IMPLEMENTATION OF WEB-STORED SYSTEM BY BUILDING A WEB PAGE RECOMMENDER SYSTEM

Phyo Phyo Kywe
Computer University (Mandalay)
prinklady.mdy@gmail.com

Abstract

As Recommender systems are an important part of recent E-commerce, this paper is proposed to recommend the items to a user based a description of the item of the user's interests and them allows the user to see more details about a selected product (computer accessories) and to recommend the best seller products according to the user's preferences. We also propose a framework for building a web-page recommender system for especially computer accessories. To construct a recommender system, information filtering is considered. By using the vector space model, the information filtering processes can be performed prior to the actual recommendation process. As a result, it generates the list of information which suits the user's most buying products/interests and it enables the increase of sales by suggesting to the users selected products on offer.

Keywords: Recommender System, Collaborative Filtering, Vector Space Model

1. Introduction

Many recommender systems have been designed and implemented for various types of items including newspaper, research papers, emails, Usenet news, books, movies, music, restaurants, web pages, and e-commerce products [2, 3, 5]. To construct a recommender system, two information filtering methods for providing the recommended information are considered:(1) by analyzing the information content, i.e., content-based filtering, and (2) by referencing other user's access behaviors, i.e., collaborative filtering technique relies on the Nearest Neighbor (NN) clustering algorithm to compare and search for other user's ratings which are closely resemble to the user's profile. The Nearest Neighbor clustering algorithm, however, suffers from the scalability problem, ie, the execution time grows linearly with the number of the users in the system. This collaborative filtering approach also suffers from the *cold-start* problem. Another alternative technique called content-based filtering has additionally been considered in the recommender system. The content-based filtering is based on the analysis of the information contents, eg; by applying a classification method on textual content [2].

In this paper, we use the collaborative filtering based on the method of mining user access patterns. The proposed system is applied to construct a recommendation of web-store, which automatically generates recommended lists of the best seller products/ interests. The selling and buying organization of computers and accessories device is chosen as case study one. Besides, the recommendation systems for choosing items are also provided.

The remainder of this paper is organized as follows. In Section 2, it describes the related works. In Section 3, it describes background theory and the Vector Space Model. Section 4 presents the proposed system design. Finally, the conclusion is presented in Section 5.

2. Related Works

In this section we briefly present some of the research literature related to vector space model, collaborative filtering, recommender systems, data mining and personalization.

Vector Space Model is first used in the SMART information retrieval system. In applications, relevancy rankings of documents in a keyword search can be calculated, using assumptions of document similarities theory, by comparing the deviation of angles between each document vector and the original query vector where the query is represented as same kind of vector as the documents [6].

A variety of collaborative filtering algorithms have been designed and deployed. The Tapestry system relied on each user to identify like-minded users manually. GroupLens and Ringo, developed independently, were the first to automate prediction. Typical algorithms compute similarity scores between all pairs of user; predictions for a given user are generated by weighting other user' ratings proportionally to their similarity to the given user. A variety of similarity metrics are possible, including correlation, mean-squared difference, vector similarity, dependency network models, clustering models, and models of how people rate items. Collaborative filtering has also been cast as a machine learning problem and as a list-ranking problem. Singular Value

Decomposition (SVD) was used to improve scalability of collaborative filtering systems by dimensionality reduction.

Pure information filtering systems use only content to make recommendations. For example, search engines recommend web pages with content similar to (e.g., containing) user queries. In contrast to collaborative methods, content-based systems can even recommend new (previously un-accessed) items to users without any history in the system. A content-based book recommender using information extraction and machine learning techniques for text categorization is proposed in [1].

We describe the proposed recommendation system framework with methods in section 3.

