Distinguishability of Structures Relative to a Class of Structures

Abstract

We investigate the distinguishability of relational structures relative to a class of structures in the following sense: for a class of structures, given two sets P and N of structures such that each structure belongs to the class, find a formula in first-order logic distinguishing these sets. The distinguishability can be seen as a declarative framework for machine learning where hypotheses are first-order formulas. We consider the following classes of structures: sets, linear orders, unary relation structures, equivalence structures, and disjoint union of linear orders. In order to define algorithms to the distinguishability problem, we use a measure of similarity defined by winning strategies in Ehrenfeucht–Fraïssé games. For the class of disjoint union of linear orders, we show a necessary and sufficient condition for a winning strategy of a player in such games. Furthermore, we also show how to define formulas with polynomial size describing the gametheoretic properties of a given structure. The distinguishability algorithms here defined run in polynomial time in the size of P, N, and in the larger structure therein. Finally, we briefly show the application of this work in learning first-order formulas to describe states in the blocks world domain.

1 Introduction

Ehrenfeucht–Fraïssé games [Ehrenfeucht, 1961] is a fundamental technique of finite model theory [Ebbinghaus and Flum, 1995; Libkin, 2004; Grädel *et al.*, 2005] in proving the inexpressibility of certain properties in first-order logic. For instance, first-order logic cannot express that a finite structure has even cardinality. The Ehrenfeucht–Fraïssé game is played on two structures by two players, the Spoiler and the Duplicator. The Spoiler tries to show that the structures are different, while the Duplicator tries to show that they are the same. The player who achieves its goal is the winner.

A player, the Spoiler or the Duplicator, has a winning strategy if it is possible for him to win each play whatever choices are made by the opponent. If the Spoiler has a winning strategy for k rounds of such a game, it means that the structures can be distinguished by a first-order sentence whose quantifier rank is at most k.

Besides providing a tool to measure the expressive power of a logic, Ehrenfeucht–Fraïssé games allow one to investigate the similarity between structures by means of the notion of remoteness. In a game played on structures $\mathcal{A}$ and $\mathcal{B}$, remoteness is the minimum number of rounds such that the Spoiler has a winning strategy [Montanari *et al.*, 2005].

Explicit conditions characterising the winning strategies for the Duplicator on some standard finite structures are provided in [Khoussainov and Liu, 2009]. Examples of such standard structures are unary relation structures, equivalence structures, and some of their natural expansions. Besides, it is well known the necessary and sufficient conditions characterising the winning strategies for the Duplicator on sets and linear orders [Libkin, 2004].

An m-Hintikka formula is a formula obtained from a structure $\mathcal{A}$ and a positive integer m that describes the gametheoretic properties of $\mathcal{A}$ on the Ehrenfeucht–Fraïssé game with m rounds [Ebbinghaus and Flum, 1995]. An m-Hintikka formula $\varphi^m_{\mathcal{A}}$ holds exactly on all structures $\mathcal{B}$ such that the Duplicator has a winning strategy for the Ehrenfeucht–Fraïssé game with m rounds in $\mathcal{A}$ and $\mathcal{B}$.

An algorithm to cope with the problem of distinguishing structures is presented in [Kaiser, 2012]. An important part of the distinguishability algorithm is the computation of m-Hintikka formulas from structures: given two sets of structures, the positive and the negative ones, it returns a formula of minimal number of variables and minimal quantifier rank that holds on all positive structures and on none of the negative ones. This algorithm works for arbitrary finite relational structures and runs in exponential time.

Kaiser applied his algorithm in a general system for learning formulas defining board game rules [Kaiser, 2012]. These results are also used in learning reductions and polynomial-time programs [Jordan and Kaiser, 2013a; 2013b; 2013c].

In this work, we introduce distinguishability algorithms on classes of structures. For first-order logic and a class of structures, given two sets of structures P and N such that each structure belongs to the class, find a formula in first-order logic distinguishing these sets.

We define distinguishability algorithms on the following classes of structures: sets, linear orders, unary relation structures, equivalence structures, and disjoint union of linear orders. We considered most of these classes of structures, but the disjoint union of linear orders, because we already have necessary and sufficient results for the Duplicator to have a winning strategy on these classes.

On disjoint union linear orders, we show a result characterising the winning strategies for the Duplicator. Disjoint union of linear orders are interesting because we can model a state of the elementary blocks worlds by using them [Cook and Liu, 2003]. See Section 6 for details.

We also show how to compute m-Hintikka formulas in all classes mentioned above. These results are important in the definition of the distinguishability algorithms on classes of structures. Our distinguishability algorithms on the classes considered run in polynomial time in the size of P, N, and in the larger structure therein.

This paper is organized as follows: in Section 2, we give the basic definitions and results of Ehrenfeucht–Fraïssé games, Hintikka formulas, and remoteness. In Section 3, we show conditions for the Duplicator to have a winning strategy on disjoint union of linear orders. In Section 4, we show results about how to compute Hintikka formulas relative to a class of structures. We propose our distinguishability algorithms and give a example in Section 5. We show an application in Section 6, and we conclude in Section 7.

In the whole text, but in Section 2, all lemmas, theorems and corollaries without explicit reference are our results.

2 Ehrenfeucht–Fraïssé Games and Remoteness

In this section, we present the basic notions about Ehrenfeucht–Fraïssé Games and remoteness with related results. We assume some background from first-order logic and complexity. For details, see [Ebbinghaus and Flum, 1995] and [Papadimitriou, 1994]. In what follows, we are concerned on relational structures.

Definition 1 (EF Game)**.** *Let r be an integer such that $r \geq 0$, τ a vocabulary, $\mathcal{A}$ and $\mathcal{B}$ two τ-structures. The Ehrenfeucht–Fraïssé game (EF game, for short) $\mathcal{G}_r(\mathcal{A}, \mathcal{B})$ is played by two players called the Spoiler and the Duplicator. Each play of the game has r rounds and, in each round, the Spoiler plays first and picks an element from the domain A of $\mathcal{A}$, or from the domain B of $\mathcal{B}$. Then, the Duplicator responds by picking an element from the domain of the other structure. Let $a_i \in A$ and $b_i \in B$ be the two elements picked by the Spoiler and the Duplicator in the ith round. The Duplicator wins the play if the mapping $(a_1, b_1), ..., (a_r, b_r)$ is an isomorphism between the substructures induced by $a_1, ..., a_r$ and $b_1, ..., b_r$, respectively. Otherwise, Spoiler wins this play. We say that a player has a winning strategy in $\mathcal{G}_r(\mathcal{A}, \mathcal{B})$ if it is possible for him to win each play whatever choices are made by the opponent.*

Example 2. *Let $\mathcal{A}$ and $\mathcal{B}$ be directed graphs such that $A = \{1, 2, 3, 4\}$, $E^{\mathcal{A}} = \{(1, 2), (2, 3), (3, 4), (4, 1)\}$, $B = \{1, 2, 3\}$, and $E^{\mathcal{B}} = \{(1, 2), (2, 3), (3, 1)\}$. The Spoiler has a winning strategy in $\mathcal{G}_2(\mathcal{A}, \mathcal{B})$. The Spoiler can win by selecting two elements in $\mathcal{A}$ with no edge between them.*

Let $\mathcal{A}$ be a structure. For $m \geq 0$, the following definition introduces a formula $\varphi^m_{\mathcal{A}}$ that describes the gametheoretic properties of $\mathcal{A}$ in any EF game.

