POLSKA AKADEMIA NAUK
Instytut Badań Systemowych

METHODS OF ESTIMATION OF RELATIONS OF: EQUIVALENCE, TOLERANCE AND PREFERENCE IN A FINITE SET

Leszek Klukowski

Warsaw 2011

SYSTEMS RESEARCH INSTITUTE
POLISH ACADEMY OF SCIENCES

Series: SYSTEMS RESEARCH
Volume 69

==

Series Editor:
Prof. dr hab. inż. Jakub Gutenbaum

Warsaw 2011

SYSTEMS RESEARCH INSTITUTE
POLISH ACADEMY OF SCIENCES

Leszek Klukowski

METHODS OF ESTIMATION OF RELATIONS OF: EQUIVALENCE TOLERANCE AND PREFERENCE IN A FINITE SET

Warsaw 2011

dr Leszek Klukowski
Systems Research Institute
Polish Academy of Sciences
Newelska 6, 01-447 Warsaw, Poland
email: Leszek.Klukowski@ibspan.waw.pl

Papers reviewers:

Prof. dr hab. inż. Ignacy Kaliszewski
Prof. dr hab. Tadeusz Trzaskalik

The work has been supported by the grant No N N111434937 of the Polish Ministry of Science and Higher Education

Printed in Polands
Systems Research Institute
Polish Academy of Sciences
Newelska 6, 01-447 Warsaw, Poland
www.ibspan.waw.pl

ISSN 0208-8029
ISBN 9788389475374

Chapter 11

Summary and conclusions

11.1. Introduction

The work here contained constitutes the synthesis of main results of the author concerning estimation of three relations – equivalence, tolerance, and preference – on the basis of pairwise comparisons with random errors. The problems of that type occur often in applications and have been investigated in statistical literature. Therefore, it appears reasonable to devote to them an entire individual work.

11.2. Achievements of the work and further researches

The following new results, presented here, should be emphasized.

1^0. Two types of data have been taken into account: binary and multivalent. Binary data reflect qualitative features of the compared pairs of elements, i.e. equivalence or direction of preference in a pair, while multivalent data – quantitative features, i.e. the number of subsets including both elements (tolerance relation) or distance between elements - in the form of difference of ranks (preference relation).

2^0. The assumptions concerning the comparison errors are weaker than those commonly used in the literature, especially:

a) expected values of comparison errors can differ from zero,
b) distributions of comparison errors may be unknown,
c) comparisons including the same element can be correlated.

Therefore, the algorithms proposed can be used in the cases, when the existing algorithms are not applicable (can produce incorrect results).

3^0. Two estimators have been examined; the first one is based on the sum of differences between the relation form and the comparison data, the second is based on differences between the relation form and the median from comparisons of each pair. The estimators have a simple intuitive form, i.e. optimization tasks, and analytical properties guaranteeing good efficiency, especially in the case of multiple comparisons of each pair. The properties indicate, in particular, that the efficiency of the first estimator is better, but involves higher cost of computations. The median estimator requires a lower amount of computations in the case of application of optimization algorithms, and is more robust (robustness is important property in the case of multivalent comparisons).

4^0. The analytical properties of the estimators have been complemented with the results of simulation study. This allows for determining of parameters, especially the number of comparisons N, guaranteeing the required precision of estimates; a definite value of N provides for the frequency of errorless result close to one or equal one. The simulation approach allows for evaluation of the distribution of frequencies of errorless solution also in the case of unknown distributions of comparison errors. Such distributions are replaced by some boundary distributions – the quasi-uniform distributions, proposed by the author. The simulation study indicates an excellent efficiency of multivalent estimators – the original concept of the author; the errorless estimate can be obtained for moderate N and the probability of errorless comparison lower than ½.

5^0. The properties of estimates can be thoroughly validated; validation comprises the fact of existence of the relation and the assumptions as to the comparison errors. The assumptions can be verified with the use of known tests and the methods proposed by the author. The establishment of existence of relation can also be based on simulation approach. It is possible, as well, to choose the relation type – equivalence or tolerance, and the type of the preference relation – strict or weak. Therefore, the approach has the features of data mining techniques.

6^0. The precision of the estimators, examined in the simulation study (Chapter 9, Klukowski, 2011a, 2012), are based on measures proposed in the work: • frequency of the errorless estimate, • average absolute, one-dimensional error, and • distribution of the average absolute, one-dimensional error. The one-dimensional error is the sum of components of the multi-dimensional error. It is an adequate measure of difference between the estimate and the relation; however, multi-dimensional error can also be subject to analysis, especially in graphical form.

7^0. The approach proposed allows for combining of comparisons obtained from different sources, e.g. statistical tests, experts, neural networks. It is also possible to combine binary and multivalent data and to apply two-stage estimators, based, in the first stage, on binary comparisons, and in the second stage – on multivalent comparisons, obtained in the first stage.

8^0. The estimates are obtained on the basis results from optimization tasks. They can be solved with the use of complete enumeration of the feasible set or the heuristic algorithms. The first approach requires fast processors, which are available currently. Heuristic algorithms can be based on random search, genetic algorithms, swarm intelligence, or hierarchical agglomeration algorithms.

9^0. The approach presented will be developed in the following directions: statistical learning, estimation of more complex structures of data (e.g. hierarchical), multidimensional (multi-criteria) pairwise comparisons, etc. An important field is also constituted by application of the estimators and tests developed.

The work here contained focuses on presentation of theoretical aspects of estimation of the relations considered. The upcoming research will concentrate on applications of the estimators proposed.

The book presents the estimators of three relations: equivalence, tolerance, and preference in a finite set of data items, based on multiple pairwise comparisons, assumed to be disturbed by random errors. The estimators were developed by the author. They can refer to binary (qualitative), multivalent (quantitative) and combined comparisons. The estimates are obtained on the basis of solutions to the discrete programming problems. The estimators have been developed under weak assumptions on the distributions of comparison errors; in particular, these distributions can have non-zero expected values. The estimators have good statistical properties, including, especially importantly, consistency. Therefore, they produce good results in cases when other methods generate incorrect estimates. The precision of the estimators has been established with the use of simulation methods. The estimates can be validated in a versatile way. The whole estimation process, i.e. comparisons, estimation and validation can be computerized. The approach allows also for inference about the relation type – equivalence or tolerance, on the basis of binary data. Thus, it has features of data mining methods.

The estimators have been applied for ranking and grouping of data from some empirical sets. In particular, estimation of the tolerance relation (overlapping classification) was applied for determination of homogenous shapes of functions expressing profitability of treasury securities and was used for forecasting purposes.

ISSN 0208-8029
ISBN 9788389475374

SYSTEMS RESEARCH INSTITUTE
POLISH ACADEMY OF SCIENCES
Phone: (+48) 22 3810246 / 22 3810277 / 22 3810241 / 22 3810273
email: biblioteka@ibspan.waw.pl