Condensed Survey on Condensed Representations of Patterns

Arnaud Soulet[1,*], François Rioult[2] and Bruno Crémilleux[2]

[1]Université de Tours, LIFAT, Blois, France
[2]UNICAEN, ENSICAEN, CNRS - UMR GREYC, Normandie Univ 14000 Caen, France

Abstract

Pattern condensed representations are an important concept in inductive databases. They aim at summarizing the information contained in a set of patterns. This paper is not an exhaustive survey of the 20 years of proposals on condensed representations but rather a general overview. We propose a simple formalism modeling a large part of the exact and approximate condensed representations. Through this modeling, we show that many of the condensed representations are based on the same foundations, we highlight the compression and generalization sides of the condensed representations and we describe several trends. Finally, we discuss the current extents of the work on condensed representations within pattern mining.

Keywords

Pattern Mining, Inductive Database, Condensed Representation

1. Introduction

Pattern condensed representations is a central concept in knowledge discovery in inductive databases [1] to cope with the "pattern flooding which follows data flooding" that is unfortunately so typical in exploratory Knowledge Discovery in Databases (KDD) processes. The key idea of pattern condensed representations is to take benefit of the redundancy of a collection of patterns to construct a concise representation of the patterns instead of mining all patterns. Given a set of patterns S, the principle of pattern condensed representations is to compute a set R of representative patterns, R being a representation of S which is lossless (Step 2 in Fig. 1) and R has to be as concise as possible. Then the whole collection of patterns S can be efficiently derived from R (Step 3 in Fig. 1) without coming back to data.

The research field on pattern condensed representations began at the origin of pattern mining. A lot of works were conducted about twenty year ago to define pattern condensed representations addressing the frequency measure [2, 3, 4, 5, 6, 7, 8] to name a few. Pattern condensed representations are a large part of the tutorial entitled "Inductive databases and Constraint-based Mining"at ECML/PKDD 2002 [9]. The paper "Mining All Non-derivable Frequent Itemsets" [10] won the award of the best paper at ECML/PKDD 2002. This paper

KDID 2022: 20th anniversary of KDID Workshop
*Corresponding author.
arnaud.soulet@univ-tours.fr (A. Soulet); francois.rioult@unicaen.fr (F. Rioult); bruno.cremilleux@unicaen.fr (B. Crémilleux)
0000-0001-8335-6069 (A. Soulet); 0000-0001-8294-9049 (B. Crémilleux)

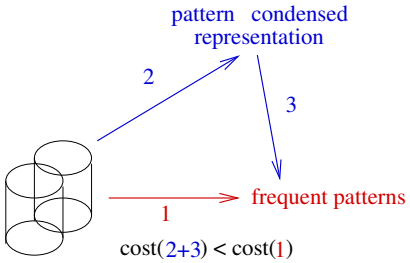

Figure 1: Principle of CR

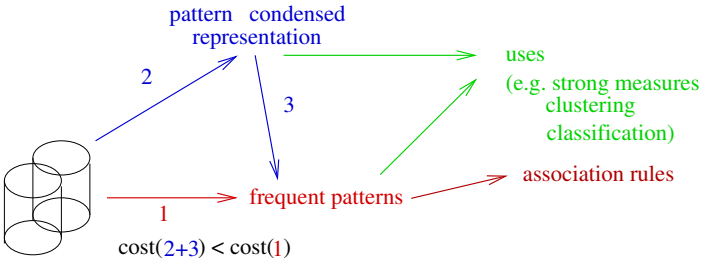

Figure 2: Examples of uses of CR

presents deduction rules to derive tight bounds on the frequency of patterns. These rules allow for constructing a minimal representation for all frequent itemsets. Ten years after, these authors won the ten year award talk with this paper.

More generally, many research directions on condensed representations have been developed. The principle has been extended from items to other pattern languages such as sequences [11, 12] or graphs [13]. In order to improve the efficiency of the approach and to get more concise representations, approximate condensed representations [3] were proposed (with approximate representations, a bounded error is tolerated when the whole collection of patterns is regenerated). On the other hand, a few works address measures other than frequency [14, 15].

At the beginning, pattern condensed representations focus on the obsession with the speed of algorithms. Literature emphasizes that the total running time of Step 1 and 2 is in general much smaller than that of Step 3 (cf. Fig. 1). Interestingly, the pattern mining community over time has witnessed a sharp shift from efficiency-based approaches to multiple uses of pattern condensed representations. This interest comes from the nice properties satisfied by the condensed representations and their representative patterns. As an example, let us consider the frequent closed patterns which are the representative patterns of condensed representations based on closed patterns. A closed pattern captures the maximum amount of similarity of a set of data, which is an essential property in conceptual clustering [16]. Several data mining methods in the clustering field are based on these patterns [17, 18]. Closed patterns also optimize frequency-based measures [19] and looking for closed patterns is enough to provide the best patterns according to these measures. Regarding associative classification [20], free or δ-free patterns are suitable to design rules minimizing the risk of overfitting by pointing the minimal information allowing to conclude on a class for a set of data [21, 22]. Fig. 2 provides a few examples of uses of representative patterns.

In this paper, we propose a simple formalism modeling a large part of the exact and approximate condensed representations. This formalism defines an abstraction on the characteristics of the condensed representations and it is based on three ingredients: a function f which represents the measure or the query according to which the condensation is performed, a function ρ_R providing the representative patterns and a binary relation $\simeq$ expressing the relationship between a pattern and its representative. Presenting pattern condensed representations through this formalism gives an unifying point of view enabling us to better understand the links between

the different types of condensed representations. Especially, we think that the relation $\simeq$ is an original characterization of condensed representations and nicely explains the differences between exact, approximate and dominant representations. In short, our formalization shows that many of the condensed representations are based on the same foundations. Especially, this formalization highlights the compression and generalization sides of the condensed representations. We propose a brief review of the 20 years of condensed representation proposals with some messages. Finally, we draw up a few research lines of pattern mining as a legacy of condensed representations.

