

# On Subgroup Discovery in Numerical Domains<sup>\*</sup>

Henrik Grosskreutz and Stefan Rüping

Fraunhofer IAIS, Schloss Birlinghoven, Sankt Augustin, Germany  
{henrik.grosskreutz, stefan.rueping}@iais.fraunhofer.de

**Abstract.** Subgroup discovery is a Knowledge Discovery task that aims at finding subgroups of a population with high generality and distributional unusualness. While several subgroup discovery algorithms have been presented in the past, they focus on databases with nominal attributes or make use of discretization to get rid of the numerical attributes. In this paper, we illustrate why the replacement of numerical attributes by nominal attributes can result in suboptimal results. Thereafter, we present a new subgroup discovery algorithm that prunes large parts of the search space by exploiting bounds between related numerical subgroup descriptions. The same algorithm can also be applied to ordinal attributes. In an experimental section, we show that the use of our new pruning scheme results in a huge performance gain when more than just a few split-points are considered for the numerical attributes.

## Reference

1. Grosskreutz, H., Rüping, S.: On Subgroup Discovery in Numerical Domains. *Data Mining and Knowledge Discovery* (2009) DOI: 10.1007/s10618-009-0136-3

---

<sup>\*</sup> This is an extended abstract of an article published in the *Data Mining and Knowledge Discovery* journal [1].