Biomedical Engineering: Applications, Basis and Communications, Vol. 26, No. 1 (2014) 1450015 (11 pages)
DOI: 10.4015/S101623721450015X

PREDICTING THE EQ-5D FROM THE PARKINSON'S DISEASE QUESTIONNAIRE PDQ-8 USING MULTI-DIMENSIONAL BAYESIAN NETWORK CLASSIFIERS

Hanen Borchani[*,‖], Concha Bielza[*,‡], Pablo Martinez-Martin[†,§] and Pedro Larrañaga[*,¶]

[*]*Computational Intelligence Group*
Departamento de Inteligencia Artificial
Facultad de Informática
Universidad Politécnica de Madrid
Boadilla del Monte, 28660, Madrid, Spain

[†]*CIEN Foundation and CIBERNED*
Carlos III Institute of Health
Alzheimer Center Reina Sofía Foundation
Valderrebollo 5, 28031, Madrid, Spain

[‡]*mcbielza@fi.upm.es*
[§]*pmartinez@fundacioncien.es*
[¶]*pedro.larranaga@fi.upm.es*
[‖]*hanen.borchani@upm.es*

Accepted 10 June 2013
Published 26 February 2014

ABSTRACT

The impact of the Parkinson's disease and its treatment on the patients' health-related quality of life can be estimated either by means of generic measures such as the european quality of Life-5 Dimensions (EQ-5D) or specific measures such as the 8-item Parkinson's disease questionnaire (PDQ-8). In clinical studies, PDQ-8 could be used in detriment of EQ-5D due to the lack of resources, time or clinical interest in generic measures. Nevertheless, PDQ-8 cannot be applied in cost-effectiveness analyses which require generic measures and quantitative utility scores, such as EQ-5D. To deal with this problem, a commonly used solution is the prediction of EQ-5D from PDQ-8. In this paper, we propose a new probabilistic method to predict EQ-5D from PDQ-8 using multi-dimensional Bayesian network classifiers. Our approach is evaluated using five-fold cross-validation experiments carried out on a Parkinson's data set containing 488 patients, and is compared with two additional Bayesian network-based approaches, two commonly used mapping methods namely, ordinary least squares and censored least absolute deviations, and a deterministic model. Experimental results are promising in terms of predictive performance as well as the identification of dependence relationships among EQ-5D and PDQ-8 items that the mapping approaches are unable to detect.

Keywords: Parkinson's disease; EQ-5D; PDQ-8; Health-related quality of life; Bayesian networks.

[‖]Corresponding author: Hanen Borchani, Computational Intelligence Group, Departamento de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, Boadilla del Monte, 28660, Madrid, Spain. Tel.: +34 91 3363675; Fax: +34 91 3524819; E-mail:hanen.borchani@upm.es

INTRODUCTION

Parkinson's Disease (PD) is a neurodegenerative disorder characterized by motor manifestations (bradykinesia, rest tremor and balance impairment) and nonmotor symptoms (depression, psychosis and sleep disturbance).[1] Health-related quality of life (HRQoL) is a patient-reported outcome reflecting the PD impact on the physical, mental, functional and social aspects of life which are important for the individual.[2]

HRQoL measures can be categorized into generic and specific. *Generic measures* are usable in general populations and in any disorder. The European Quality of Life-5 Dimensions (EQ-5D) is considered a valid generic preference-based HRQoL instrument and is recommended for evaluation of HRQoL in PD.[3–5] EQ-5D contains five items: `mobility`, `self-care`, `usual activities`, `pain/discomfort` and `anxiety/depression`, each has three options of response: no problems, some problems and severe problems. Hence, the number of all possible EQ-5D item value combinations is 243. Each possible combination corresponds to a *health state*, which can be then quantified using a *utility score*, a.k.a. *utility index*. Based on the UK scoring system,[6] this *utility score* may range from −0.594 (i.e. worse health state where all EQ-5D items report severe problems) to 1 (i.e. best health state where all EQ-5D items report no problems).[7,8]

On the contrary, *specific measures* are usable only in the population for which they were designed and cover the most important areas of interest in that setting. The 8-item Parkinson's disease questionnaire (PDQ-8) is defined as the short version of the PDQ-39 and is recommended for use in PD patients.[9] It includes 8 items represented in Table 1. Each item represents a domain of the PDQ-39 (i.e. mobility, activities of daily living, emotional well-being, social support, cognition, communication, bodily discomfort and stigma), and scores on a five-point scale: never, occasionally, sometimes, often and always.[10]

In clinical studies, PDQ-8 could be used in detriment of EQ-5D due to the excessive burden for the respondents to assess two questionnaires simultaneously, or the lack of resources and time. This also may be due to the lack of the clinical interest in generic measures, and the relative difficulty for the calculation of EQ-5D utility index and the interpretation of its outcomes. Nevertheless, PDQ-8 cannot be directly applied in cost-effectiveness analyses which require generic measures and quantitative utility scores, such as EQ-5D. To deal with this problem, a commonly used solution is the prediction of EQ-5D from disease-specific measures. For instance, several studies have been proposed to map the EQ-5D utility score from the Health Surveys SF-12[8,11–13] and SF-36.[14] Moreover, in a more related work, Cheung *et al.*[15] developed several functions for generating the EQ-5D utility index from PDQ-8.

