Contemporary Data Processing Technology (CCOD)

Lab 10(29.10.2016)

Vitaly Konovalov(AS-37)

In this work I took the english alphabet: { A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z }

Table of similarity percentage of letters:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	1	0.1	0.1	0.2	0.1	0.1	0.1	0.2	0.2	0.1	0.3	0.2	0.6	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.8	0.7	0.2	0.2	0.2
B	0.1	1	0.2	0.7	0.2	0.2	0.1	0.3	0.2	0.2	0.6	0.2	0.2	0.3	0.1	0.7	0.1	0.7	0.2	0.1	0.1	0.1	0.1	0.1	0.2	0.2
C	0.1	0.2	1	0.2	0.2	0.1	0.7	0.2	0.2	0.1	0.1	0.1	0.1	0.2	0.8	0.2	0.7	0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.1	0.2
D	0.2	0.7	0.2	1	0.1	0.2	0.2	0.2	0.1	0.2	0.5	0.2	0.2	0.1	0.1	0.7	0.2	0.7	0.2	0.2	0.2	0.1	0.2	0.1	0.1	0.1
E	0.1	0.2	0.2	0.1	1	0.8	0.2	0.7	0.2	0.1	0.2	0.3	0.2	0.2	0.1	0.2	0.1	0.3	0.1	0.2	0.2	0.1	0.2	0.2	0.2	0.1
F	0.1	0.2	0.1	0.2	0.8	1	0.1	0.8	0.2	0.2	0.2	0.3	0.3	0.2	0.1	0.2	0.1	0.3	0.1	0.2	0.1	0.2	0.2	0.1	0.1	0.1
G	0.1	0.1	0.7	0.2	0.2	0.1	1	0.2	0.1	0.3	0.2	0.2	0.2	0.1	0.7	0.2	0.7	0.1	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1
H	0.2	0.3	0.2	0.2	0.7	0.8	0.2	1	0.2	0.1	0.2	0.2	0.3	0.3	0.1	0.2	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.2	0.1	0.3
I	0.2	0.2	0.2	0.1	0.2	0.2	0.1	0.2	1	0.8	0.2	0.8	0.2	0.2	0.1	0.2	0.1	0.2	0.1	0.8	0.2	0.2	0.3	0.2	0.3	0.2
J	0.1	0.2	0.1	0.2	0.1	0.2	0.3	0.1	0.8	1	0.2	0.8	0.3	0.3	0.1	0.3	0.1	0.2	0.1	0.7	0.2	0.1	0.1	0.1	0.2	0.1
K	0.3	0.6	0.1	0.5	0.2	0.2	0.2	0.2	0.2	0.2	1	0.3	0.3	0.3	0.1	0.6	0.1	0.7	0.2	0.2	0.2	0.1	0.1	0.2	0.2	0.1
L	0.2	0.2	0.1	0.2	0.3	0.3	0.2	0.2	0.8	0.8	0.3	1	0.2	0.2	0.2	0.2	0.1	0.2	0.1	0.8	0.1	0.2	0.1	0.2	0.2	0.1
M	0.6	0.2	0.1	0.2	0.2	0.3	0.2	0.3	0.2	0.3	0.3	0.2	1	0.3	0.1	0.2	0.1	0.2	0.2	0.3	0.2	0.7	0.8	0.2	0.2	0.3
N	0.2	0.3	0.2	0.1	0.2	0.2	0.1	0.3	0.2	0.3	0.3	0.2	0.3	1	0.1	0.3	0.1	0.3	0.2	0.3	0.1	0.2	0.1	0.1	0.1	0.8
O	0.1	0.1	0.8	0.1	0.1	0.1	0.7	0.1	0.1	0.1	0.1	0.2	0.1	0.1	1	0.3	0.9	0.3	0.2	0.2	0.1	0.2	0.1	0.1	0.1	0.1
P	0.2	0.7	0.2	0.7	0.2	0.2	0.2	0.2	0.2	0.3	0.6	0.2	0.2	0.3	0.3	1	0.2	0.8	0.2	0.1	0.3	0.2	0.1	0.1	0.1	0.1
Q	0.1	0.1	0.7	0.2	0.1	0.1	0.7	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.9	0.2	1	0.1	0.2	0.1	0.1	0.2	0.1	0.2	0.1	0.2
R	0.2	0.7	0.1	0.7	0.3	0.3	0.1	0.1	0.2	0.2	0.7	0.2	0.2	0.3	0.3	0.8	0.1	1	0.2	0.3	0.1	0.2	0.1	0.2	0.2	0.1
S	0.1	0.2	0.2	0.2	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.1	0.2	0.2	0.2	0.2	0.2	0.2	1	0.1	0.2	0.1	0.2	0.1	0.1	0.2
