Exploiting Background Knowledge for Knowledge-Intensive Subgroup Discovery

Martin Atzmueller and **Frank Puppe**
Department of Computer Science
University of Würzburg
97074 Würzburg, Germany
{atzmueller, puppe}@informatik.uni-wuerzburg.de

Hans-Peter Buscher
DRK-Kliniken Berlin-Köpenick,
Clinic for Internal Medicine II,
12559 Berlin, Germany
buscher.dhp@t-online.de

Abstract

In general, knowledge-intensive data mining methods exploit background knowledge to improve the quality of their results. Then, in knowledge-rich domains often the interestingness of the mined patterns can be increased significantly.

In this paper we categorize several classes of background knowledge for subgroup discovery, and present how the necessary knowledge elements can be modelled. Furthermore, we show how subgroup discovery methods benefit from the utilization of background knowledge, and discuss its application in an incremental process-model. The context of our work is to identify interesting diagnostic patterns to supplement a medical documentation and consultation system. We provide a case study in the medical domain, using a case base from a real-world application.

1 Introduction

Knowledge-intensive learning methods usually exploit background knowledge, because this can improve the quality of their results significantly (c.f. [Richardson and Domingos, 2003]). In this paper we exploit background knowledge for subgroup discovery [Wrobel, 1997; Klösgen, 2002], applied for exploration or descriptive induction, to discover ”interesting” subgroups concerning a certain property of interest. Background knowledge can help to focus the algorithm for subgroup search on the relevant patterns according to the goals of the user, thus reducing uninteresting patterns and restricting the search space. This can increase the quality of the discovered set of subgroups significantly, as well as the efficiency of the search method. It is obvious, that the amount of available background knowledge depends on the particular application domain. For example, in the medical domain usually a lot of background knowledge is available.

Besides (simple) constraints the background knowledge for the presented knowledge-intensive process consists of two categories: a task-specific subset of derived attributes and general ontological background knowledge, that can be refined incrementally. Certain knowledge classes can also be used to derive new (lower-level) knowledge. The knowledge classes can potentially be used in other rule-based learning methods in a straightforward manner. Similar to the applied knowledge, subgroup discovery results can also be formalized incrementally and can be provided as input to the search method. We will introduce this approach in this paper.

Our implementation and evaluation is based on the knowledge-based documentation and consultation system for sonography SonoConsult [Huettig *et al.*, 2004], which is in routine use in the DRK-hospital in Berlin/Köpenick.

The rest of the paper is organized as follows: First we describe the knowledge-intensive process for subgroup discovery. We introduce the subgroup discovery method, present the different types of suitable background knowledge and discuss their characteristics in detail. After that, an experimental evaluation of the impact of the proposed method is demonstrated by a case study in the medical domain. Finally, we conclude the paper with a discussion of the presented work, and we show promising directions for future work.

2 Knowledge-Intensive Subgroup Discovery

In this section, we first give an overview of the subgroup discovery method and define the basic concepts of our knowledge representation schema. Furthermore, we present the knowledge-intensive subgroup discovery process, and describe the background knowledge in detail.

2.1 Subgroup Discovery

Subgroup discovery [Wrobel, 1997; Klösgen, 2002] is a method to discover ”interesting” subgroups of individuals, e.g., ”the subgroup of patients who are treated in a *small, understaffed* hospital are significantly more likely to suffer from complications in the future than the patients in the reference population”. Subgroups are described by relations between independent (explaining) variables and a dependent (target) variable, rated by a certain interestingness measure. For example, two possible criteria are the difference in the distribution of the target variable concerning the subgroup and the general population, and the subgroup size. The main application areas of subgroup discovery are exploration and descriptive induction, to obtain an overview of the relations between a target variable and usually many explaining variables. Subgroup discovery does not necessarily focus on finding complete relations; instead partial relations, i.e., (small) subgroups with ”interesting” characteristics can be sufficient.