Most of these studies were mainly based on ordinary least squares (`OLS`) or censored least absolute deviation (`CLAD`) regression methods. However, Le and Doctor[8] recently discussed certain limitations of these regression methods (such as predictive values outside the range of the EQ-5D utility scores and ceiling/floor effects) and proposed a probabilistic mapping of Health Surveys SF-12 into EQ-5D using Bayesian networks. Specifically, Le and Doctor proved that Bayesian networks consistently outperformed the commonly used regression methods and pointed out the merits of the Bayesian network graphical component, which may be useful for researchers in further investigating the correlational relationships among generic and specific measures.

In this study, we propose a new probabilistic approach to predict EQ-5D from PDQ-8 using multi-dimensional Bayesian network classifiers (MBCs). Contrary to Le and Doctor's method[8] that learns an independent Bayesian network for each EQ-5D item, our approach builds a single MBC identifying interactions among all variables involved in EQ-5D and PDQ-8. In fact, taking into account the dependence relationships among EQ-5D items is crucial for both better prediction performance and graphical structure interpretation.

We evaluate our approach on a PD data set containing 488 patients, and compare it against two different Bayesian network-based approaches, namely class-bridge decomposable MBC (`CB-MBC`) and independent Markov blankets (`IndepMBs`), as well as against the models proposed by Cheung *et al.*,[15] i.e. `OLS-1`, `OLS-2`, `CLAD` and the `Deterministic` model. Experimental results are promising in terms of the predictive performance and the identification of probabilistic dependence relationships between EQ-5D and PDQ-8 items that other mapping approaches are unable to detect.

MATERIALS AND METHODS

Materials

Bayesian networks

A Bayesian network[16,17] over a set of discrete variables $\mathbf{U} = \{X_1, \ldots, X_n\}$, $n \geq 1$, is a probabilistic graphical model. Its first component $\mathcal{G}$ is a directed acyclic graph (DAG) where nodes correspond to variables in $\mathbf{U}$ and arcs represent probabilistic conditional dependencies

among these variables. Its second component is the parameters that define the set of conditional probability distributions of each variable X_i given the set of its parents $\mathbf{Pa}(X_i)$ in $\mathcal{G}$, i.e. those nodes directly directed to X_i. A Bayesian network defines a joint probability distribution factorized according to its structure $\mathcal{G}$:

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

Example 1. Figure 1 shows an example of a Bayesian network including five binary variables. For instance, node B has only node A as parent, and thus its probability distribution is conditioned only on the values of A. The joint probability distribution corresponding to this Bayesian network can be then computed as: $p(A, B, C, D, E) = p(A) \cdot p(B\,|A) \cdot p(C\,|A) \cdot p(D\,|B, C) \cdot p(E\,|C)$.

Moreover, an important notion of Bayesian networks is the *Markov blanket*. The Markov blanket for a variable X, denoted $\mathrm{MB}(X)$, is the minimal set of variables conditioned by which X is conditionally independent of all the remaining variables. Graphically, $\mathrm{MB}(X)$ consists of the union of the set of parents, children and parents of children (i.e. spouses) of X in $\mathcal{G}$.[18] For instance, in Fig. 1, the Markov blanket of variable B includes its parent A, its child D and its spouse C; that is, $\mathrm{MB}(B) = \{A, D, C\}$ that renders every other variable irrelevant to B.

Multi-dimensional Bayesian network classifiers

In our study, the prediction of EQ-5D values from PDQ-8 is modeled as a multi-dimensional classification problem where each instance given by an input vector of 8 features $\mathbf{x} = (x_1, \ldots, x_8)$ (i.e. PDQ-8) has to be associated with a predicted vector of 5 class values $\mathbf{c} = (c_1, \ldots, c_5)$ (i.e. EQ-5D). To deal with this problem, we use MBCs.

An MBC[19,20] is a Bayesian network where the structure $\mathcal{G}$ has a restricted topology. The set of n nodes is partitioned into two subsets: d class variables (in our case, the EQ-5D items to be predicted), $d \geq 1$, and m feature variables (in our case, the PDQ-8 items), $m \geq 1$ and $d + m = n$. The structure $\mathcal{G}$ is partitioned 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. Similar to general Bayesian networks, the parameters of an MBC consist of the conditional probability distribution of each variable given the set of its parents in the structure $\mathcal{G}$.

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

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

Example 2. An example of an MBC structure with its different subgraphs is shown in Fig. 2. It contains three class variables $\{C_1, C_2, C_3\}$, and five features variables

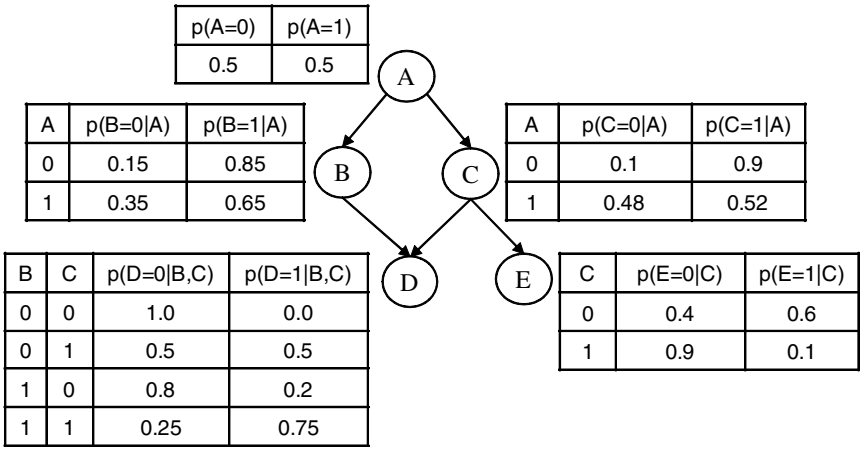

Fig. 1 An example of a Bayesian network.

