

Implementation of Recommender System for Book Store Using Decision Tree Technique

Hnin Thiri Chaw, Ah Nge Htwe
University of Computer Studies, Yangon
hninthirichaw87@gmail.com

Abstract

The proposed system is the recommender system for book store using decision tree induction technique which is the data mining technique of the recommendation to extract rules in terms of entropy and information gain and to recommend the item to the customer. There had been various kinds of recommendation systems with different techniques. The recommendation system is to recommend the item that the customers wanted to buy and easy buying for customers .In addition ,many vendors are trying to compete each other to give good services to the customers who use their system to get their satisfaction and to promote and sale their product. The proposed system can determine what kind of item the customer really want to buy .

1. Introduction

Recommender systems are important part of recent e-commerce. They enable the increase of sales by suggesting to users selected products on offer. The problem of how to choose the most suitable items, possibly with respect to the user's inclinations, is a challenging research problem that has been investigated for many years. Recommendation systems can assist and augment this natural social process. The main purpose of the recommendation systems is to provide tools for people to leverage the information hunting and gathering activities of other people or groups of people. Recommendation systems have been an important application area and the focus of considerable recent academic and commercial interest. The paper mentions about the recommendation system in section two and describes about the decision tree technique in the section three .In section four, the paper mention about the overview of the system .

2. Recommender System

A common scenario for recommendation system is a web application with which the user interacts .Normally, the systems show a summary list of items to a user, and the user selects among the items to receive more details on an item or to interact with the item in the same way.

2.1 Recommendation method

2.1.1 Four types of recommendation strategies

Normally, recommendation method can be briefly classified into four: demographic filtering, collaborative and content based recommendation, and simplified statistical approach.

(1) Demographic filtering

In demographic recommendation system, customers are classified according to their personal data, which they provided during the registration process. Alternatively, this kind of data can be extracted from the purchasing history, survey responses etc. Each product is assigned to one or more classes with certain weights and the user is attracted to items from the closet to the profile .This is attribute based recommendation.

(2) Collaborative Recommendation

Collaborative recommendation is typically based on item ratings explicitly delivered by customer .The system recommend products which have been evaluated positively by another similar user or by a set of such users, whose ratings have the strongest correlation with the current user.

(3)Content based Recommendation

Content based recommendation focuses on the similarity between products, usually taking into account their features like textual descriptions, hyperlinks, related ratings or co-occurrence in the same purchased transactions or web user session .Item that are the closet to the most recently processed(viewed) are recommended regardless of preferences. This is item to item correlation. Association rule and sequential pattern are the most interesting technique used in recommendation based on item-to-item correlation

(4)Simplified Statistical Recommendation

The customer is shown products based on some statistical factors; usually popularity measures like averages or summary rating and number of sold units.

A single recommendation method can offer either ephemeral or persistent personalization. The former is based only on current session and can deliver a different list on every page of a website but be the same for all users. Persistent personalization uses the history of user's behavior and generates a different

product list for each user, but it works only with identified logged in users.

2.2 Classification and Prediction

Classification and Prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. While Classification predicts categorical labels (classes) or to prediction models continuous-valued functions.

There are two type of learning: supervised learning in which the training sample are provided, and unsupervised learning in which the training samples are not provided(or clustering).

Preprocessing of the data in preparation for classification and prediction can involve data cleaning to reduce noise or handle missing values, relevance analysis to remove irrelevant and redundant attributes and data transformation , such as generalizing the data to higher –level concepts or normalizing the data. Classification is a two step process.

(1) Learning

Training data are analyzed by a classification algorithm .After that, the learned model or classifier is presented in the form of classification rules as shown in figure.

