ICIC Express Letters
Part B: Applications
Volume **7**, Number **3**, March **2016**
 ISSN 2185-2766
pp. 613–619

A NEW USER PERSONALIZATION MODEL FOR WEB SERVICE DISCOVERY BASED ON PRUNING STRATEGY

Hao Tian[1] and Lijun Duan[2,*]

[1]School of Information Engineering
Hubei University of Economics
No. 8, Yangqiaohu Ave., Jiangxia Dist., Wuhan 430205, P. R. China
haotian@whu.edu.cn

[2]School of Computer
Hubei University of Education
No. 129, Gaoxin Second Road, Wuhan 430205, P. R. China
*Corresponding author: dljhue@sina.com

Received September 2015; accepted December 2015

Abstract. *Personalized information service has become the hot issue in the field of Web service research. User personalized model can build different models to suit individual behavior. By studying the ontology modeling and semantic technology, we presented a new user personalization model for Web service discovery based on pruning strategy in this paper, and respectively, analyzed and introduced the model formalization and pruning strategy. An application instance illustrated the execution flow of the pruning strategy and the structure property of the model, and simulation experiments also verified the performances of the proposed model by comparing two typical algorithms. The results show that this model can reveal the points of user interest, and has ideal stability and accuracy, which can satisfy the Web service application well.*
Keywords: Ontology, Personalization model, Semantic similarity, Pruning strategy

1. **Introduction.** User Personalization Model (UPM) is an application-oriented formalism with specific data structures. Based on different user habits and characteristics, it could realize distinct modeling so as to support personalized information service. In the current research field of Web services, user personalization model is widely used in Web service discovery, selection, combination and recommendation. Development of effective techniques for user personalization model is becoming a research hotspot. Currently there are many studies and typical methods related to user personalization model, which mainly focus on the definition and the expression of model, and generally adopt weights decomposition, utility function or other ways to model user's preferences; although their function is strong, due to a lack of satisfied global performance, most of these developed methods either behave in simple functionality, poor accuracy and low reliability, or preform with high complexity, large computational cost and weak flexibility, so their practical application effect tends to be low.

In this paper, we propose a new user personalization model for Web service discovery based on pruning strategy, which adopts ontology technology and mainly focuses on solving the balance between stability and flexibility, to enhance the reliability of final discovery. The experimental results show that the proposed model has realized the anticipated targets, and it could improve the efficiency of service discovery.

This paper is organized as follows. In Section 2 we will review the related personalization methods as well as analyze their strength and weakness. Model formalization and pruning strategy are introduced in Section 3, followed by the application and experiment in Section 4. The conclusion is given in Section 5.

2. **Related Work.** Various typical techniques have been used to develop efficient and effective personalization model to provide individualized service discovery to users based on their preferences and past behavior. For instance, choice trend of user is often used to define the user's individual preferences [1]. These methods often use WSMO (Web Service Modeling Ontology) to mark the Web service description and other content, and they focus on resolving some single value preferences.

Non-functional properties of Web service such as QoS also have recently been studied as an approach to construct models for personalization [2]. Relative weights are generally used to define user individuality preference factor in such methods. As the weight distribution is usually subjective, so the applicability of these methods is not satisfied.

Utility Function has been widely applied to user's preferences modeling [3]. In these methods, Utility Function is often conducted based on the QoS property of Web service. It is a special technology that depends on the information quantity about the QoS property. As the information gathering process usually takes a large consumption and has strict requirements, this kind of method is hardly to be realized.

Adapting the Conditional Preference Network (CP-net) is currently one of the common methods [4]. CP-net can be treated as a Bayesian network, which utilizes ceteris paribus [5] to express users' preferences. However, this kind of method is strictly limited by the number and capability of expressed preferences.

As the ontology can effectively support the knowledge inference, researching UPM by ontology technology [6] has become a hot issue in Web services applications. Using ontology model to define the users' individual preferences should take full advantage of history operation record and individual attribute information; meanwhile, the user's interests and preferences should be summarized with initiative. These kinds of methods have strong inference, high effect and reliability.

3. **Personalization Based on Ontology.** In this paper, we argue that the UPM for a specific user is essentially a subtree from domain ontology, and the main structure of UPM will be obtained by reasonable formal description about this subtree. To simplify the analysis, we assume the user's interest points are only confined to a particular application domain, and we tentatively will not consider user's interdisciplinary interest.

