Explainable, Data-Efficient, Verifiable Representation Learning

Pasquale Minervini, UCL
p.minervini@ucl.ac.uk

ICL, Nov 27th
Symbolic and Sub-Symbolic AI
∀X, Y, Z : \textit{grandFather}(X, Y) \iff \textit{father}(X, Z), \textit{parent}(Z, Y)

Symbolic

Ontologies, First-Order Logic, Logic Programming, Knowledge Bases, Theorem Proving..
∀X, Y, Z : grandFather(X, Y) ⇐ father(X, Z), parent(Z, Y)

✓ Data-Efficient
✓ Interpretable and Explainable
✓ Easy to Incorporate Knowledge
✓ Verifiable

Symbolic

Ontologies, First-Order Logic, Logic Programming, Knowledge Bases, Theorem Proving..
Symbolic and Sub-Symbolic AI

Sub-Symbolic/Connectionist

Neural Representation Learning, (Deep) Latent Variable Models..

\[
f_1(f_2(\ldots f_n(\ldots))) = \begin{cases} 0.15 & \equiv \text{cat} \\ 0.85 & \equiv \text{dog} \end{cases}
\]

\[\forall X, Y, Z : \text{grandFather}(X, Y) \iff \text{father}(X, Z), \text{parent}(Z, Y)\]

✓ Data-Efficient
✓ Interpretable and Explainable
✓ Easy to Incorporate Knowledge
✓ Verifiable

Symbolic

Ontologies, First-Order Logic, Logic Programming, Knowledge Bases, Theorem Proving..
Symbolic and Sub-Symbolic AI

Sub-Symbolic/Connectionist

Neural Representation Learning, (Deep) Latent Variable Models..

\[
\begin{align*}
    f_1(f_2(\ldots f_n(\ldots))) &= \begin{cases} 
    0.15 \equiv \text{cat} \\
    0.85 \equiv \text{dog} 
\end{cases}
\end{align*}
\]

Noisy, Ambiguous, Sensory Data
Highly Parallel
High Predictive Accuracy

Symbolic

\[
\forall X, Y, Z: \text{grandFather}(X, Y) \iff \text{father}(X, Z), \text{parent}(Z, Y)
\]

Data-Efficient
Interpretable and Explainable
Easy to Incorporate Knowledge
Verifiable

Ontologies, First-Order Logic, Logic Programming, Knowledge Bases, Theorem Proving..
Symbolic and Sub-Symbolic AI

Sub-Symbolic/Connectionist

Neural Representation Learning, (Deep) Latent Variable Models..

\[
\begin{align*}
 f_1(f_2(\ldots f_n(\ldots))) &= \begin{cases} 
 0.15 \equiv \text{cat} \\
 0.85 \equiv \text{dog} 
\end{cases}
\end{align*}
\]

∀X, Y, Z : grandFather(X, Y) ⇐ father(X, Z), parent(Z, Y)

Symbolic

Data-Efficient
Interpretable and Explainable
Easy to Incorporate Knowledge
Verifiable

Ontologies, First-Order Logic, Logic Programming, Knowledge Bases, Theorem Proving..
Knowledge Graphs

Knowledge Graph — graph structured Knowledge Base, where knowledge is encoded by relationships between entities.
Knowledge Graphs

Knowledge Graph — graph structured Knowledge Base, where knowledge is encoded by relationships between entities.

In practice — set of subject-predicate-object triples, denoting a relationship of type predicate between subject and object.

<table>
<thead>
<tr>
<th>subject</th>
<th>predicate</th>
<th>object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barack Obama</td>
<td>was born in</td>
<td>Honolulu</td>
</tr>
<tr>
<td>Hawaii</td>
<td>has capital</td>
<td>Honolulu</td>
</tr>
<tr>
<td>Barack Obama</td>
<td>is politician of</td>
<td>United States</td>
</tr>
<tr>
<td>Hawaii</td>
<td>is located in</td>
<td>United States</td>
</tr>
<tr>
<td>Barack Obama</td>
<td>is married to</td>
<td>Michelle Obama</td>
</tr>
<tr>
<td>Michelle Obama</td>
<td>is a</td>
<td>Lawyer</td>
</tr>
<tr>
<td>Michelle Obama</td>
<td>lives in</td>
<td>United States</td>
</tr>
</tbody>
</table>
Knowledge Graphs

Imperial College London

Public university in London, England · 2.8 mi

Imperial College London is a public research university located in London. In 1851, Prince Albert built his vision of an area for culture, including the Victoria and Albert Museum, Natural History Museum, Royal Colleges, Royal Albert Hall, and the Imperial Institute.

Wikipedia

Address: South Kensington, London SW7 2AZ

Acceptance rate: 14.3% (2015)

Subsidiaries: Imperial College Business School, MORE
Knowledge Graphs

Imperial College London

Imperial College London is a public research university located in London. In 1851, Prince Albert built his vision of an area for culture, including the Victoria and Albert Museum, Natural History Museum, Royal Colleges, Royal Albert Hall, and the Imperial Institute.

