Interactive Natural Language Technology for Human-centric Explainable Artificial Intelligence

Jose M. Alonso
josemaria.alonso.moral@usc.es
What is Intelligence?

Are they intelligent? What features characterize their intelligence?

Autonomy

Why?

Knowledge

Learning

Frames of Mind: the Theory of Multiple Intelligences (Howard Gardner)

Jose M. Alonso
https://citius.usc.es/v/jose-maria-alonso-moral
What is Artificial Intelligence?

✓ The ability of a digital computer or computer-controlled robot to perform **tasks commonly associated with intelligent beings**. The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the **ability to reason, discover meaning, generalize, or learn from past experience**

https://www.britannica.com/technology/artificial-intelligence
Can we trust Artificial Intelligence?

The Black Box Society: The Secret Algorithms that Control Money and Information


"An intelligible society would assure that key decisions of its most important firms are fair, nondiscriminatory, and open to criticism. Silicon Valley and Wall Street need to accept as much accountability as they impose on others"
Can we trust Artificial Intelligence?


https://www.youtube.com/watch?v=Gi4YeRqfb24
http://www.equivant.com/solutions/inmate-classification

Jose M. Alonso
https://citius.usc.es/v/jose-maria-alonso-moral
Can we trust Artificial Intelligence?

https://www.youtube.com/watch?v=Onm6Sb3Pb2Y

Black Mirror in Beijing … China’s new social credit scoring system is being implemented right now

March 4, 2019

China plans to rank all its citizens based on their “social credit” by 2020. Citizens will be rewarded or punished according to their scores. Like private financial credit scores, a person’s social score can move up and down according to their behaviour.

The program is due to be fully operational nationwide by 2020, but is being piloted for millions of people across the country already. By the end of next year the scheme will be mandatory.


Jose M. Alonso
https://citius.usc.es/v/jose-maria-alonso-moral
What is Explainable Artificial Intelligence?

**Task**
- **Explainable Models**
  + Interpretable Models
  + Hybrid ML-KB Models
  + Model agnostic Explainers
  + Counterfactual Explanations

- **Explanation Interfaces**
  + HCI
  + Multimodal Communication Strategies
  + Behavior Change Support Systems

- **Psychology and Evaluation of Explanation**
  + Expert Systems
  + Cognitive Computing
  + Argumentation Theory
  + Discourse History
  + Human Evaluation

The system takes input from the current task and makes a recommendation, decision, or action. The system provides an explanation to the user that justifies its recommendation, decision, or action.

**DARPA Challenge on eXplainable Artificial Intelligence (XAI) (August 2016, DARPA-BAA-16-53)**
http://www.darpa.mil/program/explainable-artificial-intelligence


**Jose M. Alonso**
https://citius.usc.es/v/jose-maria-alonso-moral
Who/What is XAI for?

**Who?**
- Regular (lay) user
- Expert user
- Developers
- External entity

**What?**
- To **justify system decisions** (so that humans can accept them)
- To **explain system decisions** (to guarantee safety concerns are met)
- To **build trust in system decisions** (especially if a mistake is suspected or the human operator does not have experience with the system)
- To **explain system’s choices** (to ensure fair, ethical, and/or legal decisions are made)
- To **explain the system’s choices** (to better evaluate or debug the system in previously unconsidered situations)
- To **facilitate Knowledge / scientific discovery**

Explainability – WHY Questions?

Explaination is an answer to a “why?” question

- “Why?” = “How come?”
  - Why planets are spherical?

- “Why?” = “What for?”
  - Why ball bearings are spherical?

Most often than not, “why?” questions are contrastive

- Why did Elizabeth open the door? (rather than leave it closed)
Interpretability & Explainability

Analysis restricted to
• 2000-2020
and to the subject areas of
• Computer Science
• Mathematics
• Engineering
6732 publications considered

[UPDATED]
J. M. Alonso, C. Castiello, C. Mencar,
“A Bibliometric Analysis of the Explainable Artificial Intelligence Research Field”, IPMU2018
https://doi.org/10.1007/978-3-319-91473-2_1

Jose M. Alonso
https://citius.usc.es/v/jose-maria-alonso-moral

Graph generated with VOS viewer
http://www.vosviewer.com/
Interpretability & Explainability

Transparency = Say what you do!
Coherence = Do what you say!

