

Certification of Products in Ecommerce Websites using Rule Based Decision Tree

Amol Gawande

Dr. D. Y. Patil B-School Tathawade, Pune, Maharashtra, India

Email: amol.gawande@dpu.edu.in

Ranit Kishore

Tulas' Institute of Management, Dehradun, Uttarakhand

Parikshit Kala

Dept. of Mgt., Himgiri Zee University, Dehradun, Uttarakhand

Anant Sharma

Himgiri Zee University, Dehradun, Uttarakhand

Y. Suryanarayana Murthy

Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, AP

M. Govardhan Reddy

Dept. of Mgt. Studies, Aditya College of Engineering, Surampalem, AP

Abstract: From technologies used by a few e-commerce portals, recommendation services are evolving to serious business platforms that are re-shaping the landscape of e-commerce. In order to help their customers find goods to buy, many of the biggest commerce websites are now using recommendation systems. A recommender system observes from a client and recommends items from among the products available that they would find most useful. Generally ecommerce websites have common search criteria for all the products. If a specialized search is available for products based on their usage and features it would be easy for buyers. The essential aspect of recommending goods based on the needs of consumers is presented in this chapter. It allows the consumers to choose the appropriate product along with the consumer feedback. Purchasing requirements for the goods was generated based on the survey and input from the previous customers and suppliers. The rule-based decision tree is used to identify goods based on specialized use and the purchaser's choice. This is used as interface to the customers that helps the users to buy the right product.

Keywords: technologies, e – commerce, consumers, input, indentify.

JEL classification: L81, D12, O33

1. Introduction

Recommender systems were first described by its developers as Collaborative Filtering in which they address how individuals communicate and filter email documents that are important to them and of use to their audience. An overview of common properties between two or more documents was included in the filtering process (Wei *et al.*, 2001, Dadas *et al.*, 2021). The letter, reply or its annotations is included in the properties belonging to the documents examined. This was found to be more powerful than basic analysis of the content of the document offered by many other mail systems. Human involvement in the filtering process led to the selection of more interesting papers (Patil *et al.*, 2021).

The following are the list of recommendation types that are commonly used.

- **Non-personalized recommendations based:** About what other consumers have said on average about the products, non-personalized feedback programs recommend products to customers. The recommendations are independent of the client, so the same recommendations are obtained from each customer.
- **Character-based recommendations based:** Character-driven recommender programs recommend products based on the conceptual features of the products to consumers. For eg, if the consumer is searching for a mobile phones and the ecommerce platform responds with a list of mobiles, it is an example of a recommendation based on an attribute (Goldberg *et al.*, 1992, Kumar *et al.*, 2021a).
- **Item-to-item correlation:** Recommendation programs are similarity recommend items to clients based on a small range of products that consumers have shown interest in. For example, once a consumer has put a few items in their shopping cart, complementary products may be recommended by the recommender system. This may be automated as they are based on interpretations of the unchanged behavior of the client (Kumar *et al.*, 2021b).
- **People-to-people correlation:** Recommendation mechanisms for people-to - people correlation recommend goods to a buyer based on the relationships between that buyer and other consumers who ordered products from the ecommerce platform. In order to maximize both the benefit of the seller and the satisfaction of the customer, Recommender Systems (RS) offers users useful and interesting products. A very active field of research leads to the commercial success of many online projects. The main objective of this research is to provide new users

with more specific product recommendations. The decision tree is generated and rules are framed to identify the user's response and the user is advised to use the correct product (Kumar & Wanjari, 2021).

2. Methods and Models

Most of the recommender system in e-commerce website will provide common search criteria such as price, type, brand, age etc., Most of the novice users will be helped out if a personalized search or endorsement of the product is available in recommender system. Consider a novice user needs to buy laptop. As a nontechnical user, he doesn't know the exact specification that he actually need. In this case, the e-commerce website could provide the personalized search for him based on the product. The existing recommender system's query have age of the person, work type of the person, preferred price range, checking the whether he is a gamer or animation user or a researcher or a student. In addition to these, the expected software/software type used by the user, expected duration of usage, professional or personal use. etc can also be asked. These provide the user with most relevant laptop satisfying all criteria. The output of the query will be the relevant laptop for his appropriate usage. The e-commerce website can get a feedback or survey from the previous buyers and a thorough survey must be done with the manufacturer. The system will be trained with the training data and the rules are developed with them with the help of decision trees (Schafer *et al.*, 1991).

3. System Architecture

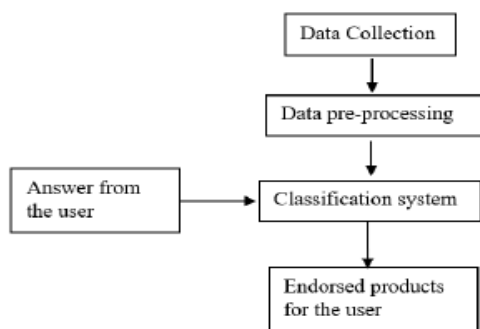


Figure 1: System Architecture

- **Data Collection:** Data for this recommender system is collected from the previous buyers and manufacturers. A survey is taken in the form of feedback from the previous buyers. The description of the products and

details of the usage are collected. Based on these data collected, recommender system helps in finding the right product for the user. Recommender system for washing machine: Questions to be answered by the novice buyer, a. Price range, b. Brand c. Family size d. Number of times machine to be used per day e. Number of children f. Child age group g. Cloth type mostly used h. Provision for drying many cloths available or not. Power consumption j. Water quality (salty or normal) etc., similarly personalized questions for the different category of the products must to do.

