Agent based discovery of web service to enhance the quality of web service selection

S.Susila,

S.Vadivel department of computer science BITS PILANI DUBAI, UAE
PDPTA

Abstract:
Web services play a major role in developing business application in this internet era. As there are many web services available for a particular function, finding appropriate web service based on our criteria is becoming a tedious process. Currently, finding the exact web service based on the keyword based search method of UDDI registry doesn't return the web service of user interest at all times. In this paper we have discussed an agent based method for web service selection, from the information that is given in the wsdl file by the web service provider. Data mining is done on those data that are collected from wsdl files and feedback taken from the web service users, by the agent to discover some interesting patterns for further users of the web service.

Keywords :
Web Service; Qos; Non-functional attribute; SOA; Weka

1. Introduction:

1.1 SOA:

Service oriented Architecture (SOA) is an architectural design for building software applications that use services available in a network such as the web. It does loose coupling between software components so that they can be reused. Services are software components with well defined interfaces that are implementation independent. The important aspect of SOA is the separation of the service interface from its implementation. Such services are used by the client without knowing how these services will execute their request. Web services have played major role in implementing business level applications based on SOA [1].

1.2 Web service:

A Web service is a software application identified by a URI, whose interfaces and binding are capable of being defined, described and discovered by XML artifacts and supports direct interactions with other software applications using XML based messages via internet-based protocols. Web service technology is finding its application in many areas such as electronic commerce, application integration, scientific research, flow management, etc. [2]. Web services realize the usage of distributed application on the internet. This is a platform to realize distributed SOA.

Web service is supported by multiple standards, such as Simple Object Access Protocol (SOAP), Web Service Description Language (WSDL), and Hyper Text Transport Protocol (HTTP). There are five basic specifications, serve as the foundation of the web service protocol.

Figure 1: protocol stack of web service

UDDI: UDDI, Universal Description, Discovery and Integration, defines a way to publish and dynamic discovery of services in the registry.
WSDL: XML based language for describing web service. This gives the abstract definition of operation and message in the web service. WSDL document is an interface contract between service provider and service consumer.
SOAP: Simple object access protocol is XML based protocol used in communication among services. This is an envelope for the messages that are exchanged between services.

1.3 Data mining:

Data mining is the process of extracting knowledge from data. That knowledge can be used to understand the nature of a business or scientific problem, or applied to new data to make predictions or classifications. It is the process of applying computer based methods for sorting through large sets of data for discovering actionable information such as patterns and trends that cannot be done by simple analysis.

Manuscript received February 5, 2011
Manuscript revised February 20, 2011

This includes techniques like preprocessing, classification, clustering etc.

- Classification: Classification analysis is the union of data in particular classes. Also known as supervised classification, the classification uses specified class labels to direct the objects in the data compilation. Classification approaches in general use a training set where every object is previously linked with known class labels. The classification algorithm discovers from the training set and constructs a model. The model is used to categorize new objects.
- Clustering: clustering also known as unsupervised classification is the process of finding groups of objects such that the objects in a group will be similar or related to one another and unrelated to the objects in other groups [3]. In simple K means algorithm, objects are clustered according to the distance calculation of defined attribute.

1.4 Web service selection:

Users developing service-oriented application and integrating existing applications are deploying huge number of web services. So day by day the number of web services available is growing in size exponentially. So selecting a quality web service is becoming an issue. The attributes of interest in a web service can come under two broad categories as functional attributes and non-functional attributes. In the case of functional attributes the attribute describes about what the web service is about with respect to domain specific details. Functional properties include the input, output, Conditional output, precondition, access condition and effect of service. These can be well understood only by the users who have the interest in the specific business domain, where as non functional (NF) attributes describes about how the web service is in terms of quality. Here web service selection done based on the NF attributes of the web service.

This paper makes mention of related work in section 2. Extended SOA is explained in section 3. The Non functional attributes that are taken for consideration are given in section 4. Section 5 gives introduction of WEKA and application of the same for service selection. Section 6 presents interesting data mining patterns discovered based on the data given by the service user after using the service. The conclusion and direction for future work is presented in section 7.

