BAYESIAN NETWORKS IN ACTION

Pedro Larrañaga

Computational Intelligence Group
Artificial Intelligence Department
Technical University of Madrid

XVIII Conferencia de la Asociación Española para la Inteligencia Artificial
Granada, October 23, 2018
The five tribes of machine learning and their master algorithms (Domingos, 2015)
Machine learning methods for approaching the eight problems in this talk
Machine learning methods for approaching the eight problems in this talk
The 23 Asilomar AI principles

<table>
<thead>
<tr>
<th>Asilomar Conference on Beneficial AI (Future of Life Institute 2017)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Research goal</td>
</tr>
<tr>
<td>2. Research funding</td>
</tr>
<tr>
<td>3. Science-policy lik</td>
</tr>
<tr>
<td>4. Research culture</td>
</tr>
<tr>
<td>5. Race avoidance</td>
</tr>
<tr>
<td>6. Safety</td>
</tr>
<tr>
<td>7. Failure transparency</td>
</tr>
<tr>
<td>8. Judicial transparency</td>
</tr>
<tr>
<td>9. Responsability</td>
</tr>
<tr>
<td>10. Value alignment</td>
</tr>
<tr>
<td>11. Human values</td>
</tr>
<tr>
<td>12. Personal privacy</td>
</tr>
<tr>
<td>13. Liberty and privacy</td>
</tr>
<tr>
<td>14. Shared Benefit</td>
</tr>
<tr>
<td>15. Shared prosperity</td>
</tr>
<tr>
<td>16. Human control</td>
</tr>
<tr>
<td>17. Non-subversion</td>
</tr>
<tr>
<td>18. AI arms race</td>
</tr>
<tr>
<td>19. Capacity caution</td>
</tr>
<tr>
<td>20. Importance</td>
</tr>
<tr>
<td>21. Risks</td>
</tr>
<tr>
<td>22. Recursive self-improvement</td>
</tr>
<tr>
<td>23. Common good</td>
</tr>
</tbody>
</table>

Run away from black-box models
Outline

1. Introduction
2. Bayesian Networks
3. Neuroscience
4. Industry
5. Sport
6. Conclusions
7. References
“Risk of dementia” (Bielza and Larrañaga, 2014a)
The Bayes net for dementia risk is characterized by the following formulas and conditional probabilities:

\[
p(A, N, S, D, P) = p(A)p(N|A)p(S|A)p(D|N, S)p(P|S)
\]

### Conditional Probabilities

- **Age (A)**
  - \(a\): 0.75
  - \(\neg a\): 0.25

- **Neuronal Atrophy (N)**
  - \(n\|a\): 0.15
  - \(a\|n\): 0.85
  - \(\neg a\|n\): 0.03
  - \(\neg a\|\neg n\): 0.97

- **Stroke (S)**
  - \(a\|s\): 0.10
  - \(a\|\neg s\): 0.90
  - \(\neg a\|s\): 0.02
  - \(\neg a\|\neg s\): 0.98

- **Dementia (D)**
  - \(n\|a\|s\): 0.96
  - \(n\|a\|\neg s\): 0.04
  - \(n\|\neg a\|s\): 0.40
  - \(n\|\neg a\|\neg s\): 0.60
  - \(\neg n\|a\|s\): 0.45
  - \(\neg n\|a\|\neg s\): 0.55
  - \(\neg n\|\neg a\|s\): 0.10
  - \(\neg n\|\neg a\|\neg s\): 0.90

- **Paralysis (P)**
  - \(s\|a\|p\): 0.75
  - \(s\|a\|\neg p\): 0.25
  - \(\neg s\|a\|p\): 0.05
  - \(\neg s\|a\|\neg p\): 0.95

These figures illustrate the interdependencies in the risk factors for dementia.
"Risk of dementia"

\[
p(A, N, S, D, P) = p(A)p(N|A)p(S|A)p(D|N, S)p(P|S)
\]
Inference (reasoning) with Bayesian networks

No evidence
Inference (reasoning) with Bayesian networks

Evidence: “Stroke = yes”
Evidence: “Stroke = yes, Neuronal Atrophy = yes”
Inference (reasoning) with Bayesian networks

Evidence: “Stroke = yes, Neuronal Atrophy=yes, Age= young”
Bayesian networks (Pearl, 1988; Koller and Friedman, 2009)

