

On Modal Logic Association Rule Mining

Mauro Milella¹, Emilio Muñoz-Velasco², Giovanni Pagliarini^{1,3}, Andrea Paradiso¹,
Guido Sciavicco^{1,†} and Ionel Eduard Stan^{1,3,*,†}

¹University of Ferrara, Italy

²University of Málaga, Spain

³University of Parma, Italy

Abstract

The most iconic duality of machine learning models is symbolic learning versus functional learning. While functional learning is based on a numerical approach to knowledge extraction and modelling, the purpose of symbolic machine learning is to extract knowledge from data in such a form that it can be understood, discussed, modified, and applied by humans, as well as serve as the basis of artificial intelligence applications. The typical problems associated with machine learning are classification and rule extraction; while classification can be dealt with using both functional and symbolic learning, rule extraction is essentially symbolic. One element common to nearly all definitions and tools for rule extraction is that they are applied to static datasets and based on propositional logic; unfortunately, very often real-world applications give rise to non-static sets of data (e.g., temporal, spatial, graph-based data) and may require more-than-propositional expressive power. In order to extract association rules from non-static data, in this paper we propose a definition of modal association rules based on modal logic, and we study how a standard rule extraction algorithm such as APRIORI can be generalized to the modal case while keeping the properties of the canonical, non-modal case, namely, correctness and completeness.

Keywords

Modal logic, Association rule mining, Modal symbolic learning

1. Introduction

The most iconic and fundamental separation between the sub-fields of machine learning is the one between *functional* and *symbolic* learning. Functional learning is the process of learning a *function* that represents the theory underlying a certain phenomenon. Symbolic learning, on the other hand, is the process of learning a *logical description* that represents that phenomenon. Whether one or the other approach should be preferred raised a long-standing debate among experts, which roots in the fact that functional methods tend to be more versatile and statistically accurate than symbolic ones, while symbolic methods are able to extract models that can be

Proceedings of the 23rd Italian Conference on Theoretical Computer Science, Rome, Italy, September 7-9, 2022

*Corresponding author.


† These authors contributed equally.

✉ mauro.milella@edu.unife.it (M. Milella); ejmunoz@uma.es (E. Muñoz-Velasco); giovanni.pagliarini@unife.it (G. Pagliarini); andrea.paradiso@edu.unife.it (A. Paradiso); guido.sciavicco@unife.it (G. Sciavicco);
ioneleduard.stan@unife.it (I. E. Stan)

🆔 0000-0001-7128-6745 (M. Milella); 0000-0002-0117-4219 (E. Muñoz-Velasco); 0000-0002-8403-3250 (G. Pagliarini);
0000-0002-3614-2487 (A. Paradiso); 0000-0002-8403-3250 (G. Sciavicco); 0000-0001-9260-102X (I. E. Stan)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

interpreted, explained, and then enhanced using human-expert knowledge. These characteristics of symbolic methods, both for *political* reasons (consider, for example, the recent General Data Protection Regulation (GDPR) [1] of the European Union, that highlights the need for interpretable/explainable automatic learning-based decision-making processes, including those involving AI technologies) and *technical* ones (interpretable models are often easier to train, explore, integrate, and implement), are sometimes used as arguments for preferring a symbolic approach over a functional one. *Classification* and *rule extraction* are, among others, typical problems of machine learning; while classification can be dealt with using both functional and symbolic learning, rule extraction is essentially symbolic.

From a logical standpoint, canonical symbolic learning methods are mostly characterized by being based on propositional logic, and by being designed for static data, in which every instance is described by the value of n attributes. Non-static data, such as dimensional (temporal, spatial, spatio-temporal) data, graph-based data, or textual data, however, cannot be successfully dealt with within the same schema in a native way. The standard approach is to pre-process such data, so that their aspect becomes static again (e.g., by averaging the value of attributes along all dimensions), and that off-the-shelf static symbolic methods can be used. Modal symbolic learning [2, 3, 4] takes a different point of view: modal symbolic methods are characterized by being based on modal logic (e.g., temporal, spatial logic) so that non-static data can be dealt with natively, and that the extracted knowledge takes the form of interpretable modal logic formulas. Modal symbolic learning has been mainly focused on classification; here, we consider the case of association rule extraction. At the propositional level, there are two main approaches to association rule extraction, or, more particularly, to its key step, frequent sets extraction, namely APRIORI [5] and FP-Growth [6]. In this paper, we consider the most basic one, APRIORI, and we generalize it to the modal level. A first approach for extracting interval temporal logic association rules, that can be considered a particular case of this work, has been proposed in [7].

This paper is organized as follows. First, in Section 2, we lay down some basic preliminaries, and we define the concept of modal datasets. Then, in Section 3, we define modal association rules from a logical viewpoint, and in Section 4 we extend APRIORI to extract modal frequent sets and we study its properties, namely, correctness and completeness. In Section 5 we give examples of possible applications of modal association rules, before concluding.

