Fast Subgroup Discovery for Continuous Target Concepts

Martin Atzmueller and Florian Lemmerich

University of Würzburg,
Department of Computer Science VI
Am Hubland, 97074 Würzburg, Germany
{atzmueller, lemmerich}@informatik.uni-wuerzburg.de

Abstract. Subgroup discovery is a flexible data mining method for a broad range of applications. It considers a given property of interest (target concept), and aims to discover interesting subgroups with respect to this concept. In this paper, we especially focus on the handling of continuous target variables and describe an approach for fast and efficient subgroup discovery for such target concepts. We propose novel formalizations of effective pruning strategies for reducing the search space, and we present the SD-Map* algorithm that enables fast subgroup discovery for continuous target concepts. The approach is evaluated using real-world data from the industrial domain.

1 Introduction

Subgroup discovery is a general knowledge discovery and data mining method that can be customized for various application scenarios. Prominent examples include knowledge discovery in medical and technical domains, e.g., [1–4]. Subgroup discovery is an undirected method for the identification of groups of individuals that deviate from the norm considering a certain property of interest [5], that is, a certain target concept. For example, the risk of coronary heart disease (target variable) is significantly higher in the subgroup of smokers with a positive family history than in the general population.

In this context, continuous target concepts have received increasing attention recently, e.g., [4, 6], especially regarding industrial applications. For example, we could investigate whether certain combinations of factors cause an increased repair and/or scrap rate in a manufacturing scenario.

Many existing approaches, e.g., [2, 5] require the discretization of continuous target attributes – with a significant loss of information. Therefore, this paper describes methods for fast and effective subgroup discovery for continuous target concepts and presents an efficient algorithm for this purpose. We first focus on pruning strategies for reducing the search space utilizing optimistic estimate functions for obtaining upper bounds for the possible quality of the discovered patterns. Specifically, we focus on the recently introduced notion of *tight optimistic estimate* functions for the case of continuous target concepts. Additionally, we show how an efficient method for subgroup discovery can be combined with the proposed pruning strategy, and present the SD-Map* algorithm as a novel adaptation of the efficient SD-Map [7] algorithm.

The rest of the paper is organized as follows: Section 2 provides the basics on subgroup discovery. After that, Section 3 introduces the challenges of handling continuous target concepts and proposes novel implementations of the tight optimistic estimate functions. Next, we propose the SD-Map* algorithm as an adaptation of the efficient SD-Map subgroup discovery method, and discuss related work. Section 4 provides an evaluation of the approach using real-world data from an exemplary industrial application. Finally, Section 5 concludes with a summary of the presented work and provides pointers for future work.

2 Preliminaries

In the following, we first introduce the necessary notions concerning the used knowledge representation, before we introduce subgroup discovery and its implementation using continuous target concepts.

2.1 Subgroup Discovery

The main application areas of subgroup discovery are exploration and descriptive induction to obtain an overview of the relations between a (dependent) target variable and a set of explaining (independent) variables. Then, the goal is to uncover properties of the selected target population of individuals featuring the given target property of interest. Specifially, these interesting subgroups should have the most unusual (distributional) characteristics with respect to the concept of interest given by the target variable [5].

A subgroup discovery task mainly relies on the following four main properties: the target variable, the subgroup description language, the quality function, and the discovery strategy. Since the search space is exponential concerning all the possible selectors of a subgroup description efficient discovery methods are necessary, e.g., beam-search or the exhaustive SD-Map algorithm [7].

First, let us introduce some basic notions: Let Ω_A denote the set of all attributes. For each attribute $a \in \Omega_A$ a range $dom(a)$ of values is defined. Let CB be the case base (data set) containing all available cases (instances). A case $c \in CB$ is given by the n-tuple $c = ((a_1 = v_1), (a_2 = v_2), \ldots, (a_n = v_n))$ of $n = |\Omega_A|$ attribute values, $v_i \in dom(a_i)$ for each a_i.