Wikipedia

Address: South Kensington, London SW7 2AZ
Acceptance rate: 14.3% (2015)
Subsidiaries: Imperial College Business School, MORE
Link Prediction in Knowledge Graphs

- Malia Ann Obama
- Sasha Obama
- Barack Obama
- Michelle Obama
- Washington

Relationships:
- Malia Ann Obama is a parent of Sasha Obama.
- Barack Obama lives in Washington.
- Michelle Obama is a parent of Malia Ann Obama and Sasha Obama.
∀X, Y, Z:
married with(X, Y) ⇐
parent of(X, Z),
parent of(Y, Z)
Rule-Based Link Prediction

∀X, Y, Z:
married with(X, Y) ⇐
parent of(X, Z),
parent of(Y, Z)

✗ Not always true
✗ Hard to learn from data
✗ Hard to formalise for other modalities
Neural Link Prediction

- Malia Ann Obama
- Sasha Obama
- Barack Obama
- Michelle Obama
- Washington

parent of
lives in
Neural Link Prediction

\[ P(\text{BO married MO}) \propto \]

\[ f_{\text{married}}(\text{, , }) \]
Neural Link Prediction

\[ P(\text{BO married MO}) \propto f_{\text{married}}(\text{BO}, \text{MO}) \]

Learning Representations

\[ \mathcal{L}(\mathcal{G} | \Theta) = \sum_{(s,p,o) \in \mathcal{G}} \log \sigma \left( f_p(e_s, e_o) \right) + \sum_{(s,p,o) \notin \mathcal{G}} \log \left[ 1 - \sigma \left( f_p(e_s, e_o) \right) \right] \]
# Neural Link Prediction — Scoring Functions

The interaction between the *latent features* is defined by the scoring function $f(\cdot)$ — several variants in the literature:

<table>
<thead>
<tr>
<th>Models</th>
<th>Scoring Functions</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESCAL [Nickel et al. 2011]</td>
<td>$\mathbf{e}_s^T \mathbf{W}_p \mathbf{e}_o$</td>
<td>$\mathbf{W}_p \in \mathbb{R}^{k \times k}$</td>
</tr>
<tr>
<td>TransE [Bordes et al. 2013]</td>
<td>$- | \mathbf{e}_s + \mathbf{r}_p - \mathbf{e}_o |_p^2$</td>
<td>$\mathbf{r}_p \in \mathbb{R}^k$</td>
</tr>
<tr>
<td>DistMult [Yang et al. 2015]</td>
<td>$\langle \mathbf{e}_s, \mathbf{r}_p, \mathbf{e}_o \rangle$</td>
<td>$\mathbf{r}_p \in \mathbb{R}^k$</td>
</tr>
<tr>
<td>HolE [Nickel et al. 2016]</td>
<td>$\mathbf{r}_p^T \left( \mathcal{F}^{-1} \left[ \mathcal{F}[\mathbf{e}_s] \odot \mathcal{F}[\mathbf{e}_o] \right] \right)$</td>
<td>$\mathbf{r}_p \in \mathbb{R}^k$</td>
</tr>
<tr>
<td>ComplEx [Trouillon et al. 2016]</td>
<td>$\text{Re} \left( \langle \mathbf{e}_s, \mathbf{r}_p, \overline{\mathbf{e}_o} \rangle \right)$</td>
<td>$\mathbf{r}_p \in \mathbb{C}^k$</td>
</tr>
<tr>
<td>ConvE [Dettmers et al. 2017]</td>
<td>$f \left( \text{vec} \left( f \left( [\mathbf{e}_s; \overline{\mathbf{r}_p}] \ast \omega \right) \right) \mathbf{W} \right) \mathbf{e}_o$</td>
<td>$\mathbf{r}_p \in \mathbb{R}^k, \mathbf{W} \in \mathbb{R}^{c \times k}$</td>
</tr>
</tbody>
</table>
# Neural Link Prediction — Scoring Functions

The interaction between the *latent features* is defined by the scoring function $f(\cdot)$ — several variants in the literature:

<table>
<thead>
<tr>
<th>Models</th>
<th>Scoring Functions</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESCAL [Nickel et al. 2011]</td>
<td>$e_s^T W_p e_o$</td>
<td>$W_p \in \mathbb{R}^{k \times k}$</td>
</tr>
<tr>
<td>TransE [Bordes et al. 2013]</td>
<td>$- | e_s + r_p - e_o |_p^2$</td>
<td>$r_p \in \mathbb{R}^k$</td>
</tr>
<tr>
<td>DistMult [Yang et al. 2015]</td>
<td>$\langle e_s, r_p, e_o \rangle$</td>
<td>$r_p \in \mathbb{R}^k$</td>
</tr>
<tr>
<td>HolE [Nickel et al. 2016]</td>
<td>$r_p^T \left( \mathcal{F}^{-1} \left[ \mathcal{F}[e_s] \odot \mathcal{F}[e_o] \right] \right)$</td>
<td>$r_p \in \mathbb{R}^k$</td>
</tr>
<tr>
<td>ComplEx [Trouillon et al. 2016]</td>
<td>$\text{Re} \left( \langle e_s, r_p, \bar{e}_o \rangle \right)$</td>
<td>$r_p \in \mathbb{C}^k$</td>
</tr>
<tr>
<td>ConvE [Dettmers et al. 2017]</td>
<td>$f \left( \text{vec} \left( f \left[ e_s, r_p \right] * \omega \right) \right) W e_o$</td>
<td>$r_p \in \mathbb{R}^k, W \in \mathbb{R}^{c \times k}$</td>
</tr>
</tbody>
</table>
Neural Link Prediction — Scoring Functions

