

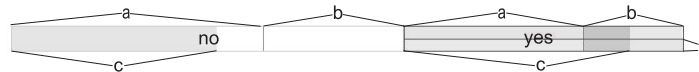
# Visualization Techniques for Semi-Automatic Subgroup Mining

Martin Atzmueller, Tobias Vogeles, and Frank Puppe

Institut für Informatik, Universität Würzburg, D-97074 Würzburg

**Abstract.** Subgroup mining, i.e., discovering "interesting" subgroups of individuals given a certain target property, is a promising approach for descriptive and explorative data mining; in such settings semi-automatic approaches often yield better results than automatic methods (c.f. Atzmueller et al. 2004). However, to effectively support the user exploring the space of interesting patterns, appropriate visualization techniques are required in order to guide the subgroup mining process.

In this paper, we introduce such visualization methods consisting of two types: 'embedded' visualizations that guide the subgroup mining task and visualizations for post-processing of the results. As a main issue, we present a new attribute-oriented and interactive visualization method to guide the semi-automatic search process, inspired by the work of (Spence (2001)): in this novel visualization, data is represented in the form of highly condensed tables; the rows correspond to specific attributes and their ranges, including useful information by several sub-bars and color encodings. Figure 1 shows a schematic example of this visualization for a binary attribute with the values *yes* and *no*. Instances within a subgroup that contain the target property are *positive* instances, otherwise negative ones. Then, one possible goal is to maximize the *positive* share. In the current subgroup  $SG_c$ , (a) indicates the (currently) positive instances, and (b) denotes the negative ones. In the 'next' subgroup  $SG_n$ , i.e., including the particular attribute value, (c) shows the positive instances for this subgroup, which can be compared to (a). Finally, (d) shows the gain in precision, comparing the subgroups  $SG_c$  and  $SG_n$ .



**Fig. 1.** Example visualization of a binary attribute

We discuss the specific advantages of the presented visualizations, and propose strategies for their application. Furthermore, we demonstrate the presented methods in the context of a fielded medical system, e.g., (Atzmueller et al. 2003, 2005).

## References

- ATZMUELLER, M., BAUMEISTER, J. and PUPPE, F. (2003): Evaluation of two Strategies for Case-Based Diagnosis handling Multiple Faults In: *Proc. 2nd Conf. Professional Knowledge Management (WM2003)*, Luzern, Switzerland
- ATZMUELLER, M., PUPPE, F. and BUSCHER, H.-P. (2005): Profiling Examiners using Intelligent Subgroup Mining In: *Proc. 10th Intl. Workshop on Intelligent Data Analysis in Medicine and Pharmacology*, Aberdeen, Scotland. 46–51
- ATZMUELLER, M., PUPPE, F., and BUSCHER, H.-P. (2004): Towards Knowledge-Intensive Subgroup Discovery In: *Proc. LWA-2004*, Berlin, 117–123
- SPENKE, M. (2001): Visualization and Interactive Analysis of Blood Parameters with InfoZoom. *Artificial Intelligence In Medicine*, 22, 159–172