The paper is not an exhaustive survey of the 20 years of proposals on condensed representations. For more details on extraction techniques and methods for condensed representations of frequent patterns, we recommend reading [23]. Note that we focus on the research contributions of the pattern mining community but there are strong links and obvious intersections with the Formal Concept Analysis field [24].

The guidelines of the paper are as follows. We start by recalling preliminaries on pattern condensed representations and we describe our formalism to model condensed representations. In Section 3, we present exact condensed representations. Approximate and dominant representations are reported in Section 4. Finally, we discuss in Section 5 the current extents and trends of the work on condensed representations within pattern mining.

2. Preliminaries

Basic definitions Condensed representations of patterns take place in the framework of inductive databases [1, 25]. Given a *database* D, there is a *language* L which describes the sentences (or subgroups or *patterns*) of D. L is provided with a specialization relation $\preceq$. For example, L can describe itemsets, sequences, trees, graphs or association rules. Note that $<$ also denotes an arbitrary total order over the language L. Pattern mining is the art of finding interesting sentences in L, and this interestingness is modeled by a function f, whose range is either a number – f is a *measure* – or a Boolean – f is a *constraint* or a *query*. In general, pattern mining methods return collections of patterns, named *theory*, that are far too large and hold many redundancies. Quite quickly, it appeared necessary to summarize (or even generalize) these patterns based on the notion of representation:

Definition 1 (Representation). *Given a set of patterns $S \subseteq L$ and a binary relation $\simeq \subseteq L \times L$, a set of patterns $R \subseteq S$ is a* representation *of S for $\simeq$ iff there exists a* surjective *function $\rho_R : S \to R$ such that $X \simeq \rho_R(X)$.*

Representations are precious because they allow to focus on particular patterns whose value of the interest function are emblematic (for an underlying relation $\simeq$ that is not necessarily symmetrical). In the following, we will wonder if the representation is exact (see Section 3), *i.e.* if the pattern in R gives the exact value of the function of the representative pattern in S ($X \simeq Y \Leftrightarrow f(X) = f(Y)$) or if it induces good properties on f (see Section 4) such as $X \simeq Y \Leftrightarrow f(X) \leq f(Y)$.

For example, let us consider the representation offered by the *closed* patterns. These patterns are useful in Boolean databases; the language here is the powerset of the set of items. In such

databases, one could be interested in finding the *frequent* itemsets. The closed patterns are the patterns whose supersets have a lower frequency, or they are the maximal elements of the equivalence class of support. Then, focusing on the only closed patterns allows to find the frequency of each pattern.

Interest of representations In order to be actionable, a condensed representation has to fulfill different requirements. First of all, the condensation has to be effective, i.e. the cardinal of R should be lower than this of S. In practice, there are orders of magnitude between these cardinals but there is no general formal relation between these cardinals. Indeed, extreme cases exist where datasets have 2^n frequent patterns but only one closed pattern (think about a database where each object is in relation with each item). *A contrario*, there are situations where each frequent pattern is closed.

Second, algorithms for computing the representative patterns *should be at least* as efficient as those used for computing the whole set of patterns. In fact, there is usually some extra cost to compute the representative patterns, but since there are far fewer of them, the overall calculation of the representation is paid for. Most of time, the representative function has nice properties: it is either *intensive* ($\rho_R(X) \subseteq X$) or *extensive* ($X \subseteq \rho_R(X)$). As the interestingness function f underlying of $\simeq$ is also increasing or decreasing according to the partial order, it gives rise to (anti-)monotone constraints which facilitate the pruning.

Third, the condensed representation should be useful and actionable. First approaches on condensed representation tried to show that it was easy to regenerate the whole collection starting only from the representation (see [26] for the example of the association rules), though the regeneration of the whole collection was not always useful. One could prefer an *inductive database approach* [27]: information on a pattern could be found using the representation function by querying the representation, without accessing the database. On the other hand, multiple direct uses of these representations were found (classification, clustering, basis of association rules) that avoid the regeneration of the whole collection.

Problem formulation Given a set of patterns S and a binary relation $\simeq$, the goal is to find a smallest representation R^* of S for $\simeq$:

$$R^* = \arg\min_{R \subseteq L} |R| \text{ subject to } (\forall X \in S)(X \simeq \rho_R(X))$$

Of course, the naive computation of this expression by enumerating all the possible set of patterns is just impossible. Hence, the idea is rather to enumerate the patterns $X \in S$ to be represented and to select for each one its representative $\rho_R(X)$. For example, in Section 3, we will define the condenser function which associates to each pattern a closed pattern to build a condensed representation of frequent patterns.

This formalization of the representation R^* puts forward an optimization problem where one of the smallest representations is sought. At first sight, it is just a compression problem where the patterns R^* allow to find all the information about S with an imprecision controlled by $\simeq$. This compression is also reminiscent of the Minimum Description Length principle where this smallest pattern set R^* would be the best code table of the pattern set S. With this perspective,

R^* appears rather as a generalization of S. In practice, we will see that this point of view is reinforced by the choice of the patterns which are the most general/specific (see the notion of condenser function in Section 3) and by the choice of the $\simeq$ relation to better generalize the data in spite of the noise or a non-compressible function f (see Section 4).

