Advanced Review

Subgroup discovery

Martin Atzmueller*

Subgroup discovery is a broadly applicable descriptive data mining technique for identifying *interesting* subgroups according to some property of interest. This article summarizes fundamentals of subgroup discovery, before that it also reviews algorithms and further advanced methodological issues. In addition, we briefly discuss tools and applications of subgroup discovery approaches. In that context, we also discuss experiences and lessons learned and outline some of the future directions in order to show the advantages and benefits of subgroup discovery. © 2015 John Wiley & Sons, Ltd.

How to cite this article:
WIREs Data Mining Knowl Discov 2015, 5:35–49. doi: 10.1002/widm.1144

INTRODUCTION

Subgroup discovery[1–5] has been established as a general and broadly applicable technique for descriptive and exploratory data mining. It aims at identifying descriptions of subsets of a dataset that show an interesting behavior with respect to certain interestingness criteria, formalized by a quality function, e.g., Ref 2. This article summarizes fundamental concepts of subgroup before it provides an advanced review on algorithms, methodological issues, and applications. Overall, subgroup discovery and analytics are important tools for descriptive data mining. They can be applied, for example, for obtaining an overview on the relations in the data, for automatic hypotheses generation, and for data exploration. Prominent application examples include knowledge discovery in medical and technical domains, e.g., Refs 3, 4, 6, 7. Typically, the discovered patterns are especially easy to interpret by the users and domain experts, cf., Refs 6, 8. Standard subgroup discovery approaches commonly focus on a *single* target concept as the property of interest,[1,3,8] while the quality function framework also enables *multi-target concepts*, e.g., Refs 9, 10. Furthermore, more complex target properties[11,12] can be formalized as *exceptional models*, cf., Ref 12.

The remainder of this article is organized as follows. First, we present fundamentals of subgroup discovery. After that, we discuss state of the art algorithms. In the following sections, we outline methods for subgroup set selection, discuss applications and experiences, and provide an outlook on future directions and challenges. Finally, we conclude with a summary.

FUNDAMENTALS OF SUBGROUP DISCOVERY

Subgroup discovery[1,2,3–5,9] has been established as a versatile and effective method in descriptive and exploratory data mining. Similar to other methods for mining supervised local patterns, e.g., discriminative patterns,[13] contrast sets,[14] and emerging patterns,[15] subgroup discovery aims at identifying *interesting* groups of individuals, where 'interestingness is defined as distributional unusualness with respect to a certain property of interest'.[2] Subgroup discovery has been well investigated concerning binary and nominal target concepts, i.e., properties of interest with a finite number of possible values.[1,2,16] Furthermore, numeric target concepts have received increasing attention in subgroup discovery recently, and several approaches for using numeric attributes have been proposed, e.g., Refs 7, 17, 18, 19. In the scope of this paper, we will adopt a broad definition of subgroup discovery, including single binary, nominal, and numeric target variables, but also extending to multi-target concepts, and to *exceptional model mining*,[11,12] as a variant of subgroup discovery that especially focuses on complex target properties.

In the remainder of this section, we first summarize the idea of local exceptionality detection

*Correspondence to: atzmueller@cs.uni-kassel.de

Knowledge and Data Engineering Group, University of Kassel, Kassel, Germany

Conflict of interest: The author has declared no conflicts of interest for this article.

employed by subgroup discovery. After that, we provide some necessary definitions and notation, before we formally tackle quality functions, optimistic estimates, and *top-k* pruning strategies.

Local Exceptionality Detection

Subgroup discovery is based on the idea of *local exceptionality detection*, i.e., how locally exceptional, relevant, and thus interesting patterns can be detected, so-called nuggets in the data (cf., Ref 1). Local pattern mining, e.g., Refs 20, 21 aims to detect such locally interesting patterns, in contrast to global models. Related approaches include, for example, frequent pattern mining,[22] mining association rules,[23,24] and closed representations.[25,26] In contrast to those, however, subgroup discovery allows for a flexible definition of the applied quality function or interestingness measure, respectively. Therefore, many of the mentioned techniques can be captured in a subgroup discovery setting, e.g., Ref 27 for a description-oriented community discovery method.

As sketched above, the exceptionality of a pattern is measured by a certain quality function. According to the type of the property of the subgroup, that we are interested in, we can distinguish between simple concepts such as a minimal frequency/size of the subgroup (also known as support for association rules), a deviating target share (confidence) of a binary target property of the subgroup, a significantly different subgroup mean of a numeric target concept, or more complex models, e.g., based on several target attributes for which their distribution significantly differs comparing the subgroup and the whole dataset. Using a quality function, a set of subgroups is then identified using a given subgroup discovery algorithm, e.g., using a heuristic or exhaustive search[1–3,16] strategy, or a direct sampling approach.[28] Typically, the top-k subgroups, or those above a minimal quality threshold are determined.

Basic Definitions

Formally, a *database* $D=(I, A)$ is given by a set of individuals I and a set of attributes A. For nominal attributes, a *selector* or *basic pattern* $(a_i = v_j)$ is a Boolean function $I \rightarrow \{0, 1\}$, that is true if the value of attribute $a_i \in A$ is equal to v_j for the respective individual. For a numeric attribute a_{num} selectors $(a_{\text{num}} \in [\min_j; \max_j])$ can be defined analogously for each interval $[\min_j; \max_j]$ in the domain of a_{num}. The Boolean function is then set to true if the value of attribute a_{num} is within the respective range. The set of all basic patterns is denoted by Σ.

A subgroup is described using a description language, cf., Ref 2 typically consisting of attribute–value pairs, for example, in conjunctive or disjunctive normal form. Below, we present an exemplary conjunctive pattern description language; internal disjunctions can also be generated by appropriate attribute–value construction methods, if necessary, cf., Ref 16. A *subgroup description* or (complex) *pattern P* is then given by a set of basic patterns $P = \{\text{sel}_1, \ldots, \text{sel}_l\}$, $\text{sel}_i \in \Sigma$, $i = 1, \ldots, l$, which is interpreted as a conjunction, that is, $P(I) = \text{sel}_1 \wedge \ldots \wedge \text{sel}_l$, with $\text{length}(P) = l$. A pattern can thus also be interpreted as the *body* of a *rule*. The rule *head* then depends on the property of interest. A *subgroup* $S_P := \text{ext}(P) := \{i \in I | P(i) = \text{true}\}$, i.e., a *pattern cover* is the set of all individuals that are covered by the subgroup description P.

The set of all possible subgroup description, and thus the possible search space is then given by 2^{Σ}, that is, all combinations of the basic patterns contained in Σ. In this context, the pattern $P = \emptyset$ covers all instances contained in the database.

Quality Functions

In general, quality and interestingness measures can be grouped into two categories: *Objective* and *subjective* measures.[29,30] Typically, a quality measure is determined according to the requirements and objectives of the analysis. Then also combinations of objective and subjective measures into hybrid quality measures are usually considered, cf., Ref 31 for rules.

Common subjective interestingness measures are understandability, unexpectedness (new knowledge or knowledge contradicting existing knowledge), interestingness templates (describing classes of interesting patterns), and actionability (patterns which can be applied by the user to his or her advantage).[32] Objective measures are data driven and are derived using structure and properties of the data, for example, based on statistical tests. In the following, we focus on such objective interestingness measures formalized by quality functions.

A *quality function*

$$q : 2^{\Sigma} \rightarrow \mathbb{R}$$

maps every pattern in the search space to a real number that reflects the interestingness of a pattern (or the pattern cover, respectively). The result of a subgroup discovery task is the set of k subgroup descriptions $\text{res}_1, \ldots, \text{res}_k$ with the highest interestingness according to the selected quality function.

In the binary, nominal and numeric setting a large number of quality functions has been proposed in literature, cf., Refs 1, 33. In general, quality functions utilize the statistical distribution of the target concept(s) to score a subgroup pattern P. More complex quality functions compare a set of distributions, for example, by utilizing the concept of exceptional models[12] discussed below. We can consider, for example, a pair of variables, or a whole set of variables arranged in a Bayesian network.

