

Research on Apparel Sales Forecast Based on ID3 Decision Tree

Algorithm

Xurigan Saiyin^{1,a}, Chenna Hu^{1,a}, Dou Tan^{1,a}, Yi Liu^{1,a*}

¹School of Information Engineering, Beijing Institute of Fashion Technology, China, 100029.

^amableliuyi@163.com

Key words: ID3 decision tree, clothing market sales impact factor.

Abstract: Today's clothing industry should get consumers' demands in a changeable market to confirm a market scheme. Thus it's important for enterprises to make a prediction of it before their production lines begin. The clothing markets are determined by a widely range of individual factors such as age, gender, personal preference and finance, consumption levels and habits and so on. In this paper, we consider all kinds of individual factors, It also introduces the idea and the concrete algorithm of the traditional ID3 decision tree algorithm for the accurate prediction of apparel sales market and the industry related clothing sales forecast of the traditional method. In the end is the comparison and summary.

Introduction

The analysis of domestic and international clothing sales forecast method and its research status

In the diversified competition of apparel market, customer demand for clothing variability also make the enterprise is often at the same time facing best-selling product shortages and unsalable product backlog. Apparel sales in addition to the region, weather, culture, economy and other factors, mainly by the consumer's own factors, distribution and other personal factors, resulting in difficult to accurately predict market demand, can not provide effective support for operational decisions. At present, the research on the forecast of sales can be divided into two categories.

One is the personal judgment method or qualitative prediction based on Delphi method. Personal judgment method, namely the sales management personnel based on personal experience, the sales forecast. Such methods are based on the subjective judgment of the sales data and the subjective factors, which are often lack of objectivity and reliability. Another kind is the prediction method on the basis of the existing mathematical model of quantitative, with in-depth study, the researchers began using different quantitative methods to achieve demand forecast, moving average method, index smoothing method, seasonal index method, a yuan return or cloudy regression method, time sequence column model prediction, Markov chain prediction model. This method has a good accuracy in predicting the future sales of apparel enterprises.

Through the analysis of domestic and international clothing sales forecast research situation, it is not difficult to find that the current research on the prediction method has been improved, but the research of forecasting methods is more emphasis on the comprehensive application of multiple forecasting methods to get the final forecast value. However, for the management of clothing enterprises, the other focus of attention is what factors affect the sales, and the above model is not related to the analysis of the factors in the sales. In this paper, according to the characteristics of the

clothing industry, the paper seeks to influence the different factors from the consumer age, personal economic situation, product range, product price range, product style, and individual year shopping. This method not only can effectively predict the demand of apparel sales, but also can be easily transformed into association rules, which helps managers to understand the main influencing factors, and to formulate rational production and marketing strategies.

ID3 decision tree algorithm

Decision tree is based on the probability of the occurrence probability of a variety of circumstances, through a decision tree to find the net present value of the expected value is greater than zero, the evaluation of project risk, to determine the feasibility of the decision analysis method, is a graphical method of using probability analysis. Because of this decision branch painted graphics like the branches of a tree, so that the decision tree. In machine learning, decision tree is a prediction model, which represents a mapping relationship between object and object value. Entropy is a disorder of the system, using the algorithm of ID3, C4.5 and C5.0 spanning tree algorithm using entropy. This metric is based on the concept of entropy in information theory. Decision tree is a tree structure, where each internal node represents a property of the test, each branch represents a test output, each leaf node represents a category.

About ID3 algorithm:

The basic decision tree algorithm is a greedy algorithm, which in a top-down recursive way to construct the decision tree crush one by one. A well-known decision tree algorithm is ID3, the basic strategy of the algorithm is as follows:

a.Create a node. If the sample is in the same class, the algorithm stops and changes the node to a leaf node, and uses the class label.

b.Otherwise, select one of the best attributes that can be classified as the training set, which is used as the test attribute of the node.

c.To create a corresponding branch, and then divide the sample into a sample.

d.Using the same process, the top-down recursive, until meet the following three conditions in a recursive stop.