The subgroup description language specifies the individuals belonging to the subgroup. For a commonly applied single-relational propositional language a subgroup description can be defined as follows:

Definition 1 (Subgroup Description). *A subgroup description $sd(s) = \{e_1, \ldots, e_n\}$ of the subgroup s is defined by the conjunction of a set of selection expressions (selectors). The individual selectors $e_i = (a_i, V_i)$ are selections on domains of attributes, $a_i \in \Omega_A, V_i \subseteq dom(a_i)$. We define Ω_E as the set of all selection expressions and Ω_{sd} as the set of all possible subgroup descriptions.*

A subgroup s described by $sd(s)$ is given by all cases $c \in CB$ covered by the subgroup description $sd(s)$. A subgroup s' is called a *refinement* of s, if $sd(s) \subset sd(s')$.

2.2 Subgroup Quality Functions

A quality function measures the interestingness of the subgroup and is used to rank these. Typical quality criteria include the difference in the distribution of the target variable concerning the subgroup and the general population, and the subgroup size.

Definition 2 (Quality Function). *Given a particular target variable $t \in \Omega_E$, a quality function $q : \Omega_{sd} \times \Omega_E \rightarrow R$ is used in order to evaluate a subgroup description $sd \in \Omega_{sd}$, and to rank the discovered subgroups during search.*

For binary target variables, examples for quality functions are given by

$$q_{WRACC} = \frac{n}{N} \cdot (p - p_0)\,, \quad q_{PS} = n \cdot (p - p_0)\,, \quad q_{LIFT} = \frac{p}{p_0}\,, n \geq \mathcal{T}_n$$

where p is the relative frequency of the target variable in the subgroup, p_0 is the relative frequency of the target variable in the total population, $N = |CB|$ is the size of the total population, n denotes the size of the subgroup, and $\mathcal{T}_n$ specifies a minimal size constraint for the subgroup. As discussed in [2] q_{WRACC} (weighted relative accuracy) trades off the increase in the target share p vs. the generality (n) of the subgroup. The Piatetsky-Shapiro, e.g., [8] quality function q_{PS} is a variation of q_{WRACC} without considering the size of the total population. Finally, the quality function q_{LIFT} focuses on the decrease/increase of the target share.

The main difference between the different types of target variables is given by disjoint sets of applicable quality functions, due to the different parameters that can be applied for estimating the subgroup quality: Target *shares* are only applicable for binary or categorical attributes, while continuous target variables require averages/aggregations of values, e.g., the *mean*. As equivalents to the quality functions for binary targets discussed above, we consider the functions *Continuous Piatetsky-Shapiro* (q_{CPS}), *Continuous LIFT* (q_{CLIFT}), and *Continuous Weighted Relative Accuracy* (q_{CWRACC}):

$$q_{CWRACC} = \frac{n}{N} \cdot (m - m_0)\,, \quad q_{CPS} = n \cdot (m - m_0)\,, \quad q_{CLIFT} = \frac{m}{m_0}\,, n \geq \mathcal{T}_n$$

where n and N denote the size of the subgroup and the size of the total population as defined above, respectively, and m specifies the mean of the target variable within the subgroup; m_0 specifies the mean of the target variable in the total population.

The *CN2-SD* algorithm [2], is a prominent example of an heuristic subgroup discovery algorithm that applies a beam-search strategy. The adaption of such an algorithm is rather simple, as in each step the quality values of the subgroup hypotheses contained in the beam are directly updated from the case base. Instead of determining the target share(s) of (binary) target variables, simply the mean values of the cases contained in the subgroup m and (once) for the total population m_0 need to be obtained. It is easy to see that the continuous case subsumes the binary one as a special case: Computing the averages includes computing the target shares – by considering the values 1 and 0 for a *true/false* target concept, respectively. The ordinal case can be captured by mapping the ordinal values to continuous values and normalizing these if necessary.

3 Adapting Subgroup Discovery for Continuous Target Concepts

In the following, we show how to efficiently adapt exhaustive subgroup discovery for continuous target concepts. We discuss tight optimistic estimate quality functions [8], and we introduce novel formalizations for the case of continuous target variables. Additionally, we describe how the efficient subgroup discovery algorithm SD-Map [7] can be combined with pruning measures using tight optimistic estimate functions resulting in the novel SD-Map* algorithm.