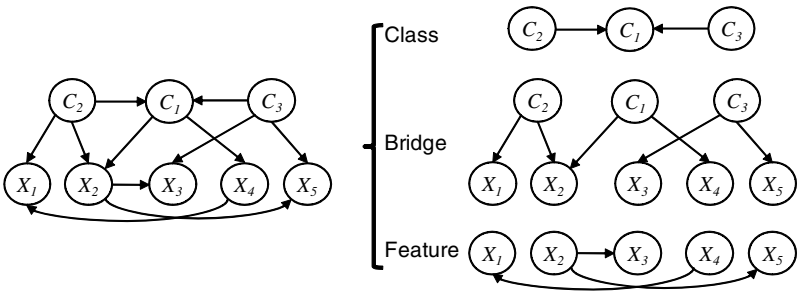

Fig. 2 An example of an MBC structure with its class, bridge and feature subgraphs.

$\{X_1, X_2, X_3, X_4, X_5\}$. Using Eq. (1), we have

$$\begin{aligned}\max_{c_1,c_2,c_3} p(C_1 = c_1, C_2 = c_2, C_3 = c_3 \,|\mathbf{x}) \\ = \max_{c_1,c_2,c_3} [p(c_1 \,|c_2, c_3)p(c_2)p(c_3) \\ \cdot p(x_1 \,|c_2, x_4)p(x_2 \,|c_1, c_2) \\ \cdot p(x_3 \,|c_3, x_2)p(x_4 \,|c_1)p(x_5 \,|c_3, x_2)]\cdot\end{aligned}$$

The MB-MBC algorithm

Let $\mathcal{D}$ be a data set of N instances containing a value assignment for each variable $X_1, \ldots, X_m, C_1, \ldots, C_d$, that is, $\mathcal{D} = \{(\mathbf{x}^{(1)}, \mathbf{c}^{(1)}), \ldots, (\mathbf{x}^{(N)}, \mathbf{c}^{(N)})\}$. Our approach, named Markov blanket MBC (`MB-MBC`), aims to find an MBC that best describes the dependence relationships in the available data set. It firstly consists of determining the Markov blanket around each class variable using the HITON algorithm and then specifying the directionality over the MBC subgraphs.

The HITON algorithm has been proposed by Aliferis *et al.*[21,22] Briefly, it identifies the Markov blanket of a variable X, based on statistical independence tests, in a two-phase scheme:

- HITON-PC: determines the parents and children of X, denoted $\mathrm{PC}(X)$.
- HITON-MB: initializes the Markov blanket set of X, denoted $\mathrm{MB}(X)$, with $\mathrm{PC}(X)$; then identifies and includes in $\mathrm{MB}(X)$ the rest of the parents of the children of X (i.e. the spouses of X).

In our case, with more than one class variable, the first step in `MB-MBC` is to apply the HITON algorithm to each class variable C_i to determine its Markov blanket $\mathrm{MB}(C_i)$. Then, the second step is to induce the MBC graphical structure 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 parents-children set $\mathrm{PC}(C_i)$. Then, we direct all these edges using the PC algorithm's edge orientation rules.[23] This allows to identify the set of class parents and class children belonging to the Markov blanket of C_i.
- *Bridge subgraph*: this is built by inserting an arc from each class variable C_i to every feature variable belonging to $\mathrm{PC}(C_i)$, and specifies thereby the set of feature children belonging to the Markov blanket of C_i.
- *Feature subgraph*: this consists of determining the set of spouses of C_i. In fact, for every feature X in the set $\mathrm{MB}(C_i)\backslash\mathrm{PC}(C_i)$, i.e. for every spouse X of C_i, we insert an arc from X to the corresponding common child given by $\mathrm{PC}(X) \cap \mathrm{PC}(C_i)$.

Example 3. Let us assume that we apply HITON algorithm to a data set coming out of the MBC structure of Fig. 2. By the end of HITON-PC and HITON-MB algorithms, we identify, respectively, the parents-children and the Markov blanket sets of each class variable:

- $\mathrm{PC}(C_1) = \{C_2, C_3, X_2, X_4\}$; $\mathrm{MB}(C_1) = \mathrm{PC}(C_1)$
- $\mathrm{PC}(C_2) = \{C_1, X_1, X_2\}$;
  $\mathrm{MB}(C_2) = \{C_1, C_3, X_1, X_2, X_4\}$
- $\mathrm{PC}(C_3) = \{C_1, X_3, X_5\}$;
  $\mathrm{MB}(C_3) = \{C_1, C_2, X_2, X_3, X_5\}$

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; and from C_3 to X_3 and X_5.
- *Feature subgraph*: given that $\mathrm{MB}(C_2)\backslash\mathrm{PC}(C_2) = \{X_4\}$, an arc is inserted from spouse X_4 to the common child X_1 determined by $\mathrm{PC}(X_4) \cap \mathrm{PC}(C_2) = \{X_1\}$. Similarly, given that $\mathrm{MB}(C_3)\backslash\mathrm{PC}(C_3) = \{X_2\}$ and $\mathrm{PC}(X_2) \cap \mathrm{PC}(C_3) = \{X_3, X_5\}$, arcs are inserted from X_2 to X_3 and X_5.

