WEB SERVICE CLASSIFICATION WITH MULTILAYER PERCEPTRON

A. Syed Mustafa*
Research Scholar, Dept. Of Computer Science and Engineering,
Sathyabama University, Chennai, India and Professor, HKBK College of Engineering, Bengaluru.
Email: Syedmustafa_blr@yahoo.com

Dr. Y.S. Kumaraswamy
Dean R&D , Dept. Of Computer Science and Engineering,
Nagarjuna College of Engineering & Technology, Bengaluru, India
Email: yskldswamy2@yahoo.com

Abstract - Web services are made of software components for expressing the application information, communicating messages and for interacting with open XML and Internet technologies. Web services are software applications accessible on the web, used for machine to machine interaction using Uniform Resource Identifier (URI) on distributed internet environment.With the explosion of Web Services accessible on the web, automatic categorization of the services to organize the data becomes essential. In this paper we use 3 different classifiers for web service classification. Multilayer perceptron is trained with back propagation and genetic algorithm.

Keywords: Web Services, Naïve Bayes, MultiLayer Perceptron, Back propagation and Genetic Algorithm (GA), QWS dataset.

1. INTRODUCTION

Web services are software applications that are available on the web and used for machine to machine interaction by usingUniform Resource Identifier (URI) on the distributedenvironment of internet. Simple Object AccessProtocol (SOAP) messages are used for communication mechanism by using Hyper Text Transfer Protocol (HTTP) protocol. Each web service has an Application Program Interface (API) thatcan be accessed over the network and executes the web serviceat host location [1]. Every service provides a role, such asservice provider, a requester or a broker. In other words, webservices make possible the effective processing of machinereadable information.

Web services are a novelset of Web applications. They are self-contained, self-describing, modular applications that can be published, located, and invoked across the Web. Web services perform functions that can be anything from simple requests for information to creating and executing complicated business processes [2]. Once a Web service is deployed, it can be discovered and invoked by other applications (or other Web services).

The key advantage of using Web services is the ability to create applications on the fly through the use of loosely coupled, reusable software components. This has fundamental implications in both technologies and business applications. Web Services can be classified as follows [3]:

User-centric Web Services: User-centric Web Services are used to provide user personalization, interface customization, andsupport forvarious languages that helps in the enhancement of user experience. Logical separation of layouts (presentation) in a particular formatlike HyperText Markup Language (HTML) andactual datapresent in Extensible Markup Language (XML) exists.

Application-centric Web Services**:** These are utilized for integration of enterprise as well as business-to-businessapplications. Application-centric Web Services enable companies to integrate applications and business processes withoutthe constraints of a proprietary infrastructure, platforms and operating systems.

Both user- and application-centric Web Services, make full use of open standards, including HTTP, Extensible Markup Language (XML), SOAP, Web Services Description Language(WSDL), and Universal Discovery, Description, and Integration (UDDI).

Operations in a Web Service Architecture [4]:

Publish - To be accessible, a service description needs to be published so that the service requestor can find it.

Find - The service requestor retrieves a service description directly or queries the service registry for the type of service required

Bind - The servicerequestor invokes or initiates an interaction with the service at runtime using the bindingdetails in the service description to locate, contact and invoke the service.

Web Service has been used more and more widely in recentyears, and with the rapid growth of web services in differentfields, it is obviously unrealistic to organize, classify andmanage web services manually. Therefore, how to makemachines recognize, manage and use web servicesautomatically has become a focus point research. And thefirst step to achieve the automated management of webservices is to classify them accurately and efficiently. SinceWSDL is an XML format language to describe the way theweb service works and communicates, it can be regarded asa basis of classification [5].

The automatic assignment of classes to Web Services is known as serviceclassification. This problem is vital in SOC, because of the increasing number ofWeb Services. Therefore, it has been investigated by several researchers in thecommunity [6]. Generally, machine learning methods are utilized to perform automatic service classification, where different approaches are based on argumentdefinitions matching, document classification techniques, or semantic annotation matching.

