



STANOVENIE MIER VZDIALENOSTÍ (NEPODOBNOTÍ) OBJEKTOV ZHLUKOVEJ ANALÝZY

DETERMINATION OF DISTANCE METRICS (DISSIMILARITY) OF OBJECTS IN CLUSTER ANALYSIS

Peter TREBUŇA - Jana HALČINOVÁ - Marek KLIMENT - Jaromír MARKOVIČ - Milan FIEO

Abstract

Cluster analysis is used for classification of objects into clusters, provided that the similarity of objects in the cluster is the maximum and the relative similarity of clusters is minimal. Similarity resp. dissimilarity of objects can be determined by the distance metric, degree of association and degree of correlation. This paper describes the method of determining the degree of distance, which are the most popular degree of similarity used in the cluster analysis.

Key words

Cluster Analysis, Degree of Dissimilarity, Distance Metric.

Úvod

Zhluková analýza patrí medzi viacrozmerné štatistické metódy a definuje sa ako súhrnné označenie pre skupinu metód, ktorých cieľom je buď zoskupiť dané objekty do zhlukov alebo vytvoriť hierarchiu zhlukov objektov, pričom zhluk predstavuje skupinu objektov, ktorých vzdialenosť tzn. nepodobnosť je menšia ako vzdialenosť objektov do zhluku nepatriacich.

Stanovenie mier vzdialenosti

Vzdialenosť medzi objektmi A a B je označovaná ako $d_{A,B}$. Miere vzdialeností majú nasledujúce vlastnosti:

1. Vzdialenosť medzi dvoma objektmi je nezáporná, teda $d_{A,B} \geq 0$. Ak je vzdialenosť medzi objektmi nulová, tak hodnoty premenných charakterizujúcich tieto dva objekty sú rovnaké.
2. Vzdialenosť medzi objektmi A a B sa rovná vzdialenosti B a A, teda $d_{A,B} = d_{B,A}$.
3. Musí platiť trojuholníková nerovnosť: $d_{A,B} \leq d_{A,i} + d_{B,i}$.

Základnou myšlienkou zhlukovej analýzy je podobnosť resp. nepodobnosť objektov, preto sú výsledky zhlukovej analýzy vo veľkej miere ovplyvnené výberom miery podobnosti resp. nepodobnosti.

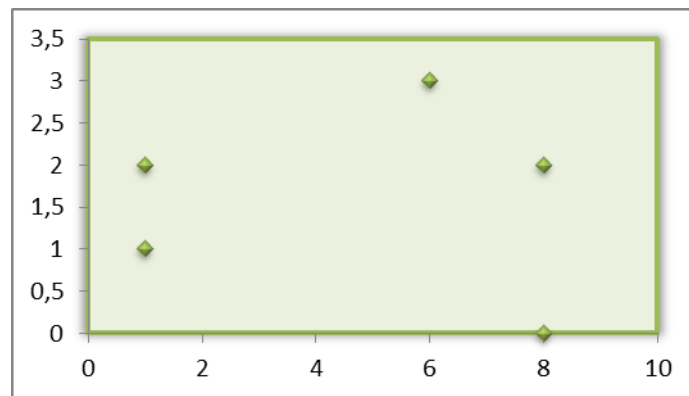
Miere nepodobnosti (vzdialenosti) sú založené na prezentácii objektov v priestore, ktorého súradnice tvoria jednotlivé premenné. Tieto miery nepodobnosti objektov sa využívajú v štatistických programoch a platí, že čím je vzdialenosť medzi objektmi menšia, tým sú si tieto objekty podobnejšie a čím je vzdialenosť medzi objektmi väčšia, tým sú si objekty nepodobnejšie (vzdialenejšie). Na konkrétnom príklade popíšeme spôsob stanovenia mier nepodobnosti (vzdialenosti) objektov zhlukovej vzdialenosti. Máme daných 5 objektov



charakterizovaných 2 znakmi ako je to uvedené v Tab.1., ktoré je možné graficky znázorniť v priestore (Obr.1).

Tab. 1 Znaký objektov zhlukovej analýzy

Objekt	1. znak	2. znak
1	1	1
2	1	2
3	6	3
4	8	2
5	8	0



Obr. 1 Znázornenie objektov zhlukovej analýzy

Najznámejšími metrikami pre vyjadrenie vzdialenosti dvoch bodov x_i a x_j v m -rozmernom priestore sú:

- *Euklidovská vzdialenosť* – nazýva sa tiež geometrická metrika a predstavuje dĺžku prepony pravouhlého trojuholníka a jej výpočet je založený na Pytagorovej vete.

$$D_E(x_i, x_j) = \sqrt{\sum_{l=1}^m (x_{il} - x_{jl})^2} = \|x_i - x_j\|,$$

Pre určenie Euklidovskej vzdialenosti 1.objektu a 4.objektu je výpočet nasledovný:

$$d(1,4) = \sqrt{(1-8)^2 + (1-2)^2} = 7,07$$

V Tab.2 sú uvedené vzájomné Euklidovské vzdialenosti všetkých existujúcich dvojíc objektov pre daný príklad zhlukovej analýzy.

