Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

OPEN ACCESS Journal of Applied Sciences

ISSN 1812-5654
DOI: 10.3923/jas.2016.223.229

Research Article
Discriminant Analysis of Web Services Successability

[1]F. Ezhil Mary Arasi, [2]S. Govindarajan and [2]A. Subbarayan

[1]Department of Computer Science Engineering, SRM University, Kattankulathur, Chennai, 603203 Tamil Nadu, India
[2]Department of Master of Computer Applications, SRM University, Kattankulathur, Chennai, 603203 Tamil Nadu, India

Abstract

Background and Objective: Web Services are e-services have become an important method for online business services. Service oriented architecture is based on combining several web services each one responsible to develop a concrete task, in order to obtain full professional software. Quality of service is an important attribute for selecting a web service during the service composition process and it encompasses a group of nonfunctional properties viz., response time, reliability, availability, throughput, latency, successabilty, best practices, documentation and compliance. **Methodology:** An attempt is made to apply discriminant analysis method for identifying the properties that are discriminating. A discriminant analysis model is constructed using successability percentage as he categorized variable and treating the predictor variables viz., availability, reliability and the compliance. **Results:** Results clearly reveal that availability is the primary one which discriminants between the percentages of successability (below 75% and above 75%). **Conclusion:** Preliminary analysis indicates that there exists significant difference in means for availability and compliance.

Key words: Web services, quality of services, discriminant function, predictor variables, Wilks' lambda

Received: March 08, 2016 Accepted: March 28, 2016 Published: April 15, 2016

Citation: F. Ezhil Mary Arasi, S. Govindarajan and A. Subbarayan, 2016. Discriminant analysis of web services successability. J. Applied Sci., 16: 223-229.

Corresponding Author: A. Subbarayan, Department of Master of Computer Applications, SRM University, Kattankulathur, Chennai, 603203 Tamil Nadu, India

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

In the recent years a new technology called web services has emerged. The highlighted characteristic of a web service is that it is a piece of software that the user can utilize but doesn't own, that is, the user doesn't install the software but uses it through the internet and standard protocols. Web services and e-services have become an important media for online business services. With this new technology, a new architecture paradigm called Service Oriented Architecture (SOA) has appeared. This architecture is based on combining several web services, each one responsible to develop a concrete task, in order to obtain full operational software. World wide web consortium has given a common definition of service: A service is an abstract resource that characterizes a capability of performing tasks that form a coherent functionality from the point of view of provider's entities and requester's entities. To be used, a Service must be realized by a concrete provider agent[1]. Web services that compose a SOA system might be able to perform a task in a certain time, might be unavailable in some cases, might have security policies, etc. All this attributes, named Quality of Service (QoS) is an important attribute for selecting a service during the service composition process. One of the key issues for adding QoS information to the web services is, in fact, to provide to the service client a way to compare and finally choose the web service that best fulfills his non functional requirements from those ones which offers the same functionality. Discriminant analysis is a technique to discriminate between two or more mutually exclusive and exhaustive groups on the basis of some explanatory variables. These groups are known as priori. When the criterion variable has two categories, the technique is known as two-group discriminant analysis. When three or more categories are involved, the technique is referred to as multiple discriminant analysis[2].

Now-a-days, as the number of functionally-equivalent services with different QoS levels is increasing rapidly, the search space size of compositions is growing dramatically. The efficiency of service composition mechanisms becomes a challenging issue according to Chen *et al.*[3]. For those web services providing the same functionality, Quality of Service (QoS) has been mostly applied to represent their nonfunctional properties and differentiate them for service composition. The QoS is a broad concept that encompasses a group of nonfunctional properties such as response time, reliability, availability, throughput, latency, successability, best practices, documentation and compliances. The problem of QoS-based service selection has attracted the attention of many researchers and was recently discussed in a number of studies. Ran[4] suggested the use of a QoS certifier that certifies the QoS claims made by service providers about their corresponding services. Zheng *et al.*[5] have proposed a personalized ranking prediction framework, named cloud rank to predict the QoS ranking of a set of cloud services without requiring additional real-world service invocations from the intended users. Al-Masri and Mahmoud[6] has outlined the approaches for dealing with QoS metrics for web services. In this context they have cautioned that relying on the Service providers to supply their QoS metrics may lead to manipulations.

