DIET RECOMMENDER SYSTEM USING WEB DATA MINING

Nandish Shah[1], Ishani Shah[2]

[1]*Department of IT Engineering, Inst. Of Tech, Nirma University, Ahmedabad (India)*

[2]*Department of IT Engineering, Inst. Of Tech, Nirma University, Ahmedabad (India)*

ABSTRACT

In this fast paced and busy scheduled life, people very seldom are giving importance to the quality of food they are eating. Fast food consumption is increasing dramatically among the people over the past few years. And this consequently, has lead to unhealthy food habits among the people of all generation. Hence it has become very essential for the people to have a good balanced nutritional healthy diet. Not only does it reduce the tendency of falling ill, by building up the immune system but also increase the resistance to diseases such as high cholesterol, obesity, high blood pressure, ulcers and diabetes and heart attacks. For a better recovery from the illness, different individuals have different needs according to their medical history ,social backgrounds and nutrient and mineral requirements. Here we present a proposal of healthy food habits and eating system based on web data mining, to discover hidden patterns and business strategies from their customer and web data, which would track your eating habit and recommend the types of food that improve your health and avoid the type of food that raise the risk of illness.

Keywords*: *Data Mining, Web Mining, Healthy Diet

1. INTRODUCTION

With the pace of life drastically accelerated in the current scenario, fast-food has become more and more popular in daily life, which has basically lead to deterioration in the eating habits of the people. An important healthy eating tip is to eat foods with antioxidant properties to defend the body from viral and bacterial attacks and to maintain sufficient water intake. In order to solve this problem, we present a proposal of healthy eating analyzing and recommender system based on web data mining, which would track the eating habits and recommend the types of foods that improve health and avoid the types of foods that increase the risk of various illness and diseases. In this paper, we introduce some basic knowledge of web mining and data mining. Then a web-based data mining solution to healthy eating analyzing and recommender system is brought up. Finally we have shown the system implementation proposal for this system which is followed b the conclusions we have drawn.

2. WEB MINING

The discovery and analysis of useful patterns and information from the World Wide Web is referred to as Web mining. Businesses might implement the concept of Web mining to understand customer behavior, evaluate the effectiveness of a specific Web site, or quantify the success of a marketing campaign. For instance, marketers use Google Trends and Google Insights for Search services, that detect the popularity of various words and phrases used in Google search queries, to learn what people are interested in and what they are interested in buying.

3. DATA MINING

Data mining provides understanding into corporate data(that cannot be achieved with OLAP), by finding concealed patterns and relationships in large databases and deducing rules from them to predict future tendencies. These patterns and relations are used to guide decision making and predict the effect of those decisions. The types of information obtained from data mining include associations, classifications, sequences, clusters, and forecasts. Here is a brief idea.

3.1. ASSOCIATIONS

Associations are occurrences linked to one event.

For instance, a study of supermarket purchasing patterns might reveal that, when chips are purchased, a coke drink is purchased 65 percent of the time, but when there is an advancement, coke is purchased 85 percent of the time. This information helps managers make better decisions because they have learned the profitability of a promotion.

3.2. SEQUENCE

In sequences, events are linked over time. For example, if a house is purchased, a new washing machine will be purchased within two weeks 64 percent of the time, and a microwave will be bought within one month of the home purchase 46 percent of the time.

3.3. CLASSIFICATION

Classification recognizes patterns which describe and define the group to which an item belongs by examining existing items that have been classified, and by deducing a set of rules. For example, businesses such as credit card or telephone companies worry about the loss of steady customers. Classification helps discover the characteristics and behavior of customers who are likely to leave and thereby provide a model to help managers predict who those customers are, so that the managers can devise special campaigns to retain such customers.

3.4. CLUSTERING

Clustering works in a manner similar to classification however no groups are yet been defined in this case. A data mining tool can discover different groupings within data ,such as finding attracted groups for bank cards or dividing and classifying a database into groups of customers based on demographics and types of personal investments.

3.5. FORECASTING

Forecasting uses the concept of predictions in a different way. It uses a series of existing values to forecast what other values will be. For example, forecasting might find patterns in data to help managers estimate the future value of constant variables, such as sales figures.

4. DATA MINING PROCESS

4.1. DATA COLLECTION

In terms of functionality, data acquisition selectively obtains data from the outside web environment to provide input and resources for the latter data mining. The data source that the web environment provided includes the web pages, hyperlinks data and the history data of user visiting log. It is basically comprised of relatively independent processes which are data search, data selection and data collection.

4.2. DATA PREPROCESSING

Data preprocessing processes and rebuilds the source data acquired in data acquisition phase and creates the data warehouse and data marts of related areas to create basic platform for data mining process. Data preprocessing is preparation for data mining and it mainly includes data scrubbing, data integration, data conversion, data reduction, etc.

