Categorization of web services based on QOS constraint using Decision tree classifier

J.Ghayathri
Ph.D Research Scholar, Department of Computer Science,
Erode Arts & Science College (Autonomous), Erode, Tamil Nadu, India.
Email: rsrajakumarid@gmail.com
Dr.S.Pannirselvam
Associate Professor & Head, Department of Computer Science,
Erode Arts & Science College (Autonomous), Erode, Tamil Nadu, India.
Email: pannirselvam08@gmail.com

Abstract— The rapid detonation of web services technology in our businesses and day-to-day lives, to satisfy user requirements the composite services are formed by combining selected web services. Web Services are emerging technologies that flows the mechanism of communication between the electronic machines and reuse of component of services over the web. The selection of web services to satisfy the requirements of the consumers and to form the composite service made using Quality of Services (QoS). There is numerous service providers provide web services in the similar context, the web service selection based on QoS classification becomes incredibly essential for the consumer. As the QoS are significant features that decide the accuracy and efficiency of the web services to be selected, they are classified accordingly. In the proposed method based on decision trees the classification of web services has been applied. The modified algorithm of C5 classifier is used to do the classification based on QoS parameters. The confusion matrix is used to compute the accuracy of classification. It gives the higher accuracy in the selection of proper web service as per the need of consumers.

Keywords- Web services, QoS, Decision Tree, Classification, Confusion Matrix.

1. INTRODUCTION

Web services are reusable software components which require less interaction from the human being. This loosely-coupled web services use the technologies such as WSDL Web Service Description Language), SOAP (Simple Object Access Protocol), BPEL (Business Process Execution Language) ,UDDI (Universal Definition and Description Language) and HTTP (Hyper Text Transfer Protocol) to transfer messages and data among the services and applications.

In a web services environment service provider publishes the service descriptions in the public or private registry using WSDL and the service consumer discovers the service from the service repository through the communication protocols SOAP and HTTP. When the user finds the right service the information regarding binding is provided to the consumer in the form of WSDL and the consumer's application is bound to the service providers' web service to obtain the required service. The service-oriented architecture facilitates the service binding through dynamic and static methods. As the static method is rigid and pre-defined, it cannot be used in on-the-fly applications. The dynamic binding of services, supports the users to find, to select and to invoke the services at runtime. To find and select the most suitable service functional and non-functional parameters are used. In addition to that various methods support the optimal selection of services. When the requirement is not fulfilled by a single service more than one service are combined to satisfy the requirement of the user. Composition of web services is made through modeling and AI planning techniques. In this research work, we aim to design a plan for decision making in choosing the web services based on QoS using the Decision trees.

Decision tree approach is a supervised classification technique. It has simple structure with non terminal nodes representing tests on one or more attributes and the terminal nodes reflect decision outcomes or class labels. In order to classify unknown sample, its values are tested against decision tree. Decision trees can be easily converted into decision rules. Unlike Neural Networks, decision tree methods are able to identify independent variables through the built tree and basic functions when many potential variables are considered [16]. When the dataset is huge they can save lots of modeling time since they do not need a long training process.

2. RELATED WORKS

The authors J. Ghayathri, S. Pannir Selvam [1] proposed Selection of paramount web services based on ranking of QoS constraint. In that the selected QoS parameters as taken into consideration against the users' constraints and based on ranking the web services are selected. In [2] Susila, S et.al measured the QoS based on ID3 algorithm and using the decision tree the services are selected. ID3 inherently uses entropy based discretization for creating pure bins out of the training dataset. But the proposed algorithm uses a variation of ID3 algorithm to induce the decision tree to enable decision tree classification for continuous datasets. In [3] Venkataiah Vaadaala et.al. have applied the J48 algorithm to select the web services based on QoS and confusion matrix analysis. In [4] A. S. Galathiya et. al., discussed the C5 classifier and its Pseudo code and the comparison with the earlier versions of the algorithm. In [5] A. S. Galathiya et. al., explains the cross validation, model complexity and decision tree induction in detail. A ranking model [8] is proposed to rank and recommend a web service using artificial neural network by measuring QoS parameters. It proposes a principal component analysis (PCA) method for initial attribute weight then gives training algorithm for weight adjusting based on neural network. Although neural networks take long time for training for large datasets but it was shown that[5] at a starting point neural networks could be used to discover and rank the web service Naïve based Bayesian network [9] can also be used for classification of the services.

