

Recommendation System for Automobile Purchasing: A Survey

Shrey Talati
BE (Comp)

Dr.D.Y.Patil Institute of Engineering & Technology
SavitribaiPhule Pune University

Anukrity
BE (Comp)

Dr.D.Y.Patil Institute of Engineering & Technology
SavitribaiPhule Pune University

Priyanka Salian
BE (Comp)

Dr.D.Y.Patil Institute of Engineering & Technology
SavitribaiPhule Pune University

Anam Hussain
BE (Comp)

Dr.D.Y.Patil Institute of Engineering & Technology
SavitribaiPhule Pune University

ABSTRACT

Consumers are adopting different search strategies, so identifying the patterns of their information search habits has become a challenge. The way in which allocation of resources is done across different sources of information depends on the understanding of the patterns of information search behavior. Cluster analysis was used to identify the distinct segments of new car buyers concurrently, and the relative importance of important variables in differentiating the segments. With the help of classification it can predict categorical class labels.

Different patterns of information search behavior were obtained across four different groups -broad temperate searchers, extreme heavy searchers, low broad searchers, and low searchers. Here both cluster analysis and classification was compared to identify consumer assortment of external pre-purchase information search behavior and the most optimized method found, out of the two was clustering.

General Terms

Information Retrieval, Data Mining, Recommender

Keywords

Clustering, Classification, Automobile, Information Search Behaviour

1. INTRODUCTION

Information search behavior is an assortment of Economic, Technological, Political and Cultural factors as well as his own characteristics which is reflected by his attitude, motivation, perception, personality, knowledge and lifestyle. Marketers can downsize their existence only when they are able to understand Information search behavior of consumers. Early research has been intended to get broaden up in a present study in a few ways. First, individual character variables are considered while finding the patterns of information search approach. In previous studies individual character variables were not taken into an account for finding different sectors of customers based on their information search method; main concentration of these studies was on search and decision time, and different origin of information. Second, the present study is being operated in India; most of the previous research was carried out in West. Third, using a cluster analysis, it suggest the major decisive factor variables which separate out the distinct groups. This will allow the practicing managers to get the significant insights to formulate

strategies to go well with different sector of consumers based on their search method. Personality related variables for finding the patterns of information search was not taken into an account in most of the studies. To know more about an individual, understanding the personality is indispensable. Temporal and situational invariant personal characteristics discern different individuals leading to consistencies in behaviour in different situations and time and this temporal and situational invariant personal characteristics is dealt by personality traits. Individuals will take up certain behaviour patterns that illustrates their personality, and as a result will adopt those behaviour patterns while buying a product.

A person's oblivious requirements are the outcome of his/her motivation and individuality. Four different kinds of individuality are: sensing-intuiting, thinking-feeling, extroversion introversion, and judging-perceiving. Two distinctly different personality characteristics that sway consumer responses to the world around them is reflected by each of these four pairs of dimensions. Trait theory hypothesized that individuals have inborn cognitive characteristics to a greater or lesser degree, which can be measured by specifically designed scales or catalogue. Lifestyle, attitude, and several other factors potentially influence personality development and the response of consumers towards their environment, but the present study does not include them as the reach is limited to appropriate personality traits. The variables taken into consideration to identify and explain the patterns of information search behavior are:

1. Shopping enjoyment

Assessment of past studies suggests that shopping enjoyment is one of the qualifications of information search. Shopping enjoyment is nothing but the degree to which performing an activity is professed to provide pleasure and joy in its own right, aside from performance consequences.

2. Subjective knowledge

Prior studies have given away differing effects of the dimensions of knowledge on information search, with the results showing positive, negative, or inverted U shaped relations. Consumer knowledge has two dimensions—subjective knowledge and objective knowledge. Objective knowledge consists of what the consumer actually knows, while subjective knowledge is the consumer's acuity of the

amount she/he knows about the product domain—what she/he thinks she/he knows.

3. Optimum stimulation level

The behavior of an individual is influenced by the vitally motivated crave to bring about a specific level of stimulation, which is known as optimum stimulation level. It has been established by consumer research that the level of optimum stimulation varies across individuals. When the stimulation obtained from the environment is too low, individuals will endeavor to increase the stimulation, and vice versa. Optimum stimulation level is a crucial factor when individuals are occupied in activities like information search and exploratory behaviour.

