Conceitos úteis

Normalização

A necessidade de harmonizar escalas pode ser suprida através da normalização dos valores dos atributos. Este processo corresponde no fundo, numericamente, à inconsciente adaptação de escalas que fazemos ao construir um gráfico XY com duas grandezas muito diferentes. Qualquer que seja a normalização utilizada, no entanto, não deve entender-se este processo como um meio de agregação num atributo único.

Esquemas típicos

As fórmulas seguintes referem-se a um atributo *k*, onde a alternativa *i* tem o valor z_i^k. As três primeiras correspondem a transformações lineares do atributo, o que não acontece com a última[1]. Note-se que existem outras possibilidades de normalização.

Fórmulas de transformação	Valores normalizados dos extremos da escala	Observações
$\left(z_i^k\right)_N = \dfrac{z_i^k - z_{min}^k}{z_{max}^k - z_{min}^k}$	Máximo = 1 Mínimo = 0	Atributos de maximização Ideal = 1
$\left(z_i^k\right)_N = \dfrac{z_{max}^k - z_i^k}{z_{max}^k - z_{min}^k}$	Máximo = 0 Mínimo = 1	Atributos de minimização Ideal = 1
$\left(z_i^k\right)_N = \dfrac{z_i^k}{z_{max}^k}$	Máximo = 1 Mínimo = z_{min}^k / z_{max}^k	Atributos de maximização Ideal = 1
$\left(z_i^k\right)_N = \dfrac{z_{min}^k}{z_i^k}$	Máximo = z_{max}^k / z_{min}^k Mínimo = 1	Atributos de minimização Transformação não-linear

Métricas

A distância entre duas alternativas (ou entre uma alternativa e o Ideal) é uma *medida de dissemelhança*, podendo usar-se uma métrica L_p para avaliar a dissemelhança entre z_i e z_j na globalidade dos critérios *k* (k=1..c):

$$L_p = \left[\sum_{k=1}^{c} \left(w_k . \left|z_i^k - z_j^k\right|\right)^p \right]^{1/p}$$

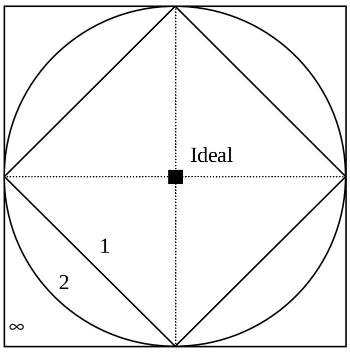
onde w_k é um factor de adequação de escalas, dispensável se se tiver procedido a normalização prévia (ou se as escalas forem semelhantes). No quadro seguinte explicita-se a fórmula geral para os valores de *p* mais interessantes, que correspondem a diferentes filosofias na valorização relativa das diferenças nos vários atributos.

[1] Por esse motivo, não é completamente correcto usar o termo "normalização" no último caso.

Fórmula	Designação
$L_1 = \sum_{k=1}^{c} w_k . \lvert z_i^k - z_j^k \rvert$	Distância de Manhattan, de Hamming ou pombalina
$L_2 = \sqrt{\sum_{k=1}^{c} \left(w_k . \left(z_i^k - z_j^k \right) \right)^2}$	Distância euclidiana
$L_\infty = \max_k \left\{ w_k . \lvert z_i^k - z_j^k \rvert \right\}$	Distância de Chebyshev

Ignorando os factores de escala para facilidade de raciocínio, não é difícil verificar a diferença entre L_1 e L_∞: enquanto que no primeiro caso há *compensação completa* entre critérios (tanto faz uma diferença de 8 como duas de 4), no segundo caso só a maior diferença conta (desde que a maior diferença seja 8, todas as outras são irrelevantes). Analisados os extremos, torna-se claro que a distância euclidiana (e qualquer métrica com $p \geq 2$) corresponde a uma situação intermédia, em que existe compensação, mas não completa (com $p=2$, duas diferenças de 2 correspondem a uma diferença de 8).

A utilização desta métricas para avaliar a distância ao Ideal presta-se a uma descrição gráfica dos diversos casos, conforme se vê na figura seguinte, onde se mostram os lugares geométricos dos pontos à distância 1, 2 e ∞ do Ideal, para o caso de dois atributos (eixos paralelos aos indicados na figura).



Também é esclarecedor, para comparar as diferentes métricas, verificar a sua influência num exemplo simples como o seguinte, onde se mostram as três distâncias para um conjunto limitado de pontos a duas dimensões (atributos *x* e *y*), num caso de maximização. As duas colunas *dx* e *dy* indicam, em cada componente, as distâncias de cada ponto ao Ideal, que é obviamente o ponto (10,10).

x	y	dx	dy	L1	L2	Linf
10	**0**	0	10	10	10,0	10
8	**2**	2	8	10	8,2	8
2	**2**	8	8	16	11,3	8
5	**5**	5	5	10	7,1	5
2	**5**	8	5	13	9,4	8
0	**5**	10	5	15	11,2	10
0	**10**	10	0	10	10,0	10