Learning Multi-Dimensional Bayesian Network Classifiers Using Markov Blankets: A Case Study in the Prediction of HIV Protease Inhibitors

Hanen Borchani, Concha Bielza, and Pedro Larrañaga

Computational Intelligence Group, Departamento de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, Boadilla del Monte, 28660, Spain.
hanen.borchani@upm.es, mcbielza@fi.upm.es, pedro.larranaga@fi.upm.es

Abstract. Multi-dimensional Bayesian network classifiers (MBCs) are Bayesian network classifiers especially designed to solve multi-dimensional classification problems, where each instance in the data set has to be assigned to one or more class variables. In this paper, we introduce a new method for learning MBCs from data basically based on determining the Markov blanket around each class variable using the HITON algorithm. Our method is applied to the human immunodeficiency virus (HIV) protease inhibitor prediction problem. The experimental study showed promising results in terms of classification accuracy, and we gained insight from the learned MBC structure into the different possible interactions among protease inhibitors and resistance mutations.

1 Introduction

Multi-dimensional classification is an extension of the classical one-dimensional classification, where each instance given by a vector of m features $\mathbf{x} = (x_1, ..., x_m)$ is associated with a vector of d class values $\mathbf{c} = (c_1, ..., c_d)$ rather than a single class value [16]. Recently, the concept of multi-dimensionality has been introduced in Bayesian network classifiers providing an accurate modelling of this emerging problem and ensuring interactions among all variables [4, 5, 9, 16–18]. In these probabilistic graphical models, known as multi-dimensional Bayesian network classifiers (MBCs), 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 [1–3], and then we determine edge directionality over all three MBC subgraphs. Thanks to this filter and local approach to MBC learning, we can lighten the computational burden of MBC learning using wrapper algorithms [4, 5, 16] and provide more accurate MBC structures.

We finally apply our Markov blanket MBC (MB-MBC) algorithm to the problem of predicting human immunodeficiency virus (HIV) protease inhibitors (PIs) given an input set of resistance mutations that an HIV patient carries. In general, a combination of several antiretroviral PI drugs should be repeatedly administered for each patient in order to prevent and treat the HIV infection. We analyze a data set obtained from the Stanford HIV protease database [13]. The class variables are eight protease inhibitor drugs (i.e., d=8) and the feature variables are 74 predefined mutations [10] associated with resistance to protease inhibitors (i.e., m=74). Experimental results were promising in terms of classification accuracy as well as of the identification of interactions among drugs and resistance mutations, which were either consistent with the current 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 protease inhibitor data set. Finally, Section 6 sums up the paper with some conclusions.

2 Background

A Bayesian network over a set of discrete random variables $\mathbf{U} = \{X_1, ..., X_n\}$, $n \geq 1$, is a pair $\mathcal{B} = (\mathcal{G}, \boldsymbol{\Theta})$. $\mathcal{G} = (V, A)$ 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. $\boldsymbol{\Theta}$ is a set of conditional probability distributions such that $\theta_{x_i|\mathbf{pa}(x_i)} = p(x_i \mid \mathbf{pa}(x_i))$ 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:

$$p(X_1, ..., X_n) = \prod_{i=1}^{n} p(X_i \mid \mathbf{Pa}(X_i)) \cdot \tag{1}$$

Definition 1. *Conditional Independence. Two variables X and Y are conditionally independent given $\boldsymbol{Z}$, denoted as $I(X, Y \mid \boldsymbol{Z})$, iff $P(X \mid Y, \boldsymbol{Z}) = P(X \mid \boldsymbol{Z})$ for all values $x, y, \boldsymbol{z}$ of $X, Y, \boldsymbol{Z}$, respectively, such that $P(\boldsymbol{Z} = \boldsymbol{z}) > 0$.*

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

In other words, $\mathrm{MB}(X)$ is a minimal set of variables conditioned by which X is conditionally independent of all the remaining variables. Under the faithfulness assumption, $\mathrm{MB}(X)$ consists of the union of the set of parents, children, and parents of children (i.e., spouses) of X [11].