2. Related work:

As the number of Web services for any specific domain increases in the Internet, QoS based Web service selection has become a hot research. In literature [3], the existing service registry is modified to include service reviews and rating details along with service information. Literature [4] provides an approach to combine Artificial Intelligence specifically data mining technologies and service rating techniques into SOA infrastructure for service mining. Literature [5] proposes a framework that utilizes user preferences as an additional input to the selection engine and system ranks the available services based on the requirement. In [6], the authors proposed a web service discovery model where functional and non-functional requirements are taken into account. However, feedback can not be collected from service users to update the QoS database value.

3. Extended Service Oriented Architecture:

The SOA is extended in such a way that there is an agent in between, who run data mining on the QoS attributes available to him to discover some interesting web service selection ways. Here for each functional domain e.g. weather forecast, there might be many number of web services available. In our example, there are going to be 10 web services for the same functional domain.

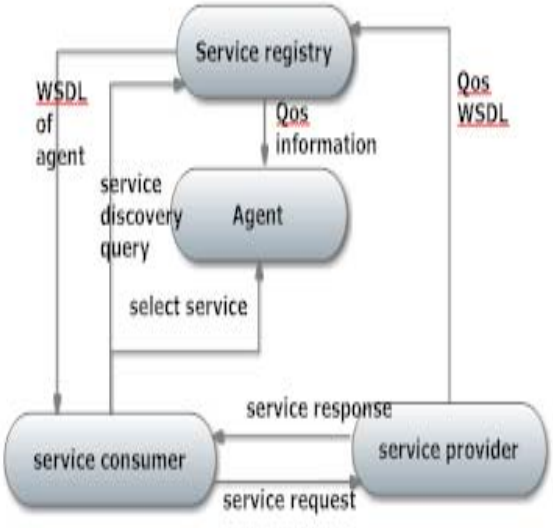

Figure 2: Extended SOA for WSS

As shown in the fig. 2, there exists an agent for this weather forecast. All the wsdl files provided by the web service provider ws1, ws2…ws10 have values for the QoS attributes that are selected. This quality metric information is extracted from the wsdl file by the agent is shown in fig 3. Web service client after using the service, logs his comments and feedbacks to this agent. Now the agent discovers very interesting pattern after mining the QoS attributes available to him.

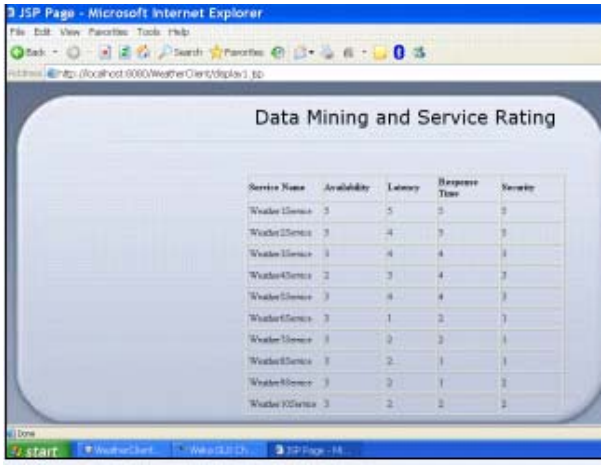

Figure 3: table maintained by the agent

4. Nonfunctional attributes that be taken for consideration for QoS of web services.

Availability: it is the degree of readiness of the service for immediate consumption.
Response Time: time elapsed from the Submission of a request to the time the response is received.
Cost: this represents price that a consumer of a Web service must pay in order to use the web service.
Security: This is ability of the web service to provide security mechanisms like encryption, authentication, and access control.
Latency: It is the time delay involved in start servicing the request of the client.
Scalability: number of throughput at given interval.
Stability: service quality and maintenance of the service and service quality [7].

The part of wsdl file given by the service provider which has got these quality metric tags as given in the fig.4

```
  <Qos>
<Availability>5</Availability>
<Latency>5</Latency>
<ResponseTime> 5</ResponseTime>
<Security>5</Security>
<Price>5</Price>
<Accessibility>5</Accessibility>
<Reliability>5</Reliability>
<Authentication>no</Authentication>
</Qos>
```