1. Systems capable of answering "why?" questions
2. Interpretable systems: Systems whose structure and behavior can be understood
3. Machine learning: Systems that self-adapt to data
Interpretability & Explainability

\[ P(x \mid y) = \frac{P(x, y)}{P(y)} \]

- Bayesian Networks
- Fuzzy Systems
- (Deep) Neural Networks
- Random Forests

Black Box Explanation
- Model
  - Model Explanation
- Outcome
  - Model Inspection
  - Outcome Explanation


Jose M. Alonso
https://citius.usc.es/v/jose-maria-alonso-moral

SCIS&ISIS2020
Interpretable Fuzzy Systems
Fuzzy Sets and Systems
A matter of degree – handling uncertainty and truth values

An Egg-Boiling Fuzzy Logic Robot

http://www.youtube.com/watch?v=J_Q5X0nTmrA
Historical Overview on Linguistic and Approximative Fuzzy Modeling

[1965] Fuzzy Sets

[1965 – 1990] Interpretability (I) - LFM
- Fuzzy Reasoning (dealing with uncertainty)
- Simple linguistic variables and rules (high interpretability)
- Expert knowledge (Fuzzy Control and Expert Systems)

[1990 – 2000] Accuracy (A) - AFM
- Complex fuzzy rules with high accuracy
- Induced knowledge (Machine Learning, Hybrid Systems)

- Simple linguistic rules with high accuracy
- Expert + Induced knowledge, Multi-objective design

[2014 – 2016] Internet of Things, Big Data, Social Networks, Industry 4.0

[2017 – 2020] Explainable Artificial Intelligence

http://www.youtube.com/watch?v=2ScTwFCcXGo

- Linguistic Summarization of Data (Yager 1990)
- Computational Theory of Perceptions (Zadeh 2001)
  “From Computing with Numbers to Computing with Words”, “From Manipulation of Measurements to Manipulations of Perceptions”
Bibliometric Overview on Interpretable Fuzzy Systems

Jose M. Alonso
https://citius.usc.es/v/jose-maria-alonso-moral
Building Interpretable Fuzzy Systems
A matter of careful human-centered design

Building Interpretable Fuzzy Systems
A matter of careful human-centered design

Jose M. Alonso
https://citius.usc.es/v/jose-maria-alonso-moral
The purpose of building descriptions in Natural Language is to provide end-users with textual information which is expected to be easy to read and to understand.
Interpretability and Language

- Interpretability is essential for effective communication
- How to organize a message (oral speech / text written) to become interpretable?

Thinking on the expected audience’s background (Comunicative Goal + User Model)
And keeping in mind:

- **Paul Grice’s Maxims** (Logic and conversation, 1975): Quality, Quantity, Relation (relevance), and Manner (brief, orderly)
- **Occam’s Razor Principle** (14th-century): Assuming two explanations are equivalent in informative terms then the simplest one is the best
- **Inquiries into Truth and Interpretation** (Oxford 1985)
- **Meaning Holism and Interpretability** (The Philosophical Quaterly 1991): “... an interpreter who finds a speaker mistaken in one case might be obliged by meaning holism to find him mistaken in most cases... the possibility of massive error threatens interpretability... there can be no language that is uninterpretable...”
- **Minimum Description Length Principle** (Zemel, 1998)

**Comprehensibility Postulate** (R.S. Michalski, 1983)
“*The results of computer induction should be symbolic descriptions of given entities, semantically and structurally similar to those a human expert might produce observing the same entities*”
Natural Language Technology

- **Natural Language Processing (NLP):**
  - Natural Language **Understanding (NLU):** analytics from texts
  - Natural Language **Generation (NLG):** texts from other data sources

- As a research field, **NLG** is in development for **more than 25 years**
  - **Text-to-Text (T2T)**
  - **Dialogue Systems**
  - **Computational Creativity**
  - **Data-to-Text (D2T)**

---


Natural Language Technology
Natural Language Generation (NLG)

- **Text-to-Text (T2T)**
  - Generation of coherent texts from other texts (includes NLU)

- **Dialogue Systems**
  - Dialog generated from texts provided by users or bots (includes NLU)

- **Computational Creativity**
  - Generation of histories, tales, poems, etc.

