Decision Support Systems 72 (2015) 97–109

Contents lists available at ScienceDirect

Decision Support Systems

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

A semantic enhanced hybrid recommendation approach: A case study of e-Government tourism service recommendation system

Malak Al-Hassan, Haiyan Lu *, Jie Lu

Center for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia

ARTICLE INFO

Article history:
Received 27 March 2014
Received in revised form 15 December 2014
Accepted 1 February 2015
Available online 14 February 2015

Keywords:
Semantic enhanced recommender systems
Collaborative filtering
Semantic similarity
e-Government tourism services

ABSTRACT

Recommender systems are effectively used as a personalized information filtering technology to automatically predict and identify a set of interesting items on behalf of users according to their personal needs and preferences. Collaborative Filtering (CF) approach is commonly used in the context of recommender systems; however, obtaining better prediction accuracy and overcoming the main limitations of the standard CF recommendation algorithms, such as sparsity and cold-start item problems, remain a significant challenge. Recent developments in personalization and recommendation techniques support the use of semantic enhanced hybrid recommender systems, which incorporate ontology-based semantic similarity measure with other recommendation approaches to improve the quality of recommendations. Consequently, this paper presents the effectiveness of utilizing semantic knowledge of items to enhance the recommendation quality. It proposes a new Inferential Ontology-based Semantic Similarity (IOBSS) measure to evaluate semantic similarity between items in a specific domain of interest by taking into account their explicit hierarchical relationships, shared attributes and implicit relationships. The paper further proposes a hybrid semantic enhanced recommendation approach by combining the new IOBSS measure and the standard item-based CF approach. A set of experiments with promising results validates the effectiveness of the proposed hybrid approach, using a case study of the Australian e-Government tourism services.

1. Introduction

Recommendation systems (RSs) are known as the most popular applications of Web personalization. The RSs aim to provide users with personalized services or products that are relevant to their needs and interests. Recent research studies show that existing personalized online services adopt several RSs approaches. These approaches are classified into four main categories, including content-based (CB) filtering, collaborative filtering, knowledge-based filtering and hybrid recommendation [1,10,40]. Although the CB filtering and CF approaches are the most popular in practical applications, both of them suffer from several limitations [23]. For instance, the CB filtering approach tends to result in overspecialization in which the diversity in the recommendation results eventually vanishes [35], while the CF approach suffers from the data sparsity problem which occurs when the ratings obtained are few compared to the number of available items. Moreover, both the CB filtering and CF approaches have difficulty offering accurate recommendations for new items as there is usually little available information about new items.

On the other hand, hybrid recommendation approaches, as a combination of two or more recommendation approaches, have been proposed to overcome the main limitations of traditional recommendation approaches and improve the quality of the recommendations offered [1,11,35]. Most of the existing hybrid recommendation approaches combine conventional CF approaches with other approaches such as CB filtering, since CF approaches are generally known to be the most promising approaches in the recommendation systems domain [1,23,45]. There has been considerable research into the hybridization of CF-based algorithms and improvements on the prediction accuracy have been made [11,12,45,50]. However, obtaining better prediction accuracy and overcoming the main limitations of the standard CF recommendation approaches remain open challenges, as no cure-all solution is yet available and many research studies have been working on solutions for each of the CF limitations [12,45].

These challenges, combined with the increasing popularity of semantic web technologies, have inspired a growing interest in semantic enhanced recommendation approaches. These approaches mainly incorporate the semantic knowledge of users and/or items within the recommendation process of conventional CF-based algorithms to accurately evaluate similarity of items and to enhance recommendation accuracy [8,36]. Most of these approaches rely on semantic knowledge extracted from a target ontology that includes the direct hierarchical (i.e. taxonomical) relationships of items and/or their shared attributes.

* Corresponding author. Tel.: +61 2 95141758; fax: +61 2 95144535 (office).
E-mail address: haiyan.lu@uts.edu.au (H. Lu).

http://dx.doi.org/10.1016/j.dss.2015.02.001
0167-9236/

However, evaluating the similarity of items is limited since ontological relationships[1] that connect the items in a target ontology are not usually handled very well [7,25,26,33,44]. Such relationships may include complex relationships between instances (i.e. items[2]) that consist of two or more relationships [3].

Even though progress is being made in developing efficient strategies for estimating the semantic similarity of items in semantic enhanced recommendation systems, this work is still in an early stage and more research is needed [3,8,13,15,25,44]. This observation, combined with the specific features of service items (e.g. services are multi-relation and highly interrelated) in a specific domain, such as services in government, has motivated the research presented in this paper. Consequently, this paper presents two contributions (i) it proposes a new IOBSS measure to evaluate the semantic similarity between instances in specific domain ontology and (ii) it develops a new semantic enhanced hybrid recommendation approach that combines the new semantic similarity measure and the item-based CF to generate accurate recommendations.

The effectiveness of the new semantic-based hybrid recommendation approach has been validated through a case study of the Australian e-Government tourism service. It achieves highly effective results in terms of prediction accuracy of generated recommendations and in alleviating data sparsity and cold-start new item problems.

The rest of the paper is organized as follows. Section 2 presents the related work. Section 3 presents the concept and calculation procedure of the new IOBSS measure with an illustrative example. Section 4 presents the new semantic-based enhanced hybrid recommendation approach, its workflow and its computation recommendation procedure. An experimental study of the new hybrid recommendation approach, in the context of recommending e-Government tourism services, is illustrated in Section 5. Finally, Section 6 concludes the paper and highlights potential future work.

2. Related work

This section reviews the literature related to this study, including semantic-based similarity and semantic-based recommendation systems.

2.1. Semantic similarity approaches

Computing semantic similarity among ontological concepts with regard to their positions in a particular taxonomy has been studied in the last decade. Semantic similarity [5,42] approaches can be classified into three main categories, namely (i) distance-based approaches, (ii) information content (IC) based approaches, and (iii) hybrid approaches.

Distance-based approaches measure the similarity between concepts in a specific taxonomy according to the distance/edge length between concepts. One of the most well-known distance-based measures is the shortest path-based approach, where the shorter the path between two concepts, the more similar they are [37]. Generally, distance-based approaches are highly dependent on the construction of the taxonomy [5,41]. The main drawback of these approaches is that they consider that the edges in a taxonomy structure represent uniform distances.

The IC-based approaches compute the similarity between two concepts based on the extent to which they share information; the more information two concepts share in common, the more similar they are [38]. These approaches avoid the unreliability of edge distance measure because they require less information about the structure of a taxonomy. According to Resnik [38], the IC of two concepts can be measured with respect to the IC of their least common ancestor in a specific taxonomy [38]. Lin [27] enhanced Resnik's IC measure based on the assumption of commonality information, i.e. the similarity between two concepts relies on the extent to which they share information. Based on Lin's assumption, the IC value of two concepts can be measured as the IC of compared concepts themselves in addition to the IC of their least common ancestor [27]. The IC-based approaches obtain the IC values of concepts by combining the knowledge of the hierarchical structure of concepts with statistics on their actual usage and are usually computationally expensive. Seco et al. [41] proposed a wholly intrinsic measure for computing the IC of a specific concept. The new metric depends on the hierarchal structure (i.e. taxonomy) alone without the need to involve statistics [41].

The hybrid semantic similarity approaches combine the features of edge-based and IC-based approaches, with the aim of producing more accurate similarity measure [22,30,42,47,49]. For instance, Jiang & Conrath [22] developed a hybrid model that uses the IC-based approach to enhance the distance-based approach. Their approach takes into account the factors of local density, node depth and link types [22]. Seddiqui & Aono [42] proposed a hybrid similarity measure which combines the intrinsic IC-based approach presented by Seco [41] and the content of concepts (attributes and relations). Their new measure is used to compute similarity between concepts for the purpose of partitioning a large taxonomy of ontology.

All the aforementioned approaches are mainly designed for computing similarity between concepts based on the relative positions of concept nodes in a semantic network[3] [16,39], with some exceptions, as in Maedche and Zacharias and Seddiqui and Aono [42,30]. The semantic similarity measures presented in Maedche and Zacharias and Seddiqui and Aono [42,30] compute similarity between concepts in the ontology environment. Unlike semantic networks [32], where concepts are only linked by “is-a” relations, ontologies are more complex and concepts are defined with sufficient datatype properties, object properties, restrictions, etc. The knowledge of content i.e. attributes and relationships can be regarded as crucial information for identifying concepts and can significantly influence similarity estimations between concepts. Therefore, existing semantic similarity measures which are designed for semantic networks can be difficult to apply to ontologies, as they cannot capture the semantics represented in ontology. Although some studies consider the content knowledge of concepts for similarity computation, they only focus on explicit relationships[4] and pay little attention to the attributes and indirect relationships between concepts [2,15,42]. Accordingly, this study develops a new approach to estimate similarity between ontological instances based on rich semantics that can be captured from ontology by taking into account not only the items' hierarchal relationships but also their ontological relationships. Moreover, a new IOBSS measure is proposed that can be utilized in this study to improve recommendation accuracy.

2.2. Semantic-based recommendation systems

Ontology is considered to be a knowledge base that enables systems to interpret, process and share information effectively [4,29]. The merit of ontology lies in its ability to provide a clear conceptual description of relationships between entities (i.e. concepts) in a specific domain. Ontology aims to support the rich variety of semantic relations among entities in a specific domain, which in turn distinguishes it from other types of representation, such as keyword-based representation [4].

Semantic-based recommendation systems have recently been developed that make use of semantics based on ontology and semantic reasoning in the recommendation process to specifically improve the

[1] Ontological relationships refer to semantic associations that link instances, examples of such relationships can be seen in object properties in OWL. Links between instances that consist of two or more relationships represent complex relationships.

[2] Henceforth, item and instance are used interchangeably.

[3] Semantic network is a graphic notation for representing knowledge in patterns of interconnected nodes (e.g. concepts) and arcs. A typical example of a semantic network is WordNet.

[4] Explicit relationships refer to taxonomical (i.e. hierarchal) relationships of instances and their attributes, such relationships also called direct relationships.

similarity estimations used in traditional CB filtering and CF approaches [36]. Based on a broad literature review, the incorporation of semantic knowledge that is formalized in the form of ontology with CF-based recommendation approaches can be summarized into three categories: (i) incorporate semantic knowledge of considered content (i.e. items) with the traditional item-based CF approach [33]; (ii) incorporate semantic knowledge of items with the user-based CF approach [14,28,31,43,48], and (iii) combine the user-based CF approach with the semantic enhanced CB filtering approach [7].