2. Preliminaries

Given a set of *propositional letters* \mathcal{P} , the well-formed formulas of the *propositional modal logic* (\mathcal{ML}) are obtained by the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \diamond\varphi,$$

where the remaining classic Boolean operators can be obtained as shortcuts. In what follows, we use $\Box\varphi$ to denote $\neg\diamond\neg\varphi$, and \equiv to denote the logical equivalence. The *modality* \diamond (resp., \Box) is usually referred to as *it is possible that* (resp., *it is necessary that*). Modal logic is considered as archetypal of temporal, spatial, and spatio-temporal logics, and it is an extension of *propositional logic* (\mathcal{PL}). Its semantics is given in terms of Kripke models. A (finite) *Kripke model* $K =$

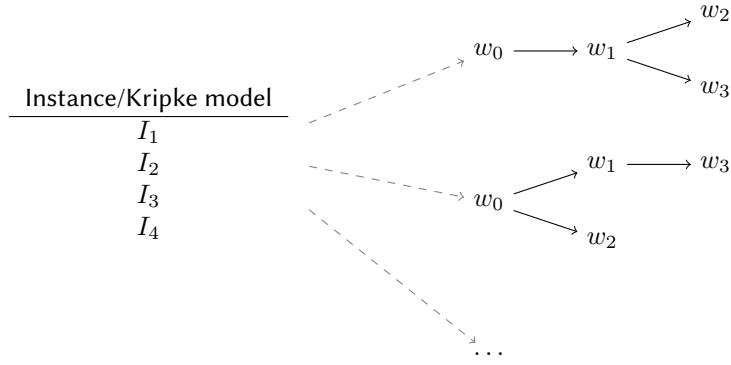


Figure 1: An example of modal dataset with 4 instances, each described by a Kripke model.

(W, R, V) is composed by a finite set of *worlds* W (which contains a distinguished world w_0 , called *initial world*), a binary *accessibility relation* $R \subseteq W \times W$, and a *valuation function* $V : W \rightarrow 2^{\mathcal{P}}$, which associates each world with the set of propositional letters that are true on it. The *truth relation* $K, w \Vdash \varphi$ for a generic model K and a generic world w in that model is given by the following clauses:

$$\begin{aligned}
 K, w \Vdash p & \quad \text{iff } p \in V(w); \\
 K, w \Vdash \neg\varphi & \quad \text{iff } K, w \not\Vdash \varphi; \\
 K, w \Vdash \varphi \wedge \psi & \quad \text{iff } K, w \Vdash \varphi \text{ and } K, w \Vdash \psi; \\
 K, w \Vdash \diamond\varphi & \quad \text{iff } \exists v \text{ s.t. } wRv \text{ and } K, v \Vdash \varphi.
 \end{aligned}$$

Most classic temporal [8, 9, 10] and spatial logics [11, 12] are specializations of the uni-modal logic with more than one (possibly with higher arities) accessibility relations (and associated modalities), subject to constraints that range from very simple and intuitive ones (e.g., transitivity, antisymmetry) to very complex ones (e.g., when worlds are assumed to be intervals in time and modalities are assumed to mimic relations between intervals).

Definition 1 (modal dataset). *Given a set of propositional letters \mathcal{P} , a modal dataset $\mathcal{I} = \{I_1, \dots, I_m\}$ is a finite collection of m instances, each of which is a finite Kripke model $I_i = (W_i, R_i, V_i)$, for each $i = 1, \dots, m$.*

The above definition generalizes the conventional notion of *propositional*, or *static*, dataset, in which each of the m instances is associated to the value of n distinct numerical attributes. An example of modal dataset is illustrated in Fig. 1. In machine learning, several problems are associated to a dataset \mathcal{I} . The most common one is the *classification/regression* problem; here, however, we are interested in the equally common learning problem known as *association rule extraction*: in a generic dataset, it is the problem of extracting meaningful associations among the values of the attributes.

In the following, we first define the concept of meaningful association rule, and then we focus our attention to the problem of extracting (all) meaningful rules from a modal dataset.

3. Modal Logic Association Rules

Rule-based methods are ubiquitous in modern machine learning and data mining applications [13], that aim at learning regularities in data which can be expressed in the form of *if-then* rules. In this paper, we are interested in *descriptive* (or *association*) rules, that collectively cover the instances' space.

Let \mathcal{P} be an alphabet on \mathcal{I} . In its simplest form, *propositional* association rules are objects of the type [14]:

$$\rho : p_1 \wedge \dots \wedge p_k \Rightarrow p_{k+1}, \quad (1)$$

where $p_1, \dots, p_k, p_{k+1} \in \mathcal{P}$ are *positive propositional literals* (or simply *propositional literals*); recalling the standard nomenclature in the association rule mining literature, propositional literals are also known as *items* [5]. We use the notation:

