Supervised Descriptive Rule Discovery: A Survey of the State-of-the-Art

C. J. Carmona
Languages and Computer Technology Systems, Department of Civil Engineering, University of Burgos, 09006, Burgos (Spain)
D. Elizondo
School of Computer Science and Informatics, De Montfort University, The Gateway, Leicester, LE1 9BH, (United Kingdom)

Abstract

The supervised descriptive rule discovery concept groups a set of data mining techniques whose objective is to describe data with respect to a property of interest. Among the techniques within this concept are the subgroup discovery, emerging patterns and contrast sets.

This contribution presents the supervised descriptive rule discovery concept within the data mining literature. Specifically, it is important to remark the main difference with respect to other existing techniques within classification or description. In addition, a a survey of the state-of-the-art about the different techniques within supervised descriptive rule discovery throughout the literature can be observed. The paper allows to the experts to analyse the compatibilities between terms and heuristics of the different data mining tasks within this concept.

1. Introduction

The knowledge discovery in databases (KDD) is a computational process for discovering knowledge in data through the use of different methodologies, technologies and systems (Fayyad et al., 1996). Within the KDD process there are two areas differentiated perfectly: predictive data mining and descriptive data mining. Former attempts to make predictions about unknown objects with respect to a class whereas the descriptive searches for relationships between data. In general, the predictive data mining process employs the supervised learning because it is necessary to have a property of interest in order to predict it. However, the descriptive data mining process is based on unsupervised learning because it is not necessary this class.

The supervised descriptive rule discovery (SDRD) concept was presented in (Kralj-Novak et al., 2009b). Its main proposal is the search for interesting descriptions in data with respect to a property or class of interest. Essentially, SDRD describes labelled data, i.e. it combines the descriptive data mining with supervised learning. The most representative techniques within SDRD are Subgroup Discovery (SD) (Herrera et al., 2011, Carmona et al., 2014), Emerging Patterns (EPs) (Dong and Li, 1999) and Contrast Sets (CSs) (Bay and Pazzani, 2001). These techniques have defined by different

These techniques have been studied and analysed at different stages by different authors. However, their main goals are very similar and it is primarily the terminology that differs as well as the quality measures used in order to analyse a given problem. This contribution presents a perfect positioning of the SDRD concept and a state-of-the-art for the different techniques grouped on the SDRD.

The paper is organised as follows: the definitions and main properties are outlined for SD are presented in Section 3, for EPs in Section 4 and for CSs in Section 5. Finally, Section 6 presents the compatibilities between the different concepts and terms within the SDRD, and Section 7 shows the main heuristics employed within the SDRD.

2. Supervised descriptive rule discovery

In data mining there are two main approaches used in order to analyse data: supervised learning (labelled data) and unsupervised learning (unlabelled data). Together with these approaches we further distinguish between predictive and descriptive induction, whereby predictive data mining methods are usually supervised (induce models from labelled data), and descriptive data mining methods are typically unsupervised (induce interesting association in unlabelled data).

The SDRD concept was introduced by Kralj-Novak et al. (Kralj-Novak et al., 2009b). It describes the group of rule based techniques used in order to obtain descriptive knowledge with respect to labelled data. All techniques represented in this concept have as their objective the understanding of underlying phenomena instead of the classification of new instances.

An illustrative example for an SDRD model:

> *A medical center wants to know in what circumstances a patient may suffer a certain type of cancer; the intention is not to predict cancer, but to understand the risk factors that lead to this.*

In Fig. 1 examples of the predictive supervised, descriptive unsupervised and SDRD tasks are presented in order to show the main differences and properties of the tasks included in the SDRD concept:

- Fig. 1(a) represents graphically the model obtained by a predictive algorithm based on the extraction of rules for classification. As can be observed, six rules (areas between dotted lines) divide the space into different areas that allow analysis of the problem in an easy way. In this way, the model is able to classify new instances of the problem with good values of precision.

- The model presented in Fig. 1(b) is an unsupervised descriptive model, e.g. clustering that groups unlabelled instances in different areas (circles). As can be observed, the model obtains three groups of instances with a soft overlapping between the lower and the remaining groups, with a simple and single interpretation for each group.

- On the other hand, Fig. 1(c) presents an SDRD model, where two rules (circles) for each value of the target variable are obtained. Rules are usually represented in a similar way to Fig. 1(a). Another important property is that the knowledge for each rule is considered as individual knowledge instead of rules dependant on one another. There is a possibility of overlapping between rules, as can be observed in the rules for the blue target value.

3. Subgroup discovery

The SD was introduced by Kloesgen (Kloesgen, 1996) and Wrobel (Wrobel, 1997). Its objective is to discover interesting relationships between different objects in a set with respect to a property of interest. The patterns extracted are normally represented through rules (Gamberger and Lavrac, 2002), such as:

$$R = Class \leftarrow Cond$$

where *Cond* is a conjunction of attribute-value pairs and *Class* the property of interest. These patterns were called subgroups by Siebes (Siebes, 1995).

There is no consensus within SD about the use of a concrete quality measure, however the weighted accuracy relative (*WRAcc*) is the one most employed in the literature. This quality measure was defined as (Lavrac et al., 1999):

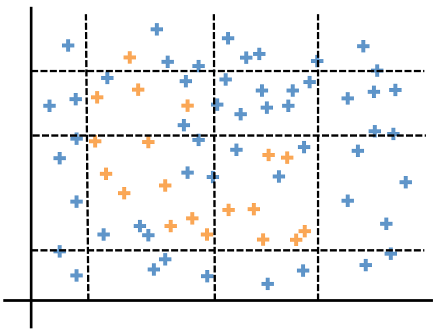
$$\begin{aligned} &WRAcc(Class \leftarrow Cond) = \\ &p(Cond) \cdot (p(Class|Cond) - p(Class)) \end{aligned} \tag{1}$$

where a balance between generality, precision and gain accuracy is considered.