3.1 Tight Optimistic Estimates

The basic principle of optimistic estimates [8] is to safely prune parts of the search space. This idea relies on the intuition that if the k best hypotheses so far have already been obtained, and the optimistic estimate of the current subgroup is below the quality of the worst subgroup contained in the k best, then the current branch of the search tree can be safely pruned. More formally, an optimistic estimate oe of a quality function qf is a function such that $s' \subseteq s \Rightarrow oe(s) > qf(s')$, i.e., that no refinement of subgroup s can exceed the quality $oe(s)$. An optimistic estimate is considered *tight* if for any database and any subgroup s, there exists a subset $s' \subseteq s$, such that $oe(s) = qf(s')$. While this definition requires the existence of a subset of s, there is not necessarily a subgroup description, that describes s', cf., [8].

For binary targets the determination of such a best subset is relatively simple using any quality function that is constructed according to the axioms postulated in [9]. The best subset is always given by the set of all cases, for which the target concept is true.

We introduce the following notation: $n(s) = |\{c \in s\}|$ specifies the size of subgroup s, $tp(s) = |\{c \in s | t(c) = true\}|$ the number of positive examples in s; $t(c)$ denotes the value of the target variable in case c and $p(s) = \frac{tp(s)}{n(s)}$ is the target share of the subgroup.

Theorem 1 *For each subgroup s with $p > p_0$ and for each boolean quality function q for which the axioms postulated in [9] apply: $s' \subseteq s \Rightarrow q(s') \leq q(s^*)$, where $s^* = \{c \in s | t(c) = true\}$*

Proof. We first show, that $q(s) \leq q(s^*)$. This means, that the quality of any subgroup with a positive quality is always lower or equal to the quality of the subset of examples, that only contains the positive examples of s. We apply the third axiom of [9]: "$q(s)$ monotonically decreases in n, when $p = c/n$ with a fixed constant c.": As fixed constant c we consider the number of positive examples tp, as $p = c/n \Leftrightarrow c = p \cdot n$ and $n(s) \cdot p(s) = tp(s) = tp(s^*) = n(s^*) \cdot p(s^*)$. So, the quality function monotonically decreases in n. As $n(s) > n(s^*)$ we conclude: $q(s) \leq q(s^*)$.

For arbitrary $s' \subseteq s$ we now need to consider two cases: If $s^* \subseteq s'$, then $q(s') \leq q(s^*)$ as shown above. If $s^* \not\subseteq s'$, then there exists a subset $s'' = \{c \in s | t(c) = true\}$, that contains only the positive examples of s'. The above proof then implies: $q(s') \leq q(s'')$. On the other hand $s'' \subseteq s^*$ is true, as s'' only consists of positive examples of s. We now apply the fourth axiom of [9]: "q(s) monotonically increases in n when $p > p_0$ is fixed.": As $p(s'') = p(s^*) = 1$ it follows that $q(s'') \leq q(s^*)$. Thus, $q(s') \leq q(s'') \leq q(s^*)$, proving the theorem. □

In contrast to the binary case, the continuous one is more challenging, since the best refinement of a subgroup is dependent on the used quality function: Given an average target value of $m_0 = 50$ and the subgroup s containing cases with values 20, 80 and 90. Then, for the quality function lift with $\mathcal{T}_n = 1$ the subset with the best quality contains only the case with value 90. On the other hand for the Pietatsky-Shapiro quality function, it contains two cases with the values 80 and 90, respectively.

Theorem 2 *For the Pietatsky-Shapiro quality function $q_{CPS}(s) = n \cdot (m - m_0)$ the tight optimistic estimate for any subgroup s is given by*

$$oe(s) = \sum_{c \in s, t(c) > 0} (t(c) - m_0),$$

.