Definition 3 (m-Hintikka Formula)**.** *Let $\mathcal{A}$ be a structure, $\overline{a} = a_1 ... a_s \in A^s$, and $\overline{v} = v_1, ..., v_s$ a tuple of variables, $\varphi^0_{\mathcal{A},\overline{a}}(\overline{v}) := \bigwedge\{\varphi(\overline{v}) \mid \varphi$ atomic or negated atomic and $\mathcal{A} \models \varphi[\overline{a}]\}$ and for $m > 0$, $\varphi^m_{\mathcal{A},\overline{a}}(\overline{v}) := \bigwedge_{a \in A} \exists v_{s+1} \varphi^{m-1}_{\mathcal{A},\overline{a}a}(\overline{v}, v_{s+1}) \wedge \forall v_{s+1}(\bigvee_{a \in A} \varphi^{m-1}_{\mathcal{A},\overline{a}a}(\overline{v}, v_{s+1}))$.*

The Hintikka formula $\varphi^m_{\mathcal{A},\overline{a}}$ describes the isomorphism type of the substructure generated by $\overline{a}$ in $\mathcal{A}$. We write $\varphi^m_{\mathcal{A}}$ whenever $s = 0$.

Example 4. *Let $\mathcal{A} = \langle A, S^{\mathcal{A}} \rangle$ be a structure such that $A = \{1, 2\}$, and $S^{\mathcal{A}} = \{2\}$. Therefore, $\varphi^1_{\mathcal{A}} = \exists v_1 \neg S(v_1) \wedge \exists v_1 S(v_1) \wedge \forall v_1(\neg S(v_1) \vee S(v_1))$.*

Given a structure $\mathcal{A}$ and a positive integer m, the size of the m-Hintikka formula $\varphi^m_{\mathcal{A}}$ is $O(2^m \times |A|^m)$. Therefore, since m is bounded by $|A|$, the size of $\varphi^m_{\mathcal{A}}$ is exponential in the size of $\mathcal{A}$. The following result asserts the relation between m-Hintikka formulas and winning strategies in EF games.

Theorem 5 (Ehrenfeucht's Theorem)**.** *[Ehrenfeucht, 1961] Given $\mathcal{A}$ and $\mathcal{B}$, and $r \geq 0$, the Duplicator has a winning strategy in $\mathcal{G}_r(\mathcal{A}, \mathcal{B})$ iff $\mathcal{B} \models \varphi^r_{\mathcal{A}}$.*

Ehrenfeucht–Fraïssé games provide information about similarity between structures. If two structures $\mathcal{A}$ and $\mathcal{B}$ are not isomorphic, then there is a r such that the Spoiler has a winning strategy in $\mathcal{G}_r(\mathcal{A}, \mathcal{B})$. This information about similiarity is represented by the notion of remoteness.

Definition 6 (Remoteness)**.** *The remoteness of two structures $\mathcal{A}$ and $\mathcal{B}$, denoted by $rem(\mathcal{A}, \mathcal{B})$, is the minimum number of rounds r such that Spoiler has a winning strategy in $\mathcal{G}_r(\mathcal{A}, \mathcal{B})$.*

Example 7. *Let $\mathcal{A}$ and $\mathcal{B}$ be the structures in Example 2. We have $rem(\mathcal{A}, \mathcal{B}) = 2$.*

In the following, we give an elementary result that provides a full solution for EF games in the case the vocabulary $\tau = \emptyset$, that is, a structure is just a set. For every $n \geq 1$, let $\mathcal{A}_n$ and $\mathcal{B}_n$ be sets with n elements.

Theorem 8 (EF Games on Sets)**.** *Let r be a positive integer, $\mathcal{A}_n$ and $\mathcal{B}_m$ be sets. The Duplicator has a winning strategy in $\mathcal{G}_r(\mathcal{A}_n, \mathcal{B}_m)$ if and only if $n = m$ or ($n \geq r$ and $m \geq r$).*

Corollary 9 (Remoteness on Sets)**.** *Let $\mathcal{A}_n$ and $\mathcal{B}_m$ be sets. $rem(\mathcal{A}_n, \mathcal{B}_m) = min(n, m) + 1$.*

The remoteness of sets $\mathcal{A}_n$ and $\mathcal{B}_m$ can be computed in constant time.

Now, we define a linear order structure as a τ-structure $\mathcal{A}$ such that a binary relation $<$ is in τ, and $<$ is a linear order on A. For every $n \geq 1$, let $\mathcal{A}_n$ and $\mathcal{B}_n$ be linear orders with n elements.

Theorem 10 (EF Games on Linear Orders)**.** *Let r be a positive integer, $\mathcal{A}_n$ and $\mathcal{B}_m$ be linear orders. The Duplicator has a winning strategy in $\mathcal{G}_r(\mathcal{A}_n, \mathcal{B}_m)$ if and only if $n = m$ or ($m \geq 2^r - 1$ and $n \geq 2^r - 1$).*

Corollary 11 (Remoteness on Linear Orders)**.** *Let $\mathcal{A}_n$ and $\mathcal{B}_m$ be linear orders. The remoteness can be computed as $rem(\mathcal{A}_n, \mathcal{B}_m) = \lfloor log_2(min(n, m) + 1) \rfloor + 1$.*

The remoteness of linear orders $\mathcal{A}_n$ and $\mathcal{B}_m$ can also be computed in constant time.

Now, let $\tau = \{P_1, ..., P_k\}$ be a vocabulary where each P_i is unary, for $i \in \{1, ..., k\}$. For a structure $\mathcal{A} = \langle A, P_1^{\mathcal{A}}, ..., P_k^{\mathcal{A}}\rangle$, we assume $P_1, ..., P_k$ be pairwise disjoint. We set $P_{k+1}^{\mathcal{A}} = \overline{(P_1^{\mathcal{A}} \cup ... \cup P_k^{\mathcal{A}})}$.

Theorem 12. *[Khoussainov and Liu, 2009] The Duplicator has a winning strategy in $\mathcal{G}_r(\mathcal{A}, \mathcal{B})$ iff for all $i \in \{1, ..., k+1\}$, ($|P_i^{\mathcal{A}}| \geq r$ and $|P_i^{\mathcal{B}}| \geq r$) or $|P_i^{\mathcal{A}}| = |P_i^{\mathcal{B}}|$.*

Corollary 13 (Remoteness on Unary Relation Structures)**.** *Let $\mathcal{A}$ and $\mathcal{B}$ be unary relation structures. The remoteness $rem(\mathcal{A}, \mathcal{B})$ can be computed as $min\{min(|P_i^{\mathcal{A}}|, |P_i^{\mathcal{B}}|) \mid |P_i^{\mathcal{A}}| \neq |P_i^{\mathcal{B}}|, 1 \leq i \leq k+1\}$.*

In a fixed vocabulary $\tau = \{P_1, ..., P_k\}$, the remoteness of unary relation structures $\mathcal{A}$ and $\mathcal{B}$ can be computed in constant time.

An equivalence structure is a structure of the form $\mathcal{A} = \langle A, E^{\mathcal{A}}\rangle$ such that $E^{\mathcal{A}}$ is an equivalence relation on A. Let $q_t^{\mathcal{A}}$ be the number of equivalence classes in $\mathcal{A}$ with size t. Let $q_{\geq t}^{\mathcal{A}}$ be the number of equivalence classes in $\mathcal{A}$ with size at least t. The following definitions and result can be found in [Khoussainov and Liu, 2009].