Experiments

Data

The used Parkinson's 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 by neurologists with expertise in movement disorders, and according to internationally recognized diagnostic criteria,[1] were followed up in movement disorder clinics. Patients in all stages of PD (Hoehn and Yahr 1 to 5) were included.

In total, the analyzed data set contains $N = 488$ patients, where 59.43% are male and 40.57% are female, and the average age for all patients is 65 years old (minimum = 30, maximum = 89). For each patient, we have information about the PDQ-8 items represented in Table 1 (i.e. 8 feature variables) with values ranging from 0 (never) to 4 (always); and the corresponding EQ-5D (i.e. 5 class variables) with values ranging from 1 (no problems) to 3 (severe problems).

Our objective is to simultaneously predict the 5 class values of EQ-5D from PDQ-8 using `MB-MBC` algorithm. Given the EQ-5D values, to complement them, the corresponding utility index could also be induced using the UK general scoring system.[6] For instance, let us 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,[6] EQ-5D utility index is $1 - 0.081 - 0.036 - 0.123 - 0.236 - 0.269 = 0.255$.

Table 1. The PDQ-8 Items.

Item	Domain of the PDQ-39
pdq1. Had problems getting around in public	Mobility
pdq2. Had difficulty dressing yourself	Activities of daily living
pdq3. Felt depressed	Emotional well-being
pdq4. Had problems with close personal relationships	Social support
pdq5. Had problems with concentration	Cognition
pdq6. Felt unable to communicate with people properly	Communications
pdq7. Had painful muscle cramps or spasms	Bodily discomfort
pdq8. Felt embarrassed in public due to having PD	Stigma

Comparing methods

We compared `MB-MBC` against the following methods:

- Class-bridge decomposable MBC (`CB-MBC`)[24]: learns MBCs based on a greedy forward selection wrapper approach optimizing the accuracy of the model given the training data set.
- Independent Markov blankets algorithm (`IndepMBs`): learns independently a Bayesian network classifier for each class variable using the same HITON algorithm[21,22] (sometimes called binary relevance in the literature). 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 of dimension 5.
- `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.[8,11,12,15] In the `OLS` model, the EQ-5D utility index is directly regressed on the PDQ-8 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.

  For comparison, we adopt the two `OLS` mapping functions proposed by Cheung *et al.*[15] In the first one (`OLS-1`), the EQ-5D utility index is estimated by:

$$\begin{cases} \text{Utility} = 1 \text{ if at least seven responses are ``never''}, \\ \text{Utility} = 1 - 0.135 - 0.052 \times pdq1 - 0.034 \times pdq2 \\ \qquad - 0.031 \times pdq3 - 0.030 \times pdq7, \\ \qquad\qquad \text{otherwise.} \end{cases}$$

  In their second `OLS` function (`OLS-2`), Cheung *et al.*[15] considered the item $pdq3$ as a categorical binary variable with value 0 if the response is equal to "never" and 1 for all other responses. The mapping function is then:

$$\begin{cases} \text{Utility} = 1 \text{ if at least seven responses are ``never''}, \\ \text{Utility} = 1 - 0.105 - 0.052 \times pdq1 - 0.037 \times pdq2 \\ \qquad - 0.031 \times [pdq3 > 0] - 0.026 \times pdq7, \\ \qquad\qquad \text{otherwise.} \end{cases}$$

- `CLAD`[25]: 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,[8,13,15] and it only estimates EQ-5D utility index without predicting the 5 class values of EQ-5D. Here as well, we use the `CLAD` model defined by

Cheung *et al.*[15] as follows:

$$\begin{cases} \text{Utility} = 1 \text{ if at least seven responses are ``never''}, \\ \text{Utility} = 1 - 0.208 - 0.037 \times pdq1 - 0.028 \times pdq2 \\ \qquad - 0.023 \times pdq3 - 0.023 \times pdq5, \\ \qquad\qquad \text{otherwise.} \end{cases}$$

- Deterministic model (`Deterministic`): is a pre-determined algorithm introduced by Cheung *et al.*[15] for predicting EQ-5D from PDQ-8. EQ-5D `mobility`, `self-care`, `pain/discomfort` and `anxiety/depression` items are predicted using PDQ-8 mobility, activities of daily living, bodily discomfort and emotional well-being items, respectively, while the EQ-5D `usual activities` item is predicted using PDQ-8 social support, cognition, communications and stigma items (see Table 1 in Ref. 15).

Figure 3 summarizes the approaches used in this paper for predicting EQ-5D from PDQ-8.

It is important to emphasize here that it is also possible to perform `OLS` and `CLAD` directly on the data set and/or test different mapping functions, however, our main objective is to compare `MB-MBC` against the existing mapping functions defined by Cheung *et al.*[15]

Note finally that, we applied `MB-MBC` and `IndepMBs` with a restriction of the Markov blanket set of each class variable $\text{MB}(C_i)$ to the set of its parents-children $\text{PC}(C_i)$. This restriction was introduced based upon the theoretical discussion introduced by Aliferis *et al.* in Ref. 22 and the empirical observation that including more spouses leads to a less accurate MBC classifier. In fact, Aliferis *et al.*[22] discussed in (see Sec. 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 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 our data set (including only 488 instances).

