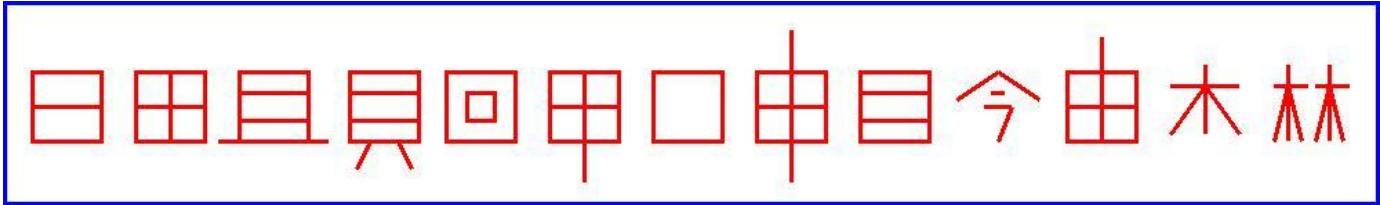


CCOD

Professor: Akira Imada

Student: Aleksandr Lishko (AS-37) 04.11.2017

Clustering of 13 Japanese characters



Here is a table of similarity percentage of letters:

	日	田	且	貝	回	甲	口	申	目	今	由	木	林
日	1	0,9	0,8	0,75	0,8	0,85	0,95	0,85	0,85	0,1	0,85	0,1	0,1
田	0,9	1	0,6	0,55	0,7	0,95	0,85	0,9	0,75	0,1	0,95	0,1	0,1
且	0,8	0,6	1	0,9	0,5	0,6	0,5	0,6	0,95	0,1	0,6	0,1	0,1
貝	0,75	0,55	0,9	1	0,4	0,5	0,5	0,5	0,95	0,1	0,5	0,2	0,15
回	0,8	0,7	0,5	0,4	1	0,5	0,9	0,75	0,7	0,1	0,7	0,1	0,1
甲	0,85	0,95	0,6	0,5	0,5	1	0,7	0,95	0,75	0,1	0,95	0,1	0,1
口	0,95	0,85	0,5	0,5	0,9	0,7	1	0,75	0,9	0,1	0,75	0,1	0,1
申	0,85	0,9	0,6	0,5	0,75	0,95	0,75	1	0,75	0,1	0,95	0,1	0,1
目	0,85	0,75	0,95	0,95	0,7	0,75	0,9	0,75	1	0,1	0,8	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	0,6	0,6
由	0,85	0,95	0,6	0,5	0,7	0,95	0,75	0,95	0,8	0,1	1	0,1	0,1
木	0,1	0,1	0,1	0,2	0,1	0,1	0,1	0,1	0,1	0,6	0,1	1	0,9
林	0,1	0,1	0,1	0,15	0,1	0,1	0,1	0,1	0,1	0,6	0,1	0,9	1

In the final table, where all values less than $\alpha=0,55$ and main diagonal will become zeros.

Final table:

														
		1	2	3	4	5	6	7	8	9	10	11	12	13
	1	0	0,9	0,9	0,9	0,9	0,9	0,95	0,9	0,9	0	0,9	0	0
	2	0,9	0	0,9	0,9	0,9	0,95	0,9	0,95	0,9	0	0,95	0	0
	3	0,9	0,9	0	0,95	0,9	0,9	0,9	0,9	0,95	0	0,9	0	0
	4	0,9	0,9	0,95	0	0,9	0,9	0,9	0,9	0,95	0	0,9	0	0
	5	0,9	0,9	0,9	0,9	0	0,9	0,9	0,9	0,9	0	0,9	0	0
	6	0,9	0,95	0,9	0,9	0,9	0	0,9	0,95	0,9	0	0,95	0	0
	7	0,95	0,9	0,9	0,9	0,9	0,9	0	0,9	0,9	0	0,9	0	0
	8	0,9	0,95	0,9	0,9	0,9	0,95	0,9	0	0,9	0	0,95	0	0
	9	0,9	0,9	0,95	0,95	0,9	0,9	0,9	0,9	0	0	0,9	0	0
	10	0	0	0	0	0	0	0	0	0	0	0	0,6	0,6
	11	0,9	0,95	0,9	0,9	0,9	0,95	0,9	0,95	0,9	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0,6	0	0	0,9
	13	0	0	0	0	0	0	0	0	0	0,6	0	0,9	0

First Iteration:

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

$C_1 = \{1, 7\}$ (selected at random from the maximum = 0,95).

Table of sum lines 1 & 7:

	2	3	4	5	6	8	9	10	11	12	13
1	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0	0,9	0	0
7	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0	0,9	0	0
sum:	1,8	1,8	1,8	1,8	1,8	1,8	1,8	0	1,8	0	0

Max sum = 1,8. Than select at random 2. Now $C_1 = \{1, 7, 2\}$.

