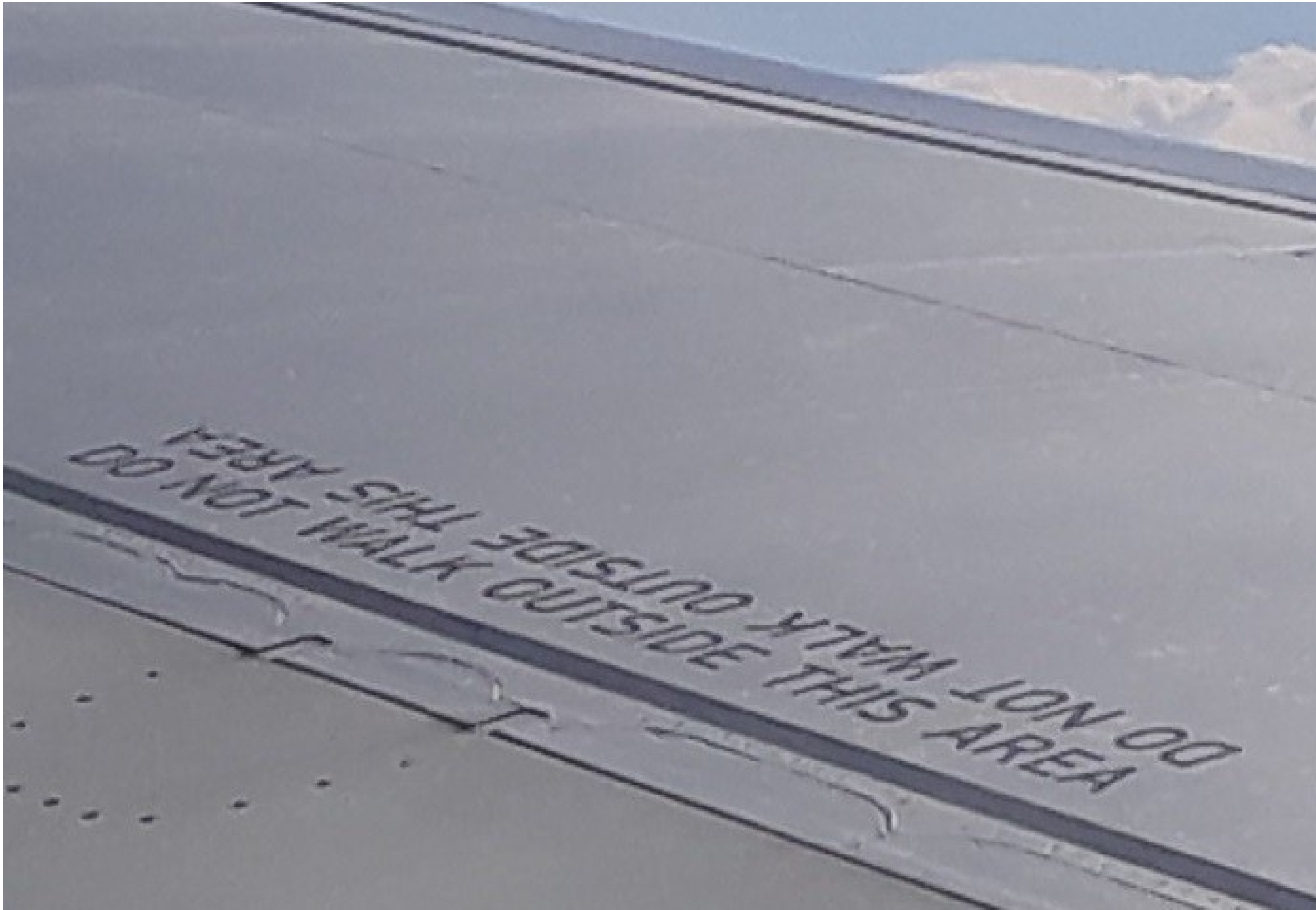
The Alan Turing Institute
1 March 2018

Neurosymbolic Computation: Thinking beyond Deep Learning

Artur d'Avila Garcez
City, University of London
a.garcez@city.ac.uk


DO NOT WALK OUTSIDE THIS AREA
DO NOT WALK OUTSIDE THIS AREA


Context is everything... commonsense!

The AI revolution...

US$40B investment in 2016 and growing, but AI adoption still low in 2017 (McKinsey)

The promise of AI:

- Education (active learning)
- Finance (time series prediction)
- Security (image and speech recognition)
- Health (IoT, companions, drug design)
- Telecom and Tech (infrastructure data analysis)
- Gaming (online learning)
- Transport (logistics, car industry)
- Manufacturing, Retail, Marketing, Energy, etc.

Brain/Mind dichotomy

Symbolic AI: a symbol system has all that is needed for general intelligence

Sub-symbolic AI: intelligence emerges from the brain (neural networks)

Neural-Symbolic Computing (NSC)

- Neural networks the best machinery for perception, learning and computation
- But perception alone is insufficient:
  Planning, Reasoning, Explanation, Transfer
- Rich knowledge representation: needed nonmonotonic, relational, variables, recursion, time, uncertainty.
- NSC = neural networks + logical structure
- **Compositionality is key**
- Within AI, NSC=ML+KR, if you accept that ML = deep learning!

Types of Machine Learning (ML)

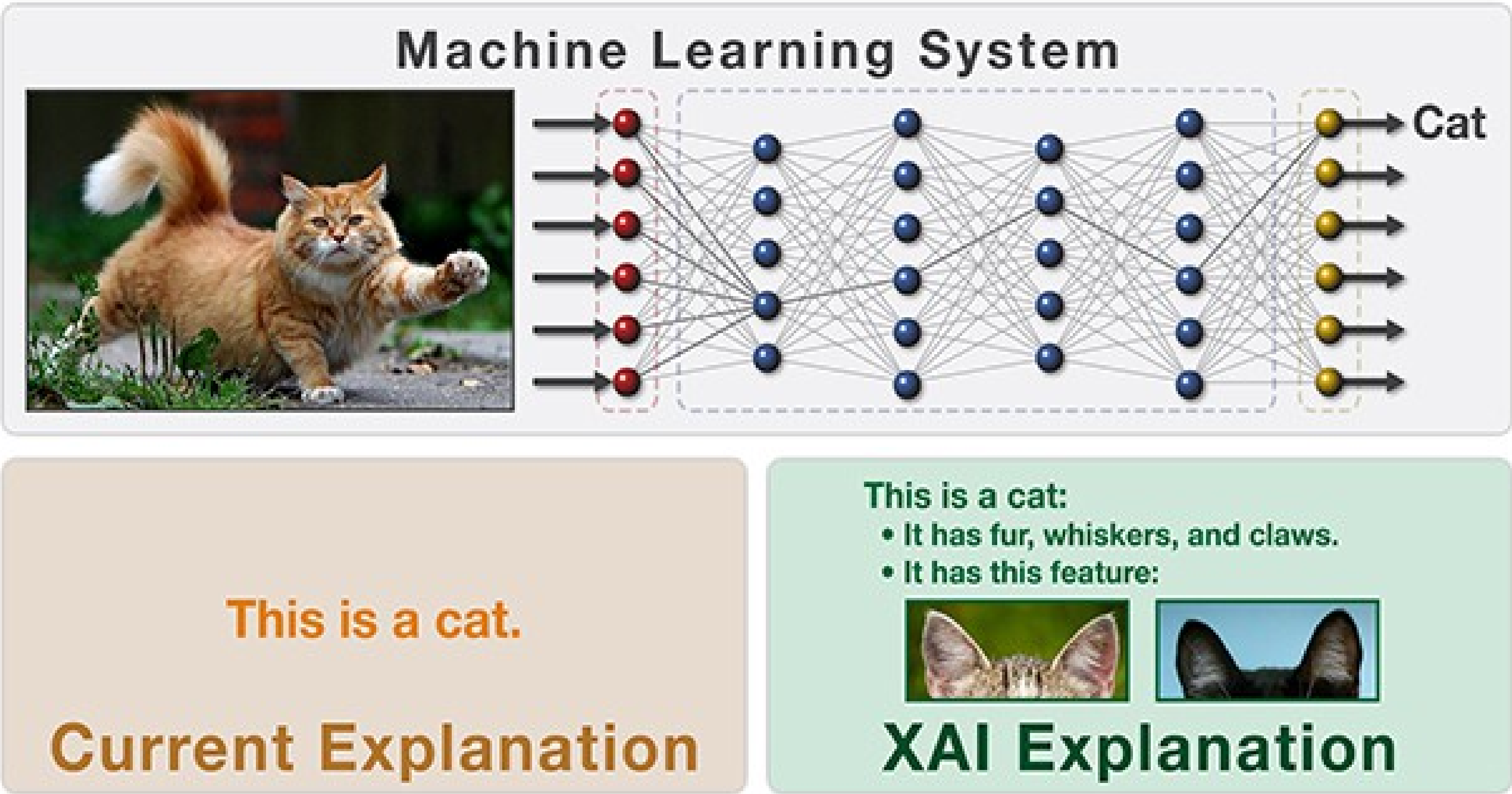
Systems that improve their own performance from experience

Systems that, in addition, enable humans to improve their own performance (human-machine interaction needed here); contrast this with AlphaGo and AlphaZero (explaining the model vs. one instance)

Explainable AI: accountability, trust and transfer learning... c.f. EU GDPR Reg. 71