Student qualitative data has been taken from educational data mining and the performance analysis of the decision tree algorithm ID3, C4.5 and CART are compared by T.Miranda Lakshmi et. al [10]. The comparison result shows that the Gini Index of CART influence information Gain Ratio of ID3 and C4.5. The classification accuracy of CART is higher than ID3 and C4.5. However the difference in classification accuracy between the decision tree algorithms is not considerably higher. The experimental results of decision tree indicate that student's performance also influenced by qualitative factors [10]. Multicriteria Evaluation Component (MEC) is added in the registry of web services architecture [11] for evaluation. A set of preference parameters are used in the evaluation to satisfy the user requirements.

The authors V. Estruch et. al. [12] presented the distance based decision tree learning algorithm (DBDT), which is used in web categorization by means of metric conditions as splitting criterion. It allows decision trees to handle structured attributes such as lists, graphs, sets along with the well-known nominal and numerical attributes. These structured attributes have been used to represent the content and the structure of the web-site [12]

The authors Mamoun Mohamad Jamous et. al. [13] classified and stored the web services into classes according to non-functional criteria. The classes are predefined and belong to different criterions. Classification attributes values are provided by web service provider during registration of the web service to the registry. A classification algorithm depends on information supplied by web service provider at the registration time is proposed. Also the usefulness and efficiency of the proposed algorithm has been proved mathematically and experimentally [13].

The authors Zeina Azmeh [14] presented the WSPAB tool for the automatic classification and selection of web services depending on an online web service repository. This tool queries the service repository to find a first set of candidate services and filters this service set according to functional and nonfunctional criteria. It extracts the operation signatures of the services from the resulting set in order to further filter them according to this syntactic information. Finally, the set of remaining services is classified into a service lattice using Formal Concept Analysis. The obtained lattice can be used to identify both the service that best adapts to the user's needs and its possible substitutes when needed. [14]

The authors Rama Kanta Mohanty et. al. [15] used Naïve Bayes, Markov blanket and Tabu search techniques to classify the web services. The average accuracy of Naïve Bayes classifier is greater than Tabu search and Markov blanket. The Back propagation trained neural network has been applied to find the importance of different attributes in web services. It is concluded that Bayesian network is a very good classifier to classify classification type of problems when compare with Marcov Blanket and Tabu serach [15].

3. METHODOLOGY

1. Web services and Quality of Services

The dynamic e-business visualization requires a perfect combination of business processes, web services, and applications over the internet. Carrying out QoS on the internet is a vital and major challenge because of its vibrant and changeable nature. The dynamic electronic business idea requires a perfect arrangement of business procedures, web-services, and functions on the web. Implementing quality of service on the web is an essential and main test due to its exciting and variable character. With web services proliferating, QoS is a major factor to differentiate web services and providers. In selecting a web service, its non-functional properties should be considered to satisfy user's requirement constraints. QoS concludes a comprehensive selection of processes that are comparable to the needs of service-requester with those of the service-publisher on the basis of the network properties available.

Web Services are outcome of the advancement of the web into a means of scientific, commercial and social exchanges. A Web service can be described as a way of calling a function which is inside software from software. The software which makes the call is called as the client and the software which services the client is called a server. The two softwares might have been programmed using different languages and could be running on different machines but have to be connected by a network. Web services have an interface expressed as the WSDL (Web services description language) file. WSDL file can be seen a contract for communications between the web service client and server. The success of web services depends on the functional and non-functional requirements of the users which are significant criterion to be fulfilled.

