A Service Retrieval Assistance Mechanism Based on Association Mining

Bin Tang, Leqiu Qian, and Yunjiao Xue
Department of Computer Science and Engineering, Fudan University, Shanghai, China
{tangbin, lqqian, yjxue }@fudan.edu.cn
Hui Tang
the Guangdong Branch of China Mobile, Guangzhou

Abstract

During the process of service retrieval, it's often dif-ficult for users to give exact retrieval requirements be-cause users are not familiar with the complex description mechanism of services. This limits the function of ontol-ogy service models and leads to lower completion, preci-sion, efficiency and easiness of service retrieval. It is ur-gent to have an efficient method to help the users. The paper introduces a self-adaptive learning algorithm based on association mining theory in data mining field to learn from the retrieval history and assist users in giving high quality retrieval requirements. The experiment results show the effectivity of the proposed algorithm.

1. Introduction

Web service is a new kind of self-contained, self-de-scribed, modular and loose-coupled web application that can be advertised, discovered and invoked through the web. In other words, web services are a new kind of web component that is advertised once and can be invoked everywhere and support service reuse and composition, which is very important to improve the production effi-ciency and quality of web application [1][2].

With increasingly abundant web service markets, it is very important that the services needed can be retrieved quickly and easily. And the quality of service retrieval has direct influence on the quality of service reuse, the match-ing of service composition and substitution, and the plug and play of services. So as an important part of service-oriented system architecture, service retrieval receives much attention from industry and academia.

The performance of a service retrieval system is meas-ured from the recall, precision, efficiency and easiness. Recall is the ratio of the number of services fit users' need in the retrieval result and in the service registry center. Precision means the ratio of the service fit users' retrieval requirement in the retrieval result and the total number of services in the result. In the paper efficiency is broadly measured by the time from the user beginning to give a retrieval requirement to he selects the services needed finally, which can be divided into three parts: the time users use to give the retrieval requirement, the time the system use to retrieve services based on retrieval require-ment, the time the user use to select the best service from the result retrieval set. In literature, the retrieval time only include the middle part time. Easiness is seldom studied in literature but it is an important problem in practice and urgently needed to be improved.

A service retrieval method may in theory have high re-call, precision and efficiency, but if it cost user too much to give an exact and precise retrieval requirement, user may not take the efforts finally to give high quality re-trieval requirement the service retrieval method supposed, which can heavily influence the quality of the retrieval result. So it is urgent to give a mechanism to help user give high quality retrieval requirements if the expected superiority of advanced but also complex ontology service model is to be encased. In the paper we give such a mechanism to help user to give high quality retrieval re-quirements, which is verified through experiments.

The second part of the paper introduces the related work on service description, the third part of the paper introduces an adaptive service retrieval process model based on association mining and the core learning algo-rithm based on association mining and analyzes how the algorithm can help to improve the recall, precision, effi-ciency and easiness of service retrieval, the fourth part demonstrates the effectivity of the algorithm of through experiment, the last part gives the conclusion and feature work of the paper.

2. related work

Proceedings of the First International Conference on Innovative Computing, Information and Control (ICICIC'06)
0-7695-2616-0/06 $20.00 © 2006 IEEE

According to the richness of the information contained in the service description model, web service description and retrieval models can be divided into two categories [2]:

- Syntactical level

In description, this category of models emphasize the Syntax of the interface of web service and lack of con-straints of behavior. In retrieval, them mostly base on key words. The representative systems based on this category of models are the UDDIs of IBM, Microsoft, SUN, which are simple to realize but the retrieval effect are not desir-able [7][9].

- Semantic level

This kind of service models introduce ontology to de-scribe services thoroughly and provide description of the function and behavior, which can be expected to give good retrieval result in theory. The matching of retrieval requirement and service description is based on descrip-tion logics and usually hundred-percent matching can not be obtained, which can be divided into five classes ac-cording to the match degree: exact match, generalized match, specialized match, partial match and no match. Recently there are researches to calculate matching degree based on similarity counting, this kind of methods calcu-late the matching degree as the powered sum of the simi-larity degree of the terms in the retrieval requirement and service advertisement [2]. The latter kind of methods for calculating matching, which is adopted in the paper, has higher efficiency and recall and can deal with incomplete retrieval requirement and can let the user better know how the retrieved components fit his need. Representative studies include the augment UDDI Registry of Carnegie Mellon University [7], the Meteor-s project University of Georgia [6], the OWL-S ontology service standard of W3C [4], WSMO working group's WSMO ontology ser-vice standard [8] and so on. Ontology service model are expected to have good retrieval result, but it is very com-plex and difficult for user to understand and use. It is nec-essary to have an auxiliary mechanism to help user to give retrieval requirement.

3. An adaptive service retrieval process model based on association mining

The attributes of a service model are often not isolated and the values of some attributes will influence the values of the values of some other attributes. This kind of asso-ciations between of attribute values can be determined by the domain experts or through association mining algo-rithms. In 3.1 we will introduce an adaptive service re-trieval process model and the association mining algo-rithm as the core of the model and in 3.2 we analyze the model..

