Neural-Symbolic Learning and Reasoning with Constraints
Tutorial at IEEE IJCNN
8 July 2018

Neural-Symbolic Computation: Thinking beyond Deep Learning

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

The AI revolution...

The promise of AI:

- Education (active learning)
- Finance (time series prediction)
- Security (image and speech recognition)
- Health (sensors, companions, drug design)
- Telecom (infrastructure data analysis)
- Games (online learning)
- Transport (logistics optimization)
- 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

- Neural networks for effective perception
- But perception alone is insufficient:
  - Reasoning, Explanation, Transfer
- Rich knowledge representation: nonmonotonic, relational, variables, recursion, time, uncertainty.
- Neural-symbolic computing: neural networks with logical structure (in particular: **compositionality**)

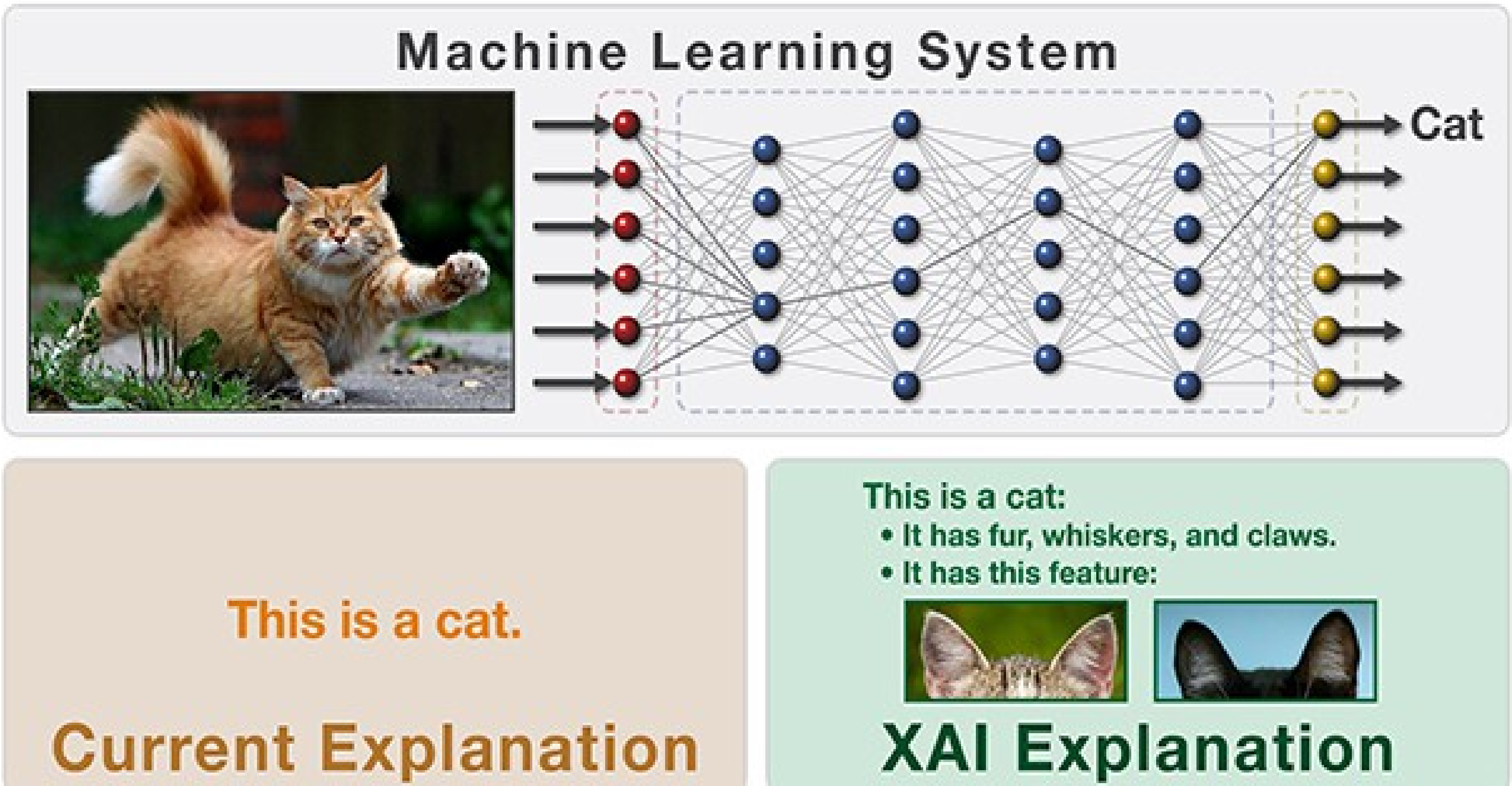
Machine Learning (ML)

Systems that improve their own performance from experience

Systems that, in addition, enable humans to improve their performance (human-machine interaction needed here)

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

DARPA's Explainable AI

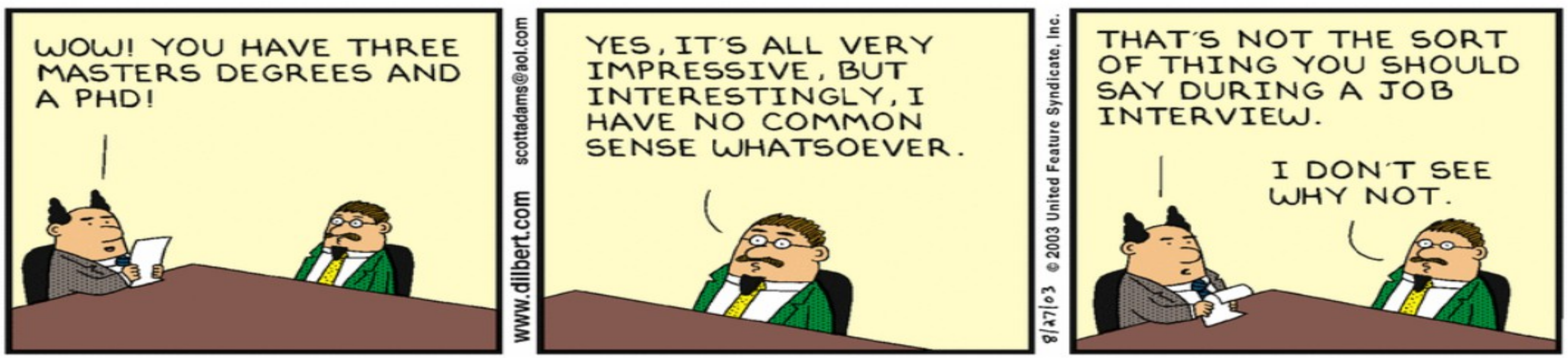


- XAI = Interpretable ML
- Explanation = proof history, not XAI

Deep Learning

Given Big Data, deep learning (based on neural nets) works better than symbolic ML!

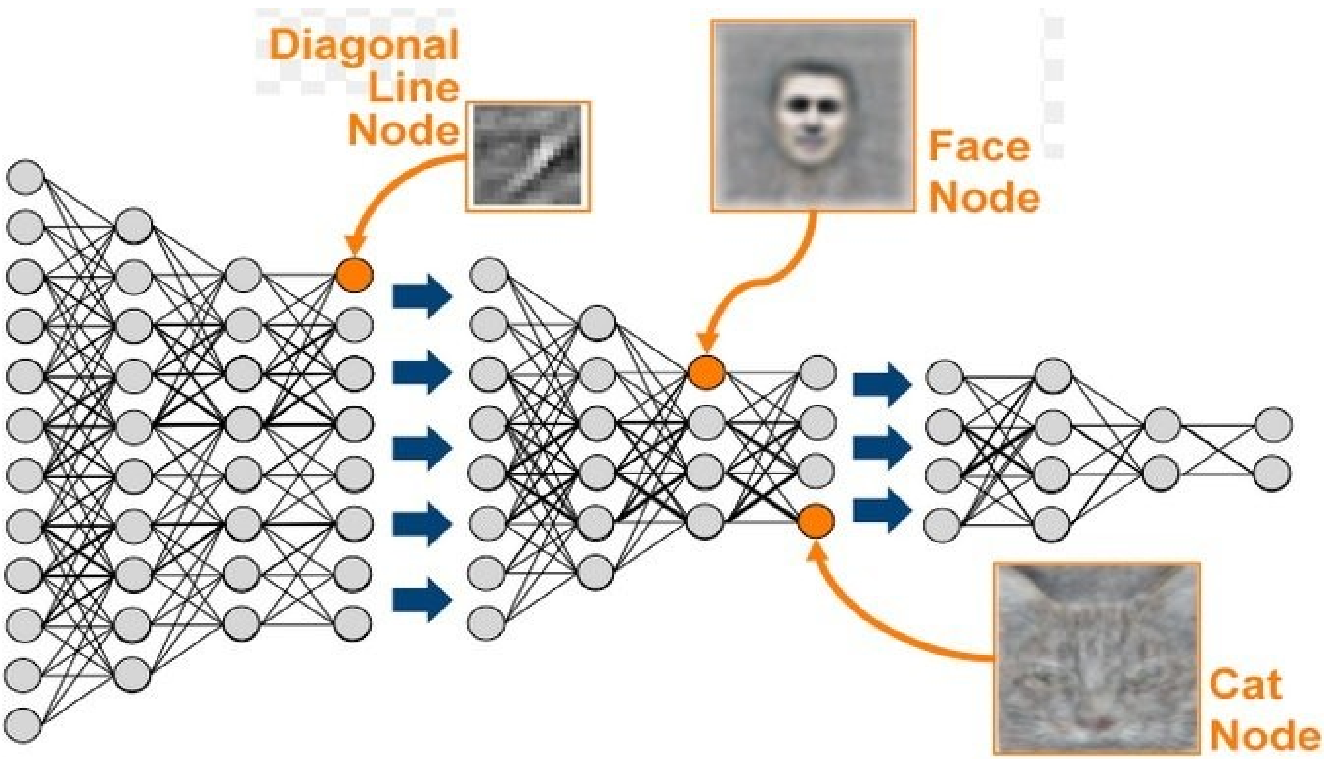
But, where is commonsense?



