World Applied Sciences Journal 29 (Data Mining and Soft Computing Techniques): 99-105, 2014
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2014.29.dmsct.18

# Domain Specific Web Service Composition by Parameter Classification Using Naïve Bayes Algorithm

[1]T.F. Michael Raj, [2]K.S. Ravichandran and [1]K. Rajesh

[1]Department of CSE, SRC, SASTRA University, Kumbakonam, Tamil Nadu, India
[2]School of Computing, SASTRA University, Thanjavur, Tamil Nadu, India

**Abstract:** Web service composition for the specific implementation or business process is typically based on the service description and its parameter constructs. Discovering and compositing the WS from the pool of web services is a challenging task and more attention has to be given on the similarities checking. In this paper we propose an approach on identifying the most appropriate service based on the user's preferences of the requested WS. The given WS description may contain the parameters, which may have relations with the requested WS of the specific domain in different aspects like name, parameters and types. The domain specific WS classification can be done using Naive Bayes classification algorithm and we have given the experimental evaluation on the sample WSDL datasets and the simulated results shows that effectiveness and efficiency of the proposed approach.

**Key words:** Web service classification · Web service similarity · Service discovery · Web service composition

## INTRODUCTION

Web service (WS) composition in any business process models is the important process; hence most of the researchers pay their attention and taking efforts on the development of the WS. The WS providers describe the functionality and its parameters in the web service description language (WSDL), consists of various terminologies. These terminologies are belongs to the particular domain and are composed by the domain experts. Business Process Models constitutes its tasks based on the discovery and the selection of the appropriate services [1]. The objective of the WS classification is to achieve the integrated interoperability between the applications. In the business process context WS has to follow the various industry standards like WSDL, UDDI, etc [2, 3].

To classify the WS in the dataset registry and conduct the experiments on it, the available WS datasets are Seekda [4] and Biocatalogue [5] can be used. The classification phase of the WS starts with the user's request, then navigating through the structure WSDL [6]. The specification of WSDL includes various structured elements based on the XML specification, been represented in the Figure 1.

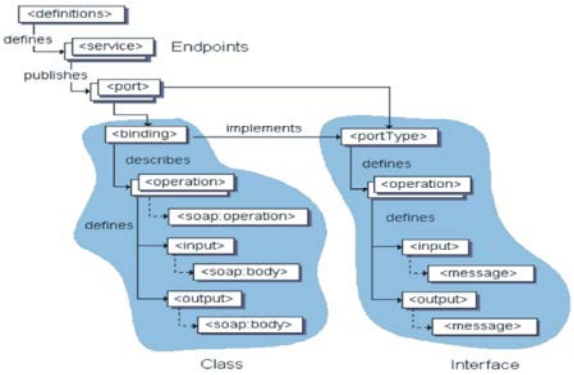


Fig. 1: Structure of WSDL

In this model we have considered the business process WS as our domain of classification. The designated WSs are distributed across the domains. Let us consider this example for this scenario.

**Example:** Let *A* be a service provider (SP) who published the numerous WSs and their descriptions are stored in the repository of different domains and *clientX* be a service requester who need to implement a task by integrating the WSs the objective of the process can be fulfilled. For that he/she need to identify the WSs and

**Corresponding Author:** T.F. Michael Raj, Department of CSE, SRC, SASTRA University, Kumbakonam, Tamil Nadu, India.

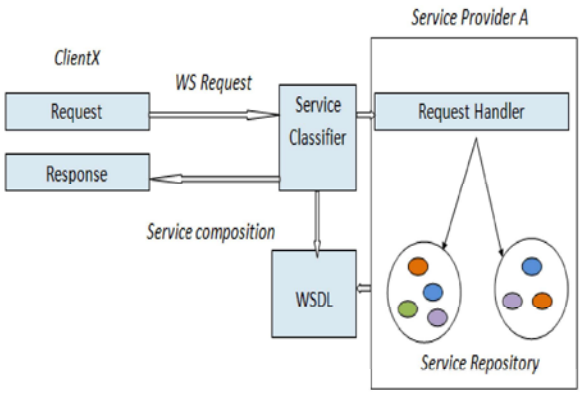


Fig. 2: Service retrieval process

make a composition which is related to his/her business domain based on his request. This process is shown in the following Figure 2.

