



HAL
open science

On the stimulation of patterns Definitions, calculation method and first usages

Ryan Bissell-Siders, Bertrand Cuissart, Bruno Crémilleux

► **To cite this version:**

Ryan Bissell-Siders, Bertrand Cuissart, Bruno Crémilleux. On the stimulation of patterns Definitions, calculation method and first usages. 18th International Conference on Conceptual Structure (ICCS'10), Conceptual Structures: from information to intelligence, Jul 2010, Kuching, Malaysia. pp.56-69. hal-01016911

HAL Id: hal-01016911

<https://hal.science/hal-01016911v1>

Submitted on 2 Dec 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the Stimulation of Patterns

Definitions, Calculation Method and First Usages

Ryan Bissell-Siders, Bertrand Cuissart, and Bruno Crémilleux

Groupe de Recherche en Électronique, Informatique et Imagerie de Caen,
CNRS-UMR6072, Université de Caen, France

ryan.bissell-siders@info.unicaen.fr, cuissart@info.unicaen.fr,
cremilleux@info.unicaen.fr
<http://www.greyc.unicaen.fr/>

Abstract. We define a class of patterns generalizing the *jumping emerging patterns* which have been used successfully for classification problems but which are often absent in complex or sparse databases and which are often very specific. In supervised learning, the objects in a database are classified a priori into one class called *positive* – a *target* class – and the remaining classes, called *negative*. Each *pattern*, or set of attributes, has support in the positive class and in the negative class, and the ratio of these is the *emergence* of that pattern; the *stimulating patterns* are those patterns a , such that for many closed patterns b , adding the attributes of a to b reduces the support in the negative class much more than in the positive class. We present methods for comparing and attributing stimulation of closed patterns. We discuss the complexity of enumerating stimulating patterns. We discuss in particular the discovery of highly stimulating patterns and the discovery of patterns which capture contrasts. We extract these two types of stimulating patterns from UCI machine learning databases.

1 Introduction

We introduce *stimulation*, a new measure of interest in the classification of objects by their attributes. We suppose here that a dataset is a finite list of descriptions of objects, where the description of each object corresponds to a list of its binary attributes. A *pattern* denominates a set of binary attributes. The *extent* of a pattern is the set of objects whose descriptions contain each attribute in the pattern. The objects are classified and we aim to predict the classification of a new object from its description. The *support* of a pattern in a class is the cardinality of the extent of the pattern, restricted to the objects in that class. The classification of a pattern is a function of its supports in the classes of the classification. The *stimulation* of a pattern captures its influence on the classification of other patterns.

When we consider whether a pattern favors a certain class, we refer to that class as the *positive* class, and to the union of the remaining classes as *negative*. If a pattern stimulates the classification of other patterns to be more positive,

by removing more (or relatively more) negative objects than positive objects from the extent of other patterns, then this is a strong correlation between the pattern and the positive class. Such a pattern stimulates a positive classification not only alone, but when mixed with any other pattern, and we observe this by considering the stimulation of a pattern on all other patterns. In addition to patterns with a constant influence, patterns that have a variable influence on the classification of other patterns are interesting too: they are useful for adding a new dimension to an existing model. Our work found inspiration in [14], which mines a dataset to find an attribute with very variable influence on a set of mutually exclusive patterns $p_0 \dots p_n$, with the intention of explaining the difference between these patterns.

The relationship between a pattern and the classification is an interesting quantitative problem. Information gain measures the amount of information in the class which is explained by a pattern, [4]. If the class and an attribute are both continuous, the Gini index measures correlation between them, [8] Chapter 9]. When we focus on a positive class, the simple odds of a positive classification, given a pattern, are called the *emergence* of that pattern [1]. Patterns with high emergence are called *emerging patterns* (or EPs), they have found wide application since their introduction in the data-mining community [2]. Mining EPs produces a flood of rules, among which may be found some rules which are valuable for constructing a classification model or for explaining the classification in a human-readable way. EPs yielded successful characterizations of biochemical properties and medical data in [11]. EPs are used in top-performing classifiers [3][10][18] and in commercial algorithms to find rules to explain separations between groups [21]. Creating a dataset of chemical graphs and subgraphs is in itself an interesting problem; once the dataset is constructed, extracting emerging patterns produces rules of interest to chemists [16]. We choose to use emergence to measure the relationship between a pattern and the classification because the notion is simple and powerful, and it maintains continuity with [14].

The influence of one pattern on another has been considered theoretically in statistics. Conditional probability is able to analyze the correlation between a pattern with the classification. *Naïve bayesian classification* then makes the assumption that the influence of each attribute on the pattern is independent of which patterns have gone before. When we restrict our attention to the extent of a pattern q , the correlation of the classification with a pattern p containing q is called the *odds ratio* between p and q . It expresses the influence of the larger pattern p on the classification of the subpattern q . Mining the influence of a single attribute on a set of patterns, or visa versa, has been carried out efficiently with *contrast sets* [14]. Contrast sets are similar to EPs and are mined so as to explain the difference between two classifications, They can detect a threshold or a fault. They are useful for refining a model of the classification of other patterns.

Mining the influence of all patterns on all attributes is inefficient in general; techniques to reduce the EPs to a readable and meaningful set of patterns make it efficient to study the influence of each EP on the rest [18]. In this text, we organize pairs of patterns into groups which have been stimulated in the same

way, so that for each pair in a group, it becomes clear which parts of the patterns are responsible for their classification. In one experiment, we extract groups with high and uniform stimulation. These patterns can explain why an object has positive classification, for if EPs cover some of the attributes of an object, then only the remaining attributes would oppose a positive classification. As predicted in [12], these groups conservatively extend the EPs. In another experiment, we extract groups with highly varying stimulation. These patterns are useful for extending a classification model.