Aruna *et al.*[7] proposed a novel algorithm for the selection of most discriminating features to improve the complexity of Geometrical Structure Anomaly Detection (GSAD) based on a linear discriminant function. De Oliveria *et al.*[8] performed cluster analysis of internet users based on hourly traffic utilization using partitioning around medoids and Ward's method being the preferred clustering methods. They have validated the cluster structure using discriminant analysis. According to Mallaya *et al.*[9] QoS parameters of web services act as a discriminator in identifying the suitable web services from the set of available web services. Cai *et al.*[10] presented a novel algorithm for discriminant analysis, called Spectral Regression Discriminant Analysis (SRDA). By using spectral graph analysis, SRDA casts discriminant analysis into a regression framework that facilitates both efficient computation and the use of regularization techniques. Garcia *et al.*[11] have proposed a new feature selection methodology. The methodology is based on the stepwise variable selection procedure, but, instead of using the traditional discriminant metrics such as Wilks' lambda, it uses an estimation of the misclassification error as the figure of merit to evaluate the introduction of new features. Omar *et al.*[12] presented the application of Discriminant Function Analysis (DFA) of stroke brainwave for ischemic stroke group level discrimination. The Relative Power Ratio (RPR) of stroke brainwave has been selected as an input for DFA. This study presented an analysis on the capability of DFA to discriminate three different stroke group levels. The classification rule acted as a pattern recognition tool in grouping the group of stroke level. In this study a lexicon driven segmentation-recognition scheme is proposed by Jayadevan *et al.*[13] for the recognition of legal amount words from Indian bank cheques written in English. A water reservoir concept is used to pre-segment the words into primitive components and the primitive components of a word are then merged into possible characters to get the best word using the lexicon of 36 different legal words of bank cheque. To merge these primitive components into characters and to get

optimum character segmentation, dynamic programming is employed using total likelihood of the characters of a word as an objective function. To calculate the likelihood of a character, Modified Quadratic Discriminant Function (MQDF) is used. Nakayama *et al.*[14] have developed a computerized scheme for detecting early-stage micro calcification clusters in mammograms. The Bayes discriminant function was employed for distinguishing among abnormal ROIs with a micro calcification cluster and two different types of normal ROIs without a micro-calcification cluster. They also evaluated the detection performance by using 600 µg. Discriminant analysis was used by Erimafa *et al.*[15] to predict the class of degree obtainable in a university system. The conditions for predictive discriminant analysis were obtained and the analysis yielded a linear discriminant function which successfully classified or predicted 87.5% of the graduating student's class of degrees.

MATERIALS AND METHODS

Motivation for the present study: Ran[4] reported that the performance of a web service represents how fast a service request can be completed. It can be measured in terms of throughput, availability, response time, latency, execution time and transaction time and so on. Most of the studies carried out have not ensured the services balance of functionality and non functionality of user preferences. The author wishes to note that practicality and versality are the prime factors in the selection of services. Quality of service (QoS) of a web service can be used as a discriminating factor that differentiates the functionally similar web service.

Methodological aspects of QoS: Web services QoS can be described as a set of non-functional attributes that may impact the quality of the service offered by a web service. Quality-of-Service (QoS), which is usually employed for describing these non-functional characteristics, has become an important differentiating point of different web services. The QoS parameters help to determine which of the available web services is the best and meets client's requirements. Because of their significance, we selected the following QoS parameters for identifying which QoS parameter is the best discriminating variable having highest correlation with discriminant function:

- **Response Time (RT):** The time taken to send a request and receive a response (msec)
- **Availability (AV):** A ratio of the time period when a web service is available (%/3-day period)
- **Throughput (TP):** The maximum requests that are handled at a given unit in time (requests min^{-1})
- **Reliability (RE):** Ratio of the number of error messages to total messages (%)
- **Compliance (CO):** The extent to which a WSDL document follows WSDL specification (%)
- **Successability (SU):** Number of response/number of request messages (%)
- **Latency (LA):** Time taken for the server to process a given request (msec)

Discriminant analysis for the study of QoS: This study investigate the following based on the information obtained through the samples of Web Services out of 2507 web service providers. A sample of thousand web services was selected randomly for identifying the predictor variables treating successability as the dependent variable. Preliminary analysis of the data indicates that the 75% of successability covers around 1900 web services out of 2507. This is the primary basis for sample selection. The sample has been divided into two groups based on two groups based on the 75% successability of the web services (i.e. 500 web services with less than 75% successability and 500 web services with more than 75% successability).

Discriminant is a statistical method that allows the researchers to study the differences among groups. This method provides a unified approach to research situations where there are several groups and it is desirable to:

- Establish significant group differences
- Study and describe the variables on which groups differ
- Classify individuals into the most appropriate group

When discriminant analysis is used to separate two groups, it is called Discriminant Factor Analysis (DFA). When we separate more than two groups, the Canonical Variable Analysis (CVA) method is used. The most common application of DFA is to include many measures of the study, in order to determine the ones that discriminate between groups.