T	0.2	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.8	0.7	0.2	0.8	0.3	0.3	0.2	0.1	0.1	0.3	0.1	1	0.2	0.2	0.2	0.1	0.2	0.2
U	0.1	0.1	0.2	0.2	0.2	0.1	0.1	0.1	0.2	0.2	0.2	0.1	0.2	0.1	0.1	0.3	0.1	0.1	0.2	0.2	1	0.2	0.2	0.1	0.2	0.1
V	0.8	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.2	0.1	0.1	0.2	0.7	0.2	0.2	0.2	0.2	0.2	0.1	0.2	0.2	1	0.7	0.3	0.2	0.1
W	0.7	0.1	0.2	0.2	0.2	0.2	0.1	0.1	0.3	0.1	0.1	0.1	0.8	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.7	1	0.2	0.2	0.1
X	0.2	0.1	0.1	0.1	0.2	0.1	0.1	0.2	0.2	0.1	0.2	0.2	0.2	0.1	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.3	0.2	1	0.8	0.1
Y	0.2	0.2	0.1	0.1	0.2	0.1	0.1	0.1	0.3	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.2	0.1	0.2	0.2	0.2	0.2	0.8	1	0.1
Z	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.3	0.2	0.1	0.1	0.1	0.3	0.8	0.1	0.1	0.2	0.1	0.2	0.2	0.1	0.1	0.1	0.1	0.1	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	14	15	16	17	18	19	20	21	22	23	24	25	26
1	0	0	0	0	0	0	0	0	0	0	0	0	0.7	0	0	0	0	0	0	0	0	0.8	0.7	0	0	0
2	0	0	0	0.7	0	0	0	0	0	0	0.7	0	0	0	0	0.7	0	0.7	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0.7	0	0	0	0	0	0	0	0.8	0	0.8	0	0	0	0	0	0	0	0	0
4	0	0.7	0	0	0	0	0	0	0	0	0.7	0	0	0	0	0.7	0	0.7	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0.8	0	0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0.8	0	0	0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0.7	0	0	0	0	0	0	0	0	0	0	0	0.7	0	0.7	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0.8	0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0.8	0	0.8	0	0	0	0	0	0	0	0.8	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0.8	0	0	0.8	0	0	0	0	0	0	0	0.8	0	0	0	0	0	0
11	0	0.7	0	0.7	0	0	0	0	0	0	0	0	0	0	0	0.7	0	0.7	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0.8	0.8	0	0	0	0	0	0	0	0	0	0.8	0	0	0	0	0	0
13	0.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.7	0.8	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8
15	0	0	0.8	0	0	0	0.7	0	0	0	0	0	0	0	0	0	0.9	0	0	0	0	0	0	0	0	0
16	0	0.7	0	0.7	0	0	0	0	0	0	0.7	0	0	0	0	0	0	0.8	0	0	0	0	0	0	0	0
17	0	0	0.8	0	0	0	0.7	0	0	0	0	0	0	0	0.9	0	0	0	0	0	0	0	0	0	0	0
18	0	0.7	0	0.7	0	0	0	0	0	0	0.7	0	0	0	0	0.8	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0.8	0.8	0	0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0.8	0	0	0	0	0	0	0	0	0	0	0	0.7	0	0	0	0	0	0	0	0	0	0.7	0	0	0
23	0.7	0	0	0	0	0	0	0	0	0	0	0	0.8	0	0	0	0	0	0	0	0	0.7	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8	0	0	0	0	0	0	0	0	0	0	0	0

