Tecnologías del lenguaje para Explainable-AI y su impacto en el soporte a la decisión. Algunas aplicaciones a salud

K. Gibert

Knowledge Engineering & Machine Learning group at Intelligent Data Science and Artificial Intelligence Res.C. (KEMLG-at-IDEAI).

Universitat Politècnica de Catalunya-BarcelonaTech, Spain

Vice dean of Big Data, Data Science and Artificial Intelligence

Official Chamber of Informatics Engineering of Catalonia

Interest Group IABIomed-Spanish Association of Artificial Intelligence (CAEPIA)

InfoDay sobre tecnologías del Lenguaje en sanidad y Biomedicina

BSC, Barcelona 2, diciembre 2019
Outline

• Introduction: Automatic interpretation of profiles

• Knowledge acquisition tools
  – Prior knowledge bases
  – Ontologies
  – Termometer
  – Super-concept based distance

• Explainability through embedded strategies in Data Science methods
  – Clustering based on rules and ontologies

• Profiles oriented Explainability tools
  – Visual: TLP, a-TLP
  – Conceptual: CCEC, CI-IMS
  – Dinamic: Trajectory map, Adherence map

• Knowledge production tools

• Other cases: Topic modelling, Explainability in ANN

• Conclusions
The Fact Gap: The Disconnect Between Data and Decisions  

[Hammond 2004]

No analysis  No understandable  No trust

Needs to be general literacy about data interpretation  [A “Sandy” Pentland]

keynote Campus Party Europa Sept 4th 2013 Head of MediaLab Enterpreneurship MIT
Data Science concept

- 2018: Gibert, Horsburg, Athanasiadis, Holmes [ENVSOFT, 2018]

Data science: emergent multidisciplinary field combining
  - Data analysis
  - Data processing
  - Domain expertise

To transform data into understandable and actionable knowledge
Relevant for informed decision making (reduces the Fact Gap)

- involves intensive consumption of available and required data
- Copes with data heterogeneity
- BigData is a tool, not the focus, but domain complexity

DOI:10.1016/j.envsoft.2018.04.005
Data Mining and Knowledge Discovery

- Knowledge Discovery System [Fayy96]:

**Focus: Clustering/Profiling**
Expert-based collaborative Analysis (EbCA)


© K. Gibert
Profiling mental health systems in LAMIC

KLASS
Clustering based on rules [Gib’96]
Ward’s criterion
Gibert’s mixed Metrics [Gib’97]

WHO-AIMS data
Rows: 42 LAMIC
Columns:
- 19 WHO-AIMS indicators
- WHO-AIMS region
- Income group (World Bank classification)

Prior Expert Knowledge

- \( r_0 \): Region=euro and income= lower
  \( \rightarrow \) PreSoviet

- \( r_1 \): Region=AMR and income = lower
  \( \rightarrow \) poorAmerica

- \( r_2 \): Region= SEAR and population < 10000000
  \( \rightarrow \) smallAsia

Profiling mental health systems in LAMIC

Prior Expert Knowledge Acquisition

Prior Expert Knowledge

- r0: Region = euro and income = lower
  -> PreSoviet
- r1: Region = AMR and income = lower
  -> poorAmerica
- R2: Region = SEAR and population < 10000000
  -> smallAsia

KLASS
Clustering based on rules [Gib'96]
Ward’s criterion
Gibert’s mixed Metrics [Gib’97]

WHO-AIMS data
Rows: 42 LAMIC
Columns:
19 WHO-AIMS indicators + WHO-AIMS region + Income group (World Bank classification)
Profiling mental health systems in LAMIC countries for healthcare policy-making at WHO

Use WHO-AIMS DB to learn a **typology** of MHS in LAMIC

- Easy **understanding** of reality
- Assessment to countries
- Intervention design: guidelines, mental health policies....

### TLP elicits clustering criteria

Conceptualization

Induces categories of variables and classes

<table>
<thead>
<tr>
<th>BLOCK</th>
<th>CLASS</th>
<th>CARE CAPACITY</th>
<th>CARE ARRANGEMENT</th>
<th>POLICY FRAMEWORK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Incom</td>
<td>HR</td>
<td>$\text{MHe}$</td>
</tr>
<tr>
<td>I</td>
<td>Cl</td>
<td>Low-Moderate</td>
<td>UpMid</td>
<td>Highst</td>
</tr>
<tr>
<td>C6</td>
<td>Mod</td>
<td>Low</td>
<td>Mod</td>
<td>Mod</td>
</tr>
<tr>
<td>C22</td>
<td>Mod</td>
<td>Mod</td>
<td>Low</td>
<td>Mod</td>
</tr>
<tr>
<td>II</td>
<td>C21</td>
<td>Mid-Limited</td>
<td>LMid</td>
<td>Low</td>
</tr>
<tr>
<td>C20</td>
<td>LMid</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>C19</td>
<td>LMid</td>
<td>Low</td>
<td>Low</td>
<td>Lowest</td>
</tr>
<tr>
<td>C18</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>

