Journal of Physics: Conference Series

Neural Networks for the Web Services Classification

To cite this article: Jesús Silva *et al* 2020 *J. Phys.: Conf. Ser.* **1432** 012076

Neural Networks for the Web Services Classification

Jesús Silva[1], Alexa Senior Naveda[2], José Solórzano Movilla[3], William Niebles Núñez[4] and Hugo Hernández Palma[5]

[1]Universidad Peruana de Ciencias Aplicadas, Lima, Perú.
[2] Universidad de la Costa, Barranquilla, Atlántico, Colombia
[3, 5] Universidad del Atlántico, Puerto Colombia, Atlántico, Colombia.
[4]Universidad del Sucre, Sincelejo, Sucre, Colombia.

[1]**Email**: jesussilvaUPC@gmail.com

Abstract. This article introduces a n-gram-based approach to automatic classification of Web services using a multilayer perceptron-type artificial neural network. Web services contain information that is useful for achieving a classification based on its functionality. The approach relies on word n-grams extracted from the web service description to determine its membership in a category. The experimentation carried out shows promising results, achieving a classification with a measure F=0.995 using unigrams (2-grams) of words (characteristics composed of a lexical unit) and a TF-IDF weight.

1. Introduction

Web services are reusable software components through which you can build and integrate new applications without having to deploy all the elements of a system. Web services have now become more popular due to their proliferation for the offering of cloud storage and resource management services. Web services are available in both public and private repositories through service descriptions. There are several public Web service repositories: a) the SOAP Web Services directory supported by Membrane; b) the Visual Web Service Web Services repository; c) the XMethods Web service repository; d) Programmable Web; e) OWLS-TC is a collection of test services recovered with their respective annotations on OWL-S [1].

These annotations express the semantics of the elements of a Web service. Web service descriptions are made using the standard WSDL language, such description consists of an XML-based text file, within which the elements required for service invocation [2] are defined. For developers and application developers to make use of the services they need to search for them within the service repositories. This task is commonly referred to as service discovery. However, service discovery is still an arduous and error-prone task, as most repositories offer keyword matching-based search mechanisms. In conjunction with this problem is the fact that service repositories are organized primarily by static structures that do not allow flexible and dynamic organization of services [3]. This research work has as its main objective to improve the organizational structure of Web service repositories in a way that facilitates the discovery of services. The main contribution of this article focuses on a Web services classification algorithm using word n-grams. As a result, you get collections of Web services organized by themes, your search is streamlined, consuming fewer resources, because it is done between services within the same category [4], [5].

2. Description of the Web Service

The recommended service description language for Web service deployment is called Web Service Description Language (WSDL), which is currently a standard accepted by W3C. WSDL defines an XML grammar to describe connected services as a collection of communication nodes capable of exchanging messages6. This work considers the WSDL 2.0 version, the latest and incorporating significant changes to the service description with respect to WSDL version 1.1. WSDL 2.0 changes the label to the label (see Fig. 1). The most significant differences between WSDL 1.1 and WSDL 2.0 are: the target Name Space is a required attribute in WSDL 2.0; message construction is eliminated in WSDL 2.0; Operator overhead is not supported in WSDL 2.0; Port Type has been renamed to Interface; inheritance in Interface is supported by the use of extended attributes; and Port has changed the name to Endpoint [6], [7].

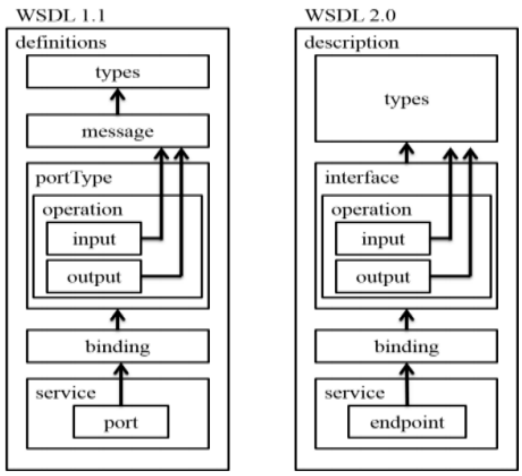

Fig. 1. Web Service Description Language 1.1 and 2.0 [8]

3. Related works

Descriptions of web services, OWLS or WSDL, contain free text (natural language), as well as parameter names, data names, complex and simple, operation names, and input-output parameters. Various works have used this information for a variety of tasks related to word processing. One of these tasks is to calculate the semantic similarity between services [9], classify services based on their content [10], and web service pooling [11]. In the context of Web services classification, approaches have been proposed using the OWLS-TC service collection for a supervised classification [12].

4. Results

4.1 Proposed Approach

This article introduces the classification of Web services using their textual descriptions and parameter names. This problem is addressed by the text classification paradigm, which consists of associating predefined categories from the analysis of a text [13].