Two existing hybrid recommendation approaches that use semantic similarity with the traditional CF approaches are closely related to this study: (i) a semantically enhanced collaborative filtering (SECF) approach proposed by Mobasher et al. [33] and (ii) a collaborative filtering with ontology-based (CFO) user profiles approach proposed by Sieg et al. [44]. The aforementioned approaches resort to semantic knowledge of items to improve the prediction accuracy of the standard CF recommendation algorithms, as well as to deal with the sparsity and cold-start new items problems. However, these approaches use the semantic knowledge of items that is extracted from item descriptions (including datatype and object properties), as in the SECF approach, or hierarchical relationships of items, as in the CFO approach. Even though the use of semantic knowledge has improved the recommendation process of the aforementioned approaches, this source of knowledge is limited and not informative in the evaluation of instances since ontological relationships between instances are not usually handled very well [7,15,28,48].

This paper proposes a new semantic-based enhanced hybrid recommendation approach that combines item-based CF similarity and an IOBSS measure to improve the prediction accuracy of recommendations. Details of the new approach will be presented in the following sections.

3. Inferential ontology-based semantic similarity

This section first introduces an ontology model and definition, and then describes the proposed inferential ontology-based semantic similarity measure.

3.1. Domain ontology model

According to Gruber [17], an ontology is a formal representation of the world. It defines a set of representational primitives that are relevant for modeling a domain of knowledge or discourse. These primitives typically consist of a set of concepts or entities within a domain, relationships among these concepts, and attributes that distinguish each concept [17]. A formal definition of an ontology structure as introduced by Maedche and Zacharias [33] is given below:

Definition 1. Ontology

An ontology structure is a six-tuple $O{:} = <C, P, A, H^c, prop, att>$, where C represents the concept set defined in O; P is a set of relationships defined in O, each $(p \in P)$ has a domain and range which are at least one concept of the set C; A is a set of attributes defined in O; H^c is a directed transitive relation $H^c \subset C \times C$ which is also called concept taxonomy, $H^c(c_2, c_1)$ means c_2 "*is-a*" c_1, or c_2 is a sub-concept of c_1; *prop* is a function, i.e. $\mathrm{prop}: P \rightarrow C \times C$, that relates concepts non-taxonomically, e.g. the function $\mathrm{prop}(p_1) = (c_1, c_2)$ means that the concept c_1 is related to concept c_2 through p_1; and *att* is a function, i.e. $\mathrm{att}: A \rightarrow C$, that relates concepts with literal values such as string, integer, boolean, etc.

In a domain ontology, concepts are linked through two kinds of relationships. One is the asserted relationships which are direct relationships between concepts that are defined by the developers of the ontology This kind of relationships includes (i) the taxonomical or hierarchical relationships, denoted by H^c as defined in Definition 1; (ii) the associations between concepts (e.g. object properties) and (iii) the attributes as special relationships of concepts (e.g. datatypes). The other type is the implicit relationships (i.e. inferred) which are the indirect relationships obtained through reasoning the asserted relationships [20]. Furthermore, ontology also includes instances of concepts, referred to as ontological instances. Based on the relationships between concepts, the relationships between instances will be automatically established when the instances are instantiated from corresponding concepts.

3.2. Terms needed to define the new semantic similarity measure

This section first introduces some terms that are needed to describe the new semantic similarity measure, including an associate relationship, an associate of an instance, an associate network of an instance and the common associate set of two instances, and then presents the IOBSS measure. Lastly, the IOBSS calculation procedure is presented using an illustrative example.

3.2.1. Association

Definition 2. Association

Association is a link between two ontological instances through an object property. Two instances are associates if they are linked through an object property in a given OWL ontology.

An association has three features: (i) self-determination, i.e., one instance is an associate of itself, (ii) reversibility, i.e. if I_x has an association with I_y via an object property *op*, denoted as $I_x \rightarrow^{op} I_y$, I_y will have an inverse association with I_x, denoted as $I_y \rightarrow^{op^{-1}} I_x$; and (iii) transitivity, i.e. for a given instance I_x, if an instance I_z is an associate of I_y which is an associate of I_x, then this instance (I_z) is also an associate of the given instance (I_x). In other words, an associate's associate is also an associate.

3.2.2. Associate network of an ontological instance

The associate network of an instance is a network of instances that are directly or indirectly linked with this instance through its object properties (i.e. associations).

Definition 3. Associate network

An associate network of an ontological instance I_x in regard to ontology O $(I_x \subset I)$ is defined as a four-tuple, denoted as $\mathrm{AssN}_{I_x}: <I_{ep1}, I_{ep2}, \mathrm{OPC}, \mathrm{Closeness}>$, where I_{ep1}, $I_{ep2} \subset I$ are two sets of instances whose elements are associated through object properties; $\mathrm{OPC} = \left\{op_i^k | \, k \in [1, N],\ i \in \left[1, N_{op_x}^k\right]\right\}$ is a collection of object properties that form the associate network of I_x, where k indicates how far an instance from the root instance in the hierarchical tree, op_i^k is the i^{th} object property at k^{th} level of the associate network of I_x, $N_{op_x}^k$ is the number of distinct object properties at the k^{th} level; N is the maximum number of associations in the associate chains of I_x; and $\mathrm{Closeness} \subset R$ is a set of real numbers indicating how close an instance $I_{x_{ij}}^k \subset I_{ep2}$ $\left(k \in [1, N], i \in \left[1, N_{op_x}^k\right], j \in \left[1, N_{ins,i}^k\right]\right)$ is to the root instance in the hierarchy of associate network of I_x, where $N_{ins,i}^k$ is the number of instances that are introduced by the object property op_i^k at the k^{th} level.

An associate network can also be represented as a tree structure in which a node represents an instance, an edge represents an association (through an object property), two directly linked nodes are associates of each other, the edge sequence that links the instance $I_{x_{ij}}^k \subset I_{ep2}$ from the root instance I_x is the associate-chain. The length of the associate-chain represents the depth of instance $I_{x_{ij}}^k$ in the tree hierarchy and determines the closeness of instance $I_{x_{ij}}^k$ to the root instance I_x, where $I_{x_{ij}}^k$ denotes the j^{th} associate of I_x at level k. Figs. 1 and 2 illustrate the associate networks (in tree structure) for instances I_x and I_y, respectively.

To describe an associate network of an instance, we introduce some symbols to represent the instances and object properties.

For the associate network of I_x, $AssN_{I_x}$ in Fig. 1, we have the maximum closeness levels $N_x = 4$, $op^k_{x_i}$ is the i^{th} association (object propriety) at the k^{th} closeness level, $i = 1, 2, \ldots, N^k_{op_x}$, where $N^k_{op_x}$ is the number of object properties at the k^{th} closeness level, and $I^k_{x_{ij}}$ for the j^{th} associate of I_x at the k^{th} closeness level that is introduced by the association $op^k_{x_i}$, $k \in [1, 4]$, $i \in \left[1, N^k_{op_x}\right]$, $j \in [1, N^k_{ins,i}]$. For example, the instance $I^2_{x_{22}}$ indicates that this instance is an associate of I_x at the closeness level 2 and it is the second associate introduced by the object property $op^2_{x_2}$.

For the associate network of I_y, $AssN_{I_y}$ in Fig. 2, we have the maximum closeness levels $N_y = 3$, $op^k_{y_i}$ is the i^{th} association at the k^{th} closeness level, $i = 1, 2, \ldots, N^k_{op_y}$, where $N^k_{op_y}$ is the number of object properties at the k^{th} closeness level, and $I^k_{y_{ij}}$ for the j^{th} associate of I_y at the k^{th} closeness level that is introduced by the association $op^k_{y_i}$, $k \in [1, 3]$,, $i \in \left[1, N^k_{op_y}\right]$, $j \in [1, N^k_{ins,i}]$. For example, the instance $I^3_{y_{11}}$ indicates that this instance is an associate of I_y at the closeness level 3 and it is the first associate introduced by the association $op^3_{y_1}$.

Table 1 lists the parameters, as defined in Definition 3, for the two associate networks of I_x and I_y as shown in Figs. 1 and 2.

Based on Definition 3, we can extract the features of an associate network of an instance I_x as follows:

(1) If I_x has no object property, its associate network is itself.
(2) There exists a function AssF^k that can retrieve the direct associates of all instances through $op^k_{x_i}$, $i \in \left[1, N^k_{op_x}\right]$, at a closeness level k in a given ontology. All these associates become instances at the level $(k + 1)$.
(3) At the k^{th} closeness level, if the number of instances is $\mathrm{Num}^k_{\mathrm{inst}}$ and their numbers of object properties are $\left(\mathrm{numProp}^k_1, \mathrm{numProp}^k_2, \ldots, \mathrm{numProp}^k_{\mathrm{Num}^k_{\mathrm{inst}}}\right)$, then the total number of object properties $\left(\sum_i^{\mathrm{Num}^k_{\mathrm{inst}}} \mathrm{numProp}^k_i\right)$ is the number of the instances at the level $(k + 1)$.

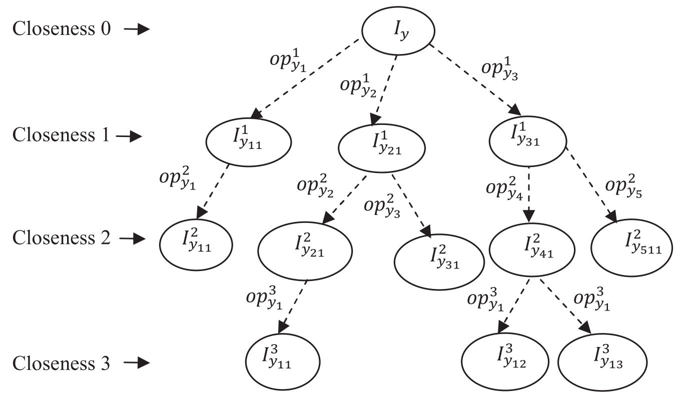

Fig. 2. The associate network of instance I_y.

(4) An instance $I^k_{x_{ij}}$ is in the associate network of instance I_x if and only if this instance $\left(I^k_{x_{ij}}\right)$ is the root instance (I_x) or an instance that has an association with another instance in the network.
(5) Each associate of an instance, say $I^k_{x_{ij}}$, has one direct predecessor which introduces it into the associate network of root instance (I_x) through an object property. It may have a number of indirect predecessors depending on its closeness level. An instance $I^k_{x_{i,j}}$, for example, is an associate at the closeness level k, has $(k - 1)$ predecessors, which is denoted by $\mathrm{supAss}^q_{I^k_{x_{i,j}}}$, where q is the level number $0 \le q < k$.
(6) If an association $op^k_{x_i}$, where $op^k_{x_i} \in \mathrm{OPC}$ of AssN_{I_x}, is in the associate network, its reversed association must be excluded from the network to avoid an infinite loop.