3.1 An adaptive service retrieval process model based on association mining

As in figure 2, the retrieval requirement of a user is sent to both the retrieval system and the learning system and the retrieval result of the user is also sent to the learning system. The learning system analyzes an amount of retrieval requirement and corresponding retrieval result and obtains association rules of the form $A \Rightarrow B$ (Due to limited space, interesting can refer to [3] for association mining), where A is the attribute-values appeared in the user's retrieval requirements and B is the attribute-values appeared in the descriptions of the retrieval result, the probability of A and B appearing in the same one retrieval exceeds the given threshold minsupp, and the probability that the retrievals requirements contain A should also contain B exceeds the given threshold minconf. B is the requirement that the user probably needed. Is/Os are seen as single attribute respectively in the paper.

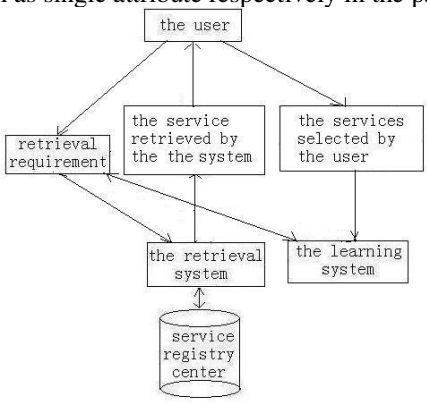

Figure 1. The model of adaptive service retrieval system

The learning system delivers the obtained association rules to the retrieval system. In later retrieval, if A appeared in user's retrieval requirement, then the retrieval system adds B to the retrieval requirement after the user confirms, which is carried out during users give their retrieval requirement.

The learning system is the core part of the adaptive service retrieval process model. The following introduces the core algorithm of the learning system—adaptive service retrieval algorithm based on association mining.

Basic concepts: The user's retrieval requirement is denoted as Q (which does not include the part of the retrieval requirement obtain by association rules), and there are m services selected by the user, which are

described by D_i respectively. The learning system maintains a database H of all past retrieval data, for every retrieval, the retrieval requirement Q and the D_i of every services selected by user are put into the database H.

The association rule of the retrieval requirement given by the user and the probable implicit requirement is of the form $A \Rightarrow B$, $A \cap B = \phi$. $A \Rightarrow B$ is satisfied in database H with support s and confidence c, the former is the percentage of the records containing $A \cup B$ in H, scilicet probability $P(A \cup B)$, the latter is the is the percentage of the records containing B in the set of records which contain A. The s should be larger than minsupp and the c larger than minconf.

The algorithm: Self-adaptive retrieval learning algorithm based on association mining adopts the Apriori to search the frequent attribute-value sets in the user's retrieval requirement and obtains information from the corresponding services selected by the user. The user's retrieval requirement Q and the description of each corresponding service selected by the user D_i are stored in the retrieval history database H and denoted as H_Q and H_D respectively. The method to obtain association rules of the algorithm is slightly different from wont in literature. The algorithm first utilizes the Apriori algorithm to select the set QL of frequent attribute-value sets contained in H_Q satisfying minsupp. For every set ql in QL, obtain the set DL of frequent attribute-value sets in the corresponding service selected by the user with frequency not less than the total number of the corresponding service selected multiply minconf. Select every set from DL and together with ql to form an association rule.

Algorithm: Self-adaptive Learning Algorithm Based On Association Mining

Input: retrieval history database H，minimum support threshold minsupp, minimum confidence threshold minconf

Output: association rule sets R

Method:

(1) Find frequent retrieval attribute-value sets QL in Q using Apriori

(2) for (i=1； QL≠ϕ， i++){

(3) fetch one frequent attribute-value set ql in QL whose corresponding user feedback retrieval result set is D

(4) find frequent attribute-value set DL whose support counting is larger than count(D)×minconf in D using Apriori

(5) Select every frequent attribute-value set dl from DL and together with ql to form a association rule $ql \Rightarrow dl$ and put into R

(6) remove ql from QL

(7) }，

Figure 2. Adaptive service retrieval algorithm based on association mining

The learning system sends the association rule set R to the retrieval system which will add the attribute-value sets on the right side of applicable association rules to the retrieval requirement when carrying out a retrieval.

