AIC Based Research on Cluster Analysis of Vehicle Equipment Technical State

Xiang-Bo ZHANG*, Hong SHEN, Lei ZHANG

Department of Technological Support Engineering, Academy of Armored Force Engineering, Beijing 100072, China
zxb2002226@shou.com
*Corresponding author

Keywords: Vehicle Equipment, Technical State Evaluation, Cluster Analysis, AIC.

Abstract. The technical state of vehicle equipment is the base of operate, manage and support. This paper researches its technical state cluster, which is based on AIC (Akaike’s information criterion). The conclusion shows that AIC is reliable and concise to the technical attributes of vehicle equipment.

Introduction

The technical evaluation of Vehicle Equipment is the base to operate, manage and support. The technology cluster is an important content during assessment [1-3]. Some authors proposes cluster methods, but those methods is not enough [4-7]. This paper gives the cluster method which is based on AIC (Akaike’s information criterion).

Cluster Analysis

Two-step Cluster

Cluster analysis is to classify something by quantity, and is a method which comprised with numerical taxonomy and multi-variables statistics. Although it is not enough well, it can be easy to be used. So it is a data tool to explore. Two-step is a new Hierarchical Algorithm. It is widely used. During cluster, the cluster number is vital. Nowadays, there are two criterions, BIC(Schwarz’s Bayesian Criterion) and AIC (Akaike’s information criterion) [8-9]. The AIC is adopted in this paper.

AIC

AIC is firstly proposed by H. Akaike. The principle of parsimony is its advantage. It can be defined as

\[ AIC = -2 \ln \left( \text{the likelihood function} \right) + 2 \left( \text{the number of free parameters} \right) \]  

(1)

For sample data, \( \{X_i | i = 1,2,\cdots,K\} \), the cluster number is \( L_m (m = 1,2,\cdots,N) \), and the central position is \( \{C_m | m = 1,2,\cdots,N\} \). The deviation is as \( \{D_m | m = 1,2,\cdots,N\} \), and the number of samples is \( \{Q(m) | m = 1,2,\cdots,N\} \). The distribution density in one cluster is as

\[ f(D_m) = \frac{Q(m)}{K} \frac{K}{d_{max} - d_{min}} = \frac{N}{K} \frac{Q(m)}{d_{max} - d_{min}} \]

(2)
In which,  
\[ d_{\text{max}} = \max \{ D_m \mid m = 1, 2, \ldots, N \} \]
\[ d_{\text{min}} = \min \{ D_m \mid m = 1, 2, 1, N \} \]

So can be obtained as

\[
l(D \mid C_1, C_2, \ldots, C_n) = \ln L(D_m \mid C_1, C_2, \ldots, C_n) \]
\[= -N \ln \frac{K}{N} - \sum_{m=1}^{N} \ln \frac{d_{\text{max}} - d_{\text{min}}}{Q(m)} \]

For AIC,

\[
AIC = -2 \left( -N \ln \frac{K}{N} - \sum_{m=1}^{N} \ln \frac{d_{\text{max}} - d_{\text{min}}}{Q(m)} \right) + 2N
\]
\[= 2 \sum_{m=1}^{N} \ln \frac{d_{\text{max}} - d_{\text{min}}}{Q(m)} + 2N \left( 1 + \ln \frac{K}{N} \right) \tag{3} \]

\[ K > N \] . When AIC is small, the cluster number is well accepted.

**Analysis Process**

Select vehicle equipments 102, which is marked with training vehicles and for training vehicle. There are 6 attributes with them, v1(operation marking), v2(vehicle body hours), v3(vehicle body kilometers), v4(motor hours), v5(logistical hours per year), v7(training hours per year).

Analyzed by SPSS software [10], The Table 1 is the auto-clustering from 15 to 1. It can be obtained that when AIC is 185.855, which is the minimum, the number of clusters is perfect. Considered the professional knowledge, when the cluster number is 2, AIC is 231.414. The former conclusion can be admitted.

<table>
<thead>
<tr>
<th>Number of Clusters</th>
<th>Akaike's Information Criterion (AIC)</th>
<th>AIC Change(a)</th>
<th>Ratio of AIC Changes(b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>505.446</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>231.414</td>
<td>-274.032</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>205.523</td>
<td>-25.89</td>
<td>0.094</td>
</tr>
<tr>
<td>4</td>
<td>185.855</td>
<td>-19.668</td>
<td>0.072</td>
</tr>
<tr>
<td>5</td>
<td>187.243</td>
<td>1.388</td>
<td>-0.005</td>
</tr>
<tr>
<td>6</td>
<td>194.492</td>
<td>7.249</td>
<td>-0.026</td>
</tr>
<tr>
<td>7</td>
<td>206.491</td>
<td>11.999</td>
<td>-0.044</td>
</tr>
</tbody>
</table>

Each cluster can be described by statistical indexes from Table 2.

Table 2. Each cluster centroids.
The frequency of each cluster is in Table 3. It shows that all training vehicles are included in cluser 2, and all the logical vehicles are in cluer 1. The conclusion is that the vehicle operation is valid and obvious.

Each variable has various contributions during clus ter. Figure 1 shows the status by t test. In cluster 1, except v4, the other variables have obvious function and the v5 is the most in Fig. 1 (a); In cluster 2, all variables have work, but the v4 is the most Fig. 1 (b).

**Table 3. Cluster frequency.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Cluster</th>
<th>1</th>
<th>2</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>v1 Mean</td>
<td></td>
<td>67.03</td>
<td>479.67</td>
<td>212.67</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td></td>
<td>38.053</td>
<td>316.132</td>
<td>273.559</td>
</tr>
<tr>
<td>v2 Mean</td>
<td></td>
<td>342.29</td>
<td>2657.25</td>
<td>1159.33</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td></td>
<td>225.413</td>
<td>2858.057</td>
<td>2024.685</td>
</tr>
<tr>
<td>v3 Mean</td>
<td></td>
<td>67.03</td>
<td>238.44</td>
<td>127.53</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td></td>
<td>38.053</td>
<td>122.392</td>
<td>113.576</td>
</tr>
<tr>
<td>v4 Mean</td>
<td></td>
<td>7.47</td>
<td>0.19</td>
<td>4.9</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td></td>
<td>9.479</td>
<td>0.822</td>
<td>8.383</td>
</tr>
<tr>
<td>v5 Mean</td>
<td></td>
<td>0.56</td>
<td>84.19</td>
<td>30.08</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td></td>
<td>4.554</td>
<td>57.926</td>
<td>52.814</td>
</tr>
</tbody>
</table>

Conclusion

This paper determines the cluster number by AIC, but the professional knowledge is also considered. Its advantage is helpful to study the variables contribution.
References


