



Predicting human immunodeficiency virus inhibitors using multi-dimensional Bayesian network classifiers

Hanen Borchani^{a,*}, Concha Bielza^a, Carlos Toro^b, Pedro Larrañaga^a

^a Computational Intelligence Group, Departamento de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, Boadilla del Monte 28660, Spain

^b Department of Microbiology, Hospital Carlos III, Madrid 28029, Spain

ARTICLE INFO

Article history:

Received 26 October 2011

Received in revised form

14 December 2012

Accepted 16 December 2012

Keywords:

Multi-dimensional Bayesian network classifiers

Markov blanket

Human immunodeficiency virus

Protease inhibitors

Reverse transcriptase inhibitors

ABSTRACT

Objective: Our aim is to use multi-dimensional Bayesian network classifiers in order to predict the human immunodeficiency virus type 1 (HIV-1) reverse transcriptase and protease inhibitors given an input set of respective resistance mutations that an HIV patient carries.

Materials and methods: Multi-dimensional Bayesian network classifiers (MBCs) are probabilistic graphical models especially designed to solve multi-dimensional classification problems, where each input instance in the data set has to be assigned simultaneously to multiple output class variables that are not necessarily binary. In this paper, we introduce a new method, named $MB-MBC$, for learning MBCs from data by determining the Markov blanket around each class variable using the HITON algorithm. Our method is applied to both reverse transcriptase and protease data sets obtained from the Stanford HIV-1 database.

Results: Regarding the prediction of antiretroviral combination therapies, the experimental study shows promising results in terms of classification accuracy compared with state-of-the-art MBC learning algorithms. For reverse transcriptase inhibitors, we get 71% and 11% in mean and global accuracy, respectively; while for protease inhibitors, we get more than 84% and 31% in mean and global accuracy, respectively. In addition, the analysis of MBC graphical structures lets us gain insight into both known and novel interactions between reverse transcriptase and protease inhibitors and their respective resistance mutations.

Conclusion: $MB-MBC$ algorithm is a valuable tool to analyze the HIV-1 reverse transcriptase and protease inhibitors prediction problem and to discover interactions within and between these two classes of inhibitors.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The multi-dimensional classification problem is an extension of the classical one-dimensional classification problem, where we have to deal with multiple output class variables rather than a single output class variable [1]. Formally, the multi-dimensional classification problem consists of finding a function f that predicts for each input instance, given by a vector of m features $\mathbf{x} = (x_1, \dots, x_m)$, a vector of d class values $\mathbf{c} = (c_1, \dots, c_d)$:

$$f: \Omega_{X_1} \times \dots \times \Omega_{X_m} \longrightarrow \Omega_{C_1} \times \dots \times \Omega_{C_d}$$

$$\mathbf{x} = (x_1, \dots, x_m) \mapsto \mathbf{c} = (c_1, \dots, c_d)$$

where Ω_{C_i} and Ω_{X_j} denote the sample spaces of each class variable C_i , for all $i \in \{1, \dots, d\}$, and each feature variable X_j , for all $j \in \{1, \dots, m\}$, respectively. Note that, we consider that all class and feature variables are discrete random variables such that $|\Omega_{C_i}|$ and $|\Omega_{X_j}|$ are greater than 1.

When $|\Omega_{C_i}| = 2$ for all $i \in \{1, \dots, d\}$, i.e., all class variables are binary, the multi-dimensional classification problem is known as a multi-label classification problem [2,3]. In general, a multi-label classification problem can be easily modeled as a multi-dimensional classification problem where each label corresponds to a binary class variable. However, modeling a multi-dimensional classification problem, that possibly includes non-binary class variables, as a multi-label classification problem may require a transformation over the data set to meet multi-label framework requirements.

In recent years, the concept of multi-dimensionality has been introduced in Bayesian network classifiers providing an accurate modeling of this emerging problem and ensuring interactions among all variables [1,4–8]. In these probabilistic graphical models, known as multi-dimensional Bayesian network classifiers (MBCs),

* Corresponding author. Tel.: +34 91 3363675; fax: +34 91 3524819.

E-mail addresses: hanen.borchani@upm.es (H. Borchani), mcbielza@fi.upm.es (C. Bielza), carlos.toro@salud.madrid.org (C. Toro), pedro.larranaga@fi.upm.es (P. Larrañaga).

the graphical structure partitions the set of class and feature variables into three different subgraphs: class subgraph, feature subgraph and bridge subgraph, and the parameter set defines the conditional probability distribution of each variable given its parents.

In this paper, we introduce a novel MBC learning algorithm based on Markov blankets. Motivated by the fact that the classification is unaffected by parts of the structure that lie outside the Markov blankets of the class variables, we first build the Markov blanket around each class variable using the well-known HITON algorithm [9–11], and then we determine edge directionality over all three MBC subgraphs. Thanks to this filter and a local approach to MBC learning, we can lighten the computational burden of MBC learning using wrapper algorithms [1,4,5] and provide more accurate MBC structures.

We finally apply our Markov blanket MBC ($MB-MBC$) algorithm to the problem of predicting human immunodeficiency virus type 1 (HIV-1) reverse transcriptase and protease inhibitors given an input set of corresponding resistance mutations that an HIV patient carries. In general, a combination of several antiretroviral drugs should be repeatedly administered for each patient in order to prevent and treat the HIV infection.

We analyze both reverse transcriptase and protease data sets obtained from the Stanford HIV-1 database [12]. In the reverse transcriptase data set (respectively, protease data set), the class variables are ten reverse transcriptase inhibitors (respectively, eight protease inhibitors) and the feature variables are 38 predefined mutations [13] associated with resistance to reverse transcriptase inhibitors (respectively, 74 predefined mutations [13] associated with resistance to protease inhibitors).

In both data sets, all class and feature variables are binary, so that the problem of predicting HIV-1 reverse transcriptase and protease inhibitors can be also viewed as a multi-label classification problem. However, since our approach is general and can be applied to additional classification problems where class variables are not necessarily binary, we opt to use the term multi-dimensional classification as a more general concept. Moreover, contrary to multi-label classification methods, our approach presents the merit of explicitly modeling the relationships between all variables through their graphical structure component which, in our study, may be useful in further investigating the interactions among the different inhibitors and resistance mutations.

Experimental results on reverse transcriptase and protease inhibitors data sets were promising in terms of classification accuracy compared with state-of-the-art MBC and multi-label classification methods, as well as regarding the identification of interactions among inhibitors and resistance mutations, which were either consistent with the latest knowledge or not previously mentioned in the literature.

The remainder of this paper is organized as follows. Section 2 introduces Bayesian networks. Section 3 presents MBCs and briefly reviews state-of-the-art MBC learning algorithms. Section 4 describes our new MBC learning approach. Section 5 presents the experimental study on the HIV-1 reverse transcriptase and protease inhibitor data sets. Finally, Section 6 sums up the paper with some conclusions.

2. Background

A Bayesian network [14,15] over a set of discrete random variables $\mathbf{U} = \{X_1, \dots, X_n\}$, $n \geq 1$, is a pair $\mathcal{B} = (\mathcal{G}, \Theta)$. $\mathcal{G} = (V, A)$ is a directed acyclic graph (DAG) whose vertices V correspond to variables in \mathbf{U} and whose arcs A represent direct dependencies between the vertices. Θ is a set of conditional probability distributions such that $\theta_{x_i|\mathbf{pa}(x_i)} = p(x_i|\mathbf{pa}(x_i))$ defines the conditional probability of

each possible value x_i of X_i given a set value $\mathbf{pa}(x_i)$ of $\mathbf{Pa}(X_i)$, where $\mathbf{Pa}(X_i)$ denotes the set of parents of X_i in \mathcal{G} .

A Bayesian network \mathcal{B} represents a joint probability distribution over \mathbf{U} factorized according to structure \mathcal{G} as follows:

$$p(X_1, \dots, X_n) = \prod_{i=1}^n p(X_i|\mathbf{Pa}(X_i)). \quad (1)$$

Definition 1 (Conditional independence [14]). Two set of variables \mathbf{X} and \mathbf{Y} are conditionally independent given some set of variables \mathbf{Z} , denoted as $I(\mathbf{X}, \mathbf{Y}|\mathbf{Z})$, iff $P(\mathbf{X}|\mathbf{Y}, \mathbf{Z}) = P(\mathbf{X}|\mathbf{Z})$ for any assignment of values $\mathbf{x}, \mathbf{y}, \mathbf{z}$ of $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, respectively, such that $P(\mathbf{Z} = \mathbf{z}) > 0$.

