

Application of Association Analysis and Visualization Methods in Car Parts Repair

Ye Li^{1, a *}, Xiaoping Du^{2, b} and Bo Yang^{3, c}

^{1,2,3} College of Software, Beihang University, Beijing, China 100191

^aleyee1991@163.com, ^bxpdu@buaa.edu.cn, ^cyangbo@buaa.edu.cn

Keywords: data mining, car parts repair, association rules, visualization

Abstract. This article combines association analysis and visualization methods, mining relations and laws for car parts repair. Firstly, process the original data into data that can be used in the association analysis, then use apriori algorithm of association rules method for mining association relationships among car parts repair, geography, months and car models, and through the establishment of a network model to visualize the association rule results dynamically and visually. By this application, we can get the relationship between car parts repair and several important attributes, and use visual tools to show the results, helping give decision proposal of car parts inventory.

Introduction

By the end of 2014, car ownership in China has reached 140million, and car parts repair data has become very huge. It can be collected that part name, part category, service areas, service time and car model information from the existing car parts repair data. Data queries, input and output, storage and statistics functions can be achieved by traditional mathematical statistics and database technology. However, we cannot predict the potential development trends and exhume the relationship between data based on available data^[1]. It is a problem needs to be solved that how to use a large amount of data, mining the potential relationship in the data and displaying the results, so as to make the management of the enterprise see the results directly.

Data mining technology can find valuable information through analyzing car parts repair data. There are a lot of applications of data mining in the automotive filed^[2-7].

Association rules technology^[8] which is proposed by R.Agrawal et al in 1993 can give relationships and reveal unknown dependencies among data items. It is applied widely in car sales customer group analysis, emission detection, car crash simulation and so on^[9-13], but the application in car parts repair is still rare.

Using the association rules technology, the relationship between car parts damage and car models, months and cities can be obtained from massive maintenance data. It can not only help to identify safety problems, but also it can provide references for car design in order to increase its life. Visualization methods have been used in automotive industry^[14-18], and combine them with car parts repair data can show the rule results intuitively.

This article combines association rules technology and visualization methods from different dimensions such as months, cities, car models and car parts to analyze car parts repair data, mining relationships among the dimensions and visualize the results.

Description of the Problem and Solutions

The scale of car repair data is growing rapidly, and traditional mathematical statistics and database technology cannot analyze the hidden relationships among the important dimensions in car parts repair data. This article expects to mine what is related in parts name and repair time, area, car models. What's more, show the rule results using visualization methods, and guide car parts inventory decisions. At present, there isn't a common data mining method, which can complete a process that analyze the correlation relationships among actual data and show the results.

This article presents a general method of data analysis. Data acquisition and data preprocessing is the first step, then select analysis method for analysis, and show the analysis results in the end. This article will accomplish the goals through three steps. The first step is data processing, which is converting the original data into an acceptable form. Then dig out the relationships among parts and repair time, cities and car models. At last, show the association rules results using network diagrams.

Data Preprocessing

Introduction of the Original Data. The data of this study used is from an automobile corporation, and it is a total of 1.7 million vehicle maintenance records within two years from 2010 to 2011. The records contain car warranty data of all the service stations all over the country of this corporation. There are more than one thousand service stations, and they are located in 32 provinces and 325 cities.

This study uses “Warranty List” and “Service Station Contracts” of the original data. The main contents of the two tables are shown in table 1 and table 2.

The “Warranty List” contains five fields, namely “maintenance number”, “part number”, “part name”, “service station number” and “car model”. “Maintenance number” marks an event that one car was repaired in the service station. It is formatted as WX+ warranty date(format YYYYMMDD)+6-digit number, such as WX201001010001. Every record in “warranty list” represents that one part is used once. If a maintenance event used two part A and one part B, there are three records in the “warranty list”. For example, from 1-3 line in table 1 we can get that the maintenance event whose maintenance number is “WX20100926000004” used two “Rotary mounted oil filter” and one “Left front plate sixth pieces of spring”.

Table1 Warranty List

Maintenance Number	Part Number	Part Name	Service Station Number	Car Model
WX20100926000004	61000070005A1307A	Rotary mounted oil filter	GUX00013	BJ4253SNFKB-1
WX20100926000004	61000070005A1307A	Rotary mounted oil filter	GUX00013	BJ4253SNFKB-1
WX20100926000004	1327129280002-6A20	Left front plate sixth pieces of spring	GUX00013	BJ4253SNFKB-1
WX20100926000003	1424235622002A1285	Trailer brake valve	ANH00020	BJ4253SNFKB-1

“Service Station Contracts” records the geographic information of the service stations. It has four fields, namely “Service Station Number”, “Service Station Name”, “Province” and “City”.

