Journal of Biomedical Informatics 45 (2012) 1175–1184

Contents lists available at SciVerse ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Markov blanket-based approach for learning multi-dimensional Bayesian network classifiers: An application to predict the European Quality of Life-5 Dimensions (EQ-5D) from the 39-item Parkinson's Disease Questionnaire (PDQ-39)

Hanen Borchani [a,*], Concha Bielza [a], Pablo Martínez-Martín [b], Pedro Larrañaga [a]

[a] Computational Intelligence Group, Departamento de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, Boadilla del Monte, Madrid 28660, Spain
[b] CIEN Foundation and CIBERNED, Carlos III Institute of Health, Alzheimer Center Reina Sofía Foundation, Calle Valderrebollo 5, Madrid 28031, Spain

ARTICLE INFO

Article history:
Received 3 February 2012
Accepted 28 July 2012
Available online 8 August 2012

Keywords:
Multi-dimensional Bayesian network classifiers
Markov blanket
Multi-dimensional classification
Parkinson's disease
EQ-5D
PDQ-39

ABSTRACT

Multi-dimensional Bayesian network classifiers (MBCs) are probabilistic graphical models recently proposed to deal with multi-dimensional classification problems, where each instance in the data set has to be assigned to more than one class variable. In this paper, we propose a Markov blanket-based approach for learning MBCs from data. Basically, it consists of determining the Markov blanket around each class variable using the HITON algorithm, then specifying the directionality over the MBC subgraphs. Our approach is applied to the prediction problem of the European Quality of Life-5 Dimensions (EQ-5D) from the 39-item Parkinson's Disease Questionnaire (PDQ-39) in order to estimate the health-related quality of life of Parkinson's patients. Fivefold cross-validation experiments were carried out on randomly generated synthetic data sets, Yeast data set, as well as on a real-world Parkinson's disease data set containing 488 patients. The experimental study, including comparison with additional Bayesian network-based approaches, back propagation for multi-label learning, multi-label *k*-nearest neighbor, multinomial logistic regression, ordinary least squares, and censored least absolute deviations, shows encouraging results in terms of predictive accuracy as well as the identification of dependence relationships among class and feature variables.
© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Multi-dimensional classification problem is an extension of the classical one-dimensional classification, where each instance given by a vector of m features $\mathbf{x} = (x_1, \ldots, x_m)$ is associated with a vector of d class values $\mathbf{c} = (c_1, \ldots, c_d)$ rather than a single class value. Multi-dimensional classification has been motivated by several real-world applications such as medical diagnosis where a patient may be suffering from multiple diseases, video or text categorization where a video or a text document may be assigned to multiple topics, etc. [1].

Multi-dimensional Bayesian network classifiers (MBCs) have been recently proposed to deal with multi-dimensional classification [2–6] providing an accurate modeling of the probabilistic dependence relationships among all variables, the class variables included. As Bayesian networks, the first component of the MBCs is the *graphical structure*. It partitions the set of class and feature (predictor) variables into three different subgraphs: *class subgraph* representing the dependence relationships between class variables, *bridge subgraph* representing the dependence relationships between class and feature variables, and *feature subgraph* representing the dependence relationships between feature variables. The second component consists of the *parameters* that define the conditional probability distribution of each variable given its parent set in the structure.

In this paper, we propose a Markov blanket-based approach for learning MBCs from data named `MB-MBC`. Based on the fact that the classification performance is unaffected by parts of the structure that lie outside the Markov blanket of the class variable, `MB-MBC` starts by building the Markov blanket around each class variable using the HITON algorithm [7,8], then determines edge directionality over all three MBC subgraphs. We evaluated the `MB-MBC` algorithm using randomly generated data sets, and then we applied it to predict the European Quality of Life-5 Dimensions (EQ-5D) from the 39-item Parkinson's Disease Questionnaire (PDQ-39).

In fact, EQ-5D is a *generic* health-related quality of life (HRQoL) measure usable in general populations and in any disorder [9,10]. It is considered a valid instrument and is recommended for evaluation of HRQoL in Parkinson's disease (PD) [11,12]. EQ-5D contains five items, namely `mobility`, `self-care`, `usual activities`, `pain/discomfort`, and `anxiety/depression`; each item has three options of response: no problems, some problems and severe problems. However, PDQ-39 is a *specific* HRQoL instrument widely used in PD and it is also recommended for use in this disorder [12]. It contains 39 questions each scoring on a five-point scale: never, occasionally, sometimes, often and always [13–16].

* Corresponding author. Fax: +34 91 3524819.
E-mail address: hanen.borchani@upm.es (H. Borchani).

1532-0464/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jbi.2012.07.010

In our study, the prediction of EQ-5D values from PDQ-39 is modeled as a multi-dimensional classification problem where each instance given by an input vector of 39 features $\mathbf{x} = (x_1, \ldots, x_{39})$ (i.e., PDQ-39) is associated with a vector of 5 class values $\mathbf{c} = (c_1, \ldots, c_5)$ (i.e., EQ-5D). For empirical evaluation, we firstly performed experiments on synthetic and Yeast data sets to assess the predictive performance of MB-MBC and its ability to efficiently recover the initial MBC graphical structure, then we tested it on a real-world Parkinson's disease data set containing 488 patients. We compared MB-MBC against three different Bayesian network-based approaches, namely class-bridge decomposable MBC (CB-MBC), independent Marokov blankets (IndepMBs), and independent PC Bayesian networks (IndepPC-BNs), as well as against back propagation for multi-label learning (BP-MLL), multi-label k-nearest neighbor (ML-kNN), and commonly used methods to predict EQ-5D from HRQol specific instruments, like multinomial logistic regression (MNL), ordinary least squares (OLS), and censored least absolute deviations (CLAD). Experimental results were promising in terms of predictive performance, the identification of dependence relationships among class variables, that other approaches were unable to detect, and the identification of dependence relationships between class and feature variables.

The remainder of this paper is organized as follows. Section 2 briefly presents Bayesian networks and multi-dimensional Bayesian network classifiers; then introduces the proposed MB-MBC learning approach. Subsequently, Section 3 describes the experimental setting including the used synthetic, Yeast and Parkinson's data sets, the approaches considered for comparison, and the evaluation metrics. Section 4 presents and discusses the experimental results, and finally, Section 5 sums up the paper with some conclusions and future works.

2. Methods

2.1. Bayesian networks

A Bayesian network [17,18] over a set of discrete random variables $\mathbf{U} = \{X_1, \ldots, X_n\}$, $n \geqslant 1$, is a pair $\mathcal{B} = (\mathcal{G}, \mathbf{\Theta})$. $\mathcal{G} = (V, A)$ is a directed acyclic graph (DAG) whose vertices V correspond to variables in $\mathbf{U}$ and whose arcs A represent direct probabilistic dependencies between the vertices. $\mathbf{\Theta}$ 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 $\mathcal{G}$:

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

The probability distribution p satisfies certain conditional independencies between the random variables in $\mathbf{U}$ as follows:

Definition 1. 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. 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 [19].