Definition 2 (Markov blanket [14]). A Markov blanket of a variable X , denoted as $MB(X)$, is a minimal set of variables with the following property: $I(X, \mathbf{S} | MB(X))$ holds for every variable subset \mathbf{S} with no variables in $MB(X) \cup X$.

In other words, $MB(X)$ is a minimal set of variables conditioned by which X is conditionally independent of all the remaining variables. Under the faithfulness assumption, ensuring that all the conditional independencies in the data distribution are strictly those entailed by \mathcal{G} , $MB(X)$ consists of the union of the set of parents, children, and parents of children (i.e., spouses) of X [16].

3. Multi-dimensional Bayesian network classifiers

In this section we present MBCs, then briefly review the state-of-the-art methods for learning these models from data.

Definition 3 (Multi-dimensional Bayesian network classifier [1]). An MBC is a Bayesian network $\mathcal{B} = (\mathcal{G}, \Theta)$ where the structure $\mathcal{G} = (V, A)$ has a restricted topology. The set of n vertices V is partitioned into two sets: $V_C = \{C_1, \dots, C_d\}$, $d \geq 1$, of class variables and $V_X = \{X_1, \dots, X_m\}$, $m \geq 1$, of feature variables ($d + m = n$). The set of arcs A is partitioned into three sets A_C , A_X and A_{CX} , such that:

- $A_C \subseteq V_C \times V_C$ is composed of the arcs between the class variables having a subgraph $\mathcal{G}_C = (V_C, A_C)$ – class subgraph – of \mathcal{G} induced by V_C .
- $A_X \subseteq V_X \times V_X$ is composed of the arcs between the feature variables having a subgraph $\mathcal{G}_X = (V_X, A_X)$ – feature subgraph – of \mathcal{G} induced by V_X .
- $A_{CX} \subseteq V_C \times V_X$ is composed of the arcs from the class variables to the feature variables having a subgraph $\mathcal{G}_{CX} = (V, A_{CX})$ – bridge subgraph – of \mathcal{G} induced by V [4].

Depending on the graphical structures of the class and feature subgraphs MBCs can be divided into several families. These families can be denoted as $\langle \text{class subgraph structure} \rangle \langle \text{feature subgraph structure} \rangle$ MBCs, where the possible structures of each subgraph are: empty, tree, polytree, or DAG [4]. In this paper, we do not consider any constraints on the subgraph structures of the learned MBCs, i.e., any possible structure type is allowed for either class or feature subgraphs.

Classification with an MBC under a 0–1 loss function is equivalent to solving the most probable explanation (MPE) problem, which consists of finding the most likely instantiation of the vector of class variables $\mathbf{c}^* = (c_1^*, \dots, c_d^*)$ given an evidence about the input vector of feature variables $\mathbf{x} = (x_1, \dots, x_m)$. More formally, for a given observed evidence \mathbf{x} , we have to determine

$$\mathbf{c}^* = (c_1^*, \dots, c_d^*) = \arg \max_{c_1, \dots, c_d} p(C_1 = c_1, \dots, C_d = c_d | \mathbf{x}). \quad (2)$$

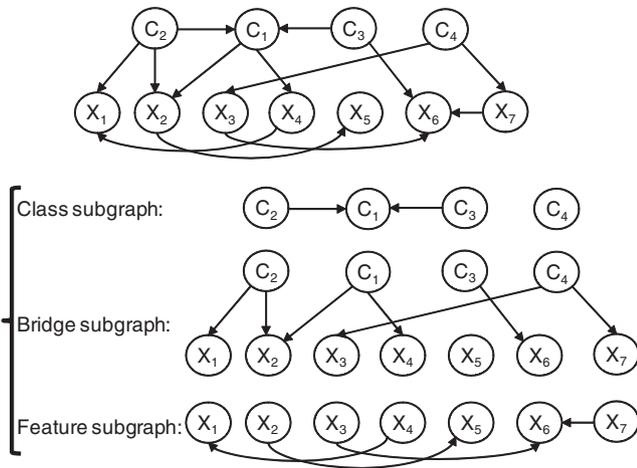


Fig. 1. An example of a multi-dimensional Bayesian network classifier structure with its class, bridge and feature subgraphs.

Example 1. An example of an MBC structure is shown in Fig. 1. V_C contains four classes, V_X includes seven features, and the structure \mathcal{G} is equal to $\mathcal{G}_C \cup \mathcal{G}_X \cup \mathcal{G}_{C,X}$. We have

$$\begin{aligned} \max_{c_1, \dots, c_4} p(C_1 = c_1, \dots, C_4 = c_4 | \mathbf{x}) &\propto \max_{c_1, \dots, c_4} p(c_1 | c_2, c_3) p(c_2) p(c_3) p(c_4) \cdot \\ &p(x_1 | c_2, x_4) p(x_2 | c_1, c_2) p(x_3 | c_4) p(x_4 | c_1) \cdot \\ &p(x_5 | x_2) p(x_6 | c_3, x_3, x_7) p(x_7 | c_4). \end{aligned}$$

Several approaches have been recently proposed to learn MBCs from data. In [1], van der Gaag and de Waal use Chow and Liu’s algorithm [17] to learn the class and feature subgraphs of a $\langle \text{tree} \rangle$ - $\langle \text{tree} \rangle$ MBC, then they greedily select the bridge subgraph, using a wrapper method, aiming to induce the most accurate classifier. De Waal and van der Gaag later present a theoretical approach for learning $\langle \text{polytree} \rangle$ - $\langle \text{polytree} \rangle$ MBCs in [6]. Class and feature subgraphs are separately generated using Rebane and Pearl’s algorithm [18]; however, the induction of the bridge subgraph was not specified.

Zaragoza et al. [19] propose a two-step method to also learn $\langle \text{polytree} \rangle$ - $\langle \text{polytree} \rangle$ MBCs. First, they build class and feature subgraphs using Chow and Liu’s algorithm [17] and generate an initial bridge subgraph based on mutual information. Then, in a second step, they refine the bridge subgraph by adding more arcs to improve MBC accuracy.

Rodríguez and Lozano [7] use a multi-objective evolutionary approach to learn $\langle \text{DAG} \rangle$ - $\langle \text{DAG} \rangle$ MBCs. Each permitted MBC structure is coded as an individual with three substrings, one per subgraph. Based on different classification measures, joint and marginal, they define the objective functions as k -fold cross-validated estimators of each class classification error. The aim is to find non-dominated structures according to the objective functions.

Bielza et al. [4] propose three MBC learning algorithms: pure filter (guided by any filter algorithm based on a fixed ordering among the variables), pure wrapper (guided by the classification accuracy) and a hybrid algorithm (a combination of pure filter and pure wrapper). None of these three algorithms places any constraints on the subgraph structures of the generated MBCs.

In [5], we propose the CB-MBC learning algorithm for class-bridge decomposable MBCs [4], instead of general MBCs, based on a greedy forward selection wrapper approach. Class or feature

subgraphs can have any type of structure. The algorithm firstly learns an initial bridge subgraph with a number of maximal connected components equal to the number of class variables, then it learns an initial feature subgraph. Next, as long as the number of maximal connected components is greater than one and there is an accuracy improvement, it iteratively and sequentially merges together the components and updates the bridge and feature subgraphs.

More recently, Zaragoza et al. present a two-step method in [8]. In the first phase, a tree-based Bayesian network that represents the dependency relations between the class variables is learned. In the second phase, several chain classifiers are built using selective naive Bayes models, such that the order of the class variables in the chain is consistent with the class subgraph. At the end, the results of the different generated orders are combined in a final ensemble model.

4. Learning multi-dimensional Bayesian network classifiers using Markov blankets

In this section we describe a new algorithm for learning MBCs from data based on Markov blanket discovery. As far as we know, this is the first paper to propose a MBC learning algorithm relying exclusively on a constraint-based approach. Our objective is to tackle the shortcomings of our previous learning method [5], mainly its computational cost, by taking advantage of the merits of a filter constraint-based approach. This should considerably lighten the computational burden, especially when the data set includes a large number of class and feature variables, while guaranteeing good performance.

Additionally, this work is motivated by its application to the HIV drug resistance problem, where it is not only important to build an MBC with a high predictive power but also to discover the resistance pathways of each HIV drug by analyzing the MBC structure. Applying our previous learning method may not always lead to an accurate MBC structure, since arcs between features are selected at random in the feature subgraph learning steps. In fact, in each feature subgraph learning step, CB-MBC iteratively selects a random arc between features, then adds it if it improves the accuracy. This means that, in each iteration, no exhaustive search is performed to add the arc that improves the most the accuracy. Such exhaustive search is avoided since, as pointed out in [5], it may be impractical and very time-consuming. In other words, these random arcs added to the feature subgraph, though they improve in each iteration the accuracy, they may not lead to the MBC with the best accuracy. Therefore, this may affect the overall quality of the learned MBC structure and consequently lead to misinterpretations.