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. *A multi-dimensional Bayesian network classifier is a Bayesian network* $\mathcal{B} = (\mathcal{G}, \boldsymbol{\Theta})$ *where the structure* $\mathcal{G} = (V, A)$ *has a restricted topology. The set of* n *vertices* V *is partitioned into two sets:* $V_C = \{C_1, ..., C_d\}, d \geq 1$*, of class variables and* $V_X = \{X_1, ..., X_m\}, m \geq 1$*, 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* $\mathcal{G}_\mathcal{C} = (V_C, A_C)$ -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* $\mathcal{G}_\mathcal{X} = (V_X, A_X)$ -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* $\mathcal{G}_{\mathcal{CX}} = (V, A_{CX})$ -bridge subgraph- *of* $\mathcal{G}$ *connecting class and feature variables.*

Depending on the graphical structures of the class and feature subgraphs MBCs can be divided into several families. These families can be denoted as `class subgraph structure-feature subgraph structure` MBCs, where the possible structures of each subgraph are: empty, tree, polytree, or DAG [4].

Classification with an MBC under a 0-1 loss function is equivalent to solving the most probable explanation (MPE) problem, i.e., for a given fact $\mathbf{x} = (x_1, ..., x_m)$ we have to obtain

$$\begin{aligned} \mathbf{c}^* &= (c_1^*, ..., c_d^*) \\ &= arg \max_{c_1,...,c_d} p(C_1 = c_1, ..., C_d = c_d \mid \mathbf{x}) \cdot \end{aligned} \tag{2}$$

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

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

Several approaches have recently been proposed to learn MBCs from data. In [16], Van der Gaag and de Waal use Chow and Liu's algorithm [6] to learn the class and feature subgraphs of a `tree-tree` 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 presented a theoretical approach for learning `polytree-polytree` MBCs in [9]. Class and feature subgraphs are separately generated using Rebane and Pearl's algorithm [12]; however, the induction of the bridge subgraph was not specified.

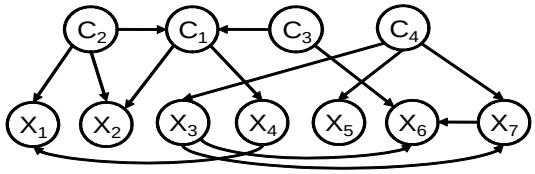

Fig. 1. An example of an MBC structure.

More recently, a two-step method was proposed by Zaragoza et al. [17] to also learn `polytree-polytree` MBCs. First, they build class and feature subgraphs using Chow and Liu's algorithm [6] 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.

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). Note that none of these algorithms places any constraints on the subgraph structures of the generated MBCs.

In [5], we propose a learning algorithm for class-bridge decomposable MBCs, instead of general MBCs, based on a greedy forward selection wrapper approach. Class or feature subgraphs can have any type of structure. Compared with prior algorithms in [4, 9, 16], our method performs better and requires less computational time than the existing wrapper algorithms.

Moreover, Zaragoza et al. present a two-step method in [18]. 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. Our objective is to tackle the shortcomings of our previous learning method [5], mainly its high computational cost, by taking advantage of the merits of a filter 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 [5] may not always lead to an accurate MBC structure, since arcs between features

are selected at random in the feature subgraph learning steps. This may affect the overall quality of the learned MBC structure and lead consequently 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, MMMB and HITON (see [2, 3] and their references for reviews). In this paper, we only consider and apply the HITON algorithm [1–3] in 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 [2].

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.

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

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

HITON-PC starts with an empty set of candidates $\mathrm{PC}(T)$, 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 $\mathrm{PC}(T)$, it checks if there is any variable inside $\mathrm{PC}(T)$ that is independent of T given some subset $\mathbf{Z}$. In this case, this variable will be removed from $\mathrm{PC}(T)$ (steps 6 to 11). These steps are iterated until there are no more variables in OPEN. Finally, $\mathrm{PC}(T)$ is filtered using the symmetry criterion (steps 13 to 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) [2] .

Algorithm 2 HITON-PC(T)

1. $PC(T) \leftarrow \emptyset$
2. OPEN $\leftarrow \mathbf{U} \setminus \{T \cup PC(T)\}$
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** $\exists\ \mathbf{Z} \subseteq PC(T) \setminus \{X\}$, s.t. $I(X, T \mid \mathbf{Z})$ **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 $\mathbf{Z}$ [2]. 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, $i = 1, ..., d$, 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 [15].
- *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)\ 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 $\mathbf{Y} = PC(X) \setminus (\{C_i\} \cup \{\text{MB}(C_i) \cap \text{PC}(X)\})$. If $\mathbf{Y} \neq \emptyset$, we insert an arc from X to every feature variable in $\mathbf{Y}$.