2.2. Multi-dimensional Bayesian network classifiers

Definition 3. An MBC [4] is a Bayesian network $\mathcal{B} = (\mathcal{G}, \mathbf{\Theta})$ where the structure $\mathcal{G} = (V, A)$ has a restricted topology. The set of n vertices V is partitioned into two subsets: $V_C = \{C_1, \ldots, C_d\}, d \geqslant 1$, of class variables and $V_X = \{X_1, \ldots, X_m\}, m \geqslant 1$, of feature variables ($d + m = n$). The set of arcs A is partitioned into three subsets 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 $\mathcal{G}$ inuced by V [2].

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 $\mathbf{x} = (x_1, \ldots, x_m)$ we have to obtain

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

Example 1. An example of an MBC structure is shown in Fig. 1. V_C contains four class variables, V_X includes seven features. Using (1), we have

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

2.3. MB-MBC: learning MBCs using Markov blankets

Let $\mathcal{D}$ be a data set of N instances containing a value assignment for each variable $X_1, \ldots, X_m, C_1, \ldots, C_d$, i.e., $\mathcal{D} = \{(\mathbf{x}^{(1)}, \mathbf{c}^{(1)}), \ldots, (\mathbf{x}^{(N)}, \mathbf{c}^{(N)})\}$. Our proposed approach aims to find an MBC that best fits the available data set and ensures afterwards an accurate and efficient classification for the new unlabeled instances.

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 [7,8] and the references therein). In this paper, we only consider and adapt the HITON algorithm [7,8] 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 [7].

Basically, the idea of our Markov blanket MBC (MB-MBC) learning algorithm consists of firstly determining the Markov blanket around each class variable using the HITON algorithm [7,8] and then specifying directionality over the MBC subgraphs. HITON identifies the Markov blanket of each class variable in a two-phase

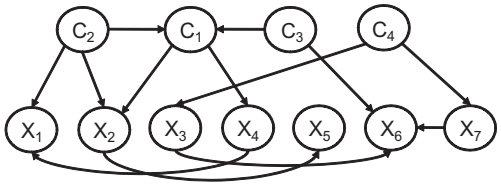

Fig. 1. Example of an MBC structure.

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. $\mathbf{S} \leftarrow \{\bigcup_{T \in PC(C_i)} PC(T)\} \setminus \{PC(C_i) \cup C_i\}$
7. **for** every variable $X \in \mathbf{S}$ **do**
8. Retrieve a subset $\mathbf{Z}$ s.t. $I(X, C_i|\mathbf{Z})$
9. **for** every variable $T \in PC(C_i)$ s.t. $X \in PC(T)$
10. **if** $\neq gI(X, C_i|\mathbf{Z} \cup \{T\})$ **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 parents-children PC set for every member T of PC(C_i) (steps 2–4). 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 the set OPEN $= \mathbf{U} \setminus \{T\}$ by priority of inclusion according to $I(X, T)$, and discards variables having $I(X, T) = 0$ (steps 3 and 4). 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) [7].

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 the 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. $I(X, T|\mathbf{Z})$**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)*.

Unlike the HITON algorithm that only determines the Markov blanket of a single target variable, `MB-MBC` considers many target variables (all the class variables), then induces the MBC graphical structure. Given the MBC definition, the direct parents of any class variable $C_i, i = 1, \ldots, d$, 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 as follows:

- *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)\ PC(C_i), i.e., for every spouse X, we insert an arc from X to the corresponding common child given by $PC(X) \cap PC(C_i)$.

Note finally that, being based on HITON algorithm, that is proved to be scalable [7,8], and being defined as a filter constraint-based approach that locally determines the Markov blanket around each class variable, `MB-MBC` is scalable with respect to data set size and dimensionality. Moreover, using independence tests instead of accuracy metrics (used by wrapper approaches such as `CB-MBC` algorithm [3]) considerably lightens the computational burden, especially when the data sets include a large number of variables or instances.

3. Experimental study

3.1. Data sets

In order to evaluate our proposed approach, we firstly performed experiments on synthetic data sets, Yeast data set, then on a real-world Parkinson's disease data set.

3.1.1. Synthetic data sets

We randomly generated five data sets using the MBC structure presented in Fig. 1 containing four class and seven feature variables. Although the structure is the same, each data set is sampled from a different, randomly defined, probability distribution using the probabilistic logic sampling method [21]. All class and feature variables are binary, and each generated data set contains 2000 instances. Fivefold cross-validation experiments were carried out on each data set and each learning algorithm (see Section 3.2).

3.1.2. Yeast data set

The Yeast data set [22] is a biological data set where genes in the yeast Saccharomyces cerevisiae are associated with several biological functions at the same time. The Yeast data set contains in total 2417 instances. Genes are represented by 103 numeric feature variables, and biological functions are represented by 14 binary class variables. Since MBCs are defined for discrete variables, we used a static, global, supervised and top-down discretization algorithm, called class-attribute contingency coefficient [23], to discretize all the continuous feature variables. After discretization, each feature variable has four possible values ranging from 0 to 3. As with synthetic data sets, fivefold cross-validation was performed on the Yeast data set.

3.1.3. Parkinson's disease data set

Parkinson's disease data set was obtained from an international multipurpose database collected by the National Center of

Epidemiology, Carlos III Institute of Health, Madrid. Patients with diagnosis of PD, in all disease stages, were included. In total, the analyzed data set contains 488 patients. For each patient, we have information about the PDQ-39 items represented in Table 6 (i.e. 39 feature variables) each with five possible values: never, occasionally, sometimes, often and always, coded in the data set using numbers ranging from 0 (never) to 4 (always); and the corresponding EQ-5D (i.e. 5 class variables): `mobility`, `self-care`, `usual activities`, `pain/discomfort` and `anxiety/depression` coded with numbers ranging from 1 (no problems) to 3 (severe problems).

The objective is to simultaneously predict the 5 class values of EQ-5D from PDQ-39 using `MB-MBC` algorithm. Given the EQ-5D values, to complement them, the corresponding utility score (a.k.a. utility index) could also be induced using the UK general scoring system [24]. For instance, let's assume that we obtain an EQ-5D equal to $\mathbf{c} = (1,1,2,2,3)$ indicating that the considered patient has no problems with `mobility` and `self-care`; some problems with `usual activities` and `pain/discomfort`; and severe problems with `anxiety/depression`. Based on UK scoring system [24], EQ-5D utility index is $1 - 0.081 - 0.036 - 0.123 - 0.236 - 0.269 = 0.255$.