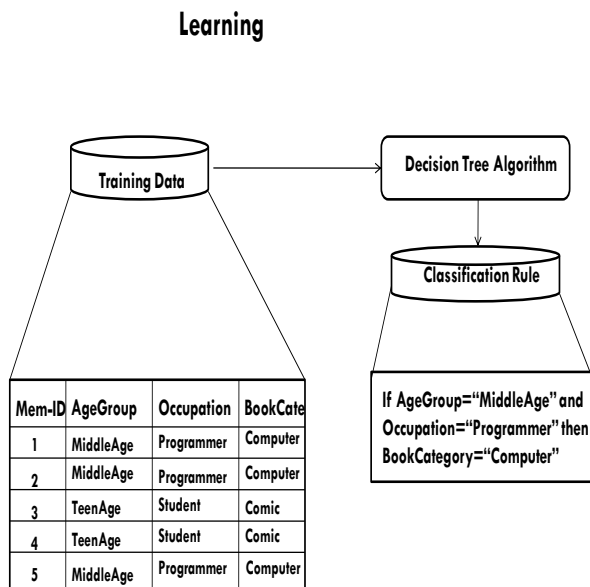


Figure 1. Learning from training data set

(2) Classification

Test data are used to estimate the accuracy of the classification rules. If the accuracy is considered

acceptable, the rules can be applied to the classification of new data tuples. The following figure is the process of classification.

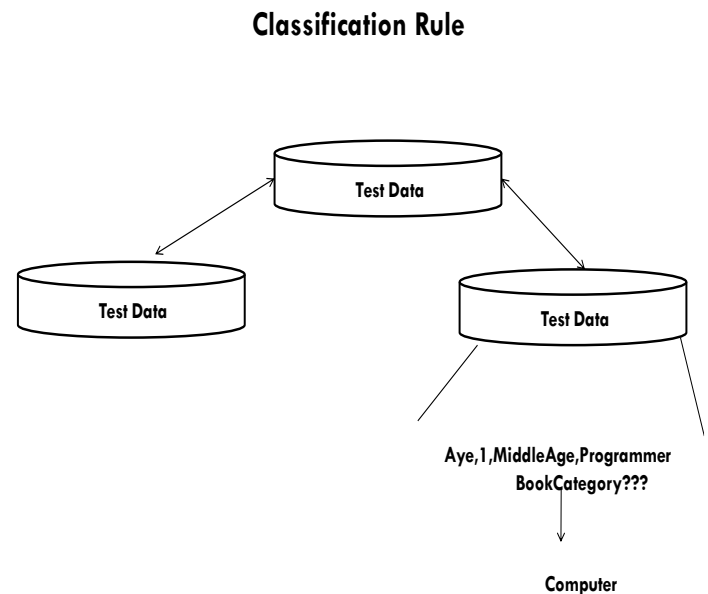


Figure 2. Classification

3. Decision Tree Induction Techniques

Decision Tree Induction is widely used practical method for inductive interface and also robust to noisy data and capable of disjunctive expression. It also approximate the target function as tree. However, learned tree can also be represented as set of if then rule to improve human readability. In addition, it can classify instances by sorting them down from the root to leaf node. Each node in a tree specifies a test of some attributes and each branch descending from that node corresponding to one of the possible values for this attributes. Attributes are classified by starting at the root node of the tree, testing the attribute specified by this node, then moving down the tree branch corresponding to the value of the attribute. This process is then repeated for the sub tree rooted at the new tree.

3.1 Step by Step Process of Decision Tree Induction Technique

Step(1): Pruning

Before using the decision tree induction technique, the training data must be preprocessed to remove the noise and redundant data. Pruning is the step of eliminating unnecessary attributes from raw data. The training dataset must be correlated with the target attribute. The step is also known as data preprocessing which is to prune noisy and irrelevant data and to reduce data volume.

Step(2): Calculating Entropy

Information gain is calculated in term of entropy .Entropy value must be always between zero and one .Entropy also known as the probability characterizes the impurity of arbitrary collection of examples according to the attribute. Entropy is classified into two :Boolean Classification and C-Wise Classification.

