Forecasting the Needs of Users and Systems

A New Approach to Web Service Mining

Juan I. Guerrero, Enrique Personal, Antonio Parejo, Antonio García and Carlos León
Department of Electronic Technology
University of Seville
Seville, Spain
email: juaguealo@us.es

***Abstract*—The Smart Grids provides a great scope of new technologies. The new technologies include the integration of heterogeneous systems which perform different task in the Smart Grid ecosystems. These new systems have their own users with different knowledge levels. It is important that the integrated system can assimilate the new systems without a complex implementations and deployments. Additionally, the users of these systems will increase their needs when new systems are integrated. The proposed framework forecasts the needs of new systems make available new Web Services (WSs) to users and systems. The standards related with Smart Grids include specifications for different Web Service Architectures. Thus, the proposed framework is based on Web Service Mining, using different modules, including semantic engine and analytic module, with time series classification and clustering. Finally, the proposed framework was applied in Smart Business Park (SBP), creating new WSs based on the analysis of Web Services activity information, forecasting the needs of users or/and systems.**

Keywords-web service mining; ant colony optimization; smart grids; computational intelligence.

I. Introduction

There are a lot of examples of technologies that makes our lives more comfortable, since robots with artificial intelligence to make different things (cooking, management, cleaning, etc.), to information technologies to optimize our task and reduce the time. Time is the only thing that you waste and never recover. Additionally, time is a very important asset for any company. Thus, all technologies that let you optimize any task are very important. There are several examples in different economic sectors. In health and e-health sectors, the new technologies related with data mining [1], biomedical imaging and image processing [2] have been reduced the diagnostic time and the quality of health services. In the power distribution sector, the emergent technologies related with Smart Grids (SGs) have provided new functionalities and services for consumers, reducing the restoring power supply time [3] and increasing the supply quality [4]. In information technology sector, the new technologies related with big data [5][6] and high performance computing [7] have improved the capabilities to store and analyze great volumes of information (applied in health, environmental control and monitoring, finance, telecommunication, and other utilities). In particular, Google [8] reduced the searching time using different techniques to rank the results of search requests. Sometimes, these systems are designed to recommend information [9] that could be interesting. This means the systems try to forecast the needs of the users in order to save time and increase the Quality of Service (QoS). Sometimes, some of the previously mentioned technologies could go unnoticed by the final users. There are other technologies as the previously mentioned ones: Web Service Mining.

The Web Service Mining [10] is an emergent technology that deals with the WS in order to discover, check, and improve service behavior, based on service composition, service pattern discovery, managing service registry-repositories, etc. There are several solutions based on process mining [11], pattern usage discovery [12], hypergraph-based matrix representation with a service set mining algorithm [13], constraint satisfaction [14], semantics-based methods [15][16], customer value analysis [17], frequent composite algorithm [18], Heterogeneous Feature Selection [19], etc. for cross-cloud environment [20], RESTful Web Services [21], etc.

This paper proposes a novel web service mining approach based on computational intelligence and data mining framework. This framework creates new WSs based on the usage of existing WSs, in order to forecast functionalities for users and systems. This framework is been researched for a Smart Business Parks (SBPs). The current version has been tested with several systems related to SGs Ecosystem. Thus, the SG ecosystem is formed by several systems to allow the distributed management of power energy, for example, energy management systems, distributed energy resource management, etc. Additionally, the proposed framework could integrate the WSs of the additional system, which may be added to the ecosystem. In this sense, the proposed framework tries to learn the WS usage pattern made by the systems and users, providing the capability to forecast the needs of systems or users creating new WSs that can be useful for them.

The proposed framework is formed by several modules which are described in this paper, each module implies different techniques. In Section II, the overview of proposed framework is included. In Section III, the description of each module is provided. In Section IV, the experimental results

Copyright (c) IARIA, 2016. ISBN: 978-1-61208-518-0

are explained. Finally, in Section V, the conclusions and future works are described.

II. FRAMEWORK OVERVIEW

The proposed framework contains several modules, as shown in Fig. 1. Additionally, the architecture of the proposed framework applied in a SBP architecture is shown in Fig. 2.

In this framework, there are several modules to add information to the database of the system. Monitoring Module stored all activity of WSs. Discovery and Check & Test modules are based on Ant Colony Optimization (ACO) Algorithms. Discovery Module performs an ACO in order to discovery new services in the ecosystem. Check & Test Module perform different tests in order to check the new or existing services. The Hybrid Web Service Engine (HWSE) provides to the previously described modules the possibility to interact with different type of services: REST and SOAP.