This paper is organized as follows. Section 2 outlines background concepts. We define stimulation in section 3. We present an algorithm to extract the stimulation measure and describe experiments in which this measure is of interest in section 4.

2 Preliminaries

2.1 Notions of Formal Concept Analysis [6, 5]

We use standard notions from Formal Concept Analysis (FCA): a formal *context* denominates the triple (M, G, I) where the binary *attributes* M and *objects* G are related by the dataset $I \subseteq G \times M$. In FCA *concepts* are the inclusion-maximal sets $a \subseteq I$ of the form $a = A \times B, A \subseteq G, B \subseteq M$. A set of attributes is named a *pattern*. The *extent* of an attribute $m \in M$ is $\{g \in G : (g, m) \in I\}$. The *extent* of a pattern p , denoted $\text{ext}(p)$, is the intersection of the extents of its attributes. Likewise, the *intent* of an object g is the set of attributes m such that $(g, m) \in I$, and the *intent* of a set $A \subseteq G$ of objects, denoted $\text{int}(A)$ is the intersection of the intents of its objects. For any two patterns a, b we write $a < b$ and say a is *more specific than* b just in case $\text{ext}(a) \subset \text{ext}(b)$.

The function taking a pattern a to the intent of its extent, denoted \bar{a} , is a *closure* function on patterns: for any two patterns a, b , $\bar{a} \subseteq \bar{b}$ holds whenever $a \subseteq b$ and $\overline{\bar{a}} = a$. The set \bar{a} is called *closed*. An elegant alternate notation [6] is to write x' for both $\text{int}(x)$ and $\text{ext}(x)$ and x'' for \bar{x} . We denote by 0 the concept that satisfies $\text{int}(0) = M$; similarly, we denote by 1 the concept that satisfies $\text{ext}(1) = G$. For any two patterns a and b , $a \vee b$ is a least upper bound – the least (most specific) closed pattern c such that $a < c$ and $b < c$. Likewise, $a \wedge b$ is a greatest lower bound – the greatest (least specific) closed pattern c below a and b . Because $\text{ext}(a \wedge b) = \text{ext}(a) \cap \text{ext}(b)$ and $\text{int}(a \vee b) = \text{int}(a) \cap \text{int}(b)$, the upper and lower bounds are unique, so the set of closed patterns inherits the structure of a lattice from the Boolean algebra of subsets of G (or of M) with \vee and \wedge defined as above.

The lattice structure on the set of closed patterns can be recovered from a single function, the *upper covers*. Defined on any lattice \mathcal{L} , the *upper covers* is the function from $a \in \mathcal{L}$ to the set of its immediate successors (those $b \in \mathcal{L}$ such that $a < b$ and there is no c such that $a < c < b$). We represent a lattice \mathcal{L} by storing only its domain, also denoted \mathcal{L} , the extent and intent and upper covers functions.

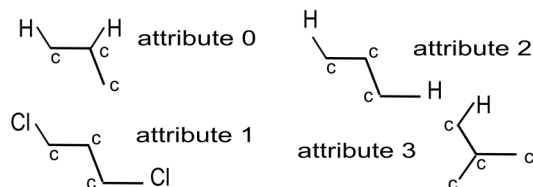


Fig. 1. The attributes used to describe the molecules in table 1 are the presence or absence of these subgraphs, each of which has 5 elements

Table 1. Polychlorinated biphenyl molecules. The first column is the molecule’s descriptive name. The next four columns indicate the presence of attributes 0, 1, 2, 3, i.e., the existence of an isomorphic copy of the corresponding subgraph from figure 1. The final column indicates which two molecules are most toxic.

The molecule’s name	attributes:				toxicity
	att. 0	att. 1	att. 2	att. 3	
3,3',4,4'-TetraCB	1	0	0	0	most toxic most toxic
3,4,4',5-TetraCB	1	1	1	0	
3,3',4,4',5-PentaCB	1	1	0	0	
3,3',4,4',5,5'-HexaCB	0	1	0	0	
2,3,3',4,4'-PentaCB	1	1	0	1	
2,3,4,4',5-PentaCB	1	1	1	1	
2,3',4,4',5-PentaCB	1	1	0	1	
2',3,4,4',5-PentaCB	1	1	1	1	
2,3,3',4,4',5-HexaCB	1	1	0	1	
2,3,3',4,4',5'-HexaCB	0	1	0	1	
2,3',4,4',5,5'-HexaCB	0	1	0	1	
2,3,3',4,4',5,5'-HeptaCB	0	1	0	1	

Illustration. Figure 1 shows four chemical graphs which we use as attributes. If a molecule contains an isomorphic copy of one of these as an induced subgraph, then we say that it contains that subgraph as an attribute. The figure lists all graphs with 5 atoms which are present in at least two polychlorinated biphenyls and not present in at least two polychlorinated biphenyl molecules (PCBs). Table 1 displays (PCBs) and their subgraph attributes. These twelve PCBs are of special concern and are regulated in international law for their toxicity. Two of them are orders of magnitude more toxic than the others. Figure 2 displays the lattice of concepts for this dataset, showing the intent of the concept as a set in each circle. The edges in the diagram represent the upper covers.