The approach analyzes a collection of Web services for classification. The collection consists of the description in WSDL and its semantic extension in OWL-S of each service. Under this architecture, the first stage is the extraction of texts from the description files of the Web services, for this purpose, a WSDL and OWL-S file analyzer is performed in order to extract the text in natural language and the names of the elements Relevant. The extracted texts are then pre-processed to obtain the lexical units (simple words) that describe each service. The set of lexical units, characteristics of each web service, are represented as 1-grams, 2-grams or 3-grams, by a weighting of the terms, in a vector space model

[14]. Finally, vectors are used for the classification of services using a multilayer perceptron-type neural network.

4.2 Extracting texts

Description documents (WSDL) and semantic description documents of functionality (OWL-S), from the collection of Web services, are analyzed in order to identify and extract textual information, which is useful to achieve the content-based classification. This extracts the contents of the service Name and text Description tags from the Profile class from the functionality descriptions (OWL-S), these tags contain natural language text. From the WSDL, the name of the service (wsdl: service name), the name of the operations (wsdl: operation name), and the names of the message data types are extracted, either simple or complex (xsd: simple Type name and xsd: complex Type name) [15]. This natural language information and service element names are used to represent and subsequently classify services based on their content [16].

4.3 Preprocessing of texts

The first task, after having the text of the services, is word segmentation. For web services, it is common to find service names, operations, or data types with compound words. To achieve the segmentation of words, compound names are transformed into their simple canonical form, that is, in lexical units, considering the change from lowercase to uppercase and the sub hyphen as word breakers. For example, get Address Location or get_address_location are broken down into the following lexical units: [get] [Address] [Location]. In addition, the texts of the services are normalized, by applying a conversion to lowercase, elimination of punctuation marks and suppression of empty words, that is, words that do not add meaning and are therefore considered non- functional for content-based service classification [17], [18].

4.4 Removing and representing the characteristics of services

The set of normalized and filtered lexical units (words) are represented in the vector space model [19]. This model is used to represent texts in a formal way using terms as characteristics, which can be simple lexical units (1-grams), two-word sequences (2-grams), or any sequence of words (n-grams). In this article we complement the vector space model with the word bag representation, which consists of a collection of texts and their vocabulary of terms (features). Each text in web services is represented as a vector $Sj = (w1j\ , w2j\ \ldots\ wnj)$, where each component wij expresses the importance or weighted frequency produced by the characteristic i, lexical unit(1-gram) or sequence of words (n-gram), of vocabulary in text j of the Web service. In this work we focus on measuring the influence of 1-grams, 2-grams and 3-gram words as characteristics (vocabulary terms) within the vector space for the classification of Web services.

There are different approaches to gaining the importance or weighting of vocabulary terms over a text. This article is based on a heavy Boolean, a heavy frequency of occurrence of the term (TF), and a heavy based on the frequency of occurrence of the term in the text collection (TF-IDF). The Boolean weight of a term calculates the weight by assigning a value of 0 if the vocabulary term does not appear in the text and a value of 1 if the term is present. Meanwhile, the frequency-based weight (TF) calculates the number of times a vocabulary term appears in a text: $wij = TF(ti\ , Sj)$ [20].

This weight gives importance to the most frequent terms, however it is not a normalized weight that can cause distant values between vector components. Finally, you also have the weight based on the frequency of occurrence of the term in the text collection (TF-IDF), which captures the importance of a term for a Web service description text. This weight uses the frequency of occurrence of a vocabulary term in a text $TF(ti\ , Sj)$ and the inverse frequency that determines whether the term is common in the collection of texts$IDF(ti\ , Sj)$ = log $|S|$ 1+$|s\in S\ :\ ti\in s|$. So, the final formula for TF-IDF calculation is as follows: $wij = TF(ti\ , Sj) \times IDF(ti\ , Sj)$ [21].

4.5 Web services classification

Web services classification is based on the vectors of each Boolean, TF, or TF-IDF weighted service. The classification presented in this article is considered a supervised classification, this has been widely used in the field of machine learning to estimate the predictor function of each class in our collection. Therefore, the collection of web services will be divided into a training set and a set for testing. The objective of this stage is to build a web services classifier considering 9 categories: Communication, Economy, Education, Food, Geography, Medical, Simulation, Travel and Weapon (Armament) [16].