A Bayesian network consists of two components

1. **Graphical structure** $G$ is a directed acyclic graph (DAG)
   - *Vertices* $\rightarrow$ variables
   - *Directed edges* $\rightarrow$ conditional (in)dependences

2. **Set of parameters** specifies the set of conditional probability distributions

   $\text{Joint probability distribution: } P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \text{pa}(x_i))$
Bayesian networks

Learning and inference

Bayesian network learning
- Structure learning
- Parameter learning

Probabilistic inference
- Compute the conditional probability $P(\text{Query} \mid \text{Evidence})$
Learning Bayesian networks from data

**STRUCTURE LEARNING**

1. Detecting conditional independencies **between triples of variables** by hypothesis tests
2. Score and search methods

**PARAMETER LEARNING**

1. Maximum likelihood estimation
2. Bayesian estimation
Outline

1. Introduction
2. Bayesian Networks

3. Neuroscience

4. Industry
5. Sport

6. Conclusions
7. References
Human Brain Project (2013-2023)

Future Medicine
- From symptom-based to biology-based classifications
- Unique biological signatures of disease
- Early diagnosis & preventive medicine
- Personalized clinical trials
- Efficient drug & clinical treatments
- Personalized medicine

Future Neuroscience
- Multi-level view of the brain
- Causal chain of events from genes to cognition
- Understanding the human brain
- Body ownership, language, emotions, consciousness
- Theory of mind

Future Computing
- Supercomputing as a scientific instrument
- Supercomputing as a commodity
- New software for interactive and considerate supercomputing
- New hardware for neuromorphic computing
- Intelligent tools for managing and making sense of big data
- Human-like intelligence

ICT Platforms

Pedro Larrañaga
The human brain

Brain lobes and layers

- **Weight** = 1.3kg, **width** = 140mm, **length** = 167mm, **height** = 93mm
The human brain

Brain at microscopic level

- Composed of neurons, blood vessels, glial cells
- Neuron is the basic structural and functional unit of the nervous system —neuron doctrine— (S. Ramón y Cajal, late 19th century)
- Just 4 microns thick → could fit 30,000 neurons on the head of a pin
- ~86,000 million neurons (more than known stars in the universe)
The neuron

3 parts of a neuron: dendrites, soma and axon

- Axons fill most of the space in the brain → >150,000 km in the human brain!!
- Each neuron connected to 1,000 neighboring neurons
- 10,000 synaptic connections each
Clustering of dendritic spines

**Mixture of Gaussian Bayesian networks learnt with the structural EM**
(Luengo-Sanchez et al., 2018)


(a) 3D reconstruction process of human dendritic spines

(b) Spine repair process and multiresolution Reeb graph computation
Clustering of dendritic spines

**Mixture of Gaussian Bayesian Networks learnt with the structural EM**
(Luengo-Sanchez et al., 2018)

The data set:
- 3D reconstruction of more than 7,900 spines from layer III pyramidal neuron human cingulate cortex (aged 40 and 85)
- Each spine is characterized with 54 morphological variables (some of them directional variables)

The model:
- **Multivariate Gaussian mixture model**: \( f(x; \theta) = \sum_{k=1}^{K} \pi_k f_k(x; \mu_k) \)
- Each mixture density is given by:
  \[ f_k(x; \mu_k, \Sigma_k) = (2\pi)^{-\frac{n}{2}} |\Sigma_k|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) \right\} \]
- **Mixture of Gaussian Bayesian networks**: Each mixture is expressed as a Gaussian Bayesian network
Clustering of dendritic spines

Mixture of Gaussian Bayesian networks learnt with the structural EM (Luengo-Sanchez et al., 2018)

Virtual spines obtained from the simulation of the mixture model with six components
**Maximizing the BIC criterion** (Rodriguez-Sanchez et al., work in progress)

- **Allen Cell Types Database**: 436 mice neurons with reconstructions of soma, dendrites and axon; 19 variables discretized into 4 bins
- **EM algorithm** (Dempster et al. 1977) adapted to structure learning of Bayesian networks (Friedman, 1998)

![Graphs of two clusters](image)

(a) First cluster

(b) Second cluster
Electrophysiological probabilistic clustering of neurons

**Maximizing the BIC criterion** (Rodriguez-Sanchez et al., work in progress)

- **Allen Cell Types Data Base**: the same 436 neurons, 42 electrophysiological variables discretized into 4 bins