(i) Boolean Classification

Given a collection of S of positive and negative sample, entropy of S according to the Boolean Classification is

$$\text{Entropy}(S) = -P^+ \log_2 P^+ - P^- \log_2 P^-$$

In this formula ,P+ represents positive sample and P- represents negative sample .For the point of sale system , Boolean classification only specify buy and not buy item, not for the specified item or specified category of the item.It is only for the Boolean-value target function.

(ii) C-Wise Classification

C-Wise classification is for multi-value target function. If the target attribute can take on c different values, the entropy of S relative to this C-Wise Classification is

$$\text{Entropy}(S) = \sum_{i=1}^c -P_i^+ \log_2 P_i^+$$

Step (3): Calculating Information Gain

Information Gain measure the expected reduction in entropy.Information gain, Gain(S,A) of attribute A is the reduction in entropy caused by partitioning the examples according to this attribute.

$$\text{Gain}(S,A) = \sum_{v \in \text{value}(A)} \text{Entropy}(S) - S_v / S \text{ Entropy}(S_v)$$

Step(4): Rule Extraction as Tree

According to the Information Gain, root node and sub root nodes are classified by selecting one of the largest value of the information gain of the specific attribute as root node and calculated the information gain iteratively to specify the new sub root nodes without including the parent node or root node .These rules are ready to recommend the category of item.



Fig 3: Rule extraction as tree

Step(5): Recommending item

The system collect the data during the registration process and then after the registration process is completed ,the system recommend the book category of book items that match with the rules extracted from the training data set. The user must fill user identification ,password, email, date of birth, marital status, gender, education ,major ,book category that the user like.Depending on these attribute, the system recommend the book item that most of the user likely to read according to the rules by using decision tree induction technique.The step is also known as classification because the new user registration can be viewed or assumed as testing data.

Book Information List

Book ID	Book Image	Book Type	Title	Author Name	Price	Detail	
14		COMPUTER	Java Programming	Berkelay	10000.00	Detail	Add To Cart
29		COMPUTER	Sams Teach Yourself Beginning Programming in 24 Ho	Greg Perry	20000.00	Detail	Add To Cart
30		COMPUTER	Beginning Programming For Dummies	Wallace Wang	12000.00	Detail	Add To Cart
31		COMPUTER	Code Complete: A Practical Handbook of Software Co	Steve McConnell	12000.00	Detail	Add To Cart

Fig 4: Recommending item to customer

4. Overview of the system

4.1 Proposed system design

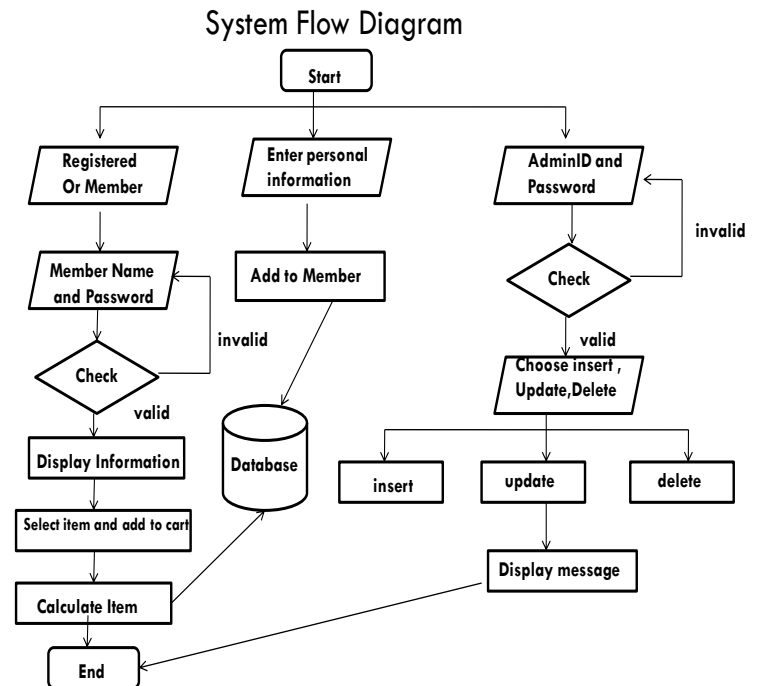


Fig 5: System Flow Diagram

4.2 Implementation of proposed system

The proposed system can be viewed as two sides: the customer and the admin. At the customer side, the system can recommend the book items that match with the rules extracted from the training data set. At the admin side, the admin can insert, update and delete the book item and see the result tree.