Computer says no! is no longer acceptable

DARPA's Explainable AI



- XAI = Interpretable ML
- Explanation = proof history, not XAI

The need for knowledge extraction

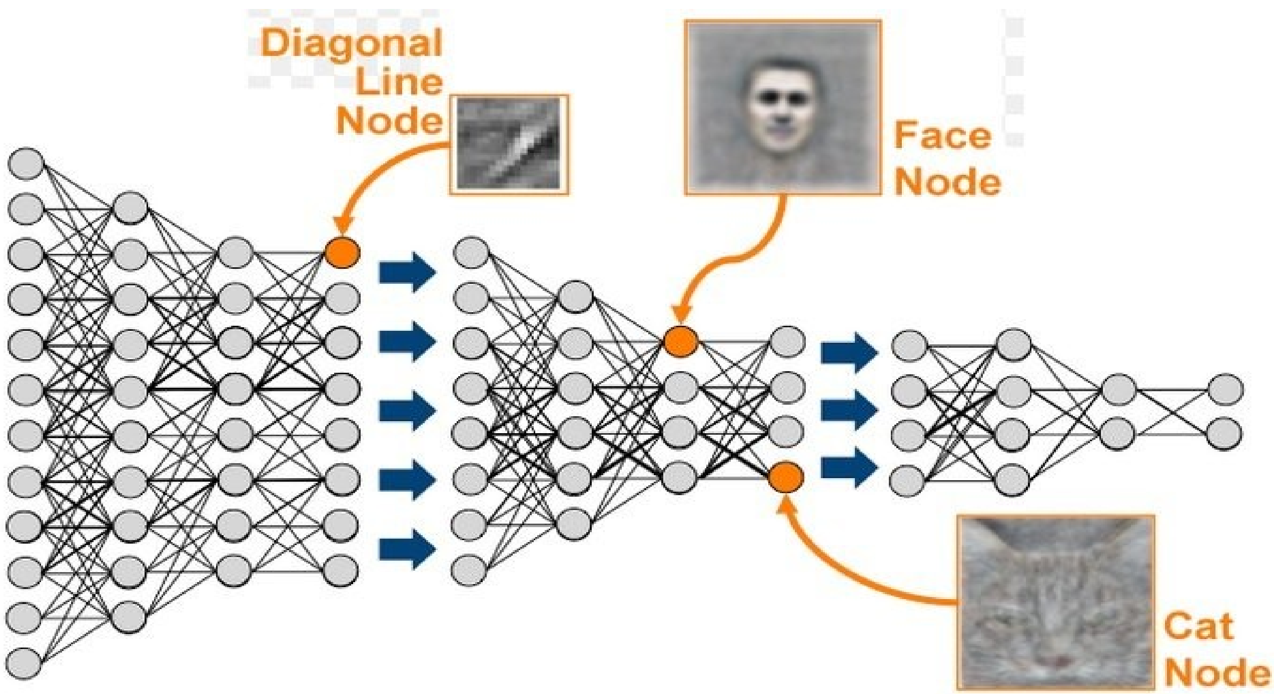
- Making sense of AlphaZero
  - Long-term positional sacrifices
  - Grand-masters now watching how it plays

- The need for Knowledge Extraction (http://www.staff.city.ac.uk/~aag/talks/GarcezDSI.pdf)
  - Correctness / soundness
  - Proof history (goal-directed reasoning)
  - Levels of abstraction (modularity)
  - Transfer learning (analogy)
  - System maintenance/improvement

Deep networks

Very nice original idea (deep Belief nets, semi-supervised learning) then turned/engineered into effective practical systems trained using backprop).

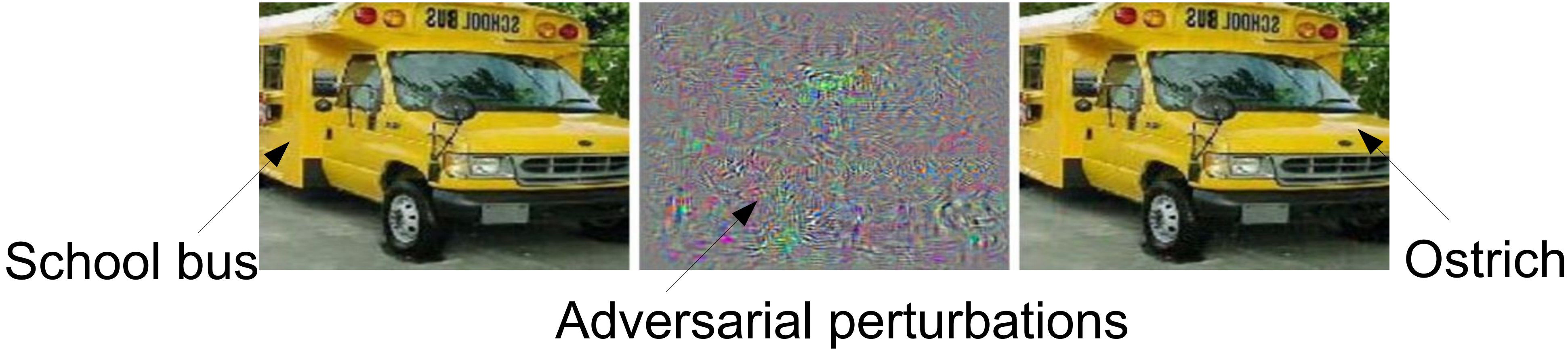
Very successful at object recognition, speech/audio and games, language translation and some video understanding



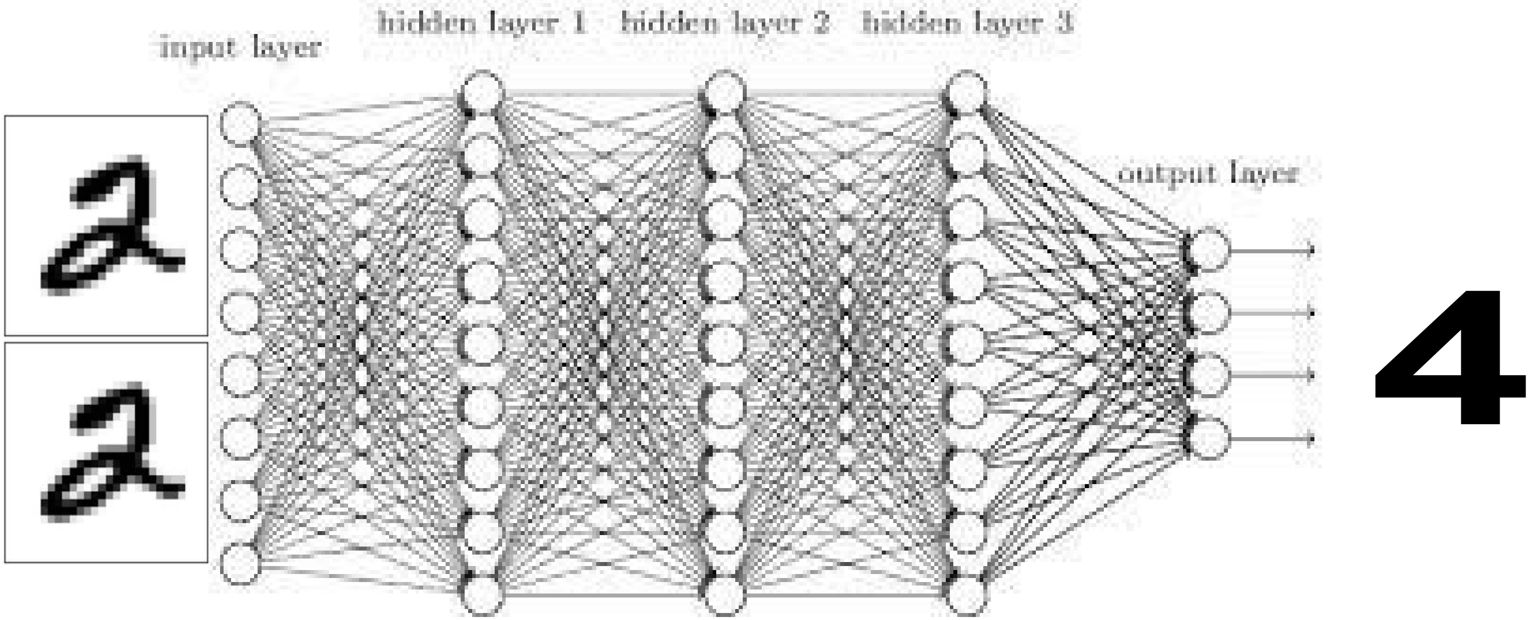
Geoff Hinton says AI needs to start over...

- Deep neural nets (CNNs) that do not recognise negative images?

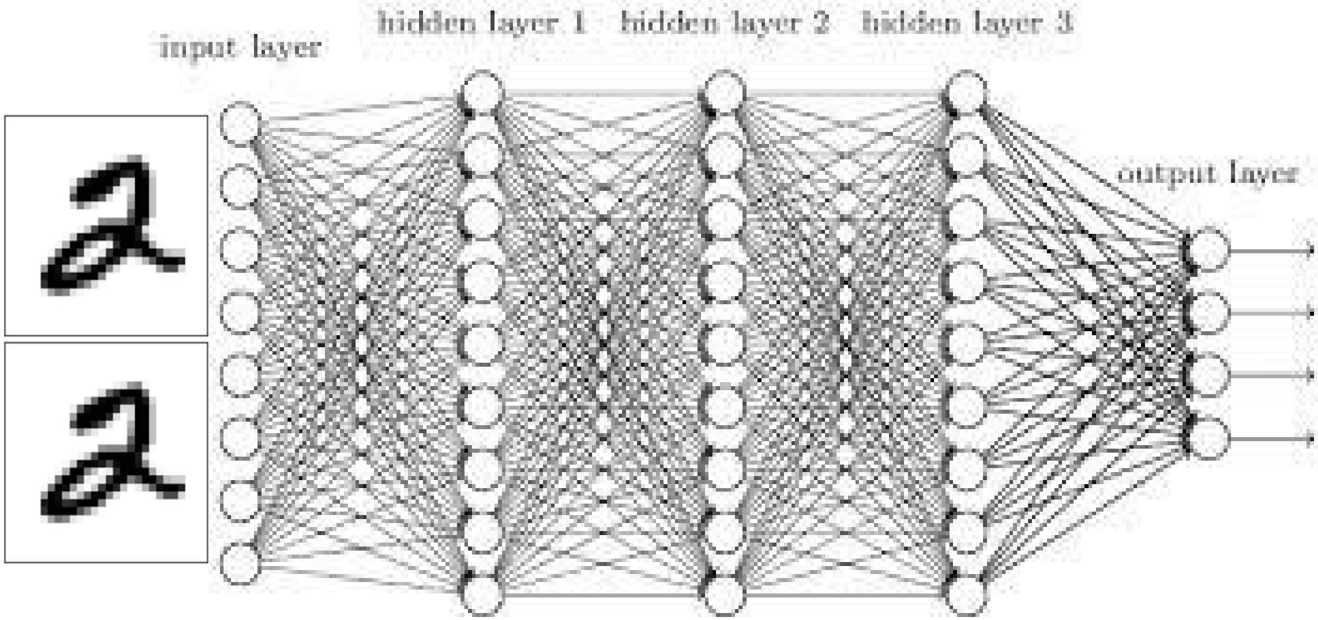
- Catastrophic failure with increase of input size (contrast with symbolic description of e.g. factorial)
- Adversarial examples that can fool the network



Knowledge Extraction from Deep Nets



Knowledge Extraction from Deep Nets



4

2+2=4

Why Neurons and Symbols?