(a) First cluster

(b) Second cluster
Morphological and electrophysiological (as independent)

Maximizing the BIC criterion (Rodriguez-Sanchez et al., work in progress)

BIC value = -23,017.83

<table>
<thead>
<tr>
<th>Cluster Morpho</th>
<th>cluster 1</th>
<th>cluster 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>cluster 1</td>
<td>104</td>
<td>129</td>
</tr>
<tr>
<td>cluster 2</td>
<td>21</td>
<td>169</td>
</tr>
<tr>
<td></td>
<td>125</td>
<td>298</td>
</tr>
</tbody>
</table>
**Maximizing the BIC criterion** (Rodriguez-Sanchez et al., work in progress)

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Morpho</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>141</td>
<td>97</td>
</tr>
<tr>
<td>cluster 1</td>
<td></td>
<td>238</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
<td>176</td>
</tr>
<tr>
<td>cluster 2</td>
<td></td>
<td>150</td>
<td>273</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>423</td>
</tr>
</tbody>
</table>

BIC value = -22,940.32
Morphological classification of interneurons

Web-Based System. 42 Experts. 320 Interneurons (DeFelipe et al., 2013)
Morphological classification of interneurons

**Web-Based System. 42 Experts. 320 Interneurons** (DeFelipe et al., 2013)

The five class variables to be predicted by the classification model
**Morphological classification of interneurons**

**Web-Based System. 42 Experts. 320 Interneurons** (DeFelipe et al., 2013)

<table>
<thead>
<tr>
<th>Neuron</th>
<th>$X_1$</th>
<th>$X_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.3</td>
<td>14.1</td>
</tr>
<tr>
<td>2</td>
<td>1.6</td>
<td>13.7</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>320</td>
<td>1.1</td>
<td>22.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$E_1$</th>
<th>...</th>
<th>$E_{42}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>0</td>
<td>...</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$C_1$</th>
<th>$C_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_1$</td>
<td>$E_{42}$</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

The dataset with the five class variables
Morphological classification of interneurons

**Web-Based System. 42 Experts. 320 Interneurons** (DeFelipe et al., 2013)

Inter-expert agreement in each of the class variables
Morphological classification of interneurons

**Web-Based System. 42 Experts. 320 Interneurons** (DeFelipe et al., 2013)

Bayesian network models of the choice behaviour of expert 16 and expert 27 when selecting **Martinotti**
Morphological classification of interneurons

**Web-Based System. 42 Experts. 320 Interneurons** (DeFelipe et al., 2013)

Bayesian network models of the choice behaviour of expert 16 and expert 27 when selecting common basket.
Morphological classification of interneurons

Three different approaches

A. Neuron label as its most frequent cell type

B. Different label reliability thresholds

C. GEM algorithm for probabilistic class labels
Morphological classification of interneurons

A. NEURON LABEL AS ITS MOST FREquent CELL TYPE (DeFelipe et al., 2013)

<table>
<thead>
<tr>
<th>Neuron</th>
<th>(X_1)</th>
<th>(X_{2886})</th>
<th>(C_5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.3</td>
<td>14.1</td>
<td>MT</td>
</tr>
<tr>
<td>2</td>
<td>1.6</td>
<td>13.7</td>
<td>LB</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>241</td>
<td>1.1</td>
<td>22.2</td>
<td>CB</td>
</tr>
</tbody>
</table>

Supervised classification problem. The labels of \(C_5\) are the modes of the assignments provided by the 42 experts.

- **Classifiers:** Gaussian naive Bayes, discrete naive Bayes, radial basis function, support vector machines, 1-nearest neighbor, \(k\)-nearest neighbors, rule induction, classification trees, random forests, and random trees.
- **Feature selection methods:** univariate filter (gain ratio), and multivariate filter (correlation feature selection).
- **Best accuracies:** 85.48% for \(C_1\), 81.33% for \(C_2\), 73.86% for \(C_3\), 60.17% for \(C_4\), and 62.24% for \(C_5\).
**Morphological classification of interneurons**

