Combining learning and reasoning: new challenges for knowledge graphs

Frank van Harmelen
Vrije Universiteit Amsterdam
Let’s first celebrate our successes!
Just some of the use cases

- **Google** = meaningful **search**
- **NXP** = data **integration**
- **BBC** = content **re-use**
- **Amazon** = product **search**
- **data.gov** = data-**publishing**
Knowledge Graphs in 4 principles

1. Give all things a name

2. Make a graph of relations between the things
   This makes a **Giant Graph**

3. Make sure all names are URIs
   This makes a **Giant Global Graph**

4. Add semantics (= predictable inference)
   This makes a **Giant Global Knowledge Graph**
P1. Give all things a name

“Now! That should clear up a few things around here!”
P3. The names are addresses on the Web

On the Web, anybody can say anything about anything

\[
\langle x \rangle \text{ IsOfTyp}\langle T \rangle
\]

different owners & locations
P3. The names are addresses on the Web
P1+P2+P3 = Giant Global Graph
P4. explicit & formal semantics

- assign types to things
- assign types to relations
- organise types in a hierarchy
- impose constraints on possible interpretations
± 45-100 billion facts & rules
But the eyes of the world are elsewhere...
Supervised learning
Unsupervised learning
Reinforcement learning

Robot Motor Skill Coordination with EM-based Reinforcement Learning

Petar Kormushev, Sylvain Calinon, and Darwin G. Caldwell

Italian Institute of Technology
A stairway?

Symbolic
Knowledge
Logic
Reasoning

Connectionist
Data
Statistics
Learning
A pendulum!

Connectionist
Data
Statistics
Learning

Symbolic
Knowledge
Logic
Reasoning

A GOOD JOB.
So let’s compare
Representation
Reasoning

- Model-based
- 13,000 variables

facts \rightarrow logicbase \rightarrow reasoning engine \rightarrow answers

Function-based
\[
\left\{ P \lor (Q \lor \neg R), P \rightarrow \neg R, Q \rightarrow \neg R \right\} \vdash \neg R
\]

1. \[ P \lor (Q \lor \neg R) \checkmark \] Ass
2. \[ P \rightarrow \neg R \checkmark \] Ass
3. \[ Q \rightarrow \neg R \checkmark \] Ass
4. \[ \neg R \rightarrow \neg R \] \neg Conc
5. \[ P \quad Q \lor \neg R \] \[ 1 \lor \text{Elim} \]
6. \[ \neg P \quad \neg R \] \[ 2 \rightarrow \text{Elim} \]
7. \[ \otimes \quad \otimes \] \[ 5 \lor \text{Elim} \]
   \[ 5,6 \quad 4,6 \]
8. \[ \otimes \quad \otimes \] \[ 3 \rightarrow \text{Elim} \]
   \[ 7,8 \quad 4,8 \]

\[
\tilde{L}^A(\theta, \phi; x^{(i)}) = \frac{1}{L} \sum_{l=1}^{L} \log p_\theta(x^{(i)}, z^{(i,l)}) - \log q_\phi(z^{(i,l)}|x^{(i)})
\]

where \[ z^{(i,l)} = g_\phi(\epsilon^{(i,l)}, x^{(i)}) \quad \text{and} \quad \epsilon^{(l)} \sim p(\epsilon) \]
## Strengths & Weaknesses

<table>
<thead>
<tr>
<th></th>
<th>Symbolic</th>
<th>Connectionist</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Construction</strong></td>
<td>Human effort</td>
<td>Data hunger</td>
</tr>
<tr>
<td><strong>Scaleable</strong></td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td><strong>Explainable</strong></td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td><strong>Generalisable</strong></td>
<td>Performance cliff</td>
<td>Performance cliff</td>
</tr>
</tbody>
</table>
## Strengths & Weaknesses

