Expert Systems With Applications 57 (2016) 139–158

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Privileged contextual information for context-aware recommender systems

Camila Vaccari Sundermann[a], Marcos Aurélio Domingues[b,*], Merley da Silva Conrado[a,1], Solange Oliveira Rezende[a]

[a] Institute of Mathematics and Computer Science, University of São Paulo, Avenida Trabalhador São Carlense, 400, São Carlos, SP, 13566-590, Brazil
[b] Department of Informatics, State University of Maringá, Avenida Colombo, 5790, Maringá, PR, 87020-900, Brazil

ARTICLE INFO

Article history:
Received 3 June 2015
Revised 18 March 2016
Accepted 19 March 2016
Available online 26 March 2016

Keywords:
Recommender systems
Topic hierarchy
Named entities
Domain terms
Contextual information
Context-aware recommender systems

ABSTRACT

A recommender system is used in various fields to recommend items of interest to the users. Most recommender approaches focus only on the users and items to make the recommendations. However, in many applications, it is also important to incorporate contextual information into the recommendation process. Although the use of contextual information has received great focus in recent years, there is a lack of automatic methods to obtain such information for context-aware recommender systems. Some works address this problem by proposing supervised methods, which require greater human effort and whose results are not so satisfactory. In this scenario, we propose an unsupervised method to extract contextual information from web page content. Our method builds topic hierarchies from page textual content considering, besides the traditional bag-of-words, valuable information of texts as named entities and domain terms (privileged information). The topics extracted from the hierarchies are used as contextual information in context-aware recommender systems. We conducted experiments by using two data sets and two baselines: the first baseline is a recommendation system that does not use contextual information and the second baseline is a method proposed in literature to extract contextual information. The results are, in general, very good and present significant gains. In conclusion, our method has advantages and innovations:(i) it is unsupervised; (ii) it considers the context of the item (Web page), instead of the context of the user as in most of the few existing methods, which is an innovation; (iii) it uses privileged information in addition to the existing technical information from pages; and (iv) it presented good and promising empirical results. This work represents an advance in the state-of-the-art in context extraction, which means an important contribution to context-aware recommender systems, a kind of specialized and intelligent system.

1. Introduction

Nowadays, most web sites offer a large number of items (e.g., movies, music, web pages, etc) to their users. Finding relevant content according to the taste of each individual has become a challenge. Recommender systems have emerged in response to this problem. A recommender system is an information filtering technology which can be used to predict preference ratings of items, not currently rated by the user, and/or to output a personalized ranking of items that are likely to be of interest to the user (Ricci, Rokach, Shapira, & Kantor, 2011). These systems have flourished on the Internet, and web sites such as Amazon[2], Netflix[3] and Last.fm[4] are good examples of recommenders that adapt recommendations to the user's personal tastes.

Traditionally, the data that are most often available for the recommender systems are web access logs which represent the interaction activity between users and items. Therefore, the most common systems focus on these two entities to build a model which is used to recommend an ordered list of N items that are expected to be of interest to a certain user.

Unlike the traditional systems, that make recommendations only by using the relation *User* × *Item*, a context-aware

* Corresponding author. Tel.: +554499183087.
E-mail addresses: camilavs@icmc.usp.br (C.V. Sundermann), maddomingues@gmail.com (M.A. Domingues), merleyc@gmail.com (M.d.S. Conrado), solange@icmc.usp.br (S.O. Rezende).
[1] The author is currently Software Engineer, Intel Corp., Hillsboro, OR, USA.

[2] http://www.amazon.com.
[3] http://www.netflix.com.
[4] http://www.last.fm.

http://dx.doi.org/10.1016/j.eswa.2016.03.036
0957-4174/

recommender system can make recommendations by incorporating available contextual information into the recommendation process as explicit additional categories of data (Adomavicius, Sankaranarayanan, Sen, & Tuzhilin, 2005):

$$User \times Item \times Context \rightarrow Recommendation,$$

where *Context* specifies the contextual information associated with the application.

Thus, a promising way to improve the accuracy of the recommender systems is to incorporate additional information, such as context, besides the typical information about users and items. There are many definitions of context in the literature considering the field of application and the available customer data (Ricci et al., 2011). In this work, context is defined as any information that can be used to characterize the situation of an entity (i.e., a web page) (Dey, 2001).

Although the use of contextual information for recommender systems has received great focus in recent years (Adomavicius, Mobasher, Ricci, & Tuzhilin, 2011; Adomavicius et al., 2005; Hariri, Mobasher, Burke, & Zheng, 2011; Li et al., 2010; Panniello & Gorgoglione, 2012), an intense and time-consuming human effort is usually necessary to identify, collect and label the contextual information in order to be properly employed in the recommender systems. Moreover, manually labeling the content becomes impracticable for large data sets. On the other hand, there is a lack of automatic methods to obtain such information for the context-aware recommender systems. For all these reasons, the acquisition of contextual information is a research area that needs to be better explored.