3.1. **Model formalization.** There are a variety of ontology formal methods [7] in current researching such as the five-tuple, six-tuple and seven-tuple. In this paper, we define the UPM P based on a previous study [8] which is in the form of three-tuples and can be shown as Equation (1):

$$P =< C, H, I > \tag{1}$$

where, C means the set of all concepts of ontology in all levels; H represents the hierarchy of C, which is used to describe specific relationships between concepts. In general, these relationships are parent-child relationships; I is used to describe the fact about user's interest in C, and it is a set of two-tuples consisting of a concept and the degree of user's interest in it. Specifically, I is shown as Equation (2):

$$I = \{(c_i, r_i)|i = 1, 2, \dots, n\}, \quad c_i \in C,\ r_i \in [0, 1] \tag{2}$$

where, c_i is a concept of C; r_i means the degree of user's interest in c_i.

In order to facilitate subsequent calculation and comparison, we first give the following definitions.

Definition 3.1. *r_i is the degree of user's interest in concept c_i, and it is the ratio of the called number of concept c_i and its child concepts to the total called numbers of all concepts of ontology. It can be calculated by Equation (3):*

$$r_i = \frac{cnt(c_i) + \sum cnt(c_j)}{\sum_{c_k \in C} cnt(c_k)}, \quad (c_i, c_j) \in H \tag{3}$$

Definition 3.2. *The distance between concept c_i and concept c_j is $dis(c_i, c_j)$, which is the number of nodes of the shortest path which connects these two concepts.*

Definition 3.3. *The depth of concept c_i is denoted by $dep(c_i)$, which is the distance between concept c_i and root node R.*

Definition 3.4. *The depth of ontology is expressed as $depO$, which is the maximum depth of concept about C.*

Definition 3.5. *The floor depth of ontologies is expressed as $FdepO_{ab}$ which is the minimum depth of ontology among ontology a and ontology b; and the upper depth of ontologies is expressed as $UdepO_{ab}$, which is the maximum depth of ontology among ontology a and ontology b.*

Definition 3.6. *$den(c_i)$ signifies the density of concept c_i, and it refers to number of child nodes which belong to concept c_i.*

Definition 3.7. *$jden(c_i, c_j)$ signifies the joint density between concept c_i and concept c_j, which can be calculated by Equation (4):*

$$jden(c_i, c_j) = \sqrt{\frac{den(c_i) \times den(c_j)}{den(c_i) + den(c_j)}} \tag{4}$$

In addition, we express the parent node of concept c_i as $par(c_i)$, express the nearest common parent node of concept c_i and concept c_j as $ncp(c_i, c_j)$, and denote the similarity degree between concept c_i and concept c_j by $sim(c_i, c_j)$.

In our rules, $den(par(R)) = 0$, and $sim(c_i, c_j)$ can be calculated by Equation (5):

$$sim(c_i, c_j) = \frac{dep(ncp(c_i, c_j)) + jden(par(c_i), par(c_j))}{dis(c_i, c_j) - 1 + dep(ncp(c_i, c_j)) + jden(par(c_i), par(c_j))} \tag{5}$$

3.2. **Pruning strategy.** In order to compare UPM efficiently, it is necessary to provide the calculation method and application of strategy. We propose a pruning strategy based on the above formalization.

Firstly, we need construct a matching tree MT to record the matching results. Each node nd_k of MT represents a concept pair cp_k and its matching result, which can be described as follows:

$$nd_k =< cp_k, sim(cp_k), r(cp_k) > \tag{6}$$

where, $cp_k = (c_i, c_j)$, $sim(cp_k) = sim(c_i, c_j)$, and $r(cp_k) = \min(r_i, r_j)$.

Suppose there are two UPM: P_a of U_a and P_b of U_b that need to be matched, and C_{am} means the set of the concepts of the m level from the root in P_a, $|C_{am}|$ means the number of concepts in C_{am}, and then the pruning strategy can be decomposed into the following steps.

(1) Match the root node R_a of P_a and root node R_b of P_b, and if they are the same then we can get the root node nd_1 of MT, namely, $nd_1 =< (R_a, R_a), 1.0, 1.0 >$; else we can make a conclusion that P_a and P_b are totally different and terminate the following steps.

