Performance Analysis of Different Classification Algorithms in Information Retrieval through Web Services

Amandeep Kaur¹ and Rachna Soni²

¹Faculty of Computer Applications, Chitkara University, Rajpura - 140401, Punjab, India; amandeep.bhullar@chitkara.edu.in
²Faculty of Computer Applications, DAV College, Yamuna Nagar - 135001, Haryana, India; sonirachna67@gmail.com

Abstract

Background/Objectives: The web client gets easily lost in the web's rich hyper structure as the utilization of web is expanding more step by step. The primary point of proprietor of the website is to give the important data in terms of satisfactory QoS (Quality of Service) factors like throughput, response time, accuracy and content availability. From the client point of view, Web Service based QoS Discovery is a multi-criteria decision mechanism that requires knowledge about the service and its QoS description. These clients are not experienced enough to acquire the best selection of web service and trust the QoS information published by the provider. Methods/Analysis: The existing t Model was used with XML based SOAP protocol in order to solve the problem of UDDI registry which holds QoS description. Findings: The new discovery approach is expected to be the solution for contemporary web service discovery problems. A comparative performance analysis of prominent page rank algorithms was made on the basis of metrics like throughput, response time, recall rate and precision rate etc. Simulation Interface has been designed for classification algorithms. The program is developed for the Fuzzy Logic, Naïve Bayes, Neural Network, Linear Discriminant Analysis and Support Vector Machine using MATLAB application. Improvements: The experiment revealed the fact that recall and precision rate are the best to predict the Quality of Service (QoS) supported by various E-Commerce web sites like Amazon, Jabong and Shop Clues etc. Detailed performance analysis further concluded that Neural Network could be the best algorithm to rate the service quality of E-Commerce websites.

Keywords: Fuzzy Classification, Linear Discriminant Analysis, Neural Network, Support Vector, Universal Discovery Description and Integration (UDDI)

1. Introduction

WWW gives Information System to achieve data representation, data recovery, assignments and information administration. With the end goal of data recovery, WWW gives Web Services and Web applications. Web applications incorporate Internet specialized tools, informal organization and E-Commerce. Web Services assumed to be a critical part in the E-Business, Computer Software and Communication Industries. Semantic Web Services (SWS) as coordinated answer for understanding the vision of the up and coming era of the web. These administrations are self-depicted and give important and interpretable data of every page content in order to work in participation by individuals and PCs. Agent is one who talks on the behalf of other entities too. Agent Systems are independent programming programs handling space information and a capacity to act with some level of freedom to complete activities to accomplish determined objectives. Web Services are application that impart utilizing open conventions, For example, Hypertext Transfer Protocol (HTTP), Extensible Mark-up Language (XML)
and Simple Object Access Protocol (SOAP)\(^4\) which are intended to bolster computer-to-computer communication over a system. Numerous companies give similar Web Services to clients. For instance, Google permits programming designers to question billions of site pages straightforwardly from their own PC programs. Ranking of Web Services requires when problem for layman in selection of similar Web Services arises. An engineer can utilize his or her most loved programming language, for example, Java or Dot Net to create any functionality that get to the Google Web Services. Functionalities created under any framework can be changed over into web applications through Web Services which can be found, distributed and utilized web services. Web Services normally comprise of XML, SOAP\(^5\) and UDDI\(^6\). XML is used for transferring the data, UDDI list down the Web Services which are available and SOAP used for shifting the data. By the extensive propagation of the Web Services, QoS (Quality of Service)\(^8\) has become an important factor in comparing the success of service providers. QoS has now a challenging task due to the dynamic and unpredictable character of the web\(^11,12\). Mostly, the issues that are not solved can create critical transactional applications to endure by the improper levels of performance deprivation. It has been shown that the QoS-aware programs can have a higher business value\(^12\).

2. Literature Review

This paper is addressing the role of web service in the E-commerce. This section explains the background research for this study.

In research paper\(^14\), the author presented the work in which collaboration of SWS (Semantic Web Services) will be done with WSMO (Web Service Modeling Ontology). In proposed work the methodology is divided into three tasks to enhance the user interaction. The main disadvantage of this work is that no explicit information is available in prior. The authors\(^15\) proposed the replication method based on UDDI replications. In this work, basically client gets connected to the UDDI based on three features from unique table. This paper is just an invention of new method irrespective of any comparison. In\(^16\) the researchers presented the analysis of Web Services based on QoS systems. The main motive of this web service in proposed model is to optimize the web services. QoS plays an important role in providing good QoS rate over time.

