Applied Soft Computing 11 (2011) 793–806

Contents lists available at ScienceDirect

Applied Soft Computing

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

NEWER: A system for NEuro-fuzzy WEb Recommendation

G. Castellano, A.M. Fanelli, M.A. Torsello*

University of Bari, Department of Informatics, Via Orabona, 4, 70126 Bari, Italy

ARTICLE INFO

Article history:
Received 24 October 2008
Accepted 31 December 2009
Available online 14 January 2010

Keywords:
Neuro-fuzzy systems
User profiling
Web recommendation
Web personalization
Web usage mining

ABSTRACT

In the era of the Web, there is urgent need for developing systems able to personalize the online experience of Web users on the basis of their needs. Web recommendation is a promising technology that attempts to predict the interests of Web users, by providing them with information and/or services that they need without explicitly asking for them. In this paper we propose NEWER, a usage-based Web recommendation system that exploits the potential of Computational Intelligence techniques to dynamically suggest interesting pages to users according to their preferences. NEWER employs a neuro-fuzzy approach in order to determine categories of users sharing similar interests and to discover a recommendation model as a set of fuzzy rules expressing the associations between user categories and relevances of pages. The discovered model is used by a online recommendation module to determine the list of links judged relevant for users. The results obtained on both synthetic and real-world data show that NEWER is effective for recommendation, leading to a quality of the generated recommendations comparable and often significantly better than those of other approaches employed for the comparison.

1. Introduction

During the past few years, the World Wide Web has become the biggest and the most popular way of communication and information dissemination. Everyday, the Web grows by roughly millions of electronic pages, adding to the hundreds of millions pages already on-line. Because of its rapid and chaotic growth, the resulting network of information lacks of organization and the structure of Web sites becomes more and more complex. When searching and browsing the Web, users are very often overwhelmed by huge amount of information and are faced with a big challenge to find the most relevant information in right time. As a consequence, users often feel disoriented and get lost in that information overload that continues to expand.

Web personalization represents one of the most promising and potent remedies against the problem of information overload. Today, the need of predicting the user preferences and customizing the interactions on a Web site according to the implicit/explicit interests and desires of users is more than ever evident. Besides, the ability of a site to engage visitors at a deeper level, and to successfully guide them to useful and pertinent information, is now viewed as one of the key factors for the site's ultimate success.

Personalization plays a fundamental role in many contexts and, dependently on the context, it may be used to achieve several goals, ranging from increasing customer loyalty in e-commerce sites to enable better search providing results satisfying user needs. Recommendation systems are one of the major examples of personalization systems. Such systems have shown to greatly help Web users in navigating the Web, locating relevant and useful information, and receiving dynamic recommendations from Web sites on possible products or services that match their interests. To build Web recommendation systems, the Web usage mining (WUM) methodology is one of the main approaches used in the literature. WUM involves the application of data mining and machine learning techniques to discover usage patterns (or build user models) through the analysis of Web users' historical navigational activities. These models may be properly exploited to realize the different personalization functions.

In this way, in the design of a Web recommendation system based on the WUM methodology, three main phases may be distinguished [10]:

- *Web data preprocessing*: usage data are collected and preprocessed to identify user sessions.
- *Knowledge discovery*: useful usage patterns and recommendation models are discovered from preprocessed data.
- *Recommendation*: the discovered models are exploited to deliver intelligent recommendations.

In the process of knowledge discovery, different facets have to be addressed in order to mine models useful for the determination of interesting and accurate recommendations. In fact, the Web is a complex and heterogeneous network of interconnected components. Moreover, Web data are highly characterized by vagueness and imprecision. Their fuzzy and uncertain nature gives rise to the necessity of relevant intelligent techniques able to process the dif-

* Corresponding author. Tel.: +39 0805442456; fax: +39 0805443196.
E-mail address: torsello@di.uniba.it (M.A. Torsello).

1568-4946/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2009.12.040

ferent kinds of Web uncertainty that cannot be heavily processed through the traditional and precise techniques. Due to their characteristics, Computational Intelligence (CI) [18] techniques reveal to be appropriate to deal with this uncertainty and to develop Web-based applications tailored to user preferences. The main reason behind this success seems to be the synergy resulting from CI paradigms, such as fuzzy logic, neural networks and genetic algorithms. Rather than being competitive, each of these computing paradigms provides complementary reasoning and searching methods that allow the use of domain knowledge and empirical data to solve complex problems. On the basis of these considerations, CI techniques have been combined together in different ways leading to several hybrid schemes which, exploiting the strengths of each involved computing paradigm, can achieve high intelligence degrees.