4.3. INFORMATION FILTERING VIA DATA MINING

In this process 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.

4.4. DATABASE DESIGN AND IMPLEMENTATION

In order to improve the efficiency of data and information access and retrieval, the database for the eating habit recommender system is designed for all related data sets including the textual content, link structure and the recommended lists of Web pages.

4.5. USER INTERFACE DESIGN AND IMPLEMENTATION

It acts as an intermediate between the users and the recommender system. This step involves the design and implementation of a Web (i.e., HTTP) server which receives the users' requests via WWW, processes the requests by accessing the database, and responds by returning the results to the users.

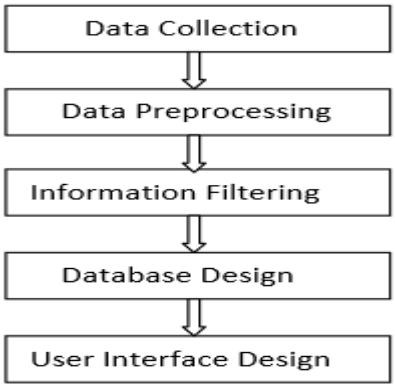

Figure1.Data Mining Process

5. DESIGN OF HEALTHY EATING SYSTEM

5.1. FUNCTION DESCRIPTION:

5.1.1 DATA ACQUISITION

We assume that there is a website where people could take their orders over the internet just like that in the restaurant. We acquire the eating data in the database which could track people's recipe record. Also they can directly input their eating data into the database through the website. This is the firsthand material cum input for this system, e.g. the eating time, food name, amount, material, etc.

5.1.2. DATA MINING

We use data mining algorithms s like classification, clustering, association rules, etc. in the data mining process to extract the useful information of people's eating habit. We analyze the nutritive structure of each kind of food and calculate the fat, energy, vitamin percentage in the recipe. Then we use the classification mining algorithm to process the composition data and give out the result whether the diet is healthy or not.

5.1.3. HEALTHY EATING RECOMMENDATION

Once data mining process is accomplished you get much useful information. For ex. What does your diet lack of, what you have in excess, potential diseases, etc. On the basis of this data, we can recommend the healthy diet menu,tips, etc. according to the personal condition, thereby raising their health standards. We can also keep a

track of clients' individual preferences and recommend the related dishes to fulfill personalized need by using association rule mining. Thus it would provide better service and experience for clients.

5.2. A SYSTEM PROTOTYPE FOR RECOMMENDING HEALTHY FOOD

5.2.1. DATA COLLECTION AND PREPROCESSING

This comprises three relatively independent processes which are data search, data selection and data collection. The Web server system contains two types of databases:

(1) The content database for providing the information content such as food name, food time amount with the images and other types of data. The number of user preference food list collected from the Web site. For this purpose we implement a crawling and parsing program using Java Programming language to collect 2 different data sets: the textual and hyperlink contents.

(2) The server log database for recording the HTTP transactions (i.e., log records). Under HTTP, each request to the Web site is recorded as a transaction on the Web server. Due to the load-Balancing configuration, The HTTP traffic volumes of both servers are closely similar. From the traffic volumes, the first observation is that both servers yield a similar amount of daily traffic. Another observation is the patterns in a weekly traffic cycle ,i.e. the traffic volume is higher during the weekdays and relatively lower during the weekends. Based on this observation, to reduce the execution time, a traffic volume of one-week period is used to represent the input to the Web mining process, which the log records from both servers are combined into a single file using the Merge Sort algorithm.

5.2.2. INFORMATION FILTERING VIA DATA MINING ALGORITHMS

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 access behavior, i.e. collaborative filtering. Content-based filtering is achieved by applying the association rule mining technique on the simple classification of the user's recipe in the database and could also find the association between users' diet composition and some potential diseases. Then according to this relation we could recommend different groups of dishes, which would be better for people's health so that each input is used to represent a record for the association rule mining algorithm. The results are a set of rules in the form of "IF (preconditioned set of user diet and diseases) THEN (post- conditioned healthy food)." In order to simplify the model, only the single-consequent rules are considered. If the same precondition occurs in more than one rule, the post-conditioned dishes are ranked based on the confidence values of the rules.

The content-based filtering rules are such that the preconditioned Web pages imply the post conditioned Web page based on the similarity in the keywords (i.e., textual content).

The collaborative or social based filtering process applies the association rule mining technique on the user access sequences in order to generate a set of rules .For the recommender system association rule learning would be effective. For example, the rule {onions, potatoes}-{beef} found in the sales data of a supermarket would indicate that if a customer buys onions and potatoes together, he or she is likely to also buy beef .According with this theory, we could process clients' history data to find out their preferences. Then the system could recommend the related dishes to clients, which would enhance the quality of personalized service.