To summarize, the details of the considered data sets are given in Table 1.

3.2. Approaches

We compared `MB-MBC` against three Bayesian network-based approaches: `CB-MBC` a general approach for learning MBCs from data, `IndepMBs` a binary relevance version of `MB-MBC`, and `IndepPC-BNs` recently proposed in [25] to predict EQ-5D from Health Surveys SF-12. We also considered for comparison the two state-of-the-art multi-label classification methods `BP-MLL` and `ML-kNN`. Note that, multi-label classification methods can only deal with multi-dimensional problems where all class variables are binary. Therefore, `BP-MLL` and `ML-kNN` were only applied to synthetic and Yeast data sets, containing binary class variables, but not used to analyze the Parkinson's disease problem since the EQ-5D class variables are not binary. Moreover, we compared `MB-MBC` against three commonly used approaches to predict EQ-5D from HRQol specific instruments, namely, `MNL`, `OLS` and `CLAD`, the last two being based on the utility score. In what follows, we briefly present more details for each considered approach:

- *Class-bridge decomposable MBC algorithm* (`CB-MBC`) [3]: learns class-bridge (CB) decomposable MBCs based on a greedy forward selection wrapper approach optimizing the accuracy of the model given the training data set. `CB-MBC` 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 accordingly.
- *Independent Markov blankets algorithm* (`IndepMBs`): learns independently a Bayesian network classifier for each class variable using the same HITON algorithm [7,8]. Therefore, there are no arcs between class variables, the classification is independently performed for each class variable, and the individual results are then aggregated to form the predicted class vector.
- *Independent PC Bayesian networks* (`IndepPC-BNs`): The PC algorithm [20] is a constraint-based approach for learning Bayesian networks from data. It starts with a fully connected DAG, then sequentially removes edges between nodes based on statistical independence tests. Similar to Le and Doctor [25] that recently applied this approach to predict EQ-5D utility index from Health Surveys SF-12, we used the PC algorithm to learn independently a Bayesian network for each class variable in EQ-5D.
- *Back propagation for multi-label learning* (`BP-MLL`) [26]: is derived from the popular back propagation algorithm through modifying its error function with a new function that takes into account the characteristics of multi-label learning, i.e., the labels belonging to an instance should be ranked higher than those not belonging to that instance.
- *Multi-label k-nearest neighbor* (`ML-kNN`) [27]: It extends the *k*-nearest neighbor lazy algorithm (*k*NN) to the multi-label framework. Basically, for each test instance, it firstly identifies the *k* nearest neighbors in the training data, then, it predicts the label set based on the statistical information gained from the label sets of the neighboring instances and the maximum a posteriori principle.
- *Multinomial logistic regression* (`MNL`) [25,28]: Using the multinomial logistic regression on an input set of feature variables, this approach returns the estimated posterior probabilities of each class value; then the class value with the highest probability is selected. Similar to `IndepMBs` and `IndepPC-BNs`, `MNL` is applied independently to each class variable, and the results are aggregated to obtain the predicted class vector.
- *Ordinary least squares* (`OLS`): is one of the mostly used methods for mapping specific HRQoL instruments such as Health Surveys SF-12 and PDQ-8 into a generic utility index [25,29,30]. In the `OLS` model, the EQ-5D utility index is directly regressed on the PDQ-39 items. In other words, `OLS` does not provide the 5 estimated class values of EQ-5D, but only returns the estimated EQ-5D utility index.
- *Censored least absolute deviation* (`CLAD`) [31]: is a generalization of the least absolute deviations method. Similar to `OLS`, `CLAD` is widely used to convert specific HRQoL instruments into a generic utility index [25,29,32], and it only estimates EQ-5D utility index without predicting the 5 class values of EQ-5D.

Table 1
Description of the data sets used in the experiments

Data set	#Classes	#Features	#Instances
Synthetic	4	7	2000
Yeast	14	103	2417
Parkinson's disease	5	39	488

All methods were run in Matlab R2010b on a laptop 2.2 GHz with 6 GB RAM using Windows operating system. The HITON and PC algorithms were run using Causal Explorer Toolkit [33] provided as compiled Matlab functions, and G^2 statistical test was used to evaluate the conditional independencies between variables with a significance level $\alpha = 0.01$. `BP-MLL` and `ML-kNN` were run using the Matlab packages available at http://lamda.nju.edu.cn/datacode/BPMLL.htm and http://lamda.nju.edu.cn/datacode/MLkNN.htm, respectively. For the `BP-MLL` algorithm, the number of training epochs was set to 20, and the number of hidden neurons was set to 3 for the synthetic data set and 20 for the Yeast data set; and for `ML-kNN` the number of clusters was to set to 3 for the synthetic data set and 7 for the Yeast data set. Note also that `OLS` and `CLAD` were only considered for comparison on Parkinson's disease data set since they do not predict class values but directly return an utility index. Fig. 2 summarizes the approaches used for predicting EQ-5D from PDQ-39.

3.3. Evaluation metrics

We used the following metrics to assess the predictive performance of the considered approaches:

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

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

where N is the size of the test set, $\hat{c}_{li}$ is the C_i class value predicted by the model for sample l, and c_{li} denotes its corresponding true value. $\delta(\hat{c}_{li}, c_{li}) = 1$ if the predicted and true class values are equal, i.e., $\hat{c}_{li} = c_{li}$, and $\delta(\hat{c}_{li}, c_{li}) = 0$ otherwise.

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

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

In this more strict case, the (d-dim) vector of predicted classes $\hat{\mathbf{c}}_l$ is compared to the vector of true classes $\mathbf{c}_l$, so that we have $\delta(\hat{\mathbf{c}}_l, \mathbf{c}_l) = 1$ if both vectors are equal in all their components, i.e., $\hat{\mathbf{c}}_l = \mathbf{c}_l$, and $\delta(\hat{\mathbf{c}}_l, \mathbf{c}_l) = 0$ otherwise.

Moreover, for the experiments over PD data set, we used the following metrics, commonly used in comparison with MNL, OLS and CLAD methods [25]:

- The mean squared error (MSE) between the true and predicted EQ-5D utility scores.
- The mean absolute error (MAE) between the true and predicted EQ-5D utility scores.
- The square of the Pearson product-moment correlation (R^2) between the true and predicted EQ-5D utility scores.
- The absolute difference (AbsDiff) between the true and predicted EQ-5D utility mean scores, i.e. the absolute difference is computed between the mean of all true EQ-5D utility scores and the mean of all predicted EQ-5D utility scores.

Notice that different metrics guide the tested approaches. In fact, mean and global accuracy guide CB-MBC, MSE guides OLS, MAE guides CLAD, conditional likelihood guides MNL, whereas the rest of the approaches, MB-MBC, Indep-MBs and IndepPC-BNs, are guided by G^2 based hypothesis test.