Before defining the subgroup discovery task, we first introduce the necessary notions concerning our knowledge representation schema. Let Ω_A the set of all attributes. For each attribute $a \in \Omega_A$ a range $dom(a)$ of values is defined. Furthermore, we assume $\mathcal{V}_A$ to be the (universal) set of attribute values of the form $(a : v)$, where $a \in \Omega_A$ is an attribute and $v \in dom(a)$ is an assignable value.

A subgroup discovery task mainly relies on the following four main properties: the target variable, the subgroup description language, the quality function, and the search strategy. The target variable may be binary, nominal or numeric. Depending on its type, there are different analytic questions, e.g., for a numeric target variable we can search for significant deviations of the mean of the target variable.

A subgroup discovery problem encapsulates the target variable, the search space of independent variables, the general population, and additional constraints.

Definition 1 (Subgroup Discovery Problem)**.** *A subgroup discovery problem SP is defined as the tuple*

$$SP = (T, A, C, CB)\,,$$

where $T \in \Omega_A \cup \mathcal{V}_A$ is a target variable. $A \subseteq \Omega_A$ is the set of attributes to be included in the subgroup discovery process. CB is the case base representing the general population used for subgroup discovery. C specifies (optional) constraints for the discovery method. We define Ω_{SP} as the set of all possible subgroup discovery problems.

The definition above allows for arbitrary target variables. However, for our analytic questions we will focus on binary target variables, i.e., $T \in \mathcal{V}_A$.
The description language specifies the individuals from the reference population belonging to the subgroup.

Definition 2 (Subgroup Description)**.** *A subgroup description $sd = \{e_i\}$ consists of a set of selection expressions (selectors) $e_i = (a_i, V_i)$ which are selections on domains of attributes, i.e., $a_i \in \Omega_A, V_i \subseteq dom(a_i)$. A subgroup description is defined as the conjunction of its contained selection expressions. We define Ω_{sd} as the set of all possible subgroup descriptions.*

A quality function measures the interestingness of the subgroup (c.f., [Klösgen, 2002] for examples).

Definition 3 (Quality Function)**.** *A quality function*

$$q : \Omega_{sd} \times \Omega_{SP} \rightarrow R$$

evaluates a subgroup description $sd \in \Omega_{sd}$ given a subgroup discovery problem $SP \in \Omega_{SP}$. It is used by the search method to rank the discovered subgroups during search.

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

$$q_{BT} = \frac{(p - p_0) \cdot \sqrt{n}}{\sqrt{p_0 \cdot (1 - p_0)}} \cdot \sqrt{\frac{N}{N - n}}\,, \quad q_{RG} = \frac{p - p_0}{p_0 \cdot (1 - p_0)}\,,$$

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, and n denotes the size of the subgroup. In contrast to the quality function q_{BT}, the quality function q_{RG} only compares the target shares of the subgroup and the total population measuring the relative gain. Therefore, a suitable support threshold is necessary to discover significant subgroups.

Considering the subgroup search strategy an efficient search is necessary, since the search space is exponential concerning all the possible selectors of a subgroup description. Commonly, a beam search strategy is used because of its efficiency [Klösgen, 2002]. For our search strategy, we use a modified beam search strategy, where an initial subgroup description can be selected as the initial value for the beam. Beam search adds a selector to the k best subgroup descriptions in each iteration. Iteration stops if the quality as evaluated by the quality function q does not improve any further.

2.2 The Knowledge-Intensive Process Model

Often suitable background knowledge can help both to focus the subgroup discovery task on the relevant patterns concerning the given analysis goals, and to restrict the search space. We propose an incremental approach in which background knowledge can be applied initially at the start, but also during the discovery process. The proposed knowledge-intensive process for subgroup discovery is depicted in Figure 1.