In\(^12\) the authors solved the problem of ranking using particle swarm optimization algorithm in proposed work. In addition to this QoS services gets also enhanced in proposed work. Also the ranking operation is being enhanced by using k-neighbor function. In\(^17\) the authors have focused on the REST (RESTful) and SOAP (Simple Object Access Protocol) web based services. The SOAP architecture is implemented in the XML encoded message which is transmitted on the HTTP (Hypertext Transfer Protocol). The WSDL file\(^18\) is written first on the server side with the XML schema types and later the WSDL (Web Services Description Language) is made public for reaching the particular public server by using SOAP enabled web based server. While in the case of Representational State Transfer Protocol\(^19\), the HTTP client library is used for interacting with the REST server instead of the SOAP client. As SOAP and REST have their own positive characteristics, therefore, they have been compared.

In\(^20\) it has shown a QoS bootstrapping solution for the Web Services and a QoS bootstrapping framework is built. SOA (Service Oriented Approach) work is being extended. A prototype is implemented for supporting QoS bootstrapping. QoS is used as a factor among the services and shows a better performance in service delivery, selection and composition. There is a need to extend the work to bootstrap other QoS in the QoS model. In\(^21\) the authors have proposed the method that uses the information extraction techniques and some transformations are being made in automating the larger amount for extraction of resource process. The information retrieval metrics are used for evaluating the extraction. This research provides a traceable process for extracting hierarchical model for generating a RESTer API for large real life applications. In this work, REST specification addresses about REST API are still missing.

In\(^22\) researchers presented a service oriented multimedia componentization model for supporting QoS (Quality of Service). The enhancement is projected to make possible Simple Object Access Protocol (SOAP) and Composite Capability/Preference Profiles (CC/PP) standards to recover their flexibility to implement multimedia web services. The application which is being selected for implementation is a distance-learning environment. The designing is such that every client can develop the threads and then can communicate with the server by requesting the multimedia services. The tests are being performed on the server machine by Multimedia Web Service Server.
Agent and without MWSS agent. The problem that arises in proposed work is that it takes longer because of the extra operation time spent on the MWSS Agent. In\textsuperscript{24} web splitter proposed the work in which web splitter has to be implemented as XML files in the network. The proposed work is dependent on the CC/PP protocol for exchanging of the files in the network. In\textsuperscript{25} author proposed 3 open tools and then compared them on the basis of basic features like usability, performance and s/w requirements. Comparison of these three is difficult as the testing criteria are not the same for three. In\textsuperscript{26} authors presented or reviewed the related work of various quality attributes like usability, performance and s/w requirements based on web services. In\textsuperscript{27} authors proposed the QoS estimation based on two methods i.e. LFM model and Latent Neighbor model.

Although the above mentioned methods have shown the improvement but they may be less practical when some additional information or the personalized knowledge made unavailable.

### 3. Web Service Model

QoS selection has become essential for real time applications. Number of web applications requires guaranteed services processing. There are various types of standards that support the Web Services that are SOAP (Simple Object Access Protocol), WSDL (Web Services Description Language), UDDI (Universal Description Discovery Integration) and XML (Extended Markup Language). The Web Services that are available these days do not support Quality of Services (QoS) or some of the non-functional aspects\textsuperscript{24}. Quality of Service with the service consumer is necessary for adopting the best quality web services. The existing approaches that have shown dissimilarity in the results of the same Web Services are in number.

Above Figure 1 shows the base model of Web Services that has the service provider that selects service by using the service provider. Service provider publishes the description of the service functions. Service consumer faced most problems in an open environment when the same function is offered by the cluster of services. XML (Extended Markup Language) and SOAP (Simple Object Access Protocol) protocols are used for the communication between Service Provider and Consumer\textsuperscript{28}. It is necessary that a service consumer should have knowledge about the working of the web services. Only syntactic information is provided by the UDDI (Universal Description Discovery Integration). But Semantic information is mandatory for the consumers that contains the response time, availability, content, throughput and many more and if the solution of this problem is provided, then one more problem can occur. The need of updating and saving information can occur in UDDI after publishing of QoS.

For storing the web services, number of systems and architecture used UDDI registry. T-model (Technical Model)\textsuperscript{21} was represented for solving the problem of UDDI that would store QoS values for every web service. But there are various problems that need to resolve yet. The UDDI was not proposed for the web service discovery in the search engine. This can make passive easily because of the voluntary registration. It doesn't have any guarantee for the information it contains. Between UDDI and Current Web, a disconnection exists and UDDI has not the capability to provide quality of service for the Web Services that are registered.