2.2 Definition of Emergence 2

Given a classification of the objects G into *positive* and *negative* classes G_0 and G_1 , the *emergence* of a pattern a compares the *frequencies* of a , where the *frequency* of a pattern a within a class C indicates the portion of the objects of C that are in relation with a . If $C \neq \emptyset$, define $\text{frequency}(a, C) = \frac{|\{g \in \text{ext}(a) : g \in C\}|}{|\{g \in C\}|}$. The emergence of a pattern is the defined as:

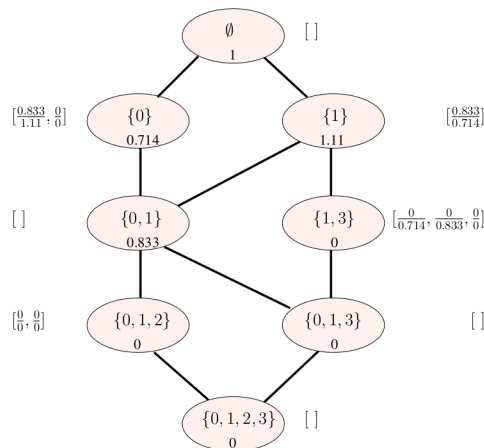


Fig. 2. The lattice of concepts of table 1 showing each concept's intent (as a set), emergence (definition 1) as a numerical value under the intent, and the stimulation set (definition 4) as a multiset $[\frac{a_0}{b_0}, \dots]$ to the side

Definition 1 (Emergence 2). Given positive class G_0 and negative class G_1 , for any pattern $a \subseteq M$, the emergence of a is

- $\text{emergence}(a) = \frac{\text{frequency}(a, G_0)}{\text{frequency}(a, G_1)}$ if $\text{frequency}(a, G_1) \neq 0$
- $\text{emergence}(a) = \infty$ if $\text{frequency}(a, G_1) = 0$ and $\text{frequency}(a, G_0) \neq 0$
- $\text{emergence}(a)$ is not defined if $\text{frequency}(a, G_1) = 0$ and $\text{frequency}(a, G_0) = 0$

If the role of the class G_0 and its complement G_1 were reversed in the above definition, the emergence of any pattern would become its inverse. We could specify the emergence defined above to be the emergence *into* G_0 , relative to G_1 , denoted $\text{emergence}_{(G_0, G_1)}(a)$. But for the sake of simplicity, we denote $\text{emergence}_{(G_0, G_1)}(a)$ as $\text{emergence}(a)$, supposing that we have already directed our attention towards G_0 as the *positive* class.

If the emergence of a is ∞ , then a is called a *jumping emerging pattern* (JEP); if the emergence of a is 0, then we call a an *anti-JEP*. We will be careful to never refer to the emergence of a pattern with empty extent, so that our definition of emergence differs in no way from that of ([2], p.45).

Illustration. Figure 2 displays the closed patterns \emptyset , $\{0\}$, $\{1\}$, $\{0,1\}$, $\{1,3\}$, $\{0,1,2\}$, $\{0,1,3\}$, $\{0,1,2,3\}$. The frequency of each of these patterns among the most-toxic molecules can be observed from table 1 to be, respectively, $\frac{2}{2}$, $\frac{1}{2}$, $\frac{2}{2}$, $\frac{1}{2}$, 0, 0, 0, 0. The frequency of each of these patterns among the rest is: $\frac{10}{10}$, $\frac{7}{10}$, $\frac{9}{10}$, $\frac{6}{10}$, $\frac{8}{10}$, $\frac{3}{10}$, $\frac{5}{10}$, $\frac{2}{10}$. The emergence of each pattern is the ratio of its frequency among the most-toxic molecules to its frequency among the rest. Figure 2 shows the emergence of each pattern in the circle, under pattern (the intent of the concept).

3 Stimulation of a Pattern

3.1 Definition of Stimulation

We define the stimulation of a pattern a on b to be the ratio of the emergence of $a \wedge b$ to the emergence of b .

Definition 2 (Stimulation). *Let (a, b) be an ordered pair of patterns. If $a \wedge b \neq 0$, the stimulation of a on b , denoted $\text{stimulation}(a, b)$, is defined to be:*

- $\text{stimulation}(a, b) = \frac{\text{emergence}(a \wedge b)}{\text{emergence}(b)}$ if $0 < \text{emergence}(a \wedge b) < \infty$,
- $\text{stimulation}(a, b) = \infty$ if $\text{emergence}(a \wedge b) = \infty$ and $\text{emergence}(b) < \infty$,
- $\text{stimulation}(a, b) = 0$ if $\text{emergence}(a \wedge b) = 0$ and $0 < \text{emergence}(b)$,
- $\text{stimulation}(a, b) = 1$ if $\text{emergence}(a \wedge b) = \text{emergence}(b) = 0$ or $\text{emergence}(a \wedge b) = \text{emergence}(b) = \infty$.

Since $\text{ext}(a \wedge b) \subseteq \text{ext}(b)$, the condition $a \wedge b \neq 0$ implies $b \neq 0$. Consequently, when $a \wedge b \neq 0$, both $\text{emergence}(a \wedge b)$ and $\text{emergence}(b)$ are defined. In particular, if b is a JEP (resp. an anti-JEP) then $(a \wedge b)$ is a JEP (resp. an anti-JEP). If $\text{stimulation}(a, b)$ is a finite fraction and we switch G_0 and G_1 , then $\text{stimulation}(a, b)$ becomes its own inverse.

Illustration. In figure 2 the pattern $\{0, 1\}$ appears, so it is a closed pattern and it is the intent of a concept. It has frequency $\frac{1}{2}$ among the most toxic molecules and frequency $\frac{6}{10}$ among the rest. Its emergence is, then, $\frac{1}{2} / \frac{6}{10}$. Its upper cover $\{0\}$ has frequency $\frac{1}{2}$ among the most toxic molecules and frequency $\frac{7}{10}$ among the rest. Its emergence is $\frac{1}{2} / \frac{7}{10}$. The fact that $\text{ext}(\{0\}) \setminus \text{ext}(\{0, 1\})$ contains a single molecule – the molecule which does not have attribute 1 – is reflected in $\text{stimulation}(\{1\}, \{0\}) = \frac{0.833}{0.714} = \frac{7}{6}$.