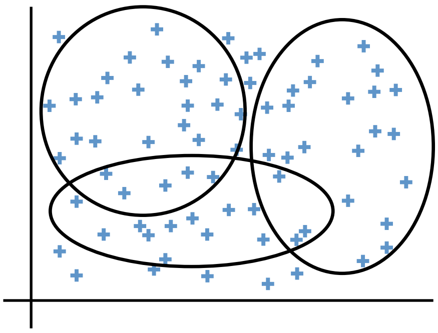
From the inception of the SD concept in 1996 there has been a wide application, especially in the last decades with the appearance of different approaches and applications in real-world problems. In fact, in (Herrera et al., 2011) a complete review of SD, its algorithms and applications was presented in order to show the community its importance; and recently, some reviews one focused on evolutionary algorithms (Carmona et al., 2014), another focused on exhaustive algorithms (Atzmueller, 2015), and in a empirical evaluation (Helal, 2016), have been presented.

In summary, SD algorithms can be classified according to the search algorithm employed in order to obtain rules, such as:

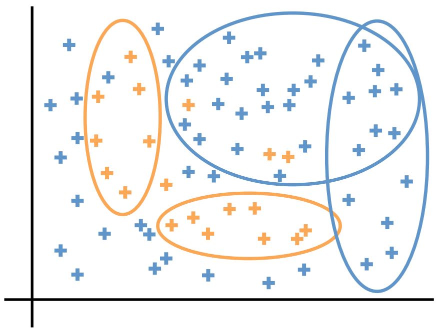
1. *Heuristic algorithms*: Within this group can be found CN2-SD (Lavrac et al., 2004), which is one of the exponent within SD, and the pioneering approaches EXPLORA (Kloesgen, 1996) and MIDOS (Wrobel, 1997). On the other hand, there are a large number of approaches based on soft computing techniques such as NMEEFSD (Carmona et al., 2010a), amongst others (Carmona et al., 2015, del Jesus et al., 2007a,b, Luna et al., 2013, Pachón et al., In Press, Rodríguez et al., 2012). Recently, a new evolutionary fuzzy system for big data environments has been presented in (Pulgar-Rubio et al., 2017) or for example, SDIGA (del Jesus et al., 2007b), MESDIF (del Jesus et al., 2007a), FuGePSD (Carmona et al., 2015), G3P (Luna et al., 2013), EDER-SD (Rodríguez et al., 2012), or GAR-SD (Pachón et al., In Press)



(a) Classification model



(b) Clustering model



(c) Supervised Descriptive Rule Model

Figure 1: Representation of data mining techniques with different types of induction

amongst others. Recently, a new evolutionary fuzzy system for big data environments called MEFASD-BD has been presented in (Pulgar-Rubio et al., 2017).

2. *Exhaustive algorithms*: In this group the most relevant algorithms are the Apriori-SD (Kavsek and Lavrac, 2006) and the SD-Map (Atzmueller and Puppe, 2006), although there are other interesting approaches such as Merge-SD (Grosskreutz and Rueping, 2009), CG (Zimmerman and de Raedt, 2009) or GP-Growth (Lemmerich et al., 2012). In addition, within this group, there are some algorithms that are able to work with numerical target variables such as SD-Map$*$ (Atzmueller and Lemmerich, 2009) and NumBSD (Lemmerich et al., 2016).

In the pioneering papers of SD, a wide applicability of the algorithms to different real-world problems such as medicine or bioinformatics was presented (Herrera et al., 2011). In fact, the interesting properties of SD have continued demonstrating its ability to obtain novelty knowledge in such disparate areas as educational data mining (Carmona et al., 2010b, 2011, Poitras et al., 2016a), bioinformatics and medicine (Carmona et al., 2015, Lavrac and Kralj-Novak, 2013, Liu et al., 2015, Poitras et al., 2016b), industry and technology (Almeida and Soares, 2013, Carmona et al., 2013, Jin et al., 2014, Konijn et al., 2013) or commerce (Brito et al., 2015, Carmona et al., 2012, Gamberger et al., 2013, Rodríguez et al., 2012, 2013), for example.

4. Emerging patterns

The EPs were defined by Dong and Li (Dong and Li, 1999, 2005) as itemsets whose support increases significantly from one dataset (D_1) to another (D_2) in order to discover trends in data, time or differentiating between features. In this way, a pattern is emerging if it has a growth rate (GR) higher than one and it is defined as (Dong and Li, 1999):

$$GR(x) = \begin{cases} 0, & IF\ Supp_{D_1}(x) = Supp_{D_2}(x) = 0, \\ \infty, & IF\ Supp_{D_2}(x) = 0 \wedge \\ & \quad Supp_{D_1}(x) \neq 0, \\ \dfrac{Supp_{D_1}(x)}{Supp_{D_2}(x)}, & another\ case \end{cases} \tag{2}$$

where $Supp_{D_1}(x)$ is the support for the pattern x in the first dataset and $Supp_{D_2}(x)$ is the support with respect to the second dataset, i.e. $Supp_{D_1}(x) = \dfrac{count_{D_1}(x)}{|D_1|}$ and $Supp_{D_2}(x) = \dfrac{count_{D_2}(x)}{|D_2|}$.

These patterns can be associated to datasets with classes and they are usually represented as pairs with a variable (Var) and a value ($value$) for this variable. Pairs are connected through conjunctions such as (Dong and Li, 1999):

$$x = \{Var_1 = value_1\}, \ldots, \{Var_n = value_n\}$$

The search space is related directly to the complexity of the dataset and in this way the number of EPs obtained by one algorithm could become huge. Throughout the literature there have been attempts to filter the number of patterns extracted with the use of different concepts or filtering operators such as jumping EPs (Dong et al., 1999a), essential EPs (Fan and Ramamohanarao, 2002), strong EPs (Fan and Ramamohanarao, 2006), maximal EPs (Wang et al., 2005) or noisy EPs (Fan and Ramamohanarao, 2006), negative EPs (Terlecki and Walczak, 2007), chi EPs (Fan and Ramamohanarao, 2004), shared EPs (Chen and Zhang, 2013), fuzzy EPs (García-Borroto et al., 2011), or disjunctive EPs (Vimieriro and Moscato, 2014), amongst others.