Figure 4: Part of wsdl file

5. WEKA:

Weka (Waikato Environment for Knowledge Analysis) is a compilation of machine learning algorithms for data mining tasks. The algorithms can be used straightforwardly to a dataset. Weka includes tools for data classification, pre-processing, clustering, regression, visualization, and association rules. It is also compatible for creating innovative machine learning schemes. It was built up at the University of Waikato, New Zealand.
The main graphical user interface, the "Explorer," has six different panels, accessed by the tabs at the top that correspond to the various data mining tasks supported. In the "Pre-process" the dataset which is in the ARFF format is read and various data preprocessing tools called filters are applied first to the data. The Explorer's second panel, called "Classify," classification and regression algorithms can be applied to the preprocessed data. Classification algorithms typically produce decision trees or rules, the third panel, "Cluster," enables users to apply clustering algorithms to the dataset.
There are 7 NF attributes taken for as quality metric for the web service. The values for the attributes might be from 1 to 5. 5 specifying the most preferred characteristic e.g. if the response time is less, it is given value 5 and if the response time is too high it is given value 1 for a particular web service. These values which are specified in the wsdl file by the web service provider is extracted by the agent and there are going to be 4 classes in our classification of web services like platinum, gold, silver and bronze. The wsdl file from ws1 is shown in fig 5. If all the attributes are having values 5, it will come under platinum class. If all the values are 4 it is gold, if the values are between 3 and 2 it will come silver and the value 1 in all attribute make the web service to come in class bronze.
Weka is applied on these data and preprocess tool in weka produces output which can be used by the classifier. Figure 6 shows the J48 classification obtained in this way. So from this pattern that we got, if we get a new instance we can only find out the class it belongs to. In weka simple K means clustering is selected to produce clusters which are shown in fig 7. From right clicking on the instance, we can find out which instance or web service is coming under what class as shown in fig8. So when ever client wants to call a web service for particular domain in our case weather forecast, we can call the agent for that and the agent will infer about quality of the entire web services available for weather forecast.

Feedback from Web service client:
The few attributes that we have discussed here like security, speed, latency, response time will be very much acceptable if these all given by the service user rather than service provider. So option is given for the client to enter his views about attributes like response