3.2.3. Common associate pair set of two ontological instances

Definition 4. Common associate pair set

A common associate pair set of two ontological instances I_x and I_y, i.e. $\mathrm{CAPSet}_{I_x I_y}$, is defined as a set of associate pairs that satisfy the following conditions: (i) the first and second elements of each pair are instances from the associate networks of I_x and I_y, respectively; (ii) the two

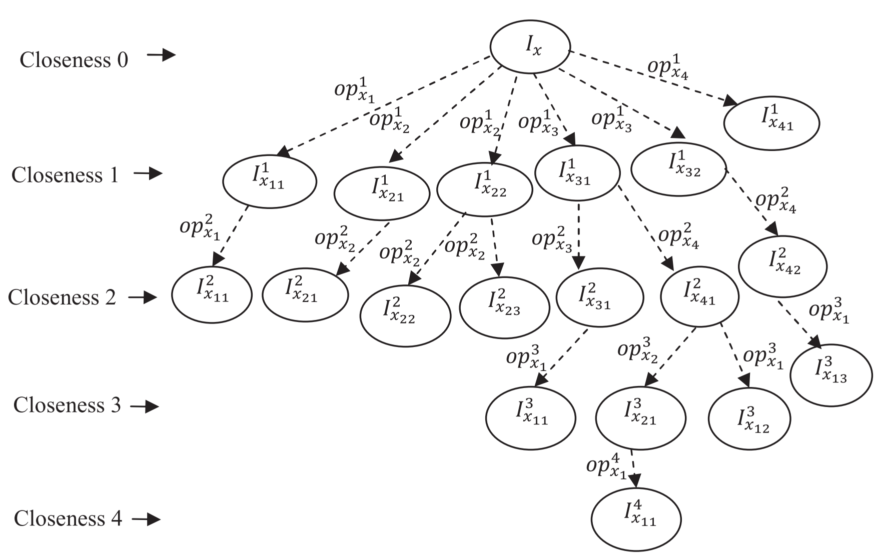

Fig. 1. The associate network of instance I_x.

Table 1
Parameters of the associate networks of instances I_x and I_y.

Root instance	N_x	$\|I_{ep2}\|$[5]	$\|OPC\|$	$N_{op_x}^k$, $k \in [1, N]$		$k \in [1, N]$, $i \in \left[1, N_{op_x}^k\right]$	
				k	$N_{op_x}^k$	op_i^k	$N_{ins,i}^k$
I_x	4	19	11	1	4	i = 1	1
						i = 2	2
						i = 3	2
						i = 4	1
				2	4	i = 1	1
						i = 2	3
						i = 3	1
						i = 4	2
				3	2	i = 1	3
						i = 2	1
				4	1	i = 1	1
Root instance	N_y	$\|I_{ep2}\|$	$\|OPC\|$	$N_{op_y}^k$, $k \in [1, N]$		$k \in [1, N]$, $i \in \left[1, N_{op_y}^k\right]$	
				k	$N_{op_y}^k$	op_i^k	$N_{ins,i}^k$
I_y	3	12	9	1	3	i = 1	1
						i = 2	1
						i = 3	1
				2	5	i = 1	1
						i = 2	1
						i = 3	1
						i = 4	1
						i = 5	1
				3	1	i = 1	3

[5] Operator | | denotes the cardinality of a set, i.e. the number of elements in a set.

elements of each pair have the same closeness level; (iii) the two elements of each pair are introduced into the corresponding associate network by their direct predecessors through the same object property; and (iv) the direct predecessors of the two elements of each pair must be a pair in the $\mathrm{CAPSet}_{I_x I_y}$.

The m^{th} individual element in the $\mathrm{CAPSet}_{I_x I_y}$, is denoted by $p_m = \left(I_{x_{wi}}^k, I_{y_{wj}}^k\right)$, where $w \in [1,\ N_{comop}^k]$ and N_{comop}^k is the number of common object properties at the closeness level k in both associate networks of I_x and I_y; $I_{x_{wi}}^k$ represents the i^{th} associate at the k^{th} level in the $AssN_{I_x}$, which is introduced into this network through an object property $op_{x_w}^k$, while $I_{y_{wj}}^k$ represents the j^{th} associate at the k^{th} level in the $AssN_{I_y}$, which is introduced into this network through an object property $op_{y_w}^k$, $i \in \left[1,\ N_{x_w}^k\right]$ and $j \in \left[1,\ N_{y_w}^k\right]$ with $N_{x_w}^k$ and $N_{y_w}^k$ being the number of instances introduced by the common object property op_w^k ($op_w^k = op_{x_w}^k = op_{y_w}^k$) at the closeness level k in the associate networks of I_x and I_y, respectively.

The predecessor of the m^{th} element in the $\mathrm{CAPSet}_{I_x I_y}$ is a pair of associates that introduces the $I_{x_{wi}}^k$ and $I_{y_{wj}}^k$ instances into AssN_{I_x} and AssN_{I_y}, respectively, through the common object property op_w^k and is denoted by $\mathrm{supAss}^{k-1}_{I^k_{x_{wi}}, I^k_{y_{wj}}}$.

As shown in Figs. 1 and 2, given that $op_{x_1}^1$ and $op_{y_1}^1$ are common object properties between the instances $I_{x_{11}}^1$ and the $I_{y_{11}}^1$, the pair $(I_{x_{11}}^1, I_{y_{11}}^1)$ is an associate pair in the $\mathrm{CAPSet}_{I_x I_y}$ because of the following factors: (i) $I_{x_{11}}^1$ and $I_{y_{11}}^1$ is from the associate networks of I_x and I_y, respectively, as shown in Figs. 1 and 2; (ii) they have the same closeness level $k = 1$; (iii) they are introduced into the corresponding associate network via the same object property $op_{x_1}^1 = op_{y_1}^1$, and (iv) the associate pair (I_x, I_y), whose elements I_x and I_y are the direct predecessors of $I_{x_{11}}^1$ and $I_{y_{11}}^1$ respectively, is an element in the $\mathrm{CAPSet}_{I_x I_y}$.

3.2.4. Weight factor

Each element (i.e. associate pair) in the $\mathrm{CAPSet}_{I_x I_y}$ has a weight factor that indicates how much the similarity of each element contributes to the semantic similarity of the two given instances I_x and I_y.

Definition 5. Weight factor

A weight factor for the m^{th} element in the $\mathrm{CAPSet}_{I_x I_y}$, denoted as F_m, is defined as follows:

$$F_m = \begin{cases} \frac{1}{3}\left(\frac{1}{3^k} * \frac{1}{\ell} * \frac{1}{\mathbb{C}}\right) & \text{non-leaf nodes} \\ \frac{1}{2}\left(\frac{1}{3^k} * \frac{1}{\ell} * \frac{1}{\mathbb{C}}\right) & \text{leaf nodes} \end{cases}, \tag{1}$$

where $m \in [1, M]$ and M is the number of elements in the $\mathrm{CAPSet}_{I_x I_y}$, k is the closeness level of the m^{th} element, ℓ and $\mathbb{C}$ are two parameters.

The rationale behind treating the weight factor of the given m^{th} element in the $\mathrm{CAPSet}_{I_x I_y}$ differently, as defined in Eq. (1), is that the leaf nodes have no further object properties to be evaluated, so that the weight of leaf nodes to the similarity is influenced by two sources of information (the structure and datatype property); while the non-leaf nodes have object properties that lead to further exploration of the associates so their weight factors are determined using three factors including their object properties, datatype properties and taxonomical relationships.

The two parameters, ℓ and $\mathbb{C}$, are determined as follows:

Determination of the ℓ parameter

The ℓ parameter for the m^{th} element, i.e. $p_m = \left(I_{x_{wi}}^k, I_{y_{wj}}^k\right)$, is defined as the product of the numbers of associates of all predecessors of the m^{th} element, which are introduced by the common object properties of all predecessors of $I_{x_{wi}}^k$ and $I_{y_{wj}}^k$ and can be calculated as follows:

$$\ell\left(I_{x_{wi}}^k, I_{y_{wj}}^k\right) = \prod_{q=k-1}^{0}\left(\left|R\left(\mathrm{supAss}^q_{I^k_{x_{wi}}}, op_w^{q+1}\right)\right| * \left|R\left(\mathrm{supAss}^q_{I^k_{y_{wj}}}, op_w^{q+1}\right)\right|\right), \tag{2}$$

where

- $R\left(\mathrm{supAss}^q_{I^k_{x_{wi}}}, op_w^{q+1}\right)$ denotes the set of associates that are introduced by $\mathrm{supAss}^q_{I^k_{x_{wi}}}$ into $AssN_{I_x}$ through op_w^{q+1}, where $\mathrm{supAss}^q_{I^k_{x_{wi}}}$ is the direct predecessor associate of $I_{x_{wi}}^k$ at level q, and the size of this set is $\left|R\left(\mathrm{supAss}^q_{I^k_{x_{wi}}}, op_w^{q+1}\right)\right|$;
- $R\left(\mathrm{supAss}^q_{I^k_{y_{wj}}}, op_w^{q+1}\right)$ denotes the set of associates that are introduced by $\mathrm{supAss}^q_{I^k_{y_{wj}}}$ into AssN_{I_y} through op_w^{q+1}, where $supAss^q_{I^k_{y_{wj}}}$ is the direct predecessor associate of $I_{y_{wj}}^k$ at level q, and the size of this set is $\left|R\left(supAss^q_{I^k_{y_{wj}}}, op_w^{q+1}\right)\right|$;
- $supAss^q_{I^k_{x_{wi}}} \in prd_{inst}\left(I_{x_{wi}}^k\right)$, where $prd_{inst}\left(I_{x_{wi}}^k\right) = \left\{supAss^q_{I^k_{x_{wi}}} \mid \forall q \in [0, k-1]\right\}$ is the set of all the predecessor associates of instance $I_{x_{wi}}^k$;
- $supAss^q_{I^k_{y_{wj}}} \in prd_{inst}\left(I_{y_{wj}}^k\right)$, where $prd_{inst}\left(I_{y_{wj}}^k\right) = \left\{supAss^q_{I^k_{y_{wj}}} \mid \forall q \in [0, k-1]\right\}$ is the set of all the predecessor associates of instance $I_{y_{wj}}^k$; and

– op_w^{q+1} is the object property of instances at the q^{th} level that introduces the associates at the level $q^{th}+1$.

As a special case, the ℓ parameter of $(I_x, I_y) = \left(I^0_{x_{wi}}, I^0_{y_{wj}}\right)$ is set to one, i.e. $\ell(I_x, I_y) = 1$.