$$\rho : X \Rightarrow Y, \quad (2)$$

where $X = \{p_1, \dots, p_k\}$ is called *antecedent*, $Y = \{p_{k+1}\}$ is called *consequent*, and $X \cap Y = \emptyset$. Association analysis must be interpreted with caution: causality is not implied in the inference process by an association rule as it simply represents strong co-occurrence relationship of the literals in the antecedent with those in the consequent [15]. Therefore, to emphasize the difference with respect to \rightarrow , which is a logical implication, in (association) rules we use the symbol \Rightarrow . In this sense, ρ is *not* a formula of propositional logic. When treated like one, however, it becomes a formula that belongs to the known subset of propositional logic, that is, Horn propositional logic; this usually happens after the extraction phase, when rules are no longer at risk of not being meaningful. Horn propositional logic is interesting from the deductive point of view (its satisfiability problem is P -complete), which becomes relevant when data-driven knowledge is mixed, in some sense, with expert, top-down knowledge in intelligent applications. Non-empty subsets of \mathcal{P} are also called *propositional patterns* [5]. Moreover, the *length* of a (propositional) pattern X is the set-theoretic cardinality of X . In this paper we focus only on Horn-like association rules; however, the fact that, in other settings, the consequent Y of a rule $\rho : X \Rightarrow Y$ could be a set with more than one literal, that is, it could have k' literals $Y = \{p_{k+1}, \dots, p_{k+k'}\}$ which, if interpreted as a disjunction, would result in being a non-Horn logic rule.

In order to formalize the modal association rule mining problem, where instances of a modal dataset are generic, finite Kripke models, we must first lift the concept of literals to the modal case.

Definition 2 (modal literals). *For a fixed alphabet \mathcal{P} , the set of positive modal literals (or simply modal literals) $\Lambda_{\mathcal{P}}$ over \mathcal{P} is defined by the grammar:*

$$\lambda ::= p \mid \diamond\lambda \mid \square\lambda,$$

where $p \in \mathcal{P}$.

Modal literals of the type p are called *propositional*, as in the propositional case, and modal literals of the type $\diamond\lambda$ (resp., $\square\lambda$) are called *existential* (resp., *universal*).

Definition 3 (modal association rules). For a fixed set of modal literals $\Lambda_{\mathcal{P}}$, modal association rules are objects of the type:

$$\rho : \lambda_1 \wedge \dots \wedge \lambda_k \Rightarrow \lambda_{k+1}, \quad (3)$$

where $\lambda_1, \dots, \lambda_k, \lambda_{k+1} \in \Lambda_{\mathcal{P}}$.

Note that (3) has the same structure as (1), except that \mathcal{P} has been replaced by $\Lambda_{\mathcal{P}}$. Non-empty subsets of $\Lambda_{\mathcal{P}}$ are called *modal patterns*, and we sometimes write a rule ρ in its set-theoretical notation, that is, $X \Rightarrow Y$, where both X and Y are modal patterns and Y is a singleton. Similarly to the propositional case, modal rules are not modal formulas; when treated like that, they belong to the Horn fragment of modal logic $\mathcal{ML}^{\text{Horn}}$. There are several, although very similar, proposals for a Horn fragment of modal logic. Bresolin et al. [16] studied $\mathcal{ML}^{\text{Horn}}$ and several of its sub-fragments from the deductive point of view; while the satisfiability problem for $\mathcal{ML}^{\text{Horn}}$ is still *PSPACE*-complete (as it is for full \mathcal{ML}), sub-fragments of it such as the one with only universal and propositional literals becomes *P*-complete.

Not all rules are interesting or meaningful. As for the propositional case, in order to capture the notion of meaningfulness in logical terms, in [5] authors implicitly defined a notion of *ρ holding* on an instance, and then on a dataset. They did so by introducing two parameters, namely support and confidence, which they used to modify the notion of truth of literal, first, and of a rule, then. After the first introduction of such concepts, and specifically of the concept of confidence, alternative measures of meaningfulness have been proposed, among which we mention lift, and conviction (see, e.g., [15]). In the end, as we shall see, the key observation is that their definitions are all consequences of the definition of support; our aim is to generalize the latter to the modal case, so that all other measures can then be re-defined accordingly. If the notion of propositional support was originally designed to capture the idea of how frequent was a literal (resp., a set of literals) across the static dataset, we now have to capture the additional idea of how frequent a modal literal (resp., a set of modal literals) is across a single instance, as well. Such two notions are only partially orthogonal: locally, we want to capture the number of different worlds in which a certain phenomenon happens, in order to measure its interestingness, while, globally, we want to capture the number of different instances in which such an interesting phenomenon occurs. In the following, if X is a set of literals, $I = (W, R, V)$ an instance, and $w \in W$ a world in I , we use $I, w \Vdash X$ as a shortcut for $I, w \Vdash \bigwedge_{\lambda \in X} \lambda$.

Definition 4 (support). Let \mathcal{I} be a modal dataset and $X \subseteq \Lambda_{\mathcal{P}}$ be a set of literals. The local support of X on some instance $I \in \mathcal{I}$, being $I = (W, R, V)$, is defined as:

$$lsupp(I, X) = \frac{|\{w \in W \mid I, w \Vdash X\}|}{|W|},$$

and, given a certain local support $s_l \in (0, 1] \subset \mathbb{R}$, the global support of X on \mathcal{I} relatively to s_l is defined as:

$$gsupp_{s_l}(\mathcal{I}, X) = \frac{|\{I \in \mathcal{I} \mid lsupp(I, X) \geq s_l\}|}{|\mathcal{I}|}.$$

Similarly to the propositional case, we can define popular interestingness measures for modal rules in such a way to be (local and global) support-dependent.