To deal with this issue, we make use of Markov blankets. In recent years, several specialized Markov blanket learning methods have been proposed in the literature, such as GS, TPDA, IAMB and its variants, MMHC, MMBB and HITON (see [10,11] and their references for reviews). In this paper, we only consider and adapt the HITON algorithm [9–11] extended to the context of multi-dimensional Bayesian network classifiers. In fact, the HITON algorithm was empirically proven to outperform most of the state-of-the-art Markov blanket discovery algorithms in terms of combined classification performance and feature set parsimony [10].

The idea of our Markov blanket MBC (MB-MBC) learning algorithm is simple and consists of applying the HITON algorithm to each class variable and then specifying directionality over the MBC subgraphs. HITON identifies the Markov blanket of each class variable in a two-phase scheme, HITON-MB and HITON-PC, outlined, respectively, in Algorithms 1 and 2.

Algorithm 1 HITON-MB(C_i)

```

1.  $PC(C_i) \leftarrow$  HITON-PC( $C_i$ )
2. for every variable  $T \in PC(C_i)$  do
3.    $PC(T) \leftarrow$  HITON-PC( $T$ )
4. end for
5.  $MB(C_i) \leftarrow PC(C_i)$ 
6.  $S \leftarrow (\bigcup_{T \in PC(C_i)} PC(T)) \setminus \{PC(C_i) \cup C_i\}$ 
7. for every variable  $X \in S$  do
8.   Retrieve a subset  $Z$  s.t.  $I(X, C_i|Z)$ 
9.   for every variable  $T \in PC(C_i)$  s.t.  $X \in PC(T)$  do
10.    if  $\neg I(X, C_i|Z \cup \{T\})$  then
11.      Insert  $X$  into  $MB(C_i)$ 
12.    end if
13.  end for
14. end for
15. return  $MB(C_i)$ 

```

Step 1 of HITON-MB identifies the parents and children of each class variable C_i , denoted $PC(C_i)$, by calling the HITON-PC algorithm. Then, it determines the PC set for every member T of $PC(C_i)$ (steps 2 to 4). The Markov blanket set $MB(C_i)$ is initialized with $PC(C_i)$ (step 5) and set S includes potential spouses of C_i (step 6). From steps 7 to 14, HITON-MB loops over all members of S to identify correct spouses of C_i . $MB(C_i)$ is finally returned in step 15.

HITON-PC starts with an empty set of candidates $PC(T)$, ranks the variables X in OPEN by priority of inclusion according to $I(X, T)$ and discards variables having $I(X, T) = 0$. Then, for every new variable inserted into $PC(T)$, it checks if there is any variable inside $PC(T)$ that is independent of T given some subset Z . In this case, this variable will be removed from $PC(T)$ (steps 6–11). These steps are iterated until there are no more variables in OPEN. Finally, $PC(T)$ is filtered using the symmetry criterion (steps 13–17). In fact, for every $X \in PC(T)$, the symmetrical relation holds iff $T \in PC(X)$. Otherwise, i.e., if $T \notin PC(X)$, X will be removed from $PC(T)$. At the end of this step, we obtain $PC(T)$ [10].

Algorithm 2 HITON-PC(T)

```

1.  $PC(T) \leftarrow \emptyset$ 
2. OPEN  $\leftarrow U \setminus \{T \cup PC(T)\}$ 
3. Sort the variables  $X$  in OPEN in descending order according to  $I(X, T)$ 
4. Remove from OPEN variables  $X$  having  $I(X, T) = 0$ 
5. repeat
6.   Insert at end of  $PC(T)$  the first variable in OPEN and remove it from OPEN
7.   for every variable  $X \in PC(T)$  do
8.     if  $\exists Z \subseteq PC(T) \setminus \{X\}$ , s.t.  $I(X, T|Z)$  then
9.       Remove  $X$  from  $PC(T)$ .
10.    end if
11.  end for
12. until OPEN =  $\emptyset$ 
13. for every variable  $X \in PC(T)$  do
14.  if  $T \notin PC(X)$  then
15.    Remove  $X$  from  $PC(T)$ 
16.  end if
17. end for
18. return  $PC(T)$ .

```

Note that the complexity of both algorithms could be controlled using a parameter \max_{CS} restricting the maximum number of elements in the conditioning sets Z [10]. In our experiments, we use the G^2 statistical test to evaluate the conditional independencies between variables with a threshold significance level of $\alpha = 0.05$, and we consider different values of $\max_{CS} = 1, 2, 3, 4, 5$.

Unlike the HITON algorithm that only determines the Markov blanket of a single target variable for solving the variable selection problem, our algorithm considers many target variables, then induces the MBC graphical structure. Given the MBC definition, direct parents of any class variable C_i , $i = 1, \dots, d$, can only be among the remaining class variables, whereas direct children or spouses of C_i can include either class or feature variables. We can then easily deduce the different MBC subgraphs based on the results of the HITON algorithm:

- **Class subgraph:** we firstly insert an edge between each class variable C_i and any class variable belonging to its corresponding parents–children set $PC(C_i)$. Then, we direct all these edges using the PC algorithm's edge orientation rules [20].
- **Bridge subgraph:** this is built by inserting an arc from each class variable C_i to every feature variable belonging to $PC(C_i)$.
- **Feature subgraph:** for every feature X in the set $MB(C_i) \setminus PC(C_i)$, i.e., for every spouse X , we insert an arc from X to the corresponding common child given by $PC(X) \cap PC(C_i)$. Moreover, more arcs can be added especially to discover additional dependency relationships among features. In fact, for every feature X , child of C_i , we determine the set $Y = PC(X) \setminus (\{C_i\} \cup \{MB(C_i) \cap PC(X)\})$. If $Y \neq \emptyset$, we insert an arc from X to every feature variable in Y .

Example 2. Let us assume that we apply HITON algorithm to a data set coming out of the MBC structure of Fig. 1. By the end of HITON-PC and HITON-MB algorithms, we identify, respectively, the parents–children sets and the Markov blanket sets of each class variable:

- $PC(C_1) = \{C_2, C_3, X_2, X_4\}$; $MB(C_1) = PC(C_1)$
- $PC(C_2) = \{C_1, X_1, X_2\}$; $MB(C_2) = \{C_1, C_3, X_1, X_2, X_4\}$
- $PC(C_3) = \{C_1, X_6\}$; $MB(C_3) = \{C_1, C_2, X_6, X_3, X_7\}$
- $PC(C_4) = \{X_3, X_7\}$; $MB(C_4) = PC(C_4)$

Next, we specify the three MBC subgraphs as follows:

- **Class subgraph:** edges are inserted between the class variables C_1 , C_2 and C_3 . Then, using the PC algorithm's edge orientation rules, these edges are directed from C_2 and C_3 to C_1 .
- **Bridge subgraph:** arcs are inserted from C_1 to X_2 and X_4 ; from C_2 to X_1 and X_2 ; from C_3 to X_6 ; and from C_4 to X_3 , X_5 and X_7 .
- **Feature subgraph:** given that $MB(C_2) \setminus PC(C_2) = \{X_4\}$, an arc is inserted from spouse X_4 to the common child X_1 determined by $PC(X_4) \cap PC(C_2) = \{X_1\}$. Similarly, given that $MB(C_3) \setminus PC(C_3) = \{X_3, X_7\}$ and $PC(X_3) \cap PC(C_3) = PC(X_7) \cap PC(C_3) = \{X_6\}$, arcs are inserted from X_3 and X_7 to X_6 . For additional dependency relationships among features, given that $PC(X_2) = \{C_1, C_2, X_5\}$, we determine the set $Y = PC(X_2) \setminus (\{C_1\} \cup \{MB(C_1) \cap PC(X_2)\}) = PC(X_2) \setminus (\{C_2\} \cup \{MB(C_2) \cap PC(X_2)\}) = \{C_1, C_2, X_5\} \setminus \{C_1 \cup C_2\} = \{X_5\}$. Thus, an arc is inserted from X_2 to X_5 .

5. Experimental study**5.1. Data sets**

Commonly used therapies for human immunodeficiency virus type 1 (HIV-1) are combinations or cocktails of antiretroviral drugs. Typically, these drugs may belong to one or more different drug groups that target different stages of the viral HIV-1 life cycle. In order to deal with the HIV-1 therapy prediction problem and gain insight into the different interactions between drugs and resistance mutations, we analyzed reverse transcriptase and protease data sets obtained from the online Stanford HIV-1 database [12].

