

Performance Analysis of Healthy Diet Recommendation System using Web Data Mining

Shilpa Dharkar, Anand Rajavat

Abstract— Medical study has revealed that people set a bigger possibility of countering free radicals and warding off illness by consumption of healthy foods and by increasing their resistant system. Due to the poor eating habits people suffer from many diseases. In the current scenario fast food become important food in daily routine because it is effortlessly available but taking fast food in routine may cause for disease like heart attack, diabetics etc. Healthier diets help us to maintain our health and keep us away from many diseases. For better recovery from diseases or surgery etc individual have special needs according to their medical profile, cultural backgrounds and nutrient requirements. Design and implementation of healthy diet recommendation system is based on web data mining which is the application of data mining technique help us to determine pattern from web. In terms of accuracy and time performance analysis of recommendation system using two decision tree learning algorithm ID3 and C4.5 and apply it on healthy diet application.

Index Terms— Data mining, decision tree, eating system, healthy diet, web data mining, ID3, C4.5, bagging.

1 INTRODUCTION

In a machine learning process the classification can be described as a supervised learning algorithm. Data records are belong to class on the bases of knowledge of class it assign a class labels to data to co- design and co develop software and hardware, and hence, such components. However, incorporation of that deal with knowledge extraction from database records and prediction of class label from unknown data set of records. We can define classification is a development in which specified set of data records is separated into training and test data sets. For validating the model we required the test data record and for constructing the classification model training data set is required. The constructed classification model is used for classifying and predicting new data set records. These new data set records are different from training and test data set. For getting higher classification accuracy or accurate prediction we required a prior knowledge of the class label data record which makes attribute selection effortless. For higher classification accuracy supervised learning algorithm (like classification) is preferred to unsupervised learning algorithm (like clustering). In current scenario, data mining technology has been widely used in education, real estate, stocks, health care and other fields. A number of widespread classification algorithms used in data mining and decision support systems is: neural networks, logistic regression, Decision trees etc. Among these classification algorithms decision-tree algorithms is the most frequently used because of it is effortless to understand and cheap to implement

the article is being submitted to and the manuscript identification number. Click the forward arrow in the pop-up tool bar to mod

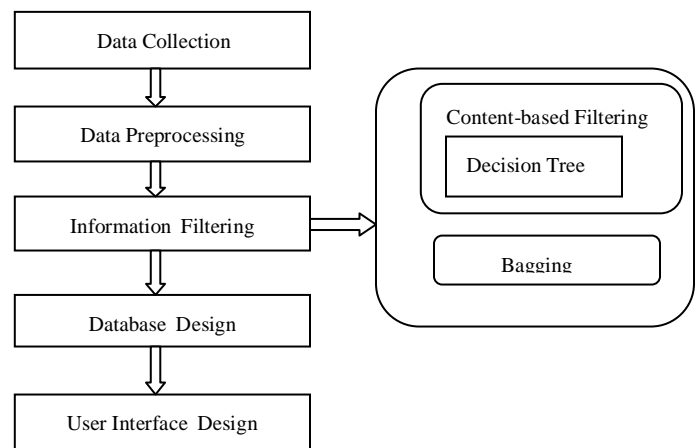


Fig1. Healthy diet system architecture

2 HEALTHY DIET RECOMMENDATION SYSTEM

2.1 Data Collection

In the data collection and preprocessing web server data base contains two types of data base one is content data base that contain the information like user information and other types of data and second is the server log data base for recording the HTTP transaction (log records).Data collection or data acquisition module collect data from the external web atmosphere to provide resources and material for the latter data mining. From the web environment the data source we get the web

pages data, hyperlinks data and history data of user visiting log. Data collection module composed by three independent processes that are data collection, data selection, data search.

2.2 Data Preprocessing

Data preprocessing mainly renovate and progression the source data acquired in data collection phase and construct the data warehouse of associated themes to generate 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... Basically in the data preprocessing step convert the data into the form which is accepted by the data mining algorithm.