“We need a language for describing the alternative algorithms that a network of neurons may be implementing” L. Valiant

(New) Logic + Neural Computation

GOAL: Learning from experience and reasoning about what has been learned in an uncertain environment in a computationally efficient way

A Compiler for Neural Nets

high-level symbolic representations
(abstraction, recursion, relations, modalities)

translations

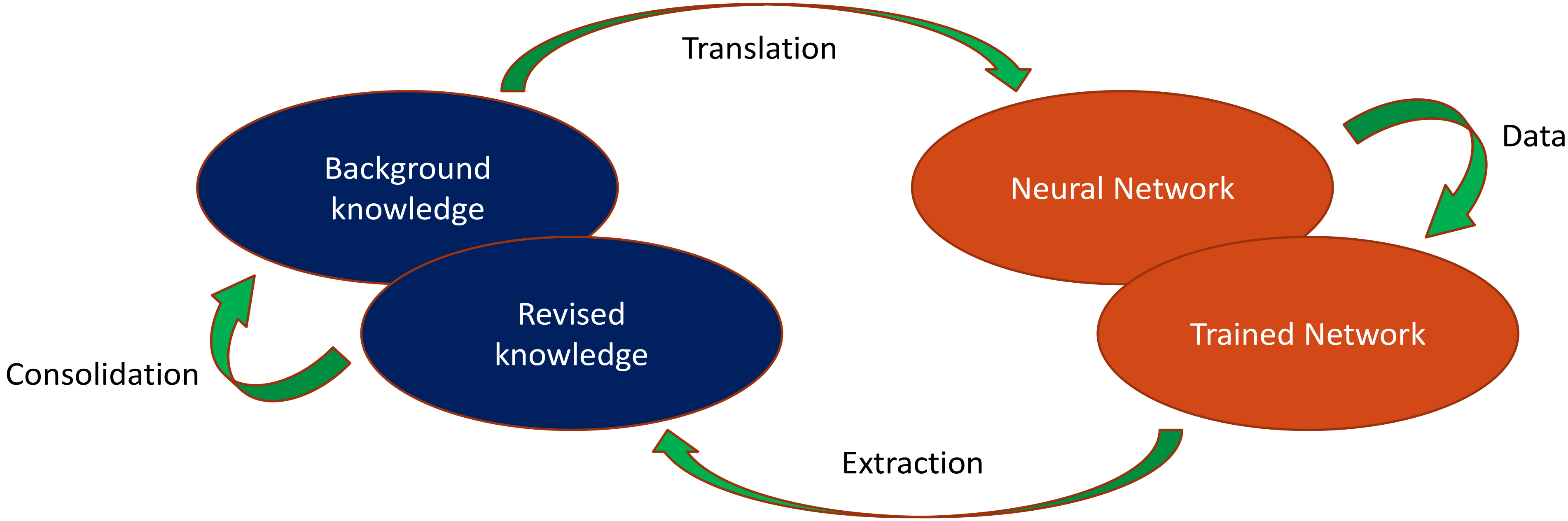
low level, efficient neural structures
(with the same, simple architecture throughout)

Analogy: low-level implementation (machine code) of high-level representations (e.g. java, requirement specs)

Neural-symbolic systems were applied to:

Training in simulators
Robotics (robocup)
Evolution of software models
Protein classification
Power systems fault diagnosis
Semantic web (ontology learning)
General game playing
Visual intelligence
Business compliance

Neural-Symbolic Learning Cycle



Connectionist Inductive Logic Programming (CILP System)

Neural Nets + Logic Programming (rules with exceptions)

Using Background Knowledge

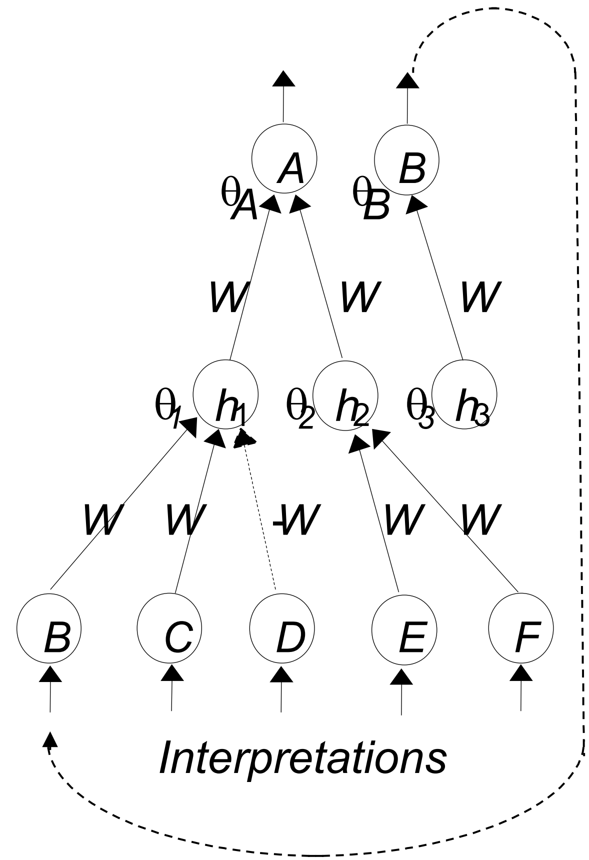
Learning with Backpropagation

Knowledge Extraction

r_1: $A \leftarrow B, C, \sim D;$

r_2 : $A \leftarrow E, F;$

r_3 : $B \leftarrow$

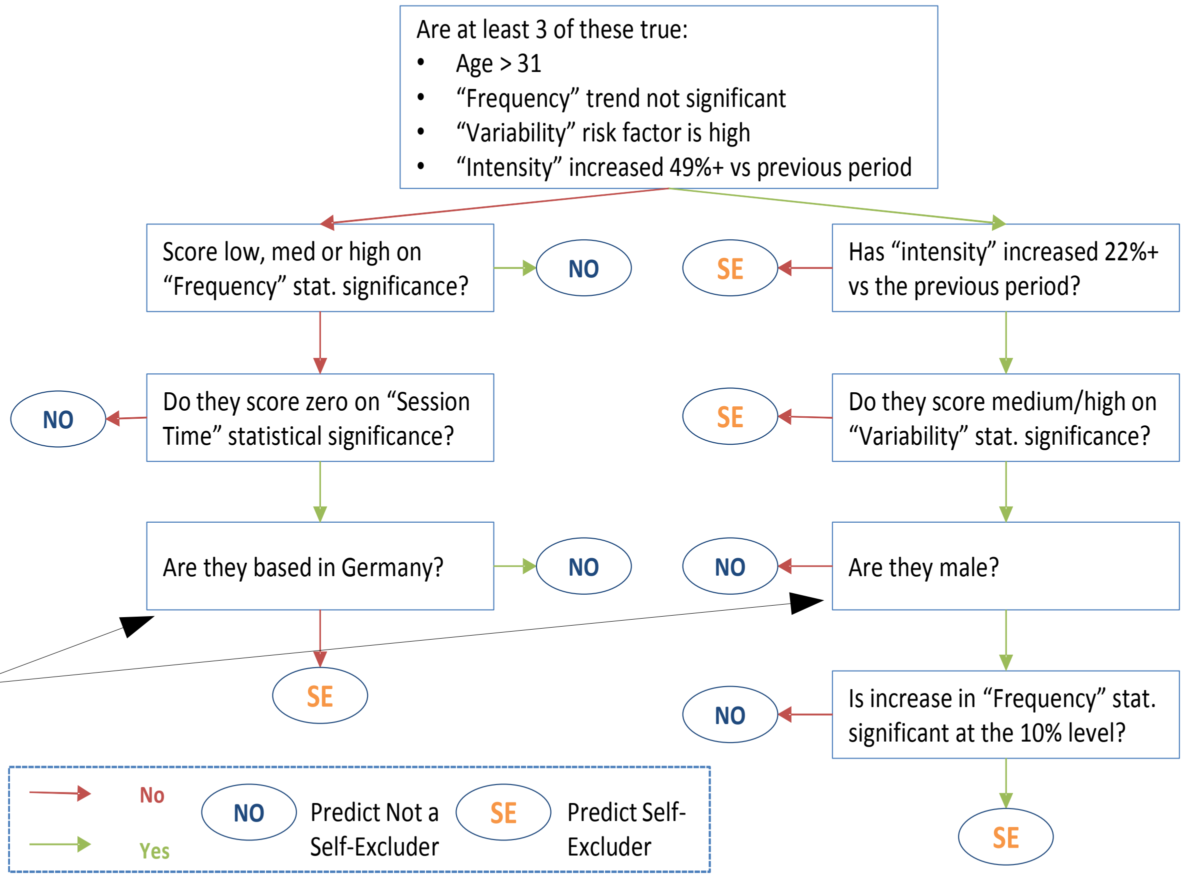


Rule Extraction: Neural Net = Black Box?

- Extracted rules can be visualized in the form of a state transition diagram

- Alternatively, use TREPAN-like rule extraction and variations...

