Future Generation Computer Systems ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Cultural distance for service composition in cyber–physical–social systems

Shangguang Wang [a], Yan Guo [a], Yan Li [b], Ching-Hsien Hsu [c,*]

[a] *State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China*
[b] *School of Business and Management, Shanghai International Studies University, Shanghai, China*
[c] *School of Mathematics and Big Data, Foshan University, China*

HIGHLIGHTS

- This paper proposes a service composition approach based on the cultural distance.
- We propose a novel concept called the preference degree to represent the degree of preference of the users to the services.
- The proposed approach is evaluated using a real-world service QoS dataset.
- Experimental results show that our approach can obtain composition services that satisfy specific requirements in a short time.

ARTICLE INFO

Article history:
Received 23 May 2017
Received in revised form 16 March 2018
Accepted 10 June 2018
Available online xxxx

Keywords:
Service composition
User preference
Cultural distance
Cyber–physical–social systems

ABSTRACT

Cyber–physical–social systems can be organized as workflows that interconnect the resources in physical, cyber, and social worlds in real time. This integration of these worlds with various resource as services requires service composition approaches that can integrate essential components and share relevant information in these worlds. Although numerous service composition approaches have been proposed, there still exist many challenges in the representation of user preference of service composition in cyber–physical–social systems. In this paper, a service composition approach based on the cultural distance is proposed to improve the reliability and satisfaction. This approach employs the cultural distance to measure the user preference quantitatively in a simple mathematical form. The user preference degree and the user preference vector are defined to select the service from a global view and an accurate point separately. Then, a 0-1 mixed-integer programming algorithm is used to identify the most suitable services for composition. The experimental results based on two real-world datasets show that the proposed approach for representing the user preference quantitatively is more simple and effective than other approaches. In a word, the proposed approach yields satisfactory performance of service composition in CPSS with regard to the user preference and efficiency.

1. Introduction

Cyber–physical–social systems (CPSS) is the next generation of intelligent systems which aim at interconnecting cyber, physical and social worlds by monitoring and controlling these worlds [1], and it will lead us to an era of intelligent enterprises and industries [2]. It focuses on coordination of the human, computing and physical resources, and will be applied in intelligent enterprise, intelligent transportation, smart home, intelligent medical and other fields. Currently, CPSS is organized as workflows that are composed of multiple, heterogeneous elements and data types from cyber, physical and social worlds, such as smart phones, tablets, cameras, vehicle, energy meters, computational models, programming abstraction models and social media [3–9]. To facilitate the integration tasks in heterogeneous CPSS workflows, service composition approach that can be used to integrate essential components in these world is required [10].

Service-oriented architecture (SOA) enables the service composition in a loosely coupled way in order to achieve complex functionality by combining basic services [11]. It is well-known that with the rapid development of CPSS, the number of essential components published in the three worlds is increasing rapidly. At the same time, the requirements of CPSS users are becoming increasingly complex and personalized. Therefore, although SOA technologies provide an easy way to integrate components within and across organizational boundaries, it is difficult for CPSS users

* Corresponding author.
E-mail addresses: sgwang@bupt.edu.cn (S. Wang), guoyan@bupt.edu.cn (Y. Guo), liyan0417@shisu.edu.cn (Y. Li), chh@chu.edu.tw (C.-H. Hsu).

https://doi.org/10.1016/j.future.2018.06.012
0167-739X/© 2018 Published by Elsevier B.V.

to obtain a service composition that satisfies their specific requirements for quality of service (QoS) [12].

Most existing service composition algorithms considered the user preference are based on the Semantic Web environment, where a user agent works on behalf of its owner and knows the personal preference of users. With the rapid increase in the number of components, it is critical for the agent to identify the service that best matches the user preference in CPSS. Thus, the representation of the user preference and the service composition according to the representation are important research problems in CPSS.

Although numerous service composition approaches based on the user preference have been proposed, for example, researchers have applied a hierarchical task network [13] and fuzzy set theory [14–16] to represent the requirements and preference of users, these approaches have the following two limitations.

First, the previous service composition approaches seek to represent the user preference in as accurate and detailed a manner as possible. However, some users do not realize their real preference clearly or cannot express their preference clearly; therefore, the accurate and detailed representation of the user preference may lead to an inaccurate representation of the user preference in CPSS. Additionally, the result of the service composition is very sensitive to the preference weight in CPSS. In this case, the too accurate representation of the user preference may lead to an unreliable service composition result. Moreover, the accurate and detailed representation of the user preference costs additional time and affects the real-time nature of the service composition in CPSS.

Second, the existing methods are complicated because they always concentrate on obtaining a preference representation for each user and ignore the group or regional characteristics of CPSS users. In many domains, it is desirable to assess such a preference qualitatively rather than quantitatively [17]. Thus, it is not necessary to give the preference representation for each user, only to give the preference representation for each user group formed according to some characteristics.

To meet the personalized requirements of CPSS users and improve the reliability and satisfaction of service composition results in CPSS, this paper aims to propose a new service composition approach considered the user preference. In response to these limitations of existing approaches, the proposed approach does not concentrate on obtaining an accurate preference representation for every user but on finding a simple and powerful representation. Considering that there is a close relationship between the user preference and their cultural background, this paper introduces an important concept in the study of cultural differences, called cultural distance, employs the cultural distance to express the user preference and propose a service composition approach based on cultural distance. In the proposed service composition approach, the representation of the user preference based on the cultural distance groups the CPSS users according to their country/region. Users from the same region have the same preference value. Hence, it does not matter if the preference of some CPSS users is unclear. Thus, the proposed approach in this paper can avoid the two aforementioned problems.

The main contributions of this paper are summarized as follows:

- This paper proposes a novel representation of the user preference based on the cultural distance, which is a concept used in cultural-difference measurement systems [18]. Compared with the existing representations, the representation of the preference in this paper is simpler and more effective for CPSS.
- This paper proposes a novel concept called the preference degree to represent the degree of preference of the users to the services. Services with a low preference degree are eliminated to improve the user satisfaction and the real-time nature of the service composition in CPSS.

The remainder of the paper is organized as follows: Section 2 discusses the background of service composition; Section 3 provides a novel representation of the user preference and describes the proposed service composition approach; Section 4 presents the experimental results; And Section 5 concludes the paper and discusses future work.

2. Background

Service composition technology is the core technology of service-oriented architecture and service-oriented computing and can quickly satisfy the requirements of complex, dynamic, and inter-organizational businesses. In this section, the service composition problem and related research are introduced.

2.1. Concept of service composition

Service composition is a process of dynamically selecting, integrating, and calling candidate services through workflows according to specific requirements. It can combine basic candidate services to provide a new complex value-added service that satisfies specific requirements of users.