The model shown in Figure 2 is basically helpful for those customers who evaluate and compare the quality factors of a particular thing for choosing the best one depends on QoS requirements from different shopping sites. Different websites produce similar results will differ according to their QoS factors. For example: A customer wants to buy a mobile cover and requests through the search engine. The broker receives the request and interacts with the service provider and checks whether the response exists or not in UDDI registry. If the customer makes number of requests, then with the number of responses it will return to the client. The objective of the research is to find out the best response, according to the QoS factor. By adding the functionalities to search agent, the performance of the broker has been enhanced in this model. From database to search a particular query requested by the customer is the task of the customer. To
check the existence of results in the registry, the search agent interacts with the service provider. To search a particular request, fetch and display of multiple results in front of the user is the limited role of the search agent. The efficiency of the search will increase by adding other functionalities. Through suggest an action procedure; users will get some tips for requests.

When something is entered in the search box by the user then based on the particular keyword some suggestions will come. HTML parsing is done by an HTML DOM parser method in order to retrieve meaningful information or plain text. The conversion results in RDF (Resource Description Framework) format that provides semantic interpretability of data. To scrape the data of shopping sites HTML (Hypertext Markup Language) DOM (Document Object Model) PARSER is used which manipulates HTML elements. All the web pages return similar results linked to a website are fetched by the web crawler under Search Engine Optimization. Then indexing of all the fetched web pages is done and then placed in a large database and later from there it can be retrieved. To identify the words and expressions that describe and assigns the page to a particular keyword the process of indexing is used. In service broker, the SEO (Search Engine Optimization) plays a very important role. All the web pages are being forwarded by the service provider equivalent to the user demand to quality manager for finding the best quality factors. Later the results which are calculated by the priority means through the service provider are being forwarded to the broker with quality manager and to the same client at the end.

![Figure 2. Proposed web service model.](image)

4. Fuzzy Model

Fuzzy Model is the generalized model of previous classic models. As the output is not limited to only 0 and 1, so the theory of Fuzzy Logic is introduced. It is also known as diffuse logic. Difference between Fuzzy Logic and Classical Model is introduced using membership functions. Consider a finite set,

\[ C = \{c_1, c_2, c_3, \ldots, c_n\} \]

It is the universal set. Now according to graphical representation, suppose fuzzy sets have only two elements c1 and c2. So the degree of fuzzification can be called as entropy. Therefore entropy can be shown as:

\[ E = \frac{f_1}{f_2^2} \]

Where f1 and f2 are the distances.

- Procedures on Fuzzy Set
  - The intersection of D and M is defined as:
    \[ (D \cap M)(u) = \min\{D(u), M(u)\} = D(u) \land M(u) \]
    For all \( u \in V \).
  - The union of D and M is defined as \( (D \cup M)(u) = \max\{D(u), M(u)\} = D(u) \lor M(u) \) for all \( u \in V \).
  - The complement of Fuzzy set is fuzzy set S is defined as \( (\neg D)(u) = 1 - D(u) \).
  - Law of excluded middle can be shown as:
    Let \( D(u) = 1/2, \forall u \in Y \), then it is easy to see that
    \[ (\neg D \lor D)(u) = \max \{\neg D(u), D(u)\} \]
    Law of Contradiction can be shown as: Let \( D(u) = 1/2, \forall u \in Y \), then it is easy to see that
    \[ (\neg D \land D)(u) = \min \{\neg D(u), D(u)\} \]
    Performance of Fuzzy Logic using MATLAB is shown in Figure 3.

![Figure 3. Fuzzy logic in MATLAB.](image)
5. Support Vector Machine (SVM)

Support Vector Machine (SVMs) is a binary classification algorithm developed by Vapnik. The main features of SVM are shown below, due to which its applications are quite important:

- Robust to large number of variables.
- It can learn complex and simple learning models.
- It avoid overfitting.

Support Vector Machines (SVMs) have the Hyperplane that classifies the various variables as shown below:

Equation of hyperplane can be written as below:

\[ w \cdot x + b = 0 \]  
(3)

(Recall that \( w \) is in fact the vector orthogonal to the hyperplane.) Given such a hyperplane \((e, n)\) that separates the data, this gives the function \( f(c) = \text{sign}(e \cdot c + n) \).