<table>
<thead>
<tr>
<th></th>
<th>Symbolic</th>
<th>Connectionist</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Construction</strong></td>
<td>Human effort</td>
<td>Data hunger</td>
</tr>
<tr>
<td><strong>Scaleable</strong></td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td><strong>Explainable</strong></td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td><strong>Generalisable</strong></td>
<td>Performance cliff</td>
<td>Performance cliff</td>
</tr>
</tbody>
</table>

40 years of effort, 10,000 updates every years
## Strengths & Weaknesses

<table>
<thead>
<tr>
<th></th>
<th>Symbolic</th>
<th>Connectionist</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Construction</strong></td>
<td>Human effort</td>
<td>Data hunger</td>
</tr>
<tr>
<td><strong>Scalable</strong></td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td><strong>Explainable</strong></td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td><strong>Generalisable</strong></td>
<td>Performance cliff</td>
<td>Performance cliff</td>
</tr>
</tbody>
</table>

10M training samples

4.8M training games
<table>
<thead>
<tr>
<th></th>
<th>Symbolic</th>
<th>Connectionist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construction</td>
<td>Human effort</td>
<td>Data hunger</td>
</tr>
<tr>
<td>Scaleable</td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td>Explainable</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Generalisable</td>
<td>Performance cliff</td>
<td>Performance cliff</td>
</tr>
</tbody>
</table>

- worse with **more** data
- worse with **less** data
## Strengths & Weaknesses

<table>
<thead>
<tr>
<th></th>
<th>Symbolic</th>
<th>Connectionist</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Construction</strong></td>
<td>Human effort</td>
<td>Data hunger</td>
</tr>
<tr>
<td><strong>Scaleable</strong></td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td><strong>Explainable</strong></td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td><strong>Generalisable</strong></td>
<td>Performance cliff</td>
<td>Performance cliff</td>
</tr>
</tbody>
</table>

![Image of panda and gibbon with confidence scores]
### Strengths & Weaknesses

<table>
<thead>
<tr>
<th></th>
<th>Symbolic</th>
<th>Connectionist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construction</td>
<td>Human effort</td>
<td>Data hunger</td>
</tr>
<tr>
<td>Scaleable</td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td>Explainable</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Generalisable</td>
<td>Performance cliff</td>
<td>Performance cliff</td>
</tr>
</tbody>
</table>
## Strengths & Weaknesses

<table>
<thead>
<tr>
<th></th>
<th>Symbolic</th>
<th>Connectionist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construction</td>
<td>Human effort</td>
<td>Data hunger</td>
</tr>
<tr>
<td>Scaleable</td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td>Explainable</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Generalisable</td>
<td><strong>Performance cliff</strong></td>
<td><strong>Performance cliff</strong></td>
</tr>
</tbody>
</table>

![Graph showing quality and generality](image)

↑ quality

generality →

- **Symbolic**
  - Construction: Human effort
  - Scaleable: +/-
  - Explainable: +
  - Generalisable: **Performance cliff**

- **Connectionist**
  - Construction: Data hunger
  - Scaleable: +/-
  - Explainable: -
  - Generalisable: **Performance cliff**
## Strengths & Weaknesses

<table>
<thead>
<tr>
<th></th>
<th>Symbolic</th>
<th>Connectionist</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Construction</strong></td>
<td>Human effort</td>
<td>Data hunger</td>
</tr>
<tr>
<td><strong>Scaleable</strong></td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td><strong>Explainable</strong></td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td><strong>Generalisable</strong></td>
<td>Performance cliff</td>
<td><strong>Performance cliff</strong></td>
</tr>
</tbody>
</table>

Class: 793  
Label: n04209133 *(shower cap)*  
Certainty: 99.7%
Can we get them to collaborate?
Plan:

loose coupling + compositional patterns

\[
sym \xrightarrow{KR} sym
\]

\[
data \xrightarrow{ML} data
\]
Simplest unification: Learning on symbols

- Inductive Logic Programming
- Probabilistic Soft Logic
- Markov Logic Networks

Ignored by too many of us

All of these need Knowledge Engineering
From data to symbols

- Ontology learning from text
  - Learning class-instance distinctions from text
  - Knowledge graph learning from text
- Conceptual spaces