The Web Service Database (WSDB) stores information about WSs registered in the ecosystem, adding metadata and analytic information, using the Analytic Module and Semantic Engine.

The Analytic Module is based on time series. This module performs an analysis of all available information. The objectives of this module are: establish the importance of service, order the service by importance level and identify the WS groups. Additionally, the module establishes the time usage pattern of each WS and group.

The Web Service Engine (WS Engine or WSE) creates the new WSs by composition of WS groups, registering in Web Service Hybrid Repository (WS Hybrid Repository or WSHR).

The WSHR stores the information of each service in different WS standards.

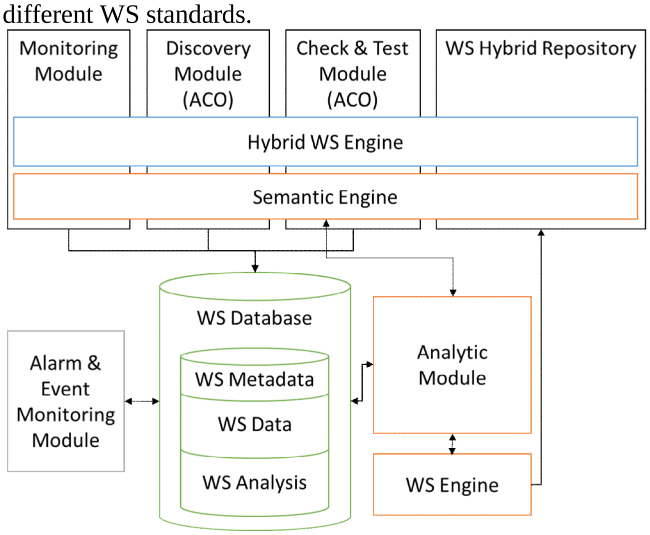

Figure 1. Framework overview

The SBP ecosystem has several systems: Public Lighting Management System, Smart building Management System, KPI Monitoring System, Photovoltaic Management System, and other systems. The ecosystem includes a layer for system integration. This layer is the middleware or gateway to access to Intelligent Electronic Devices (IEDs), which is based on several standards like IEC 61850 and other standards related with SG Standard European Roadmap [22].

In this sense, the Alarm and Event Monitoring Module registers the information of events and alarms generated in the system, at high level in the application layer, and in low level in the IED layer.

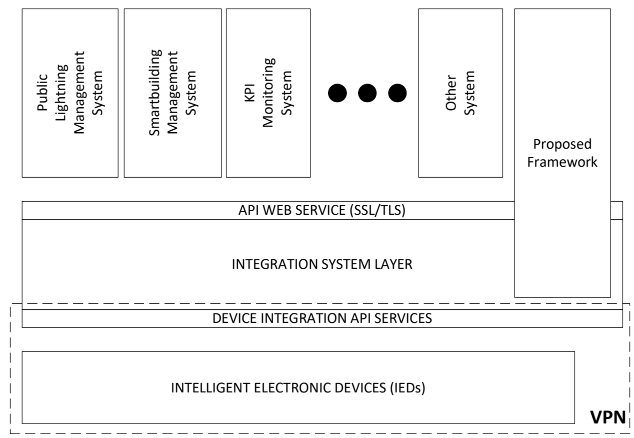

Figure 2. Proposed Framework in SBP Ecosystem

One of the most important requirements of this SBP ecosystem is the capability to assimilate new systems that complains with the established standards. Thus, the proposed framework is an intelligent framework to integrate the services of new systems in the SBP ecosystem.

III. DESCRIPTION OF MODULES

The proposed framework is composed of several modules (see Fig. 1), each of which are responsible of following tasks:

A. *Hybrid WS Engine (HWSE)*

The Hybrid WS Engine is a parser of WSs. This module can extract information from messages in different WS standards: SOAP and REST. The information extracted includes data and metadata.

B. *Monitoring Module*

The Monitoring Module gathers the information of all WS usage, using the HWSE in order to parse the messages. This module is always monitoring and registering the message traffic. The module stores the information of WSs (data and metadata) in the WSDB. Additionally, the module stores information about timestamps, response periods, preconditions, information to calculate frequencies (usage, requester, and provider), results, effects, average size, etc. This information is gathered for Analytic Module.

C. *Discovery Module*

The Discovery Module takes advantage from different WS technologies: Universal Description Discovery and Integration (UDDI), Web-Service discovery (WS-Discovery), electronic business using eXtensible Markup

Language (ebXML), Domain Name System (DNS), etc. Additionally, the Discovery Module uses the information stored in WSDB to improve the WS discovery. This information is loaded in an ACO algorithm. The ACO algorithm has two main objectives: discover the WSs which are not accessible to traditional WS discovery technologies; and extracting the functional information of WS.

