ModalFP-Growth: Efficient Extraction of Modal Association Rules from Non-Tabular Data

Mauro Milella[1,*], Giovanni Pagliarini[1], Guido Sciavicco[1] and Ionel Eduard Stan[1,2,*]

[1]Department of Mathematics and Computer Science, University of Ferrara, Italy
[2]Faculty of Engineering, Free University of Bozen-Bolzano, Italy

Abstract

This paper explores the extraction of modal association rules from non-tabular data using a novel algorithm, ModalFP-Growth. By extending the FP-Growth algorithm to modal logic, ModalFP-Growth processes instances represented as Kripke models, facilitating efficient rule extraction from temporal, spatial, and spatio-temporal datasets. Each instance is transformed into a tabular form where worlds correspond to rows and literals to columns, enabling the application of the original FP-Growth. The algorithm, then, aggregates locally frequent itemsets from individual instances to identify globally supported itemsets across the dataset. We prove the soundness and completeness of ModalFP-Growth, ensuring that all and only frequent itemsets are included in the final output. Additionally, we present an open-source implementation within the Sole learning and reasoning suite. Experimental evaluations using Halpern and Shoham's Interval Temporal Logic on a public temporal dataset demonstrate the algorithm's practical efficiency and the interpretability of the extracted rules.

Keywords

Modal logic, Association rule mining, Modal symbolic learning

1. Introduction

The distinction between *sub-symbolic* and *symbolic* learning is a fundamental separation in machine learning. Sub-symbolic learning involves learning a *function* to represent a phenomenon, offering versatility and statistical accuracy. Conversely, symbolic learning creates a *logical description* of the phenomenon, valued for its interpretability and explainability. This interpretability is crucial for both *political* reasons, such as compliance with the EU's General Data Protection Regulation (GDPR) [1],that highlights the need for interpretable/explainable automatic learning-based decision-making processes [2, 3], and *technical* reasons, as symbolic models are easier to train, explore, integrate, and implement. In machine learning, *classification* and *rule extraction* are common tasks. While classification can utilize both learning types, rule extraction is inherently symbolic. Traditional symbolic methods are based on propositional logic and designed for tabular data, with propositions typically expressed as $A \bowtie a$ or $a \bowtie A \bowtie b$ ($\bowtie \in \{<, \leq\}$) and rules formulated as $p_1 \wedge \ldots \wedge p_k \Rightarrow p_{k+1}$, where $\Rightarrow$ denotes a strong co-occurrence relationship between the antecedent and the consequent.

For temporal or spatial data, which are non-tabular, a pre-processing step is usually required to transform the data into a tabular format. However, recent research suggests that native learning methods may yield better results. *Modal symbolic learning* [4, 5, 6] uses modal logic to process non-tabular data directly, resulting in interpretable modal logic formulas.

Modal symbolic learning has primarily focused on classification, but the concept of *modal* association rules has been formalized in [7, 8], introducing the *ModalApriori* algorithm based on the well-known standard algorithm Apriori. Modal association rules of the type $\lambda_1 \wedge \ldots \wedge \lambda_k \Rightarrow \lambda_{k+1}$ where λ_i are positive *modal literals*, generalize propositional rules, and can easily cover the cases of both spatial and temporal data, among many others. Allowing rules to work in dimensions such as space and time is a

ICTCS'24: Italian Conference on Theoretical Computer Science, September 11–13, 2024, Torino, Italy

[†]These authors contributed equally.

mauro.milella@edu.unife.it (M. Milella); giovanni.pagliarini@unife.it (G. Pagliarini); guido.sciavicco@unife.it (G. Sciavicco); ioneleduard.stan@unibz.it (I. E. Stan)
0000-0001-7128-6745 (M. Milella); 0000-0002-8403-3250 (G. Pagliarini); 0000-0002-9221-879X (G. Sciavicco); 0000-0001-9260-102X (I. E. Stan)

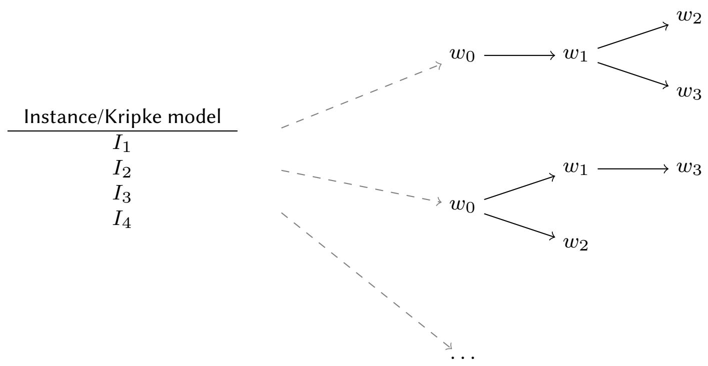


Figure 1: An example of modal dataset with 4 instances, each described by a Kripke model.

natural extension that has already been proposed in prototypical form, for example for dealing with images and text (see, e.g. [9, 10]).

FP-Growth [11] is a more efficient alternative to Apriori for generating frequent patterns from tabular data. This paper extends FP-Growth to the modal case, introducing the *ModalFP-Growth* algorithm and proving its soundness and completeness. We provide an open-source implementation of both ModalApriori and ModalFP-Growth within the Sole learning and reasoning suite [12]. Finally, we propose a customizable algorithm for probing rules from frequent patterns, and we test our approach in the particular case of temporal rule extraction from a public dataset, discussing the results regarding practical efficiency and the meaningfulness of the extracted rules.

2. Modal Logic, Frequent Patterns, and Association Rules

Given a set of *propositions* $\mathcal{P}$, the well-formed formulas of *propositional modal logic* are constructed using the following grammar:

$$\varphi \quad ::= \quad p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Diamond\varphi,$$

where the remaining classic Boolean operators can be derived as shortcuts. In this context, we use $\Box\varphi$ to denote $\neg\Diamond\neg\varphi$, and $\equiv$ to indicate logical equivalence. The *modality* $\Diamond$ (resp., $\Box$) is typically interpreted as *it is possible that* (resp., *it is necessary that*). Modal logic is considered archetypical of propositional temporal, spatial, and spatio-temporal logics and is a non-conservative extension of *propositional logic*.

