IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 25, Issue 6, Ser. 1 (Nov. – Dec. 2023), PP 28-31
www.iosrjournals.org

Research On The Application Of Cluster Analysis In The Automotive Industry

Chu Fang
College of Economics and Management, Zhaoqing University, Zhaoqing City, Guangdong, China

Abstract:
In recent years, with the improvement of people's living standards, cars have gradually entered people's daily lives, thus driving the rapid development of the automotive industry. At the same time, there are many types of cars in the market, and there are various performance indicators to measure car quality. Therefore, this study will explore the inherent relationship between car types and their performance indicators through multiple statistical methods, and cluster various cars based on these indicators, such as dividing them into three levels of quality: excellent, medium, and good, which can provide guidance for car buyers and sellers.

Date of Submission: 03-11-2023 Date of Acceptance: 13-11-2023

I. Data Introduction

This data mainly comes from "19 sets of data and their explanations", which records data on several vehicle performance indicators, including 23 samples. Each sample is mainly described by 8 variable indicators, namely economy, service, value, price, design, sport, safety, and ease.

II. Data Processing

Before conducting data analysis, first use descriptive statistical analysis methods to determine the general level of performance indicators for automobiles in the market. This mainly starts with the calculation and description of basic statistics (such as mean, variance, standard deviation, maximum/minimum, skewness, kurtosis, etc.), and is assisted by the graphical functions provided by SPSS to grasp the basic characteristics of the data and the overall distribution characteristics.

In this case, by comparing the average and maximum/minimum performance indicators of different vehicle models (such as A100, BMW 3, CiAX, etc.), it is possible to determine which ones have better performance and which ones are worse overall.

III. Research Method

Cluster analysis is a method of classifying research objects based on their individual characteristics. Classification is widely used in fields such as economics, management, sociology, medicine, etc. Cluster analysis can automatically classify a batch of sample (or variable) data based on its many characteristics and their degree of familiarity in nature without prior knowledge, resulting in multiple classification results. There is similarity between individual features within a class, and there are significant differences in individual features between different classes.

After conducting basic descriptive statistics on the data, we need to conduct cluster analysis on 23 vehicle models, roughly divided into 3-5 categories, with performance ranging from poor to excellent. In this case, we will use two methods for clustering: system clustering and K-means (fast clustering).

The basic principle of system clustering method: first, a certain number of samples or indicators are considered as one category, and then based on the degree of familiarity of the samples (or indicators), the two categories with the highest degree of familiarity are merged. Then, the degree of familiarity between the merged class and other classes is considered before merging. Repeat this process until all samples (or indicators) are merged into one category.

System clustering can be divided into two types: Q-type clustering and R-type clustering. Q-type clustering is the clustering of samples, which gathers samples with similar features and separates samples with significant differences; R-type clustering is the clustering of variables, which separates variables with significant differences and gathers similar variables together. This allows for the selection of a few representative variables from similar variables to participate in other analyses, achieving the goal of reducing the number of variables and reducing the dimensionality of variables.

In this example, Q-type clustering is performed.

There are several main methods for calculating the distance between classes:
(1) The Nearest Neighbor method refers to the minimum distance between each individual of two classes;
(2) The farthest neighbor method refers to the maximum distance between each individual of two classes;
(3) Between groups Linkage refers to the average distance between individuals of two categories;
(4) Within group linkage refers to taking into account the distance between all individuals of two categories;
(5) The Centroid clustering method refers to the distance between two class center points;
(6) The Ward method requires that the sum of squares of deviations for similar samples should be smaller, and the sum of squares of deviations between classes should be larger.

The K-means method (also known as fast clustering method) was proposed by MacQueen in 1967. It treats data as points in a K-dimensional space, uses distance as an indicator to measure individual "familiarity", and sacrifices multiple solutions for high execution efficiency. However, the K-means method can only produce

clustering results with a specified number of classes, and the determination of the number of classes cannot be separated from the accumulation of practical experience.

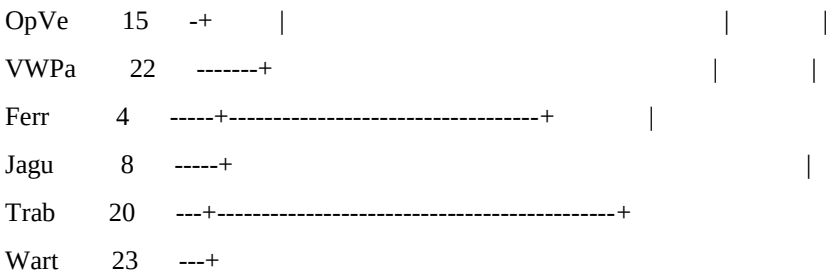
The basic idea of fast clustering analysis is to first select a batch of aggregation points (centers) according to a certain method, then let the samples condense towards the nearest aggregation point to form an initial classification, and then modify the unreasonable classification according to the principle of nearest distance until it is reasonable. Therefore, in fast clustering, users should first be required to provide their own information on how many classes they need to cluster into, and ultimately can only output unique solutions about it. Fast clustering is an iterative classification process in which the class to which the sample belongs is continuously adjusted until it reaches stability.