Functional requirements describe how the system behaves within the problem domain. Non-functional requirements describe how the system behaves from a technical perspective. They are independent of the problem domain. For any application non-functional requirements are expressed in the same terms. If data needs to be replicated to a different location, that's a non-functional requirement. When a new page is added in a new web site, it is changing the non-functional requirements which includes the quality attributes such as response time, reliability, scalability, throughput, robustness, success ability, exception handling, reliability, accuracy, integrity, accessibility, availability, interoperability, security, and network-related QoS requirements. The aim of this system is to classify the web services based on the response time, availability, throughput, success ability and reliability and hence help the service requestor choose a web service which best suits the users' requirements. However, QoS categorization is very helpful for web services clustering and filtering, which highly helps end user on making decision of what web service to choose among a group of similar functionality web services. Classification of web services is the act of grouping similar web services into groups. The similarity among a group of web services depends on different criteria. Classification enhances the speed of web service discovery process. Moreover, classification of web services increases the accuracy of discovering the right service for the specified need. Web services can be classified in different criteria [13]. In this research, the web services are classified based on their non-functional qualities.

The explanation on QoS parameters and its measurements used in this proposed system are specified in the Table-1.

S.No	QoS Param.	Description	Computation Formula
1	Response Time	The time taken by the web service for responding the given request. It is used to grade the web service. The lesser response latency is preferred by web service consumers. (Milliseconds)	Time taken to complete the response – Time taken for user request
2	Availability	This is the probability that the system is up and read for immediate consumption when the service is invoked. Service providers should provide their web service with high availability ratio so as to satisfy the customer. -(Percentage)	1 – (Down time/ Unit time)
3	Success ability	Ability of the web service to give the service to the consumers' requests. -(Percentage)	Number Of responses / Number Of requests
4	Reliability	Ability of a web service to execute its required functions under the given conditions for a particular time interval.-(Percentage)	f(aC,bF,cT,dI,eA,fP) Accuracy(C),fault-tolerance(F), Testability(T), interoperability (I), availability (A), and performance (P). a, b, c, d, e, and f are the weights of each attribute.

Table -1

2. Categorization of web services

In the world of internet, web services plays vital role in providing services to the users. When the service providing activity is automated there is a need for any organization to publish the corresponding service in the public or private repositories for the consumption. This makes it to enhance the activities of transaction between the provider and the consumer. However, when numerous services are published in the web, selecting the correct service for a process becomes a big issue and complex for the consumer. Even though number of methods available for selection of services, the availability of large volume of services formulates the challenge in selection of services. To overcome this issue the decision tree which is one of the artificial intelligence approaches for classification is used.

The decision tree learning is one of the most widely used and practical methods for inductive inference over supervised data. A decision tree represents a procedure for classifying categorical data based on their attributes. It is also efficient for processing large amount of data, so is it is used in service mining. The construction of decision tree does not require any domain knowledge or parameter setting, and therefore appropriate for exploratory knowledge discovery. Their representation of acquired knowledge in tree form is intuitive and easy to assimilate by humans. The Figure-1 depicts the frame work of classification.

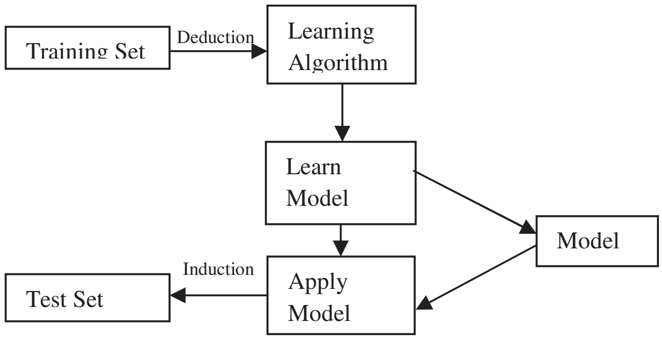

Figure-1

There are some flaws in Decision trees such as irrelevant attributes, decision making boundaries, replication of sub-trees, continuous class attribute, focusing on relevant attributes, missing values of attributes. The proposed algorithm addresses these problems by using feature selection, error pruning, cross validation and model checking. It also determines the depth of the decision tree.

3. Decision Tree Construction

Classification trees are decision trees derived using recursive partitioning data algorithms that classify each incoming data into one of the class labels for the outcome. A classification tree consists of root node, split node and terminal node. The root node is the top node of the tree that consists of all the data. A node that assigns data to a subgroup is splitting node. The node which is having the final decision is known as terminal node.