D. Check & Test Module

The check and test module performs an ACO in order to optimize the created or composed WSs. The non-functional attributes of WS limit the usage of this module. The security level and the critical nature of WS could avoid the WS check or test.

This module has two main objectives:

- Check and test the new WS registered in WSHR.
- Check and test the existing WS registered in other repositories.

This module support SOAP and REST.

E. Alarm & Event Monitoring Module

This module was specially designed for SGs ecosystems. The module registers and stores information about any event, warning, or alarm fired in any system or in any device of SG. Alarms, warnings, and events are considered as information entities. These entities are one of the most important sources to improve the web service mining. Entities are usually related with the WS usage in a SG ecosystem. An entity has usually associated several WS usage patterns. The end of pattern is often marked by another entity. Thus, this module stores in WSDB all information about these entities: timestamp, source, etc. This information is gathered for Analytic Module.

F. Analytic Module

This Analytic Module applies several data mining techniques to provide additional information about the WSs. This module was implemented in R. This new information is stored in WSDB. The objectives of this module are:

- Composite services according to WS semantic analysis. This analysis is based on the results of Semantic Engine combined with the results of time series classification.
- Analyze services according to the relation between WS invocations, alarms, warnings, and events (from Alarm & Event Module). This analysis is based on time series clustering.

Both objectives are based on time series. Thus, the first step is to translate the data in the same temporal scale. The time series extract and build features from this information. According to the features of the time series and the information stored in WSDB:

- The time series classification applies Support Vector Machine (SVM) and decision tree, in order to aggregate or compose all the WS with similar pattern behavior.
- The time series clustering starts selecting the appropriate distance/similarity metric based on time series features and the results of semantic engine. Then this module applies several clustering techniques: K-means, hierarchical clustering, and density-based clustering, in order to get the best cluster.

This module calculates several Key Performance Indicators (KPI). This KPIs measures the performance and the QoS [23]: throughput, availability, response time, interoperability, accessibility, and cost.

G. Semantic Engine

The Semantic Engine provides additional information about WSs. This engine is based on Ontology Web Language – Services (OWL-S) that is a semantic markup for WS. Additionally, Semantic Web Rule Language (SWRL) is used to represent rules.

This module has three main objectives:

- Identification of the service profile, describing the signature of the service in terms of its input, output, parameters, preconditions, service name, service type (stateless, state-based, etc.), provider, business domain, etc.
- Description of the service process model, describing how the service works in terms of the interplay between data and control flow.
- Description of the non-functional service semantics. In this case, the non-functional attributes are related to a QoS model: availability, delivery constraints, etc. Additionally, the semantic engine identifies: the security level, the critical nature of service, the level of service (software, middleware, or hardware), etc.

H. WS Engine

This module gathers the information from the analytic module in order to create or composite the new WS. These new WSs are registered in WSHR, in order to start the test & check stage.

I. WS Database (WSDB)

The WSDB stores WSs data and metadata. Additionally, the WSDB registers all information about the data and metadata analysis of WSs.

The data model is based on Common Information Model (CIM) from Distributed Management Task Force (DMTF). Nevertheless, some extensions and modifications have been applied in order to store analytic information and speeding up the information exchange.

J. WS Hybrid Repository (WSHR)

This module implements several repositories based on ebXML, UDDI, and DNS. The repository stores information of all composed and created WS. Additionally, there is a table in WSDB, which represents the stage of new WSs. There are several stages in the test process: NEW, CHECK, TEST, VALID, NON-VALID, BROADCAST, and USING.