Formulas are interpreted on Kripke models. A *Kripke model* $K = (W, R, v)$ consists of a finite set of *worlds* W (including a distinguished world w_0, called the *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 propositions true in that world. The pair (W, R) is known as the *frame*. The *truth* relation $K, w \Vdash \varphi$ for a model and a world within that model is defined by the following clauses:

$$\begin{array}{lll}
K, w \Vdash p & \text{iff} & p \in v(w); \\
K, w \Vdash \neg\varphi & \text{iff} & K, w \not\Vdash \varphi; \\
K, w \Vdash \varphi \wedge \psi & \text{iff} & K, w \Vdash \varphi \text{ and } K, w \Vdash \psi; \\
K, w \Vdash \Diamond\varphi & \text{iff} & \exists w' \text{ s.t. } wRw' \text{ and } K, w' \Vdash \varphi.
\end{array}$$

Note that the truth relation is inherently internal, as formulas are evaluated within models at specific worlds, and local, since a world is compared only with those accessible through the relation R.

We use modal logic to express properties of non-tabular data. A non-tabular instance can be seen as a finite Kripke model, and a set of non-tabular instances as a modal dataset.

Definition 1 (modal dataset)**.** *Given a set of propositions $\mathcal{P}$, a* modal dataset $\mathcal{I} = \{I_1, \ldots, I_m\}$ *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 $i = 1, \ldots, m$.*

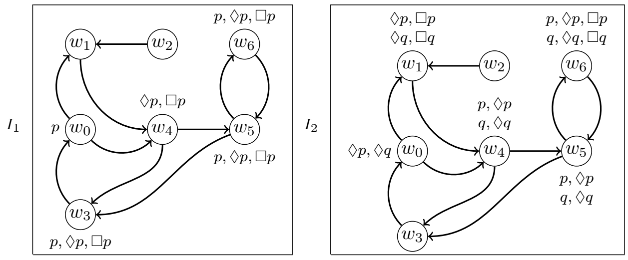


Figure 2: Example of modal dataset consisting of two instances I_1 and I_2. Each world $w_i \in W$ is labelled with the literals in $\Lambda_{\mathcal{P}} = \{p, \Diamond p, \Box p, q, \Diamond q, \Box q\}$ that are true on it.

Fig. 1 illustrates an example of modal dataset.

A modal association rule is informally defined as a pair of modal formulas that show a statistically interesting association pattern on a modal dataset. For a rigorous definition, we begin with the atomic concept of modal literal.

Definition 2 (modal literals, modal rules)**.** *For a fixed alphabet $\mathcal{P}$, a* positive modal literal *(or, simply, a* literal*) is an object λ of the type*

$$\lambda ::= p \mid \Diamond\lambda \mid \Box\lambda.$$

The set of all modal literals over $\mathcal{P}$ is denoted by $\Lambda_{\mathcal{P}}$. A modal rule *(or, simply, a* rule*) is an object of the type*

$$\rho : \lambda_1 \wedge \ldots \wedge \lambda_k \Rightarrow \lambda_{k+1},$$

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

Modal literals of the type p are called *propositional*, as in the propositional case, and modal literals of the type $\Diamond\lambda$ (resp., $\Box\lambda$) are called *existential* (resp., *universal*). Non-empty subsets of $\Lambda_{\mathcal{P}}$ are called *modal* patterns. In a rule ρ, $X = \{p_1, \ldots, p_k\}$ is called *antecedent*, and it is also denoted by $ant(\rho)$, $Y = \{p_{k+1}\}$ is called *consequent*, also denoted by $con(\rho)$, and $X \cap Y = \emptyset$. The *length* of a (propositional) pattern X is the set-theoretic cardinality of X. An example of modal dataset with six instances is shown in Fig. 2. Observe that in the example, all instances have the same frame; although this is not a requirement in modal association rule extraction, it is often the case.

Association analysis should be interpreted with caution. Causality is not implied by an association rule: it simply indicates a strong co-occurrence of the literals in the antecedent with those in the consequent [13]. To emphasize the distinction from logical implication (denoted $\rightarrow$), we use the symbol $\Rightarrow$ in rules. Consequently, the conjunction in antecedents should not be interpreted as a logical conjunction. Thus, ρ is *not* a modal logic formula. After the extraction phase, however, rules are considered meaningful and can be treated as logical formulas. A modal association rule then becomes a formula of the Horn fragment of modal logic [14, 15] (similar to the propositional case, where an association rule becomes a Horn formula [16]). Horn propositional and modal logics are interesting from a deductive perspective, as satisfiability often becomes computationally simpler. This is relevant when data-driven knowledge is integrated with expert, top-down knowledge in intelligent applications.

Not all rules are interesting or meaningful. Similar to the propositional case, in order to capture the notion of meaningfulness in logical terms, in [17] the authors implicitly defined the notion of a rule ρ holding on an instance, and then on a dataset, by introducing two parameters: support and confidence. These parameters modify the notion of truth of a literal, and subsequently of a rule. Following the introduction of these concepts, especially confidence, alternative measures of meaningfulness have been proposed, including lift and conviction (see, e.g., [13]). More recently, additional measures of rule interestingness have been introduced to avoid extracting trivially true rules.

Generalizing these concepts to the modal case requires generalizing the notion of a set of literals, that is, a pattern, being frequent in a dataset. At the propositional level, a literal p is considered to hold on some instance if it is true in it; at the modal level, a literal λ should be considered interesting if it is both *locally* and *globally* frequent: via local frequency we aim to evaluate on how many worlds of a given instance λ occurs, while via global frequency we aim to evaluate in how many instances this happens. This notion, that requires two parameters s_l, s_g (resp., the local and the global support), induces a generalized notion of a pattern X *holding* on some instance I and on some dataset $\mathcal{I}$.

