M. Krmelova: Factor Analysis of Sports Data

Department of Computer Science, Palacky University, Olomouc (17. listopadu 12, CZ–77146 Olomouc, Czech Republic)

Introduction

Factor analysis is method of finding new variables which are hidden in data and which may data explain. This variables are also known as factors. In our work we suppose data with fuzzy attributes. Especially we focus on data from sports events.

Factor Analysis of Data with Fuzzy Attributes

:: input: relational object-attribute table describing objects (rows), their attributes/features (columns) and relationship expressing grades to which object has attribute

:: output: an object-factor matrix and a factor-attribute matrix

:: grades are taken from a bounded scale L (we assume L=$\{0, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, 1\}$ due to Miller's 7 ± 2 phenomenon)

:: five-element scale (0 = not at all, 0.25 = little bit, 0.5 = half, 0.75 = quite, 1 = fully)

:: it could be illustrated as color boxes, darker blue means higher degree, lighter means lower degree

Decomposition and Factors

:: consider an $n \times m$ object-attribute matrix I

:: decompose I into an $n \times k$ object-factor matrix A and a $k \times m$ factor-attribute matrix B

:: goal: find k as small as possible

:: for a factor l there is a degree A_{il} to which l applies to i and a degree B_{lj} to which j is a manifestation of l

:: degree a to which "l applies to i and j is a manifestation of l":

$$a = A_{il} \otimes B_{lj}$$

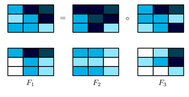
:: $I = A \circ B$ is defined:

$$(A \circ B)_{ij} = \bigvee_{l=1}^{k} A_{il} \otimes B_{lj}$$

:: reasonable choice is Łukasiewicz pair of adjoint operations

$$a \otimes b = max(a + b - 1, 0)$$
$$a \to b = min(1 - a + b, 1)$$

:: example:



Decathlon Examples

:: first example was presented in [2]

:: the decathlon is a combined event in athletics consisting of ten track and field events and it's great example, because

- the winners are determined by the combined performance in all events
- performance is judged on a points system in each event, not by the position achieved
- all events have the same point scale

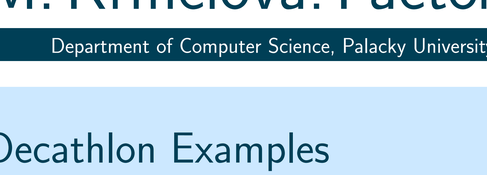
2004 Olympic Games Decathlon

The results of top five athletes in points which are obtained using the IAAF Scoring Tables for Combined Events

	10	lj	sp	hj	40	hu	dt	pv	jt	15
Sebrle	894	1020	873	915	892	968	844	910	897	680
Clay	989	1050	804	859	852	958	873	880	885	668
Karpov	975	1012	847	887	968	978	905	790	671	692
Macey	885	927	835	944	863	903	836	731	715	775
Warners	947	995	758	776	911	973	941	880	669	693

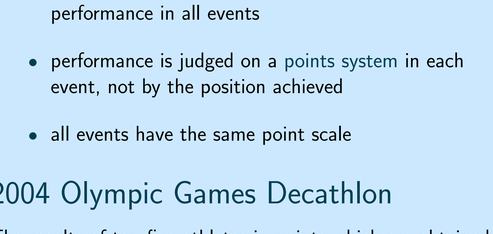
10 = 100m sprint race, lj = long jump, sp = shot put, hj = high jump, 40 = 400m sprint race, hu = 110m hurdles, dt = discus throw, pv = pole vault, jt = javelin throw, 15 = 1500m run

Data table with graded attributes



transformed data into a five-element scale by a natural transformation and rounding

Results



:: factor F_1 applies to Sebrle and Macey to degree 0.5, to both Clay and Karpov to degree 1, and to Warners to degree 0.75

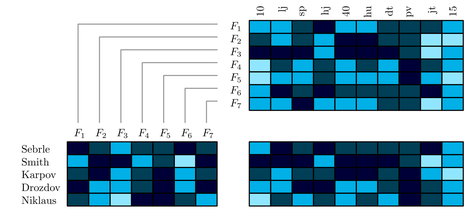
:: factor F_1 applies to attribute 10 (100 m) to degree 1, to attribute lj (long jump) to degree 1, to attribute sp (shot put) to degree 0.75, etc.

:: relatively weak performance (degree 0.25) in javelin throw and pole vault

:: factor F_1 could be understood as ability to run fast for short distances

:: factor F_2 could be understood as explosiveness, which is known for Sebrle and Clay

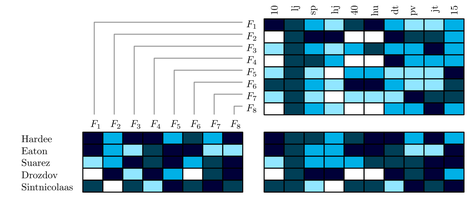
2007 World Championships in Athletics



:: factor F_6 is nearly the same as F_2 in '2004 Olympic Games Decathlon' example (they are equal with degree 0.925)

:: other factors are also similar

2011 World Championships in Athletics



:: in this example are also almost the same factors like factor F_6 and F_6 in '2004 Olympic Games Decathlon' example (with degree 0.9) and factor F_1 and F_1 in '2004 Olympic Games Decathlon' example (with degree 0.875)

:: different factors are caused by big differences between top 5 athletes in this championship

Ordinal Scaling

:: scaling - process of transformation of data tables with general attributes to data tables with yes/no attributes

:: misunderstanding is that ordinal scaling yields the same results as the above approach

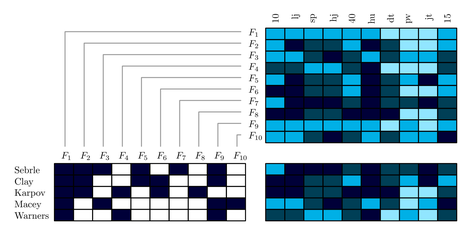
:: ordinal scaling:

- smaller # of formal concepts
- therefore, possibly larger # of factors necessary for decomposition

compared to the above approach

:: for example we scale all attributes from '2004 Olympic Games Decathlon' example into new five attributes (e.g. 10 → 10 not at all, 10 little bit, 10 half, 10 quite, 10 fully) and process factor analysis with crisp attributes

:: we get 10 factors



:: F_7 could be understood as explosiveness like in '2004 Olympic Games Decathlon' example with fuzzy attributes, we get Sebrle, but no information about another athletes (e.g. Clay is also quite explosive)

Future Research

:: experiments with another sports, where we can use factor analysis of data with fuzzy attributes (Pentathlon, Heptathlon, Figure skating, Hockey, ...)

:: robustness of the method

References

[1] Belohlavek R.: Optimal decompositions of matrices with entries from residuated lattices. Journal of Logic and Computation, 2010.

[2] Belohlavek R., Vychodil V.: Factor Analysis of Incidence Data via Novel Decomposition of Matrices. Proceedings, Lecture Notes in Artificial Intelligence, 2009.

International Center for Information and Uncertainty
SSIU 2012: International Summer School "Information and Uncertainty"
http://mcin.upol.cz/SSWIU-2012/

esf european social fund in the czech republic
EUROPEAN UNION
MINISTRY OF EDUCATION, YOUTH AND SPORTS
OP Education for Competitiveness