Contemporary Data Processing Technology (CCOD)

Laboratory Work(4.11.2016)

Vitaly Konovalov(AS-37)

Clustering of Japanese Characters

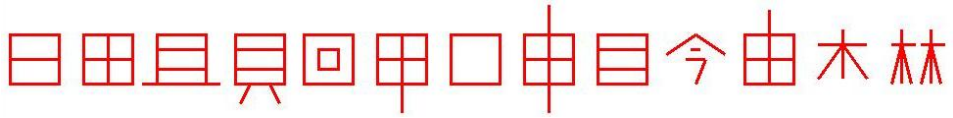

Table of similarity percentage of letters:

		1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	1	0.7	0.3	0.2	0.5	0.4	0.5	0.4	0.4	0.2	0.3	0.1	0.1
2	2	0.7	1	0.4	0.3	0.5	0.4	0.5	0.4	0.4	0.2	0.1	0.1	0.1
3	3	0.3	0.4	1	0.6	0.3	0.4	0.3	0.3	0.8	0.1	0.1	0.1	0.1
4	4	0.2	0.3	0.6	1	0.4	0.3	0.1	0.1	0.7	0.1	0.2	0.1	0.1
5	5	0.5	0.5	0.3	0.4	1	0.4	0.6	0.4	0.4	0.1	0.1	0.1	0.1
6	6	0.4	0.4	0.4	0.3	0.4	1	0.4	0.7	0.4	0.1	0.9	0.2	0.1
7	7	0.5	0.5	0.3	0.1	0.6	0.4	1	0.4	0.4	0.2	0.1	0.2	0.2
8	8	0.4	0.4	0.3	0.1	0.4	0.7	0.4	1	0.4	0.1	0.7	0.2	0.2
9	9	0.4	0.4	0.8	0.7	0.5	0.4	0.4	0.4	1	0.1	0.4	0.1	0.1
10	10	0.2	0.2	0.1	0.1	0.1	0.1	0.2	0.1	0.1	1	0.1	0.4	0.2
11	11	0.3	0.1	0.1	0.2	0.1	0.9	0.1	0.7	0.4	0.1	1	0.1	0.1
12	12	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.1	0.4	0.1	1	0.5
13	13	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.1	0.2	0.1	0.5	1

In the final table, where all values less than α and main diagonal will become zeros.

Final table:

		1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	0	0.7	0	0	0	0	0	0	0	0	0	0	0
2	2	0.7	0	0	0	0	0	0	0	0	0	0	0	0
3	3	0	0	0	0.7	0	0	0	0	0.8	0	0	0	0
4	4	0	0	0.7	0	0	0	0	0	0.7	0	0	0	0
5	5	0	0	0	0	0	0	0.6	0	0	0	0	0	0
6	6	0	0	0	0	0	0	0	0.7	0	0	0.9	0	0
7	7	0	0	0	0	0.6	0	0	0	0	0	0	0	0
8	8	0	0	0	0	0	0.7	0	0	0	0	0.7	0	0
9	9	0	0	0.8	0.7	0	0	0	0	0	0	0	0	0
10	10	0	0	0	0	0	0	0	0	0	0	0	0	0
11	11	0	0	0	0	0	0.9	0	0.7	0	0	0	0	0
12	12	0	0	0	0	0	0	0	0	0	0	0	0	0
13	13	0	0	0	0	0	0	0	0	0	0	0	0	0

First iteration

First, set I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13} and C_1 = { }.

$a_{6\,11}$ = 0.9 are maximum. Then C_1= {6, 11}.

$a_{6\,8} + a_{11\,8}$ = 1.4 are maximum. Then C_1= {6, 11, 8}.

There are no j such that $a_{6\,j} + a_{11\,j} + a_{8\,j}$ is maximum. Then final C_1= {6, 11, 8} = { , , }.

After deleting **(6, 11, 8)** rows and columns the table has become a:

		1	2	3	4	5	7	9	10	12	13
日	1	0	0.7	0	0	0	0	0	0	0	0
田	2	0.7	0	0	0	0	0	0	0	0	0
且	3	0	0	0	0.7	0	0	0.8	0	0	0
貝	4	0	0	0.7	0	0	0	0.7	0	0	0
回	5	0	0	0	0	0	0.6	0	0	0	0
口	7	0	0	0	0	0.6	0	0	0	0	0
目	9	0	0	0.8	0.7	0	0	0	0	0	0
今	10	0	0	0	0	0	0	0	0	0	0
木	12	0	0	0	0	0	0	0	0	0	0
林	13	0	0	0	0	0	0	0	0	0	0

Second iteration

I = {1, 2, 3, 4, 5, 7, 9, 10, 12, 13}, C_2 = { }.

$a_{3\,9}$ = 0.8 are maximum. Then C_2= {3, 9}.

$a_{3\,4} + a_{9\,4}$ = 1.4 are maximum. Then C_2= {3, 9, 4}.

There are no j such that $a_{3\,j} + a_{9\,j} + a_{4\,j}$ is maximum. Then final C_2= {3, 9, 4} = { 且, 目, 貝 }.

After deleting **(3, 9, 4)** rows and columns the table has become a:

		1	2	5	7	10	12	13
1	1	0	0.7	0	0	0	0	0
2	2	0.7	0	0	0	0	0	0
5	5	0	0	0	0.6	0	0	0
7	7	0	0	0.6	0	0	0	0
10	10	0	0	0	0	0	0	0
12	12	0	0	0	0	0	0	0
13	13	0	0	0	0	0	0	0

The third iteration

I = {1, 2, 5, 7, 10, 12, 13}, C_3 = { }.

$a_{1\,2}$ = 0.7 are maximum. Then C_3= {1, 2}.

There are no j such that $a_{1\,j} + a_{2\,j}$ is maximum. Then final C_3= {1, 2} = { 日, 田 }.

After deleting **(1, 2)** rows and columns the table has become a:

		5	7	10	12	13
5	5	0	0.6	0	0	0
7	7	0.6	0	0	0	0
10	10	0	0	0	0	0
12	12	0	0	0	0	0
13	13	0	0	0	0	0

Fourth iteration

I = {5, 7, 10, 12, 13}, C_4 = { }.

$a_{5\,7}$ = 0.6 are maximum. Then C_4= {5, 7}.

There are no j such that $a_{5\,j} + a_{7\,j}$ is maximum. Then final C_4= {5, 7} = { , }.

After deleting **(5, 7)** rows and columns the table has become a:

		10	12	13
	10	0	0	0
	12	0	0	0
	13	0	0	0

Now $a_{10\,12} = a_{10\,13} = a_{12\,13}$ = 0. Then {10}, {12}, {13} are three separated clusters.

C_5 = {10} = { }.

C_6 = {12} = { }.

C_7 = {13} = { }.

In this way, when α = 0.55, we have **7 clusters:**

{ , , }

{ , , }

{  ,  }

{  ,  }

{  }

{  }

{  }