Association Analysis

Rabnawaz jansher(rabsh696) & Saman Zahid(samza595)

7 March 2018

Describing clustering through association analysis

In this part we have added cluster attribute from filters.Then select the Clustering algorithm which is SimpleKMean with different number of cluster and seed value 10.After that we apply discretization with different number of bins and than run Apriori algorithm by setting a $minimum\ \ support = 0.05$ and $Rules = 150$.

Experiment 1

1. Clustring Algorithm = SimpleKMean
2. Number of Cluster = 3
3. Number of bins = 3

```
 5. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 15 ==> cluster=cluster1 15    <conf:(1)>
14. sepallength='(6.7-inf)' sepalwidth='(2.8-3.6]' 14 ==> cluster=cluster3 14    <conf:(1)> lift:(3.85) lev:(0.07) [10] conv:(10.36)
29. sepalwidth='(3.6-inf)' petallength='(-inf-2.966667]' 13 ==> cluster=cluster2 13    <conf:(1)> lift:(3) lev:(0.06) [8] conv:(8.67)
33. sepalwidth='(3.6-inf)' petalwidth='(-inf-0.9]' 13 ==> cluster=cluster2 13    <conf:(1)> lift:(3) lev:(0.06) [8] conv:(8.67)
77. sepallength='(-inf-5.5]' petallength='(2.966667-4.933333]' 12 ==> cluster=cluster1 12    <conf:(1)> lift:(2.46) lev:(0.05) [7] conv:(7.12)
79. sepallength='(-inf-5.5]' petalwidth='(0.9-1.7]' 12 ==> cluster=cluster1 12    <conf:(1)> lift:(2.46) lev:(0.05) [7] conv:(7.12)
82. sepallength='(-inf-5.5]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 12 ==> cluster=cluster1 12    <conf:(1)> lift:(2.46)
89. sepallength='(6.7-inf)' sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' 12 ==> cluster=cluster3 12    <conf:(1)> lift:(3.85)
104. sepallength='(-inf-5.5]' sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' 11 ==> cluster=cluster1 11    <conf:(1)> lift:(2.46)
106. sepallength='(-inf-5.5]' sepalwidth='(-inf-2.8]' petalwidth='(0.9-1.7]' 11 ==> cluster=cluster1 11    <conf:(1)> lift:(2.46)
118. sepallength='(6.7-inf)' sepalwidth='(2.8-3.6]' petalwidth='(1.7-inf)' 11 ==> cluster=cluster3 11    <conf:(1)> lift:(3.85)
121. sepallength='(-inf-5.5]' sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 11 ==> cluster=cluster1 11    <conf:(1)>
136. sepallength='(6.7-inf)' sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 11 ==> cluster=cluster3 11    <conf:(1)>
```

Above are the Number of rules selected which satisfy the given condition.

Best Rule For Cluster 1 :

```
 5. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 15 ==> cluster=cluster1 15    <conf:(1)>
```