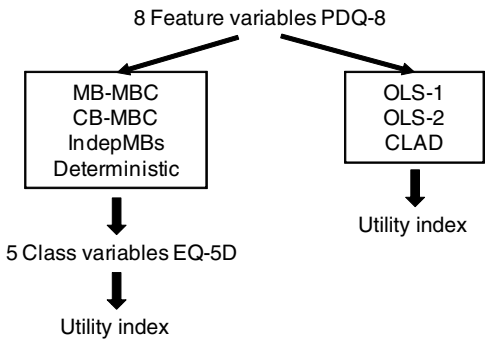

Fig. 3 Used approaches for predicting EQ-5D from PDQ-8.

All methods were run in Matlab R2010b. For `CB-MBC`, we used the Matlab implementation from Ref. 24, and for both `MB-MBC` and `IndepMBs` approaches, the HITON algorithm was run using Causal Explorer Toolkit[26] provided as compiled Matlab functions. The G^2 statistical test was used to evaluate the conditional independencies between variables with a significance level $\alpha = 0.01$.

Evaluation metrics

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

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

$$\text{Acc}_m = \frac{1}{d} \sum_{i=1}^{d} \frac{1}{N} \sum_{l=1}^{N} \delta(\hat{c}_{li}, c_{li}), \tag{3}$$

  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:

$$\text{Acc}_g = \frac{1}{N} \sum_{l=1}^{N} \delta(\hat{\mathbf{c}}_l, \mathbf{c}_l). \tag{4}$$

  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, we considered the following metrics, commonly used in comparison with `OLS` and `CLAD`[8,15]:

- The mean, the first quartile, the median and the third quartile of the predicted EQ-5D utility scores. Obviously, the closer to the corresponding values of the true EQ-5D utility scores, the better.
- 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 mean of all true EQ-5D utility scores and the mean of all predicted EQ-5D utility scores.

For MSE, MAE and AbsDiff, the lower the values, the better. However, for R^2, the higher, the better.

RESULTS

Table 2 shows the estimated classification results over five-fold cross-validation experiments performed on five different data sets randomly generated without replacement from the original PD data set. The mean values and standard deviations for each metric and each method over all five-fold cross-validation experiments are reported, and the best result for each metric is written in bold. Recall that `OLS` and `CLAD` only return the utility index; thus, in order to compute the mean and global accuracies for `OLS` and `CLAD` mapping functions, we proceeded by retrieving the EQ-5D 5 class values as follows: first, we look for the utility index from the UK scoring list[6] closest to the one returned by `OLS-1`, `OLS-2` and `CLAD`, then we determine the EQ-5D vector corresponding to that index.

In Table 2, `MB-MBC` presented the best global accuracy, while `IndepMBs` presented the best mean accuracy. `OLS-1`, `OLS-2` and `CLAD` performed poorly for both mean and global accuracies. In addition, `Deterministic` had the lowest mean and global accuracies compared to all other approaches. 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 equal to 0.05. It turns out that (1) for mean accuracy, `IndepMBs` is significantly better than `OLS-1`, `OLS-2` and `CLAD`, while `MB-MBC` is only significantly better than `OLS-1`; and (2) for global accuracy, both `MB-MBC` and `IndepMBs` are significantly better than `OLS-1` and `Deterministic`. For all remaining algorithms, the differences in predictive performance are not statistically significant.

Table 2. Estimated Accuracies (Mean ± Standard Deviation).

Learning Method	Mean Accuracy	Global Accuracy
`MB-MBC`	0.684 ± 0.003	**0.188 ± 0.005**
`CB-MBC`	0.661 ± 0.013	0.176 ± 0.014
`IndepMBs`	**0.687 ± 0.007**	0.184 ± 0.009
`OLS-1`	0.435 ± 0.022	0.071 ± 0.018
`OLS-2`	0.434 ± 0.011	0.075 ± 0.020
`CLAD`	0.429 ± 0.008	0.088 ± 0.026
`Deterministic`	0.428 ± 0.052	0.033 ± 0.018

In addition, Table 3 shows the values of the mean, the first quartile, the median and the third quartile of the observed (first line) and the predicted utility scores by the different evaluated approaches (rest of the lines). We can see that `MB-MBC` was the best to predict the mean with 0.634 followed by `IndepMBs` with 0.637. `Deterministic` resulted in the best estimated first quartile value equal to 0.516, `IndepMBs` predicted better the median, whereas `MB-MBC` predicted better the third quartile. As previously, we ran a multiple comparison using the Friedman test followed by the Tukey–Kramer post-hoc test with a significance level equal to 0.05. We conclude that (1) for the mean, `MB-MBC` and `IndepMBs` turn out to be only significantly better than `OLS-2`; (2) for the first quartile, `Deterministic`, `IndepMBs` and `MB-MBC` are also significantly better than `OLS-2`; (3) for the median, `IndepMBs` is significantly better than `OLS-2` and `Deterministic`; and (4) for the third quartile, `MB-MBC` is only significantly better than `CLAD`. For all other algorithms, the differences in performance results are not statistically significant.