Determination of the $\mathbb{C}$ parameter

The $\mathbb{C}$ parameter of the m^{th} element in the $CAPSet_{I_xI_y}$, i.e. the associate pair $\left(I^k_{x_{wi}}, I^k_{y_{wj}}\right)$, is defined as the product of numbers of common object properties of its all predecessor pairs and can be calculated as follows:

$$\mathbb{C}\left(I^k_{x_{wi}}, I^k_{y_{wj}}\right) = \prod_{q=k-1}^{0} N_{cop}(p^q), \tag{3}$$

where

– $N_{cop}(p^q)$ denotes the number of common object properties with respect to a predecessor pair (p^q), i.e. $\left(I^q_{x_{wi}}, I^q_{y_{wj}}\right)$, at the q^{th} closeness level of the associate pair $\left(I^k_{x_{wi}}, I^k_{y_{wj}}\right)$;
– $p^q \in prd_{pair}\left(I^k_{x_{wi}}, I^k_{y_{wj}}\right)$, where $prd_{pair}\left(I^k_{x_{wi}}, I^k_{y_{wj}}\right) = \left\{supAss^q_{I^k_{x_{wi}}, I^k_{y_{wj}}} \middle| \forall q \in [0, k-1]\right\}$ is the set of all the predecessor pairs of the element $\left(I^k_{x_{wi}}, I^k_{y_{wj}}\right)$, and $supAss^q_{I^k_{x_{wi}}, I^k_{y_{wj}}}$ is the predecessor pair of the $\left(I^k_{x_{wi}}, I^k_{y_{wj}}\right)$ pair at the closeness level q; $supAss^0_{I^k_{x_{wi}}, I^k_{y_{wj}}}$ represents the given pair of instances (i.e. I_x and I_y).

As a special case, the $\mathbb{C}$ parameter of $(I_x, I_y) = \left(I^0_{x_{wi}}, I^0_{y_{wj}}\right)$ is set to one, i.e. $\mathbb{C}(I_x, I_y) = 1$.

As an example of calculating the weight factor of an associate pair, consider the associate pair $\left(I^2_{x_{41}}, I^2_{y_{41}}\right)$ in the $CAPSet_{I_xI_y}$, the common object property op^k_w between $I^2_{x_{41}}$ and $I^2_{y_{41}}$ is $op^2_3 = op^2_{x_4} = op^2_{y_4}$, as shown in Figs. 1 and 2. In view of that, the ℓ and $\mathbb{C}$ parameters of the given associate pair $\left(I^2_{x_{41}}, I^2_{y_{41}}\right)$ are calculated as follows:

$$\begin{aligned}
&\ell\left(I^2_{x_{41}}, I^2_{y_{41}}\right) : prd_{inst}\left(I^2_{x_{41}}\right) = \left\{I^1_{x_{31}}, I^0_x\right\}; prd_{inst}\left(I^2_{y_{41}}\right) = \left\{I^1_{y_{31}}, I^0_y\right\}\\
&\text{Thus, } \ell\left(I^2_{x_{41}}, I^2_{y_{41}}\right) = \prod_{q=k-1}^{0}\left(\left|R\left(supAss^q_{I^2_{x_{41}}}, op_w^{q+1}\right)\right| * \left|R\left(supAss^q_{I^2_{y_{41}}}, op_w^{q+1}\right)\right|\right)\\
&= \left(\left|R\left(I^1_{x_{31}}, op^2_3\right)\right| * \left|R\left(I^1_{y_{31}}, op^2_3\right)\right|\right) * \left(\left|R\left(I^0_x, op^1_3\right)\right| * \left|R\left(I^0_y, op^1_3\right)\right|\right)\\
&= (1*1)*(2*1) = 2
\end{aligned}$$

$$\begin{aligned}
&\mathbb{C}\left(I^2_{x_{41}}, I^2_{y_{41}}\right) : prd_{pair}\left(I^2_{x_{41}}, I^2_{y_{41}}\right) = \left\{\left(I^1_{x_{31}}, I^1_{y_{31}}\right), \left(I^0_x, I^0_y\right)\right\}\\
&\text{Thus, } \mathbb{C}\left(I^2_{x_{41}}, I^2_{y_{41}}\right) = \prod_{q=1}^{0} N_{cop}(p^q) = N_{cop}\left(I^1_{x_{31}}, I^1_{y_{31}}\right) * N_{cop}\left(I^0_x, I^0_y\right)\\
&= 1*3 = 3
\end{aligned}$$

Since the instances of the pair $\left(I^2_{x_{41}}, I^2_{y_{41}}\right)$ are not leaf-nodes, the weight factor of this pair is calculated as: $F_{\left(I^2_{x_{41}}, I^2_{y_{41}}\right)} = \frac{1}{3}\left(\frac{1}{3^k} * \frac{1}{\ell} * \frac{1}{\mathbb{C}}\right) = \frac{1}{3}\left(\frac{1}{3^2} * \frac{1}{2} * \frac{1}{3}\right) = \frac{1}{162}$

3.3. Definition of the semantic similarity (IOBSS) measure

Given two instances I_x and I_y, the new semantic similarity (IOBSS) measure of I_x and I_y, denoted as $OntSemSim(I_x, I_y) : I \times I \to [0, 1]$, can be expressed as follows:

$$OntSemSim\left(I_x, I_y\right) = \sum_{m=1}^{M} F_m * \left(Sim_{str}(p_m) + Sim_{dt}(p_m)\right), \tag{4}$$

where, F_m is the weight factor of the m^{th} element in the $CAPSet_{I_xI_y}$, which is determined using Eq. (1); M is the number of elements in the $CAPSet_{I_xI_y}$; $Sim_{Str}(p_m)$ and $Sim_{dt}(p_m)$ is the structure-based similarity and datatype-based similarity of the m^{th} element, respectively. The structure-based similarity and datatype-based similarity are illustrated in the following sections.

3.4. Structure-based similarity of two ontological instances

The structure-based similarity between two ontological instances compares two instances in terms of concepts that they belong to in the hierarchical structure H^c. Given two instances I_x and I_y, the structure-based similarity between two instances, denoted as $Sim_{Str}(I_x, I_y)$, is calculated as follows [41]:

$$Sim_{Str}\left(I_x, I_y\right) = 1 - \left(\frac{IC(I_x) + IC\left(I_y\right) - 2 \times IC\left(LCA_{I_x, I_y}\right)}{2}\right), \tag{5}$$

where $IC(I_x)$ and $IC(I_y)$ are the intrinsic IC of I_x and I_y respectively; $IC\left(LCA_{I_x I_y}\right)$ denotes the intrinsic IC of given two instances I_x and I_y, which is obtained with regard to their Least Common Ancestor (LCA) of the concepts that subsumes them in the considered H^c. The intrinsic IC of a specific instance, I_x, is assigned as the intrinsic IC of the concept that it belongs to in H^c, as follows:

$$IC(I_x) = IC(c_1), \quad c_1 \in \{C\}, \tag{6}$$

where c_1 is the concept that I_x belongs to in H^c (the concept that the instance I_x is instantiated from).

Since the parent concept of any given instance will be the leaf concept in H^c [20,41], and the intrinsic IC values of leaf concepts are assigned to their maximum values of one according to Seco's IC metric [41], we assume that the intrinsic IC value of an instance I_x would always be one, i.e. $IC(I_x) = IC(I_y) = 1$. Substituting these values into Eq. (6), we can simplify Eq. (5) as follows:

$$Sim_{Str}\left(I_x, I_y\right) = IC\left(LCA_{I_x, I_y}\right), \tag{7}$$

Considering that the instances in OWL ontology may have more than one parent concept [20], we define LCA_{I_x, I_y} as the most informative LCA for I_x and I_y, which is the pair of parent concepts that has the highest IC. For example, if the parent set of two given instances I_x and I_y is $\{c_1, c_2\}$ and $\{c_3\}$, respectively, the LCA_{I_x, I_y} can be expressed as follows:

$$max\left(IC\left(LCA_{c_1, c_2}\right), IC\left(LCA_{c_1, c_3}\right)\right), \tag{8}$$

The IC of a concept can be calculated using the metric proposed by Seco et al. [41] as follows:

$$IC(c) = 1 - \frac{\log(hypo(c) + 1)}{\log(max_{cons})}, \quad 0 \le IC(c) \le 1 \tag{9}$$

where c is a concept in H^c, *hypo* is a function that returns the number of hyponyms[5] of a given concept (c) and max_{cons} is the number of concepts that exist in the taxonomy under consideration H^c.

Based on Eq. (9), it can be seen that the *IC* value decreases monotonically as we traverse from leaf nodes up to the root node in the taxonomy. Hence, the *IC* value of a leaf-node concept will have an *IC* value of one, which indicates that the concept has been maximally expressed and cannot be further differentiated. In contrast, the *IC* values of concepts that are at the upper levels are less than one because they have many hyponyms. In particular, the root node concept will have an *IC* value of zero.

3.4. Datatype-based semantic similarity of two ontological instances

Datatype-based semantic similarity describes the similarity of two instances based on their common datatype properties with respect to a domain ontology. Datatype properties connect an instance to an XML schema datatype value or an RDF literal. The XML schema datatypes include interval-scaled, binary, nominal, ordinal, and/or ratio. The similarity between two instances connected to each datatype needs to be treated differently. A detailed description of similarity metrics that suits each type can be found in Han and Kamber [18].

Given two instances, I_x and I_y, let N be the set of their common datatype properties. The datatype-based similarity of these two instances, denoted as $Sim_{dt}(I_x, I_y)$, is defined as follows:

$$Sim_{dt}\left(I_x, I_y\right) = \frac{\sum_{i=1}^{N} DtSim_{p_i}\left(I_x, I_y\right)}{N}, \tag{10}$$

where $p_i \in N$ is the i^{th} common datatype property, and $DtSim_{p_i}\left(I_x, I_y\right)$ denotes the datatype similarity between I_x and I_y for the property p_i. If two given instances I_x and I_y do not share any datatype property, their $Sim_{dt}(I_x, I_y) = 0$.

The datatype-based similarity shown in Eq. (10) differs according to the type of datatype property p_i. Since the type of datatype properties that may be involved in similarity computation in our case study in the tourism domain is mainly nominal (or categorical), we adopted the *Jaccard coefficient*[6] [34], to compute the datatype similarity between two instances with regard to their common categorical properties. Suppose p_i is a common datatype property between I_x and I_y, v_x^n and v_y^m are the sets of values that the I_x and I_y can take for p_i, respectively. Then, datatype-based similarity between I_x and I_y for a categorical property p_i, $DtSim_{p_i}(I_x, I_y)$, is defined using the *Jaccard coefficient* as follows:

$$DtSim_{p_i}\left(I_x, I_y\right) = \frac{\#\left(v_x^n \cap v_y^m\right)}{\#\left(v_x^n \cup v_y^m\right)}, \tag{11}$$

where $\#\,(v_x^n \cap v_y^m)$ is the cardinality of positive matching values between I_x and I_y for p_i, $\#\,(v_x^n \cup v_y^m)$ is the cardinality of union of none zero values between I_x and I_y for p_i.

3.5. Algorithmic procedure of the IOBSS measure

Having presented the new IOBSS measure, this subsection describes the algorithmic procedure to calculate the semantic similarity of any two instances in a given OWL domain ontology. The two instances I_x and I_y and their associate networks as shown in Figs. 1 and 2, respectively, are used as examples. The procedure consists of three steps as listed below:

Step 1. Determine the associate networks of the two given instances (I_x and I_y)
Determine the associate network of each instance by finding all its associates through tracing its object property chains. Starting from the given instance, e.g. I_x or I_y, with closeness level $= 0$, retrieve all the object properties of this instance. For each object property, find the linked instances as its associates at the next level; this process continues until the last closeness level where the instances have no object properties (leaf-nodes).

