pattern.GDM2(clusterSim)

An application of GDM2 distance for ordinal data to compute the distances of objects from the upper (ideal point co-ordinates) or lower (anti-ideal point co-ordinates) pattern object

The main goal of the linear ordering methods is to identify the order of the objects with respect to predetermined criterion. Usually the synthetic measure, which aggregates the partial information contained in the variables, is used.

The GDM2 distance measure can be applied for computing distances from the pattern object in the linear ordering methods. Here:

1. We start with data matrix $[x_{ij}]$, where x_{ij} denotes *i*-th observation on *j*-th variable.

Table 1 Data matrix (27 residential housing properties described by six variables)

No.	x1	x2	x3	x4	x5	x6
1	5	3	1	3	1	3
2	3	3	3	3	2	2
3	5	4	3	4	1	2
4	2	3	1	3	2	3
5	5	4	2	4	1	2
6	4	3	2	3	1	3
7	3	4	3	3	2	2
8	4	4	3	4	1	1
9	5	3	2	4	1	2
10	4	2	1	3	1	3
11	5	4	3	4	1	4
12	4	3	1	4	1	2
13	4	4	3	3	1	1
14	4	4	3	3	2	3
15	5	4	2	3	2	4
16	3	3	2	3	1	1
17	4	2	1	3	2	3
18	4	1	2	4	1	2
19	3	3	2	3	2	4
20	3	2	1	3	1	3
21	4	3	2	3	1	1
22	5	3	2	4	1	2
23	5	4	3	4	1	2
24	4	2	2	3	1	2
25	3	2	1	2	2	3
26	3	3	1	1	2	3
27	2	3	1	1	2	3

Residential housing properties were described by the following variables:

x1 Location of environmental land, which is linked to a dwelling (1 – poor, 2 – inadequate, 3 – satisfactory, 4 – good, 5 – very good).

x2 Standard utility of a dwelling (1 – bad, 2 – low, 3 – average, 4 – high).

x3 Living conditions occurring on the land, which is linked to a dwelling (1 – bad, 2 – average, 3 – good).

x4 Location of land, which is related to dwelling in the area of the city (1 – central, 2 – downtown, 3 – intermediate, 4 – peripheral).
x5 Type of condominium (1 – low, 2 – large).
x6 Area of land, which is related to dwelling (1 – below the contour of the building, 2 – outline of the building, 3 – the outline of the building with the environment acceptable, such as parking, playground, 4 – the outline of the building with the environment too much).

2. Three types of performance variables are distinguished:
– stimulants – where higher value means better performance,
– destimulants – where low values indicate better performance,
– nominants – where the best value is implied. Object performance is positively assessed if the measure has implied value.

Types of performance variables:
x1, x2, x3 – stimulants,
x4, x5 – destimulants,
x6 – nominant (the nominal category: 3).

In `performanceVariable` we give information about performance variables – `c("s","s","s","d","d","n")`.

In `nomOptValues` vector we give information about nominal values of nominants – `c(NA,NA,NA,NA,NA,3)`.

3. The co-ordinates of pattern object consist of the best variables' values (for ideal point co-ordinates) or consist of the worst variables' values (anti-ideal point co-ordinates).

4. Upper pattern – ideal point co-ordinates consists of the best variables' values. Two types of construction `upper` pattern are distinguished in `patternType`:
a) `"dataBounds"` – pattern should be calculated as following: maximum for stimulants, minimum for destimulants, nominal value for nominants,
b) `"manual"` – pattern should be given in `patternManual` and pattern co-ordinates contain:
   – real numbers,
   – "min" – for minimal value of variable (for destimulants),
   – "max" – for maximal value of variable (for stimulants),
   – "nom" – for nominal value of variable (given in `nomOptValues` vector) – for nominants.