Supports data-driven Ontologies

Knowledge Production

MHS for LAMIC ontology

Used by WHO for recommendations to LAMIC countries

Intervention plans designed for each type
The KLASS thermometer-tool

- hr/100000pop
- US$ per capita
- treated/100000pop
- % (treated pacs/prev)
- (%f totbeds) per cap
- % from MH budget
- % from pacs in MHos
- % from treated pacs
- Outcontacts/daysHsp
## TLP elicits clustering criteria

**Conceptualization**

Induces categories of variables and classes

<table>
<thead>
<tr>
<th>Block</th>
<th>Class</th>
<th>Care Capacity</th>
<th>Care Arrangement</th>
<th>Policy Framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Upper-Moderate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td></td>
<td>Incom</td>
<td>HR</td>
<td>$%MHe$</td>
</tr>
<tr>
<td>C6</td>
<td></td>
<td>UpMid</td>
<td>Mod</td>
<td>Mod</td>
</tr>
<tr>
<td>C22</td>
<td></td>
<td>LMid</td>
<td>Mod</td>
<td>Low</td>
</tr>
<tr>
<td>II</td>
<td>Mid-Limited</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C21</td>
<td></td>
<td>LMid</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>C20</td>
<td></td>
<td>LMid</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>C19</td>
<td></td>
<td>LMid</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>C18</td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>

**Supports data-driven Ontologies**

Semantic Distances

$X_i$  $X_k$

$i_1$  $i_n$

Numerical  Categorical  Linguistic

Semantic Similarity

Distance for numeric vars

Categorical distance

Semantic Similarity

Distance Matrix

Combination

Distance Matrix

Clustering

Ontology

Ontology 1

Ontology 2

Ontology N

WordNet

a-TLP: going further (WWTP case)

<table>
<thead>
<tr>
<th>Class</th>
<th>nc</th>
<th>Influent</th>
<th>2nd An-T</th>
<th>1st A-T</th>
<th>2nd A-T</th>
<th>Effluent</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Q</td>
<td>NH₄</td>
<td>TN</td>
<td>TOC</td>
<td>Ni</td>
<td>Tri tox</td>
</tr>
<tr>
<td>C360</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C358</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C353</td>
<td>122</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C357</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CCEC: Conceptual Caracterization by Embedded Conditioning

Exploits dendrogramm structure to induce classification rules

\[ r1.BC0.\neg r50r0: ((\text{treatpre} \in [18, 57, 172, 77]) \land (\text{comcarewor} \in [0, 0197, 0, 1098])) \land (\text{Region} \in \{\text{AFR}\}) \rightarrow (\text{NovaClasseBLN7})C18 \]

\[ r2.BC1.\neg r2-r46-r50r0-r35-r37-r39: (((\text{treatpre} \in [18, 57, 172, 77]) \land (\text{comcarewor} \in [0, 0197, 0, 1098])) \land (((\text{Region} \in \{\text{SEAR}\}) \lor (\text{lundpararectrail} \in [0, 49, 0, 53])) \lor (\text{comcarewor} \in [0, 0197, 0, 0255]))) \land (((\text{Region} \in \{\text{SEAR}\}) \land (\text{treatpre} \in [31, 81, 87, 59]))) \land (\text{lundpararectrail} = 0, 49) \land (\text{comcarewor} = 0, 0197) \rightarrow (\text{NovaClasseBLN7})C19 \]

\[ r4.BC3.\neg r46r53: (\text{treatpre} \in [18, 57, 172, 77]) \land (\text{comcarewor} \in [0, 1313, 0, 624]) \rightarrow (\text{NovaClasseBLN7})C20 \]
CIMS: Cluster Interpretation based on Integrated Marginal Significance

Same differences with same conceptualizations in all classes

**Consistency Inter Classes: Generalized Test – Value**

### Numerical

\[
\tau_v = \frac{\bar{X}^c - \bar{X}}{\sqrt{\left(1 - \frac{n_c}{n}\right)s^2}} \sim t_{v-1}
\]

### Qualitative

\[
\pi_v = \frac{p_{sc} - p_s}{\sqrt{\left(1 - \frac{n_c}{n}\right)p_s(1 - p_s)}} \sim z
\]

#### Sensitivity Analysis

\[\downarrow \nu \rightarrow \uparrow p\text{-value}\]

<table>
<thead>
<tr>
<th>Description-Power ((\Pi))</th>
<th>0.5(n)</th>
<th>0.7(n)</th>
<th>(n)</th>
<th>1.3(n)</th>
<th>1.5(n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robust Non-Significant ((R))</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\times)</td>
</tr>
<tr>
<td>Moderate Non-Significant ((M))</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\checkmark)</td>
<td></td>
</tr>
<tr>
<td>Weak Non-Significant ((W))</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>Weak Significant ((W))</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>Moderate Significant ((M))</td>
<td>(\times)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>Robust Significant ((R))</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>Basic Descriptor ((B))</td>
<td>(\text{B})</td>
<td>(\text{B})</td>
<td>(\text{B})</td>
<td>(\text{B})</td>
<td>(\text{B})</td>
</tr>
</tbody>
</table>