Factorization of stimulation. Stimulation allows us to factor “the odds of a and b ” into “the odds of a ” and “the stimulation of b on a ”; likewise, we can factor the “stimulation of a_0 and a_1 on b ” into “the stimulation of a_0 on b ” and “the stimulation of a_1 on $(a_0 \wedge b)$.” Stimulation thus transforms the interaction of patterns into multiplication:

Proposition 3. $\text{stimulation}(a, b) \times \text{stimulation}(c, a \wedge b) = \text{stimulation}(a \wedge c, b)$

Proof. The emergence of $a \wedge b$ is the numerator in the first multiplicand and the denominator in second multiplicand. \square

As a corollary: if a stimulates b to a degree > 1 and c stimulates $a \wedge b$ to a degree > 1 , then $a \wedge c$ stimulates b to a degree > 1 .

3.2 Stimulation Set

Reducing the domain of stimulation. Since the emergence of a pattern is defined from its extent, the emergence of a set of attributes is the same as the

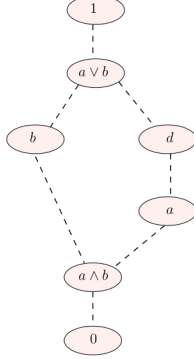


Fig. 3. a is not responsible for the change of emergence between b and $a \wedge b$

emergence of its closure. Let a and b be patterns. As $\text{ext}(a) = \text{ext}(\bar{a})$ and $\text{ext}(b) = \text{ext}(\bar{b})$, we have $\text{ext}(a \wedge b) = \text{ext}(\bar{a} \wedge \bar{b})$. Furthermore, by definition, we have $\text{ext}(a \wedge b) = \text{ext}(\overline{a \wedge b})$. It results $\bar{a} \wedge \bar{b} = \overline{a \wedge b}$. In our case, this equality implies that $\text{stimulation}(a, b)$ is defined if $\text{stimulation}(\bar{a}, \bar{b})$ is defined. Moreover, if $\text{stimulation}(a, b)$ is defined, we have $\text{stimulation}(a, b) = \text{stimulation}(\bar{a}, \bar{b})$. Consequently, the domain of the second argument of stimulation is naturally narrowed to closed patterns, and to the concepts which have these closed patterns as intents. The domain of the first argument could be left as patterns, but we choose to restrict it as well to closed patterns, or their concepts.

The responsibility of a stimulation. We classify pairs of concepts $\{(b, c) : b > c\}$ so that (b, c) and (f, g) are in the same group if there exists any concept e such that $b \wedge e = c$ and $f \wedge e = g$. In this case, we say that e might be *responsible* for the difference between b and c . So, while stimulation captures a change in emergence, we want *stimulation sets* to classify $\text{stimulation}(a, b)$ by the proper argument a which is really responsible for the stimulation from b to $c = b \wedge a$. The change in emergence from pattern b to pattern $b \wedge a$ can be attributed to different patterns, and not only to a itself. Suppose there exists $d \in \mathcal{L}$ such that: $d < a$ and $\text{ext}(d \wedge b) = \text{ext}(a \wedge b)$ (see Figure 3). Consequently $d \wedge b \neq 0$ is equivalent to $a \wedge b \neq 0$, $\text{stimulation}(d, b)$ is defined if $\text{stimulation}(a, b)$ is defined. Moreover, when $\text{stimulation}(a, b)$ is defined we have $\text{stimulation}(d, b) = \text{stimulation}(a, b)$.

We decide to attribute $\text{stimulation}(a, b)$ to a if a is the smallest pattern d such that $\text{ext}(d \wedge b) = \text{ext}(a \wedge b)$. If there exists some $d < a$ such that $\text{ext}(d \wedge b) = \text{ext}(a \wedge b)$, then we consider that the stimulation should be attributed to d rather than to a . We write $[f(a) : a \in U]$ for a multiset of values.

Definition 4 (Stimulation set). Let $a \in \mathcal{L}$. The stimulation set of a , denoted $SS(a)$, is the multi-set which is a subset of $[\text{stimulation}(a, b) : a \wedge b \neq 0]$ for which $\text{stimulation}(a, b)$ belongs to $SS(a)$ just in case :

- i) $a \not\leq b$ and
- ii) there is no $d \in \mathcal{L}$ such that $d \geq a$ and $d \wedge b = a \wedge b$.

If $b \leq a$, we suppress $\text{stimulation}(a, b)$ when printing $SS(a)$, as in figure 2

Illustration. The stimulation sets shown in figure 2 are in brackets to the left and right of the concepts in the lattice. Let us see why $SS(\{0, 1\})$ is empty. By the first condition of definition 4, each pattern stimulates only concepts on the other side of the diagram. Further, $\{0, 1\}$ can only stimulate concepts which are not related to it by $<$. This leaves only $\{1, 3\}$. But taking $a = \{0, 1\}$ and $b = \{1, 3\}$, let $d = \{0\}$. $\text{stimulation}(\{0, 1\}, \{1, 3\}) = \text{stimulation}(\{0, 1\}, \{1, 3\})$, so by the second condition in definition 4 $\{0, 1\}$ is not responsible for this stimulation value. The edges and paths in figure 2 which can be attributed to changes off the path itself are partitioned into $SS(a)$. The following lemma shows that it is always the case that $\{(a, b) : a < b\}$ is almost-partitioned into the stimulation sets. An edge will be present in two stimulation sets $SS(a), SS(b)$ if both a and b are minimal explanations responsible for that edge.