Proof. We reformulate the Pietatsky-Shapiro quality function:

$$\begin{aligned} q_{CPS}(s) &= n \cdot (m - m_0) \\ &= n \cdot (\frac{\sum_{c \in s} t(c)}{n} - \frac{n \cdot m_0}{n}) \\ &= \sum_{c \in s} t(c) - n \cdot m_0 \\ &= \sum_{c \in s} (t(c) - m_0) \end{aligned}$$

For all subsets $s' \subseteq s$, this sum reaches its maximum for the subgroup s^*, that contains all cases with larger target values than the average of the population, since it contains only positive summands, but no negatives. The quality of s^* is given by $q_{CPS}(s^*) = oe(s)$ using the above formula. As no other subset of s can exceed this quality the $oe(s)$ is an optimistic estimate. Since for any given subgroup the estimate is reached by one of its subsets, the estimate is tight. □

Please note, that the tight optimistic estimate for the binary case provided in [8], i.e., $np(1 - p_0)$, can be seen as special case of this formula, considering $t(c) = 1$ for *true* target concepts and $t(c) = 0$ for *false* target concepts:

$$\begin{aligned} oe(s) &= \sum_{c \in s, t(c) > 0} (t(c) - m_0) \\ &= \sum_{c \in s, t(c) = 1} (1 - p_0) \\ &= np(1 - p_0). \end{aligned}$$

Theorem 3 *Considering the quality function Weighted Relative Accuracy $q_{CWRACC}(s) = \frac{n}{N} \cdot (m - m_0)$ the tight optimistic estimate for any subgroup s is given by*

$$oe(s) = \frac{1}{N} \sum_{c \in s, t(c) > 0} (t(c) - m_0),$$

where t(c) is the value of the target variable for the case c.

Proof. q_{CWRACC} differs by the factor $\frac{1}{N}$ from the Pietatsky-Shapiro function. The population size can be considered as a constant, so the proof proceeds analogously. □

Theorem 4 *For the quality function Lift with a minimum subgroup size $\mathcal{T}_n$ the optimistic estimate is given by $oe(s) = \sum_{i=1}^{\mathcal{T}_n}(v_i - m_0)$, where v_i is the value i-th highest in the subgroup in respect of the target variable.*

Proof. Since the size of the subgroup is not relevant for these quality functions, the best possible subset is always the subset with the highest average of the target attribute with size k. The quality of this subset is given by the above formula. □

3.2 Efficient Subgroup Discovery with SD-Map*

SD-Map [7] is based on the efficient FP-growth [10] algorithm for mining frequent patterns. As a special data structure, the frequent pattern tree or FP-tree is used which is implemented as an extended prefix-tree-structure that stores count information about the frequent patterns. FP-growth applies a divide and conquer method, first mining frequent patterns containing one selector and then recursively mining patterns of size 1 conditioned on the occurrence of a (prefix) 1-selector. For the recursive step, a conditional FP-tree is constructed, given the conditional pattern base of a frequent selector (node). Due to the limited space we refer to Han et al. [10] for more details.

SD-Map utilizes the FP-tree structure built in one database pass to efficiently compute quality functions for all subgroups. For the binary case, an FP-tree node stores the subgroup size and the true positive count of the respective subgroup description. In the case of a continuous target variable, we consider the sum of values of the target variable, enabling us to compute the respective quality functions value accordingly. Therefore, all the necessary information is locally available in the FP-tree structure. Please note, that the adaptions for numeric target variables includes the case of a binary variable as a special case, where the value of the target variable is 1, if the target concept is *true* and 0, otherwise.

SD-Map* extends SD-Map by including (optional) pruning strategies and utilizes quality functions with tight optimistic estimates for this purpose: For embedding (tight) optimistic estimate pruning into the SD-Map algorithm, we basically only need to consider three options for pruning and reordering/sorting according to the current (tight) optimistic estimates: (1) **Pruning**: In the recursive step when building a conditional FP-tree, we omit a (conditioned) branch, if the optimistic estimate for the conditioning selector is below the threshold given by the k best subgroup qualities. (2) **Pruning**: When building a (conditional) frequent pattern tree, we can omit all the nodes with an optimistic estimate below the mentioned quality threshold. (3) **Reordering/Sorting**: During the iteration on the currently active selector queue when processing a (conditional) FP-tree, we can dynamically reorder the selectors that have not been evaluated so far by their optimistic estimate value. In this way, we evaluate the *more promising* selectors first. This heuristic can help to obtain higher values for the pruning threshold early in the process, a way to prune more often earlier. Additionally, this step implements a modified depth-first search guided by the current optimistic estimates.

