xtractis®
Augmented Fuzzy Cognitivist AI Robots
for Robust Predictive Knowledge Discovery
Is it still possible to challenge them?
Medicine of 21st century

- the Right Diagnosis
- the Right Drug in the Right Dose
to the Right Patient

Epigenetics, Oncogenetics
Cancer: 2nd / 3rd fatal disease worldwide

Hugely complex process

- ≈25K genes (often weak expression intensities)
- ≈1K metabolic & environmental variables
  (food, stress, location, happiness, work, well-being, …)
- ∝ interactions (genes ↔ genes, genes ↔ environment, …)

Unattainable for a human brain

Limitation of human comprehension
(1-3 to 7-9 variables simultaneously)
Robust Predictive Analytics at everyone’s fingertips

New Scientific Approach: automatic design of most efficient learning strategies for automatic induction of decision rules

- Find out w/o *a priori*, w/o decomposition, w/o a high-level in math
- Co-operate with your Exobrain

Discover the real complexity of your process

- Make explicit the tacit implicit knowledge
- Complex ≠ Complicated  ([Descartes 1637] was wrong!)
- Complexity cannot be reduced to few dimensions

Hire your AI Robots!

- No best Human experts could beat best AI Robots
- AI Robot can reason 24/7/365 and solve very complex problems faster and better than a Human
  - no tiredness, no holidays, no strike, at a low hourly cost
AI Robots

what ethical/unethical social impacts?

1/2

on Employment

- Less specialist jobs: drivers, doctors, actuaries, traders…
- Always manual expert jobs: building workers, surgeons…
- More AI scientists jobs (Masters, PhD) ➔ to invent new generation of Robots

on Sciences (Virtual Scientist)

- New discoveries ➔ next Nobel Prize, an AI Robot for sure!
- PhD (experimental sciences) in few months instead of 3-4 years

on Finance/Insurance (Virtual Banker/Insurer)

- Bank advisor available 24/7/365 ➔ mandatory for the digital natives!
- Credit/Insurance approval in few seconds, at the lowest price (digital service)

on Transport (Virtual Driver)

- More safety and comfort ➔ “Rolling Lounge”
- Fewer deaths and injuries BUT potentially: vehicle hacking (terrorist attacks)
AI Robots

**what ethical/unethical social impacts?**

2/2

**on Marketing/Industry** (Virtual Marketer/Engineer)
- New optimal products, fitting customer preferences

**on Health** (Virtual Doctor)
- Early diagnosis of pathologies (cancer, diabetes, Alzheimer…)
  - **BUT** potentially: refusal of insurance/credit/hiring… and eugenics risk!!
- Predictive Personalized Medicine (drug discovery)
  - **BUT** potentially: design of chemical weapons

**on Defense** (Virtual Soldier)
- Armed drones to save soldiers’ life
  - **BUT** potentially: killing authorization without remorse (= video game)

**on Justice/Security/Cybersecurity** (Virtual Judge/Policeman)
- Guarantee of an impartial decision
- Early detection of fraudulency and malicious behaviors
  - **BUT** potentially: hacking of the defender’s strategy (reverse engineering)
Certifying AI Robots

- To ensure their efficiency and ethics/honesty
  - ID cards and diplomas for AI Robots given by Human Regulator
    (DeepMind AlphaGo won the "Divine" 9th Dan in Go Game - 03/2016)
  - Universities for AI Robots? Do they still need Humans to learn?
    (xtractis® Robot builds autonomously its efficient learning strategies for each new predictive problem it has to solve)

- Trust and use AI Robots with the most renowned diplomas

Paradox: to certify complex strategies, Human must rely on an AI Robot Regulator BUT who will certify the AI Regulator?

Legal Recognition of AI Robots

- Rights & Duties, Tax on the value created by AI Robots

Paradox: IP of discoveries made by an AI Robot: granted to the Robot, to the User of the Robot or to the inventor of the Robot?