Algorithm 1. Construction of the common associate pair set of two instances.

Input: associate networks of two instances I_x and I_y, $AssN_{I_x}$ and $AssN_{I_y}$
Output: common associate pair set of these two instances, $CAPSet_{I_xI_y}$
Process:
  Declare a Set, "$CAPSet_{I_xI_y}$"
  Declare a Queue, "*ComInsPairQ*"
  Declare Lists: "InstList1", "InstList2", "PairInstList"
  Add the given element (I_x, I_y) to the queue *ComInsPairQ*
  $k = 0$
  While-loop (condition) {
  $k = +1$
    For each level k of both $AssN_{I_x}$ and $AssN_{I_y}$
      For each common object property op_w at level k of $AssN_{I_x}$ and $AssN_{I_y}$
        InstList1← GetConnectedInstances($op_w^k, AssN_{I_x}$)
        InstList2← GetConnectedInstances($op_w^k, AssN_{I_y}$)
        //each list contains instances that are associates of the op_w^k in $AssN_{I_x}$ and $AssN_{I_y}$
        PairInstList← GetPairOfInstances(PairInstList1, PairInstList2)
        //each element in the this list represents two instances that are introduced via an op_w^k
        For each element in the PairInstList
          Add this element to the queue *ComInsPairQ*
        end for
      end for
    end for
  end while
  For each element in the *ComInstPairQ*
    k ← GetClosnessLevel(element, $AssN_{I_x}$, $AssN_{I_y}$)
    Determine ℓ using Eq. 2
    Determine $\mathbb{C}$ using Eq.3
    Form a five-tuple consisting of the two associates in the element, k, ℓ and $\mathbb{C}$.
    Add this tuple to the set $CAPSet_{I_xI_y}$
  end for

Step 2. Construct the common associate pair set of two given instances I_x and I_y
Step 2.1 Determine the common associate pairs
Given the associate networks of two given instances, (I_x and I_y), $AssN_{I_x}$ and $AssN_{I_y}$, as shown in Figs. 1 and 2, go through all the closeness levels from top to the bottom, for each level k of $AssN_{I_x}$ and $AssN_{I_y}$, find the common object properties, then retrieve the linked instances for each common object property to form the common pairs of instances of I_x and I_y. The common associate pair set can be viewed as a set of five-tuple, i.e. (element 1, element 2, k, ℓ, $\mathbb{C}$). The algorithmic procedure of this Step 2.1 is presented in Algorithm 1.
Step 2.2. Calculate the weight factor for each common associate pair
In this step, the weight factor is calculated for each common associate pair (*element 1*, *element 2*), thus, we need to determine its parameters ℓ (Eq.(2)), $\mathbb{C}$ (Eq.(3)) and k and calculate its weight factor using Eq. (1).
Step 3. Calculate the semantic similarity (IOBSS) of the two instances I_x and I_y
For each element in the common associate pair set of two given instances, I_x and I_y, first calculate the structure-based and datatype-based similarities of the pair of instances of each element using Eqs. (7) and (10), respectively, and then calculate the IOBSS similarity value using Eq. (4).

[5] *Hyponymy* involves specific instantiations of a more general concept. On another word, the *hypo* of a concept c denotes the number of its direct subclasses.
[6] which is frequently used as a similarity measure for asymmetric information on binary and non-binary variables.

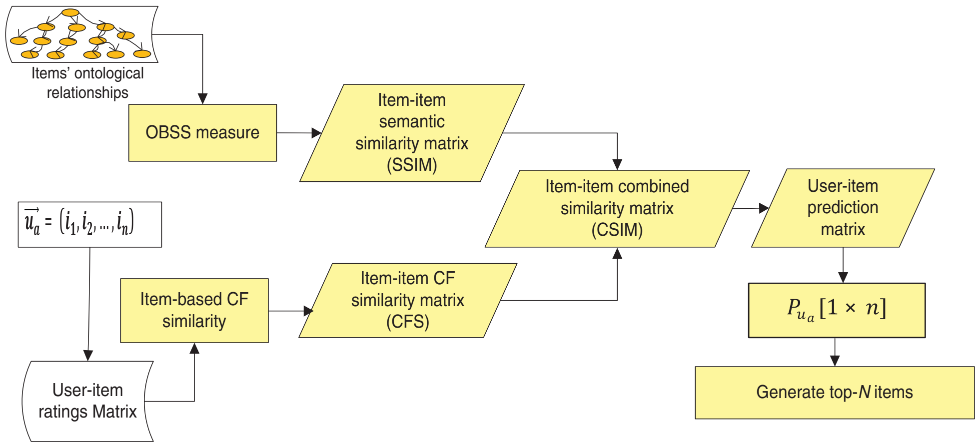

Fig. 3. The workflow of the computational recommendation procedure steps of the SBCF-IOBSS approach.

4. The semantic-based enhanced hybrid recommendation approach

With the aim of recommending the most appropriate items to users, we propose a semantic-based enhanced hybrid recommendation approach (SBCF-IOBSS) by combining the new IOBSS measure of items with the item-based CF framework. The rationale for this combination is twofold: (i) the IOBSS measure can enhance the similarity of items so that the accuracy of recommendation can be improved, and (ii) the hybrid approach can alleviate the sparsity and new item problems, because it captures additional knowledge by using the IOBSS measure.

4.1. Procedure of generating top-N recommendations

Fig. 3 shows the workflow of generating recommendations with this new hybrid approach, where the inputs of this approach include the user-item ratings matrix, denoted by $R[m \times n]$, where m represents the number of users and n represents the number of items; the target domain ontology schema and data (instances or items); and a given user, u_a, with his or her ratings of some of items (indicated by a vector of ratings). The output of this approach is the top-N recommendations to the given user (u_a).

Details of the workflow of the proposed SBCF-IOBSS approach are described as follows:

Step 1: Compute the item-based CF similarity of items

We adopted item-based CF similarity to calculate the similarity of each pair of items because it is superior in performance to other similarity measures, according to previous research [1,6]. The Pearson Correlation coefficient [40] is used to calculate the item-based CF similarity, based on the given user-item ratings matrix $R[m \times n]$. Formally, given the user-item ratings matrix $R[m \times n]$, the item-based CF similarity value between two items I_i and I_j, denoted as $CFSim_{I_i,I_j} : I \times I \rightarrow [-1,\ 1]$, is calculated as follows [24]:

$$CFSim_{I_i,I_j} = \frac{\sum_{u=1}^{|U_{ij}|}\left(r_{u,\,I_i} - \bar{r}_{I_i}\right)\left(r_{u,\,I_j} - \bar{r}_{I_j}\right)}{\sqrt{\sum_{u=1}^{|U_{ij}|}\left(r_{u,\,I_i} - \bar{r}_{I_i}\right)^2}\sqrt{\sum_{u=1}^{|U_{ij}|}\left(r_{u,\,I_j} - \bar{r}_{I_j}\right)^2}}, \tag{12}$$

where U_{ij} is the set of users who rated the items I_i and I_j together, $|U_{ij}|$ is the number of users in U_{ij}, $r_{u,\,I_i}$ and $r_{u,\,I_j}$ represents the rating given by user $u \in U_{ij}$ on service items I_i and I_j, respectively, and $\bar{r}_{I_i}$ and $\bar{r}_{I_j}$ is the average ratings of all users who have rated the item I_i and I_j, respectively.

The resultant item-based CF similarity of each pair of items is stored in an item–item similarity matrix, denoted by $CFS[n \times n]$.

Step 2: Compute the ontology-based semantic similarity (IOBSS)

The semantic similarity between each pair of items is calculated based on the IOBSS measure using Eq. (4), and stored in an item–item semantic similarity matrix, denoted by $SSIM[n \times n]$, where n is the number of items in the ontology dataset.

Step 3: Integrate the item-based CF and ontology-based semantic similarities

We calculate the semantic enhanced item–item similarity by linearly combining the item-based CF and IOBSS similarities. The combined similarity of instances I_i and I_j, denoted as $CombSim_{I_i,I_j} : I \times I \rightarrow [-1, 1]$, is computed as follows:

$$CombSim_{I_i,I_j} = \alpha \times OntSemSim_{I_i,I_j} + (1-\alpha) \times CFSim_{I_i,I_j}, \tag{13}$$

where α is a semantic combination parameter which specifies the weight of IOBSS in the combined similarity. If $\alpha = 0$, then the combined similarity represents only the respective item-based CF similarity of I_i and I_j; if $\alpha = 1$, then the combined similarity represents only the respective IOBSS similarity of I_i and I_j. Finding the proper value of α is not a trivial task and is usually highly dependent on the characteristics of the data. Thus, a sensitive analysis of the different values of α parameter is necessary to choose an appropriate α value that achieves the best performance for a given dataset. The combined similarity of each pair of items is stored in a new item–item similarity matrix, denoted as $CSIM[n \times n]$.

Step 4: Generate top-N recommendations for an active user

This step aims to generate the most relevant items that an active user might be interested in. First, we predict the user's ratings (i.e. rating values between 1 and 5) on all unseen items, and then generate the top-N items for the active user based on his/her predicted ratings.

To predict the ratings of the active user for the unseen items, the weighted sum method is employed as it commonly used in studies of recommendation systems [40]. With this method, first, we determine the neighborhood of each un-rated item (e.g. I_i), denoted as K_i[7]; then we calculate the predicted rating value for an active user (u_a), on the target item I_i, $P_{u_a,I_i} : U \times I \rightarrow [0, 5]$, using the following formula:

$$P_{u_a,I_i} = \frac{\sum_{q=1}^{K_i} r_{u_a,I_q} \times CombSim\left(I_i, I_q\right)}{\sum_{q=1}^{K_i} CombSim\left(I_i, I_q\right)}, \tag{14}$$

[7] K_i denotes the service items that are most similar to the un-rated item I_i.