Definition 5 (confidence, lift). Let \mathcal{I} be a modal dataset, and $\rho : X \Rightarrow Y$ a modal rule. The local confidence of ρ on some instance $I \in \mathcal{I}$ is defined as:

$$lconf(I, X \Rightarrow Y) = \frac{lsupp(I, X \cup Y)}{lsupp(I, X)},$$

and, given a certain local support $s_l \in (0, 1] \subset \mathbb{R}$, the global confidence of ρ on \mathcal{I} relatively to s_l is defined as:

$$gconf_{s_l}(\mathcal{I}, X \Rightarrow Y) = \frac{gsupp_{s_l}(\mathcal{I}, X \cup Y)}{gsupp_{s_l}(\mathcal{I}, X)}.$$

Similarly, the local lift of ρ on some instance $I \in \mathcal{I}$ is defined as:

$$llift(I, X \Rightarrow Y) = \frac{lsupp(I, X \cup Y)}{lsupp(I, X) \cdot lsupp(I, Y)},$$

and, given a certain local support $s_l \in (0, 1] \subset \mathbb{R}$, the global lift of ρ on \mathcal{I} relatively to s_l is defined as:

$$glift_{s_l}(\mathcal{I}, X \Rightarrow Y) = \frac{gsupp_{s_l}(\mathcal{I}, X \cup Y)}{gsupp_{s_l}(\mathcal{I}, X) \cdot gsupp_{s_l}(\mathcal{I}, Y)}.$$

Confidence measures how frequently items in Y appear in instances (in the global setting) or worlds (in the local setting) that contain X . Similarly, lift measures the dependence between X and Y . From the above examples, we can derive a general notion of support-dependent interestingness measure; we say that a measure is *local support-dependent* (resp., *global support-dependent*) if and only if it can be computed by a function of the local support (resp., global support) of its set of worlds (resp., instances). We denote by *lmeas* (resp., *gmeas_{s_l}*) a generic local (resp., global) support-dependent measure.

Definition 6 (holding). Let \mathcal{I} be a modal dataset, $\rho : X \Rightarrow Y$ a modal rule, θ_l a $(z + 1)$ -vector of values in $(0, 1] \subset \mathbb{R}$ (indexed from 0 to z), and let $lmeas^1, \dots, lmeas^z$ be z local support-dependent measures. We say that ρ θ_l -holds on $I \in \mathcal{I}$ if and only if: (i) $lsupp(I, X \cup Y) \geq \theta_l[0]$, and (ii) for each $1 \leq i \leq z$, $lmeas^i(I, \rho) \geq \theta_l[i]$. We denote such a situation by:

$$I \Vdash_{\theta_l} \rho.$$

Similarly, given $s_l \in (0, 1] \subset \mathbb{R}$, if $\theta_g^{s_l}$ is a $(z + 1)$ -vector of values in $(0, 1] \subset \mathbb{R}$ (indexed from 0 to z), and $gmeas^1, \dots, gmeas^z$ are z global support-dependent measures, we say that ρ $\theta_g^{s_l}$ -holds on \mathcal{I} if and only if: (i) $gsupp_{s_l}(\mathcal{I}, X \cup Y) \geq \theta_g^{s_l}[0]$, and (ii) for each $1 \leq i \leq z$, $gmeas_{s_l}^i(\mathcal{I}, \rho) \geq \theta_g^{s_l}[i]$. We denote such a situation by:

$$\mathcal{I} \Vdash_{\theta_g^{s_l}} \rho.$$

As defined in the literature, the association rules mining problem has two parts: frequent patterns mining and extraction of interesting rules. Therefore, in both settings, local and global, the first conditions, $lsupp(I, X \cup Y) \geq \theta_l[0]$ and $gsupp_{s_l}(\mathcal{I}, X \cup Y) \geq \theta_g^{s_l}[0]$, enforces a rule to be frequent in terms of their supports, locally and globally, respectively. The remaining conditions can be arbitrarily combined but they cannot be separated from imposing a minimal

support. Observe that, the notion of a rule holding on a single instance collapses to the standard notion of truth when instances are propositional (that is, with a single world); similarly, in the same case, the notion of a rule holding on a dataset collapses to the (implicit) notion of propositional rule holding on a dataset (driven by the classical support and confidence/lift/...). The definition of such generalized notions is not unique, as different constraints could be enforced. However, not every combination makes sense. For example, forcing a rule to be globally supported if and only if it is locally supported on every instance would imply requiring perfect global support, which is rather restrictive and possibly non-adherent to the intended use of rules.

In the next section we shall focus on solving the following natural problem: given a modal dataset \mathcal{I} and a vector $\theta_g^{s_l}$ of $(z + 1)$ values in $(0, 1] \subset \mathbb{R}$ for some $s_l \in (0, 1] \subset \mathbb{R}$, extract (or mine) all and only modal rules that $\theta_g^{s_l}$ -hold on \mathcal{I} , which we regard to as *interesting*.