Teaching Ethics

- To AI students/specialists (≈ Hippocratic/judicial oath ≈ Do no harm!)
xtractis®
Augmented Fuzzy Cognitivist AI Robots
for Robust Predictive Knowledge Discovery
Is it still possible to challenge them?

www.xtractis.ai
xtractis@intellitech.fr

Warning
The entirety of this document is protected by copyright. Reproduction rights are reserved. Quotations from any part of the document must necessarily include the following reference:

**Fuzzy model**

**classification – Prostate Cancer**

1/2

---

**Prostate Cancer Diagnosis**

102 cases

12,600 gene expression levels

Variable to predict: 52 (51%) Tumor (1) / 50 (49%) Normal (0) diagnosis

Independent testing set from a different experiment: 25 tumor and 9 normal samples

Database source: D. Singh & al., Department of Adult Oncology, Brigham and Women’s Hospital, Harvard Medical School, [http://www-genome.wi.mit.edu/mpr/prostate](http://www-genome.wi.mit.edu/mpr/prostate)

---

**xtractis® model**

type 5 Combined model (CB5) – 500 fuzzy models

- 70 variables (2 to 11 variables per model)
- 1,000 rules (2 rules per model)

---

**Performances**

<table>
<thead>
<tr>
<th>Actual class</th>
<th>CB5 (majority voting)</th>
<th>Predicted class</th>
<th>Classification error</th>
<th>Matthews Correlation</th>
<th>Min. Sensitivity</th>
<th>Specificity</th>
<th>Unavailable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>50 (100.00%)</td>
<td>0 (0.00%)</td>
<td>0.00%</td>
<td>1.000</td>
<td>100.00%</td>
<td>0 (0.00%)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0 (0.00%)</td>
<td>52 (100.00%)</td>
<td>1.96%</td>
<td>0.961</td>
<td>98.00%</td>
<td>0 (0.00%)</td>
<td></td>
</tr>
<tr>
<td>Unavailable</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
<td>3.92%</td>
<td>0.922</td>
<td>96.00%</td>
<td>0 (0.00%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.94%</td>
<td>0.930</td>
<td>96.00%</td>
<td>0 (0.00%)</td>
<td></td>
</tr>
</tbody>
</table>

---

Results from xtractis® Generate 9.2.19175
Fuzzy model vs. KSVM classification – Prostate Cancer 2/2

KSVM model performance

<table>
<thead>
<tr>
<th>Decision</th>
<th>Classification error</th>
<th>Matthews Correlation</th>
<th>Minimum Sensitivity Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.00%</td>
<td>1.000</td>
<td>100.00%</td>
</tr>
<tr>
<td>1,000 x MC 15%</td>
<td>8.23%</td>
<td>0.836</td>
<td>89.67%</td>
</tr>
<tr>
<td>External Testing</td>
<td>26.47%</td>
<td>0.566</td>
<td>64.00%</td>
</tr>
</tbody>
</table>

Comparison

⇒ xtractis® beats KSVM by 800%
(based on increase of External Testing Error)

Results from R 3.2.0 packages: mlr 1.1.18, imputeR 1.0.0, kernlab
Tuning and Robustness Assessment modules by intellitech

Results from xtractis® Generate 9.2.19175

External Testing

<table>
<thead>
<tr>
<th>Decision</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>9 (100.00%)</td>
<td>9 (36.00%)</td>
</tr>
<tr>
<td>1</td>
<td>0 (0.00%)</td>
<td>16 (64.00%)</td>
</tr>
</tbody>
</table>

error = 26.47%

Training

<table>
<thead>
<tr>
<th>Decision</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50 (100.00%)</td>
<td>0 (0.00%)</td>
</tr>
<tr>
<td>1</td>
<td>0 (0.00%)</td>
<td>52 (100.00%)</td>
</tr>
</tbody>
</table>

error = 0.00%

Actual class

predicted class

<table>
<thead>
<tr>
<th>Decision</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50 (100.00%)</td>
<td>0 (0.00%)</td>
</tr>
<tr>
<td>1</td>
<td>0 (0.00%)</td>
<td>52 (100.00%)</td>
</tr>
</tbody>
</table>

CBS
(majority voting)

<table>
<thead>
<tr>
<th>Decision</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>9 (100.00%)</td>
<td>1 (4.00%)</td>
</tr>
<tr>
<td>1</td>
<td>0 (0.00%)</td>
<td>24 (96.00%)</td>
</tr>
</tbody>
</table>

Unavailable

0 (0.00%) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) |

error = 2.94%
**xtractis® Robot**
*a smart knowledge discoverer*

---

**Robust Predictive Modeling (KDD, DDM)**
- Automatic discovery of hidden laws ruling the Real World from big multidimensional, heterogeneous, structured datasets
- Proprietary AI algorithms: Fuzzy Theory + Machine Learning
- Holistic approach handling **weak signals** & non-linearity