So, we can define the separation using hyperplane as below having distance gap of 1.

\[ c_i \cdot e + n \geq \pm 1 \text{ when } u_i = \pm 1 \]  
(4)

\[ c_i \cdot e + n \leq \mp 1 \text{ when } u_i = \mp 1 \]  
(5)

or more compactly:

\[ u_i (c_i \cdot e + n) \geq 1 \forall i \]  
(6)

5.1 Basic Operations Utilize the SVM

5.1.1 Multiplication by Scalar

Consider a vector \( a = \{ a_1, a_2, \ldots, a_n \} \) and a scalar \( \alpha \), then multiplication can be shown as:

\[ \{ \alpha a_1, \alpha a_2, \ldots, \alpha a_n \} \quad (\text{7}) \]

5.1.2 Addition of Vectors

Consider a vector \( a = \{ a_1, a_2, \ldots, a_n \} \) and \( b = \{ b_1, b_2, \ldots, b_n \} \), then addition can be shown as:

\[ \{ a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n \} \quad (\text{8}) \]

5.1.3 Subtraction of Vectors

Consider a vector \( a = \{ a_1, a_2, \ldots, a_n \} \) and \( b = \{ b_1, b_2, \ldots, b_n \} \), then subtraction can be shown as:

\[ \{ a_1 - b_1, a_2 - b_2, \ldots, a_n - b_n \} \quad (\text{9}) \]

5.1.4 Euclidean Distance

Consider a vector \( a = \{ a_1, a_2, \ldots, a_n \} \)

Then it can be writing as:

\[ ||a|| = \sqrt{a_1^2 + a_2^2 + \ldots + a_n^2} \quad (\text{10}) \]

5.1.5 Dot Product

Consider a vector \( a = \{ a_1, a_2, \ldots, a_n \} \) and \( b = \{ b_1, b_2, \ldots, b_n \} \), then dot product can be shown as:

\[ \{ a_1 b_1, a_2 b_2, \ldots, a_n b_n \} \quad (\text{11}) \]

6. Neural Network

Research in the field of neural networks has been growing from past very years. Neural Network structure is similar to human brain consists of small neurons. The Neural Networks are developed to work like human brain to reduce the computational tasks. Basic elements of Neural Network in MATLAB are epochs, gradient, validation step as shown below in Figure 5 and Figure 6.

There are three types of layers i.e. input layer, hidden layer and output layer as shown below in mathematical model. The structure of neuron can be shown as below:

The net input can be shown as:

\[ n_1^{k_1+1}(i) = \sum_{j=1}^{L_1} w_1^{k_1+1}(i,j) a_1^{k_1}(j) + b_1^{k_1+1}(i) \]  
(12)

The unit is given by:

\[ a_1^{k_1+1}(i) = f_1^{k_1+1}(n_1^{k_1+1}(i)) \]  
(13)

This recurrence relation is executed at the final layer:

\[ -F_1^{N_1}(n_1^{N_1})(a_1^{N_1} - a_1^{q_1}) \]  
(14)
7. Results and Discussions

The Table 1 shows the recall rate and precision rate value based on SVM classifier and the obtained values are Recall rate = 91.271, 92.089, 92.521, 91.898, 91.232, 91.037, 91.2241, 92.0054, 91.561 and 91.890 and Precision rate = 91.003, 91.24, 91.22, 91.34, 91.56, 91.46, 91.25, 91.29, 91.31 and 91.51 respectively.

Table 1. Recall and precision rate values w.r.t. SVM

<table>
<thead>
<tr>
<th>Sr.No</th>
<th>Recall Rate (%)</th>
<th>Precision Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>91.271</td>
<td>91.003</td>
</tr>
<tr>
<td>2</td>
<td>92.089</td>
<td>91.24</td>
</tr>
<tr>
<td>3</td>
<td>92.521</td>
<td>91.22</td>
</tr>
<tr>
<td>4</td>
<td>91.898</td>
<td>91.34</td>
</tr>
<tr>
<td>5</td>
<td>91.232</td>
<td>91.56</td>
</tr>
<tr>
<td>6</td>
<td>91.037</td>
<td>91.46</td>
</tr>
<tr>
<td>7</td>
<td>91.2241</td>
<td>91.25</td>
</tr>
<tr>
<td>8</td>
<td>92.0054</td>
<td>91.29</td>
</tr>
<tr>
<td>9</td>
<td>91.561</td>
<td>91.31</td>
</tr>
<tr>
<td>10</td>
<td>91.890</td>
<td>91.51</td>
</tr>
</tbody>
</table>

Figure 7. Recall and precision rate graph.

