Mining Emerging Gradual Patterns

Anne Laurent, Marie-Jeanne Lesot, Maria Rifqi

Mining Emerging Gradual Patterns

Anne Laurent[1] **Marie-Jeanne Lesot**[2] **Maria Rifqi**[3]

[1]LIRMM - Université Montpellier 2, Montpellier, France
[2] Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris, France
CNRS, UMR 7606, LIP6, F-75005, Paris, France
[3]LEMMA - Université Panthéon-Assas, Paris, France

Abstract

Mining emerging patterns aims at contrasting data sets and identifying itemsets that characterise a data set by contrast to a reference data set, so as to capture and highlight their differences. This paper considers the case of emerging gradual patterns, to extract discriminant attribute co-variations. It discusses the specific features of these gradual patterns and proposes to transpose an efficient border-based algorithm, justifying its applicability to the gradual case. Illustrative results obtained from a UCI data set are described.

Keywords: Gradual Patterns, Emerging Patterns, Discriminant Characterisation

1. Introduction

Gradual patterns [1, 2, 3] extract knowledge from numerical data bases as attribute co-variations allowing linguistic representations of the form "the more A increases, the more B increases", where A and B are numerical attributes. This paper addresses the task of mining emerging gradual patterns, defined as gradual patterns that describe a data set *by contrast to* a reference data set, *i.e.* occur in a data set but not the other one. Such patterns aim at characterising the specificity of the considered data in terms of attribute co-variations and at highlighting its differences with respect to reference data. In the case where the two data sets correspond to different dates, emerging gradual patterns allow to adapt to the data evolution over time and underline their changes.

As an example, one can consider the case of data that describe users of social networks through their age and the social tools they favour: comparing data after 2014 to data before 2014, emerging gradual patterns may for instance indicate that the pattern *the younger, the higher the use of Facebook* holds before 2014 but is then replaced by the *the younger, the higher the use of Snapchat*. Such an emerging gradual pattern could reflect the observation of the recent Global Social Media Impact Study based on the UCL Department of Anthropology and funded by the European union, according to which it has recently appeared that young people are less present on Facebook and prefer tools such as Snapchat.

This data mining task extends, to the case of gradual patterns, the notion of emerging patterns [4], which is intrinsically related to that of contrast set [5]: in both cases, the aim can be described as the identification of discriminant itemsets in transactional data sets, that occur in subsets of the data but not others.

Mining emerging patterns raises computational issues mainly due to non-monotone nature of the emergence property: the subpatterns of an emerging pattern are not necessarily emerging. The absence of monotonicity hinders the application of a classic generate-and-filter approach and it requires the development of dedicated approaches.

This paper considers the extraction of emerging patterns in the case of gradual patterns: it proposes to adapt an efficient border-based algorithm that has been proposed for classic patterns [4, 6] and that exploits a compact representation based on maximal frequent patterns. The paper justifies its applicability to the case of gradual patterns, discussing their specificity and making it possible to extract emerging gradual patterns. It illustrates the results obtained on the UCI vehicle data set [7].

The paper is organised as follows: Section 2 recalls the definitions of gradual patterns and of emerging patterns, as well as the efficient MBD-LL BORDER algorithm [4]. Section 3 justifies the transposition of the latter to the case of gradual patterns and Section 4 illustrates the obtained experimental results.

2. Gradual and emerging patterns

This section successively discusses two types of specific patterns that extract different types of knowledge from different types of data, namely gradual patterns and emerging patterns.

2.1. Gradual patterns

Whereas classic pattern mining applies to transactional data, described by binary attributes denoting the presence or absence of each item, gradual patterns (GP) [1, 2, 3] are extracted from numerical data, described by real values associated to numerical features. GP are linguistically expressed in the form "the more A increases, the more B increases", or equivalently, "the higher A, the higher B". They impose constraints across the whole data set and

must be distinguished from fuzzy gradual rules. Indeed, the latter impose constraints on the attribute values for each data point individually [8, 9, 10]. These rules are not considered further in this paper.

In the following, $\mathcal{D}$ denotes the considered data set, that contains n objects described by m numerical attributes; A denoting an attribute, for any object $x \in \mathcal{D}$, $A(x)$ denotes the value A takes for x.