4. Results and discussion

4.1. Synthetic data sets

Table 2 shows the classification performance results for the fivefold cross-validation experiments performed over the five synthetic data sets, with mean values and standard deviations for the mean and global accuracy and each method. The best result for each metric is written in bold. For the mean accuracy, MB-MBC and IndepMBs present a similar predictive performance (72%), slightly better than other approaches. For the global accuracy, MB-MBC outperforms the remaining approaches with more than 31%. Using the Friedman test followed by the Tukey–Kramer post hoc test with a significance level equal to 0.05, it turns out that MB-MBC is significantly better than BP-MLL and ML-kNN only for the global accuracy. For all remaining methods and for the mean accuracy, the differences in classification performance are not statistically significant.

Table 2
Estimated accuracies (mean ± std. deviation) over synthetic data set.

Method	Mean accuracy	Global accuracy
MB-MBC	**0.7204 ± 0.0313**	**0.3159 ± 0.0502**
CB-MBC	0.7145 ± 0.0321	0.2878 ± 0.0590
IndepMBs	0.7202 ± 0.0323	0.2903 ± 0.0535
IndepPC-BNs	0.7160 ± 0.0400	0.2724 ± 0.0626
BP-MLL	0.7119 ± 0.0617	0.2533 ± 0.1200
ML-kNN	0.6934 ± 0.0376	0.2505 ± 0.0450
MNL	0.7056 ± 0.0392	0.2739 ± 0.0464

4.2. Yeast data set

Table 3 reports the classification performance results for the fivefold cross-validation experiments performed over Yeast data set. The best result for each metric is written in bold. ML-kNN presents the best mean accuracy (78%) followed by MNL and CB-MBC with a slightly lower predictive performance. For the global accuracy, MB-MBC outperforms the remaining approaches with 14%. Using the Friedman test followed by the Tukey–Kramer post hoc test with a significance level equal to 0.05, it turns out that (1) for mean accuracy, ML-kNN and MNL are only significantly better than BP-MLL, and (3) for global accuracy, MB-MBC and MNL are only significantly better than IndepPC-BNs and BP-MLL. For all remaining methods, the differences in predictive performance are not statistically significant.

4.3. Parkinson's disease data set

For the experiments over the Parkinson's disease data set, we applied MB-MBC and IndepMBs with a restriction of the Markov blanket set of each class variable $MB(C_i)$ to the set of its parents-children $PC(C_i)$. This restriction was introduced based upon the theoretical discussion introduced by Aliferis et al. [8] and the empirical observation that including more spouses leads to a less accurate MBC classifier. In fact, Aliferis et al. [8] discussed in Section 4.6 five plausible scenarios explaining the better performance of substituting the PC set in place of the MB set. The third scenario applies in

Table 3
Estimated accuracies (mean ± std. deviation) over Yeast data set.

Method	Mean accuracy	Global accuracy
MB-MBC	0.7602 ± 0.0208	**0.1423 ± 0.0249**
CB-MBC	0.7762 ± 0.0067	0.0741 ± 0.0133
IndepMBs	0.7624 ± 0.0035	0.0608 ± 0.0156
IndepPC-BNs	0.7716 ± 0.0013	0.0240 ± 0.0197
BP-MLL	0.7417 ± 0.0180	0.0414 ± 0.0269
ML-kNN	**0.7815 ± 0.0061**	0.1018 ± 0.0098
MNL	0.7810 ± 0.0066	0.1262 ± 0.0163

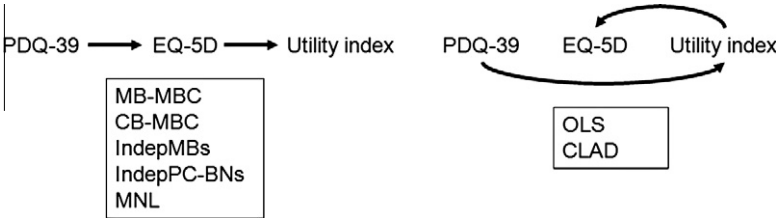

Fig. 2. Approached used for predicting EQ-5D from PDQ-39.

our case, where the spouses have connecting paths to the class variables that cannot be blocked due to the small sample size, i.e., the conditional independencies between the spouses and the class variables could not be established due to the small number of instances in the PD data set (including only 488 instances).

Table 4 shows the classification performance results for the fivefold cross-validation experiments performed on PD data set with mean values and standard deviations for each metric and each method. Recall that OLS and CLAD only return the utility index; thus, in order to compute the mean and global accuracies for OLS and CLAD, we proceeded by retrieving the EQ-5D class values as follows: first, we look for the utility index from the UK scoring list [24] closest to the one returned by OLS and CLAD, then we determine the EQ-5D vector corresponding to that index.

In Table 4, MB-MBC presents the best mean accuracy (71%), whereas surprisingly IndepMBs outperforms in the global accuracy the remaining approaches with 20%. As with the synthetic data sets, we ran a multiple comparison test of all method performance using the Friedman test followed by the Tukey–Kramer post hoc test with a significance level equal to 0.05. It turns out that, for both mean and global accuracy, MB-MBC and IndepMBs results are only significantly better than OLS and CLAD methods.

In addition, Table 5 presents results for MSE, MAE, R^2 and AbsDiff metrics. The best result for each metric is written in bold. Once again, MB-MBC outperforms other predictive approaches in terms of MSE and MAE. IndepMBs presents the best R^2 and MNL produces the best AbsDiff.

Note that, both OLS and CLAD methods performed poorly for all the performance metrics in Tables 4 and 5. As pointed out by Le and Doctor [25], this may be due to certain limitations of these regression methods such as predictive values that are outside the domain of the preference-based target, ceiling/floor effects, and assignment to health states that are not defined in the UK scoring list. Previous studies testing OLS and CLAD for predicting the EQ-5D utility index from the Health Surveys SF-12 [25,32], or from PDQ-8, the short version of PDQ-39 [10], proved that OLS and CLAD methods induce very similar results with a possible better performance of the simple OLS over the more theoretically justifiable CLAD. In our case, OLS method resulted in a slightly better MSE, MAE and R^2 than CLAD, but for the absolute difference between the true and the predicted EQ-5D mean scores, CLAD performed better.

Table 4
Estimated accuracies (mean ± std. deviation) over Parkinson's data set.