During service classification, classifiers need tobe trained previously. When registering a new Web service, the functional descriptions, suchas the input and output information, will be extracted from theservice document to map to a feature vector. Then the featurevector is input into the classifier to determine which categorythe service belongs to. In order to decrease the complexity of theproblem, web services are categorized top to bottom. When category is obtained, users will confirm if theservice is classified into the proper category, if not, users canclassify the service manually [7]. At last, the service is added tothe service list of the category.

The classification of web service composition is donebased on technology such as workflow based,mathematical based, model driven and AI planning and on the basis ofQoSfeatures [8].Web services comprise several atomic web services, while choosing, services are to address several service provides. For providing seamless services from encapsulated services, effective selection method is needed. Optimization protocols are capable of reaching optimal solutions, optimizing real-life issues is a difficult task due to the vast domain.

Evolutionary algorithms are stochastic search mechanisms which imitate natural as well as social behavior of various species of animals. The benefits over evolutionary algorithms are that they are robust as well as easy to utilize as global optimization technique. They also work on sets of solutions at a time rather than single solutions and thereby facilitating them to search entire problem space [9]. The main difference between evolutionary algorithms and other optimization algorithms are EA’s work at every step with a group of solutions called population. This population yields a set of solutions known as offspring by performing an evolution process known as crossover and mutation..

In this paper, we used 3 classifiers are Naïve Bayes, MLP BPP and MLP GA. Remaining sections are as follows: Section 2 reviews related work. Section 3 explains methodology. Section 4 discusses the experiment results and Section 5 concludes the proposed work.

2. RELATED WORK

A new method was proposed by Yuan-jie& Jian [5] which applies automatic web service semantic annotation and used three classification method: Naïve Bayes, SVM and REP Tree, furthermore ensemble learning was employed. As per the experiment carried out on 951 WSDL files and19 categories, accuracy was 87.39%.

Multi-Layer Perceptron optimized with Tabu search (MLP-TS) for learning was proposed by Syed Mustafa&Swamy [10]. Experimental results demonstrate that the proposed MLP-TS outperforms Multi-Layer Perceptron-Levenberg-Marquardt (MLP-LM) and Multi-Layer Perceptron Back Propagation (MLP-BPP) for web service classification.

Qi et al., [11] summarized existing web service composition approaches and then presented a service classification management mechanism to organize web services more efficiently and accurately. Finally, an automated web services composition system was designed, which consists of two main parts: service management subsystem and service provision subsystem. The service management part is based on the service classification management mechanism, and service provision part is to meet the need of users' request by AI Planning.

Sowmya Kamath et al., [12] proposed an approach for web service classification based on conversion of services into a class dependent vector by applying the concept of semantic relatedness and to generate classes of services ranked by their semantic relatedness to a given query. The OWLS-tc service dataset was used for evaluating our approach and the experimental results were presented in the proposed work.

Song,& Tang [13] presented an approach based on semantic reasoning and ontology techniques in order to organize web services automatically and accurately. And the proposed approach verified by comparing with METEOR-S's classification results.

Raj et al., [14] proposed a method for the effective web service selection based on the QOS parameters using the K-Nearest Neighbor algorithm. Implementation of the classification algorithm over the large dataset has some performance limitations. Addition of new parallel classification model will improve the performance. The evaluation reports showed that the effectiveness of the proposed method for service classification and selection.

Gowri&Lavanya[7] proposed a novel classification matrix for Web service composition that distinguishes between the context and technology dimension. The context dimension is aimed at analyzing the QoS influence on the effort of Web service composition, while the technology dimension focuses on the technique influence on the effort. Finally, a suggestion provided to improve the quality of service selection those participates in the composition process with Cskyline approach using agents.

Li et al., [15] proposed an effort-oriented classification matrix for Web service composition, which distinguishes between the context and technology dimension. The context dimension was focused on analysis of environmental influence on the efforts of web service composition, and the technological dimension focused on the method influence on effort. Subsequently, apart from the conventional classification advantages, matrix may be utilized for building the basis of cost prediction for web service composition in future work.