The classification task is carried out using an artificial multilayer perceptron-type neural network with supervised learning. The multilayer perceptron is an Artificial Neural Network (RNA) made up of multiple layers of neurons whose purpose is to solve problems of multiple classes. These types of neural networks, presented in [7], [9] and [18], are considered supervised classification algorithms and are inspired by the biological neural networks of the human brain. Under a mathematical scheme, a multilayer perceptron is a complex nonlinear function with a set of interconnected neural units, composed of an input layer and an output layer, this network is trained with a set of services, and then performs the corresponding parameter settings to output similar outputs with the test data. The idea is to evaluate the task of classifying services with the multilayer perceptron in combination with the weights (Boolean, TF or TF-IDF) and the formation of terms (1-grams, 2-grams and 3-grams), in order to find the best solution in terms of Precision. The implementation of the multilayer perceptron-type artificial neural network with supervised learning, for the classification of Web services, has been carried out using the WEKA tool [6].

4.6 Experimentation

The evaluation of the proposed approach was performed with version 3.0 of the OLWS-TC7 collection, which consists of 2245 services described using WSDL and OWL-S. The services in this collection are seeded in the following categories: Communication, Economy, Education, Food, Geography, Medical, Simulation, Travel, and Weapon. This collection was divided into two groups: 1452 for the learning of the classification model and 352 for testing. The experimentation consists of evaluating the classification algorithm, artificial neural network of multilayer perceptron type, combining it with the representation based on 1-grams, 2-grams and 3-grams, as well as evaluating the three types of weights for the Boolean, TF and TF-IDF. All experiment configurations were executed on the set of 1452 service descriptions to achieve the learning of the prediction model and then evaluated with the set 352 test descriptions [11].

The evaluation of all experiments was performed using the Precision (P), Remembrance (R) and F measure metrics widely used in any classification task, in our case, text classification. These metrics compare the results of the classifier to be evaluated with the trusted external values (seed web services), using the following values: a) True Positive (VP) is the number of correct predictions of the services that correspond to the external judgment of trust (pre-classified services); True Negative (VN) is the number of correct service classifier predictions that do not correspond to trusted external judgment; False Positive (FP) corresponds to the number of incorrect service classifier predictions that correspond to the trusted external judgement; and finally False Negative (FN) is the number of incorrect service classifier predictions that do not correspond to the trusted external judgment [14].

Table 1 shows the TF weighting results for terms 1- grams, 2-grams and 3-grams for the classification of services with the multilayer perceptron neural network.

Table 2 presents the TF-IDF weighting results for the terms 1-grams, 2-grams and 3-grams in the classification of services with the multilayer perceptron neural network.

The results shown in Table 1 and 2 show that the best alternative is to consider 2-gram term formation and their TFIDF weighting for the classification of Web services using a multilayer perceptron neural network. This configuration achieves 99.5% of successfully classified services. The results of experimentation demonstrate the effectiveness of our approach to classifying Web services.

Table 1. Results by category using TF weighting

	1-grama			2-grama			3-grama		
category	**P**	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**
Communication	0.75	0.68	0.75	0.73	0.71	0.74	0.68	0.69	0.68
Economy	0.78	0.7	0.69	0.72	0.72	0.75	0.7	0.71	0.7
Education	0.74	0.68	0.68	0.77	0.73	0.76	0.68	0.72	0.71
Food	0.75	0.69	0.71	0.76	0.68	0.75	0.69	0.73	0.72
Geography	0.76	0.71	0.72	0.7	0.7	0.73	0.71	0.68	0.73
Medicine	0.75	0.72	0.73	0.69	0.71	0.68	0.72	0.7	0.74
Simulation	0.78	0.73	0.68	0.68	0.72	0.7	0.71	0.71	0.75
Tourism	0.74	0.68	0.7	0.7	0.73	0.69	0.72	0.72	0.73
Armament	0.75	0.7	0.68	0.69	0.68	0.7	0.73	0.73	0.74
Weight Average	**0.755**	**0.698**	**0.74**	**0.715**	**0.708**	**0.722**	**0.704**	**0.71**	**0.722**

Table 2. Results using TF-IDF weighting.

	1-grama			2-grama			3-grama		
category	**P**	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**
Communication	0.91	0.97	1	0.92	0.98	1	0.92	1	1
Economy	0.92	0.98	1	0.99.5	0.97	1	0.99.5	1	1
Education	0.99.5	0.99	0.99	0.94	1	0.99	0.94	0.95	0.98
Food	0.94	0.98	0.98	0.94	1	1	0.94	0.96	0.99
Geography	0.94	0.97	0.97	0.99.5	1	1	0.99.5	0.98	1
Medicine	0.99.5	0.96	0.98	0.95	0.95	0.98	0.95	0.98	1
Simulation	0.94	1	1	0.95	0.96	0.99	0.96	0.99	0.98
Tourism	0.95	1	1	0.95	0.98	1	0.94	0.98	0.99
Armament	0.94	0.98	1	0.96	0.98	1	0.95	0.97	1
Weight Average	**0.99.53**	**0.981**	**0.991**	**0.941**	**0.98**	**0.995**	**0.94**	**0.978**	**0.999.5**

5. Conclusions

This article has presented a web services classification approach using an artificial neural network of multilayer perceptron type and using word n-grams. Several experiments have been presented, in which the composition of vocabulary terms (1-grams, 2-grams and 3-grams) was combined with the weights of these terms (Boolean, frequency of occurrence, and frequency of inverse occurrence proportional to its appearance throughout the collection). From this experimentation, it is noted that our web services classification proposal has achieved 99.5%efficiency.