Method	Mean accuracy	Global accuracy
MB-MBC	**0.7119 ± 0.0338**	0.2030 ± 0.0718
CB-MBC	0.6807 ± 0.0285	0.1865 ± 0.0429
IndepMBs	0.7009 ± 0.0427	**0.2051 ± 0.0835**
IndepPC-BNs	0.6587 ± 0.0636	0.1867 ± 0.0937
MNL	0.6926 ± 0.0430	0.1802 ± 0.0713
OLS	0.4201 ± 0.0252	0.0123 ± 0.0046
CLAD	0.4254 ± 0.0488	0.0143 ± 0.0171

MNL performed quite well compared to OLS and CLAD as well as compared to the Bayesian network-based approaches. For instance, it had better results for mean accuracy, MSE, MAE, and R^2 than IndepPC-BNs and CB-MBC; it also resulted in the best AbsDiff compared to all the remaining approaches. However, MNL presented a lower global accuracy compared to all Bayesian network-based approaches. This can be explained by the fact that taking into account the probabilistic dependence relationships among class and feature variables ensures a better predictive performance, and in this context, MB-MBC and IndepMBs performed better than CB-MBC and IndepPC-BNs through determining the Markov blanket around each class variable.

Moreover, contrary to MNL, OLS and CLAD, Bayesian network-based approaches present also the merit of representing the relationships between all variables through their graphical structure component. In our study, in order to investigate the dependence relationships among EQ-5D and PDQ-39 variables, we first examine in Fig. 3 the graphical structure of the MBC network learnt by the MB-MBC algorithm from the full PD data set, then compare it to the graphical structures learnt by CB-MBC, IndepMBs and IndepPC-BNs.

Firstly, the class subgraph in Fig. 3 (red arcs) shows associations between the three class variables mobility, self-care and usual activities which may reveal the strong relevance between these classes. Pain/discomfort is not directly related to any other class variable, but its Markov blanket includes the class usual activities which proves as well the strong relevance between both classes. Anxiety/depression has no direct connections with the remaining classes. This can be explained by the fact that anxiety/depression is more related to emotional problems rather than physical health problems (i.e. mobility, self-care, usual activities and pain/discomfort).

Secondly, the bridge subgraph (blue arcs) reveals direct dependence relationships between EQ-5D classes and PDQ-39 features. Table 6 lists the PDQ-39 features grouped into eight domains: Mobility, activities of daily living, emotional well-being, stigma, social support, cognitions, communication and bodily discomfort. Each domain is depicted in Fig. 3 with a different color. We have the following dependence relationships from EQ-5D to PDQ-39:

- Mobility is directly associated with pdq1, pdq4, pdq6, and pdq7.
- Self-care is directly associated with pdq12.
- Usual activities is directly associated with pdq1, pdq2, pdq3, pdq5, pdq6, pdq8, pdq10, pdq12 and pdq15.
- Pain/discomfort is directly associated with pdq2, pdq3, pdq38.
- Anxiety/depression is directly associated with pdq17, pdq18.

Note that the detected associations are very appropriate and clearly related to mobility, self-care and usual activities. The selected pdq variables associated with these three class variables exclusively pertain to *mobility* and *activities of daily living*

Table 5
MSE, MAE, R^2 and AbsDiff (mean ± std. deviation) over Parkinson's data set.

Method	MSE	MAE	R^2	AbsDiff
MB-MBC	**0.0650 ± 0.0156**	**0.1737 ± 0.0316**	0.5996 ± 0.0683	0.0659 ± 0.0373
CB-MBC	0.0905 ± 0.0167	0.1973 ± 0.0298	0.4094 ± 0.0860	0.0790 ± 0.0541
IndepMBs	0.0699 ± 0.0188	0.1784 ± 0.0328	**0.6026 ± 0.0653**	0.0737 ± 0.0315
IndepPC-BNs	0.0909 ± 0.0909	0.2026 ± 0.0391	0.4602 ± 0.1379	0.0863 ± 0.0670
MNL	0.0759 ± 0.0152	0.1922 ± 0.0284	0.4935 ± 0.0961	**0.0503 ± 0.0331**
OLS	0.1832 ± 0.0373	0.3502 ± 0.0422	0.0186 ± 0.0177	0.0943 ± 0.0388
CLAD	0.1962 ± 0.0360	0.3583 ± 0.0395	0.0165 ± 0.0155	0.0779 ± 0.0278

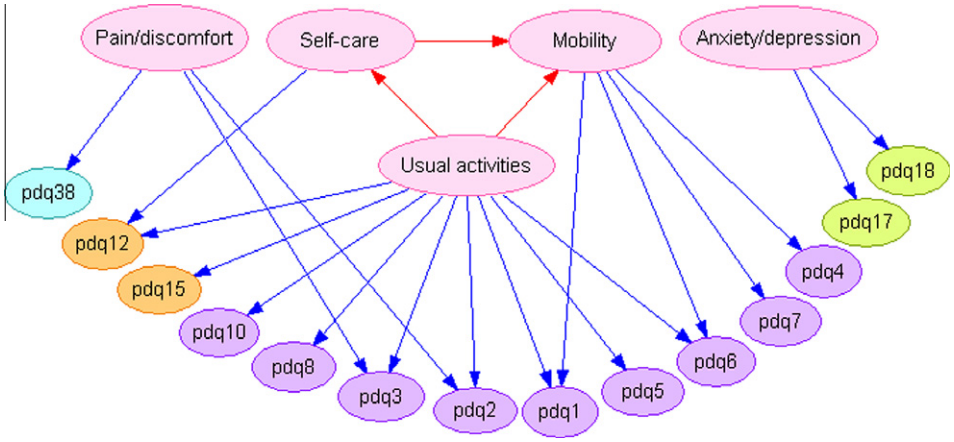

Fig. 3. MBC graphical structure with 5 classes (EQ-5D) and 39 features (PDQ-39) learnt by `MB-MBC`. The class subgraph (red arcs) shows probabilistic dependence relationships between classes, the bridge subgraph (blue arcs) shows dependence relationships from classes to features, and the feature subgraph is empty. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6
The Parkinson's Disease Questionnaire (PDQ-39) items.