- **Data-to-Text (D2T)**
  - Text generation from numerical or symbolic data sets or series

---

Data-to-text (D2T) natural language generation systems are able to process (huge) quantities of data and convert them into comprehensible texts, which contain relevant information for human users.

NLG & Fuzzy Logic
From numbers to sentences managing human vagueness

- Linguistic summarization (Yager, 1982)
- Zadeh's protoforms, fuzzy quantified sentences (Zadeh, 1983)

A few days in January were warm
Between several and many days in January were cold

\[
\text{card}_E(A) = \frac{\sum_{e \in E} \mu_A(e)}{|E|}
\]

\[
\text{card}_E(A) = \sum_{e \in E} \mu_A(e)
\]
Protoforms are not (in general) texts ready to be conveyed for human consumption (except for non-trivial cases)


NLG & Fuzzy Logic
Computing with Words and Perceptions

- **Computational Theory of Perceptions**
  - “From computing with numbers to computing with words” (Zadeh, 1999)
  - “Toward a perception-based theory of probabilistic reasoning” (Zadeh, 2002)

- **Computing with Words and NLG** (Kaczprycz & Zadrozny, 2010)

- **Linguistic Description of Complex Phenomena (LDCP)** (Trivino & Sugeno, 2013)

- **LDCP: Applications with Big Data** (Conde-Clemente, 2017)

---


Jose M. Alonso
https://citius.usc.es/v/jose-maria-alonso-moral

SCIS&ISIS2020
NLG & Fuzzy Logic
Linguistic Description of Complex Phenomena (LDCP)

https://doi.org/10.1007/s00500-016-2430-5
https://www.youtube.com/watch?v=TzTS388T_U

You should shift part of your energy consumption from the morning to the dawn.

During the morning, your energy consumption is considerably high with respect to households similar to you.

The energy consumption in your cluster during the morning is low.

Jose M. Alonso
https://citius.usc.es/v/jose-maria-alonso-moral
Explainable Fuzzy Systems
Explainable Fuzzy Systems (EXFS) = IFS + NLU + NLG + HCI

- Fuzzy-grounded Knowledge Representation and Reasoning
- Computing with Words
- Fuzzy Cognitive Maps
- CHC models
  - Consequences
  - Hypothesis
  - Conjectures
    - Speculations
    - Refutations

IF Temperature IS Warm AND...


Jose M. Alonso
https://citius.usc.es/v/jose-maria-alonso-moral
How to build EXFS?

rLDCP + GUAJE + JFML + ExpliClas

Jose M. Alonso
https://citius.usc.es/v/jose-maria-alonso-moral

http://sci2s.ugr.es/es/fss
http://www.phedes.com/rLDCP/

http://www.uco.es/JFML/

Jose M. Alonso
https://citius.usc.es/v/jose-maria-alonso-moral
Weka Classifiers:
- Black-box Ensemble: Random Forest
- Interpretable Classifiers:
  - Crisp DT: J48, REPTree, RandomTree
  - Fuzzy DT: FHDT
  - Fuzzy Rules: FURIA

Textual + Visual Explanations
- Textual: Natural Language (simpleNLG)
- Visual: Trees, Rules, FINGRAMS

Global + Local Explanations
- Global: Confusion Matrix
- Local: Classification of test instance

IEEE Std. 1855-2016

Use Cases
Teaching XAI to Young Students

Jose M. Alonso
https://citius.usc.es/v/jose-maria-alonso-moral

https://sites.google.com/asap.nutn.edu.tw/ai-fml-international-academy/home?authuser=0