All the samples of a given node belong to the same class.

No remaining attributes can be used to divide.

Branch no sample.

Let S be a collection of s data samples. If the class label attribute has different values, the m C_i ($i = 1, 2, \dots, M$). Let S_i be the number of samples in the class C_i . The desired information is given by the next type of a given sample classification:

$$E(A) = \sum_{j=1}^v \frac{s_{1j} + s_{2j} + \dots + s_{mj}}{s} I(s_{1j}, s_{2j}, \dots, s_{mj}) \quad (1)$$

Where P_i is an arbitrary sample of the probability of belonging to the C_i , and s_i/s estimates.

A V has a different set of attribute values $\{A_1, a_2, \dots, A_V\}$. Can use A attribute of S is divided into v sub set $\{S_1, S_2, \dots, S_v\}$; Among them, S_j, S_v ; S contains some samples, they have a value of A_j in A . If A is chosen as the test attribute (i.e., the best splitting attribute), the subset corresponds to a branch that grows out of the nodes that contain the S .

Let S_{ij} be the sample number of C_i in S_j . According to the entropy or the expected information of the subsets by A , the information is given by the following formula:

$$I(s_1, s_2, \dots, s_m) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (2)$$

Among them, $\frac{s_{1j} + s_{2j} + \dots + s_{mj}}{s}$ is the right of the j sub set, and is equal to the number of

samples in the subset (i.e., A value AJ) divided by the total number of samples in S. The smaller the entropy, the higher the purity of subset partition. Note that for a given subset Sj,

Among them, $p_{ij} = \frac{s_{ij}}{|s_j|}$ is the Sj of the sample belongs to the class Ci probability. The encoding

information on the A will be obtained by the:

$$Gain(A) = I(s_1, s_2, \dots, s_m) - E(A) \quad (3)$$

Gain (A) is known as the information gain, which is due to the entropy of the expected value of the A. Attribute selection with the highest information gain for a given set of S test properties. Create a node, and mark the property, create a branch for each value of the property, and then divide the sample.

Forecast model of apparel sales based on clustering and decision tree.

In order to allocate the existing resources more effectively, realize the maximization of enterprise profit, it is needed to analyze the specific market environment of different sales outlets by means of scientific method. Decision tree algorithm can be used to analyze the input data. So, we can find out the influence factors of the external factors and the internal factors, which are the input of the decision tree, and use the decision tree algorithm to form the classification rules between the factors and target sales. At the same time, we can use the existing rules to predict the future market. Based on the above analysis, this paper proposes a sales forecasting model of ID3 decision tree algorithm. This method first from the overall sales of clothing, through clustering analysis, the historical records of consumer age "25 years of age, 25-35 years old, 35 years old," the three major groups, combined with the actual analysis requirements, and then set up a series of factors including geographical factors, economic factors, cultural factors and enterprise itself, and so on. In this paper, we select the "consumer gender, economic situation, the purchase price range, clothing style, the annual average number of shopping" several major factors affecting the analysis, which means the economic situation is divided into "rich, ordinary, poverty" 3 levels. Buy clothing price range is divided into "200 the following, 200-600, 600-1000", clothing style divided into "fashion models, the basic models, sports leisure," the three categories, the annual average shopping number is divided into "5-10, 10-20, 0-5," decision tree will be these attributes and three age range of the Association rules; finally through the decision tree can be used to forecast the future sales of properties, which can guide the management of scientific decision-making.