Definition 3 (support)**.** *Let $\mathcal{I}$ be a modal dataset and $X \subseteq \Lambda_{\mathcal{P}}$ be a set of modal 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 $s_l \in (0, 1]$, the global support *of X on $\mathcal{I}$* relatively to s_l *is defined as:*

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

Notation-wise, given certain local and global supports $s_l, s_g \in (0, 1]$, we write $I \Vdash^{s_l} X$ (resp., $\mathcal{I} \Vdash^{s_l, s_g} X$) to denote the fact that $lsupp(I, X) \geq s_l$ (resp., $gsupp_{s_l}(\mathcal{I}, X) \geq s_g$), and we say that X *locally holds* (resp., *globally holds*) on I (resp., $\mathcal{I}$). A set of literals that globally holds on a dataset $\mathcal{I}$ is said to be *frequent*.

Rule extraction relies on frequent pattern extraction, which is determined by support. Rules are derived from frequent patterns using a probing algorithm that assesses the interestingness of a potential rule in terms of its confidence, lift, or other support-dependent indices. These indices must also be generalized to the modal case.

Definition 4 (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 $s_l \in (0, 1]$, 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 $s_l \in (0, 1]$, the global lift *of ρ on $\mathcal{I}$* relatively to s_l *is defined as:*

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

As before, and again generalizing the propositional case, local and global confidence (as well as lift and the other interestingness measures) induce a notion of a rule *holding* on an instance and on a dataset; for example, given a certain local and global support, and global confidence $s_l, s_g, c_g \in (0, 1] \subset \mathbb{R}$, for a rule $\rho : X \Rightarrow Y$ we write $\mathcal{I} \Vdash^{s_l, s_g, c_g} \rho$ to denote the fact that $\mathcal{I} \Vdash^{s_l, s_g} X \cup Y$ and that $gconf_{s_l}(\mathcal{I}, X \Rightarrow Y) \geq c_g$. This is not the only possible notion of a rule holding on a dataset; different ones depend on the subset of measures that are considered in a particular case, and both a local (to a single instance) and a global (to a whole dataset) analysis can be performed. To ease the notation, in general we write $\mathcal{I} \Vdash^{\theta} \rho$ (or $I \Vdash^{\theta} \rho$, if the analysis is purely local), where θ collects all chosen thresholds.

In summary, a proper modal association rule extraction process is characterized by the following choices: (i) the modal logic of reference: deciding how to interpret the instances of a non-tabular dataset as modal instances, identifying worlds, relations, and propositions; (ii) the learning parameters: setting appropriate parameters to determine minimal local and global support; and (iii) the rule probing algorithm: listing all frequent sets of literals, probing different rules on each, and accepting them if they meet minimal confidence and lift, and possibly other specific conditions relevant to the case at hand.

3. Frequent Modal Pattern Extraction with ModalFP-Growth and Association Rule Mining

The FP-Growth algorithm is designed to efficiently extract frequent itemsets from a propositional dataset by compressing the dataset into an FP-Tree structure. This compression enables efficient reading and processing of the crucial information needed for mining frequent itemsets. We start by summarizing the essential characteristics of an FP-Tree and the operation of FP-Growth as originally introduced by Han et al. [11]. Subsequently, we generalize this approach to handle modal datasets, which we call ModalFP-Growth, following the pseudocode in Alg. 1. Finally, we prove both ModalFP-Growth's soundness and completeness, before presenting the association rule mining algorithm.

3.1. Overview of FP-Growth

An FP-Tree T is a data structure composed of two main components: $T.tree$, a prefix tree, and $T.htable$, a hash table known as the *header* table. Each node $\eta \in T.tree$ has five attributes: $\eta.content$, a literal from a fixed alphabet $\mathcal{P}$, $\eta.children$, a collection of child nodes, $\eta.parent$, a reference to the parent node (possibly null), $\eta.count$, a counter indicating the number of identical nodes represented by η, and $\eta.link$, a pointer to next node $\nu \in T.tree$ with the same content ($\eta.content = \nu.content$).

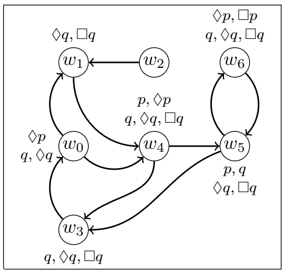
The header table $T.htable$ maps each literal $p \in \mathcal{P}$ to a node $\eta \in T.tree$ where $\eta.content = p$. This mapping facilitates efficient horizontal traversal of the tree, which is frequently required during the construction of *conditional pattern bases* in FP-Growth.

The first step in FP-Growth is to compress the initial dataset into an FP-Tree. This is achieved by inserting each instance into the tree as a branch, considering items in order of decreasing frequency to maximize compression. Nodes representing non-frequent items are pruned to ensure the tree contains only necessary information. The process involves two iterative phases. First, it extracts a conditional pattern base for a frequent item p, which is a projection of the original dataset retaining only relevant information for mining itemsets containing p. Then, it builds a new FP-Tree from the conditional pattern base. This iterative process stops when the generated FP-Tree degenerates into a list, at which point all itemsets and their combinations are mined from the list. The reader can refer to FP-Growth's original paper by Han et al. [11] for comprehensive details.

3.2. Transition to ModalFP-Growth

To handle modal datasets, we extend the FP-Growth algorithm to the ModalFP-Growth algorithm. A modal dataset consists of instances represented as Kripke models. From an operational perspective, each Kripke model is transformed into a tabular form: rows of the table correspond to worlds in the Kripke model, and the columns correspond to literals (see Fig. 3).