Definition 14. *The pair of structures $(\mathcal{A}, \mathcal{B})$ has an r-small disparity if there is a $t < n$ such that $q_t^{\mathcal{A}} \neq q_t^{\mathcal{B}}$ and $r > min\{q_t^{\mathcal{A}}, q_t^{\mathcal{B}}\} + t$.*

Definition 15. *The pair of structures $(\mathcal{A}, \mathcal{B})$ has an r-large disparity if there is a $t \leq n$ such that $q_{\geq t}^{\mathcal{A}} \neq q_{\geq t}^{\mathcal{B}}$ and $r \geq min\{q_{\geq t}^{\mathcal{A}}, q_{\geq t}^{\mathcal{B}}\} + t$.*

Theorem 16 (EF Games on Equivalence Structures)**.** *[Khoussainov and Liu, 2009] Let r be a positive integer, $\mathcal{A}$ and $\mathcal{B}$ be equivalence structures. The Duplicator has a winning strategy in $\mathcal{G}_r(\mathcal{A}, \mathcal{B})$ iff $(\mathcal{A}, \mathcal{B})$ has neither r-small disparity nor r-large disparity.*

Corollary 17 (Remoteness on Equivalence Structures)**.** *Let $\mathcal{A}$ and $\mathcal{B}$ be equivalence structures. $rem(\mathcal{A}, \mathcal{B}) = min(min\{min(q_t^{\mathcal{A}}, q_t^{\mathcal{B}}) + t \mid q_t^{\mathcal{A}} \neq q_t^{\mathcal{B}}\} + 1, min\{min(q_{\geq t}^{\mathcal{A}}, q_{\geq t}^{\mathcal{B}}) + t \mid q_{\geq t}^{\mathcal{A}} \neq q_{\geq t}^{\mathcal{B}}\})$.*

The remoteness of equivalence structures $\mathcal{A}$ and $\mathcal{B}$ can be computed in $O(|A| + |B|)$ time.

3 EF Game on Disjoint Union of Linear Orders

In this section, we show conditions for the Duplicator to have a winning strategy on disjoint union of linear orders. We also determine the remoteness in this context.

For a relational vocabulary τ, we introduce the disjoint union of τ-structures. Assume that $\mathcal{A}$ and $\mathcal{B}$ are τ-structures with $A \cap B = \emptyset$. Then, $\mathcal{A} \uplus \mathcal{B}$, the disjoint union of $\mathcal{A}$ and $\mathcal{B}$, is the τ-structure with domain $A \cup B$ and $R^{\mathcal{A} \uplus \mathcal{B}} = R^{\mathcal{A}} \cup R^{\mathcal{B}}$ for any R in τ.

Now, we consider the class of structures $\mathcal{A} = \langle A, <\rangle$ such that $\mathcal{A}$ is a disjoint union of linear orders. We represent a disjoint union of linear orders $\mathcal{A}$ by a disjoint union $(...(\mathcal{A}_{i_1} \uplus \mathcal{A}_{i_2}) \uplus ... \uplus \mathcal{A}_{i_k})$ where i_l is the size of the linear order $\mathcal{A}_{i_l}$, for $l \in \{1, ..., k\}$.

In what follows, for a disjoint union of linear orders $\mathcal{A}$, let $q_t^{\mathcal{A}}$ be the number of linear orders in $\mathcal{A}$ with size t, and $q_{\geq t}^{\mathcal{A}}$ be the number of linear orders in $\mathcal{A}$ with size at least t.

Definition 18 ($(<, r)$-Small Disparity)**.** *Let r be a positive integer, $\mathcal{A}$ and $\mathcal{B}$ be disjoint union of linear orders. We say that the pair $(\mathcal{A}, \mathcal{B})$ has a $(<, r)$-small disparity if there is a $t < 2^r - 1$ such that $q_t^{\mathcal{A}} \neq q_t^{\mathcal{B}}$ and $r > min\{q_t^{\mathcal{A}}, q_t^{\mathcal{B}}\} + \lfloor log_2(t+1)\rfloor$.*

Lemma 19. *Let r be a positive integer, $\mathcal{A}$ and $\mathcal{B}$ be two disjoint union of linear orders. If $(\mathcal{A}, \mathcal{B})$ has a $(<, r)$-small disparity, then the Spoiler has a winning strategy for $\mathcal{G}_r(\mathcal{A}, \mathcal{B})$.*

Proof. Assume that $q_t^{\mathcal{A}} > q_t^{\mathcal{B}}$ and $r > q_t^{\mathcal{B}} + \lfloor log(t+1)\rfloor$. The Spoiler's strategy is the following. The Spoiler chooses elements $a_1, a_2, ..., a_{q_t^{\mathcal{B}}}$ from distinct linear orders of size t in $\mathcal{A}$. Duplicator must choose elements $b_1, ..., b_{q_t^{\mathcal{B}}}$ from distinct linear order of size t in $\mathcal{B}$. Next, the Spoiler chooses an element in a distinct linear order of size t in $\mathcal{A}$. The Duplicator must select an element in a linear order of size different from t. Therefore, the Spoiler has a winning strategy in the game between these linear orders with $\lfloor log_2(t+1)\rfloor + 1$ rounds. □

Example 20. *Let $\mathcal{A}$ and $\mathcal{B}$ be two disjoint union of linear orders such that $\mathcal{A} = \mathcal{A}_6 \uplus \mathcal{A}_7 \uplus \mathcal{A}_7 \uplus \mathcal{A}_7$, and $\mathcal{B} = \mathcal{B}_6 \uplus \mathcal{B}_6 \uplus \mathcal{B}_7 \uplus \mathcal{B}_7$. $(\mathcal{A}, \mathcal{B})$ has a $(<, 4)$-small disparity because, for $t = 6$, we have $t < 2^r - 1$, $q_t^{\mathcal{A}} = 1$, $q_t^{\mathcal{B}} = 2$, $q_t^{\mathcal{A}} \neq q_t^{\mathcal{B}}$, and $r > 1 + \lfloor log(t+1)\rfloor$. For $\mathcal{G}_4(\mathcal{A}, \mathcal{B})$, the Spoiler has the following winning strategy. The Spoiler picks two elements b_1, b_2 from two distinct linear orders with size 6 in $\mathcal{B}$. The Spoiler ensures his winning strategy by choosing b_1, b_2 to be the middle elements. The Duplicator must select an element in a linear order with size different from 6 in $\mathcal{A}$. Now, the Spoiler follows his winning strategy in $\mathcal{G}_3(\mathcal{A}_6, \mathcal{B}_7)$.*

Definition 21 ($(<, r)$-Large Disparity)**.** *Let r be a positive integer, $\mathcal{A}$ and $\mathcal{B}$ be disjoint union of linear orders. We say that the pair $(\mathcal{A}, \mathcal{B})$ has a $(<, r)$-large disparity if there is a $t < 2^r - 1$ such that $q_{\geq t}^{\mathcal{A}} \neq q_{\geq t}^{\mathcal{B}}$ and $r \geq min\{q_{\geq t}^{\mathcal{A}}, q_{\geq t}^{\mathcal{B}}\} + \lfloor log(t+1)\rfloor$.*

Lemma 22. *Let r be a positive integer, $\mathcal{A}$ and $\mathcal{B}$ be disjoint union of linear orders. If $(\mathcal{A}, \mathcal{B})$ has a $(<, r)$-large disparity, then the Spoiler has a winning strategy in the game $\mathcal{G}_r(\mathcal{A}, \mathcal{B})$.*