---

**Advantages**
- **Interpretability** ("If…Then" decision rules) ➔ human validation, certification
- Better **predictive capacity** (robustness) vs. Open Source algorithms (Random Forest, CART Decision Tree, Boosted Trees, SVM, Neural Networks/Deep Learning, Polyn. Regression, Logistic Regression, PLS)
- **Fully automated** ➔ neither coding, nor framework

---

**Clients & Awards**
- Deloitte Technology Fast 500 – EMEA 2011 & 2012 Winner
xtractis®
4 software Robots for an “in-house” complete solution

1 - LB Building
FILTERING AGREGATION SEGMENTATION

2 - Modeling
LD : Learning Datasets
INDUCTION VALIDATION

3 - Deployment
DEDUCTION
new cases
Top-models Delivery
predictions

4 - Monitoring
SUPERVISION
perf. reporting
YES
models have expired?
NO
new Data Retrieval
optimal solutions
new requests

xtractis MONITOR

Copyright © intellitech [intelligent technologies]. 2002-2017 All rights reserved
Fuzzy Mathematics
suitable for real world data

Fuzzy set

Crisp set: Young traveler
(for airlines)

Fuzzy set: Young traveler
(for Customs & Border Police)

μ_{Young}(age)

age ∈ [0, 130] years

μ_{Young}(age)

age ∈ [0, 130] years

μ_{Young}(age)

age ∈ [0, 130] years

μ_{Young}(age)

age ∈ [0, 130] years

μ_{Young}(age)

age ∈ [0, 130] years

X = [0,130] years; ∀x ∈ X, μ_{Young}(x) = \begin{cases} 
1, & \text{if } x \in [0,20] \\
\frac{40 - x}{20}, & \text{if } x \in [20,40] \\
0, & \text{otherwise}
\end{cases}
Fuzzy model

regression – interpretable model

Sweet perception of a fresh tomato: 2 variables, 4 rules  (complexity 33.0)

Input 1: Total Acidity

Input 2: Sum of Sugars

Rule 1:
If Total Acidity is rather low
And Sum of Sugars is rather low
Then Sweet equals 3.39

Rule 2:
If Total Acidity is rather low
And Sum of Sugars is medium
Then Sweet equals 7.19

Rule 3:
If Total Acidity is high
And Sum of Sugars is above average
Then Sweet equals 3.30

Rule 4:
If Total Acidity is high
And Sum of Sugars is very high
Then Sweet equals 7.49

Data: INRA (Institut National de la Recherche Agronomique) et CTIFL (Centre Technique Interprofessionnel des Fruits et Légumes) – 7th Sensometrics Conf., July 2004, Davis, CA, USA
**Fuzzy model**

*regression – complex model*

Pesticide toxicity on trout: 11 variables, 4 rules (complexity 99.3)

---

**Partitions**

Input 1: \(nCOOR\)

Input 2: \(\log DpH7\)

---

**Rules**

1. **Rule 1**
   If \(nCOOR\) is very low
   And \(\log DpH7\) is medium
   And …
   then Toxicity_Trout equals -1.78

2. **Rule 2**
   If \(nCOOR\) is medium
   And \(\log DpH7\) is very high
   And …
   then Toxicity_Trout equals 5.96

3. **Rule 3**
   If \(nCOOR\) is very low
   And \(\log DpH7\) is high
   And …
   then Toxicity_Trout equals 6.28

4. **Rule 4**
   If \(nCOOR\) is very high
   And \(\log DpH7\) is high
   And …
   then Toxicity_Trout equals 7.74

---

**Inference**

Mapping

Decision surface (cross sections)

Data: US EPA, German BBA, EU-project SEEM
Fuzzy model classification

Breast cancer diagnosis
- 569 patient images
- 30 potential predictors
- Variable to predict: 357 (62.7%) Malignant diagnosis (1)
- 212 (37.3%) Benign diagnosis (0)
- Data: Dr. William H. Wolberg, W. Nick Street, Olvi L. Mangasarian – University of Wisconsin [UCI Machine Learning Repository]

xtractis® model
- type 4 Combined model (CB4) – 1,000 models
  - 30 variables
  - 5,159 rules

Performances

Decision surface (cross section on variables texture cell 2 & symmetry cell 3)