4. Need for cognition

The level of cognitive resources available is related to a consumer's need for cognition, which indicates a person's motivation to engage in elaborate cognitive activities. Consumers with a high NFC have potent internal motivations to process or cognize information, and so they have more cognitive resources available. The different external information search behavior can be explained with the help of Need for cognition.

5. Technology readiness

Technology will trigger both positive and negative feelings. The positive feelings would impel people towards new technologies, while the negative feelings may hold them back. The relative dominance of these two types of feelings is likely to differ across individuals. Technology readiness has four dimensions, reflecting two positive and two negative feelings: optimism—a positive view of technology, and a belief that it offers people increased control, flexibility, and efficiency in their life; innovativeness—a tendency to be a technology pioneer and thought leader; discomfort—a perceived lack of control over technology, and a feeling of being overwhelmed by it; and insecurity—a distrust of technology, and skepticism about its ability to work properly. As companies have increasingly begun to use the Internet to broadcast information about their products, it becomes important to understand the tendency of new car buyers towards new technology.

6. LITERATURE REVIEW

In the online world where there is huge number of data, the need for making an efficient decision or choice among them is the most challenging task. In order to do this, first need to filter the data, prioritize them and efficiently hand over the relevant information so that the problem of information overload is lessened to an extent. This problem is solved with the help of recommender system as it searches through dynamically generated large volume of data and gives the customer a personalized solution. Different Characteristics and potential of various prediction techniques in recommendation system has been explored [1, 9].

Due to evolution of industrialization in Indian Automobile markets buying car is a necessity rather than luxury. As there is a stiff competition between different car industries it has become essential to know the approach and view point of the expected customers and record the influential factor that helps in making decision [2, 7, 8].

Since huge amount of data flows in, it needs some efficient and accurate data mining techniques that can recommend good and useful solution to individual user. So this calls for the importance of understanding of different data mining techniques like clustering and classification in order to find patterns useful in decision making [3].

Riu Dongyan Xuan [4] claims that china is viewed to have a fast growing economy in the automobile sector by the western and European industries, this research questions the different factors considered while purchasing a car and the decision process involved in purchasing the car.

Previous recommendation system used prediction and rating based methods in order to show the suggestion that the user might likely use. But, the ranking based recommendation algorithms has shown better results than the rating based ones and has attracted more attention in research [5,10].

In the online market as there are numerous choices available, customers has to put in lots of effort to search and find the product of his/her desire [6]. Focuses on the analysis of consumer's online information search behavior.

Data in large amount are already being collected and refined by most of the companies. The value of existing resources available can be enhanced by implementing data mining techniques on existing hardware and software platform. As now there is a strong competitive pressure in commercial field, data mining provide customized and better services for an edge. Classification is the process of learning a model that describes different classes of data .The model that is produced is usually in the form of a decision tree. Clustering can be considered the most important unsupervised learning problem; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. Clustering is the process of organizing objects into groups whose members are similar in some way [11, 12].

7. MOTIVATION

With the rise in globalization and industrialization, there is a cut-throat competition among the automobile industries. Main focus is on cars that are no longer viewed as luxury, it is now a part of day-today life and have become a need. Car purchasing comes second in the list of most required and expensive decision after house purchasing.

Those who are involved in car manufacturing gets an opportunity to build positive image of the brand from the first time car buyers. Market is now a very competitive and important place to observe the consumer's buying behavior and obtain useful insight about consumer's need and preferences regarding any product in a constant dynamic environment.

In today's world automobile industries is one of the most profitable industries. In the present scenario, the personal income in both rural and urban area has been appreciably increased and there is an easy availability of finance and all these are the main influential factors of car segment that are purchased on large scale.

Influential factors for Car Purchasing

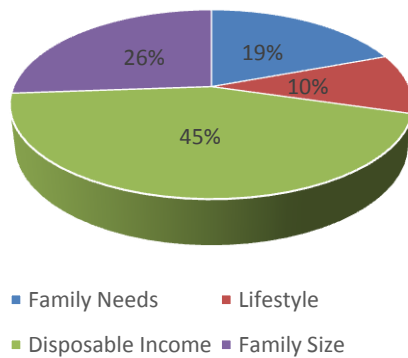


Fig 1: Pie Chart depicting the influential factors

A humungous growth of Indian automobile market has been observed which will continue to become an important market for worldwide auto giant.

With the identification of factors that spur consumer's preference and need for particular section of cars. An attempt can be made to combine result and give recommendation to resolve the current situation of stagnancy in sales and improving demand for car market in future to some appreciable extent.

