

# Augmenting Labeled Probabilistic Topic Model for Web Service Classification

Shengye Pang, School of Computer Engineering and Science, Shanghai University, Shanghai, China

Guobing Zou, School of Computer Engineering and Science, Shanghai University, Shanghai, China

Yanglan Gan, School of Computer Science and Technology, Donghua University, Shanghai, China

Sen Niu, School of Computer and Information Engineering, Shanghai Polytechnic University, Shanghai, China

Bofeng Zhang, School of Computer Engineering and Science, Shanghai University, Shanghai, China

## ABSTRACT

Web service classification has become an urgent demand on service-oriented applications. Most existing classification algorithms mainly rely on the original service descriptions. That leads to low classification accuracy, since it cannot fully reflect the semantic feature specific to a service category. To solve the issue, this article proposes a novel approach for web service classification, including service topic feature extraction, service functionality augmentation, and service classification model learning. The characteristic is that the original service descriptions can be semantically augmented, which is fed to deriving a service classifier via labeled probabilistic topic model. A benefit from this approach is that it can be applied to an online service management platform, where it assists service providers to facilitate the registration process. Extensive experiments have been conducted on a large-scale real-world data set crawled from ProgrammableWeb. The results demonstrate that it outperforms state-of-the-art methods in terms of service classification accuracy and convergence speed.

## KEYWORDS

*Labeled LDA, Service Classification, Service Feature Extraction, Service Functionality Augmentation, Web Service*

## 1. INTRODUCTION

Due to the fast advancement of Web 2.0 technologies and service-oriented computing, more and more service providers publish their services on the internet mainly in the form of web APIs. They can be more easily organized and manipulated in a loosely coupled style for creating service mashups to fulfill comprehensive functional requirements and offer value-added integrated software systems with complex business processes. As the rapid increase in the number and diversity of web services, it accelerates the interoperable machine-to-machine interaction and greatly promotes the procedure of service discovery, optimum selection, automatic composition and recommendation (Xia, Luo, Li, & Zhu, 2013; Xia, Liu, Liu, & Zhu, 2012; Li, Luo, Xia, Han, & Zhu, 2015). However, with the boom of overwhelming number of functional characteristics of the published web services, there are

DOI: 10.4018/IJWSR.2019010105

always hundreds of categories in an online RESTful service repository. As a result, it tends to be a labor-intensive challenging task for service providers to search and find an appropriate category from diverse registered ones, when publishing their API services on a service management platform. For example, ProgrammableWeb.com, which is the largest online RESTful service repository (APIs and mashups), collects over 19,000 APIs and 7000 mashups with more than 400 diverse categories on the web. In addition to providing basic registration information when service providers register their API services on ProgrammableWeb, it needs to further manually choose at least one desired category from more than 400 categories so that it can match corresponding service functional description. Therefore, how to design an effective approach that can classify web services and recommend an accurate category has become a critical research issue to be addressed (Ames & Naaman, 2007).

In recent years, correlative research efforts have been posed on web service classification (Tsoumakas, Katakis, & Taniar, 2008). These existing approaches achieve the goal of web service classification and service tag recommendation by training traditional supervised learning model (e.g., SVM) (Lopez & Maldonado, 2016; Wang, Shy, Zhou, & Bouguettaya, 2010), active learning-based supervised learning model (Tong & Koller, 2001; Liu, Agarwal, Ding, & Yu, 2016; Shi, Liu, & Yu, 2017), or a comprehensive supervised learning model where unlabeled probabilistic topic model (e.g., LDA) (Krestel, Fankauer, & Nejdli, 2009) has been applied to extract semantic feature of web services. Some of the works generally learn a classification model under an existing labeled service repository, while active learning method was taken into account for boosting the learned service classifier, where the most informative services are intellectually selected at each iteration and manually labeled with human efforts to enrich the quality of small scale training data. Although they take advantage of the existing service repository as training data to derive a service classifier which can be easily deployed and applied, it is still unsatisfactory for service providers' demands on high accuracy of web service classification.