The interaction between the *latent features* is defined by the scoring function $f(\cdot)$ — several variants in the literature:

<table>
<thead>
<tr>
<th>Models</th>
<th>Scoring Functions</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESCAL [Nickel et al. 2011]</td>
<td>$e_s^T W_p e_o$</td>
<td>$W_p \in \mathbb{R}^{k \times k}$</td>
</tr>
<tr>
<td>TransE [Bordes et al. 2013]</td>
<td>$- | e_s + r_p - e_o |^2_p$</td>
<td>$r_p \in \mathbb{R}^k$</td>
</tr>
<tr>
<td>DistMult [Yang et al. 2015]</td>
<td>$\langle e_s, r_p, e_o \rangle$</td>
<td>$r_p \in \mathbb{R}^k$</td>
</tr>
<tr>
<td>HolE [Nickel et al. 2016]</td>
<td>$r_p^T \left( \mathcal{F}^{-1} \left[ \mathcal{F}[e_s] \odot \mathcal{F}[e_o] \right] \right)$</td>
<td>$r_p \in \mathbb{R}^k$</td>
</tr>
<tr>
<td>ComplEx [Trouillon et al. 2016]</td>
<td>$\text{Re} \left( \langle e_s, r_p, \overline{e_o} \rangle \right)$</td>
<td>$r_p \in \mathbb{C}^k$</td>
</tr>
<tr>
<td>ConvE [Dettmers et al. 2017]</td>
<td>$f \left( \text{vec} \left( f \left( [e_s; r_p] \ast \omega \right) \right) W \right) e_o$</td>
<td>$r_p \in \mathbb{R}^k, W \in \mathbb{R}^{c \times k}$</td>
</tr>
</tbody>
</table>
Neural Link Prediction — Scoring Functions

The interaction between the *latent features* is defined by the scoring function \( f(\cdot) \) — several variants in the literature:

<table>
<thead>
<tr>
<th>Models</th>
<th>Scoring Functions</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESCAL [Nickel et al. 2011]</td>
<td>( \mathbf{e}_s^T \mathbf{W}_p \mathbf{e}_o )</td>
<td>( \mathbf{W}_p \in \mathbb{R}^{k \times k} )</td>
</tr>
<tr>
<td>TransE [Bordes et al. 2013]</td>
<td>(- | \mathbf{e}_s + \mathbf{r}_p - \mathbf{e}_o |^2_p)</td>
<td>( \mathbf{r}_p \in \mathbb{R}^k )</td>
</tr>
<tr>
<td>DistMult [Yang et al. 2015]</td>
<td>( \langle \mathbf{e}_s, \mathbf{r}_p, \mathbf{e}_o \rangle )</td>
<td>( \mathbf{r}_p \in \mathbb{R}^k )</td>
</tr>
<tr>
<td>HolE [Nickel et al. 2016]</td>
<td>( \mathbf{r}_p^T \left( \mathcal{F}^{-1} \left[ \mathcal{F}[\mathbf{e}_s] \odot \mathcal{F}[\mathbf{e}_o] \right] \right) )</td>
<td>( \mathbf{r}_p \in \mathbb{R}^k )</td>
</tr>
<tr>
<td>ComplEx [Trouillon et al. 2016]</td>
<td>( \text{Re} \left( \langle \mathbf{e}_s, \mathbf{r}_p, \overline{\mathbf{e}_o} \rangle \right) )</td>
<td>( \mathbf{r}_p \in \mathbb{C}^k )</td>
</tr>
<tr>
<td>ConvE [Dettmers et al. 2017]</td>
<td>( f \left( \text{vec} \left( f \left( [\overline{\mathbf{e}_s}; \overline{\mathbf{r}_p}]^\ast \omega \right) \right) \mathbf{W} \right) \mathbf{e}_o )</td>
<td>( \mathbf{r}_p \in \mathbb{R}^k, \mathbf{W} \in \mathbb{R}^{c \times k} )</td>
</tr>
</tbody>
</table>
Neural Link Prediction — Accuracy

Evaluation Metrics — Area Under the Precision-Recall Curve (AUC-PR), Mean Reciprocal Rank (MRR), Hits@k. In MRR and Hits@k, for each test triple:

- Modify its subject with all the entities in the Knowledge Graph,
- Score all the triple variants, and compute the rank of the original test triple,
- Repeat for the object.

$$MRR = \frac{1}{|\mathcal{T}|} \sum_{i=1}^{|\mathcal{T}|} \frac{1}{\text{rank}_i}, \quad \text{HITS} @ k = \frac{|\{\text{rank}_i \leq 10\}|}{|\mathcal{T}|}$$

From [Lacroix et al. ICML 2018]