- Decide on whether/how to intervene (improve the system; make it ethical)



C. Percy, A. S. d'Avila Garcez, S. Dragicevic, M. Franca, G. Slabaugh and T. Weyde. The Need for Knowledge Extraction: Understanding Harmful Gambling Behavior with Neural Networks, ECAI 2016, The Hague, September 2016.

Frosst and Hinton: Distilling a Neural Network Into a Soft Decision Tree, AI-IA CEX workshop, Bari, September 2017.

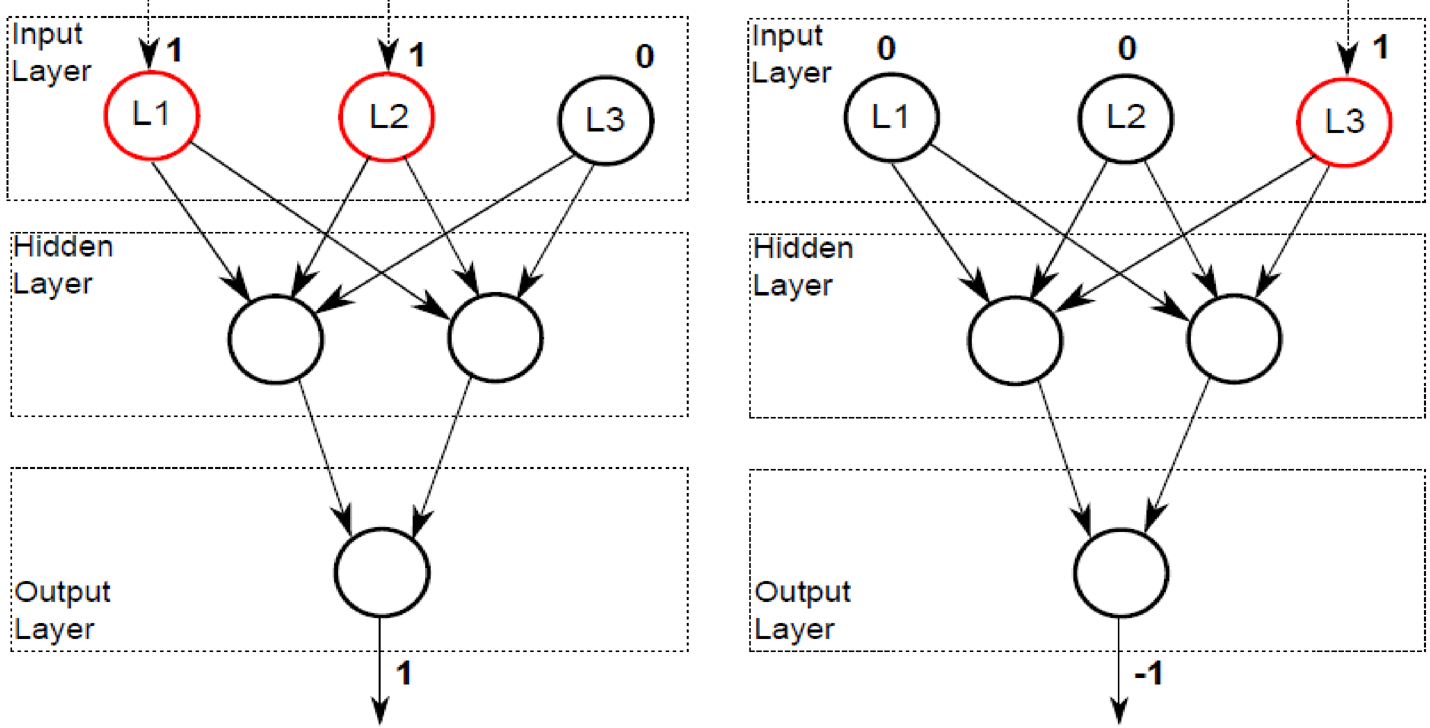
Relational Learning (CILP++)

- Extension of CILP allowing the system to be applied directly to ILP problems.
- Each neuron now denotes a first-order literal.
- Data for training the neural network is obtained by propositionalization
- Choice of literals to use is important, e.g. consider macro-operators (c.f. Mooney)
- TREPAN-like extraction of first-order rules possible

CILP++

motherInLaw(A,B) :- mother(A,C), wife(C,B)   2) ~motherInLaw(A,B) :- wife(A,C)

L1: mother(A,C)
L2: wife(C,B)
L3: wife(A,C)

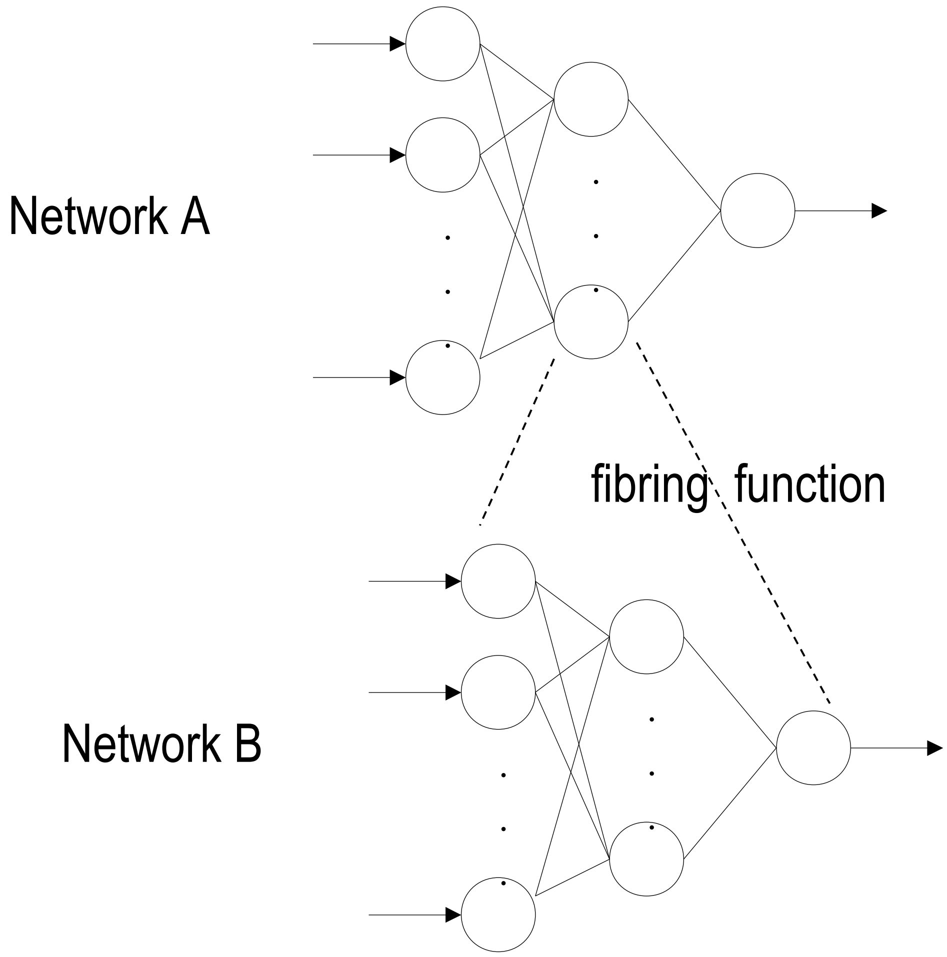


Franca, Zaverucha and d'Avila Garcez. Fast relational learning using bottom clause propositionalization with artificial neural networks. Machine Learning 94(1):81–104, 2014

Muggleton and Tamaddoni-Nezhad. QG/GA: A Stochastic Search for Progol. Machine Learning 70(2-3):121-133, 2008.

Richer structures: Fibring of Networks

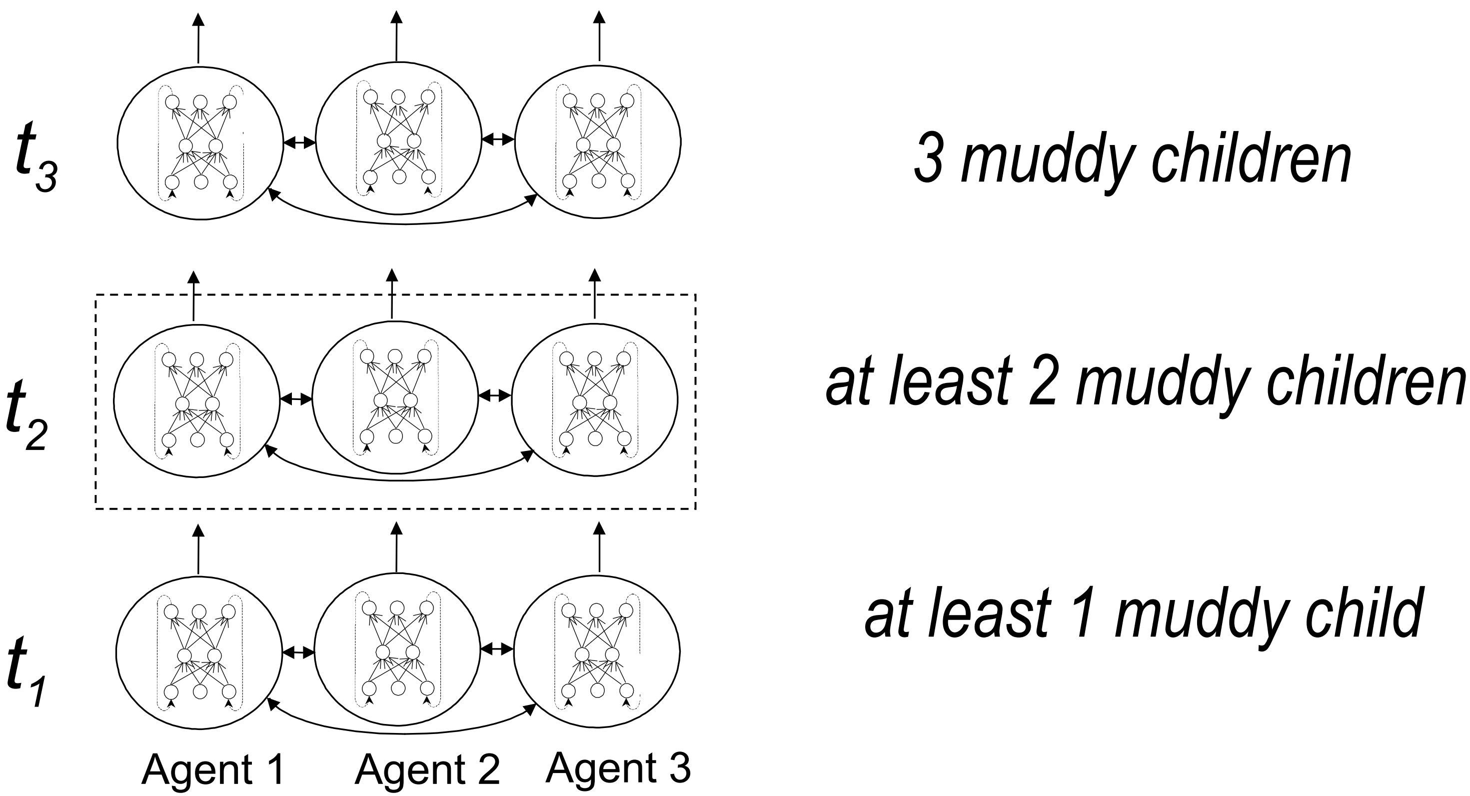
A neuron that is a network!
Early form of (modular) deep network
Strictly more expressive than shallow nets