Finally, Table 4 presents the obtained average results for MSE, MAE, R^2 and AbsDiff metrics over the five-fold cross-validation performed experiments. The best result for each metric is written in bold. `MB-MBC` outperformed other predictive approaches in terms of MSE and MAE. `OLS-1` presented the best R^2 and `IndepMBs` produced the best AbsDiff. Using the Friedman test followed by the Tukey–Kramer post-hoc test with a

Table 3. Mean, 1st Quartile, Median and 3rd Quartile Over Five-Fold Cross-Validation Estimation of the Utility Scores (Mean ± Standard Deviation).

	Mean	1st Quartile	Median	3rd Quartile
Observed	0.565 ± 0.000	0.402 ± 0.037	0.672 ± 0.004	0.796 ± 0.003
`MB-MBC`	**0.634 ± 0.004**	0.524 ± 0.011	0.675 ± 0.008	**0.804 ± 0.011**
`CB-MBC`	0.658 ± 0.015	0.537 ± 0.017	0.684 ± 0.015	0.827 ± 0.019
`IndepMBs`	0.637 ± 0.012	0.522 ± 0.008	**0.674 ± 0.017**	0.826 ± 0.012
`OLS-1`	0.683 ± 0.000	0.567 ± 0.002	0.678 ± 0.003	0.781 ± 0.001
`OLS-2`	0.733 ± 0.000	0.630 ± 0.005	0.730 ± 0.002	0.824 ± 0.002
`CLAD`	0.668 ± 0.000	0.566 ± 0.005	0.653 ± 0.005	0.731 ± 0.001
`Deterministic`	0.640 ± 0.000	**0.516 ± 0.037**	0.725 ± 0.004	0.852 ± 0.008

Table 4. MSE, MAE, R^2 and AbsDiff Over Five-fold Cross-Validation Estimation of the Utility Scores (Mean ± Standard Deviation).

Method	MSE	MAE	R^2	AbsDiff
MB-MBC	**0.074 ± 0.003**	**0.186 ± 0.004**	0.444 ± 0.025	0.068 ± 0.004
CB-MBC	0.094 ± 0.008	0.205 ± 0.009	0.348 ± 0.045	0.092 ± 0.015
IndepMBs	0.075 ± 0.003	0.191 ± 0.004	0.458 ± 0.025	**0.072 ± 0.010**
OLS-1	0.080 ± 0.000	0.195 ± 0.000	**0.486 ± 0.005**	0.117 ± 0.000
OLS-2	0.100 ± 0.000	0.215 ± 0.000	0.454 ± 0.005	0.168 ± 0.000
CLAD	0.084 ± 0.000	0.201 ± 0.000	0.425 ± 0.006	0.103 ± 0.000
Deterministic	0.081 ± 0.000	0.202 ± 0.000	0.451 ± 0.009	0.075 ± 0.000

significance level equal to 0.05, it turns out that (1) for MSE, MB-MBC and IndepMBs are significantly better than CB-MBC and OLS-2; (2) for MAE, MB-MBC is significantly better than CB-MBC, OLS-2 and Deterministic, whereas IndepMBs is only significantly better than OLS-2; (3) for R^2, OLS-1, OLS-2 and IndepMBs are significantly better than CB-MBC; and (4) for AbsDiff, MB-MBC and IndepMBs are only significantly better than OLS-2. For all remaining methods, the differences are not statistically significant.

DISCUSSION

As shown in Tables 2 to 4, OLS-1, OLS-2 and CLAD methods performed worse than MB-MBC for almost all the considered evaluation metrics. In particular, they presented worse results for MSE, MAE and AbsDiff, even having being designed for optimizing those criteria. As pointed out by Le and Doctor,[8] this may be due to certain limitations of these regression methods such as predictive values of EQ-5D utility scores that are not defined in the UK scoring list, which correspond as well to undefined health states. Previous studies testing OLS and CLAD for predicting the EQ-5D utility index from the Health Surveys SF-12[8,13] 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-1 mapping function resulted in a better MSE, MAE, R^2 than CLAD, but for the absolute difference between the true and the predicted EQ-5D mean scores, CLAD performed better. OLS-2 performed better than CLAD only for R^2.

Moreover, contrary to OLS-1, OLS-2, CLAD and Deterministic, 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-8 variables, we first examine in Fig. 4 the graphical structure of the MBC network learnt by the MB-MBC algorithm, then compare it to the graphical structures learnt by CB-MBC and IndepMBs. The interpretation and the medical significance of the obtained MBCs have been ensured by the neurologist specialist in PD Pablo Martínez-Martín.

Firstly, the *class subgraph* in Fig. 4 (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 variables usual activities and self-care which proves as well the strong relevance between these class variables. 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 probabilistic dependence relationships between EQ-5D classes and PDQ-8 features, the latter listed in Table 1. We have the following dependence relationships from EQ-5D to PDQ-8:

- Mobility is directly associated with pdq1.
- Self-care is directly associated with pdq2.
- Usual activities is directly associated with pdq2.
- Pain/discomfort is directly associated with pdq2 and pdq7.
- Anxiety/depression is directly associated with pdq3.

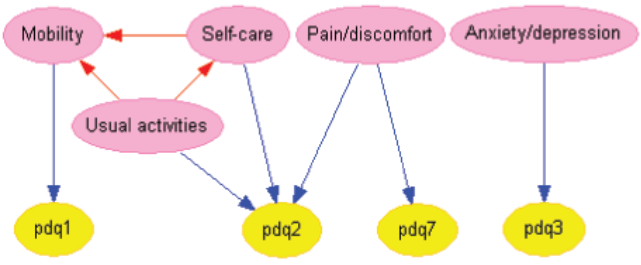

Fig. 4 MBC graphical structure learnt by MB-MBC.