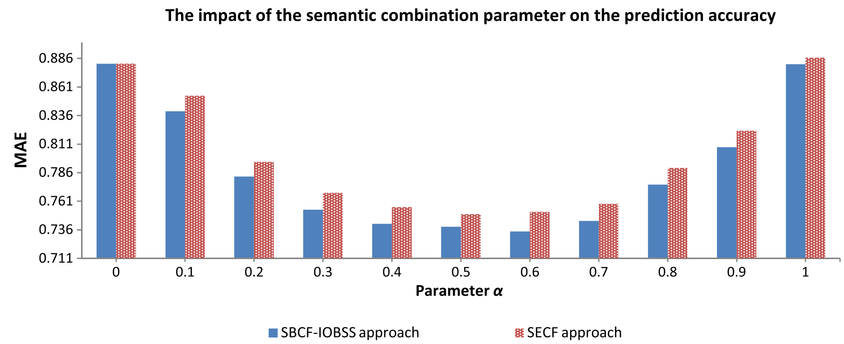

Fig. 4. The impact of the integration of the item-based CF and IOBSS measure on prediction accuracy.

where I_q belongs to the neighborhood of I_i and should be rated by the active user u_a, r_{u,I_q} denotes the rating of an item I_q by the user u_a, $CombSim(I_i, I_q)$ denotes the combined similarity value of the target item I_i and I_q which can be calculated by Eq. (13). The predicted rating values of unseen items for the user u_a are stored as a vector in the prediction matrix $P_{u_a}[1 \times n]$. Based on $P_{u_a}[1 \times n]$, we sort all unseen items according to the predicted rating values and then choose the top-*N* service items as the top-*N* recommendations for the given user.

4.2. Computational complexity analysis

The computational complexity of the proposed SBCF-IOBSS approach is the combination of the computational complexities of calculating similarity of items and predictions. The computational complexity of calculating similarity of items includes the time required to calculate both the item-based CF and the IOBSS similarities. The item-based CF similarity requires $O(n^2)$ for calculating the item–item similarity of n items. This step can be accomplished offline.

On the other hand, the time required to calculate the item–item similarity using the IOBSS measure is divided into three components, including the time required to build the associate networks, find the common associate sets and calculate item–item semantic similarity. First, the time required to build the associate networks of all available items defined in the ontology is $O(n \times (OP + C))$, where $O(OP + C)$ is the time required to build the associate network of each item, OP is the number of object properties defined in the ontology and C is the number of concepts in the target ontology. Second, the time required to find all the common associate sets is $O(n^2 \times (COP + C))$, where $O(COP + C)$ is the time required to find the common associate network of a pair of items, COP is the number of common object properties between any two associate networks. Third, the time required to compute the IOBSS similarity for n items, as defined in Eq. (4), is $O(n^2 + n^2N + 2n^2 \log COP + n)$, n), where, n^2 is needed to calculate the structure similarity, n^2N is needed to calculate the datatype similarity (N is the number of common datatype property between any two items), $2n^2 \log COP$ is for computing ℓ and $\mathbb{C}$ parameters and lastly n is needed for calculating the factor F. Therefore, the overall computational complexity of calculating the IOBSS similarity measure is $O(n \times (OP + C)) + O(n^2 \times (COP + C)) + O(n^2 + n^2N + 2n^2 \log COP + n) \approx O(n \times (OP + C)) + O(n^2 \times (COP + C))$. The IOBSS measure can be calculated offline. Finally, $O(n)$ is required to predict all unrated items for an active user; hence the overall computational complexity of the hybrid SBCF-IOBSS recommendation approach in the worst case becomes $O(m(n \times (OP + C))) + O(m(n^2 \times (COP + C)))$.

Although the proposed SBCF-IOBSS approach is computationally more expensive than classical item-based CF recommendation approaches (i.e. $O(n^2m)$), the calculations in the SBCF-IOBSS recommendation approach will be conducted at the beginning and when a new item is added to the ontology. In addition, all these calculations can be done offline. Therefore this approach is computationally feasible.

5. Experimental validation

To validate the effectiveness of the proposed SBCF-IOBSS recommendation approach, this section presents the experimental validation through conducting comparisons with three competing approaches based on a case study.

5.1. A case study: Australian e-Government tourism service

One of the main directions in the e-Government development strategy is to provide better online services to citizens such that the required information can be located with less time and search effort [21]. Tourism is one of the focused domains of e-Government service development strategies as it represents 11% of the worldwide GDP. Many governments around the world have devoted considerable time and energy to promote the tourism industry through non-profit services [46]. In the tourism domain, a government usually provides information about tourism entities including destinations, attractions that can be visited, activities that can be taken and events that can be attended at different destinations within the corresponding country. In this study, the Australian e-Government tourism service domain is utilized to validate the effectiveness of the new SBCF-IOBSS recommendation approach.

The experimental validation is conducted on a real-world dataset of Australian tourism services, extracted from two main sources: (i) the official NSW tourism service websites, and (ii) the Australian Tourism Data Warehouse (ATDW) (http://www.atdw.com.au/). The tourism service dataset consists of a total of 500 Australian tourism service items that include different attractions, activities, events, and destinations. This dataset is used to construct an Australian e-Government tourism ontology, which represents the semantic knowledge of Australian tourism e-Government service items.

The Australian e-Government tourism ontology was formalized using Protégé (http://protege.stanford.edu/) based on the Australian tourism knowledge. The knowledge formalized in ontology for the e-Government tourism domain provides a detailed semantic description of the entities in the domain, such as tourist attractions, and events or activities that are associated with a specific attraction. These entities are formalized as concepts. Each concept can have attributes and relationships with other concepts. The knowledge in the Australian tourism service ontology is utilized for the purpose of computing item similarity using the proposed IOBSS measure as well as to build the user-item ratings matrix. The columns of user-item matrix represent tourism service items which reference their corresponding items in the tourism ontology. The rows of the user-item ratings matrix represent user ratings information, where each data entry of each row represents a user's rating score which is either a rating value that ranges from 1 to 5, or zero (for entries in which the items have not been rated by the corresponding users). The user ratings information about preferred tourism items is retrieved from the ATDW.

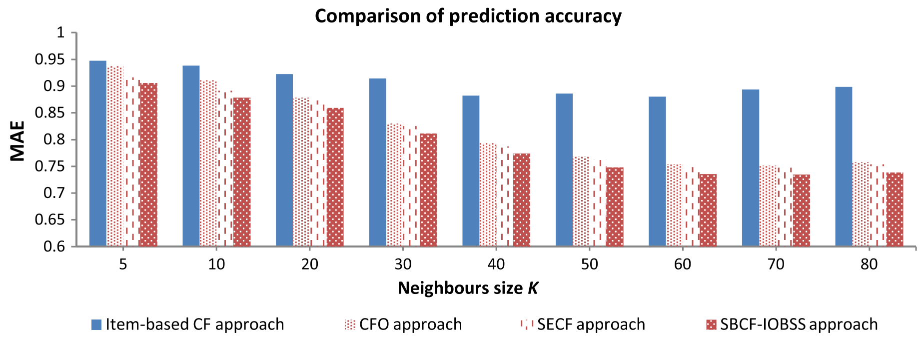

Fig. 5. Comparison of prediction accuracy between the new SBCF-IOBSS approach and the three competing approaches.

The user-item ratings matrix, of 400 users and 500 tourism items, is split into a training set and a test set using a specific parameter called training/test ratio (x). A value of $x = 0.8$ indicates that 80% of all the ratings of the entire dataset will be randomly selected as a training set, while the remaining 20% of ratings data will be used as the test set. The training set will be used to construct the required similarity matrix (the item–item similarity for the standard item-based CF approach, SECF and our new hybrid SBCF-IOBSS approach or the user–user similarity for the CFO approach) while the test dataset will be used to validate the predicted ratings of unseen items (i.e. the hidden portion of the rated tourism items).

5.2. Experimental design

To validate the performance of the new semantic-based enhanced hybrid recommendation approach (SBCF-IOBSS), three approaches were chosen as competing approaches for the experimental comparison which are, the standard item-based CF approach proposed by Sarwar et al. [40] and two state-of-the-art semantic enhanced recommendation approaches as mentioned in Section 2. One is the semantically enhanced CF (SECF) approach proposed by Mobasher et al. [38], and the other is the user-based CF with ontology-based approach (CFO) proposed by Sieg et al. [44]. The reasons for selection of these three as competing approaches are that the standard item-based CF approach has been widely exploited as a benchmark approach for its effective performance results, while the two advanced semantic enhanced hybrid recommendation approaches — the SECF and CFO — are closely related to the work presented in this study.

The experimental evaluation was conducted based on the dataset from the case study to generate the top-N most-liked service items, such as destinations, attractions, activities or events, to a given user using the new hybrid recommendation approach. The results were compared with the ones obtained from the three competing approaches which were run in the same environment. The platform used for the implementation is the Java NetBeans. The OWLModel and Jena OntModel were employed to facilitate and manage the communication between the OWL ontology of the tourism data and the Java NetBeans platform.

5.3. Experimental evaluation metric

The Mean Absolute Error (*MAE*) metric is used to evaluate the accuracy and quality of generated recommendations, as it is widely used in the recommendation research field [9,19,40]. The *MAE* is a measure of the deviation of predicted values of recommendations from their true user-specified values. This metric determines recommendation accuracy by computing the mean absolute deviation of the predicted rating values of unseen items compared to their actual ratings. For a given set of n items, the *MAE* metric is given by:

$$MAE = \frac{\sum_{i=1}^{n} |a_i - p_i|}{n}, \tag{15}$$

where p_i is the predicted rating and a_i is the actual rating of a hidden item i in the test dataset. Note that, a lower *MAE* value represents a higher prediction accuracy of generated recommendations.

To validate the performance of the new SBCF-IOBSS approach and to eliminate the potential bias of training/test sets in calculating the recommendation accuracy, ten-fold cross validation is conducted for each experiment. At each fold, 80% of rated tourism service items of the entire user-item ratings matrix will be randomly selected as training dataset. The remaining 20% of the rated items will be included in the test dataset. The *MAE* was computed and recorded at each fold and the overall *MAE* value then obtained as the averaged value. The *MAE* in the following experiments represents the overall *MAE*.

5.4. Determination of experimental parameters

In this study, there are three parameters that have a noticeable impact on the prediction accuracy of the new hybrid recommendation approach, namely the neighborhood size (K) (Step 4 in the sub-Section 4.1), the semantic combination parameter (α) (step 3 in the sub-Section 4.1), and the sparsity level. The values of K and α were determined based on the sensitivity analysis of these two parameters to the recommendation accuracy. In this case study, we run the experiments using the new hybrid SBCF-IOBSS and SECF[8] approaches by varying the semantic combination parameter (α) and neighborhood size K values. For each α value within the range from 0 to 1 with an increment of 0.1, we run the experiment by varying the neighborhood size K from 5 to 80, and the neighborhood size K with the minimum *MAE* is then recorded. Fig. 4 plots the minimum value of *MAE* for each parameter α value with the best neighborhood size K.

It can be seen from Fig. 4 that the integration of the semantic similarity with traditional item-based CF yields substantial improvement to the accuracy. The best prediction accuracy result is obtained when the α parameter equals 0.6 and 0.5 by the proposed hybrid SBCF-IOBSS approach and the SECF approach, respectively. Fig. 4 also shows that the proposed approach has outperformed the SECF approach by achieving better prediction accuracy at different α values.

Furthermore, the improvement of the SBCF-IOBSS approach compared to the SECF approach has been verified statistically using the paired t-test statistical measure. Using this test, it has been found that

[8] The ESCF approach is sensitive to parameter alpha as it linearly combines the item-based CF similarity and the semantic similarity. The combined similarities are used to generate predictions of unseen items.

