Semantic Web-based Interoperability for Intelligent Agents with PSyKE

Federico Sabbatini Giovanni Ciatto Andrea Omicini
{*f.sabbatini*, giovanni.ciatto, andrea.omicini}@unibo.it

Dipartimento di Informatica – Scienza e Ingegneria (DISI)
Alma Mater Studiorum – Università di Bologna

4th International Workshop on EXplainable and TRAnsparent AI and Multi-Agent Systems (EXTRAAMAS)
May 9, 2022, Auckland, NZ (fully online)

Next in Line...

Context I

Need of cooperation between heterogeneous agents

- distributed systems are composed of agents
  - they may have different nature
- necessity to agree on a shared, formal semantics
  - for the data they exchange and operate upon

Solution

→ the Semantic Web (SW) offers tools to encode semantic concepts

i.e. ontologies, where data are represented as knowledge graphs

Context II

Need to understand the agents' behaviour

- intelligent agents are based on ML models
  - their nature is opaque, i.e., they behave as black boxes (BB)
  - they are unintelligible from the human perspective
- necessity to inspect their internal functioning

  e.g. in critical applications where interpretability is mandatory

Solution

→ the XAI community suggests various methods

i.e. apply symbolic knowledge extraction (SKE) techniques to the BB

Motivation

Need to have an extraction framework with SW features

- allow ML predictors to be trained upon knowledge graphs
  - instead of only tensors
  - ! expand the horizon of the possible applications
- allow SKE techniques to extract semantic agent-interpretable rules
  - and not only human-readable logic rules
  - ! augment the interoperability between agents

Solution

→ enhance the PSyKE framework with SW features

Some state of the art I

Symbolic knowledge extraction

- ML-based predictors detect patterns and relationships buried in data
- the acquired knowledge is:
  - reused in similar applications, but
  - sub-symbolically represented—i.e., stored as internal parameters
  - not suitable for human comprehension (BB behaviour [Lipton, 2018])

  → thus unreliable in critical applications

  e.g. involving human health, wealth, freedom, etc.

→ need to rely on more interpretable predictors [Rudin, 2019]

→ need to adopt inspection techniques [Guidotti et al., 2018]

Some state of the art II

The PSyKE framework [Sabbatini et al., 2021]

- PSyKE is a general purpose framework
- it provides a unified API interface for SKE algorithms
- it supports several extraction techniques for different kinds of BB
  - currently, 5 pedagogical extractors
  - applicable to both BB classifiers and regressors
- until now, it only extracted first-order logic rules

  i.e. according to the Prolog syntax and semantics
- its main implementation is in Python language
  - relying on the Scikit-learn library

Some state of the art III

The Semantic Web [Berners-Lee et al., 2001]

- considered since its birth as a tool for interoperability
  - between humans and software agents
  - between software agents
- provides methods to formalise data
  - together with their implicit semantics
- provides inference rules useful to reason with the data
- enabling technologies:
  - → RDF [Manola et al., 2004] to represent objects and relationships
  - → OWL [Hitzler et al., 2012] to extend RDF with FOL expressiveness
  - → SWRL [Horrocks et al., 2004] to represent inductive rules on SW entities
  - → automated reasoners to perform reasoning
    - e.g. HermiT [Glimm et al., 2014, Motik et al., 2009, Shearer et al., 2008]
    - e.g. Pellet [Sirin and Parsia, 2004, Sirin et al., 2007]

Contribution of the paper

Making PSyKE SW-compatible to

- train ML predictors upon knowledge graphs
  - related to an ontology
- extract SWRL rules from BB predictors
  - possibly adhering to an ontology

As a result

→ ontologies and KG become both input and output of the SKE phase

→ implementation of new framework features to perform

- propositionalisation of KG into tabular data
- relationalisation of tabular data into KG
- conversion of output rules into SWRL format

→ PSyKE as a tool for agent-interoperability

- other than human-interpretability

Next in Line...