<table>
<thead>
<tr>
<th>Model</th>
<th>WN18</th>
<th>WN18RR</th>
<th>FB15K</th>
<th>FB15K-237</th>
<th>YAGO3-10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MRR</td>
<td>H@10</td>
<td>MRR</td>
<td>H@10</td>
<td>MRR</td>
</tr>
<tr>
<td>Reciprocal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP-FRO</td>
<td>0.95</td>
<td>0.95</td>
<td>0.46</td>
<td>0.48</td>
<td>0.86</td>
</tr>
<tr>
<td>CP-N3</td>
<td>0.95</td>
<td>0.96</td>
<td>0.47</td>
<td>0.54</td>
<td>0.86</td>
</tr>
<tr>
<td>ComplEx-FRO</td>
<td>0.95</td>
<td>0.96</td>
<td>0.47</td>
<td>0.54</td>
<td>0.86</td>
</tr>
<tr>
<td>ComplEx-N3</td>
<td>0.95</td>
<td>0.96</td>
<td>0.48</td>
<td>0.57</td>
<td>0.86</td>
</tr>
</tbody>
</table>
Convolutional 2D Knowledge Graph Embeddings

Idea — use ideas from computer vision for modeling the interactions between latent features.

Subject Embedding

Predicate Embedding
**Convolutional 2D Knowledge Graph Embeddings**

**Idea** — use ideas from *computer vision* for modeling the interactions between latent features.
Convolutional 2D Knowledge Graph Embeddings

Idea — use ideas from computer vision for modeling the interactions between latent features.
Convolutional 2D Knowledge Graph Embeddings

**Idea** — use ideas from *computer vision* for modeling the interactions between latent features.

![Diagram of Convolutional 2D Knowledge Graph Embeddings]

- **Scalable**
- **State-of-the-art Results**

[AAAI 2018]
Convolutional 2D Knowledge Graph Embeddings

**Idea** — use ideas from *computer vision* for modeling the interactions between latent features.

- Efficiency via parameter sharing
- State-of-the-art Results

[AAAI 2018]
Interpreting Knowledge Graph Embeddings

Quite hard to understand the *semantics* of the learned representations..

<table>
<thead>
<tr>
<th></th>
<th>Real Part</th>
<th>Imaginary Part</th>
</tr>
</thead>
<tbody>
<tr>
<td>hypernym</td>
<td>1.0</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>-3.1</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>-3.0</td>
</tr>
<tr>
<td></td>
<td>-2.7</td>
<td>-3.0</td>
</tr>
</tbody>
</table>

[Minervini et al. ECML 2017]
Interpreting Knowledge Graph Embeddings

Quite hard to understand the *semantics* of the learned representations..

.. but we can use their geometric relationships for identifying — and *incorporating* — semantic relationships between them.

[Minervini et al. ECML 2017]
Regularising Knowledge Graph Embeddings

Quite hard to understand the *semantics* of the learned representations...

.. but we can use their geometric relationships for identifying — and *incorporating* — semantic relationships between them.

[Minervini et al. ECML 2017]
Regularising Knowledge Graph Embeddings

Quite hard to understand the *semantics* of the learned representations.

\[
\begin{array}{c|cccc|cccc}
 & \text{Real Part} & & & \text{Imaginary Part} & & & \\
\hline
\text{hypernym} & 1.0 & 3.0 & -3.1 & 2.5 & -2.7 & 3.2 & 2.9 & 1.7 & -3.0 & -3.0 \\
\text{hyponym} & 1.0 & 3.1 & -3.1 & 2.6 & -2.7 & -3.4 & -2.8 & -1.7 & 2.9 & 3.0 \\
\text{instance hypernym} & -1.1 & -2.8 & 1.6 & 2.7 & -2.5 & 3.0 & -2.6 & 2.6 & -1.1 & -2.8 \\
\text{instance hyponym} & -1.0 & -2.9 & 1.5 & 2.9 & -2.4 & -2.9 & 2.8 & -2.6 & 1.1 & 2.8 \\
\text{part of} & -2.4 & 3.2 & 2.7 & -1.5 & 3.0 & -2.4 & -0.6 & -2.6 & 2.9 & -1.9 \\
\text{has part} & -2.5 & 3.2 & 2.9 & -1.5 & 3.0 & 2.4 & 0.7 & 2.8 & -3.0 & 1.9 \\
\end{array}
\]

.. but we can use their geometric relationships for identifying — and *incorporating* — semantic relationships between them.

\[\times \text{ is } a(x, y) \land \text{ is } a(y, z) \Rightarrow \text{ is } a(x, z)\]

[Minervini et al. ECML 2017]
Incorporating Background Knowledge via Adversarial Training

Idea — adversarial training process where, iteratively:

[Minervini et al. UAI 2017]
Incorporating Background Knowledge via Adversarial Training

Idea — adversarial training process where, iteratively:

• An adversary searches for inputs where the model violates constraints
e.g. $x, y, z$ such that

$$\text{is } a(x, y) \land \text{is } a(y, z) \land \lnot \text{is } a(x, z)$$
Incorporating Background Knowledge via Adversarial Training

Idea — adversarial training process where, iteratively:

- An adversary searches for inputs where the model violates constraints
- The model is regularised to correct such violations.