Proof. Assume that $q_{\geq t}^{\mathcal{A}} > q_{\geq t}^{\mathcal{B}}$ and $r \geq q_{\geq t}^{\mathcal{B}} + \lfloor log(t+1)\rfloor$. The Spoiler selects $q_{\geq t}^{\mathcal{B}}$ elements from distinct linear orders of size at least t in $\mathcal{A}$. The Duplicator must choose elements $b_1, ..., b_{q_t^{\mathcal{B}}}$ from distinct linear order of size at least t in $\mathcal{B}$. Next, the Spoiler selects an element in a distinct linear order of size at least t in $\mathcal{A}$. The Duplicator must select an element in a linear order of size less than t. Therefore, the Spoiler has a winning strategy in the game between these linear orders with $\lfloor log_2(t+1)\rfloor$ rounds. □

Example 23. *Let $\mathcal{A}$ and $\mathcal{B}$ be two disjoint union of linear orders such that $\mathcal{A} = \mathcal{A}_5 \uplus \mathcal{A}_5 \uplus \mathcal{A}_5 \uplus \mathcal{A}_6$, and $\mathcal{B} = \mathcal{B}_5 \uplus \mathcal{B}_5 \uplus \mathcal{B}_5 \uplus \mathcal{A}_6 \uplus \mathcal{A}_{15}$. $(\mathcal{A}, \mathcal{B})$ does not have a $(<, 4)$-small disparity but it has a $(<, 4)$-large disparity. For $t = 6$, we have $q_{\geq t}^{\mathcal{A}} = 1$, $q_{\geq t}^{\mathcal{B}} = 2$, $q_{\geq t}^{\mathcal{A}} \neq q_{\geq t}^{\mathcal{B}}$, and $r \geq 1 + \lfloor log(t+1)\rfloor$. The Spoiler has the following winning strategy. The Spoiler* *selects elements from distinct linear orders with size* ≥ 6 *in* $\mathcal{B}$. *The Duplicator must select an element from a linear order with size* < 6 *in* $\mathcal{A}$. *Next, the Spoiler follows his winning strategy in three rounds in a linear order with size* ≥ 6, *and in another linear order with size* < 6.

Theorem 24 (EF Games on Disjoint Union of Linear Orders)**.** *Let r be a positive integer, $\mathcal{A}$ and $\mathcal{B}$ be disjoint union of linear orders. The Duplicator has a winning strategy in $\mathcal{G}_r(\mathcal{A}, \mathcal{B})$ if and only if $(\mathcal{A}, \mathcal{B})$ has neither $(<, r)$-small disparity nor $(<, r)$-large disparity.*

Proof. On one hand, we apply Lemmas 19 and 22
On the other hand, assume that the Spoiler selects an element a_l in a linear order $\mathcal{A}_t$ such that $q_t^{\mathcal{A}} \neq q_t^{\mathcal{B}}$. If there is no element already chosen in $\mathcal{A}_t$, then the Duplicator must choose an element in a linear order with no element already chosen. If there is a linear order $\mathcal{B}_t$ without selected elements, then the Duplicator chooses an element b_l in this linear order. Otherwise, since $r \leq min\{q_t^{\mathcal{A}}, q_t^{\mathcal{B}}\} + \lfloor log(t+1) \rfloor$, the Duplicator can choose in a linear order with size different from t. If there is an element a_i already chosen in $\mathcal{A}_t$, then the Duplicator must choose an element in a linear order $\mathcal{B}_k$ with b_i already chosen. The Duplicator must follow his strategy in a game on structures $\mathcal{A}_t$ and $\mathcal{B}_k$. This is possible even if $t \neq k$, since $r \leq min\{q_t^{\mathcal{A}}, q_t^{\mathcal{B}}\} + \lfloor log(t+1) \rfloor$. The case such that $q_{\geq t}^{\mathcal{A}} \neq q_{\geq t}^{\mathcal{B}}$ is analogous.

□

Corollary 25 (Remoteness on Disjoint Union of Linear Orders)**.** *Let $\mathcal{A}$ and $\mathcal{B}$ be two disjoint union of linear orders. The remoteness $rem(\mathcal{A}, \mathcal{B})$ can be computed as* $min(min\{min(q_t^{\mathcal{A}}, q_t^{\mathcal{B}}) + \lfloor log(t+1) \rfloor \mid q_t^{\mathcal{A}} \neq q_t^{\mathcal{B}}\} + 1, min\{min(q_{\geq t}^{\mathcal{A}}, q_{\geq t}^{\mathcal{B}}) + \lfloor log(t+1) \rfloor \mid q_{\geq t}^{\mathcal{A}} \neq q_{\geq t}^{\mathcal{B}}\})$.

We can represent the number of linear orders in $\mathcal{A}$ with size t and the number of linear orders in $\mathcal{A}$ with size at least t in two lists. The remoteness of disjoint union of linear orders $\mathcal{A}$ and $\mathcal{B}$ can be computed in $O(|A| + |B|)$ time. For each r, the remoteness algorithm checks whether, or not, r-small or r-large disparity occurs in the lists from $\mathcal{A}$ and $\mathcal{B}$. The algorithm returns the least value in which a r-small or r-large disparity occurs.

4 Hintikka Formulas on Classes of Structures

In this section, given a structure $\mathcal{A}$ on a class of structures $\mathcal{C}$ and r a positive integer, we define a formula $\varphi_{\mathcal{A}}^{m,\mathcal{C}}$ that describes the gametheoretic properties of $\mathcal{A}$ in any EF game $\mathcal{G}_r(\mathcal{A}, \mathcal{B})$ such that $\mathcal{A}$ and $\mathcal{B}$ belong to $\mathcal{C}$. To simplify the presentation, we define $\varphi_{\geq n} := \exists x_1 ... \exists x_n (\bigwedge_{i \neq j}^n x_i \neq x_j)$ and $\varphi_{=n} := \exists x_1 ... \exists x_n (\bigwedge_{i \neq j}^n x_i \neq x_j \wedge \forall x_{n+1} (\bigvee_{i=1}^n x_i = x_{n+1}))$.

Definition 26 (r-Hintikka Formula on Sets)**.** *Let $\mathcal{C}$ be the class of sets, $\mathcal{A}$ be a set and r a positive integer.*

$$\varphi_{\mathcal{A}}^{r,\mathcal{C}} := \begin{cases} \varphi_{=|A|}, & \textit{if } |A| < r \\ \varphi_{\geq r}, & \textit{otherwise.} \end{cases}$$

Example 27. *Let $\mathcal{A}_2$ be a set and $r = 3$.* $\varphi_{\mathcal{A}}^{r,\mathcal{C}} = \varphi_{=2}$.

Given a set $\mathcal{A}$, a positive integer r, and the class of sets $\mathcal{C}$, the size of $\varphi_{\mathcal{A}}^{r,\mathcal{C}}$ is $O(|A|^2)$.

Now, consider $\mathcal{C}$ as the class of linear orders. The following formula describes the gametheoretic properties of a linear order $\mathcal{A}$ in any r-round of the EF game on linear orders.

Definition 28 (r-Hintikka Formula on Linear Orders)**.** *Let $\mathcal{C}$ be the class of linear orders, $\mathcal{A}$ be a linear order and r a positive integer.*

$$\varphi_{\mathcal{A}}^{r,\mathcal{C}} := \begin{cases} \varphi_{=|A|}, & \textit{if } |A| < 2^r - 1 \\ \varphi_{\geq 2^r - 1}, & \textit{otherwise.} \end{cases}$$

Example 29. *Let $\mathcal{A}_9$ be a linear order and $r = 3$. Thus,* $\varphi_{\mathcal{A}_9}^{r,\mathcal{C}} = \varphi_{\geq 7}$.

