Bridging Symbolic and Sub-Symbolic AI: Towards Cooperative Transfer Learning in Multi-Agent Systems

Matteo Magnini* Giovanni Ciatto* Andrea Omicini*

*Dipartimento di Informatica – Scienza e Ingegneria (DISI)
Alma Mater Studiorum – Università di Bologna
{matteo.magnini, giovanni.ciatto, andrea.omicini}@unibo.it

21st International Conference of the
Italian Association for Artificial Intelligence (AIxIA 2022)
28th, November – 2nd December, 2022, Udine, IT

Next in Line...

1. Motivation & Context
2. Symbolic and Sub-Symbolic Tools
3. Cooperative (Transfer) Learning
4. Conclusions & future works

Motivation

The three ways of reasoning

- induction: a kind of reasoning that uses particular examples in order to reach a general conclusion about something
  $\rightarrow$ machine learning (e.g., neural networks);
- deduction: the act or process of using logic or reason to form a conclusion or opinion about something
  $\rightarrow$ symbolic artificial intelligence (e.g., logic programs);
- abduction: the forming and accepting on probation of a hypothesis to explain surprising facts
  $\rightarrow$ abductive logic programming.

Mimic the human society

- do not be bound for one single type of reasoning;
- cooperation! Knowledge sharing and explanation. [Omicini, 2020]

Context

Knowledge representation

- symbolic representation → formalism consisting of: [van Gelder, 1990]
  - set of symbols;
  - set of rules enabling possibly infinite combinations of symbols;
  - each (composed) symbol comes with its *meaning*;
  - human readable and interpretable;
  - intensional and extensional.
- sub-symbolic representation → the knowledge is represented via *numeric* data structures:
  - usually with fixed size;
  - a number does not have a particular meaning *per se*;
  - a number could be meaningful only considering its local context;
  - obscure, difficult to interpret.

Next in Line...

Symbolic Knowledge Extraction

Symbolic Knowledge Extraction (SKE)

[d'Avila Garcez et al., 2001, Hailesilassie, 2016, Zilke et al., 2016, Guidotti et al., 2018]

*any **algorithmic** procedure accepting **trained** **sub-symbolic predictors** as input and producing **symbolic knowledge** as output, in such a way that the extracted knowledge **reflects** the behaviour of the predictor with high fidelity.*

SKE advantages

- *post-hoc* explanation of sub-symbolic predictors;
- generates a symbolic representation of the predictor's behaviour, usually a logic formalism
  → much more concise w.r.t. the predictor, less space cost;
  → formal logic can be used as *lingua franca* in knowledge sharing.

Symbolic Knowledge Injection

Symbolic Knowledge Injection (SKI)

[Besold et al., 2017, Xie et al., 2019, Calegari et al., 2020]

any algorithmic procedure affecting how sub-symbolic predictors draw their inferences in such a way that predictions are either computed as a function of, or made consistent with, some given symbolic knowledge.

SKI advantages

- improve predictor's metrics (e.g., accuracy, f1-score, r2, etc.);
- reduce learning time;
- loosening the need of big datasets;
- prevent the predictor to become a black-box;
- independent from the source of the knowledge.

Transfer Learning

Transfer Learning (TL) [Pan and Yang, 2010]

is the set of **techniques** *aimed at letting a predictor P, targetting task T, take advantage from the knowledge acquired by some prior predictor P', trained on some other task T' (usually similar to T).*

TL properties

- similarly to SKI it improves the predictor's performance
  $\rightarrow$ inference ability (e.g., accuracy, f1-score, r2, ect.);
  $\rightarrow$ external metrics (e.g., training time, dataset size, etc.).
- constrained by the (sub-symbolic) knowledge of the first predictor
  $\rightarrow$ for instance, if we are doing TL from a NN to another, we have to replicate part of it – contiguous layers and weights – to the new one;
  $\rightarrow$ it is obscure for humans.