The remaining paper has been organised as follows that Section-II related concepts and literature survey of the various classifications, Section-III gives the proposed architecture, Section-IV provides the evaluation results and Section-V concludes the model with the future enhancements.

**Related Concepts and Literature Survey:** The service oriented architecture is a platform for web services based on the SOAP [8]. The business organization wants to implement the features of interoperability, reusability and flexibility among the B2B (business to business) enterprise applications. Business processes can be succeeded and completed if they are incorporating the knowledge of WS and its customization [9]. This customization can be enabled by classifying the WS and dynamically composing and invoking them through the machine learning. There is numerous classification methods available and are employed to categorize the Wss.

**Web Services:** W3C defines the term web services as a heterogeneous software component provides the interoperability and dynamically exchanges the information [10]. The dynamic discovery and composition is depends upon the processing the contents of the WSDL. The WSDL datasets are processed to identify the set of classes and groups related to the requirements. Due to the easy implementation and cost effectiveness of the Naive Bayes classification it can be used to categorise the WSs because it counts only the class distribution [11].

**WS Classification Models:** WS classification can be done in two ways Functional classification and Non-Functional classification. The functional classification based on the process carried over by the WS. It is very difficult to do such classifications. Since it's a black box to the clients, the functionality of the WS cannot be identified but a generic classifications can done according to the usage of the service namely elementary services and composite services.

**Elementary Services:** It's a smallest component used to achieve short term objectives. For example, a Currency converter service, Temperature converter service.

**Composite Services:** A Business model like online shopping, bank transactions uses the composite WSs to provide integrated service to their consumer. It can be done in two ways that dynamic and static composition. These composite services also provide the interoperability with the environment.

Non-Functional Classifications (NFC) based on the criteria that suit the requirements and data mining concepts. There are NFC can be done in various ways and are represented as a Model. The NFC properties are hierarchical representation of properties, classification of user preferences, evaluation of properties, user's behaviour, dynamic aggregation and automation [12]. Jamous said that NFC can produce the results in a beneficial way to classify the WS, which is based on the three level classifications [13]. The distribution of WS will be most important factor of classification in NFC. The NFP method is based on the policies, which are specified in the QOS attributes of WS given by the [14, 15].

The SVM technique based domain specific service classification is implemented with LIBSVM tool [16]. This model has some limitations that SVM needs to be trained before the classification hence it does not classify the WSs dynamically that is newly added WSs could not be considered for the classification and also it requires user guidance for the classification [17].

**Naive Bayes Classification:** Naive Bayes theorem provides the probabilistic classifier approach capable of producing the results interpreting the user queries. Bayes' theorem is dealt with the conditional probabilities which influence the event on the probability on another event. The terminologies prior and posterior probabilities are associated with it. It's also provides the feature adding dynamic data to the class. Naive Bayes classifies

the data for any case, if the class is given. The classes and attributes are independently processed by the Naïve Bayes Classifier and this phenomenon of the class is called class independence [18]. This approach has given the results which are effective in most of the applications [19-21].

"Machine learning (ML) is programming computers to optimize a performance criterion using example data or past experience" [22]. ML processes the web request and identifies the relationships between the WS and its parameters. The computer program makes this learning process and optimises the parameters of datasets and generates the better results using model. The learning process starts with the training datasets. One of the machine learning algorithms are supervised learning, where pre-processed classes of the training data set are identified.

Bayes' Theorem can be written as:

$$P\left(\frac{A}{B}\right) = \frac{P\left(\frac{B}{A}\right)p(A)}{P(B)} \qquad \text{Eq - 1}$$

where,

P (A) is the prior probability of A
P (B) is the prior probability of B
P (A|B) is the posterior probability of A given B
P (B|A) is the posterior probability of B given A

Since the denominator P (B) in Eq. 3 is the probability of the evidence without any knowledge of the event A and since the hypothesis A can be true or false, Bayes' theorem can also be written as

$$P\left(\frac{A}{B}\right) = \frac{P\left(\frac{B}{A}\right)p(A)}{P\left(\frac{B}{A}\right)XP(\alpha) + \mathrm{P}(\mathrm{B}|\neg\mathrm{A})\,X\ \mathrm{p}\,(\neg\mathrm{A})} \qquad \text{Eq -2}$$

where,

P (¬A) is the probability of A being false
P (B|¬A) is the probability of B given A is false