These concepts have been joined with different search strategies in order to obtain efficient EPs. In summary, the algorithms for extracting EPs can be grouped into:

- *Algorithms based on borders*: A border defines a pair of minimal and maximal patterns $< L, R >$ in order to represent all the patterns within this border (Dong and Li, 1999). Each element of L is a subset of some element in R and each element of R is a superset of some element in L. The pioneering algorithms of EPs employ this concept in order to discover all the EPs of a problem, for example, DeEPs (Li et al., 2004), CAEP(Dong et al., 1999b), BCEP (Ramamohanarao and Fan, 2007) and JEPC algorithm (Ramamohanarao et al., 2001).

- *Algorithms based on trees:* These algorithms employ this type of structure in order to optimise the complexity related to the search space with the border concept. Within this group there are two different subtypes: trees used to mine association rules such as Strong-JEP (Fan and Ramamohanarao, 2006), Tree-based JEP (Bailey et al., 2002), Top-k minimals JEP (Terlecki and Walczak, 2008), DCGP-Tree (Liu et al., 2014), or DFP-SEPSF (Alvai and Hashemi, 2015). On the other hand, there is another group of algorithms based on decision trees such as Fuzzy-EP (García-Borroto et al., 2011) and LCMine (García-Borroto et al., 2010).

- *Evolutionary approaches:* Only one preliminary proposal has been presented in (García-Vico et al., 2016), the EvAEP algorithm which is an evolutionary fuzzy system for extracting EPs.

A complete review of these strategies is presented in (García-Borroto et al., 2014). As can be observed, the use of EPs is mainly focused on the classification task because this type of methodology has a very interesting differentiating character in spite of the fact that the EP concept was defined for descriptive problems. In recent years, there has been increasing interest in the analysis of real-world problems based on EPs from a predictive point of view such as streaming data (Akhriza et al., 2015, Alavi and Hashemi, 2014, Park et al., 2010, Yu et al., 2015, 2012), sequential data (Barreto and Antunes, 2014, Desai and Ganatra, 2015, Nofong et al., 2014), technology (Acosta-Mendoza et al., 2016, Ding et al., 2010, Gu et al., 2011a,b, Kobylinski and Walczak, 2010, Li and Zhou, 2016, Yu et al., 2014a,b) and bioinformatics (Asses et al., 2012, Chen and Chen, 2011, Gardiner and Gillet, 2015, Loglisci et al., 2015, Métivier et al., 2015, Sherhod et al., 2013, 2012, Tzanis et al., 2011), amongst others.

5. Contrast sets

The CS technique was defined by Bay and Pazzani (Bay and Pazzani, 2001) as finding patterns as conjunctions of attributes and values that differ meaningfully in their distributions across groups (G_1, G_2, ..., G_i). It is important to remark that the groups must be exclusive among them, i.e. the instances can only belong to one group.

A pattern (x) is considered as CS if there is a significant difference of support (DS) between the support of the groups (Bay and Pazzani, 2001):

$$\begin{array}{l} \exists_{ij} \ where \ P(x = True \mid G_i) \ \neq \ P(x = True \mid G_j) \\ DS(x) = max_{ij} \ |Sup(x, G_i) - Sup(x, G_j)| \ \geq \ \delta \end{array} \quad (3)$$

where $Sup(x, G_i)$ is the support for the pattern x in the i^{th} group and $Sup(x, G_j)$ for the j^{th} group. The δ value is the minimum threshold (minimum difference needed) in order to consider a pattern as contrast. The 0.10 value is usually employed.

The CSs are represented as conjunctions of pairs variable-value ($Var = value$) such as (Bay and Pazzani, 2001):

$$x = \{Var_1 = value_1\} \wedge \ldots \wedge \{Var_n = value_n\}$$

CSs have traditionally been the least extensive task within SDRD. Nonetheless, there is a large number of algorithms throughout the literature (Boettcher, 2011). With respect to the algorithms presented, they can be classified into:

- *Algorithms based on trees:* Within this group are the more well-known in CSs such as STUCCO (Bay and Pazzani, 2001), CIGAR (Hilderman and Peckham, 2005, 2007) and Magnum-Opus (Webb et al., 2003). In (Webb, 2007) different concepts were presented in order to improve the quality of the patterns extracted in these algorithms such as productivity or the importance of avoiding false discoveries. Other approaches with tree structures are (Morita et al., 2009, Simeon and Hilderman, 2007, 2011a,b, Simeon et al., 2012, Zhu et al., 2015).

- *Approaches based on association rules:* The most relevant algorithm within this group is the algorithm RCS (Azevedo, 2010). It employs statistical tests and a pruning procedure based on preservation of support in order to obtain the most significant patterns.

The search for CSs in real-world problems has not been extended in the literature and it has been focused mainly on medicine (Kralj-Novak et al., 2009a, Li and Yang, 2007, Reps et al., 2015), enterprises (Wei et al., 2013) and social studies (Magalhaes and Azevedo, 2015). However, it is very important to note the existing relations between CS and other very close (although relatively novel) concepts, such as change mining (Liu et al., 2001), discriminative patterns (Fang et al., 2012, He et al., 2017, Kameya and Sato, 2012), correlated pattern (Morishita and Sese, 2000), closed sets (Garriga et al., 2008) and collaborative patterns (Zhu et al., 2011), amongst others.