Lemma 5. *For each $b, c \in \mathcal{L}$ such that $b < c$, there is at least one $a \in \mathcal{L}$ such that $\frac{\text{emergence}(c)}{\text{emergence}(b)} \in SS(a)$.*

Proof: $b \wedge c = c$, so $\{a \in \mathcal{L} : a \wedge b = c\}$ is not empty. This set has at least one minimal element a' . For each such minimal a' , $SS(a')$ contains the desired fraction. \square

Now we use the stimulation values to characterize which concepts consistently stimulate other concepts. Let $MS(a)$ be the minimal stimulation of a , i.e., the minimal value in $SS(a)$. We can bound $MS(a \wedge b)$ by $MS(a)$ and $MS(b)$:

Proposition 6. *There is an injection from $SS(a \wedge b)$ into $SS(a) \times SS(b)$ such that when $\text{stimulation}(a \wedge b, c) \mapsto (r_0, r_1)$, $\text{stimulation}(a \wedge b, c) > r_0 \times r_1$.*

Proof. If $p_0 \subseteq p_1 \subseteq p_2$ are closed patterns, and $\text{ext}(a) \cap \text{ext}(p_0) = \text{ext}(p_1)$ and $\text{ext}(b) \cap \text{ext}(p_1) = \text{ext}(p_2)$, then $\text{ext}(b) \cap \text{ext}(a) \cap \text{ext}(p_0) = \text{ext}(p_2)$. If a is not a pattern of minimal intent such that $\text{ext}(a) \cap \text{ext}(p_0) = \text{ext}(p_1)$, say, $d \subseteq a$ and $\text{ext}(d) \cap \text{ext}(p_0) = \text{ext}(p_1)$, then $(\text{ext}(b) \cap \text{ext}(d)) \cap \text{ext}(p_0) = \text{ext}(p_2)$, so that $\text{ext}(b) \cap \text{ext}(a)$ is not the minimal c such that $c \cap \text{ext}(p_0) = \text{ext}(p_2)$. Likewise, if b is not the pattern of minimal intent such that $\text{ext}(b) \cap \text{ext}(p_1) = \text{ext}(p_2)$, then $\text{ext}(b) \cap \text{ext}(a)$ is not the minimal c such that $c \cap \text{ext}(p_0) = \text{ext}(p_2)$. \square

Application of the lemma and proposition. Definition 4 prunes some emergence ratios from the notion of stimulation. Lemma 5 shows that the pruning is conservative, preserving the emergence ratios for any interval in the lattice. Proposition 6 shows that if the minimal value in $SS(a)$ is $MS(a)$ and the minimal value in $SS(b)$ is $MS(b)$, then the minimal value in $SS(a \cup b)$ is $MS(a \cup b) \geq MS(a) \times MS(b)$. Thus the set of patterns with uniform, high stimulation is join-closed. We can then consider only the boundary of this set, when searching for highly stimulating patterns.

In this section, we have defined a new measure of interaction for any ordered pair of patterns which captures how emergence changes under additional information or in a restricted situation. We have introduced the notion of the set of stimulation values for which a pattern a is responsible. In the next section,

Algorithm 1. Enumerate $SS(a)$ of all closed sets of objects a

Input: A Galois lattice \mathcal{L} with extent, upper covers; a classification of the objects $G = G_0 \cup G_1$ into positive class G_1 and negative class G_0

Output: $\{SS(a) : a \in \mathcal{L}\}$

Let T order \mathcal{L} from the concept with the largest extent to the concept with the least extent, so that $a <_T b$ holds just in case the support of a in $G_0 \cup G_1$ is \geq the support of b in $G_0 \cup G_1$.

Let L order \mathcal{L} (arbitrarily).

foreach $b \in L$ **do**

 compute the support of $\text{ext}(b)$ in both G_0 and G_1 .

foreach $a \in T$ **do**

if $a = b$ **then**

 write $a \leq b$.

 compute the support of $\text{ext}(b) \cap \text{ext}(a)$ in both G_0 and G_1 .

if for each upper cover a' of a :

$b \geq a'$ fails (else, save $b \geq a$) and

$\text{ext}(b) \cap \text{ext}(a') = \text{ext}(b) \cap \text{ext}(a)$ fails **then**

 add $\text{stimulation}(a, b)$, as a 4-tuple of supports, to $SS(a)$.

we address the problem of calculating the stimulation sets, and we describe experiments extracting stimulation sets which mostly contain large values and stimulation sets which contain as widely varying a set of values as possible.

4 Computing Stimulation

4.1 Calculation of the Stimulation Sets

A naïve search through the lattice finds all pairs $b \geq c$, and assigns the ratio of their emergence to some other pattern a , thus computing a matrix of stimulations for $a, b \in \mathcal{L}$. See algorithm [1](#) for the pseudocode. For some uses of this stimulation matrix, it may be possible to achieve that use without the naïve time-complexity factor $|\mathcal{L}|^2$.

Sound and complete. This algorithm computes $SS(a)$ as in definition [4](#). The first condition, that $a \not\leq b$, is enforced by storing $f(a) = \{b : a \leq b\}$, which can be computed from the set of $f(a')$ for which a' is an upper cover of a , since $a \leq b$ holds just in case $a = b$ or for some upper cover a' of a we have $a < a'$ and $a' \leq b$. The second condition is enforced by comparing the support of $a \wedge b$ with the support of $a' \wedge b$ for each upper cover a' of a . If for no upper cover $a' > a$ do we have $a' \wedge b = a \wedge b$, then for no $d > a$ do we have $d \wedge b = a \wedge b$, because the function $a \mapsto \text{ext}(a)$ is monotonic, taking \leq to \subseteq , so the function $a \mapsto \text{ext}(a) \cap \text{ext}(b)$ is monotonic, too. Thus, this algorithm searches all triples $(a, b, a \wedge b)$ and assigns the value $\text{stimulation}(a, b)$ to $SS(a)$ just in case the two conditions in definition [4](#) hold.