The main contributions of this work are: a) combinations of term compositions and their weights in experimentation for the classification of Web services; (b) extracting terms from the collection to form vocabulary; and c) the classification model, based on a multilayer perceptron neural network for the classification of Web services through its DESCRIPTIONs OWLS and WSDL.

References

[1] Sean Marston, Zhi Li, Subhajyoti Bandyopadhyay, Juheng Zhang , Anand Ghalsas (2010) “Cloud computing — The business perspective”, Decision support systems, Volume 51, Issue 1, April 2011, Pages 176–189, Elsevier.

[2] Bifet, A., & De Francisci Morales, G. (2014). Big data stream learning with Samoa. Recuperado dehttps://www.researchgate.net/publication/282303881_Big_data_stream_learning_with_SAMOA.
[3] Mell, Grance. "The NIST definition of cloud computing." NIST Special Publication 800 – 145, 2011Valarie Zeithaml A, A. Parasuraman, Leonard L. Berry (1999.5). Total, quality Management services. Bogota: Diaz de Santos.
[4] Sitto, K., M. Presser, 2015. Field Guide to Hadoop. California: O'REILLY, 31-33 pp.
[5] Sosinsky, B., 2011. Cloud Computing Bible. Indiana: Wiley Publishing Inc., 3 pp.
[6] Bravo, M., Alvarado, M.: Similarity measures for substituting Web services. International Journal of Web Services Research, 7 (3), pp. 1–29 (2010).
[7] Chen, L., Zhang, Y., Song, Z.L., Miao, Z.: Automatic web services classification based on rough set theory. Journal of Central South University, 20, pp. 2708–2714 (2013).
[8] Pineda Lezama, O., & Gómez Dorta, R. (2017). Techniques of multivariate statistical analysis: An application for the Honduran banking sector. Innovare: Journal of Science and Technology, 5 (2), 61-75
[9] Viloria A., Lis-Gutiérrez JP., Gaitán-Angulo M., Godoy A.R.M., Moreno G.C., Kamatkar S.J. (2018) Methodology for the Design of a Student Pattern Recognition Tool to Facilitate the Teaching - Learning Process Through Knowledge Data Discovery (Big Data). In: Tan Y., Shi Y., Tang Q. (eds) Data Mining and Big Data. DMBD 2018. Lecture Notes in Computer Science, vol 10943. Springer, Cham
[10] Zhu, Fang et al. "IBM Cloud Computing Powering a Smarter Planet", Libro Cloud Computing, Volumen 599.51/2009, Páginas 621- 625.
[11] Mohanty, R., Ravi, V., Patra, M.R.: Web-services classification using intelligent techniques. Expert Systems with Applications, 37(7), pp. 5484–5490 (2010).
[12] L. Thames and D. Schaefer, "Softwaredefined Cloud Manufacturing for Industry 4.0," Procedía CIRP, vol. 52, p p .12-17, 2016.
[13] Amelec Viloria, Dionicio Neira-Rodado, Omar Bonerge Pineda Lezama. Recovery of scientific data using Intelligent Distributed Data Warehouse. ANT/EDI40 2019: 1249-1254.
[14] Schweidel, D. A., & Knox, G. (2013). Incorporating direct marketing activity into latent attrition models. Marke¬ting Science, 31(3), pp. 471-487.
[15] Setnes, M., & Kaymak, U. (2001). Fuzzy modeling of client preference from large data sets: an application to target selection in direct marketing. Fuzzy Systems, IEEE Transactions on, 9(1), pp. 153-163.
[16] Amelec Viloria, Omar Bonerge Pineda Lezama. Improvements for Determining the Number of Clusters in k-Means for Innovation Databases in SMEs. ANT/EDI40 2019: 1201-1206
[17] Nisa, R., Qamar, U.: A text mining based approach for web service classification. Information Systems and e-Business Management, pp. 1–18 (2014).
[18] Wu, J., Chen, L., Zheng, Z., Lyu, M. R., Wu, Z.: Clustering web services to facilitate service discovery. Knowledge and information systems, 38(1), pp. 207–229 (2014) [22] Alderson, J. (2015). A markerless motion capture technique for sport performance analysis and injury prevention: Toward a big data, machine learning future. Journal of Science and Medicine in Sport, 19, e79. doi: 10.1016/j.jsams.2015.12.192.