We start with a defined population given as a case base CB and existing background knowledge, if available. For the analysis task defined by a *subgroup discovery problem* the subgroup discovery method generates a set of subgroups. If these are considered as interesting by the user, the results are presented, and the process is finished. Otherwise the subgroups are analyzed, and background knowledge including constraints can be added to the subgroup discovery problem. Additionally, selected subgroups can be used as the basis for further refinement. Then the process continues with a new iteration: all constraints of the current subgroup discovery problem are applied for the search process. In summary, the knowledge-intensive process consists of three main steps:

1. DISCOVER: Discover subgroups – result set $\mathcal{SG}$
2. INSPECT: If $\mathcal{SG}$ is interesting – *STOP* and present the subgroup discovery results; otherwise
3. REFINE: Analyze the discovered subgroups; adapt the subgroup discovery problem, i.e., extend or modify background knowledge and/or use interesting subgroups as a starting point for step 1. *GOTO* step 1

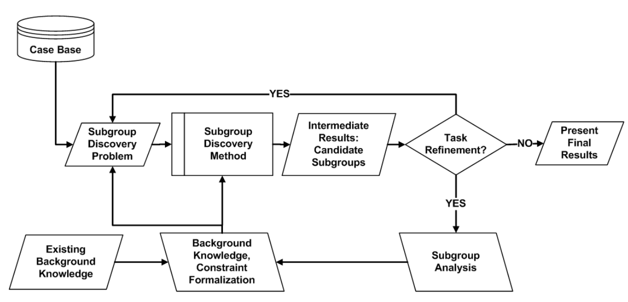

Figure 1: The Knowledge-Intensive Process Model

Step 3 is performed by the domain specialist according to the analysis goals. Examples for the adaptation of the subgroup discovery problem are given in the case study in Section 3.

2.3 Background Knowledge for Subgroup Discovery

We want to make the acquisition of helpful background knowledge as easy as possible. Therefore, we try utilize knowledge known from other knowledge systems, that the domain specialist is already familiar with. There are different classes of background knowledge which can be used in the knowledge-intensive process for subgroup discovery: constraints, ontological knowledge, abstraction knowledge, and pattern knowledge. Knowledge acquisition is always expensive, so its costs should be minimized. Sometimes knowledge can be derived from already formalized knowledge, e.g., we can derive constraints from ontological knowledge, and thus reduce its acquisition costs.

In the following table, we summarize the different classes and types of background knowledge (CK = constraint knowledge, OK = ontological knowledge, PK = pattern knowledge, AK = abstraction knowledge). We show their characteristics in terms of the 'derivable knowledge' if applicable, their costs, and their potential contribution to restricting the search space and/or focusing the search process. Considering the costs and the impact of the knowledge types on the search space, the label - indicates no cost/impact; the labels +, ++, and +++ indicate increasing costs and impact. A +(+) signifies, that the respective element has low costs if it can be derived/learned, and moderate costs otherwise. Similarly ++(+) indicates this for moderate and high costs, respectively.

Knowledge		*Derivable*	*Cost*	Search Space	
Class	*Type*	*Knowledge*		Restr.	Focus
CK	Syntactical Constr.	–	+	+	+
CK	**Quality Constr.**	–	+	++	++
CK	Attr. Values Constr.	–	+(+)	+	+
CK	Meta Values Constr.	–	+(+)	–	++
CK	**Attributes Constr.**	–	+(+)	++	++
OK	**Normality** Info	Attr. Val. Constr.	+	+	++
OK	Abnormality Info	Attr. Val. Constr.	++	+	++
		Meta Val. Constr.		–	++
OK	Similarity Info	Meta Val. Constr.	++(+)	–	++
OK	**Ordinality Info**	Meta Val. Constr.	+	++	+++
OK	Attr. Weights	Attr. Constraints	+(+)	+	++
AK	***Derived Attributes***	Derived Attributes	+++	+++	+++
PK	Subgroup Pattern	Derived Attributes	+(+)	–	+
PK	Priority Groups	–	++	–	+

The most important types of background knowledge with an especially good cost/benefit ratio are indicated in bold type. Derived attributes are a special case with potentially high benefit as discussed below. Then, the need for derived attributes depends on the specific application domain.

Constraint Knowledge Constraints are a class of background knowledge, which is especially simple to apply. The different constraint types can be described as follows:

- Attribute values constraints: the attribute values can be restricted to the set of relevant values.
- Meta values constraints: specific value groups defining an abstracted 'meta value' can be specified, e.g., intervals for ordinal values. Meta values are not restricted to intervals, but can cover any combination of values.
- Constraints for attributes: specific attributes and/or combinations of attributes can be excluded.
- General constraints restricting the syntactical form of the discovered subgroups, or quality constraints for the discovered subgroups can be applied.