3. Background Theory

3.1. Proposed Recommender-System Framework

Data mining or knowledge discovery in database (KDD) has emerged recently as an active research area for extracting implicit, previously unknown, and potentially useful information from large database, mining techniques into the IR context, specifically as the information filtering tools for the recommender system framework. Two types of information filtering can be accomplished by using data mining: content-based filtering and collaborative filtering. In the proposed framework, the association rules mining technique is applied as the content-based filtering where the data set is the keyword matrix. Collaborative filtering retrieves the information for a particular user by referring to other user evaluations on the information content. The method of mining is applied as the collaborative filtering technique. The overall process for designing and implementing a recommender system is illustrated in Figure 1. The process involves the following five steps.

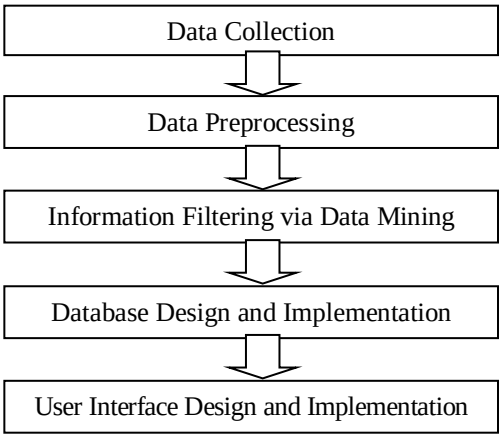

Figure 1. Process for designing and implementing a recommender system

- Data collection: This initial step involves the collection of data sets for executing the data mining algorithms.
- Data Preprocessing: This step is required to clean and transform the collected data sets into the formats which are suitable for the data mining algorithms. This step includes the data reduction and selection techniques to improve the efficiency of the data mining algorithms.
- Information Filtering via Data Mining: This step is the core process of the recommender system framework, where the data sets are analyzed and the data mining algorithms are applied as the information filtering tools to generate and discover any useful and interesting recommended outputs.
- Database Design and Implementation: To improve the efficiency of data and information access and retrieval, the database for the recommender system is designed and implemented for all related data sets including the textual content, link structure, and the recommended lists of Web pages.
- User Interface Design and Implementation: The user interface acts as an intermediary between the user and the recommender system. This step involves the design and implementation of a Web (i.e, HTTP) server which receives the users' requests via the WWW, processes the requests by accessing the database, and responds by returning the results to the users. The user interface provides a recommendation function with the user personalization technique by requiring each user to log into the system in order to keep track of the preferences.

3.2. Nearest Neighbor Methods

The nearest neighbor algorithm simply stores all of its training data, here textual descriptions of implicitly or explicitly labeled items, in memory. In order to classify a new or unlabeled item. The algorithm compares it to all stored items using a similarity function and determines the “nearest neighbor” or the k nearest neighbors. The class labels of the nearest neighbors.

The similarity function used by the nearest neighbor algorithm depends on the type of data. For structured data, a Euclidean distance metric is often used. When using the vector space model, the cosine similarity measure is often used. In the Euclidean distance function, the same having a small value in two examples is treated the same as that feature having a large value in both examples. In contrast, the cosine similarity function will not have a large value if corresponding features of two examples have small values. As a consequence, it is appropriate for text when we want two documents to be similar when they are about the same topic, but not when they are both not about a topic.

The vector space approach and the cosine similarity function have been applied to several text classification applications and, despite the algorithm's unquestionable simplicity; it performs competitively within more complex algorithms [4].

3.3. Vector Space Model

Vector Space Model is an algebraic model for representing text documents as vectors of identifiers. It is used in information filtering, retrieval, indexing and relevancy rankings.

A document is represented as a vector. Each dimension corresponds to a separate term. If a term occurs in the document, its value in the vector is non-zero. Several different ways of computing these values, have been developed. One of the best known schemes is TF/IDF weighting [6].

In this system, Vector Space Model uses weight each term based on its inverse document frequency (IDF) in the document collection. The weight of a term t_i in a document is given by

$$W_i = tf_i * IDF$$

Where tf_i, $i=1,....,n$ is the term frequency of the term t_i in the document . The term frequency can also use to the normalization of term frequency that is defined as

$$Tf_i = 0.5 + 0.5\ (tf_i)/\ max\ (tf_i)$$

IDF is inverse document frequency which is defined by

$$IDF = log\ (N/\ df_i)$$

Where df_i is the number of documents in a collection of N documents in which term t_i occurs.

