Assignment Report

Machine Learning (CS60050) Assignment 3

Arijit Kar

31-03-2019
16CS30005

(1) Learned Clusters for each Clustering Method:

(1.a) Bottom Up Hierarchical Clustering - Single Linkage

Final number of Clusters = 9

The Final Clusters are:

Cluster 1

Number of Papers in Cluster = 1

[52]

Cluster 2

Number of Papers in Cluster = 1

[75]

Cluster 3

Number of Papers in Cluster = 1

[77]

Cluster 4

Number of Papers in Cluster = 1

[107]

Cluster 5

Number of Papers in Cluster = 1

[128]

Cluster 6

Number of Papers in Cluster = 45

[0, 1, 2, 4, 8, 10, 17, 18, 19, 20, 24, 28, 29, 31, 37, 39, 49, 50, 55, 67, 68, 70, 74, 76, 79, 81, 86, 87, 95, 96, 106, 112, 116, 117, 121, 123, 126, 131, 132, 133, 135, 136, 137, 138, 146]

Cluster 7

Number of Papers in Cluster = 64

[3, 5, 6, 9, 11, 14, 16, 21, 22, 25, 27, 30, 32, 33, 36, 40, 41, 42, 43, 44, 45, 46, 51, 54, 56, 57, 60, 61, 63, 64, 65, 69, 72, 78, 80, 84, 89, 90, 91, 93, 94, 97, 98, 99, 100, 101, 102, 103, 104, 109, 110, 111, 114, 115, 119, 130, 140, 141, 143, 144, 145, 147, 148, 149]

Cluster 8

Number of Papers in Cluster = 22

[7, 12, 13, 15, 23, 26, 35, 47, 48, 53, 59, 62, 66, 73, 83, 85, 88, 92, 113, 118, 127, 134]

Cluster 9

Number of Papers in Cluster = 14

[34, 38, 58, 71, 82, 105, 108, 120, 122, 124, 125, 129, 139, 142]

(1.b) Bottom Up Hierarchical Clustering - Complete Linkage

Final number of Clusters = 9

The Final Clusters are:

Cluster 1

Number of Papers in Cluster = 11

[46, 57, 60, 62, 78, 84, 90, 92, 104, 130, 145]

Cluster 2

Number of Papers in Cluster = 11

[6, 7, 11, 15, 32, 35, 59, 94, 97, 101, 143]

Cluster 3

Number of Papers in Cluster = 14

[24, 25, 37, 39, 41, 50, 70, 89, 91, 100, 111, 116, 121, 144]

Cluster 4

Number of Papers in Cluster = 12

[0, 4, 9, 10, 18, 27, 44, 47, 53, 76, 83, 140]

Cluster 5

Number of Papers in Cluster = 16

[12, 23, 34, 61, 63, 64, 65, 85, 107, 115, 118, 119, 139, 141, 147, 148]

Cluster 6

Number of Papers in Cluster = 18

[1, 8, 19, 22, 36, 45, 52, 69, 72, 79, 106, 109, 110, 117, 128, 133, 137, 138]

Cluster 7

Number of Papers in Cluster = 16

[5, 28, 33, 56, 66, 73, 74, 75, 77, 81, 88, 93, 112, 114, 131, 149]

Cluster 8

Number of Papers in Cluster = 23

[3, 13, 14, 26, 30, 38, 40, 42, 48, 54, 58, 71, 82, 99, 103, 105, 113, 120, 122, 127, 129, 134, 142]

Cluster 9

Number of Papers in Cluster = 29

[2, 16, 17, 20, 21, 29, 31, 43, 49, 51, 55, 67, 68, 80, 86, 87, 95, 96, 98, 102, 108, 123, 124, 125, 126, 132, 135, 136, 146]

(1.c) Girvan Newman Clustering

Final number of Clusters = 9

The Final Clusters are:

Cluster 1

Number of Papers in Cluster = 40

[0, 68, 2, 67, 4, 133, 70, 135, 136, 137, 138, 76, 79, 17, 18, 19, 20, 86, 87, 132, 49, 27, 29, 31, 96, 123, 121, 39, 106, 146, 8, 50, 52, 117, 55, 116, 95, 124, 125, 126]

Cluster 2

Number of Papers in Cluster = 39

[128, 1, 69, 100, 65, 72, 9, 12, 14, 143, 144, 84, 21, 22, 89, 110, 25, 91, 94, 32, 16, 98, 36, 102, 80, 40, 41, 42, 103, 44, 45, 46, 111, 104, 51, 54, 60, 109, 63]

Cluster 3

Number of Papers in Cluster = 22

[129, 66, 3, 134, 71, 73, 74, 13, 81, 131, 88, 28, 30, 48, 99, 43, 47, 112, 113, 83, 53, 127]

Cluster 4

Number of Papers in Cluster = 19

[64, 33, 130, 11, 5, 75, 141, 61, 115, 145, 114, 147, 148, 149, 119, 56, 57, 90, 93]

Cluster 5

Number of Papers in Cluster = 7

[97, 101, 6, 24, 10, 140, 37]

Cluster 6

Number of Papers in Cluster = 11

[59, 35, 85, 118, 7, 26, 23, 92, 77, 62, 15]

Cluster 7

Number of Papers in Cluster = 1

[34]

Cluster 8

Number of Papers in Cluster = 10

[82, 139, 38, 120, 105, 122, 107, 108, 58, 142]

Cluster 9

Number of Papers in Cluster = 1

[78]

(2) NMI Values for each Clustering:

Bottom Up Hierarchical Clustering - Single Linkage

NMI Value = 0.5066227644244067

Bottom Up Hierarchical Clustering - Complete Linkage

NMI Value = 0.38725253245576413

Girvan Newman Clustering

NMI Value = 0.6274579407960715

(3) Threshold Value:

I have used the threshold value of 0.15 in the Girvan Newman Clustering implementation. I tested the quality of the clusters for threshold values ranging from 0.05 to 0.3 and found the best clusters for 0.15.