Constraints can significantly restrict the search space and focus the search process. The knowledge acquisition cost for constraints is moderate, depending on the number of relations that need to be modeled; the costs can also be decreased utilizing ontological knowledge as described below.

Pattern Knowledge In general, pattern knowledge specifies the kinds of patterns that the user is especially interested in. For example, pattern knowledge can be used to define already known subgroups, that can be directly applied in the discovery process for subgroup refinement. As another kind of pattern knowledge, the user can specify *priority groups* which are partial disjunctive sets of attributes, that have an assigned priority, e.g., depending on the association strength between an attribute and the target variable. Then, we can start with the set of attributes with the highest priority. If the discovered subgroups cannot be improved any further by the attributes in the current set, then we take the attributes in the next prioritized attribute set into account, in the next iteration. The costs for pattern knowledge are not too high, if the patterns are discovered automatically. For priority knowledge, the cost is encoded in the partially disjunctive separation of attributes. Pattern knowledge does not really restrict the search space, but can help to focus the search.

Ontological Knowledge We can utilize ontological knowledge which is commonly used in the development of knowledge systems, e.g., in case-based reasoning systems. The following elements need to be defined by the domain specialist if they cannot be learned (semi-)automatically, e.g., [Baumeister *et al.*, 2002].

- weights of attributes denoting their importance
- similarity information about the relative similarity between attribute values
- abnormality information about attribute values
- ordinality information about attributes

Abnormality/Normality Information If abnormality or normality information about attribute values is available, then each value $v \in dom(a)$ of an attribute a is attached with a label that explains, if v is describing a normal or an abnormal state of the attribute. Normality information only requires a binary label while abnormality information defines several categories. For example, consider the attribute temperature with the value range $dom(\text{temperature}) = \{\text{normal}, \text{marginal}, \text{high}, \text{very high}\}$. The values *normal* and *marginal* denote normal states of the attribute, while the values *high* and *very high* describe abnormal states. Several categories can be defined according to the degree of abnormality. We use five degrees of abnormality given by the symbolic values $A1, A2, A3, A4, A5$. Category $A1$ denotes a normal value; the categories $\{A2, A3, A4, A5\}$ denote abnormal values in ascending order. Using abnormality/normality knowledge we can constrain the value range of an attribute: e.g., either all values corresponding to selected *abnormal* categories or the values marked with the *normal* category can be excluded from the domain of an attribute used in the subgroup discovery method, by a global exclusion condition. Then, attribute values constraints are derived accordingly.

Meta Values Abnormality information and similarity information concerning attribute values can be used to define additional *meta values*: e.g., if the similarity between two attribute values is very high, then they can potentially be analyzed as one value, thus forming a disjunctive selection expression on the value range of an attribute. Likewise, global abnormality groups can be defined by providing groups of abnormality degrees that specify which values should be combined. This is especially relevant in the medical domain with attribute values such as *probable*, *possible*, and *unverifiable*. In diagnostics the value *probable* contributes more evidence to the represented concept than the value *possible*. However, for analysis often the values *probable* and *possible* have an almost equivalent meaning and can be considered as one value.

Ordinality information that specifies if an attribute is ordinal can be used to define meta values relating to ordinal groups: we can generate meta values covering all adjacent combinations of attribute values, or all ascending/descending combinations of values, starting with the minimum or maximum, respectively. If abnormality information is available, then we partition the value range by the given *normal* value and only start with the most extreme value. For example, for the ordinal attribute *liver size* with the values *1:smaller than normal*, *2:normal*, *3:marginally increased*, *4:slightly increased*, *5:moderately increased* and *6:highly increased*, we partition by the *normal* value (2) and obtain the following meta values: (1), (3, 4, 5, 6), (4, 5, 6), (5, 6) and (6).