Table 2. For the PCB dataset in table 1 these concepts were consistently stimulating (or anti-stimulating) and had no upper cover with the same stimulation

Intent	stimulation at 10 th or 90 th percentile	stimulation set
{1}	stimulates ≥ 1.1	$SS\{1\} = [7/6]$
{1, 3}	is anti-JEP	$SS(\{1, 3\}) = [0, 0, \frac{0}{0}]$
{0}	stimulates $\leq 1/1.2$	$SS\{0\} = [3/4, \frac{0}{0}]$
{0, 1, 2}	is anti-JEP	$SS(\{0, 1, 2\}) = [0, 0, \frac{0}{1.11}]$

Complexity. The runtime of the algorithm as given is dominated by its obviously nested loops, and is bounded by $|\mathcal{L}|^2 \times \sup_a |\text{upper cover}(a)| \times \sup_{a,b} |\text{ext}(a) \cap \text{ext}(b)|$, where $\sup_a |f(a)| = \sup\{f(a) : a \in \mathcal{L}\}$ denotes the maximum cardinality of $f(a)$ as a varies over \mathcal{L} .

Lower complexity. Let L be a traversal of \mathcal{L} , ascending from the concept with minimal extent and stepping always from a closed pattern b_0 to a new pattern b such that $\text{ext}(b) \setminus \text{ext}(b_0)$ is minimal; then we can check whether $\text{ext}(b) \cap \text{ext}(a') \neq \text{ext}(b) \cap \text{ext}(a)$ by examining only elements of $\text{ext}(b) \setminus \text{ext}(b_0)$, which reduces the last factor in the runtime to $\sup\{|\text{ext}(b) \setminus \text{ext}(b_0) \cap \text{ext}(a)| : b \text{ is an upper cover of } b_0 \text{ and } a, b, b_0 \in \mathcal{L}\}$.

4.2 Experiment

We implemented the algorithm introduced in the previous subsection. From the archive of datasets stored at UCI (www.ics.uci.edu/~mlearn/), we extracted concept lattices using the Galicia suite of programs [20]; we subsequently extracted $\{SS(a) : a \in \mathcal{L}\}$. We found that whenever Galicia could extract a concept lattice without overflowing memory, we were able to extract the stimulation: during our experience, the factor of $|\mathcal{L}|^2$ in the runtime is not an order of magnitude more prohibitive to computation than the decision to operate with a concept lattice.

Our goal was to discover 1. highly stimulating sets that are not JEPs, and 2. to discover non-homogeneity. Drawbacks to a purely JEP-based classification are discussed in [13]. For any dataset in which every rule $\text{ext}(p) \subseteq G_i$ has exceptions, there are no JEPs. For many datasets, there are a flood of JEPs. In some fuzzy datasets such as census and satellite images (according to a study of their second-order properties in [14]), emerging patterns fail to extract certain important properties, and logically more flexible rules are desirable. However, we chose to extract stimulating sets first in contexts where JEPs classify *well*, so as to evaluate the “border” of new information which they add to the emerging patterns.

If we discover the highly stimulating patterns in the PCB example, we get the list in table 2. The value of stimulation(a, b) is ∞/∞ when b is a JEP; the stimulation is $0/0$ when b is an anti-JEP. These values are reported as 1 in the table above, but do not count against the designation of a pattern as highly stimulating. Thus pattern $\overline{\{2\}}$ has $SS(\overline{\{2\}}) = [0, 0, 1, 1]$ and is called an anti-JEP even though 1 occurs with frequency $> 10\%$ because these values are not

Table 3. For the dataset shuttle-landing-control, these are the patterns with greatest extent which stimulated almost all other concepts, and which had no upper cover with (roughly) the same average stimulation

Intent	stimulation at 10 th or 90 th percentile	stimulation set
≥ 3	{VISIBILITY:yes}	$[1^7 1^2 2^3 5^3 3^{23} 9.3^4 12^{20} 20^6 \dots \infty^{502}]$
$\leq 1/\infty$	{VISIBILITY:no}	$[0^6 35^1 699]$
$\geq \infty$	{STABILITY:xstab,VISIBILITY:yes}	$[\infty^{222}]$
≥ 2	{SIGN:pp,VISIBILITY:yes}	$[0^5 1^9 9.3^2 10^3 12^6 15^3 22^1 \dots \infty^{165}]$
$\geq \infty$	{MAGNITUDE:OutOfRange,VISIBILITY:yes}	$[\infty^{134}]$
$\geq \infty$	{ERROR:XL,VISIBILITY:yes}	$[\infty^{135}]$
$\geq \infty$	{VISIBILITY:yes,ERROR:LX}	$[\infty^{135}]$
≥ 1.5	{SIGN:pp,VISIBILITY:yes,WIND:tail}	$[0^3 1^3 37.3^1 48^3 50^2 54^1 63^1 \infty^{58}]$
≥ 1.5	{VISIBILITY:yes,ERROR:MM}	$[0^9 1^3 55^4 60^5 88^1 99^2 240^2 \infty^{99}]$
≥ 1.5	{MAGNITUDE:Strong,VISIBILITY:yes}	$[0^{10} 1^1 39^1 66^6 264^5 378^1 456^1 \dots \infty^{103}]$
≥ 1.1	{MAGNITUDE:Low,VISIBILITY:yes}	$[0^{13} 1^1 66^6 \dots \infty^{98}]$
≥ 1.1	{MAGNITUDE:Medium,VISIBILITY:yes}	$[0^{13} 1^1 66^6 70^1 94.5^1 \dots \infty^{98}]$
≥ 1.2	{VISIBILITY:yes,WIND:tail,ERROR:MM}	$[0^3 198^2 210^1 240^2 247^1 358.3^1 \infty^{32}]$
$\geq \infty$	{VISIBILITY:yes,SIGN:nn,ERROR:MM}	$[1^3 \infty^{42}]$
≥ 1.5	{MAGNITUDE:Low,VISIBILITY:yes,WIND:tail}	$[0^4 189^1 264^2 273^1 1056^1 1075^1 \infty^{34}]$
≥ 1.1	{MAGNITUDE:Strong,SIGN:pp,VISIBILITY:yes}	$[0^5 189^1 264^2 304^1 1056^1 1148^1 \infty^{33}]$
≥ 1.5	{MAGNITUDE:Medium,VISIBILITY:yes,WIND:tail}	$[0^4 189^1 264^2 273^1 1056^1 1075^1 \infty^{34}]$