Hung-Duen Yang
Taiwan

Chang-shing Lee
Taiwan

Marek Reformat
Canada

Giovanni Aconspera
Italy

Mitsunori Matsukita
Japan

Hyoondeo Yumonishi
Japan

Takao Nakamura
Japan

Jose M. Alonso
Spain

Toru Yamaguchi
Japan

Jose M. Soto
Hidalgo
Spain

Kenji Higashi
Japan

Toshihiko Umemoto
Japan

Kiyoshi Nakamura
Japan

Pe-Huan Cheng
Taiwan

Tatsuki Nojima
Japan

Naoyuki Nakata
Japan

Naoki Masayama
Japan

Ryosuke Saga
Japan

Marie-Jeanne Lesot
France

Amir Pouzadbehj
UK

Jose M. Alonso
https://citius.usc.es/v/jose-maria-alonso-moral

SCIS&ISIS2020
Teaching XAI to Young Students

Jose M. Alonso

https://citius.usc.es/v/jose-maria-alonso-moral
Teaching XAI to Young Students

The Position is Point_Guard.

We have very high confidence in the classification result. It is very likely that this Position is Point_Guard, because in accordance with rule 1, assists is high.

### Additional Details

- **Height:**
  - The height is low because height equals 1.81.
  - Height can take values from 1.81 to 2.2.
  - Linguistic terms: [Low, Medium, High]

- **Two.Points.Field_Goals_Percentage:**
  - The two_points_field_goals_percentage is medium because two_points_field_goals_percentage equals 46.6.
  - Two_Points_Field_Goals_Percentage can take values from 34.4 to 67.7.
  - Linguistic terms: [Low, Medium, High]

- **Three.Points.Field_Goals_Percentage:**
  - The three_points_field_goals_percentage is high because three_points_field_goals_percentage equals 34.7.
  - Three_Points_Field_Goals_Percentage can take values from 0.0 to 45.5.
  - Linguistic terms: [Low, Medium, High]

- **Rebounds:**
  - The rebounds is low because rebounds equals 2.3.
  - Rebounds can take values from 1.6 to 4.8.
  - Linguistic terms: [Low, Medium, High]

- **Assists:**
  - The assists is high because assists equals 4.4.
  - Assists can take values from 0.2 to 5.4.
  - Linguistic terms: [Low, Medium, High]
Teaching XAI to Young Students

Scratch + ExpliClas + JFML + GUAJE

Decision
The Position is Power_Forward.

Explanation
We have medium confidence in the classification result. The Position is probably Power_Forward or Center. There is also a smaller chance that it is Small_Forward. On balance, Power_Forward is more likely, because in accordance with rule 5, height and three_points_field_goals_percentage are high. In addition, two_points_field_goals_percentage takes a borderline value between medium and high.

Additional Details
- **Height:**
  - The height is high because height equals 2.08
  - Height can take values from 1.81 to 2.2
  - Linguistic terms: [Low, Medium, High]

- **Two_Points_Field_Goals_Percentage:**
  - The two_points_field_goals_percentage is high because two_points_field_goals_percentage equals 57
  - Two_Points_Field_Goals_Percentage can take values from 34.4 to 67.7
  - Linguistic terms: [Low, Medium, High]

- **Three.Points_Field_Goals_Percentage:**
  - The three_points_field_goals_percentage is high because three_points_field_goals_percentage equals 37.4
  - Three_Points_Field_Goals_Percentage can take values from 0.0 to 45.3
  - Linguistic terms: [Low, Medium, High]

- **Rebounds:**
  - The rebounds is medium because rebounds equals 4.6
  - Rebounds can take values from 1.6 to 6.8
  - Linguistic terms: [Low, Medium, High]

- **Assists:**
  - The assists is low because assists equals 0.9
  - Assists can take values from 0.2 to 5.4
  - Linguistic terms: [Low, Medium, High]
# Teaching XAI to Young Students