Because it has higer support and confidence 1

Experiment 2

1. Clustring Algorithm = SimpleKMean
2. Number of Cluster = 3
3. Number of bins = 6

```
4. sepallength='(-inf-4.9]' sepalwidth='(2.8-3.2]' 15 ==> cluster=cluster2 15    <conf:(1)> lift:(3) lev:(0.07) [10] conv:(10)
7. sepallength='(4.9-5.5]' sepalwidth='(3.2-3.6]' 15 ==> cluster=cluster2 15    <conf:(1)> lift:(3) lev:(0.07) [10] conv:(10)
9. sepallength='(5.5-6.1]' petalwidth='(0.9-1.3]' 15 ==> cluster=cluster1 15    <conf:(1)> lift:(2.46) lev:(0.06) [8] conv:(8.9)
17. sepallength='(-inf-4.9]' sepalwidth='(2.8-3.2]' petallength='(-inf-1.983333]' 15 ==> cluster=cluster2 15    <conf:(1)> lift:(3) lev:(0.07)
23. sepallength='(-inf-4.9]' sepalwidth='(2.8-3.2]' petalwidth='(-inf-0.5]' 15 ==> cluster=cluster2 15    <conf:(1)> lift:(3) lev:(0.07)
32. sepallength='(4.9-5.5]' sepalwidth='(3.2-3.6]' petallength='(-inf-1.983333]' 15 ==> cluster=cluster2 15    <conf:(1)> lift:(3) lev:(0.07)
49. sepallength='(-inf-4.9]' sepalwidth='(2.8-3.2]' petallength='(-inf-1.983333]' petalwidth='(-inf-0.5]' 15 ==> cluster=cluster2 15    <conf:(1)>
90. sepallength='(6.1-6.7]' petallength='(3.95-4.933333]' 14 ==> cluster=cluster1 14    <conf:(1)> lift:(2.46) lev:(0.06) [8] conv:(8.31)
93. sepallength='(4.9-5.5]' sepalwidth='(3.2-3.6]' petalwidth='(-inf-0.5]' 14 ==> cluster=cluster2 14    <conf:(1)> lift:(3) lev:(0.06) [9]
98. sepallength='(4.9-5.5]' sepalwidth='(3.2-3.6]' petallength='(-inf-1.983333]' petalwidth='(-inf-0.5]' 14 ==> cluster=cluster2 14    <conf:(1)>
110. sepallength='(5.5-6.1]' sepalwidth='(2.8-3.2]' 13 ==> cluster=cluster1 13    <conf:(1)> lift:(2.46) lev:(0.05) [7] conv:(7.71)
112. sepalwidth='(2.4-2.8]' petalwidth='(0.9-1.3]' 13 ==> cluster=cluster1 13    <conf:(1)> lift:(2.46) lev:(0.05) [7] conv:(7.71)
117. sepalwidth='(2.8-3.2]' petalwidth='(2.1-inf)' 12 ==> cluster=cluster3 12    <conf:(1)> lift:(3.85) lev:(0.06) [8] conv:(8.88)
128. petallength='(2.966667-3.95]' 11 ==> cluster=cluster1 11    <conf:(1)> lift:(2.46) lev:(0.04) [6] conv:(6.53)
130. petallength='(5.916667-inf)' 11 ==> cluster=cluster3 11    <conf:(1)> lift:(3.85) lev:(0.05) [8] conv:(8.14)
132. sepallength='(6.1-6.7]' petalwidth='(2.1-inf)' 11 ==> cluster=cluster3 11    <conf:(1)> lift:(3.85) lev:(0.05) [8] conv:(8.14)
133. sepallength='(6.7-7.3]' sepalwidth='(2.8-3.2]' 11 ==> cluster=cluster3 11    <conf:(1)> lift:(3.85) lev:(0.05) [8] conv:(8.14)
142. sepallength='(5.5-6.1]' sepalwidth='(2.8-3.2]' petallength='(3.95-4.933333]' 11 ==> cluster=cluster1 11    <conf:(1)> lift:(2.46)
144. sepallength='(5.5-6.1]' petallength='(3.95-4.933333]' petalwidth='(0.9-1.3]' 11 ==> cluster=cluster1 11    <conf:(1)> lift:(2.46)
```

Best Rule For Cluster 1 :

```
9. sepallength='(5.5-6.1]' petalwidth='(0.9-1.3]' 15 ==> cluster=cluster1 15    <conf:(1)> lift:(2.46) lev:(0.06) [8] conv:(8.9)
```

Because it has higer support and confidence 1

Experiment 3

1. Clustring Algorithm = SimpleKMean
2. Number of Cluster = 6
3. Number of bins = 3

```
19. sepalwidth='(3.6-inf)' petallength='(-inf-2.966667]' 13 ==> cluster=cluster4 13    <conf:(1)> lift:(3) lev:(0.06) [8] conv:(8.67)
23. sepalwidth='(3.6-inf)' petalwidth='(-inf-0.9]' 13 ==> cluster=cluster4 13    <conf:(1)> lift:(3) lev:(0.06) [8] conv:(8.67)
38. sepalwidth='(3.6-inf)' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 13 ==> cluster=cluster4 13    <conf:(1)> lift:(3) lev:(0.06)
79. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petalwidth='(1.7-inf)' 12 ==> cluster=cluster1 12    <conf:(1)> lift:(6.82) lev:(0.07)
109. sepallength='(-inf-5.5]' sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' 11 ==> cluster=cluster2 11    <conf:(1)> lift:(7.89) lev:(0.06)
113. sepallength='(-inf-5.5]' sepalwidth='(-inf-2.8]' petalwidth='(0.9-1.7]' 11 ==> cluster=cluster2 11    <conf:(1)> lift:(7.89) lev:(0.06)
125. sepallength='(6.7-inf)' sepalwidth='(2.8-3.6]' petalwidth='(1.7-inf)' 11 ==> cluster=cluster3 11    <conf:(1)> lift:(6) lev:(0.06)
130. sepallength='(-inf-5.5]' sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 11 ==> cluster=cluster2 11    <conf:(1)>
143. sepallength='(6.7-inf)' sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 11 ==> cluster=cluster3 11    <conf:(1)>
```