Let S be a composition service process in a CPSS that contains n service classes with different functions, i.e., $S = \{S_1, \ldots, S_n\}$. $S_i (0 < i \leqslant n)$ represents a concrete service class, which is a set composed of l service candidates with the same functions but different non-functional attribute values, i.e., $S_i = \{s_{i1}, \ldots, s_{il}\}$. $s_{(ij)} (0 < j \leqslant l)$ represents a concrete service candidate, which is a service that can satisfy a particular functional demand. In CPSS, the service candidates represent the multiple, heterogeneous elements and data types from cyber, physical and social worlds, and a concrete service class represents an abstract task.

Service selection involves choosing the most suitable candidate service to satisfy the functional and non-functional requirements of CPSS users from one or more service classes, that is, selecting a s_{ij} from each S_i to integrate a composition service under the global constraints.

To select a candidate service from a service class, an indicator is needed to measure the merits of multiple candidate services with the same function. The QoS is the commonest index in the current service composition research based on non-functional attributes [19,20]. For example, a candidate service s has r QoS attributes, and the QoS vector of s is $Qs = \{q_1(s), \ldots, q_r(s)\}$, where $q_k(s) (0 < k \leqslant r)$ represents the kth attribute value of s. Similarly, the QoS vector of composition service S is $QS = \{q_1(S), \ldots, q_r(S)\}$, where $q_k(S) (0 < k \leqslant r)$ represents the aggregation value of the kth attribute value of the candidate services selected from the respective service classes.

The QoS attributes of services are generally divided into two categories: positive QoS attributes and negative QoS attributes [21]. The values of positive QoS attributes need to be maximized, such as the reliability, reputation, availability, and throughput. In contrast, the values of negative QoS attributes need to be minimized, such as the response time, delay time, and price. Note that, this paper only considers the positive QoS attributes for the sake of convenience in this study, as the positive attribute values can easily be converted into negative attribute values.

Generally, it is difficult to compare and sort service compositions solely using the QoS, because the service composition has several QoS attributes with different units and scopes in CPSS. A utility function is designed to map the vector of QoS values into a real value to compare and sort the candidate services and composition services in CPSS. The utility functions of a candidate service and a composition service in the sequential composition model can be computed as follows:

$$U(s) = \sum_{k=1}^{r} \omega_k \cdot \frac{q_k(s) - Q_{i,k}^{min}}{Q_{i,k}^{max} - Q_{i,k}^{min}}, \tag{1}$$

Please cite this article in press as: S. Wang, et al., Cultural distance for service composition in cyber–physical–social systems, Future Generation Computer Systems (2018), https://doi.org/10.1016/j.future.2018.06.012.

$$U(S)=\sum_{k=1}^{r}\omega_k\cdot\frac{q_k(S)-Q_k^{min}}{Q_k^{max}-Q_k^{min}}, \tag{2}$$

$$\begin{cases} Q_k^{max}=\sum_{i=1}^{n}Q_{i,k}^{max} \quad (Q_{i,k}^{max}=\max\limits_{\forall s_{ij}\in S_i} q_k(s_{ij})) \\ Q_k^{min}=\sum_{i=1}^{n}Q_{i,k}^{min} \quad (Q_{i,k}^{min}=\min\limits_{\forall s_{ij}\in S_i} q_k(s_{ij})) \end{cases} \tag{3}$$

where r is the number of QoS attributes, $\omega_k(0 < \omega_k < 1, \sum_{k=1}^{r}\omega_k = 1)$ is the weight of the kth QoS attribute representing the preference of users, $Q_{i,k}^{max}$ is the maximum value of the kth QoS attribute in all the candidate services of the ith service class, Q_k^{max} is the maximum value of the kth QoS attribute in all the composition services S and is the sum of $Q_{i,k}^{max}$, similarly, $Q_{i,k}^{min}$ is the minimum value of the kth QoS attribute in the ith service class, and Q_k^{min} is the sum of $Q_{i,k}^{min}$.

The optimal service composition in CPSS must satisfy not only the maximum utility function value but also the given vector of global QoS constraints; i.e., $C = \{C_1, \ldots, C_m\}(0 \le m \le r)$, $q(S) \le C(\forall C_k \in C)$, where $C_k(0 \le k \le m)$ represents the maximum or minimum boundary of an aggregated QoS attribute, $q(S)$ is the aggregated QoS value of the composition service.

2.2. Related work

Numerous service composition approaches have been proposed in the literature. Most of the optimization of service composition approaches is based on QoS, and some of them mainly focus on the reliability [22–27], some of them mainly focus on the efficiency and real time [28–34], some of them mainly focus on the quality of user experience [35]. This paper mainly focuses on the user preference on QoS and reviews some notable ones here.

To date, research on user preference of service composition has mainly focused on measurements for quantitative preference rather than qualitative preference, although the latter has attracted considerable attention recently [36].

For example, C. Boutilier et al. [37] proposed a qualitative graphical representation of preferences that reflects the conditional dependence and independence of preference statements under a ceteris paribus (all else being equal) interpretation. Such a representation is often compact and arguably quite natural in many circumstances. They provided formal semantics for this model and describe how the structure of the network can be exploited in several inference tasks, such as determining whether one outcome dominates (is preferred to) another, ordering a set outcomes according to the preference relation, and constructing the best outcome subject according to the available evidence.

Wang et al. [17] used a qualitative graphical representation tool to describe qualitative preference relationships in a compact, intuitive, and structured manner under conditional ceteris paribus (all else being equal) preference statements.

In some studies, fuzzy set theory was used to represent the needs and preference of users. Wen et al. [16] modeled the user preference and satisfaction as a fuzzy constraint satisfaction problem. On this basis of this model, the hierarchical task network is used to compose services via the branch and bound method.

Reformat et al. [14] proposed an approach for providing the capability to mimic human behavior in the case of a multi-criteria decision-making process and applied fuzziness and an approximate reasoning methodology in a Semantic Web environment. An ontology with fuzziness was used to represent human needs and preference and contained information about different acceptance levels that users may have depending on responses obtained from different service providers.

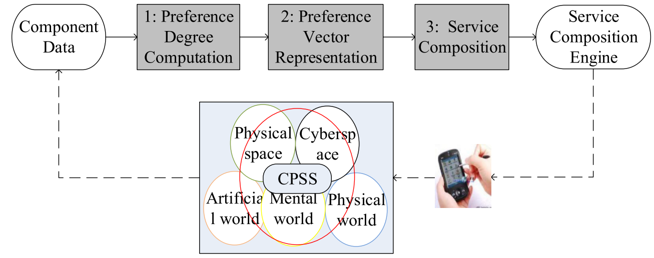


Fig. 1. Procedures of the service composition approach based on cultural distance.

