

Ordinal Data Analysis

Gerd Stumme

Knowledge & Data Engineering Group, Dept. of Electrical Engineering and
Computer Science & Research Center for Information System Design (ITeG),
University of Kassel, Germany

stumme@cs.uni-kassel.de

Orders are ruling our live: social hierarchies, rankings in online shops, classification systems, waiting queues and many more. However, the majority of data analysis approaches has been developed for numerical features. There are – at least – two reasons for this: *i*) These features exist in many datasets, because many of the properties can be measured by real numbers, and *ii*), they allow to make use of a rich set of mathematical tools, such as computing differences/distances, means, deviations, weighted averages, etc.

In 1946, S. S. Stevens aimed at providing a solid mathematical foundation to the question whether and how it is possible to measure human sensation. To this end, he introduced four different levels of measurement - nominal, ordinal, interval, and ratio. This triggered a discussion about the meaning of ‘measurement’ in the various cases and the statistical manipulations that can legitimately be applied. This discussion has been very productive, even though vicious at times, and is still going on today.

In the meanwhile, many analysis methods have been developed for data on the nominal, interval and ratio levels. However, there exists up to now no comprehensive theory for analysing ordinal data. There are only few data science/data analysis/machine learning techniques that are particularly suited for ordinal data (e. g., decision trees work well for ordinal data). Because of the appeal of analysis and machine learning techniques for numerical data, many scientists also apply these techniques to ordinal data. In clustering, for instance, ranks are frequently treated as being interval-scaled, so that distances one is used to (e. g., Euclidean distance) can be applied. However, this may lead to significant misinterpretations of the data.

The aim of this talk is to inspire the audience to venture into the development of tools for ordinal data analysis. To this end, we will illustrate various roles of orders in human live by real-world examples and summarize the main issues of the discussion about the various levels of measurement before discussing different types of ordinal data analysis tasks. We conclude the talk by presenting ongoing work of our group on ordinal data analysis.