2.1.1. Definitions

As given by [1, 2], the formal definitions of gradual items and gradual patterns are as follows:

Definition 1 *A* gradual item*, denoted $A^{\geq}$ or $A^{\leq}$, is a pair made of an attribute A and a variation denoted by $\geq$ or $\leq$.*

If A is an attribute corresponding to the user age for instance, $A^{\geq}$ and $A^{\leq}$ are the gradual items that can be linguistically expressed as *the older* and *the younger* respectively.

Definition 2 *A* gradual pattern *M is a set of gradual items, denoted $M = \{A_i^{*_i}, i = 1..k\}$, where $*_i \in \{\geq, \leq\}$ for all $i \in [1, k]$.*

The number of attributes M involves, k, is called its length.

A gradual pattern is semantically interpreted as the conjunction of its gradual items: for instance $M = A^{\geq}B^{\leq}$ is interpreted as *the more A and the less B*.

A gradual pattern $M = \{A_i^{*_i}, i = 1..k\}$ therefore imposes a variation constraint on several attributes simultaneously. It induces an order on objects, denoted $\preceq_M$, defined as $o \preceq_M o'$ iff $\forall i \in [1, k]$, $A_i(o) *_i A_i(o')$.

2.1.2. Quality criterion

The quality of a pattern is measured as the extent to which it holds for a given data set, and is assessed as its *support*.

Two main approaches for the support definition can be distinguished in the case of gradual patterns. The first interpretation takes into account attribute values and for instance relies on a linear regression analysis [11]: the support of a gradual pattern is then measured as the quality of the regression, combined to the slope of the line. This approach requires to define numerical combinations of attribute values and in particular applies to fuzzy data, where the features correspond to membership degrees to various fuzzy modalities.

A second interpretation only considers the order induced by the attribute values, ignoring their values. It can, in turn, be decomposed into two main approaches. The compliant subset approach [12, 2] identifies data subsets $\mathcal{D}^*$ that can be ordered so that all couples from $\mathcal{D}^*$ satisfy the order induced by the pattern. Formally, the support is defined as

$$supp(M) = \frac{1}{|\mathcal{D}|} \max_{\mathcal{D}^* \in \mathcal{L}(M)} |\mathcal{D}^*| \qquad (1)$$

where $\mathcal{L}(M)$ denotes the set of all maximal subsets $\mathcal{D}^* = \{x_1, ... x_m\} \subseteq \mathcal{D}$ for which there exists a permutation π such that $\forall l \in [1, m-1]$, $x_{\pi_l} \preceq_M x_{\pi_{l+1}}$.

The rank correlation approach [1, 3] considers a more local view, focused on data couples instead of data subsets: it counts the number of data couples that satisfy the order induced by the pattern. Formally, its support is defined as

$$supp(M) = \frac{|\{(x, x') \in \mathcal{D}^2 / x \preceq_M x'\}|}{|\mathcal{D}|(|\mathcal{D}| - 1)/2} \qquad (2)$$

Despite their interpretation differences [13], all these support definitions satisfy the classic anti-monotony property, allowing for efficient algorithms to extract frequent gradual patterns.

2.1.3. Specific features

Two specific features of gradual patterns as opposed to classic patterns must be underlined: both in terms of data and attributes, they focus on pairs and not individuals.

Indeed, as can be seen from the order they induce, gradual patterns apply to data pairs, which significantly increases the computational complexity of their processing: mining gradual patterns can be interpreted as mining classic patterns in a rewritten data base, transforming the data to a transactional form that contains a transaction for each data couple [1]. The approach explicitly building this transformed data set requires approximations to keep tractable extraction processes [1]. An alternative approach solves the crucial data representation issue exploiting a representation by means of concordance matrices, that indicate for each data couple whether it satisfies a considered gradual pattern. These matrices allow highly efficient processing through bitmap operations [2, 3].

In addition to this data pair specificity, gradual patterns also focus on attribute pairs: elementary gradual patterns are actually of length 2. Indeed gradual patterns of length 1 do not impose constraints, as any object pair can be trivially ordered to satisfy them. As a consequence, in the explicit data transformation approach [1], items are built for all pairs of gradual items. This approach therefore altogether leads to a transformed transaction base with $n(n-1)/2$ rows, one for each data pair, and $m(m-1)$ columns, to represent all possible 2-gradual items[1] $A^{\geq}B^{\geq}$ and $A^{\geq}B^{\leq}$ for each pair of attributes AB.

