5] C. Y. Chan, H. V. Jagadish, K.-L. Tan, A. K. H. Tung, and Z. Zhang, "Finding k-dominant skylines in high dimensional space," in *SIGMOD Conference*, 2006, pp. 503–514.