5.2.3. DATABASE DESIGN AND IMPLEMENTATION WITH A WEB-BASED USER INTERFACE

In our proposed framework, a database using the Relational Database Management System (RDBMS) is designed and implemented. This database stores the URLs (i.e., Web pages), keywords for the Web pages, the recommended set of rules from content-based filtering, the recommended set of rules from social-based filtering, user login information, and user profiles. The current version of the recommender system prototype uses MySQL as the choice for database implementation. MySQL provides a multi-threaded, multi-user, and robust SQL (Structured Query Language) database management system, which is suitable for the application of recommender systems. To provide a Web-based interface, the database is stored and linked to a Web server. The Web server is implemented on the H'ITP specification and has the following functions: listening for HTTP requests on a network, receiving HTTP requests made by user agents (usually Web browsers), serving the requests (accessing the database) and returning H'ITP response that contains the requested resources. The user accesses the recommender system by using a Web browser. The communication between the user and the system is carried out on the Internet via the HTTP request and response functions. The recommender system server provides the database which contains the processed information such as the recommended lists. The user profile resides on the server side to keep track of the user's preferences. A logical path exists between the user and his/her profile.

6. IMPLEMENTATION PROPOSAL

In this system, a website as the raw data source should be implemented, which is also used as an e-commerce platform for clients.

6.1. E-COMMERCE PLATFORM

The first problem we meet in constructing this system is how to acquire the firsthand users recipe data. To solve this realistic problem, a consumer-to-consumer e-commerce platform is needed, e.g. taobao.com. This C2C ecommerce platform would connect the buyers and sellers over the internet, which would create the opportunity to acquire the raw data of users. The e-commerce web site should let their clients take orders over the internet just like that in the restaurant. Then the web server could send clients' orders for food to the restaurants and later clients could have their meals in the booked restaurant.

6.2. USER INTERFACE

If Users don't take orders over the internet and they just want to input what they have during a day, the web site should fulfill this requirement.

6.3. LOG INFORMATION

The web site should take down the log information of user's action over the browser. These data could be used as data resource for data mining.

6.4. DATA PROCESSING IN BACKGROUND

Data processing engine is an indispensable part of this web site. First, the web server should collect user's diets data through the database and history log. Then the firsthand data should be handled by special data mining algorithms which could give out the evaluated results and recommended suggestions.

7. CONCLUSIONS

In this paper, a new framework based on data mining techniques is proposed to improve your health and avoid the types of foods that raise your risk for illnesses. The proposed framework is designed to enhance this interaction by analyzing user access behaviors on the system. In addition to the content analysis (i.e., content-based filtering)information is also retrieved according to each individual's preferences (i.e. user personalization) and by recommendation from other users (i.e. collaborative filtering).We suppose that there is a website where people could take their orders over the internet just like that in the restaurant. We acquire people eating habit data in the database which could track people's recipe record. Also people could input their eating data into the database through the website. Then we introduce a web data mining solution to e-commerce to discover hidden patterns and business strategies from their customer and web data, propose a new framework based on data mining technology for building a Web-page recommender system, which would be used as the basic frame work for healthy eating system. Finally we give out personalized recommendations for each person.

REFERENCES

[1.] A Survey on Web Data Mining Applications by Dhanashree S. Deshpande Emerging Trends in Computer Science and Information Technology -2012(ETCSIT2012) Proceedings published in International Journal of Computer Applications® (IJCA)

[2.] Prof. Sheetal Khanore, Prof. Rajendra Patil and Prof. Hiren Dand "Management Information System" in 2011

[3.] Srinivas Nowduri "Management Information systems and business decision making: review, analysis, and recommendations"

[4.] Data mining system and applications: A review. Mr. S. P. Deshpande, Dr. V. M. Thakare, Department of MCA, Amravati, India, International Journal of Distributed and Parallel systems (IJDPS) Vol. 1, No. 1, September 2010.

[5.] Design of Healthy Eating System based on Web Data Mining. 2010 WASE International Conference on Information Engineering, Xiaocheng Li, Xinliu, Zengjie Zhang, Yongming Xia, Songrong Qian, Shanghai, China

[6.] A Data Mining Framework for Building A Web-Page Recommender System Choochart Haruechaiyasak , Klong Luang, Pathumthani, Mei-Ling Shy, Shu-Ching Chen

[7.] Ken Laudon, Jane Laudon and Rajanish Dass "Management Information System" (11th Edition)

Biographical Notes

Mr. Nandish Shah is presently pursuing B. Tech. third year IT Engineering Department from Nirma University, Ahmedabad, India.

Ms. Ishani Shah is presently pursuing B. Tech. third year IT Engineering Department from Nirma University, Ahmedabad, India.