We summarize how constraints can be derived from ontological knowledge: The set of relevant attributes can be constrained using attribute weights, with a suitable threshold. Given similarity and abnormality/normality information about attribute values we can model/restrict the value ranges of attributes. Ordinality information about attribute values can be easily used to construct ordinal grouped meta values, which are often more meaningful for the domain specialist. Then, the original values can be replaced, optionally.

Abstraction Knowledge This class of background knowledge is given by derived attributes that are inferred from basic attributes or other derived attributes. Derived attributes often correspond to certain known dependencies between attributes, or known intermediate concepts, that are not stored as basic attributes. For example, in the medical domain, we can infer the derived attribute *body mass index*, given the attributes *height* and *weight*. Additionally, if there are a lot of basic attributes in the case base, (known) multi-correlations between basic attributes can cause unstructured subgroup discovery results. Then, data abstraction can increase the interpretability of the subgroup discovery results significantly. Simple concepts can be aggregated to intermediate concepts to form more potentially meaningful and interesting selectors. A nominal derived attribute $a \in \Omega_A$ is defined using abstraction rules, which are utilized to derive the values $v_{i_a} \in \mathcal{V}_A$ concerning attribute a. A rule of the form $r_{v_a} = cond(r_{v_a}) \rightarrow v_a$ is used for a value v_a of attribute a, where the rule condition $cond(r_{v_a})$ contains conjunctions and/or disjunctions of (negated) attribute values $v_i \in \mathcal{V}_A$. Furthermore, derived attributes with a numerical value range can be defined by algebraic formula expressions.

Improving the Handling of Missing Values Considering the quality functions abstraction knowledge contributes to one major point – handling *missing values*. Missing values in cases are a significant problem for subgroup discovery and machine learning in general. For example, in medical case bases for a specific patient usually only a subset of the possible examinations is performed, given a structured data gathering strategy in real-world applications. Then, only the relevant questions for the diagnostic tasks are presented to the user. This results in reduced costs for the examiner, however then a specific instance of the data set concerning the basic attributes may be quite sparse.

There exist several strategies for dealing with missing values: a common strategy [Tsumoto, 2002] removes objects (cases) with missing values from the set of analyzed objects. Other strategies try to fill in the missing values according to statistical evaluations, or try to model the distribution [Ragel and Créwmilleux, 1999]. The subgroup quality functions basically perform a kind of statistical hypothesis testing given a subgroup description, the target variable and the general population. For such a test only the cases of the population can be considered in which all variables have defined values. The power of the test is decreased significantly, if many analysis objects are removed due to missing values.

In general, we cannot simply apply the ”closed-world assumption”, i.e., that missing values of a concept indicate the non-existence or negation of the concept. For example, in the medical domain, a diagnosis may be missing, because either all its relevant observations are missing or they are known but denote the normal, i.e., the non-pathological state. Consequently the diagnosis is not inferred. If we construct a derived attribute to indicate the cases when the relevant observations are missing, then we can use this derived ”helper” attribute to indicate when the diagnosis really has a missing value. Additionally, derived attributes besides the described ”helper” attributes can also be constructed accordingly to minimize missing values themselves, such that a certain default value is provided which denotes the normal category. So, the derived attributes serve three purposes:

- They focus the subgroup discovery method on the relevant analysis objects,
- They decrease multi-correlations between attributes that are not interesting,
- Derived attributes can reduce missing values for a given concept, since they can be constructed such that a defined value is more often computed if the respective concept would have a missing value otherwise.

The derived attributes can either be constructed based on expert knowledge, or on specific discovered subgroups. A subgroup description is a set of selectors for a specific target concept that are highly correlated with the concept. If the selectors can be abstracted into a derived attribute, then it can be used as potential background knowledge as well. Furthermore, derived attributes can potentially be refined according to the subgroup discovery results, by specialization or generalization. Abstraction knowledge is probably the most costly class of background knowledge in the knowledge-intensive process. If the abstractions are not based on discovery results, then they have to be formalized manually by the expert.