Note that the detected associations are very appropriate and a direct connection can be clearly observed between the domains of both instruments PDQ-8 and EQ-5D. In fact, the item pdq1 (having problems getting around public), pertaining to mobility domain, is associated with the class variable `mobility`. The item pdq2 (having difficulty dressing yourself), pertaining to activities of daily living domain, is simultaneously associated with `self-care`, `usual activities` and `pain/discomfort`. The item pdq3 (felt depressed), pertaining to emotional well-being domain, is directly associated with `anxiety/depression`, and finally the item pdq7 (having painful muscle cramps and spasms), pertaining to bodily discomfort domain, is related to `pain/discomfort`. Moreover, we may notice that the selected pdq items here (i.e. pdq1, pdq2, pdq3 and pdq7) are exactly the same kept by the `OLS-1` and `OLS-2` mapping functions proposed by Cheung *et al.*[15] The remaining features (pdq4, pdq5, pdq6 and pdq8) are absent in Fig. 4 since no associations were detected between them and the EQ-5D class variables.

These findings have sense from a clinical point of view. In fact, PD is a complex disorder with many aspects to be considered: motor impairment, disability, motor complications (dyskinesia and fluctuations) and a wide array of nonmotor symptoms (depression, hallucinations, cognitive decline, dribbling saliva, for example). For the huge majority of PD patients, from earliest to most advanced stages, the common perceived health problems are reflected as limitations for mobility (pdq1) and activities of daily life (pdq2), whereas the most prevalent nonmotor symptoms are associated with the impact on patients' health perception, namely depression (pdq3) and pain (pdq7). As a matter of fact, disability, depression and pain are between commonest determinants of HRQoL in general population[27] and also in PD.[28,29] Additional motor complications and nonmotor symptoms (for instance those related to pdq4, pdq5, pdq6 and pdq8) are also relevant in PD, however, they usually present a problem only in moderate/advanced phases of the disease but not in early/mild disease, and are not constant (neither in presence nor severity) in all patients.

Taking the previous arguments into account, we may conclude that, from a clinical point of view, the selected variables in the network are relevant and cover the most important areas of PD including both motor and nonmotor symptoms.

Finally, the feature subgraph is obviously empty with no direct dependence relationships between the different PDQ-8 items, because the third step in the `MB-MBC` algorithm has no effect at all. This is reflected as well through the construction process of the PDQ-8 questionnaire. In fact, as shown in Table 1, each item in the PDQ-8 represents a single domain of the PDQ-39, and therefore no inter-domain relationships were taken into account for the construction of the PDQ-8.

For structural comparison, we depict in Figs. 5 and 6 the graphical structures learnt by `CB-MBC` and `IndepMBs`, respectively.

As shown in Fig. 5, `CB-MBC` only detects a direct dependence relationship among the EQ-5D class variables `self-care` and `usual activities`. All remaining classes are kept independent. For the bridge subgraph, and similar to `MB-MBC`, the following dependence relationships are detected: `mobility` is associated with pdq1, `self-care` with pdq2, `usual activities` with pdq2,

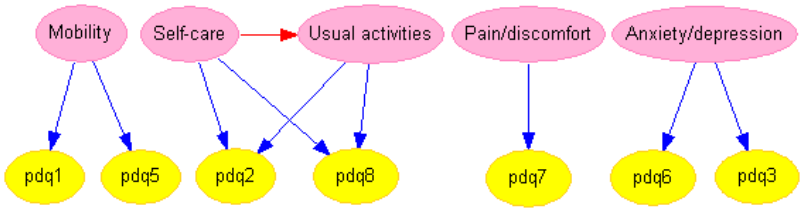

Fig. 5 MBC graphical structure learnt by `CB-MBC`.

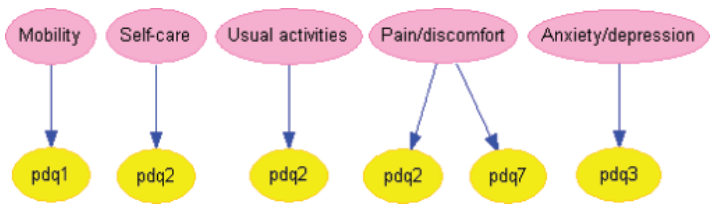

Fig. 6 Graphical structures learnt by `IndepMBs`.

pain/discomfort with pdq7, and anxiety/depression with pdq3. Moreover, additional arcs are discovered in the bridge subgraph between the EQ-5D class variables and pdq5, pdq6 and pdq8. As before, the feature subgraph is empty.

Figure 6 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,[21,22] 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. As shown in Table 3, this also leads to a lower predictive performance comparing to MB-MBC.

CONCLUSIONS

This paper proposed MB-MBC learning algorithm to predict EuroQol EQ-5D health states from the disease-specific HRQoL measure PDQ-8. Experimental results on a PD data set including 488 patients were promising in comparison with two Bayesian network-based approaches, CB-MBC and IndepMBs, as well as three state-of-the-art mapping approaches OLS-1, OLS-2 and CLAD and a deterministic model. The learned MBC graphical structure allowed also the identification of the dependence relationships among EQ-5D and PDQ-8 items, and proved, as expected, the strong relevance between EQ-5D class variables and selected PDQ-8 feature variables. Note finally that, a limitation of our study is that the considered PD data set contained only few patients with severe problems. In the future, it will be interesting to consider a larger and diverse PD data set to perform more analyses and additionally prove the merits of our approach. Another possibility as well is to use a bootstrap resampling method to deal with the small number of instances and obtain more robust Bayesian network classifiers.[30] Moreover, another direction for future research is to apply our approach to different medical or biological multi-dimensional classification problems, where an instance has to be assigned to more than one class variable.