Mobility	
pdq1	Had difficulty doing the leisure activities you would like to do
pdq2	Had difficulty after your home e.g. DIY, housework, cooking
pdq3	Had difficulty carrying bags of shopping
pdq4	Had problems walking half a mile
pdq5	Had problems walking 100 yards
pdq6	Had problems getting around the house as easily as you would like
pdq7	Had problems getting around in public
pdq8	Needed someone else to accompany you when you went out
pdq9	Felt frightened or worried about falling over in public
pdq10	Been confined to the house more than you would like
Activities of daily living	
pdq11	Had difficulty washing yourself
pdq12	Had difficulty dressing yourself
pdq13	Had problems doing up buttons or shoe laces
pdq14	Had problems writing clearly
pdq15	Had difficulty cutting up your food
pdq16	Had difficulty holding a drink without spilling it
Emotional well-being	
pdq17	Felt depressed
pdq18	Felt isolated and lonely
pdq19	Felt weepy or tearful
pdq20	Felt angry or bitter
pdq21	Felt anxious
pdq22	Felt worried about your future
Stigma	
pdq23	Felt you had to conceal you Parkinson's from people
pdq24	Avoided situations which involve eating or drinking in public
pdq25	Felt embarrassed in public due to having Parkinson's disease
pdq26	Felt worried by other people's reaction to you
Social support	
pdq27	Had problems with your close personal relationships
pdq28	Lacked support in the ways you need from your spouse or partner
pdq29	Lacked support in the ways you need from your family or close friends
Cognitions	
pdq30	Unexpectedly fallen asleep during the day
pdq31	Had problems with your concentration, e.g. when reading or watching TV
pdq32	Felt your memory was bad
pdq33	Had distressing dreams or hallucinations
Communication	
pdq34	Had difficulty with your speech
pdq35	Felt unable to communicate with people properly
pdq36	Felt ignored by people
Bodily discomfort	
pdq37	Had painful muscle cramps or spasms
pdq38	Had aches and pains in your joints or body
pdq39	Felt unpleasantly hot or cold

domains, which cover the main information about PD. In fact, PD is a neurodegenerative disorder characterized by motor manifestations (bradykinesia, rest tremor and balance impairment) and non-motor symptoms (depression, psychosis and sleep disturbance) [34]; and for the huge majority of PD patients, from earliest to most advanced stages, the common perceived health problems are reflected as limitations for *mobility* and *activities of daily living*, whereas the most prevalent non-motor symptoms are associated with the impact on patients' health status perception (pain and depression, for example).

For `pain/discomfort`, the associations are well explained, as a whole, from the point of view of bone and joint disorders (pdq38), mainly. There are other types of pain in PD, more difficult to associate with these findings. Perhaps, pain due to dystonia in off state can also be related, in some way, with those selected PDQ-39 items. Distinction between the types of pain that may be present in PD is not easy.

For `anxiety/depression`, it seems that depression is quite well represented by the detected pdq items, but not the anxiety. In fact, there are PDQ-39 items about anxiety (pdq21 about feeling anxious, pdq22 about feeling worried about the future), but they do not appear in association with the EQ-5D class variable `anxiety/depression`. This can be explained by the well-known close relationship between depression and anxiety. A useful representation of this connection is made evident with the Hospital Anxiety and Depression scale, a measure furnishing scores for rating anxiety and depression separately but also usable as a unique score of emotional distress [35,36].

Taking the previous arguments into account, the findings in this study have sense from a clinical point of view. Moreover, EQ-5D is more restricted in content than the PDQ-39, explaining why several components of the PDQ-39, in addition to the most immediately relatable (for example, pain), can converge on a domain of the EQ-5D. Therefore, we may conclude that the combination of the selected variables in the network properly represents the relationships between the generic (EQ-5D) and specific (PDQ-39) instruments, and covers both motor and non-motor symptoms of PD.

The feature subgraph is empty due to the restriction of the Markov blanket set of each class variable $MB(C_i)$ to the set of its parents children $PC(C_i)$, that is, no feature spouses are allowed and thereby the parents of each feature variable can be only among class variables. Finally, notice that several features are also not present in Fig. 3 since no associations were detected between them and the EQ-5D class variables. These features are considered irrelevant

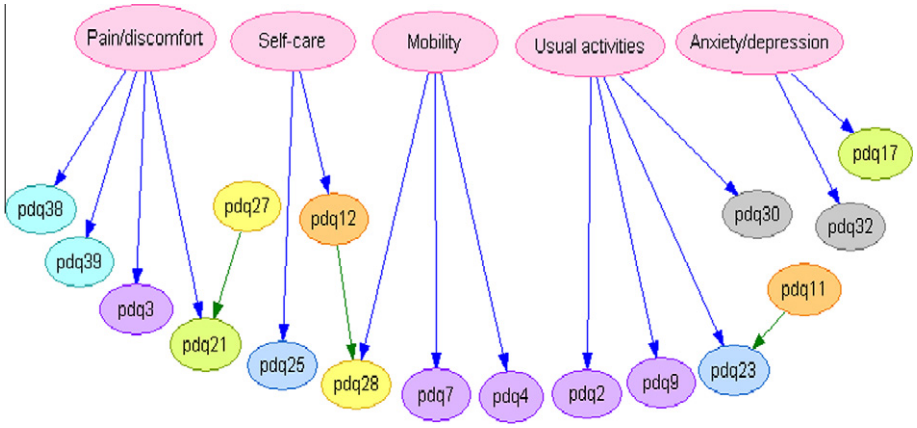

Fig. 4. MBC graphical structure learnt by CB-MBC.

and this may be due to the lack of instances of these features and/or their low interactions with the other variables in the analyzed data set.

For structural comparison, we depict in Figs. 4–6, the graphical structures learnt by CB-MBC, IndepMBs and IndepPC-BNs from the full PD data set, respectively.

As shown in Fig. 4, CB-MBC fails to detect any direct dependence relationship among the EQ-5D class variables, i.e., the class subgraph is empty. For the bridge subgraph, and similar to MB-MBC, the following dependence relationships are detected: mobility is associated with pdq4 and pdq7, self-care with pdq12, usual activities with pdq2, pain/discomfort with pdq3 and pdq38, and anxiety/depression with pdq17. Moreover, additional arcs are discovered in the bridge subgraph between the EQ-5D class variables and pdq9, pdq21, pdq23, pdq25, pdq28, pdq30, pdq32 and pdq39. Regarding the feature subgraph (green arcs), only three arcs were added between pdq11 and pdq23, pdq12 and pdq28, and pdq21 and pdq27.

Fig. 5 shows the five Markov blanket-based Bayesian network classifiers learnt independently for each EQ-5D class variable by IndepMBs. Being based on the same HITON algorithm [7,8], IndepMBs discovered similar dependence relationships between the EQ-5D and the pdq items, as MB-MBC does. However, as it can be observed, the main drawback of IndepMBs is its inability to detect the dependence relationships between the different EQ-5D class variables and their simultaneous interactions with the pdq items.

Additionally, Fig. 6 presents the graphical structure of the Bayesian network learnt by IndepPC-BNs for the mobility class variable. Many dependence relationships are added between the pdq items, and as determined by MB-MBC and Indep-MBs, mobility is only directly associated with pdq1, pdq4, pdq6 and pdq7. Similar conclusions are obtained for the Bayesian network graphical structures learnt by IndepPC-BNs for the class variables self-care, usual activities, pain/discomfort, and anxiety/depression (details and graphs not shown). As in IndepMBs, the main drawback of IndepPC-BNs is that each network is learnt independently for each class variable, and thus interactions between class variables could not be detected.