Alg. 1 is the generalization of FP-Growth to modal datasets. The transition from FP-Growth to ModalFP-Growth involves the following key modifications. Each instance in the modal dataset is represented as a Kripke model. These models are transformed into a tabular form where rows correspond to worlds and columns to literals (*modal data representation*). The literals $\Lambda_{\mathcal{P}}$ are generated based on the propositions $\mathcal{P}$ and the maximum modal depth δ (*literal generation*). For each world $w \in W$ in the Kripke model $I = (W, R, v) \in \mathcal{I}$, locally frequent itemsets $F_w \subseteq \Lambda_{\mathcal{P}}$ are identified based on a local support threshold s_l (*local support calculation*). These locally frequent itemsets F_w are inserted into an FP-Tree, which is then used to extract (locally) frequent patterns ($Itemsets$) through the FP-Growth process



can be seen as

	p	$\Diamond p$	$\Box p$	q	$\Diamond q$	$\Box q$
w_0	$\bot$	$\top$	$\bot$	$\top$	$\top$	$\bot$
w_1	$\bot$	$\bot$	$\bot$	$\bot$	$\top$	$\top$
w_2	$\bot$	$\bot$	$\bot$	$\bot$	$\bot$	$\bot$
w_3	$\bot$	$\bot$	$\bot$	$\top$	$\top$	$\top$
w_4	$\top$	$\top$	$\bot$	$\top$	$\top$	$\top$
w_5	$\top$	$\bot$	$\bot$	$\top$	$\top$	$\top$
w_6	$\bot$	$\top$	$\top$	$\top$	$\top$	$\top$

Figure 3: Instance $I \in \mathcal{I}$ of a modal dataset $\mathcal{I}$ in its tabular form.

ALGORITHM 1: ModalFP-Growth algorithm for mining frequent modal patterns.

input : Modal dataset $\mathcal{I}$, propositions $\mathcal{P}$, maximum modal depth δ, user-specific parameterization θ
output: Set of globally frequent modal itemsets

function $ModalFP\text{-}Growth(\mathcal{I}, \mathcal{P}, \delta, \theta)$:
    $\Lambda_{\mathcal{P}} \leftarrow GenerateLiterals(\mathcal{P}, \delta)$
    $s_l \leftarrow LocalSupportThreshold(\theta)$
    $s_g \leftarrow GlobalSupportThreshold(\theta)$
    $LocalItemsets \leftarrow \emptyset$
    foreach $I = (W, R, v) \in \mathcal{I}$ **do**
        $T \leftarrow NewFP\text{-}Tree()$
        foreach $w \in W$ **do**
            $F_w \leftarrow \{X \in \Lambda_{\mathcal{P}} \mid I, w \Vdash X$ and $lsupp(I, X) \geq s_l\}$
            $F_w \leftarrow SortDecreasinglyByLocalSupport(F_w)$
            $InsertTree(T.tree, F_w)$
        end
        $PopulateHeaderTable(T.htable)$
        $Itemsets \leftarrow FP\text{-}Growth(T, s_l, \emptyset)$
        $LocalItemsets \leftarrow LocalItemsets \cup Itemsets$
    end
    $GlobalItemsets \leftarrow FilterFrequentItemsets(LocalItemsets, \mathcal{I}, s_g)$
    return $GlobalItemsets$
end

function $InsertTree(\eta, X)$:
    if $|X| = 0$ **then return**
    $\lambda, \tilde{X} \leftarrow PopFirst(X)$
    if $\exists \nu \in \eta.children$ **and** $\nu.content = \lambda$ **then**
        $\eta \leftarrow \nu$
        $\eta.count \leftarrow \eta.count + 1$
    else
        $\nu \leftarrow NewNode(content = \lambda, parent = \eta, count = 1)$
        $AddChild(\eta, \nu)$
        $\eta \leftarrow \nu$
    end
    $InsertTree(\eta, \tilde{X})$
end

function $FP\text{-}Growth(T, s_l, X)$:
    if $IsSinglePath(T.tree)$ **then return** $\{X \cup Y \mid Y \in combinations(T.tree)\}$
    $FrequentItemsets \leftarrow \emptyset$
    foreach $\lambda \in T.htable$ **do**
        $PatternBase \leftarrow ConditionalPatternBase(T, \lambda)$
        $ConditionalTree \leftarrow BuildConditionalTree(PatternBase, s_l)$
        $FrequentItemsets \leftarrow FrequentItemsets \cup FP\text{-}Growth(ConditionalTree, s_l, \{\lambda\})$
    end
    return $FrequentItemsets$
end

function $FilterFrequentItemsets(GlobalItemsets, \mathcal{I}, s_g)$:
    $H \leftarrow CountMap(GlobalItemsets)$
    $FrequentItemsets \leftarrow \{X \in H.keys \mid \frac{H[X]}{|\mathcal{I}|} \geq s_g\}$
    return $FrequentItemsets$
end

(*FP-Tree construction*). The globally frequent itemsets ($GlobalItemsets$) are identified by filtering the combined locally frequent itemsets ($LocalItemsets$) based on the global support threshold s_g (*global filtering*). The ModalFP-Growth algorithm ensures the efficient extraction of frequent modal itemsets, maintaining the interpretability and explainability of the symbolic models.

3.3. Soundness and Completeness

The ModalFP-Growth algorithm is both sound and complete, ensuring that all and only the frequent modal itemsets are identified.

Theorem 1 (Soundness of ModalFP-Growth)**.** *Let $\mathcal{I} = \{I_1, \ldots, I_m\}$ be a modal dataset, where each I_i is a Kripke model represented in tabular form. Then, if ModalFP-Growth returns an $X \subseteq \Lambda_{\mathcal{P}}$ then X is frequent.*

Proof. To prove the soundness of the ModalFP-Growth algorithm, we need to show that every itemset X included in the final output is frequent, meaning that it meets the global support criterion.

First, we consider the FP-Growth algorithm applied to the tabular representation of each Kripke model instance $I_i = (W_i, R_i, v_i)$. This algorithm correctly identifies all itemsets that are locally frequent. Formally, for an itemset X to be considered locally frequent in an instance I_i, it must satisfy:

$$lsupp(I_i, X) = \frac{|\{w \in W_i \mid I_i, w \Vdash X\}|}{|W_i|} \geq s_l,$$

where s_l is the local support threshold.

After running FP-Growth on each instance $I_i \in \mathcal{I}$, we obtain a collection of locally frequent itemsets for each instance. These itemsets are then combined to form a global collection. An itemset X is included in the final result if and only if it appears in a sufficient number of instances with the required local support. Specifically, X is included if:

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

where s_g is the global support threshold.