Tab. 2 Matica Euklidovskej vzdialenosti objektov E_1

	1	2	3	4	5
1	0	1,00	5,39	7,07	7,07
2	1,00	0	5,10	7,00	7,28
3	5,39	5,10	0	2,24	3,61
4	7,07	7,00	2,24	0	2,00
5	7,07	7,28	3,61	2,00	0

Podobným spôsobom sú stanovené ostatné miery nepodobnosti (vzdialenosti) objektov.



- Štvorcová euklidovská vzdialenosť – je základom Wardovej metódy zhlukovej analýzy.

$$D_{ES}(x_i, x_j) = \sum_{l=1}^m (x_{il} - x_{jl})^2,$$

$$d(1, 4) = (1 - 8)^2 + (1 - 2)^2 = 50$$

Tab. 3 Matica Štvorcovej euklidovskej vzdialenosti objektov E_2

	1	2	3	4	5
1	0	1	29	50	50
2	1	0	26	49	53
3	29	26	0	5	13
4	50	49	5	0	4
5	50	53	13	4	0

- *Manhattanská (mestských blokov) vzdialenosť* - nazýva sa tiež taxi metrika, pretože je to najkratšia vzdialenosť, akú môže taxík prejsť medzi mestskými blokmi na Manhattane.

$$D_B(x_i, x_j) = \sum_{l=1}^m |x_{il} - x_{jl}| = |x_i - x_j|,$$

$$d(1, 4) = |1 - 8| + |1 - 2| = 8$$

Tab. 4 Matica Manhattanskej vzdialenosti objektov E_3

	1	2	3	4	5
1	0	1	7	8	6
2	1	0	6	7	5
3	7	6	0	1	1
4	8	7	1	0	2
5	6	5	1	2	0

- Čebyševova vzdialenosť - napodobňuje pohyb kráľa po šachovnici.

$$D_C(x_i, x_j) = \max_l (|x_{il} - x_{jl}|)$$

$$d(1, 4) = \max \{|1 - 8|; |1 - 2|\} = \max \{7; 1\} = 7$$

Tab. 5 Matica Čebyševovej vzdialenosti objektov E_4

	1	2	3	4	5
1	0	1	5	7	7
2	1	0	5	7	7
3	5	5	0	2	3
4	7	7	2	0	2
5	7	7	3	2	0



Súhrn

Príspevok sa venuje stanoveniu vzdialenosti (nepodobnosti) objektov zhlukovej analýzy aplikovanom na konkrétnom príklade. Hodnoty jednotlivých mier vzdialenosti sú usporiadané v matici. Takáto matica vzdialenosti predstavuje vstupnú maticu pre nasledujúce kroky zhlukovej analýzy využitím hierarchických alebo nehierarchických metód zhlukovania.

Príspevok bol pripravený v rámci riešenia grantového projektu KEGA 004TUKE-4/2013 Intenzifikácia modelovania vo výučbe II. a III. stupňa v študijnom odbore 5.2.52 Priemyselné inžinierstvo.

Kľúčové slová

Zhluková analýza, miery nepodobnosti, miery vzdialenosti.

Použitá literatúra

- [1] BUDA, J., FILO, M.: Manažment operácií, Prešov: Vydavateľstvo Michala Vaška, 2008. 152s., ISBN 978-80-7165-710-1,
- [2] KALL, F. –BARTÁNUSOVÁ, M.: Using Tracking Systems for Ergonomic Analysis. In. Transcom 2013. 2013. s 157-160. ISBN 978-80-554-0695-4,
- [3] MARKOVIČ, J., JAMRICOVÁ, S.: Prosperujúci podnik v globálnom prostredí, Invent, Žilina 2008,
- [4] MELOUN, M. – MILITKÝ, J. – HILL, M.: Počítačová analýza vícerozmerných dat v príkladoch. Praha: Academia. 2005. 449 s. ISBN 80-200-1335-0,
- [5] ŘEZANKOVÁ, H. – HÚSEK, D. – SNÁŠEL, V.: Shluková analýza dat. Praha: Professional Publishing. 2009. 218 s. ISBN 978-80-86946-81-8,
- [6] STANKOVIČOVÁ, I. – VOJTKOVÁ, M.: Viacrozmerné štatistické metódy s aplikáciami. Bratislava: IURA Edition, 2007, 261 s. ISBN 978-80-8078-152-1.

Kontaktná adresa

doc. Ing. Peter Trebuňa, PhD.

Ing. Jana Halčinová

Ing. Marek Kliment

Technická univerzita v Košiciach

Strojnícka fakulta

Katedra priemyselného inžinierstva a manažmentu

Němcovej 32, 040 01 Košice

e-mail: peter.trebuna@tuke.sk, jana.halcinova@tuke.sk, marek.kliment@tuke.sk

Ing. Jaromír MARKOVIČ, PhD.

Slovak Legal Metrology

Hviezdoslavova 1124/31

97401 Banská Bystrica

e-mail: slm@slm.sk

Dr.h.c. Ing. Milan Fiľo, PhD

ECO-INVEST a.s. Bratislava