2.3 Information Filtering

Information filtering is the main step of the recommendation system. In the existing association rules are applied in the content base filter. In the performance analysis of healthy diet recommendation system, introduced a new architecture based on data mining algorithm for constructing a healthy diet recommender system. A healthy diet recommender system is an intermediary program (or an agent) with a user interface that automatically and intelligently extracts the useful information of people's eating habit which suits an individual's needs. Figure shows the process in the information filter.

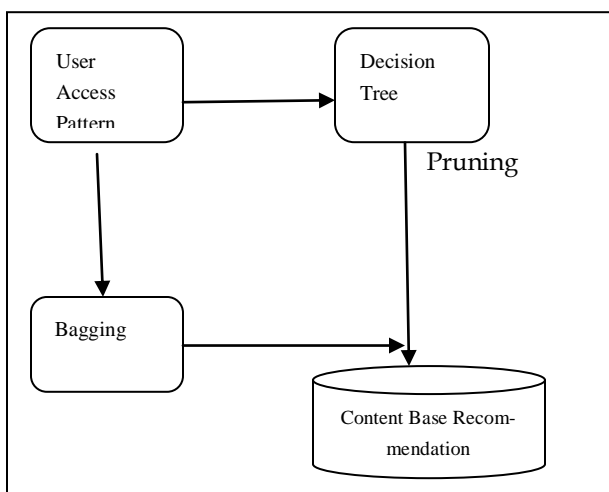
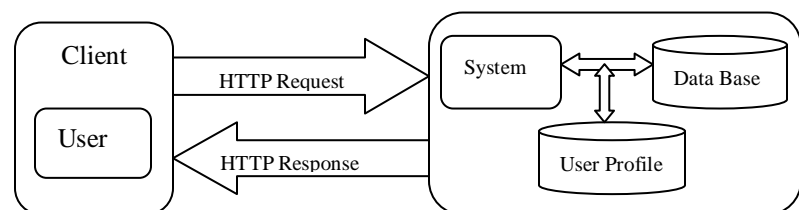


Fig 2 Information Filtering

CONTENT BASED FILTER: The content-based filtering (CBF) is a consequence and persistence of information filtering research. It constructs the recommendation based on the correlation between difference resources. In content-based recommendation systems, resources are described as a vector of attributes. The system then learns a profile of the user's interests based on the features presented in the objects the user has rated. When making a prediction on the customers' preferences, the system analyzes the relationship between the products rated by the users and other products by calculating the similarity between their attribute vectors. In our healthy eating recommendation system the healthy eating dataset first apply to the content base filter it analysis the user behavior or the content of dataset for example the whether the user is vegetarian or suffering from some kind of diseases. The content base filters analysis the user profile. For classifying data we apply the decision rule mining on user access pattern. We apply the ID3 algorithm for classify the data. Decision rule mining construct the rule that is apply on user access pattern and generate the result. The output of the content base filter is the food that is beneficial for your health. For improving the accuracy of the system we apply bagging on user access patter.

2.4 Data Base Design and Implementation with Web Based User Interface

In healthy diet recommendation system framework, database using the Relational Database Management System (RDBMS) is designed and constructed. This database stores the URLs (i.e., Web pages), keywords for the Web pages, the recommended set of rules from content-based filtering, user login information, and user profiles. 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.



The User and Recommender system interaction of via HTTP

Decision Rule Mining

Recommendation systems are used to predict the desire value. By applying the data mining algorithm on data set in recommendation system predict the data according to the user preference. Prediction can be categorized into: classification, density estimation and regression. In classification, the predicted variable is a binary or categorical variable. Various well-liked decision tree classification methods include decision trees, logistic regression and support vector machines.