In addition to testing the (statistical) validity of the patterns, the (syntactical) complexity and simplicity of a pattern can also be considered. Commonly, simpler patterns are easier to understand and to interpret.[34] Then, Atzmueller et al.[31] describes a combination of quality measures for rules concerning the validity, that is, the accuracy and the simplicity of the contained patterns. Furthermore, cost-based quality functions, e.g., Ref 35 and cost-sensitive approaches[36] allow the modeling of costs for the quality assessment.

Binary and Nominal Target Quality Functions

Most of the quality functions for binary target concepts are based on the parameters contained in a four-fold table, e.g., Ref 1 covering the positive/negative instances for the pattern P, its complement, and the general population, respectively. Many of the quality measures proposed in Ref 1 trade-off the size $n = |\mathrm{ext}(P)|$ of a subgroup and the deviation $t_P - t_0$, where t_P is the average value of a given target concept in the subgroup identified by the pattern P and t_0 the average value of the target concept in the general population. For binary (and nominal value) target concepts this relates to the *share* of the target concept in the subgroup and the general population. Thus, typical quality functions are of the form

$$q_S^a(P) = n^a \cdot (t_P - t_0), a \in [0; 1]. \tag{1}$$

For binary target concepts, this includes for example the *weighted relative accuracy* (q_S^1) for the size parameter $a = 1$, a simplified binomial function ($q_S^{0.5}$), for $a = 0.5$, or the *added value* function (q_S^0) for $a = 0$, which is order-equivalent to the lift[19] and the relative gain quality[4] function. Further examples for quality functions are given by the binominal test quality function q_B and the χ^2 quality function q_C:

$$q_B = \frac{(t_P - t_0) \cdot \sqrt{n}}{\sqrt{t_0 \cdot (1 - t_0)}} \cdot \sqrt{\frac{N}{N - n}},$$

$$q_C = \frac{n}{N - n} \cdot (t_P - t_0)^2,$$

where $N = |D|$ denotes the size of the database (general population).

Nominal valued target concepts (given by basic patterns) can be analyzed as in the binary case (one versus all). For nominal attributes for which the set of the different nominal values needs to be analyzed, the binary case can be generalized analogously to multi-class settings, such that the whole distribution of the different values is assessed, cf., Refs 1, 37. As an example, the quality function q_S^a can be generalized as follows for the general nominal setting:

$$q_M^a(P) = n^a \cdot \sum_{v_i} \left(t_P^{v_i} - t_0^{v_i}\right)^2, \ a \in [0; 1], \tag{2}$$

where $t_P^{v_i}$ and $t_0^{v_i}$ denote the target shares in the subgroup and the general population, respectively, for each of the respective nominal values v_i contained in the value domain of the nominal target concept.

Other alternatives to the quality functions presented above include, for example, functions adapted from the area of *association rules*, e.g., Ref 23 concerning the support and confidence parameters, as well as adaptations of measures from *information retrieval*, for example, precision and recall and their combination in the F-measure, cf., Ref 38. For more details, we refer to, e.g., Ref 8 which provides a broad overview on quality functions used for subgroup discovery. In principle, many measures for *subgroup analysis* in epidemiology can also be utilized for subgroup discovery, especially in the medical domain. For example, the Odds Ratio function, sensitivity, specificity, significancy, false alarm rate etc., see e.g., Refs 3, 39, 40 for a survey and discussion. Furthermore,[41] provide an in-depth discussion for using the odds ratio and define statistically non-redundant subgroups utilizing the error bounds of the odds ratio measure.

Numeric Target Quality Functions

Quality functions for numeric target concepts, i.e., numeric attributes can be formalized by slightly adapting the quality functions q_a for binary targets presented above, cf., Ref 1. The target shares t_P, t_0 of the subgroup and the general population are replaced by the mean values of the target variable m_P, m_0, respectively.

For the analog to the quality function q_S^a this results in:

$$q_M^a(P) = n^a (m_P - m_0), \ a \in [0; 1], \tag{3}$$

It is easy to see that this function includes the binary formalization as a special case when we set $m = 1$ if the Boolean target concept is *true* and $m = 0$, if it is *false*.

Using the parameter a, Eq. (3) can be utilized for formalizing (order-) equivalent functions for several typically applied quality functions:

- The *mean gain* function q_M^0 ranks subgroups by the respective means m_P (order-equivalently) of the target concept, without considering the size of the subgroup. Therefore, a suitable minimal subgroup size threshold is usually required.

- Another simple example is given by the *mean test*[1,42] with $q_M^{0.5}$. Furthermore, $q_M^{0.5}$ is also order-equivalent to the *z-score* quality function,[43] given by $q_M^{0.5} \cdot \sigma_0$, where σ_0 is the standard deviation in the total population.
- Analogously to the weighted relative accuracy, the *impact* quality function[44] is given by q_M^1.

Further quality functions consider, for example, the median[43] or the variance[45] of the target concepts in the subgroup. For more details on quality functions based on statistical tests (e.g., Student *t*-test or Mann–Whitney *U*-test) we refer to Refs 1, 43.

Multi-Target Quality Functions

For multi-target quality functions, we consider functions that take into account a set of target concepts, e.g., Ref 9. It is possible to extend single-target quality functions accordingly, for example, by extending an univariate statistical test to the multivariate case, e.g., Ref 10. We then need to compare the multivariate distributions of a subgroup and the general population in order to identify interesting (and exceptional) patterns. For comparing multivariate means, for example, for a set of m numeric attributes T_M, with $m = |T_M|$ we can make use of Hotelling's T-squared test,[46] for the quality measure q_H:

$$q_H = \frac{n(n-m)}{m(n-1)} \left(\mu_P^{T_M} - \mu_0^{T_M}\right)^\tau \mathrm{CV}_P^{T_M\,-1} \left(\mu_P^{T_M} - \mu_0^{T_M}\right),$$

where $\mu_P^{T_M}$ is the vector of the means of the model attribute in the subgroup S_P, $CV_P^{T_M}$ is the respective covariance matrix, and $\mu_0^{T_M}$ is the vector of the means of the (numeric) target concepts in D.

As another option for a disjunction of target concepts,[47] propose convex quality functions for discovering cluster groups. Their approach does not use a single target concept, but allows for a disjunction of several target concepts/variables.

A more general framework for multi-target quality functions is given by *exceptional model mining*.[12] It tries to identify interesting patterns with respect to a local model derived from *a set* of attributes. The interestingness can be defined, for example, by a significant deviation from a model that is derived from the total population or the respective complement set of instances within the population. In general, a model consists of a specific *model class* and *model parameters* which depend on the values of the model attributes in the instances of the respective pattern cover. The quality measure q then determines the interestingness of a pattern according to its model parameters. Following Lemmerich et al.[48], we outline some examples below, and refer to Ref 12 for a detailed description.

- A simple example for an exceptionality measure for a set of attributes considers the task of identifying subgroups in which the correlation between two numeric attributes is especially strong, for example, as measured by the Pearson correlation coefficient. This *correlation model class* has exactly one parameter, that is, the correlation coefficient.
- Furthermore, using a *simple linear regression model*, we can compare the slopes of the regression lines of the subgroup to the general population or the subgroups' complement. This *simple linear regression model* shows the dependency between two numeric variables x and y. It is built by fitting a straight line in the two dimensional space by minimizing the squared residuals e_j of the model:

  $$y_i = a + b \cdot x_i + e_j$$

  As proposed in Ref 12, the slope $b = \mathrm{cov}(x, y)/\mathrm{var}(x)$ computed given the covariance $\mathrm{cov}(x, y)$ of x and y, and the variance $\mathrm{var}(x)$ of x can then be used for identifying interesting patterns.
- The *logistic regression model* is used for the classification of a binary target attribute $y \in T$ from a set of independent binary attributes $x_j \in T \backslash y, j = 1, \ldots, |T| - 1$.
  The model is given by:

  $$y = \frac{1}{1 + e^{-z}}, \quad z = b_0 + \sum_j b_j x_j.$$

  Interesting patterns are then those, for example, for which the model parameters b_j differ significantly from those derived from the total population.
- Another example is given by a *Bayesian network* as a rather complex target model. Then, a quality function for assessing the differences between two Bayesian networks can be defined. As proposed in Ref 11, for example, models can then be compared based on the edit distance.[49] Then networks induced by a subgroup and the general population (or the subgroups' complement), respectively, can be analyzed for identifying interesting patterns.