6. Compatibilities of concepts and terms for supervised descriptive rule discovery tasks

All of these different techniques grouped by SDRD concept were developed in different communities, each developing their own terminology. However, CSs, EPs and SD show that terms used in different communities are compatible, according to the their definitions. These compatibilities were defined in (Kralj-Novak et al., 2009b):

Specifically, Table 1 provides a dictionary of equivalent terms from CSs, EPs and SD, in a unifying terminology of classification rule learning, and in particular of concept learning (considering class C_i as the concept

Table 1: Table of synonyms from different communities, showing the compatibility of terms

Contrast Sets	Emerging Patterns	Subgroup Discovery	Rule Learning
contrast set	itemset	subgroup description	rule condition
groups $G_1, \dots, G_n$	datasets D_1 and D_2	class/property C	class/concept C_i
attribute value pair	item	binary feature	condition
examples in groups	transactions in datasets	examples	examples of $C_1, \dots, C_n$
examples for which the CS is true	transactions containing the itemset	subgroup of instances	covered examples
support of CS on G_i and G_j	support of EP in D_1 and D_2	true/false positive rate	true/false positive rate

to be learned from the positive examples of this concept, and the negative examples formed of examples of all other classes).

Once, we have established the compatibility among the terminologies, now it is provided the similarities between the definitions of these techniques. As we have defined previously:

- **CSs**. Let $A_1, A2, \dots, A_k$ be a set of k variables called attributes. Each A_i can take values from the set $\{v_{i1}, v_{i2}, \dots, v_{im}\}$. Given a set of user defined groups $G_1, G_2, \dots, G_n$ of data instances, a contrast set is a conjunction of attribute-value pairs, defining a pattern that best discriminates the instances of different user-defined groups. A special case of contrast set mining considers only two contrasting groups (G_1 and G_2). In such cases, we wish to find characteristics of one group discriminating it from the other and vice versa.

- **SD**. A subgroup is described as conjunctions of features. Given the property of interest C, and the population of examples of C and $\overline{C}$, the SD aims at finding population subgroups that are as large as possible and have the most unusual statistical (distributional) characteristics with respect to the property of interest C.

- **EPs**. Let $I = \{i_1, i_2, \dots, i_N\}$ be a set of items (equivalent to a binary feature in SD, and an individual attribute-value pair in CSs). A transaction is a subset T of I. A dataset is a set D of transactions. A subset X of I is called an itemset. Transaction T contains an itemset X in a dataset D, if $X \subseteq T$. For two data sets D_1 and D_2, EPs aims at discovering itemsets whose support increases significantly from one data set to another.

Instead of the definitions appear different we can observe that the goals of these descriptive data mining techniques are similar: *to search for discriminating characteristics, emerging trends and/or subgroup descriptions between different values of a class property*. In this way, there is a compatibility between concepts as can be observed in Table 2.

The knowledge is represented through patterns (x) or rules (R) and in general they are described through pairs attribute-value and a class or target variable. In summary, the rule R can be formally defined as:

$$R : Cond \ \rightarrow \ Class$$

$Cond$ is the conjunction of features (attribute-value pairs), and $Class$ is a value for the variable of interest or class. This $Class$ is represented through G_1, D_1 and C, i.e. the positive examples (P) of the problem, and G_2, D_2 and $\overline{C}$ represents the negative examples (N).

7. Heuristics in supervised descriptive rule discovery tasks

The main quality measures used throughout the literature for the different techniques within the SDRD employ different nomenclatures and it is necessary to homogenize them in order to analyse in a better way. However, most measures are derived by analysing the covering properties of the rule and the class in the rule consequent considered as positive. This relationship can be depicted by a confusion matrix as can be observed in Table 3.

Table 3: Confusion matrix for a rule

	Predicted condition		
True condition	Positive	Negative	
Positive	$p = tp$	$\overline{p} = fn$	$p + \overline{p} = P$
Negative	$n = fp$	$\overline{n} = tn$	$n + \overline{n} = N$
	$p + n$	$\overline{p} + \overline{n}$	$P + N = T$

Table 2: Compatibility of definitions

	CSs	EPs	SD	SDRD
Knowledge	pattern (x)	pattern (x)	subgroup (R)	rule (R)
Cond	itemsets	itemsets	pairs att-val	pairs att-val
Class	$G_1, \ldots, G_i$	D_1, D_2	$C, \overline{C}$	P, N

The confusion matrix for a rule represents the following information:

- p number of examples correctly covered,
- $\overline{p}$ number of examples for the class not covered,
- n number of examples incorrectly covered,
- $\overline{n}$ number of examples not covered for the non-class,
- $p + n$ number of examples covered for the rule,
- $\overline{p} + \overline{n}$ number of examples not covered for the rule, and
- $P = p + \overline{p}$ number of examples for the positive class. Examples containing a concrete value for the target variable are considered.
- $N = n + \overline{n}$ number of examples for the negative class. Examples for the remaining values of the target variable are included.
- $T = P + N$ number of examples for the whole dataset.

Next, the original quality measures presented in the pioneering papers for each approach are shown:

- Contrast Sets:

$$DS(x) = max_{ij} \ |Sup(x, G_i) - Sup(x, G_j)| \ \geq \ \delta \tag{4}$$

- Emerging Patterns:

$$GR(x) = \frac{supp_{D1}(x)}{supp_{D2}(x)} \tag{5}$$

- Subgroup Discovery:

$$WRAcc(Class \leftarrow Cond) = p(Cond) \cdot (p(Class|Cond) - p(Class)) \tag{6}$$

As can be observed, there are different nomenclatures for each one but nevertheless they employ very close concepts. Next, we present the quality measures modified with the use of the matrix presented in Table 3:

- Contrast Sets:

$$\begin{aligned} DS(R) &= |Sup(R, PIS) - Sup(R, NIS)| \ \geq \ \delta \\ DS(R) &= \left|\frac{p}{P} - \frac{n}{N}\right| \ \geq \ \delta \end{aligned} \tag{7}$$

- Emerging Patterns:

$$\begin{aligned} GR(R) &= \frac{Sup(R, PIS)}{Sup(R, NIS)} \\ GR(R) &= \frac{\frac{p}{P}}{\frac{n}{N}} > 1 \end{aligned} \tag{8}$$

- Subgroup Discovery:

$$\begin{aligned} WRAcc(R) &= p(Cond) \cdot (p(Class \cdot Cond) - p(Class)) \\ WRAcc(R) &= \frac{p+n}{P+N}\left(\frac{p}{p+n} - \frac{P}{P+N}\right) \end{aligned} \tag{9}$$

These equations present some interesting assertions regarding the possible future study and analysis of the different techniques from different point of view:

1. The GR and the DS are directly related because when the DS is positive the GR is upper than one. In this way, we could confirm that a CSs is emerging, but an EPs is contrasting rule only when the differences between the positives rate and negative rate is upper than α.
2. The $WRAcc$ is a more complicated quality measure but it is interesting to see that the value of this measure is positive when the accuracy of the rule is upper than the percentage of the examples of the class.