(2) Continue to match next level nodes of P_a and P_b and get MT nodes directly by recording the concepts in the set C_m, where $C_m = C_{am} \cap C_{bm}$.

(3) If $|C_{am} - C_m| \neq |C_{bm} - C_m|$, it is necessary to prune the UPM which has the bigger number. Specifically, if $|C_{am} - C_m| > |C_{bm} - C_m|$, then rank the concepts in $|C_{am} - C_m|$ by $intr_i$ and delete the concepts which have low-ranking values and all their child nodes to make the number of left concepts in new set $C'_{am} - C_m$ equal to $|C_{bm} - C_m|$.

(4) Also rank the concepts in set $C_{bm} - C_m$ by the degree of user's interest, then select a concept from set $C'_{am} - C_m$ and a concept from set $C_{bm} - C_m$ separately in the same rank order, and calculate the pairwise similarity degrees.

(5) Record the concept pairs and their matching results whose similarity degrees are not smaller than a threshold T.
(6) Delete the concepts in the concept pairs whose similarity degrees are smaller than T and all their child nodes.
(7) Repeat step (2) to step (6) in the pruned P_a and pruned P_b until the level n, where $n = FdepO_{ab}$, which is the floor depth of pruned P_a and pruned P_b.
(8) After getting the complete MT, it is possible to calculate the similarity degree of these two UPM by Equation (7):

$$Psim_{ab} = \frac{2 \times FdepO_{ab} \times \sum_{cp_k \in MT} \left(r(cp_k) \times sim(cp_k) \times \frac{dep(cp_k)}{depO_{MT}} \right)}{UdepO_{ab} \times \left(\sum_{c_i \in P_a} \left(r_i \times \frac{dep(c_i)}{depO_a} \right) + \sum_{c_j \in P_b} \left(r_j \times \frac{dep(c_j)}{depO_b} \right) \right)} \tag{7}$$

4. **Application and Experiment.**

4.1. **Application instance.** To further illustrate the above strategy and the meaning of the formulas, in this paper, we firstly describe and analyze the related process through an application instance. We assume that users' interests focus in the field of tourism, and adopt the travel ontology in OWLS-TC3 as the standard ontology.

Suppose that user U_a has called the concepts of travel ontology 100 times in his operation history, and the specific distribution is shown in Figure 1. Then, according to the proposed methods, we can get the UPM P_a of U_a shown as Figure 2.

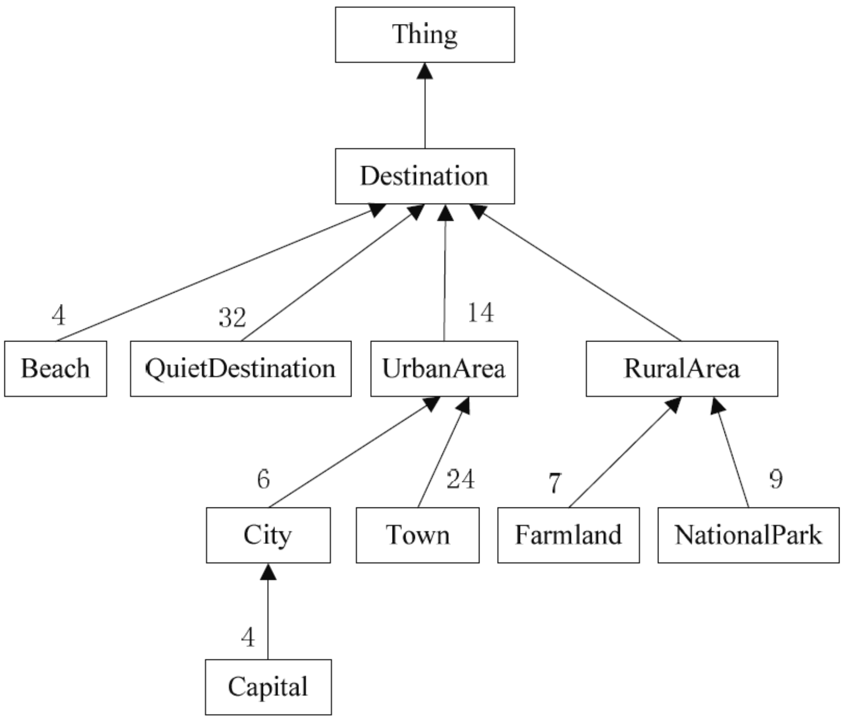

FIGURE 1. Distribution of ontology concepts called by U_a

Also suppose that another UPM P_b of user U_b is shown as Figure 3. In order to calculate the similarity degree of these two UPM, based on the proposed pruning strategy, we generate the matching tree MT_{ab} that is shown in Figure 4, where the threshold $T = 0.6$.