Since the FP-Growth algorithm ensures that any itemset X identified for an instance I_i satisfies $lsupp(I_i, X) \geq s_l$, the final step of ModalFP-Growth ensures that an itemset X is included in the output if and only if it satisfies the global support criterion. By the definition of global support, for an itemset X to be included in the final output, it must satisfy:

$$gsupp_{s_l}(\mathcal{I}, X) \geq s_g.$$

This means that X is frequent in the modal dataset $\mathcal{I}$ according to the defined global support criterion. Therefore, every itemset X included in the final output of the ModalFP-Growth algorithm is frequent with respect to the specified support thresholds. □

Theorem 2 (Completeness of ModalFP-Growth)**.** *Let $\mathcal{I} = \{I_1, \ldots, I_m\}$ be a modal dataset, where each I_i is a Kripke model represented in tabular form. Then, if $X \subseteq \Lambda_{\mathcal{P}}$ is frequent, then ModalFP-Growth returns it.*

Proof. To prove the completeness of the ModalFP-Growth algorithm, we need to show that every itemset X that is frequent, meaning it meets the global support criterion $gsupp_{s_l}(\mathcal{I}, X) \geq s_g$, is included in the final output of the algorithm.

First, consider an itemset X that is frequent in the modal dataset $\mathcal{I}$. By definition, this means:

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

ALGORITHM 2: Association rule mining from a set of frequent itemsets.

input : Modal dataset $\mathcal{I}$, set of frequent itemsets $\mathcal{S}$, user-specific parameterization θ
output: Set of globally confident association rules

function $AssociationRuleMining(\mathcal{I}, \mathcal{S}, \theta)$:
    $s_l \leftarrow LocalSupportThreshold(\theta)$
    $\mu \leftarrow InterestingnessMetric(\theta)$
    $\tau \leftarrow InterestingnessMetricThreshold(\theta)$
    $Rules \leftarrow \emptyset$
    foreach $Z \in \mathcal{S}$ **do**
        foreach *non-empty set* $X \subset Z$ **do**
            $Y \leftarrow Z \setminus X$
            if $ChecksOut(X \Rightarrow Y, \theta)$ **and** $\mu_{s_l}(\mathcal{I}, X \Rightarrow Y) \geq \tau$ **then**
                $Rules \leftarrow Rules \cup \{X \Rightarrow Y\}$
            end
        end
    end
    return $Rules$
end

implying that X has a local support of at least s_l in at least a fraction s_g of the instances in $\mathcal{I}$. Let $\mathcal{J} \subseteq \mathcal{I}$ be the subset of instances where X has a local support of at least s_l:

$$\mathcal{J} = \{I_i \in \mathcal{I} \mid lsupp(I_i, X) \geq s_l\}.$$

By the definition of global support, we have:

$$\frac{|\mathcal{J}|}{|\mathcal{I}|} \geq s_g.$$

Next, consider the FP-Growth algorithm applied to each instance $I_i \in \mathcal{J}$. Since X has a local support of at least s_l in each $I_i \in \mathcal{J}$, the FP-Growth algorithm will correctly identify X as a locally frequent itemset for these instances. ModalFP-Growth then combines the locally frequent itemsets identified by FP-Growth from each instance. Since X is identified as locally frequent in each instance of $\mathcal{J}$, it will be included in the collection of itemsets combined by ModalFP-Growth.

Finally, ModalFP-Growth checks if X meets the global support criterion. Since X is frequent by assumption, it satisfies $gsupp_{s_l}(\mathcal{I}, X) \geq s_g$. Therefore, X will be included in the final output of the ModalFP-Growth algorithm. $\square$

3.4. Association Rule Mining

Alg. 2 extracts significant relationships from frequent itemsets using a user-defined interestingness measure, generically denoted by μ (e.g., confidence or lift, among others) and its threshold τ. The algorithm processes the modal dataset $\mathcal{I}$ and the set of frequent itemsets $\mathcal{S}$, by using parameters from θ, which is supposed to include the value of μ.

For each frequent itemset $Z \in \mathcal{S}$, the algorithm considers all non-empty subsets $X \subset Z$, generating potential rules of the type $X \Rightarrow Y$, where $Y = Z \setminus X$. Each rule is evaluated based on μ, which is parametric in s_l (recall Definition 4), and those meeting or exceeding τ are retained. Generating these potential rules follows specific policies defined by the user in θ and checked by the function $ChecksOut$. Such policies range from rather standard ones, such as imposing that Y is always a singleton, to specific ones to lower the probability of getting insignificant results. To understand this, recall that at the propositional level a literal p is considered to hold on some instance if it is true on it, but at the modal level literals may sometimes be trivially true on a world of a finite Kripke structure, such as in the case of $\Box p$ on a world without successors; another representative example of potential problem is that of a proposition p being true on some world which is the successor of many worlds, forcing $\Diamond p$ to hold on all of them and making it difficult to establish its interesting degree only by the cardinality of its support set. Rule selection policies may partially address these problems; forcing an antecedent to include at least one non-modal literal, and avoiding specific combinations of propositions between antecedent and consequent of a rule are two among many possible examples.

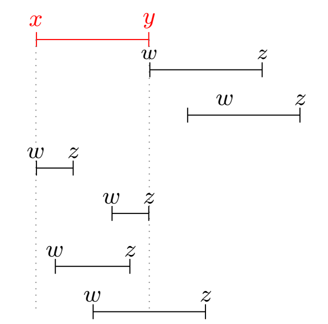
HS modality	Definition w.r.t. the interval structure	Example
$\langle A \rangle$ (after)	$[x,y]R_A[w,z] \quad \Leftrightarrow \quad y = w$	
$\langle L \rangle$ (later)	$[x,y]R_L[w,z] \quad \Leftrightarrow \quad y < w$	
$\langle B \rangle$ (begins)	$[x,y]R_B[w,z] \quad \Leftrightarrow \quad x = w \wedge z < y$	
$\langle E \rangle$ (ends)	$[x,y]R_E[w,z] \quad \Leftrightarrow \quad y = z \wedge x < w$	
$\langle D \rangle$ (during)	$[x,y]R_D[w,z] \quad \Leftrightarrow \quad x < w \wedge z < y$	
$\langle O \rangle$ (overlaps)	$[x,y]R_O[w,z] \quad \Leftrightarrow \quad x < w < y < z$	

Table 1
Allen's interval relations and HS modalities.