**Web Service Composition:** WS discovery and composition is important task of the business process models. The composition of WSs starts with matchmaking process and it relies on UDDI. The publishing, finding and compositing the WS depends on the user's requirements and the appropriate query includes the input and output parameters. IOPE model also provide simple match making process [23]. QOS based service composition relies on the various parameters response time, availability, price, reputation, integrity and so on. PROMOTHEE deals with QOS composition [24].

**Research Challenges:** In syntactic based environment identifying appropriate WS is a very difficult task. The description may not be clear to represent the WSs. The internal behaviour may not be well defined. An NFC property plays the important role in the service selection and composition. Here we have considered WS parameters name and type as the NFC properties and are defined in WSDL. The NFC of WS also depends upon the other issues of selection environment.

- Service requirement of the request
- Discovering the WS related to the request
- Evaluate the WSDL of the WS to select the appropriate one
- Handling the missing features
- Make a composition of WS which will be used to implement the business task

**Proposed Model:** The main objective of this model is to provide generic solution for this problem; probabilistic text categorization can be used. It provides the best way of categorization of the WS based on the user's request. The following Eq-3 shows the probabilistic equation of WS category

$$\text{Probability of WS category based on the parameter} = (\mathrm{WS}_{Category} / \mathrm{WSDL}_{Document}) \qquad \text{Eq-3}$$

Here we have implemented NB classifier for WSDL parsing and the model consists of the following steps

- Process the user's request to identify the i/o parameters
- Develop a naïve Bayes classifier to process the WSDL data set for parameter based classification
- Categorise WS according to i/o parameters
- Compose the WSs to fulfil the requirements

**User's Request Process:** The initial and important step of this model is Stemming. It is the process of tuning the query to populate the good results; it improves the performance of information retrieval systems and process the query from the business model's interface. Stemming process incorporates the binary weighting scheme finds the similarities of the given parameters between the queries Q in the queryset and WSDL document $D_{WSDL}$ given by the Eq. 4. The similarities can be calculated follows [25].

$$Similarity(Q, D_{WSDL}) = \frac{(Q \cap D_{WSDL})}{|Q|^{1/2} \cdot |D_{WSDL}|^{1/2}} \qquad \text{Eq} - 4.$$

where,

$|Q|$ is number of terms in the query ($Q$)

$|D_{WSDL}|$ is number of terms in WSDL document ($D_{WSDL}$) and

($Q$ n $D_{WSDL}$) number of terms exists in both query ($Q$) and WSDL document ($D_{WSDL}$)

**Implementation of NB Classifier:** After normalising the query Q can be used for the parameter based WS composition. If the parameters of Q matched with the parameters of WS methods then can be grouped. The selected WSs are evaluated for the effective composition. WSDL document consists of various elements types, operations and binding. Types can be defined in XML schemas it consists of parameter name and type of the web methods. The pre-processed query Q is in the form of Q={$WS_{Methods}$, $WS_{Parameters}$[ ]}and passed as an argument to WSDL dataset, which are in the following format

$D_{WSDL}$= {$WSDL_{Name}$, $_{WSMethod}$, $WS_{Parameters}$[ ], $WS_{ParametersTypes}$[ ], Class}

This model considers the class (C) for the domain specific classification and composition of WS as C={Authentication, Bankdetails, Account, Customercare}. The Classifier procedure produces the result of related WSs to compose.

```
Procedure WSDLProcessor()
{

    Input     : WSDL dataset
    Output    : Contents of D_WSDL= { WSDL_Name, WS_Method, WS_Parameters[ ], WS_ParametersTypes[ ], C }
    Foreach Uri in WebRequest(WSDL dataset)
    {
            Read (Uri)
            Get the service description (SD) <list> SD
            Process the class types (CT) <list> CT
    }
    Foreach MethodInfo in <list> SD
    {
            Get the Method name (MN) <list> MN
            Get the parameters (P) <list> P
    }
}
Procedure Classifier()
{
    Input : IntClassCount, IntTotalMethods, D_WSDL
    Output: <list> Prior probability (PP) of the C, <list> Posterior Probability (POP) of each C
    PP = (IntClassCount / IntTotalMethods )
    Foreach i in PP
    {
```

```
            POP_C= i
            Foreach j in POP_C
            {
                POP_C = POP_C * j
                <list> POP.add(POP_C)
            }
    }
}
```

