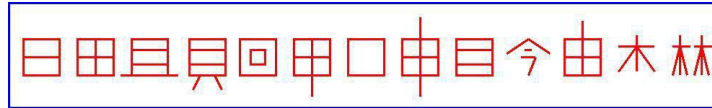


CCOD

Professor: Akira Imada

Student: Nikolay Siliuk (AS-37) 04.11.2017

Clustering of 13 Japanese characters



The table below shows the similarity of characters:

	日	田	且	貝	回	甲	口	申	目	今	由	木	林
日	1	0,9	0,5	0,45	0,5	0,75	0,9	0,65	0,6	0,1	0,7	0,1	0,1
田	0,9	1	0,8	0,3	0,45	0,9	0,7	0,85	0,5	0,1	0,9	0,1	0,1
且	0,5	0,8	1	0,5	0,35	0,3	0,4	0,6	0,85	0,1	0,7	0,1	0,1
貝	0,45	0,3	0,5	1	0,2	0,4	0,45	0,55	0,8	0,1	0,6	0,1	0,1
回	0,5	0,45	0,35	0,2	1	0,5	0,75	0,6	0,7	0,1	0,65	0,1	0,1
甲	0,75	0,9	0,3	0,4	0,5	1	0,55	0,9	0,45	0,1	0,9	0,1	0,1
口	0,9	0,7	0,4	0,45	0,75	0,55	1	0,5	0,6	0,1	0,55	0,1	0,1
申	0,65	0,85	0,6	0,55	0,6	0,9	0,5	1	0,4	0,1	0,9	0,1	0,1
目	0,6	0,5	0,85	0,8	0,7	0,45	0,6	0,4	1	0,1	0,7	0,1	0,1
今	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	1	0,1	0,3	0,2
由	0,7	0,9	0,7	0,6	0,65	0,9	0,55	0,9	0,7	0,1	1	0,1	0,1
木	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,3	0,1	1	0,7
林	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,7	1

First Iteration:

First, set $I = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}$ and $C1 = \{\}$.

$C1 = \{1, 2\}$ (selected at random from the maximum = 0,9).

1	2	3	4	5	6	7	8	9	10	11	12	13
-	-	1,6	1,6	1,5	1,8	1,8	1,8	1,6	-	1,8	-	-

Max sum = 1,8. Than select at random 6. Now $C1 = \{1, 2, 6\}$.

1	2	3	4	5	6	7	8	9	10	11	12	13
-	-	2,4	2,4	2,25	-	2,7	2,7	2,4	-	2,7	-	-

Max sum = 2,7. Than select at random 7. Now $C1 = \{1, 2, 6, 7\}$.

1	2	3	4	5	6	7	8	9	10	11	12	13
-	-	3,2	3,2	3	-	-	3,6	3,2	-	3,6	-	-

Max sum = 3,6. Than select at random 8. Now $C1 = \{1, 2, 6, 7, 8\}$.

1	2	3	4	5	6	7	8	9	10	11	12	13
-	-	4	4	3,75	-	-	-	4	-	4,5	-	-

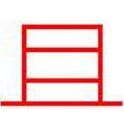
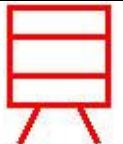
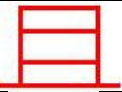
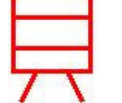
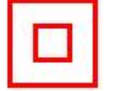
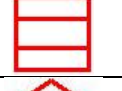
Max sum = 4,5. Than select 11. Now $C1 = \{1, 2, 6, 7, 8, 11\}$.

There is no j such that $a_{1j} + a_{2j} + a_{6j} + a_{7j} + a_{8j} + a_{11j}$ is maximum. The final $C1 = \{1, 2, 6, 7, 8, 11\}$.

After we delete 1, 2, 6, 7, 8, 11 rows and columns from the final table and go to the second iteration.

Second Iteration:

New table after deleting rows and columns:

		3	4	5	9	10	12	13
								
3		0	0,8	0,75	0,85	0	0	0
4		0,8	0	0,75	0,8	0	0	0
5		0,75	0,75	0	0,75	0	0	0
9		0,85	0,8	0,75	0	0	0	0
10		0	0	0	0	0	0	0
12		0	0	0	0	0	0	0,7
13		0	0	0	0	0	0,7	0

Set I = {3, 4, 5, 9, 10, 12, 13}, C2 = { }.
C2 = {3, 9}.

3	4	5	9	10	12	13
-	1,6	1,5	-	-	-	-

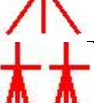
Max sum = 1,6. Than select 4. Now C2 = {3, 4, 9}.

There is no j such that $a_{3j} + a_{4j} + a_{9j}$ is maximum. The final C2 = {3, 4, 9}.

After we delete 3, 4, 9 rows and columns from the previous table and go to the third iteration.

Third Iteration:

New table after deleting rows and columns:

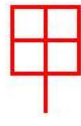
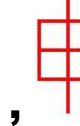
		5	10	12	13
					
5		0	0	0	0
10		0	0	0	0
12		0	0	0	0,7
13		0	0	0,7	0

Set I = {5, 10, 12, 13}, C3 = { }.
C3 = {12, 13}.

5	10	12	13
-	-	-	-

Now C3 = {12, 13}.
The final C3 = {12, 13}.

As a result, we have 5 clusters: C = {{1, 2, 6, 7, 8, 11}, {5}, {3, 4, 9}, {10}, {12, 13}}.
Or:

      }, {  }, {    }, {  }, {   }.