#### Class Descriptor

\(< W, C, \text{description-power, sense}>\)

\[W = \begin{cases} <X, s> & \text{if } X \text{ numerical} \\ X & \text{if } X \text{ categorical} \land s \text{ category } \in D_X \end{cases}\]

\[\text{sense } \in \{\uparrow, \downarrow\}\]

#### Regular Expressions

- Proportion of Smokers (Tobacco) is higher in C1
- Weight is high in class C2
- Age is lowest in class C1

K. Gibert
Interpreting X in Nested Partitions

\[ C = \bigcup_{c' \in S_C} c' \in P' \]

Relationship between interpretation of X in C and \( S_C \)

<table>
<thead>
<tr>
<th>Super Class</th>
<th>Non-Significant</th>
<th>Significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub Classes</td>
<td>Irrelevance ( \forall c' \in S_C ): ( R(C, c') ) = Irrelevance</td>
<td>Inconsistency ( \forall c' \in S_C ): ( R(C, c') ) = Inconsistency</td>
</tr>
<tr>
<td>Non-Significant</td>
<td>Specification ( \exists c' \in S_C ): ( R(C, c') ) = Specification</td>
<td>Inheritance ( \exists c' \in S_C ): ( R(C, c') ) = Inheritance</td>
</tr>
<tr>
<td>Significant</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**NCI-IMS:** *Cluster Interpretation based on Integrated Marginal Significance for Nested partitions*

Table Ą : Actions associated to Table $R$

<table>
<thead>
<tr>
<th>Super Class SubClass</th>
<th>$R$</th>
<th>$M$</th>
<th>$W$</th>
<th>$W$ in description of $C$ and $C' \in S_C$</th>
<th>$M$ in description of $C$ and $C' \in S_C$</th>
<th>$W$ in description of $C' \in S_C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$W$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$W$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R/B$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Trajectory maps

More typical patterns ($\gamma \geq 0.05$)

Knowledge Discovery about Quality of Life changes of Spinal cord Injury patients: Clustering based on rules by States. In Studies in Health Technology and Informatics, v150: 579—583. IOSSPress
Expert’s conceptualization of patterns


Physical autonomy and psychological wellness maintained over time.


High impairment. Starting with different coping strategies. Long term adaptation to moderate distress, no anxiety.
Trajectory Characterization. Adherence

Assignment of the profile of a new patient

Given a new patient:

Estimate $\pi_{High}$ by applying equation 1

If $p_{High} \geq \xi$ then assign patient to High profile

Else, Estimate $\pi_{IntII}$ by applying equation 2

If $p_{IntII} \geq \xi$ then assign patient to IntermediateII profile.

Else Estimate $\pi_{IntI}$ by applying equation 3.

If $p_{IntI} \geq \xi$

then assign patient to IntermediateI profile.

Else assign patient to Low profile.


https://doi.org/10.1016/j.envsoft.2009.11.004
The Profile Assessment Grid

Identifying the profile of a new patient

Misclassification Rate (ε=0.5): 8.3%

\[ P(\text{High}) = e^{-35.93 + 1.70B2 + 3.35B4 + 3.98B9 + 2.20S4} \cdot (1 + e^{-35.93 + 1.70B2 + 3.35B4 + 3.98B9 + 2.20S4}) \]

<table>
<thead>
<tr>
<th>B2</th>
<th>B4</th>
<th>B9</th>
<th>S4</th>
<th>S5</th>
<th>S9</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

P(\text{High}) | P(\text{Int-I}) | P(\text{Int-II})
---|---|---
0.1 | 0.07 | 0.20
PCA for topic modelling

Find terms with significant contributions to axes

Generalize in the reference ontology (Wordnet by default)
Discover the latent variables (automatic interpretation of axes)
Interpreting ANN

Visualization of Input Effect: VEC curve (Bank Marketing)

Conclusions

• Explainable models required for trust and decisions
• Post-processing provide explainability
  • Visual tools: TLP/a-TLP (profiling), PAG (predictive models)
  • Conceptual: CCEC, CI-MIS (machine readable)
  • Dynamics: trajectory maps, adherence maps
• Prior knowledge transfer to models increase explainability
  • Termometers (semantics of variables, polarities)
  • Prior Knowledge Bases
  • Ontologies (semantic relations between terms)
• Language technologies play a relevant role in building these tools
Tecnologías del lenguaje para Explainable-AI y su impacto en el soporte a la decisión
Algunas aplicaciones a salud

K. Gibert
karina.gibert@upc.edu
https://www.eio.upc.edu/en/homepages/karina

KEMLG-@-IDEAI: Knowledge Engineering and Machine Learning group at Intelligent Data Science and Artificial Intelligence Research Center

Universitat Politècnica de Catalunya, Barcelona

Are there any questions?...

InfoDay sobre tecnologías del Lenguaje en sanidad y Biomedicina
BSC, Barcelona 2, diciembre 2019