

# An Analysis of Web Mining-based Recommender Systems for E-commerce

Luo Ya

Guizhou University of Finance and Economics

**Abstract**-This article proposes a framework of Web mining-based recommender systems for e-commerce. Building on clustering analysis of data involving Web usage, content and structure, the author demonstrates how to provide users with effective recommender services according to the mining results obtained by recommender engine. Finally, the author reaches his conclusion of the advantages and practicalities of Web mining-based recommender systems for e-commerce.

**Keywords**-E-commerce, Clustering, Recommender Systems, and Personalized Services

## I. INTRODUCTION

As the Internet and E-commerce grows popular, E-commerce systems provide users with vast commodity information and increasing choices. Meanwhile, lost in such vast information, users often find it difficult to locate satisfactory items, for e-commerce systems cannot offer one-to-one recommendation like a salesperson.

Recommender systems for e-commerce can simulate salespersons to communicate directly with users and help them find satisfactory commodities, successfully getting browsers into buying, effectively developing user relationships and increasing sales volume. Currently, almost all the large-scale e-commerce systems, such as that of Amazon, eBay, Dangdang and CDNOW, use various recommender systems in different degrees to provide their users with personalized services. Thanks to great progress in theory and practice, successful recommender systems for e-commerce will create enormous economic benefits.

## II. RECOMMENDER SYSTEMS FOR E-COMMERCE AND RELATED TECHNOLOGIES

### A. Data Mining and Recommender Systems

Data mining refers to the process of extracting out of a great deal of data, underlying information unknown yet but potentially useful. Recommender systems are an example how data mining can serve e-commerce. Data mining-based recommender systems can study user behavior and attributes; can establish user profile based on long-term user behavior analysis of, among others, browser record, trading details, registered record, ratings, and voting data etc; or can generate recommendation through the useful information obtained out of current behavior of a user, such as current sessions, shopping cart information, and browser record of interested items etc.

### B. Recommender Systems for E-commerce

#### 1) Recommender Systems and Personalized Service

Personalized services cater for each user's interests as much as possible while he or she is surfing the websites, with constant adjustment to adapt to the user's interest, making him/her feel like being served as the only VIP customer. Personalized service features enable recommender systems to adapt to individual needs and create an electronic shop offering unique, personalized shopping experience.

#### 2) Components of Recommender Systems

A recommender system can consist of three modules:

- **Input module:** Collecting and update user information. Inputs can come from individual users or groups. The inputs by individuals refer to the ones by target users, i.e. customers who desire to have access to recommendation. They are required to make comments and express their preferences. Those inputs include explicit and implicit browser record, keywords or commercial property, user ratings, user text comments, editor's recommendation and user purchase history etc. The inputs by groups refer to the comments data in collective form, including item property, group purchase history, text comments, and ratings. Different recommender systems have different input information, which, in turn, will create different recommendation.
- **Output module:** deliver recommendation to users after the system has processed the inputs. Different recommender systems have different outputs.
- **Recommendation methodology module:** process the inputs and generate outputs. It is the core of recommender systems and determines the system's performance. It is supported by recommendation techniques and algorithms.

#### 3) Classification of Recommender Systems

User-centered recommender systems can be divided into:

- **Non-personalized Recommender Systems:** The recommendation offered to a user may be based on average comments by other users, sales ranking or editor's recommendation. This recommendation technique is independent from all the users, which means the recommendation to users is identical.
- **Features-based Recommender Systems:** They can create a recommendation list for users based on commodity features. Users can use such a system like a search engine: manually input desired features and click "Go".
- **Commodity Relativity Recommender Systems:** create recommendation based on commodity relativity.
- **User Relativity Recommender Systems** (also known as

Recommender Systems with Collaborative Filtering): search neighbors nearest to current user, and then create recommendation for current user based on the purchase history and ratings of the neighbors.