Acknowledgment

This work was partially supported by the Spanish Ministry of Economy and Competitiveness under the project TIN2015-68454-R (FEDER Founds).

References

Acosta-Mendoza, N., Gago-Alonso, A., Carrasco-Ochoa, J.A., Martnez-Trinidad, J.F., E.Medina-Pagola, J., 2016. Improving graph-based image classification by using emerging patterns as attributes. Engineering Applications of Artificial Intelligence 50, 215–225.

Akhriza, T.M., Ma, Y., Li, J., 2015. A novel fibonacci windows model for finding emerging patterns over online data stream, in: Proc. of the 2015 International Conference on Cyber Security of Smart cities, Industrial Control System and Communications, pp. 1–8.

Alavi, F., Hashemi, S., 2014. Mining jumping emerging patterns by streaming feature selection., in: Proc. of the 5th International Conference Knowledge and Systems Engineering, Springer. pp. 337–349.

Almeida, S., Soares, C., 2013. CN2-SD for Subgroup Discovery in a Highly Customized Textile Industry: A Case Study, in: Advances in Sustainable and Competitive Manufacturing Systems, Springer. pp. 585–595.

Alvai, F., Hashemi, S., 2015. DFP-SEPSF: A dynamic frequent pattern tree to mine strong emerging patterns in streamwise features. Engineering Applications of Artificial Intelligence 37, 54–70.

Asses, Y., Buzmakov, A., Bourquard, T., Kuznetsov, S., Sergei, O., Napoli, A., 2012. A hybrid classification approach based on fca and emerging patterns - an application for the classification of biological inhibitors, pp. 211–222.

Atzmueller, M., 2015. Subgroup discovery. WIREs Data Mining and Knowledge Discovery 5, 35–49.

Atzmueller, M., Lemmerich, F., 2009. Fast Subgroup Discovery for Continuous Target Concepts, in: Proc. of the 18th International Symposium on Methodologies for Intelligent Systems, Springer. pp. 35–44.

Atzmueller, M., Puppe, F., 2006. SD-Map - A Fast Algorithm for Exhaustive Subgroup Discovery, in: Proc. of the 17th European Conference on Machine Learning and 10th European Conference on Principles and Practice of Knowledge Discovery in Databases, Springer. pp. 6–17.

Azevedo, P.J., 2010. Rules for contrast sets. Intelligent Data Analysis 14, 623–640.

Bailey, J., Manoukian, T., Ramamohanarao, K., 2002. Fast Algorithms for Mining Emerging Patterns, in: Principles of Data Mining and Knowledge Discovery. Springer. volume 2431, pp. 187–208.

Barreto, A., Antunes, C., 2014. Finding cyclic patterns on sequential data, in: Proc. of the 10th International Conference on Active Media Technology, pp. 110–121.

Bay, S.D., Pazzani, M.J., 2001. Detecting group differences: Mining contrast sets. Data Mining and Knowledge Discovery 5, 213–246.

Boettcher, M., 2011. Contrast and change mining. WIREs Data Mining and Knowledge Discovery 1, 215–230.

Brito, P., Soares, C., Almeida, S., Monte, A., Byvoet, M., 2015. Customer Segmentation in a Large Database of an Online Customized Fashion Business. Robotics and Computer-Integrated Manufacturing 36, 93–100.

Carmona, C.J., González, P., García-Domingo, B., del Jesus, M.J., Aguilera, J., 2013. MEFES: An evolutionary proposal for the detection of exceptions in subgroup discovery. An application to Concentrating Photovoltaic Technology. Knowledge-Based Systems 54, 73–85.

Carmona, C.J., González, P., del Jesus, M.J., Herrera, F., 2010a. NMEEF-SD: Non-dominated Multi-objective Evolutionary algorithm for Extracting Fuzzy rules in Subgroup Discovery. IEEE Transactions on Fuzzy Systems 18, 958–970.

Carmona, C.J., González, P., del Jesus, M.J., Herrera, F., 2014. Overview on evolutionary subgroup discovery: analysis of the suitability and potential of the search performed by evolutionary algorithms. WIREs Data Mining and Knowledge Discovery 4, 87–103.

Carmona, C.J., González, P., del Jesus, M.J., Romero, C., Ventura, S., 2010b. Evolutionary algorithms for subgroup discovery applied to e-learning data, in: Proc. of the IEEE International Education Engineering, pp. 983–990.

Carmona, C.J., González, P., del Jesus, M.J., Ventura, S., 2011. Subgroup discovery in an e-learning usage study based on Moodle, in: Proc. of the International Conference of European Transnational Education, pp. 446–451.

Carmona, C.J., Ramírez-Gallego, S., Torres, F., Bernal, E., del Jesus, M.J., García, S., 2012. Web usage mining to improve the design of an e-commerce website: OrOliveSur.com. Expert Systems with Applications 39, 11243–11249.

Carmona, C.J., Ruiz-Rodado, V., del Jesus, M.J., Weber, A., Grootveld, M., González, P., Elizondo, D., 2015. A fuzzy genetic programming-based algorithm for subgroup discovery and the application to one problem of pathogenesis of acute sore throat conditions in humans. Information Sciences 298, 180–197.

Chen, X., Chen, J., 2011. Emerging patterns and classification algorithms for dna sequence. Journal of Software 6, 985–992.

Chen, X., Zhang, X., 2013. Similarity measure by aggregating shared emerging pattern, in: Proc. of the 5th International Conference on Computational and Information Sciences, IEEE. pp. 802–805.