This approach offers flexibility by allowing any user-defined interestingness measure to be used to ensure that the algorithm adapts to various analytical needs. Additionally, it supports local mining within individual modal instances, leveraging local definitions of support, confidence, lift, and other measures to generate significant rules. These locally mined rules can then be assessed for their global significance.

4. Experiments

As we have explained, modal association rules can be extracted from temporal, spatial, or other types of non-tabular data; to show the effectiveness of our approach, we focus here on a temporal case.

4.1. Interval Temporal Logic and HS Modalities

A temporal dataset is generally presented as dataset of multivariate time series. We choose to describe temporal logical patterns using interval temporal logic, a specialization of modal logic; among the various interval temporal logics proposed in recent literature [18], *Halpern and Shoham's Modal Logic for Time Intervals* (*HS*) [19] has received significant attention.

Let $\mathbb{D} = \langle D, < \rangle$ be a linear order with domain D. A strict interval over $\mathbb{D}$ is an ordered pair $[x, y]$, where $x, y \in \mathbb{D}$ and $x < y$. Excluding the identity relation, there are 12 different binary ordering relations between two strict intervals on a linear order, often called Allen's interval relations [20], and depicted in Tab. 1. Interval structures are interpreted as Kripke structures, with Allen's relations serving as accessibility relations. Each Allen's relation R_X is associated with an existential modality $\langle X \rangle$. Additionally, for each $X \in \{A, L, B, E, D, O\}$, the transpose of modality $\langle X \rangle$ is $\langle \overline{X} \rangle$, corresponding to the inverse relation $R_{\overline{X}}$.

Well-formed HS formulas are built from a set of propositions $\mathcal{P}$, classical connectives $\wedge$ and $\neg$, and a modality for each Allen's interval relation:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \langle X \rangle \varphi,$$

where $p \in \mathcal{P}$ and $X \in \mathcal{X}$. Syntactically, HS is a modal logic with 12 diamond operators and their corresponding box versions.

The strict semantics of HS is defined in terms of interval models $M = \langle \mathbb{I}(\mathbb{D}), v \rangle$, where $\mathbb{I}(\mathbb{D})$ is the set of all strict intervals over $\mathbb{D}$, and $v : \mathbb{I}(\mathbb{D}) \to 2^{\mathcal{P}}$, which associates each interval with the set of propositions true in that world. The truth of a formula φ on an interval $[x, y]$ in an interval model M,

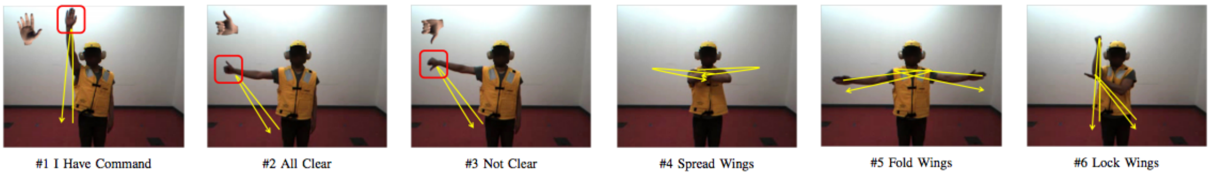


Figure 4: The six movements of interest, referred to as $1, \ldots, 6$ in the text, extracted from NATOPS dataset.

denoted $M, [x, y] \Vdash \varphi$, is defined by structural induction:

$$
\begin{array}{lll}
M, [x, y] \Vdash p & \text{iff} & p \in v([x, y]), \text{ for each } p \in \mathcal{AP}, \\
M, [x, y] \Vdash \neg\psi & \text{iff} & M, [x, y] \not\Vdash \psi, \\
M, [x, y] \Vdash \psi_1 \wedge \psi_2 & \text{iff} & M, [x, y] \Vdash \psi_1 \text{ and } M, [x, y] \Vdash \psi_2, \\
M, [x, y] \Vdash \langle X \rangle \psi & \text{iff} & \text{there exists } [w, z] \text{ s.t. } [x, y] R_X [w, z] \text{ and } M, [w, z] \Vdash \psi,
\end{array}
$$

where $X \in \mathcal{X}$.

4.2. Time Series as Interval Models

A single multivariate time series with variables $\mathcal{V} = \{V_1, \ldots, V_n\}$, where each temporal variable is defined over N points, can be interpreted as an interval model. To this end, we use a set of *feature extraction functions* $\mathcal{F} = \{F_1, \ldots, F_k\}$, where each function F is defined as $F : \mathbb{R}^d \to \mathbb{R}$ for some natural value $d \leq N$; examples include *maximum* and *mean*. To interpret a time series as an interval model, we fix $\mathbb{D} = \langle \{1, \ldots, N\}, < \rangle$, compute the set of strict intervals $\mathbb{I}(\mathbb{D})$, and define the set of propositions

$$\mathcal{P} = \{a \leq F(V) \leq b \mid F \in \mathcal{F}, V \in \mathcal{V}, a \in \mathbb{R} \cup \{-\infty\}, b \in \mathbb{R} \cup \{+\infty\}\}.$$

In a way, feature extraction functions reduce the dimensionality of time series by summarizing information over intervals, allowing for more sophisticated analysis of the temporal relationships. By applying feature extraction functions to intervals within a time series, we can study their values and their relative qualitative positions. Such transformation enables expressing complex temporal relationships through propositional interval temporal logic, such as, for instance, describing an interval where *the average of* $V_1 \geq 4$ *and such that it is overlapped by an interval where the maximum of* $V_2 \leq 12$.