[1]The gradual itemsets $A^{\leq}B^{\leq}$ and $A^{\leq}B^{\geq}$ can be considered as equivalent to $A^{\geq}B^{\geq}$ and $A^{\geq}B^{\leq}$ respectively, as they induce the reverse orders and are supported by the same data pairs.

These specific features impose constraints when considering the transposition of algorithms to mine emerging patterns from the classic to the gradual case, as discussed in Section 3.

2.2. Emerging patterns

In the framework of transactional data mining, emerging patterns (EP) are defined as a specific case of classic itemsets: as introduced in [4], they are defined as itemsets characterising a data set *by contrast to* a reference data set. They are helpful to capture discriminant characteristics between categorical data. Thanks to this ability to extract distinctions between classes, EPs have successfully been applied to classification problems [14, 15, 16].

2.2.1. Definition

Formally, denoting $\mathcal{D}_1$ and $\mathcal{D}_2$ the two considered data sets and ρ a numerical threshold, a pattern P *emerges from* $\mathcal{D}_1$ *to* $\mathcal{D}_2$ if its support significantly increases [4]:

$$\frac{supp_{\mathcal{D}_2}(P)}{supp_{\mathcal{D}_1}(P)} \geq \rho \tag{3}$$

This fraction is called the *growth rate*. In the case where the support of P on both $\mathcal{D}_1$ and $\mathcal{D}_2$ is 0, it is considered as 0.

In the case where the fraction is infinite, *i.e.* if P does not occur at all in $\mathcal{D}_1$, P is called a *jumping* emerging pattern.

2.2.2. Specific features

Extracting emerging patterns is a challenging task for several intertwined reasons. First, their definition involves two data sets, which increases the complexity as compared to classic itemsets. Second, the above mentioned growth rate criterion does not satisfy an anti-monotony property, which hinders approaches relying on extensions of classic itemset mining algorithms, such as APRIORI or FP-growth: knowing that a pattern P is not emerging gives no information regarding longer patterns of the form PP', hindering candidate pruning. It must be underlined that this absence of monotony generally comes from the definition of emerging patterns, beyond that of the growth rate criterion.

Emerging patterns can be interpreted as itemsets that are rare in $\mathcal{D}_1$ but frequent in $\mathcal{D}_2$. An approach consisting in filtering out candidates defined as rare itemsets in $\mathcal{D}_1$ is not tractable, both because of the very high number of rare itemsets and the absence of monotony.

Another approach consists in extracting frequent gradual patterns from the two data sets $\mathcal{D}_1$ and $\mathcal{D}_2$ and then keeping the ones present in $\mathcal{D}_2$ results but not in $\mathcal{D}_1$ ones, *i.e.* computing the set difference of these results. Due to the high number of frequent gradual patterns, this approach requires both a compact representation of the frequent itemsets from $\mathcal{D}_1$ and $\mathcal{D}_2$ and a highly efficient computation of their set difference.

2.2.3. Emerging pattern mining algorithms

The first compact representation of patterns used to mine emerging patterns relies on maximal patterns [4, 6] and makes it possible to compute the set difference of the results obtained from the two considered data sets. Because the method we propose is based on this approach, it is described in more details in Section 2.3. This method has been extended in the specific case of jumping emerging patterns, combining it with an efficient tree representation in the spirit of FP-trees [17].

An alternative compact representation is based on closed patterns [18, 19] and makes it possible to provide the growth rate of the extracted emerging patterns, contrary to the previous methods. Exploiting the fact that any itemset and its closure have the same support, it consists in extracting the closed patterns (*e.g.* using the algorithm proposed by [20]), from both $\mathcal{D}_1$ and $\mathcal{D}_2$, computing their growth rate and keeping the ones above the ρ threshold.

Methods proposed in the domain of contrast sets follow a somewhat different approach: expressing the data mining aim as characterising differences between the pattern distributions in the two data sets, they rely on statistical tests [5, 21, 22], to identify discriminant patterns.