Also covers many “classical” learning algorithms:
- Decision tree learning
- Rule learning

All of these need Knowledge Engineering
From symbols to data and back again

Knowledge Graph completion

sym → ML → data → ML → sym
From symbols to data and back again

Knowledge Graph completion

Entity Space

Relation Space of $r$
From symbols to data and back again

Knowledge Graph completion

from a graph representation ... to real vector representation

embedding algorithm

prediction algorithm
Generating an explanation

data -> ML -> sym -> KR -> sym

crown

wears

queen
Ranking hypotheses
(≈ explaining why not)
Symbolic prior

\[ \forall x, y \ \text{chair}(x) \land \text{partOf}(y, x) \rightarrow \text{cushion}(y) \lor \text{armRest}(y) \]

\[ \text{P(cushion|chair)} \gg \text{P(flower|chair)} \]
Learning intermediate abstractions

Example: Reinforcement learning for spatial navigation
Learning intermediate abstractions

Example: Reinforcement learning for spatial navigation

Unsupervised learning of mapping from high-dimensional raw input to low-dimensional symbolic representation

Compositionally structured symbolic representation

Reinforcement learning of mapping from symbolic representation to action with maximum expected reward over time

Faster adaptation to changes; Better transfer learning
Learning intermediate abstractions

Example:
Try to learn early symptoms of CRC from GP data: (symptoms, life style, diagnoses, drugs)
1. Raw data \(\rightarrow\) no meaningful signal
2. Raw data + KG = abstracted data \(\rightarrow\) significant results
Learning intermediate abstractions

data $\rightarrow$ ML $\rightarrow$ sym $\rightarrow$ KR $\rightarrow$ sym

Tree search

Monte Carlo search
Learning ("imitating?") to reason

![Diagram](image)

<table>
<thead>
<tr>
<th>Test Dataset</th>
<th>Hop 0</th>
<th>Hop 1</th>
<th>Hop 2</th>
<th>Hop 3</th>
<th>Hop 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F</td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>Linked Data(^a)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>80</td>
<td>99</td>
</tr>
<tr>
<td>Linked Data(^b)</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>82</td>
<td>91</td>
</tr>
<tr>
<td>OWL-Centric (^c)</td>
<td>19</td>
<td>5</td>
<td>9</td>
<td>31</td>
<td>75</td>
</tr>
<tr>
<td>Synthetic</td>
<td>32</td>
<td>46</td>
<td>33</td>
<td>31</td>
<td>87</td>
</tr>
</tbody>
</table>

\(^a\) LemonUby Ontology
\(^b\) Agrovoc Ontology
\(^c\) Completely Different Domain
Learning ("imitating?") to reason

Reasoning transfers to other KG’s!
Knowledge-based auto-ML

- Algorithmic configuration
- Hyperparameter tuning
- Selection of training examples
Concluding remarks
Plan:

loose coupling
+
compositional patterns

\[
\begin{array}{c}
\text{sym} \rightarrow \text{KR} \rightarrow \text{sym} \\
\text{data} \rightarrow \text{ML} \rightarrow \text{data}
\end{array}
\]
1. The future of AI depends on combining
   Thinking fast and thinking slow, or
   Perception and cognition, or
   Learning and reasoning

2. Lots of useful results by “loose coupling”

3. Develop a theory of reusable components and compositional patterns

4. In the style of knowledge engineering & software engineering
1. **The future of AI** depends on combining
   - Thinking fast and thinking slow
   - Perception and cognition
   - Learning and reasoning

2. Lots of useful results by **“loose coupling”**

3. Develop a science of **reusable components** and **compositional patterns**

4. In the style of **knowledge engineering**
   - software engineering
   - “X” engineering