Deep networks

Very successful on handwritten digit classification, object recognition, speech/audio and games

How about language? Video understanding?



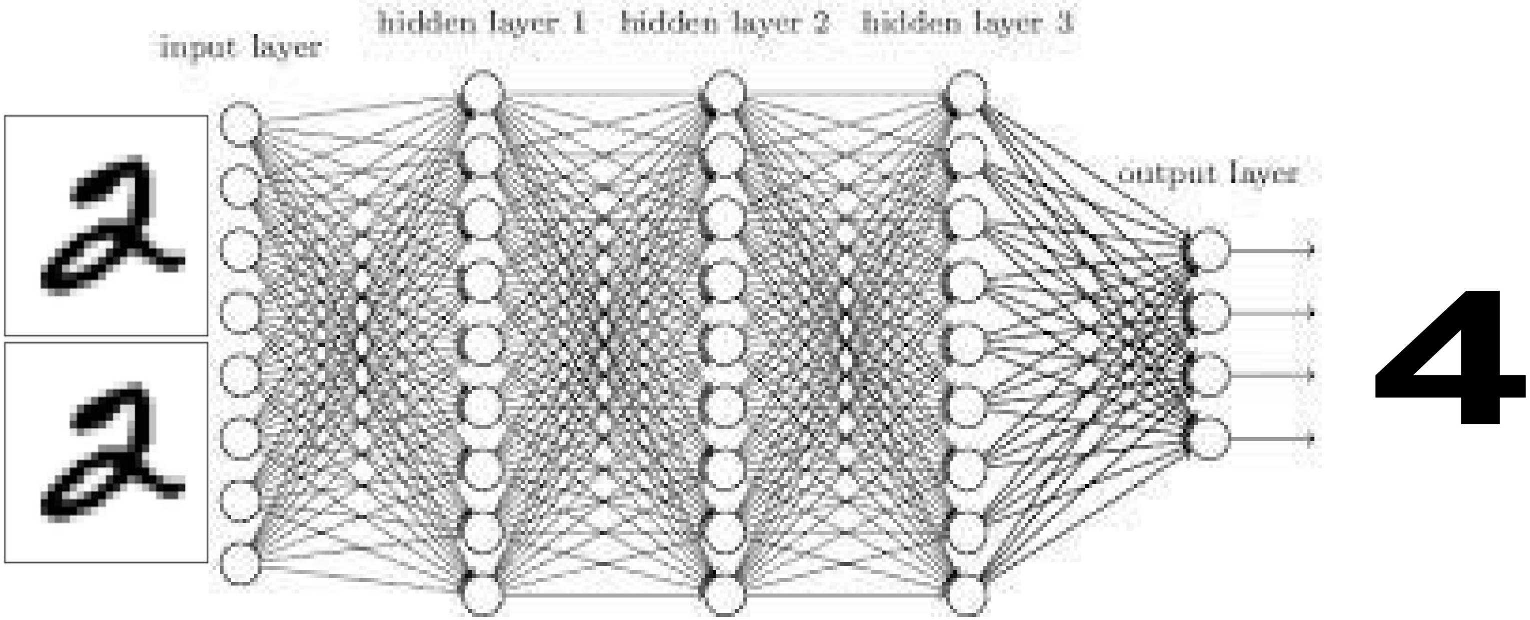
Geoff Hinton says AI needs to start over...

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

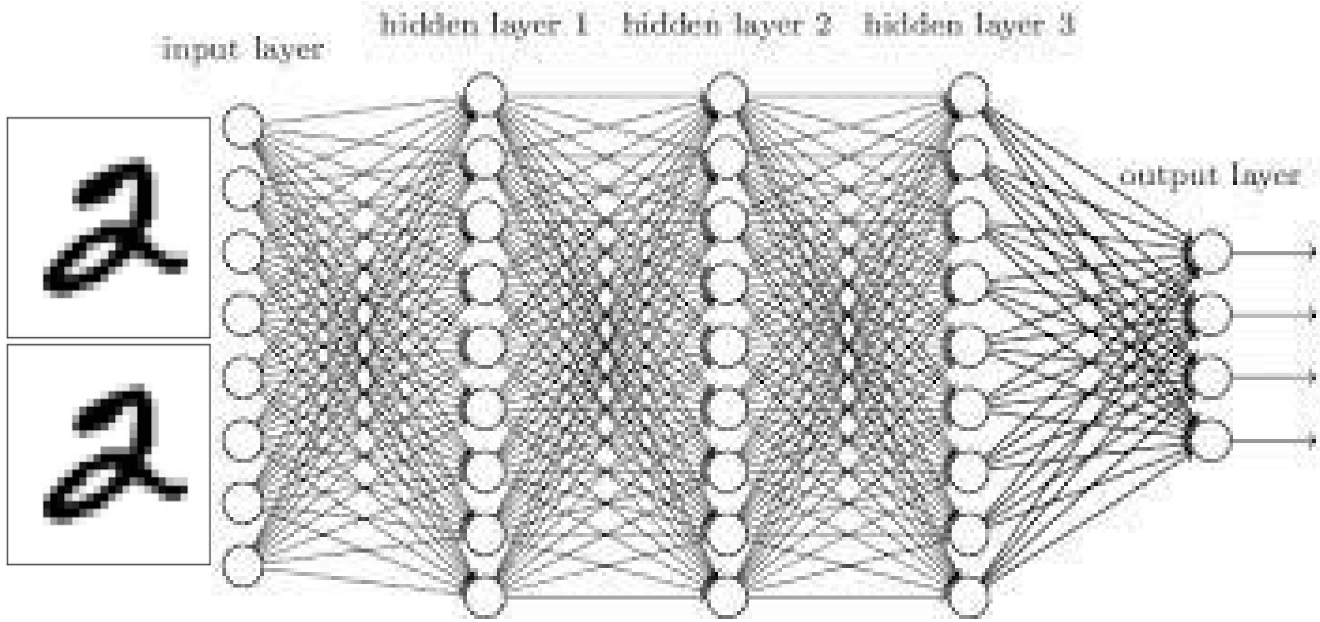
- Adversarial networks are not doing much better

School bus Adversarial perturbations Ostrich

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

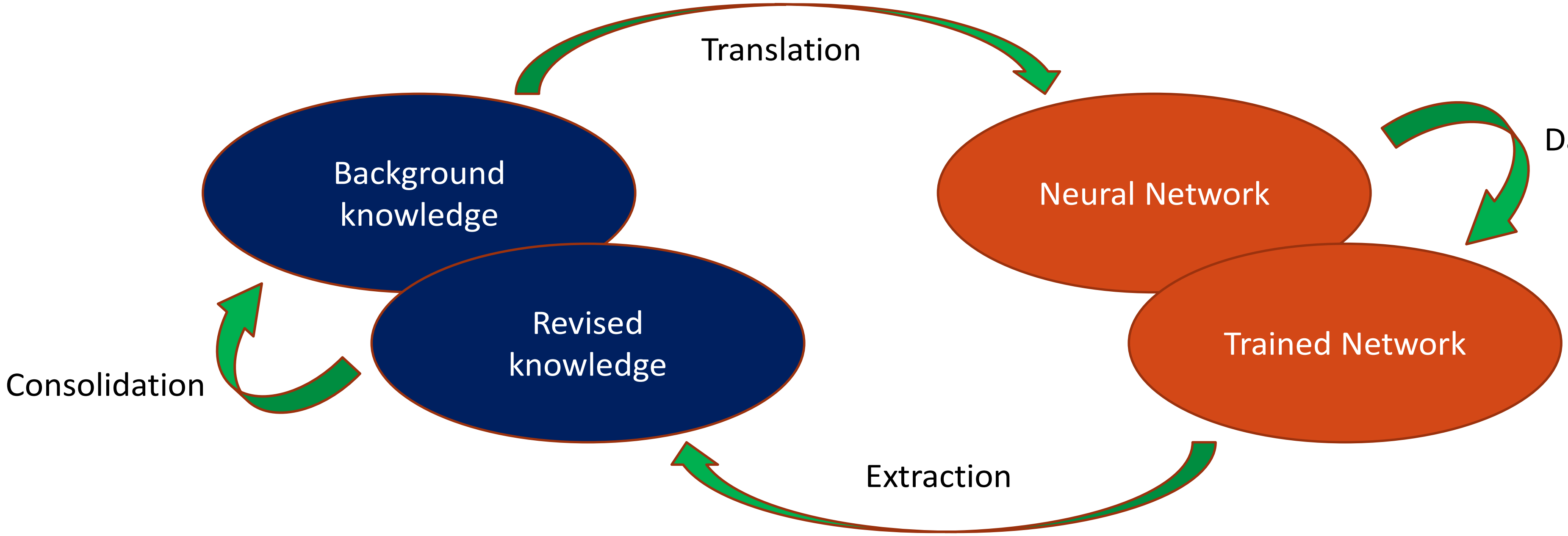
low level, efficient neural structures
(with the same, simple architecture throughout)