4. Mining Modal Logic Rules

As a consequence of focusing on support-dependent interestingness measures only, it turns out that extracting interesting rules can be achieved by simply extracting sets of literals with a relatively high support, which in the classic terminology are called *frequent* sets. To see this, consider a rule $\rho : X \Rightarrow Y$, and assume that we want to check if it $[s_g, c_g]_g^{s_l}$ -holds on a certain dataset \mathcal{I} , where c_g is the minimum requested global confidence. To this end, it must hold that $gsupp_{s_l}(\mathcal{I}, X \cup Y) \geq s_g$ and that $gconf_{s_l}(\mathcal{I}, \rho) \geq c_g$. To compute the latter, one has to compute, first, $gsupp_{s_l}(\mathcal{I}, X \cup Y)$ and $gsupp_{s_l}(\mathcal{I}, X)$, which are both by-products of computing $gsupp_{s_l}(\mathcal{I}, X \cup Y)$. Therefore, being able to extract *all* frequent sets of literals is necessary and sufficient to be able to extract *all* interesting rules as we have defined them. This phenomenon occurred at the propositional level too, obviously; it is interesting to observe that it still holds at the modal one by inheriting exactly the same procedure (see again, e.g., [5]).

The problem of enumerating all sets of frequent literals has been studied in the literature [17]; in general, enumerating frequent sets has several applications, not necessarily related to rule extraction. The first key observation is that so-called, and well-known *APRIORI principle*.

Proposition 1 (APRIORI principle [5]). *If a set of literals is frequent, then all of its subsets must be frequent.*

Naturally, there are a potentially exponential number of frequent sets of a given dataset \mathcal{I} , even at the propositional level. Focusing on *maximal* frequent sets of literals does not change this, and the complexity of the problem of extracting them becomes therefore interesting. As proved in [17], the counting problem associated to the problem of extracting all maximal frequent sets is already $\#P$ -hard [18]. In most cases, the interesting problems in machine learning are computationally hard, thus the go-to strategy is turning to approximation algorithms. This is not the case with frequent set extraction: in many practical cases, indeed, there are only *a few* maximal frequent sets in a given dataset, and enumerating them in a correct and complete fashion is the typical approach. There are two classical approaches to this problem, called, respectively, APRIORI [5] and FP-Growth [6]. In the rest of this section, we consider the former one, and we study its modal generalization.

ALGORITHM 1: Modal extension of the APRIORI algorithm for mining frequent modal patterns.

```
1 function ModalAPRIORI( $\mathcal{I}, \mathcal{P}, s_g, s_l, \delta$ ):
2    $\Lambda_{\mathcal{P}} \leftarrow \text{Literals}(\mathcal{P}, \delta)$ 
3    $C_1 \leftarrow \{\{\lambda\} \mid \lambda \in \Lambda_{\mathcal{P}}\}$ 
4    $F_1 \leftarrow \{X \in C_1 \mid \text{gsupp}_{s_l}(\mathcal{I}, X) \geq s_g\}$ 
5    $k \leftarrow 1$ 
6   repeat
7      $k \leftarrow k + 1$ 
8      $C_k \leftarrow \text{CandidateGeneration}(F_{k-1})$ 
9      $C_k \leftarrow \text{CandidatePruning}(C_k, F_{k-1})$ 
10     $F_k \leftarrow \{X \in C_k \mid \text{gsupp}_{s_l}(\mathcal{I}, X) \geq s_g\}$ 
11  until  $F_k = \emptyset$ 
12  return  $\bigcup_{i=1}^k F_i$ 
13 end

14 function CandidateGeneration( $F_{k-1}$ ):
15   $C_k \leftarrow \{\{\lambda_1, \dots, \lambda_{k-2}, \lambda_{k-1}, \lambda'_{k-1}\} \mid \{\lambda_1, \dots, \lambda_{k-2}, \lambda_{k-1}\}, \{\lambda_1, \dots, \lambda_{k-2}, \lambda'_{k-1}\} \in$ 
16     $F_{k-1} \text{ and } \lambda_{k-1} < \lambda'_{k-1}\}$ 
17  return  $C_k$ 
18 end

18 function CandidatePruning( $C_k, F_{k-1}$ ):
19  foreach  $X \in C_k$  do
20    forall  $(k-1)$ -length subsets  $S$  of  $X$  do
21      if  $S \notin F_{k-1}$  then  $C_k \leftarrow C_k \setminus X$ 
22    end
23  end
24  return  $C_k$ 
25 end
```

APRIORI has been the first pruning-based algorithm for mining sets of propositional literals. Borrowing the same concept from the propositional level, the global support models the (relative) frequency of instances in \mathcal{I} that satisfy the modal pattern X , whereas the local support, in the more general modal setting that we present here, models the frequency of worlds in $I \in \mathcal{I}$ that satisfy the modal pattern X . Observe that the anti-monotonic property (the APRIORI principle) also holds for the local support, as longer modal patterns hold on fewer worlds, which is a consequence of the logical conjunction in the definition of the truth $I, w \Vdash X$. The modal extension of APRIORI is based on the following assumption, which also holds in its propositional version of APRIORI: modal literals in modal patterns can be lexicographically ordered. The reason behind such an assumption, that works in synergy with the APRIORI principle, is to avoid useless computation during the learning routine. Algorithm 1 generalizes the APRIORI algorithm to the modal case and takes as input a modal dataset \mathcal{I} , an alphabet \mathcal{P} defined on \mathcal{I} , three user-specified thresholds $s_l, s_g \in (0, 1] \subset \mathbb{R}$, and $\delta \in \mathbb{N}$ for the minimum local support, minimum global support, and maximum modal depth of modal literals, respectively. Setting a maximum modal depth is a necessary step for the algorithm to work and terminate, as there is no anti-monotonic property that holds for modal literals; it should be noticed, however, that modal rules with too long modal literals would disfavour their interpretability. In Algorithm 1, F_k and C_k denote the set of frequent modal patterns and the set of generated candidates of