Shafiqet al., [16] presented a hybrid approach towards enabling dynamic Web service discovery which is based on Bayesian Classification mechanism that classifies different available Web services, representing service providers, based on light-weight semantic descriptions.

Sawant&Ghorpade [17] proposed a framework for automatic service classification and categorization of web service process in digital environment. The proposed framework semantically performed automated service discovery and domain selection using domain-knowledge ontology based classification in a digital environment to improvise the service categorization. It is efficiently able to classify and annotated service information by means of specific service domain knowledge. In order to thoroughly evaluate the performance of the proposed semantic based crawlers for automatic service discovery, the Precision, Mean Average Precision, Recall and F-measure Rates were measured.

Web services are formulated for access by other applications and differ in complexity from simple operations like checking bank account balances online to complicated processes running Customer Relationship Management (CRM) or Enterprise Resource Planning (ERP) systems. Since they are based on open standards such as HTTP and XML-based protocols including SOAP and WSDL, Web services are hardware, programming language, and operating system independent. Mustafa&Kumaraswamy [18] used Naïve Bayes, C4.5 and Random forest methods as classifiers for the efficiency of web services classification.

Li et al., [19] shown that how to classify modulus according to the data characteristics and retrieve data through multi-threading respectively in batches are shown in details. The experimental results showed that the proposed method based on modular classification and multi-threading transmission can respond client requests quickly when large amounts of data need to be transmitted.

3. METHODOLOGY

QWS dataset is used in the proposed work. Naïve Bayes andMultiLayer Perceptron are discussed. MLP is trained with Back propagation algorithm and Genetic algorithm.

3.1 QWS dataset

QWS dataset consists of different web service implementations and their attributes. The classification is measured based on the overall quality rating provided by all the attributes. The functionality of the web services can be helpful to differentiate between various services. The web services in the QWS dataset are classified into four categories, such as: 1) Platinum (high quality); 2) gold; 3) silver and 4) bronze (low quality). The classification is measured based on the overall quality rating provided by WSRF. It is grouped into a particular web service based on classification [20]. The functionality of the web services can be helpful to differentiate between various services.

Updated QWS Dataset Version 2.0 has a set of 2,507 Web services and QWS measurements conducted in March 2008 using a Web Service Broker (WSB) framework. Every row in the dataset represents a Web service and its corresponding nine QWS measurements (separated by commas). The first nine elements were QWS metrics measured with multiple Web service benchmark tools over six-days. QWS values represent measurements averages collected during this period [21]. The last 2 parameters represent service name and reference to WSDL document.

3.2 Naïve Bayes

Naïve Bayes theoremprovides the probabilistic classifier approach capableof producing the results interpreting the user queries.Bayes' theorem deals with the conditional probabilitieswhich influence the event on the probability on another revent. The terminologies prior and posterior probabilitiesare associated with it. It's also provides the featureadding dynamic data to the class. Naïve Bayes classifiesthe data for any case, if the class is given. The classesand attributes are independently processed by the NaïveBayes Classifier and this phenomenon of the class iscalled class independence [22]. This approach has giventhe results which are effective in most of the applications.

NaïveBayes classifieris simply to classify collection based on Bayes'theorem in supervised classification. NaïveBayes assumes that all properties of a category are strongly mutual independent [23]. But this supposition is not existent in a real world. NaïveBayes'principle is that, it will compute the probability of unclassified data which belongs toeachcategory; category of the most probability is the one unclassified data belongs to. NaïveBayes is provenby many experiments that it has a good effect.

Bayes' theorem:

$$P(c_i \mid D) = \frac{P(D \mid c_i)P(c_i)}{P(D)}$$

3.3 Proposed MultiLayer Perceptron

Multilayer Perceptron (MLP) also known as Back Propagation Neural Network, is a feed forward multilayer artificial neural network which is based upon extended gradient-descent based Delta learning rule, commonly known as Back Propagation rule. In this algorithm, error signal between desired output and actual output is being propagated in backward direction from output to hidden layer and then to input layer in order to train the algorithm and to use it as classifier [24].