Among various algorithms used in the classification C5.0 algorithm is an extension of C4.5 algorithm. C5.0 is the classification algorithm which applies in the big data set. [4]. C5.0 is better than C4.5 on the efficiency and the memory. C5.0 model works by splitting the sample based on the field that provides the maximum information gain. The C5.0 model can split samples on the basis of the biggest information gain field. But in this there are some difficulties in learning decision trees. It is difficult to take a decision that how deeply to grow the decision tree. It is also difficult to choose an appropriate attribute selection measure and manage training data with missing attribute values. The modeling complexity also analysed in the proposed algorithm.

Information Gain:

Gain [8] is computed to estimate the gain produced by a split over an attribute. Let S be the sample: Ci is Class I; i = 1,2,…,m

$I(s1, s2, \ldots, sm) = -\Sigma\, pi \log2 (pi)$

si is the number of samples in class I

pi = si /S, log2 is the binary logarithm.

Entropy provides an information-theoretic approach to measure the goodness of a split. It measures the amount of information in an attribute. Let Attribute A has v distinct values. Entropy = E(A) is

$\Sigma\{(S1j+S2j+..+Smj)/S\}*I(s1j,..smj)$

$I(S1j,S2j,..Smj) = -\Sigma\, pij \log2 (pij)$

Gain (A) =I (s1,s2,..,sm) - E(A)

Gain ratio then chooses, from among the tests with at least average gain, If the Gain Ratio= P (A) then

Gain Ratio (A) = Gain (A)/P (A)

The service classification characterizes different levels of service contributing qualities. There are four service classifications: 1.Excellent 2.Good 3.Average and 4. Poor. The classification is differentiated on the on the whole quality evaluation of the selected parameters of QoS and the normalized values of the parameters [1].

The proposed algorithm classifies and selects the most relevant services in the tree arrangement. The classifier is trained and tested first. Then the resulting decision tree is used to classify unseen data. It is having only focus with the relevant attributes through Feature selection.

The following steps and Algorithm-1 are carrying out to classify the decision tree methods.

Input Parameter: WSQA (Web Service Quality Attributes) that are input data to be classified.

Attributes (A): Input to algorithm consists of a collection of training cases, each having a tuple of values for a fixed set of attributes or independent variables A = {A1, A2,....,Ak} and a class attribute(or dependent variable.

Target attributes (TA): The class attribute C is discrete and has values C1, C2,...,Cx.

Algorithm-1
Input: WSQA, A, TA
Output: classified decision tree CDT
Generate_Tree (WSQA, A, TA, CDT)

Step 1: Create a root node R for the tree
Step 2: If all cases of WSQA belong to same class then returns leaf node with label C_j ; Exit.
Step 3: If A ={ } then returns leaf node as failure; Exit..
Step 4: If TA = { } then returns leaf node with label of majority class in WSQA; Exit.
Step 5: Select quality attributes using Genetic Search modified Wrapper Method (Algorithm-2)
Step 6: BestTree = Construct a DT using training data
Step 7: Perform Cross validation
a. Divide all WSQA into N disjoint subsets,
$WSQA= WSQA_1, WSQA_2, \; , WSQA_N$
b. For each i = 1, ..., N do
i. Test set = $WSQA_i$
ii. Training set = WSQA - $WSQA_i$
iii. Compute decision tree using Training set
iv. Determine performance accuracy Pi using Test set
c. Compute N-fold cross-validation estimate of performance = $(P_1 + P_2 + ... + P_N)/N$
Step 8: Perform Reduced Error Pruning technique
Step 9: Perform Model complexity
Step 10: Find the attribute with the highest info gain (A_Best)
Step 11: Partition S (Service) into S_1,S_2,S_3... according to the value of A_Best.
Step 12: Repeat the steps for S_1, S_2, S_3…
Step 13: Classification: For each outcome ϵ WSQA, apply the CDT to determine its class; if all are same class then return as leaf node else go to Step 4.
Step 14: Return the decision tree.

Feature selection selects a subset of features from the original feature set without any transformation, and maintains the physical meanings of the original features. Feature Selection used dimensionality reduction technique in machine learning and data mining. Feature Selection builds the faster model by reducing the number of features, and also helps remove irrelevant, redundant and noisy features.