To efficiently compute the (tight) optimistic estimates we store additional information in the nodes of the FP-Tree, depending on the used quality function. For example, for the Piatetsky-Shapiro quality function we add the value $max(0, t(c) - p_0)$ for each case c to a field in the respective node during the construction of the FP-Tree. This field can also be propagated recursively – analogously to the sum of target values when building the conditional trees. It directly reflects the optimistic estimate of each node and can be immediately evaluated whenever it is needed for pruning.

3.3 Related Work and Discussion

In this paper, we propose novel formalizations of tight optimistic estimates for numeric quality functions. Additionally, we present the SD-Map* algorithm that enables efficient subgroup discovery for continuous target concepts. By utilizing these novel quality functions, SD-Map* shows a significant decrease in the number of examined states of the search space, and therefore also a significant reduction concerning the runtime and space requirements of the algorithm, as shown in the evaluation in Section 4. The techniques were implemented in the data mining environment VIKAMINE, and evaluation showed its benefit for the intended application in the industrial domain.

Handling numeric target concepts in the context of subgroup discovery has been first discussed by Kloesgen [9, 11] in the EXPLORA system. Kloesgen applied both heuristic and exhaustive subgroup discovery strategies without pruning. An improvement was proposed by Wrobel [5], presenting optimistic estimate functions for binary target variables. Recently, Grosskreutz et al. [8] introduced tight optimistic estimate quality functions as a further improvement on optimistic estimate quality functions for binary and nominal target variables.

Jorge et al. [4] introduced an approach for subgroup discovery with continuous target concepts applying special visualization techniques for the interactive discovery step. In contrast to the presented approach, the methods focus on a different rule representation, i.e., distribution rules, not on deviations of the target concept averages in the subgroup vs. the total population. Furthermore, an adapted standard algorithm for discovering frequent sets is applied, so there are no pruning options for enabling a more efficient discovery process.

Grosskreutz et al. [12] proposed the DpSubgroup algorithm that also incorporates tight optimistic estimate pruning. While their algorithm is somehow similar to the SD-Map algorithm, since also a frequent pattern tree is used for efficiently obtaining the subgroup counts, the DpSubgroup algorithm focuses on binary and categorical target concepts only, and lacks the efficient propagation method of SD-Map* when computing the tight optimistic estimates in the FP-tree. In contrast, SD-Map* is applicable for binary, categorical, and continuous target concepts. Additionally, DpSubgroup uses an explicit depth-first search step for evaluating the subgroup hypotheses while this step is implicitly included in the divide-and-conquer frequent pattern growth method of SD-Map* directly (that is, by the reordering/sorting optimization).

4 Evaluation

In the following, we outline the application scenario of the presented approach. After that, we present an evaluation using real-world data, and show, how the proposed approach helps to prune the search space significantly.

4.1 Application Scenario

The development of the presented techniques was mainly motivated by industrial applications in the service support and in the manufacturing domain that required fast responsiveness for interactive scenarios. In general, industrial applications of subgroup discovery often require the utilization of continuous parameters, for example, certain measurements of machines or production conditions. Then, the target concepts can often not be analyzed sufficiently using the standard techniques for binary/nominal subgroup discovery, since the discretization of the variables causes a loss of information. As a consequence, the interpretation of the results is often difficult.

Therefore, the presented techniques provide a robust alternative: In our applications, one important goal was the identification of subgroups (as combination of certain factors) that cause a significant increase/decrease in certain parameters, for example, the number of service requests for a certain technical component, or the fault/repair rate of a certain manufactured product. For a comprehensive analysis, the presented pruning technique was important to enable a semi-automatic involvement of the domain experts in order to effectively contribute in a discovery session.