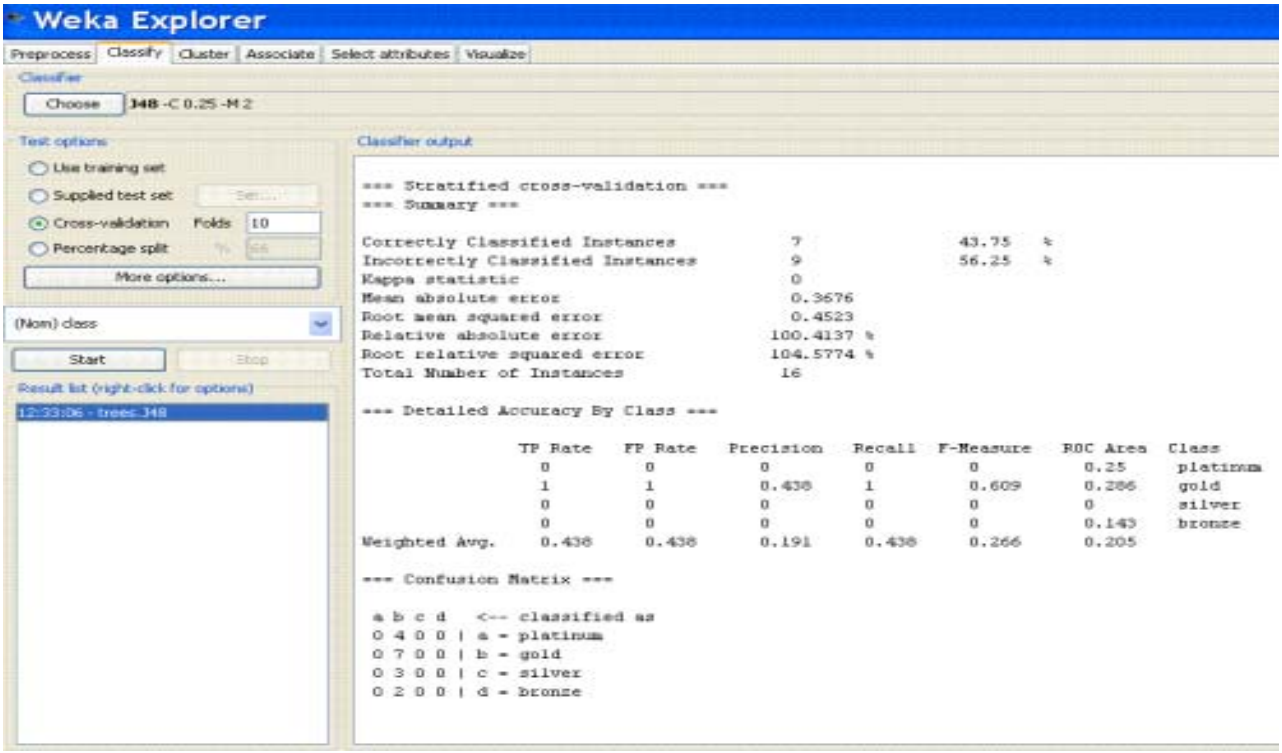

Figure 6: Classification output

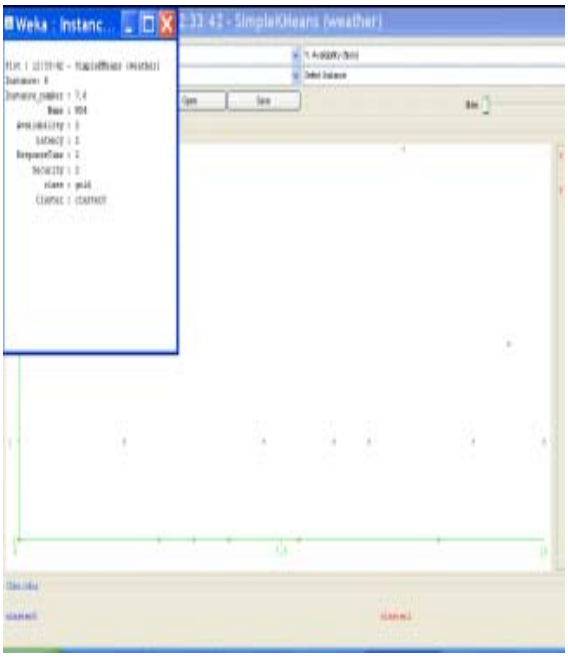

Figure 7: Cluster output

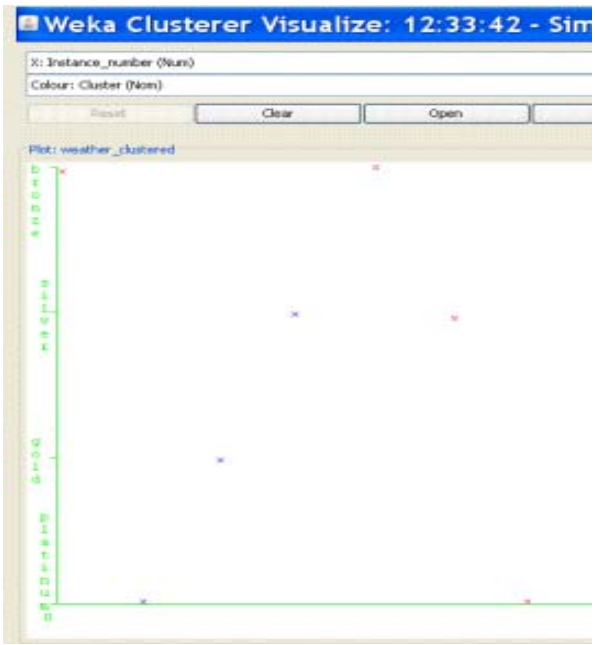

Figure 8: Clusters showing four groups

time, latency, throughput and speed. So he is directed how to evaluate the web service to give values 1 to 5. The feed back that the user wants to give about these attributes are collected in a user log access file. Then data mining can be done on these, to extract the information that is required by other service users.

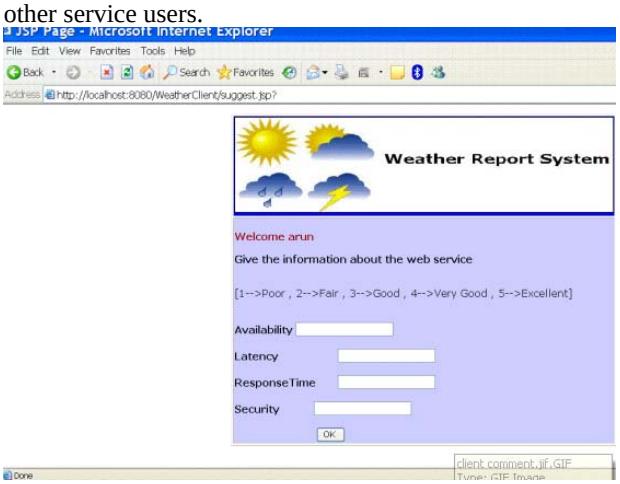

Figure 9: client feedback form

Agent does analysis done on these feedback given by the service user and produces bar chart output the for each attribute and so, the further users which ever attribute they are particular about can have a view that bar chart and accordingly they can invoke the service.