3. Exact Condensed Representations

Definition of exact condensed representation Intuitively, a condensed representation is just a set of patterns extracted individually from groups of contiguous patterns having the same value for a given function f. Considering Definition 1, each pattern X in S and its representative pattern $\rho_R(X)$ have the same value for f meaning that $X \simeq Y \Leftrightarrow f(X) = f(Y)$:

Definition 2 (Exact Condensed Representation). *Given a set of patterns $S \subseteq L$ and a function f, a representation R of S is an exact condensed representation adequate to f iff* $(\forall X \in S)(f(X) = f(\rho_R(X)))$.

Thanks to an exact condensed representation, it is possible to find exactly the value of the function f for any pattern of S without the original database D (considering of course that we keep the value of f for the representative patterns of R). For example, it will be possible for a pattern to know if it is frequent (f is then $X \mapsto \text{freq}(X, D) \geq \gamma$, see Section 3.1) or more precisely, how frequent it is (f is then the frequency freq, see Section 3.2). Note that only the pattern set S is represented. In practice, if we query the representative function ρ_R with a pattern outside of S, we will not find any representative. In such a situation, we may associate a default value (for instance, false for a predicate or 0 for the frequency). Other approaches recommend adding parts of the negative border to the condensed representation [23, 28].

Table 1 gives details about the most common condensed representations including maximal, closed, (δ-)free and strong patterns. For each representation, it precises the adequate function f being used, how is defined the representation and what is the representative function ρ_R. This table suggests several remarks. First of all, the representation function ρ_R is generic when the representation is known: it links a pattern with this of the representation which is directly above or below (according to the direction of the condenser function, see below). Note that when several representative patterns in R are available for a given pattern, it is necessary to use an arbitrary order $<$ in order to retain only one pattern. Second, the representative function ρ_R uses its image set. That means that the entire image set has to be computed before being able to compute the representative.

Condenser function To compute the condensed representation, it is therefore not possible to take advantage of the representative function ρ_R which is based on R. The direct computation of the definition of the condensed representation R as defined in Table 1 is not very interesting because it relies on the complete extraction of the set of patterns S. Another way to avoid this would be to use the adequate function to compute the representative pattern. In fact, there is a strong connection or even a duality between the definition of the representation and the representative function. For example, for the closed patterns, one could benefit from $X \mapsto \max_{\subseteq}\{Y \in S \ : \ X \subseteq Y \wedge f(Y) = f(X)\}$. It is then possible to go through the patterns and

to focus on the representative ones. More generally, most exact condensed representations will just retain, for each pattern, the largest or the smallest contiguous pattern with respect to an order relation $\preceq$ having the same value for the function f:

Definition 3 (Condenser function). *Given a function f and an order relation $\preceq$, the condenser function, denoted by $\kappa^{f,\preceq}$, is defined as $X \mapsto \max_{\preceq}\{Y \in L \ : \ (\forall X \preceq Z \preceq Y)(f(Z) = f(X))\}$.*

Definition 3 means that the condenser function returns the most specialized patterns w.r.t $\preceq$ having the same value for f. Note that for some order relation or some languages (e.g, for the graphs [29]), the condenser function returns several patterns, among which one will choose in an arbitrary way, for example according to an enumeration order.

The condenser function is useful for computing the condensed representation R by computing the representative of each pattern in S. In contrast, it cannot be used as a representative function because it involves the adequate function f. Instead, it is enough to take the closest pattern with respect to the order relation $\preceq$. The definition of the function $\kappa^{f,\preceq}$ is simple but its computation may be costly in practice. However, in many cases, its use is efficient, especially when it returns only one representative pattern for any pattern in L. Indeed, in such a situation, the condenser function is a closure operator leading to nice pruning properties of the search space or to transposition trick [30].

3.1. Borders, frontiers and version spaces

The first proposals of condensed representations [25] were just interested in summarizing a predicate q. Typically, they aimed at representing concisely by R the set S of all frequent patterns. In practice, a pattern X not represented in R means that X does not belong to S and that it does not satisfy q.

The maximal (resp. minimal) border of patterns generalizes this principle to the set of anti-monotone (resp. monotone) Boolean predicates q i.e., having a downward (resp. upward) closed solution space. Consequently, it is possible by using a maximal border and a minimal border together to represent any predicate generating a convex solution space [31]. For more complex constraints, it is possible to decompose a theory into a set of convex spaces [32]. Note that the maximal (resp. minimal) border can be calculated by applying the condenser function $\kappa^{q,\preceq}$ (resp. $\kappa^{q,\succeq}$) to all patterns in S. Dually, it is also possible to consider $L \setminus S$ by considering negative borders [25]. Negative borders represent rather the patterns not belonging to the set of patterns S. This is one of the rare cases where R is not included in S. Figure 3 illustrates the maximal patterns and the negative minimal border for a downward closed space S. Note that the most general patterns w.r.t $\preceq$ are at the top of the lattice represented as a diamond.

These boundary techniques compress very strongly the set of patterns satisfying a monotone and/or anti-monotone constraint. Beyond that, these borders delimiting a concept in the language is similar to the work on generalization in symbolic machine leaning [33] with version spaces.

From a compression perspective, all the patterns verifying an (anti-)monotone constraint can be regenerated without loss. But, it may be necessary to have more precise information on each pattern with respect to an interestingness measure motivating the condensed representations introduced in the next section.