modal patterns of length k , for $k \geq 1 \in \mathbb{N}$, respectively. The algorithm starts by computing the set of modal literals $\Lambda_{\mathcal{P}}$ from \mathcal{P} and δ . Then, it computes F_1 whose patterns are frequent from C_1 . Next, the repeat-until loop block of the algorithm computes iteratively F_k , with $k > 1$. The *CandidateGeneration*(\cdot) procedure generates candidate modal patterns C_k by joining¹ F_{k-1} with itself: two lexicographically ordered frequent $(k-1)$ -length modal patterns $\{\lambda_1, \dots, \lambda_{k-1}, \lambda_{k-1}\}$ and $\{\lambda_1, \dots, \lambda_{k-2}, \lambda'_{k-1}\}$ of F_{k-1} are merged together if they have the same modal patterns in the first $k-2$ positions and if $\lambda_{k-1} < \lambda'_{k-1}$ to form a candidate k -length modal pattern $\{\lambda_1, \dots, \lambda_{k-2}, \lambda_{k-1}, \lambda'_{k-1}\}$. The candidates from C_k are then *pruned* by the *CandidatePruning*(\cdot) function based on the APRIORI principle. It is important to stress the fact that, both *CandidateGeneration*(\cdot) and *CandidatePruning*(\cdot) are identical as in the canonical APRIORI algorithm. At this stage, the algorithm produces F_k whose elements are from C_k which are frequent. Termination is guaranteed by the anti-monotonic property of the support.

We now study correctness and completeness of our approach. It is important to stress the fact that completeness is parametric in δ .

Theorem 1 (correctness). *ModalAPRIORI is correct.*

Proof. We want to show that F_k , for each $k \geq 1$, is the set of frequent k -length modal patterns. The proof is by induction on k . If $k = 1$, then F_1 contains singleton sets of modal literals of the type $\{\lambda\}$, where $\lambda \in \Lambda_{\mathcal{P}}$, that are frequent. Thus, by construction, F_1 is made of frequent 1-length modal patterns. If $k > 1$, to obtain the set of frequent sets of modal literals F_k , the algorithm generates candidates by combining elements of F_{k-1} , which are frequent by inductive hypothesis. Let $\{\lambda_1, \dots, \lambda_{k-2}, \lambda_{k-1}\}$ and $\{\lambda_1, \dots, \lambda_{k-2}, \lambda'_{k-1}\}$ be elements of F_{k-1} used by the candidate generation procedure to generate $X \in F_k$, such that $X = \{\lambda_1, \dots, \lambda_{k-2}, \lambda_{k-1}, \lambda'_{k-1}\}$. If X is frequent, that is, meets the support, then $X \in F_k$. Therefore, F_k is made of k -length frequent patterns. \square

Theorem 2 (completeness). *ModalAPRIORI is complete.*

Proof. Let F_k be the set of frequent sets of modal literals and let C_k be the set of candidate frequent sets of modal literals, whose modal patterns have length k . We want to show that $F_k \subseteq C_k$, for all $k \geq 1$. The proof is by induction on k . If $k = 1$, then C_1 is the set of singleton sets of the type $\{\lambda\}$, and $F_1 \subseteq C_1$ holds since F_1 is defined as the subset of C_1 that are frequent. If $k > 1$, assume, by inductive hypothesis, that $F_{k-1} \subseteq C_{k-1}$. We must show that $F_k \subseteq C_k$. Let $X = \{\lambda_1, \dots, \lambda_{k-2}, \lambda_{k-1}, \lambda'_{k-1}\}$. The pattern X is generated by two $(k-1)$ -length modal patterns, say $\{\lambda_1, \dots, \lambda_{k-2}, \lambda_{k-1}\}$ and $\{\lambda_1, \dots, \lambda_{k-2}, \lambda'_{k-1}\}$, such that $\lambda_{k-1} < \lambda'_{k-1}$, which are necessarily frequent, by inductive hypothesis. Thus, $X \in C_k$ must hold and if X meets the support, then $X \in F_k$, obtaining the thesis. \square

5. Applications

Modal logics are, as we have observed, archetypal of many temporal, spatial, and spatio-temporal logics. The purpose of modal symbolic learning is designing symbolic machine learning

¹This is the original nomenclature.