We defined decision tree is a tree in which each branch node symbolize a preference between a number of substitute, and each leaf node correspond to a decision. Decision tree are generally used for gaining information for the reason of decision - making. It starts with a root node on which it is for users to acquire actions. From this node, users split each node recursively according to decision tree learning algorithm. The final result is a decision tree in which each branch represents a possible scenario of decision and its outcome. There are various decision tree classification algorithm are used like ID3, C4.5, C5.0 etc we work on ID3 and C4.5 the basic decision tree learning algorithm used for classify data.

1. Apply Decision Tree Rule Mining on Recommendation System

The performance of healthy diet recommendation system used the ID3 and C4.5 decision tree classification algorithm for classify the healthy diet data set. First the content base filters analysis the user access pattern. Content base filter analyzed the user profile whether the user vegetarian or non vegetarian, suffering from some kind of diseases etc are analyzed.

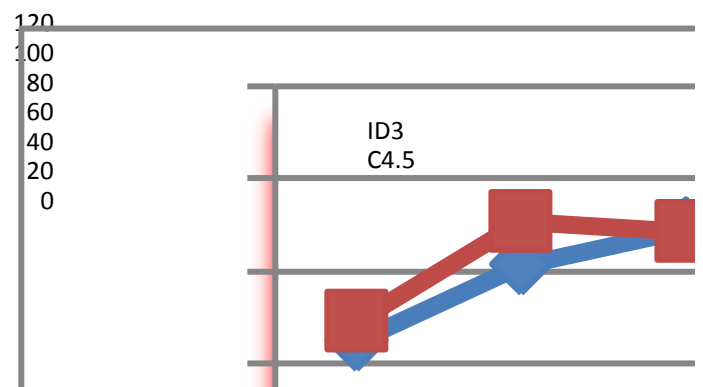
Then according to the user profile healthy diet data set is classified by the decision rule mining. It trains the data set and generate rule according to the user access pattern. In recommendation system we use the ID3 decision rule mining for

mining the data and generate rule. These rules are applied on healthy diet data set and suggest food which is beneficial for your health. For performance analysis we calculate the accuracy of the system with ID3 and then compare the accuracy of ID3 with C4.5. For improving the performance of the system we apply bagging with ID3.

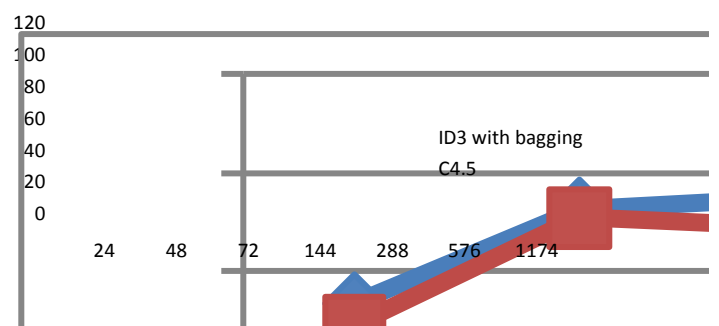
RESULT ANALYSIS

In the performance analysis of healthy diet recommendation system decision tree first get the data from content base filter. In the implementation phase we first select the data set then the generated rule. Then these rules are applied into the healthy diet recommendation data set. After applying the rule admin selects the profile where we want to apply rule. Once the profile selected the rules are applied and according to the user profile the food is suggested. Then we apply the rules on and analysis the system.

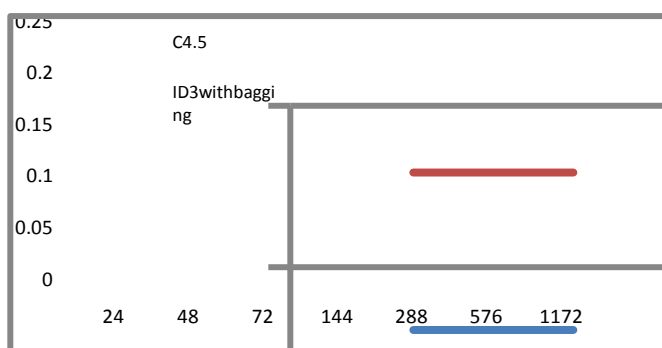
Comparative Analysis of ID3 and C4.5 In terms of Accuracy