Table2 Service Station Contracts

Service Station Number	Service Station Name	Province	City
GUX00013	Liuzhou YuanXiang Auto Trade Co., Ltd	Guangxi	Liuzhou
ANH00018	Fuyang Anhui Auto Sales Service Co	Anhui	Fuyang
ANH00020	Anhui spring rain Auto Sales Service Co	Anhui	Haozhou

Generation of the Analysis Data Set. After analysis, we choose four dimensions, namely months, cities, car models and parts to analyze the relationship among them in Car Parts Repair. The result cannot be got from the original data directly, and it is needed that reorganizing the original data to get the input data set of the association analysis, shown in table 3.

Table3 Service Station Contracts

Transaction ID	Item set
maintenance number	Part name, month, city, car model

There are two fields which are part name and part number can represent the part in the original data. The same parts may come from different manufacturers, so part number of the same part name may be different. In the study, the analyzed object is what part is easy to be damaged, and don't care about the quality difference between the same parts from different manufacturers, so select the part name as an input field of the association analysis.

The association analysis expects to get the relationship among parts, time, region and car models. As there is no time properties in the original data and the “maintenance number” field contains the

data information, extract the 7-8 bit character of the “maintenance number” to generate month field. As for region information, we choose “city” as the analyzing property. Car model can be got from the “Warranty List” table directly.

Connect table “Warranty List” and table “Service Station Contracts” with the "maintenance number" field, extract the month field from the "maintenance number" field, and filter out the irrelevant field. "Maintenance number" needs to be kept as it is the transaction ID for the association analysis. If a part is repaired more than once in the same repair event, two or more identical records will appear in the “Warranty List”. This condition needs to remove the duplicate records. Data after processed is shown in table 4, and there are five fields, namely “Maintenance Number”, ” Part Name”, ” Car Model”, ”Month” and “City”.

Table4 Service Station Contracts

Maintenance Number	Part Name	Car Model	Month	City
WX20100926000004	Rotary mounted oil filter	BJ4253SNFKB-1	09	Liuzhou
WX20100926000004	Left front plate sixth pieces of spring	BJ4253SNFKB-1	09	Liuzhou
WX20100926000001	Left front plate seventh pieces of spring	BJ3251DLPJB-2	09	Fuyang

Association Rule Mining

Selection of Algorithm. Apriori algorithm is the most classical association rule algorithm, which is proposed by Agrawal Rama Krishnan and Skrikan Rakesh in 1994. Association analysis is divided into two steps. The first step is to generate frequent item sets. The goal is to find all the sets that satisfy the minimum support threshold. The second step is to extract all the rules that satisfy the minimum conference from the frequent item sets.

Association Rules Mining. In order to complete the rule mining process, use the typical data mining tool SPSS Clementine in this study. Use variable file node to read the original data from free-field text files, and use filter node to filter out the fields that we don’t concern about. Use type node to set the type and direction of the fields. For the storage type of the “month” field is integer, in order to process easily in the association analysis, use a filler node to convert it to string. There are two types of data that the association rules model can receive in SPSS Clementine, that is, the transaction table and the fact table. The data this study uses is too large to use fact table in the association analysis, so convert it to form of transaction table, see table 5.

Table5 Maintenance Transaction Table

Maintenance Number	Maintenance Item
WX20100926000004	Rotary mounted oil filter
WX20100926000004	Left front plate sixth pieces of spring
WX20100926000004	BJ4253SNFKB-1
WX20100926000004	09
WX20100926000004	Liuzhou
WX20100926000001	Left front plate seventh pieces of spring
WX20100926000001	BJ3251DLPJB-2
WX20100926000001	09
WX20100926000001	Fuyang

The process is shown in figure 1.