Feed-forward structured networks do nothave connections between units in the same layer. These networks usually comprisesof input, hidden and output layers, all of which are interconnected with respect to thehidden layer.The training of these networks is accomplished through backpropagation and acomplex nonlinear hidden as well as output weights optimization. At iterations, theerror of the network is assessed by forward propagating the inputs through thenetwork and the derivative of this error is calculated with respect to each weightwithin the network [25].

Neurons for MLP input layer pattern classification isdetermined by features representing relevant feature spacepatterns. Input layer neurons, acting as sensory units, computeidentity function, y = x. A hidden layer neuron and outputlayers compute sigmoidal function of sum of products of inputvalues and weight values of corresponding connections [10].

Back Propagation algorithm

For the MLP neural network trained with Back Propagation algorithm, acomplex network's error surface is full of hills and valleys.Due to gradient descent a network can be trapped in localminimum when a deeper minimum is nearby. Probabilisticmethods help avoid this trap, but they are slow. Anotherpossibility is increasing the hidden units. Though this worksbecause of error space's higher dimensionality and chances ofbeing trapped is smaller, there is an upper limit to hidden unitswhich, when exceeded, result in system being trapped in localminima.

The backpropagation algorithm employs a gradient descentmethod. Once the momentum term is added, the algorithmgives the weight change of a connection band *j* as follows [26],

$$\vartriangle w_{ji}(k) = \eta \delta_j x_i + \alpha \vartriangle w_{ji}(k-1)$$

Whereηis a learning rate parameter, α is the momentum coefficient, and δ_j is a factor depending on whether neuron j is an output neuron or a hidden neuron. For output neurons

$$\delta_j = \left(\frac{\partial f}{\partial net_j} \right) (y_j^{(k)} - y_j)$$

where $net_j \equiv \sum x_i w_{ji}$ *and* $y_j^{(k)}$ is the desired output for neuron j.

Genetic Algorithm (GA)

The evolutionary based Genetic Algorithm (GA) was formulated on the basis of the Darwinian principle of the survival of the fittest as well as the natural process of evolution via reproduction. It possessesthe capability to reach near optimal solutions for huge problems. GA based learning is utilized to discover near-optimal solutions globally from search space without computing gradient data. GAs originally uses binary string vector representations such as the chromosome structure of biology. But, binary GA has drawbacks as its learning primary objective is to enhance accuracy of the network and no thought is given to the speed of convergence and local error.

Theemphasis is to first apply GA learning and next the gradient descent algorithm, inMLP to optimize the best connection weights and minimize local errors with fast convergence [27]. Genetic algorithms imitate nature and through this, obtain powerful optimization algorithms for computation of global optima of given functions. In theory, gene pool is represented by solution space, nature is denoted by the function to be optimized (objective or cost function) while the organisms are given by trial functions (individuals) utilized for working out solutions. One ought to understand the mechanics of natural evolution and later interpret them into the scientific domain. This ultimately leads to a powerful optimization method.

Both protocols have same running time of O(n) [nn5] in theory Both protocols have other similar features; both are valid training protocols for MLP networks. Generally the Back Propagation algorithm tends to be more reliable at finding a similar accuracy in each run of the algorithm.

4. EXPERIMENTAL RESULTS

The experiments are conducted using QWS data set and the results are compared.

Table 1 Experimental Results

	Naïve Bayes	MLP - BPP	MLP-GA
Classification Accuracy	77.81	96.99	97.81
Avg Precision	0.7783	0.9701	0.9782
Avg Recall	0.7795	0.97	0.9781
RMSE	0.2912	0.1328	0.1062