4. Analysis of the assistance and the goodness of the learning algorithm

The assistance and the goodness to the quality of re-trieval result of the algorithm introduced in the paper are analyzed as following:

- Because of the better completion of the retrieval requirement based on association rules, the precision of the retrieval result is improved.
- Because with the assistance of the system, the retrieval requirement is refined and of high quality, and the matching of retrieval requirement and the advertise-ment of services based on similarity computing, which will set a threshold of similarity for the service to be in-clude into the result set, the refinement of the retrieval requirement means the recall of the retrieval will be im-proved.
- The easiness of the retrieval is improved because some attributes of the retrieval requirements especially the difficult ones for users can be automatically obtained by the association rules and the precision of the retrieval re-sult is improved and the base of the result set is reduced so it is easy for the user to select the services needed.
- If define the service retrieval as three phrases in a broad sense: the phrase the user gives an initial retrieval requirement and refine it with the help of the system, the phase the system retrieves service based on the retrieval requirement, the phrase the user selects the service needed. The improvement of the efficiency is come from the first and third phrase with the same reason as above

5. Experiment

We have implemented a prototype service retrieval sys-tem X-com and verify the algorithm introduced in the paper. We select two hundred and eighty services such as mobile phone message services, data mining services and calculate the matching of retrieval requirement and the description of services based on similarity computing. One hundred service retrievals are carried out without the learning algorithm

introduced in the paper, the recall, pre-cision and efficiency is calculated for every ten times. The average recall, precision and efficiency, which is defined as the time sum of the three phrase, of these service re-trievals are around 0.69, 0.62 and five minutes and ten seconds respectively. When adopting the learning algo-rithm and the same experiment is carried out again, the recall, precision and efficiency of service retrieval in-crease with the learning of the retrieval history as the fig-ure 4, 5, 6 respectively,

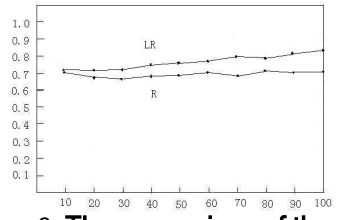

Figure 3. The comparison of the recall

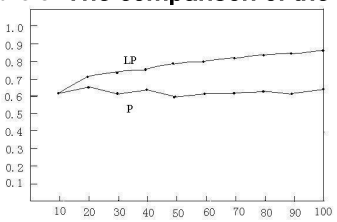

Figure 4. The comparison of the precision

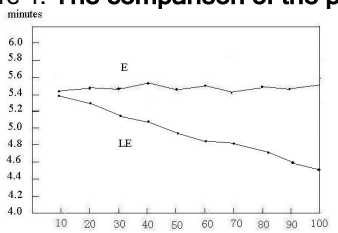

Figure 5. The comparison of the efficiency

where R, P, E Represent recall, precision and efficiency before adopting the learning algorithm respectively and LR, LP, LE Represent recall, precision and efficiency after adopting the learning algorithm respectively. The satisfaction of the user to the prototype system X-com is increased.

6. Conclusion and feature work

The description and retrieval of services on syntactical level is easy to realize but the effect leave much to be ex-pected. Ontology service models attempt to overcome this by utilizing ontology technology and there are many stud-ies devote this. But because of the complexity of the on-tology service models, it is difficult for users to under-stand and apply them. It is urgent to have an auxiliary mechanism to assist users to give high quality retrieval requirement to bring ontology service model into play really.

The paper introduces a self-adaptive learning algorithm based on association mining theory in data mining field to obtain the internal relationships between users' explicit retrieval conditions and their implicit requirements through mining in the retrieval history and assist users in giving high quality retrieval requirements The experimen-tal results show the effectivity of the proposed algorithm. The feature work is to optimize the efficiency of the algo-rithm and verify and improve the algorithm further.

Acknowledgements

This paper is supported by the National Natural Science Foundation of China under Grant No. 60473062.

7. References

[1] Yang F.Q., Mei H. The software technology of the Internet era. chinese Sciencepaper Online, http://www.paper.edu.cn/scholar/paper.jsp?name=meihong

[2] Hu JQ,Zhou P,Wang HM,Zhou B. Research on Web Service Description Language QWSDL and Service Matching Model. Chinese Journal of Computer, Vol.28, No.4, pp.505-513, 2005,

[3] Jiawei Han, Micheline Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, 2000.

[4] W3C. OWL-S: Semantic Markup for Web Services. http://www.w3.org/Submission/OWL-S/ , 2004.11.

[5] Mark H. Burstein. Dynamic invocation of Semantic Web Services that use unfamiliar ontologies. Intelligent Systems, IEEE. Volume 19, Issue 4, pp.67-73, Jul-Aug 2004.

[6] Large Scale Distributed Information Systems Lab of the Georgia University, METEOR-S, http://lsdis.cs.uga.edu/Projects/METEOR-S.

[7] Massimo Paolucci, Takahiro Kawamura, Terry R. Payne, Katia Sycara. Importing the Semantic Web in UDDI. In Proceedings of Web Services, E-business and Semantic Web Workshop, Toronto, Canada, pp.225~236, 2002,.

[8] WSMO working group. D2v1.2. Web Service Modeling Ontology (WSMO). www.wsmo.org/TR/d2/v1.2/D2v1-2_20050414.pdf . 2005,4.

[9] Dogac A, Kabak Y, Laleci GB. Enriching ebXML registries with OWL ontologies for efficient service discovery. Proceed-ings of the 14th International Workshop on Research Issues on Data Engineering, pp.69–76, 2004..