### B. Different Label Reliability Thresholds (Mihaljević et al., 2015)

<table>
<thead>
<tr>
<th>$D_{25}^{5, X}$</th>
<th>$X_1, \ldots, X_{214}$</th>
<th>$C_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.2, \ldots, 4.1</td>
<td>MA</td>
</tr>
<tr>
<td>3</td>
<td>1.2, \ldots, 4.2</td>
<td>HT</td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
</tr>
<tr>
<td>237</td>
<td>1.0, \ldots, 2.2</td>
<td>CB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$D_{25,21}^{5, 1234}$</th>
<th>$C_1, \ldots, C_4$</th>
<th>$C_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>T1, \ldots, As.</td>
<td>MA</td>
</tr>
<tr>
<td>3</td>
<td>T1, \ldots, Ds.</td>
<td>HT</td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$D_{25, \ldots, 21}^{5, X, 1234}$</th>
<th>$X_1, \ldots, X_{214}, C_1, \ldots, C_4$</th>
<th>$C_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.2, \ldots, 4.1, T1, \ldots, As.</td>
<td>MA</td>
</tr>
<tr>
<td>3</td>
<td>1.2, \ldots, 4.2, T1, \ldots, Ds.</td>
<td>HT</td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$D$</th>
<th>$X_1, \ldots, X_{214}$</th>
<th>$C_1$</th>
<th>$C_2$</th>
<th>$C_3$</th>
<th>$C_4$</th>
<th>$C_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5, \ldots, 2.1</td>
<td>I1. (30)</td>
<td>Ic. (30)</td>
<td>Ce. (30)</td>
<td>Bo. (10)</td>
<td>CT (21)</td>
</tr>
<tr>
<td>2</td>
<td>0.2, \ldots, 4.1</td>
<td>T1. (40)</td>
<td>Tc. (29)</td>
<td>De. (40)</td>
<td>As. (39)</td>
<td>MA (40)</td>
</tr>
<tr>
<td>3</td>
<td>1.2, \ldots, 4.2</td>
<td>T1. (40)</td>
<td>Ic. (39)</td>
<td>De. (40)</td>
<td>Ds. (40)</td>
<td>HT (40)</td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
</tr>
<tr>
<td>237</td>
<td>1.0, \ldots, 2.2</td>
<td>I1. (21)</td>
<td>Ic. (24)</td>
<td>Ce. (35)</td>
<td>As. (7)</td>
<td>CB (26)</td>
</tr>
</tbody>
</table>

A schematic overview of our automatic categorization of interneurons according to 214 morphological variables and 4 high-level axonal features $C_1 - C_4$
Morphological classification of interneurons

B. Different label reliability thresholds (Mihaljević et al., 2015)

Number of neurons of different types of $C_5$ versus label reliability threshold ($C_1$–$C_4$ with a threshold of 21)
B. Different label reliability thresholds (Mihaljević et al., 2015)

Bayesian network classifiers (Bielza and Larrañaga, 2014b)

(a) Naive Bayes (NB) structure

\[ p(x, c) = p(c)p(x_1 | c)p(x_2 | c)p(x_3 | c)p(x_4 | c)p(x_5 | c) \]

(b) Selective naive Bayes (NB-FSS) structure

\[ p(x, c) = p(c)p(x_1 | c)p(x_2 | c)p(x_3 | c)p(x_5 | c) \]

(a) Attribute weighted naive Bayes (AWNB) structure

\[ p(x, c) = p(c)p(x_1 | c)^{0.9}p(x_2 | c)^1p(x_3 | c)^{0.4}p(x_4 | c)^0\]
\[ p(x_5 | c)^{0.8} \]

(b) Tree augmented naive Bayes (TAN) structure

\[ p(x_4 | x_3, c)p(x_5 | x_4, c) \]
Morphological classification of interneurons

B. Different label reliability thresholds (Mihaljević et al., 2015)

From $D_5^{5,\mathcal{X}}$, until $D_5^{5,\mathcal{X}}$

From $D_5^{5,1234,10,21}$, until $D_5^{5,1234,28,21}$

From $D_5^{5,\mathcal{X},1234,10,-,21}$, until $D_5^{5,\mathcal{X},1234,28,-,21}$
B. Different label reliability thresholds (Mihaljević et al., 2015)
Morphological classification of interneurons

B. Different label reliability thresholds (Mihaljević et al., 2015)
B. Different label reliability thresholds (Mihaljević et al., 2015)
B. Different label reliability thresholds (Mihaljević et al., 2015)
Morphological classification of interneurons