Then, according to the structure of MT_{ab} and Equation (7), we can get the similarity degree of P_a and P_b, namely, $Psim_{ab} = 0.712$.

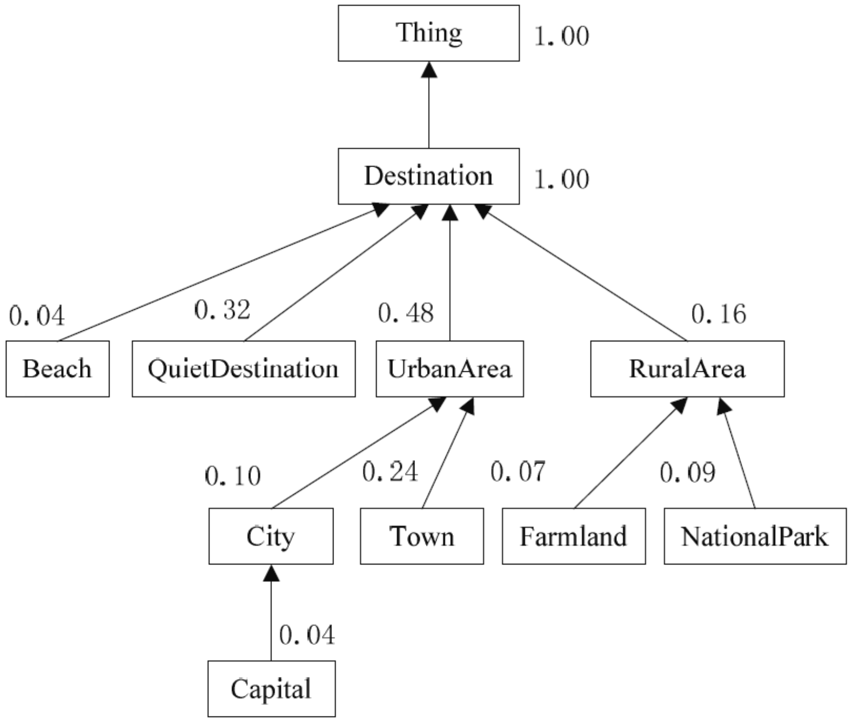

Figure 2. UPM P_a of U_a

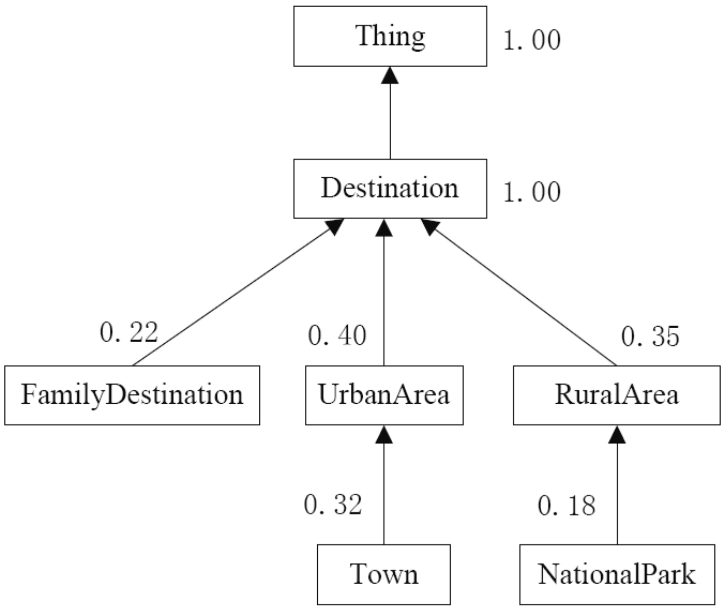

Figure 3. UPM P_b of U_b

4.2. **Experiment and analysis.** In order to verify the performances of the UPM proposed in this paper, we also conducted following simulation experiments. In a previous work [9] we have proposed a semantic Web service discovery with stage-based matching (DSM). And the proposed UPM can reveal the points of user's interest, so it can be used to narrow the matching range in discovery process. Therefore, we propose to improve DSM by adding a Web service filtering procedure before its initialization stage, and name the new algorithm advanced DSM (ADSM). To be specific, in the Web service filtering procedure, all available Web services will be ranked by the degree of correlation of user's UPM, which can be calculated by the method proposed in [8]. The procedure is usually performed when users are offline so as to improve the efficiency of ADSM.