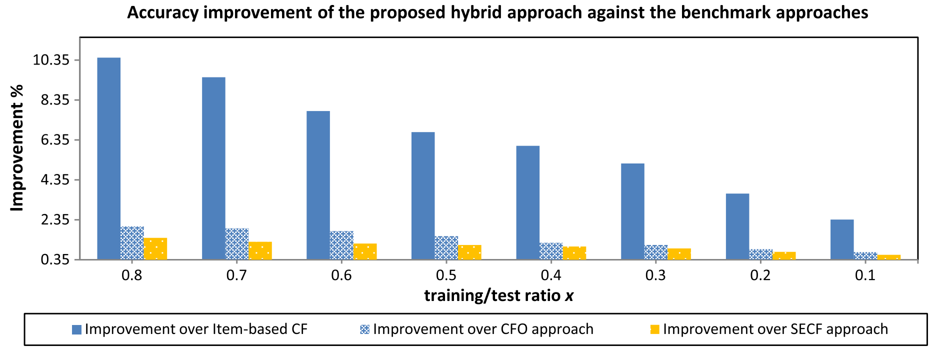

Fig. 6. Improvement in prediction accuracy of the SBCF-IOBSS approach over competing approaches at different sparsity levels.

the obtained p-value is 1.1148e-05, which is significant (i.e. $p < 0.05$), thus, the null hypothesis of mean equality is rejected and a meaningful difference in the prediction accuracy exists.

The sparsity level of a user-item ratings matrix is defined as:

$$\text{Sparsity} = 1 - \text{density}, \tag{16}$$

where density is the density of the user-item ratings matrix, which is defined as the ratio of the number of non-zero elements to the total number of elements in the matrix.

For instance, the density of the user-item ratings matrix that used in this study is 0.0577, then the sparsity level of this matrix is $1 - 0.0577 = 0.9423$.

5.5. Experimental results

This section presents the results of the experiments in terms of the prediction accuracy of the recommendations.

5.5.1. Effectiveness of the new hybrid approach on prediction accuracy

A number of experiments are conducted using different K values from 5 to 80 and the optimal α is set to values for the hybrid SBCF-IOBSS and SECF approaches. Fig. 5 shows the best prediction accuracy values of all considered approaches with different values of parameter K. It can be seen from Fig. 5 that the proposed hybrid approach reveals substantially better prediction accuracy than the three competing approaches for all values of parameter K under consideration. It can also be clearly seen that prediction accuracy increases as parameter K increases and reaches the optimal value, which is around $K = 70$ for all approaches except for the traditional item-based CF.

To justify the differences of *MAE* values of the proposed approach from other competing approaches on the prediction accuracy, the paired t-test statistical measure has been applied. The reported p-values are 9.58051e-05, 3.66739e-06 and 2.15778e-09 for the proposed approach in comparison with the SECF, CFO and item-based CF approaches, respectively. Therefore, the null hypothesis of mean equality is rejected and meaningful differences in prediction accuracy of the proposed approach are proven against all other competing approaches.

5.5.2. Effectiveness of the SBCF-IOBSS approach in dealing with the sparsity problem

Sparsity is one of the main problems that negatively affect the prediction accuracy. It occurs when the obtained ratings are few compared to the number of available items. For testing the effectiveness of the SBCF-IOBSS approach in handling the sparsity problem, we conduct a number of experiments using all the considered approaches with several datasets which were formed based on the same Australian tourism dataset. Each new dataset has a sparsity levels. Fig. 6 plots the *MAE* improvement of the proposed approach against the three competing approaches. It can be seen from Fig. 6 that the *MAE* of the proposed hybrid approach has achieved better improvement than the traditional item-based CF and other two competing approaches at all sparsity levels. Nevertheless, the achieved improvement in the prediction accuracy by the SBCF-IOBSS approach clearly declines as the proportion of the training data is reduced (the sparsity is increased), and as might be expected this improvement tends to converge to zero for very sparse datasets. This is because for very sparse data, neither approach can generate a reasonable recommendation. However, the shown result in Fig. 6 indicates that the new approach performs better in handling the sparsity problem than the competing approaches even when the data is very sparse.

To verify the differences of the *MAE* values of the proposed approach from other competing approaches on the prediction accuracy, the paired t-test statistical measure has been applied. The reported p-values are 0.000375, 0.000394 and 2.59216e-05 for the proposed approach in comparison with the SECF, CFO and item-based CF approaches, respectively. Therefore, the null hypothesis of mean equality is rejected and meaningful differences in prediction accuracy of the proposed approach are proven against all other competing approaches.

5.5.3. Effectiveness of the SBCF-IOBSS in dealing with the cold-start item problem

As reported by other studies (Schafer et al. 2007), it is difficult to give accurate recommendations for new items, because high-quality recommendations can only be obtained with sufficient data ratings. To validate the effectiveness of the proposed SBCF-IOBSS approach in dealing with the new items problem, we form a new dataset based on the Australian tourism dataset by purposely adding a number of new items to the test set, which are the items that have been rated only once in the training dataset. Using this new dataset, we conducted a number of experiments in which the K parameter is varied from 5 to 80 and α parameter is set to 1 using only the proposed SBCF-IOBSS and SECF approaches, the item-based CF and the CFO are excluded from the experiments as they cannot make recommendations for new items. Fig. 7 plots the *MAE* values for the two approaches. It can be seen that the proposed approach gives better prediction accuracy for new items than the SECF approach at all values of parameter K under consideration. This indicates that the new hybrid SBCF-IOBSS approach can better deal with the new item problem than the SECF approach. This result has been confirmed through conducting a paired t-test statistical measure. According to this test, the null hypothesis of mean equality is rejected and a meaningful difference in the prediction accuracy is proven with a p-value of 8.50e-06 less than the significance value.

5.6. Discussion of the results

The achieved improvement by the proposed hybrid approach, as presented in the sub-Section 5.5, can be explained by following factors: (i) the proposed IOBSS measure explores the implicit semantics of instances by inferring rich semantic knowledge through semantic

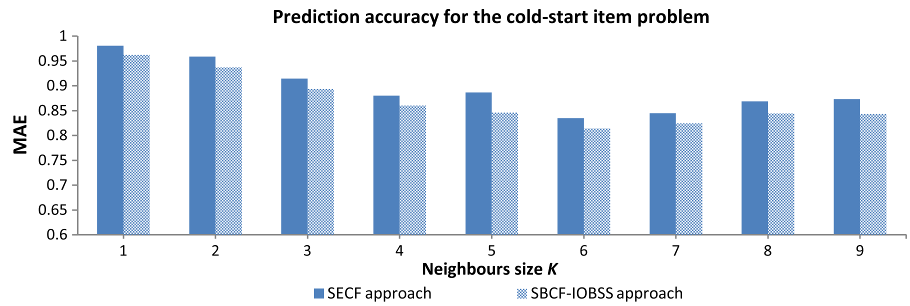

Fig. 7. Improvement in prediction accuracy of the SBCF-IOBSS approach against the SECF approach on new items problem.

associations; (ii) the IOBSS measure can handle complex relationships well by the new inference mechanism, termed associate networks. By means of associate networks, relationship chains that span several instances become a very useful approach for discovering hidden links between seemingly disparate instances; (iii) the associate networks can support the semantic analytic of heterogeneous content which in turn can reveal useful insights into the similarity of ontological instances. This improves the existing semantic similarity measures which mainly focus on direct relationships of instances and pay less attention to indirect ontological relationships.

5.7. Concerns about computational feasibility and flexibility

Even though the proposed SBCF-IOBSS hybrid approach is mainly validated using the case study of the Australian e-Government tourism service dataset, several facts reveal that the proposed approach is also computationally viable and scalable in more complex environments with a greater number of users and items. The first fact is that the proposed SBCF-IOBSS approach has no real-time requirements, as the calculation of the semantic similarity of instances can be done offline and updated only when new instances are added to the system. Secondly, the similarity computation using the IOBSS measure is within the items space, which has less chance to change, compared to the users' space. Since, the user's preferences and the items similarity are known in advance, the computational complexity of the proposed approach does not cause an unacceptable delay in the delivery of recommendations. Lastly, the semantic similarity measure provides valuable information in improving the estimation of item–item similarity which in turn contributes positively towards generating more accurate and high quality recommendations, especially in the cases where the sparsity problem and/or new item problem are present.

Regarding the generalization of the proposed approach, although the IOBSS measure, related terms and calculation procedure were validated using a case study, its inferential mechanism and the steps of calculating the IOBSS measure can be used in any domain as long as the domain knowledge can be modeled and formalized as an ontology and the type of datatype properties are known. Therefore there is no limitation to practical scope for extending the framework to different domains. However, since the IOBSS measure aims to capture both the direct and implicit relationships to compute semantic similarity between any pair of available items in the considered domain, if the given domain ontology has no much implicit relationships, the effects of using the IOBSS measure would be not significant.

6. Conclusion and future work

This paper proposes a new hybrid semantic-based enhanced recommendation approach that can be used to effectively offer items tailored to users' needs and preferences. The proposed approach integrates semantic similarity of items with the traditional item-based CF approach to enhance the personalization capabilities of existing recommendation approaches. A new IOBSS measure is proposed to accurately estimate semantic similarity among instances. The performance of the new recommendation approach has been validated using a real world dataset from the Australian tourism domain and has been compared with the traditional item-based CF as a baseline approach and two advanced semantic-enhanced CF. The experimental evaluation results demonstrate that the proposed approach outperforms the three competing approaches in terms of recommendation accuracy and capability to deal with the sparsity and new-item problems. Furthermore, it has been shown that the SBCF-IOBSS recommendation approach is feasible and practical for use in real world e-Government recommendation systems.

Some future work could be (i) to apply the SBCF-IOBSS approach in other e-Government service domains, such as Education, Medicare and Welfare; (ii) to develop an e-Government tourism service recommendation system using the proposed approach.

Acknowledgment

The authors would like to thank the Australian Tourism Data Warehouse (ATDW) for providing the dataset used in this study.