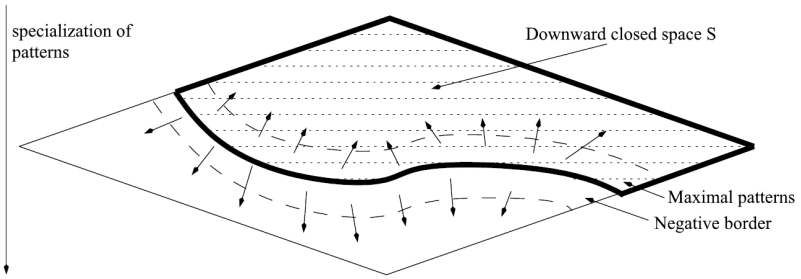

Figure 3: Illustration of borders for a downward closed space S

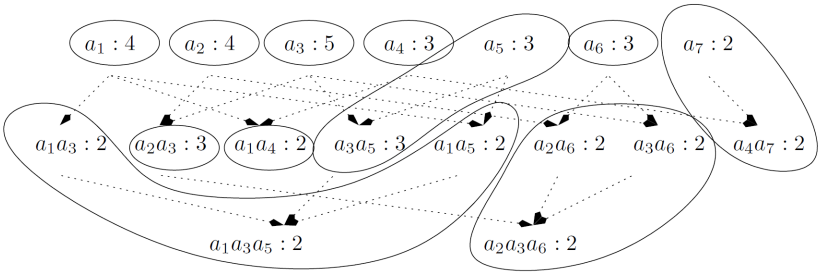

Figure 4: Illustration of minimal/maximal patterns of equivalence classes for the itemsets

3.2. Equivalence classes

Exact condensed representations for measures are based on the concept of equivalence class. All patterns that are comparable with respect to the order relation $\preceq$ and have the same value for the measure m belong to the same equivalence class. For example, in the case of frequency, this means that two patterns in the same equivalence class share the same transactions in the dataset. Then, a representative is chosen from each of the equivalence classes. In order to easily find the representative of a pattern, the simplest way is to choose the minimal patterns (named free patterns[1]) or maximal patterns (named closed patterns). It turns out that these patterns also have good properties to be computed efficiently and to be used in machine learning. In the literature, maximal patterns of equivalence classes are clearly the most used. A literature review in 2014 indicated that there were 4 times more papers dedicated to closed patterns than dedicated to free patterns [34]. Figure 4 illustrates the closed and free patterns for the itemset language. Each equivalence class is represented by a shape where all patterns are comparable w.r.t of inclusion and of equal frequency. For each equivalence class, the most general itemsets are free while the most specific itemset is closed.

Introduced in [6], the frequent closed itemsets are the largest patterns with equal frequency. They correspond exactly to the condensed representation computed by $\kappa^{\text{freq},\subseteq}$, which is the main closure operator used in Formal Concept Analysis. Interestingly, it has been shown that

[1]Also named key patterns or generators.

this representation of closed patterns is also adequate for all measures involving sub-datasets $D_i \subseteq D$ (e.g., confidence or growth rate) [35, 36]. In contrast, specific condensed representations are necessary for other interestingness measures. For example, it is possible to use the $\kappa^{f,\subseteq}$ operator for aggregate functions min, max or sum [15]. The notion of closed patterns has mostly been extended to a large set of languages including sequences [37], trees [38], graphs [39], and so on. There are also works that can take into account various languages by exploiting the notion of set systems [40].

As mentioned before, the free patterns [4] are the minimal patterns of the equivalence classes. They correspond exactly to the patterns computed with $\kappa^{\mathsf{freq},\supseteq}$. There are also proposals for other interestingness measures including the disjunctive frequency [41] and the bond measure [42]. For most languages, several minimal patterns may exist for the same equivalence class leading to difficulties during their computation. However, pattern mining methods have been proposed for sequential patterns [11, 43], graphs [13] and set systems [44]. It should be noted that several generalizations of free patterns exploit the inclusion-exclusion principle such as non-derivable itemsets [10] (with a non trivial extension to sequences [45]) or k-free [46], which is very specific to frequency[2]. This principle allows to eliminate some free itemsets whose frequency can be deduced from the others [23]. Free patterns are often used as left-hand sides for association rules [12] as a free pattern captures the shortest information to infer a rule conclusion and therefore reducing the risk of overfitting. Apart from this usage, these patterns and their generalization have received much less attention in the literature than closed patterns. One possible explanation is that it rather captures a lack of correlation between the items belonging to the pattern.

4. Non-Exact Condensed Representations

4.1. Approximate condensed representations

The cost of extracting exact condensed representations may remain high (especially in noisy data). This observation explains why from the origin of condensed representations researchers are looking for approximate condensed representation frameworks [27]. In short, an approximate condensed representation is a set of patterns that enables to recover the values of the function f on the patterns of S but eventually with some loss of precision (bounded by an error ϵ). Considering Definition 1, it means that given the adequate function f, the similarity $X \simeq Y$ equals to $|f(X) - f(Y)| \leq \epsilon$ leading to the following definition:

Definition 4 (Approximate Condensed Representation). *Given a set of patterns $S \subseteq L$ and a function f, a representation R of S is an approximate condensed representation adequate to f with an error $\epsilon \geq 0$ iff $(\forall X \in S)(|f(X) - f(\rho_R(X))| \leq \epsilon)$.*

δ-free [3] itemsets form one of the most famous approximate condensed representations. The key intuition of a δ-free itemset X is the absence of rule between X and its subsets with at most δ exceptions. The higher δ, the more concise the representation. The freeness of itemsets is

[2] We think that these condensed representations are purely (nice) compression techniques, but the representative patterns have a lower ability to generalize.

anti-monotone in that sense that if X is not a free itemset then none of its supersets can be a free itemsets, leading to efficient mining algorithms. Beyond a higher conciseness and a lower mining time, this representation allowed to induce classification rules with a controlled and high confidence [22]. Again, the compression initially sought has given way to uses (without regeneration) in symbolic machine learning.