C. GEM ALGORITHM FOR PROBABILISTIC CLASS LABELS (López-Cruz, 2013)

<table>
<thead>
<tr>
<th>Morpho var.</th>
<th>$C_5$ class variable values</th>
</tr>
</thead>
<tbody>
<tr>
<td>cell</td>
<td>$X_1$</td>
</tr>
<tr>
<td>1</td>
<td>10.1</td>
</tr>
<tr>
<td>2</td>
<td>3.7</td>
</tr>
<tr>
<td>3</td>
<td>5.9</td>
</tr>
<tr>
<td>4</td>
<td>11.2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>236</td>
<td>13.6</td>
</tr>
<tr>
<td>237</td>
<td>8.9</td>
</tr>
</tbody>
</table>

Probabilistic class labels for each cell according to the votes received for each interneuron type from the 42 experts. The ten values of $C_5$ are denoted by: (1) arcade, (2) Cajal-Retzius, (3) Chandelier, (4) common basket, (5) horse-tail, (6) large basket, (7) Martinotti, (8) neurogliaform, (9) common type, (10) other

The generalized expectation maximization (GEM) algorithm is particularized to learn Bayesian classifiers with different structural complexities

$$\ell(D_{X,\Pi}|\Theta) = \sum_{i=1}^{N} \ln \left( \sum_{c\in\Omega_C} \pi_{ic} p_C(c; \theta_C) f_{X|C}(x_i|c; \theta_{X|C}) \right)$$
EQ-5D health states from PDQ-39 in Parkinson disease

Parkinson disease motor symptoms
EQ-5D health states from PDQ-39 in Parkinson disease

PDQ-39 and EQ-5D: quality of life instruments to measure the degree of disability in PD

39-item Parkinson’s Disease Questionnaire: a specific instrument

PDQ-39 captures patient’s perception of his illness covering 8 dimensions:

1. Mobility
2. Activities of daily living
3. Emotional well-being
4. Stigma
5. Social support
6. Cognitions
7. Communication
8. Bodily discomfort

PDQ-39 QUESTIONNAIRE

Please complete the following

Due to having Parkinson’s disease, how often during the last month have you...

<table>
<thead>
<tr>
<th>Question</th>
<th>Never</th>
<th>Occasionally</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always or cannot do at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Had difficulty doing the leisure activities which you would like to do?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Had difficulty looking after your home, e.g. DIY, housework, cooking?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Had difficulty carrying bags of shopping?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Had problems walking half a mile?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Had problems walking 100 yards?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Had problems getting around the house as easily as you would like?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**European Quality of Life - 5 Dimensions: a generic instrument**

**EQ-5D** is a generic **measure of health for clinical and economic appraisal**

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mobility</strong></td>
<td>I have no problems in walking about</td>
</tr>
<tr>
<td></td>
<td>I have some problems in walking about</td>
</tr>
<tr>
<td></td>
<td>I am confined to bed</td>
</tr>
<tr>
<td><strong>Self-care</strong></td>
<td>I have no problems with self-care</td>
</tr>
<tr>
<td></td>
<td>I have some problems washing and dressing myself</td>
</tr>
<tr>
<td></td>
<td>I am unable to wash and dress myself</td>
</tr>
<tr>
<td><strong>Usual activities</strong></td>
<td>(eg. work, study, housework, family or leisure activities)</td>
</tr>
<tr>
<td></td>
<td>I have no problems with performing my usual activities</td>
</tr>
<tr>
<td></td>
<td>I have some problems with performing my usual activities</td>
</tr>
<tr>
<td></td>
<td>I am unable to perform my usual activities</td>
</tr>
<tr>
<td><strong>Pain/discomfort</strong></td>
<td>I have no pain or discomfort</td>
</tr>
<tr>
<td></td>
<td>I have moderate pain or discomfort</td>
</tr>
<tr>
<td></td>
<td>I have extreme pain or discomfort</td>
</tr>
<tr>
<td><strong>Anxiety/depression</strong></td>
<td>I am not anxious or depressed</td>
</tr>
<tr>
<td></td>
<td>I am moderately anxious or depressed</td>
</tr>
<tr>
<td></td>
<td>I am extremely anxious or depressed</td>
</tr>
</tbody>
</table>
EQ-5D health states from PDQ-39 in Parkinson disease