RESULTS

Discriminant analysis model and description: The mathematical form of the discriminant analysis model is:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \ldots\ldots\ldots\ldots + b_kX_k \quad (1)$$

Where:
Y = Dependent variable
b_s = Coefficients of independent variables
X_s = Predictor or independent variables

It may be kept in mind that the dependent variable Y should be a categorized variable, whereas the independent variables X_s should be continuous. As the dependent variable is a categorized variable, it should be coded as 0, 1 or 1, 2 and 3, similar due to the dummy variable coding. The method of estimating b_s is based on the principle that the ratio between group sum of squares to within group sum of squares be maximized. This will make the groups differ as much as possible on the values of the discriminant function. After having estimated the model, the b_s coefficients are used to calculate Y, the discriminant score by substituting the values of X_s in the estimated discriminant model.

The relative importance of the independent variables could be determined from the standardized discriminant function coefficient and the structure matrix. The difference between the standardized and unstandardized discriminant function is that in the unstandardized discriminant function have a constant term, whereas in the standardized discriminant function, there is no constant term.

The criteria for the selection of the variables under study are:

- For all the predictor variables Mean>Standard Deviation
- The variables are in the same unit (%)

A descriptive account of the variables relating to QoS is given in the starting of this session. Among the list of variables specified, to examine the impact of predictor variables on the dependent variable:

- **The predictor variables:** Availability, reliability and compliance
- **The dependent variable:** Successability

Empirical analysis based on discriminant function: The discriminant analysis involves the following:

- The percentage of sample which helps in correct classification
- To study the statistical significance of the discriminant function
- Identifying variables which are relatively better in discriminating between two groups

For comparative purposes we have presented the mean and standard deviation in respect of the predictor variables in Table 1.

The mean percentage of availability (in respect of successability <75%) is compared with the percentage of availability (in respect of successability >75%). This clearly indicates this predictor variable plays major role compared to the other two predictor variables namely reliability and compliance.

Test of equality of group Means table is given in the following Table 2:

- From the above Table 2 we observed that there exists significant difference for availability (AV) and compliance (CO) for which the p-values are 0.00 which is less than 0.05, the assumed level of significance
- There does not seem to be any significant difference in the Mean in respect of the predictor variable reliability as the p-value is >0.05

The correlation matrix for the data set is presented in Table 3.

The value of correlation coefficient does not indicate any multicollinearity since none of the correlation coefficient is greater than 0.75.

Unstandardized discriminant function: Canonical discriminant function coefficients are presented in below Table 4.

Table 1: Group statistics

Categorical variable	Predictor variable	Mean	Standard deviation
1.0 (Successability <75%)	AV	49.736	17.3153
	RE	68.396	10.0577
	CO	84.134	8.6622
2.0 (Successability >75%)	AV	89.184	4.0929
	RE	68.932	7.9552
	CO	89.204	11.1340
Overall	AV	69.460	23.3998
	RE	68.664	9.0670
	CO	86.669	10.2875

Table 2: Test of equality of group means

Predictor variable	Wilks' lambda	F	df1	df2	Significance
AV	0.289	2457.824	1	998	0.000
RE	0.999	0.874	1	998	0.350
CO	0.939	64.585	1	998	0.000

Table 3: Correlation matrix

Predictor variables	AV	RE	CO
AV	1.000	0.297	0.030
RE	0.297	1.000	0.029
CO	0.030	0.029	1.000

Table 4: Canonical discriminant function coefficients

Predictor variables	Function 1
AV	0.082
RE	-0.032
CO	0.013
(Constant)	-4.640

Table 5: Eigen values

Function	Eigen value	Variance (%)	Cumulative (%)	Canonical correlation
1	2.719[a]	100.0	100.0	0.855

Table 6: ANOVA

Sources of variation	df	Sum of squares	Mean Square	F	Significance
Between groups	1	2713.971	2713.971	2713.97	0.000
Within groups	998	998.000	1.000		
Total	999	3711.971			

Table 7: Wilks' lambda

Wilks' lambda	Chi-square	df	Significance
0.269	1308.967	3	0.000

Table 8: Standardized canonical discriminant function coefficients

Predictor variables	Coefficients
AV	1.035
RE	-0.293
CO	0.132

Table 9: Structure matrix

Predictor variables	Coefficients
AV	0.952
CO	0.154
RE	0.018

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions variables ordered by absolute size of correlation within function

The discriminant function is given by:

$$Y = -4.640+0.082\ X_1-0.032\ X_2+0.013\ X_3 \quad (2)$$

Substituting the values of X_1, X_2 and X_3 in respect of each sample in the above equation we can compute the discriminant values using the SPSS.