counted against 2 in determining that it is infinitely stimulating. Ignoring the third line, we find the obvious classification that a PCB congener which is in the list of 12 toxins of concern for international control is highly toxic if: it contains attribute 1 but not attribute 2 or 3.

To test the algorithm in a “contrary” domain, we chose shuttle-landing-control from the UCI Machine Learning datasets. This dataset is presented as a set of 15 JEPS. The dataset is a rule base, where each rule has one of two forms: $\text{ext}(a) \subseteq G_0$, which indicates that if pattern a obtains, then G_0 : the shuttle should be landed by a human; or $\text{ext}(a) \subseteq G_1$, meaning that if a obtains, then the shuttle should be landed by autopilot. We expected to recover these 15 rules as highly stimulating rules as stimulating patterns. In addition, we found 7 other rules and 16 highly stimulating patterns. Table 3 displays the first 17 of these patterns, those with the smallest intent and largest extent, ranked by increasing intent and decreasing extent. The resulting 38 patterns are of interest in guiding the classification problem, as they are not numerous, and the new patterns behave as “fuzzy” JEPs.

Non-homogeneity. Even in the small dataset of shuttle-landing-control, with four multivariate attributes, one can find pairs of attributes (a, b) such that the matrix $\{\text{stimulation}(a_i, b_j) : i < n_a, j < n_b\}$ exhibits contrast behavior (it is “twisted”). In table 4 the values in each column are comparable, whereas the values in the last column are very different. Whether STABILITY:stab or STABILITY:xstab is more highly correlated with the positive classification also varies across columns. Thus, in order to explain the effect of the attribute STABILITY on the classification, one must discuss the attribute ERROR.

The runtime grows exponentially with even a trivial increase in the number of objects and attributes, a typical feature of operations on lattices. Runtimes

Table 4. Two attributes from the dataset shuttle-landing-control, and their matrix of emergences. If a is the attribute labeling the row and b is the attribute labeling the column, then the first two rows of the table list the supports of $\overline{\{a, b\}}$ in positive and negative classes, and the final two lines evaluate the emergence of the pattern $\overline{\{a, b\}}$.

	ERROR:XL	ERROR:LX	ERROR:MM	ERROR:SS
STABILITY:stab	(2025 / 2279)	(2025 / 2120)	(2025 / 848)	(4293 / 901)
STABILITY:xstab	(640 / 640)	(640 / 640)	(512 / 320)	(1280 / 0)
STABILITY:stab	0.9	1	2.4	5
STABILITY:xstab	1	1	1.6	∞

were calculated on a Dell Latitude D610: Pentium M, 800MHz, 1GB RAM. The example of 12 PCBs had 4 binary attributes, 12 objects, 8 concepts, 4 highly-stimulating concepts, and the highly stimulating concepts were extracted in < 1 second, producing a matrix $\{SS(a) : a \in \mathcal{L}\}$ with size 458b. The example of shuttle-landing-control had 16 binary attributes, 253 objects, 2040 concepts, of which 38 were extracted as highly-stimulating concepts; the highly stimulating concepts were extracted in 211 seconds, producing a matrix $\{SS(a) : a \in \mathcal{L}\}$ with size 1.98 Mb.

In conclusion, by testing the implemented algorithm on some small databases, we were able to discover homogeneously and highly stimulating patterns, which were a useful generalization of the jumping emerging patterns and yet were not too numerous. On the other hand, we were able to discover non-homogeneous interaction between attributes. The notion of stimulation captures some notions already studied in the literature. We generalize this notion to the interaction between any pair of patterns; the implementation allows for the extraction of patterns with either contrast-discriminative capacity or high, homogeneous stimulation, which should prove to be as useful as the contrast-discriminative pairs of attributes and as measures of confidence in further applications.

5 Further Work and Conclusion

There are interesting theoretical directions for further work, as well. We can discuss the stimulation of patterns on other patterns with respect to any concept-evaluating or pattern-evaluating measure, not only emergence. A second direction for further theoretical expansion is to consider how a single variable affects a pattern. If the variable a has attribute a_i and a_i is in the intent of a pattern c , then to evaluate the degree to which a_i influences c , we must consider an attribute c' as close to c as possible, and yet in which a_i is replaced by another value $a_j, j \neq i$ of the variable a . c' will be $\overline{d \cup \{a_j\}}$ for some $d > c$ for which $a_i \notin \overline{d}$. In this way, we can answer the question of how the attribute a_i in the intent of a pattern influences the pattern's emergence.

