



Gower's Similarity Coefficient...

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

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

A

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 $s(p_d, q_d) = 0$, $w(p_d, q_d) = 0$.

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

ld	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	

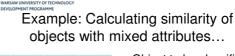
6	20	Sport	?			
$G(6,0) = \frac{1 * \left(1 - \frac{ 20 - 23 }{100}\right) + 1 * 0}{1 + 1}$						

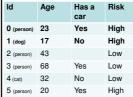
$$G(p,q) = \frac{\sum_{d=1}^{m} w(p_d, p_d) \times s(p_d, p_d)}{\sum_{d=1}^{m} w(p_d, p_d)}$$

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

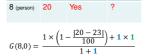
Example: Calculating similarity of objects with mixed attributes...

ld	Age	Car	Risk	Object to be classified:				
		type			ld	Age	Car type	Risk
0	23	Family	High			Age	our type	11131
1	17	Sport	High		7		Sport	?
2	43	Sport	Low					
3	68	Family	Low					
4	32	Truck	Low	$G(7,0) = \frac{0 * undefined}{0 + 1}$				-1*0
5	20	Family	High	$G(7,0) = \frac{0 * undefined +}{0+1} $ = $\frac{0}{1} = 0$;				
G(p,q)	$= \frac{\sum_{d=1}^{m} w}{\sum}$	$(p_d, p_d) \times \dots $ $m \atop d=1 $ $w(p_d, p_d)$	(p_d, p_d)				$\frac{undefined +}{0+1}$	





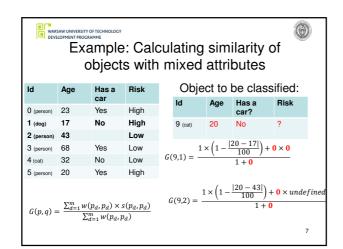
Obje	ot to i	oe class	itiea:
d	Age	Has a	Risk



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

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

6



References

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

8