Desai, N., Ganatra, A., 2015. Incorporating boundary value concept and recency constraint to capture emerging trends in time stamp based sequence dataset, in: Proc. of the 2015 International Conference on Communication, Information and Computing Technology, pp. 1–7.

Ding, G., Wang, J., Qin, K., 2010. A visual word weighting scheme based on emerging itemsets for video annotation. Information Processing Letters 110, 692–696.

Dong, G.Z., Li, J.Y., 1999. Efficient Mining of Emerging Patterns: Discovering Trends and Differences, in: Proc. of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press. pp. 43–52.

Dong, G.Z., Li, J.Y., 2005. Mining border descriptions of emerging patterns from dataset pairs. Knowledge and Information Systems 8, 178–202.

Dong, G.Z., Li, J.Y., Zhang, X., 1999a. Discovering jumping emerging patterns and experiments on real datasets, in: Proc. on International Database Conference Heterogeneous and Internet Databases, pp. 155–168.

Dong, G.Z., Zhang, X., Wong, L., Li, J.Y., 1999b. CAEP: Classification by Aggregating Emerging Patterns, in: Proc. of the Discovery Science, Springer. pp. 30–42.

Fan, H., Ramamohanarao, K., 2002. An efficient single-scan algorithm for mining essential jumping emerging patterns for classification, in: Proc. on the 6th Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 456–462.

Fan, H., Ramamohanarao, K., 2004. Noise Tolerant Classification by Chi Emerging Patterns, in: Proc. of the 8th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer. pp. 201–206.

Fan, H., Ramamohanarao, K., 2006. Fast discovery and the generalization of strong jumping emerging patterns for building compact and accurate classifiers. IEEE Transactions on Knowledge and Data Engineering 18, 721–737.

Fang, G., Pandey, G., Wang, W., Gupta, M., Steinbach, M., Kumar, V., 2012. Mining low-support discriminative patterns from dense and high-dimensional data. IEEE Transactions on Knowledge Data Engineering 24, 279–294.

Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., 1996. From data mining to knowledge discovery: an overview, in: Advances in knowledge discovery and data mining. AAAI/MIT Press, pp. 1–34.

Gamberger, D., Lavrac, N., 2002. Expert-Guided Subgroup Discovery: Methodology and Application. Journal Artificial Intelligence Research 17, 501–527.

Gamberger, D., Lucanin, D., Smuc, T., 2013. Analysis of World Bank Indicators for Countries with Banking Crises by Subgroup Discovery Induction, in: Proc. on the 36th International Convention on Information and Communication Technology, Electronics and Microelectronics, IEEE. pp. 1138–1142.

García-Borroto, M., Martínez-Trinidad, J., Carrasco-Ochoa, J., 2011. Fuzzy emerging patterns for classifying hard domains. Knowledge and Information Systems 28, 473–489.

García-Borroto, M., Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A., 2014. A survery of emerging patters for supervised classification. Artificial Intelligence Review 42, 705–721.

García-Borroto, M., Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A., Medina-Pérez, M.A., Ruiz-Shulcloper, J., 2010. LCMine: An efficient algorithm for mining discriminative regularities and its application in supervised classification. Pattern Recognition 43, 3025–3034.

García-Vico, A.M., Montes, J., Aguilera, J., Carmona, C.J., del Jesus, M.J., 2016. Analysing Concentrating Photovoltaics Technology through the use of Emerging Pattern Mining, in: Proc. of the 11th International Conference on Soft Computing Models in Industrial and Environmental Applications, Springer. pp. 1–8.

Gardiner, E., Gillet, V.J., 2015. Perspectives on knowledge discovery algorithms recently introduced in chemoinformatics: Rough set theory, association rule mining, emerging patterns, and formal concept analysis. Journal of Chemical Information and Modeling 55, 1781–1803.

Garriga, G., Kralj-Novk, P., Lavrac, N., 2008. Closed sets for labeled data. Journal of Machiche Learning Research 9, 559–580.

Grosskreutz, H., Rueping, S., 2009. On Subgroup Discovery in Numerical Domains. Data Mining and Knowledge Discovery 19, 210–216.

Gu, T., Wang, L., Chen, H., Tao, X., Lu, J., 2011a. Recognizing multiuser activities using wireless body sensor networks. IEEE Transactions on Mobile Computing 10, 1618–1631.

Gu, T., Wang, L., Wu, Z., Tao, X., Lu, J., 2011b. A Pattern Mining Approach to Sensor-Based Human Activity Recognition. IEEE Transactions on Knowledge and Data Engineering 23, 1359–1372.

He, Z., Gu, F., Zhao, C., Liu, X., Wu, J., Wang, J., 2017. Conditional discriminative pattern mining: Concepts and algorithms. Information Sciences 375, 1–25.

Helal, S., 2016. Subgroup Discovery Algorithms: A Survey and Empirical Evaluation. Journal of Computer Science and Technology 31, 561–576.

Herrera, F., Carmona, C.J., González, P., del Jesus, M.J., 2011. An overview on Subgroup Discovery: Foundations and Applications. Knowledge and Information Systems 29, 495–525.

Hilderman, R.J., Peckham, T., 2005. A Statistically Sound Alternative Approach to Mining Contrast Sets, in: Proc. of the fourth Australasian Data Mining Conference, pp. 157–172.

Hilderman, R.J., Peckham, T., 2007. Statistical methodologies for mining potentially interesting contrast sets, in: Quality Measures in Data Mining, pp. 153–177.

del Jesus, M.J., González, P., Herrera, F., 2007a. Fuzzy Sets and Their Extensions: Representation, Aggregation and Models. Springer. volume 220. chapter Subgroup Discovery with Linguistic Rules. pp. 411–430.

del Jesus, M.J., González, P., Herrera, F., Mesonero, M., 2007b. Evolutionary Fuzzy Rule Induction Process for Subgroup Discovery: A case study in marketing. IEEE Transactions on Fuzzy Systems 15, 578–592.

Jin, N., Flach, P., Wilcox, T., Sellman, R., Thumim, J., Knobbe, A., 2014. Subgroup discovery in smart electricity meter data. IEEE Transactions on Industrial Informatics 10, 1327–1336.