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

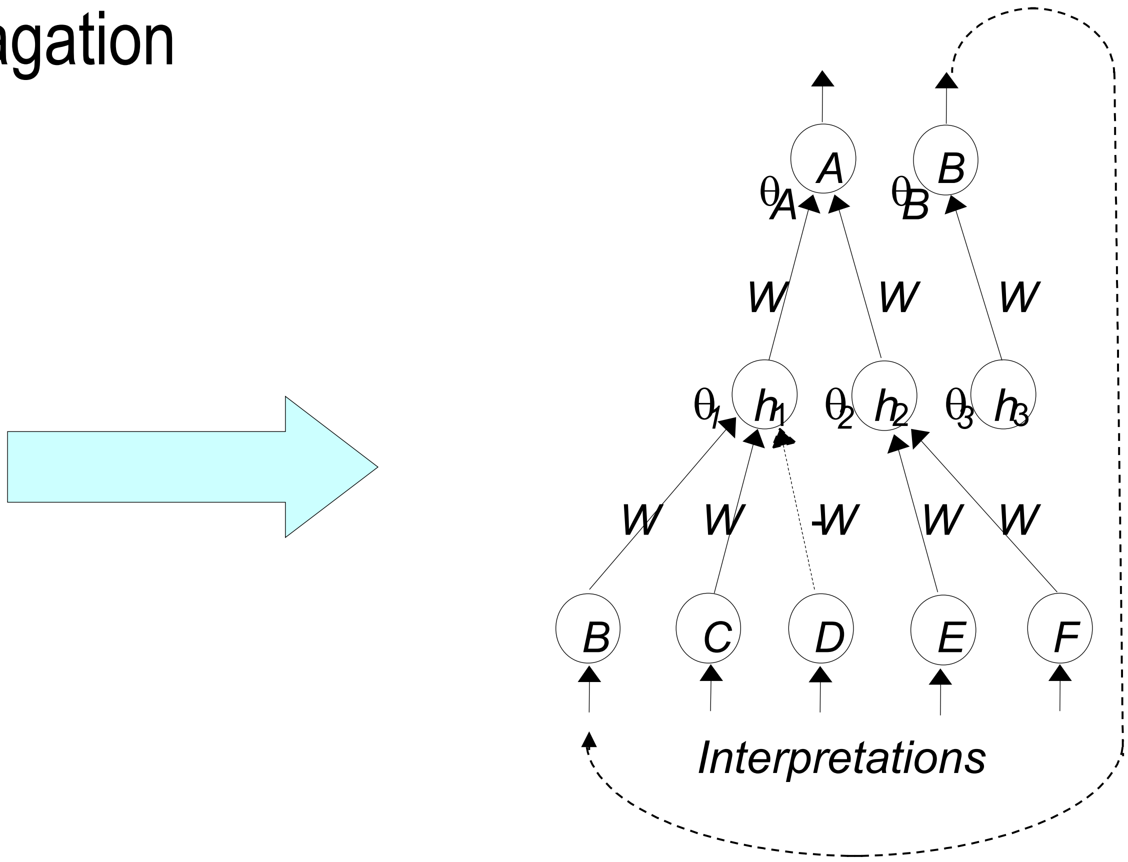
Learning with Backpropagation

Knowledge Extraction

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

r_2 : $A \leftarrow E, F;$

r_3 : $B \leftarrow$



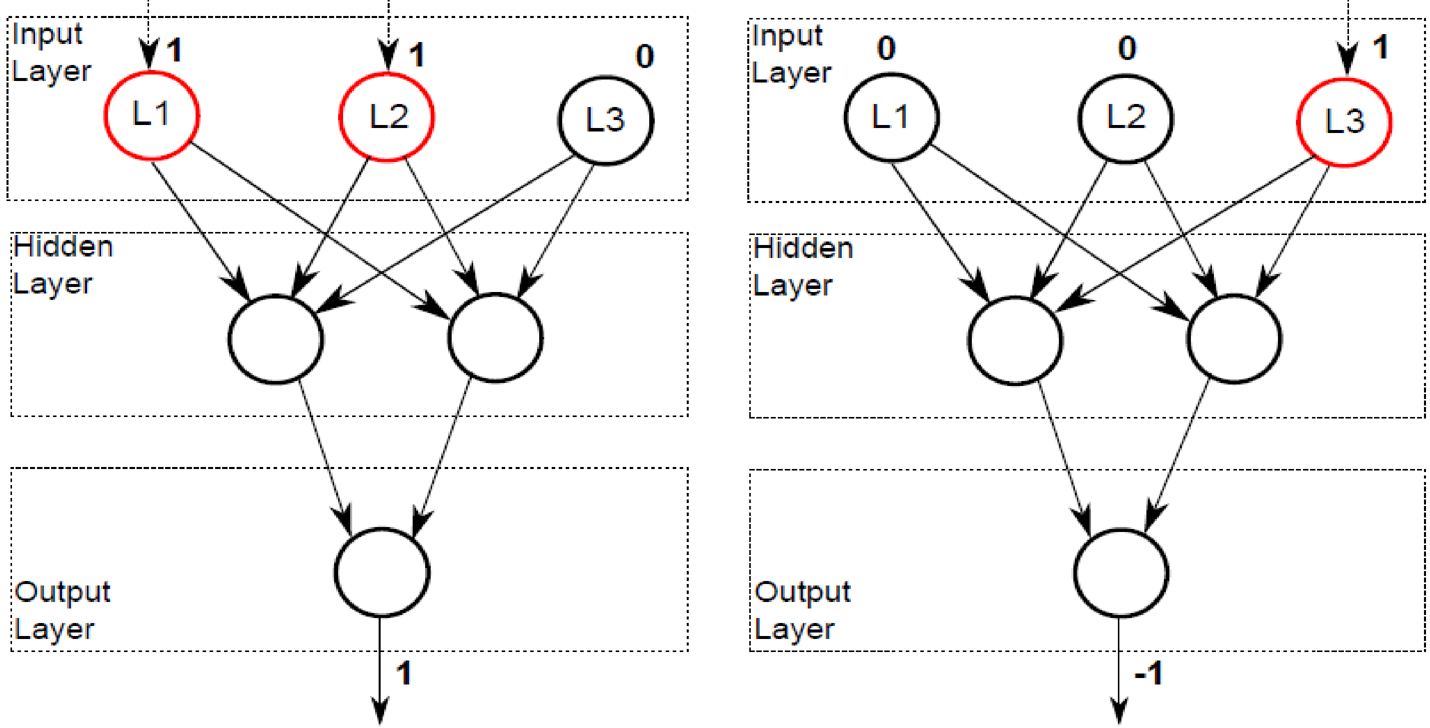
Relational Learning (CILP++)

- 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)

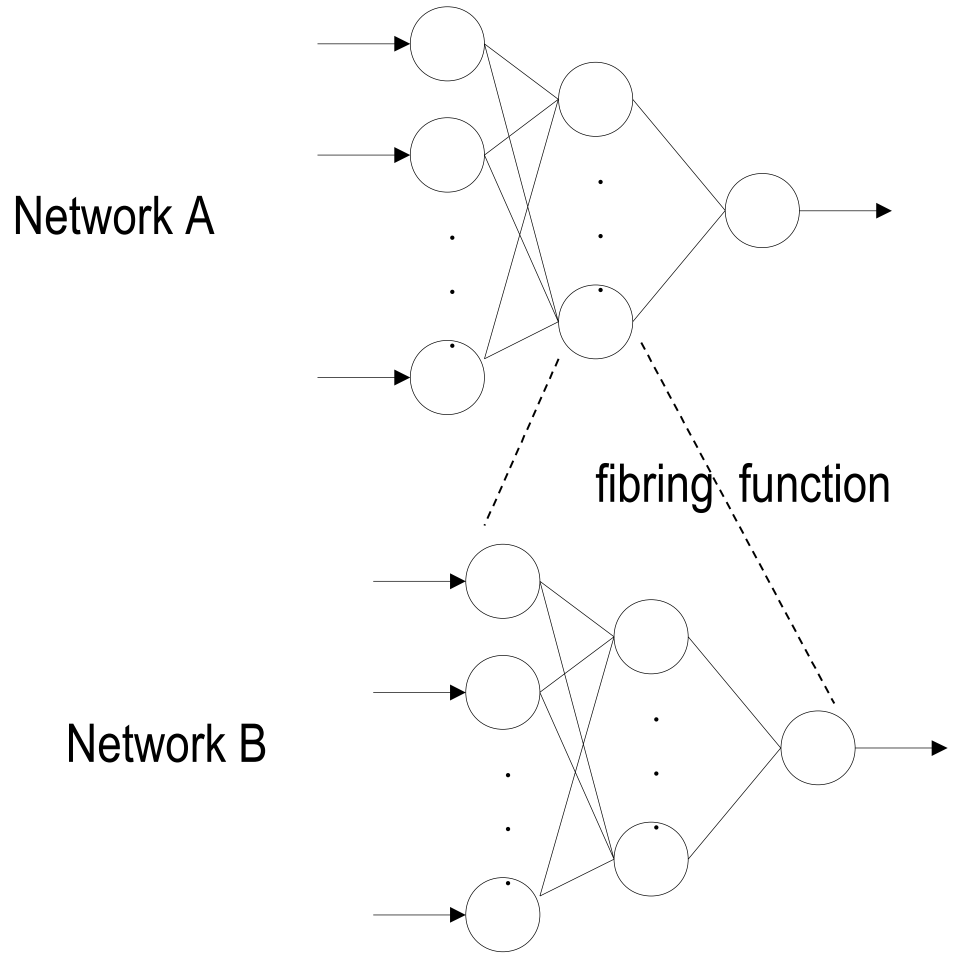


Franca, Zaverucha and d'Avila Garcez. Fast relational learning using bottom clause propositionalization with artificial neural networks. Machine Learning 94(1):81–104, 2014