For the class of linear orders $\mathcal{C}$, given a linear order $\mathcal{A}$ and a positive integer r, the size of $\varphi_{\mathcal{A}}^{r,\mathcal{C}}$ is $O(|A|^2)$ since $2^r - 1$ is bounded by $|A|$.

In what follows, let $\tau = \{P_1, ..., P_k\}$ be a vocabulary such that each P_i is unary, for $i \in \{1, ..., k\}$. Let $\mathcal{C}$ be the class of τ-structures. We can assume $P_1, ..., P_k$ pairwise disjoint. Let $\varphi_{|P_i|=n}$, for $i \in \{1, ..., k\}$, be a formula describing that the number of elements in P_i is n, i.e., $\varphi_{|P_i|=n} := \exists x_1 ... \exists x_n (\bigwedge_{l \neq j}^n x_l \neq x_j \wedge \bigwedge_{l=1}^n P_i(x_l) \wedge \forall x_{n+1} (P_i(x_{n+1}) \rightarrow \bigvee_{l=1}^n x_l = x_{n+1}))$ and let $\varphi_{|P_i| \geq n}$, for $i \in \{1, ..., k\}$, be a formula expressing that the number of elements in P_i is at least n, i.e., $\varphi_{|P_i| \geq n} := \exists x_1 ... \exists x_n (\bigwedge_{l \neq j}^n x_l \neq x_j \wedge \bigwedge_{l=1}^n P_i(x_l))$.

Definition 30 (r-Hintikka Formula on Unary Relation Structures)**.** *Let $\mathcal{C}$ be the class of unary relation structures, $\mathcal{A}$ be a unary relation structure and r a positive integer. We define* $\varphi_{\mathcal{A}}^{r,\mathcal{C}} := \bigwedge_{i=1}^k \varphi_{\mathcal{A}}^{r,i}$, *such that*

$$\varphi_{\mathcal{A}}^{r,i} := \begin{cases} \varphi_{|P_i|=|P_i^{\mathcal{A}}|}, & \textit{if } |P_i^{\mathcal{A}}| < r \\ \varphi_{|P_i| \geq r}, & \textit{otherwise.} \end{cases}$$

Example 31. *Let $\mathcal{A} = (\{1, 2, 3\}, P_1, P_2)$ be a unary relation structure such that $P_1 = \{1, 2\}$ and $P_2 = \{3\}$. Hence,* $\varphi_{\mathcal{A}}^2 = \varphi_{|P_1| \geq 2} \wedge \varphi_{|P_2|=1}$.

For the class of unary relation structures $\mathcal{C}$, given a unary relation structure $\mathcal{A}$ and a positive integer r, the size of $\varphi_{\mathcal{A}}^{r,\mathcal{C}}$ is $O(k \times |A|^2)$ such that k is constant. Thus, the size of $\varphi_{\mathcal{A}}^{r,\mathcal{C}}$ is polynomial in the size of $\mathcal{A}$.

The r-Hintikka formulas on disjoint union of linear orders are analogous to r-Hintikka formulas on equivalence structures. We omit the case for equivalence structures.

For the case in which $\mathcal{C}$ is the class of disjoint union of linear orders, we need a formula $\varphi_{<, q_t=n}$ defining that the number q_t of linear orders with size t in a disjoint union of linear orders is n. First, we introduce a formula $\varphi_{R,n,t} := \bigwedge_{l \neq j} x_l \neq x_j \wedge (\bigwedge_{k=1}^n (\bigwedge_{j=2}^t R(x_{1k}, x_{jk}) \wedge \forall z (R(x_{1k}, z) \rightarrow (\bigvee_{j=1}^t z = x_{jk}))))$ stating that $x_{11}, ..., x_{t1}, ..., x_{1n}, ..., x_{tn}$ are pairwise different and, for $l \in \{1, ..., n\}$, x_{1l} is related by a binary relation R to $x_{2l}, ..., x_{tl}$ and to no other element. We also define

$$\gamma_{R,n,t} := \begin{array}{l} \forall z_1 ... \forall z_t ((\bigwedge_{l \neq j} z_l \neq z_j \wedge \bigwedge_{j=2}^t R(z_1, z_j) \wedge \\ \forall z_{t+1} (R(z_1, z_{t+1}) \rightarrow \bigvee_{j=1}^t z_j = z_{t+1})) \rightarrow \\ (\bigvee_{k=1}^n \bigvee_{j=1}^t z_1 = x_{jk})). \end{array}$$

stating that any tuple of exactly t elements related by R is a single tuple of $x_{1l}, ..., x_{tl}$, for $l \in \{1, ..., n\}$. Now, we can define $\varphi_{<,q_t=n} := \exists x_{11} ... \exists x_{t1} ... \exists x_{1n} ... \exists x_{tn}(\varphi_{<,n,t} \wedge \gamma_{<,n,t})$.

We also need a formula $\varphi_{<,q_t \geq p}$ describing that the number q_t of linear orders with size t in a disjoint union of linear orders is at least p:

$$\varphi_{<,q_t \geq p} := \exists x_{11} ... \exists x_{t1} ... \exists x_{1p} ... \exists x_{tp} \varphi_{<,p,t}.$$

Definition 32 ($(<, r)$-Small Disparity Formula)**.** *Let $\mathcal{A}$ be a disjoint union of linear orders, and r a positive integer. We define $\varphi^r_{\mathcal{A},<,small} := \bigwedge_{t=1}^{2^r-2} \varphi^{r,t}_{\mathcal{A},<}$, such that*

$$\varphi^{r,t}_{\mathcal{A},<} := \begin{cases} \varphi_{<,q_t=q_t^{\mathcal{A}}}, & \text{if } q_t^{\mathcal{A}} + \lfloor log(t+1) \rfloor < r \\ \varphi_{<,q_t \geq r-\lfloor log(t+1) \rfloor}, & \text{otherwise.} \end{cases}$$

Example 33. *Let $\mathcal{A}$ be a disjoint union of linear orders such that $\mathcal{A} = \mathcal{A}_2 \uplus \mathcal{A}_3 \uplus \mathcal{A}_3$. Thus, $\varphi^3_{\mathcal{A}} = \varphi_{<,q_1=0} \wedge \varphi_{<,q_2=1} \wedge \varphi_{<,q_3 \geq 1} \wedge \varphi_{<,q_4=0} \wedge \varphi_{<,q_5=0} \wedge \varphi_{<,q_6=0}$.*

The size of $\varphi^r_{\mathcal{A},<,small}$ is $O(2^r \times |A|^2)$. The size of $\varphi^r_{\mathcal{A},<,small}$ is $O(|A|^3)$ since 2^r is bounded by $|A|$.

Lemma 34. *Let $\mathcal{A}$ and $\mathcal{B}$ be two disjoint union of linear orders, and r a positive integer. $(\mathcal{A}, \mathcal{B})$ does not have a $(<, r)$-small disparity if and only if $\mathcal{B} \models \varphi^r_{\mathcal{A},<,small}$.*

Proof. By Definition 18, it suffices to show that $\mathcal{B} \not\models \varphi^r_{\mathcal{A},E,small}$ iff there is a $t < 2^r - 1$ such that $q_t^{\mathcal{A}} \neq q_t^{\mathcal{B}}$ and $r > min\{q_t^{\mathcal{A}}, q_t^{\mathcal{B}}\} + \lfloor log_2(t+1) \rfloor$.

