

STATISTICAL ANALYSIS AND DATA MINING

Research Article

Brief overview of symbolic data and analytic issuesL. Billard 

First published: 08 March 2011

<https://doi.org/10.1002/sam.10115>

Cited by: 10



About



Access

JLS



ire

Abstract

With the advent of contemporary computers, datasets can be massively huge, too large for direct analysis. One of the many approaches to this problem of size is to aggregate the data according to some appropriate scientific question of interest, with the resulting dataset performance being one with symbolic - valued observations such as lists, intervals, histograms, and the like. Other datasets, small or large, are naturally symbolic in nature. One aim here is to provide a brief nontechnical overview of symbolic data and discuss how they arise. We also provide brief insights into some of the issues that arise in their analyses. These include the need to take into account the internal variations inherent in symbolic data but not present in classical data. Another issue is that, by the nature of the aggregation, resulting datasets can contain “holes” or regions that are not possible; thus, accommodation for these need to be taken into account, when, e.g. seemingly interval data are actually some other form of symbolic data (such as histogram data). Also, we show how other forms of complex data differ from symbolic data; so, e.g. fuzzy data are a different domain than that for symbolic data. Finally, we look at further research needs for the subject. A more technical introduction to symbolic data and available analytic methodology is given by Noirhomme and Brito. © 2011 Wiley Periodicals, Inc. *Statistical Analysis and Data Mining* 4: 149–156, 2011

Citing Literature



About Wiley Online Library



Help & Support



Opportunities



Connect with Wiley



Copyright © 1999-2018 John Wiley & Sons, Inc. All rights reserved

WILEY