4.3. Experimental Setup

In our experiments we use a well-known public dataset, namely NATOPS, introduced in [21]. Each instance in this dataset is a time series representing the x, y, z coordinates of sensors placed on various body parts of subjects performing aircraft handling signals. These signals are standardized in the Naval Air Training and Operating Procedures Standardization (NATOPS) manual. The dataset, initially designed for classification, includes several signals such as "I have command", "All clear", "Not clear", "Spread wings", "Fold wings", and "Lock wings", as depicted in Fig. 4.

Our objective is to describe common patterns within the same family of movements and highlight unique patterns across different classes. We focus on the six classes in Fig. 4, so that our actual dataset comprises 360 instances balanced across classes (i.e., 60 instances per class).

4.4. Experimental Procedure

The experiments are organized as follows. First, we decide the policy following which the set $\mathcal{P}$ is built; we analyze the variable behaviour via a pre-processing step to identify the informative ones and the significant thresholds. Then, we set the local and global support thresholds; in our case, $s_l = s_g = 0.1$, and $c_g = 0.3$. Finally, we set the policy for rule extraction; in particular, we choose to

Rule	Target class	Measures	$1-S$
$\langle E\rangle min(x_{rh}) \geq 1 \wedge min(z_{rh}) \geq -0.5 \Rightarrow [D]min(y_{rh}) \geq 0$	I have command	$s_g = 0.47$ $c_g = 0.61$	0.37
$min(y_{rh}) \geq 1 \Rightarrow [B]min(z_{rh}) \geq -0.5$	I have command	$s_g = 0.20$ $c_g = 1.00$	0.61
$min(x_{rh}) \geq 1 \wedge \langle O\rangle min(y_{rh}) \geq 1 \Rightarrow [D]min(z_{rh}) \geq -0.5$	I have command	$s_g = 0.10$ $c_g = 1.00$	1.00
$min(y_{rh}) \geq -0.5 \Rightarrow [O]max(\Delta_{rhrt}) \leq 0.0$	Not clear	$s_g = 0.80$ $c_g = 0.80$	0.54
$max(\Delta_{rhrt}) \leq 0.0 \Rightarrow [O]min(y_{rh}) \geq -0.5$	Not clear	$s_g = 1.00$ $c_g = 1.00$	1.00
$min(y_{lh}) \geq -1.0 \Rightarrow max(z_{le}) \leq -0.25$	Lock wings	$s_g = 0.23$ $c_g = 0.87$	0.39
$\langle O\rangle min(y_{lh}) \geq -1.0 \wedge max(z_{le}) \leq -0.25 \wedge \langle O\rangle min(y_{re}) \geq -0.5$ $\wedge \langle O\rangle max(z_{re}) \geq -0.3 \Rightarrow \langle O\rangle min(y_{rh}) \geq 0.5$	Lock wings	$s_g = 0.23$ $c_g = 0.78$	0.64

Table 2
Experiment results showing the association rules extracted, target class, measures of global support s_g and global confidence c_g, and the entropy S of the confidence across all classes except the target. Variables represent coordinates x, y, z, with subscripts indicating body parts (r: *r*ight, l: *l*eft, h: *h*and, e: *e*lbow). The variable Δ_{rhrt} indicates the difference between the height of the *r*ight *h*and and the *r*ight *t*humb. A $1-S$ value close to 1 indicates high specificity of the rule to the target class.

extract only rules with a singleton consequent, to exclude rules where the same variable appears in both antecedent and consequent (*self-absorbing rules*), and to exclude rules with antecedents containing only non-propositional literals (*non-anchored rules*). After mining the most interesting association rules for the target class, we compute the entropy S of the set $\{c_1, \ldots, c_6\}$, where c_i is the global confidence of the rule on the instances of class i. Setting $C = \sum_{i=1}^{6} c_i$ and $\pi_i = c_i/C$, the entropy is defined as $S = -\sum_{i=1}^{6} \pi_i log_2(\pi_i)$, and it is one way to determine a rule's effectiveness in describing only the target class. Finally, we interpret the results by reading the rules in natural language, in order to understand and to explain, whenever possible, the extracted patterns.

4.5. Results and Analysis

We conducted three experiments, with results summarized in Tab. 2. Each rule is associated with a target class, and its global support s_g and confidence c_g are reported. The entropy S of the confidence across other classes is also included; a $1-S$ value close to 1 indicates that the rule is very specific to the target class. Variables in the rules represent coordinates x, y, z with subscripts indicating the body part (e.g., rh for *r*ight *h*and, le for *l*eft *e*lbow). The variable Δ_{rhrt} denotes the difference between the height of the *r*ight *h*and and the *r*ight *t*humb.

Experiment 1: "I have command". The first two rules are common between "I have command" (resp., $c_g = 0.61$ and $c_g = 1.00$) and "Lock wings" (resp., $c_g = 0.30$ and $c_g = 1.00$) classes. The second rule, for instance, translates to *when the right hand is above the ears, it is also slightly to the right of the body.* The third rule uniquely identifies a pattern typical to the target class, translating to *when the right hand is distant from the body forward and at ear height, it is also to the right.*

Experiment 2: "Not clear". The fourth rule is common between "All clear" ($c_g = 0.42$) and "Not clear" ($c_g = 0.80$) classes, describing a less precise movement. It translates to *when the arm is just below shoulder height, the right thumb points downward.* The fifth rule uniquely identifies the target class, translating to *when the right thumb points downward, the right hand is just below shoulder height,* indicating that the thumb is pointed downward during the hand's ascent.

Experiment 3: "Lock wings". The second-to-last rule captures a common behavior across three classes: "Spread wings", "Fold wings", and "Lock wings." The final rule uniquely describes the target class, translating to *when the left hand is at chest height and the left elbow is retracted towards the sternum, while the right elbow is raised and to the right, the right hand is above the chin.*

5. Conclusions

This paper introduced a novel approach for extracting modal association rules from non-tabular data. Our extension of the FP-Growth algorithm to handle modal data allows for mining frequent patterns and interpretable association rules, which may be crucial in certain applications. We demonstrated the effect of non-propositional rule extraction on a public dataset in the temporal case.