PSyKE design

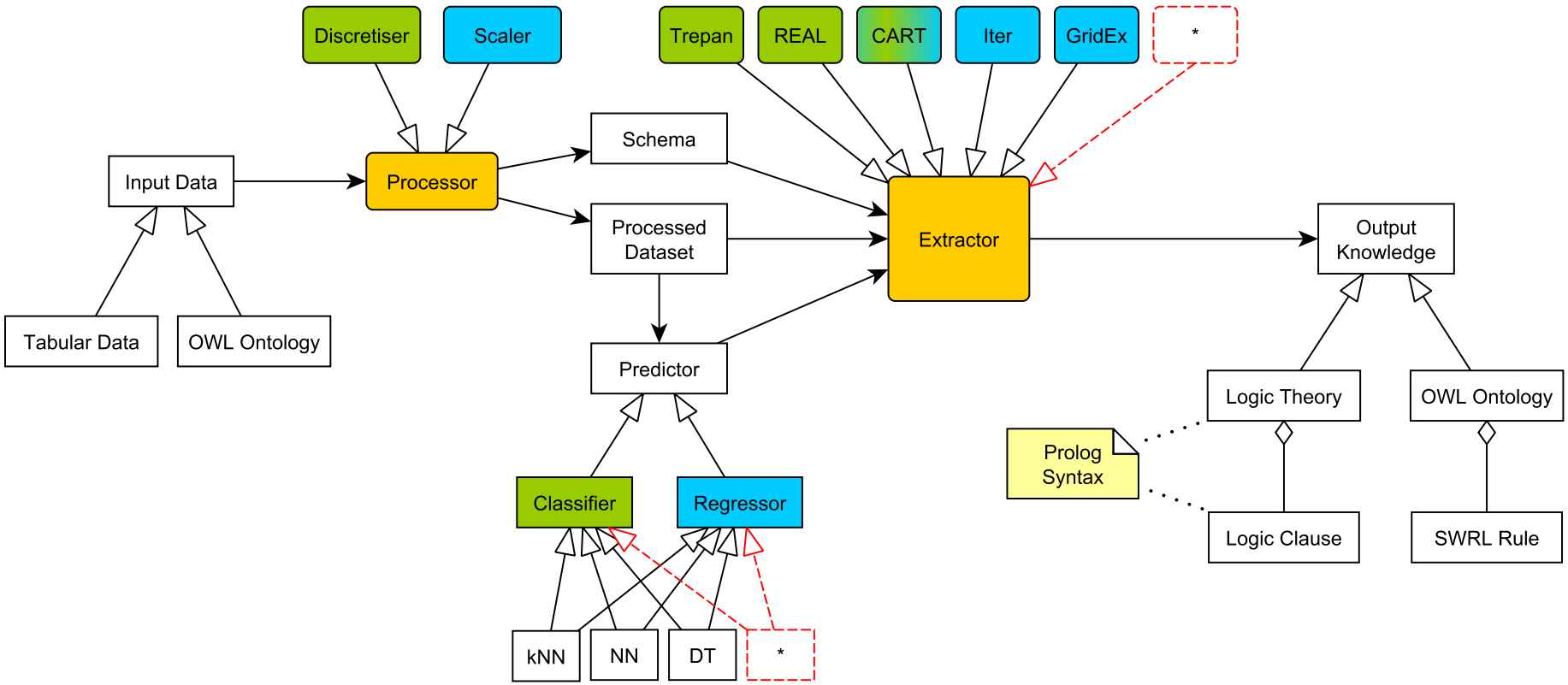
