Attributes of mixed types:

ID	Name	Age	Gender	Grade	Sum
1	A	20	Male	A 2	10
2	B	21	Male	A+ 3	20
3	C	35	Female	A- 1	15
4	D	18	Female	A+ 3	5

⇒ For name:

$$d(i,j) = \frac{p-m}{p}$$

p = total no of nominal attribute

m = match.

$\therefore d(1,2) = \frac{1-0}{1} = 1,$

$\therefore d(1,3) = \frac{1-0}{1} = 1,$

$\therefore d(1,4) = \frac{1-0}{1} = 1,$

$\therefore d(2,3) = \frac{1-0}{1} = 1$

$\therefore d(2,4) = \frac{1-0}{1} = 1$

$\therefore d(3,4) = \frac{1-0}{1} = 1$

P = 1

$\delta = 0 \rightarrow$ Null / 0

$\delta = 1 \rightarrow$

for age:

$d(1,2) = \frac{|21-20|}{35-18} = 0.06$, $d(2,3) = \frac{|35-21|}{35-18} = 0.82$

$d(1,3) = \frac{|35-20|}{35-18} = 0.88$, $d(2,4) = \frac{|18-21|}{35-18} = 0.18$

$d(1,4) = \frac{|18-20|}{35-18} = 0.12$, $d(3,4) = \frac{|18-35|}{35-18} = 1$

for Gender: Binary

$$d(i,j) = \frac{r+s}{q+r+s+t}$$

i ↑ \ j →	1	0
1	q	r
0	s	t

	1	0
1	1	0
2	1	0
3	0	1
4	0	0

$\therefore d(1,2) = \frac{0+0}{1+0+0+0} = 0$

$\therefore d(1,3) = \frac{1+0}{0+1+0+0} = 1$

$\therefore d(1,4) = \frac{1+0}{0+1+0+0} = 1$

$\therefore d(2,3) = \frac{1+0}{0+1+0+0} = 1$

$\therefore d(2,4) = \frac{1+0}{0+1+0+0} = 1$

$\therefore d(3,4) = \frac{0+0}{0+0+0+1} = 0$

for grade:

Let

ID	Grade
1	A
2	A+
3	A-
4	A+

→

ID	Grade
1	2
2	3
3	1
4	3

Now

$$z_{if} = \frac{r_{if} - 1}{M_f - 1}$$

r_{if} = value of particular cell.

M_f = Highest level.

$\therefore z_1 = \frac{2-1}{3-1} = 0.5$, $z_2 = \frac{3-1}{3-1} = 1$

$z_3 = \frac{1-1}{3-1} = 0$, $z_4 = \frac{3-1}{3-1} = 1$

Now,

ID	Grade
1	0.5
2	1
3	0
4	1

$\therefore d(1,2) = |1-0.5| = 0.5$, $d(2,3) = |1-0| = 1$

$\therefore d(1,3) = |0-0.5| = 0.5$, $d(2,4) = |1-1| = 0$

$\therefore d(1,4) = |1-0.5| = 0.5$, $d(3,4) = |1-0| = 1$

$8 \times d(1,2) + 8 \times d(1,2)$

4

Now,

dissimilarity of mixed attribute :

$$d(i,j) = \frac{\sum_{f=1}^{p} \delta_{ij}^{(f)} d_{ij}^{(f)}}{\sum_{f=1}^{p} \delta_{ij}^{(f)}}$$

num — nu

চারটির attribute এর value গুলোর dissimilarity কে একত্রে combine করে একটির মান পাওয়া।

এখানে, d(1,2) এর ক্ষেত্রে প্রতিটি attribute এর d(1,2) value কে s(ij) দিয়ে গুণ করে, sij হল ... তারপর যোগ

$s_{ij} = 0 \rightarrow$ যদি missing

$s_{ij} = 1 \rightarrow$ " " 1.

0+1

$$d(1,2) = \frac{(1 \times 1) + (1 \times 0.06) + (0 \times 0) + (1 \times 0.5)}{1+1+0+1}$$

$= 0.52$

$$\therefore d(1,3) = \frac{(1 \times 1) + (0.88 \times 1) + (1 \times 1) + (0.5 \times 1)}{1+1+1+1}$$

$= 0.845$

$$d(1,4) = \frac{(1 \times 1) + (0.12 \times 1) + (1 \times 1) + (0.5 \times 1)}{1+1+1+1}$$

$= 0.655$

$$d(2,3) = \frac{(1 \times 1) + (0.82 \times 1) + (1 \times 1) + (1 \times 1)}{1+1+1+1}$$

$= 0.955$

$$d(2,4) = \frac{(1\times1)+(0.38\times1)+(1\times1)+(0\times0)}{1+1+1+0}$$

$$= 0.73$$

$$d(3,4) = \frac{(1\times1)+(1\times1)+(0\times0)+(1\times1)}{1+1+0+1}$$

$$= 1$$

∴ dissimilarity matrix is:

	1	2	3	4
1	0			
2	0.52	0		
3	0.845	0.955	0	
4	0.655	0.73	1	0

Cosine Similarity

Formula

Document	team	Coach	hockey	baseball
D_1	5	0	3	0
D_2	3	0	2	0
D_3	0	7	0	2
D_4	0	1	0	0

formula; $\text{cosine similarity} = \frac{\sum D_1 \cdot D_2}{\sqrt{\sum (D_1)^2} \sqrt{\sum (D_2)^2}}$

$$\therefore \cos(d_1, d_2) = \frac{(5\times3)+(0\times0)+(3\times2)+(0\times0)}{\sqrt{(5)^2+0^2+3^2+0^2} \times \sqrt{3^2+0^2+2^2+0^2}}$$

$$= \frac{15+0+6}{5.83 \times 3.61}$$

$$= 1$$

$$\cos(d_2, d_3) = \frac{(3\times0)+(0\times7)+(2\times0)+(0\times2)}{\sqrt{3^2+0^2+2^2+0^2} \times \sqrt{0^2+7^2+0^2+2^2}}$$

$$= \frac{0}{}$$

$$= 0$$

Same way $\cos(d_1, d_3)$, $\cos(d_1, d_4)$

$\cos(d_3, d_4)$ এভাবে করে পাওয়া যাবে.