Finally, the computation times consumed by each method on each data set are reported in Table 7, measured in seconds for synthetic and Parkinson's disease data sets, and in minutes for Yeast data set. BP-MLL and ML-*k*NN were not applied to Parkinson's disease data set, and thus no computation times are available for them on that data set. In addition, no computation times are available for OLS and CLAD for synthetic and Yeast data sets because both OLS and CLAD were only applied to Parkinson's disease data set.

As shown in Table 7, MNL is the fastest on synthetic data set, ML-*k*NN is the fastest on Yeast data set, while OLS is the fastest on Parkinson's disease data set. MB-MBC is quite efficient and requires less time than IndepMBs and IndepPC-BNs. IndepPC-BNs may take more time since it builds a full Bayesian network for each class variable whereas MB-MBC and IndepMBs only determine the Markov blanket around each class variable; this is mainly noticed on Yeast data set that contains a larger number of variables (14 class variables and 103 feature variables). CB-MBC takes always the longest

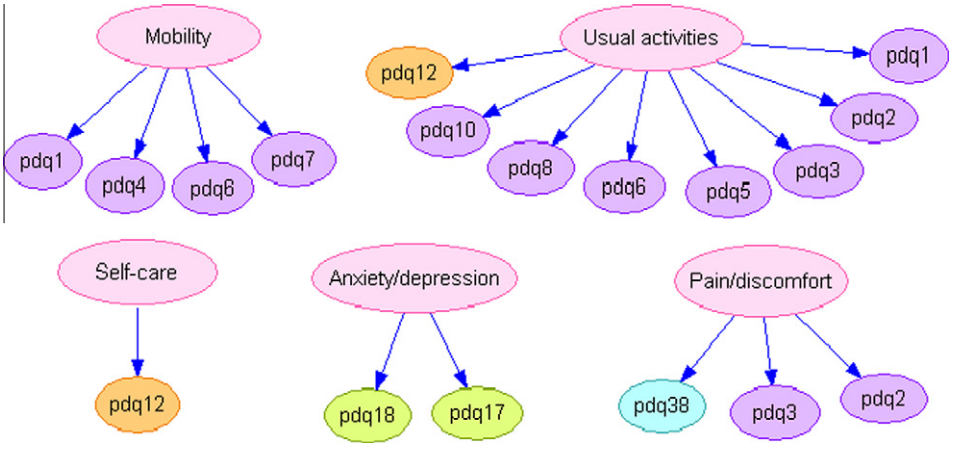

Fig. 5. Graphical structures learnt by IndepMBs for the 5 classes EQ-5D.

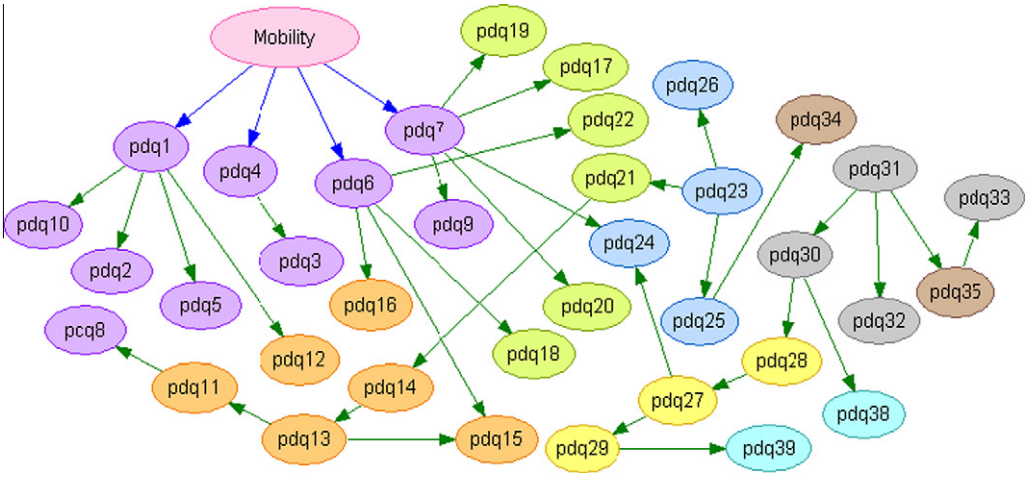

Fig. 6. Graphical structure of the Bayesian network learnt by `IndepPC-BNs` for the `mobility` class variable.

Table 7
Computation times of the learning process of each method and each data set.

Method	Synthetic (s)	Yeast (min)	Parkinson (s)
`MB-MBC`	0.57 ± 0.10	10.02 ± 1.99	13.38 ± 0.72
`CB-MBC`	44.35 ± 9.11	2370.69 ± 69.93	757.51 ± 47.91
`IndepMBs`	0.60 ± 0.08	17.00 ± 3.57	34.98 ± 1.57
`IndepPC-BNs`	0.67 ± 0.06	119.72 ± 17.68	23.40 ± 0.38
`BP-MLL`	2562.06 ± 171.61	69.61 ± 1.57	
`ML-`*k*`NN`	3.20 ± 0.09	**0.10 ± 0.01**	
`MNL`	**0.54 ± 0.04**	1.07 ± 0.30	10.56 ± 4.06
`OLS`			**0.72 ± 0.08**
`CLAD`			0.75 ± 0.21

computation time compared to the other Bayesian network-based methods since it is based on a wrapper approach. `BP-MLL` consumes more computation times than `MB-MBC`, `IndepMBs`, `MNL`, and `ML-`*k*`NN` on both synthetic and Yeast data sets, and this is mainly due to its complex error function which needs to be optimized through an iterative learning process [26]. Generally speaking, `ML-`*k*`NN`, regression and filter approaches require less computation time, whereas the wrapper approach always requires more computation time.

5. Conclusion

We proposed `MB-MBC`, a new approach for learning multi-dimensional Bayesian network classifiers from data. Our approach is general and can be applied to any multi-dimensional classification problem where an instance has to be assigned to more than one class variable. In this study, we presented its application to the problem of predicting the European Quality of Life-5 Dimensions EQ-5D from the PDQ-39 items. Experimental results on synthetic and Yeast data sets, as well as a real-world Parkinson's disease data set were encouraging in terms of the predictive performance and the identification of dependence relationships among class and feature variables, compared with state-of-the-art approaches.

In the future, we intend to carry out a more extensive experimental study using additional synthetic and real data sets. Moreover, it will be interesting to extend `MB-MBC` to deal with the challenging task of multi-dimensional classification for evolving data streams. This may require the development of an appropriate method to detect the changes in the stream, then extending `MB-MBC` to ensure a local updating of the current MBC over time, without a need to relearn it from scratch.