References

[1] G. Adomavicius, A. Tuzhilin, Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions, IEEE Transactions on Knowledge and Data Engineering 17 (6) (2005) 734–749.
[2] R. Albertoni, M.D. Martino, Asymmetric and Context-Dependent Semantic Similarity Among Ontology Instances, in: S. Spaccapietra (Ed.), Journal on Data Semantics X, Springer, Berlin/Heidelberg, 2008, pp. 1–30.
[3] B. Aleman-Meza, I.B. Arpinar, M.V. Nural, A.P. Sheth, Ranking documents semantically using ontological relationships, In Fourth International Conference on Semantic Computing (ICSC), IEEE, 2010, pp. 299–304.
[4] Berners-Lee, The semantic web, Scientific American 284 (5) (2001) 34–43.
[5] A. Bernstein, E. Kaufmann, C. Burki, M. Klein, How similar is it? Towards personalized similarity measures in ontologies, 7th International Conference Tagung Wirtschaftsinformatik, 2005, pp. 1345–1366 (Germany).
[6] C. Birtolo, D. Ronca, Advances in clustering collaborative filtering by means of fuzzy C-means and trust, Expert Systems with Applications 40 (17) (2013) 6997–7009.
[7] Y. Blanco-Fernández, J.J. Pazos-Arias, A. Gil-Solla, M. Ramos-Cabrer, M. López-Nores, J. García-Duque, A. Fernández-Vilas, R.P. Díaz-Redondo, Exploiting synergies between semantic reasoning and personalization strategies in intelligent recommender systems: a case study, Journal of Systems and Software 81 (12) (2008) 2371–2385.
[8] Y. Blanco-Fernández, J.J. Pazos-Arias, A. Gil-Solla, M. Ramos-Cabrer, M. López-Nores, J. García-Duque, A. Fernández-Vilas, R.P. Díaz-Redondo, J. Bermejo-Muñoz, A flexible semantic inference methodology to reason about user preferences in knowledge-based recommender systems, Knowledge-Based Systems 21 (4) (2008) 305–320.
[9] J.S. Breese, D. Heckerman, C. Kadie, Empirical analysis of predictive algorithms for collaborative filtering, Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, 1998, pp. 43–52.
[10] R. Burke, Hybrid recommender systems: survey and experiments, User Modeling and User-Adapted Interaction 12 (4) (2002) 331–370.
[11] R. Burke, Hybrid web recommender systems, in: P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.), The Adaptive Web, Springer, Berlin/Heidelberg, 2007, pp. 377–408.
[12] F. Cacheda, V. Carneiro, D. Fernández, V. Formoso, Comparison of collaborative filtering algorithms: limitations of current techniques and proposals for scalable, high-performance recommender systems, ACM Transactions on the Web 5 (1) (2011) 1–33.
[13] I. Cantador, An Enhanced Semantic Layer for Hybrid Recommender Systems, Semantic Web: Ontology and Knowledge Base Enabled Tools, Services, and Applications, 2013.
[14] I. Cantador, A. Bellogín, P. Castells, Ontology-based personalised and context-aware recommendations of news items, ACM International Conference on Web Intelligence and Intelligent Agent Technology, ACM, Sydney, 2008, pp. 562–565.

[15] H. Dong, F.K. Hussain, E. Chang, A service concept recommendation system for enhancing the dependability of semantic service matchmakers in the service ecosystem environment, Journal of Network and Computer Applications 34 (2) (2011) 619–631.
[16] M. Gan, X. Dou, R. Jiang, From ontology to semantic similarity: calculation of ontology-based semantic similarity, The Scientific World Journal 2013 (2013).
[17] T.R. Gruber, Toward principles for the design of ontologies used for knowledge sharing, International Journal of Human-Computer Studies 43 (5–6) (1995) 907–928.
[18] J. Han, M. Kamber, Data Mining: Concepts and Techniques, 2006.
[19] J.L. Herlocker, A.K. Joseph, B. Al, R. John, An algorithmic framework for performing collaborative filtering, Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, Berkeley, California, United States, 1999.
[20] M. Horridge, S. Jupp, G. Moulton, A. Rector, R. Stevens, C. Wroe, A Practical Guide to Building OWL Ontologies Using Protégé 4 and CO-ODE Tools, University Of Manchester, 2007.
[21] P.T. Jaeger, J.C. Bertot, Designing, implementing, and evaluating user-centered and citizen-centered e-Government, International Journal of Electronic Government Research (IJEGR) 6 (2) (2010) 1–17.
[22] J.J. Jiang, D.W. Conrath, Semantic similarity based on corpus statistics and lexical taxonomy, Proceedings of the 10th International Conference on Research on Computational Linguistics, 1997 (Taiwan).
[23] H.-N. Kim, A. El-Saddik, G.-S. Jo, Collaborative error-reflected models for cold-start recommender systems, Decision Support Systems 51 (3) (2011) 519–531.
[24] U. Shardanand, P. Maes, Social information filtering: algorithms for automating "word of mouth", Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM Press/Addison-Wesley, 1995, pp. 210–217.
[25] F. Lecue, Combining collaborative filtering and semantic content-based approaches to recommend web services, IEEE Fourth International Conference on Semantic Computing (ICSC), IEEE, 2010, pp. 200–205.
[26] F. Li, T.C. Du, Who is talking? An ontology-based opinion leader identification framework for word-of-mouth marketing in online social blogs, Decision Support Systems 51 (1) (2011) 190–197.
[27] D. Lin, An information-theoretic definition of similarity, Proceedings of the 15th International Conf. on Machine Learning, Morgan Kaufmann, San Francisco, CA, 1998, pp. 296–304.
[28] P. Liu, G. Nie, D. Chen, Exploiting semantic descriptions of products and user profiles for recommender systems, IEEE Symposium on Computational Intelligence and Data Mining, IEEE, 2007, pp. 179–185.
[29] A. Maedche, S. Staab, Ontology Learning, Springer, 2004.
[30] A. Maedche, V. Zacharias, Clustering ontology-based metadata in the semantic web, in: J.G. Carbonell, J.O. Siekmann (Eds.), Principles of Data Mining and Knowledge Discovery, Springer, 2002.
[31] S.E. Middleton, N.R. Shadbolt, D.C.D. Roure, Ontological user profiling in recommender systems, ACM Transactions on Information Systems 22 (1) (2004).
[32] G.A. Miller, W.G. Charles, Contextual correlates of semantic similarity, Language and Cognitive Processes 6 (1) (1991) 1–28.
[33] B. Mobasher, X. Jin, Y. Zhou, Semantically enhanced collaborative filtering on the web, in: B. Berendt, A. Hotho, D. Mladenic, M.V. Someren, M. Spiliopoulou, G. Stumme (Eds.), Web Mining: From Web to Semantic Web, Springer, 2004, pp. 57–76.
[34] T. Pang-Ning, M. Steinbach, V. Kumar, Introduction to Data Mining, Addison-Wesley, 2005.
[35] M.J. Pazzani, D. Billsus, Content-based recommendation systems, in: P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.), The Adaptive Web, Springer, 2007, pp. 325–341.
[36] E. Peis, J.M. Morales-del-Castillo, J.A. Delgado-López, Semantic recommender systems, Analysis of the State of the Topic, 62008. 1–9 (Hipertext.net).
[37] R. Rada, H. Mili, E. Bicknell, M. Blettner, Development and application of a metric on semantic nets, IEEE Transactions on Systems, Man, and Cybernetics 19 (1) (1989) 17–30.
[38] P. Resnik, Using information content to evaluate semantic similarity in a taxonomy, 14th International Joint Conference on Artificial Intelligence, 1995, pp. 448–453 (Montreal).
[39] D. Sánchez, M. Batet, D. Isern, A. Valls, Ontology-based semantic similarity: a new feature-based approach, Expert Systems with Applications 39 (9) (2012) 7718–7728.
[40] B. Sarwar, G. Karypis, J. Konstan, J. Reidl, Item-based collaborative filtering recommendation algorithms, Proceedings of the 10th International Conference on World Wide Web, 2001, pp. 285–295.
[41] N. Seco, T. Veale, J. Hayes, An intrinsic information content metric for semantic similarity in WordNet, Proc. of the European Conference on Artificial Intelligence (ECAI), 2004, pp. 1089–1090.
[42] M.H. Seddiqui, M. Aono, Metric of intrinsic information content for measuring semantic similarity in an ontology, Proceedings of the Seventh Asia-Pacific Conference on Conceptual Modelling (APCCM), ACM, Brisbane-Australia, 2010, pp. 89–96.
[43] Q. Shambour, J. Lu, A trust-semantic fusion-based recommendation approach for e-business applications, Decision Support Systems 54 (1) (2012) 768–780.
[44] A. Sieg, B. Mobasher, R. Burke, Improving the effectiveness of collaborative recommendation with ontology-based user profiles, Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems, ACM, Spain, 2010, pp. 39–46.
[45] X. Su, T.M. Khoshgoftaar, A survey of collaborative filtering techniques, Advances in Artificial Intelligence 2009 (1) (2009) 1687–7470.
[46] United Nations, The 2012 global e-government survey: e-Government for the people, 2012.
[47] C. Wang, J. Lu, G. Zhang, Integration of ontology data through learning instance matching, IEEE/WIC/ACM International Conference on Web, Intelligence, 2006, pp. 536–539.
[48] R.-Q. Wang, F.-S. Kong, Semantic-enhanced personalized recommender system, Proceeding of the International Conference on Machine Learning and, Cybernetics, 2007, pp. 4069–4074.
[49] J. Wu, Z. Wu, Similarity-based web service matchmaking, Proceedings of IEEE International Conference on Services, Computing, 2005, pp. 287–294.
[50] Y. Xu, X. Guo, J. Hao, J. Ma, R.Y.K. Lau, W. Xu, Combining social network and semantic concept analysis for personalized academic researcher recommendation, Decision Support Systems 54 (1) (2012) 564–573.

Malak Al-Hassan has completed her Ph.D. in the School of Software, Faculty of Engineering and Information Technology, at the University of Technology Sydney (UTS)/Australia. She is a member at the Decision Systems & e-Service Intelligence (DeSI) Research Lab, Center for Quantum Computation & Intelligent Systems (QCIS), School of Software, and Faculty of Engineering and Information Technology, University of Technology Sydney. Her research interest includes Intelligent E-service system, E-government services, web personalization, recommendation systems and ontology.

Haiyan Lu received her B. Eng. and M. Eng. from the Harbin Institute of Technology, Harbin, China, in 1985 and 1988, respectively, and the Ph.D. degree in engineering from the University of Technology, Sydney, Australia, in 2002.
She is currently with the Decision Systems & e-Service Intelligence (DeSI) Lab in the Center for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia. She has published over 70 refereed journal and conference papers. Her research interests include heuristic optimization algorithms, time series forecasting, ontology, and recommendation techniques and their applications in business and engineering.

Professor **Jie Lu** is the Associate Dean Research (Acting) of Faculty of Engineering and Information Technology, and the Director of the Decision Systems and e-Service Intelligence Research Laboratory in the Center for Quantum Computation & Intelligent Systems at the University of Technology, Sydney (UTS). Her research interests lie in the area of decision support systems and uncertain information processing. She has published five research books and 300 papers, won five Australian Research Council discovery grants and 10 other grants. She received a University Research Excellent Medal in 2010. She serves as Editor-In-Chief for Knowledge-Based Systems (Elsevier), Editor-In-Chief for International Journal of Computational Intelligence Systems (Atlantis), editor for book series on Intelligent Information Systems (World Scientific) and guest editor of six special issues for international journals, as well as delivered six keynote speeches at international conferences.