Kameya, Y., Sato, T., 2012. RP-growth: Top-k Mining of Relevant Patterns with Minimum Support Raising, in: Proc. of the 2012 SIAM International Conference on Data Mining, pp. 816–827.

Kavsek, B., Lavrac, N., 2006. APRIORI-SD: Adapting association rule learning to subgroup discovery. Applied Artificial Intelligence 20, 543–583.

Kloesgen, W., 1996. Explora: A Multipattern and Multistrategy Discovery Assistant, in: Advances in Knowledge Discovery and Data Mining. American Association for Artificial Intelligence, pp. 249–271.

Kobylinski, L., Walczak, K., 2010. Spatial Emerging Patterns for Scene Classification, in: Proc. of the 10th International Conference on Artificial Intelligence and Soft Computing, Elsevier. pp. 515–522.

Konijn, R., Duivesteijn, W., Kowalczyk, W., Knobbe, A.., 2013. Discovering Local Subgroups, with an Application to Fraud Detection, in: 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 1–12.

Kralj-Novak, P., Lavrac, N., Gamberger, D., Krstacic, A., 2009a. Csm-sd: Methodology for contrast set mining through subgroup discovery. Journal of Biomedical Informatics 42, 113–122.

Kralj-Novak, P., Lavrac, N., Webb, G.I., 2009b. Supervised Descriptive Rule Discovery: A Unifying Survey of Constrast Set, Emerging Pateern and Subgroup Mining. Journal of Machine Learning Research 10, 377–403.

Lavrac, N., Cestnik, B., Gamberger, D., Flach, P.A., 2004. Decision Support Through Subgroup Discovery: Three Case Studies and the Lessons Learned. Machine Learning 57, 115–143.

Lavrac, N., Flach, P.A., Zupan, B., 1999. Rule Evaluation Measures: A Unifying View, in: Proc. of the 9th International Workshop on Inductive Logic Programming, Springer. pp. 174–185.

Lavrac, N., Kralj-Novak, P., 2013. Relational and Semantic Data Mining for Biomedical Research. Informatica 37, 35–39.

Lemmerich, F., Atzmueller, M., Puppe, F., 2016. Fast exhaustive subgroup discovery with numerical target concepts. Data Mining and Knowledge Discovery 30, 711–762.

Lemmerich, F., Becker, M., Atzmueller, M., 2012. Generic pattern trees for exhaustive exceptional model mining, in: Proc. of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Springer. pp. 277–292.

Li, J., Yang, Q., 2007. Strong compound-risk factors: Efficient discovery through emerging patterns and contrast sets. IEEE Transactions on Information Technology in Biomedicine 11, 544–552.

Li, J.Y., Dong, G.Z., Ramamohanarao, K., Wong, L., 2004. DeEPs: A New Instance-Based Lazy Discovery and Classification System. Machine Learning 54, 99–124.

Li, Y., Zhou, H., 2016. Bagging eEP-basecl classifiers for junk mail classification. International Journal of Security and its Applications 10, 121–128.

Liu, B., Hsu, W., Ma, Y., 2001. Discovering the set of fundamental rule changes, in: Proc. of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 335–340.

Liu, Q., Shi, P., Hu, Z., Zhang, Y., 2014. A novel approach of mining strong jumping emerging patterns based on BSC-tree. International Journal of Systems Science 45, 598–615.

Liu, X., Wu, J., Gu, F., Wang, J., He, Z., 2015. Discriminative pattern mining and its applications in bioinformatics. Briefings in Bioinformatics 16, 884–900.

Loglisci, C., Balech, B., Malerba, D., 2015. Discovering variability patterns for change detection in complex phenotype data, in: Proc. of the 22nd International Symposium on Methodologies for Intelligent Systems, pp. 9–18.

Luna, J.M., Romero, J.R., Romero, C., Ventura, S., 2013. Discovering

Subgroups by Means of Genetic Programming, in: Proc. of the 16th European Conference on Genetic Programming, Springer. pp. 121–132.

Magalhaes, A., Azevedo, P.J., 2015. Contrast set mining in temporal databases. Expert Systems 32, 435–443.

Métivier, J.P., Lepailleur, A., Buzmakov, A., Poezevara, G., Crémilleux, B., Kuznetsov, S.O., Goff, J., Napoli, A., Bureau, R., Cuissart, B., 2015. Journal of Chemical Information and Modeling 55, 925–940.

Morishita, S., Sese, J., 2000. Traversing itemset lattices with statistical metric pruning, in: Proc. of the 19th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, pp. 226–236.

Morita, H., Nakahara, T., Hamuro, Y., Yamamoto, S., 2009. Decision tree-based classifier incorporating contrast patterns, in: IEEE International Symposium on Consumer Electronics.

Nofong, V.M., Liu, J., Li, J., 2014. A Study on the Applications of Emerging Sequential Patterns, in: Proc. of the 25th Australasian Database Conference, Springer. pp. 62–73.

Pachón, V., Mata, J., Domínguez, J., In Press. Searching for the most significant rules: an evolutionary approach for subgroup discovery. Soft Computing , 1–10.

Park, J., Lee, H., Park, J., 2010. Real-time Diagnosis System Using Incremental Emerging Pattern Mining, in: Proc. of the 5th International Conference on Ubiquitous Information Technologies and Applications, pp. 1–5.

Poitras, E.G., Lajoie, S.P., Doleck, T., Jarrel, A., 2016a. Subgroup discovery with user interaction data: An empirically guided approach to improving intelligent tutoring systems. Educational Technology and Society 19, 204–214.

Poitras, E.G., Naismith, L.M., Doleck, T., Lajoie, S.P., 2016b. Using learning analytics to identify medical student misconceptions in an online virtual patient environment. Journal of Asynchronous Learning Network 20, 1–12.