Mapping PDQ-39 to EQ-5D

<table>
<thead>
<tr>
<th>PDQ_1</th>
<th>PDQ_2</th>
<th>...</th>
<th>...</th>
<th>PDQ_{39}</th>
<th>EQ_1</th>
<th>EQ_2</th>
<th>EQ_3</th>
<th>EQ_4</th>
<th>EQ_5</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>...</td>
<td>...</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>...</td>
<td>...</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>...</td>
<td>...</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>...</td>
<td>...</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>...</td>
<td>...</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>...</td>
<td>...</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

\[ \phi : (PDQ_1, \ldots, PDQ_{39}) \rightarrow (EQ_1, \ldots, EQ_5) \]
**EQ-5D health states from PDQ-39 in Parkinson disease**

**MULTIPLE DIAGNOSIS PROBLEM. DIRECT APPROACH** (Peng and Reggia, 1987a, 1987b)

<table>
<thead>
<tr>
<th></th>
<th>$X_1$</th>
<th>...</th>
<th>$X_m$</th>
<th></th>
<th>$C_1$</th>
<th>...</th>
<th>$C_d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(X^{(1)}, C^{(1)})$</td>
<td>$x_1^{(1)}$</td>
<td>...</td>
<td>$x_m^{(1)}$</td>
<td></td>
<td>$c_1^{(1)}$</td>
<td>...</td>
<td>$c_d^{(1)}$</td>
</tr>
<tr>
<td>$(X^{(2)}, C^{(2)})$</td>
<td>$x_1^{(2)}$</td>
<td>...</td>
<td>$x_m^{(2)}$</td>
<td></td>
<td>$c_1^{(2)}$</td>
<td>...</td>
<td>$c_d^{(2)}$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$(X^{(N)}, C^{(N)})$</td>
<td>$x_1^{(N)}$</td>
<td>...</td>
<td>$x_m^{(N)}$</td>
<td></td>
<td>$c_1^{(N)}$</td>
<td>...</td>
<td>$c_d^{(N)}$</td>
</tr>
</tbody>
</table>

Optimal diagnosis as abductive inference: searching for the most probable explanation (MPE)

$$(c_1^*, \ldots, c_d^*)$$

$$= \arg \max_{(c_1, \ldots, c_d)} p(C_1 = c_1, \ldots, C_d = c_d | X_1 = x_1, \ldots, X_m = x_m)$$

$$= \arg \max_{(c_1, \ldots, c_d)} p(C_1 = c_1, \ldots, C_d = c_d) p(X_1 = x_1, \ldots, X_m = x_m | C_1 = c_1, \ldots, C_d = c_d)$$

Number of parameters to be estimated for the case of binary predictors and classes:

$$2^d - 1 + 2^d (2^m - 1)$$
MULTIPLE DIAGNOSIS WITH MULTI-DIMENSIONAL BAYESIAN NETWORK CLASSIFIERS (MBCs) (van der Gaag and de Waal, 2006)

- In a multi-dimensional Bayesian network classifier (MBC) the set of vertices $\mathcal{V}$ is partitioned into:
  - $\mathcal{V}_C = \{C_1, \ldots, C_d\}$ of class variables and
  - $\mathcal{V}_X = \{X_1, \ldots, X_m\}$ of feature variables
  with $(d + m = n)$

- Three subgraphs in the structure of a multi-dimensional Bayesian network classifier: **Class subgraph**: $G_C$, **Bridge subgraph**: $G_{CX}$ and **Feature subgraph**: $G_X$
**TRACTABILITY OF MPE IN MBCs WITH CLASS BRIDGE DECOMPOSABLE MBCs** (Bielza et al., 2011)

MPE is generally **NP-hard** in Bayesian networks (Kwisthout, 2011)

An MBC is **class-bridge decomposable MBC** if:

1. \( G_C \cup G_{C_X} \) can be decomposed as: \( G_C \cup G_{C_X} = \bigcup_{i=1}^{r} (G_{C_i} \cup G_{(C_X)_i}) \) where \( G_{C_i} \cup G_{(C_X)_i} \) with \( i = 1, \ldots, r \) are its maximal connected components

2. Non-shared children: \( Ch(V_{C_i}) \cap Ch(V_{C_j}) = \emptyset \), with \( i, j = 1, \ldots, r \) and \( i \neq j \), where \( Ch(V_{C_i}) \) denotes the children of all the variables in \( V_{C_i} \)

---

(a) A CB-decomposable MBC

(b) Its two maximal connected components
EQ-5D health states from PDQ-39 in Parkinson disease