3. Service composition approach

Aiming to identify the services that best satisfy the needs of users, cultural distance is introduced as a new QoS attribute. Based on this attribute, this paper proposes a simple service composition approach in CPSS that can satisfy the user preference.

As shown in Fig. 1, the proposed approach can be divided into three phases. The first phase is the preference degree computation, where the user preference degree to services is computed and the services with a low preference degree are filtered out in order to save the time cost of service composition in CPSS. The second phase is the preference vector representation, where a simpler and more effective method is used to represent the user preference. The first phase involves selecting services from a global view, and the second phase involves selecting services from an accurate point. The final phase is service composition, where a 0–1 MIP algorithm is adopted to identify the most suitable service in each service class for composition.

3.1. Cultural distance

The main idea of the proposed approach is the use of the cultural distance, and the first two phases of the approach refer to the cultural distance. Thus, cultural distance is introduced in detail here.

Cultural distance is an important concept in the study of cultural differences and is commonly used in cross-cultural study and practice, such as enterprise internationalization [38]. However, all computation of cultural distance is based on research of the cultural dimension.

The most famous research on the cultural dimension is Hofstedes cultural dimension theory, which provided the first quantitative description of the abstract complicated concept of culture [14]. This cultural dimension theory allows researchers to express culture as data and intuitively compare the differences between cultures and various behaviors due to cultural differences.

Hofstedes cultural dimension theory consists of six cultural dimensions, which represent independent preferences for one state of affairs over another that distinguish countries (rather than individuals) from each other. The country scores for the dimensions are relative. That is, the scores can be only used meaningfully by comparison. The six dimensions are as follows.

(1) ***Power distance index (PDI)***. This dimension expresses the degree to which the less powerful members of a society accept and expect that power is distributed unequally. Owing to different understandings of power, there is a large disparity in this dimension between countries.
(2) ***Individualism versus collectivism (IDV)***. Individualism can be defined as a preference for a loosely-knit social framework where individuals are expected to take care of only themselves and their immediate families. Its opposite, collectivism, represents a preference for a tightly-knit social

framework where individuals focus on ethnic relations and the interests of the collective.

(3) ***Masculinity versus femininity (MAS)***. Masculinity represents a societal preference for achievement, heroism, assertiveness, and material rewards for success. Its opposite, femininity, represents a preference for cooperation, modesty, caring for the weak, and quality of life.

(4) ***Uncertainty avoidance index (UAI)***. This dimension expresses the degree to which the members of a society feel uncomfortable with uncertainty and ambiguity. Countries with a large UAI maintain rigid codes of belief and behavior and are intolerant of unorthodox behavior and ideas. Countries with a small UAI maintain a more relaxed attitude, where practice is held in higher regard than principles.

(5) ***Long-term orientation versus short-term normative orientation (LTO)***. Countries that score low on this dimension maintain time-honored traditions and norms while viewing societal change with suspicion. On the other hand, countries that score high adopt a more pragmatic approach: they encourage thrift and efforts in modern education to prepare for the future.

(6) ***Indulgence versus restraint (IND)***. Indulgence refers to a society that allows the relatively free gratification of basic and natural human drives related to enjoying life and having fun. Restraint refers to a society that suppresses the gratification of needs and regulates it via strict social norms.

In Hofstedes analysis of culture, values are considered to be the core part of culture. They directly impact peoples beliefs, thoughts, ideas, actions, and social behavior. Although some people are unaware of this impact, culture and values have taken root in peoples minds and are the decisive factors on our social behavior. Hofstedes research shows that people always habitually think and act according to their experience and that the thoughts, feelings, and behaviors of people in different cultures differ, even with regard to the most basic things, such as work, study, and communication.

In accordance with Hofstedes cultural dimension theory, the cultural distance was first defined as a simple mathematical formula to simply describe cultural differences [39], as follows:

$$CD_j = \frac{1}{m}\sum_{i=1}^{m}\frac{(I_{ij}-I_{ik})^2}{V_i}, \tag{4}$$

where m is the number of cultural dimensions, $CD_j(j = 1, 2, \ldots)$ is the cultural distance of the jth country, $I_{ij}(i = 1, \ldots, m)$ is the ith Hofstede cultural dimension score of the jth country, the kth country is the host country, and V_i is the variance of the ith cultural dimension score.

3.2. Preference degree computation

On the national/regional level, the user preference in the service composition in CPSS can be interpreted as a reflection of the culture. For example, to improve user satisfaction, international companies can provide users from different regions with different services that are consistent with the local characteristics and preferences. Accordingly, cultural distance is introduced to represent the user preference in the service composition approach for CPSS.

Everything in the world has a cultural distance. In contrast to other QoS attributes, the cultural distance is not an attribute that only services can have; the users also have a cultural distance. According to this feature, a relationship between the users and services that represents the user preference to different services can be established.

Definition 1 (*Preference Degree*)**.** The preference degree represents the degree of the user preference to the services and is defined as the difference between the culture distance of user and the culture distance of service. The formula is as follows:

$$D_{us} = -\left|CD_u - CD_s\right|, \tag{5}$$

where CD_u represents the culture distance of the user u, CD_s represents the culture distance of the service s, and D_{us} represents the preference degree. This definition indicates that a smaller cultural difference yields a higher preference degree.

To improve the user satisfaction and reduce the time required for composition in CPSS, filtering is performed in this paper. Specifically, the services with a low preference degree are filtered out, and the filter factor is a number between 0 and 1 that can be adjusted according to the actual situation.

3.3. Preference vector representation

The preference degree is a rough description of the user preference to select services from a global view, and in this section, an accurate description of the user preference to select services from an accurate point is proposed.

Definition 2 (*Preference Vector*)**.** The preference vector represents the user preference to the services, more specifically, the preference to the QoS attributes of services. The formula is as follows:

$$\omega = (\omega_1, \ldots, \omega_k, \ldots, \omega_r), \tag{6}$$

$$\begin{cases} \omega_k = \dfrac{\sum_{i=1}^{n} q_k(ds_i)}{\sum_{k=1}^{r}\sum_{i=1}^{n} q_k(ds_i)} \\ ds_i = arg\max\limits_{s\in S_i}\{D_{us}\} \end{cases} \tag{7}$$

where ω represents the preference vector, r represents the number of QoS attributes, $\omega_k(k = 1, \ldots, r)$ represents the preference value of the kth QoS attribute, n represents the number of service classes, and ds_i represents the service with the highest preference degree in the ith service class. This definition indicates that $0 < \omega_k < 1$, and $\sum_{k=1}^{r}\omega_k = 1$. Note that $q_k(ds_i)$ in this formula is the standardized value.