5.1.1. Reverse transcriptase inhibitors

Reverse transcriptase inhibitors (RTIs) consist of two groups of antiretroviral drugs preventing HIV-1 replication, namely nucleoside and nucleotide reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs). NRTIs inhibit reverse transcription by being incorporated into the newly synthesized viral deoxyribonucleic acid (DNA) and preventing its further elongation [21]. We study seven drugs in this group: Abacavir (ABC), Didanosine (DDI), Emtricitabine (FTC), Lamivudine (3TC), Stavudine (D4T), Tenofovir (TDF), and Zidovudine (AZT).

NNRTIs inhibit reverse transcriptase directly by binding to the enzyme, restricting its mobility and making it unable to function [21]. We consider three drugs in this group: Efavirenz (EFV), Nevirapine (NVP), and Delavirdine (DLV).

We studied a data set obtained from the Stanford HIV-1 reverse transcriptase database [12] containing treatment histories from 2855 patients that received either NRTIs, NNRTIs or both. These treatment histories were collected from previously published studies and all belonged to subtype B. There may be one or multiple isolates for the same patient. Each isolate corresponds to a sample in the data set, including a list of resistance mutations and a combination of RTIs administered to a patient at a specified time point during his or her course of RTI treatment. Only samples where no drug was administered were discarded. Accordingly, the final data set contained a total of 4884 samples. Note that the number of RTIs in the administered combinations varies from 1 to 8 drugs, such that the highest number of samples comprises 5 RTIs (1852 samples) and 6 RTIs (1600 samples). There are only 698 samples including 4 RTIs, 483 samples including 7 RTIs, 157 samples including 3 RTIs, 56 samples including 8 RTIs, and finally, we have only 17 and 25 samples, respectively, for 1 and 2 RTIs.

Additionally, we considered a total of 38 mutations associated with resistance to RTIs and defined in the latest International AIDS Society-USA resistance mutation list [13]. There are no common resistance mutations between the two RTI groups; in fact, 22 mutations are associated with NNRTs and 16 mutations are associated with NNRTIs.

5.1.2. Protease inhibitors

Protease inhibitors (PIs) represent the third group of antiretroviral drugs. They bind to the protein cleavage site, and therefore prevent the enzyme from releasing the individual core proteins and virus particles from subsequently maturing as infections [22]. We considered eight PI drugs: Atazanavir (ATV), Darunavir (DRV), Fosamprenavir (FPV), Indinavir (IDV), Lopinavir (LPV), Nelfinavir (NFV), Saquinavir (SQV) and Tipranavir (TPV), and we analyzed a data set obtained from the Stanford HIV-1 protease database [12] containing antiretroviral PI treatment histories from 1255 patients. As with the RTI data set, the treatment histories were collected from previously published studies. There may be one or more patient instances in the data set such that each instance includes a list of resistance mutations and a set of administered PIs drugs. Only samples where no drug was administered were discarded.

The final data set contained a total of 4341 samples belonging mainly to subtype B (subtype B: 92%, subtype C: 3%, and other subtypes (A, D, F, G, H, J, K, CRF01_AE, CRF02_AG): 5%). Note also that the number of PI combinations is not evenly represented; in fact, there are 3256 samples including only 1 PI, 862 samples including 2 PIs, 213 samples including 3 PIs, and only 10 samples containing 4 PIs.

Using the International AIDS Society-USA resistance mutation list [13], we also considered a set of established PI resistance mutations. The total number of mutations in the protease gene associated with resistance to PIs is 74, where 23 are classified as major and the remainder as minor mutations. *Major mutations* are defined as mutations selected first in the presence of the drug or mutations substantially reducing drug susceptibility, whereas *minor mutations* generally emerge later than major mutations and, by themselves, do not have a substantial effect [13].

In both the RTI and PI data sets, drug combinations (respectively, resistance mutations) were represented using binary vectors such that every value indicates either the presence, 1, or absence, 0, of an individual drug (respectively, an individual resistance mutation) in the corresponding sample of each data set. Using two multi-dimensional Bayesian network classifiers learned separately from each data set, we were able to predict the antiretroviral

combination of RTI and PI therapies given sets of corresponding input resistance mutations. Thanks to the graphical structure of the learned MBCs, we were also able to investigate dependencies among classes (i.e., RTI or PI drugs), features (i.e., RTI or PI resistance mutations) and between classes and features (i.e., interactions between RTI or PI drugs and their respective resistance mutations).

5.2. Evaluation

We compare our MB-MBC algorithm with what is defined as a *independent classifiers* method (sometimes called binary relevance in the literature on multi-label classification) where each classifier with one class variable is learned independently using the same HITON algorithm [10,11]. In what follows, we denote *independent classifiers* method as *IndepMBCs*. Additionally, we compare MB-MBC with five other MBC learning algorithms, namely, *Tree-Tree* [1], *Polytree-Polytree* [6], *Pure Filter* [4], *Pure Wrapper* [4], and *class-bridge decomposable MBC (CB-MBC)* [5].

As non Bayesian network-based approaches, we consider for comparison three different methods: multi-label *k*-nearest neighbor (*ML-kNN*) [23], back propagation for multi-label learning (*BP-MLL*) [24], and multinomial logistic regression (*MNL*) [25]. *ML-kNN* extends the *k*-nearest neighbor lazy algorithm to a multi-label version and uses the maximum a posteriori principle to predict the label set; *BP-MLL* is derived from the popular back propagation algorithm by modifying its error function with a new function that takes into account the characteristics of multi-label learning; and *MNL* uses the multinomial logistic regression on an input set of feature variables, and returns the class value with the highest posterior probability. Similar to *IndepMBCs*, *MNL* is applied independently to each class variable, and the results are then concatenated to obtain the predicted class vector.

All methods were run in Matlab R2010b on a laptop 2.2 GHz with 6 GB RAM using Windows operating system. The HITON algorithm was run using Causal Explorer toolkit [26] provided as compiled Matlab functions, and G^2 statistical test was used to evaluate the conditional independencies between variables with a significance level $\alpha = 0.05$. *ML-kNN* and *BP-MLL* were run using the Matlab packages available at <http://lamda.nju.edu.cn/datacode/MLkNN.htm> and <http://lamda.nju.edu.cn/datacode/BPMLL.htm>, respectively. For the *ML-kNN* algorithm, the number of clusters was set to 4 for both RTI and PI data sets, and for *BP-MLL* the number of training epochs was set to 20, and the number of hidden neurons was set to 7 for the RTI data set and 14 for the PI data set. For the remaining MBC learning approaches and *CB-MBC*, Matlab implementations from [4] and [5] were used, respectively.

Five 10-fold cross-validation experiments were run for each learning algorithm for both RTI and PI data sets. Bayesian network-based methods all start from an empty structure. For MB-MBC and *IndepMBCs* methods, the five 10-fold cross-validation experiments were run with five different conditioning set size values, i.e., with $\max_{CS} = 1, 2, 3, 4, 5$.

In order to evaluate the performance of the learned MBCs, we use two performance metrics [4], namely:

- The *mean accuracy* over the *d* class variables:

$$Acc_m = \frac{1}{d} \sum_{i=1}^d \frac{1}{N} \sum_{l=1}^N \delta(c'_i, c_{li}), \quad (3)$$

where *N* is the size of the test set, c'_i is the C_i class value predicted by the MBC for sample *l*, and c_{li} denotes its corresponding

Table 1
Estimated performance metrics (mean ± std. deviation) for reverse transcriptase inhibitors data set.

		Mean accuracy	Global accuracy
MB-MBC	max _{CS} = 1	0.7108 ± 0.0221	0.1151 ± 0.0466
	max _{CS} = 2	0.7062 ± 0.0191	0.0881 ± 0.0403
	max _{CS} = 3	0.7019 ± 0.0153	0.0780 ± 0.0363
	max _{CS} = 4	0.6995 ± 0.0145	0.0701 ± 0.0336
	max _{CS} = 5	0.6978 ± 0.0106	0.0646 ± 0.0241
IndepMBs	max _{CS} = 1	0.7331 ± 0.0178	0.0561 ± 0.0199
	max _{CS} = 2	0.7344 ± 0.0143	0.0484 ± 0.0142
	max _{CS} = 3	0.7314 ± 0.0141	0.0398 ± 0.0101
	max _{CS} = 4	0.7316 ± 0.0141	0.0380 ± 0.0098
	max _{CS} = 5	0.7315 ± 0.0141	0.0377 ± 0.0099
Tree-Tree		0.6968 ± 0.0163	0.0364 ± 0.0101
Polytree-Polytree		0.6999 ± 0.0147	0.0299 ± 0.0062
Pure Filter		0.7074 ± 0.0063	0.0240 ± 0.0066
Pure Wrapper		0.7095 ± 0.0040	0.0291 ± 0.0008
CB-MBC		0.7261 ± 0.0113	0.0382 ± 0.0105
ML-kNN		0.7373 ± 0.0180	0.0729 ± 0.0259
BP-MLL		0.7189 ± 0.0095	0.0428 ± 0.0165
MNL		0.7365 ± 0.0159	0.0595 ± 0.0203

real value. $\delta(c'_{li}, c_{li}) = 1$ if the predicted and real class values are equal, i.e., $c'_{li} = c_{li}$, and 0 otherwise.