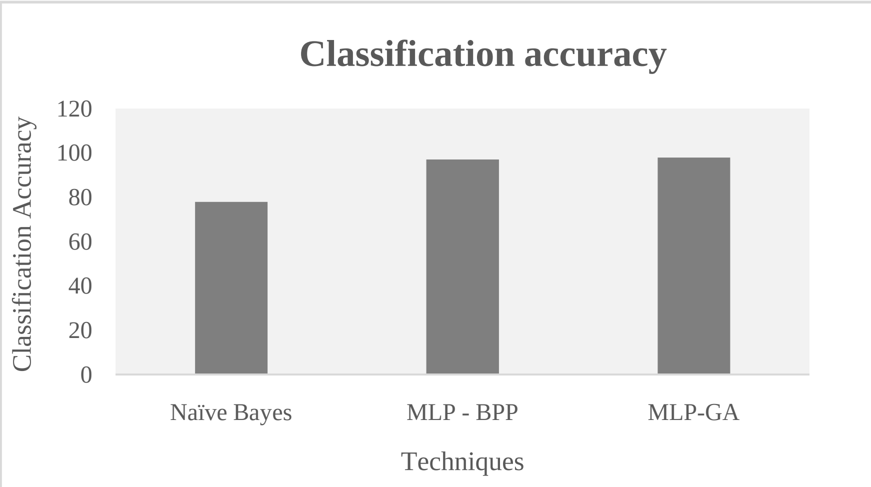

Figure 1 Classification Accuracy

From the figure 1 it is observed that the proposed method MLP with GA improved the accuracy by 22.77% when compared with Naïve Bayes.

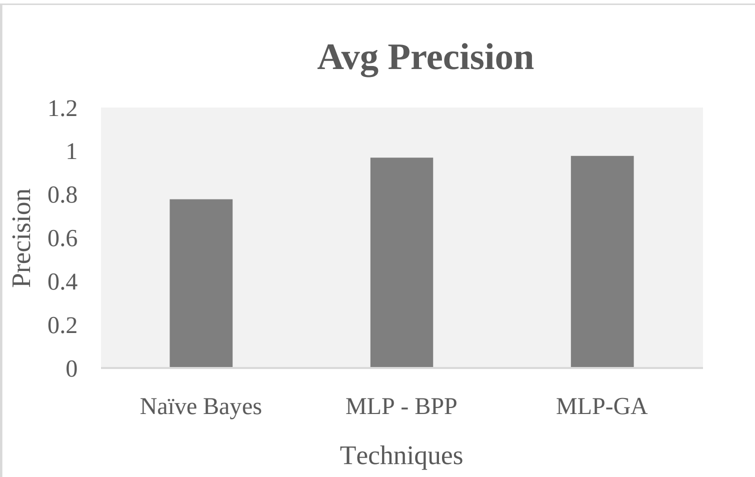

Figure 2 Average Precision

From the figure 2 it is observed that the proposed method MLP with GA improved the average precision by 22.76% when compared with Naïve Bayes. Also MLP BPPimproved the average precision by 21.94% when compared with Naïve Bayes.

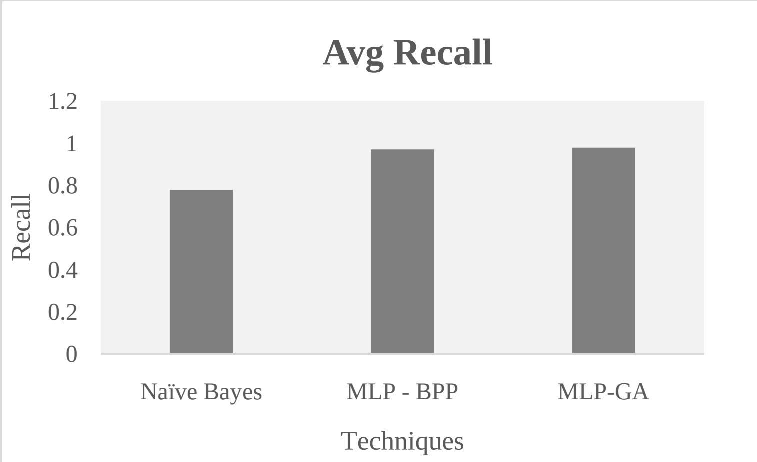

Figure 3 Average Recall

From the figure 3 it is observed that the proposed method MLP with GA improved the average recall by 22.59% when compared with Naïve Bayes. And MLP BPP improved the average recall by 21.77% when compared with Naïve Bayes. .