8. RECOMMENDATION SYSTEM & TECHNIQUES USED

Absence of solution for prioritization and personalization of information led to an increased demand for recommendation system.

Recommendation System filters the important information portion out of the huge amount of information being dynamically generated on the basis of user's preferences, curiosity or observed habit.

Base on the user's profile this recommendation system can make the prediction as to whether a particular user will choose or prefer an item or not.

Both service providers and users are benefitted by recommendation system, it also causes an improvement in decision making process and quality.

Recommendation System assists and augments the social process of making use of other's recommendation for making choice in the absence of making sufficient personal knowledge or experience of the alternatives.

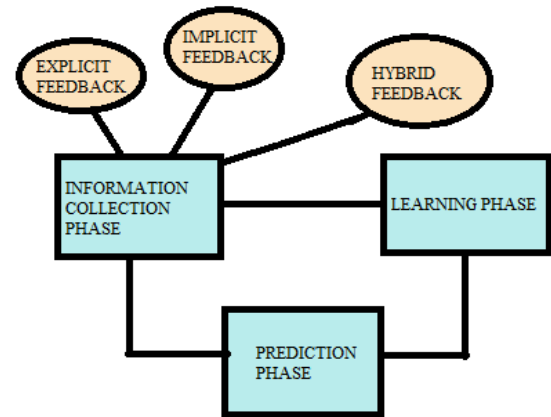


Fig2: Phases of Recommendation System

The focus will be on model based filtering a collaborative filtering technique. It identifies other users with similar taste and recommends current active users using their opinion

In different application areas this collaborative recommendation system have been employed, like GroupLens is a news-based architecture which assists users to locate articles from voluminous news database by implementing collaborative method.

In model based technique previous/past/prior rating is employed to learn a model with the purpose of improving the performance of collaborative filtering technique. Data mining techniques can be used for model building process. Machine learning can also be considered.

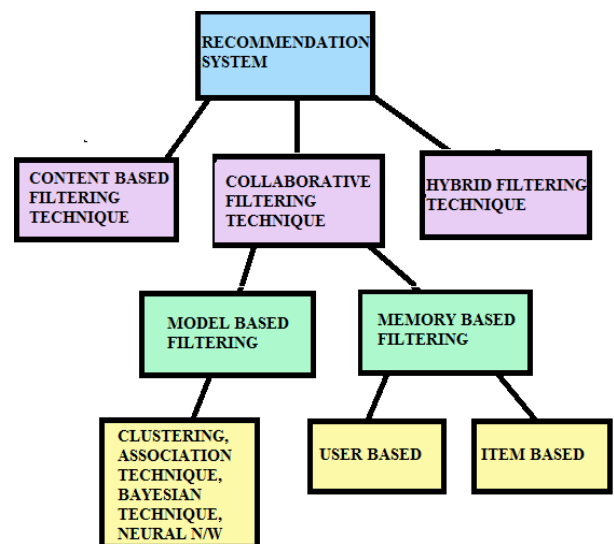


Fig 3: Recommendation System

Model Based Techniques includes:

1. Clustering
2. Bayesian Classifier
3. Matrix Completion Technique-KNN
4. Association Rule

Clustering has application in different domains like pattern recognition, image processing, knowledge discovery and statistical data analysis. High quality clusters with high intra-clusters similarity is formed by the good clustering method. Commonly used methods of clustering are:

1. K-means: There are always k clusters formed, each cluster contains at least one item.
2. SOM (Self Organizing Map): Method for an unsupervised learning.

Often this clustering is one of the first step in data mining analysis. The development of population segmentation models like demographic-based customer segmentation is supported by clustering technique and the characteristics of these segments can be determined with respect to some desired result through additional analysis using standard analytical and other data mining techniques

Bayesian Classification problems are solved by this probabilistic framework. It is based on Bayes Theorem and conditional probability. Naïve Bayes Classifier is commonly used Bayes Classifier. Naïve Bayes Classifier are robust to noisy data and irrelevant attributes and missing values are handled by them by ignoring the instant during probability estimate calculations.

Matrix-Completion Technique (KNN), prediction of unknown values within the user item matrix is done by matrix completion techniques. KNN is a correlation based major technique implemented in collaborative filtering recommendation system, large dependences is on historical rating data on items.