d'Avila Garcez and Gabbay, Fibring Neural Networks, In Proc. AAAI 2004

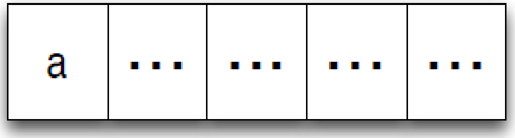
Richer structures: Temporal Reasoning

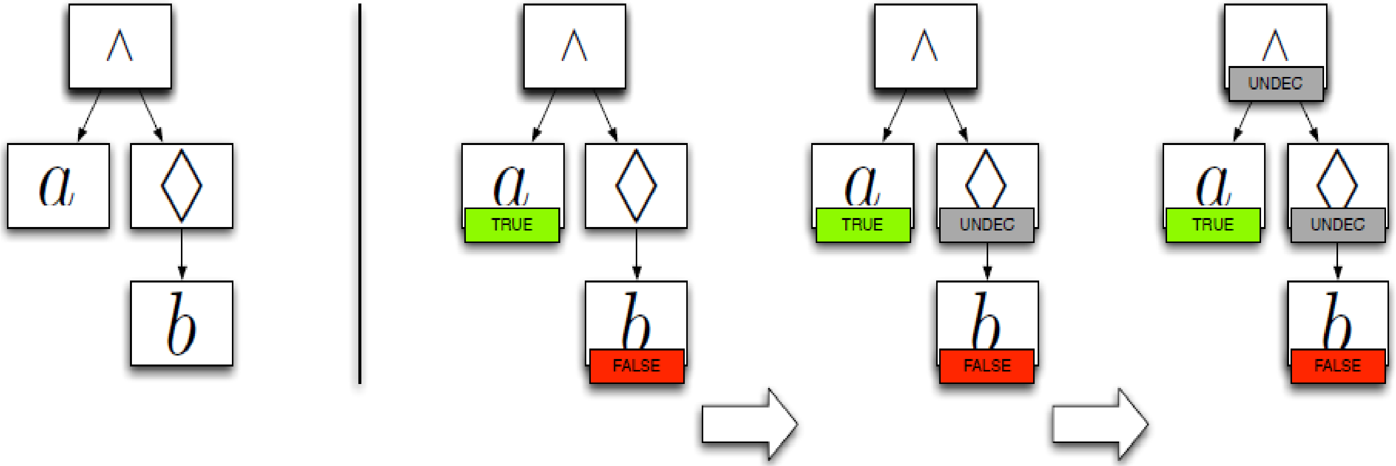
E.g. Muddy children puzzle (full solution)



d'Avila Garcez and Lamb. Reasoning about Time and Knowledge in Neural-Symbolic Learning Systems. NIPS 2003, Vancouver, Canada.

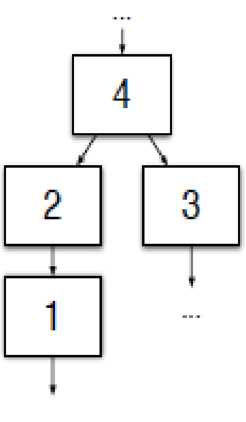
Run-time Monitoring (Propositional Modal Logic)

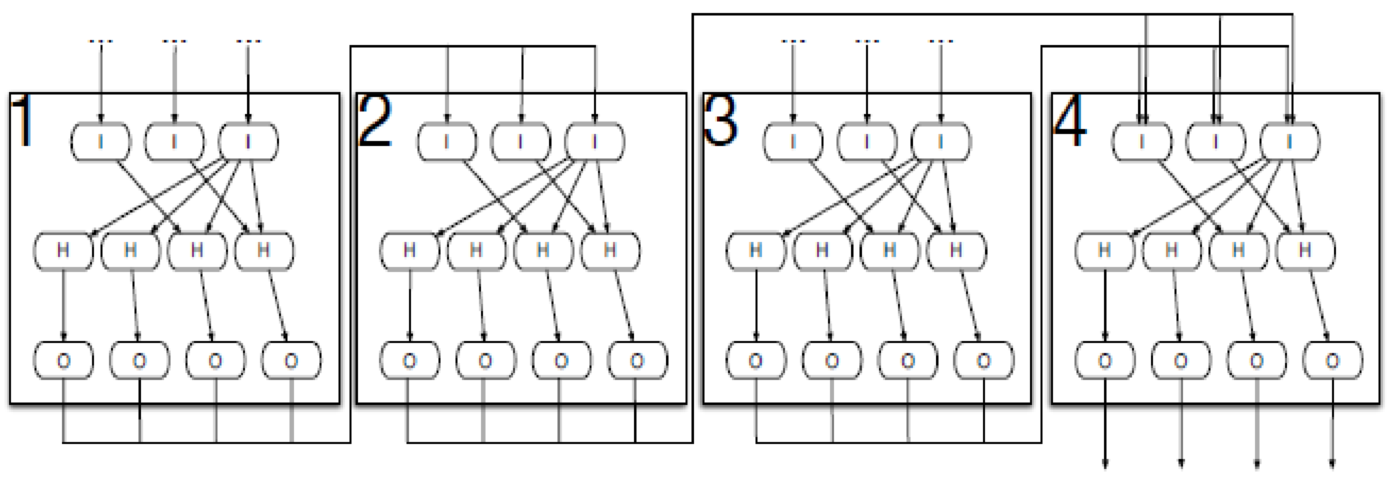




Run-Time Neural Monitor

The tree structure is “flattened” into an ensemble of CILP networks



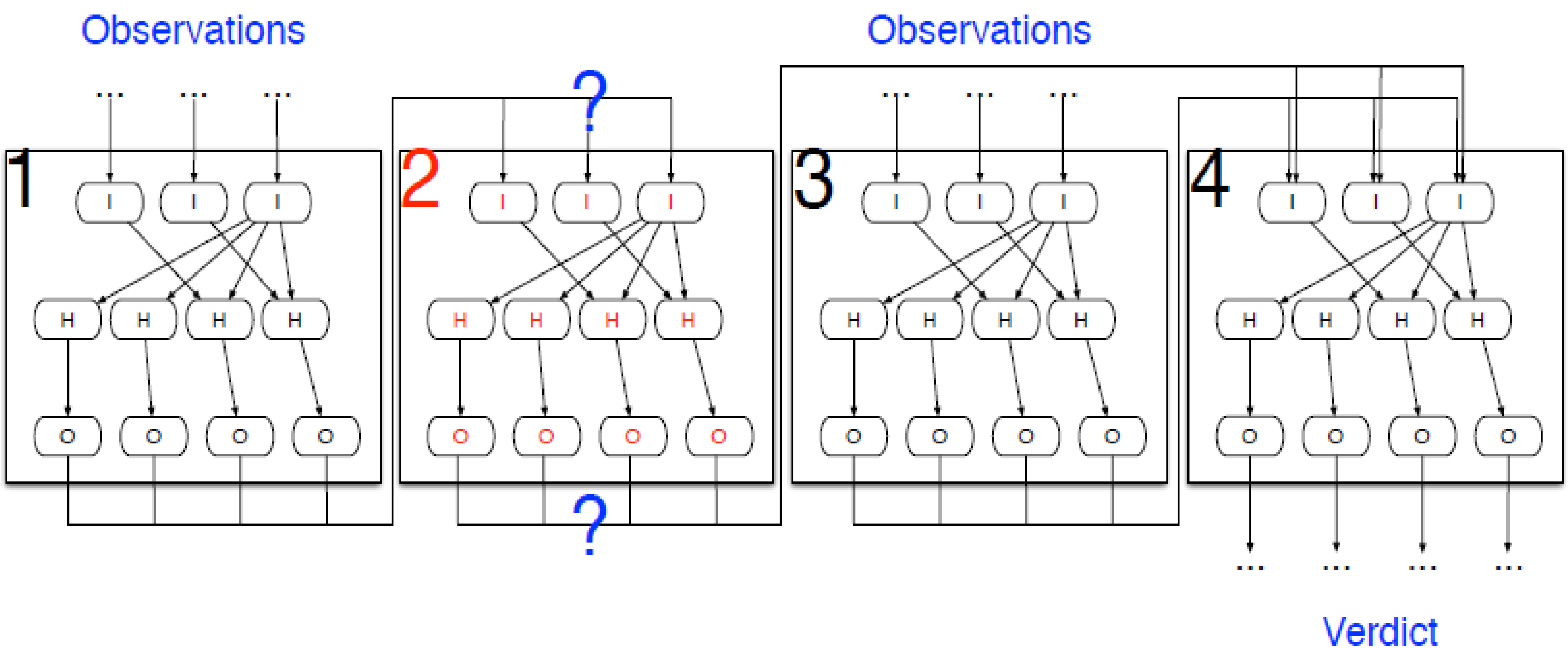


Learning = property adaptation

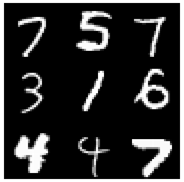
A. Perotti, A. S. d'Avila Garcez and Guido Boella. Neural-Symbolic Monitoring and Adaptation. In Proc. IEEE/INNS IJCNN 2015, Killarney, Ireland, July 2015.