Pulgar-Rubio, F., Rivera-Rivas, A.J., Pérez-Godoy, M.D., González, P., Carmona, C.J., del Jesus, M.J., 2017. MEFASD-BD: Multi-Objective Evolutionary Fuzzy Algorithm for Subgroup Discovery in Big Data Environments - A MapReduce Solution. Knowledge-Based Systems 117, 70–78.

Ramamohanarao, K., Dong, G.Z., Li, J.Y., 2001. Making Use of the Most Expressive Jumping Emerging Patterns for Classification. Knowledge and Information Systems 3, 131–145.

Ramamohanarao, K., Fan, H., 2007. Patterns Based Classifiers. World Wide Web 10, 71–83.

Reps, J., Guo, Z., Zhu, H., Aickelin, U., 2015. Identifying candidate risk factors for prescription drug side effects using causal contrast set mining, in: Proc. of the 4th International Conference on Health Information Science, pp. 45–55.

Rodríguez, D., Ruíz, R., Riquelme, J.C., Aguilar-Ruiz, J.S., 2012. Searching for rules to detect defective modules: A subgroup discovery approach. Information Sciences 191, 14–30.

Rodríguez, D., Ruíz, R., Riquelme, J.C., Harrison, R., 2013. A study of subgroup discovery approaches for defect prediction. Information and Software Technology 55, 1810–1822.

Sherhod, R., Gillet, V.J., Hanser, T., Judson, P.N., Vessey, J.D., 2013. Toxicological knowledge discovery by mining emerging patterns from toxicity data. Journal of Chemical Information and Modeling 5, 9.

Sherhod, R., Gillet, V.J., Judson, P.N., Vessey, J.D., 2012. Automating knowledge discovery for toxicity prediction using jumping emerging pattern mining. Journal of Chemical Information and Modeling 52, 3074–3087.

Siebes, A., 1995. Data Surveying: Foundations of an Inductive Query Language, in: Proc. of the 1st International Conference on Knowledge Discovery and Data Mining, AAAI Press. pp. 269–274.

Simeon, M., Hilderman, R.J., 2007. Exploratory quantitative contrast set mining: A discretization approach, in: Proc. of the 14th IEEE International Conference of Tools with Artificial Intelligence, pp. 124–131.

Simeon, M., Hilderman, R.J., 2011a. COSINE: A vertical group difference approach to contrast set mining, in: Proc. of the 24th Canadian Conference on Advances in Artificial Intelligence, pp. 359–371.

Simeon, M., Hilderman, R.J., 2011b. GENCCS: A correlated group difference approach to contrast set mining, in: Proc. of the 7th International Machine Learning and Data Mining in Pattern Recognition, pp. 140–154.

Simeon, M., Hilderman, R.J., Hamilton, H.J., 2012. Mining interesting correlated contrast sets, in: Proc. of the 32th International Conference on Innovative Techniques and Applications of Artificial Intelligence, pp. 49–62.

Terlecki, P., Walczak, K., 2007. Jumping emerging patterns with negation in transaction databases classification and discovery. Information Sciences 177, 5675–5690.

Terlecki, P., Walczak, K., 2008. Efficient Discovery of Top-K Minimal Jumping Emerging Patterns, in: Proc. of the 6th International Conference Rough Sets and Current Trends in Computing, Elsevier. pp. 438–447.

Tzanis, G., Kavakiotis, I., Vlahavas, I.P., 2011. Polya-iep: A data mining method for the effective prediction of polyadenylation sites. Expert Systems with Applications 38, 12398–12408.

Vimieriro, R., Moscato, P., 2014. A new method for mining disjunctive emerging patterns in high-dimensional datasets using hypergraphs. Information Sciences 40, 1–10.

Wang, Z., Fan, H., Ramamohanarao, K., 2005. Exploiting Maximal Emerging Patterns for Classification, in: Proc. of the 17th Australian Joint Conference on Artificial Intelligence, Springer. pp. 1062–1068.

Webb, G.I., 2007. Discovering Significant Patterns. Machine Learning 67, 1–33.

Webb, G.I., Butler, S., Newlands, D., 2003. On detecting differences between groups, in: Proc. of the ninth International Conference on Knowledge Discovery and Data Mining, pp. 256–265.

Wei, W., Li, J., Cao, L., Ou, Y., Chen, J., 2013. Effective detection of sophisticated online banking fraud on extremely imbalanced data. World Wide Web 16, 449–475.

Wrobel, S., 1997. An Algorithm for Multi-relational Discovery of Subgroups, in: Proc. of the 1st European Symposium on Principles of Data Mining and Knowledge Discovery, Springer. pp. 78–87.

Yu, K., Ding, W., Simovici, D., Wang, H., Pei, J., Wu, X., 2015. Classification with streaming features: An emerging-pattern mining approach. ACM Transactions on Knowledge Discovery Data 9, 30:1–30:31.

Yu, K., Ding, W., Simovici, D., Wu, X., 2012. Mining emerging patterns by streaming feature selection, in: Proc. of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM. pp. 60–68.

Yu, Y., Yan, K., Zhu, X., Wang, G., 2014a. Detecting of PIU Behaviors Based on Discovered Generators and Emerging Patterns from Computer-Mediated Interaction Events, in: Proc. of the 15th International Conference on Web-Age Information Management, Elsevier. pp. 277–293.

Yu, Y., Yan, K., Zhu, X., Wang, G., Luo, D., Sood, S., 2014b. Mining Emerging Patterns of PIU from Computer-Mediated Interaction Events, in: Proc. of the 9th International Workshop on Agents and Data Mining Interaction, Elsevier. pp. 66–78.

Zhu, G., Wang, Y., Agrawal, G., 2015. Scicsm: novel contrast set mining over scientific datasets using bitmap indices, in: Proc. of the 27th International Conference on Scientific and Statistical Database Management, pp. 38:1–38:6.

Zhu, X., Li, B., Wu, X., He, D., Zhang, C., 2011. CLAP: collaborative pattern mining for distributed information systems. Decision Support Systems 52, 40–51.

Zimmerman, A., de Raedt, L., 2009. Cluster-grouping: from subgroup discovery to clustering. Machine Learning 77, 125–159.