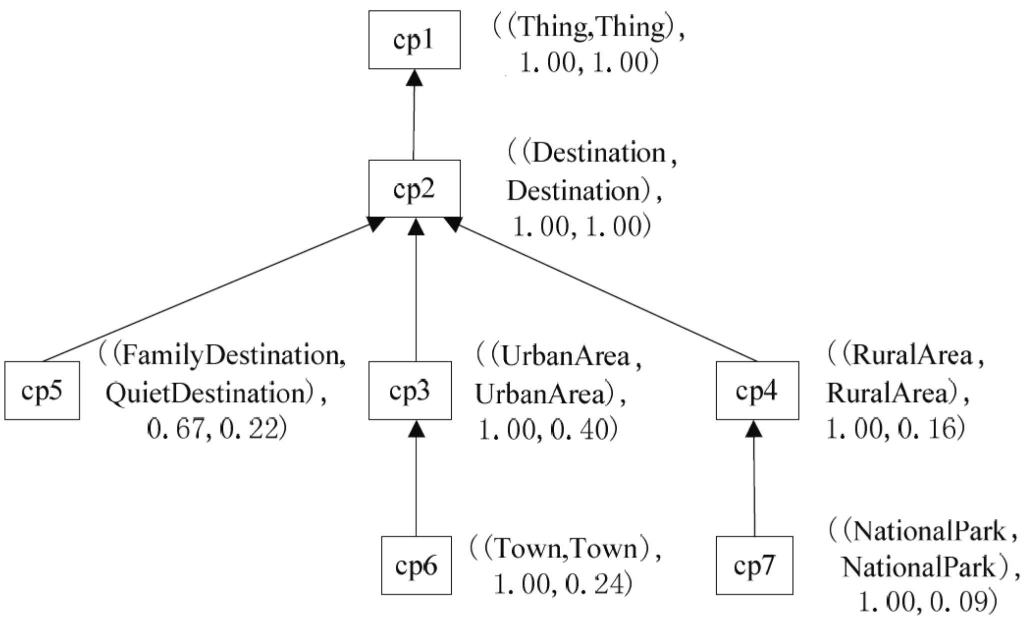

FIGURE 4. Matching tree MT_{ab} of P_a and P_b

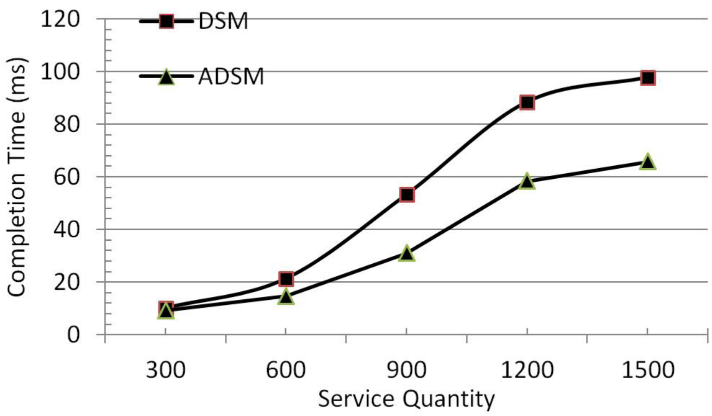

FIGURE 5. Comparison of discovery efficiency

We compare these two algorithms in terms of success rate and discovery efficiency, and adopt the definitions about the two parameters which are defined in [8]. The experimental tools are the same as before: Net Beans 5.5.1, Protégé 3.5, Jena 2.5.1 and Apache Tomcat 6.0.14. We take the OWLS-TC3 test set of queries as service requests and use the standard service descriptions of OWLS-TC3 and WS-DREAM set [10] involved in travel field as the original models of service advertisement, and create 1500 service description files with QoS level. The experiments compare the two algorithms while the number of available Web services is separately 300, 600, 900, 1200 and 1500.