Learning is local (modularity)

Propagate from observations to verdict and backpropagate label to abduce local input-output patterns (e.g. for network 2).

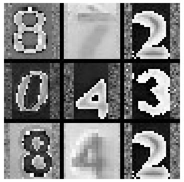


Transfer Learning (1)



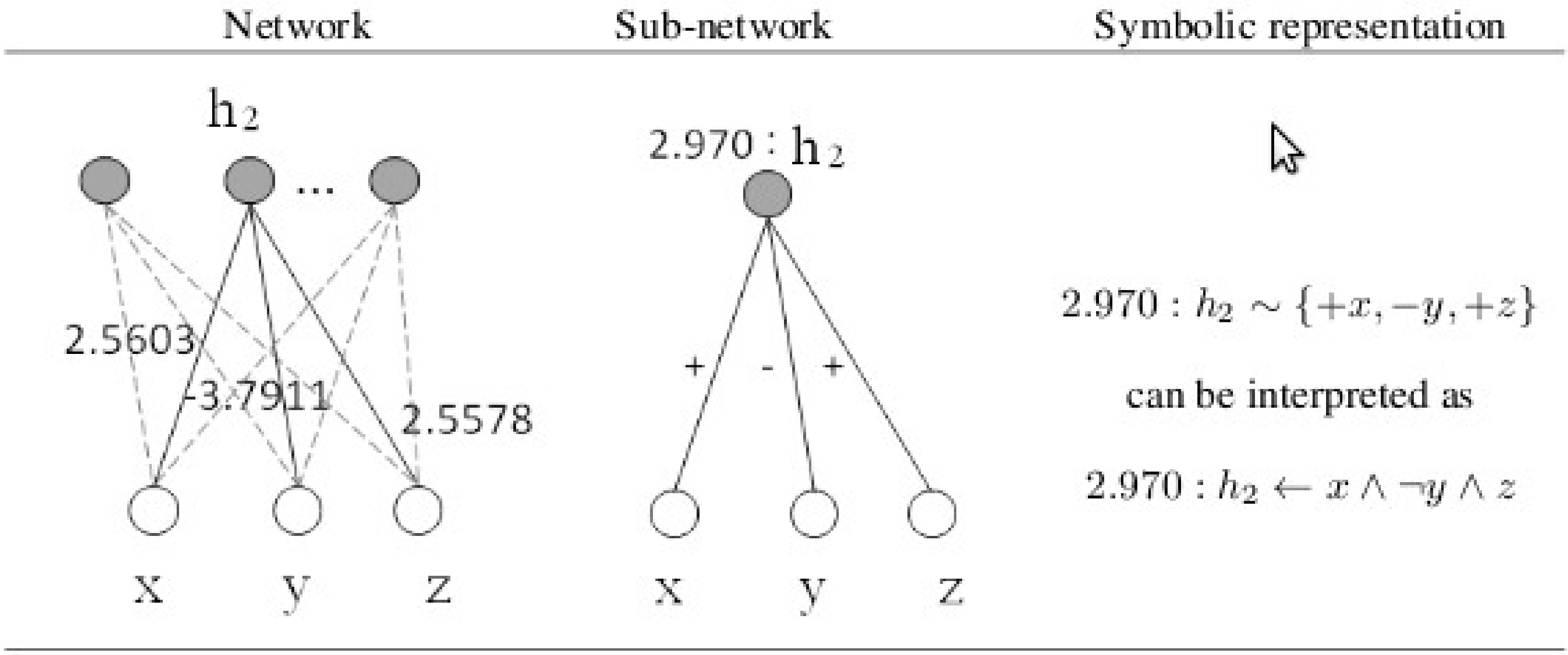
MNIST

TiCC



ICDAR

Knowledge extraction from RBMs (originally the building block of (modular) deep nets, c.f. Hinton's DBNs)

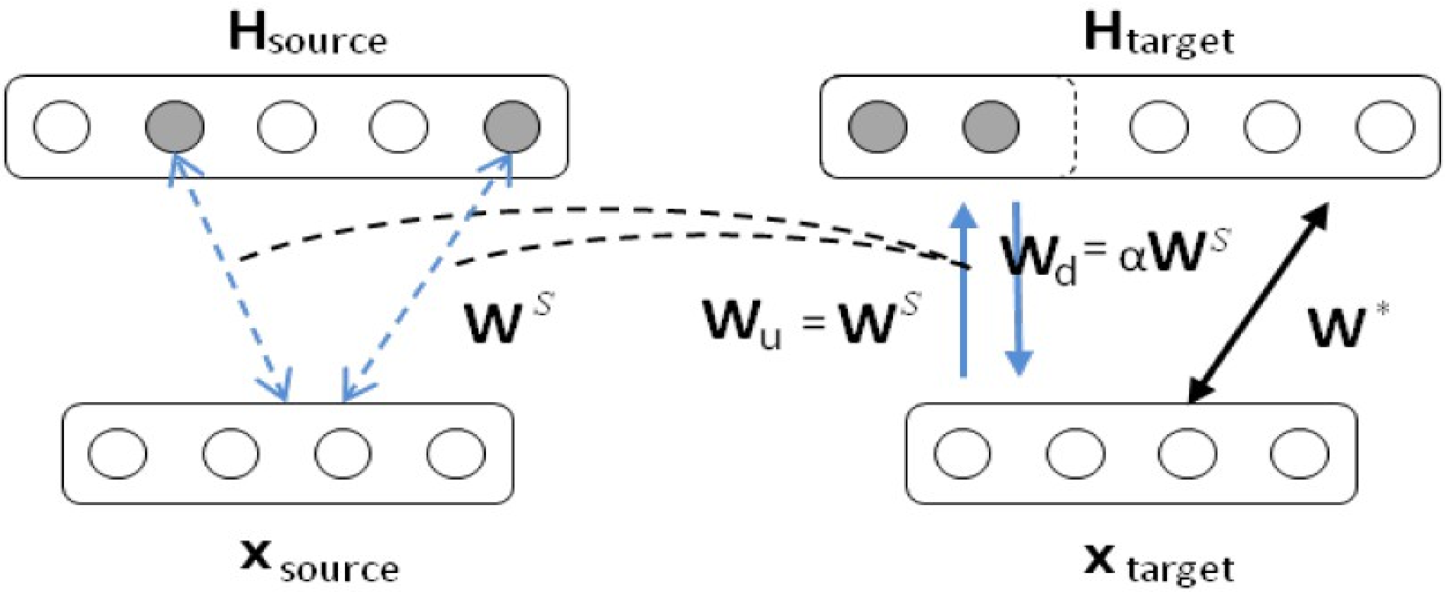


$2.970 : h_2 \sim \{+x, -y, +z\}$

can be interpreted as

$2.970 : h_2 \leftarrow x \wedge \neg y \wedge z$

Transfer Learning (2)



S. Tran and A. S. d'Avila Garcez. Deep Logic Networks: Inserting and Extracting Knowledge from Deep Belief Networks. IEEE Transactions NNLS, Nov, 2016

Logic Tensor Networks (LTNs)

- Neural nets with rich structure can represent more than classical propositional logic
- But... neural nets are essentially propositional (John McCarthy was right)
- To take advantage of full FOL, a more **hybrid** approach is needed
- One needs to get the representation right first: the logical statements act as (soft) **constraints** on the neural network...

Semantic Image Interpretation (1)

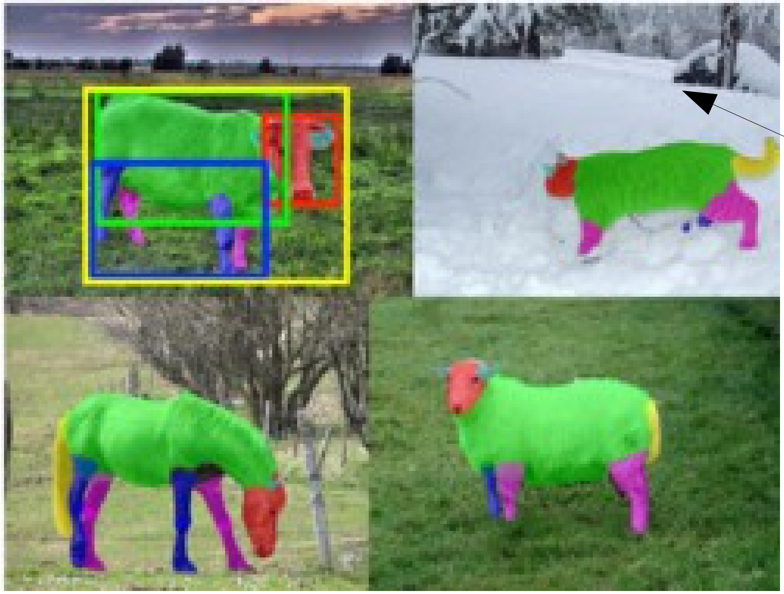
Given an image, extract a graph that describes its semantic content

You may want to be able to say: Normally, every cat has a tail, or having learned how to recognise horses very well from millions of images, you may now want to simply define zebra = horse with stripes

Or you may want to communicate with the system:

Q. Get me the red thing next to the sheep

A. The horse's muzzle? Yes.



Make sure your system does not distinguish cats from wolves 99% correctly because of the snow in the background...