It can be noted that the notions of emerging patterns and contrast sets are related to that of characterised gradual patterns [23]: the latter automatically extract data subsets on which the support of a gradual pattern is greater than on the whole data. Two differences must be underlined: first it is specific to the interpretation of gradual patterns in terms of ranking compliant data subsets [2]. Second it considers as input a gradual pattern and extracts data subsets, whereas emerging pattern mining considers the reverse process.

2.3. Border-based EP extraction

This section details the EP mining algorithm based on the compact representation of frequent patterns as maximal patterns, in the form of so-called borders, which we propose to adapt to the gradual case in Section 3.

Border representation A *border* is defined [4, 6] as a couple $\langle \mathcal{L}, \mathcal{R} \rangle$ of antichain[2] collections of sets such that each element of $\mathcal{L}$ is a subset of some element in $\mathcal{R}$ and each element of $\mathcal{R}$ is a superset of some element in $\mathcal{L}$. A border is said to be *left-rooted* if $\mathcal{L} = \{\emptyset\}$.

A border $\langle \mathcal{L}, \mathcal{R} \rangle$ compactly represents the collection of sets $\{Y | \exists X \in \mathcal{L}, \exists Z \in \mathcal{R}$ such that $X \subseteq$

[2] A collection of sets $\mathcal{S}$ is an antichain if $\forall X, Y \in \mathcal{S}$, $X \not\subseteq Y$ and $Y \not\subseteq X$.

$Y \subseteq Z\}$: it is a concise representation that avoids enumerating all these sets explicitly.

Borders are especially useful in the case of convex[3] collections: the key property of a convex collection of sets $\mathcal{S}$ is that it can be uniquely described by a border. More precisely, its border $\langle \mathcal{L}, \mathcal{R} \rangle$ is such that $\mathcal{L}$ (resp. $\mathcal{R}$) is the collection of minimal (resp. maximal) sets in $\mathcal{S}$. (It can be noted that a non-convex collection can be decomposed into a union of convex collections [6]).

Border exploitation The compact representation as borders is useful because several set operations applied to collections can be expressed exclusively in terms of borders, without requiring an explicit enumeration of all sets the collections contain.

In particular, the BORDER-DIFF algorithm [4], further optimised in [6], computes the set difference of two set collections represented by left-rooted borders, only exploiting this representation. Moreover, the resulting collection is also compactly represented as a border (usually not left-rooted).

Application to emerging pattern mining The relevance of the border representation for EP mining comes from the fact that a set of frequent itemsets is a convex collection of sets. It can thus be represented as a border. Moreover, due to anti-monotony, it can be shown that this border is left-rooted [4]: a set of frequent patterns can be represented as the border $\langle \{\emptyset\}, \mathcal{R} \rangle$ where $\mathcal{R}$ is the set of the maximal frequent patterns it contains. As a consequence, the set difference between two sets of frequent patterns can be computed efficiently.

The MBD-LL-BORDER algorithm [4] mines the emerging patterns characterising a data set $\mathcal{D}_1$ by contrast to $\mathcal{D}_2$ in 2 steps: it first extracts their respective frequent patterns, with respect to 2 support thresholds s_1 and s_2, using any itemset mining classical algorithm and it represents them through their borders. It then computes their set difference applying BORDER-DIFF, yielding as a result a border representation of all EP. Some pre-computations are proposed so as to reduce the cost of the whole process [4].

3. Extraction of gradual emerging patterns

This section describes the approach proposed to mine gradual emerging patterns (GEP), based on the MBD-LL-BORDER algorithm [4]: after discussing the properties of gradual patterns that justify this choice, it describes the considered gradual pattern representation and the MBD-LL-BORDER transposition to the case of gradual patterns.

[3] A collection of sets $\mathcal{S}$ is convex if $\forall X, Z \in \mathcal{S}$, for all Y such that $X \subseteq Y \subseteq Z$, it holds that $Y \in \mathcal{S}$.

3.1. Motivation of the selected border-based approach

In order to adapt emerging pattern mining algorithms, as the ones sketched in Section 2.2.3, to the gradual case, it is necessary to examine whether gradual patterns satisfy the respective requirements of the algorithms, taking into account the specific features of gradual patterns, as recalled in Section 2.1.3.