2.4 Related Work

Using background knowledge to constrain the search space and pruning hypotheses during the search process has been proposed in ILP approaches. [Weber, 2000] proposes *require-* and *exclude*-constraints for attribute – value pairs, in order to prune the search space. [Zelezny *et al.*, 2003] integrate constraints into an ILP approach as well; the used constraints are mainly concerned with syntactical restrictions and constraints relating to the quality of the discovered subgroups.

The main difference between the presented approach and the existing approaches is the fact, that we are able to integrate several new types of additional background knowledge. This additional background knowledge can be refined incrementally according to the requirements of the discovery task, and can additionally be used to infer new background knowledge on the fly, e.g., constraints. As a major point we apply special abstraction knowledge, which can be defined by the expert, or can be constructed semi-automatically using the subgroup discovery results. This type of knowledge can be applied dynamically in the process and does not rely on a static data-preprocessing and cleaning task, for example.

3 Case Study

In this section we describe a case study for the application of the knowledge-intensive process. We use cases taken from the SONOCONSULT system [Huettig *et al.*, 2004] – a medical documentation and consultation system for sonography – which has been developed with the knowledge system D3 [Puppe, 1998]. The system is in routine use in the DRK-hospital in Berlin/Köpenick and documents an average of about 300 cases per month. These are detailed descriptions of findings of the examination(s), together with the inferred diagnoses (binary attributes). The derived diagnoses are usually correct as shown in a medical evaluation (c.f. [Huettig *et al.*, 2004]), resulting in a high-quality case base with detailed case descriptions. The applied SONOCONSULT case base contains 4358 cases. The domain ontology contains 427 basic attributes with about 5 symbolic values on average, 133 symptom interpretations, which are rule-based abstractions of the basic attributes, and 221 diagnoses. This indicates the potentially huge search space for subgroup discovery.

Subgroup discovery was performed using the [VIKAMINE, 2005] (*Visual, Interactive and Knowledge-Intensive Analysis and Mining Environment*) system, developed at the Department of Computer Science VI of the University of Würzburg. We used beam search with a beam size of 10 as the search strategy, and the quality function q_{RG} as defined in Section 2.1. The discovered subgroups were evaluated by domain specialists according to (clinical) novelty, interestingness, and actionability aspects.

First, we performed subgroup discovery only using basic attributes and general background knowledge. We used attribute weights for feature subset selection. The subgroup discovery algorithm presented many significant subgroups, that supported the validity of the subgroup discovery techniques. However the results at this stage were not really novel, since they indicated mostly already known dependencies, e.g., the relation between *relative body weight* and *body-mass index*.

In the next stage, the expert decided to define new attributes, i.e., abstracted attributes which described interesting concepts for analysis. The expert provided 45 derived attributes, that consist of symptom interpretations directly indicating a diagnosis and intermediate concepts which are used in clinical practice, for example *pleura-effusion*, *portal hypertension*, or *pathological gallbladder status*. Furthermore, constraints were formalized, that prevented the combination of certain attributes, to restrict the search space. Some examples are depicted in the Table 1, where the dependent and independent variables are directly related to each other. The

Target Variable	*Independent Variable*
Chronic Pancreatitis	Pancreas Disease
Pancreas Disease	Carcinoma of the Pancreas
Body Mass Index	Relative Body Weight

Table 1: Examples of known/uninteresting relations

newly defined abstraction knowledge was applied extending the search space to the expert-defined attributes. For each attribute a in the set of derived attributes and each value $v_i \in dom(a)$, a subgroup discovery problem $SP_{ai} \in \Omega_{SP}$ was generated using the binary target variable $(a = v_i)$.

The impact of the added background knowledge was proven by a greater acceptance of the subgroup discovery results by the expert. However still too many subgroups were not interesting for the expert, because too many *normal* values were included in the results, e.g., *liver-size* = *normal*, or *fatty liver* = *unlikely*. This motivated the application of abnormality information to constrain the value space to the set of *abnormal* values of the attributes. Additionally, the expert suggested to group sets of values into disjunctive value sets defined by abnormality groups, e.g., grouping the values *possible* and *probable* for some attributes. Furthermore, ordinality information was applied to construct meta values of ordinal attributes like *age* or *liver size*.