Algorithm-2
Genetic Algorithm with random Search

Step 1: Consider the original feature set.
Step 2: Generate initial population (t).
Step 3: Repeat Step 4 to Step 7 Until generation count reached.
Step 4: Perform crossover on parents creating population (t+1).
Step 5: Perform mutation of population (t+1).
Step 6: Determine fitness computation of population using decision tree (t).
Step 7: Select the new population.
Step 8: Best feature is selected and validate using decision tree.

Reduced Error Pruning is a technique in machine learning that reduces the size of decision trees by removing sections of the tree that provide little power to classify instances. The dual goal of pruning is reduced complexity of the final classifier as well as better predictive accuracy by the reduction of over fitting and removal of sections of a classifier that may be based on noisy or erroneous data.

Cross-Validation is the method of evaluating and comparing learning algorithms by dividing data into two segments: one used to learn or train a model and the other used to validate the model.

By increasing the complexity of the model, classification accuracy is increased. Complexity of Model is increased by changing parameters.

4. PERFORMANCE

The proposed algorithm Modified C5 decision tree algorithm is used classification of web services. Classification accuracy is usually calculated by determining the percentage of tuples placed in a correct class. This ignores the fact that there also may be a cost associated with an incorrect assignment to the wrong class.

The accuracy of the solution to a classification problem can be determined using the confusion matrix. The confusion matrix is also called as contingency table. Given n classes a confusion matrix is a m x n matrix, where Ci,j indicates the number of tuples from D that were assign to class Ci,j but where the correct class is Ci. Obviously the best solution will have only zero values outside the diagonal.

A confusion matrix contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using the data in the matrix. The Table-2 shows the confusion matrix for a two class classifier. The entries in the confusion matrix have the following meaning.

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Predicted</td></tr>
<tr><td>Negative</td><td>Positive</td></tr>
<tr><td rowspan="2">Actual examples</td><td>Negative</td><td>a
True Negative (tn)
Correct Rejection</td><td>b
False Positive (fp)
false alarms (Error of First kind – false hit)</td></tr>
<tr><td>Positive</td><td>c
False Negative (fn)
(Error of Second kind – a miss)</td><td>d
True Positive (tp)
Correct Inference</td></tr>
</table>

Table-2

When the prediction p matches with the actual value then it is called a true positive (tp) and if it does not match then it is said to be a false positive (fp). The precision and recall gives the measure of relevance. The fraction of retrieved instances that are relevant are called precision where as the fraction of relevant instances that are retrieved is known as recall. Precision can be seen as a measure of exactness or quality, whereas recall is a measure of completeness or quantity. Recall is the true positive rate for the class. [1]

The precision is (tp)/(tp+fp), which is the proportion of positive predictions that are actual positives. The recall or true-positive rate is (tp)/(tp+fn), which is the proportion of actual positives that are predicted to be positive. The false-positive error rate is (fp)/(fp+tn), which is the proportion of actual negatives predicted to be positive

Service	RT	Avail.	Per.	Relia.	Classification
S1	105.00	80	55	62	G
S2	320.50	95	78	60	G
S3	780.81	93	80	88	G
S4	520.11	87	68	75	G
S5	536.50	72	79	66	G
S6	247.00	99	100	72	G
S7	73.00	70	96	82	E
S8	525.12	67	60	78	P
S9	709.40	87	75	73	G
S10	147.44	94	97	60	G

Table-3

This web service relation consists of attributes Response time, Availability, Success ability and Reliability.

5. EXPERIMENTAL RESULTS

The results of reliability of classification are obtained through the Modified C5 Classifier. The attributes have been chosen randomly for given data set. The confusion matrix is used to assess the accuracy of the model being used. At this point the confusion matrix given in Table-4 is generated for class gender having two possible values i.e. YES or NO.

	Predicted Services a (YES)	Predicted Services b (NO)	
Actual YES	54	2	56
Actual NO	1	52	53
	55	54	

Table-4

For above confusion matrix, true positives for class a='YES' is 54 while false positives is 2 whereas, for class b='NO', true positives is 52 and false positives is 1 i.e. diagonal elements of matrix 54+52=106 represents the correct instances classified and other elements2+1=3 represents the incorrect instances.

