


**Gower's Similarity Coefficient**

Marzena Kryszkiewicz




**Gower's Similarity Coefficient...**

Gower's similarity coefficient for objects  $p$  and  $q$ :

$$G(p, q) = \frac{\sum_{d=1}^m w(p_d, q_d) \times s(p_d, q_d)}{\sum_{d=1}^m w(p_d, q_d)}, \text{ where}$$

- $m$  is the number of attributes;
- $s(p_d, q_d)$  is the *similarity* of objects  $p$  and  $q$  on attribute  $d$ ;
- $w(p_d, q_d)$  is the *weight* of attribute  $d$  for objects  $p$  and  $q$ . If  $p$  and  $q$  are *incomparable* on  $d$  (e.g.  $p_d$  or  $q_d$  is unknown), then  $w(p_d, q_d) = 0$ . Otherwise, it is typically equal to 1.

2



**Gower's Similarity Coefficient**

- For numerical attributes:  $s(p_d, q_d) = 1 - \frac{|p_d - q_d|}{range_d}$ .
- For nominal attributes:  $s(p_d, q_d) = 1$  if  $p_d = q_d$ ;  $s(p_d, q_d) = 0$ , otherwise.
- For binary (dychotomous) attributes:
  - If  $p_d = 1$  and  $q_d = 1$ , then  $s(p_d, q_d) = 1$ ,  $w(p_d, q_d) = 1$ .
  - If  $p_d = 1$  and  $q_d = 0$ , then  $s(p_d, q_d) = 0$ ,  $w(p_d, q_d) = 1$ .
  - If  $p_d = 0$  and  $q_d = 1$ , then  $s(p_d, q_d) = 0$ ,  $w(p_d, q_d) = 1$ .
  - If  $p_d = 0$  and  $q_d = 0$ , then  $s(p_d, q_d) = 0$ ,  $w(p_d, q_d) = 0$ .

3

**Example: Calculating similarity of objects with mixed attributes...**

Id	Age	Car type	Risk
0	23	Family	High
1	17	Sport	High
2	43	Sport	Low
3	68	Family	Low
4	32	Truck	Low
5	20	Family	High

Object to be classified:

Id	Age	Car type	Risk
6	20	Sport	?

$$G(6,0) = \frac{1 \times \left(1 - \frac{|20 - 23|}{100}\right) + 1 \times 0}{1 + 1} = \frac{0.97}{2} = 0.485;$$

$$G(6,1) = \frac{1 \times \left(1 - \frac{|20 - 17|}{100}\right) + 1 \times 1}{1 + 1} = \frac{1.97}{2} = 0.985$$

4

**Example: Calculating similarity of objects with mixed attributes...**

Id	Age	Car type	Risk
0	23	Family	High
1	17	Sport	High
2	43	Sport	Low
3	68	Family	Low
4	32	Truck	Low
5	20	Family	High

Object to be classified:


Id	Age	Car type	Risk
7		Sport	?

$$G(7,0) = \frac{0 \times \text{undefined} + 1 \times 0}{0 + 1} = \frac{0}{1} = 0;$$

$$G(7,1) = \frac{0 \times \text{undefined} + 1 \times 1}{0 + 1} = \frac{1}{1} = 1.$$

$$G(p, q) = \frac{\sum_{d=1}^m w(p_d, p_d) \times s(p_d, p_d)}{\sum_{d=1}^m w(p_d, p_d)}$$

5



**Example: Calculating similarity of objects with mixed attributes...**

Id	Age	Has a car	Risk
0 (person)	23	Yes	High
1 (dog)	17	No	High
2 (person)	43		Low
3 (person)	68	Yes	Low
4 (cat)	32	No	Low
5 (person)	20	Yes	High

Object to be classified:

Id	Age	Has a car?	Risk
8 (person)	20	Yes	?

$$G(8,0) = \frac{1 \times \left(1 - \frac{|20 - 23|}{100}\right) + 1 \times 1}{1 + 1}$$

$$G(8,1) = \frac{1 \times \left(1 - \frac{|20 - 17|}{100}\right) + 1 \times 0}{1 + 1}$$

$$G(p, q) = \frac{\sum_{d=1}^m w(p_d, p_d) \times s(p_d, p_d)}{\sum_{d=1}^m w(p_d, p_d)}$$

6

WARSAW UNIVERSITY OF TECHNOLOGY  
DEVELOPMENT PROGRAMME

### Example: Calculating similarity of objects with mixed attributes

Id	Age	Has a car	Risk
0 (person)	23	Yes	High
1 (dog)	17	No	High
2 (person)	43		Low
3 (person)	68	Yes	Low
4 (cat)	32	No	Low
5 (person)	20	Yes	High

Object to be classified:

Id	Age	Has a car?	Risk
9 (cat)	20	No	?

$$G(9,1) = \frac{1 \times \left(1 - \frac{|20 - 17|}{100}\right) + 0 \times 0}{1 + 0}$$
$$G(9,2) = \frac{1 \times \left(1 - \frac{|20 - 43|}{100}\right) + 0 \times undefined}{1 + 0}$$
$$G(p,q) = \frac{\sum_{d=1}^m w(p_d, p_d) \times s(p_d, p_d)}{\sum_{d=1}^m w(p_d, p_d)}$$

7

### References

- J. C. Gower, A general coefficient of similarity and some of its properties. Biometrics 27, 857-874 (1971)

8