Above are the Number of rules selected which satisfy the given condition.

Best Rule For Cluster 1 :

```
79. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petalwidth='(1.7-inf)' 12 ==> cluster=cluster1 12    <conf:(1)> lift:(6.82) lev:(0.07)
```

Because it has higer support and confidence 1

Experiment 4

1. Clustring Algorithm = SimpleKMean
2. Number of Cluster = 6
3. Number of bins = 6

```
4. sepallength='(-inf-4.9]' sepalwidth='(2.8-3.2]' 15 ==> cluster=cluster4 15    <conf:(1)> lift:(3) lev:(0.07) [10] conv:(10)
7. sepallength='(4.9-5.5]' sepalwidth='(3.2-3.6]' 15 ==> cluster=cluster4 15    <conf:(1)> lift:(3) lev:(0.07) [10] conv:(10)
17. sepallength='(-inf-4.9]' sepalwidth='(2.8-3.2]' petallength='(-inf-1.983333]' 15 ==> cluster=cluster4 15    <conf:(1)> lift:(3) lev:(0.07)
23. sepallength='(-inf-4.9]' sepalwidth='(2.8-3.2]' petalwidth='(-inf-0.5]' 15 ==> cluster=cluster4 15    <conf:(1)> lift:(3) lev:(0.07)
32. sepallength='(4.9-5.5]' sepalwidth='(3.2-3.6]' petallength='(-inf-1.983333]' 15 ==> cluster=cluster4 15    <conf:(1)> lift:(3)
46. sepallength='(-inf-4.9]' sepalwidth='(2.8-3.2]' petallength='(-inf-1.983333]' petalwidth='(-inf-0.5]' 15 ==> cluster=cluster4 15    <conf:(1)>
93. sepallength='(4.9-5.5]' sepalwidth='(3.2-3.6]' petalwidth='(-inf-0.5]' 14 ==> cluster=cluster4 14    <conf:(1)> lift:(3) lev:(0.06)
102. sepallength='(4.9-5.5]' sepalwidth='(3.2-3.6]' petallength='(-inf-1.983333]' petalwidth='(-inf-0.5]' 14 ==> cluster=cluster4 14    <conf:(1)>
117. sepalwidth='(2.8-3.2]' petalwidth='(2.1-inf)' 12 ==> cluster=cluster3 12    <conf:(1)> lift:(6) lev:(0.07) [10] conv:(10)
126. petallength='(5.916667-inf)' 11 ==> cluster=cluster3 11    <conf:(1)> lift:(6) lev:(0.06) [9] conv:(9.17)
140. sepalwidth='(3.6-4]' petallength='(-inf-1.983333]' 10 ==> cluster=cluster4 10    <conf:(1)> lift:(3) lev:(0.04) [6] conv:(6.67)
144. sepalwidth='(3.6-4]' petalwidth='(-inf-0.5]' 10 ==> cluster=cluster4 10    <conf:(1)> lift:(3) lev:(0.04) [6] conv:(6.67)
```

Above are the Number of rules selected which satisfy the given condition.

Best Rule For Cluster 1 :

There is no such a rule which satisfy Cluster 1

Conclusion

From these 4 experiments, we can see that in experiment 3, there is one rule which satisfy the cluster 1 but, In experiment 4, when we increase number of clusters as well as number of bins we found no such a rule which satisfy the cluster 1. In order to get the best result, we should select the appropriate number of clusters as well as number of bins.