	Class a	Class b
TP rate	0.964286	0.981132
FP Rate	0.018868	0.035714
Precision	0.981818	0.962963
F-Measure	0.972973	0.971963
Accuracy	0.972477	

Table-5

6. CONCLUSION

In selecting a pertinent web service for use, to satisfy the constraints or requirements of users is necessary to use non functional parameters of the corresponding web services. This work presents web services quality prediction model, which takes non-functional qualities in account. The classification accuracy based on the proposed algorithm is 97%. This improves the selection of web services more efficient one to make composite web services for business processes which requires more than one services to be combined to complete a customer request.

7. REFERENCES

[1] J. Ghayathri and Dr.S. Pannir Selvam, "Selection of paramount web service using sentient QoS parameters", International Journal of Computer Applications, 2015.

[2] Susila, S. Vadivel , "QoS Measurement Tool for Web Service Selection", International Journal of Web Engineering, 2014.

[3] Venkataiah Vaadaala Hyderabad, India, Dr. R. Rajeswara Rao, Venkateswara Rao, "Classification of Web Services Using JForty Eight", International Journal of Electronics communication and Computer Engineering, 2013.

[4] A. S. Galathiya, A. P. Ganatra Charotar, "Classification with an improved Decision Tree Algorithm", International Journal of Computer Applications (0975 – 8887), Volume 46– No.23, May 2012.

[5] A. S. Galathiya, A. P. Ganatra and C. K. Bhensdadia, "Improved Decision Tree Induction Algorithm with Feature Selection, Cross Validation, Model Complexity and Reduced Error Pruning ", International Journal of Computer Science and Information Technologies, 2012.

[6] Bruno,M, Benevento, Di Penta, M. ; Scognamiglio, R." An Approach to support Web Service Classification and Annotation, e-Technology", e-Commerce and e-Service, 2005. EEE '05. Proceedings.

[7] Shilpa Sonawani and Debajyoti Mukhopadhyay, "A Decision Tree Approach to Classify Web Services using Quality Parameters", ICWA 2013 International Conference.

[8] Guo, L.Y., Chen, H.P., Yang, G., Fei, R.Y: A "QoS evaluation algorithm for ranking based on artificial neural network",. International Conference on Computer Science and Software Engineering, CSSE (2) 381-384(2008)

[9] M.Swami Das, Ramakanta Mohanty, D.Vijayalakshmi: A.Govardhan, "Application of data mining using Bayesian Belief Network To classify quality of web services", International Journal of Computer Science and Informatics (IJCSI), ISSN (*PRINT*): 2231 –5292 Volume. 2.

[10] T.Miranda Lakshmi , A.Martin ,R.Mumtaj Begum , V.Prasanna Venkatesan , " An Analysis on Performance of Decision Tree Algorithms using Student's Qualitative Data " MECS I.J. Modern Education and Computer Science, 2013, 5, 18-27.

[11] Serge Haddad, Lynda Mokdad, Samir Youcef, " Selection of the best composite Web service based on quality of service',International Conference on Business Process and Services Computing (BPSC'10), Lecture Notes in Informatics 177, Leipzig Germany 2010.

[12] V. Estruch C. Ferri J. Hern´andez-Orallo, M.J. Ram´ırez-Quintana, "Web Categorisation Using Distance-Based Decision Trees", Elsevier, Electronic Notes in Theoretical Computer Science 157 (2006) 35–40, 2006.

[13] Mamoun Mohamad Jamous, Safaai Bin Deris, "Web Services Non-Functional Classification to Enhance Discovery Speed", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 4, July 2011

[14] Zeina Azmeh, Marianne Huchard, Chouki Tibermacine, " WSPAB: A Tool for Automatic Classification & Selection of Web Services Using Formal Concept Analysis "

[15] Rama Kanta Mohanty, "Classification of Web Services Using Bayesian Network", Journal Of Software Engineering And Applications, January 2012

[16] Yongheng Zhao, Yanxia Zhang, " Comparison of decision tree methods for finding active objects ", Advances in Space Research, 41: 1955–1959, 2008.