The following figure is a tree like clustering graph generated using the "inter group connection" clustering method. If all samples are divided into three categories, as shown in the figure, the first category includes 14 models including A100, FoFi, Hyun, Mazd, Mits, NiSu, OpCo, OpVe, P306, Re19, Rove, ToCo, VWGo, VWPa, etc. The second category includes 4 models including BMW 3, Ferr, Jagu, M200, etc. The remaining 5 belong to the third category.

Dendrogram using Average Linkage (Between Groups)
Rescaled Distance Cluster Combine

```
C A S E      0              5             10             15             20             25
Label   Num  +---------+---------+---------+---------+---------+

P306     16   -+
Re19     17   -+
FoFi      6   -+
ToCo     19   -+-+
Mazd     10   -+ |
Hyun      7   -+ +---+
NiSu     13   -+ |   |
OpCo     14   ---+   +-----------+
CiAX      3   -+-+   |           |
Lada      9   -+ +---+           |
FiUn      5   ---+               +----------------------+
M200     11   -+---+             |                      |
VWGo     21   -+   +---+         |                      |
BMW3      2   -----+   |         |                      |
A100      1   -+-+     +---------+                      +-------+
Rove     18   -+ +---+ |                                |       |
Mits     12   -+-+   +-+                                |       |
```



IV. Result

Conclusion 1: Comparison of average levels of different performance indicators

The three indicators with higher average performance level are power (3.4652), economy (3.293), and safety (3.2870); The lowest average performance level is comfort (2.7870). The indicators with significant fluctuations in performance include price (1.412) and safety (1.259), while comfort (0.318) has the best stability.

Conclusion 2: Comparison of average performance levels of different vehicle models

Comparing the output results of the system clustering method and the K-means clustering method, it can be seen that their clustering results are roughly the same. A more reasonable clustering method is to divide all samples into three categories. The first category includes 14 car models, including A100, FoFi, Hyun, Mazd, Mits, NiSu, OpCo, OpVe, P306, Re19, Rove, ToCo, VWGo, VWPa, etc. The second category includes 4 car models, including BMW 3, Ferr, Jagu, M200, etc. The remaining other car models belong to the third category. Obviously, the performance level of the third category is the best, with the first category being in the middle and the second category taking the second place.

References

[1]. Gurley, J.W. The One Internet Metric That Really Matters. Fortune 2000, 2, 141–392.

[2]. Bickerton, P. 7 Technologies That Are Transforming The Hospitality Industry. Hosp. Mark. 2015, 1, 14–28.

[3]. Pui, L.T.; Chechen, L.; Tzu, H.L. Shopping Motivations On Internet: A Study Based On Utilitarian And Hedonic Value. Technovation 2007, 27, 774–787.

[4]. Lee, E.Y.; Soo, B.L.; Yu, J.J. Factors Influencing The Behavioral Intention To Use Food Delivery. Apps. Soc. Behav. Pers. 2017, 45, 1461–1474. [Crossref]

[5]. Armstrong, G.; Kotler, P. Marketing; Prentice-Hall: Englewood Cliffs, NJ, USA, 2000; Pp. 1–98.

[6]. Ozkara, B.Y.; Ozmen, M.; Kim, J.W. Examining The Effect Of Flow Experience On Online Purchase: A Novel Approach To The Flow Theory Based On Hedonic And Utilitarian Value. J. Retail. Consum. Serv. 2017, 37, 119–131. [Crossref]

[7]. Park, J.; Ha, S. Co-Creation Of Service Recovery: Utilitarian And Hedonic Value And Post-Recovery Responses. J. Retail. Consum. Serv. 2016, 28, 310–316. [Crossref]

[8]. Anderson, K.C.; Knight, D.K.; Pookulangara, S.; Josiam, B. Influence Of Hedonic And Utilitarian Motivations On Retailer Loyalty And Purchase Intention: A Facebook Perspective. J. Retail. Consum. Serv. 2014, 21, 773–779. [Crossref]

[9]. Chiu, C.M.; Wang, E.T.; Fang, Y.H.; Huang, H.Y. Understanding Customers' Repeat Purchase Intentions In B2C E-Commerce: The Roles Of Utilitarian Value, Hedonic Value And Perceived Risk. Inf. Syst. J. 2014, 24, 85–114. [Crossref]