($\Rightarrow$) Assume that $\mathcal{B} \not\models \varphi^r_{\mathcal{A},<,small}$, that is, there is a t such that $\mathcal{B} \not\models \varphi^{r,t}_{\mathcal{A},<}$. We have two cases depending on this particular t:

1) $\mathcal{B} \not\models \varphi_{<,q_t=q_t^{\mathcal{A}}}$, that is, $q_t^{\mathcal{A}} + \lfloor log(t+1) \rfloor < r$. Thus, $q_t^{\mathcal{A}} \neq q_t^{\mathcal{B}}$. Therefore, $q_t^{\mathcal{A}} \neq q_t^{\mathcal{B}}$ and $r > min\{q_t^{\mathcal{A}}, q_t^{\mathcal{B}}\} + \lfloor log(t+1) \rfloor$.

2) $\mathcal{B} \not\models \varphi_{<,q_t \geq r-\lfloor log(t+1) \rfloor}$. Then $q_t^{\mathcal{B}} + \lfloor log(t+1) \rfloor < r \leq q_t^{\mathcal{A}} + \lfloor log(t+1) \rfloor$. Therefore, $q_t^{\mathcal{A}} \neq q_t^{\mathcal{B}}$ and $r > min\{q_t^{\mathcal{A}}, q_t^{\mathcal{B}}\} + \lfloor log(t+1) \rfloor$.

($\Leftarrow$) Suppose $\mathcal{B} \models \varphi^r_{\mathcal{A},<,small}$. Assume that there is a t such that $t < 2^r - 1$, $q_t^{\mathcal{A}} \neq q_t^{\mathcal{B}}$ and $r > min\{q_t^{\mathcal{A}}, q_t^{\mathcal{B}}\} + \lfloor log(t+1) \rfloor$. We have two cases depending on this particular t:

1) $\mathcal{B} \models \varphi_{<,q_t=q_t^{\mathcal{A}}}$. Thus, $q_t^{\mathcal{A}} = q_t^{\mathcal{B}}$, which contradicts our assumption.

2) $\mathcal{B} \models \varphi_{<,q_t \geq r-\lfloor log(t+1) \rfloor}$. Hence, $r \leq q_t^{\mathcal{B}} + \lfloor log(t+1) \rfloor$ and $r \leq q_t^{\mathcal{A}} + \lfloor log(t+1) \rfloor$. Therefore, $r \leq min\{q_t^{\mathcal{A}}, q_t^{\mathcal{B}}\} + \lfloor log(t+1) \rfloor$ and we reach to a contradiction. $\square$

Now, we handle formulas defining the $(<, r)$-large disparity. First, we define

$$\varphi_{R,\geq n,t} := \bigwedge_{l \neq j} x_l \neq x_j \wedge (\bigwedge_{k=1}^{n} (\bigwedge_{j=2}^{t} R(x_{1k}, x_{jk}))),$$

stating that, for $l \in \{1, ..., n\}$, x_{1l} is related to $x_{2l}, ..., x_{tl}$ by relation R. Next, we define

$$\gamma_{R,\geq n,t} := \begin{array}{l} \forall z_1 ... \forall z_t (\bigwedge_{l \neq j} z_l \neq z_j \wedge \\ (\bigwedge_{j=2}^{t} R(z_1, z_j)) \rightarrow (\bigvee_{k=1}^{n} \bigvee_{j=1}^{t} z_1 = x_{jk})) \end{array}$$

as a formula expressing that any linear order of size t is a linear order with elements in $x_{1l}, ..., x_{tl}$, for $l \in \{1, ..., n\}$. Now, we can define

$$\varphi_{R,q_{\geq t}=n} := \exists x_{11} ... \exists x_{t1} ... \exists x_{1n} ... \exists x_{tn}(\varphi_{R,\geq n,t} \wedge \gamma_{R,\geq n,t}),$$
$$\varphi_{R,q_{\geq t} \geq p} := \exists x_{11} ... \exists x_{t1} ... \exists x_{1p} ... \exists x_{tp} \varphi_{R,\geq n,t}.$$

Definition 35 ($(<, r)$-Large Disparity Formula)**.** *Let $\mathcal{A}$ be a disjoint union of linear orders, and r a positive integer. We define $\varphi^r_{\mathcal{A},<,large} := \bigwedge_{t=1}^{2^r-2} \psi^{r,t}_{\mathcal{A},<}$, such that*

$$\psi^{r,t}_{\mathcal{A},<} := \begin{cases} \varphi_{<,q_{\geq t}=q^{\mathcal{A}}_{\geq t}}, & \text{if } q^{\mathcal{A}}_{\geq t} + \lfloor log(t+1) \rfloor \leq r \\ \varphi_{<,q_{\geq t} \geq r-\lfloor log(t+1) \rfloor}, & \text{otherwise.} \end{cases}$$

Lemma 36. *Let $\mathcal{A}$ and $\mathcal{B}$ be two disjoint union of linear orders, and r a positive integer. $(\mathcal{A}, \mathcal{B})$ does not have a r-large disparity iff $\mathcal{B} \models \varphi^r_{\mathcal{A},<,large}$.*

The proof of this lemma is analogous of the proof of Lemma 34.

Now, we introduce Hintikka formulas on the class of disjoint union of linear orders.

Definition 37 (r-Hintikka Formulas on Disjoint Union of Linear Orders)**.** *Let $\mathcal{C}$ be the class of disjoint union of linear order, $\mathcal{A}$ be a disjoint union of linear order, and r a positive integer. $\varphi^{r,\mathcal{C}}_{\mathcal{A}} := \varphi^r_{\mathcal{A},<,small} \wedge \varphi^r_{\mathcal{A},<,large}$.*

Theorem 38. *Let $\mathcal{C}$ be the class of disjoint union of linear orders, $\mathcal{A}$ and $\mathcal{B}$ two disjoint union of linear orders, and r a positive integer. The Duplicator has a winning strategy in $\mathcal{G}_r(\mathcal{A}, \mathcal{B})$ if and only if $\mathcal{B} \models \varphi^{r,\mathcal{C}}_{\mathcal{A}}$.*

Proof. The Duplicator has a winning strategy in $\mathcal{G}_r(\mathcal{A}, \mathcal{B})$ if and only if $(\mathcal{A}, \mathcal{B})$ has neither $(<, r)$-small disparity nor $(<, r)$-large disparity by Theorem 24. By Lemma 19 and Lemma 22, $(\mathcal{A}, \mathcal{B})$ has neither $(<, r)$-small disparity nor $(<, r)$-large disparity iff $\mathcal{B} \models \varphi^r_{\mathcal{A},<,small}$ and $\mathcal{B} \models \varphi^r_{\mathcal{A},<,large}$. Therefore, the Duplicator has a winning strategy in $\mathcal{G}_r(\mathcal{A}, \mathcal{B})$ iff $\mathcal{B} \models \varphi^{r,\mathcal{C}}_{\mathcal{A}}$. $\square$

5 Distinguishability of Structures Relative to a Class of Structures

Now we define the Distinguishability Problem and we show how to solve it in polynomial time.