In the simulation, the first comparison is about the discovery efficiency of ADSM and DSM, and the experimental results are shown in Figure 5.

As can be seen from Figure 5, when the number of services is around 300, the discovery completion time of both algorithms are roughly equal; with the increasing of service quantity, the discovery efficiency of ADSM is significantly better than the efficiency of DSM. And with the growth of the number of available services, the difference is more and more obvious. The reason is that, with the deepening of the operation of user, the UPM of user is increasingly clear and abundant, and the corresponding services which do not match user's preference will be filtered more thoroughly, so the follow-up matching range will be smaller and clearer.

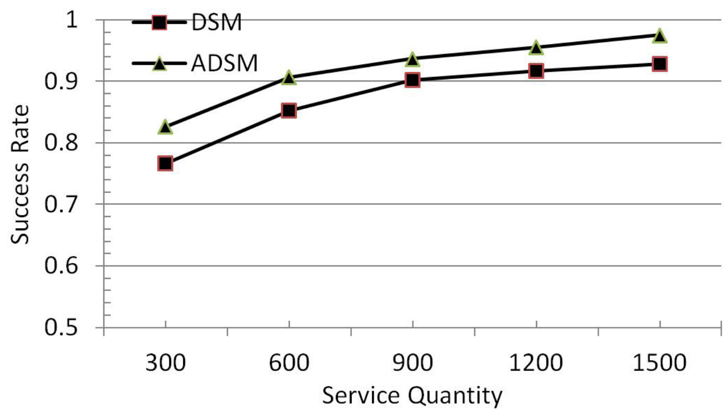

Figure 6. Comparison of success rate

Figure 6 shows that, compared with DSM, ADSM has a higher discovery success rate throughout. The main reason is that ADSM can match preferentially the available service related to user's interest with the help of the proposed UPM, and therefore, its discovery results can meet user's request better.

5. **Conclusions.** In this paper we proposed a new user personalization model for Web service discovery based on pruning strategy. We formalized the model on the basis of ontology technology, as well as provided the pruning strategy. Through comparing the completion time and success rate of two typical algorithms, experimental results show that the proposed model is better adapted to meet user's preference, and can satisfy the requirements of the application of service discovery well.

Acknowledgment. This work is supported by the Scientific and Technological Research Project Foundation of Educational Ministry of Hubei Province, China under Grant No. B2 015017.

REFERENCES

[1] S. G. Wang, X. L. Zhu, Q. B. Sun, H. Zou and F. C. Yang, Low-cost web service discovery based on distributed decision tree in P2P environments, *Wireless Personal Communications*, vol.73, no.4, pp.1477-1493, 2013.
[2] J. Z. Feng and L. F. Kong, Research on web service composition based on fuzzy QoS and preference weight, *Journal of Chinese Computer Systems*, vol.33, no.7, pp.1516-1521, 2012.
[3] C. Berry and H. Wang, Product architecting for personalization, *Journal of Manufacturing Systems*, vol.32, no.3, pp.404-411, 2013.
[4] P. T. Moore and H. V. Pham, Personalization and rule strategies in data-intensive intelligent contextaware systems, *Knowledge Engineering Review*, vol.30, no.2, pp.140-156, 2015.
[5] J. van Benthem, P. Girard and O. Roy, Everything else being equal: A modal logic for ceteris paribus preferences, *Journal of Philosophical Logic*, vol.38, no.1, pp.83-125, 2009.
[6] Y. Zhang, B. Fang and C. Y. Xu, Preference ontology-oriented metric model for trustworthy web services, *International Journal of Intelligent Systems*, vol.26, no.2, pp.158-168, 2011.
[7] N. Lingutla, J. Preece, S. Todorovic, L. Cooper, L. Moore and P. Jaiswal, AISO: Annotation of image segments with ontologies, *Journal of Biomedical Semantics*, vol.5, pp.50-55, 2014.
[8] H. Tian, *Research on User-Centered Web Service Discovery and Its Application in Financial Services*, Wuhan University, Wuhan, 2014.
[9] H. Tian and H. Fan, Distinct formalisms and stage-based matching for Web service discovery, *Journal of Computational Information Systems*, vol.8, no.22, pp.9377-9385, 2012.
[10] Z. B. Zheng, *WS-DREAM*, http://www.wsdream.net, 2015.