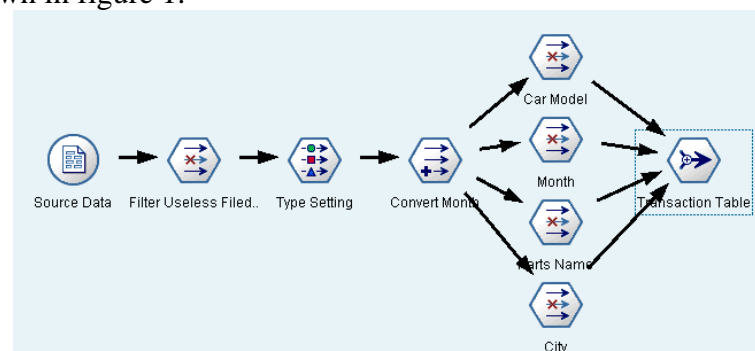


Fig. 1 Transaction table genaretion process

Use Apriori node to establish the association analysis, see figure 2.

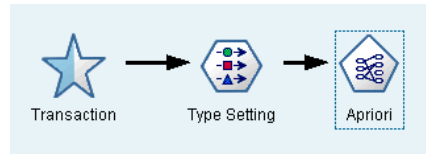


Fig. 2 Association rule mining model

Association Rules Results. By setting different support and confidence threshold, we can generate different association rule sets. Set the minimum support 0.2%, and the minimum confidence 20%, and 571 rules are obtained. Among them, the maximum support is 2.898%, the minimum support is 0.201%, the maximum confidence is 93.131%, the minimum confidence is 20%.

All the association rules have one consequent. 422 rules have one antecedent, 139 rules have two antecedents and 10 rules have three antecedents. Table 6 shows parts of the rule results.

Table6 Association rule results

Antecedent 1	Antecedent 2	Antecedent 3	consequent	support	confidence
Balanced axle housing	Spherical bearing	Shaft sealing ring	Bearing oil seal cover	0.25	42.49
Sync assembly	BJ4253SMFKB-S		Vice box drive gear	0.22	35.77
Yinchuan			BJ3313DMPKJ-S	0.30	47.62
Compressor assembly			09	0.31	45.95

As is shown in table 6, the first rule has three antecedents, and it represents when balanced axle housing, Spherical bearing and Shaft sealing ring are used in the same repair event, the possibility of bearing oil seal cover used is 42.49%. The second rule means that when a car whose model is "BJ4253SMFKB-S" replaces Sync assembly, there are 35.77 percent possibilities that vice box drive gear is replaced together. The third rule means that 47.62% of the repair events' car model in Yinchuan is "BJ3313DMPKJ-S". The fourth rule means that 45.9 percent of Compressor assemblies are repaired in month 9.

The above rules can effectively guide the storage of automobile parts. Such as Yinchuan should store more cars' parts whose model are "BJ3313DMPKJ-S". The storage of compressor assembly in September should be increased.

The choice of regional and time properties can be adjusted freely according to the original data. For example, month can be replaced by week, quarter of the year, year and so on. As to region, this study chooses city that each service station located in, it can also be reduced to towns, urban areas, or the rise for provinces, state and so on.

Rules Visualization

In order to display the results of association rules visually, select parts of the rules results as a data source, and use the network diagram to visualize it.

In SPSS Clementine, use a select node to select a subset of the rules results as the source data of the network graph. Rules of the subset have only one antecedent, and the minimum support is 0.4%, the minimum confidence is 40%. Use a web node to show the network graph. Choose the consequents and antecedents as the nodes of the graph and rules as the edges, confidences as the edge weights, and we can obtain the following figure.

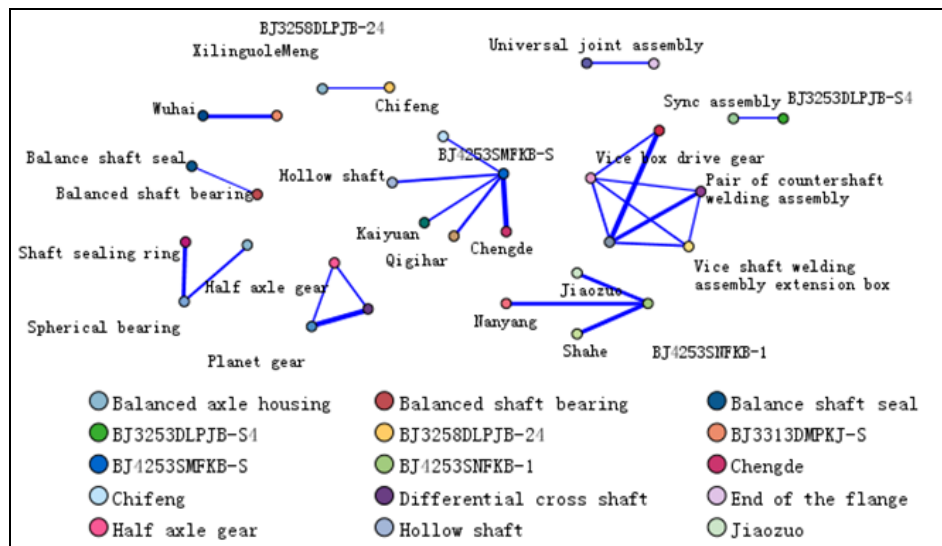


Fig. 3 Network Graph

In Figure 3, each side represents a rule. The thicker the edge is, the greater the confidence of the rule is. It indicates that the association relationship is more likely to occur in the future.