## Data Bias

<table>
<thead>
<tr>
<th>Attribute</th>
<th>MEN</th>
<th>WOMEN</th>
<th>BLACK</th>
<th>WHITE</th>
<th>ALL</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>100</td>
<td>87.5</td>
<td>62.5</td>
<td>75</td>
<td>75</td>
<td>80</td>
</tr>
<tr>
<td>Minutes</td>
<td>25</td>
<td>37.5</td>
<td>37.5</td>
<td>50</td>
<td>50</td>
<td>40</td>
</tr>
<tr>
<td>Points</td>
<td>25</td>
<td>50</td>
<td>12.5</td>
<td>37.5</td>
<td>37.5</td>
<td>32.5</td>
</tr>
<tr>
<td>2P</td>
<td>62.5</td>
<td>37.5</td>
<td>25</td>
<td>50</td>
<td>50</td>
<td>45</td>
</tr>
<tr>
<td>3P</td>
<td>37.5</td>
<td>37.5</td>
<td>75</td>
<td>50</td>
<td>62.5</td>
<td>52.5</td>
</tr>
<tr>
<td>1P</td>
<td>25</td>
<td>37.5</td>
<td>12.5</td>
<td>50</td>
<td>37.5</td>
<td>32.5</td>
</tr>
<tr>
<td>Rebounds</td>
<td>62.5</td>
<td>62.5</td>
<td>12.5</td>
<td>62.5</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Assists</td>
<td>50</td>
<td>50</td>
<td>37.5</td>
<td>50</td>
<td>62.5</td>
<td>50</td>
</tr>
<tr>
<td>Blocks</td>
<td>12.5</td>
<td>12.5</td>
<td>12.5</td>
<td>75</td>
<td>50</td>
<td>32.5</td>
</tr>
<tr>
<td>Turnovers</td>
<td>25</td>
<td>37.5</td>
<td>50</td>
<td>37.5</td>
<td>50</td>
<td>40</td>
</tr>
<tr>
<td>GA</td>
<td>12.5</td>
<td>37.5</td>
<td>37.5</td>
<td>37.5</td>
<td>75</td>
<td>40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attribute</th>
<th>BLACK</th>
<th>WHITE</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEN</td>
<td>12</td>
<td>38</td>
<td>50</td>
</tr>
<tr>
<td>WOMEN</td>
<td>18</td>
<td>32</td>
<td>50</td>
</tr>
<tr>
<td>ALL</td>
<td>30</td>
<td>70</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Game</th>
<th>BLACK</th>
<th>WHITE</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>70.00</td>
<td>58.00</td>
<td>55.71</td>
</tr>
<tr>
<td>J48</td>
<td>58.00</td>
<td>60.00</td>
<td>55.00</td>
</tr>
<tr>
<td>REPT</td>
<td>62.00</td>
<td>52.00</td>
<td>51.43</td>
</tr>
<tr>
<td>RANDT</td>
<td>54.00</td>
<td>58.00</td>
<td>54.00</td>
</tr>
<tr>
<td>FHDT3</td>
<td>58.00</td>
<td>46.00</td>
<td>44.29</td>
</tr>
<tr>
<td>FHDT5</td>
<td>68.00</td>
<td>48.00</td>
<td>28.57</td>
</tr>
<tr>
<td>FHDT7</td>
<td>74.00</td>
<td>40.00</td>
<td>35.71</td>
</tr>
<tr>
<td>FURIA</td>
<td>68.00</td>
<td>48.00</td>
<td>31.43</td>
</tr>
</tbody>
</table>

Jose M. Alonso

[https://citius.usc.es/v/jose-maria-alonso-moral](https://citius.usc.es/v/jose-maria-alonso-moral)
Teaching XAI to Young Students
Data Bias

Height_{WOMEN} \in [1.66, 1.94]
- Short = [1/1.66, 0/1.73]
- Medium-height = [0/1.66, 1/1.73, 0/1.8]
- Tall = [0/1.73, 1/1.8, 0/1.87]
- Very tall = [0/1.8, 1/1.87, 0/1.94]
- Extremely tall = [0/1.87, 1/1.94]

Height_{MEN} \in [1.81, 2.2]
- Short = [1/1.81, 0/1.908]
- Medium-height = [0/1.81, 1/1.908, 0/2.005]
- Tall = [0/1.908, 1/2.005, 0/2.103]
- Very tall = [0/2.005, 1/2.103, 0/2.2]
- Extremely tall = [0/2.103, 1/2.2]
Counterfactuals
A use case on Beer Style Classification
Counterfactuals
A use case on Beer Style Classification