Acknowledgments

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

References

[1] Tsoumakas G, Katakis I, Vlahavas I. Mining multi-label data. In: Data mining and knowledge discovery handbook. Springer; 2010.
[2] Bielza C, Li G, Larrañaga P. Multi-dimensional classification with Bayesian networks. Int J Approx Reason 2011;52(6):705–27.
[3] 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; 2010. p. 25–32.
[4] van der Gaag LC, de Waal PR. Muti-dimensional Bayesian network classifiers. In: Proceedings of the third European conference on probabilistic, graphical models; 2006. p. 107–14.
[5] De Waal PR, van der Gaag LC. Inference and learning in multi-dimensional Bayesian network classifiers. In: Proceedings of the ninth European conference on symbolic and quantitative approaches to reasoning with uncertainty. Lecture notes in computer science; 2007. p. 501–11.
[6] Zaragoza JH, Sucar LE, Morales EF, Larrañaga P, Bielza C. Bayesian chain classifiers for multidimensional classification. In: Proceedings of the twenty-second international joint conference on artificial intelligence; 2011. p. 2192–97.
[7] 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. J Mach Learn Res 2010;11:171–234.
[8] 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. J Mach Learn Res 2010;11:235–84.
[9] Group EuroQol. EuroQol – a new facility for the measurement of health-related quality of life. The EuroQol group. Health Policy 1990;16(3):199–208.
[10] Cheung K, Oemar M, Oppe M, Rabin R. EQ-5D user guide. Basic information on how to use the EQ-5D, version 2.0. EuroQol Group 2009. <http://www.euroqol.org>.
[11] Schrag A, Selai C, Jahanshahi M, Quinn NP. The EQ-5D – a generic quality of life measure – is a useful instrument to measure quality of life in patients with Parkinson's disease. J Neurol Neurosurg Psychiatry 2000;69(1):67–73.
[12] Martínez-Martín P, Jeukens-Visser M, Lyons KE, Rodriguez-Blazquez C, Selai C, Siderowf A, et al. Health-related quality of life scales in Parkinson's disease: Critique and recommendations. Mov Disord 2011;26(13):2371–80.
[13] Peto V, Jenkinson C, Fitzpatrick R, Greenhall R. The development and validation of a short measure of functioning and well being for individuals with Parkinson's disease. Qual Life Res 1995;4(3):241–8.
[14] Peto V, Jenkinson C, Fitzpatrick R. PDQ-39: a review of the development, validation and application of a Parkinson's disease quality of life questionnaire and its associated measures. J Neurol 1998;245(Suppl. 1):S10–4.

[15] Jenkinson C, Fitzpatrick R, Peto V, Greenhall R, Hyman N. The Parkinson's Disease Questionnaire (PDQ-39): development and validation of a Parkinson's disease summary index score. Age Ageing 1997;26(5):353–7.
[16] Jenkinson C, Fitzpatrick R, Norquist J, Findley L, Hughes K. Cross-cultural evaluation of the Parkinson's Disease Questionnaire: tests of data quality, score reliability, response rate, and scaling assumptions in the United States, Canada, Japan, Italy, and Spain. J Clin Epidemiol 2003;56(9):843–7.
[17] Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann Publishers; 1988.
[18] Koller D, Friedman N. Probabilistic graphical models principles and techniques. Cambridge: MIT Press; 2009.
[19] Pearl J, Verma TS. Equivalence and synthesis of causal models. In: Proceedings of the sixth conference on uncertainty in artificial intelligence; 1990. p. 220–7.
[20] Spirtes P, Glymour C, Scheines R. Causation, prediction, and search. 2nd ed. Cambridge: MIT Press; 2000.
[21] Henrion M. Propagating uncertainty in Bayesian networks by probabilistic logic sampling. In: Proceedings of the fourth conference on the uncertainty in artificial intelligence; 1988. p. 14–163.
[22] Elisseeff A, Weston JA. kernel method for multi-labelled classification. Adv Neural Inf Process Syst 2002;14:681–7.
[23] Cheng-Jung T, Chien-I L, Wei-Pang YA. discretization algorithm based on class-attribute contingency coefficient. Inform Sci 2008;178(3):714–31.
[24] EuroQuol Group. A1 tariff based on UK survey. EuroQuol EQ-5D user guide; 1993.
[25] Le QA, Doctor JN. Probabilistic mapping of descriptive health status responses onto health state utilities using Bayesian networks: an empirical analysis converting SF-12 into EQ-5D utility index in a national US sample. Med Care 2011;49(5):451–60.
[26] Zhang ML, Zhou ZH. Multi-label neural networks with applications to functional genomics and text categorization. IEEE Trans Knowledge Data Eng 2006;18(10):1338–51.
[27] Zhang ML, Zhou ZH. ML-kNN: a lazy learning approach to multi-label learning. Pattern Recognit 2007;40(7):2038–48.
[28] Gray A, Rivero-Arias O, Clarke P. Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. Med Decis Making 2006;26(1):18–29.
[29] Cheung YB, Tan LC, Lau PN, Au WL, Luo N. Mapping the eight-item Parkinson's Disease Questionnaire (PDQ-8) to the EQ-5D utility index. Qual Life Res 2008;17(9):1173–81.
[30] Franks P, Lubetkin DI, Gold MR, Tancredi DJ, Jia H. Mapping the SF-12 to the EuroQoL EQ-5D index in a national US sample. Med Decis Making 2004;24(3):247–54.
[31] Powell JL. Least absolute deviations estimation for the censored regression model. J Economet 1984;25:303–25.
[32] Sullivan PW, Ghushchyan V. Mapping the EQ-5D index from the SF-12: US general population preferences in a nationally representative sample. Med Decis Making 2006;26(4):401–9.
[33] Aliferis CF, Tsamardinos I, Statnikov A. Causal explorer: a probabilistic network learning toolkit for discovery; 2010. <http://www.discover.mc.vanderbilt.edu/discover/public/causal_explorer/>.
[34] Lees AJ, Hardy J, Revesz T. Parkinson's disease. Lancet 2009;373(9680):2055–66.
[35] Rodriguez-Blazquez C, Frades-Payo B, Forjaz MJ, de Pedro-Cuesta J, Martínez-Martín P. On behalf of the Longitudinal Parkinson's Disease Patient Study (Estudio longitudinal de pacientes con enfermedad de Parkinson-ELEP) Group. Psychometric attributes of the hospital anxiety and depression scale in Parkinson's disease. Mov Disord 2009;24(4):519–25.
[36] Forjaz MJ, Rodriguez-Blazquez C, Martínez-Martín P. On behalf of the Longitudinal Parkinson's Disease Patient Study (Estudio longitudinal de pacientes con enfermedad de Parkinson-ELEP) Group. Rasch analysis of the hospital anxiety and depression scale in Parkinson's disease. Mov Disord 2009;24(4):526–32.