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

Richer structures: Fibring of Networks

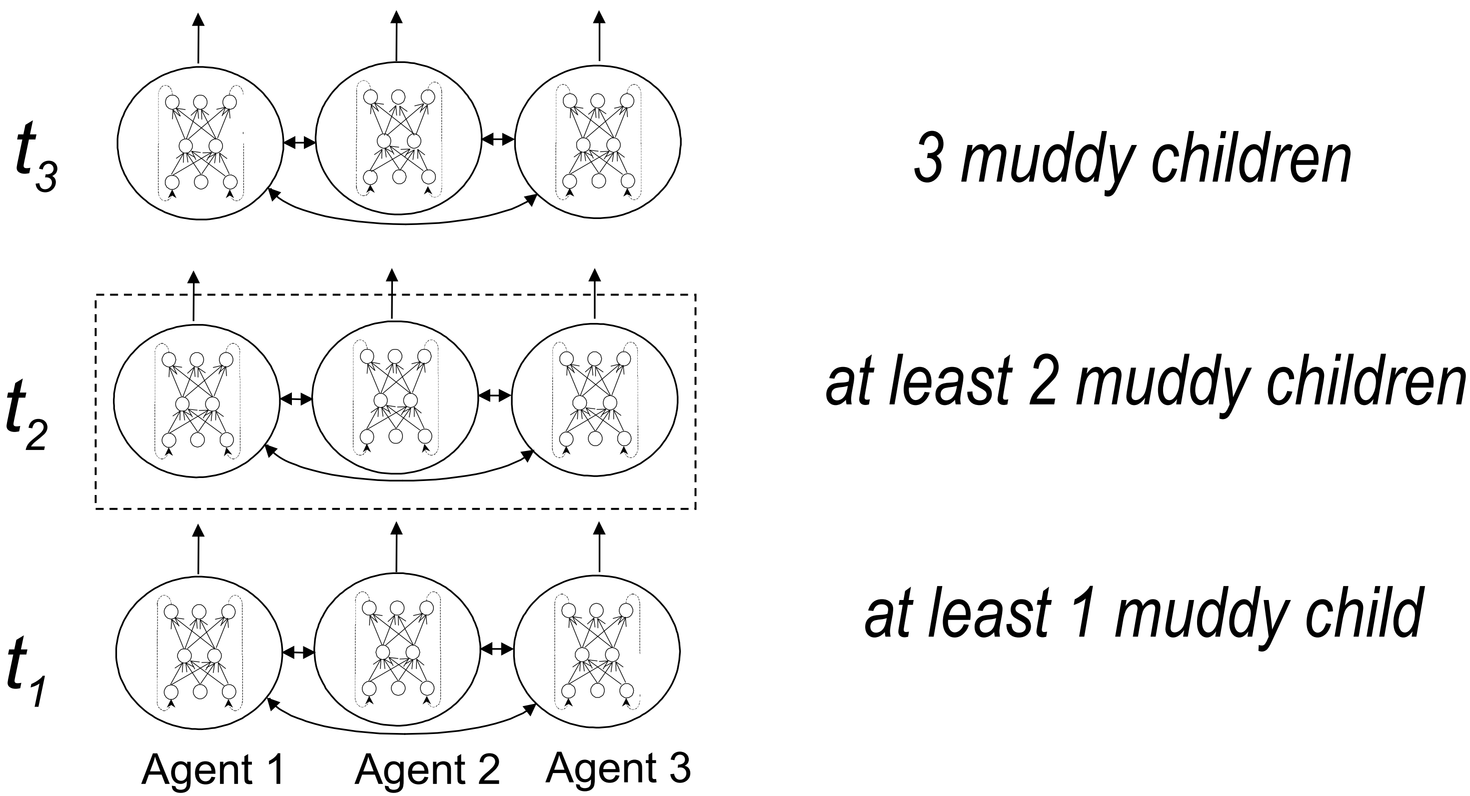
A neuron that is a network!
Early form of (modular) deep network
Strictly more expressive than shallow nets



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.

Why Knowledge Extraction?

Correctness / soundness

Proof history (goal-directed reasoning)

Levels of abstraction (modularity)

Transfer learning (analogy)

System maintenance/improvement

Knowledge Extraction techniques

- Soundness is important!
- Pedagogical vs Decompositional
- Early methods: MofN, CILP
- Decision tree extraction - TREPAN
- Automata extraction - recurrent networks
- Reducing harm from gambling: a practical application of knowledge extraction
- Current work: extraction from deep nets, soft decision trees, probabilistic MofN, distilling...

Soundness

- A guarantee that the explanation extracted reflects the behavior/semantics of the neural network
- Sound/complete extraction implies a loss in performance (guarantee in the limit only)
- Be suspicious of knowledge extraction that produce higher accuracy than the neural net
- In practice, efficient extraction may be unsound (and work more like a learning algorithm)
- Soundness is needed e.g. if neural net is used in a safety-critical domain, e.g. self-driving car...

Extraction methods

Algorithms:

- Pedagogical: treat network as an oracle to query input/output patterns
- Decompositional: inspect the internal structure of the network
- Eclectic: consider doing both of the above

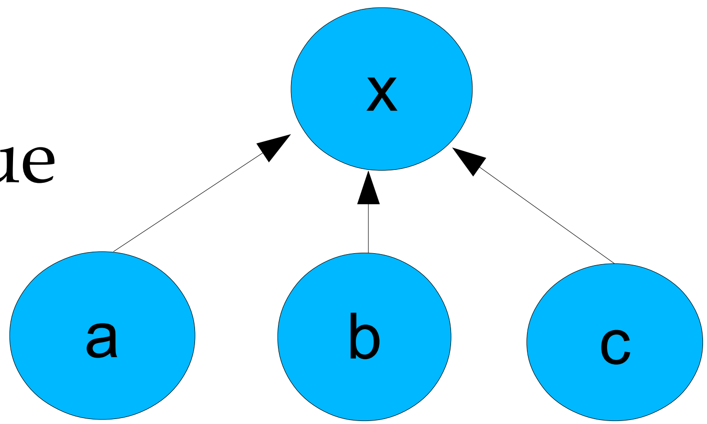
Explanation:

- Explanation of a case or instance (distilling, feature importance ranking, visualization)
- Model description (knowledge extraction)

MofN and CILP extraction algorithms

- MofN [1]: realization that the building block of a neural net is very good at learning/ representing MofN rules:

  If 2 of (a,b,c) are true then x is true

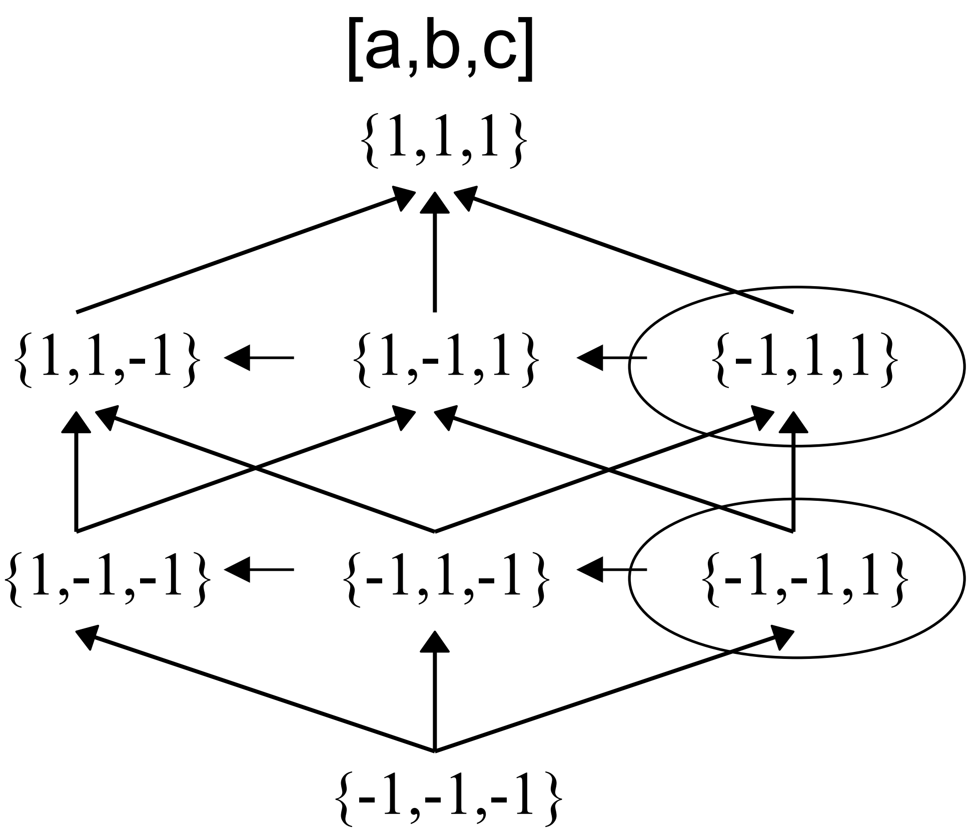


- CILP [2] sound extraction algorithm

  [1] Knowledge-based artificial neural networks, G. Towell and J. Shavlik, AIJ, 1994

  [2] Symbolic knowledge extraction from trained neural networks: A sound approach, A. d'Avila Garcez, K. Broda, D. Gabbay, AIJ, 2001.