First it can be observed that the transposition of the closed pattern notion to the gradual case is not obvious: the closure of a classic pattern is defined as the intersection of all transactions that contain it. This definition can obviously be adapted to the gradual case if the data set is transformed to the transactional form, as proposed in [1]; however this approach has a high computational cost. It may be possible to study an approach based on ranking compliant data subsets [12, 2]: the latter identify all data that satisfy the considered gradual pattern, in the same way as the closure definition first identifies the transaction subsets that contain the considered classic pattern. However, the definition and computation of their intersection, to build the set of all gradual patterns these data subsets share, is more complex than in the set interpretation of classic patterns.

Along the same lines, the contrast set mining approach, based on the notion of pattern distribution, can obviously be transposed to the transformed transactional view of the numerical data set. However, it raises the same computational concerns.

In both cases, the difficulty comes from the fact that gradual patterns apply to data pairs, leading to a quadratic computational complexity.

On the contrary, the border approach appears to be more tractable, due to the properties of gradual itemsets. Indeed, it holds that

Property 1 *The collection of frequent gradual patterns is convex.*

This property directly derives from the anti-monotony feature of all three definitions of support recalled in Section 2.1. As a consequence, in all three cases, the set of frequent gradual patterns can be represented as borders and the MBD-LL-BORDER algorithm [4] can be applied to extract gradual emerging patterns.

3.2. Border representation of GP

Due to the previous key property, gradual patterns can be both compactly represented and processed through the border associated with their maximal patterns.

3.2.1. Gradual border representation

In order to extract and ease the manipulation of borders associated with maximal gradual patterns,

we propose to use the following explicit transformation: a gradual pattern $M = \{A_i^{*_i}, i = 1..k\}$ is represented as the set of all gradual patterns of length 2 it contains, $\{(A_{i_1}^{*_{i_1}}, A_{i_2}^{*_{i_2}}), (i_1, i_2) \in \{1..k\}^2, i_1 < i_2\}$. For instance, the 4-gradual pattern $A^{\geq}B^{\geq}C^{\leq}D^{\geq}$ is represented as the set of 6 gradual patterns $\{A^{\geq}B^{\geq}$, $A^{\geq}C^{\leq}$, $A^{\geq}D^{\geq}$, $B^{\geq}C^{\leq}$, $B^{\geq}D^{\geq}$, $C^{\leq}D^{\geq}\}$. In the classic itemset case, the 4-pattern $ABCD$ is decomposed into its 4 items A, B, C and D, which again illustrates the increased complexity of gradual patterns.

More generally, a gradual pattern of length k is then represented by the set of $k(k-1)/2$ gradual patterns of length 2, each of them encoded with a single identifier. This step has a non negligible cost, but it is much less expensive than performing the transactional transformation of the whole data set. Moreover, it eases the application of the MBD-LL-BORDER algorithm.

Combining the principles of this section and the previous one, we consider the following border representation of gradual patterns:

Property 2 *A collection $\mathcal{S}$ of frequent gradual patterns is represented as the left-rooted border $\langle\{\emptyset\}, \mathcal{R}\rangle$, where $\mathcal{R}$ is the set of the maximal gradual itemsets in $\mathcal{S}$, represented as the sets of their gradual patterns of length 2.*

3.2.2. Gradual border handling

In the gradual pattern case, a specificity in handling the representation introduced in Property 2 must be underlined, regarding the computation of the set union. Indeed, the union of two gradual patterns is not obtained as the set union of their components of length 2, due to the fact that the latter are not independent: some implicitly implied ones must be explicitly added.

For instance, in the basic case of gradual patterns of length 2, the union of the $A^{\geq}B^{\leq}$ with $B^{\leq}C^{\geq}$ is the gradual pattern of length 3 $A^{\geq}B^{\leq}C^{\geq}$, corresponding to 3 gradual patterns of length 2: it does not only contains the union of the 2 considered patterns, but also the implicit $A^{\geq}C^{\geq}$ one.