Semantic Image Interpretation (2)

In LTN, we build the graph by predicting facts given the bounding boxes, e.g.: Cow(b1), PartOf(b2,b1), Head(b2), etc.

In LTN, an object is described by a vector of features: e.g. John = (NI number, age, height, 3x4 picture, etc.)

Object detection (bounding box detection and labeling) is performed by an object detector (Fast RCNN)

LTN assigns a degree of truth (the grounding G) to atomic formulas: G(Cow(b1)) = 0.65, G(PartOf(b2,b1)) = 0.79...

$$G(bi) = <\underbrace{\text{score(Cow), score(Leg)} \ldots \text{score(Head)}}_{\substack{\text{Semantic features: the score of the} \\ \text{bounding box detector on } b_i \\ \text{for each class of objects}}}, \underbrace{x, y, x', y'}_{\substack{\text{Geometric features:} \\ \text{the coordinates of } b_i}}>$$

LTN in action

1. $\forall x(\neg PartOf(x, x))$
2. $\forall xy(PartOf(x, y) \rightarrow \neg PartOf(y, x))$
3. $\forall xy(Cow(x) \wedge PartOf(x, y) \rightarrow Leg(y) \vee Neck(y) \vee Torso(y) \vee Head(y))$
4. $\forall xy(Cow(x) \rightarrow \neg PartOf(x, y))$
5. $\forall xy(Torso(x) \rightarrow \neg PartOf(y, x))$.

- Grounding for PartOf is given by the % of intersection between two bounding boxes
- One can query the knowledge-base (KB) to obtain further groundings for training
- Learning is... maximizing satisfiability!

Learning in LTNs...

Given KB and data, LTN calculates a grounding for both in the "usual ways", i.e. compositionally

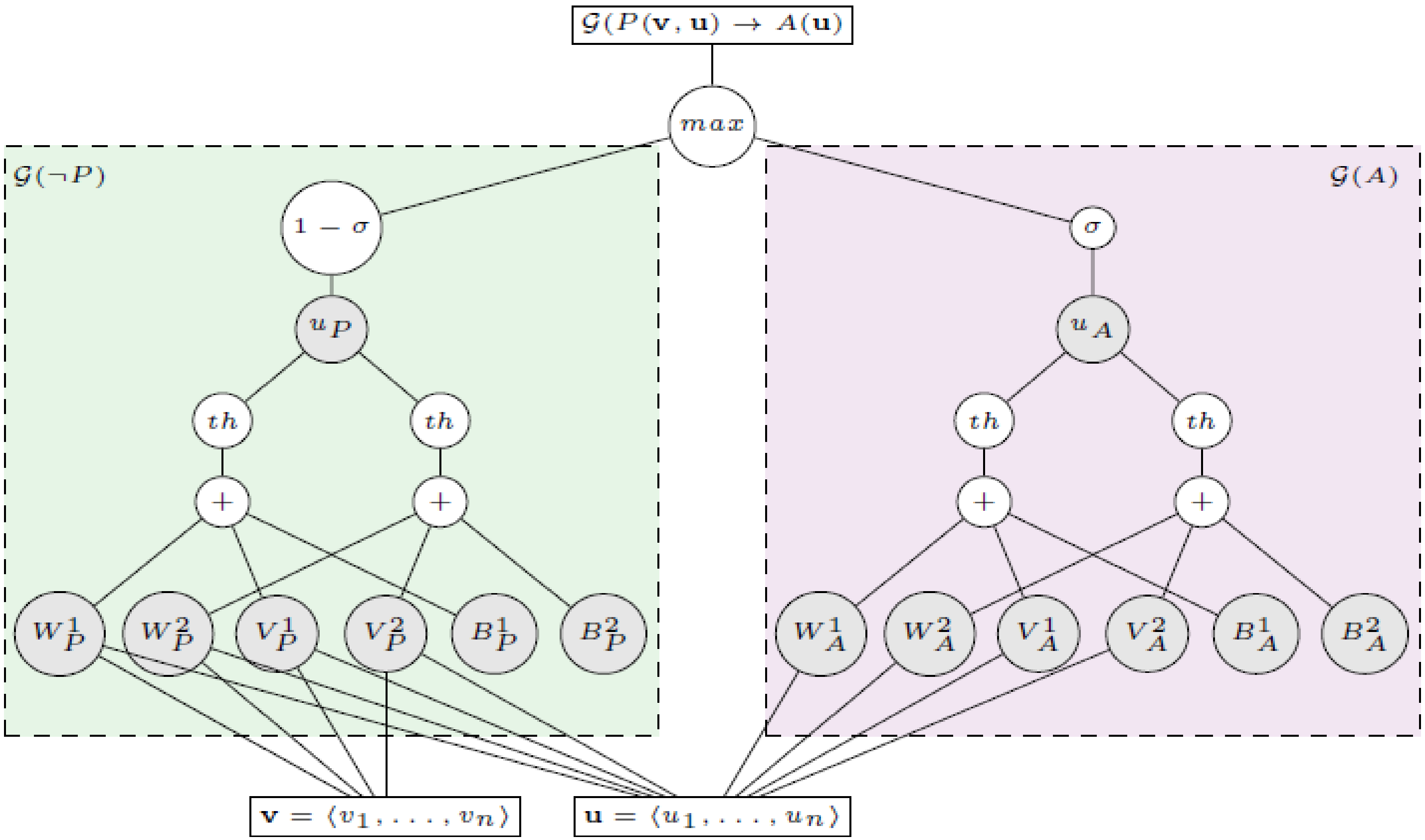


Fig. 1. Tensor net for $P(x, y) \rightarrow A(y)$, with $\mathcal{G}(x) = \mathbf{v}$ and $\mathcal{G}(y) = \mathbf{u}$ and $k = 2$.

The Tensor Network...

LTN + Fast RCNN improves on Fast RCNN (state of the art at the time) at object type classification:

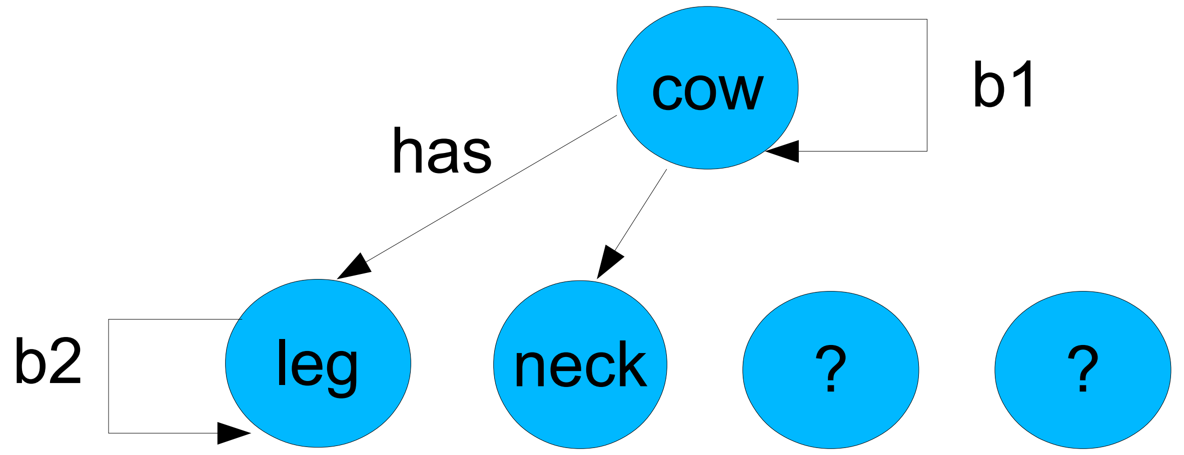
$$\mathcal{G}(P) = \sigma\left(u_P^T \tanh\left(\mathrm{v}^T W_P^{[1:k]} \mathrm{v} + V_P \mathrm{v} + B_P\right)\right)$$

$$\mathcal{G}^* = \operatorname*{argmin}_{\hat{\mathcal{G}} \subseteq \mathcal{G} \in \mathbb{G}} \sum_{\langle [v,w], \phi(\mathrm{t}) \rangle \in \mathcal{K}_0} Loss(\mathcal{G}, \langle [v,w], \phi(\mathrm{t}) \rangle)$$