Target Variable	*Def. Pop.*	*Subgroup Description*
Aorta-sclerosis = calcified	1418	Pancreas Disease={probable;possible}
Aorta-sclerosis = calcified	75	Pancreas Disease={probable;possible} AND Ascites = present

Table 2: Example: Missing Value Problem

Further investigation showed that missing values play a central role in the discovery process. Sometimes the defined population significantly decreased, when adding a selection expression to a subgroup description, compared to the parent subgroup. An example is given in Table 2; the defined population is indicated in the second column. In this example, the attribute *ascites* has many missing values. This problem was solved by adapting the derived attributes to indicate when a missing value corresponds to the value ”disease not present”. Then, the final set of interesting subgroups was obtained.

Figure 2 shows exemplary discovered subgroups concerning the target variable *gallstones*, that were considered as interesting for clinical practice (lines 1-6). All these subgroups were at least significant at the 10^{-6} level. The individual subgroups are shown in the rows of the table. Subgroup parameters given in the columns are: (Subgroup) Size, TP (true positives), FP (false positives), Pop. (defined population size), RG (relative gain) and the value of the binomial quality function q_{BT} (Bin. QF), c.f., Section 2.1. The domain specialists evaluated the discovered subgroups concerning interestingness for clinical practice: for them, the relative gain was the primary criterion to rank the subgroups in a first step. Then, as a combined (helper) measure for subgroup size and gain quality, the value of the binomial quality function q_{BT} was used for post-processing the set of subgroups.

Target Variable: Gallstones																					
#	Age			Sex		Liver size						Aorta sclerosis									
	1	2	3	m	f	1	2	3	4	5	6	n	c	Size	TP	FP	Pop.	p0	p	RG	Bin. QF
1			X		X				X	X	X		X	89	37	52	3171	0.172	0.416	1.71	6.17
2			X		X				X	X	X	X	X	119	46	73	3171	0.172	0.387	1.5	6.31
3		X	X		X				X	X	X		X	132	51	81	3171	0.172	0.386	1.5	6.66
4					X			X	X	X	X		X	190	68	122	3177	0.172	0.358	1.3	6.99
5			X		X				X	X	X			207	72	135	3171	0.172	0.348	1.23	6.92
6		X	X		X					X	X	X	X	64	22	42	3171	0.172	0.344	1.2	3.67
7			X											1651	414	1237	3743	0.177	0.251	0.51	10.57
8												X	X	1334	310	1024	3749	0.177	0.232	0.38	6.66
9					X									1776	408	1368	3749	0.177	0.23	0.37	8.1
10									X	X	X			894	178	716	3177	0.172	0.199	0.19	2.52

Age:
1 = <50
2 = 50-69
3 = >=70

Sex:
m = male
f = female

Liver size:
1 = smaller than normal
2 = normal
3 = marginally increased
4 = slightly increased
5 = moderately increased
6 = highly increased

Aorta sclerosis:
n = not calcified
c = calcified

Figure 2: Examples of discovered subgroups (clinically interesting: lines 1-6): e.g., the first line depicts the subgroup (89 cases) described by *Age* ≥ 70 *AND Sex=female AND Liver size=*$\{$*slightly or moderately or highly increased*$\}$ *and Aorta sclerosis=calcified* with a target share (gallstones) of 41.6% (p) compared to 17.2% (p_0) in the general population, with a relative gain of 171% (RG).

Especially interesting for the expert proved the situations, when a specialization of a subgroup significantly improved its quality (#5 vs. #2). Additionally, an important criterion to determine sound subgroups was the comparison of possible value combinations for ordinal and nominal attributes (e.g., #6 vs. #2 or #3 vs. #2).

For clinical practice, the expert preferred small subgroup descriptions if the subgroups were comparable concerning the relative gain measure. This is in line with the heuristic of preferring simpler knowledge for actionability. However, significant improvements in a subgroup specialization countered the increase in the length of the subgroup description, e.g., #7, #9, #10 vs. #1. Furthermore, the baseline qualities of known factors (e.g., age, gender) (c.f., #6 to #10) were also very important for the domain specialist as a reference for subgroup comparison. In summary, the interpretation and judgment of the subgroup discovery results ultimately depends on the expert: even if a suitable quality function is applied, the expert still needs to semantically interpret the subgroup descriptions for the final assessment of a subgroup.