Approximate condensed representations have clearly received less attention than exact condensed representations because the latter do not generalize naturally, especially for maximal patterns of equivalence classes such as [47]. However, a generalization framework to different interestingness measures exists for free patterns [48] assuming that the language is representable by a set system. As for the δ-free, it is then possible to tolerate an error ϵ for aggregate functions and disjunctive function.

4.2. Dominant Representations

Constraint-based pattern mining just returns too many patterns even by using condensed representations. It then appears necessary to switch to optimization techniques by selecting only the best patterns. It is relevant to look at representations where we no longer have an equality between the represented pattern $X \in S$ and its representative $\rho_R(X)$, but only a upper bound.

Considering Definition 1, it means that given the adequate function f, the similarity $X \simeq Y$ equals to the non-symmetrical relation $f(X) \leq f(Y)$ leading to the following definition:

Definition 5 (Dominant Representation). *Given a set of patterns $S \subseteq L$ and a function f, a representation R of S is a dominant representation for f iff $(\forall X \in S)(f(X) \leq f(\rho_R(X)))$.*

It is easy to see that any exact condensed representation adequate to f is also a dominant representation for f. But, more interestingly, there are dominant representations with situations where $f(X)$ is strictly less than $f(\rho_R(X))$. For example, δ-free forms a dominant representation for frequency. There are even more favorable situations where a condensed representation adequate to f can serve as dominant representation for another function g (whereas R is not a condensed representation adequate to g). For instance, let us consider a dataset D where $D_1 \subseteq D$, the set of closed patterns in D_1 forms a condensed representation adequate to $\text{freq}(X, D_1)$ (i.e., $\text{freq}(X, D_1) = \text{freq}(\kappa^{\text{freq}_{D_1}, \subseteq}(X), D_1)$ where $\kappa^{\text{freq}_{D_1}, \subseteq}$ is the closure operator in D_1). This same representation is a dominant representation with respect to the frequency-based measures because the closed pattern of X in D is a subset of the closed pattern of X in D_1 implying, $\text{freq}(\kappa^{\text{freq}_{D}, \subseteq}(X), D) \geq \text{freq}(\kappa^{\text{freq}_{D_1}, \subseteq}(X), D)$. This dominant representation is relevant because it retains the patterns characteristic of D_1 by maximizing many frequency-based interestingness measures such as the growth rate[35] or confidence [36, 49].

This principle can be generalized to all condensed representations:

Property 1. *Given an exact condensed representation R of a pattern set $S \subseteq L$ adequate to a function f with ρ_R as representative function, R is also a dominant representation for g if (i) ρ_R is extensive w.r.t $\preceq$ and (ii) $f(X) = f(Y)$ implies $g(X) \leq g(Y)$ for all $X \preceq Y$.*

This property is straightforward, but very relevant in practice. For instance, the area measure $X \mapsto |X| \times \text{freq}(X, D)$ has no interesting condensed representation because of the length. Nevertheless, the condensed representation of closed patterns is a dominant representation because, with equal frequency between two patterns $X \subseteq Y$, Y has a larger area. The computation of skypatterns with respect to a set of measure M benefits from a dominant relation for the function $X \mapsto (m^{(1)}(X), \dots, m^{(k)}(X))$ [50]. Indeed, given a set of measures $M = \{m^{(1)}, \dots, m^{(k)}\}$, a pattern in L is a Pareto-optimal pattern (or skypattern) with respect to M iff there is no other pattern Y such that $m^{(i)}(X) \leq m^{(i)}(Y)$ for $i \in \{1, \dots, k\}$ with at least one measure $m^{(j)}$ where $m^{(j)}(X) < m^{(j)}(Y)$.

In the end, dominant condensed representations have a twofold advantage when a task requires a complex function f for which no exact condensed representations exist. First, they allow to build very concise representations adequate to the targeted task. Secondly, the selected patterns are a good compromise between diversity and quality. Indeed, the dominant condensed representations cover well the whole language L unlike the top-k patterns, which often focus on a small part of the language leading to strong redundancies between patterns.

5. Discussion

In this paper, we have proposed a simple formalism modeling a large part of the exact and approximate condensed representations. This formalization shows that many of the condensed representations are based on the same foundations. It characterizes differences between exact, approximate and dominant representations. Whereas a condensed representation is often seen as a compression tool, the proposed framework clarifies the compression and generalization sides of the condensed representations.

The end of the paradigm of the exhaustive search. There was an active research around pattern condensed representations during the "golden age" of pattern mining (2000-2005) with notably two consecutive awards at ECML/PKDD in 2002 [10] and 2003 [46]. However, ten years after these successes, when he received at ECML/PKDD 2012 the award for the most influential paper, Toon Calders gave the following message: "Please, please stop making new algorithms for mining all patterns". Clearly, we observe that works on pattern mining move from an exhaustive search to non-exhaustive methods. Is the end of the paradigm of the exhaustive search means that this paper is the last survey on condensed representations? We think that ideas and principles of condensed representations are disseminated in several research directions that we sketch below.