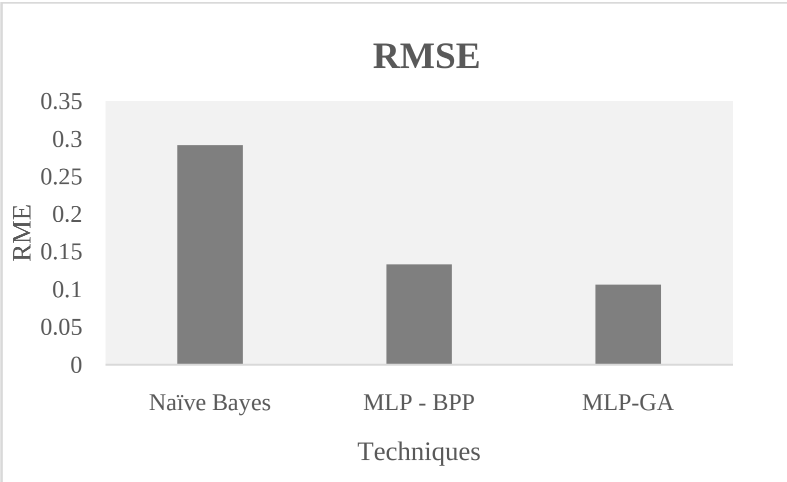

Figure 4 RMSE

From the figure 4 it is observed that the proposed method MLP with GA reduced the RMSE by 93.10% when compared with Naïve Bayes.

5. CONCLUSION

A Web service is an interface defining a set of operations which are network accessible through standardized XML messaging. This work focuses on the classification of Web services. MultiLayer Perceptron is a classifier using back propagation to sort instances. Optimizing the number of hidden layer neurons to establish a MLP to solve problems is an unsolved task. So we proposed Back propagation to train MLP and Genetic algorithm to optimize MLP. Experiments conducted and the results of the three classifiers compared with each other. The proposed method improved accuracy, precision, recall and reduced the RMSE when compared with Naïve Bayes classifier.

REFERENCES

[1] Khan, F. H., Bashir, S., Javed, M. Y., Khan, A., &Khiyal, M. S. H. (2010). QoS based dynamic web services composition & execution. arXiv preprint arXiv:1003.1502.
[2] Wang, H., Huang, J. Z., Qu, Y., &Xie, J. (2004). Web services: problems and future directions. Web Semantics: Science, Services and Agents on the World Wide Web, 1(3), 309-320.
[3] Samtani, G., &Sadhwani, D. (2002). Web Services and Application Frameworks (. NET and J2EE). In Web Services Business Strategies and Architectures (pp. 273-289). Apress.
[4] Kreger, H. (2001). Web services conceptual architecture (WSCA 1.0). IBM Software Group, 5, 6-7.