- The global accuracy over the d -dimensional class variable:

$$Acc_g = \frac{1}{N} \sum_{l=1}^N \delta(\mathbf{c}'_l, \mathbf{c}_l). \tag{4}$$

In this more strict case, the (d -dim) vector of predicted classes \mathbf{c}'_l is compared to the vector of real classes \mathbf{c}_l , so that we have $\delta(\mathbf{c}'_l, \mathbf{c}_l) = 1$ if there is a complete equality between both vectors, i.e., $\mathbf{c}'_l = \mathbf{c}_l$, and 0 otherwise.

5.3. Experimental results

5.3.1. Reverse transcriptase inhibitors

Table 1 shows the prediction results for the RTI data set with mean values and standard deviations for each metric and each learning method. The best result of each metric is written in

bold. ML-kNN presents the best mean accuracy (73%), whereas MB-MBC outperforms the remaining approaches in the global accuracy with 11%. Note that the best results for the MB-MBC algorithm are obtained with max_{CS} = 1, and as max_{CS} grows, the overall mean and global accuracies decrease. We performed a multiple comparison of all algorithm performances using the Friedman test followed by the Tukey-Kramer post hoc test with a significance level of $\alpha = 0.05$. For the mean accuracy, it turns out that (1) ML-kNN and MNL are significantly better than MB-MBC with max_{CS} = 4, MB-MBC with max_{CS} = 5, Tree-Tree, and Polytree-Polytree; (2) IndepMBs with max_{CS} = 2 are significantly better than MB-MBC with max_{CS} = 4, MB-MBC with max_{CS} = 5, and Tree-Tree; and (3) IndepMBs with max_{CS} = 1 is only significantly better than Tree-Tree. For the global accuracy, it turns out that MB-MBC with max_{CS} = 1 is significantly better than IndepMBs with max_{CS} = 5, Tree-Tree, Polytree-Polytree, Pure Filter, and Pure Wrapper. For all remaining algorithms, the differences in classification performance are not statistically significant.

Using the learned graphical structure of the most accurate MBC, shown in Fig. 2, we identified and analyzed the different interactions in the RTI data set between drugs belonging to both the NRTI and NNRTI groups and established resistance mutations.

Firstly, the class subgraph (red arcs) in the RTI network shows associations between the following NRTI drugs: AZT, ABC, 3TC, TDF and DDI, which may reveal the extent of cross-resistance between each related pair of these drugs. The NRTI drug D4T has a unique association with the NRTI drug AZT, and two associations with the NNRTI drugs EFV and NVP. Note that these identified dependence relationships are partially consistent with the previous work by Deforche et al. [27] that proved the existence of direct influences between the NRTI drugs AZT, 3TC, ABC, DDI, and D4T, as well as direct influences between the NNRTI drugs EFV and NVP and D4T. Deforche et al. also used Bayesian networks to discover interactions between drugs and resistance mutations within and between NRTIs and NNRTIs; however, contrary to our approach, they do not deal with the HIV treatment prediction problem and they learned just two Bayesian networks separately for two NNRTI drugs, namely EFV and NVP, to investigate resistance pathways [27].

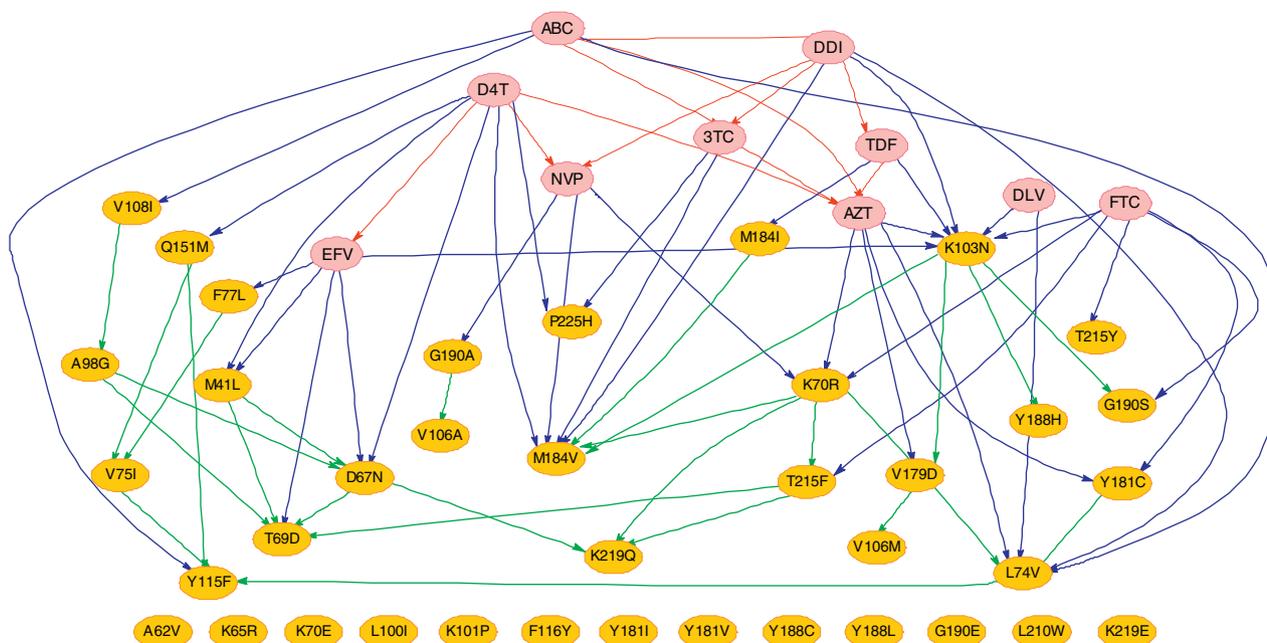


Fig. 2. The graphical structure of the multi-dimensional Bayesian network classifier learnt using reverse transcriptase inhibitors data set.

The unique NRTI drug that has no dependency relationships with the other drugs is FTC. This can be attributed either to the fact that there is not enough data on this drug since it appears in only 205 samples, or to their being no influence between FTC and all other drugs in this data set.

In the class subgraph, we also find that no dependency relationships are detected within the NNRTI group, that is, there is no dependency relationships between the three NNRTI drugs: NVP, EFV and DLV. In fact, NVP only has associations with two NRTI drugs D4T and DDI, EFV has a unique association with the NRTI drug D4T, whereas DLV has no associations with any other drugs. The same finding as for FTC may also apply to DLV. However, the second hypothesis, i.e., absence of influence between DLV and all other drugs, is more likely since there is a greater frequency of appearance of DLV in the data set (1990 samples) than FTC.

Secondly, the bridge subgraph (blue arcs) reveals dependency relationships between RTI drugs and resistance mutations. For the first group of NRTIs, we find that each NRTI drug, except TDF, was directly related to at least one of its established resistance mutations, which confirms the current knowledge on the role of these mutations in relation to corresponding NRTI drugs [13]. For instance, ABC was associated with mutations L74V and Y115F; D4T was associated with mutations M41L and D67N; FTC was associated with mutations K70R and T215F/Y; and DDI was associated with mutation L74V. Additionally, AZT and 3TC were directly connected to mutations K70R and 184V, respectively. This result was also confirmed in recent work by Theys et al. [28], where the K70R and 184V mutations were identified as two major resistance mutations to the combination of AZT and 3TC.

In general, two basic types of NRTI-resistance mechanism are known for HIV-1. The first resistance mechanism is *exclusion* and involves enhanced discrimination at the time the NRTI is incorporated. One example is the M184-V/I mutation that reduces the incorporation of 3TC and FTC by steric hindrance. The second mechanism is *excision* and involves the selective removal of NRTI from the end of the viral DNA after it has been incorporated by RT. This is, for instance, the excision mechanism involved in AZT resistance and is achieved through the accumulation of a specific set of mutations including M41L, D67N, K70R, L210W, T215F, and K219E/Q. Interestingly, the same set of mutations is also selected in viruses from patients under D4T therapy and are commonly designated as TAMs (i.e., thymidine analog resistance mutations) [29]. Cross-resistance due to the presence of TAMs affects all NRTI drugs to some extent [13,30].