### III. WEB MINING-BASED, PERSONALIZED RECOMMENDER SYSTEMS

#### A. Systems Components

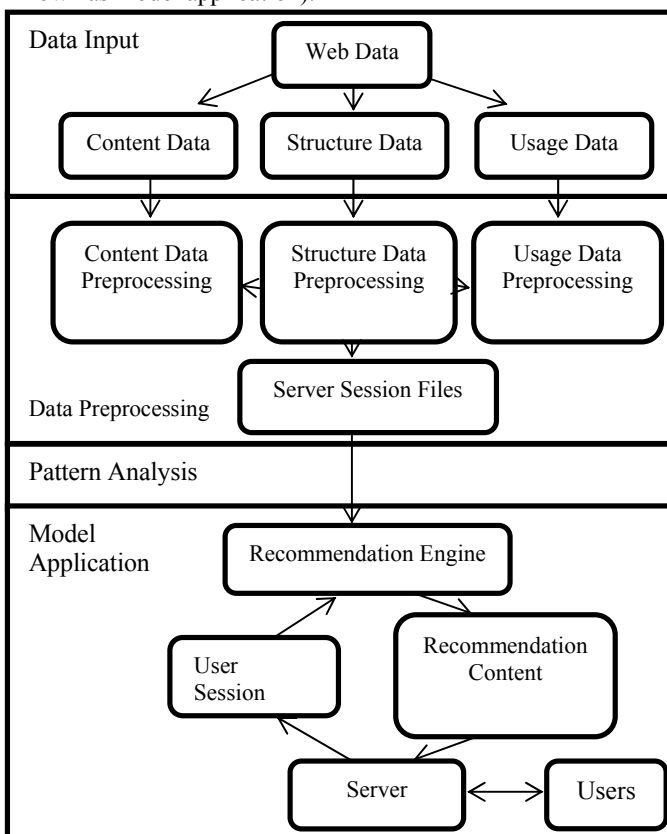
The key issue involving personalized recommendation is to have deep understanding of a great number of nonregistered user behavior models. As conventional collaborative filtering hardly deals with nonregistered user information, Web mining-based personalized recommender systems work it out effectively.

Web mining-based recommender systems for e-commerce consists of the following components:

- Web usage mining: find useful information out of a user's commercial website visit record.
- Web content mining: analyze Web contents, such as texts, sound, images, video, and etc., extracting useful information.
- Web structure mining: find useful models out of topologic structures that form Web hyperlinks.

#### B. Systems Architecture

Web mining-based personalized recommender systems are running in two processes: offline and online ones. The offline process deals with preprocessing and Web mining (also known as model acquisition); whereas, the online process works with real-time recommender engine (also known as model application).



#### 1) Model Acquisition

##### (1) Data Input

In traditional recommendation methods, data inputs come from either individual users or groups. "Individual user inputs" refers to inputs from target users who, in order to have access to recommendation, have to make comments and express their preferences, including explicit browser records, keywords and item property, and user purchase history; "group inputs" refers to comment data in collective form, including item property, group purchase history, text comments, and ratings. Obviously, in traditional methods, the side effects of requiring the user to make comments may undermine the usability of the whole system.

Unlike traditional recommendation methods, the Web mining-based recommender system does not interfere with browsing; its data inputs primarily consists of Web content data, Web structure data, and Web usage data of a website. Among them, the Web content data is the major information from an e-commerce site, i.e., a collection of network objects the e-commerce site has passed on to the user and their relations; the Web structure data includes structure data of external links on the web pages and the data of internal structure as presented by the contents on each page and arranged with HTML or XML tags; Web usage data is the data that describes the pattern in which the users visits the site, and can be collected by means of Web server log configuration among others.

##### (2) Data Preprocessing

The purpose of data preprocessing is to convert data from various sources into data abstraction concepts that are suitable and necessary for data mining. Data preprocessing basically includes Web content data preprocessing, Web structure data preprocessing, and Web usage data preprocessing. Both Web usage preprocessing and Web content preprocessing are based on preprocessing of Web structure data, the task of which is to describe the topological structure diagram and directional webpage file link diagram of the site, identify the Frame structure, Head page, Navigation page, and Content pages, as well as the methods to request a link to the files on each page, for example, GET, POST, and Head. Server session ① will arise from preprocessed Web usage data. And if the preprocessed Web content data can be integrated with session ①, session ② will arise; the two types of server sessions provide subsequent Web mining algorithms with data that can be processed directly.

- Of Web usage data: create a Web transaction file that consists of a collection of all the pages and user transactions after data cleansing, session identification, page identification, and etc. In this transaction file, pages collection (P) denotes the collection of URL addresses; each transaction (t) of user transactions collection (T) is represented by a multi-dimensional vector in space P; the vector value can be 0/1 or weighted by the time window of the transaction on the page to indicate the degree of the user's interest in this

page.