5 Experimental Study

5.1 Data set

Treatments for human immunodeficiency virus (HIV) mostly involve 18 antiretroviral drugs grouped into three classes: nucleoside and nucleotide reverse transcriptase inhibitors (NRTIs) including seven drugs, non-nucleoside reverse transcriptase inhibitors (NNRTIs) including three drugs, and protease inhibitors (PIs) containing eight drugs. In this paper, we studied PIs only, but we plan to extend our study to both NRTIs and NNRTIs in the future.

We analyzed a data set obtained from the Stanford HIV protease database [13] containing antiretroviral PI treatment histories from 1255 patients. These treatment histories were collected from previously published studies. Eight PI drugs (i.e., d=8) are considered: Atazanavir (ATV), Darunavir (DRV), Fosamprenavir (FPV), Indinavir (IDV), Lopinavir (LPV), Nelfinavir (NFV), Saquinavir (SQV) and Tipranavir (TPV). 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 PIs administered to a patient at a specified time point during his or her course of PI treatment. Only samples where no drug was administered were discarded. Accordingly, the final data set contained a total of 4341 samples. However, 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.

Moreover, we considered established drug resistance mutations that were defined in the last International AIDS Society-USA resistance mutation list [10]. The total number of mutations in the protease gene associated with resistance to PIs is 74 (i.e., m=74), where 23 are classified as major and the remaining as minor mutations. *Major mutations* are defined as mutations selected first in the presence of the drug or mutations substantially reducing drug susceptibility. *Minor mutations* generally emerge later than major mutations and by themselves do not have a substantial effect [10].

PI 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 PI drug (respectively an individual resistance mutation) in the corresponding sample of the data set. Using a multi-dimensional Bayesian network classifier learned from this data we were able to predict antiretroviral combination PI therapies given sets of input mutations. Thanks to its graphical structure, we were also able to investigate dependencies among classes (i.e., PI drugs), features (i.e., mutations) and between classes and features (i.e., interactions between PI drugs and mutations).

5.2 Experimental Results

We compare our MB-MBC algorithm with what is defined as a multiple classifier method, where each classifier is learned independently (sometimes called binary relevance in the literature on multi-label classification) using the same HITON

approach with just a single class variable. In order to evaluate the performance of the learned MBCs, five 10-fold cross-validation experiments are run for each classifier and each conditioning set size value, i.e., with $max_{CS} = 1, 2, 3, 4, 5$. We use two performance metrics [4], namely:

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

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

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

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

  In this case, the 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.

Table 1 shows the prediction results with mean values and standard deviations for each metric and each method. Note that the best results are obtained with $max_{CS} = 1$ (94% mean accuracy and 71% global accuracy), and as max_{CS} grows, the overall mean and global accuracies decrease. As expected, without exception, MB-MBC outperforms the independent classifier model notably with respect to global accuracy.

Table 1. Estimated performance metrics (mean ± standard deviation).

	MB-MBC		**Independent classifiers**	
$\mathbf{max_{CS}}$	**Mean accuracy**	**Global accuracy**	**Mean accuracy**	**Global accuracy**
1	0.9416 ± 0.0049	0.7188 ± 0.0250	0.9339 ± 0.0019	0.7035 ± 0.0054
2	0.9330 ± 0.0033	0.6868 ± 0.0075	0.9247 ± 0.0017	0.5994 ± 0.0185
3	0.9193 ± 0.0031	0.6338 ± 0.0083	0.9156 ± 0.0039	0.4960 ± 0.0153
4	0.8890 ± 0.0108	0.5153 ± 0.0321	0.8775 ± 0.0091	0.4071 ± 0.0296
5	0.8641 ± 0.0201	0.4266 ± 0.0568	0.8438 ± 0.0107	0.3551 ± 0.0328

In addition, we examined the graphical structure of the most accurate learned MBC, shown in Figure 2, in order to evaluate the usefulness of the proposed learning algorithm in identifying the different interactions between drugs and mutations in the HIV 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. [14] recently proved in their PIs 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 [14]. Similarly, based on their study, IDV and ATV, ATV and NFV as well as NFV and IDV had high correlation coefficients. Nevertheless, correlation coefficients between LPV and both drugs NFV and SQV were lower, equal to 0.14 and 0.05 respectively. This goes to confirm then that the dependency relationships identified in the network among the above PI drugs are consistent with Rhee et al.'s study [14].