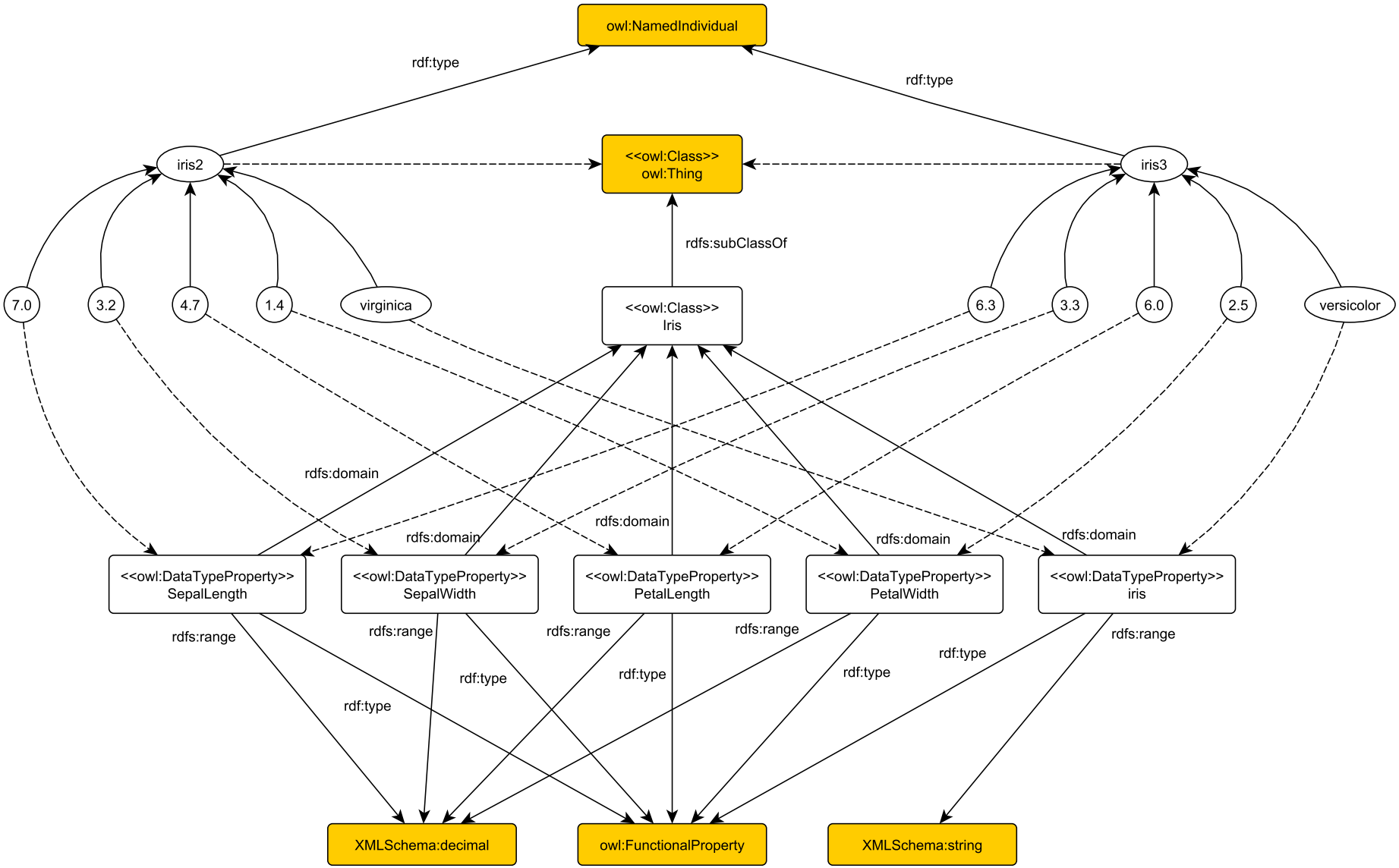