CILP Extraction Algorithm (discrete case)



b, c → x
a, c → x
a, b → x

2(a, b, c) → x

a → x
b → x
c → x

1 (a, b, c) → x

THEOREM: CILP rule extraction is sound

Challenge: efficient extraction of sound, readable knowledge from large-scale networks (100's of neurons; 1000's of connections)

TREPAN

Extracts decision trees from trained neural networks:

- Treats neural net as black-box (oracle) from which to query for input/output patterns
- Samples data from the training set or synthetic data to generate examples for the decision tree training
- Simplifies the rules in the trained decision tree into MofN rules

Extracting tree-structured representations of trained networks, Mark W. Craven and Jude W. Shavlik, NIPS 1995

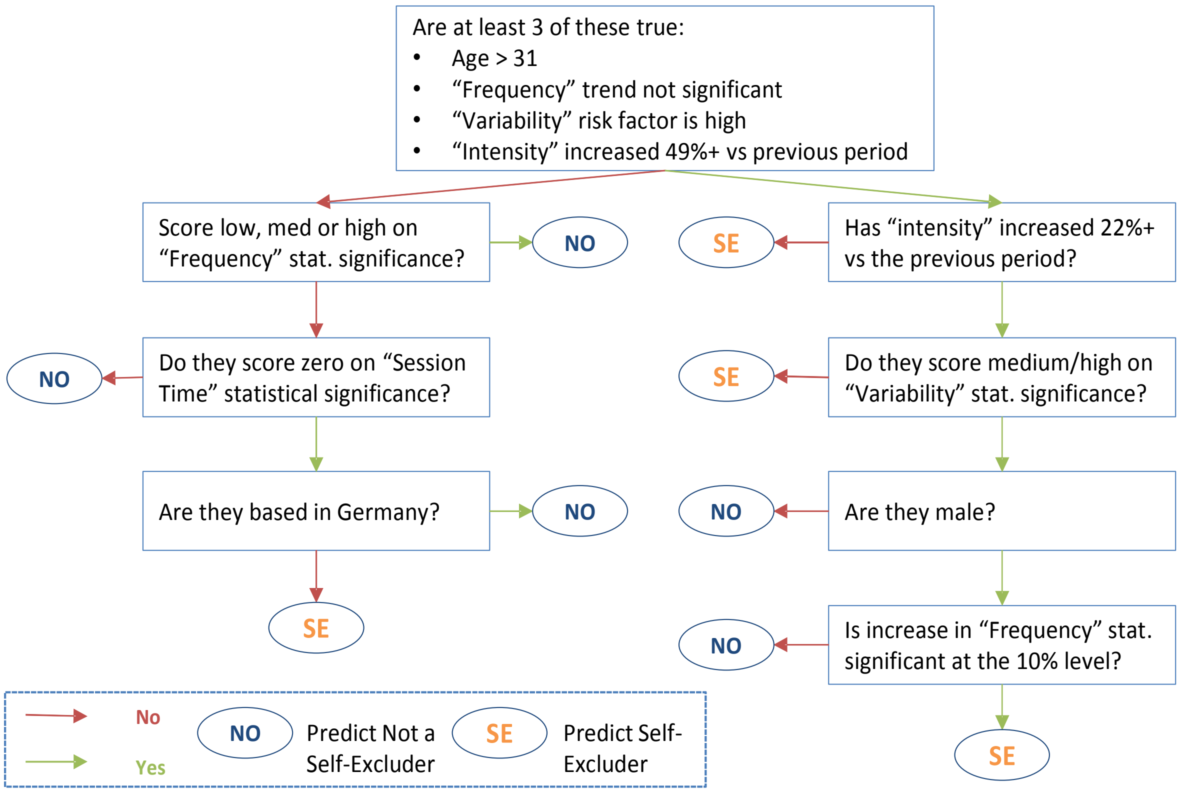
Recent application: Reducing harm from gambling

- 2014-16 EPSRC/InnovateUK project with BetBuddy ltd.
- Trained a neural net to predict whether someone should self-exclude from the game based on transaction data: frequency of play, betting intensity, variation, etc. (altogether some 25 markers)
- Used self-exclusion as a proxy for potential harm (avoids use of much more complex model of addiction)

Reducing harm from gambling

- Neural nets and Random Forests performed considerably better than logistic regression and Bayesian nets
- BetBuddy ltd. system is required to provide explanation to the regulator, gambling operator and to the player!
- Extracted decision tree can help debug the system and improve results too: "Are they based in Germany?"

TREPAN variations:



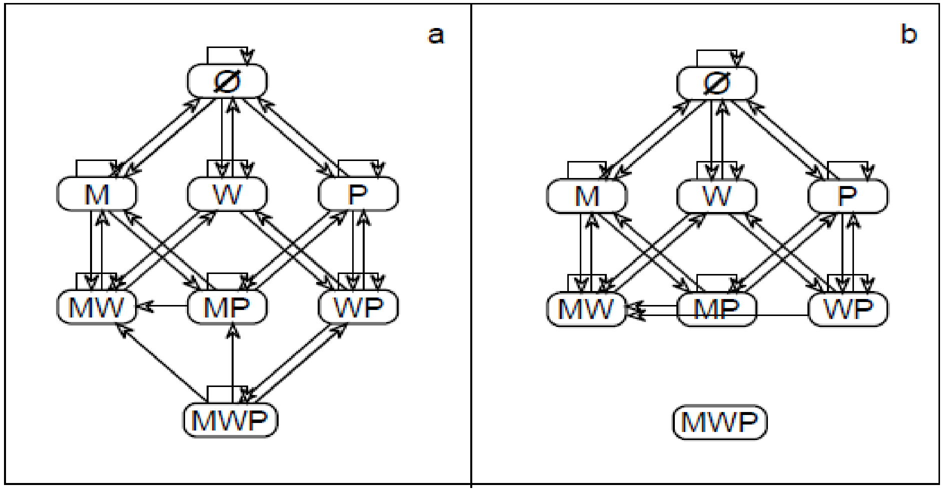
C. Percy, A. S. d'Avila Garcez, S. Dragicevic, M. Franca, G. Slabaugh and T. Weyde. The Need for Knowledge Extraction: Understanding Harmful Gambling Behavior with Neural Networks, In Proc. ECAI 2016, The Hague, September 2016.

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

Recurrent networks

- Extraction of state transition diagrams...

  CrMeth = M (level of methane is critical)
  HiWat = W (level of water is high)
  PumpOn = P (pump is turned on)

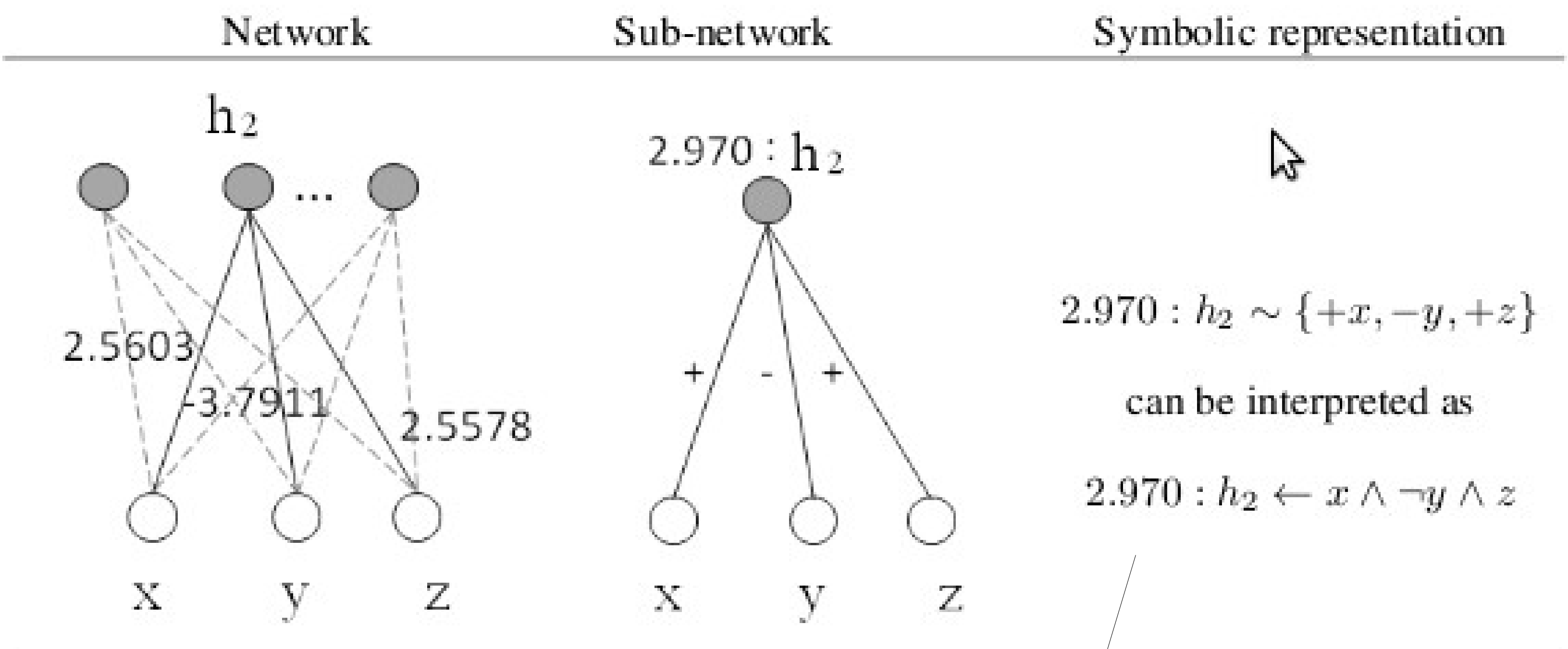


Extracting Automata from Recurrent Neural Networks Using Queries and Counterexamples, Gail Weiss, Yoav Goldberg, Eran Yahav, 2017
https://arxiv.org/abs/1711.09576