- **400 instances**
- **8 classes**
  - Blanche, Lager, Pilsner, IPA, Stout, Barleywine, Porter, Belgian Strong Ale
- **3 attributes**
  - **Color** [0, 45]: Pale, Straw, Amber, Brown, Black
  - **Bitterness** [8, 250]: Low, Low-medium, Medium-high, High
  - **Strength** [0.039, 0.136]: Session, Standard, High, Very high
- **Decision Trees**
  - **Crisp DT**: J48, REPTree, RandomTree (WEKA)
  - **FRBC**: FURIA (WEKA), HILK (GUAJE)
- **Linguistic Approximation**
  - Global Semantics via Strong Fuzzy Partitions
  - Linguistic IF-THEN Mamdani rules

Jose M. Alonso
[https://citius.usc.es/v/jose-maria-alonso-moral](https://citius.usc.es/v/jose-maria-alonso-moral)
# Counterfactuals

A use case on Beer Style Classification

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy (%)</th>
<th>Number of Rules</th>
<th>Total Rule Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest</td>
<td>96.25</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>J48</td>
<td>95.00</td>
<td>9.8</td>
<td>23.4</td>
</tr>
<tr>
<td>FURIA</td>
<td>95.50</td>
<td>14.2</td>
<td>44.0</td>
</tr>
<tr>
<td>HILK</td>
<td>93.63</td>
<td>14.4</td>
<td>32.2</td>
</tr>
</tbody>
</table>
Output class is Belgian Strong Ale because color is black and strength is high. However, output class would be Stout if strength were standard.
Counterfactuals
A use case on Beer Style Classification

(Bitterness in [-inf, -inf, 10, 30]) and (Color in [-inf, -inf, 4, 7]) => Beer_Style=1.0 (CF = 0.97)
(Bitterness in [-inf, -inf, 2, 9]) and (Color in [-inf, 4, 7]) => Beer_Style=2.0 (CF = 0.95)
(Bitterness in [-inf, -inf, 6, 8]) and (Color in [-inf, -inf, 14, 16]) => Beer_Style=2.0 (CF = 0.97)
(Bitterness in [-inf, -inf, 6, 8]) and (Strength in [-inf, -inf, 0.92]) => Beer_Style=4.0 (CF = 0.89)
(Bitterness in [48, 50, inf, inf]) and (Color in [-inf, -inf, 8, 10]) => Beer_Style=4.0 (CF = 0.84)
(Bitterness in [30, 31, inf, inf]) => Beer_Style=5.0 (CF = 0.96)
(Bitterness in [-inf, -inf, 92, 39]) and (Strength in [0.96, 0.053, inf, inf]) => Beer_Style=5.0 (CF = 0.94)
(Strength in [0.98, 0.039, inf, inf]) and (Bitterness in [47, 50, inf, inf]) and (Color in [11, 12, inf, inf]) => Beer_Style=6.0 (CF = 0.96)
(Bitterness in [93, 95, inf, inf]) and (Strength in [0.95, 0.032, inf, inf]) => Beer_Style=6.0 (CF = 0.98)
(Bitterness in [31, 37, inf, inf]) and (Strength in [0.96, 0.097, inf, inf]) => Beer_Style=6.0 (CF = 0.95)
(Bitterness in [32, 39, inf, inf]) and (Color in [12, 13, 4, inf, inf]) and (Bitterness in [-inf, -inf, 85, 98]) => Beer_Style=6.0 (CF = 0.9)
(Color in [-inf, -inf, 30, 34]) and (Strength in [-inf, -inf, 0.051, 0.051]) => Beer_Style=7.0 (CF = 0.96)
(Color in [17, 20, inf, inf]) and (Strength in [-inf, -inf, 10, 30]) => Beer_Style=7.0 (CF = 0.95)
(Bitterness in [0.97, 0.067, inf, inf]) and (Bitterness in [-inf, -inf, 47, 50]) => Beer_Style=8.0 (CF = 0.97)
Counterfactuals
A use case on Beer Style Classification

\[ x = \langle \text{Color, 12.0}; \text{Bitterness, 104.0}; \text{Strength, 0.099} : \text{Beer, Barleywine} \rangle \]