Definition 39 (The Distinguishability Problem)**.** *Let $\mathcal{L}$ be a logic and $\mathcal{C}$ be a class of structures. Given two sets of structures $P = \{\mathcal{A}_1, ..., \mathcal{A}_k\}$ and $N = \{\mathcal{B}_1, ..., \mathcal{B}_l\}$ such that each structure belongs to $\mathcal{C}$, find a formula φ in $\mathcal{L}$ such that $\mathcal{A} \models \varphi$ for all $\mathcal{A} \in P$ and $\mathcal{B} \not\models \varphi$ for all $\mathcal{B} \in N$.*

The decision version of the above problem is hard, $PSPACE$-complete [Pezzoli, 1999], for first-order logic, the class of all structures, and unary sets P and N. In the following, we show that this problem is solvable in polynomial time for first-order logic when the class of structures $\mathcal{C}$ is one of the classes of structures we have considered.

Definition 40 (Remoteness Between Sets of Structures)**.** *Let P and N be sets of structures beloging to a class of structures. The remoteness between P and N, denoted by $rem(P, N)$ is $max\{rem(\mathcal{A}, \mathcal{B}) \mid \mathcal{A} \in P \text{ and } \mathcal{B} \in N\}$.*

Let P and N be sets of structures beloging to a class of structures $\mathcal{C}$. The complexity to compute $rem(P, N)$ is $O(|P| \times |N| \times n^c)$ such that n^c depends on $\mathcal{C}$.

Definition 41 (The Distinguishability Sentence)**.** *Let P and N be sets of structures in a class $\mathcal{C}$ of structures. The distinguishability sencence between P and N on $\mathcal{C}$, denoted by $\varphi^{\mathcal{C}}_{P,N}$ is $\bigvee\{\varphi^{rem(P,N),\mathcal{C}}_{\mathcal{A}} \mid \mathcal{A} \in P\}$.*

Example 42. *Let $\mathcal{C}$ be the class of unary relation structures. A course can be represented by a unary relation structure $\mathcal{A} = \langle A, P_1^{\mathcal{A}} \rangle$ where A is the domain that represents the students attending the course and $P_1^{\mathcal{A}}$ is the set of women taking the course. Let $P = \{\mathcal{A}_1, \mathcal{A}_2\}$ and $N = \{\mathcal{B}_1, \mathcal{B}_2\}$ where $|A_1| = 12$, $|P_1^{\mathcal{A}_1}| = 2$, $|A_2| = 9$, $|P_1^{\mathcal{A}_2}| = 1$, $|B_1| = 13$, $|P_1^{\mathcal{B}_1}| = 9$, $|B_2| = 14$, and $|P_1^{\mathcal{B}_2}| = 12$. Then $rem(P, N) = 3$ and $\varphi_{P,N}^{\mathcal{C}} = (\varphi_{|P_1|=2} \wedge \varphi_{|P_2|\geq 3}) \vee (\varphi_{|P_1|=1} \wedge \varphi_{|P_2|\geq 3})$.*

Theorem 43 (Distinguishability Relative to a Class of Structures)**.** *Let $\mathcal{C}$ be one of the classes of structures we have considered. Let P and N be sets of structures such that each structure belongs to $\mathcal{C}$. For all $\mathcal{A} \in P$, $\mathcal{A} \models \varphi_{P,N}^{\mathcal{C}}$ and for all $\mathcal{B} \in N$, $\mathcal{B} \not\models \varphi_{P,N}^{\mathcal{C}}$. Besides, the size of $\varphi_{P,N}^{\mathcal{C}}$ is polynomial in the size of P and the large structure therein.*

Proof. Let $\mathcal{A}$ be a structure in P. Thus, $\varphi_{P,N}^{\mathcal{C}}$ has a disjunct $\varphi_{\mathcal{A}}^{rem(P,N),\mathcal{C}}$. It is easy to see that $\mathcal{A} \models \varphi_{\mathcal{A}}^{rem(P,N),\mathcal{C}}$.
Let $\mathcal{B}$ be a structure in N. Suppose $\mathcal{B} \models \varphi_{\mathcal{A}}^{rem(P,N),\mathcal{C}}$ for a disjunct in $\varphi_{P,N}$. We have $rem(\mathcal{A}, \mathcal{B}) \leq rem(P, N)$. If the Spoiler has a winning strategy in $\mathcal{G}_{rem(\mathcal{A},\mathcal{B})}(\mathcal{A}, \mathcal{B})$, then he has a winning strategy in $\mathcal{G}_{rem(P,N)}(\mathcal{A}, \mathcal{B})$. Therefore, $\mathcal{B} \not\models \varphi_{\mathcal{A}}^{rem(P,N),\mathcal{C}}$, which contradicts the fact that $\mathcal{B} \models \varphi_{\mathcal{A}}^{rem(P,N),\mathcal{C}}$. The size of $\varphi_{P,N}^{\mathcal{C}}$ is $O(|P| \times n^c)$ such that n^c depends on the class $\mathcal{C}$ and on the size of the larger structure in P. □

6 An Application in the Elementary Blocks World

In machine learning, classification is the task that, given a set of categorized examples, one searches for a hypothesis identifying to which category a new example belongs. The distinguishability problem, in Definition 39, can be seen as a classification problem in which hypothesis are formulas in a logic $\mathcal{L}$, and examples are structures beloging to a class $\mathcal{C}$ and categorized in P (for Positive) or N (for Negative).

The distinguishability framework differs from Inductive Logic Programming (ILP) [Muggleton, 1991; Muggleton *et al.*, 1992; Lavrac and Dzeroski, 1994] in that the latter uses logic programming as a uniform representation for examples, background knowledge and hypotheses, while the former uses formulas for hypotheses, relational structures for examples and there is no explicit background knowledge.

In what follows, we show an application of the distinguishability of disjoint union of linear orders in learning first-order formulas that classifies states of the elementary blocks world. The elementary blocks world [Gupta and Nau, 1992; Sierra-Santibáñez, 1998] consists of a set of cubic blocks, with the same size and color, sitting on a table. A robot can pick up a block and moves it to another position, either onto the table or on the top of some other block.

Let $\mathcal{C}$ be the class of disjoint union of linear orders. The formula $\varphi_{P,N}^{\mathcal{C}}$ specifies a description of states in the elementary blocks world. Let $\mathcal{A} = \mathcal{A}_3$, $\mathcal{B} = \mathcal{B}_1 \uplus \mathcal{B}_1 \uplus \mathcal{B}_1$, and $\mathcal{B}' = \mathcal{B}_1 \uplus \mathcal{B}_2$ representing the states of the blocks world, as in Figure 1, such that $\mathcal{A}$ represents the goal state.

Let $P = \{\mathcal{A}\}$ and $N = \{\mathcal{B}, \mathcal{B}'\}$. Therefore $rem(P, N) = 3$ and $\varphi_{P,N}^{\mathcal{C}} = \varphi_{<,q_1=0} \wedge \varphi_{<,q_2=0} \wedge \varphi_{<,q_{\geq 1}=1} \wedge \varphi_{<,q_{\geq 2}=1}$. See that $\mathcal{A} \models \varphi_{P,N}^{\mathcal{C}}$, $\mathcal{B} \not\models \varphi_{P,N}^{\mathcal{C}}$ and $\mathcal{B}' \not\models \varphi_{P,N}^{\mathcal{C}}$.

In a planning framework with an initial, intermediaries and final states, we may use formulas like $\varphi_{P,N}^{\mathcal{C}}$ to characterize these states. We aim to apply this logical setting to discover the moves that link states, and to reason about planning [Ghallab *et al.*, 2004].