[Minervini et al. UAI 2017]
Incorporating Background Knowledge via Adversarial Training

Idea — adversarial training process where, iteratively:
• An adversary searches for inputs where the model violates constraints,
• The model is regularised to correct such violations.

Formally:

\[
\min_{\Theta} \mathcal{L}_{\text{data}}(D \mid \Theta) + \lambda \max_{S} \mathcal{L}_{\text{violation}}(S, D \mid \Theta)
\]

[e.g. \(S = \{x, y, z\}\) such that \(\text{is } a(x, y) \land \text{is } a(y, z) \land \neg \text{is } a(x, z)\)]

[Minervini et al. UAI 2017]
Incorporating Background Knowledge via Adversarial Training

Idea — adversarial training process where, iteratively:
• An adversary searches for inputs where the model violates constraints,
• The model is regularised to correct such violations.

Formally:
\[
\min_{\Theta} \mathcal{L}_{\text{data}}(D \mid \Theta) + \lambda \max_{S} \mathcal{L}_{\text{violation}}(S, D \mid \Theta)
\]

• Inputs S can be either input space or embedding space
• In most interesting cases, max has closed form solutions
• Constraints are guaranteed to hold everywhere in embedding space.

[minervini et al. UAI 2017]
Incorporating Background Knowledge via Adversarial Training

**Idea** — *adversarial training* process where, iteratively:
- An *adversary* searches for *inputs* where the model violates constraints,
- The model is *regularised* to correct such violations.

Formally:

$$\min_{\Theta} \mathcal{L}_{\text{data}}(D \mid \Theta) + \lambda \max_{S} \mathcal{L}_{\text{violation}}(S, D \mid \Theta)$$

- Incorporates Background Knowledge
- Verifiable

*Minervini et al. UAI 2017*
Incorporating Background Knowledge via Adversarial Training

[Minervini et al. UAI 2017]
Incorporating Background Knowledge via Adversarial Training

[Minervini et al. UAI 2017]
Incorporating Background Knowledge in Natural Language Inference Models

Natural Language Inference — detect the type of relationship, i.e. entailment, contradiction, neutral, between two sentences.

[Minervini et al. CoNLL 2018]
Incorporating Background Knowledge in Natural Language Inference Models

Natural Language Inference — detect the type of relationship, i.e. entailment, contradiction, neutral, between two sentences.

If a sentence $x$ contradicts $y$, then also $y$ contradicts $x$.
If $x$ entails $y$, and $y$ entails $z$, then $x$ also entails $z$.

[Minervini et al. CoNLL 2018]
Incorporating Background Knowledge in Natural Language Inference Models

Natural Language Inference — detect the type of relationship, i.e. entailment, contradiction, neutral, between two sentences.

If a sentence \( x \) contradicts \( y \), then also \( y \) contradicts \( x \).
If \( x \) entails \( y \), and \( y \) entails \( z \), then \( x \) also entails \( z \).

\( x \) A man in uniform is pushing a medical bed.
\( y \) A man is pushing *carrying* something.
Incorporating Background Knowledge in Natural Language Inference Models

Natural Language Inference — detect the type of relationship, i.e. entailment, contradiction, neutral, between two sentences.

If a sentence $\mathbf{x}$ contradicts $\mathbf{y}$, then also $\mathbf{y}$ contradicts $\mathbf{x}$.
If $\mathbf{x}$ entails $\mathbf{y}$, and $\mathbf{y}$ entails $\mathbf{z}$, then $\mathbf{x}$ also entails $\mathbf{z}$.

$x)$ A man in uniform is pushing a medical bed.
$y)$ A man is pushing carrying something.

$P(\mathbf{x} \xrightarrow{\text{entails}} \mathbf{y}) = 0.72$
$P(\mathbf{y} \xrightarrow{\text{contradicts}} \mathbf{x}) = 0.93$

$\mathcal{L}_{\text{violation}}(\{\mathbf{x}, \mathbf{y}\}) : 0.01 \Rightarrow 0.92$

[Minervini et al. CoNLL 2018]
Incorporating Background Knowledge in Natural Language Inference Models

Number of violations (%) made by ESIM

- con($X_1, X_2$) $\Rightarrow$ con($X_2, X_1$)
- ent($X_1, X_2$) $\Rightarrow$ $\neg$ con($X_2, X_1$)
- neut($X_1, X_2$) $\Rightarrow$ $\neg$ con($X_2, X_1$)
- $\top$ $\Rightarrow$ ent($X_1, X_1$)

Violations (%)

Regularisation Parameter $\lambda$

[Minervini et al. CoNLL 2018]
End-to-End Differentiable Reasoning
End-to-End Differentiable Reasoning

Core idea — we can combine *neural networks* and *symbolic models* by re-implementing classic reasoning algorithms using end-to-end differentiable (neural) architectures.