3.3. Border-based gradual emerging pattern mining

The straightforward adaptation of the MBD-LL-BORDER algorithm [4] to the case of gradual patterns therefore takes the following simple form:

1. Extract frequent gradual patterns from $\mathcal{D}_1$ and $\mathcal{D}_2$ with respective support thresholds s_1 and s_2, using any support definition (see Section 2.1.2).
2. Build the border representation of all maximal gradual patterns from the results of the first step, as described in the previous subsection.
3. Apply MBD-LL-BORDER to the borders obtained in step 2, with the modified union operation described in the previous subsection.

This method yields as a result a border representation of all emerging gradual patterns.

1	compactness
2	circularity
3	distance circularity
4	radius ratio
5	pr.axis aspect ratio
6	max.length aspect ratio
7	scatter ratio
8	elongatedness
9	pr.axis rectangularity
10	max.length rectangularity
11	scaled variance along major axis
12	scaled variance along minor axis
13	scaled radius of gyration
14	skewness about major axis
15	skewness about minor axis
16	kurtosis about minor axis
17	kurtosis about major axis
18	hollows ratio

Table 1: Shape descriptive attributes for the vehicle data set (for their formal definition, see [7])

4. Experimental results

In order to illustrate the proposed method, this section presents the results obtained when applying it to an extract of the UCI vehicle data set [7]. The latter describes views of two types of vehicles, more precisely vans vs buses. 18 shape features are used, their names are given in Table 1. The aim is to extract feature co-variation constraints that apply to one of the vehicle type as opposed to the other type.

For the first step of the method, to extract frequent gradual itemsets, we apply the GRAANK algorithm [3] that implements the rank correlation interpretation of gradual patterns, using the efficient data representation as binary concordance matrices.

We extract gradual emerging patterns, considering for $\mathcal{D}_1$ the data subset containing the 198 van views and for $\mathcal{D}_2$ the subset describing 217 bus views. Table 2 presents the obtained borders, first ordered by $\mathcal{R}$ length and second by the items occurring in $\mathcal{L}$; in order to simplify notations, $A^{\geq}$ (resp. $A^{\leq}$) is denoted $A+$ (resp. A-).

When characterising buses as opposed to vans with support threshold 0.75 in both data sets, 3 borders are obtained. For each of them $\mathcal{L} = \mathcal{R}$: each of these borders actually contains a single gradual pattern. All are of length 2. In particular, it can be observed that buses satisfy the relation "the more compact, the less elongated", whereas it does not hold (at a support threshold $s_2 = 0.75$) for vans.

Buses as opposed to vans, $s_1 = s_2 = 0.75$	
$\mathcal{L}_1 = \{(1+\ 8-)\}$	$\mathcal{R}_1 = \{(1+\ 8-)\}$
$\mathcal{L}_2 = \{(1+\ 12+)\}$	$\mathcal{R}_2 = \{(1+\ 12+)\}$
$\mathcal{L}_3 = \{(4+\ 12+)\}$	$\mathcal{R}_3 = \{(4+\ 12+)\}$
Vans as opposed to buses, $s_1 = s_2 = 0.75$	
$\mathcal{L}_1 = \{(2+\ 11+)\}$	$\mathcal{R}_1 = \{(2+\ 11+)\}$
$\mathcal{L}_2 = \{(4+\ 5+)\}$	$\mathcal{R}_2 = \{(4+\ 5+)\}$
$\mathcal{L}_3 = \{(10+\ 11+)\}$	$\mathcal{R}_3 = \{(10+\ 11+)\}$
$\mathcal{L}_4 = \{(2+\ 7+),\ (2+\ 12+)\}$	$\mathcal{R}_4 = \{(2+\ 7+\ 12+)\}$
$\mathcal{L}_5 = \{(6+\ 7+),\ (6+\ 12+)\}$	$\mathcal{R}_5 = \{(6+\ 7+\ 12+)\}$
$\mathcal{L}_6 = \{(8-\ 10+),\ (7+\ 10+),\ (10+\ 12+)\}$	
	$\mathcal{R}_6 = \{(7+\ 8-\ 10+\ 12+)\}$
$\mathcal{L}_7 = \{(3+\ 7+\ 8-)\}$	$\mathcal{R}_7 = \{(3+\ 7+\ 8-\ 12+)\}$
Vans as opposed to buses, $s_1 = 0.5$, $s_2 = 0.75$	
$\mathcal{L}_1 = \{(6+\ 7+),\ (6+\ 12+)\}$	$\mathcal{R}_1 = \{(6+\ 7+\ 12+)\}$

Table 2: Border representation of the gradual emerging patterns. To simplify notations, $A^{\geq}$ (resp. $A^{\leq}$) is denoted $A+$ (resp. $A-$). The attribute meaning is given in Table 1.