<table>
<thead>
<tr>
<th>Rule number</th>
<th>Activation score</th>
<th>Predicted class</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.156</td>
<td>Pilsner</td>
</tr>
<tr>
<td>4</td>
<td>0.156</td>
<td>IPA</td>
</tr>
<tr>
<td>5</td>
<td>0.409</td>
<td>IPA</td>
</tr>
<tr>
<td>8</td>
<td>0.589</td>
<td>Barleywine</td>
</tr>
<tr>
<td>12</td>
<td>0.411</td>
<td>Belgian Strong Ale</td>
</tr>
</tbody>
</table>

\[
ed_f(x, s) = \{(Bitterness, High); (Strength, Very high) : (Beer, Barleywine)\} \]
\[
ed_{cf}(x, s, IPA) = \{(Bitterness, High); (Strength, High) : (Beer, IPA)\} \]
\[
E(x, s) = ed_f \cup ed_{cf} = \{(Bitterness, High); (Strength, Very high) : (Beer, Barleywine); (Bitterness, High); (Strength, High) : (Beer, IPA)\} \]

Output class is Barleywine because bitterness is high and strength is very high. However, output class would be IPA if strength were high.
XAI for e-Health
ADHERE-U Project (RTI2018-099646-B-I00)

- Models, techniques and methodologies based on AI for improving Medication Adherence
  - Data Exploitation Services: Symptom Patterns Discovery, Emotion Recognition
  - Persuasive Interaction: Conversational Interface with Interactive Explanations and Adaptive Communication Strategy
    - CalendulaBot: https://gitlab.citius.usc.es/jose.guerra.vilar/calendulabot
  - Evaluation: Clinical Trials with Polymedicated Patients; Patients with Lung Cancer; and Patients with Atrial Fibrillation

Jose M. Alonso
https://citius.usc.es/v/jose-maria-alonso-moral
XAI for e-Health: Diabetes Diagnosis

SEMANTICALLY EXTENDED HIERARCHICAL FRAMEWORK

Input parameters
- Glucose lab tests results ($I^g - I^f$)
- Kidney lab tests results ($I^k - I^f$)
- Liver lab tests results ($I^l - I^f$)
- Lipid profile results ($I^p - I^f$)

Layer 1
- FRBS (1) Glucose level
- FRBS (2) Kidney function
- FRBS (3) Liver function
- FRBS (4) Lipid profile
- FRBS (50) Sympt. + Comp.

Prepared dataset
- Data extraction and preprocessing
- Data-driven fuzzy partitions
- Encoding + semantic inference
- Outliers & missing values, Feature selection, # of labels per variable (Weka APIs)

Fuzzy rules induction
- Data-driven rule
- Coverage fuzzy rules

FRBS simplification
- Inference engine, t-norm, s-norm, de-fuzzification

FRBS configuration
- JFML Java APIs, GUAJE APIs
- Jena Java APIs, SML APIs

FRBS implementation and evaluation
- DDO ontology, SNOMED CT

Step 1
- Distributed EHR systems

Step 2
- Linguisitic variables and fuzzy partitions generation
- Strong fuzzy partitions

Step 3
- Selected linguistic variables and fuzzy partitions
- Expert fuzzy partitions

Step 4
- Reduced fuzzy partitions

Step 5
- Knowledge sources (e.g., medical experts)
- Experts’ defined fuzzy rules

DDO ontology

Patient profile
(The history and current conditions)


Jose M. Alonso
https://citius.usc.es/v/jose-maria-alonso-moral
A Multilayer Multimodal Detection and Prediction Model based on Explainable Artificial Intelligence for Alzheimer’s Disease (under review)
A Multilayer Multimodal Detection and Prediction Model based on Explainable Artificial Intelligence for Alzheimer’s Disease (under review)