Dependency relationships identified for NNRTI group are also consistent with current knowledge [27,13] as EFV and DLV were directly associated with the established resistance mutation K103N, and NVP was associated with G190A.

The bridge subgraph indicated that all NRTI drugs were directly associated with some NNRTI resistance mutations, such as K103N (associated with AZT, FTC, DDI and TDF), Y181C (associated with AZT and FTC), P225H (associated with D4T and 3TC), V108I, V179D and G190S (associated, respectively, with ABC, AZT and FTC). Similarly, all NNRTI drugs were directly associated with some NRTI resistance mutations, such as M41L, D67N, T69D and F77L (associated with EFV), K70R and M184V (associated with NVP), and L74V (associated with DLV). This certainly reveals the extent of inter-group interactions between NRTIs and NNRTIs.

Finally, the feature subgraph (green arcs) allowed us to identify interactions between NRTI and NNRTI resistance mutations. The mutations with the greatest number of dependence relationships were mutations D67N (4 connections: M41L, T69D, A98G, K219Q), T69D (4 connections: M41L, D67N, A98G, T215F), K70R (4 connections: L74V, M184V, T215F, K219Q) and K103N (4 connections: V179D, M184V, Y188H, G190G). Notice that there are 13 resistance mutations (at the bottom) that do not interact with drugs

Table 2

Estimated performance metrics (mean \pm std. deviation) for protease inhibitors data set.

		Mean accuracy	Global accuracy
MB-MBC	max _{CS} = 1	0.8476 \pm 0.0072	0.3187 \pm 0.0304
	max _{CS} = 2	0.8463 \pm 0.0086	0.3123 \pm 0.0412
	max _{CS} = 3	0.8449 \pm 0.0070	0.3035 \pm 0.0298
	max _{CS} = 4	0.8408 \pm 0.0072	0.2886 \pm 0.0329
	max _{CS} = 5	0.8407 \pm 0.0069	0.2904 \pm 0.0333
IndepMBs	max _{CS} = 1	0.8518 \pm 0.0051	0.2726 \pm 0.0256
	max _{CS} = 2	0.8563 \pm 0.0054	0.2231 \pm 0.0134
	max _{CS} = 3	0.8594 \pm 0.0036	0.2010 \pm 0.0047
	max _{CS} = 4	0.8593 \pm 0.0038	0.1896 \pm 0.0099
	max _{CS} = 5	0.8610 \pm 0.0021	0.1912 \pm 0.0060
Tree-Tree		0.8399 \pm 0.0019	0.1931 \pm 0.0256
Polytree-Polytree		0.8432 \pm 0.0050	0.1509 \pm 0.0220
Pure Filter		0.8527 \pm 0.0030	0.0966 \pm 0.0157
Pure Wrapper		0.8530 \pm 0.0006	0.0998 \pm 0.0038
CB-MBC		0.8552 \pm 0.0037	0.2224 \pm 0.0232
ML-kNN		0.8612 \pm 0.0069	0.2279 \pm 0.0364
BP-MLL		0.8080 \pm 0.0283	0.1473 \pm 0.0328
MNL		0.8682 \pm 0.0051	0.2551 \pm 0.0174

or features. A possible explanation is the lack of instances of such mutations and/or their low interactions with the other variables in the data set.

5.3.2. Protease inhibitors

Table 2 presents the prediction results for the PI data set with mean values and standard deviations for each metric and each learning method. The best result of each metric is written in bold. MNL presents the best mean accuracy (86%), whereas MB-MBC induces the best global accuracy (31%). As with the RTI data set, we ran a multiple comparison of all algorithm performances using the Friedman test followed by the Tukey-Kramer post hoc test with a significance level of $\alpha = 0.05$. For the mean accuracy, it turns out that (1) MNL is significantly better than MB-MBC with max_{CS} = 3, MB-MBC with max_{CS} = 4, MB-MBC with max_{CS} = 5, Tree-Tree, Polytree-Polytree and BP-MLL; (2) ML-kNN and IndepMBs with max_{CS} = 5 are significantly better than MB-MBC with max_{CS} = 4, MB-MBC max_{CS} = 5, Tree-Tree and BP-MLL; and (3) IndepMBs with max_{CS} = 3 and IndepMBs with max_{CS} = 4 are only significantly better than BP-MLL. For the global accuracy, it turns out that (1) MB-MBC with max_{CS} = 1, MB-MBC max_{CS} = 2, and MB-MBC with max_{CS} = 3 are significantly better than Polytree-Polytree, Pure Filter, Pure Wrapper, and BP-MLL; and (2) MB-MBC with max_{CS} = 4 and MB-MBC with max_{CS} = 5 are only significantly better than Pure Filter and Pure Wrapper. For all remaining algorithms, the differences in classification performance are not statistically significant.

Note finally that the performance results are better over the PI than the RTI data set; this can be explained by the fact that the number of classes (8 in PI and 10 in RTI) and the number of possible class combinations (256 in PI and 1024 in RTI) are lower in PI than in the RTI data set.

In addition, we examined the graphical structure of the most accurate learned MBC, shown in Fig. 3, in order to evaluate the usefulness of the proposed learning algorithm for identifying the different interactions between drugs and mutations in the HIV-1 protease data set.

Firstly, the learned network, specifically the class subgraph (red arcs), shows dependency relationships between the following drugs IDV, ATV, NFV, LPV and SQV, which may reveal the extent of cross-resistance between each related pair of these drugs. Notice that, for IDV, which has associations with LPV, ATV and NFV, Rhee et al. [31] recently proved in their PI cross-resistance study that IDV and LPV are among the most strongly correlated PIs. In fact, these two drugs had a correlation coefficient value equal to 0.57

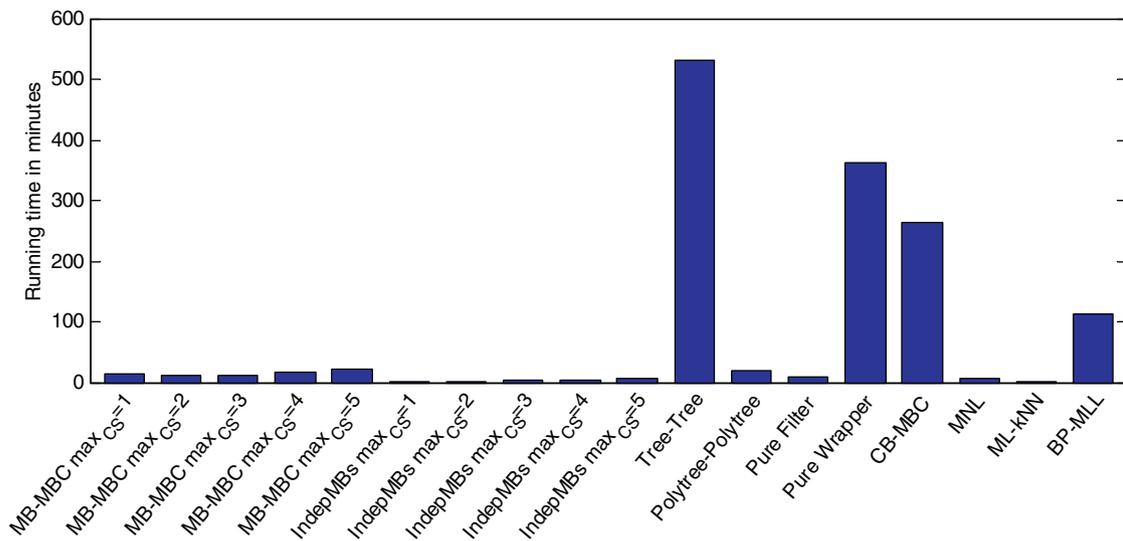


Fig. 4. Computation times of the learning process of each algorithm on reverse transcriptase inhibitors data set.

N88D, L89V, L90M), M46I (8 connections: L10F, L10I, K20I, V32I, M46L, I64L, V77I, N88S), and 7 connections for L33F (L10F, K43T, M46L, I50V, I54L, A71L, V82L) and G48V (L10F, L10I, L24I, D30N, I54A, I54S, V77I).

Note that, only three of the 19 mutations that present no interactions with other drugs or features (at the bottom), are major ones, namely T74P, V82F and N83D. As with RTIs, a possible explanation is the low number of occurrences of these mutations and/or their low interactions with the other variables in the data set.

In summary, for both RTI and PI data sets, the identified dependence relationships were proved to be consistent with the current knowledge, and were also verified by the third author, who is a medical doctor specialist in the HIV problem. However, for the variables (inhibitors or resistance mutations) that do not present any interactions with the rest of variables, larger and more diverse RTI and PI data sets need to be considered, and additional analyses need to be performed to thoroughly prove the current findings.