In contrast to the existing approach, the representation of the user preference based on cultural distance is a quantitative description and can be seen as a combination of a quantitative description and a qualitative description because the users are grouped according to regions. From this perspective, the representation is more flexible and practical than the existing approach.

3.4. Service composition

Following the computation of the preference degree and the preference vector, the preference value of each QoS attribute and the services with a high preference degree are identified. Then, a service composition algorithm is used to identify the most appropriate services in each service class under the global QoS constraints. In this paper, the 0–1 MIP algorithm is adopted to solve this optimization problem in CPSS [25].

By re-writing (2), the overall utility function is obtained as follows:

$$F(S) = \sum_{k=1}^{r}\omega_k \cdot \frac{\sum_{i=1}^{n}\sum_{j=1}^{l} x_{ij}\cdot q_k\left(s_{ij}\right) - Q_k^{min}}{Q_k^{max} - Q_k^{min}}, \tag{8}$$

where x_{ij} is a binary decision variable indicating whether a service candidate is selected. If x_{ij} is 1, the corresponding candidate service s_{ij} has been selected; otherwise, it has been discarded.

Please cite this article in press as: S. Wang, et al., Cultural distance for service composition in cyber–physical–social systems, Future Generation Computer Systems (2018), https://doi.org/10.1016/j.future.2018.06.012.

Table 1
Candidate services of each service class.

	Response time	Video quality	Cultural distance
s_{11}	10	12	0.3
s_{12}	5	8	2.8
s_{13}	8	10	1.6
	Response time	Video quality	Cultural distance
s_{21}	6	10	3.3
s_{22}	8	15	1.2

This service composition problem can be formulated as follows:

$$Max \quad F(S) \tag{9}$$

s.t.

$$\sum_{i=1}^{n}\sum_{j=1}^{l} q_k(s_{ij}) \cdot x_{ij} \leq C_k, \quad 1 \leq k \leq m \leq r \tag{10}$$

$$\sum_{i=1}^{n} x_{ij} = 1, \quad 1 \leq i \leq n \tag{11}$$

$$x_{ij} \in \{0, 1\} \tag{12}$$

There are many ways to solve this 0–1 MIP problem, such as Enumeration algorithm, Cutting Plane algorithm, Branch and Bound algorithm, Heuristic methods and so on [40]. In this paper, the 0–1 MIP problem is solved by using IBM-CPLEX solver [41] which applies Branch and Cut algorithm, and then the service composition satisfying the above constraints and the CPSS users preference can be obtained.

4. Experiment

This paper compares the proposed service-composition approach, which is called SIR, with the MIP approach [42] and two variational approaches of SIR with regard to the user satisfaction and real time using two real-world datasets. The experimental results show that SIR can find a more satisfactory service composition solution in a shorter computation time than other approaches.

Additionally, this paper analyzes the effect of the filtration ratio in SIR and discuss the relationship between the cultural distance and the preference to further investigate the computation of the user preference and the service composition.

4.1. Case study

As case study, this paper considers an online educational video application, which is composed of two types of atom services. The first type of atom service (denoted by S_1) enables the selection of the video source; The second type of atom service (denoted by S_2) is a compression service, which is used to adapt the video content to the wireless link. Each service class often consists of multiple candidate services. For example, many video companies may have the same video.

There are three candidate services for S_1, which is denoted by $\{s_{11}, s_{12}, s_{13}\}$. There are two candidate services for S_2, which is denoted by $\{s_{21}, s_{22}\}$. In the service composition process, a subset of the atom services is need to be selected to form the composition service. As shown in Table 1, the candidate services offer different QoS values (e.g. response time and video quality) and cultural distance, which leads to the composition video service with different sharpness and fluency. The traditional service composition approaches always set the preference vector for each QoS attribute roughly, or describe the preference in a complicated way.

Table 2
Utility values of all the composition services.

	s_{12}	s_{13}
s_{21}	0.8	0.5
s_{22}	0.68	0.38

Assume that the cultural distance of the user is 2.4, the preference degree of the candidate services can be calculated through the above definition, the preference degree of $\{s_{11}, s_{12}, s_{13}\}$ and $\{s_{21}, s_{22}\}$ are −2.1, −0.4, −0.8, and −0.9, −1.2, respectively. To reduce the computation time, the service with a low preference degree: s_{11} is filtered. According to the definition, ds_1 is s_{12} and ds_2 is s_{21}, the preference vector is (0.86, 0.14). And then compute the utility value of all the composition services, the results are as Table 2 shown. The results show that the best composition service is $\{s_{12}, s_{21}\}$.

4.2. Experiment setup

To ensure the reliability of the experiment, two real-world datasets are used: the WS-Dream dataset [43] and Hofstedes cultural dimension dataset [18].

The WS-Dream dataset describes nearly 2 million real-world QoS evaluation results for 5,825 services from 339 users. It contains the location information (e.g., IP address, latitude, longitude) for the users and services, and each invocation record contains two QoS attributes: the response time and throughout.

Hofstedes cultural dimension dataset has been published and contains the scores for the six cultural dimensions. The PDI and the UAI values for 76 countries and regions are obtained according to three items in the IBM database plus extensions, the IDV and the MAS values for 76 countries and regions are obtained according to factor scores from 14 items in the IBM database plus extensions, and the LTO and IVR index values for 93 countries and regions are obtained according to factor scores from three items in the World Values Survey.

To evaluation the performance of the approach, a series of experiments are conducted to compare SIR with three other approaches with regard to the user satisfaction and the real time: SIR-, SIR*, and MIP.

- **SIR**. The approach as described in Section 3, involving filtration, the accurate representation of the preference, and the MIP selection algorithm.
- **SIR**$^{-}$. A variation of SIR that involves the accurate representation of the preference and the MIP selection algorithm but not the filtration.
- **SIR***. A variation of SIR that involves the filtration and the MIP selection algorithm but not the accurate representation of the preference.
- **MIP**. The simple MIP approach, which only involves the MIP selection algorithm and does not involve the filtration or the accurate representation of the preference.

In the experiments, unless otherwise specified, the number of service classes n is 5, the number of candidate services in each service class l is 200, the number of QoS attributes r is 2, the number of historical records for each candidate service is 250, and the number of global constraints m is 2. In SIR and SIR*, the filtration ratio is 0.5. In MIP and SIR*, the weights of the two QoS attributes are 0.5. All experiments are conducted on the same computer, which has an Intel(R) Core(TM) 2.5 GHz processor and 4.0 GB of RAM, using the Windows 10 operating system and the MATLAB R2015a software.