[5] Yuan-jie, L., & Jian, C. (2012, May). Web service classification based on automatic semantic annotation and ensemble learning. In Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 2012 IEEE 26th International (pp. 2274-2279). IEEE.
[6] AbuJarour, M., Naumann, F., &Craculeac, M. (2010). Collecting, annotating, and classifying public web services. In On the Move to Meaningful Internet Systems: OTM 2010 (pp. 256-272). Springer Berlin Heidelberg.
[7] Wang, H., Shi, Y., Zhou, X., Zhou, Q., Shao, S., &Bouguettaya, A. (2010, October). Web service classification using support vector machine. In Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on (Vol. 1, pp. 3-6). IEEE.
[8] Gowri, R., &Lavanya, R. (2013, December). A novel classification of web service composition and optimization approach using skyline algorithm integrated with agents. In Computational Intelligence and Computing Research (ICCIC), 2013 IEEE International Conference on (pp. 1-8). IEEE.
[9] Rajeswary, C. (2012). A survey on efficient evolutionary algorithms for web service selection. International Journal of Management, IT and Engineering, 2(9), 177-191.
[10] Syed Mustafa, A., &Swamy, K. (2015, June). Web Service classification using Multi-Layer Perceptron optimized with Tabu search. In Advance Computing Conference (IACC), 2015 IEEE International (pp. 290-294). IEEE.
[11] Qi, S., Tang, X., & Chen, D. (2012, November). An Automated Web Services Composition System Based on Service Classification and AI Planning. In Cloud and Green Computing (CGC), 2012 Second International Conference on (pp. 537-540). IEEE.
[12] Sowmya Kamath, S., Ahmed, A., & Shankar, M. (2015, June). A composite classification model for web services based on semantic & syntactic information integration. In Advance Computing Conference (IACC), 2015 IEEE International (pp. 1169-1173). IEEE.
[13] Song, Z., & Tang, X. (2011, December). An automated approach to Web Service Classification based on Semantic. In 2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing (pp. 1207-1210). IEEE.
[14] Raj, T. F., SivaPragasam, P., BalaKrishnan, R., Lalithambal, G., &Ragasubha, S. (2015, March). QoS based classification using K-Nearest Neighbor algorithm for effective web service selection. In Electrical, Computer and Communication Technologies (ICECCT), 2015 IEEE International Conference on (pp. 1-4). IEEE.
[15] Li, Z., O'Brien, L., Keung, J., & Xu, X. (2010, May). Effort-oriented classification matrix of web service composition. In Internet and Web Applications and Services (ICIW), 2010 Fifth International Conference on (pp. 357-362). IEEE.
[16] Shafiq, O., Alhajj, R., Rokne, J. G., &Toma, I. (2010, August). Light-weight semantics and bayesian classification: A hybrid technique for dynamic web service discovery. In Information Reuse and Integration (IRI), 2010 IEEE International Conference on (pp. 121-125). IEEE.
[17] Sawant, V. V., &Ghorpade, V. R. (2014, December). Automatic semantic classification and categorization of web services in digital environment. In Computer and Communications Technologies (ICCCT), 2014 International Conference on (pp. 1-6). IEEE.
[18] Mustafa, A. S., &Kumaraswamy, Y. S. (2014, November). Data mining algorithms for Web-services classification. In Contemporary Computing and Informatics (IC3I), 2014 International Conference on (pp. 951-956). IEEE.
[19] Li, G., Zhang, J., Peng, A., & Wang, R. (2010, August). Web Services System Optimization Based on Modular Classification and Multi-Threading. In Management and Service Science (MASS), 2010 International Conference on (pp. 1-4). IEEE.
[20] Mohanty, R., Ravi, V., & Patra, M. R. (2012). Classification of web services using Bayesian network.
[21] Mustafa, A. S., &Kumaraswamy, Y. S. (2014). Performance evaluation of web-services classification. Indian Journal of Science and Technology, 7(10), 1674-1681.
[22] RAJ, T., Michael, T. F., RAVICHANDRAN, K., & RAJESH, K. (2014). Domain Specific Web Service Composition by Parameter Classification Using Naïve Bayes Algorithm. World Applied Sciences Journal, 29, 99-105.
[23] Yang, J., & Zhou, X. (2015). Semi-automatic Web Service Classification Using Machine Learning. International Journal of u-and e-Service, Science and Technology, 8(4), 339-348.
[24] Singh, K., & Agrawal, S. (2011). Performance Evaluation of Five Machine Learning Algorithms and Three Feature Selection Algorithms for IP Traffic Classification. IJCA Special Issue on Evolution in Networks and Computer Communications (1), 25-32.
[25] Patel, P. B. (2009). An IVR call performance classification system using computational intelligent techniques (Doctoral dissertation, Faculty of Engineering and the Built Environment, University of the Witwatersrand).
[26] Rodrak, S., &Wongsawat, Y. (2012, December). On the classification of EEG/HEG-based attention levels via time-frequency selective multilayer perceptron for BCI-based neurofeedback system. In Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific (pp. 1-4). IEEE.
[27] Jayaraj, V., & Raman, N. S. A Genetic Algorithm Optimized Multi-Layer Perceptron for Software Defect Prediction.