In a non symmetrical way, vans appear to have more specific characteristics as opposed to buses, as their emerging gradual patterns build a more complex picture: a total number of 7 borders is obtained, for several of which $\mathcal{L} \neq \mathcal{R}$. More precisely, 3 borders with $\mathcal{R}$ of length 2 are obtained as well, but also 2 borders with $\mathcal{R}$ of length 3 and 2 borders with $\mathcal{R}$ of length 4. Interpreting border $\langle \mathcal{L}_7, \mathcal{R}_7 \rangle$ for instance, $\mathcal{R}_7$ indicates that the gradual pattern $(3 + 7 + 8 - 12+)$ has support greater than 0.75 for vans but lower for buses. Moreover, none of its subpatterns of length 3 is emerging for vans, except $(3 + 7 + 8-)$, as they are all excluded from $\mathcal{L}_7$: the specificity of vans comes from the combination of these 3 items. None of the subpatterns of length 2 is specific for vans. The same type of comments apply to the other borders.

In order to increase the discrimination power of the gradual emerging patterns, the support threshold in $\mathcal{D}_1$, s_1, can be decreased, so as to focus on patterns that are more rare in $\mathcal{D}_1$ and thus more emerging in $\mathcal{D}_2$. The bottom part of Table 2 shows that for $s_1 = 0.5$ and $s_2 = 0.75$, a single border is observed, containing the pattern with length 3 (6+7+12+) and all its subpatterns except (7+12+). Of course, this border is also present in the results obtained with $s_1 = s_2 = 0.75$.

When conversely applying the algorithm to buses as opposed to vans with the same support thresholds, no border, and thus no emerging pattern is obtained. This is consistent with the reduced number of borders observed for $s_1 = s_2 = 0.75$: it appears to be more difficult to oppose buses to vans in terms of attribute co-variation than reciprocally. This result may be interpreted in terms of compactness and separability of these two classes, or in terms of typicality [24].

The originality of emerging gradual patterns as opposed to other discriminant characterisations of classes, *e.g.* classification approaches, comes from the specific considered comparison between classes: it relies on the orders induced by the attributes *i.e.* global trends, and not on classical similarity or distance measures between class instances nor on attribute value distributions.

5. Conclusion and future works

This paper presented an approach to extract emerging gradual patterns, making it possible to contrast data sets in terms of attribute co-variations. It relies on the transposition, to the case of gradual patterns, of the compact border representation for set collection and the efficient computation of set difference it brings about.

Future works aim at enriching the experimental study of the proposed method, in terms of scalability, to check its applicability to large data sets. The main concern should lie on interpretability more than computational complexity: a crucial issue is to define a representation of the possibly huge collection of emerging gradual patterns, so as to enable data experts to understand the extracted knowledge, for example by defining dedicated visualisation tools.

References

[1] F. Berzal, J.-C. Cubero, D. Sanchez, M.-A. Vila, and J. M. Serrano. An alternative approach to discover gradual dependencies. *Int. Journal of Uncertainty, Fuzziness and Knowledge-Based Systems (IJUFKS)*, 15(5):559–570, 2007.

[2] L. Di Jorio, A. Laurent, and M. Teisseire. Mining frequent gradual itemsets from large databases. In *Proc. of the Int. Conf. on Intelligent Data Analysis, IDA'09*, 2009.

[3] A. Laurent, M.-J. Lesot, and M. Rifqi. GRAANK: exploiting rank correlations for extracting gradual itemsets. In *Proc. of FQAS*, pages 382–393, 2009.

[4] G. Dong and J. Li. Efficient mining of emerging patterns: Discovering trends and differences. In *Proc. of KDD'99*, 1999.

[5] S. Bay and M. Pazzani. Detecting group differences: mining contrast sets. *Data Mining and Knowledge Discovery*, 5(3):213–246, 2001.