First, set I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26} and C_1 = { }.

First iteration

$a_{15\,17} = a_{17\,15} = 0.9$ are maximum and $a_{15\,17}$ is randomly selected. Then C_1= {15, 17}.

$a_{15\,3} + a_{17\,3} = 1.6$ are maximum. Then C_1= {15, 17, 3}

$a_{3\,7} + a_{15\,7} + a_{17\,7} = 2.1$ are maximum. Then C_1= {15, 17, 3, 7}.

There are no j such that $a_{15\,j} + a_{17\,j} + a_{3\,j} + a_{7\,j}$ is maximum. Then final C_1= {15, 17, 3, 7}.

After deleting rows and columns the table has become a:

	1	2	4	5	6	8	9	10	11	12	13	14	16	18	19	20	21	22	23	24	25	26
1	0	0	0	0	0	0	0	0	0	0	0.7	0	0	0	0	0	0	0.8	0.7	0	0	0
2	0	0	0.7	0	0	0	0	0	0.7	0	0	0	0.7	0.7	0	0	0	0	0	0	0	0
4	0	0.7	0	0	0	0	0	0	0.7	0	0	0	0.7	0.7	0	0	0	0	0	0	0	0
5	0	0	0	0	0.8	0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0.8	0	0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0.8	0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0.8	0	0.8	0	0	0	0	0	0.8	0	0	0	0	0	0
10	0	0	0	0	0	0	0.8	0	0	0.8	0	0	0	0	0	0.8	0	0	0	0	0	0
11	0	0.7	0.7	0	0	0	0	0	0	0	0	0	0.7	0.7	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0.8	0.8	0	0	0	0	0	0	0	0.8	0	0	0	0	0	0
13	0.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.7	0.8	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8
16	0	0.7	0.7	0	0	0	0	0	0.7	0	0	0	0	0.8	0	0	0	0	0	0	0	0
18	0	0.7	0.7	0	0	0	0	0	0.7	0	0	0	0.8	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0.8	0.8	0	0.8	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0.8	0	0	0	0	0	0	0	0	0	0.7	0	0	0	0	0	0	0	0.7	0	0	0
23	0.7	0	0	0	0	0	0	0	0	0	0.8	0	0	0	0	0	0	0.7	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0.8	0	0	0	0	0	0	0	0	0	0

Second iteration

I = {1, 2, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 16, 18, 19, 20, 21, 22, 23, 24, 25, 26}, C_2 = { }.

$a_{26\,14}$ = 0.8 are maximum. Then C_2= {26, 14}.

There are no j such that $a_{26\,j} + a_{14\,j}$ is maximum. Then final C_2= {26, 14}.

After deleting rows and columns the table has become a:

	1	2	4	5	6	8	9	10	11	12	13	16	18	19	20	21	22	23	24	25
1	0	0	0	0	0	0	0	0	0	0	0.7	0	0	0	0	0	0.8	0.7	0	0
2	0	0	0.7	0	0	0	0	0	0.7	0	0	0.7	0.7	0	0	0	0	0	0	0
4	0	0.7	0	0	0	0	0	0	0.7	0	0	0.7	0.7	0	0	0	0	0	0	0
5	0	0	0	0	0.8	0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0.8	0	0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0.8	0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0.8	0	0.8	0	0	0	0	0.8	0	0	0	0	0
10	0	0	0	0	0	0	0.8	0	0	0.8	0	0	0	0	0.8	0	0	0	0	0
11	0	0.7	0.7	0	0	0	0	0	0	0	0	0.7	0.7	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0.8	0.8	0	0	0	0	0	0	0.8	0	0	0	0	0
13	0.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.7	0.8	0	0
16	0	0.7	0.7	0	0	0	0	0	0.7	0	0	0	0.8	0	0	0	0	0	0	0
18	0	0.7	0.7	0	0	0	0	0	0.7	0	0	0.8	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0.8	0.8	0	0.8	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0.8	0	0	0	0	0	0	0	0	0	0.7	0	0	0	0	0	0	0.7	0	0
23	0.7	0	0	0	0	0	0	0	0	0	0.8	0	0	0	0	0	0.7	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8	0

The third iteration

I = {1, 2, 4, 5, 6, 8, 9, 10, 11, 12, 13, 16, 18, 19, 20, 21, 22, 23, 24, 25 }, C_3 = { }.

$a_{18\,16}$ = 0.8 are maximum. Then C_3= {18, 16}.

$a_{18\,4} + a_{16\,4}$ = 1.4 are maximum. Then C_3= {18, 16, 4}

$a_{18\,2} + a_{16\,2} + a_{4\,2}$ = 2.1 are maximum. Then C_3= {18, 16, 4, 2}.

$a_{18\,11} + a_{16\,11} + a_{4\,11} + a_{2\,11}$ = 2.8 are maximum. Then C_3= {18, 16, 4, 2, 11}.

There are no j such that $a_{18\,j} + a_{16\,j} + a_{4\,j} + a_{2\,j}$ is maximum. Then final C_3= {18, 16, 4, 2, 11}.

After deleting rows and columns the table has become a:

	1	5	6	8	9	10	12	13	19	20	21	22	23	24	25
1	0	0	0	0	0	0	0	0.7	0	0	0	0.8	0.7	0	0
5	0	0	0.8	0.8	0	0	0	0	0	0	0	0	0	0	0
6	0	0.8	0	0.8	0	0	0	0	0	0	0	0	0	0	0
8	0	0.8	0.8	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0.8	0.8	0	0	0.8	0	0	0	0	0
10	0	0	0	0	0.8	0	0.8	0	0	0.8	0	0	0	0	0
12	0	0	0	0	0.8	0.8	0	0	0	0.8	0	0	0	0	0
13	0.7	0	0	0	0	0	0	0	0	0	0	0.7	0.8	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0.8	0.8	0.8	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0.8	0	0	0	0	0	0	0.7	0	0	0	0	0.7	0	0
23	0.7	0	0	0	0	0	0	0.8	0	0	0	0.7	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8	0

Fourth iteration

I = {1, 5, 6, 8, 9, 10, 12, 13, 19, 20, 21, 22, 23, 24, 25 }, C_4 = { }.

