EH 2745 - Computer Application in Power System

Assignment II - K-Means Clustering & K-NN Classifications

Group 19

Aiman Setiawan

Heraldus Panji Arikson

Data Clustering by Using K-Means Method

K-means is one of machine learning method for data clustering. In this assignment, we need to do data clustering for power system data. As mentioned in description of this assignment, the system is a power system which consists of 9 buses.

Before we do data clustering we import data from mySQL database, in the data base we were given data of voltage magnitude (PU) and voltage angle. For data clustering we use data from 'measurement' table and set it as learning set. The table shows data of 200 states of time of nine buses, every bus has its own magnitude voltage and voltage angle.

Algorithm for the K-Means is descripted as follows:

1. Store voltage and voltage angle data to two-dimensional array (200 rows, 18 column)
2. Set four arrays with size of (1 x 18) which called old cluster center and new cluster center
3. Take different values of voltage and voltage angle of different state of time and set it to old cluster center
4. Create loop which end if distance between new cluster centers and old cluster centers are equal to zero
5. In this loop create loop which go through all 200 different states of times (learning set) and calculate Euclidian distance of the learning set to all four cluster centers
6. Compare the Euclidian distance of learning set to four clusters and store the learning set to the cluster with smallest distance
7. Calculate average value of data in every cluster
8. Set the average value as new cluster center
9. Compare new cluster center and old cluster center, if the results equal to zero, then clustering process is finished.

We did the clustering process the results is presented as below:

Figure 1 Data clustering summary by using K-means method

```
================================================================================
================================================================================
================================CLUSTERING SUMMARY==============================
================================================================================
================================================================================


Data which belong to cluster 1 are 49 data
Data which belong to cluster 2 are 51 data
Data which belong to cluster 3 are 47 data
Data which belong to cluster 4 are 53 data
```

We did further analysis to determine which power system state presented by every cluster, by taking look to each cluster's cluster center.

Table 1 Cluster centers of the data

	Cluster 1		Cluster 2		Cluster 3		Cluster 4	
Bus	Voltage (pu)	Angle	Voltage (pu)	Angle	Voltage (pu)	Angle	Voltage (pu)	Angle
1	1.020408163	0	1.019608	0	1.0212766	0	1.020408	0
2	1.020408163	-21.1792	1.019608	25.0876571	1.0212766	18.93443	1.020408	-21.1792
3	1.020408163	-25.6157	1.019608	19.73782295	1.0212766	20.0956	1.020408	-25.6157
4	0.913869954	-15.526	1.009961	4.30104778	0.99294664	-2.7565	0.91387	-15.526
5	0.867185177	-28.0793	1.013316	7.601314931	0.96499862	-7.99338	0.867185	-28.0793
6	0.973664277	-28.6691	1.034196	16.86785086	1.02710877	17.19634	0.973664	-28.6691
7	0.922468511	-33.5865	1.031126	17.36975866	1.00287604	12.24377	0.922469	-33.5865
8	0.95738403	-27.5402	1.029899	19.1856225	1.01038299	12.89826	0.957384	-27.5402
9	0.826730115	-30.4707	1.002223	7.872306213	0.95439253	-1.77272	0.82673	-30.4707

1. Cluster 1 (High Load Rate During Peak Hours)

   There are three busses supply the load, which are bus 5, bus 7 and bus 9. We know that when there is high load, the corresponding bus voltage will be decreased. Observe the voltage at bus 5, 7 and 9, it is the lowest of all the clusters, as we can see from the table.

2. Cluster 2 (Low Load Rate During Night)
   By also referring to table 1, the opposite of the first cluster, when there is low load (at night), the voltage at corresponding bus will be increased. As we can see from the same table above, cluster 2 has the highest voltage at corresponding bus.

3. Cluster 3 (Cluster Disconnection of a line for maintenance)
   When there is line disconnected, the power from generator will still reach the consumer eventually. Although, It will travel a little bit far, and by using this logic, and by knowing that the longer the distance, the higher the voltage will drop, we calculated all the voltage decrease from generator to each bus load. And the cluster 3 has higher voltage drop.

4. Cluster 4 (Shut Down a Generator for Maintenance)
   Based on Annex 1, three generators on the power system are located at bus 1, 2 and 3. These generators bus also connected to the corresponding bus, 1 to 4, 2 to 8 and 3 to 6. We calculated all the voltage drop from these bus connections. And by our calculation, the bus generator 3 to bus 6 has the unusual voltage drop among others cluster, which is 0,000390966 per unit value. This voltage-drop we set is too small so that we conclude there is no real connection between these bus because the generator goes for maintenance.

Data Classifying by Using KNN Method

After we finish the data clustering, we use the data learning set to do data classification by using KNN method. In this assignment, we have 20 different states of times as test set. Algorithm for the KNN is descripted as follows (in this assignment we set K as 5):

1. Store voltage and voltage angle data to two-dimensional array (20 rows, 18 column)
2. Create loop which go through all test set
3. In this loop calculate Euclidian distance of the test set to all learning set
4. Store the calculation results and the corresponding cluster number to an array
5. Sort the array based on Euclidian distance from the smallest to highest value
6. Check value of the K-smallest distance from array
7. Store the test set to cluster with the most frequent among K values

The classification results is shown by figure 2.

Table 2 Classification results by using KNN method