4.2 Results

As discussed above, we provide an evaluation using exemplary real-world data from the industrial domain using the adapted SD-Map algorithm with pruning and the ’standard’ SD-Map algorithm as a reference, using the Piatetsky-Shapiro quality function. The applied data set contains about 20 attributes and a relatively large number of cases. Figure 1 shows the number of hypotheses that were considered during the discovery process. As discussed above, the complexity grows exponentially with the number of attributes and attribute values. It is easy to see, that the optimistic estimate pruning approach shows a significant decrease in the number of hypotheses considered, and thus also in the runtime of the algorithm. The ’full’ optimistic estimate pruning approach using dynamic reordering strategy (optimistic estimate pruning and dynamic sorting of the FP-Tree-header nodes) also shows a improvement by a factor of two compared to the approach using only optimistic estimate pruning.

As a further benchmark, Figure 2 shows results of applying the approach on the credit-g data set from the UCI [13] repository. We considered the target concept *credit amount*, and provided the nominal attributes of the data set subsequently. The results confirm the observation for the industrial data, and the decrease of considered hypotheses/running time is even more significant. Similar to the industrial data set, the ’full’ pruning/sort strategy of SD-Map* shows slight ’variations’ with respect to the number of considered hypotheses (cf., lines for 9 and 12 attributes of credit-g): This can be explained by the fact that the ordering strategy can yield better subgroup qualities earlier in the process.

These results clearly indicate the benefit and broad applicability of the approach: The pruning strategies enable fast (automatic) subgroup discovery for continuous target concepts, that can then be the starting point for further analysis.

#Attributes	w/o OE-Pruning	w/OE-Pruning	w/OE-Sort-Pruning
2	23	23	23
3	83	61	50
4	117	57	41
5	157	67	48
6	390	101	55
7	464	109	50
8	544	112	43
9	1546	131	48
10	2763	181	69
11	4333	170	84
12	4486	173	87
13	7567	219	110
14	9123	275	155
15	14057	325	186
16	20112	432	240
17	29708	548	335
18	43517	583	390
19	58139	612	381
20	65307	800	469

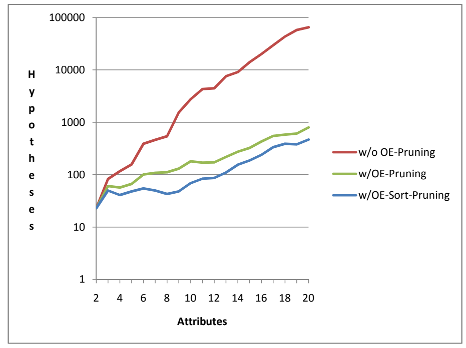

Fig. 1. Evaluation: Industrial Data. The x-axis of the graph shows the number of attributes provided to the discovery method, the y-axis shows the resulting number of considered hypotheses. The columns of the table indicate the number of hypotheses without pruning (*w/o OE-Pruning*), with optimistic-estimate pruning and no sorting (*w/ OE-Pruning*), and with the full optimistic-estimate/sorting strategy (*w/OE-Sort-Pruning*).

#Attributes	w/o OE-Pruning	w/OE-Pruning	w/OE-Sort-Pruning
2	27	27	27
3	215	119	68
4	652	130	102
5	2019	181	158
6	4756	238	170
7	7342	240	177
8	16779	260	168
9	28770	347	212
10	50015	411	307
11	93626	448	343
12	168818	500	212
13	224087	558	229

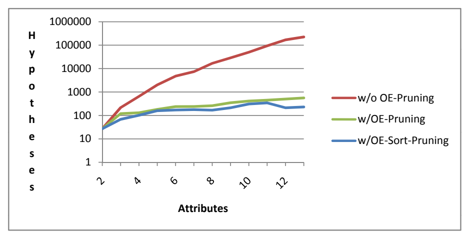

Fig. 2. Evaluation: Credit-G Data Set. The x-axis of the table shows the number of attributes provided to the discovery method, the y-axis shows the resulting number of considered hypotheses. The columns of the table indicate the number of hypotheses without pruning (*w/o OE-Pruning*), with optimistic-estimate pruning and no sorting (*w/ OE-Pruning*), and with the full optimistic-estimate/sorting strategy (*w/OE-Sort-Pruning*).