$a_{25\,24}$ = 0.8 are maximum. Then C_4= {25, 24}.

There are no j such that $a_{25\,j} + a_{24\,j}$ is maximum. Then final C_4= {25, 24}.

After deleting rows and columns the table has become a:

	1	5	6	8	9	10	12	13	19	20	21	22	23
1	0	0	0	0	0	0	0	0.7	0	0	0	0.8	0.7
5	0	0	0.8	0.8	0	0	0	0	0	0	0	0	0
6	0	0.8	0	0.8	0	0	0	0	0	0	0	0	0
8	0	0.8	0.8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0.8	0.8	0	0	0.8	0	0	0
10	0	0	0	0	0.8	0	0.8	0	0	0.8	0	0	0
12	0	0	0	0	0.8	0.8	0	0	0	0.8	0	0	0
13	0.7	0	0	0	0	0	0	0	0	0	0	0.7	0.8
19	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0.8	0.8	0.8	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0.8	0	0	0	0	0	0	0.7	0	0	0	0	0.7
23	0.7	0	0	0	0	0	0	0.8	0	0	0	0.7	0

Fifth iteration

I = {1, 5, 6, 8, 9, 10, 12, 13, 19, 20, 21, 22, 23}, C_5 = { }.

$a_{5\,8}$ = 0.8 are maximum. Then C_5= {5, 8}.

$a_{5\,6} + a_{8\,6}$ = 1.6 are maximum. Then C_5= {5, 8, 6}.

There are no j such that $a_{5\,j} + a_{8\,j} + a_{6\,j}$ is maximum. Then final C_5= {5, 8, 6}.

After deleting rows and columns the table has become a:

	1	9	10	12	13	19	20	21	22	23
1	0	0	0	0	0.7	0	0	0	0.8	0.7
9	0	0	0.8	0.8	0	0	0.8	0	0	0
10	0	0.8	0	0.8	0	0	0.8	0	0	0
12	0	0.8	0.8	0	0	0	0.8	0	0	0
13	0.7	0	0	0	0	0	0	0	0.7	0.8
19	0	0	0	0	0	0	0	0	0	0
20	0	0.8	0.8	0.8	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0
22	0.8	0	0	0	0.7	0	0	0	0	0.7
23	0.7	0	0	0	0.8	0	0	0	0.7	0

Sixth iteration

I = {1, 9, 10, 12, 13, 19, 20, 21, 22, 23}, C_6 = { }.

$a_{9\,10}$ = 0.8 are maximum. Then C_6= {9, 10}.

$a_{9\,12} + a_{10\,12}$ = 1.6 are maximum. Then C_6= {9, 10, 12}

$a_{9\,20} + a_{10\,20} + a_{12\,20}$ = 2.4 are maximum. Then C_6= {9, 10, 12, 20}.

There are no j such that $a_{9\,j} + a_{10\,j} + a_{12\,j} + a_{20\,j}$ is maximum. Then final C_6= {9, 10, 12, 20}.

After deleting rows and columns the table has become a:

	1	13	19	21	22	23
1	0	0.7	0	0	0.8	0.7
13	0.7	0	0	0	0.7	0.8
19	0	0	0	0	0	0
21	0	0	0	0	0	0
22	0.8	0.7	0	0	0	0.7
23	0.7	0.8	0	0	0.7	0

Seventh iteration

I = {1, 13, 19, 21, 22, 23}, C_7 = { }.

$a_{13\,23}$ = 0.8 are maximum. Then C_7= {13, 23}.

$a_{13\,1} + a_{23\,1}$ = 1.4 are maximum. Then C_7= {13, 23,1}

$a_{13\,22} + a_{23\,22} + a_{1\,22}$ = 2.2 are maximum. Then C_7= {13, 23, 1, 22}.

There are no j such that $a_{13\,j} + a_{23\,j} + a_{1\,j} + a_{22\,j}$ is maximum. Then final C_7= {13, 23, 1, 22}.

After deleting rows and columns the table has become a:

	19	21
19	0	0
21	0	0

Now $a_{19\ 19} = a_{12\ 21} = a_{21\ 21} = 0$. Then {19}, {21} are two separated clusters.

In this way, when $\alpha = 0.55$, we have **9 clusters:**

C_1= {15, 17, 3, 7}, C_2= {26, 14}, C_3= {18, 16, 4, 2, 11}, C_4= {25, 24}, C_5= {5, 8, 6}, C_6= {9, 10, 12, 20}, C_7= {13, 23, 1, 22}, C_8={19}, C_9={21}

OR

{ {O, Q, C, G} {Z, N} {R, P, D, B, K} {Y,X} {E, H, F} {I, J, L, T} {M, W, A, V} {S} {U} }