

CCOD
Harhun Tsimafei (AS-37)
11.11.2016

Clusterisation of 13 Japanese characters (kanji)

Process OF CLUSTERISATION:

Initial data:

1	2	3	4	5	6	7	8	9	10	11	12	13
日	田	且	貝	回	甲	口	申	目	今	由	木	林

RESULT OF CLUSTERISATION: { {1, 2, 3, 4, 6, 7, 8, 9, 11}, {12, 13}, {5}, {10}}

OR

RESULT OF CLUSTERISATION JAPANESE CHARACTERS:

日	田	且	貝	甲	口	申	目	由
林	木							
今								
回								

Process OF CLUSTERISATION:

Here is a table of similarity percentage of characters:

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

1 ITERATION

Final table, where all values less than **and main diagonal** will become zeros.

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

I = {2, 6}

Sum Table lines 2 and 6:

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

Max sum = 8 => {2, 6, 8}

Sum Table lines 2, 6, 8:

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

Max sum = 11 => {2, 6, 8, 11}

Sum Table lines 2, 6, 8, 11:

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

Max sum = 1 => {1, 2, 6, 8, 11}

Sum Table lines 1, 2, 6, 8, 11:

3	4	5	7	9	10	12	13
3,5	3,5	0	3,5	3,5	0	0	0

Max sum = 3 => {1, 2, 3, 6, 8, 11}

Sum Table lines 1, 2, 3, 6, 8, 11:

4	5	7	9	10	12	13
4,3	0	4,2	4,3	0	0	0

Max sum = 4 => {1, 2, 3, 4, 6, 8, 11}

Sum Table lines 1, 2, 3, 4, 6, 8, 11:

5	7	9	10	12	13
0	4,9	5,1	0	0	0

Max sum = 9 => {1, 2, 3, 4, 6, 8, 9, 11}

Sum Table lines 1, 2, 3, 4, 6, 8, 9, 11:

5	7	10	12	13
0	5,6	0	0	0

Max sum = 7 => {1, 2, 3, 4, 6, 7, 8, 9, 11}

Sum Table lines 1, 2, 3, 4, 6, 7, 8, 9, 11:

5	10	12	13
0	0	0	0

2 ITERATION (Remove 1, 2, 3, 4, 6, 7, 8, 9, 11 lines and columns)

	5	10	12	13
5	0	0	0	0
10	0	0	0	0
12	0	0	0	0,8
13	0	0	0,8	0

$I = \{12, 13\}$

Sum Table lines 2 and 6:

5	10
0	0

$I = \{12, 13\}$

3 ITERATION (Remove 12, 13 lines and columns)

	5	10
5	0	0
10	0	0

Matrix contain only two letters, we can decide that this letters belong to different classes.

$I = \{5\}$

$I = \{10\}$

RESULT OF CLUSTERISATION: { {1, 2, 3, 4, 6, 7, 8, 9, 11}, {12, 13}, {5}, {10}}

OR

RESULT OF CLUSTERISATION JAPANESE CHARACTERS:

								
								
								
								