Table 1: Standard format of the data

| | Data in standard format | | | |
|---|---|---|---|---|
| Case | Feature 1 | Feature 2 | … | Feature n |
| 1 | XX | X | XXX | Yes |
| 2 | XX | X | XXX | No |
| 3 | XX | X | XXX | Yes |
| .. | .. | .. | .. | .. |

Table 2: Sample WSDL Dataset

| WSDL name | WS | Method | Parameters | No of Param |
|---|---|---|---|---|
| *BusinessDataCatalog.wsdl* | *BdcWS* | *GetMethodForMethodInstance* | *lobSystemInstanceId entityId, entityId, methodInstanceId* | *4* |
| *BusinessDataCatalog.wsdl* | *BdcWS* | *GetFilterDescriptorsForMethod* | *lobSystemInstanceId entityId, entityId, methodInstanceId, methodId* | *5* |
| *Authentication.wsdl* | *Authentication* | *Login* | *username, password* | *2* |
| *Authentication.wsdl* | *Authentication* | *Mode* | *username, password* | *2* |
| *AutoComplete.wsdl* | *AutoComplete* | *GetCompletionList* | *prefixText, count* | *2* |
| *BusinessDataCatalog.wsdl* | *BdcWS* | *GetLobSystemInstances* | *-* | *0* |

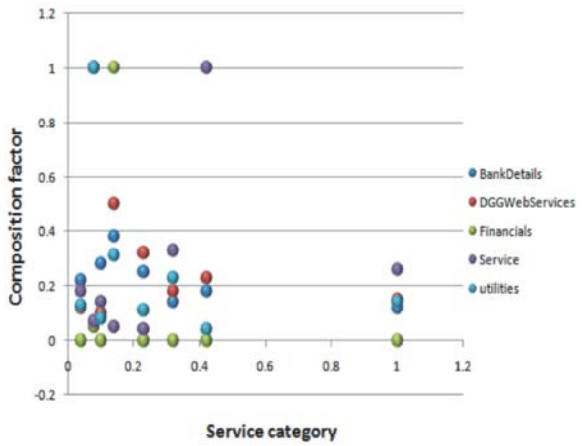


Fig. 3: Simulated results of WS composition

To implement this model we have used the dataset which consists of 3738 WSDL documents [26] and more than 10000 record-sets for the parameter classification. There may be the web methods with missing parameters or parameter-less in the WS. To manage the missing features in the request, ranking concepts and union of all features are used and provides the control over the situation of producing zero probability.

The quality of the query and relevance in the parameters are the important factors of the performance. The following Table 2 shows sample dataset generated by the procedure *WSDLProcessor*.

The model is implemented in C# dot net and SQL server. The composition factor is determined to make domain specific composition. Each composition request consists set of queries called query-set can be the input and it generates the composition list comprises the WS information. Figure 3 shows the performance of this model, proves that the classification WSs and composition of this NB approach is better for the domain specific WS composition.

## CONCLUSION

Web service composition is very essential in all the business models. Since we have proposed Naive Bayes classifier model for the domain specific composition. This model provides the better results than the other models by considering the QOS factors, cost and implementation of WS. But this method has some limitations of not considering functionality based classification. The coupling a nd cohesion properties are

not considered in the composed WSs. Future work of this model can be extended by incorporating concepts of semantic web and providing the ontology support for the WS classification and also to integrate the WS which resides in multiple servers with multi-domain support.

## ACKNOWLEDGMENT

Authors want to thank friends and colleagues at Srinivasa Ramanujan Centre (SRC) SASTRA University for their suggestions and encouragement.