Please cite this article in press as: S. Wang, et al., Cultural distance for service composition in cyber–physical–social systems, Future Generation Computer Systems (2018), https://doi.org/10.1016/j.future.2018.06.012.

4.3. User satisfaction

The proposed approach employs a novel representation of the user preference for selecting the services satisfying the user preference and improving the user satisfaction. In this section, to compare the performance of the four approaches, the user satisfaction in CPSS is defined.

Definition 3 (*User Satisfaction*)**.** The user satisfaction is the utility value of the service composition result considering the real user preference vector and can be expressed by the following formula:

$$Sa = \sum_{k=1}^{r} \omega_k \cdot \frac{\sum_{i=1}^{n} q_k(s_i) - Q_k^{min}}{Q_k^{max} - Q_k^{min}} \tag{13}$$

where $\{s_i\}$ $(i = 1, \ldots, n)$ are the services selected from each service class, and $\{\omega_k\}$ $(k = 1, \ldots, r)$ is the user preference vector defined in Eq. (6).

Fig. 2(a) shows the comparison results in terms of the user satisfaction with different numbers of candidate services ranging from 100 to 900. Similarly, Fig. 2(b) shows the comparison results in terms of the user satisfaction with different numbers of service classes ranging from 2 to 10. As shown in Fig. 2(a) and 2(b), the user satisfaction for SIR and SIR$^-$ are superior to that for the other two approaches. That is, the representation of the user preference in this paper can improve the user satisfaction of the service composition. In addition, the user satisfaction for SIR$^-$ is higher than that for SIR in some cases, because some services are filtered out according to the preference degree, and the selection space is smaller in SIR. However, the difference between SIR and SIR$^-$ is very small and can be ignored in most cases. It is difficult to compare the user satisfaction between MIP and SIR*, as the real user satisfaction with the selected services cannot be obtained under the assumption that the user preference vector is (0.5, 0.5).

Note that in Fig. 2(a), in contrast to the SIR* and MIP approaches, the utility values for SIR- and SIR do not increase with the number of candidate services, because the user preference vector may change when the number of candidate services increases, which can affect the utility value of the service composition.

4.4. Real time

In this fast-paced world, CPSS users always want to obtain the service composition as quickly as possible. However, the existing service composition approaches based on the user preference usually require extra time owing to the complicated computation procedure of the user preference. In this section, the real time of the four approaches are compared.

Fig. 3(a) shows the comparison results in terms of the computation time with different numbers of candidate services ranging from 100 to 900. Fig. 3(b) shows the comparison results in terms of the computation time for different numbers of service classes ranging from 2 to 10. As shown in these figures, except for cases where the number of service classes or candidate services is too small (n = 2, l = 200; n = 3, l = 200; n = 5, l = 100), the computation time for SIR and SIR* are superior to that for the other two approaches. Thus, the filtration can reduce the computation time and improve the real-time nature of the service composition. In addition, the computation time for SIR is similar to that for SIR*, and the computation time for SIR$^-$ is similar to that for MIP. That is, the computation of the user preference vector hardly affects the time cost of the service composition.

4.5. Parameter analysis of filtration ratio

In SIR, the preference degree is defined to filter out the redundant services and improve the real-time nature of the service composition in CPSS. In previous experiments, 50% of the candidate services were filtered out according to the preference degree. Obviously, the filtration ratio can affect the results of the experiment, thus, this section analyzes the effects of the filtration ratio on the user satisfaction and real time of SIR.

Fig. 4(a) shows the user satisfaction of SIR for different numbers of candidate services with different filtration ratios ranging from 0.1 to 0.9, and Fig. 4(b) shows the user satisfaction of SIR for different numbers of service classes with different filtration ratios ranging from 0.1 to 0.9. As shown in Fig. 4(a) and 4(b), regardless of the numbers of service classes and candidate services, the user satisfaction decreases with the increase of the filtration ratio. This is easy to understand because a higher filtration ratio causes a larger number of the candidate services to be filtered out and a worse best service to be selected. Expect for the case of l = 300, there is nearly no reduction of the user satisfaction for the service composition solution when 50% (or even 60%) of the candidate services are filtered out.

Fig. 5(a) shows the computation time of SIR for different numbers of candidate services with different filtration ratios ranging from 0.1 to 0.9, and Fig. 5(b) shows the computation of SIR for different numbers of service classes with different filtration ratios ranging from 0.1 to 0.9. As shown in these figures, regardless of the numbers of service classes and candidate services, the computation time decreases with the increase of the filtration ratio. This is easy to understand because a higher filtration ratio yields a lower amount of remaining candidate services after filtration and a lower computation time. In most cases, when 50% of the candidate services are filtered out, the computation time is reduced by up to 40%.

4.6. Parameter analysis of preference

This paper uses the cultural distance data for users and services to compute the user preference. However, in reality, there are numerous services and not all of them have cultural distance, thus, this paper attempts to determine the relationship between the user cultural distance and the user preference.

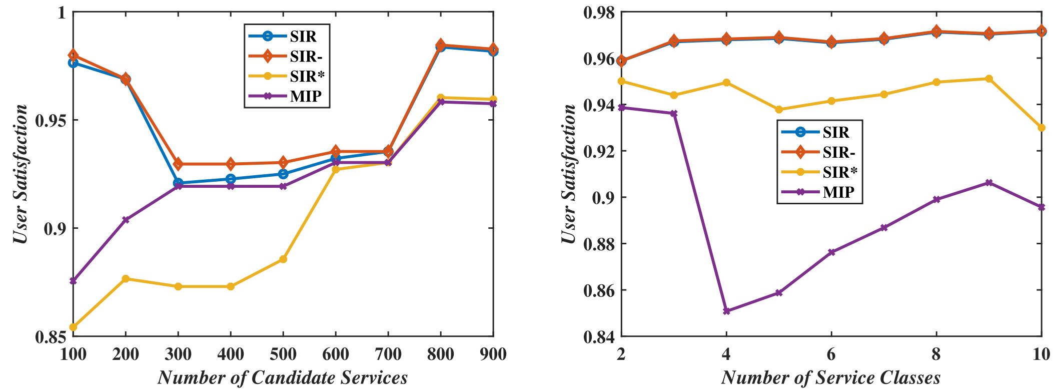
Fig. 6 shows that when the user cultural distance is between 0 and 0.5, the preference value is approximately 0.965; when the user cultural distance is between 1 and 2, the preference value is approximately 0.885; and when the user cultural distance is between 2 and 5, the preference value is approximately 0.98. Note that, the preference value indicates the preference of the user to response-time attribute.