One of the important thing is to register the user information if the user is not a member. When the user log into our system, the system check the username and password to distinguish between user and admin. If the user is a member, the system check member name and password to sign in to the book recommendation system using decision tree technique. If the user name and password is correct, the system recommends books to the user according to the rule in the system. After that, the users can select and buy the book items that was recommended by the system. In addition, the user can select the other book items according to the book category, and then they can also buy whatever they like.

When the user is admin, admin can insert, update and delete books information to recommend the book items to the member and users according to the book category. By and large, the admin can also view what is the rule in the system as tree and also as text to become user friendly.

In the proposed system, the class label or target is book category and the attributes are education, major,

occupation, gender ,age group and marital status to calculate information gain in term of entropy .Education are classified into primary school education, middle school education , high school education, undergraduate, graduate and postgraduate. Age group can be classified into child age(5yrs to 10yrs), teenage , middle age(20yrs-50yrs) and old age(over 50yrs) and marital status into single and married and, gender into male and female.

4.2.1 Sample Rule of the system

Rules are extracted from the training data set using Decision Tree Technique in term of entropy and information gain. The following rules can be seen at the admin side by calculating dynamically. The class label for the proposed system is Book Category.

Rule1-->Student-->HighSchoolEducation-->Male-->**HUMOUR(COMEDY)**

Rule2-->Student-->MiddleSchoolEducation-->FEMALE-->**SCIENCE FICTION**

Rule3-->Student-->MiddleSchoolEducation-->MALE-->**SCIENCE FICTION**

Rule4-->Student-->PrimaryEducation-->FEMALE-->**CHILDRENS CHAPTER**

Rule5-->Student-->PrimaryEducation-->MALE-->Music-->**STORY**

Fig 6 :Rule extracted from the system



Fig 7:Sample result tree using decision tree algorithm

4.3 Recommend book at the customer side

According to the above figure,a female customer with primary education will be recommended the books item of children's chapter book category.

Book Information List			
Book ID	Book Image	Book Type	Title
56		CHILDRENS CHAPTER	Eclipse
57		CHILDRENS CHAPTER	New Moon

Fig 8 :Recommended book items of children's chapter book category

5. Conclusion

In summary, the system based on user profile during the registration process .The proposed system use entropy of C-wise classification for determining more specific recommended item rather than using Boolean classification to calculate the information gain in term of entropy to determine root node and sub root nodes .A root node of the attributes must be chosen with highest information gain and that root node attribute must be omitted to calculate the information gain iteratively for its sub root nodes to get the efficient rules that can be use to recommend the book item for a customer ,but it is really need to get the enough data from the user to become more effective and accurate to recommend item category to the customer. The decision tree approach to recommendation is useful to recommend the book item to the user according to user inclination.

References

- [1] Badrul Sarwar, George Karypis, Joseph Konstan, ,John Riedl "Item-based Collaborative Filtering Recommendation Algorithms"
- [2] Jiasei Han and Micheline Kamber"Data Mining :Concepts and Techniques"
- [3] Michael J.Pazzani and Daniel Billsus"Content based Recommendation Systems"
- [4] Przemyslaw Kazienko and Pawel Kolodziejski "Personalized Integration of Recommendation Methods for E-commerce"
- [5] Sarabjot Singhan and and Bamshad Mobashers, "Intelligent Technique for Web Personalization"