Advance uses of condensed representations. Condensed representations are used for their good generalization properties which are further developed with approximate and dominant representations. A lot of data mining tasks such as the search of contrast patterns, the design of classification systems and more generally the design of models take benefit of condensed representations. This trend has been accompanied by works on pattern set mining [51, 52] which, even better, guarantee a complementarity between patterns in order to reduce redundancies. On the other hand, for the past few years, optimization techniques are

introduced in pattern mining to select the best patterns according to given criteria. Dominant representations are especially useful in this context (cf. Section 4.2). Condensed representations are at the basis of a recent method of multi-objective optimization of several functions [53]. Contrary to pattern sets, optimization in condensed representations are local (the quality of a pattern in R does not depend on the other patterns of R).

Modern compression techniques for pattern mining. The idea of compressing to keep the best patterns is still at the heart of most of the recent pattern mining methods using notably the Minimum Description Length principle [54, 55] or Boolean matrix factorization [56]. A major difference is that these works are based on the compression of the dataset, and not on a collection of patterns extracted from the dataset. The MDL principle is looking for the simple model which fits the data. In practice, a trade-off is done between the complexity of the model and the fit of the model to the data. Compression is used as a tool to compare models. Many data mining tasks (e.g. segmentation, classification, missing values) have been revisited through MDL [57, 58]. The main objective when applying the MDL principle is to moved from mining collections of good patterns to mining good collections of patterns.

Interactive pattern mining. From the perspective of inductive databases, condensed representations were seen as caches of patterns to speed up answers to future queries. The idea was really to extract as much information as possible to be able to answer any query. Nowadays, interactive methods dedicated to pattern mining rather exploit fast and repeated extractions by benefiting from Beam Search [59], Monte Carlo Tree Search [60] or even, sampling [61] methods. Indeed, these non-exhaustive techniques have the advantage of being fast enough to extract patterns on the fly while relying on complex interestingness measures, updated at each iteration by integrating user feedback.

References

[1] T. Imielinski, H. Mannila, A database perspective on knowledge discovery, Commun. ACM 39 (1996) 58–64.

[2] Y. Bastide, R. Taouil, N. Pasquier, G. Stumme, L. Lakhal, Mining frequent patterns with counting inference, SIGKDD Explor. 2 (2000) 66–75.

[3] J.-F. Boulicaut, A. Bykowski, C. Rigotti, Approximation of frequency queries by means of free-sets, in: European Conference on Principles of Data Mining and Knowledge Discovery, Springer, 2000, pp. 75–85.

[4] J.-F. Boulicaut, A. Bykowski, C. Rigotti, Free-sets: a condensed representation of boolean data for the approximation of frequency queries, Data Mining and Knowledge Discovery 7 (2003) 5–22.

[5] A. Bykowski, C. Rigotti, A condensed representation to find frequent patterns, in: P. Buneman (Ed.), Proceedings of the Twentieth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, May 21-23, 2001, Santa Barbara, California, USA, ACM, 2001.

[6] N. Pasquier, Y. Bastide, R. Taouil, L. Lakhal, Discovering frequent closed itemsets for association rules, in: International Conference on Database Theory, Springer, 1999, pp. 398–416.
[7] J. Wang, J. Han, J. Pei, CLOSET+: searching for the best strategies for mining frequent closed itemsets, in: L. Getoor, T. E. Senator, P. M. Domingos, C. Faloutsos (Eds.), Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 24 - 27, 2003, ACM, 2003, pp. 236–245.
[8] M. J. Zaki, C.-J. Hsiao, CHARM: An efficient algorithm for closed itemset mining, in: 2nd SIAM International Conference on Data Mining, 2002.
[9] J.-F. Boulicaut, L. De Raedt, Inductive databases and constraint-based mining, tutorial co-located with ECML/PKDD 2002, 2002. https://www.cs.helsinki.fi/events/ecmlpkdd/pdf/deraedt.pdf.
[10] T. Calders, B. Goethals, Mining all non-derivable frequent itemsets, in: European Conference on Principles of Data Mining and Knowledge Discovery, Springer, 2002, pp. 74–86.
[11] C. Gao, J. Wang, Y. He, L. Zhou, Efficient mining of frequent sequence generators, in: Proceedings of the 17th international conference on World Wide Web, 2008, pp. 1051–1052.
[12] D. Lo, S.-C. Khoo, L. Wong, Non-redundant sequential rules—theory and algorithm, Information Systems 34 (2009) 438–453.
[13] Z. Zeng, J. Wang, J. Zhang, L. Zhou, Fogger: an algorithm for graph generator discovery, in: Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology, 2009, pp. 517–528.
[14] A. Giacometti, D. Laurent, C. T. Diop, Condensed representations for sets of mining queries, in: M. Klemettinen, R. Meo (Eds.), Proceedings of the First International Workshop on Inductive Databases, 20 August 2002, Helsinki, Finland, Helsinki University Printing House, Helsinki, 2002, pp. 5–19.
[15] A. Soulet, B. Crémilleux, Adequate condensed representations of patterns, Data mining and knowledge discovery 17 (2008) 94–110.
[16] R. S. Michalski, R. E. Stepp, Automated construction of classifications: Conceptual clustering versus numerical taxonomy, IEEE Trans. Pattern Anal. Mach. Intell. 5 (1983) 396–410.
[17] T. Dao, C. Kuo, S. S. Ravi, C. Vrain, I. Davidson, Descriptive clustering: ILP and CP formulations with applications, in: J. Lang (Ed.), Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden, ijcai.org, 2018, pp. 1263–1269.
[18] N. Durand, B. Crémilleux, ECCLAT: a New Approach of Clusters Discovery in Categorical Data, in: 22nd Int. Conf. on Knowledge Based Systems and Applied Artificial Intelligence (ES'02), Cambridge, UK, 2002, pp. 177–190.
[19] J. Li, G. Liu, L. Wong, Mining statistically important equivalence classes and delta-discriminative emerging patterns, in: KDD, 2007, pp. 430–439.
[20] B. Liu, W. Hsu, Y. Ma, Integrating classification and association rule mining, in: R. Agrawal, P. E. Stolorz, G. Piatetsky-Shapiro (Eds.), Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98), New York City, New York, USA, August 27-31, 1998, AAAI Press, 1998, pp. 80–86.
[21] R. J. Bayardo, The hows, whys, and whens of constraints in itemset and rule discovery, in: Constraint-Based Mining and Inductive Databases, European Workshop on Inductive

Databases and Constraint Based Mining, Hinterzarten, Germany, March 11-13, 2004, Revised Selected Papers, volume 3848 of *Lecture Notes in Computer Science*, Springer, 2004, pp. 1–13.