The primary value of our work lies in its ability to process non-tabular data natively, maintaining interpretability and relevance for domain experts. A significant challenge identified is the selection of appropriate modal relations and alphabets for specific problems, which we propose as a future research direction. We have open-sourced our implementation to promote transparency and reproducibility, enabling others to apply and extend our methods.

Our research provides a robust framework for modal association rule mining and is a first step toward showing its applicability in practice to capture temporal relationships. In future work we will optimize language selection, integrate additional interestingness measures, and extend the framework to other forms of modal logics and relational data.

Acknowledgments

We acknowledge the support of the FIRD project *Methodological Developments in Modal Symbolic Geometric Learning*, funded by the University of Ferrara, and the INDAM-GNCS project *Symbolic and Numerical Analysis of Cyberphysical Systems* (code CUP_E53C23001670001), funded by INDAM; Giovanni Pagliarini, Guido Sciavicco, and Ionel Eduard Stan are GNCS-INdAM members. Moreover, this research has also been funded by the Italian Ministry of University and Research through PNRR - M4C2 - Investimento 1.3 (Decreto Direttoriale MUR n. 341 del 15/03/2022), Partenariato Esteso PE00000013 - "FAIR - Future Artificial Intelligence Research" - Spoke 8 "Pervasive AI", funded by the European Union under the NextGeneration EU programme".

References

[1] Parliament and Council of the European Union, General Data Protection Regulation, 2016. URL: https://gdpr-info.eu/.

[2] A. F. Markus, J. A. Kors, P. R. Rijnbeek, The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies, J. Biomed. Informatics 113 (2021) 103655.

[3] O. Vermesan, V. Piuri, F. Scotti, A. Genovese, R. Labati, P. Coscia, Explainability and Interpretability Concepts for Edge AI Systems, 2024, pp. 197–227.

[4] F. Manzella, G. Pagliarini, G. Sciavicco, I. E. Stan, Interval Temporal Random Forests with an Application to COVID-19 Diagnosis, in: Proc. of the 28th International Symposium on Temporal Representation and Reasoning (TIME), volume 206 of *LIPIcs*, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021, pp. 7:1–7:18.

[5] G. Pagliarini, G. Sciavicco, Decision Tree Learning with Spatial Modal Logics, in: Proc. of the 12th International Symposium on Games, Automata, Logics, and Formal Verification (GandALF), volume 346 of *EPCTS*, 2021, pp. 273–290.

[6] G. Sciavicco, I. E. Stan, Knowledge Extraction with Interval Temporal Logic Decision Trees, in: Proc. of the 27th International Symposium on Temporal Representation and Reasoning (TIME), volume 178 of *LIPIcs*, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020, pp. 9:1–9:16.

[7] D. Bresolin, E. Cominato, S. Gnani, E. Muñoz-Velasco, G. Sciavicco, Extracting Interval Temporal Logic Rules: A First Approach, in: Proc. of the 25th International Symposium on Temporal Representation and Reasoning (TIME), volume 120 of *LIPIcs*, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018, pp. 7:1–7:15.
[8] M. Milella, E. Muñoz-Velasco, G. Pagliarini, A. Paradiso, G. Sciavicco, I. E. Stan, On Modal Logic Association Rule Mining, in: Proc. of the 23rd Italian Conference on Theoretical Computer Science (ICTCS), volume 3284 of *CEUR*, CEUR-WS.org, 2022, pp. 53–65.
[9] J. Martinet, S. Satoh, A study of intra-modal association rules for visual modality representation, 2007, pp. 344 – 350.
[10] M. Ceci, C. Loglisci, L. Rudd, D. Malerba, Discovering knowledge through multi-modal association rule mining for document image analysis, in: Proceedings of 1st AI*IA Workshop on Intelligent Techniques At LIbraries and Archives co-located with XIV Conference of the Italian Association for Artificial Intelligence, IT@LIA@AI*IA 2015, Ferrara, Italy, September 22, 2015, volume 1509 of *CEUR Workshop Proceedings*, 2015.
[11] J. Han, J. Pei, Y. Yin, Mining Frequent Patterns without Candidate Generation, in: Proc. of the 2000 ACM SIGMOD International Conference on Management of Data, 2000, pp. 1–12.
[12] F. Manzella, G. Pagliarini, A. Paparella, G. Sciavicco, I. E. Stan, Sole.jl – Symbolic Learning in Julia, https://github.com/aclai-lab/Sole.jl, 2024.
[13] P. N. Tan, M. S. Steinbach, A. Karpatne, V. Kumar, Introduction to Data Mining, 2nd ed., Pearson, 2019.
[14] L. Fariñas Del Cerro, M. Penttonen, A Note on the Complexity of the Satisfiability of Modal Horn Clauses, The Journal of Logic Programming 4 (1987) 1–10.
[15] D. Bresolin, E. Muñoz-Velasco, G. Sciavicco, On Sub-Propositional Fragments of Modal Logic, Logical Methods in Computer Science 14 (2018).
[16] A. Horn, On Sentences Which are True of Direct Unions of Algebras, Journal of Symbolic Logic 16 (1951) 14–21.
[17] R. Agrawal, R. Srikant, Fast Algorithms for Mining Association Rules in Large Databases, in: Proc. of 20th International Conference on Very Large Data Bases (VLDB), 1994, pp. 487–499.
[18] V. Goranko, A. Montanari, G. Sciavicco, A Road Map of Interval Temporal Logics and Duration Calculi, J. of Applied Non-Classical Logics 14 (2004) 9–54.
[19] J. Y. Halpern, Y. Shoham, A Propositional Modal Logic of Time Intervals, Journal of the ACM 38 (1991) 935–962.
[20] J. Allen, Maintaining knowledge about temporal intervals, Communications of the ACM 26 (1983) 832–843.
[21] Y. Song, D. Demirdjian, R. Davis, Tracking body and hands for gesture recognition: NATOPS aircraft handling signals database, in: Proc. of the 9th IEEE International Conference on Automatic Face and Gesture Recognition (FG), 2011, pp. 500 – 506.