5. Lower pattern – anti-ideal point co-ordinates consists of the worst variables' values.

Firstly the variable observations, which have nominant character of type, are transformed into destimulant variables via `nominalTransfMethod` using formula:
a) `"database"` – for each nominant separately GDM2 distance is calculated between each nominant observation (with repetitions – all variable values are used in calculation) and nominal value. Next the variable observations are replaced by those distances,
b) `"symmetrical"` – for each nominant separately GDM2 distance is calculated between each nominant observation (without repetition – each observation is used once) and nominal value. Next the variable observations are replaced by those distances.

Secondly two types of construction `lower` pattern are distinguished in `patternType`:
a) `"dataBounds"` – pattern should be calculated as following: minimum for stimulants, maximum for destimulants,
b) `"manual"` – pattern should be given in `patternManual` and pattern co-ordinates contain:
   – real numbers,
   – "min" – for minimal value of variable (for stimulants),
   – "max" – for maximal value of variable (for destimulants).

6. If the weights are not equal in GDM2 it is necessary to give the weights w_j in `weights`:

a) `"different1"` – vector of different weights should satisfy conditions: each weight takes value from interval [0; 1] and sum of weights equals one – e.g. `c(0.4,0.1,0.2,0.15,0.05,0.1)`,
b) `"different2"` – vector of different weights should satisfy conditions: each weight takes value from interval [0; m] and sum of weights equals *m* (*m* – the number of variables) – (e.g. `c(0.4,1.3,1.4,0.2,1.5,1.2)`.

7. For each object the GDM2 distance from the pattern object (ideal or anti-ideal point) is determined.

8. We sort the objects based on ascending order of GDM2 distances from pattern object (upper pattern) or descending order (lower pattern).

9. Graphical presentation of results.

References

Jajuga, K., Walesiak, M., Bak, A. (2003), *On the general distance measure*, In: M. Schwaiger, O. Opitz (Eds.), *Exploratory data analysis in empirical research*, Springer-Verlag, Berlin, Heidelberg, 104-109.

Walesiak, M. (1993), *Statystyczna analiza wielowymiarowa w badaniach marketingowych* [*Multivariate Statistical Analysis in Marketing Research*], Wroclaw University of Economics, Research Papers no. 654.

Walesiak, M. (1999), *Distance Measure for Ordinal Data*, Argumenta Oeconomica, No. 2 (8), 167-173.

Walesiak, M. (2002), *Propozycja uogolnionej miary odleglosci w statystycznej analizie wielowymiarowej*, In: J. Paradysz (Ed.), *Statystyka regionalna w służbie samorzadu lokalnego i biznesu*, Internetowa Oficyna Wydawnicza, Centrum Statystyki Regionalnej, AE w Poznaniu, Poznan, 115-121.

Walesiak, M. (2006), *Uogólniona miara odległości w statystycznej analizie wielowymiarowej* [*The Generalized Distance Measure in multivariate statistical analysis]*, Wydanie drugie rozszerzone, Wydawnictwo AE, Wroclaw.

Walesiak, M. (2008), *Cluster analysis with clusterSim computer program and R environment*, Acta Universitatis Lodziensis. Folia Oeconomica, No. 216, 303-311.

Walesiak, M. (2009), *Analiza skupien [Cluster analysis]*, In: M. Walesiak, E. Gatnar (Eds.), *Statystyczna analiza danych z wykorzystaniem programu R [Statistical data analysis with R program]*, WN PWN, Warszawa, 407-433.

Walesiak, M. (2011), *Uogólniona miara odległości GDM w statystycznej analizie wielowymiarowej z wykorzystaniem programu R* [*The Generalized Distance Measure GDM in multivariate statistical analysis with R]*, Wydawnictwo UE, Wroclaw.

Walesiak, M., Dziechciarz, J., Bąk, A. (1998), *Ordinal variables in the segmentation of advertisement receivers*, In: A. Rizzi, N. Vichi, H.H. Bock (Eds.), *Advances in Data Science and Classification*, Springer, Heidelberg, 655-662.