Collaboration with Abbott Laboratories. 488 Parkinson’s patients (Borchani et al., 2012)

Empirical comparison with state of the art methods in a 5-fold cross-validation

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean accuracy</th>
<th>Exact match</th>
</tr>
</thead>
<tbody>
<tr>
<td>MB-MBC</td>
<td>0.7119 ± 0.0338</td>
<td>0.2030 ± 0.0718</td>
</tr>
<tr>
<td>CB-MBC</td>
<td>0.6807 ± 0.0285</td>
<td>0.1865 ± 0.0429</td>
</tr>
<tr>
<td>MNL</td>
<td>0.6926 ± 0.0430</td>
<td>0.1802 ± 0.0713</td>
</tr>
<tr>
<td>OLS</td>
<td>0.4201 ± 0.0252</td>
<td>0.0123 ± 0.0046</td>
</tr>
<tr>
<td>CLAD</td>
<td>0.4254 ± 0.0488</td>
<td>0.0143 ± 0.0171</td>
</tr>
</tbody>
</table>
EQ-5D health states from PDQ-39 in Parkinson disease

MB–MBC graphical structure

Pain/discomfort → Usual activities → Self-care → Mobility → Anxiety/depression

pdq38 → pdq12 → pdq15 → pdq10 → pdq8 → pdq3 → pdq2 → pdq1 → pdq5 → pdq6 → pdq7 → pdq4 → pdq17 → pdq18
Our book: Bielza and Larrañaga, 2019

Computational Neuroscience
The Statistical and Machine Learning Approach

Concha Bielza, Pedro Larrañaga

I Introduction
1. Neuroscience

II Statistics
2. Exploratory data analysis
3. Probability theory and random variables
4. Probabilistic inference

III Supervised Pattern Recognition
5. Performance evaluation
6. Feature subset selection
7. Non-probabilistic classifiers
8. Probabilistic classifiers
9. Metaclassifiers
10. Multi-dimensional classifiers

IV Unsupervised Pattern Recognition
11. Non-probabilistic clustering
12. Probabilistic clustering

V Probabilistic Graphical Models
13. Bayesian networks
14. Markov networks

VI Spatial Statistics
15. Spatial statistics

> 700 pages, > 1,300 references, 117 tables, 249 figures, ...
The fourth industrial revolution

Industrial technology shifts (Larrañaga et al. (2018))
The fourth industrial revolution

Different levels of industrial smartization: component, machine, production and distribution (Larrañaga et al. (2018))
The nine technologies transforming industrial production

- Autonomous robots
- Simulation
- Horizontal and vertical system integration
- The Industrial Internet of Things
- Cybersecurity
- The cloud
- Additive manufacturing
- Augmented reality
- Big data and analytics

Fundamental technologies for Industry 4.0. Taken from Industry4Magazine journal
Laser surface heat treatment

Time-temperature-transformation curve with a possible cooling trajectory of a hardening process (Larrañaga et al. (2018))
Monitoring the laser surface heat treatment

Physical arrangement of the different elements used to carry out and monitor the laser surface heat treatment of the steel cylinders (Gabilondo et al. (2015))
During the heat treatment the spot was programmed to avoid an obstacle on the surface of the cylinders. There were three different variants: (a) at the top, (b) in the middle, or (c) at the bottom (Larrañaga et al. (2018))
Segmentation of the frame into 14 regions

The 14 regions into which the frame was segmented. The regions adjacent to the edges were considered to be background (Larrañaga et al. (2018))
Dimensionality reduction of the feature vector $\mathbf{R}[t]$ to $\mathbf{Q}[t]$ based on segmenting the frames into $k$ different regions and extracting $s$ statistical measures from their pixel values (Larrañaga et al. (2018))
Anomalous detection method by likelihood

1. **Compute a probabilistic model** (based on dynamic Bayesian networks) for the normal instances. 2. **Establish a threshold** in this joint probability distribution. 3. **Compare the likelihood of the new instance** with the likelihood threshold.
1. **Compute a probabilistic model** (based on dynamic Bayesian networks) for the normal instances. 2. **Establish a threshold** in this joint probability distribution. 3. **Compare the likelihood of the new instance** with the likelihood threshold.
Anomaly detection method by likelihood

1. Compute a probabilistic model (based on dynamic Bayesian networks) for the normal instances. 2. Establish a threshold in this joint probability distribution. 3. Compare the likelihood of the new instance with the likelihood threshold.
Dynamic Bayesian networks (Dean and Kanazawa, 1989) are the temporal extension of Bayesian networks to stochastic processes observed at discrete time periods.