3.4 Collaborative Filtering

Figure (2) illustrates the collaborative filtering process of proposed system.

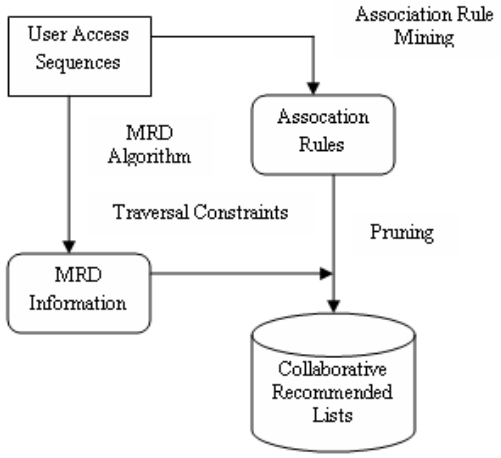

Figure 2. Information Filtering by applying Data Mining Algorithms

This process applies the association rule mining technique on the user access sequences in order to generate a set of rules. Then by using the user access sequences, the traversal constraint called the Minimum Reaching Distance (MRD) information is constructed based on the shortest-path in graph theory The MRD information captures the actual user access behaviors on the Web site. The MRD information is then used to prune the rules obtained from the association rule mining process in order to increase the prediction accuracy and reduce the space complexity.

To determine the most similar match for a given item, the algorithm builds a similar items table by finding items that customers tend to purchase together. We could build a product-to-product matrix by iterating through all item pairs and computing a similarity metric for each pair. However, many product pairs have no common customers, and thus the approach is inefficient in terms of processing time and memory usage. The following iterative algorithm provides a better approach by calculating the similarity between a single product and all related products.

For each item in product catalog, $\boldsymbol{I_1}$
  For each customer **C** who purchased $\boldsymbol{I_1}$
    For each item $\boldsymbol{I_2}$ purchased by
      Customer **C**
      Record that a customer purchased $\boldsymbol{I_1}$ and $\boldsymbol{I_2}$
  For each item $\boldsymbol{I_2}$
    Compute the similarity between $\boldsymbol{I_1}$ and $\boldsymbol{I_2}$

4. The Departmental Store Implementation and Design Approach

In this section, we explain each parts of web-store recommender system. First, we describe the use case diagram for departmental store in Figure 3. There are three levels: guest, user and administrator. The guest can view the recommended products of the system and go to register. And them the user who has already registered can go to login process. The administrator, the authorized person, has a chance for addition, updating and deletion if necessary.

4.1 Use Case Diagram for Departmental Store

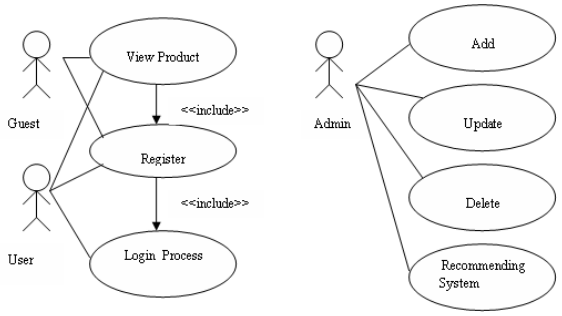

Figure 3. Use Case Diagram for Departmental Store

4.2 System Flow Diagram for Recommendation System

In Figure 4, the administrator wants to update products (computer accessories), add new products (computer accessories), delete products (computer accessories) and constructs a recommender system using TF-IDF described in Section 3.3.

Moreover, a user can look for view product and login register, and then he can see best seller product (computer accessories) for recommendation size. The user is looking for view products, choose purchase items. User favorites are either calculated by keeping track of the categories of the items purchased by users or may be set manually by the user.