<table>
<thead>
<tr>
<th>Decision</th>
<th>Classification error</th>
<th>Matthews Correlation</th>
<th>Min. Sensitivity</th>
<th>Specificity</th>
<th>Refused</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training (100s x 1,000g x70%)</td>
<td>1.41%</td>
<td>0.970</td>
<td>98.11%</td>
<td>0.00%</td>
<td></td>
</tr>
<tr>
<td>Validation (100s x 1,000g x15%)</td>
<td>1.23%</td>
<td>0.974</td>
<td>98.10%</td>
<td>0.47%</td>
<td></td>
</tr>
<tr>
<td>Testing (100s x 1,000g x15%)</td>
<td>1.93%</td>
<td>0.959</td>
<td>97.17%</td>
<td>0.00%</td>
<td></td>
</tr>
</tbody>
</table>

Results by xtractis® Generate 9.1.16139
Specificities of xtractis®
robustness through cross-validation

Evaluation of the generalization capacity of predictive models
- Accurate model is not necessarily robust
- Robust model: difficult to obtain but mandatory

77 variables, 16 rules
(complexity 2,672)

Accuracy
\( r = 1.000 \)
RMSE = 0.001 (0.01%)

Type 1 Combined
(10 models: 20 rules, 21 variables, complexity \( \in \{88 ; 107\} \))

Accuracy
\( r = 0.970 \)
RMSE = 0.414 (3.94%)

Robustness (10,000 x MC 15%)
\( r = 0.849 \)
RMSE = 0.904 (8.61%)

Risk analysis of financial investment
Data: EU-SUN & SUP

Type 1 Combined
(10 models: 20 rules, 21 variables, complexity \( \in \{88 ; 107\} \))

Accuracy
\( r = 0.970 \)
RMSE = 0.414 (3.94%)

Robustness (10,000 x MC 15%)
\( r = 0.849 \)
RMSE = 0.904 (8.61%)

Trivial result:
Accurate model but non-robust

Use of robustness estimators:
Less accurate model but robust

Results by xtractis® Generate 9.1.16139
Robustness

Training/Validation cross validation (Monte-Carlo)

N cycles of y%

(100-y)% training pts

y% validation pts

(unknown situations)

Learning Points

Random draw #1

Random draw #N

Learning Strategy S

Strategy S

Model M

y% predictions

Model M

y% predictions

Model M

y% predictions

Model M

y% predictions

Performances of (N x y%) predictions made in unknown situations = Model M Robustness

What robustness?
Specificities of xtractis® noise detection

Gaussian noise, \( \sigma = 20\% \times \mathbf{1}[0;4] \)
- RMSE (learning points / hidden values) = 0.74 (18.39%)
- Correlation (learning points / hidden values) = 0.753

Robustness 10,000 x MC 25%
- Accuracy: \( r = 0.995 \), RMSE = 0.12 (3.00%)
- Robustness: \( r = 0.643 \), RMSE = 0.93 (23.25%)

Robustness 10,000 x MC 25%
- Accuracy: \( r = 0.792 \), RMSE = 0.69 (17.25%)
- Robustness: \( r = 0.706 \), RMSE = 0.79 (19.75%)

Modeling with noisy data (without and with robustness analysis)
Weak signals

importance of predictors with weak individual influence

Breast Cancer Diagnosis
(30 potential predictors)

Top-model: 21 predictors, 3 rules

Predictors with weak individual influence

Best learning strategy applied on the 2 predictors presenting the strongest individual influences (Cartesian approach)

2 predictors, 8 rules
Data for xtractis®
structured, quantitative/qualitative

<table>
<thead>
<tr>
<th>Label</th>
<th>Binary Variable</th>
<th>Nominal Variable</th>
<th>Numerical Variable</th>
<th>...</th>
<th>Var 1 to predict</th>
<th>Var 2 to predict</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation 1</td>
<td>0</td>
<td>Modality A</td>
<td>1.2558</td>
<td></td>
<td>2.35</td>
<td>1</td>
</tr>
<tr>
<td>Observation 2</td>
<td>1</td>
<td>Modality C</td>
<td>0.2356</td>
<td></td>
<td>1.256</td>
<td>0</td>
</tr>
<tr>
<td>Observation 3</td>
<td>1</td>
<td>Modality D</td>
<td>4.568</td>
<td></td>
<td>12.03</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>1</td>
<td></td>
<td>125.36</td>
<td></td>
<td>1.002</td>
<td>1</td>
</tr>
<tr>
<td>Observation n</td>
<td>0</td>
<td>Modality A</td>
<td></td>
<td></td>
<td>7.056</td>
<td>0</td>
</tr>
</tbody>
</table>

Potential predictors

Variables to predict

- Regression
- Classification
- Scoring

Observations

Missing values allowed

Copyright © intellitech [intelligent technologies]. 2002-2017  All rights reserved