Top-*k* Pruning

As the result of subgroup discovery, the applied subgroup discovery algorithm can return a result set containing those subgroups above a certain minimal quality threshold, or only the top-*k* subgroups, that can then also be postprocessed further. While both options have their relevance depending on the analysis goals, the top-*k* approach provides more flexibility for applying different pruning options in the subgroup discovery process.

Basically, in a top-*k* setting, the set of the top-*k* subgroups is determined according to a given quality function. Then different pruning strategies can be applied for restricting the search space of a subgroup discovery algorithm. A simple option is given by *minimal support pruning* based on the antimonotone constraint of the subgroup size analogously to the Apriori algorithm for mining association rules, cf., Ref 23. Furthermore, properties of certain quality functions enable more powerful approaches.

For several quality functions, for example, *optimistic estimates*[19,50] can be applied for determining upper quality bounds. Consider the search for the k best subgroups. If it can be proven, that no subset of the currently investigated hypothesis is interesting enough to be included in the result set of k subgroups, then we can skip the evaluation of any subsets of this hypothesis, but can still guarantee the optimality of the result.

Another pruning mechanism is given by *generalization-aware pruning*,[51] such that the quality of a subgroup is estimated against the qualities of its generalizations. Below, we discuss these two options in more detail.

Optimistic Estimate Pruning

The basic principle of *optimistic estimates* is to safely prune parts of the search space, for example, as proposed in Ref 2 for binary target variables. 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 $P' \supset P \Rightarrow oe(P) > qf(P')$, that is, that no refinement P' of the pattern P can exceed the quality $oe(P)$.

An optimistic estimate is considered *tight* if there is a subset $S' \subseteq D$, such that $oe(P) = q(P'_S)$. While this definition requires the existence of a subset S', there is not necessarily a pattern P'_S that describes S', cf., Ref 50.

For several quality functions, including many for the binary and numeric case described above, there exist (tight) optimistic estimates that can be applied for pruning. For a detailed overview, we refer to Refs 19, 50.

Generalization-Aware Pruning

In general, a pattern can be compared to its generalizations, for example, in order to fulfill a minimal improvement constraint,[52] such that subgroups with a lower target share than any of its generalizations are removed. This can analogously be applied for mean-based numeric target measures,[45] such that the mean observed in a subgroup needs to deviate significantly from the mean values induced by its generalizations. Accordingly, Lemmerich et al.[51] propose a pruning technique utilizing generalization-aware techniques, and present a set of optimistic estimates in this setting. These estimates take into account subgroup covers of the generalizations of a pattern, and allow for a rather efficient pruning approach for generalization-aware techniques.

ALGORITHMS FOR SUBGROUP DISCOVERY

Many algorithms for subgroup discovery focus on binary or nominal target attributes, for which heuristic (cf., Refs 1, 6, 53) as well as exhaustive methods (cf., Refs 1, 2, 16, 19, 54) are applied. Algorithms for multi-relational data include those by the *MIDOS*[2] and *SubgroupMiner*[55] system, for which the latter also includes analysis options for spatiotemporal data. Essentially, heuristic approaches like beam search trade agility for completeness and are often applied in certain domains for which exhaustive methods take rather long to explore the complete search space, for example, in dense numerical data for which numeric features are discretized on the fly such that numeric subspaces can be explored heuristically, cf., Ref 56. However, due to efficient pruning techniques exhaustive methods can both achieve sufficiently good runtimes and guarantee completeness even in complex domains like ubiquitous social data, e.g., Ref 57.

Furthermore, for the multi-target concept setting including exceptional model mining, there exist heuristic[58,59] approaches as well efficient exhaustive discovery algorithms.[48] As another option, sampling (e.g., Refs 2, 60, 61) can be used for estimating the quality of a subgroup on a (potentially significantly) smaller subset of the case base. The sequential sampling algorithm *GSS*,[62] for example, discovers the k best subgroups according to a given confidence level, for quality functions that can be estimated with bounded error. Also, local pattern sampling methods, e.g., Ref 28 can be seen as an alternative to exhaustive mining approaches utilizing direct sampling procedures with confidence bounds.

Further subgroup discovery approaches apply evolutionary techniques, i.e., genetic algorithms, e.g., Refs 63–66 for their discovery and refinement strategy. For removing uninteresting subgroups, expectation-driven measures, e.g., Ref 67 can be considered such that (expected) interactions between variables are captured.

In general, exhaustive algorithms typically make use of the proposed pruning options for an efficient processing of the search space. Combining intelligent sampling approaches with fast exhaustive methods, for example, with SD-Map[16] or SD-Map*[19] can then be seen as a promising option for efficiently mining potentially arbitrarily large databases. Below, we summarize the main characteristics of heuristic and exhaustive approaches, and discuss exemplary algorithms.

Heuristic Algorithms

For heuristic approaches, commonly a beam search[68] strategy is used because of its efficiency. The search starts with a list of subgroup hypotheses of size w (corresponding to the *beam width*), which may be initially empty. The w subgroup hypotheses contained in the beam are then expanded iteratively, and only the best w expanded subgroups are kept implementing a hill-climbing greedy search. Lavrac et al.,[3] for example, describe the application of the beam-search based *CN2-SD* algorithm adapted for subgroup discovery. To improve upon simple greedy approaches, other alternatives such as the *PRIM* algorithm[69] have been proposed, which employ a *patient* search strategy.

Beam search traverses the search space non-exhaustively and thus does not guarantee to discover the complete set of the top-k subgroups, or all subgroups above a minimal quality threshold. It can also be regarded as a variant of an anytime algorithm, since the search process can be stopped at any point such that the *currently best* subgroups are available. It is also possible to apply beam search to larger description spaces, for example, including richer descriptions for numeric attributes, cf., Ref 56. Furthermore, subgroup set selection can also be integrated into such heuristic approaches[59] as described below.

Alternatives include genetic algorithms, e.g., Refs 63–66, that cast the subgroup discovery task into an evolutionary optimization problem. This can also be applied for subgroup discovery in continuous domains, e.g., Ref 70.

Efficient Exhaustive Algorithms

In contrast to heuristic methods, exhaustive approaches guarantee to discover the best solutions. However, the runtime costs of a (naive) exhaustive algorithm usually prohibit its application for larger search spaces. Examples of exhaustive algorithms include Apriori-based methods,[23] for example, the Apriori-SD[54] algorithm; more alternatives are mentioned below.

Depending on the applied algorithm, there are different pruning options that can be applied. Many state-of-the-art algorithms apply extensions of frequent pattern trees[71] (FP-trees) in a pattern-growth fashion. Then, typically optimistic estimate pruning is applied, while generalization-aware pruning is better supported by layer-wise algorithms based on the Apriori[23] principle. Furthermore, Zimmermann and De Raedt[47,72] have proposed branch-and-bound algorithms that require special (convex) quality functions for pruning the search space.

As efficient exhaustive algorithms, the *BSD* and the *SD-Map** algorithms, for example, allow the efficient handling of binary, nominal and numeric target properties. Both algorithms apply optimistic estimate pruning, but utilize different core data structures, bitset representations versus extended FP-trees, cf., Ref 71. FP-trees are also used in other subgroup discovery algorithms, e.g., by DpSubgroup.[18,50] As an extension of SD-Map*, the GP-Growth algorithm[48] allows subgroup discovery for single target and multi-target concepts, for example, for exceptional model mining; several model classes and quality functions can be implemented using the algorithm. In the following, we briefly review those algorithms in more detail.

SD-Map, SD-Map, and GP-Growth*

SD-Map*[19] is based on the efficient SD-Map[16] algorithm utilizing an FP-tree data structure, cf., Ref 71, i.e., an extended prefix-tree-structure that stores information for pattern refinement and evaluation. SD-Map* applies a divide and conquer method, first mining patterns containing one selector and then recursively mining patterns of size 1 conditioned on the occurrence of a (prefix) 1-selector. For the binary case, an FP-tree node stores the subgroup size and the true positive count of the respective subgroup description. In the continuous case, it considers 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.

For extending the FP-tree structure toward multi-target concepts, we utilize the concept of *evaluation bases* introduced by Lemmerich et al.[48] Then, all information required for the *evaluation* of the respective quality functions is stored in the nodes of the FP-tree, as the basis of the GP-Growth algorithm extending SD-Map/SD-Map*. With this technique, a large number of single and multi-target concept quality functions can be implemented, cf., Ref 48.