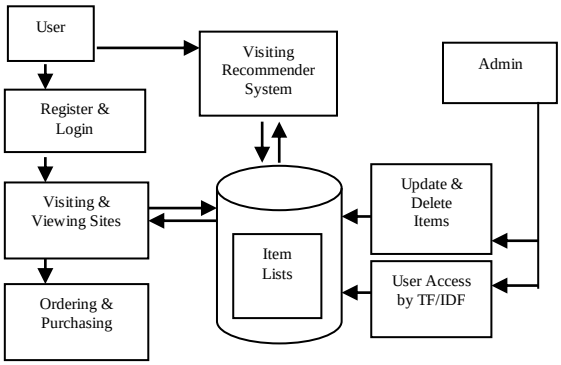

Figure 4. System Flow Diagram for Recommendation System

4.3 Interface Design of Showing Best Sell Products

The interface design in Figure 5 is created by C# Language with SQL server.

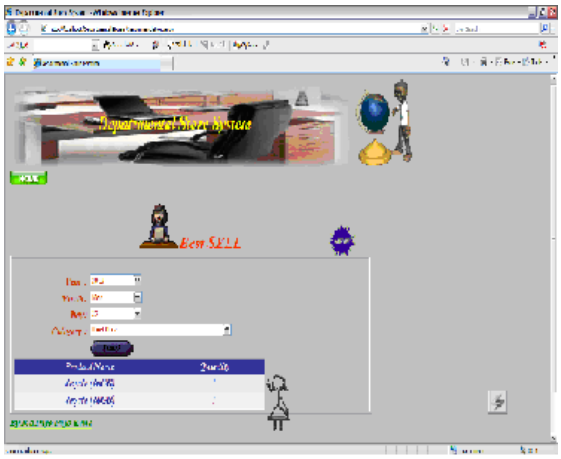

Figure 5. Showing Best Sell Products

The system shows the user interested items (especially computer accessories) for each day. Therefore user can choose for most interested items for computer accessories to support their online shopping. Moreover this system provides an effective form of targeted marking by creating a personalized shopping experience for each customer. This system also shows the updated price of computer accessories and the new products of computer accessories. Using this system, the users can know more about computer accessories.

5. Conclusion

Recommendation algorithms provide an effective form of targeted marking by creating a personalized shopping experience for each customer. A good recommendation algorithm is scalable over very large customer bases and product catalogs, requires only sub second processing time to generate online recommendations, is able to react immediately to changes in a user's data, and makes compelling recommendations for all users regardless of the number of purchases and ratings.

Many people now use the Internet to gain access to external services such as online e-store, online education and online library. Anyone with an access to Internet connection can take advantages of an online E-store. It can offer short time for shopping and searching the required items. This paper is proposed to improve the efficiency and effectiveness of the information retrieval process. This paper is designed to enhance the interaction by analyzing user access behaviors on the system. In addition to the content analysis, information is also retrieved according to each individual's preferences and by recommendation to other users.

6. References

[1] A. Pppescual, L. H. Ungar, D. M. Pennock and S. Laswrence, "Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments", In proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (UAI. 2001), pp. 437-444, Morgan Kaufmann, San Franciso, 2001.

[2] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. T. Riedl, " Item-based Collaborative Filtering Recommendation Algorithms", In proceeding of the Tenth International Conference, pp. 285-295, 2001.

[3] J. B. Schafer, J. A. Konstan and J. Riedl, "E-commerce Recommendation Application", Data Mining and Knowledge Discovery, 5(1/2): pp. 115-153, 2001.

[4] M. J. Pazzani and D. Billsus, "Content-based Recommendation Systems", In proceeding of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Seattle, pp. 394-401, WA(2004).

[5] R. J. Ney and L. Roy, "Content-based Book Recommending using Learing for Text Categorization", In proceedings of the Fifth ACM Conference on Digital Labraries, pp. 195-204, 2000.

[6]http://en.wikipedia.org/wiki/Vector_space_model#cite_note-0