(Black-Box) Neural Models

- Can generalise from noisy and ambiguous modalities
- Can learn representations from data
- SOTA on a number of tasks

Symbolic Reasoning Models

- Data efficient
- Interpretable
- Explainable
- Verifiable
- Can incorporate background knowledge and constraints
Reasoning via Backward Chaining

Backward Chaining — start with a list of goals, and work backwards from the consequent $Q$ to the antecedent $P$ to see if any data supports any of the consequents.

$q(X) \leftarrow p(X)$

$p(a)$
$q(a)$?

$p(b)$

$p(c)$

You can see backward chaining as a query reformulation strategy.
Reasoning via Backward Chaining

**Backward Chaining** — start with a list of *goals*, and work backwards from the *consequent* $Q$ to the *antecedent* $P$ to see if any data supports any of the consequents.

\[
q(X) \leftarrow p(X)
\]

\[
p(a) \quad q(a)\ ?
\]

\[
p(b)
\]

\[
p(c)
\]

\[
\ldots
\]

You can see backward chaining as a *query reformulation strategy*. 
Reasoning via Backward Chaining

**Backward Chaining** — start with a list of *goals*, and work backwards from the *consequent* $Q$ to the *antecedent* $P$ to see if any data supports any of the consequents.

$q(X) \leftarrow p(X)$

$p(a)$  \hspace{1cm} $q(a)$?  \hspace{1cm} $p(b)$  \hspace{1cm} $p(c)$  \hspace{1cm} $\ldots$

You can see backward chaining as a *query reformulation strategy*. 
End-to-End Differentiable Reasoning

\[ \text{grandPaOf (abe, bart)} \]

\[ \begin{align*}
\text{sim} &= 0.9 \\
\text{sim} &= 1 \\
\text{sim} &= 1
\end{align*} \]

\[ \text{grandFatherOf (abe, bart)} \]
End-to-End Differentiable Reasoning

Knowledge Base:
fatherOf(abe, homer)
parentOf(homer, bart)
grandFatherOf(X, Y) ⇐
fatherOf(X, Z),
parentOf(Z, Y).

proof score $S_1$

proof score $S_2$
End-to-End Differentiable Reasoning

Knowledge Base:

fatherOf(abe, homer)
parentOf(homer, bart)
grandFatherOf(X, Y) ← fatherOf(X, Z), parentOf(Z, Y).

Subgoals:

fatherOf(abe, Z)
parentOf(Z, bart)
End-to-End Differentiable Reasoning

Knowledge Base:

fatherOf(abe, homer)
parentOf(homer, bart)
grandFatherOf(X, Y) ⇐
fatherOf(X, Z),
parentOf(Z, Y).

Subgoals:
fatherOf(abe, Z)
parentOf(Z, bart)
End-to-End Differentiable Reasoning

Knowledge Base:
fatherOf(abe, homer)
parentOf(homer, bart)
$\theta_1(X, Y) \iff \theta_2(X, Z), \theta_3(Z, Y)$

Train via Self-Supervision:

$$
\sum_{F \in K} \log p^{KB\setminus F}(F)
- \sum_{\tilde{F} \sim \text{corr}(F)} \log p^{KB}(\tilde{F})
$$
End-to-End Differentiable Reasoning

Knowledge Base:

\[
\begin{align*}
\text{fatherOf}(\text{abe}, \text{homer}) \\
\text{parentOf}(\text{homer}, \text{bart}) \\
\text{grandFatherOf}(X, Y) & \iff \\
& \text{fatherOf}(X, Z), \text{parentOf}(Z, Y). \\
\end{align*}
\]

Subgoals:

\[
\begin{align*}
\text{fatherOf}(\text{abe}, Z) & \iff \\
& \text{grandFatherOf}(X, Y) \iff \\
& \text{fatherOf}(Z, \text{bart}). \\
\end{align*}
\]
End-to-End Differentiable Reasoning

Knowledge Base:

\[
\begin{align*}
\text{fatherOf}(\text{abe}, \text{homer}) \\
\text{parentOf}(\text{homer}, \text{bart}) \\
\text{grandFatherOf}(X, Y) &\iff \\
&\text{fatherOf}(X, Z), \\
&\text{parentOf}(Z, Y).
\end{align*}
\]

Subgoals:

\[
\begin{align*}
\text{proof score } S_3 \\
\text{grandFatherOf}(X, Y) \\
\text{fatherOf}(\text{abe}, Z) \\
\text{parentOf}(\text{homer}, \text{bart}) \\
\text{proof score } S_4 \\
\text{fatherOf}(\text{abe}, \text{homer}) \\
\text{proof score } S_5
\end{align*}
\]