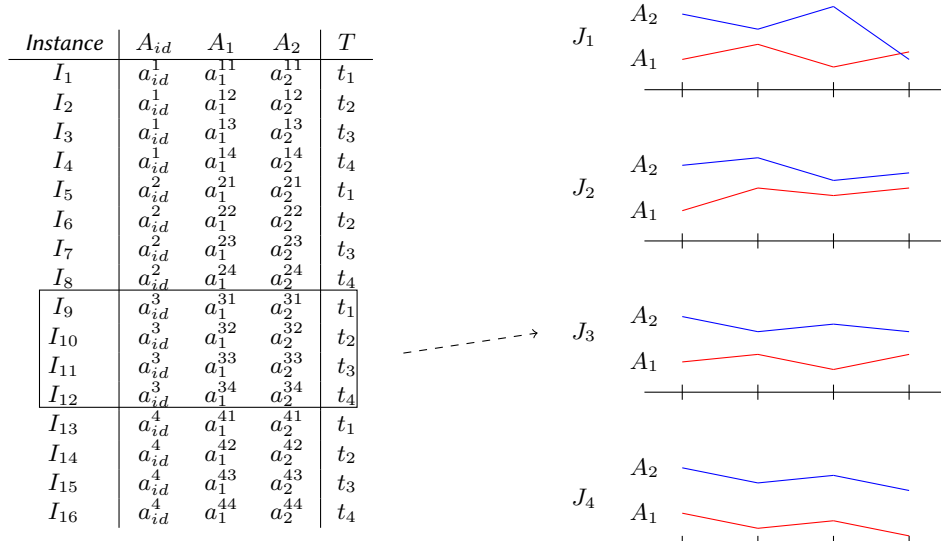


Figure 2: An example of dimensional dataset.

techniques to deal with non-static data. Non-static data include, among others, *dimensional* data (that is, data with temporal and/or spatial components), *graph-based* data (that is, data natively in the form of graphs, such as social networks, or gene expression networks), and *textual* data (that is, data in the form of natural language text). The most basic advantage of modal symbolic learning is that of reuniting all such non-static data under the same viewpoint, that is, seeing them all as *modal* data. How this is done in all cases is outside the scope of this paper; by way of example we take here the dimensional, temporal case to show how the modal framework emerges naturally.

Dimensional datasets can be regarded as modal datasets in which the Kripke model associated to an instance is implicit and presented in tabular form; this corresponds to the typical presentation of dimensional data, as in Fig. 2, left-hand side. While in the classical sense instances correspond to lines, more than one line make up for a *real-world* instance. So, for example, each instance of the left-hand side of the picture may be the status of a patient under observation, each being described by two attributes A_1 and A_2 , over the course of four time points t_1 through t_4 . Thus, the intended meaning of such a dataset is that we have in fact four patients (being distinguished by their names a_{id}^1 through a_{id}^4), which become four *multivariate time series*, denoted by J_1 through J_4 , each composed by four values per attribute, as in Fig. 2, right-hand side. By regarding each interval in the temporal domain t_1, \dots, t_4 as a world, one ends up building a Kripke structure from each time series (in this example, each structure has 6 worlds, that is, 6 non-instantaneous intervals). Connecting each pair of such intervals there is exactly one of 12 Allen's relations [19]; the modal logic that emerges from this preprocessing, and that allows one to express how the properties of a time series relate to each other through time is known as *interval temporal logic of Allen's relations*, or \mathcal{HS} (from Halpern and Shoham, who first introduced it [9]).

Observe, now, that each interval in this setting accounts for a set of values for each attribute; for example, in Fig 2, the instance J_1 has values $a_1^{11}, a_1^{12}, a_1^{13}$ for A_1 in the interval $[t_1, t_3]$. By applying suitable *feature extraction function(s)* one can, at this point, aggregate such set of values in a single one; typical aggregations include the average, the maximum, and the minimum, but in the literature of learning from temporal data one finds several more, and less trivial, examples. Let us call f any such aggregation function; by applying f to every possible interval for the attribute A , one ends up with a new set of values, that is, the domain $dom(f(A))$. At this point, it is possible to extract a propositional alphabet, of the type:

$$a_1 \leq f(A) \leq a_2,$$

with $a_1, a_2 \in dom(f(A))$. By labeling each world (interval) with the truth value of each of the propositional letters, we extract from a dimensional, temporal dataset a modal one.

The explicit knowledge that can be extracted from electroencephalogram recording (EEG) of subjects that are in certain conditions or are carrying on certain activities is of uttermost importance in the scientific enquiry of those conditions and activities. The current approaches to this problem are entirely functional, but, despite the terminological affinity, black-box classification models such as neural networks cannot be considered models of the human mind. EEG is a real-world example of dimensional data that could be used for modal association rule extraction. A typical EEG is performed with a head-set that accommodates several electrodes (typically, in the fifties), each one of which records, during a single experiment, the voltage that is expressed by a specific area of the human brain; the signal of each electrode is then preprocessed through Fourier-like transformations, so that each signal gives rise to several components at different frequency intervals (typically, in the tens). All in all, a single EEG recording generates anywhere between 500 and 2000 temporal attributes (pairs electrode-frequency) over the course of the experiment. A recent trend in neurobiology is focusing on how different experimental situations can be distinguished by the form and the values of the corresponding EEG recording, which helps neurobiologists to understand the nature of the different brain areas in different settings. While most often these problems are designed as classification problems, understanding how the several attributes relate to each other through time in a specific situation may turn out to be very interesting. Temporal relationships can be extracted as modal rules, specifically \mathcal{HS} rules, which would have the form of assertions of the type: *if the value of the voltage in this sensor at this frequency is within this interval of values during an interval of time, then the value of the voltage in this other sensor at this other frequency must be within this other interval immediately after it.*