If the relationship can be further understood, the user preference can be conveniently and efficiently computed only using the data for the user cultural distance when the data for the service cultural distance are incomplete. Research on the relationship between the user cultural distance and the user preference is useful not only for the service-composition in CPSS but also for recommendation systems, quality of experience, and other fields related to user preference in CPSS.

4.7. Results analysis

The performance of the four algorithms are illustrated in Figs. 2–3, the parameter analysis of SIR is illustrated in Figs. 4–6. These results reveal a number of interesting points.

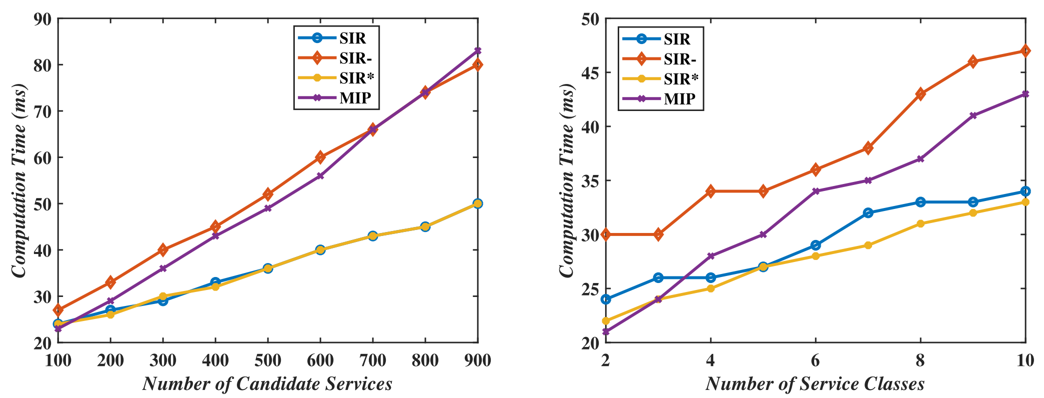
(1) Considering the user satisfaction, Fig. 2 illustrates the results of the four algorithms. SIR and SIR$^-$ are superior to SIP* and MIP, which shows that the representation of the user preference in this



(a) User satisfaction for the service composition result with respect to the number of candidate services

(b) User satisfaction for the service composition result with respect to the number of service classes

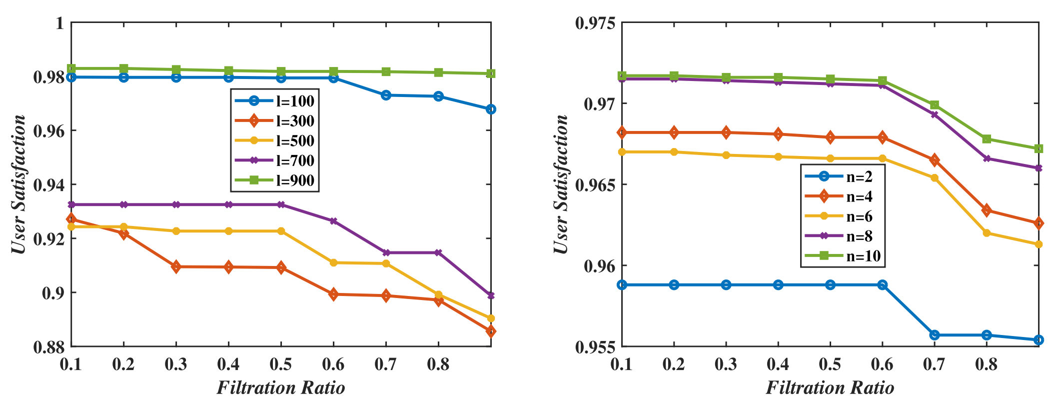
Fig. 2. Comparison results in terms of the user satisfaction.



(a) Computation time for service composition with respect to the number of candidate services

(b) Computation time for service composition with respect to the number of service classes

Fig. 3. Comparison results in terms of the computation time.



(a) User satisfaction for the service composition result with respect to the number of candidate services

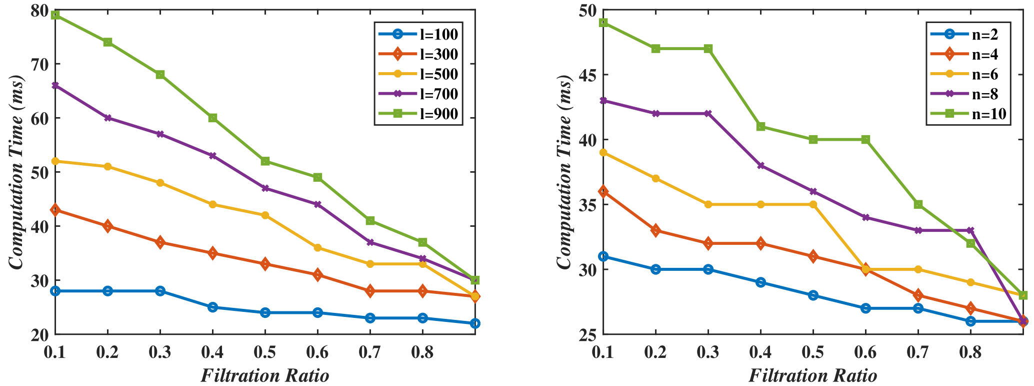
(b) User satisfaction for the service composition result with respect to the number of service classes

Fig. 4. Effect of the filtration ratio on the user satisfaction.

paper can improve the user satisfaction of the service composition. Furthermore, SIR^{-} is similar with SIR, which shows that the services filtered out according to the preference degree are not very critical.

(2) Considering the real time, Fig. 3 illustrates the results. SIR and SIR* are superior to SIR^{-} and MIP, which shows that the filtration can reduce the computation time and improve the real-time nature of the service composition. In addition, SIR is similar with SIR*, and SIR^{-} is similar with MIP, which shows that the computation of the user preference vector hardly affects the time cost of the service composition.

(3) Compared with SIR^{-}, SIR filters out the services with a low preference degree and improves the real-time nature of the service composition in CPSS, however, the utility values of these two approaches are similar. Compared with SIR*, SIR computes the user preference vector and improves the user satisfaction of the service composition in CPSS, however, the real time of these two approaches are similar. Compared with MIP, SIR improves the



(a) Computation time for the service composition result with respect to the number of candidate services

(b) Computation time for the service composition result with respect to the number of service classes

Fig. 5. Effect of the filtration ratio on the computation time.