The essential reason is that existing approaches have deficiencies on their effectiveness and efficiency. More specially, the disadvantages of current paradigm for web service classification are twofold. (1) On one hand, they mainly rely on the original service descriptions for learning a service classifier, where each functional description of a RESTful web service only consists of a bunch of short text (e.g., 10 to 20 words), failing to be fully understood on its corresponding category. Furthermore, it is observed that some words are frequently repeated with high occurrence across different service descriptions, which obviously disturbs the purity of differentiating its category of web services. Therefore, it is crucially harmful to affect the classification accuracy. (2) On the other hand, most existing approaches leverage traditional classification algorithm (e.g., SVM), where multiple basic models need to be trained as a whole to perform web service classification, because each of them is a dichotomous classifier that cannot directly solve a multi-class problem. As a result, they accomplish the task with high complexity both on huge space consumption and slow convergence speed, when training a service classifier on a large-scale web service repository.

To handle above two research challenges, this article proposes a novel framework for effective and efficient web service classification. The authors mainly focus on solving two issues: (1) How to enrich short text description of web services towards more accurate correlation with its category? (2) How to implement service classification efficiently? To address the first issue, the authors first eliminate those words (Del-List) from original service descriptions, since they cannot make differences among different service categories. Then, the authors derive keywords (Key-List) that can affect the capability of service topic recognition. Finally, the authors propose an expansion strategy to enrich service functionality description of web services. To address the second issue, the authors train augmented service descriptions by using labeled probabilistic topic model L-LDA which mines the implicit semantic relationships throughout service repository. The authors assign all the services categories as feature topics when training an L-LDA model to generate probabilistic topic distribution. That greatly enhances the accuracy and efficiency of service classification work. The main contributions of the work is as follows:

19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:

[www.igi-global.com/article/augmenting-labeled-probabilistic-topic-model-for-web-service-classification/220391?camid=4v1](http://www.igi-global.com/article/augmenting-labeled-probabilistic-topic-model-for-web-service-classification/220391?camid=4v1)

This title is available in InfoSci-Journals, InfoSci-Journal Disciplines Business, Administration, and Management, InfoSci-Journal Disciplines Computer Science, Security, and Information Technology, InfoSci-Networking, Mobile Applications, and Web Technologies eJournal Collection, InfoSci-Digital Marketing, E-Business, and E-Services eJournal Collection. Recommend this product to your

librarian:

[www.igi-global.com/e-resources/library-recommendation/?id=2](http://www.igi-global.com/e-resources/library-recommendation/?id=2)

## Related Content

---

### Semantic Search Engines Based on Data Integration Systems

D. Beneventano (2007). *Semantic Web Services: Theory, Tools and Applications* (pp. 317-342).

[www.igi-global.com/chapter/semantic-search-engines-based-data/28889?camid=4v1a](http://www.igi-global.com/chapter/semantic-search-engines-based-data/28889?camid=4v1a)

### Construction of Traceability System by Using Simple and Handy Type RFID Reader

Rei Itsuki (2011). *E-Activity and Intelligent Web Construction: Effects of Social Design* (pp. 35-46).

[www.igi-global.com/chapter/construction-traceability-system-using-simple/53272?camid=4v1a](http://www.igi-global.com/chapter/construction-traceability-system-using-simple/53272?camid=4v1a)

## Parameter Tuning for S-ABCPK: An Improved Service Composition Algorithm Considering Prior Knowledge

Ruilin Liu, Zhongjie Wang and Xiaofei Xu (2019). *International Journal of Web Services Research* (pp. 88-109).

[www.igi-global.com/article/parameter-tuning-for-s-abcpk/226291?camid=4v1a](http://www.igi-global.com/article/parameter-tuning-for-s-abcpk/226291?camid=4v1a)

## An E-negotiation Agent for an E-tourism Platform

Fayçal M'hamed Bouyakoub, Abdelkader Belkhir, Amina Belkacemnacer and Sara Harfouche (2019). *International Journal of Web Services Research* (pp. 65-87).

[www.igi-global.com/article/an-e-negotiation-agent-for-an-e-tourism-platform/226290?camid=4v1a](http://www.igi-global.com/article/an-e-negotiation-agent-for-an-e-tourism-platform/226290?camid=4v1a)