[6] G. Dong and J. Li. Mining border descriptions of emerging patterns from dataset pairs. *Knowledge and Information Systems*, 5:178–202, 2005.

[7] J. Siebert. Vehicle recognition using rule based methods. Technical report, Turing Institute, Glasgow., 1987.

[8] B. Bouchon-Meunier and S. Desprès. Acquisition numérique / symbolique de connaissances graduelles. In *3èmes Journées Nationales du*

PRC Intelligence Artificielle, pages 127–138. Hermès, 1990.

[9] D. Dubois and H. Prade. Gradual inference rules in approximate reasoning. *Information Sciences*, 61(1-2):103–122, 1992.

[10] E. Hüllermeier. Implication-based fuzzy association rules. In *Proc. of PKDD'01*, pages 241–252, 2001.

[11] E. Hüllermeier. Association rules for expressing gradual dependencies. In *Proc. of the 6th European Conf. on Principles of Data Mining and Knowledge Processing, PKDD'02*, pages 200–211. Springer-Verlag, 2002.

[12] L. Di Jorio, A. Laurent, and M. Teisseire. Fast extraction of gradual association rules: a heuristic based method. In *Proc. of the IEEE/ACM Int. Conf. on Soft Computing as a Transdisciplinary Science and Technology, CSTST'08*, 2008.

[13] B. Bouchon-Meunier, A. Laurent, M.-J. Lesot, and M. Rifqi. Strengthening fuzzy gradual rules through "all the more" clauses. In *Proc. of the IEEE Int. Conference on Fuzzy Systems, fuzzIEEE'10*, pages 2940–2946, 2010.

[14] G. Dong, X. Zhang, W. Wong, and J. Li. Caep: classification by aggregating emerging patterns. In *Proc. of the Int. Conf. on Discovery Sciences*, pages 30–42, 1999.

[15] J. Li, G. Dong, and K. Ramamohanarao. Making use of the most expressive jumping emergining patterns for classification. In *Proc. of the Pacific Asia Conf. on Knowledge Discovery and Data Mining*, 2000.

[16] J. Li and L. Wong. Emerging patterns and gene expression data. *Genome Informatics*, 12:3–13, 2001.

[17] J. Bailey, T. Manoukian, and K. Ramamohanarao. Fast algorithms for mining emerging patterns. In *Proc. of PKDD'02*, volume 2431 of *LNAI*, pages 39–50. Springer, 2002.

[18] A. Soulet, B. Crémilleux, and F. Rioult. Condensed representation of emerging patterns. In Honghua Dai, Ramakrishnan Srikant, and Chengqi Zhang, editors, *Proc. of the 8th Asia Conf. on Knowledge Discovery and Data Mining (PAKDD04)*, volume 3056 of *LNCS*, pages 127–132. Springer, 2004.

[19] A. Soulet, B. Crémilleux, and F. Rioult. Condensed representation of eps and patterns quantified by frequency based measures. In B. Goethals and A. Siebes, editors, *Proc. of the KDID04*, volume 3377 of *LCNS*, pages 173–189. Springer, 2005.

[20] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal. Efficient mining of association rules using closed itemset lattices. *Information Systems*, 24:25–46, 1999.

[21] R. Hilderman and T. Peckham. A statistically sound alternative approach to mining contrast sets. In *Proc. of the 4th Australia Data Mining Conf.*, pages 157–172, 2005.

[22] J. Lin and E. Keogh. Group SAX: extending the notion of contrast sets to time series and multimedia data. In *Proc. of the 10th European Conf. on Principles and Practices of Knowledge Discovery in Databases, PKDD'06*, pages 284–296, 2006.

[23] A. Oudni, M.-J. Lesot, and M. Rifqi. Characterisation of gradual itemsets through "especially if" clauses based on mathematical morphology tools. In *Proc. of the 8th Conference of the European Society for Fuzzy Logic and Technology, EUSFLAT'13*, pages 826–833, 2013.

[24] M.-J. Lesot, M. Rifqi, and B. Bouchon-Meunier. Fuzzy prototypes: From a cognitive view to a machine learning principle. In H. Bustince, F. Herrera, and J. Montero, editors, *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, pages 431–452. Springer, 2007.