The Eigen value of the above estimated discriminant function is 2.719 from Table 5. The canonical correlation value is 0.855. The square of the correlation coefficient is 0.7310, which means 73% of the variance in the discriminant model between two groups is due to changes in the predictor variables listed in the analysis. The one-way ANOVA carried out for the sampled data is given in Table 6.

Wilks' lambda and Chi-square values are presented in the following Table 7.

The computed value of Wilks' lambda is 0.269 which is in the range of 0 and 1. This clearly indicates the significance of the discriminant function.

Standardized discriminant function: The standardized canonical discriminant function coefficients are given in Table 8.

The discriminant function is given by:

$$Y = 1.035\ X_1-0.293\ X_2+0.132\ X_3 \quad (3)$$

The discriminant function clearly indicates that the predictor variable availability (AV) is the primary one which discriminates between the percentages of successability (below 75% and above 75%). The structure matrix is given in the following Table 9.

The result of the above structure matrix indicates that the predictor variable availability (AV) is the most important in this analysis for which the correlation coefficient is 0.952. This correlation coefficient is larger than the other two correlation coefficients namely compliance (CA) and reliability (RE).

DISCUSSION

The preliminary analysis has shown that the predictor variable viz., availability plays an important role compared to other two predictor variables viz., reliability and compliance. The correlation matrix clearly shows that none of the correlation coefficient is greater than 0.75. This indicates the absence of multicollinearity and the same is essential for carrying discriminant analysis. The squared value of the coefficient of canonical correlation is 0.7310. The unstandardized discriminant model shows that 73% of the variance in the discriminant model between the two groups (successability above 75% and successabilty below 75%) is due to changes in the predictor variables. Deepa and Sathiaseelan[16] have attempted an extensive study on QoS based web services selection and composition. This study stated that the need for novel techniques for evaluating QoS based web services. This study also pointed out to use QoS parameters for evaluation. The authors have identified discriminant function methodology as an alternative one for analyzing the sucessabilty as the dependent variable and a host of predictor variables viz., availability, reliability and compliance. Alaget *et al.*[17] have brought out an excellent review of framework and quality of service based web services discovery. They have stated that there is a need for some methods capable to evaluate and compare different services providing the same functionalities. The method developed by the authors of this study is well addressed. Khan *et al.*[18] studied dynamic web services composition and execution based on QoS. The authors have suggested crawling the web for searching web services instead of querying the UDDI registries. It is important

to note that we have collected information on QoS using crawling approach and developed a method for evaluating web services. This complements the study carried out by the above authors. Kumar and Zayaraz[19] developed a QoS aware quantitative web services selection model based on sample data and used statistical procedures. In this study, taken QoS data from the original sources (http://www.uoguelph.ca/~qmahmoud/qws/) and carried disciminant analysis. The authors feel that this is more realistic in nature because the source of data is original.

Susila and Vadivel[20] developed procedures using entropy technique for evaluating the best web service. This is another approach which is equally preferred by web services users. This method involves certain specification in quantitative terms than the entropy method in finding predictor variables. Zhang[21] have used particle swarm optimization method for the web service selection based on QoS requirements. They have concluded that this method does not clearly indicate the best choice of parameters in web services selection. This method of selection of parameters of web services is clear in differentiating the variables selection. Wang *et al.*[22] used fuzzy linear programming technologies for QoS aware service selection model to identify the variables which will lead to optimal solution. This technique requires computational complexities and in this sense our proposed method leads to easy computation.

CONCLUSION

The standardized discriminant model clearly indicates that predictor variables viz., availability is the primary one which discriminate between the two groups and the correlation for availability is larger than the correlation coefficients viz., compliance and reliability. The results of the study may help in formulating policies for effective usage of web services. Effective implementation of best policies will in turn result in optimization of resources.

ACKNOWLEDGEMENTS

We are grateful to the authorities of SRM University, 603203 Kattankulathur, India for providing the necessary facilities for carrying out this study work. The views, findings and interpretations presented in this study are the sole responsibility of the authors.