Copyright (c) IARIA, 2016. ISBN: 978-1-61208-518-0

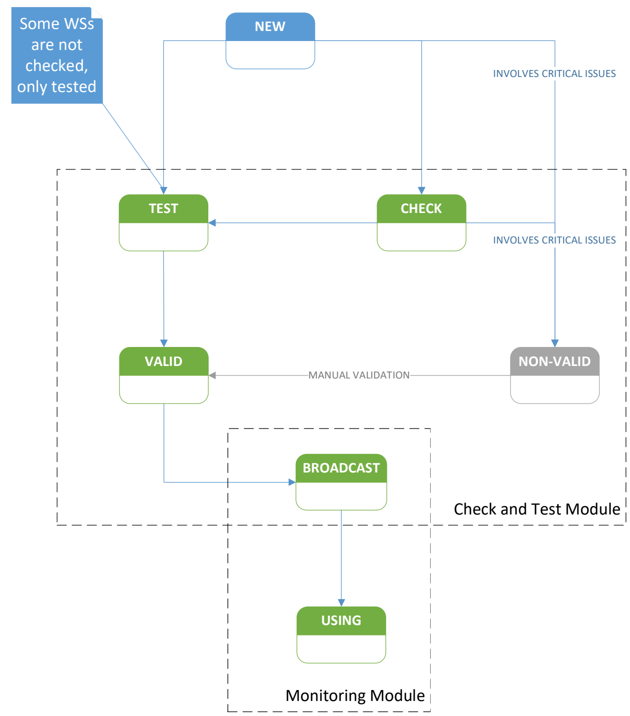

Figure 3. WSs transition status diagram

When the new WS is created the state is NEW. Then the Check and Test Module and the Monitoring Module change the status according to the results of different transitions showed in Fig. 3. Check and Test Module manage the transitions between stages: TEST, CHECK VALID, NON-VALID and BROADCAST. Monitoring Module manages the transitions between stages: BROADCAST and USING. If the WS involves critical services or devices, it will be validated and tested by an expert.

IV. Experimental Results

The first prototype of the proposed framework was developed as part of SBP. The main objective of the proposed framework was to detect the WS usage patterns to create new WS, which makes the system reliable and robust. The second objective was to make easy the integration of new systems in SBP.

The SBP ecosystem has several systems (see Fig. 2). Each system provides its own set of WSs, although the systems have several WSs in common:

- Authentication.
- General services and metadata access.
- Monitoring services.

These systems include services to interact with IEDs, except the KPI monitoring system, which interacts with other systems to extract information for KPIs.

The Monitoring Module of the proposed framework was the first module of the implementation. This module was registering information during two months. The information is stored in a data base according to Distributed Management Task Force (DMTF) Common Information Model (CIM) Standard. This information is applied to test the Analytic Module. Additionally, this information was used to check and test the Semantic Engine and the HWSE.

In case of Analytic Module, this module composed several WSs:

- WSs related with KPI Monitoring System. System were associated with renewable energy resources. The KPI Monitoring System gathered information from different type of renewable energy resources, in order to summarize all consumptions. The system found several cases: photovoltaic farms, wind farms, battery, and electric vehicles. The new WSs reduced time access to information and increased the efficiency.
- Some patterns related with alarms generated by some IEDs in substations. The new WSs involved critical devices. Critical devices can take effect over the continuity of supply. The management of these devices involves important protocols and methodologies designed by companies. The created Adapters for new WSs were successfully evaluated by companies.

However, the systems in the SBP needed an adapter to take advantage from the new WSs created by the proposed framework. The created WSs were manually checked and tested successfully.

V. Conclusion and Future Works

The proposed framework can compose the WSs according to the discovery usage patterns and the relation of WS to alarms, warnings and events on the SBP ecosystem. The lack of adapters to take advantage of these new WSs makes difficult the automatic integration of new systems. Thus, the proposed framework can create composed WS which offers the aggregated functionality of several WSs, reducing the communications and speeding up the final effects. In this way, the system forecasts the needs of users and systems based on the usage data and metadata patterns.

Additionally, the future works to improve the proposed framework are:

- Design and implementation of adapters which takes advantage of the new WS.
- Application of new techniques in the Analytic Module based on multivariable inference.
- Research of new module to composite WS from different WS standards (SOAP, REST, etc.)

Acknowledgment

The authors would like to thank the Smart Business Project (SBP) (Reference Number: P011-13/E24), which provided data sources. Additionally, the authors would like to thank the IDEA Agency for providing the funds for the project.

The authors are also appreciative of the backing of the SIIAM project (Reference Number: TEC2013-40767-R), which is funded by the Ministry of Economy and Competitiveness of Spain.

References

[1] H. Yang and Y.-P. P. Chen, “Data mining in lung cancer pathologic staging diagnosis: Correlation between clinical and pathology information,” *Expert Syst. Appl.*, vol. 42, no. 15–16, pp. 6168–6176, pp. 6168–6176, Sep. 2015.

[2] M. O. Visscher, S. A. Burkes, R. Randall Wickett, and K. P. Eaton, “Chapter 38 - From Image to Information: Image Processing in Dermatology and Cutaneous Biology A2 - Hamblin, Michael R.,” in *Imaging in Dermatology*, P. Avci and G. K. Gupta, Eds. Boston: Academic Press, pp. 519–535, 2016.

[3] F. Mekic, Z. Wang, V. Donde, F. Yang, and J. Stoupis, “Disributed automation for back-feed network power restoration,” in *20th International Conference and Exhibition on Electricity Distribution - Part 1, 2009. CIRED 2009*, 2009.