Finally, the computation times consumed by each algorithm on RTI and PI data sets are plotted in Figs. 4 and 5, respectively. As observed, ML-kNN is always the fastest, followed by MNL and the filter approaches, namely, Pure Filter and Polytree-Polytree.

For RTI data set, *IndepMBs* is also quite efficient and requires less computation time than *MB-MBC*. However, for PI data set, including a larger number of variables (82 variables instead of 48 variables in RTI data set), *IndepMBs* with $\max_{CS} = 1$ takes a similar computation time than *MB-MBC* with $\max_{CS} = 1$, but more computation times than *MB-MBC* for $\max_{CS} = \{2, 3, 4, 5\}$. Note here that, for both *MB-MBC* and *IndepMBs*, the consumed computation times increase as \max_{CS} grows, because the number of executed statistical independence tests increases as \max_{CS} grows. This is mainly noticed over PI data set that contains a larger number of variables. Moreover, as pointed out by Aliferis et al. [11], as \max_{CS} grows, the overall power decreases. This is also verified in our case, especially for the global accuracy values in Tables 1 and 2, that drop for both *MB-MBC* and *IndepMBs* as \max_{CS} grows. Therefore, using smaller \max_{CS} avoids excessive computations while producing better predictive performance.

In addition, as shown in Figs. 4 and 5, *BP-MLL* consumes more computation times than *ML-kNN*, *MNL* and the filter approaches mainly due to its complex error function which needs to be optimized through an iterative learning process [24]. *CB-MBC*, *Tree-Tree* and *Pure Wrapper* take always the longest

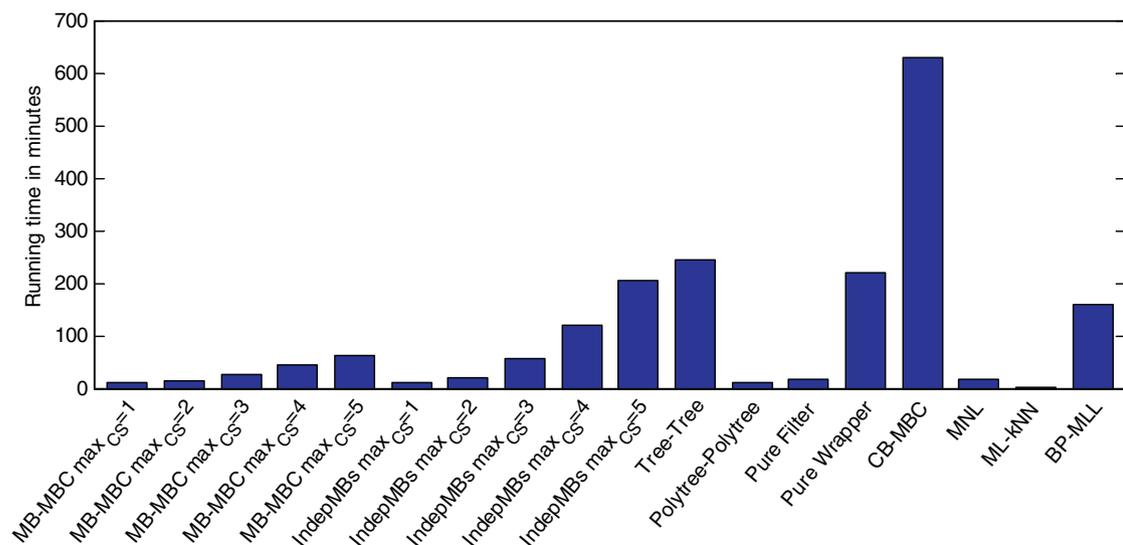


Fig. 5. Computation times of the learning process of each algorithm on protease inhibitors data set.

computation times comparing to the rest of the methods since they are all based on wrapper approaches that involve time-consuming MPE computations.

Tree-Tree is the slowest over RTI data set, whereas *CB-MBC* is the slowest over PI data set; and this can be explained by the different learning strategies behind both algorithms. In fact, the learning process of the bridge subgraph for *Tree-Tree* algorithm requires, in each iteration, an evaluation of the global accuracy of each possible MBC candidate using MPE computations. These computations mainly depend on the number of class variables and their value combinations; so that, as the number of class variables grows, *Tree-Tree* running time increases. In our case, PI data set contains 8 binary class variables, i.e., 256 class value combinations, however, RTI data set contains 10 binary class variables, i.e., 1024 class value combinations, which increases considerably *Tree-Tree* computation time over the RTI data set. The same observation applies as well for *Pure Wrapper* that iteratively evaluates and selects the MBC candidates using MPE computations.

CB-MBC is based on a different learning wrapper strategy. It first learns an initial bridge subgraph by building a selective naive Bayes for each class variable, defines the feature subgraph by randomly selecting arcs between features and adding them if they improve the accuracy, then iteratively updates MBC subgraphs as long as the number of maximal connected components is greater than one and there is an accuracy improvement. The first step in *CB-MBC* depends on the number of class and feature variables, and may require the longest computation time during *CB-MBC* learning process. In fact, for each class variable, it iterates over all feature variables to select, in each iteration, the feature that improves the accuracy the most. This, indeed, explains the highest computation time consumed by *CB-MBC* over PI data set (including 82 variables) compared to the one consumed over RTI data set (including only 48 variables).

Note finally that, being defined as a filter constraint-based approach, *MB-MBC* requires less computational time than all existing wrapper algorithms, since its learning process is based on statistical independence tests instead of accuracy metrics. Generally speaking, *MNL*, *ML-kNN* and the filter approaches require shorter computation times, whereas the wrapper approaches always take longer.

6. Conclusion

This paper proposed a novel MBC learning approach using Markov blankets, then presented its application to the HIV-1 reverse transcriptase and protease inhibitors prediction problem. Preliminary experimental results on HIV-1 Stanford data sets are promising compared with state-of-the-art MBC learning algorithms. As a constraint-based approach, *MB-MBC* ensured, through the induced MBC graphical structures, an accurate identification of the probabilistic dependence relationships among RTI and PI drugs and their corresponding resistance mutations, which was largely consistent with the latest knowledge.

It is important to emphasize that, though only applied to the HIV-1 problem, our proposed approach is general and can be applied to additional medical or biological multi-dimensional classification problems. For instance, it can be applied to coronary heart disease [36] to predict heart wall motion for the 16 segments of the heart (i.e., 16 class variables), or to the benchmark Medical data set [37] to assign clinical text reports to a subset of 45 diseases. Another example is the biological Yeast data set [38] where genes in the yeast *Saccharomyces cerevisiae* have to be associated with one or more out of 14 biological functions. Generally speaking, our approach can be applied to any multi-dimensional classification problem where an instance has to be assigned to more than one class variable.

Note however that, our approach has two main limitations. First, it cannot deal directly with continuous variables and requires a discretization pre-processing step before its application to continuous data. So, in the future, it will be an interesting issue to extend *MB-MBC* to allow the combination of both discrete and continuous predictor variables. Second, our approach cannot handle instances with missing values. To overcome this limitation, we have to perform missing values imputation before using *MB-MBC*, or adapt for example the more sophisticated Expectation-Maximization algorithm [39] to enable parameter estimations or structure learning from incomplete data [40].

In the future, we also intend to carry out a more extensive experimental study using additional synthetic and real data sets in order to prove the merits of our approach. Furthermore, since the class distributions in both the RTI and PI data sets were imbalanced, it would be worthwhile to extend our approach to deal with the challenging task of imbalanced multi-dimensional data sets. This might improve learning and classification performance results.

Finally, another line for future research is to enable our approach to adapt the learned MBCs over time as new incoming data become available. As regards non-stationary domains, this may also require the definition of concept drift for the multi-dimensional stream classification problem, as well as the development of appropriate methods to monitor concept drift and adjust the current models.

Acknowledgements

This work has been supported by projects TIN2010-20900-C04-04, Consolider Ingenio 2010-CSD2007-00018, Cajal Blue Brain, and Dynamo (FONCICYT, European Union and Mexico). Hanan Borchani is supported by an FPI fellowship from the Spanish Ministry of Economy and Competitiveness (BES-2008-003901).