Summary

This article processes two-year car parts repair data of an automobile corporation, screening four properties such as month, city, car types and car parts and mining the relationship between car parts repair data and the month, the city and the model of the car. Setting the minimum support degree as 0.2%, the minimum confidence as 20%, 571 rules are obtained. Then filter out the rules of single antecedent, setting the antecedent and the consequent as nodes and the confidence as edge weights. Finally construct the network model and show the results of association rules dynamically and visually.

References

- [1] Yu Yan. Multidimensional Association Rules Studying and Applying in Students Information System[D]. Shenyang University of Technology, 2003.
- [2] Huibing L I. Urban travel time prediction model based on floating car data mining[J]. Computer Engineering & Applications, 2012.
- [3] Buddhakulsomsiri J, Zakarian A. Sequential pattern mining algorithm for automotive warranty data[J]. Computers & Industrial Engineering, 2009, 57(1):137–147.
- [4] Liang Y H. Integration of data mining technologies to analyze customer value for the automotive maintenance industry.[J]. Expert Systems with Applications An International Journal, 2010, 37(12):7489–7496.
- [5] Kargari M, Sepehri M M. Stores clustering using a data mining approach for distributing automotive spare-parts to reduce transportation costs[J]. Expert Systems with Applications, 2012, 39(5):4740–4748.
- [6] Bao-Lin L I, Chen Y S. Data Mining Technology in the Field of Applied Research Car Sales[J]. Auto Mobile Science & Technology, 2011.
- [7] Wang X F, Miao L F, Yan L I. Data Mining in Obtaining Potential Car Consumers[J]. Mathematics in Practice & Theory, 2013.

- [8] Agrawal R, Imieliński T, Swami A. Mining Association Rules Between Sets Of Items In Large Databases[C]// SIGMOD '93 Proceedings of the 1993 ACM SIGMOD international conference on Management of data1993:207--216.
- [9] Xuemei Yang, Yingjun Li, Rong Long. Application of Data Mining in the Analysis of Automobile Sales[J]. Shandong Science, 2006, 19(3):75.
- [10] Fei Liu. Application of Association Rules in Automobile Sales[J]. Science and Technology Innovation Herald, 2013(1): 6.
- [11] Meichun Peng, Wentao Wang, Yiqing Lin. Study of Association Rule of BASM Test Method for Gasoline Vehicle Exhaust Emission[J]. Environmental monitoring management and technology, 2008, 20(1):42.
- [12] Hong Liu, Xiang Li. Simulation of Car Crash Test in C-NCAP Analysis System Based on an Improved Apriori Algorithm[J]. Journal of Huaihai Institute of Technology, 2012, 21(2): 15.
- [13] Hang X, Dai H, Zhang Y. A Web content based data mining for car consumption preference in China[C]// Information Reuse and Integration, 2003. IRI 2003. IEEE International Conference onIEEE, 2003:235-242.
- [14] Xiaojian Liu, Zhixue Wang, Jianxin Li, Xuejun Liu. A Model-driven Visual Approach to Representing Automotive Software Requirements[J]. Shandong Science, 2009,5: 23.
- [15] Mingfei Zhou. Research on Visual Design and Application of Automobile Storage[D]. HeFei University of Technology,2008.
- [16] Libiao Jiang, Pan Dai, Zemao Chen. Analysis on the Vehicle Steering stability Visualization Simulation Technology[J]. Journal of Chongqing University, 2012, 35(9): 30.
- [17] Ziteng Wang, Diange Yang, Bing Li. Comparative Study of Visual Measurement Method for Moving Vehicle Noise[J].Journal of Vibration Engineering, 2011, 24(5): 578.
- [18] Xiaokang Cui, Rumin Zhu. Visualization Investigation of Automobile Performance Analysis Based on MATLAB[J]. Auto Parts,2013(4):71.