[4] G. Blajszczak, P. Antos, and M. Wasiluk-Hassa, “Smart grid accommodation of distributed generation for more reliable quality of supply,” in *2011 11th International Conference on Electrical Power Quality and Utilisation (EPQU)*, 2011.

[5] Y. Demchenko, C. de Laat, and P. Membrey, “Defining architecture components of the Big Data Ecosystem,” in *2014 International Conference on Collaboration Technologies and Systems (CTS)*, 2014.

[6] C. Wu, R. Buyya, and K. Ramamohanarao, “Chapter 1 - Big Data Analytics = Machine Learning + Cloud Computing,” in *Big Data*, Morgan Kaufmann, pp. 3–38, 2016.

[7] A. Ali and K. S. Syed, “Chapter 3 - An Outlook of High Performance Computing Infrastructures for Scientific Computing,” in *Advances in Computers*, vol. 91, A. Memon, Ed. Elsevier, pp. 87–118, 2013.

[8] X. Wang, M. Bendersky, D. Metzler, and M. Najork, “Learning to Rank with Selection Bias in Personal Search,” *Proc. SIGIR 2016 ACM*, 2016.

[9] H.-T. Cheng et al., “Wide & Deep Learning for Recommender Systems,” *ArXiv160607792 Cs Stat*, Jun. 2016.

[10] G. Zheng and A. Bouguettaya, *Web Service Mining: Application to Discoveries of Biological Pathways*. Boston, MA: Springer Science+Business Media, LLC, 2010.

[11] W. v d Aalst, “Service Mining: Using Process Mining to Discover, Check, and Improve Service Behavior,” *IEEE Trans. Serv. Comput.*, vol. 6, no. 4, pp. 525–535, pp. 525–535, Oct. 2013.

[12] Q. A. Liang, J. y Chung, S. Miller, and Y. Ouyang, “Service Pattern Discovery of Web Service Mining in Web Service Registry-Repository,” in *2006 IEEE International Conference on e-Business Engineering (ICEBE’06)*, 2006.

[13] A. Zhao, X. Wang, K. Ren, and Y. Qiu, “Semantic Message Link Based Service Set Mining for Service Composition,” in *Fifth International Conference on Semantics, Knowledge and Grid, 2009. SKG 2009*, 2009.

[14] Q. A. Liang, S. Miller, and J. Y. Chung, “Service mining for Web service composition,” in *IRI -2005 IEEE International Conference on Information Reuse and Integration, Conf, 2005.*, 2005.

[15] H. Luo, L. Liu, and Y. Sun, “Semantics-Based Service Mining Method in Wireless Sensor Networks,” in *2011 Seventh International Conference on Mobile Ad-hoc and Sensor Networks (MSN)*, 2011.

[16] A. Zhao, K. Ren, X. Wang, and Y. Qiu, “An approach to service set mining for improving SWS based supply chains coordination,” in *2010 International Conference on Logistics Systems and Intelligent Management*, vol. 3, 2010.

[17] W. Hu and Y. Hui, “Service-mining Based on Customer Value Analysis,” in *2007 International Conference on Management Science and Engineering*, 2007.

[18] H. Meng, L. Wu, T. Zhang, G. Chen, and D. Li, “Mining Frequent Composite Service Patterns,” in *2008 Seventh International Conference on Grid and Cooperative Computing*, 2008.

[19] L. Chen, Q. Yu, P. S. Yu, and J. Wu, “WS-HFS: A Heterogeneous Feature Selection Framework for Web Services Mining,” in *2015 IEEE International Conference on Web Services (ICWS)*, 2015.

[20] T. Wu, W. Dou, C. Hu, and J. Chen, “Service Mining for Trusted Service Composition in Cross-Cloud Environment,” *IEEE Syst. J.*, vol. PP, no. 99, pp. 1–12, pp. 1–12, 2014.

[21] A. Stroinski, D. Dwornikowski, and J. Brzezinski, “RESTful Web Service Mining: Simple Algorithm Supporting Resource-Oriented Systems,” in *2014 IEEE International Conference on Web Services (ICWS)*, 2014.

[22] “SGCG/M490/G_Smart Grid Set of Standards. Version 3.1,” CEN-CENELEC-ETSI Smart Grid Coordination Group, SGCG/M490/G-version 3.1, Oct. 2014.

[23] L. K. Goel, S. K. Kottayil, M. Suchithra, and M. Ramakrishnan, “Efficient Discovery and Ranking of Web Services Using Non-functional QoS Requirements for Smart Grid Applications,” *Procedia Technol.*, vol. 21, pp. 82–87, pp. 82–87, Jan. 2015.