Learning and Representing Temporal Knowledge in Recurrent Networks, Rafael V. Borges, Artur d'Avila Garcez, Luis C. Lamb, IEEE TNNLS, 2011

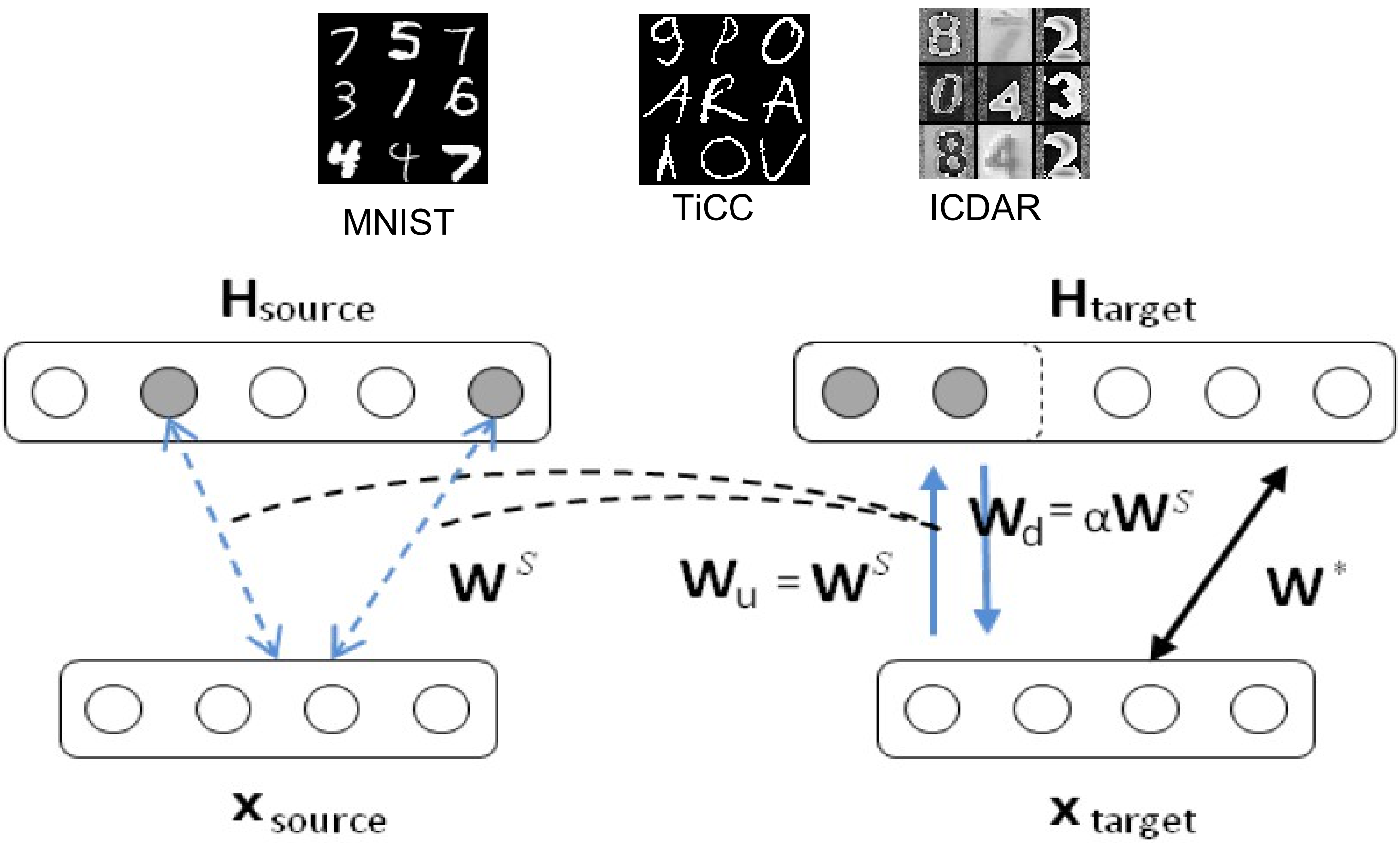
Extraction from RBMs and DBN

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



Each rule has a confidence value $\sum||w||/n$

Transfer Learning



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

Probabilistic MofN

- We can improve the accuracy of rules extracted from RBMs by extracting MofN rules
- Search values for M given extracted rules, e.g. M=0,1,2,3 in

$$2.970 : h_2 \leftarrow M \text{ of } \{x, \sim y, z\}$$

Extracting M of N Rules from Restricted Boltzmann Machines, Simon Odense and Artur S. d'Avila Garcez, ICANN 2017

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)

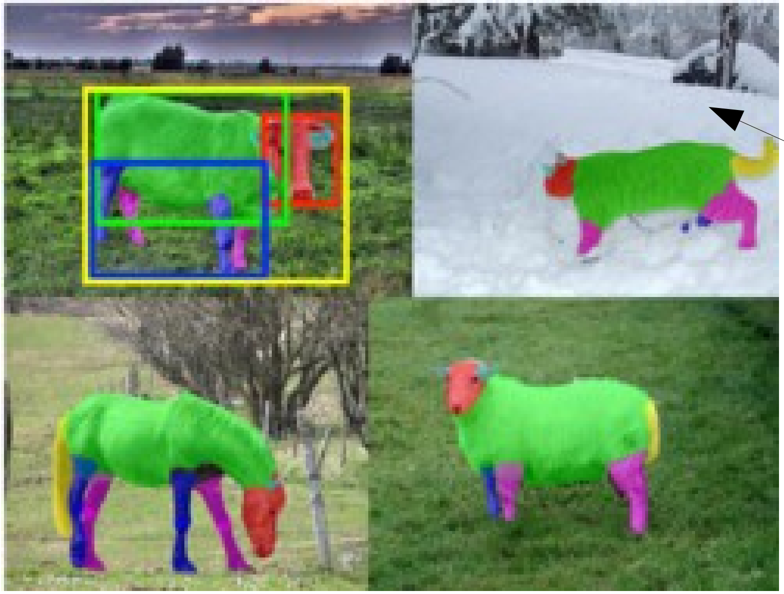
Given a picture extract a graph that describes its semantic content

Normally, every cat has a tail

Q. Get me the red thing next to the sheep

A. The horse's muzzle? Yes.

$$\forall xy(\mathsf{partOf}(x,y) \rightarrow \neg\mathsf{partOf}(y,x))$$



Make sure your system does not distinguish cats from leopards 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...

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

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 a KB and groundings, LTN calculates a grounding for the entire KB compositionally in the "usual ways"...

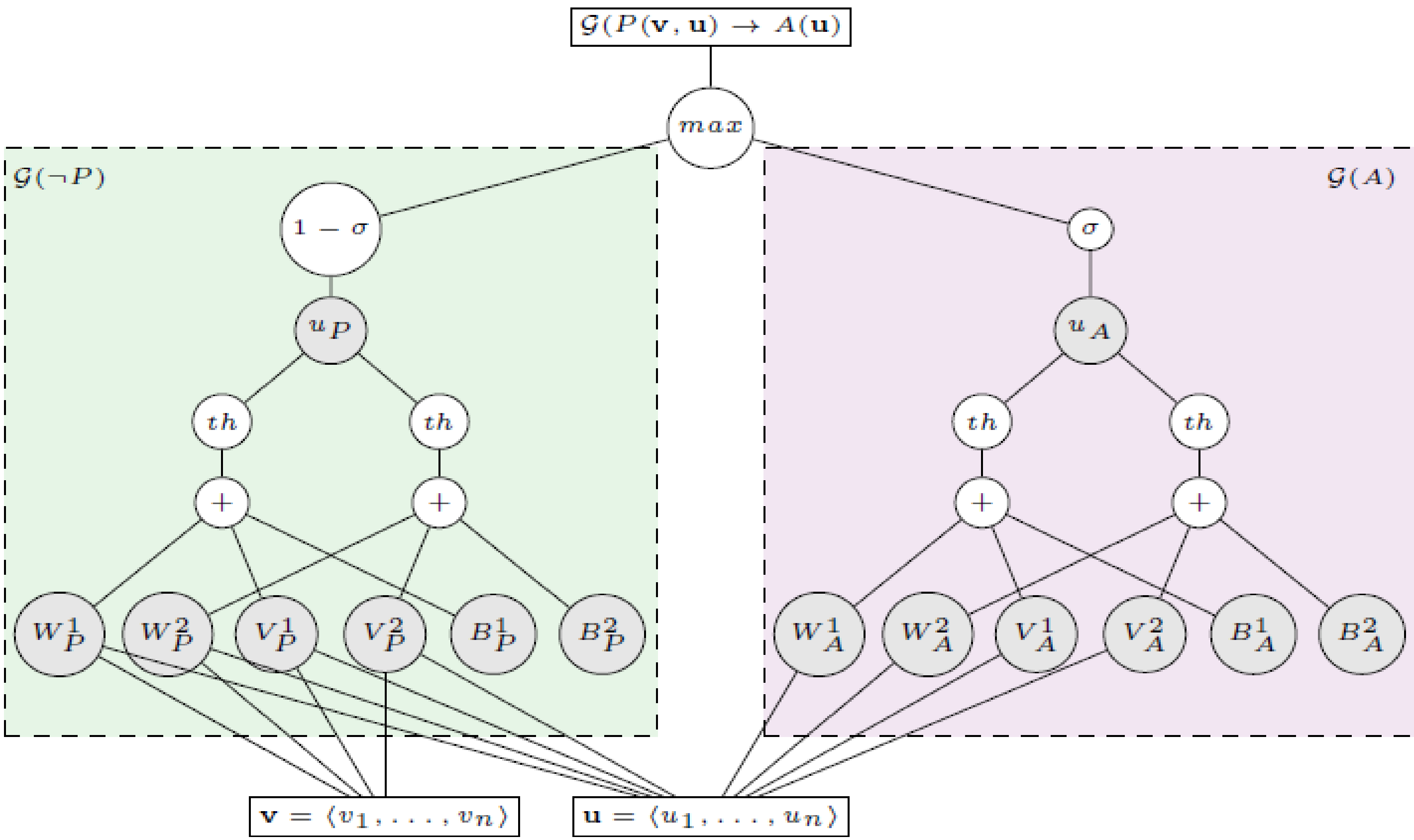


Fig. 1. Tensor net for $P(x, y) \rightarrow A(y)$, with $\mathcal{G}(x) = \mathbf{v}$ and $\mathcal{G}(y) = \mathbf{u}$ and $k = 2$.