[22] B. Crémilleux, J.-F. Boulicaut, Simplest rules characterizing classes generated by δ-free sets, in: Research and development in intelligent systems XIX, Springer, 2003, pp. 33–46.

[23] T. Calders, C. Rigotti, J.-F. Boulicaut, A survey on condensed representations for frequent sets, Constraint-based mining and inductive databases (2006) 64–80.

[24] B. Ganter, R. Wille, Formal concept analysis: mathematical foundations, Springer Science & Business Media, 2012.

[25] H. Mannila, H. Toivonen, Levelwise search and borders of theories in knowledge discovery, Data mining and knowledge discovery 1 (1997) 241–258.

[26] N. Pasquier, R. Taouil, y. Bastide, G. Stumme, L. Lakhal, Generating a condensed representation for association rules, Journal of Intelligent Information Systems 24 (2005) 29–60.

[27] H. Mannila, H. Toivonen, Multiple uses of frequent sets and condensed representations (extended abstract), in: E. Simoudis, J. Han, U. M. Fayyad (Eds.), Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, Oregon, USA, AAAI Press, 1996, pp. 189–194.

[28] M. Kryszkiewicz, Reducing borders of k-disjunction free representations of frequent patterns, in: ACM Symposium on Applied Computing (SAC'04), Nicosia, Cyprus, 2004, pp. 559–563.

[29] G. C. Garriga, R. Khardon, L. De Raedt, Mining closed patterns in relational, graph and network data, Annals of mathematics and artificial intelligence 69 (2013) 315–342.

[30] F. Rioult, J.-F. Boulicaut, B. Crémilleux, J. Besson, Using transposition for pattern discovery from microarray data, in: Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery, 2003, pp. 73–79.

[31] L. De Raedt, S. Kramer, The levelwise version space algorithm and its application to molecular fragment finding, in: International Joint Conference on Artificial Intelligence, volume 17, Citeseer, 2001, pp. 853–862.

[32] L. D. Raedt, M. Jaeger, S. D. Lee, H. Mannila, A theory of inductive query answering, in: Inductive databases and constraint-based data mining, Springer, 2010, pp. 79–103.

[33] T. M. Mitchell, Generalization as search, Artificial intelligence 18 (1982) 203–226.

[34] A. Giacometti, D. H. Li, P. Marcel, A. Soulet, 20 years of pattern mining: a bibliometric survey, ACM SIGKDD Explorations Newsletter 15 (2014) 41–50.

[35] A. Soulet, B. Crémilleux, F. Rioult, Condensed representation of emerging patterns, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, 2004, pp. 127–132.

[36] A. Soulet, B. Crémilleux, F. Rioult, Condensed representation of EPs and patterns quantified by frequency-based measures, in: International Workshop on Knowledge Discovery in Inductive Databases, Springer, 2004, pp. 173–189.

[37] X. Yan, J. Han, R. Afshar, Clospan: Mining: Closed sequential patterns in large datasets, in: Proceedings of the 2003 SIAM international conference on data mining, SIAM, 2003, pp. 166–177.

[38] A. Termier, M.-C. Rousset, M. Sebag, Dryade: a new approach for discovering closed