Table1.Method of sample classification

Consumer age class	Consumer gender	Economic situation	Purchase price range	Clothing style	Average shopping frequency
25 years old below	Male	rich	200 below	fashion models	0-5
25-35 years old	Female	ordinary	200-600	the basic models	5-10
Over 35 years old		poor	600-1000	sports leisure	10-20
			1000 above		

Decision-making process:

We first identify the clothing styles for the class, that is, three types: S_1 =basic models, S_2 = fashion models, S_3 =sports and leisure paragraph, and calculating the entropy of I, $I = S_1/100 * \text{Log}_2 S_1/100 + S_2/100 * \text{Log}_2 S_2/100 + S_3/100 \text{Log}_2 S_3/100$. After the calculation of other conditions, such as the condition of gender, the total number of samples is 100 people, including 49 men, 51 women. In the 49 men, we can know which number of men's preference of the basic models for S_{11} , number of men's fashion Preferences section for the S_{12} , number of men preference for leisure sport models for S_{13} , the entropy I_1 . The lady clothing preference entropy I_2 easily come to the conditional expectation value of information E, $E = 49/100 * I_1 + 51/100 * I_2$. The information gain value of the condition is $\text{Gain} = I - E$. In the same way we calculate the information gain values for all other conditions. By comparing these Gain values, we obtain the "economic conditions" condition of the Gain value, so we choose the "economic situation" in the initial node of the decision tree, and then calculate the nodes in the first node. A total of three calculations were carried out, and the rough decision tree structure was obtained as follows.

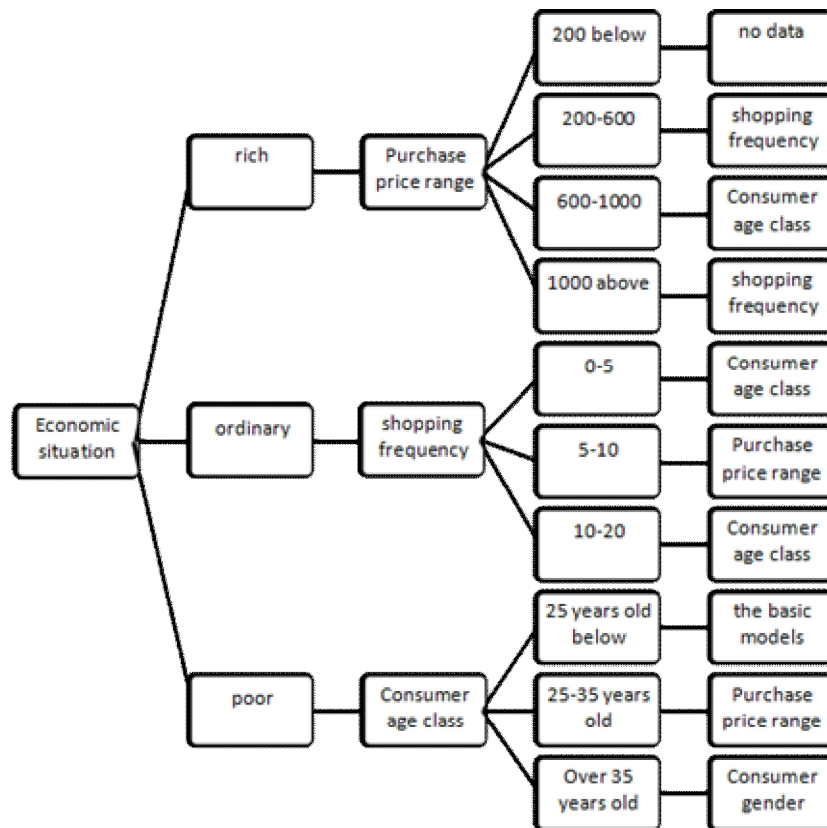


Figure1. The Resulting Tree

The detailed data of the final result is given in Appendix excel, including the original sample data, the basic frame data of decision tree, and the detailed data of the decision tree.

The whole model mainly includes the cluster analysis, the establishment of index system and the establishment of a decision tree model 3 steps, the main process is shown in the following figure.

Summary:

Through this way, we can get the people in different age and different economic situation to be more inclined to those grades and styles.