The Tensor Network...

$$\mathcal{G}(f)(\mathbf{v}_1,\ldots,\mathbf{v}_m) = M_f\mathbf{v} + N_f$$

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

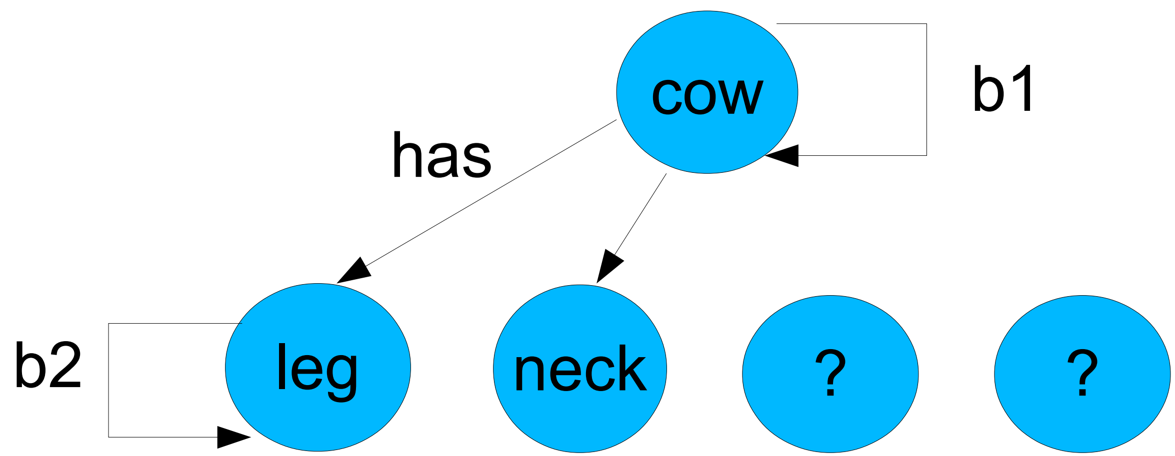
$$\mathcal{G}^* = \operatorname*{argmin}_{\hat{\mathcal{G}}\subseteq\mathcal{G}\in\mathbb{G}} \sum_{\langle[v,w],\phi(\mathbf{t})\rangle\in\mathcal{K}_0} Loss(\mathcal{G},\langle[v,w],\phi(\mathbf{t})\rangle)$$

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

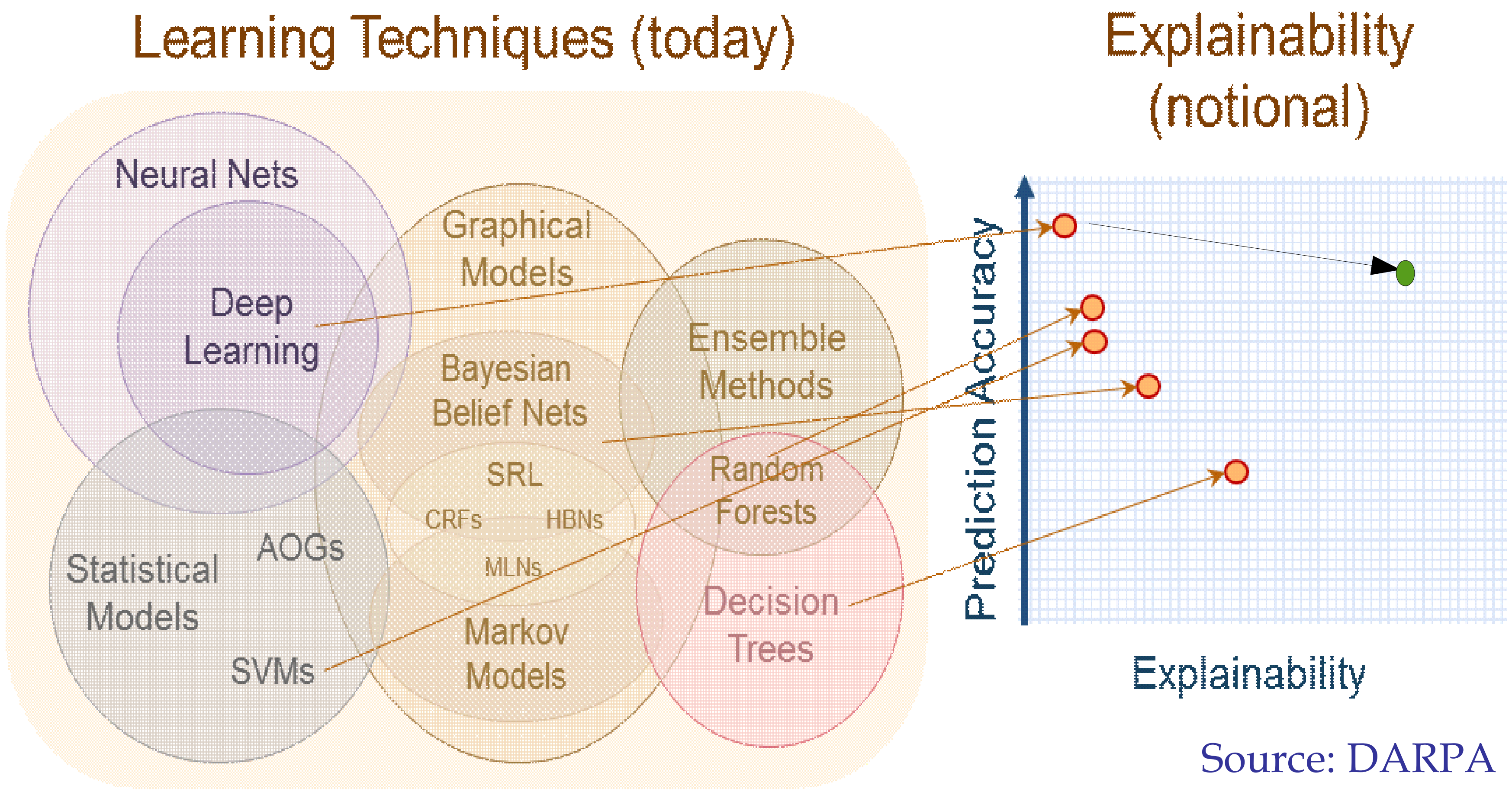
I. Donadello, L. Serafini and A. S. d'Avila Garcez. Logic Tensor Networks for Semantic Image Interpretation. In Proc. IJCAI'17, Melbourne, Australia, Aug 2017.