## REFERENCES

1. Tony Andrews, Microsoft, Francisco Curbera, IBM, 2003. Business Process Execution Language for Web Services Specification.
2. http://en.wikipedia.org/wiki/Web_Services_Description_ Language
3. http://en.wikipedia.org/wiki/Universal_Description_Discovery_and_Integration
4. Kritikos, K. and D. Plexousakis, 2007. Semantic QoS-based webservice discovery algorithms. In Proceedings of the 5th IEEE European Conference on Web Services, ECOWS, 07: 181-190.
5. Bhagat, J., 2010. BioCatalogue: a universal catalogue of web services for the life sciences. Nucleic Acids Research, 38(suppl 2): W689-W694.
6. http://download.oracle.com/otn_hosted_doc/jdeveloper/1012/web _ services / ws_wsdlstructure.html
7. Zhang H. and J. Su, 2004. Naive Bayesian classifiers for ranking. European Conference on Machine Learning (ECML2004), Pisa, pp: 501-512.
8. Martin Gudgin, M.H., Noah Mendelsohn, Jean-Jacques Moreau, Henrik Frystyk Nielsen, Anish Karmarkar and Yves Lafon, 2007. SOAP Version 1.2. 27.
9. Qianhui Liang, Xindong Wu, E.K. Park, Taghi M. Khoshgoftaar and Chi-Hung Chi, 2011. Ontology-Based Business Process Customization for Composite Web Services, IEEE Transactions on Systems, Man and Cybernetics, 41(4): 719-729.
10. Dustdar, S. and W. Schreiner, 2005. A survey on web services composition, International Journal Web Grid Services, pp: 1-30.
11. Anshul Goyal and Rajni Mehta, 2012. Performance Comparison of Naïve Bayes and J48 Classification Algorithms, International Journal of Applied Engineering Research, 7(11): 1389-1393.
12. Hong Qing Yu, Stephan Reiff-marganiec, 2008. Non-functional property based service selection: A survey and classification of approaches, Sun SITE Central Europe.
13. Jamous M.M., 2011. Web Services Non-Functional Classification to Enhance Discovery Speed, IJCSI International Journal of Computer Science Issues, 8(5): 150-155.
14. Liu, Y., A.H. Ngu and L.Z. Zeng, 2004. Qos computation and policing in dynamic web service selection. In WWW Alt. '04: Proceedings of the 13th international WorldWide Web conference on Alternate track papers & posters, ACM, pp: 66-73
15. Janicke, H. and M. Solanki, 2007. Policy-driven service discovery. In Proceedings of the 2nd European Young Researchers Workshop on Service Oriented Computing, pp: 56-62.
16. Libsvm tool. http:// www.csie.ntu.edu.tw/ cjlin/libsvm
17. Bruno M. Benevento, G. Canfora, M. Di Penta and R. Scognamiglio, 2005. An Approach to support Web Service Classification and Annotation, IEEE International Conference on e-Technology, e-Commerce and e-Service, pp: 138-143.
18. Bandana Garg, 2013. Design and Development of Naïve Bayes Classifier.
19. Friedman, N., D. Geiger and M. Goldszmidt, 1997. Bayesian Network Classifiers, Machine Learning, 29: 131-163.
20. Langley, P., W. Iba and K. Thompson, 1992. An analysis of Bayesian Classifiers, in Proceedings of the Tenth National Conference on Artificial Intelligence, San Jose, CA.
21. Rish, I., 2001. An empirical study of the naive Bayes classifier, IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, 22: 41-46.
22. Ayodele, T.O., 2010. Types of Machine Learning Algorithms, in New Advances in Machine Learning, Y. Zhang, Ed., InTech, pp; 19-30.
23. Evren Sirin, Bijan Parsia and James Hendler, 2004. Filtering and Selecting Semantic Web Services with Interactive Composition Techniques. IEEE Intelligent Systems, 19(4): 42-49.
24. Brans and J.P. Vincke, 1985. A preference ranking organization method (the PROMETHEE method for multiple criteria decision-making), Management Science, 31(6): 647-656.

25. Hani Abu-Salem, 1992. A microcomputer based Arabic bibliographic information retrieval system with relational thesauri. Ph.D. dissertation, Illinois Institute of Technology, University Microfilm, Ann Arbor, MI.
26. Al-Masri, E. and Q.H. Mahmoud, 2007. Discovering the best web service, (poster) 16th International Conference on World Wide Web (WWW): 1257-1258.