It is advantageous to the clothing enterprise to carry on the more scale economy to the product domain. It is advantageous to the clothing market specialization, the plate development. It is advantageous to save the cost of the cost of the cost of the clothing enterprises to explore the market. Improve the efficiency of enterprise inventory management, improve the security of short term liquidity, and stabilize the assets structure of the enterprise. Taking into account the complexity of the apparel sales market, the decision tree algorithm can be other algorithms, such as genetic algorithms, neural networks, and other intelligence can be considered, this method pending further discussion.

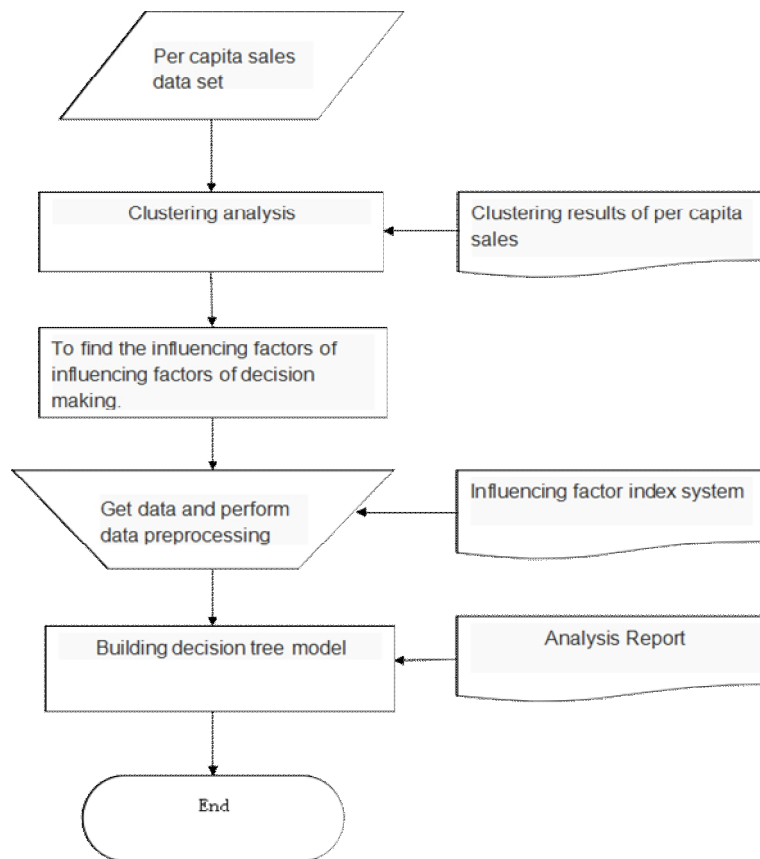


Figure2. Modeling Flow Chart

Acknowledgments

This work is supported by General program of science and technology development project of Beijing Municipal Education Commission (KM201510012005), Beijing College Student Research Training Program (2015) , Beijing Institute of Fashion Technology Teaching reform and innovation team project (JGTD-1404), Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality, PHR(IHLB) of Innovative Research Team PTTBIFT_TD_002,and Project of Beijing Institute of Fashion Technology Specialty construction(RCPY20150024).

References:

- [1] Celia frank, Ashish Garg, Amar raheja, et al. Forecasting women 's apparel sales using mathematical modeling [J]. International Journal of clothing science and technology, 2003, 15 (2):107-125.
- [2] Yanmin Wan, Chen Sheng, Shu Jiao Dai. Based on time series and pert of clothing sales forecasting method study [J]. Silk, 2006 (11).
- [3] Linyan Yu. Stochastic demand of clothing products of grey prediction model [J]. International textile bulletin, 2006 (8): 78 - 80.
- [4]Xue Meijun, Shen Jianjian, Yang Yixiong. Quantitative forecasting method for apparel sales (J). Journal of Donghua University, 2004, 30 (6): 77 - 75