Figure: The new design of PSyKE.

Relationalisation



Output SWRL rules I

SWRL rules

- more agent-interpretable than Prolog rules
- structured as logical implications
  - a list of preconditions imply a postcondition
  - the postcondition is true if all the preconditions are satisfied
- conditions expressed as triples

  i.e. subject, predicate, object

  e.g. data set instance, "has-a" relationship, property

  e.g. property, relational operator, constant value
- provide predictive capabilities
  - if associated to an automated reasoner

Output SWRL rules II

Example of SWRL rule

Mono-dimensional classification task on a data set having m input features

$$\begin{aligned} &\texttt{Object(?o)},\\ &\texttt{Prop\_1(?o, ?p1)}, \ldots, \texttt{Prop\_m(?o, ?pm)},\\ &\quad \texttt{Cond(?p1, c11)}, \ldots, \texttt{Cond(?p1, c1j)},\\ &\qquad\qquad\qquad\qquad\qquad \ldots,\\ &\quad \texttt{Cond(?pm, cm1)}, \ldots, \texttt{Cond(?pm, cmk)} \quad \Longrightarrow \texttt{Output(?o, out)} \end{aligned}$$

Next in Line...

Case study

- We tested PSyKE in a simple scenario (Iris data set)
  - classification task
  - 4 continuous input features
  - 1 output class
  - 150 instances split into train and test sets (50% + 50%)
- We then relationalised it to obtain the equivalent KG
- We trained a k-NN black-box classifier
- We applied the CART [Breiman et al., 1984] extractor
  - to obtain a tree representation of the BB behaviour
  - and then a set of SWRL rules from the extracted tree
- Finally, we merged the extracted rules to the relationalised KG

Results

Example of SWRL rules for the Iris data set

```
1   Iris(?iris), SepalLength(?iris, ?sepalLength),
2   SepalWidth(?iris, ?sepalWidth), PetalLegth(?iris, ?petalLength),
3   PetalWidth(?iris, ?petalWidth), lessThanOrEqual(?petalLength, 2.75)
4   -> iris(?iris, "setosa")
5
6   Iris(?iris), SepalLength(?iris, ?sepalLength),
7   SepalWidth(?iris, ?sepalWidth), PetalLegth(?iris, ?petalLength),
8   PetalWidth(?iris, ?petalWidth), greaterThan(?petalLength, 2.75),
9   greaterThan(?petalWidth, 1.68) -> iris(?iris, "virginica")
10
11  Iris(?iris), SepalLength(?iris, ?sepalLength),
12  SepalWidth(?iris, ?sepalWidth), PetalLegth(?iris, ?petalLength),
13  PetalWidth(?iris, ?petalWidth), greaterThan(?petalLength, 2.75),
14  lessThanOrEqual(?petalWidth, 1.68) -> iris(?iris, "versicolor")
```

Pros & cons

Pros

- Direct interoperability between heterogeneous intelligent agents
- Draw predictions from ontologies and KG via inference mechanisms
- Consistency check
  - contradictions between extracted SWRL rules
  - contradictions between individuals and (all) rules

Cons

- SKE algorithms extracting overlapping rules
  - an individual may match rules having different postconditions
- Merge the extracted rules into an ontology
  - inconsistencies deriving from BB/SWRL rules misclassifications
- SWRL rules' reduced expressiveness w.r.t. Prolog
  - inability to have linear combinations of input features as postconditions

Next in Line...

Conclusions

The new PSyKE platform can be exploited to:

- combine Semantic Web tools with SKE techniques
- apply ML and SKE techniques on KG and ontologies
  - rather than tabular inputs
- obtain output rules in agent-interpretable format
  i.e. SWRL rules
- include output SWRL in ontologies
  e.g. the input ontology
- relationalise tabular data into KG
- propositionalise KG into tabular data

→ start and terminate the SKE in the SW domain

→ promote intelligent agent interoperability

Future works

Our next research efforts will be focused on

- extending the Semantic Web functionalities to regression tasks
  - i.e. linear combinations of input features as output rules' postconditions
- resolving consistency issues
  - e.g. arising from overlapping output rules
  - e.g. caused by conflicts between data set and BB misclassifications

Semantic Web-based Interoperability for Intelligent Agents with PSyKE

Federico Sabbatini Giovanni Ciatto Andrea Omicini
{*f.sabbatini*, giovanni.ciatto, andrea.omicini}@unibo.it

Dipartimento di Informatica – Scienza e Ingegneria (DISI)
Alma Mater Studiorum – Università di Bologna

4th International Workshop on EXplainable and TRAnsparent AI and Multi-Agent Systems (EXTRAAMAS)
May 9, 2022, Auckland, NZ (fully online)

References I

[Berners-Lee et al., 2001] Berners-Lee, T., Hendler, J., and Lassila, O. (2001).
The semantic web.
Scientific american, 284(5):34–43.

