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.