## End-to-End Differentiable Reasoning

<table>
<thead>
<tr>
<th>Query</th>
<th>Score $S_p$</th>
<th>Proofs / Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>part_of(CONGO.N.03, AFRICA.N.01)</code></td>
<td>0.995</td>
<td><code>part_of(X, Y)</code>:: <code>has_part(Y, X)</code>&lt;br&gt;<code>has_part(AFRICA.N.01, CONGO.N.03)</code>&lt;br&gt;<code>part_of(X, Y)</code>:: <code>instance_hyponym(Y, X)</code>&lt;br&gt;<code>instance_hyponym(AFRICAN_COUNTRY.N.01, CONGO.N.03)</code></td>
</tr>
<tr>
<td><code>hyponym(EXTINGUISH.V.04, DECOUPLE.V.03)</code></td>
<td>0.987</td>
<td><code>hyponym(X, Y)</code>:: <code>hyponym(Y, X)</code>&lt;br&gt;<code>hypernym(DECOUPLE.V.03, EXTINGUISH.V.04)</code>&lt;br&gt;<code>hypernym(SNUFF_OUT.V.01, EXTINGUISH.V.04)</code></td>
</tr>
<tr>
<td><code>part_of(PITUITARY.N.01, DIENCEPHALON.N.01)</code></td>
<td>0.995</td>
<td><code>has_part(DIENCEPHALON.N.01, PITUITARY.N.01)</code></td>
</tr>
<tr>
<td><code>has_part(TEXAS.N.01, ODESSA.N.02)</code></td>
<td>0.961</td>
<td><code>has_part(X, Y)</code>:: <code>part_of(Y, X)</code>&lt;br&gt;<code>part_of(ODESSA.N.02, TEXAS.N.01)</code></td>
</tr>
<tr>
<td><code>hyponym(SKELETAL_MUSCLE, ARTICULAR_MUSCLE)</code></td>
<td>0.987</td>
<td><code>hypernym(ARTICULAR_MUSCLE, SKELETAL_MUSCLE)</code></td>
</tr>
<tr>
<td><code>deriv_related_form(REWRITE, REWRITING)</code></td>
<td>0.809</td>
<td><code>deriv_related_form(X, Y)</code>:: <code>hypernym(Y, X)</code>&lt;br&gt;<code>hypernym(REVISE, REWRITE)</code></td>
</tr>
<tr>
<td><code>also_see(TRUE.A.01, FAITHFUL.A.01)</code></td>
<td>0.962</td>
<td><code>also_see(X, Y)</code>:: <code>also_see(Y, X)</code>&lt;br&gt;<code>also_see(FAITHFUL.A.01, TRUE.A.01)</code>&lt;br&gt;<code>also_see(CONSTANT.A.02, FAITHFUL.A.01)</code></td>
</tr>
<tr>
<td><code>also_see(GOOD.A.03, VIRTUOUS.A.01)</code></td>
<td>0.962</td>
<td><code>also_see(VIRTUOUS.A.01, GOOD.A.03)</code>&lt;br&gt;<code>also_see(RIGHTEOUS.A.01, VIRTUOUS.A.01)</code></td>
</tr>
<tr>
<td><code>instance_hypernym(CHAPLIN, FILM MAKER)</code></td>
<td>0.812</td>
<td><code>instance_hypernym(CHAPLIN, COMEDIAN)</code></td>
</tr>
</tbody>
</table>

[Minervini et al. AAAI 2020]
## End-to-End Differentiable Reasoning

<table>
<thead>
<tr>
<th>Rules</th>
<th>Test-I</th>
<th></th>
<th></th>
<th>Test-II</th>
<th></th>
<th></th>
<th>Test-ALL</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hits@N (%)</td>
<td></td>
<td></td>
<td>MRR</td>
<td>Hits@N (%)</td>
<td></td>
<td></td>
<td>MRR</td>
<td>Hits@N (%)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.291</td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.713</td>
<td>3</td>
</tr>
<tr>
<td>KALE-Pre (Guo et al. 2016)</td>
<td>35.8</td>
<td>41.9</td>
<td>49.8</td>
<td></td>
<td>82.9</td>
<td>86.1</td>
<td>89.9</td>
<td></td>
<td>61.7</td>
</tr>
<tr>
<td>KALE-Joint (Guo et al. 2016)</td>
<td>38.4</td>
<td>44.7</td>
<td>52.2</td>
<td>0.325</td>
<td>79.7</td>
<td>84.1</td>
<td>89.6</td>
<td>0.684</td>
<td>61.2</td>
</tr>
<tr>
<td>ASR-DistMult (Minervini et al. 2017)</td>
<td>36.3</td>
<td>40.3</td>
<td>44.9</td>
<td>0.330</td>
<td>98.0</td>
<td>99.0</td>
<td>99.2</td>
<td>0.948</td>
<td>70.7</td>
</tr>
<tr>
<td>ASR-ComplEx (Minervini et al. 2017)</td>
<td>37.3</td>
<td>41.0</td>
<td>45.9</td>
<td>0.338</td>
<td>99.2</td>
<td>99.3</td>
<td>99.4</td>
<td>0.984</td>
<td>71.7</td>
</tr>
<tr>
<td>KBLR (García-Durán and Niepert 2018)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.261</td>
<td>77.5</td>
<td>82.8</td>
<td>88.4</td>
<td>0.630</td>
<td>58.9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.313</td>
<td>92.3</td>
<td>93.8</td>
<td>94.7</td>
<td>0.874</td>
<td>67.4</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.329</td>
<td>91.4</td>
<td>91.9</td>
<td>92.4</td>
<td>0.887</td>
<td>67.3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.313</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.261</td>
<td>77.5</td>
<td>82.8</td>
<td>88.4</td>
<td>0.630</td>
<td>58.9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.313</td>
<td>92.3</td>
<td>93.8</td>
<td>94.7</td>
<td>0.874</td>
<td>67.4</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.329</td>
<td>91.4</td>
<td>91.9</td>
<td>92.4</td>
<td>0.887</td>
<td>67.3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.313</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.261</td>
<td>77.5</td>
<td>82.8</td>
<td>88.4</td>
<td>0.630</td>
<td>58.9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.313</td>
<td>92.3</td>
<td>93.8</td>
<td>94.7</td>
<td>0.874</td>
<td>67.4</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.329</td>
<td>91.4</td>
<td>91.9</td>
<td>92.4</td>
<td>0.887</td>
<td>67.3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.313</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.261</td>
<td>77.5</td>
<td>82.8</td>
<td>88.4</td>
<td>0.630</td>
<td>58.9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.313</td>
<td>92.3</td>
<td>93.8</td>
<td>94.7</td>
<td>0.874</td>
<td>67.4</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.329</td>
<td>91.4</td>
<td>91.9</td>
<td>92.4</td>
<td>0.887</td>
<td>67.3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.313</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.261</td>
<td>77.5</td>
<td>82.8</td>
<td>88.4</td>
<td>0.630</td>
<td>58.9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.313</td>
<td>92.3</td>
<td>93.8</td>
<td>94.7</td>
<td>0.874</td>
<td>67.4</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.329</td>
<td>91.4</td>
<td>91.9</td>
<td>92.4</td>
<td>0.887</td>
<td>67.3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.313</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.261</td>
<td>77.5</td>
<td>82.8</td>
<td>88.4</td>
<td>0.630</td>
<td>58.9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.313</td>
<td>92.3</td>
<td>93.8</td>
<td>94.7</td>
<td>0.874</td>
<td>67.4</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.329</td>
<td>91.4</td>
<td>91.9</td>
<td>92.4</td>
<td>0.887</td>
<td>67.3</td>
</tr>
</tbody>
</table>