BSD

The BSD algorithm[73] utilizes a vertical bitset (bitvector) based data structure. Vertical data representations have been proposed for other data mining tasks, e.g., by Zaki.[74] Burdick et al.[75] used a bitset representation for maximal frequent itemset mining. As a general search strategy, BSD uses a depth-first-search approach with one level look-ahead (similar to the DpSubgroup[18,50] algorithm). BSD uses a vertical data layout utilizing bitsets (vectors of bits), for the selectors, the instances reflecting the current subgroup hypothesis, and an additional array for the (numeric) values of the target variable. Then, the search, i. e., the refinement of the patterns can be efficiently implemented using logical *AND* operations on the respective bitsets, such that the target values can be directly retrieved.

SUBGROUP SET SELECTION

Due to multi-correlations between the selectors, some of the subgroups contained in the result set can overlap significantly. Therefore, usually a *high-quality* set of *diverse* subgroups should be retrieved. Subgroup set selection is one of the critical issues for removing redundancy and improving the interestingness of the overall subgroup discovery result, for example, as first described in Ref 1. Constraints denoting redundancy filters, for example, can be used to prune large regions of the search space. This is especially important for certain search strategies, which do not constrain the search space themselves, for example, exhaustive search compared to beam search. Klösgen[1] distinguishes two types of redundancy filters: logical and heuristic filters. The filters include either logical or heuristic implications for the truth value of a constraint condition with respect to a predecessor/successor pair of subgroups. Logical filters can be described as *strong filters*; they can be used to definitely exclude a region of the search space. Heuristic filters can be used as *weak filters*; these are applied as a first step in a brute force search, where the excluded regions of the search space can be refined later.

The general task of diverse subgroup set discovery is described in Ref 59, for which different types of redundancy and according selection heuristics are proposed. There are several heuristic approaches for pattern set selection in general, e.g., Ref 76, as well as for subgroup discovery.[59,77] In particular, several quality functions for selecting *pattern teams* are proposed in Ref 77, which can be applied in a post-processing step. Furthermore, a method for semi-automatic retrieval of a set of subgroups is described in Ref 78, for which mixed-initiative case-based techniques are applied. In the following, we outline several options for *diversity-aware* and *redundancy-aware* subgroup set discovery and selection in more detail, focusing on condensed representations, relevance criteria, covering approaches, as well as causal subgroup analysis for identifying causally related sets of subgroups.

Condensed Representations

In the field of association rules, *condensed* representations of frequent item sets have been developed for reducing the size of the set of association rules that are generated, e.g., Refs 79, 80. These representations are used for the (implicit) redundancy management, since then the condensed patterns also describe the specifically interesting patterns, and can significantly reduce the size of the result sets. The efficiency of the association rule discovery method is also increased significantly. Such techniques can also be generalized for frequent patterns (cf., Refs 81, 82). For subgroup discovery, target-closed representations can be formalized, cf., Ref 83 for details. In that case, also an implicit redundancy management based on the subgroup descriptions is performed.

Relevance of Subgroups

As a quite simple method for redundancy management of subgroups for binary targets we can consider the *(ir-)relevance* (e.g., Ref 6) of a subgroup with respect to a set of subgroups. A (specialized) subgroup hypothesis S_N is *irrelevant* if there exists a (generalized) subgroup hypothesis S_P such that the true positives of S_N are a subset of the true positive of S_P and the false positives of S_N are a superset of the false positives of S_P. This concept is also closely related to the concept of closed patterns[84] and according relevance criteria.[84,85]

Embedding this redundancy management technique into the search process is straight-forward. When considering a subgroup hypothesis for inclusion into the set of the k best subgroups, the test for (strict) irrelevancy can be applied. In addition, such a method can also be implemented in an optional post-processing step. Furthermore, Großkreutz et al.[83] introduces delta-relevance which provides a more relaxed definition of coverage (essentially trading off precision versus simplicity) with the overall goal of summarizing relevant patterns even more.

Covering Approaches

Similar to covering approaches for rule learning, a subgroup set can also be selected according to its overall coverage of the dataset. The *weighted covering algorithm*[3,6] is such an approach that works by example reweighting. It iteratively focuses the subgroup selection method on the space of target records not covered so far, by reducing the weights of the already covered data records. As discussed in Ref 86, example reweighting can also be used as a search heuristic—in combination with a suitable quality function. In this way, weighted covering is integrated in the subgroup discovery algorithm, that is, the search step directly (e.g., Ref 6). In each search iteration only the best subgroup is considered, then the instances are reweighted, focusing the subgroup discovery method on the not yet covered target class cases.

Causal Subgroup Analysis

For identifying causal subgroups efficiently, constraint-based causal discovery algorithms, e.g., Ref 87 can be applied. These try to limit the possible causal models by analyzing observational data. In causal subgroup analysis (e.g., Refs 88, 89) subgroups are identified which are causal for the target concept; for a causal subgroup, the manipulation of an instance to belong to the subgroup would also affect the probability of the instance to belong to the target group.[87] In accordance with the general principle that correlation does not imply causation, constraint-based algorithms apply statistical tests, for example, the χ^2-test for independence in order to test the (conditional) dependence and independence of variables to *exclude* certain causal relations. After causal subgroups have been detected, the user can retain these (important) subgroups, which have a direct dependency relation to the target concept, in contrast to the remaining non-causal subgroups, which are often redundant given the causal subgroups.

TOOLS AND APPLICATIONS

Subgroup discovery is a powerful and broadly applicable data mining approach, in particular, for descriptive data mining tasks. It is typically applied, for example, in order to obtain an overview on the relations in the data and for automatic hypotheses generation. Furthermore, also predictive tasks can be tackled, for example, by stacking approaches,[90] or by applying the *LeGo* framework for combining local patterns into global models.

From a tool perspective, there exist several software packages for subgroup discovery, e.g., Refs 91–94. As open source options, there are, for example, subgroup discovery modules for the data mining systems *Orange*[94] and *RapidMiner*,[95] the *Cortana*[92] system for discovering local patterns in data, as well as the specialized subgroup discovery and analytics system VIKAMINE.[91,93] Using the latter a number of successful real-world subgroup discovery applications have been implemented. These cover, for example, knowledge discovery and quality control setting in the medical domain,[4,96,97] industrial applications,[19] as well as pattern analytics in the social media domain.[57] The system is targeted at a broad range of users, from industrial practitioners to ML/KDD researchers, students, and users interested in knowledge discovery and data analysis in general. Especially the visual mining methods enable the direct integration of the user to overcome major problems of automatic data mining methods, cf., Refs 91, 93.

Applications include, for example, knowledge discovery in the medical domain, technical (fault) analysis, e.g., Refs 19, 98, or mining social data, e.g., Refs 10, 27, 99. We discuss these exemplarily below. Furthermore, for heterogenous data exceptional model mining (cf., Refs 11,12,48,100) opens up a wide range of options. There are also applications in related fields, for example, in software engineering[101] for requirements engineering and design.

Knowledge Discovery in the Medical Domain

Subgroup discovery is a prominent approach for mining medical data, e.g., Refs 4, 6, 91, 96, 97, 102, 103. Using the VIKAMINE system, for example, subgroup discovery has been applied for large-scale knowledge discovery and quality control in the clinical application SonoConsult, cf., Ref 97. For this, several data sources including structured clinical documentation and unstructured documents, e.g., Ref 104, were integrated. The main data source was given by the SonoConsult system, which has been in routine use since 2002 as the only documentation system for ultrasound examinations in the DRK-hospital of Berlin-Köpenick; since 2005, it is in routine use at the university hospital of Würzburg. The physicians considered statistical analysis as one of the most desirable features. In the analysis and results, e.g., Refs 4, 96, 105, subgroup discovery was applied on a large set of clinical features together with laboratory values and diagnostic information from several systems.

According to the physicians, subgroup discovery and analysis was quite suitable for examining common medical questions, for example whether a certain pathological state is significantly more frequent if combinations of other pathological states exist, or if there are diagnoses, which one physician documents significantly more or less frequently than the average. Furthermore, VIKAMINE also provided an intuitive overview on the data, in addition to the knowledge discovery and quality control functions. Then, subgroup discovery can be performed in a semi-automatic approach, first generating hypothesis using automatic methods that are then inspected and refined using visual analytics techniques, cf., Refs 91, 106.