frequent trees in heterogeneous tree databases, in: Fourth IEEE International Conference on Data Mining (ICDM'04), IEEE, 2004, pp. 543–546.
[39] X. Yan, J. Han, Closegraph: mining closed frequent graph patterns, in: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, 2003, pp. 286–295.
[40] H. Arimura, T. Uno, Polynomial-delay and polynomial-space algorithms for mining closed sequences, graphs, and pictures in accessible set systems, in: Proceedings of the 2009 SIAM International Conference on Data Mining, SIAM, 2009, pp. 1088–1099.
[41] A. Casali, R. Cicchetti, L. Lakhal, Essential patterns: A perfect cover of frequent patterns, in: International Conference on Data Warehousing and Knowledge Discovery, Springer, 2005, pp. 428–437.
[42] S. Bouasker, T. Hamrouni, S. B. Yahia, Efficient mining of new concise representations of rare correlated patterns, Intelligent Data Analysis 19 (2015) 359–390.
[43] D. Lo, S.-C. Khoo, J. Li, Mining and ranking generators of sequential patterns, in: Proceedings of the 2008 SIAM International Conference on Data Mining, SIAM, 2008, pp. 553–564.
[44] A. Soulet, F. Rioult, Efficiently depth-first minimal pattern mining, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, 2014, pp. 28–39.
[45] C. Raïssi, T. Calders, P. Poncelet, Mining conjunctive sequential patterns, Data Min. Knowl. Discov. 17 (2008) 77–93.
[46] T. Calders, B. Goethals, Minimal k-free representations of frequent sets, in: European Conference on Principles of Data Mining and Knowledge Discovery, Springer, 2003, pp. 71–82.
[47] H. Cheng, S. Y. Philip, J. Han, AC-Close: Efficiently mining approximate closed itemsets by core pattern recovery, in: Sixth International Conference on Data Mining (ICDM'06), IEEE, 2006, pp. 839–844.
[48] A. Soulet, F. Rioult, Exact and approximate minimal pattern mining, in: Advances in Knowledge Discovery and Management, Springer, 2017, pp. 61–81.
[49] P. K. Novak, N. Lavrač, G. I. Webb, Supervised descriptive rule discovery: A unifying survey of contrast set, emerging pattern and subgroup mining., Journal of Machine Learning Research 10 (2009).
[50] W. Ugarte, P. Boizumault, B. Crémilleux, A. Lepailleur, S. Loudni, M. Plantevit, C. Raïssi, A. Soulet, Skypattern mining: From pattern condensed representations to dynamic constraint satisfaction problems, Artificial Intelligence 244 (2017) 48–69.
[51] A. Knobbe, B. Crémilleux, J. Fürnkranz, M. Scholz, From local patterns to global models: The lego approach to data mining, in: Int. Workshop "From Local Patterns to Global Models" co-located with ECML/PKDD'08, Antwerp, Belgium, 2008, pp. 1–16.
[52] L. D. Raedt, A. Zimmermann, Constraint-based pattern set mining, in: proceedings of the 2007 SIAM International conference on Data Mining, SIAM, 2007, pp. 237–248.
[53] C. Vernerey, S. Loudni, N. Aribi, L. Yahia, Threshold-free pattern mining meets multi-objective optimization: Application to association rules, in: Proceedings of the 31 Int. Joint Conf. on Artificial Intelligence, IJCAI 2022, Vienne, Austria, July, 2022.
[54] E. Galbrun, The minimum description length principle for pattern mining: A survey, Data mining and knowledge discovery (2022). https://doi.org/10.1007/s10618-022-00846-z.

[55] J. Vreeken, M. Van Leeuwen, A. Siebes, Krimp: mining itemsets that compress, Data Mining and Knowledge Discovery 23 (2011) 169–214.
[56] P. Miettinen, Boolean tensor factorizations, in: 2011 IEEE 11th International Conference on Data Mining, IEEE, 2011, pp. 447–456.
[57] M. van Leeuwen, J. Vreeken, A. Siebes, Compression picks item sets that matter, in: PKDD 2006, 10th European Conference on Principles and Practice of Knowledge Discovery in Databases, Berlin, Germany, September 18-22, 2006, Proceedings, volume 4213 of *Lecture Notes in Computer Science*, Springer, 2006, pp. 585–592.
[58] J. Vreeken, A. Siebes, Filling in the blanks - krimp minimisation for missing data, in: Proceedings of the 8th IEEE International Conference on Data Mining (ICDM 2008), December 15-19, 2008, Pisa, Italy, IEEE Computer Society, 2008, pp. 1067–1072.
[59] M. v. Leeuwen, Interactive data exploration using pattern mining, in: Interactive knowledge discovery and data mining in biomedical informatics, Springer, 2014, pp. 169–182.
[60] G. Bosc, J.-F. Boulicaut, C. Raïssi, M. Kaytoue, Anytime discovery of a diverse set of patterns with monte carlo tree search, Data mining and knowledge discovery 32 (2018) 604–650.
[61] M. Boley, C. Lucchese, D. Paurat, T. Gärtner, Direct local pattern sampling by efficient two-step random procedures, in: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, 2011, pp. 582–590.

Table 1
Examples of representations

Name	Adequate function f	Representation R	Representative function ρ_R
exact condensed representations $X \simeq Y \Leftrightarrow f(X) = f(Y)$			
Maximal patterns	anti-monotone const. q	$\max_{\preceq} S$	$X \mapsto \min_{\preceq}\{Y \in R \ : \ X \preceq Y\}$
Minimal patterns	monotone const. q	$\min_{\preceq} S$	$X \mapsto \min_{\preceq}\{Y \in R \ : \ Y \preceq X\}$
Negative min. border	anti-monotone const. q	$\min_{\preceq} L \setminus S$	$X \mapsto \min_{\preceq}\{Y \in R \ : \ Y \preceq X\}$
Negative max. border	monotone const. q	$\max_{\preceq} L \setminus S$	$X \mapsto \min_{\preceq}\{Y \in R \ : \ X \preceq Y\}$
Closed patterns	f	$\{X \in S \ : \ (\forall Y \succ X)(f(Y) \neq f(X))\}$	$X \mapsto \min_{\preceq \wedge \subseteq}\{Y \in R \ : \ Y \succeq X\}$
Free patterns	f	$\{X \in S \ : \ (\forall Y \prec X)(f(Y) \neq f(X))\}$	$X \mapsto \max_{\preceq \wedge \subseteq}\{Y \in R \ : \ Y \preceq X\}$
approximate condensed representations $X \simeq Y \Leftrightarrow \lvert f(X) - f(Y)\rvert \leq \epsilon$			
δ-Free patterns	freq	$\{X \in S \ : \ (\forall Y \subset X)(\text{freq}(Y, D) > \text{freq}(X, D) + \delta)\}$	$X \mapsto \max_{\subseteq}\{Y \in R \ : \ Y \subseteq X\}$
dominant representations $X \simeq Y \Leftrightarrow f(X) \leq f(Y)$			
Strong patterns	frequency-based measure m_i	$\{X \in S \ : \ (\forall Y \supset X)(\text{freq}(Y, D_i) \neq \text{freq}(X, D_i))\}$	$X \mapsto \min_{\subseteq}\{Y \in R \ : \ Y \supseteq X\}$