This paper[1] is intended to propose NEWER (NEuro-fuzzy WEb Recommendation), a Web recommendation system that exploits CI techniques to dynamically suggest users interesting links according to their interests. In NEWER, a neuro-fuzzy methodology is implemented in order to discover interesting user navigational patterns and to derive a recommendation model useful for the suggestion of links considered relevant for the users. According to the WUM methodology, in the proposed system, the interests of users are implicitly derived in form of user categories by analyzing the usage data stored by the Web server in log files. Then, on the basis of the derived knowledge about user interests, a recommendation model is discovered via CI techniques to generate intelligent suggestions.

Specifically, as concerns the identification of user categories, in NEWER, a fuzzy clustering technique is employed. The use of this technique enables the generation of overlapping clusters, so that a user can belong to more than one category capturing in this way the overlapping interests of users.

To create the recommendation model which is exploited to provide intelligent predictions about Web pages to be suggested, a neuro-fuzzy approach is employed. This hybrid approach permits to exploit the learning capabilities of neural networks to derive a recommendation model expressed in a comprehensible form, as a set of fuzzy rules which can be easily understood by humans.

The rest of the paper is organized as follows. Section 2 provides a review of works that exploit the WUM methodology in the process of knowledge discovery from Web data. In Section 3, the architecture of the NEWER system is presented. Section 4 describes the step of log file preprocessing. In Section 5, the knowledge discovery step is detailed and in section 6, the ultimate recommendation step is described. Section 7 provides results obtained by testing the NEWER system both on synthetic data and real world data. A comparison between NEWER and other recommendation approaches is presented in Section 8. Finally, in Section 9 some conclusions and future research directions are drawn.

2. Review of related usage mining approaches

A WUM methodology provides a complete process for the extraction of models from usage data encoding the behavior and the interests of users. These models may be automatically exploited by a personalization system to personalize its services.

In this section, we give some examples of personalization systems that exploit techniques underlying the WUM methodology for mining knowledge from Web usage data.

Analog [20] was one of the pioneer personalization systems based on the WUM methodology. In such a system, the mining process is organized into two main components performed offline and online with respect to the server activity. In the offline component, past user activity stored in log files is processed by a geometrical clustering approach to create clusters of user sessions. Then, the online component creates active user sessions which are classified into one of the clusters previously identified. This permits to identify pages related to those in the active session and to return the requested page with the list of related documents.

Mobasher et al. [9] propose a usage-based Web personalization system taking into account both the offline tasks related to the mining of usage data, and the online process of automatic Web page customization based on the mined knowledge. In particular, usage mining tasks involved the discovery of association rules and the derivation of URL clusters. Once the mining tasks have been accomplished, the frequent itemsets and the URL clusters are used to provide dynamic recommendations to users based on their current navigational activity.

A WUM approach has been also exploited in SiteHelper [14], a system that has been designed to adapt Web pages to the user needs. In this system, usage mining techniques are employed to build a set of rules that represent the user interests. The discovered rules are used by the system to recommend new or update Web pages to the users according to their interests.

WUM techniques have been employed in KOINOTITES [17] in order to customize information to the needs of individual users. More specifically, such system identifies user communities which model groups of visitors in a Web site having similar interests and navigational behavior. These communities are exploited by the administrator of the site to improve the organization of the site or as input to a personalization system to dynamically make recommendations to Web users. The mining component of KOINOTITES includes four steps that perform the main functions: data preprocessing, session identification, pattern recognition and knowledge presentation. In pattern discovery, usage models of users are extracted by a variation of *Cluster Mining*, a simple graph-based clustering algorithm.

SUGGEST [2] is another example of WUM system designed to provide useful information to make easier the Web user navigation. SUGGEST adopts a two-level architecture composed by a offline creation of historical knowledge and a online engine that understands user behavior. It creates clusters of related pages based on user past activity and then classifies new users by comparing pages in their active sessions with pages inside the clusters created. A set of suggestions is then obtained for each request.

In [12], Nasraoui and Petenes present an intelligent Web recommendation system based on WUM to discover useful knowledge about user access patterns. In the mining process, log files are processed to identify user sessions. Then, user profiles are extracted by categorizing the identified sessions through the Hierarchical Unsupervised Niche Clustering algorithm. Finally, in the recommendation engine based on a fuzzy approximate reasoning, the mined knowledge is exploited to determine a set of links to be recommended.

Albanese et al. [1] proposed a Web personalization system based on a usage mining strategy consisting in two phases. In the first phase, a fuzzy unsupervised clustering algorithm is used to classify users by deriving groups of users appearing to be similar. The second phase performs the reclassification of users by taking advantage from the interactions of each user with the Web site.

In [8], the authors proposed a Web recommendation system based on a maximum entropy model. In such a system, different levels of knowledge about the user navigational behavior are combined to generate recommendations for new users. This knowledge includes page-level clickstream statistics about the past navigations of users with the aggregate usage patterns discovered through the WUM methodology.

[1] A preliminary version of this paper has appeared in the Proceeding of the 1st International Conference on the Application of Digital Information and Web Technologies (ICADIWT 2008).