- **Data pre-processing:** In order to be used by machine learning algorithms in the research method, real-life information usually has to be preprocessed (e.g. washed, sorted, and transformed). The first step is to analyze various similarities or range measurements in the nature of a RS, and then the problem of sampling, a means of minimizing the number (Sarwar *et al.*, 2000).
- **Classification System:** A classifier is a projection between the set attribute and the class mark in the classification, where the attribute represents the attributes of the items to be categorized and the labels denote the classes. To find a model that best matches the interaction between the attribute collection and the class mark of the data input, each technology utilizes a learning algorithm. The model created by a learning algorithm should therefore both match the input data well and forecast the target class of unidentified documents accurately.

4. Decision Tree

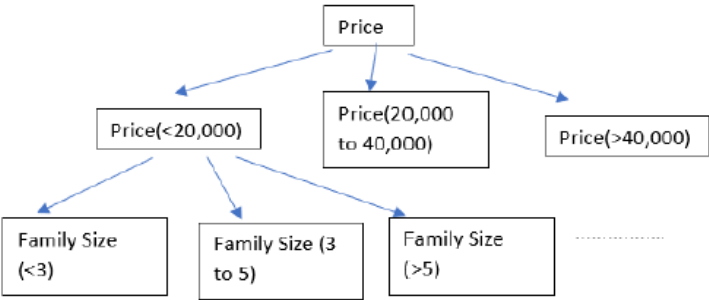


Figure 2: Decision Tree Sample

In the background of a tree system, the Decision Tree establishes models of classification or regression. A dataset is split into smaller subsets, while a random forest connected with it is progressively created at around the same

time. The end result is a tree with decision nodes and leaf nodes. There are two or three parts of a judgment node. A grouping or judgment is defined by the leaf node. The uppermost decision node in a tree that correlates to the root node of the best predictor. Both numeric and categorical data can be treated by decision trees. Decision tree sample is shown in figure 2. To divide the problem search space into subsets, the decision tree uses a divide conquer approach. The most critical aspect of the classifier of the decision tree is its ability to break down a complicated decision making process into a simpler decision-making process, offering a solution that is easier to interpret (Doha *et al.*, 2019, Ngai *et al.*, 2009).

Tree nodes may be a) decision nodes, where a single attribute-value is evaluated in these nodes to decide which sub tree branch applies. Or b) leaf nodes that show the goal attribute's value. In this analysis, the C4.5 Algorithm was selected to be used. The qualified trees are translated by C4.5 into sets of if-then laws. In order to identify the order in which they've been applied, the reliability of each law is then measured. Pruning is achieved by eliminating the prerequisite of a rule if the rule's consistency increases without it. The qualified trees then turn into sets of if-then laws (Li *et al.*, 2019, Khan *et al.*, 2018, Gawande *et al.*, 2021).

5. Algorithm

- C4.5 is a decision tree based classification algorithm proposed by Ross Quinlan. 1. M is considered as the training data. It is the mark of the class and $j = 1, 2, \dots, N$. Let $\text{freq}(M)$ represent the number of M samples in the class (out of N possible classes) and $|T|$ indicates the number of M samples in the training set.
- Base Case Check is followed by finding the best split attribute that gives optimum information gain for splitting.
- Two tests are used to find the right division. Entropy: used to assess impurity, i.e. to quantify a sample's homogeneity. The entropy of the set M is then measured as:

$$\text{info } D = - \sum_{i=1}^m p_i \log_2(p_i)$$

$$=1$$

$$(1)$$

$$\text{info}(D) = - \sum_{j=1}^D p_j \log_2(p_j)$$

$$D$$

$$* \text{info}(D_j) = - \sum_{v=1}^V p_{vj} \log_2(p_{vj})$$

$$=1$$

$$(2)$$

Information Gain: The gain of information informs us how important a particular feature is. $\text{Gain}(A) = \text{Info}(D) - \text{Info}(A|D)$ (3)

- The best characteristic for splitting is the one that offers optimum data gain. The decision tree is produced using the attributes that provides the highest data gain (Osadchiy *et al.*, 2019, Portugal *et al.*, 2018).
- And the same processes are performed recursively to each of the impure tree nodes. When all nodes are clear, the algorithm then ceases. One rule is generated for each way from the source to a leaf node to remove rules from a decision tree. Logically, each dividing criterion across a predetermined route is AND ends to form the precedent of the law ("IF" part). The leaf node retains the judgment of the class, defining the consequent rule ("THEN" part). By analyzing the pattern from the top node to each of the leaf nodes, the decision tree can be translated to IF-THEN classification rules. Based on the rules generated from the decision tree, the product is suggested to the user.

