Topography and Topic-Aware Service Clustering

Weifeng Pan, Zhejiang Gongshang University, Hangzhou, China
Jilei Dong, University of Connecticut, Storrs, USA
Kun Liu, Hubei University of Economics, Wuhan, China
Jing Wang, Jiangxi University of Finance and Economics, Nanchang, China

ABSTRACT

This article describes how the number of services and their types being so numerous makes accurately discovering desired services become a problem. Service clustering is an effective way to facilitate service discovery. However, the existing approaches are usually designed for a single type of service documents, neglecting to fully use the topic and topological information in service profiles and usage histories. To avoid these limitations, this article presents a novel service clustering approach. It adopts a bipartite network to describe the topological structure of service usage histories and uses a SimRank algorithm to measure the topological similarity of services; It applies Latent Dirichlet Allocation to extract topics from service profiles and further quantifies the topic similarity of services; It quantifies the similarity of services by integrating topological and topic similarities; It uses the Chameleon clustering algorithm to cluster the services. The empirical evaluation on real-world data set highlights the benefits provided by the combination of topological and topic similarities.

KEYWORDS

Service Clustering, Service Network, Service Usage History, SimRank, Topic Model

INTRODUCTION

Service-Oriented Computing (SOC) is a new computing paradigm that provides a distributed computing infrastructure for both intra- and cross-enterprise application integration and collaboration (Papazoglou, 2003). It has become one of the hottest topics in recent years, receiving much attention from research community and industry. SOC advocates building distributed applications by means of the composition of services, which greatly changes the way that software applications are designed, architected, delivered and consumed (Zhang & Zhang, 2013). Software is now exposed to an open and collaborative environment constituted by services. With the prevalence of SOC, the number of services and their types published on the Internet have been rapidly growing (Chen et al., 2015). The internet of services is being formed (Tan et al., 2010). The number of services and their types being so large makes how to accurately and efficiently discover the desired services become a problem facing many service consumers (Zhang et al., 2016).

DOI: 10.4018/IJWSR.2018070102

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
Service clustering is considered to be an effective way to facilitate service discovery by organizing services into clusters of similar services to aid the pruning of the query space (Xia et al., 2015; Zhang et al., 2009). In recent years, several kinds of approaches have been proposed for clustering services to improve the performance of service discovery, mainly including functional-attributes-based (FAB) approaches (Dasgupta et al., 2011) and nonfunctional-attributes-based (NFAB) approaches (Chen et al., 2015; Xia et al., 2011; Zhu et al., 2012). In the current work, we focus on FAB approaches and aim at improving the quality of functionally similar clusters. The previous work on FAB approaches has made some great achievements, but there are still the following problems:

- Most prior FAB approaches usually designed for a single type of service documents and make use of the information in WSDL (Web Service Description Language) documents or OWL-S (Ontology Web Language for Services) documents to compute service similarity. However, some new types of services such as RESTful services, Web APIs, and Mashups lack of such a WSDL document or OWL-S document to be used;
- Most prior FAB approaches use the space vector model (Salton et al., 1975) to reduce the dimensions of the service documents. Only a few approaches take the topic of services into consideration to compute service similarity;
- Most prior FAB approaches neglect to utilize the topological information in service usage histories to compute service similarity. However, the service usage histories also include rich information about the similarity of services which is one of the main issues for better clustering performances.

Thus, questions such as “how to organize services into clusters in the absence of WSDL documents or OWL-S documents” and “how to group services into clusters by using information in the topic of services and topology of service usage histories” are natural to ask.

To solve the problems in existing FAB approaches and answer the above two questions, in this paper, we present TASC (Topology- and Topic-Aware Service Clustering), a novel service clustering approach to organize services into clusters by using meta-information such as (atomic/composite) service description and atomic service usage histories. First, TASC adopts a bipartite network (Pan et al., 2011) to formally describe the topological structure of service usage histories between composite services and atomic services. Based on the bipartite network, we apply the SimRank algorithm (Jeh & Widom, 2002) to measure the topological similarity of services which results from the topological context in which services occur. Second, TASC applies the Latent Dirichlet Allocation (LDA) (David et al., 2003) to extract topics from the profiles (the service description documents) of the atomic services and further to quantify the topic similarity of services. Third, we propose a new metric, service similarity (SS), by integrating topological similarity and topic similarity together, to quantify the total similarity between every pair of services. Finally, the proposed metric SS will be used in the Chameleon clustering algorithm (Karypis et al., 2008) to organize services into clusters. We implement the proposed approach on (atomic/composite) services collected from the real-world service repository, ProgrammableWeb (PWeb) (ProgrammableWeb, 2017), and demonstrate the usefulness of the combination of topological and topic similarities for clustering services.

In particular, the major contributions of our approach can be summarized as follows:

- We propose a more general service clustering approach TASC, which is not limited to the services with WSDL or OWL-S documents since it only uses the meta-information in service profiles;
- TASC is the first approach to utilize both the topological similarity and topic similarity to quantify the service similarity;
- We propose a novel approach to compute the topological similarity between every pair of services by using a bipartite network and the SimRank algorithm;
Related Content

Automated Situation-Aware Service Composition in Service-Oriented Computing
[www.igi-global.com/article/automated-situation-aware-service-composition/3109?camid=4v1a](www.igi-global.com/article/automated-situation-aware-service-composition/3109?camid=4v1a)

Workflow Discovery: Requirements from E-Science and a Graph-Based Solution
[www.igi-global.com/article/workflow-discovery-requirements-science-graph/3127?camid=4v1a](www.igi-global.com/article/workflow-discovery-requirements-science-graph/3127?camid=4v1a)
Security and Licensing for Geospatial Web Services
Bastian Schäffer and Rüdiger Gartmann (2011). Geospatial Web Services: Advances in Information Interoperability (pp. 64-95).
www.igi-global.com/chapter/security-licensing-geospatial-web-services/51483?camid=4v1a

A Specification Framework for Big Data Initiatives
www.igi-global.com/chapter/a-specification-framework-for-big-data-initiatives/217854?camid=4v1a