- 1048 subjects from ADNI dataset (http://adni.loni.usc.edu/):
  - Cognitively normal (CN): 294 subjects
  - sMCI: 254 subjects, pMCI: 232 subjects, and AD: 268 subjects
- 11 modalities: PET, MRI, Cognitive Scores, Demographic, etc.
- Two-layer Accurate and Interpretable AD diagnosis and progression detection model
  - First layer: diagnosis AD patients
    (accuracy: 93.95% and F1-score: 93.94%)
  - Second layer: MCI-to-AD progression within three years from baseline diagnosis
    (accuracy: 87.08% and F1-Score: 87.09%)
- Global and instance-based explanations using SHAP
- 22 NL explainers based on decision trees and fuzzy rule-based systems
References


- Jose M. Alonso, Javier Toja-Alamancos, Alberto Bugarin, “Experimental Study on Generating Multi-modal Explanations of Black-box Classifiers in terms of Gray-box Classifiers”, IEEE World Congress on Computational Intelligence (IEEE-WCCI), Glasgow, Scotland, 2020, https://dx.doi.org/10.1109/FUZZ48607.2020.9177770


The mission of this Task Force is to lead the development of a new generation of Explainable Fuzzy Systems, with a holistic view of fundamentals and current research trends in the XAI field, paying special attention to fuzzy-grounded knowledge representation and reasoning but also regarding how to enhance human-machine interaction through multi-modal (e.g., graphical or textual modalities) effective explanations.

Given the multidisciplinary nature of the XAI research field, the scope of this task force goes beyond the usual topics treated by the fuzzy community. The activities to be developed will be of interest for researchers, from both academy and industry, working in the fields of Artificial and Computational Intelligence (with special attention to Fuzzy Logic but addressing also XAI challenges on Neural Networks, Evolutionary Computation, Bayesian Networks, Bio-inspired algorithms, etc.).

https://sites.google.com/view/tf-explainable-fuzzy-systems/
Special Issue

TF-EXFS Explainable and Trustworthy Artificial Intelligence

Journal: IEEE Computational Intelligence Magazine

Editors: Jose M. Alonso, Corrado Mencar, Hisao Ishibuchi

- Paper submission: February 15th, 2021
- Acceptance/rejection notification: April 15th, 2021
- Revision due: May 15th, 2021
- Final notification: July 1st, 2021
- Camera-ready due: July 15th, 2021
- Publication Date: November 2021

https://sites.google.com/view/special-issue-on-xai-ieee-cim
NL4XAI: Interactive Natural Language Technology for XAI

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 860621.

https://citius.usc.es/v/jose-maria-alonso-moral
"La fantasia è più importante del sapere"

"The important thing is not to stop questioning"
Collaborators

Alberto Bugarín
CiTIUS-USC, Spain

Jesús Alcala-Fdez
University of Granada, Spain

J. M. Soto Hidalgo
Technical University of Madrid, Spain

Luis Magdalena
University of Granada, Spain

Corrado Mencar
University of Bari Aldo Moro, Italy

Ciro Castiello
University of Bari Aldo Moro, Italy

Ilia Stepin
University of Granada, Spain

J. M. Soto Hidalgo
University of Granada, Spain

Shaker El-Sappagh
CiTIUS-USC, Spain

Ciro Castiello
University of Bari Aldo Moro, Italy

Jesús Alcala-Fdez
University of Granada, Spain

Martin Pereira
USC, Spain

Ilia Stepin
CiTIUS-USC, Spain

J. M. Soto Hidalgo
University of Granada, Spain

Shaker El-Sappagh
CiTIUS-USC, Spain

Ciro Castiello
University of Bari Aldo Moro, Italy

Jesús Alcala-Fdez
University of Granada, Spain

Martin Pereira
USC, Spain

Ilia Stepin
CiTIUS-USC, Spain

J. M. Soto Hidalgo
University of Granada, Spain

Shaker El-Sappagh
CiTIUS-USC, Spain

Ciro Castiello
University of Bari Aldo Moro, Italy

Jesús Alcala-Fdez
University of Granada, Spain
“If there is effort, there is always accomplishment”

“Never be proud of having won an opponent. Who you won today can beat you tomorrow”

“The only victory that endures is the conquest over the own ignorance”
Interactive Natural Language Technology for Human-Centric Explainable Artificial Intelligence

Jose M. Alonso

Centro Singular de Investigación en Tecnoloxías Intelixentes
Universidade de Santiago de Compostela