Multi-task Learning

Multi-task Learning (MTL) [Caruana, 1997, Zhang and Yang, 2022]

is a set of **mechanisms** *aimed to improve the performance of a predictor via TL. More precisely, given a set of* **similar** *tasks* $\{T_1, \ldots, T_m\}$*, MTL aims at learning the* m *tasks altogether, by training as many* ***predictors*** $P_1, \ldots, P_m$*. In doing so, MTL attempts to improve the performance of each* P_i*, by taking advantage of the* ***knowledge*** *while training the other predictors.*

MTL properties

- similarly to TL it improves the predictors' performance;
- unlike TL where there is one task that receives the knowledge from the other(s), all tasks *simultaneously* receive knowledge from the others
  $\rightarrow$ the simultaneous training has pro and cons;
  $\rightarrow$ still not interpretable.

Next in Line. . .

Cooperative Learning I

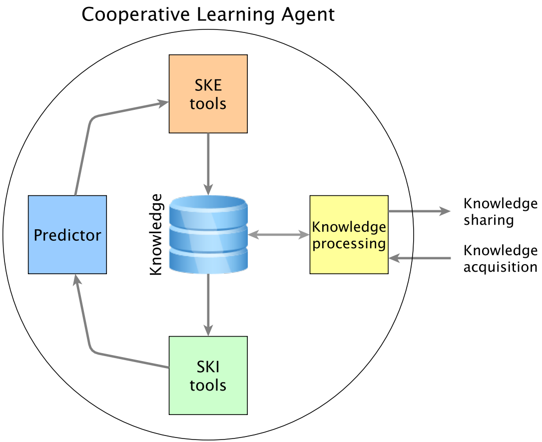
Cooperative Learning (CoL)

a CoL system consists of a multi-agent system (MAS) where agents can retrieve and exploit knowledge – especially symbolic – about a task from other agents and provide it to others when requested.

CoL characteristics

- agents should agree on common, shared symbolic representation means by which behavioural specifications could be described—and later exchanged;
- agents should come with symbolic (and sub-symbolic) tools for knowledge manipulation.

Cooperative Learning II



Cooperative Transfer Learning I

Cooperative Transfer Learning (CoTL)

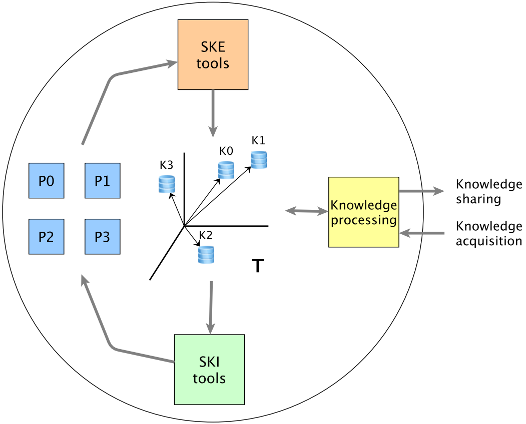
*a CoTL system consists of a multi-agent system (MAS) where **agents** can **retrieve** and **exploit** knowledge about several tasks from other **agents** and provide it to others when requested.*

CoTL characteristics

- unlike simple CoL systems, agents in CoTL systems may exploit knowledge (either their own, or other agents' one) about related tasks to learn novel tasks they were not originally designed for
  $\rightarrow$ find ways to exploit the knowledge K of the task T for a similar – yet not the same – task T' could be not trivial.

Cooperative Transfer Learning II

Cooperative Transfer Learning Agent



Next in Line. . .

Conclusions & (future) works I

Summing up

Cooperative Learning and Cooperative Transfer Learning are MAS where cooperation and sharing of (symbolic) knowledge are fundamental inter agents operations. Internally, an agent should be able to manipulate the knowledge in order to exploit it in solving tasks. Key aspects are:

- agreement upon a logic formalism as *lingua franca* for knowledge exchange;
- agents should come with algorithms able to manipulate knowledge (e.g., SKE, SKI, logic engine, etc.);
- how to exploit knowledge in scenarios of *heterogeneous* tasks.