REFERENCES

1. World Wide Web Consortium(W3C), 2004. Web services glossary. W3C Working Group Note February 11, 2004. http://www.w3.org/TR/ws-gloss
2. Kothari, C.R. and G. Garg, 2014. Research Methodology-Methods and Techniques. 3rd Edn., New Age International Publishers, India, pp: 373-383.
3. Chen, Y., J. Huang, C. Lin and J. Hu, 2015. A partial selection methodology for efficient qos-aware service composition. IEEE Trans. Serv. Comput., 8: 384-397.
4. Ran, S., 2003. A model for web services discovery with QoS. ACM SIGecom Exchanges, 4: 1-10.
5. Zheng, Z., X. Wu, Y. Zhang, M.R. Lyu and J. Wang, 2013. QoS ranking prediction for cloud services. IEEE Trans. Parallel Distrib. Syst., 24: 1213-1222.
6. Al-Masri, E. and Q.H. Mahmoud, 2007. A framework for efficient discovery of web services across heterogeneous registries. Proceedings of the 4th IEEEE Consumer Communications and Networking Conference, January 11-13, 2007, Las Vegas, NV., USA., pp: 415-419.
7. Jamdagnil, A., Z. Tan, P. Nanda, X. He and R.P. Liu, 2010. Mahalanobis distance map approach for anomaly detection. Proceedings of the 8th Australian Information Security Mangement Conference, November 30, 2010, Edith Cowan University, Perth Western Australia, pp: 8-15.
8. De Oliveira, M.R., R. Valadas, A. Pacheco and P. Salvador, 2007. Cluster analysis of internet users based on hourly traffic utilization. IEICE Trans. Commun., 90: 1594-1607.
9. Mallayya, D., B. Ramachandran and S. Viswanathan, 2015. An automatic web service composition framework using QoS-based web service ranking algorithm. Scient. World J. 10.1155/2015/207174
10. Cai, D., X. He and J. Han, 2008. SRDA: An efficient algorithm for large-scale discriminant analysis. IEEE Trans. Knowledge Data Eng., 20: 1-12.
11. Garcia, H.C., J.R. Villalobos, R. Pan and G.C. Runger, 2009. A novel feature selection methodology for automated inspection systems. IEEE Trans. Pattern Anal. Mach. Intell., 31: 1338-1344.
12. Omar, W.R.W., M.N. Taib, R. Jailani, Z. Mohamad, A.H. Jahidin and Z. Sharif, 2014. Application of Discriminant function analysis in ischemic stroke group level discrimination. Proceedings of the IEEE 10th International Colloquium Signal Processing and its Applications, March 7-9, 2014, Kuala Lumpur, pp: 229-232.
13. Jayadevan, R., U. Pal and F. Kimura, 2010. Recognition of Words from legal amounts of Indian bank cheques. Proceedings of the International Conference on Frontiers in Handwriting Recognition, November 16-18, 2010, Kolkata, pp: 166-171.

14. Nakayama, R., Y. Uchiyama, K. Yamamoto, R. Watanabe and K. Namba, 2006. Computer-aided diagnosis scheme using a filter bank for detection of microcalcification clusters in mammograms. IEEE Trans. Biomed. Eng., 53: 273-283.
15. Erimafa, J.T., A. Iduseri and I.W. Edokpa, 2009. Application of discriminant analysis to predict the class of degree for graduating students in a university system. Int. J. Phys. Sci., 4: 16-21.
16. Deepa, A.F. and J.G.R. Sathiaseelan, 2015. An empirical study on QoS-based web services selection and composition. Proceedings of the International Conference on Innovations in Information, Embedded and Communication Systems, March 19-20, 2015, Coimbatore, pp: 1-5.
17. Algaet, M.A., Z.A.B. Muhamad Noh, A.S. Shibghatullah and A.A. Milad, 2013. A review on framework and quality of service based web services discovery. Eur. J. Scient. Res., 111: 300-322.
18. Khan, F.H., S. Bashir, M.Y. Javed, A. Khan and M.S.H. Khiyal, 2010. QoS based dynamic web services composition and execution. Int. J. Comput. Sci. Inform. Secur., 7: 147-152.
19. Kumar, R.D. and G. Zayaraz, 2011. A Qos aware quantitative web service selection model. Int. J. Comput. Sci. Eng., 3: 1534-1538.
20. Susila, S. and S. Vadivel, 2011. Web service selection based on QoS attributes using entropy discretization method. Int. J. Comput. Applic., 30: 47-53.
21. Zhang, T., 2014. QoS-aware web service selection based on particle swarm optimization. J. Networks, 9: 565-570.
22. Wang, P., K.M. Chao and C.C. Lo, 2010. On optimal decision for QoS-aware composite service selection. Expert Syst. Applic., 37: 440-449.