6. Conclusions

Symbolic learning is the sub-field of machine learning that deals with symbolic algorithms and models, which have been known for decades and successfully applied to a variety of contexts. The main limitation of current symbolic models is the fact that they are essentially based on classical propositional logic, which implies that dimensional data, graph-based data, or textual data must be pre-processed in order to be dealt with within the classic framework, at the expenses of decreasing the interpretability of the result and losing information.

Modal symbolic learning is the modal logic generalization of symbolic learning; in this paper we considered the problem of defining the concept of modal association rules and their logical characterization, and we devised a correct and complete algorithm, *ModalAPRIORI*, which is able to extract frequent sets of modal literals both at the global and local level, and therefore modal symbolic rules. We also discussed potential applications of modal symbolic rules.

References

- [1] Parliament and Council of the European Union, General Data Protection Regulation, 2016. URL: <https://gdpr-info.eu/>.
- [2] F. Manzella, G. Pagliarini, G. Sciavicco, I. E. Stan, Interval Temporal Random Forests with an Application to COVID-19 Diagnosis, in: Proceedings of the 28th International Symposium on Temporal Representation and Reasoning (TIME), volume 206 of *Leibniz International Proceedings in Informatics*, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021, pp. 7:1–7:18.
- [3] G. Pagliarini, G. Sciavicco, Decision Tree Learning with Spatial Modal Logics, in: Proceedings of the 12th International Symposium on Games, Automata, Logics, and Formal Verification (GandALF), volume 346 of *Electronic Proceedings in Theoretical Computer Science*, 2021, pp. 273–290.
- [4] G. Sciavicco, I. E. Stan, Knowledge extraction with interval temporal logic decision trees, in: Proceedings of the 27th International Symposium on Temporal Representation and Reasoning (TIME), volume 178 of *Leibniz International Proceedings in Informatics*, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020, pp. 9:1–9:16.
- [5] R. Agrawal, R. Srikant, Fast Algorithms for Mining Association Rules in Large Databases, in: Proceedings of 20th International Conference on Very Large Data Bases (VLDB), 1994, pp. 487–499.
- [6] J. Han, J. Pei, Y. Yin, Mining Frequent Patterns without Candidate Generation, in: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, 2000, pp. 1–12.
- [7] D. Bresolin, E. Cominato, S. Gnani, E. Muñoz-Velasco, G. Sciavicco, Extracting Interval Temporal Logic Rules: A First Approach, in: Proceedings of the 25th International Symposium on Temporal Representation and Reasoning (TIME), volume 120 of *Leibniz International Proceedings in Informatics*, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018, pp. 7:1–7:15.
- [8] E. M. Clarke, E. A. Emerson, Design and Synthesis of Synchronization Skeletons Using Branching-Time Temporal Logic, in: Proceedings of the Workshop on Logics of Programs (LOP), volume 131 of *Lecture Notes in Computer Science*, Springer, 1981, pp. 52–71.
- [9] J. Y. Halpern, Y. Shoham, A Propositional Modal Logic of Time Intervals, *Journal of the ACM* 38 (1991) 935–962.
- [10] A. Pnueli, The Temporal Logic of Programs, in: Proceedings of the 18th Annual Symposium on Foundations of Computer Science (FOCS), IEEE Computer Society, 1977, p. 46–57.
- [11] M. Aiello, J. van Benthem, A Modal Walk Through Space, *Journal of Applied Non-Classical Logics* 12 (2002) 319–364.

- [12] C. Lutz, F. Wolter, Modal Logics of Topological Relations, *Logical Methods in Computer Science* 2 (2006) 1–41.
- [13] J. Fürnkranz, D. Gamberger, N. Lavrac, *Foundations of Rule Learning*, Cognitive Technologies, Springer, 2012.
- [14] A. Horn, On Sentences Which are True of Direct Unions of Algebras, *Journal of Symbolic Logic* 16 (1951) 14–21.
- [15] P. N. Tan, M. S. Steinbach, A. Karpatne, V. Kumar, *Introduction to Data Mining*, 2nd ed., Pearson, 2019.
- [16] D. Bresolin, E. Muñoz-Velasco, G. Sciavicco, On Sub-Propositional Fragments of Modal Logic, *Logical Methods in Computer Science* 14 (2018).
- [17] G. Yang, The complexity of mining maximal frequent itemsets and maximal frequent patterns, in: *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2004, pp. 344–353.
- [18] L. G. Valiant, The Complexity of Enumeration and Reliability Problems, *SIAM Journal on Computing* 8 (1979) 410–421.
- [19] J. F. Allen, Maintaining Knowledge about Temporal Intervals, *Communication of the ACM* 26 (1983) 832–843.