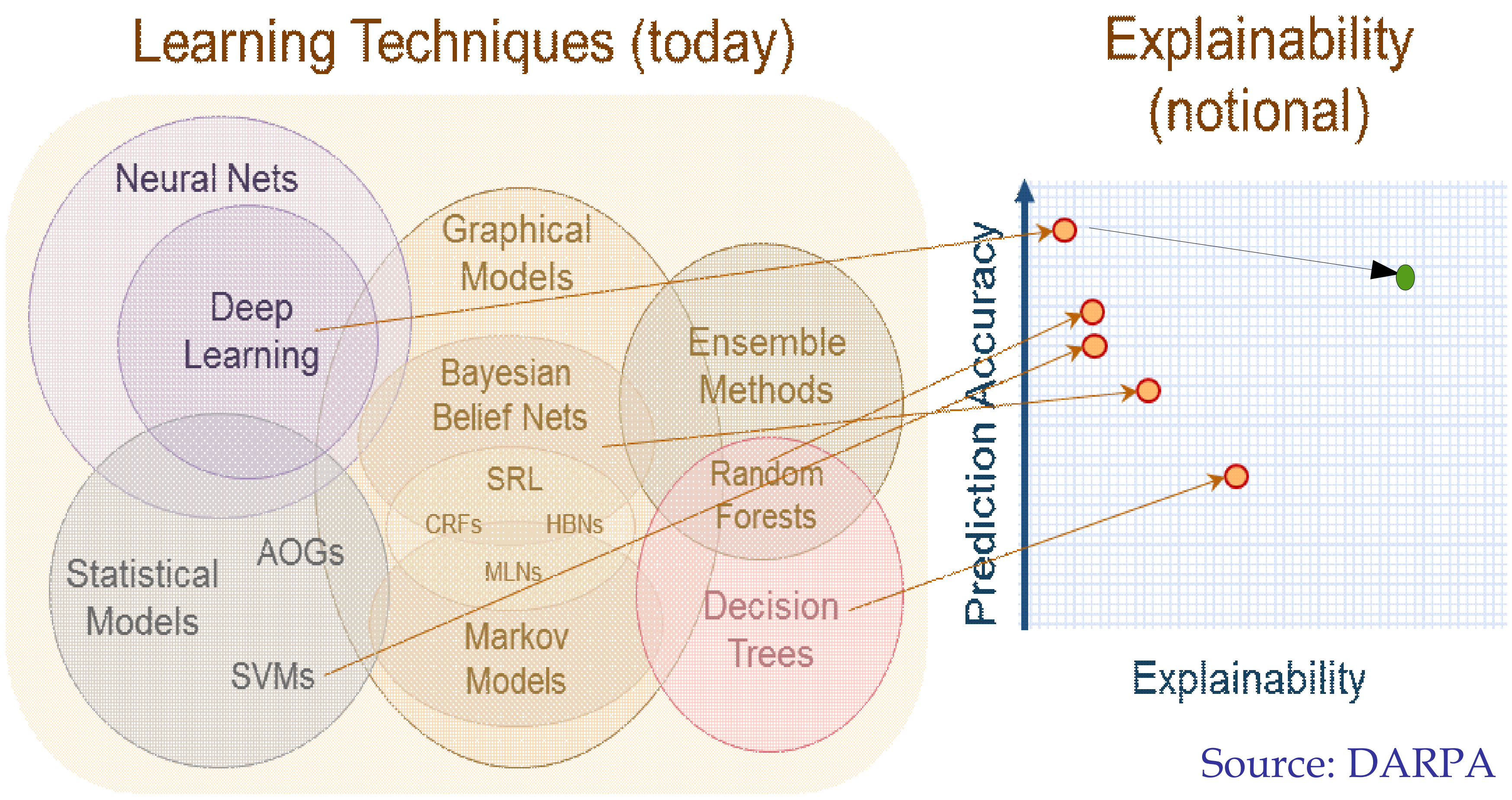
I. Donadello, L. Serafini and A. S. d'Avila Garcez. Logic Tensor Networks for Semantic Image Interpretation. In Proc. IJCAI'17, Melbourne, Australia, Aug 2017.

And finally, the knowledge graph...

- Given a trained LTN, start with an unlabeled graph.
- For every bounding box bi ask the LTN for the set of facts {Cow(bi), Leg(bi), Neck(bi), Torso(bi),...} and select the facts with grounding larger than a threshold.
- For every bounding box bi ask the LTN for the set of facts {PartOf(bi, bj)} with j = 1,…,n. Then, select the facts with grounding larger than a threshold.



Explainable AI = ML + KR



Source: DARPA

- What do I need to change in order to have my credit application accepted the next time?

Verification of Neural Nets

Whose fault is it when a self-driving car gets into an accident? How about millions of self-driving cars?

In the news recently: Tesla self-driving system cleared in deadly crash

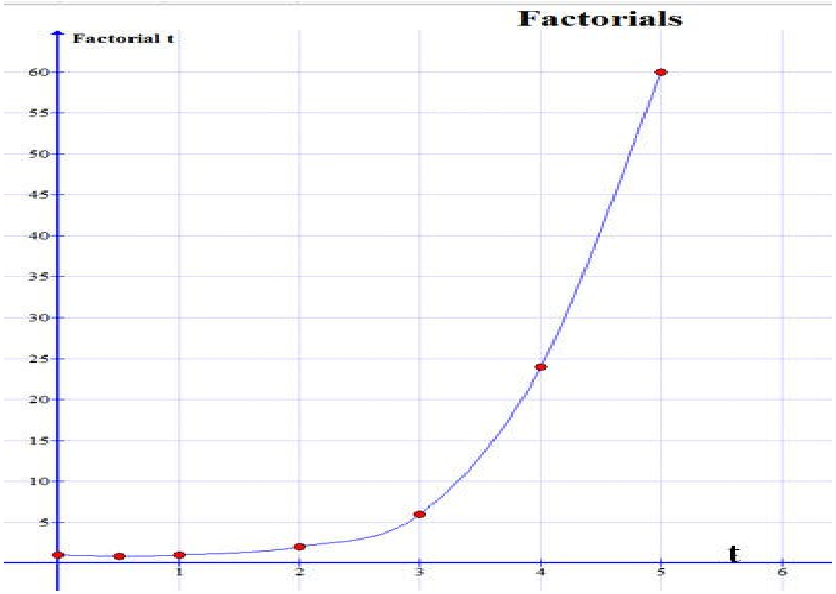
Is the following argument sufficient? There will be fewer deaths from self-driving cars!

Or do we need to be able to provide explanations and guarantees?

Deep Learning + Symbolic ML

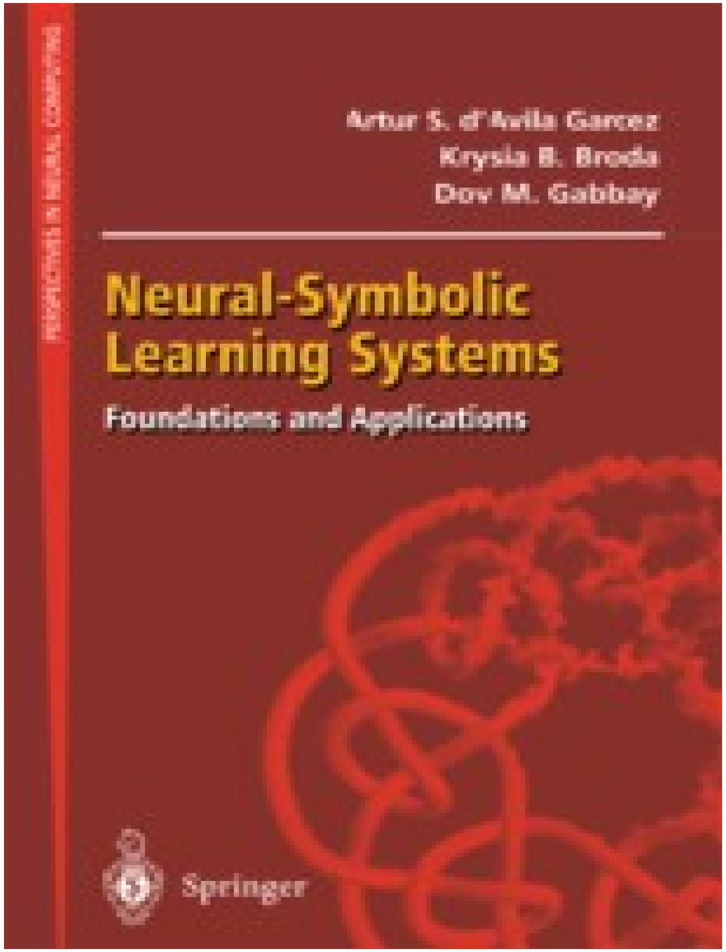
- We want good classification and prediction but also useful descriptions… e.g. learning factorial (*n!*)

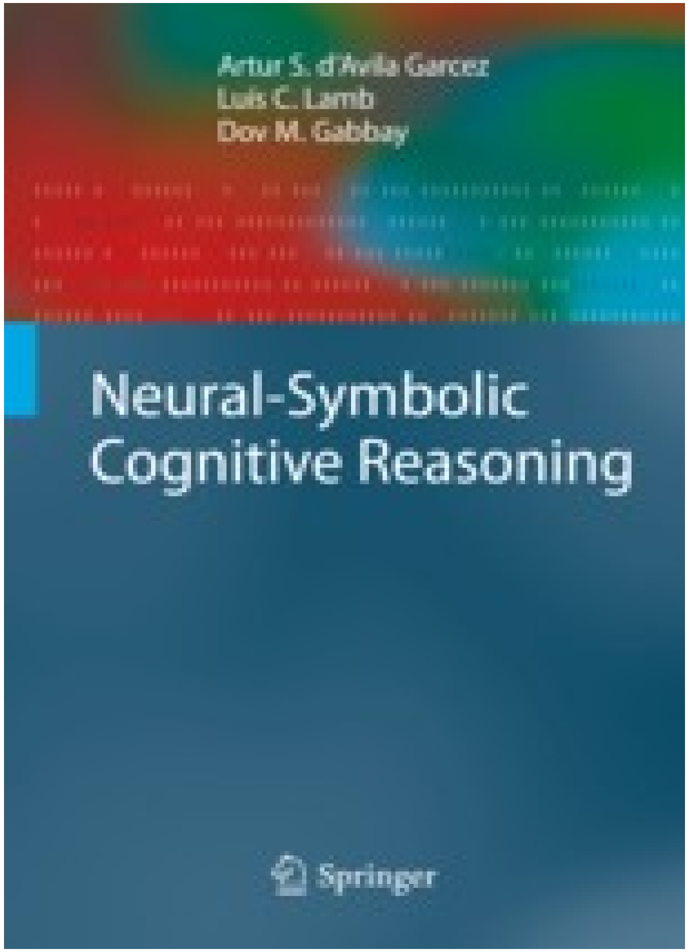
$$
\begin{aligned}
0! &= 1\\
1! &= 1\\
2! &= 1 \cdot 2 = 2\\
3! &= 1 \cdot 2 \cdot 3 = 6\\
4! &= 1 \cdot 2 \cdot 3 \cdot 4 = 24\\
5! &= 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120\\
6! &= 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = 720
\end{aligned}
$$




```
factorial(0,1).
factorial(N,F) :- N>0, N1 is N-1,
                  factorial(N1,F1),
                  F is N * F1.
```

For more information...





Conclusion: Why Neurons and Symbols

To study the statistical nature of learning and the logical nature of reasoning.

To provide a unifying foundation for robust learning and efficient reasoning.

To develop effective computational systems for AI applications.

Thank you!

Throw away your paradigm...

neurons

symbols

The future is neural-symbolic

Paraphrased from
Murray Shanahan