Technical Fault Analysis

Technical applications of subgroup discovery include, for example, mining service processes,[107] analysis of smart electrical meter data,[98] or fault analysis of production processes.[89] The latter, for example, has been implemented using VIKAMINE.[19] It aimed at large-scale fault detection and analysis using subgroup discovery. Specifically, the task required the identification of subgroups (as combination of certain factors) that cause a significant increase/decrease in, for example, the fault/repair rates of certain products. Similar problems in industry concern, for example, the number of service requests for a certain technical component, or the number of calls of customers to service support.

Such applications of subgroup discovery often require the utilization of continuous parameters. Then, the target concepts can often not be analyzed sufficiently using the standard discretization techniques, since the discretization of the variables causes a loss of information. As a consequence, the interpretation of the results is often difficult using standard data mining tools. In this context, VIKAMINE provided state-of-the-art algorithmic implementations, cf., Ref 19, and enabled a semi-automatic involvement of the domain experts for effectively contributing in a discovery session.

Subgroup Discovery in Social Media

In mining social media,[108] subgroup discovery methods are a versatile tool for focusing on different facets of the application domain. Subgroup discovery was applied, for example, for obtaining descriptive profiles of spammers in social media, specifically social bookmarking systems. Here, subgroup discovery was applied for the characterization of spammers, i. e., to describe them by their most prominent features.[38] The mined patterns capturing certain spammer subgroups provide explanations and justifications for marking or resolving spammer candidates.

In addition, in social bookmarking systems, it is usually useful to identify high-quality tags, that is, tags with a certain maturity, cf., Ref 109. Subgroup discovery was applied for obtaining maturity profiles of tags based on a set of graph centrality features on the tag–tag co-occurrence graph, which are simple to compute and to assess. Then, they can be applied for tag recommendations, faceted browsing, or for improving search.

Furthermore, subgroup discovery has been utilized for community pattern analytics in social media, e.g., Ref 110, as well as semi-automatic approaches for pattern detection in geo-referenced tagging data, exemplified by an application using Flickr data, cf., Ref 57. In this domain of 'Big Data', subgroup discovery can also provide suitable solutions using the efficient automatic algorithms; the combination of automatic and interactive visualization methods complemented each other for a successful discovery process. Especially subgroup introspection, and pattern explanation capabilities, e.g., Refs 78, 106, 111 proved essential during pattern analytics and assessment.

FUTURE DIRECTIONS AND CHALLENGES

Overall, there is already a large body of works on algorithmic as well as methodological issues on subgroup discovery. Major challenging points include the algorithmic performance, the redundancy of the result set of subgroups, adequate comprehensive visualization, and the processing and integration of heterogenous data. Larger search spaces, like those encountered for numerical data, (complex) multi-relational datasets, for example, encountered in social networks, or spatiotemporal data require efficient algorithms that can handle those different types of data, e.g., Refs 10, 112. Also combinations of such different data characteristics, for example, temporal pattern mining for event detection,[113] or temporal subgroup analytics[114] provide further challenges, especially considering sophisticated exceptional model classes in that area.

Typically, heuristic approaches are first established before advanced processing methods like sophisticated pruning and suppression heuristics enable exhaustive subgroup discovery techniques. Furthermore, processing large volumes of data (i.e., big data) is another challenge. In that area, extensions of techniques for parallelizing the computation, e.g., Ref 73 or techniques from the field of association rules, e.g., Ref 115 can provide interesting options for further improvements in that area. Also, sampling approaches, e.g., Refs 28, 62, 116 can be applied for addressing these issues.

In addition, the integration of (rich) background knowledge in a knowledge-intensive approach, e.g., Refs 4, 117, 118, 119, is a prerequisite for the analysis of large datasets for which relations and prior information needs to be utilized. This also tackles the area of automatic subgroup discovery, recent search strategies, e.g., Ref 120, and the applied significance filtering in many methods and tools.[121] For a full-scale approach, these issues need to be addressed such that suitable methods can be integrated comprehensively, from automatic to interactive approaches, e.g., Refs 91, 93, 122, which can also be applied for generating appropriate explanations.[123] Then, the combination, integration, and further development of such techniques and methods will lead to novel techniques embedded in a comprehensive approach for subgroup discovery and analytics, toward robust tool implementations, and finally to further successful applications of subgroup discovery.

CONCLUSIONS

In this paper, we have reviewed key areas of subgroup discovery considering algorithmic issues for discovering subgroups as well as for refining the discovery results. Thus, we covered the fundamentals of subgroup discovery, provided an overview from an algorithmic point of view, briefly discussed efficient algorithms, and summarized approaches for selecting a final subgroup set. Furthermore, we presented different tools and applications of subgroup discovery and outlined several interesting and challenging directions for future work. Overall, subgroup discovery is a versatile and powerful method that can be tuned to many application characteristics. It provides a comprehensive approach integrating different techniques for providing solutions in many domains.

REFERENCES

1. Klösgen W. Explora: a multipattern and multistrategy discovery assistant. In: Fayyad UM, Piatetsky-Shapiro G, Smyth P, Uthurusamy R, eds. *Advances in Knowledge Discovery and Data Mining*. Palo Alto, CA: AAAI Press; 1996, 249–271.
2. Wrobel S. An algorithm for multi-relational discovery of subgroups. In: *Proceedings of the 1st European Symposium on Principles of Data Mining and Knowledge Discovery*. Heidelberg: Springer Verlag; 1997, 78–87.
3. Lavrac N, Kavsek B, Flach P, Todorovski L. Subgroup discovery with CN2-SD. *J Mach Learn Res* 2004, 5:153–188.
4. Atzmueller M, Puppe F, Buscher H-P. Exploiting background knowledge for knowledge-intensive subgroup discovery. In: *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-05)*, Edinburgh, Scotland, 2005, 647–652.
5. Novak PK, Lavrač N, Webb GI. Supervised descriptive rule discovery: a unifying survey of contrast set, emerging pattern and subgroup mining. *J Mach Learn Res* 2009, 10:377–403.
6. Gamberger D, Lavrac N. Expert-guided subgroup discovery: methodology and application. *J Artif Intell Res* 2002, 17:501–527.
7. Jorge AM, Pereira F, Azevedo PJ. Visual interactive subgroup discovery with numerical properties of interest. In: *Proceedings of the 9th International Conference on Discovery Science (DS 2006)*. Lecture Notes in Artificial Intelligence, vol. 4265. Heidelberg: Springer Verlag; 2006, 301–305.
8. Herrera F, Carmona C, Gonzalez P, del Jesus M. An overview on subgroup discovery: foundations and applications. *Knowl Inf Syst* 2011, 29:495–525.
9. Klösgen W. Chapter 16.3: Subgroup discovery. In: *Handbook of Data Mining and Knowledge Discovery*. New York: Oxford University Press; 2002.
10. Atzmueller M, Mueller J, Becker M. Exploratory subgroup analytics on ubiquitous data. In: *Mining, Modeling and Recommending 'Things' in Social Media*. Lecture Notes in Artificial Intelligence, vol. 8940. Heidelberg: Springer Verlag; 2015.
11. Duivesteijn W, Knobbe A, Feelders A, van Leeuwen M. Subgroup discovery meets Bayesian networks – an exceptional model mining approach. In: *Proceedings of the IEEE International Conference on Data Mining (ICDM)*. Washington, DC: IEEE Computer Society; 2010, 158–167.
12. Leman D, Feelders A, Knobbe A. Exceptional model mining. In: *Proceedings of ECML/PKDD*. Lecture Notes in Computer Science, vol. 5212. Heidelberg: Springer Verlag; 2008, 1–16.
13. Cheng H, Yan X, Han J, Yu PS. Direct discriminative pattern mining for effective classification. In: *Proceedings of the 24th International IEEE Conference on*

Data Engineering. Washington, DC: IEEE Computer Society; 2008, 169–178.