However, our results were less conclusive for other drugs (DRV, FPV and TPV) since no associations are detected between them or between them and the other drugs. A possible explanation is the lack of available data, as there were fewer than 30 samples for each of these drugs. On this ground, we would require a larger and diverse data set for our future analysis in order to investigate possible interactions between these drugs and the other variables in the network.

Concerning relationships between PI drugs and mutations, visualized by the bridge subgraph (blue arcs), let us first discuss the two possible types of mutations, major and minor, and then how their associations with PI drugs have been previously interpreted in the literature in the context of Bayesian networks. As Defroche et al. found [7, 8], a major mutation actually plays a key role in drug resistance, and thus, should have an unconditional dependency on the drug, and this is indicated in the network graphical structure by the presence of an arc between the major mutation and the drug.

In contrast, a minor mutation further increases drug resistance mostly only in the presence of major mutations. Thus, it is expected to be conditionally independent of the drug but dependent on other major resistance mutations. This is indicated in the network by the presence of an arc between major or minor mutations instead of an arc between the minor mutation and the drug node. Even so, as claimed by Defroche et al. [7], a minor mutation may still be connected to the drug.

Notice that the conditional independencies revealed in our bridge subgraph in Figure 2 are largely consistent with the above definitions, since most of the major mutations are directly connected to one or more drug nodes. For instance, on the left, D30N (which is defined in [10] as a major mutation of NFV) was not only associated with NFV but also with IDV, LPV and SQV, proving again the extent of cross-resistance between these drugs. Similarly, on the right, L76V (which is defined in [10] as a major mutation of LPV) was directly associated with LPV, SQV and NFV. At the center bottom of the network, G48V (major mutation of SQV [10]) was directly associated with SQV and NFV. L90M (another major mutation of SQV [10]) was also directly associated with SQV. I47A, I50L, V82A, V82L, defined in [10] as major mutations of LPV, ATV, IDV and TPV, respectively, were directly associated with the right drugs in the MBC graphical structure.

An important number of minor mutations were also directly connected to drug nodes. L10I and L33F seem to be the main minor mutations: they have the highest number of connections (3) with PI drugs, followed by the minor

mutations L10F and I54V. L10I was associated with IDV, NFV and SQV; L33F with LPV, IDV and NFV; L10F with ATV and IDV, and I54V with LPV and NFV. Additionally, consistently with the latest knowledge in [10], more minor mutations, namely V82A/T, I84V, N88D/S, were associated directly with NFV. Also in agreement with [10], the minor mutation K20R was associated with LPV and the minor mutation I84V was associated with SQV.

From the feature subgraph (green arcs) of the learned MBC we were able to identify interactions among different protease mutations. The mutations with the greatest number of dependency relationships were L10I (21 connections: L10F, L10R, K20R, D30N, M46L, M46I, K43T, G48V, I50V, F53L , I54A, I54T, I62V, A71I, A71V, G73S, V82A, I84V, I85V, L90M, I93L), L10F (15 connections: L10I, L10V, V11I, K20T, L33F, M46I, G48V, I54L, I54V, L63P, I84V, I85V, 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).

Finally, of the 19 mutations that present no interactions with other drugs or features (at the bottom), only three are major ones, namely T74P, V82F and N83D. As they have no dependency relationships with any drug, these mutations are completely irrelevant.

6 Conclusion

This paper proposed a novel MBC learning approach using Markov blankets, then presented its application to the HIV protease inhibitors prediction problem. A preliminary experimental analysis showed that our approach performed well and confirmed current knowledge about different interactions among PI drugs and their resistance mutations.