References

- [1] van der Gaag LC, de Waal PR. Multi-dimensional Bayesian network classifiers. In: Studený M, Vomlel J, editors. Third European conference on probabilistic graphical models. Prague, Czech Republic: Electronic Proceedings; 2006. p. 107–14.
- [2] Tsoumakas G, Katakis I. Multi-label classification: an overview. *International Journal of Data Warehousing and Mining* 2007;3(3):1–13.
- [3] Madjarov G, Kocev D, Gjorgjevikja D, Džeroskib S. An extensive experimental comparison of methods for multi-label learning. *Pattern Recognition* 2012;45(9):705–27.
- [4] Bielza C, Li G, Larrañaga P. Multi-dimensional classification with Bayesian networks. *International Journal of Approximate Reasoning* 2011;52(6):705–27.
- [5] Borchani H, Bielza C, Larrañaga P. Learning CB-decomposable multi-dimensional Bayesian network classifiers. In: Myllymäki P, Roos T, Jaakkola T, editors. Fifth European workshop on probabilistic graphical models. Electronic Proceedings; 2010. p. 25–32.
- [6] De Waal PR, van der Gaag LC. Inference and learning in multi-dimensional Bayesian network classifiers. In: Mellouli K, editor. Ninth European conference on symbolic and quantitative approaches to reasoning with uncertainty. Lecture Notes in Computer Science, vol. 4724. Berlin/Heidelberg: Springer-Verlag; 2007. p. 501–11.
- [7] Rodríguez JD, Lozano JA. Multi-objective learning of multi-dimensional Bayesian classifiers. In: Xhafa F, Herrera F, Abraham A, Köppen M, Bénéitez JM, editors. Eighth international conference on hybrid intelligent systems. IEEE Computer Society; 2008. p. 501–6.
- [8] Zaragoza JH, Sucar LE, Morales EF, Larrañaga P, Bielza C. Bayesian chain classifiers for multidimensional classification. In: Walsh T, NICTA, University of New South Wales editors, editors. Twenty-second international joint conference on artificial intelligence. AAAI Press; 2011. p. 2192–7.
- [9] Aliferis CF, Tsamardinos I, Statnikov A. HITON: A novel Markov blanket algorithm for optimal variable selection. In: Proceedings of the American Medical Informatics Association annual symposium. American Medical Informatics Association; 2003. p. 21–5.
- [10] Aliferis CF, Statnikov A, Tsamardinos I, Mani S, Koutsoukos XD. Local causal and Markov blanket induction for causal discovery and feature selection for classification. Part I: Algorithms and empirical evaluation. *Journal of Machine Learning Research* 2010;11:171–234.
- [11] Aliferis CF, Statnikov A, Tsamardinos I, Mani S, Koutsoukos XD. Local causal and Markov blanket induction for causal discovery and feature selection for classification. Part II: Analysis and extensions. *Journal of Machine Learning Research* 2010;11:235–84.

- [12] Rhee SY, Gonzales MJ, Kantor R, Betts J, Ravela J, Shafer RW. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Research* 2003;31(1):298–303.
- [13] Johnson VA, Brun-Vézinet F, Clotet B, Günthard HF, Kuritzkes DR, Pillay D, et al. Update of the drug resistance mutations in HIV-1: December 2010. *International AIDS Society-USA, Topics in HIV Medicine* 2010;18(5):156–63.
- [14] Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference. San Francisco, CA, USA: Morgan Kaufmann Publishers; 1988.
- [15] Koller D, Friedman N. Probabilistic graphical models. Principles and techniques. Cambridge, MA, USA: MIT Press; 2009.
- [16] Pearl J, Verma TS. Equivalence and synthesis of causal models. In: Bonissone PP, Henrion M, Kanal LN, Lemmer JF, editors. Sixth annual conference on uncertainty in artificial intelligence. Cambridge, MA, USA: Elsevier; 1990. p. 220–7.
- [17] Chow C, Liu C. Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory* 1968;14(3):462–7.
- [18] Rebane G, Pearl J. The recovery of causal polytrees from statistical data. In: Kanal LN, Levitt TS, Lemmer JF, editors. Third annual conference on uncertainty in artificial intelligence. Washington, USA: Elsevier; 1987. p. 222–8.
- [19] Zaragoza JH, Sucar LE, Morales EF. A two-step method to learn multidimensional Bayesian network classifiers based on mutual information measures. In: Murray RC, McCarthy PM, editors. Twenty-fourth international Florida Artificial Intelligence Research Society Conference. Palm Beach, FL, USA: AAAI Press; 2011. p. 644–9.
- [20] Spirtes P, Glymour C, Scheines R. Causation, prediction, and search. Cambridge, MA, USA: MIT Press; 2000.
- [21] Altmann A, Beerenwinkel N, Sing T, Savenkov I, Doumer M, Kaiser R, et al. Improved prediction of response to antiretroviral combination therapy using the genetic barrier to drug resistance. *Antiviral Therapy* 2007;12(2):169–78.
- [22] von Kleist M, Menz S, Huisinga W. Drug-class specific impact of antivirals on the reproductive capacity of HIV. *PLoS Computational Biology* 2010;6(3), <http://dx.doi.org/10.1371/journal.pcbi.1000720>.
- [23] Zhang ML, Zhou ZH. ML-kNN: a lazy learning approach to multi-label learning. *Pattern Recognition* 2007;40(7):2038–48.
- [24] Zhang ML, Zhou ZH. Multi-label neural networks with applications to functional genomics and text categorization. *IEEE Transactions on Knowledge and Data Engineering* 2006;18(10):1338–51.
- [25] Greene WH. Econometric analysis. Upper Saddle River, NJ, USA: Prentice Hall; 1997.
- [26] Aliferis CF, Tsamardinos I, Statnikov A. Causal explorer: a probabilistic network learning toolkit for discovery. Available from: discover.mc.vanderbilt.edu/discover/public/causalexplorer/ [accessed 10.12.12].
- [27] Deforche K, Camacho R, Grossman Z, Soares MA, Van Laethem K, Katzenstein DA, et al. Bayesian network analyses of resistance pathways against efavirenz and nevirapine. *AIDS* 2008;22:2107–15.
- [28] Theys K, Deforche K, Libin P, Camacho R, Van Laethem K, Vandamme AM. Resistance pathways of human immunodeficiency virus type 1 against the combination of zidovudine and lamivudine. *Journal of General Virology* 2010;91(8):1898–908.
- [29] Sarafianos SG, Marchand B, Das K, Himmel DM, Parniak MA, Hughes SH, et al. Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *Journal of Molecular Biology* 2009;385(3):693–713.
- [30] Whitcomb JM, Parkin NT, Chappey C, Hellmann NS, Petropoulos CJ. Broad nucleoside reverse-transcriptase inhibitor cross-resistance in human immunodeficiency virus type 1 clinical isolates. *Journal of Infectious Diseases* 2003;188(7):992–1000.
- [31] Rhee SY, Taylor J, Fessel WJ, Kaufman D, Towner W, Troia P, et al. HIV-1 protease mutations and protease inhibitor cross-resistance. *Antimicrobial Agents and Chemotherapy* 2010;54(10):4253–61.
- [32] Lambert-Niclot S, Flandre P, Canestri A, Peytavin G, Blanc C, Agher R, et al. Factors associated with the selection of mutations conferring resistance to protease inhibitors (PIs) in PI-experienced patients displaying treatment failure on darunavir. *Antimicrobial Agents and Chemotherapy* 2008;52(2):491–6.
- [33] Marcelin AG, Masquelier B, Descamps D, Izopet J, Charpentier C, Alloui C, et al. Tipranavir-ritonavir genotypic resistance score in protease inhibitor-experienced patients. *Antimicrobial Agents and Chemotherapy* 2008;52(9):3237–43.
- [34] Deforche K, Silander T, Camacho R, Grossman Z, Soares MA, Van Laethem K, et al. Analysis of HIV-1 pol sequences using Bayesian networks: implications for drug resistance. *Bioinformatics* 2006;22(24):2975–9.
- [35] Deforche K, Camacho R, Grossman Z, Silander T, Soares MA, Moreau Y, et al. Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors. *Infection Genetics and Evolution* 2007;7:382–90.
- [36] Qazi M, Fung G, Krishnan S, Rosales R, Steck H, Rao R, et al. Automated heart wall motion abnormality detection from ultrasound images using Bayesian networks. In: Veloso MM, editor. Twentieth international joint conference on artificial intelligence. Hyderabad, India: AAAI Press; 2007. p. 519–25.
- [37] Medical Dataset Mulan: a Java library for multi-label learning. Datasets. Available from: <http://mulan.sourceforge.net/datasets.html> [accessed 10.12.12].
- [38] Elisseeff A, Weston J. A kernel method for multi-labelled classification. *Advances in Neural Information Processing Systems* 2002;14:681–7.
- [39] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* 1977;39:1–38.
- [40] Friedman N. The Bayesian structural EM algorithm. In: Cooper GF, Moral S, editors. Fourteenth conference on uncertainty in artificial intelligence. Madison, WI, USA: Morgan Kaufmann Publishers; 1998. p. 129–38.