14. Bay SD, Pazzani MJ. Detecting group differences: mining contrast sets. *Data Min Knowl Disc* 2001, 5:213–246.
15. Dong G., Li J. Efficient mining of emerging patterns: discovering trends and differences. *Proceedings of ACM SIGKDD*, 43–52, New York, 1999. ACM Press.
16. Atzmueller M, Puppe F. SD-Map – a fast algorithm for exhaustive subgroup discovery. In: *Proceedings of PKDD*. Lecture Notes in Artificial Intelligence, vol. 4213. Heidelberg: Springer Verlag; 2006, 6–17.
17. Moreland K, Truemper K. Discretization of target attributes for subgroup discovery. In: Perner P, ed. *International conference on Machine Learning and Data Mining*. Lecture Notes in Computer Science, vol. 5632. Heidelberg: Springer Verlag; 2009, 44–52.
18. Grosskreutz H, Rüping S. On subgroup discovery in numerical domains. *Data Min Knowl Disc* 2009, 19:210–226.
19. Atzmueller M, Lemmerich F. Fast subgroup discovery for continuous target concepts. In: *Proceedings of the 18th International Symposium on Methodologies for Intelligent Systems (ISMIS 2009)*. Lecture Notes in Computer Science, vol. 5722. Heidelberg: Springer Verlag; 2009, 1–15.
20. Morik K, Boulicaut J-F, Siebes A, eds. *Local Pattern Detection*. Heidelberg: Springer Verlag; 2005.
21. Knobbe AJ, Cremilleux B, Fürnkranz J, Scholz M. From local patterns to global models: the LeGo approach to data mining. In: *From Local Patterns to Global Models: Proceedings of the ECML/PKDD-08 Workshop (LeGo-08)*, Antwerp, Belgium, 2008, 1–16.
22. Han J, Cheng H, Xin D, Yan X. Frequent pattern mining: current status and future directions. *Data Min Knowl Disc* 2007, 15:55–86.
23. Agrawal R, Srikant R. Fast algorithms for mining association rules. In: Bocca JB, Jarke M, Zaniolo C, eds. *Proceedings of the 20th International Conference on Very Large Data Bases, (VLDB)*. San Francisco, CA: Morgan Kaufmann; 1994, 487–499.
24. Lakhal L, Stumme G. Efficient mining of association rules based on formal concept analysis. In: *Formal Concept Analysis*. Heidelberg: Springer Verlag; 2005, 180–195.
25. Boley M, Grosskreutz H. Non-redundant subgroup discovery using a closure system. In: *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases*. Heidelberg: Springer Verlag; 2009, 179–194.
26. Boley M, Horvath T, Poigné A, Wrobel S. Listing closed sets of strongly accessible set systems with applications to data mining. *Theor Comput Sci* 2010, 411:691–700.
27. Atzmueller M, Mitzlaff F. Efficient descriptive community mining. In: *Proceedings of the 24th International FLAIRS Conference*. Palo Alto, CA: AAAI Press; 2011, 459–464.
28. Boley M, Lucchese C, Paurat D, Gärtner T. Direct local pattern sampling by efficient two-step random procedures. In: *Proceedings of the ACM SIGKDD, KDD '11*. New York: ACM Press; 2011, 582–590.
29. Tuzhilin A. Chapter 19.2.2: usefulness, novelty, and integration of interestingness measures. In: *Handbook of Data Mining and Knowledge Discovery*. New York: Oxford University Press; 2002.
30. Freitas AA. On rule interestingness measures. *Knowl-Based Syst* 1999, 12:309–325.
31. Atzmueller M, Baumeister J, Puppe F. Quality measures and semi-automatic mining of diagnostic rule bases (extended version). In: *Proceedings of the 15th International Conference on Applications of Declarative Programming and Knowledge Management (INAP 2004)*. Lecture Notes in Artificial Intelligence, vol. 3392. Heidelberg: Springer Verlag; 2005, 65–78.
32. Piatetsky-Shapiro G, Matheus CJ. The interestingness of deviations. In: *Proceedings of the AAAI-94 Workshop on Knowledge Discovery in Databases (KDD-94)*. New York: ACM Press; 1994, 25–36.
33. Geng L, Hamilton HJ. Interestingness measures for data mining: a survey. *ACM Comput Surv* 2006, 38:Article no 9.
34. Romao W, Freitas AA, Gimenes IMdS. Discovering interesting knowledge from a science & technology database with a genetic algorithm. *Appl Soft Comput* 2004, 4:121–137.
35. Konijn RM, Duivesteijn W, Meeng M, Knobbe AJ. Cost-based quality measures in subgroup discovery. In: Li J, Cao L, Wang C, Tan KC, Liu B, Pei J, Tseng VS, eds. *Trends and Applications in Knowledge Discovery and Data Mining – PAKDD 2013 International Workshops: DMApps, DANTH, QIMIE, BDM, CDA, CloudSD*. Revised Selected Papers of Lecture Notes in Computer Science, vol. 7867. Heidelberg: Springer Verlag; 2013, 404–415.
36. M. Müller, R. Rosales, H. Steck, S. Krishnan, B. Rao, and S. Kramer. Subgroup discovery for test selection: a novel approach and its application to breast cancer diagnosis. N. M. Adams, C. Robardet, A. Siebes, J.-F. Boulicaut, *Advances in Intelligent Data Analysis VIII, 8th International Symposium on Intelligent Data Analysis, IDA 2009*, 5772 Lecture Notes in Computer Science 119–130. Springer Verlag, Heidelberg, 2009.
37. Abudawood T, Flach P. Evaluation measures for multi-class subgroup discovery. In: *Proceedings of the ECML/PKDD*. Lecture Notes in Artificial Intelligence, vol. 5782. Heidelberg: Springer Verlag; 2009, 35–50.
38. Atzmueller M, Lemmerich F, Krause B, Hotho A. Who are the spammers? Understandable local patterns for concept description. In: *Proceedings of the 7th Conference on Computer Methods and Systems*. Krakow, Poland: Oprogramowanie Nauko-Techniczne; 2009.

39. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003, 56:1129–1135.

40. Lavrac N, Flach PA, Kasek B, Todorovski L. Rule induction for subgroup discovery with CN2-SD. In: Bohanec M, Kasek B, Lavrac N, Maldenic D, eds. *ECML/PKDD'02 Workshop on Integration and Collaboration Aspects of Data Mining, Decision Support and Meta-Learning*. Helsinki University Printing House, Helsinki, Finland, 2002, 77–87.

41. Li J, Liu J, Toivonen H, Satou K, Sun Y, Sun B. Discovering statistically non-redundant subgroups. *Knowl-Based Syst* 2014, 67:315–327.

42. Grosskreutz H. Cascaded subgroup discovery with an application to regression. *Proceedings of the ECML/PKDD*, 5211 Lecture Notes in Artificial Intelligence, Heidelberg, 2008. Springer Verlag.

43. Pieters B, Knobbe A, Dzeroski S. Subgroup discovery in ranked data, with an application to gene set enrichment. In: *Proceedings of the Preference Learning Workshop (PL2010) at ECML/PKDD*, Barcelona, Spain, 2010.

44. Webb GI. Discovering associations with numeric variables. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '01*. New York: ACM Press; 2001, 383–388.

45. Aumann Y, Lindell Y. A statistical theory for quantitative association rules. *J Intell Inf Syst* 2003, 20:255–283.

46. Hotelling H. The generalization of student's ratio. *Ann Math Statist* 1931, 2:360–378.

47. Zimmermann A, De Raedt L. Cluster-grouping: from subgroup discovery to clustering. Machine Learning. October 2009, Volume 77, Issue 1, pp 125–159.

48. Lemmerich F, Becker M, Atzmueller M. Generic pattern trees for exhaustive exceptional model mining. In: *Proceedings of ECML/PKDD*. Lecture Notes in Artificial Intelligence, vol. 7524. Heidelberg: Springer Verlag; 2012, 277–292.

49. Shapiro LG, Haralick RM. A metric for comparing relational descriptions. *IEEE Trans Pattern Anal Mach Intell* 1985, 7:90–94.

50. Grosskreutz H, Rüping S, Wrobel S. Tight optimistic estimates for fast subgroup discovery. In: *Proceedings of the ECML/PKDD*. Lecture Notes in Artificial Intelligence, vol. 5211. Heidelberg: Springer Verlag; 2008, 440–456.

51. Lemmerich F, Becker M, Puppe F. Difference-based estimates for generalization-aware subgroup discovery. In: *Proceedings of ECML/PKDD*. Lecture Notes in Computer Science, vol. 8190. Heidelberg: Springer Verlag; 2013, 288–303.