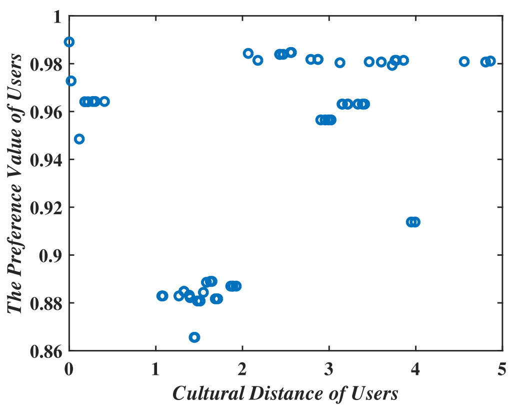


Fig. 6. Relationship between user cultural distance and preference.

user satisfaction and real-time nature of the service composition in CPSS. In a word, among the four approaches, the proposed SIR approach exhibits the best performance with regard to the user satisfaction and real time.

5. Conclusion

Cultural distance quantitatively describes the cultural difference in cross-culture study and cross-culture practice. To simply represent a user preference and guarantee the user satisfaction and real-time nature of the service composition in CPSS, this paper proposes a service composition approach based on cultural distance. The service composition approach in this paper uses the cultural distance to compute the preference degree and the preference vector and employs 0–1 MIP to select the best services. The service composition approach based on the cultural distance is evaluated experimentally using a real-world service QoS dataset and Hofstedes cultural dimension dataset. The experimental results show that the proposed approach can obtain composition services that satisfy specific requirements in a short time.

However, in this paper, the CPSS users are grouped according to regions because previous studies on cultural distance are at the regional/national level. This is not suitable for service composition in CPSS which requires a narrower grouping. Consequently, the next study in this area could focus on the presentation of the user preference in a narrower group.

Acknowledgment

This work was supported in part by the National Science Foundation of China (Grant No. 61472047).

References

[1] Z. Liu, D.S. Yang, D. Wen, W.M. Zhang, W. Mao, Cyber-physical-social systems for command and control, IEEE Intell. Syst. 26 (4) (2011) 92–96.
[2] F.Y. Wang, The emergence of intelligent enterprises: from CPS to CPSS, IEEE Intell. Syst. 25 (4) (2010) 85–88.
[3] G. Xiong, F. Zhu, X. Liu, X. Dong, W. Huang, S. Chen, K. Zhao, Cyber-physical-social system in intelligent transportation, IEEE/CAA J. Automat. Sin. 2 (3) (2015) 320–333.
[4] C. Chen, X. Liu, T. Qiu, A.K. Sangaiah, A short-term traffic prediction model in the vehicular cyber Cphysical systems, Future Gener. Comput. Syst. (2017).
[5] A. Khoshkbarforoushha, A. Khosravian, R. Ranjan, Elasticity management of streaming data analytics flows on clouds, J. Comput. System Sci. 89 (2017) 24–40.
[6] A. Khoshkbarforoushha, R. Ranjan, R. Gaire, E. Abbasnejad, L. Wang, A.Y. Zomaya, Distribution based workload modelling of continuous queries in clouds, IEEE Trans. Emerging Topics Comput. 5 (1) (2017) 120–133.
[7] Y. Ma, L. Wang, A.Y. Zomaya, D. Chen, R. Ranjan, Task-tree based large-scale mosaicking for massive remote sensed imageries with dynamic dag scheduling, IEEE Trans. Parallel Distrib. Syst. 25 (8) (2014) 2126–2137.
[8] L. Wang, Y. Ma, A.Y. Zomaya, R. Ranjan, D. Chen, A parallel file system with application-aware data layout policies for massive remote sensing image processing in digital earth, IEEE Trans. Parallel Distrib. Syst. 26 (6) (2015) 1497–1508.
[9] M. Menzel, R. Ranjan, L. Wang, S.U. Khan, J. Chen, Cloudgenius: a hybrid decision support method for automating the migration of web application clusters to public clouds, IEEE Trans. Comput. 64 (5) (2015) 1336–1348.
[10] S.G. Wang, A. Zhou, M.Z. Yang, L. Sun, C.H. Hsu, F.C. Yang, Service composition in cyber-physical-social systems, IEEE Trans. Emerging Topics Comput. 12 (99) (2017) 1.
[11] A. Klein, F. Ishikawa, S. Honiden, Towards network-aware service composition in the cloud, in: Proceedings of the international Conference on World Wide Web, 2012, pp. 959–968.
[12] S.G. Wang, L. Huang, C.-H. Hsu, F.C. Yang, Collaboration reputation for trustworthy Web service selection in social networks, J. Comput. System Sci. 82 (1) (2016) 130–143.
[13] D. Wu, B. Parsia, E. Sirin, J. Hendler, D. Nau, Automating DAML-S web services composition using SHOP2, in: Proceedings of the Second International Semantic Web Conference, Sanibel Island, FL, USA, pp. 195–210, October 2003.
[14] M. Reformat, D.M. Li, C. Ly, Approximate reasoning and semantic web services, in: Processing of the IEEE Annual Meeting of the Fuzzy Information, vol. 1, 2004, pp. 413–418.
[15] M.S. Lin, J.S. Xie, H.Q. Guo, H. Wang, Solving QoS-driven Web service dynamic composition as fuzzy constraint satisfaction, in: Proceedings of the IEEE International Conference on e-Technology, e-Commerce and e-ServicE, April 2005, pp. 9–14.
[16] J.J. Wen, J.L. Chen, Y. Peng, A user preference aware approach to Web services composition, High Technol. Lett. 17 (3) (2007) 243–248.
[17] H. Wang, J. Xu, P. Li, Incomplete Preference-driven Web Service Selection, in: Proceedings of the IEEE International Conference on Service Computing, 2008, pp. 75–82.