5 Conclusions

In this paper, we have presented techniques for fast subgroup discovery with continuous target concepts of interest: These feature novel formalizations of tight optimistic estimate quality functions and the SD-Map* algorithm for enabling efficient subgroup discovery for continuous target concepts. The applied pruning techniques for safely removing areas of the search space are based on utilizing these tight optimistic estimate functions for continuous target concepts.

The evaluation of the approach was performed using real-world data from industrial applications and showed significant improvements concerning the efficiency of the subgroup discovery approach since large areas of the search space could be safely pruned in the experiments. This enables a seamless application of the algorithm even in rather interactive contexts.

For future work, we aim to assess a combination of tight (continuous) optimistic estimates with sampling techniques and methods for distributed subgroup discovery in order to optimize the efficiency of the subgroup discovery method even more. Another promising direction for future research is given by effective visualization techniques for continuous target concepts.

Acknowledgements

This work has been partially supported by the German Research Council (DFG) under grant Pu 129/8-2.

References

1. Gamberger, D., Lavrac, N.: Expert-Guided Subgroup Discovery: Methodology and Application. Journal of Artificial Intelligence Research **17** (2002) 501–527
2. Lavrac, N., Kavsek, B., Flach, P., Todorovski, L.: Subgroup Discovery with CN2-SD. Journal of Machine Learning Research **5** (2004) 153–188
3. Atzmueller, M., Puppe, F., Buscher, H.P.: Exploiting Background Knowledge for Knowledge-Intensive Subgroup Discovery. In: Proc. 19th Intl. Joint Conference on Artificial Intelligence (IJCAI-05), Edinburgh, Scotland (2005) 647–652
4. Jorge, A.M., Pereira, F., Azevedo, P.J.: Visual interactive subgroup discovery with numerical properties of interest. In Todorovski L, Lavrac N, J.K., ed.: Proceedings of the 9th International Conference on Discovery Science (DS 2006). Volume 4265 of Lecture Notes in Artificial Intelligence., Barcelona, Spain, Springer (October 2006) 301–305 ISI, ISIProc.
5. Wrobel, S.: An Algorithm for Multi-Relational Discovery of Subgroups. In: Proc. 1st European Symposium on Principles of Data Mining and Knowledge Discovery (PKDD-97), Berlin, Springer Verlag (1997) 78–87
6. Aumann, Y., Lindell, Y.: A Statistical Theory for Quantitative Association Rules. Journal of Intelligent Information Systems **20**(3) (2003) 255–283
7. Atzmueller, M., Puppe, F.: SD-Map – A Fast Algorithm for Exhaustive Subgroup Discovery. In: Proc. 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2006). Number 4213 in LNAI, Berlin, Springer Verlag (2006) 6–17
8. Grosskreutz, H., Rüping, S., Wrobel, S.: Tight optimistic estimates for fast subgroup discovery. In: ECML PKDD '08: Proc. 2008 European Conference on Machine Learning and Knowledge Discovery in Databases - Part I, Berlin, Springer Verlag (2008) 440–456
9. Klösgen, W.: Explora: A Multipattern and Multistrategy Discovery Assistant. In Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R., eds.: Advances in Knowledge Discovery and Data Mining. AAAI Press (1996) 249–271
10. Han, J., Pei, J., Yin, Y.: Mining Frequent Patterns Without Candidate Generation. In Chen, W., Naughton, J., Bernstein, P.A., eds.: 2000 ACM SIGMOD Intl. Conference on Management of Data, ACM Press (05 2000) 1–12
11. Klösgen, W.: Applications and Research Problems of Subgroup Mining. In: Proc. 11th International Symposium on Foundations of Intelligent Systems (ISMIS '99), Berlin, Springer Verlag (1999) 1–15
12. Grosskreutz, H., Rüping, S., Shaabani, N., Wrobel, S.: Optimistic estimate pruning strategies for fast exhaustive subgroup discovery. Technical report, Fraunhofer Institute IAIS, http://publica. fraunhofer.de/eprints/urn:nbn:de:0011-n-723406.pdf (2008)
13. Newman, D., Hettich, S., Blake, C., Merz, C.: UCI Repository of Machine Learning Databases, http://www.ics.uci.edu/∼mlearn/mlrepository.html (1998)