4 Summary and Future Work

In this paper we presented how exploiting background knowledge can help to improve subgroup discovery results in a knowledge-intensive approach. We described several classes of background knowledge in detail, that can potentially be used in other rule-based learning methods besides subgroup discovery as well. A case study using cases from a real-world application showed that applying background knowledge helped to focus the discovery algorithm on the interesting subspace of subgroup hypotheses, increasing the quality of the discovery results. Furthermore, we discussed how applying abstraction knowledge can help to handle the problem of missing values.

In the future we are planning to consider appropriate quality measures concerning the simplicity of the discovered subgroups. Primary work for learning rule bases was presented e.g., in [Atzmueller *et al.*, 2004]. We will especially focus on quality measures which are easy to interpret and tunable to the analysis goals.

References

[Atzmueller *et al.*, 2004] Martin Atzmueller, Joachim Baumeister, and Frank Puppe. Quality Measures for Semi-Automatic Learning of Simple Diagnostic Rule Bases. In *Proc. 15th Intl. Conference on Applications of Declarative Programming and Knowledge Management (INAP 2004)*, pages 203–213, Potsdam, Germany, 2004.

[Baumeister *et al.*, 2002] Joachim Baumeister, Martin Atzmueller, and Frank Puppe. Inductive Learning for Case-Based Diagnosis with Multiple Faults. In *Advances in Case-Based Reasoning*, volume 2416 of *LNAI*, pages 28–42, Berlin, 2002. Springer Verlag.

[Huettig *et al.*, 2004] Matthias Huettig, Georg Buscher, Thomas Menzel, Wolfgang Scheppach, Frank Puppe, and Hans-Peter Buscher. A Diagnostic Expert System for Structured Reports, Quality Assessment, and Training of Residents in Sonography. *Medizinische Klinik*, 99(3):117–122, 2004.

[Klösgen, 2002] Willi Klösgen. *Handbook of Data Mining and Knowledge Discovery*, chapter 16.3: Subgroup Discovery. Oxford University Press, New York, 2002.

[Puppe, 1998] Frank Puppe. Knowledge Reuse among Diagnostic Problem-Solving Methods in the Shell-Kit D3. *Intl. Journal of Human-Computer Studies*, 49:627–649, 1998.

[Ragel and Créwmilleux, 1999] Arnaud Ragel and Bruno Créwmilleux. MVC–a Preprocessing Method to Deal with Missing Values. *Knowledge-Based Systems*, 12(5–6):285–291, October 1999.

[Richardson and Domingos, 2003] Matthew Richardson and Pedro Domingos. Learning with Knowledge from Multiple Experts. In *Proc. 20th International Conference on Machine Learning (ICML-2003)*, 2003.

[Tsumoto, 2002] Shusaku Tsumoto. *Handbook of Data Mining and Knowledge Discovery*, chapter Chapter 43: Medicine. Oxford University Press, New York, 2002.

[VIKAMINE, 2005] http://ki.informatik.uni-wuerzburg.de/projects/dm, 2005.

[Weber, 2000] Irene Weber. Levelwise search and Pruning Strategies for First-Order Hypothesis Spaces. *Journal of Intelligent Information Systems*, 14:217–239, 2000.

[Wrobel, 1997] Stefan Wrobel. An Algorithm for Multi-Relational Discovery of Subgroups. In *Proc. 1st European Symposion on Principles of Data Mining and Knowledge Discovery*, pages 78–87, Berlin, 1997. Springer Verlag.

[Zelezny *et al.*, 2003] Filip Zelezny, Nada Lavrac, and Saso Dzeroski. Using Constraints in Relational Subgroup Discovery. In *International Conference on Methodology and Statistics*, pages 78–81. University of Ljubljana, 2003.