Conclusions & (future) works II

Current works

- PSyKE & PSyKI [Sabbatini et al., 2021a, Magnini et al., 2022b]
- SKI algorithms [Magnini et al., 2022a, Magnini et al., 2022c]
- SKE algorithms [Sabbatini et al., 2021b, Sabbatini and Calegari, 2022]

Future works

- test the *train-extract-fix-inject* workflow;
- create and test CoL systems;
- investigate how to handle knowledge for *heterogeneous* tasks;
- finally, create and test CoTL systems.

Bridging Symbolic and Sub-Symbolic AI: Towards Cooperative Transfer Learning in Multi-Agent Systems

Matteo Magnini* Giovanni Ciatto* Andrea Omicini*

*Dipartimento di Informatica – Scienza e Ingegneria (DISI)
Alma Mater Studiorum – Università di Bologna
{matteo.magnini, giovanni.ciatto, andrea.omicini}@unibo.it

21st International Conference of the
Italian Association for Artificial Intelligence (AIxIA 2022)
28th, November – 2nd December, 2022, Udine, IT

References I

[Besold et al., 2017] Besold, T. R., d'Avila Garcez, A. S., Bader, S., Bowman, H., Domingos, P. M., Hitzler, P., Kühnberger, K., Lamb, L. C., Lowd, D., Lima, P. M. V., de Penning, L., Pinkas, G., Poon, H., and Zaverucha, G. (2017).
Neural-symbolic learning and reasoning: A survey and interpretation.
CoRR, abs/1711.03902
http://arxiv.org/abs/1711.03902.

[Calegari et al., 2020] Calegari, R., Ciatto, G., and Omicini, A. (2020).
On the integration of symbolic and sub-symbolic techniques for XAI: A survey.
Intelligenza Artificiale, 14(1):7–32
DOI:10.3233/IA-190036.

[Caruana, 1997] Caruana, R. (1997).
Multitask learning.
Mach. Learn., 28(1):41–75
DOI:10.1023/A:1007379606734.

References II

[d'Avila Garcez et al., 2001] d'Avila Garcez, A. S., Broda, K., and Gabbay, D. M. (2001).
Symbolic knowledge extraction from trained neural networks: A sound approach.
Artif. Intell., 125(1-2):155–207
DOI:10.1016/S0004-3702(00)00077-1.

[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.

[Hailesilassie, 2016] Hailesilassie, T. (2016).
Rule extraction algorithm for deep neural networks: A review.
CoRR, abs/1610.05267
http://arxiv.org/abs/1610.05267.

References III

[Magnini et al., 2022a] Magnini, M., Ciatto, G., and Omicini, A. (2022a).
KINS: Knowledge injection via network structuring.
In Calegari, R., Ciatto, G., and Omicini, A., editors, *CILC 2022 – Italian Conference on Computational Logic*, volume 3204 of *ceurws*, pages 254–267. CEUR-WS
http://ceur-ws.org/Vol-3204/paper_25.pdf.

[Magnini et al., 2022b] Magnini, M., Ciatto, G., and Omicini, A. (2022b).
On the design of PSyKI: a platform for symbolic knowledge injection into sub-symbolic predictors.
In Calvaresi, D., Najjar, A., Winikoff, M., and Främling, K., editors, *Proceedings of the 4th International Workshop on EXplainable and TRAnsparent AI and Multi-Agent Systems*, volume 13283 of *Lecture Notes in Computer Science*, chapter 6, pages 90–108. Springer
https://link.springer.com/chapter/10.1007/978-3-031-15565-9_6.

References IV

[Magnini et al., 2022c] Magnini, M., Ciatto, G., and Omicini, A. (2022c).
A view to a KILL: Knowledge injection via lambda layer.
In Ferrando, A. and Mascardi, V., editors, *WOA 2022 – 23rd Workshop "From Objects to Agents"*, volume 3261 of *CEUR Workshop Proceedings*, pages 61–76. Sun SITE Central Europe, RWTH Aachen University
http://ceur-ws.org/Vol-3261/paper5.pdf.