We hope to apply the stimulating patterns practically as well. First, we will study precisely the scalability of the algorithm to larger datasets. We will apply stimulating patterns to datasets where “factor interplay” is high, and the effect

of any one attribute depends on other attributes. We plan experiments on larger databases. We will explore applications to the presentation and compression of emerging patterns, to classification, and to supervision of rule-finding algorithms.

In this article we have introduced the notion of stimulating patterns in the context of formal concept analysis. Using the lattice-theoretic framework, we have generalized some notions which are known to be valuable in data mining to a new notion, the stimulation of a pattern. This notion has some attractive theoretical properties. We have implemented an algorithm to compute stimulation throughout a dataset, and we have extracted from the stimulation matrix two types of information which are already of known interest in the data-mining community.

Acknowledgement. we thank the regional council of Basse-Normandie for financial support (“programme emergence”).

References

1. Dong, G., Li, J.: Applications of Emerging Patterns for Microarray Gene Expression Data Analysis. In: Liu, L., Tamer Özsu, M. (eds.) Encyclopedia of Database Systems, vol. 107. Springer, Heidelberg (2009)
2. Dong, G., Li, J.: Efficient mining of emerging patterns: discovering trends and differences. In: KDD 1999: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 43–52. ACM, New York (1999)
3. Dong, G., Zhang, X., Wong, L., Li, J.: CAEP: Classification by Aggregating Emerging Patterns. In: Arikawa, S., Furukawa, K. (eds.) DS 1999. LNCS (LNAI), vol. 1721, pp. 30–42. Springer, Heidelberg (1999)
4. Fayyad, U.M., Irani, K.B.: The Attribute Selection Problem in Decision Tree Generation. In: AAAI, pp. 104–110 (1992)
5. Ganter, B., Stumme, G., Wille, R.: Formal Concept Analysis: Foundations and Applications. LNCS (LNAI). Springer, New York (2005)
6. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. In: Trans. C. Franzke. Springer, New York (1997)
7. Harrell Jr., Frank, E.: Regression Modeling Strategies. Springer, New York (2006)
8. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer Series in Statistics (2001)
9. Huang, H.-j., Qin, Y., Zhu, X., Zhang, J., Zhang, S.: Difference Detection Between Two Contrast Sets. In: Tjoa, A.M., Trujillo, J. (eds.) DaWaK 2006. LNCS, vol. 4081, pp. 481–490. Springer, Heidelberg (2006)
10. Li, J., Dong, G., Ramamohanarao, K.: Making use of the most expressive jumping emerging patterns for classification. Knowledge and Information Systems 3(2), 131–145 (2001)
11. Li, J., Wong, L.: Emerging patterns and gene expression data. Genome Informatics 12, 3–13 (2001)
12. Li, J., Yang, Q.: Strong Compound-Risk Factors: Efficient Discovery Through Emerging Patterns and Contrast Sets. IEEE Transactions on Information Technology in Biomedicine 5(11), 544–552 (2007)

13. Loekito, E., Bailey, J.: Using Highly Expressive Contrast Patterns for Classification - Is It Worthwhile? In: Theeramunkong, T., Kijssirikul, B., Cercone, N., Ho, T.-B. (eds.) PAKDD 2009. LNCS, vol. 5476, pp. 483–490. Springer, Heidelberg (2009)
14. Loekito, E., Bailey, J.: Mining influential attributes that capture class and group contrast behaviour. In: Shanahan, J.G., Amer-Yahia, S., Manolescu, I., Zhang, Y., Evans, D.A., Kolcz, A., Choi, K.-S., Chowdhury, A. (eds.) Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM, Napa CA, USA, pp. 971–980 (2008)
15. Loekito, E., Bailey, J.: Fast mining of high dimensional expressive contrast patterns using zero-suppressed binary decision diagrams. In: KDD 2006: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 307–316, 1-59593-339-5 (2006)
16. Poezevara, G., Cuissart, B., Crémilleux, B.: Discovering Emerging Graph Patterns from Chemicals. In: Rauch, J., Raś, Z.W., Berka, P., Elomaa, T. (eds.) Foundations of Intelligent Systems. LNCS, vol. 5722, pp. 45–55. Springer, Heidelberg (2009)
17. Ramamohanarao, K., Bailey, J., Fan, H.: Efficient Mining of Contrast Patterns and Their Applications to Classification. In: ICISIP 2005: Proceedings of the 2005 3rd International Conference on Intelligent Sensing and Information Processing, Washington, DC, USA, pp. 39–47, 0-7803-9588-3. IEEE Computer Society, Los Alamitos (2005)
18. Ramamohanarao, K., Fan, H.: Patterns Based Classifiers. In: World Wide Web, Hingham, MA, USA, vol. 1(10), pp. 71–83, 1386-145X. Kluwer Academic Publishers, Dordrecht (2007)
19. Ting, R.M.H., Bailey, J.: In: Ghosh, J., Lambert, D., Skillicorn, D.B., Srivastava, J. (eds.) Proceedings of the Sixth SIAM International Conference on Data Mining, SDM, Bethesda, MD, USA, April 20-22. SIAM, Philadelphia (2006)
20. Valtchev, P., Grosser, D., Roume, C., Hacene, M.R.: Galicia: An Open Platform for Lattices. In: Using Conceptual Structures: Contributions to the 11th Intl. Conference on Conceptual Structures, pp. 241–254 (2003)
21. Webb, G., Butler, S., Newlands, D.: On detecting differences between groups. In: KDD 2003: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 256–265. ACM, New York (2003)