In the near future, we intend to carry out a more extensive experimental study including the comparison of our approach with state-of-the-art MBC learning algorithms, using additional synthetic and real data sets in order to prove the merits of our approach. As regards the HIV drug prediction problem, we plan to apply our approach to the other two HIV drug groups: NRTI and NNRTI. Similarly, two MBCs could be learned separately for each group. However, it would be more interesting to build a single MBC including all the drugs in the PI, NRTI and NNRTI categories. This way, we will be able not only to investigate interactions among drugs and resistance mutations belonging to the same group but also to identify the potential inter-group interactions.

Acknowledgements

The authors would like to thank Dr. Carlos Toro Rueda researcher at Hospital Carlos III de Madrid for his valuable comments. 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). Hanen Borchani is supported by an FPI fellowship from the Spanish Ministry of Science and Innovation (BES-2008-003901).

References

1. Aliferis, C.F., Tsamardinos, I., Statnikov, A.: HITON: A novel Markov blanket algorithm for optimal variable selection. *AMIA*, 21-25 (2003)
2. Aliferis, C.F., Statnikov, A., Tsamardinos, I., Mani, S., Koutsoukos, X.D.: 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* 11, 171-234 (2010)
3. Aliferis, C.F., Statnikov, A., Tsamardinos, I., Mani, S., Koutsoukos, X.D.: Local causal and Markov blanket induction for causal discovery and feature selection for classification. Part II: Analysis and extensions. *Journal of Machine Learning Research* 11, 235-284 (2010)
4. Bielza, C., Li, G., Larrañaga, P.: Multi-dimensional classification with Bayesian networks. *International Journal of Approximate Reasoning*, In press, doi: 10.1016/j.ijar.2011.01.007 (2011)
5. Borchani, H., Bielza, C., Larrañaga, P.: Learning CB-decomposable multi-dimensional Bayesian network classifiers. *In Proceedings of the Fifth European Workshop on Probabilistic Graphical Models*, 25-32 (2010)
6. Chow, C., Liu, C.: Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, 14(3):462-467 (1968)
7. Deforche, K., Silander, T., Camacho, R., Grossman, Z., Soares, M.A. et al.: Analysis of HIV-1 pol sequences using Bayesian networks: Implications for drug resistance. *Bioinformatics* 22(24), 2975-2979 (2006)
8. Deforche, K., Camacho, R., Grossman, Z., Silander, T., Soares, M.A. et al.: Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors. *Infection, Genetics and Evolution* 7, 382-390 (2007)
9. De Waal, P.R., van der Gaag, L.C.: Inference and learning in multi-dimensional Bayesian network classifiers. *ECSQARU*, 4724:501-511 (2007)
10. Johnson, V. A., Brun-Vezinet, F., Clotet, B., Gunthard, H.F., Kuritzkes, D.R., et al.: Update of the drug resistance mutations in HIV-1: December 2010. *International AIDS Society-USA, Topics in HIV Medicine* 18(5), 156-163 (2010)
11. Pearl, J., Verma, T.S.: Equivalence and synthesis of causal models. *In Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*, 220-227 (1990)
12. Rebane, G., Pearl, J.: The recovery of causal polytrees from statistical data. *UAI*, 222-228 (1989)
13. Rhee, S.Y., Gonzales, M.J., Kantor, R., Betts, J., Ravela, J., Shafer, R.W.: Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Research* 31(1), 298-303 (2003)
14. Rhee, S.Y., Taylor, J., Fessel, W.J., Kaufman, D., Towner, W. et al.: HIV-1 protease mutations and protease inhibitor cross-resistance. *Antimicrobial Agents and Chemotherapy* 54(10), 4253-4261 (2010)
15. Spirtes, P., Glymour, C. and Scheines, R.: *Causation, Prediction, and Search*. MIT Press, 2nd edition, Cambridge, MA (2000)
16. van der Gaag, L.C., de Waal, P.R.: Muti-dimensional Bayesian network classifiers. *In Proceedings of the Third European Conference on Probabilistic Graphical Models*, 107-114 (2006)
17. Zaragoza, J.H., Sucar, L.E., Morales, E.F.: A two-step method to learn multidimensional Bayesian network classifiers based on mutual information measures. *In Proceedings of the Twenty-Fourth International FLAIRS Conference*, 644-649 (2011)
18. Zaragoza, J.H., Sucar, L.E., Morales, E.F., Larrañaga, P., Bielza, C.: Bayesian chain classifiers for multidimensional classification. *IJCAI*, In press (2011)