The above Figure 7 shows the comparable values of the graph using SVM on the basis of Recall rate and Precision rate. It has been seen that Recall rate value is higher than the Precision rate value in terms of percentage and different values is around .876%.

Table 2. Recall and Precision rate values w.r.t. NN

<table>
<thead>
<tr>
<th>Sr.No</th>
<th>Recall rate (%)</th>
<th>Precision Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>98.271</td>
<td>99.003</td>
</tr>
<tr>
<td>2</td>
<td>98.089</td>
<td>98.24</td>
</tr>
<tr>
<td>3</td>
<td>99.521</td>
<td>98.22</td>
</tr>
<tr>
<td>4</td>
<td>98.898</td>
<td>98.34</td>
</tr>
<tr>
<td>5</td>
<td>98.232</td>
<td>98.56</td>
</tr>
<tr>
<td>6</td>
<td>98.037</td>
<td>98.46</td>
</tr>
</tbody>
</table>
From the result evaluation as shown in Table 2, it has been seen that the Recall rate and Precision rate value based on NN classifier came out to be Recall rate = 98.271, 98.089, 99.521, 98.898, 98.232, 98.037, 98.2241, 98.0054, 99.561 and 98.890. Precision rate = 99.003, 98.24, 98.22, 98.34, 98.56, 98.46, 98.25, 98.29, 98.31 and 98.51 respectively.

Table 3. Recall and Precision rate values w.r.t. Fuzzy Logic

<table>
<thead>
<tr>
<th>Sr.No</th>
<th>Recall rate (%)</th>
<th>Precision Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>94.371</td>
<td>94.003</td>
</tr>
<tr>
<td>2</td>
<td>95.589</td>
<td>94.124</td>
</tr>
<tr>
<td>3</td>
<td>94.421</td>
<td>94.122</td>
</tr>
<tr>
<td>4</td>
<td>94.898</td>
<td>94.134</td>
</tr>
<tr>
<td>5</td>
<td>94.632</td>
<td>95.156</td>
</tr>
<tr>
<td>6</td>
<td>95.937</td>
<td>94.146</td>
</tr>
<tr>
<td>7</td>
<td>94.2241</td>
<td>94.125</td>
</tr>
<tr>
<td>8</td>
<td>94.3054</td>
<td>94.129</td>
</tr>
<tr>
<td>9</td>
<td>94.661</td>
<td>94.131</td>
</tr>
</tbody>
</table>

Figure 8. Recall and Precision rate graph using NN.

Above graph in Figure 8 describes the comparison values graph using NN on the basis of Recall rate and Precision rate. It has been noticed that Neural Network Method has highest recall rate and the precision rate in the proposed work. As Neural Network has a good learning rate so it has a better recall rate and precision rate having values 99 and 98 respectively.

Table 3. Recall and Precision rate values w.r.t. Fuzzy Logic

Recall rate and Precision rate values are shown in the above Table 3 and the obtained values of Recall rate = 94.371, 95.589, 94.421, 94.898, 94.632, 95.937, 94.2241, 94.3054 94.661 and 94.790. And of Precision rate = 94.003, 94.124, 94.122, 94.134, 95.156, 94.146, 94.125, 94.129, 94.131 and 94.151.

Figure 9. Recall and Precision rate graph using Fuzzy Logic.

In the above graph in Figure 9, Fuzzy Logic based comparison has been made using two parameters, i.e. Recall rate and Precision rate. It has been noticed that Fuzzy Logic Method has intermediate recall rate and the precision rate in the proposed work w.r.t SVM and neural network classifier.

Table 4. Comparison between traditional methods

<table>
<thead>
<tr>
<th>Technique</th>
<th>Precision rate</th>
<th>Recall rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Existing Work(^{22})</td>
<td>78%</td>
<td>79%</td>
</tr>
<tr>
<td>Proposed Work with NN</td>
<td>99%</td>
<td>98%</td>
</tr>
</tbody>
</table>

From the Table 4, it has been noted that the Precision rate and Recall rate of the proposed algorithm is good than that of traditional method and we can conclude that the proposed method is able to generate an acceptable result.

8. Conclusion

Number of web applications requires guaranteed services processing. There are a number of standards that support the Web Services but there is need of technique that can guarantee the results. So, in research paper three tech-
Techniques have been compared on the basis of Recall rate and Precision rate for E-commerce usage. From the result analysis it has been seen that Recall rate and Precision rate has best value using Neural Network and has worst values using SVM method.

9. Reference