The contextual information, in addition to various definitions, can also be structured in various ways. In some works, such information is organized in a hierarchical structure, which can be represented as trees (Adomavicius et al., 2005; Panniello & Gorgoglione, 2012). Given this possibility of hierarchical organization of context, in this work topic hierarchies are used as a way to organize and extract the context of the textual content of Web pages.

Many existing methods in the literature to build topic hierarchies represent the text collections using only the traditional bag-of-words. Such representation considers the terms or attributes of texts as a disordered set of words and does not consider the relationship among them (Feldman & Sanger, 2007). However it is possible, in some cases, to extract from texts a more valuable kind of information known as privileged information (Vapnik & Vashist, 2009). In this paper we built topic hierarchies from web page textual content, using the method *LUPI-based Incremental Hierarchical Clustering* (LIHC) (Marcacini & Rezende, 2013), which considers besides the bag-of-words (technical information), also the privileged information extracted from the texts. The topics obtained from the topic hierarchies are used as contextual information by context-aware recommender systems.

In this work we consider the named entities and the domain terms from the texts as privileged information, because these words represent rich information about the web page content. They are available only for a fraction of the documents and are not represented explicitly in the texts.

As part of this work, we made a depth empirical evaluation of our method, including a comparison of our proposal against some state-of-the-art methods proposed in the literature that extract contextual information from textual data. The evaluation was performed using two real world data sets and the results demonstrated that the LIHC is a promising method to extract contextual information for context-aware recommender systems.

Our method has advantages and innovations:(i) it is unsupervised; (ii) it considers the context of the item (Web page), not the context of the user as most of the few existing methods, which is an innovation; (iii) it uses privileged information in addition to the existing technical information from pages; and (iv) it presented good and promising empirical results.

This paper is structured as follows: Section 2 discusses some definitions of context. The context-aware recommender systems used in this work are presented in Section 3. Our proposal, which consists of extracting contextual information from topic hierarchies, is described in Section 4. The privileged information, Named Entities and Domain Terms, are presented in Subsections 4.1 and 4.2. We evaluate our proposal in Section 5. In Section 6 we descute some related work. Finally, in Section 7, we present the conclusion and future work.

2. Contextual information

According to Adomavicius et al. (2011), the concept of context has been studied extensively in computing areas and other disciplines. As already mentioned before, context can be defined in many ways, depending on the field of application. After examining 150 different definitions of context from different fields, Bazire and Brézillon (2005) concluded that itis difficult to find a unifying definition. They raised some questions, such as: “Is context a frame for a given object? Is it the set of elements that have any influence on the object? Is it something static or dynamic?”. For Dourish (2004), there are two different views of context: the interactional view and the representational view. In the interactional view, the context is defined dynamically and there is a relationship between context and activity, in which the activity gives rise to the context and the context influences the activity. In contrast, in the representational view, context can be described as a set of known attributes, whose structure does not change through the time.

The most widely accepted definition of context and that is used in this work was proposed by Dey (2001): “Context is any information that can be used to characterize the situation of an entity”. The entities are, in our work, web pages. According to Adomavicius et al. (2011), the contextual information can be of different types. As an example, they mentioned an application for recommending movies to the users. Besides the attributes of the users and the movies, there is also the contextual information. This information consists of three types: “Theater”, “Time” and “Companion”. Each type has some attributes/values. The type “Companion”, for example, has the attributes “alone”, “family”, “co-workers” and so on. For Adomavicius et al. (2011), this is a way of representing the context for the application. The contextual information can also be organized as a hierarchical structure that can be represented as trees (Adomavicius et al., 2005; Palmisano, Tuzhilin, & Gorgoglione, 2008; Panniello & Gorgoglione, 2012). In this way, the contextual information is a set of contextual dimensions C, where each dimension C is defined by a set of f attributes/values, i.e., $C = \{c_1, c_2, ..., c_f\}$. These attributes have a hierarchical structure. The values taken by attribute c_f define more granular levels, while c_1 defines less granular levels of the contextual information. For example, Panniello and Gorgoglione (2012) represent the contextual attribute “period of the year” as in the hierarchical structure illustrated in Fig. 1.

As mentioned in the introduction of this paper, one challenge regarding context-aware recommenders is how to extract the contextual information from the users, items and their relation. Information such as illustrated in Fig. 1 has to be obtained either manually or automatically. While manual methods are time-consuming and error-prone, automatic approaches require the development of algorithms and strategies to extract contextual information from the textual content. In Section 4, we describe our proposal, which suggests using text mining techniques applied to textual data (i.e., web pages) to obtain contextual information. Before presenting our proposal, we review, in the next section, some context-aware recommender systems used in our work.