LAB3 Association Rules 2

Wuhao Wang(wuhwa469) Hong Zhang(honzh073)

2/28/2022

# **HierarchicalClusterer**

## 1. Hierarchical clusterer with 2 clusters, single as linkType.

The number of incorrectly clustered instances is 61, which is 49.2%.

Graphical user interface, text, application

Description automatically generated

## 2. Hierarchical clusterer with 4 clusters, single as linkType.

The number of incorrectly clustered instances is 48, which is 38.7%.

Graphical user interface, application

Description automatically generated

## 3. Hierarchical clusterer with 6 clusters, single as linkType.

The number of incorrectly clustered instances is 57, which is 45.97%.

Graphical user interface, application

Description automatically generated

**SimpleKmeans algorithm**

## 1. Cluster data using the k means algorithm, 2 clusters, number of seed is 10.

The number of incorrectly clustered instance is 59, which is 47.58%.

Graphical user interface, application

Description automatically generated

## 2. Cluster data using the k means algorithm, 4 clusters, number of seed is 10.

The number of incorrectly clustered instance is 76, which is 61.29%.

Graphical user interface, text, application

Description automatically generated

## 3. Cluster data using the k means algorithm, 6 clusters, number of seed is 10.

The number of incorrectly clustered instance is 96, which is 77.42%.

Graphical user interface, application

Description automatically generated

By using the simpleKmeans and HierarchicalClusterer algorithms, we assume there are 6 variations. Compared with these variations, the best one is Hierarchical clustering with 4 clusters, single as linkType. In this variation, the number of incorrectly clustered instances is 48, which is 38.7%.

The clustering algorithms do not find a clustering that matches the class division in the database. Since simpleKmeans algorithm is suitable for clustering of convex sample sets. Meanwhile, in hierarchical clustering, it creates nonconvex clusters, it uses nearest neighbors’ methods for linkage. So, the shape of the cluster does not matter.

# **2 Use of association rules**

rules found for class 1:

1 attribute#5=1 29 ==> class=1 29 <conf:(1)>

2 attribute#1=3 attribute#2=3 17 ==> class=1 17 <conf:(1)>

3 attribute#1=2 attribute#2=2 15 ==> class=1 15 <conf:(1)>

4 attribute#1=1 attribute#2=1 9 ==> class=1 9 <conf:(1)>

From the rules 1, we can tell that attribute#5 =1 is assigned to class 1. From the rules 2-4, we can summarize more abstract rules: attribute#1 == attribute#2 is assigned to class 1. Finally, we think those clustering algorithms can not work well on monk1. Since all these algorithms are based on calculating ‘distance’ on each feature and combine them together. And for the rules we found, ’attribute#1 = attribute#2’ , it can not be spotted by calculating distance unless we add a new feature which tells if attribute#1 equals attribute#2.