6. Algorithm

By calculating the end survey provided by the consumer, the analysis of this scheme will be found. In order to verify the operation of the classification scheme, the accuracy and execution time values are measured. One of the most intuitive and simplest metrics used for finding the model's correctness and accuracy is the Confusion matrix. It is used for the problem of classification, where the output may be of two or more class types. Class labels are specified based on the feedback of the commodity and its producers (Wang *et al.*, 2018).

7. Conclusion

There is a need to process, prioritize and effectively de-furnish related content in order to mitigate the issue of overwhelming information on the Internet. Through filtering through a vast amount of vigorously generated knowledge to provide consumers with customized content and server resources, recommendation systems solve the issue. In this analysis, the suggested model using the decision tree method needs little data planning, the ability to manage both numerical and categorical data and the ability to verify a model using statistical testing since it is easy to grasp and analyze. This system also helps the novice user get the relevant product and thereby the trust in the e-commerce website gets increased.

References

B. Sarwar, G. Karypis, J. Konstan and J. Riedl (2000). "Analysis of recommendation algorithms for e-commerce," in ACM, New York,.

- B. Schafer, J. Konstan and J. Riedl (1999). "Recommender systems in ecommerce," in ACM, New York Cit.
- C. P. Wei, M. Shaw and R. Easley (2001). "A Survey of Recommendation Systems in Electronic Commerce," National Sun Yat-Sen University, Kaohsiung.
- D. Goldberg, D. Nichols, B. Oki and D. Terry (1992). "Using Collaborative Filtering to Weave an Information Tapestry," in ACM, New York.
- Dadas, A. B., Kumar, A., & Brar, V. (2021, March 26). System and method for selling project based service application software (Registered Copyright No. L-101028/2021). Copyright Office, Department for Promotion of Industry & Internal Trade Ministry of Commerce and Industry, India.
- Doha, A., Elnahla, N., & McShane, L. (2019). Social commerce as social networking. *Journal of Retailing and Consumer Services*, 47, 307-321.
- Gawande, A., Kumar, A., & Darekar, S. (2021). Studies on e-commerce business model innovation. In P. M. B. Saleem, S. Maganti, P. Ganguly, V. R. R. Gandreti & M. Neelam, *Role of Human Resource and Customer Relationship Management in the Current Scenario* (1st ed., pp. 464-476). AGAR Publications. <https://doi.org/10.5281/zenodo.6625036>
- Khan, Z. A., Zubair, S., Imran, K., Ahmad, R., Butt, S. A., & Chaudhary, N. I. (2019). A New Users Rating-Trend Based Collaborative Denoising Auto-Encoder for Top-N Recommender Systems. *IEEE Access*, 7, 141287-141310.
- Kumar, A., & Wanjari, S. (2021). HUL: E-commerce in FMCG. In S. Patil, A. Gawande & A. Kumar, *Caselets in Business* (1st ed., pp. 43-46). Dr. D. Y. Patil B-School, Pune. <https://doi.org/10.5281/zenodo.6739590>
- Kumar, A., Clementking, A., Zaman, S. M., Ramkumar, J., Karthikeyan, R., & Tiwari, M. (2021b). *Recent Trends and Innovations in Information Technology*. Association of Global Academicians and Researchers Publications, Tamil Nadu, India. <https://doi.org/10.5281/zenodo.6665571>
- Kumar, A., Hemalatha, S., Saleem, P. M. B., & Paxleal, J. S. (2021a). *E-Governance - A Comprehensive Framework with Case Study*. Saliha Publications, Tamil Nadu, India. <https://doi.org/10.5281/zenodo.6660232>
- Li, W., Bai, Q., Zhang, M., & Nguyen, T. D. (2018). Automated influence maintenance in social networks: an agent based approach. *IEEE Transactions on Knowledge and Data Engineering*, 31(10), 1884-1897.

Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, 36(2), 2592-2602.

Osadchiy, T., Poliakov, I., Olivier, P., Rowland, M., & Foster, E. (2019). Recommender system based on pairwise association rules. *Expert Systems with Applications*, 115, 535-542.

Patil, S., Gawande, A., & Kumar, A. (2021, June 03). E-commerce management system (Registered Copyright No. L-104051/2021). Copyright Office, Department for Promotion of Industry & Internal Trade Ministry of Commerce and Industry, India.

Portugal, I., Alencar, P., & Cowan, D. (2018). The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications*, 97, 205-227.

Salvatore Ruggieri, —Efficient C4.5l, *IEEE transaction on knowledge and data engineering*, Vol. 14, N0. 2 March/April 2012. Parker, Christopher J.; Wenyu, Lu (13 May 2019). "What influences Chinese fashion retail? Shopping motivations, demographics and spending". *Journal of Fashion Marketing and Management*. 23 (2): 158–175. doi:10.1108/jfmm-09-2017-0093.

Wang, W., Zhang, G., & Lu, J. (2017). Hierarchy visualization for group recommender systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(6), 1152-1163.