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

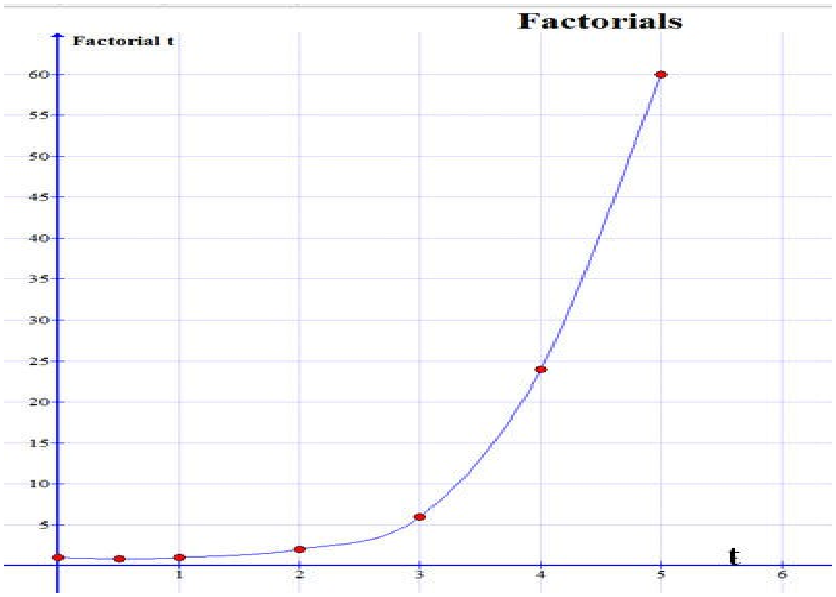


Source: DARPA

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

Deep Learning extrapolation

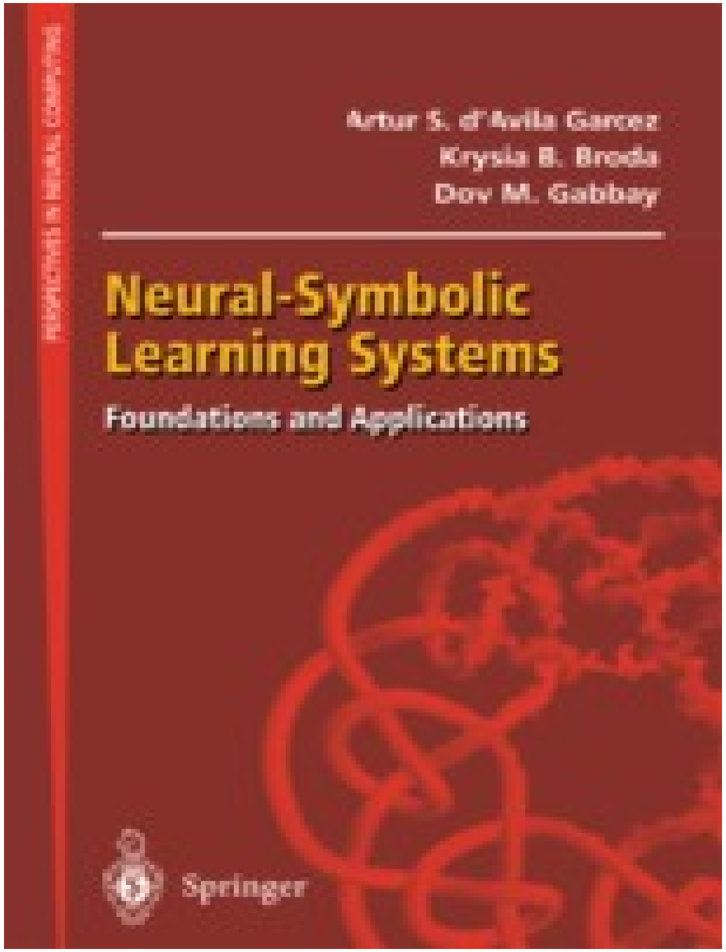
- We want good classification and prediction but also useful descriptions... e.g.: learning factorial (*n!*)

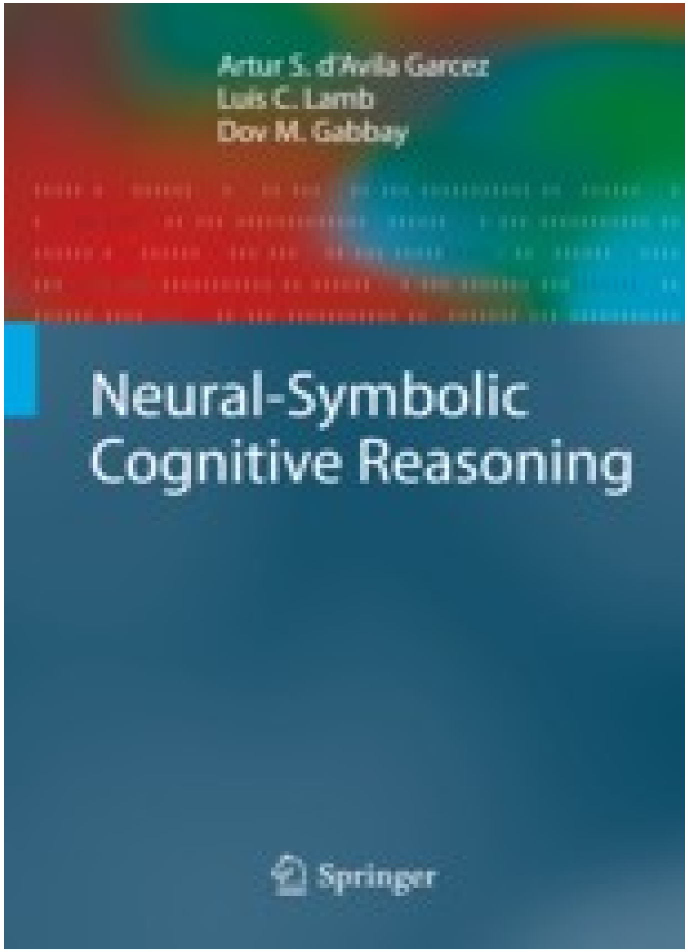


$$
\begin{gathered}
0! = 1 \\
1! = 1 \\
2! = 1 \cdot 2 = 2 \\
3! = 1 \cdot 2 \cdot 3 = 6 \\
4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24 \\
5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120 \\
6! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = 720
\end{gathered}
$$

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

For more information...





Related Work

Compare and contrast with Markov Logic Nets (MLNs), Inductive Logic Programming ILP-based approaches (e.g. ProbLog), Probabilistic Programming (WebPPL), lifted statistical relational AI...

See also:

Distillation as a Defense to Adversarial Perturbations against Deep Neural Networks, Nicolas Papernot, Patrick McDaniel, Xi Wu, Somesh Jha, Ananthram Swami, https://arxiv.org/abs/1511.04508

Ethical issues

Recall our extracted decision tree: Are they male? Yes/No

This is apparently illegal; gender cannot be a feature of the decision

Much recent work on “which features to keep out so that ML system is ethical?”

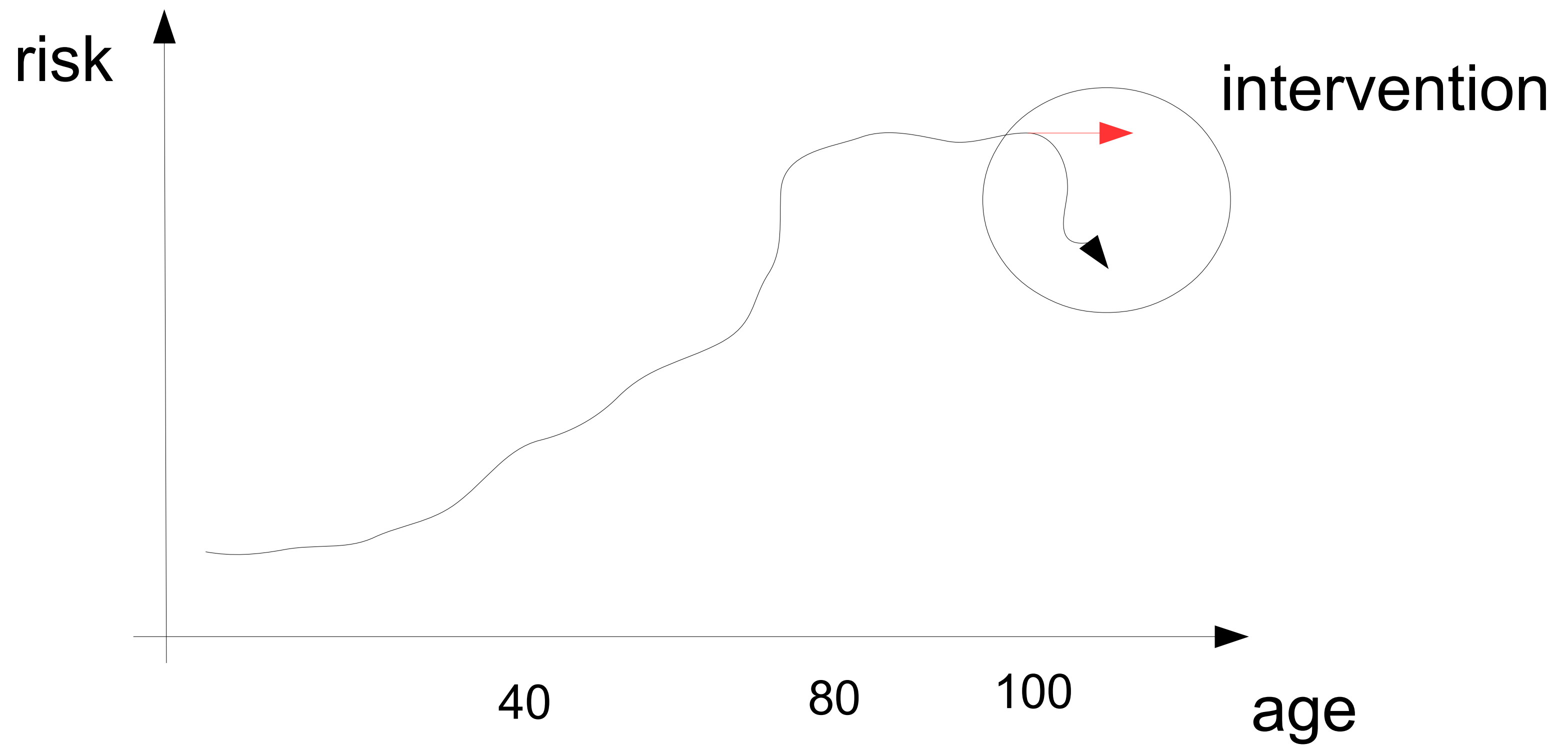
This is the wrong question... there are many unknown proxies in the data

Make system interpretable instead and decide on whether or not to intervene!

c.f. Rich Caruana's NIPS 2017 talks

How to intervene

- E.g. in healthcare, this may depend on whether you're the hospital or the insurance company
- Suppose this is your interpretable model:



Challenges

- Extraction from CNNs... c.f. Relating Input Concepts to Convolutional Neural Network Decisions, Ning Xie, Md Kamruzzaman Sarker, Derek Doran, Pascal Hitzler, Michael Raymer, NIPS workshop, 2017
- Nothing for LSTMs, GRUs other than visualizations c.f. On the memory properties of recurrent neural models, Jack Russell, Artur d'Avila Garcez and Emmanouil Benetos, IEEE IJCNN 2017
- Extraction of FOL from neural nets at different levels of abstraction (requires modularity)
- Distilling video/game analysis e.g. AlphaZero (may try and explain an instance, e.g. long-term sacrifices, but not the entire model)

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