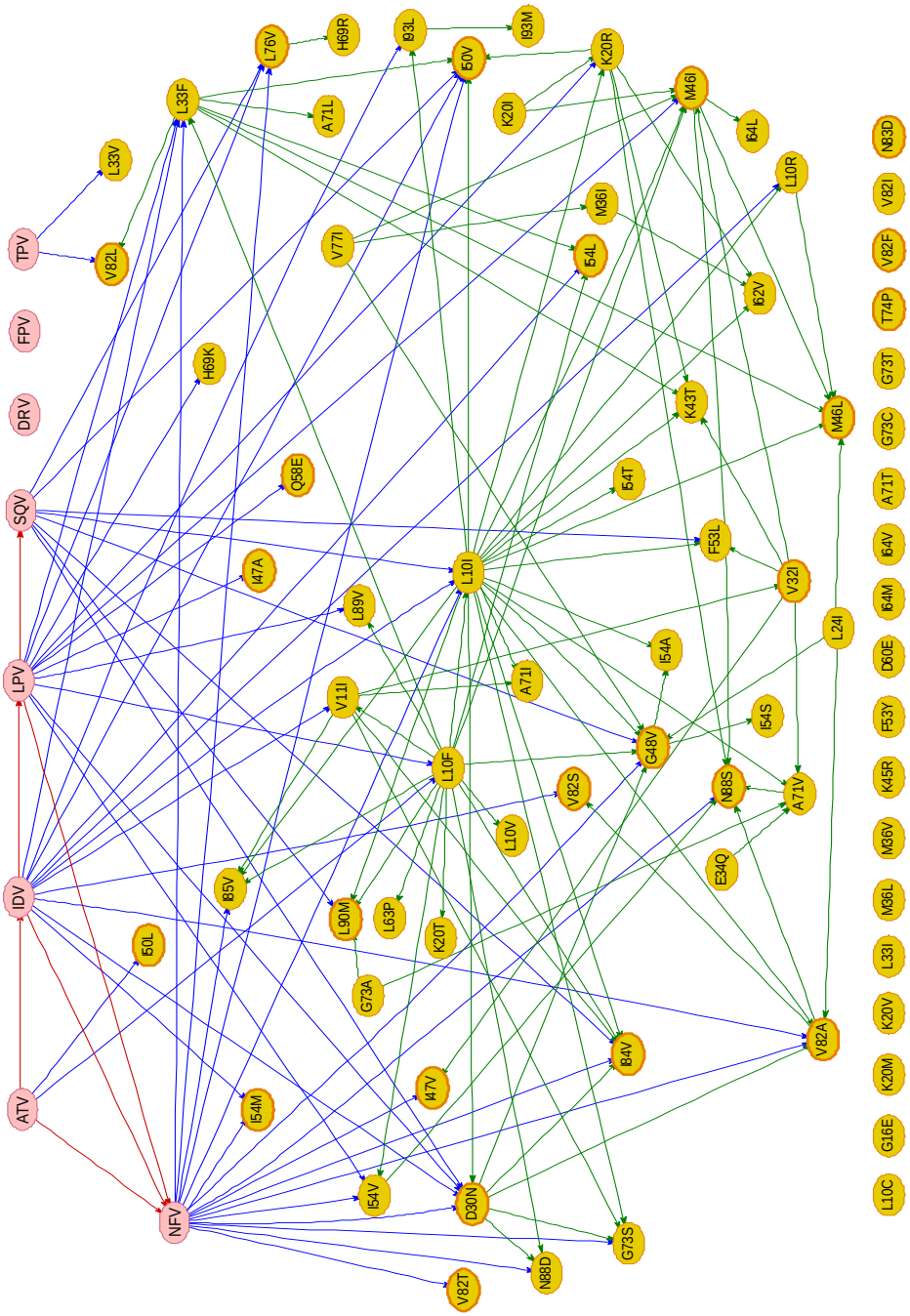

Fig. 2. The learned multi-dimensional Bayesian network classifier.