52. Roberto J, Bayardo J. Efficiently mining long patterns from databases. In: *SIGMOD '98: Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data*. New York: ACM Press; 1998, 85–93.

53. Lavrac N, Cestnik B, Gamberger D, Flach P. Decision support through subgroup discovery: three case studies and the lessons learned. *Mach Learn* 2004, 57:115–143.

54. Kavsek B, Lavrac N, Jovanoski V. APRIORI-SD: adapting association rule learning to subgroup discovery. In: *Proceedings of the 5th International Symposium on Intelligent Data Analysis*. Heidelberg: Springer Verlag; 2003, 230–241.

55. Klösgen W, May M. Census data mining – an application. D. Malerba, P. Brito, *Proceedings of the Workshop Mining Official Data, 6th European Conference, PKDD 2002*, Helsinki, 2002. Helsinki University Printing House.

56. Mampaey M, Nijssen S, Feelders A, Knobbe AJ. Efficient algorithms for finding richer subgroup descriptions in numeric and nominal data. In: *Proceedings of the IEEE International Conference on Data Mining (ICDM)*. Washington, DC: IEEE Computer Society; 2012, 499–508.

57. Atzmueller M, Lemmerich F. Exploratory pattern mining on social media using geo-references and social tagging information. *Int J Web Sci* 2013, 2:80–112.

58. Knobbe A, Feelders A, Leman D. Exceptional model mining. In: *Data Mining: Foundations and Intelligent Paradigms*, vol. 2. Heidelberg: Springer Verlag; 2011, 183–198.

59. van Leeuwen M, Knobbe AJ. Diverse subgroup set discovery. *Data Min Knowl Discov* 2012, 25:208–242.

60. Toivonen H. Sampling large databases for association rules. T. M. Vijayaraman, A. P. Buchmann, C. Mohan, and N. L. Sarda, *Proceedings of the 1996 International Conference on Very Large Data Bases*, 134–145. Morgan Kaufman, San Francisco, CA, 1996.

61. Freund Y. Self bounding learning algorithms. In: *COLT: Proceedings of the 11th Annual Conference on Computational Learning Theory*. New York: ACM Press; 1998.

62. Scheffer T, Wrobel S. Finding the most interesting patterns in a database quickly by using sequential sampling. *J Mach Learn Res* 2002, 3:833–862.

63. del Jesus MJ, González P, Herrera F, Mesonero M. Evolutionary fuzzy rule induction process for subgroup discovery: a case study in marketing. *IEEE Trans Fuzzy Syst* 2007, 15:578–592.

64. Carmona CJ, González P, del Jesús MJ, Herrera F. NMEEF-SD: non-dominated multiobjective evolutionary algorithm for extracting fuzzy rules in subgroup discovery. *IEEE Trans Fuzzy Syst* 2010, 18: 958–970.

65. Luna JM, Romero JR, Romero C, Ventura S. Discovering subgroups by means of genetic programming. Krawiec K, Moraglio A, Hu T, Etaner-Uyar AS, Hu

B, eds. *EuroGP*, 7831 Lecture Notes in Computer Science, 121–132. Springer Verlag, Heidelberg, 2013.

66. Carmona CJ, González P, del Jesus MJ, Herrera F. Overview on evolutionary subgroup discovery: analysis of the suitability and potential of the search performed by evolutionary algorithms. *WIREs Data Min Knowl Discov* 2014, 4:87–103.
67. Lemmerich F, Puppe F. Local models for expectation-driven subgroup discovery. In: *Proceedings of the IEEE International Conference on Data Mining (ICDM)*. Washington, DC: IEEE Computer Society; 2011, 360–369.
68. Lowerre BT. The Harpy speech recognition system. PhD Thesis, Carnegie Mellon University, Pittsburgh, PA, 1976.
69. Friedman JH, Fisher NI. Bump hunting in high-dimensional data. *Stat Comput* 1999, 9:123–143.
70. Rodríguez D, Ruiz R, Riquelme JC, Aguilar-Ruiz JS. Searching for rules to detect defective modules: a subgroup discovery approach. *Inf Sci* 2012, 191:14–30.
71. Han J, Pei J, Yin Y. Mining frequent patterns without candidate generation. In: Chen W, Naughton J, Bernstein PA, eds. *2000 ACM SIGMOD International Conference on Management of Data*. New York: ACM Press; 2000, 1–12.
72. Zimmermann A, Raedt LD. CorClass: correlated association rule mining for classification. E. Suzuki and S. Arikawa, *Proceedings of the 7th International Conference on Discovery Science*, 3245 Lecture Notes in Computer Science, 60–72, Springer Verlag, Heidelberg, 2004.
73. Lemmerich F, Rohlfs M, Atzmueller M. Fast discovery of relevant subgroup patterns. In: *Proceedings of the 23rd International FLAIRS Conference*. Palo Alto, CA: AAAI Press; 2010, 428–433.
74. Zaki MJ. Efficient enumeration of frequent sequences. In: *CIKM '98: Proceedings of the Seventh International Conference on Information and Knowledge Management*. New York: ACM Press; 1998, 68–75.
75. Burdick D, Calimlim M, Gehrke J. MAFIA: a maximal frequent itemset algorithm for transactional databases. In: *Proceedings of the 17th International Conference on Data Engineering (ICDE'01)*, Heidelberg, Germany, 2001, 443–452.
76. Bringmann B, Zimmermann A. The chosen few: on identifying valuable patterns. In: *Proceedings of the IEEE International Conference on Data Mining*. Washington, DC: IEEE Computer Society; 2007, 63–72.
77. Knobbe A, Ho E. Pattern teams. In: *Knowledge Discovery in Databases: PKDD 2006*. Heidelberg: Springer Verlag; 2006, 577–584.
78. Atzmueller M, Puppe F. A case-based approach for characterization and analysis of subgroup patterns. *J Appl Intell* 2008, 28:210–221.
79. Pasquier N, Bastide Y, Taouil R, Lakhal L. Discovering frequent closed itemsets for association rules. C. Beeri and P. Buneman, *Proceedings of the 7th International Conference on Database Theory (ICDT 99)*, 1540 Lecture Notes in Computer Science, 398–416. Springer Verlag, Heidelberg, 1999.
80. Bastide Y, Pasquier N, Taouil R, Stumme G, Lakhal L. Mining minimal non-redundant association rules using frequent closed itemsets. In: Lloyd J, Dahl V, Furbach U, Kerber M, Lau K-K, Palamidessi C, Pereira LM, Sagiv Y, Stuckey PJ, eds. *Computational Logic – CL 2000*. Lecture Notes in Computer Science. Heidelberg: Springer Verlag; 2000, 972–986.
81. Mielikäinen T. Summarization techniques for pattern collections in data mining. PhD Thesis, University of Helsinki, Helsinki, May 2005.
82. Boulicaut J-F. Condensed representations for data mining. In: *Encyclopedia of Data Warehousing and Mining*. Hershey, PA: Idea Group; 2006, 37–79.
83. Großkreutz H, Paurat D, Rüping S. An enhanced relevance criterion for more concise supervised pattern discovery. In: *Proceedings of the ACM SIGKDD, KDD '12*. New York: ACM Press; 2012, 1442–1450.
84. Garriga GC, Kralj P, Lavrac N. Closed sets for labeled data. *J Mach Learn Res* 2008, 9:559–580.
85. Lavrac N, Gamberger D. Relevancy in constraint-based subgroup discovery. In: Jean-Francois Boulicaut HM, Luc de Raedt , eds. *Constraint-Based Mining and Inductive Databases*. Lecture Notes in Computer Science, vol. 3848. Heidelberg: Springer Verlag; 2004.
86. Kavsek B, Lavrac N. Analysis of example weighting in subgroup discovery by comparison of three algorithms on a real-life data set. In: *Proceedings of the Workshop on Advances in Inductive Rule Learning, at ECML/PKDD*, Pisa, Italy, 2004.
87. Cooper GF. A simple constraint-based algorithm for efficiently mining observational databases for causal relationships. *Data Min Knowl Discov* 1997, 1:203–224.
88. Klösgen W, May M. Spatial subgroup mining integrated in an object-relational spatial database. In: Elomaa T, Mannila H, Toivonen H, eds. *Proceedings of the PKDD*. Lecture Notes in Computer Science, vol. 2431. Heidelberg: Springer Verlag; 2002, 275–286.
89. Atzmueller M, Puppe F. *Knowledge discovery enhanced with semantic and social information, chapter a knowledge-intensive approach for semi-automatic causal subgroup discovery*. Heidelberg: Springer Verlag; 2009.
90. Klügl P, Toepfer M, Lemmerich F, Hotho A, Puppe F. Collective information extraction with context-specific consistencies. In: *Proceedings of ECML/PKDD*. Lecture Notes in Artificial Intelligence, vol. 7523. Heidelberg: Springer Verlag; 2012, 728–743.