[18] G. Hofstede, Cultures and Organizations: Software of the Mind, Business Expert Press, 2010.
[19] Y. Ma, S.G. Wang, P.C.K. Hung, C.-H. Hsu, Q.B. Sun, F.C. Yang, A highly accurate prediction algorithm for unknown web service qos value, IEEE Trans. Serv. Comput. 9 (4) (2017) 511–523.
[20] S.G. Wang, L. Sun, Q.B. Sun, J. Wei, F.C. Yang, Reputation measurement of cloud services based on unstable feedback ratings, Internat. J. Web Grid Serv. 11 (4) (2015) 362–376.
[21] W.M. Lin, W.C. Dou, X.F. L, J. Chen, A history record-based service optimization approach for QoS-aware service composition, in: Proceedings of the IEEE International Conference on Web Services, 2011, pp. 666–673.
[22] L.Z. Zeng, B. Benatallah, M. Dumas, J. Kalagnanam, Q.Z. Sheng, Quality driven web services composition, in: Proceedings of the 12th International Conference on World Wide Web, Budapest, Hungary, 2003, pp. 411–421.
[23] Y. Gao, J. Na, B. Zhang, L. Yang, Q. Gong, Optimal web services selection using dynamic programming, in: Proceedings of the 11th IEEE Symposium on Computers and Communications, 2006, pp. 365–370.
[24] K. Ding, B. Deng, X.Y. Zhang, L. Ge, Optimization of service selection algorithm for complex event processing in Enterprise Service Bus platform, in: Proceedings of the 4th International Conference on Computer Science and Education, 2009, pp. 582–586.
[25] S.G. Wang, Z.B. Zheng, Q.B. Sun, H. Zou, F.C. Yang, Reliable web service selection via QoS uncertainty computing, Internat. J. Web Grid Serv. 7 (4) (2011) 410–426.
[26] L. Sun, S.G. Wang, J.L. Li, Q.B. Sun, QoS uncertainty filtering for fast and reliable web service selection, in: Proceedings of the IEEE International Conference on Web Services, 2014, pp. 550–557.
[27] S.M. Balakrishnan, A.K. Sangaiah, MIFIM: Middleware solution for service centric anomaly in future internet models, Future Gener. Comput. Syst. (2016).
[28] C.L. Wan, C. Ullrich, L.M. Chen, R. Huang, J.M. Luo, Z.Z. Shi, On solving QoS-aware service selection problem with service composition, in: Proceedings of the 7th International Conference on Grid and Cooperative Computing, 2008, pp. 467–474.
[29] M. Alrifai, T. Risse, Combining global optimization with local selection for efficient QoS-aware service composition, in: Proceedings of the 18th international conference on World Wide Web, 2009, pp. 881–890.
[30] M.L. Tang, L.F. Ai, A hybrid genetic algorithm for the optimal constrained web service selection problem in web service composition, in: Proceedings of the IEEE Congress on Evolutionary Computation, 2010, pp. 1–8.
[31] F. Li, Y. Huang, An web service selection optimization approach based on particle swarm optimization, in: Proceedings of the International Conference on Computer Design and Applications, vol. 2, 2010, pp. 477–481.
[32] S.G. Wang, Q.B. Sun, H. Zou, F.C. Yang, Particle swarm optimization with skyline operator for fast cloud-based web service composition, Mobile Netw. Appl. 18 (1) (2013) 116–121.
[33] M. Alrifai, D. Skoutas, T. Risse, Selecting skyline services for QoS-based web service composition, in: Proceedings of the 19th international conference on World Wide Web, 2010, pp. 11–20.
[34] N. Hiratsuka, F. Ishikawa, S. Honiden, Service selection with combinational use of functionally-equivalent services, in: Proceedings of the IEEE International Conference on Web Services, 2011, pp. 97–104.
[35] S.M. Balakrishnan, A.K. Sangaiah, Integrated QoUE and QoS approach for optimal service composition selection in internet of services, Multimedia Tools Appl. (2016) 1–28.
[36] H. Wang, J. Zhang, H. Wang, Y. Tang, G. Guo, Service selection based on similarity measurement for conditional qualitative preference, in: Proceedings of the IEEE International Conference on Service Computing, 2012, pp. 612–619.
[37] C. Boutilier, R.I. Brafman, C. Domshlak, H.H. Hoos, D. Poole, CP-nets: A tool for representing and reasoning withconditional ceteris paribus preference statements, J. Artificial Intelligence Res. 21 (1) (2011) 135–191.
[38] L. Tihanyi, D.A. Griffith, C.J. Russell, The effect of cultural distance on entry mode choice, international diversification, and MNE performance: a meta-analysis, J. Internat. Business Stud. 36 (3) (2005) 270–283.
[39] B. Kogut, H. Singh, The effect of national culture on the choice of entry mode, J. Internat. Bus. Stud. 19 (3) (1988) 411–432.
[40] Laurence A. Wolsey, Integer Programming, Vol. 20 (32), John Wiley and Sons Inc., New York, 1998, pp. 116–134.
[41] C. Bliek1, P. Bonami, A. Lodi, Solving mixed-integer quadratic programming problems with IBM-CPLEX: a progress report, 2014.
[42] D. Ardagna, B. Pernici, Adaptive service composition in flexible processes, JAMA 144 (18) (2007) 1540–1543.
[43] Z.B. Zheng, Y.L. Zhang, Michael R. Lyu, Investigating QoS of real-world web services, IEEE Trans. Serv. Comput. 7 (1) (2014) 32–39.

Shangguang Wang received his PhD degree at Beijing University of Posts and Telecommunications in 2011. He is an associate professor at the State Key Laboratory of Networking and Switching Technology (BUPT). He has published more than 100 papers, and played a key role at many international conferences, such as general chair and PC chair. His research interests include service computing, cloud computing, and mobile edge computing. He is a senior member of the IEEE, and the Editor-in-Chief of the International Journal of Web Science.

Yan Guo received her bachelor degree in mathematics and applied mathematics from Beijing University of Posts and Telecommunications in 2016. She is a first-year Master's student at Beijing University of Posts and Telecommunications, and will be a PhD candidate next term. Her research interests include service computing and mobile edge computing.

Yan Li received his PhD degree at Shanghai University of Finance and Economics in 2012. She is an associate professor of department of information management and information systems in Shanghai International Studies University. In the past 12 years, she devoted to the research of information management and system development. She has published more than 30 papers, and played a key role at many project, such as National Natural Science Foundation of China, National Planning Office of Philosophy and Social Science. Her research interests include service computing, cloud computing, financial crisis and cross-culture studies.

Ching-Hsien Hsu is a professor and the chairman in the CSIE department at Chung Hua University, Taiwan; He was distinguished chair professor at Tianjin University of Technology, China, during 2012–2016. His research includes high performance computing, cloud computing, parallel and distributed systems, big data analytics and intelligence. He has published 200 papers in these areas, including top journals such as IEEE TPDS, IEEE TSC, IEEE TCC, IEEE TETC, IEEE T-SUSC, IEEE Systems, IEEE Network, IEEE Communications, ACM TOMM. Dr. Hsu is serving as editorial board for a number of prestigious journals, including IEEE TSC, IEEE TCC, IJCS, JoCS. He has been acting as an author/co-author or an editor/co-editor of 10 books from Elsevier, Springer, IGI Global, World Scientific and McGraw-Hill. Dr. Hsu was awarded nine times distinguished award for excellence in research from Chung Hua University. He is vice chair of IEEE TCCLD, executive committee of IEEE TCSC, Taiwan Association of Cloud Computing and an IEEE senior member.

Please cite this article in press as: S. Wang, et al., Cultural distance for service composition in cyber–physical–social systems, Future Generation Computer Systems (2018), https://doi.org/10.1016/j.future.2018.06.012.