Bayes Rule

$$p(c|x) = \frac{p(x|c)p(c)}{p(x)} \quad \text{-- (1)}$$

$$p(c|X) = p(x_1|c) * p(x_2|c) * ... * p(x_m|c) * p(c) \text{-- (2)}$$

9. EVALUATION METRICS

Table 1. Comparison of Techniques

Technique	Speed	Performance Metrics	Usage
K-Means	Fastest	Good	High
Naive Bayes	Fast	Best	Medium
KNN	Average	Average	Low

***Performance Metrics include-** Precision, Recall, F-measure, and Accuracy.

For classification tasks, the standard predefined metrics true positives(tp), true negatives (tn), false positives (fp) and false negatives (fn) are used to compare the result of classifier.

The terms positive and negative refers to prediction or expectation and the terms true and false refer to judgment or observation.

Therefore these tp, tn, fp, fn are used to define the formulation of precision, recall, accuracy and F-measure.

The formulations are as follows:

$$\text{Precision} = \frac{tp}{tp+fp} \quad \text{-- (3)}$$

$$\text{Recall} = \frac{tp}{tp+fn} \quad \text{-- (4)}$$

$$\text{Accuracy} = \frac{tp+tn}{tp+tn+fp+fn} \quad \text{-- (5)}$$

$$\text{F-Measure} = 2 \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad \text{-- (6)}$$

Advantages and Disadvantages of the following techniques used for recommendation system are as follows:

Table 2. Advantages and Disadvantages

Technique	Advantages	Disadvantages
K-Means	If variables are huge than k-mean most of the times is computationally faster if we keep k small	Difficult to predict the k value
Naive Bayes	Fast to train and space efficient. Handles real and discrete data	Assumes independence of features
KNN	Cost of learning process is zero	It is computationally expensive to find k nearest neighbor when dataset is very large

10. CONCLUSION

After comparing both clustering analysis and classification the conclusion made is that clustering analysis is much more optimized or rather useful method in data mining. The study makes the contribution by focusing on rural consumers as well which was absent and limited to urban new car buyers only. Clustering method is fast, robust and easier to understand. It gives best result when data set are distinct or well separated from each other.

Based on the findings, the organization can design their marketing strategies in a particular way in order to influence the customer behavior. Future studies could focus on external information search behavior as well. In order to minimize the extent of errors a compact set of data was considered, whereas this can be overcome in future using enormous amount of previous data, which would help in assessing the behavior patterns even more accurately.

11. REFERENCES

- [1] F.O. Isinkaye, Y.O. Fojimi, B.A. Ojokoh. Recommendation Systems: Principles, Methods an evaluation, Volume 16, Issue 3, pp: 261–273, 2015
- [2] VikramShinde, Analysis of Research in Consumer Behavior of Automobile Passenger Car Customer. International Journal of Scientific and Research Publications, Volume 4, Issue 2, 2014.
- [3] L.V. Bijuraj. Clustering and its Applications. Proceedings of National Conference on New Horizons in IT-NCNHIT 2013.

- [4] Liu Dongyan, Bai Xuan. Car Purchasing Behavior in Beijing: An Empirical Investigation. Umea School of business and economics, University of Umea, Master Thesis, Spring Semester, 2008.
- [5] Ying Liua,b,* Jiajun Yanga. Improving Ranking-based Recommendation by Social Information and Negative Similarity, Volume 55, pp: 732–740, 2015.
- [6] Guan Hai-ling, Cheng Bao-dong. Analysis on the Consumer's online Information Search Behavior. E-Business and Information System Security, pp1 – 4, 2009.
- [7] RamitaVerma, ShubhkamanaRathore, PEST Analysis for Indian Luxurious Car Market, IJMSSR Volume 2, pp 22-26, 2013.
- [8] Nikhil Monga, Bhuvender Chaudhary, Car Market and buying behavior - A study on Consumer Perception, IJRMEC Volume 2, Issue 2, pp: 44-63, 2012.
- [9] Konstan JA, Riedl J. Recommender systems: from algorithms to user experience. User Model User-Adapt Interact Volume 22, Issue 1-2, and pp: 101-123, 2012.
- [10] Klein, L. R., & Ford, G. T. Consumer Search for information in the digital age: An empirical study of pre-purchase search for automobiles. Journal of Interactive Marketing, Volume 17, Issue 3, pp: 29–49, 2003.
- [11] Raj Kumar,Dr. Rajesh VermaClassification Algorithms for Data Mining:A Survey,International Journal of Innovations in Engineering and Technology (IJJET), 2012.
- [12] AnoopJain, ArunaBajpai, Manish Kumar Rohila,International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com, Volume 2, Issue 6, 2012.