91. Atzmueller M, Puppe F. Semi-automatic visual subgroup mining using VIKAMINE. *J Univer Comp Sci* 2005, 11:1752–1765.

92. Meeng M, Knobbe AJ. Flexible enrichment with Cortana – Software Demo. In: *Proceedings of Benelearn*, The Hague, The Netherlands, 2011, 117–119.

93. Atzmueller M, Lemmerich F. VIKAMINE – open-source subgroup discovery, pattern mining, and analytics. In: *Proceedings of ECML/PKDD*. Lecture Notes in Artificial Intelligence, vol. 7524. Heidelberg: Springer Verlag; 2012.

94. Demšar J, Curk T, Erjavec A, Gorup Č, Hočevar T, Milutinovič M, Možina M, Polajnar M, Toplak M, Starič A, et al. Orange: data mining toolbox in python. *J Mach Learn Res* 2013, 14:2349–2353.

95. Mierswa I, Wurst M, Klinkenberg R, Scholz M, Euler T. YALE: rapid prototyping for complex data mining tasks. In: *Proceedings of ACM SIGKDD, KDD '06*. New York: ACM Press; 2006, 935–940.

96. Atzmueller M, Puppe F, Buscher H-P. Profiling examiners using intelligent subgroup mining. In: *Proceedings of the 10th International Workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP-2005)*, Aberdeen, Scotland, 2005, 46–51.

97. Puppe F, Atzmueller M, Buscher G, Huettig M, Lührs H, Buscher H-P. Application and evaluation of a medical knowledge-system in sonography (SonoConsult). In: *Proceedings of the 18th European Conference on Artificial Intelligence (ECAI 20008)*, Patras, Greece, 2008, 683–687.

98. Jin N, Flach P, Wilcox T, Sellman R, Thumim J, Knobbe A. Subgroup discovery in smart electricity meter data. *IEEE Trans Industr Inform* May 2014, 10:1327–1336.

99. Atzmueller M. Data mining on social interaction networks. *J Data Min Digit Human* 2014, 1.

100. Duivesteijn W., Feelders A. and Knobbe A. J. Different slopes for different folks: mining for exceptional regression models with cook's distance. *Proceedings of ACM SIGKDD*, 868–876, New York, 2012. ACM Press.

101. Behrenbruch K, Atzmueller M, Evers C, Schmidt L, Stumme G, Geihs K. A personality based design approach using subgroup discovery. In: *Human-Centred Software Engineering*. Lecture Notes in Computer Science, vol. 7623. Heidelberg: Springer Verlag; 2012, 259–266.

102. Lavrac N, Gamberger D, Flach P. Subgroup discovery for actionable knowledge generation: shortcomings of classification rule learning and the lessons learned. In: Lavrac N, Motoda H, Fawcett T, eds. *Proceedings of the ICML 2002 Workshop on Data Mining: Lessons Learned*, Sydney, Australia, 2002.

103. Gamberger D, Lavrac N, Krstacic G. Active subgroup mining: a case study in coronary heart disease risk group detection. *Artif Intell Med* 2003, 28:27–57.

104. Atzmueller M, Beer S, Puppe F. Data mining, validation and collaborative knowledge capture. In: Brüggemann S, d'Amato C, eds. *Collaboration and the Semantic Web: Social Networks, Knowledge Networks and Knowledge Resources*. Hershey, PA: IGI Global; 2012.

105. Atzmueller M, Puppe F, Buscher H-P. A semi-automatic approach for confounding-aware subgroup discovery. *Int J Artif Intell Tool* 2009, 18:1–18.

106. Atzmueller M, Puppe F. Semi-automatic refinement and assessment of subgroup patterns. In: *Proceedings of the 21st International Florida Artificial Intelligence Research Society Conference (FLAIRS-2008)*. Palo Alto, CA: AAAI Press; 2008, 518–523.

107. Natu M, Palshikar G. Interesting subset discovery and its application on service processes. In: Yada K, ed. *Data Mining for Service*. Studies in Big Data, vol. 3. Heidelberg: Springer Verlag; 2014, 245–269.

108. Atzmueller M. Mining social media: key players, sentiments, and communities. *WIREs Data Min Knowl Discov* 2012, 2:411–419.

109. Atzmueller M, Benz D, Hotho A, Stumme G. Towards mining semantic maturity in social bookmarking systems. In: *Proceedings of the Workshop Social Data on the Web, 10th International Semantic Web Conference*, Bonn, Germany, 2011.

110. Atzmueller M. Social behavior in mobile social networks: characterizing links, roles and communities. In: Chin A, Zhang D, eds. *Mobile Social Networking: An Innovative Approach, Computational Social Sciences*. Heidelberg: Springer Verlag; 2014, 65–78.

111. Atzmueller M, Roth-Berghofer T. The mining and analysis continuum of explaining uncovered. In: *Proceedings of the 30th SGAI International Conference on Artificial Intelligence (AI-2010)*, Cambridge, UK, 2010.

112. Magalhães A, Azevedo PJ. Contrast set mining in temporal databases. *Expert Syst* 2014.

113. Batal I, Fradkin D, Harrison J, Moerchen F, Hauskrecht M. Mining recent temporal patterns for event detection in multivariate time series data. In: *Proceedings of ACM SIGKDD, KDD '12*. New York: ACM Press; 2012, 280–288.

114. Sáez C, Rodrigues PP, Gama J, Robles M, García-Gómez JM. Probabilistic change detection and visualization methods for the assessment of temporal stability in biomedical data quality. *Data Min Knowl Disc*, 1–26, 2014. doi: 10.1007/s10618-014-0378-6.

115. Li H, Wang Y, Zhang D, Zhang M, Chang EY. PFP: parallel Fp-growth for query recommendation. In: *Proceedings of RecSys*. New York, NY: ACM Press; 2008, 107–114.

116. Scheffer T, Wrobel S. A scalable constant-memory sampling algorithm for pattern discovery in large

databases. In: *Proceedings of PKDD*. Heidelberg: Springer Verlag; 2002, 397–409.

117. Atzmueller M, Puppe F, Buscher H-P. Towards knowledge-intensive subgroup discovery. In: *Proceedings of the LWA 2004 Workshop*, Berlin, Germany, 2004, 117–123.

118. Atzmueller M, Puppe F. A methodological view on knowledge-intensive subgroup discovery. In: *Proceedings of the 15th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2006)*. Lecture Notes in Artificial Intelligence, vol. 4248. Heidelberg: Springer Verlag; 2006, 318–325.

119. Vavpetic A, Lavrac N. Semantic subgroup discovery systems and workflows in the SDM-toolkit. *Comput J* 2013, 56:304–320.

120. Meeng M, Duivesteijn W, Knobbe AJ. ROC-search – an ROC-guided search strategy for subgroup discovery. In: Zaki MJ, Obradovic Z, Tan P-N, Banerjee A, Kamath C, Parthasarathy S, eds. *Proceedings of the SIAM International Conference on Data Mining*. Philadelphia, PA: SIAM; 2014, 704–712.

121. Lemmerich F, Puppe F. A critical view on automatic significance-filtering in pattern mining. In: *Proceedings of the Workshop Statistically Sound Data Mining, ECML/PKDD 2014*, Nancy, France, 2014.

122. van Leeuwen M. Interactive data exploration using pattern mining. A. Holzinger and I. Jurisica, *Lecture Notes in Computer Science*. Springer Verlag, Heidelberg, 2014.

123. Vavpetic A, Podpecan V, Lavrac N. Semantic subgroup explanations. *J Intell Inf Syst* 2014, 42:233–254.