Comparison Result of C4.5 and ID3 with Bagging in terms of Accuracy



Compare Time Complexity of ID3 and C4.5



4.1 Reason behind taking ID3 algorithm for classifying data

The result analysis shows that ID3 works in each instance of data it's also work properly when the number of instance are increase. As compare to C4.5 ID3 with bagging provide more accurate result .Classification accuracy is higher. C4.5 construct tree in less time as compare to ID3 but will not work on each instance. ID3 work on each instance and gives more accurate result after applying bagging accuracy is increased.

4.2 Conclusion of the result analysis

First the recommendation system suggests the food that is beneficial for your health then show the comparative analysis of two decision tree classification algorithms in terms of accuracy. For improving the performances of the system bagging is applied. The comparative study of the system shows that after applying bagging it gives more accurate result.

CONCLUSION

Performance analysis of healthy diet recommendation system recommends the food that is beneficial for your health. We acquire people eating habit data in the database which could track people's recipe record. Then we introduce a web data mining solution to e-commerce to discover hidden patterns and business strategies from their customer and web data, implement a new framework based on data mining technology

and build healthy eating recommendation system. Finally we give out personalized recommendations for each person. The system contains two domains (Administrator and member). At the administrator's end we design a model that helps to get rules for the large data set for food test it over random values. Administrator applies these rules over the all members (user's profile) to get suggestion for healthy food. User manage their account, health and medical profile(that are input attributes for our data mining rules set)Finally get suggested diet .If member update their health and medical profile the diet suggestion also update according to attributes.

REFERENCES

- [1] Xiaocheng Li, Xinliu, Zengjie Zhang, Yongming Xia, Songrong Qian Design of Healthy Eating System based on Web Data Mining 2010 WASE International Conference on Information Engineering .
- [2]A Data Mining Framework for Building A Web-Page Recommender System Choochart Haruechaiyasak , Klong Luang, Pathumthani, Mei-Ling Shy, Shu-Ching Chen
- [3] Sun, Jinhua; Xie, Yanqi, "A Web Data Mining Framework for Ecommerce Recommender Systems", Computational Intelligence and Software Engineering, CISE 2009.
- [4] Adomavicius G. and Tuzhilin A. (2005) 'Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions', *IEEE Transactions on Knowledge and Data Engineering.*
- [5]Bagging one class classification Chen Li, Yang Zhang Fuzzy Systems and Knowledge Discovery, 2008. FSKD '08. Fifth International Conference (IEEE2008).
- [6]Re Optimization of ID3 and C4.5 Decision Tree Devashish Thakur, Nisarga Markandaiah, Sharan Raj Int'l Conf. on Computer & Communication Technology (IEEE2010).
- [7]A more efficient classification scheme for ID3 CHAI Rui-min, WANG Miao Computer Engineering and Technology (ICCET), 2010 2nd International Conference .
- [8]An Extended ID3 Decision Tree Algorithm for Spatial Data Sitanggang, I.S.; Yaakob, R.; Mustapha, N.; Nuruddin, A.A.B.;[IEEE2011] .
- [9]A Comparison of Decision Tree Ensemble Creation Techniques Robert E. Banfield and Lawrence O. Hall, Kevin W. Bowyer, W. Philip Kegelmeyer
- [10]Comparison of ID3 and Its Generalized Version Tzung-Pei Hong, Shian-Shyong Tseng Computing and Information, 1992. Proceedings. ICCI '92., Fourth International Conference [IEEE1992].
- [11]"An Implementation of ID3 Decision Tree Learning Algorithm" Wei Peng, Juhua Chen and Haiping Zhou[1997].