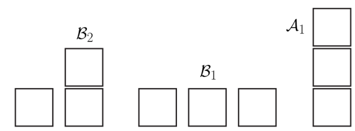

Figure 1: States of the Blocks World

7 Conclusions and Future Work

Motivated by the algorithm defined in [Kaiser, 2012], this work presents distinguishability algorithms on particular classes of structures. These algorithms run in polynomial time and return a sentence φ in first-order logic distinguishing two sets of structures, the positive and the negative ones, on a class of structures. The algorithm in [Kaiser, 2012] runs in exponential time for arbitrary finite relational structures.

Our approach consists of computing the remoteness between the given two sets of structures P and N. We compute the Hintikka formulas for all structures in P. The Hintikka formula is obtained from structures $\mathcal{A} \in P$ and the remoteness between P and N.

We used results characterising the winning strategies for the Duplicator on a particular class of structures [Libkin, 2004; Khoussainov and Liu, 2009]. We also showed our result of necessary and suficient condition for a winning strategy of the Duplicator on the class of disjoint linear orders. We need these results to derive procedures to compute Hintikka formulas on classes of structures. Besides, this procedure runs in polynomial time in the size of a given structure.

As future work, we want to investigate distinguishability algorithms on extensions of the classes considered. For example, equivalence structures with colors, and embedded equivalence structures [Khoussainov and Liu, 2009].

Besides, we intend to consider distinguishability on the class of strings [Montanari *et al.*, 2005; Maria *et al.*, 2009]. This kind of result is interesting because first-order logic on the class of strings captures star-free languages [Thomas, 1997]. A distinguishability algorithm on the class of strings can be used to derive a recognizer from positive and negative strings.

Finally, we plan to examine distinguishability with other logics as monadic second order logic. On the class of strings, monadic second order logic captures regular languages [Büchi, 1960; Ladner, 1977]. A distinguishability algorithm returning a sentence in monadic second order logic that distinguishes positive and negative strings can be used in automata learning [Grinchtein *et al.*, 2006; Heule and Verwer, 2010; Neider and Jansen, 2013].

References

[Büchi, 1960] Richard J. Büchi. Weak second-order arithmetic and finite automata. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 6(1-6):66–92, 1960.

[Cook and Liu, 2003] Stephen A. Cook and Yongmei Liu. A complete axiomatization for blocks world. *Journal of Logic and Computation*, 13(4):581–594, 2003.

[Ebbinghaus and Flum, 1995] Heinz-Dieter Ebbinghaus and Jörg Flum. *Finite model theory*. Perspectives in Mathematical Logic. Springer, 1995.

[Ehrenfeucht, 1961] Andrzej Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fundamenta Mathematicae*, 49(2):129–141, 1961.

[Ghallab *et al.*, 2004] Malik Ghallab, Dana Nau, and Paolo Traverso. *Automated Planning: theory and practice*. Elsevier, 2004.

[Grädel *et al.*, 2005] Erich Grädel, P. G. Kolaitis, L. Libkin, M. Marx, J. Spencer, Moshe Y. Vardi, Y. Venema, and Scott Weinstein. *Finite Model Theory and Its Applications (Texts in Theoretical Computer Science. An EATCS Series)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[Grinchtein *et al.*, 2006] Olga Grinchtein, Martin Leucker, and Nir Piterman. *Inferring Network Invariants Automatically*, pages 483–497. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[Gupta and Nau, 1992] Naresh Gupta and Dana S. Nau. On the complexity of blocks-world planning. *Artif. Intell.*, 56(2-3):223–254, 1992.

[Heule and Verwer, 2010] Marijn JH Heule and Sicco Verwer. Exact dfa identification using sat solvers. In *International Colloquium on Grammatical Inference*, pages 66–79. Springer, 2010.

[Jordan and Kaiser, 2013a] Charles Jordan and Lukasz Kaiser. Benchmarks from reduction finding. In *International Workshop on Quantified Boolean Formulas 2013, Informal Workshop Report*, page 40, 2013.

[Jordan and Kaiser, 2013b] Charles Jordan and Lukasz Kaiser. Experiments with reduction finding. In Matti Järvisalo and Allen Van Gelder, editors, *SAT*, volume 7962 of *Lecture Notes in Computer Science*, pages 192–207. Springer, 2013.

[Jordan and Kaiser, 2013c] Charles Jordan and Lukasz Kaiser. Learning programs as logical queries. In *Workshop on Learning Theory and Complexity, LTC*. Citeseer, 2013.

[Kaiser, 2012] Lukasz Kaiser. Learning games from videos guided by descriptive complexity. In *AAAI*, 2012.

[Khoussainov and Liu, 2009] Bakhadyr Khoussainov and Jiamou Liu. On complexity of ehrenfeucht–fraïssé games. *Annals of Pure and Applied Logic*, 161(3):404 – 415, 2009.

[Ladner, 1977] Richard E. Ladner. Application of model theoretic games to discrete linear orders and finite automata. *Information and Control*, 33(4):281 – 303, 1977.

[Lavrac and Dzeroski, 1994] Nada Lavrac and Saso Dzeroski. Inductive logic programming. In *WLP*, pages 146–160. Springer, 1994.

[Libkin, 2004] Leonid Libkin. *Elements Of Finite Model Theory (Texts in Theoretical Computer Science. An Eatcs Series)*. SpringerVerlag, 2004.

[Maria *et al.*, 2009] Elisabetta De Maria, Angelo Montanari, and Nicola Vitacolonna. Games on strings with a limited order relation. In *Logical Foundations of Computer Science, International Symposium, LFCS 2009, Deerfield Beach, FL, USA, January 3-6, 2009. Proceedings*, pages 164–179, 2009.

[Montanari *et al.*, 2005] Angelo Montanari, Alberto Policriti, and Nicola Vitacolonna. An algorithmic account of ehrenfeucht games on labeled successor structures. In Geoff Sutcliffe and Andrei Voronkov, editors, *Logic for Programming, Artificial Intelligence, and Reasoning*, volume 3835 of *Lecture Notes in Computer Science*, pages 139–153. Springer Berlin Heidelberg, 2005.

[Muggleton *et al.*, 1992] Stephen Muggleton, Ramon Otero, and Alireza Tamaddoni-Nezhad. *Inductive logic programming*, volume 38. Springer, 1992.

[Muggleton, 1991] Stephen Muggleton. Inductive logic programming. *New generation computing*, 8(4):295–318, 1991.

[Neider and Jansen, 2013] Daniel Neider and Nils Jansen. *Regular Model Checking Using Solver Technologies and Automata Learning*, pages 16–31. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[Papadimitriou, 1994] Christos M. Papadimitriou. *Computational Complexity*. Addison-Wesley, 1994.

[Pezzoli, 1999] Elena Pezzoli. Computational complexity of ehrenfeucht-fraïssé games on finite structures. In Georg Gottlob, Etienne Grandjean, and Katrin Seyr, editors, *Computer Science Logic*, volume 1584 of *Lecture Notes in Computer Science*, pages 159–170. Springer Berlin Heidelberg, 1999.

[Sierra-Santibáñez, 1998] Josefina Sierra-Santibáñez. Declarative formalization of strategies for action selection. In *Proceedings 7th Workshop on Nonmonotonic Reasoning (NMR-2000), Action and Causality track, Trento, Italy*, 1998.

[Thomas, 1997] Wolfgang Thomas. Handbook of formal languages, vol. 3. chapter Languages, Automata, and Logic, pages 389–455. Springer-Verlag New York, Inc., New York, NY, USA, 1997.