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. Lees AJ, Hardy J, Revesz T, Parkinson's disease, *Lancet* **373**:2055, 2009.
2. Karlsen K, Tandberg E, Aarsland D *et al.*, Health-related quality of life in Parkinson's disease: A prospective longitudinal study, *J Neurol Neurosurg Psychiatry* **69**:584, 2000.
3. EuroQol Group, EuroQol — a new facility for the measurement of health-related quality of life, *Health Policy* **16**:199, 1990.
4. Cheung K, Oemar M, Oppe M *et al.*, EQ-5D User Guide. Basic information on how to use the EQ-5D, version 2.0. EuroQol Group, Available at www.euroqol.org, 2009.
5. Schrag A, Selai C, Jahanshahi M *et al.*, 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* **69**:67, 2000.
6. EuroQuol Group, A1 tariff based on UK survey, EuroQuol EQ-5D user guide, 1990.
7. Brooks R, Rabin R, Charro F, *The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective*, Kluwer Academic, Dordrecht, 2003.
8. 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* **49**:451, 2011.
9. Martínez-Martín P, Jeukens-Visser M, Lyons KE *et al.*, Health-related quality of life scales in Parkinson's disease: Critique and recommendations, *Mov Disord* **26**:2371, 2011.
10. Jenkinson C, Fitzpatrick R, Petoa V *et al.*, The PDQ-8: Development and validation of a short-form Parkinson's disease questionnaire, *Psychol Health* **12**:805, 1997.
11. Franks P, Lubetkin DI, Gold MR *et al.*, Mapping the SF-12 to the EuroQoL EQ-5D index in a national US sample, *Med Decis Making* **24**:247, 2004.
12. Lawrence WF, Fleishman JA, Predicting EuroQoL EQ-5D preference scores from the SF-12 health survey in a nationally representative sample, *Med Decis Making* **24**:160, 2004.
13. 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* **26**:401, 2006.
14. Rowen D, Brazier J, Roberts J, Mapping SF-36 onto the EQ-5D index: How reliable is the relationship? *Health Qual Life Outcomes* **31**:7, 2009.
15. Cheung YB, Tan LC, Lau PN *et al.*, Mapping the eight-item Parkinson's disease questionnaire (PDQ-8) to the EQ-5D utility index, *Qual Life Res* **17**:1173, 2008.
16. Pearl J, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, San Francisco, 1988.

17. Koller D, Friedman N, *Probabilistic Graphical Models. Principles and Techniques*, MIT Press, Cambridge, 2009.
18. Pearl J, Verma TS, Equivalence and synthesis of causal models, *Sixth Conf on Uncertainty in Artificial Intelligence*, pp. 220–227, 1990.
19. van der Gaag LC, de Waal PR, Muti-dimensional Bayesian network classifiers, *Third European Workshop on Probabilistic Graphical Models*, pp. 107–114, 2006.
20. Bielza C, Li G, Larrañaga P, Multi-dimensional classification with Bayesian networks, *Int J Approx Reasoning* **52**:705, 2011.
21. Aliferis CF, Statnikov A, Tsamardinos I *et al.*, Local causal and Markov blanket induction for causal discovery and feature selection for classification. Part I: Algorithms and empirical evaluation, *J Mach Learn Res* **11**:171, 2010.
22. Aliferis CF, Statnikov A, Tsamardinos I *et al.*, Local causal and Markov blanket induction for causal discovery and feature selection for classification. Part II: Analysis and extensions, *J Mach Learn Res* **11**:235, 2010.
23. Spirtes P, Glymour C, Scheines R, *Causation, Prediction, and Search*, MIT Press, Cambridge, 2000.
24. Borchani H, Bielza C, Larrañaga P, Learning CB-decomposable multi-dimensional Bayesian network classifiers, *Fifth European Workshop on Probabilistic Graphical Models*, pp. 25–32, 2010.
25. Powell JL, Least absolute deviations estimation for the censored regression model, *J Econometr* **25**:303, 1984.
26. Aliferis CF, Tsamardinos I, Statnikov A, Causal Explorer: A probabilistic network learning toolkit for biomedical discovery, *Int Conf Mathematics and Engineering Techniques in Medicine and Biological Sciences*, pp. 371–376, 2003.
27. Martínez-Martín P, Prieto-Flores ME, Forjaz MJ *et al.*, Components and determinants of quality of life in community-dwelling older adults, *Eur J Ageing* **9**:255, 2012.
28. Schrag A, Quality of life and depression in Parkinson's disease, *J Neurol Sci* **248**:151, 2006.
29. Quittenbaum BH, Grahn B, Quality of life and pain in Parkinson's disease: A controlled cross-sectional study, *Parkinsonism Relat Disord* **10**:129, 2004.
30. Armañanzas R, Larrañaga P, Bielza C, Ensemble transcript interaction networks: A case study on Alzheimer's disease, *Comput Methods Programs Biomed* **108**:442, 2012.