- Of Web content data: create a Web content file through feature extraction and feature selection from text data and metadata. In this content file, contents collection ( $F$ ) denotes the collection of page contents extracted; each page ( $p$ ) can be represented by a multi-dimensional vector in space  $F$ ; each vector denotes the weight of page in this content dimension, so as to help analyze the similarity between different page contents. Metadata can be weighted by the factor provided by designers of the commercial website, for example, by TFIDF (term frequency-inverse document frequency) in regard to text content features.
- Of Web structure data: acquire, through link extraction and link weighting, Web structures file that presents the interlink between different commercial website pages. In this structures file, each page ( $p$ ) in pages collection ( $P$ ) can be represented by a multi-dimensional vector of the space of all the pages; each vector is weighted by 1 if there is a linkage to  $p$  or by 0 if otherwise.

### (3) Pattern Analysis

Pattern analysis has the following three tasks: 1. identify various patterns in Web data with Web mining algorithms; 2. use appropriate tools and techniques to analyze, interpret, and visualize these patterns, and pick out useful and intelligible knowledge rules from them; 3. store these knowledge rules on appropriate media, so as to support the recommender engine.

- Of Usage features: the transaction file created after Web usage data preprocessing can be used as input for data mining algorithm. Various transaction clusters can be acquired by cluster analysis with different clustering algorithm, so as to create a scenario as much as possible where browse patterns of users in the same transaction cluster are similar and that of users in different transaction clusters are different. With transaction cluster analysis, a series of useful usage features can be acquired by mining the transaction file; each usage feature indicates the behavior feature of a certain user group, where the members have similar interests and information requirements.
- Of Content features: Be consistent with the representation of usage features. As clustering algorithm directly deals with extracted contents other than pages, interchanging the rows and columns of page-content matrix acquired after data preprocessing. During clustering, pages with similar contents are grouped into a cluster as much as possible. More, different user groups may have interests in different parts of one page. Therefore, content features can indicate overlapped interests and information requirement.
- Of Structure features: Be consistent with the representation of usage and content features. During mining, pages with similar links are grouped into a structures cluster as much as possible, so as to obtain

the weight of all the pages in this structures cluster. The weight of different pages in the structures cluster can indicate the value of this page for users in this structures cluster.

Consistent representation of usage, content, and structure features (all represented by multi-dimensional vectors in original page space) enables the online recommender engine to collect together easily these features to offer effective personalized recommender services.

### 2) Model Application

In the online process, the recommendation engine generates real-time recommendation collection based on current user sessions. Current user sessions ( $S$ ) are represented by a multi-dimensional vector in the pages space ( $P$ ); 1 is assigned to the vector value if the user has visited this page, and 0 is assigned if otherwise. Usage, content, and structure features are also represented by multi-dimensional vectors in the pages space. A cosine similarity function calculates the matching coefficient between  $C$  and  $S$ , generating a recommendation collection that may include hyperlinks, advertisement, texts, commodities, and etc., with recommended items added as hyperlinks to the last page of a current session.

In a Web mining based personalized e-commerce recommender system, the recommender engine matches current user sessions with the various knowledge rules extracted in the pattern analysis phase, which in turn generates personalized recommendation pages collections. The pages collections will be attached as hyperlinks to the latest page request from the user. The separation of the offline module and online module enables the recommender engine to handle a large amount of data, significantly improving the scalability and response speed of the personalized e-commerce recommender system.

## IV. SUMMARIES

Successfully separating model acquisition from model application, Web mining-based recommender systems for e-commerce can deal with large data volume, improving greatly scalability and response speed. They integrate all the results from Web usage, content, and structure mining and can provide quality personalized recommender services in the cases where there is less usage data or frequent change of website content.

## REFERENCES

- [1] Qing, Hai. A Study on the Core Technologies of Recommender Systems for E-commerce [D]. Beijing University of Technology, 2009.
- [2] Deng, Ailin. A Study on Key Technologies of Recommender Systems for E-commerce [D]. Fudan University, 2003.
- [3] Han, Huijiun. A Study on Personalized Recommender Systems for E-commerce [D]. Shanghai Jiaotong University, 2007.
- [4] Deng, Xiuqin, and Jiang, Lianhua. A Study on Recommender Systems for E-commerce [J]. Journal of Liaodong University, 2005, 7(4): 38-42.

- [5] Li, Jianxia. A Study on Intelligent Recommender Systems Technology and Its Application for E-commerce [D]. Chongqing University, 2009.
- [6] Li, Ting. A Study on User Interests and Cooperation-based, Personalized Recommender Systems for E-commerce [D]. Harbin Institute of Technology, 2006.