SMTDA

BOOK OF ABSTRACTS

2014

3rd Stochastic Modeling Techniques and Data Analysis International Conference

Editor

Christos H. Skiadas

June 11-14, 2014

Lisbon Portugal

Imprint
Book of Abstracts of the 3rd Stochastic Modeling Techniques and Data Analysis International Conference
Lisbon, Portugal: 11-14 June, 2014
Published by: ISAST: International Society for the Advancement of Science and Technology.
Editor: Christos H Skiadas

Book ISBN: 978-618-81257-5-9
e-Book ISBN: 978-618-81257-6-6

Printed in Athens, Greece

SMTDA 2014

Stochastic Modeling Techniques and Data Analysis International Conference and Demographics Workshop Lisbon, Portugal, 11-14 June 2014

Plenary and Keynote Talks

On Cluster Analysis of Complex and Heterogeneous Data

Helena Bacelar-Nicolau[1], Fernando C. Nicolau[2], Áurea Sousa[3], Leonor Bacelar-Nicolau[4]

[1]Faculty of Psychology, LEAD; ISAMB, CEA; University of Lisbon, [2]FCT, Department of Mathematics, New University of Lisbon. [3]Dep. of Math., CEEAplA, CMATI, University of Azores. [4]Faculty of Medicine, Institute of Preventive Medicine, ISAMB, University of Lisbon, Portugal

Classical clustering methods usually work with a set of objects as statistical data units described by a set of homogeneous (that is, of the same type) variables in a two-way framework. This paradigm can be extended in such way that data units may be either simple / first-order elements (e.g., subjects, individuals from a basic population) or groups of / second-order or more objects from some population (e.g., subsets of the population, subsamples of a sample, classes of a partition) and/or descriptive variables may simultaneously be of different (e.g., binary, multi-valued, histogram or interval) types. Therefore, one has a complex and/or heterogeneous data set under analysis. In that case classification will often be carried out by using a three-way or a symbolic/complex approach.

The present work synthesizes previous methodological results and shows several developments mostly regarding hierarchical cluster analysis of three-way and of symbolic data, where statistical data units are described by either a homogeneous or a heterogeneous set of complex variables. We illustrate that approach on a few examples from the statistical literature. The methodology has been applied with success in data mining context, concerning multivariate analysis of real-life data bases issued from economy, medicine and social sciences.

Keywords: Cluster analysis, Similarity coefficient, Hierarchical clustering model, Three-way data, Symbolic data, Heterogeneous data.