### Datasets

<table>
<thead>
<tr>
<th>Metrics</th>
<th>NTP ³</th>
<th>GNTP</th>
<th>NeuralLP</th>
<th>MINERVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard</td>
<td>Attention</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Countries

| S1 | 90.83 ± 15.4 | 99.98 ± 0.05 |
| S2 | 87.40 ± 11.7 | 90.82 ± 0.88 |
| S3 | 56.68 ± 17.6 | 87.70 ± 4.79 |

### Kinship

<table>
<thead>
<tr>
<th>MRR</th>
<th>0.35</th>
<th>0.719</th>
<th>0.759</th>
<th>0.619</th>
<th>0.720</th>
</tr>
</thead>
<tbody>
<tr>
<td>HITS@1</td>
<td>0.24</td>
<td>0.586</td>
<td>0.642</td>
<td>0.475</td>
<td>0.605</td>
</tr>
<tr>
<td>HITS@3</td>
<td>0.37</td>
<td>0.815</td>
<td>0.850</td>
<td>0.707</td>
<td>0.812</td>
</tr>
<tr>
<td>HITS@10</td>
<td>0.57</td>
<td>0.958</td>
<td>0.959</td>
<td>0.912</td>
<td>0.924</td>
</tr>
</tbody>
</table>

### Rules Learned by GNTP

- `locatedIn(X,Y) :- locatedIn(X,Z), locatedIn(Z,Y)`
- `neighborOf(X,Y) :- neighborOf(X,Z), locatedIn(Z,Y)`
- `neighborOf(X,Y) :- neighborOf(Y,X)`
- `term0(X,Y) :- term0(Y,X)`
- `term4(X,Y) :- term4(Y,X)`
- `term3(X,Y) :- term3(X,Z), term10(Z,Y)`
- `term2(X,Y) :- term4(X,Z), term7(Z,Y)`
End-to-End Differentiable Reasoning with Natural Language

We can embed facts from the KG and facts from text in a *shared embedding space*, and learn to reason over them *jointly*:
End-to-End Differentiable Reasoning with Natural Language

We can embed facts from the KG and facts from text in a *shared embedding space*, and learn to reason over them *jointly*:

Control Myself \text{record\_label} Jam Recordings

\text{record\_label}(X, Z) \leftarrow p_1(X, Y)

p_1(X, Z) \leftarrow p_2(X,Y) \land p_3(Y,Z)

Control Myself [...] is a song by american rapper [...] Ell Cools 1989 album [...] was released by [...] Jam Recordings

[Welbl et al. ACL 2019, Minervini et al. AAAI 2020]
End-to-End Differentiable Reasoning with Natural Language

We can embed facts from the KG and facts from text in a *shared embedding space*, and learn to reason over them *jointly*:

Thrasyvoulos F.C. `country` Greece

country\( (X, Z) \leftarrow p_1 (X, Y) \)

\[ p_1 (X, Z) \leftarrow p_2 (X,Y) \land p_3 (Y,Z) \]

Thrasyvoulos Fylis is a football club based in Fyli, Attica [...] 
Fyli is a town and a municipality in the northwestern part of Attica, Greece

[Welbl et al. ACL 2019, Minervini et al. AAAI 2020]
Thank you!