Table of sum lines 1 & 7& 2:

	3	4	5	6	8	9	10	11	12	13
1	0,9	0,9	0,9	0,9	0,9	0,9	0	0,9	0	0
7	0,9	0,9	0,9	0,9	0,9	0,9	0	0,9	0	0
2	0,9	0,9	0,9	0,95	0,95	0,9	0	0,95	0	0
sum:	2,7	2,7	2,7	2,75	2,75	2,7	0	2,75	0	0

Max sum = 2,75. Than select at random 6. Now $C_1 = \{1, 7, 2, 6\}$.

Table of sum lines 1 & 7& 2& 6:

	3	4	5	8	9	10	11	12	13
1	0,9	0,9	0,9	0,9	0,9	0	0,9	0	0
7	0,9	0,9	0,9	0,9	0,9	0	0,9	0	0
2	0,9	0,9	0,9	0,95	0,9	0	0,95	0	0
6	0,9	0,9	0,9	0,95	0,9	0	0,95	0	0
sum:	3,6	3,6	3,6	3,7	3,6	0	3,7	0	0

Max sum = 3,7. Than select at random 8. Now $C_1 = \{1, 7, 2, 6, 8\}$.

Table of sum lines 1 & 7& 2& 6 & 8:

	3	4	5	9	10	11	12	13
1	0,9	0,9	0,9	0,9	0	0,9	0	0
7	0,9	0,9	0,9	0,9	0	0,9	0	0
2	0,9	0,9	0,9	0,9	0	0,95	0	0
6	0,9	0,9	0,9	0,9	0	0,95	0	0
8	0,9	0,9	0,9	0,9	0	0,95	0	0
sum:	4,5	4,5	4,5	4,5	0	4,65	0	0

Max sum = 4,65. Than select 11. Now $C_1 = \{1, 7, 2, 6, 8, 11\}$.

Table of sum lines 1 & 7& 2& 6 & 8 & 11:

	3	4	5	9	10	12	13
1	0,9	0,9	0,9	0,9	0	0	0
7	0,9	0,9	0,9	0,9	0	0	0
2	0,9	0,9	0,9	0,9	0	0	0
6	0,9	0,9	0,9	0,9	0	0	0
8	0,9	0,9	0,9	0,9	0	0	0
11	0,9	0,9	0,9	0,9	0	0	0
sum:	5,4	5,4	5,4	5,4	0	0	0

Max sum = 5,4. Than select at random 3. Now $C_1 = \{1, 7, 2, 6, 8, 11, 3\}$.

Table of sum lines 1 & 7& 2& 6 & 8 & 11 & 3:

	4	5	9	10	12	13
1	0,9	0,9	0,9	0	0	0
7	0,9	0,9	0,9	0	0	0
2	0,9	0,9	0,9	0	0	0
6	0,9	0,9	0,9	0	0	0
8	0,9	0,9	0,9	0	0	0
11	0,9	0,9	0,9	0	0	0
3	0,95	0,9	0,95	0	0	0
sum:	6,35	6,3	6,35	0	0	0

Max sum = 6,35. Than select at random 4. Now $C_1 = \{1, 7, 2, 6, 8, 11, 3, 4\}$.

Table of sum lines 1 & 7& 2& 6 & 8 & 11 & 3 & 4:

	5	9	10	12	13
1	0,9	0,9	0	0	0
7	0,9	0,9	0	0	0
2	0,9	0,9	0	0	0
6	0,9	0,9	0	0	0
8	0,9	0,9	0	0	0
11	0,9	0,9	0	0	0
3	0,9	0,95	0	0	0
4	0,9	0,95	0	0	0
sum:	7,2	7,3	0	0	0

Max sum = 7,3. Than select 9. Now $C_1 = \{1, 7, 2, 6, 8, 11, 3, 4, 9\}$.

Table of sum lines 1 & 7& 2& 6 & 8 & 11 & 3 & 4:

	5	10	12	13
1	0,9	0	0	0
7	0,9	0	0	0
2	0,9	0	0	0
6	0,9	0	0	0
8	0,9	0	0	0
11	0,9	0	0	0
3	0,9	0	0	0
4	0,9	0	0	0
9	0,9	0	0	0
sum:	8,1	0	0	0

Max sum = 8,1. Than select 5. Now $C_1 = \{1, 7, 2, 6, 8, 11, 3, 4, 9, 5\}$.

The final $C_1 = \{1, 7, 2, 6, 8, 11, 3, 4, 9, 5\}$.

After we deleting 1, 7, 2, 6, 8, 11, 3, 4, 9, 5 rows and columns from the **“Final table”** and go to the second iteration.

Second Iteration:

New table after deleting rows and columns:

		𠂇	木	林
		10	12	13
𠂇	10	0	0,6	0,6
木	12	0,6	0	0,9
林	13	0,6	0,9	0

Set I = {10, 12, 13}, C₂ = { }.

C₂ = {12, 13}.

Table of sum lines 12&13:

	10
12	0,6
13	0,6
sum:	1,2

Max sum = 1,2. Than select 10. Now C₂ = {12, 13, 10}.

Result:

At the end of a task we have **2 clusters**: C₁ = {1, 2, 3, 4, 5, 6, 7, 8, 9, 11}, C₂ = {10, 12, 3}.

Or:

{𠂇, 木, 林}, {𠂇, 木, 林}.