[Omicini, 2020] Omicini, A. (2020).
Not just for humans: Explanation for agent-to-agent communication.
In Vizzari, G., Palmonari, M., and Orlandini, A., editors, *Proceedings of the AIxIA 2020 Discussion Papers Workshop co-located with the the 19th International Conference of the Italian Association for Artificial Intelligence (AIxIA2020), Anywhere, November 27th, 2020*, volume 2776 of *CEUR Workshop Proceedings*, pages 1–11. CEUR-WS.org
http://ceur-ws.org/Vol-2776/paper-1.pdf.

References V

[Pan and Yang, 2010] Pan, S. J. and Yang, Q. (2010).
A survey on transfer learning.
IEEE Transactions on Knowledge and Data Engineering, 22(10):1345–1359
DOI:10.1109/TKDE.2009.191.

[Sabbatini and Calegari, 2022] Sabbatini, F. and Calegari, R. (2022).
Symbolic knowledge extraction from opaque machine learning predictors: GridREx & PEDRO.
In Kern-Isberner, G., Lakemeyer, G., and Meyer, T., editors, *19th International Conference on Principles of Knowledge Representation and Reasoning (KR 2022)*, pages 554–563, Haifa, Israel. IJCAI Organization
DOI:10.24963/kr.2022/57.

References VI

[Sabbatini et al., 2021a] Sabbatini, F., Ciatto, G., Calegari, R., and Omicini, A. (2021a).

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 VII

[Sabbatini et al., 2021b] Sabbatini, F., Ciatto, G., and Omicini, A. (2021b).
GridEx: An algorithm for knowledge extraction from black-box regressors.
In Calvaresi, D., Najjar, A., Winikoff, M., and Främling, K., editors, *Explainable and Transparent AI and Multi-Agent Systems. Third International Workshop, EXTRAAMAS 2021, Virtual Event, May 3–7, 2021, Revised Selected Papers*, volume 12688 of *Lecture Notes in Computer Science*, pages 18–38. Springer Nature, Cham, Switzerland

DOI:10.1007/978-3-030-82017-6_2.

[van Gelder, 1990] van Gelder, T. (1990).
Why distributed representation is inherently non-symbolic.
In Dorffner, G., editor, *Konnektionismus in Artificial Intelligence und Kognitionsforschung. Proceedings 6. Österreichische Artificial Intelligence-Tagung (KONNAI), Salzburg, Österreich, 18. bis 21. September 1990*, volume 252 of *Informatik-Fachberichte*, pages 58–66. Springer

DOI:10.1007/978-3-642-76070-9_6.

References VIII

[Xie et al., 2019] Xie, Y., Xu, Z., Meel, K. S., Kankanhalli, M. S., and Soh, H. (2019).
Embedding symbolic knowledge into deep networks.
In Wallach, H. M., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E. B., and Garnett, R., editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 4235–4245
https://proceedings.neurips.cc/paper/2019/hash/7b66b4fd401a271a1c7224027ce111bc-Abstract.html.

[Zhang and Yang, 2022] Zhang, Y. and Yang, Q. (2022).
A survey on multi-task learning.
IEEE Transactions on Knowledge and Data Engineering, 34(12):5586–5609
DOI:10.1109/TKDE.2021.3070203.

References IX

[Zilke et al., 2016] Zilke, J. R., Mencía, E. L., and Janssen, F. (2016).
DeepRED – Rule extraction from deep neural networks.
In Calders, T., Ceci, M., and Malerba, D., editors, *Discovery Science - 19th International Conference, DS 2016, Bari, Italy, October 19-21, 2016, Proceedings*, volume 9956 of *Lecture Notes in Computer Science*, pages 457–473
DOI:10.1007/978-3-319-46307-0_29.