Two assumptions: first-order Markov process and stationarity.

\[
P(X[0], ..., X[T]) = \prod_{i=1}^{n} P(X_i[0] | Pa(X_i[0])) \prod_{t=1}^{T} \prod_{i=1}^{n} P(X_i[t] | Pa(X_i[t]))
\]
Transition network of the dynamic Bayesian network

Transition network learned with the DHC algorithm (Trabelsi, 2013). A vertical line separates the past and the present frames (Larrañaga et al. (2018))
Each region and its parents

Illustration of the nine regions with variables that were parents (in green) of at least one variable of the target region (in yellow). There were two types of parent regions: regions that produced instantaneous influences only (light green) and regions that produced temporal influences only (dark green) (Larrañaga et al. (2018)).
Conditional probability tables

\[ p(X|Y, Z) = \theta_X \]

Example of a conditional probability table of variable \( X \) that has two parents, \( \text{Pa}(X) = \{ Y, Z \} \) (Larrañaga et al. 2018)
INDUSTRIAL APPLICATIONS OF MACHINE LEARNING

Pedro Larrañaga
David Atienza
Javier Diaz-Rozo
Alberto Ogbechie
Carlos Puerto-Santana
Concha Bielza

Chapman & Hall/CRC Data Mining and Knowledge Discovery Series

Our book: Larrañaga et al., 2018
TCT Coach

Partido

Eventos del partido

Probabilidad de Gol

Instrucciones

Fijar objetivos y variables

Probabilidad de Gol

50% 30%

Getafe CF SD Elbar

Instrucciones

<table>
<thead>
<tr>
<th>Minuto</th>
<th>Instrucciones</th>
<th>Valor actual</th>
<th>Valor recomendado</th>
<th>Mejora en el objetivo</th>
</tr>
</thead>
<tbody>
<tr>
<td>53:00</td>
<td>Lateralización derecha</td>
<td>BAJO</td>
<td>MEDIO</td>
<td>+8%</td>
</tr>
<tr>
<td>53:00</td>
<td>Líneas de pase</td>
<td>BAJO</td>
<td>ALTO</td>
<td>+7%</td>
</tr>
<tr>
<td>53:00</td>
<td>Líneas derechitas</td>
<td>RETRASADO</td>
<td>ADELANTE</td>
<td>-2%</td>
</tr>
<tr>
<td>54:00</td>
<td>Lateralización derecha</td>
<td>BAJO</td>
<td>MEDIO</td>
<td>+8%</td>
</tr>
<tr>
<td>54:00</td>
<td>Líneas de pase</td>
<td>BAJO</td>
<td>ALTO</td>
<td>+7%</td>
</tr>
<tr>
<td>54:00</td>
<td>Líneas derechitas</td>
<td>RETRASADO</td>
<td>ADELANTE</td>
<td>-2%</td>
</tr>
</tbody>
</table>
Outline

1. Introduction
2. Bayesian Networks
3. Neuroscience
4. Industry
5. Sport
6. Conclusions
7. References
Bayesian Networks: a machine learning paradigm based on probability theory and graph theory

Advantages:
- Transparency, interpretability
- Exact inference algorithms (reasoning) for predictive reasoning, diagnostic reasoning, intercausal reasoning and abductive inference
- Learning algorithms from data for clustering, multi-view clustering, supervised classification, multi-dimensional classification, anomaly detection, discovery associations, feature subset selection, multi-output regression, ...
- Inference and learning algorithms for temporal and data stream scenarios
References


S. Luengo-Sanchez, I. Fernaud-Espinosa, C. Bielza, R. Benavides-Piccione, P. Larrañaga, and J. DeFelipe (2018). 3D morphology-based clustering and simulation of human pyramidal cell dendritic spines. *PLOS Computational Biology*, 14(6), e1006221


BAYESIAN NETWORKS IN ACTION

Pedro Larrañaga

Computational Intelligence Group
Artificial Intelligence Department
Technical University of Madrid

XVIII Conferencia de la Asociación Española para la Inteligencia Artificial
Granada, October 23, 2018