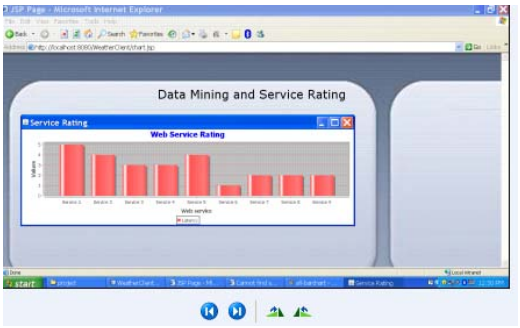

Figure 10: Bar chart view of Qos information

Future work:

Here ,there are only values from 1 to 5 is entered after using the service arbitrarily by the client .for example if the response time ids good ,client is asked to enter 1,if he feels it is taking long time he can select any value between 3 and 5 which involves too much of approximation. But if one could actually measure each attribute like response time, latency etc and give exact time taken to further improve the quality of service selection.

Conclusion:

Eyhab Al-Masri counted and statistically analyzed the growth rate of web services. Finding web services through |WSDL reference had grown by 231% while 63% of the available web services on the internet were considered active [8, 9, 10]. In this way instead of going through the UDDI, for all web services available for the weather forecast, we can get the address of the agent for weather forecast which will guide us further for web service selection.

References

[1] D. Krafzig, K. Banke, and D. Slama. Enterprise SOA: Service-Oriented Architecture Best Practices (The Coad Series). Prentice Hall PTR, Upper Saddle River, NJ, USA, 2004

[2] K. Yue, X. Wang and A. Zhou, “Underlying techniques for web services: A Survey,” journal of software, vol. 15, no.3, 2004, pp.428-442

[3] Tewari, V.; Dagdee, N.; Singh, I.; Garg, N.; Soni, P.; , "An Improved Discovery Engine for Efficient and Intelligent Discovery of Web Service with Publication Facility," Services - II, 2009. SERVICES-2 '09. World Conference on , vol., no., pp.63-70, 21-25 Sept. 2009

[4] SungHeun Nam; YunHee Kang; , "XML Schema Design for Web Service Quality Management," Future Generation Communication and Networking, 2008. FGCN '08. Second International Conference on , vol.2, no., pp.99-102, 13-15 Dec. 2008

[5] Badr, Y.; Abraham, A.; Biennier, F.; Grosan, C.; , "Enhancing Web Service Selection by User Preferences of Non-functional Features," Next Generation Web Services Practices, 2008. NWESP '08. 4th International Conference on , vol., no., pp.60-65, 20-22 Oct. 2008

[6] V. Diamadopoulou, C. Makris, Y. Panagis and E. Sakkopoulos, “Techniques to support web service selection and consumption with QoS characteristics”. Journal of Network and Computer Applications, vol. 31, 2008, pp. 108-130.

[7] SungHeun Nam; YunHee Kang; , "XML Schema Design for Web Service Quality Management," Future Generation Communication and Networking, 2008. FGCN '08. Second International Conference on, vol.2, no., pp.99-102, 13-15 Dec. 2008

[8] Al-Masri, E., and Mahmoud, Q. H., "Discovering the best web service", (poster) 16th International Conference on World Wide Web (WWW), 2007, pp. 1257-1258. (for QWS Dataset Version 1.0 or QWS Dataset Version 2.0)

[9] Al-Masri, E., and Mahmoud, Q. H., "QoS-based Discovery and Ranking of Web Services", IEEE 16th International Conference on Computer Communications and Networks (ICCCN), 2007, pp. 529-534. (for QWS Dataset Version 1.0 or QWS Dataset Version 2.0)

[10] Al-Masri, E., and Mahmoud, Q.H., "Investigating Web Services on the World Wide Web", 17th International Conference on World Wide Web (WWW), Beijing, April 2008, pp. 795-804. (for QWS-WSDLs Dataset Version 1.0)