[Breiman et al., 1984] Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984).
Classification and Regression Trees.
CRC Press.

[Glimm et al., 2014] Glimm, B., Horrocks, I., Motik, B., Stoilos, G., and Wang, Z. (2014).
HermiT: An OWL 2 reasoner.
Journal of Automated Reasoning, 53(3):245–269
DOI:10.1007/s10817-014-9305-1.

[Guidotti et al., 2018] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. (2018).
A survey of methods for explaining black box models.
ACM Computing Surveys, 51(5):1–42
DOI:10.1145/3236009.

References II

[Hitzler et al., 2012] Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P. F., and Rudolph, S. (2012).
OWL 2 Web Ontology Language Primer (Second Edition).
W3C Recommendation 11 December 2012
https://www.w3.org/TR/owl2-primer.

[Horrocks et al., 2004] Horrocks, I., Patel-Schneider, P. F., Boley, H., Tabet, S., Grosof, B., and Dean, M. (2004).
SWRL: A Semantic Web Rule Language Combining OWL and RuleML.
W3C Member Submission 21 May 2004
https://www.w3.org/Submission/SWRL.

[Lipton, 2018] Lipton, Z. C. (2018).
The mythos of model interpretability.
Queue, 16(3):31–57
DOI:10.1145/3236386.3241340.

[Manola et al., 2004] Manola, F., Miller, E., and McBride, B. (2004).
Resource Description Framework (RDF) Primer.
W3C Recommendation 10 February 2004
https://www.w3.org/TR/rdf-primer.

References III

[Motik et al., 2009] Motik, B., Shearer, R. D. C., and Horrocks, I. (2009).
Hypertableau reasoning for description logics.
Journal of Artificial Intelligence Research, 36:165–228
DOI:10.1613/jair.2811.

[Rudin, 2019] Rudin, C. (2019).
Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead.
Nature Machine Intelligence, 1(5):206–215
DOI:10.1038/s42256-019-0048-x.

[Sabbatini et al., 2021] Sabbatini, F., Ciatto, G., Calegari, R., and Omicini, A. (2021).
On the design of PSyKE: A platform for symbolic knowledge extraction.
In Calegari, R., Ciatto, G., Denti, E., Omicini, A., and Sartor, G., editors, *WOA 2021 – 22nd Workshop "From Objects to Agents"*, volume 2963 of *CEUR Workshop Proceedings*, pages 29–48. Sun SITE Central Europe, RWTH Aachen University.
22nd Workshop "From Objects to Agents" (WOA 2021), Bologna, Italy, 1–3 September 2021. Proceedings
http://ceur-ws.org/Vol-2963/paper14.pdf.

References IV

[Shearer et al., 2008] Shearer, R. D. C., Motik, B., and Horrocks, I. (2008).
HermiT: A highly-efficient OWL reasoner.
In Dolbear, C., Ruttenberg, A., and Sattler, U., editors, *Proceedings of the Fifth OWLED Workshop on OWL: Experiences and Directions, collocated with the 7th International Semantic Web Conference (ISWC-2008), Karlsruhe, Germany, October 26-27, 2008*, volume 432 of *CEUR Workshop Proceedings*. CEUR-WS.org
http://ceur-ws.org/Vol-432/owled2008eu_submission_12.pdf.

[Sirin and Parsia, 2004] Sirin, E. and Parsia, B. (2004).
Pellet: An OWL DL reasoner.
In Haarslev, V. and Möller, R., editors, *Proceedings of the 2004 International Workshop on Description Logics (DL2004), Whistler, British Columbia, Canada, June 6-8, 2004*, volume 104 of *CEUR Workshop Proceedings*. CEUR-WS.org
http://ceur-ws.org/Vol-104/30Sirin-Parsia.pdf.

[Sirin et al., 2007] Sirin, E., Parsia, B., Cuenca Grau, B., Kalyanpur, A., and Katz, Y. (2007).
Pellet: A practical OWL-DL reasoner.
Journal of Web Semantics, 5(2):51–53
DOI:10.1016/j.websem.2007.03.004.