Part II: Models based on extra information

TOPICS:

INCORPORATING ENTITY TYPES
INCORPORATING TEXTUAL DESCRIPTIONS
INCORPORATING RELATION PATHS
INCORPORATING LOGIC RULES
Part II: Models based on extra information

TOPICS:

- INCORPORATING ENTITY TYPES
- INCORPORATING TEXTUAL DESCRIPTIONS
- INCORPORATING RELATION PATHS
- INCORPORATING LOGIC RULES
Entity types

- Semantic categories to which entities belong

  - author
  - written_work

  (William Shakespeare, book/author/works_written, Romeo and Juliet)

- Each entity may have **multiple type labels**, and the types could also be **hierarchical**

Semantically smooth embedding *(Guo et al., 2015)*

- **Key idea**
  - Entities of the same type should stay close in the embedding space

- **Modeling semantic smoothness**
  - **Laplacian eigenmaps (LE):** If two entities belong to the same type, they will have their embeddings similar to each other
    \[
    \mathcal{R}_1 = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \| e_i - e_j \|^2 w_{i,j}^1, \quad w_{i,j}^1 = \begin{cases} 
    1, & \text{type}(e_i) = \text{type}(e_j) \\
    0, & \text{otherwise}
    \end{cases}
    \]

  - **Locally linear embedding (LLE):** An entity can be reconstructed from its near neighbors (entities of the same type) in the embedding space
    \[
    \mathcal{R}_2 = \sum_{i=1}^{n} \| e_i - \sum_{e_j \in \mathcal{N}(e_i)} w_{i,j}^2 e_j \|^2, \quad w_{i,j}^2 = \begin{cases} 
    \frac{1}{|\mathcal{N}(e_i)|}, & e_j \in \mathcal{N}(e_i) \\
    0, & \text{otherwise}
    \end{cases}
    \]
Semantically smooth embedding (cont.)

- Visualization of entity vectors learned by semantically smooth embedding (Guo et al., 2015)

(a) TransE.
(b) TransE-Cat.
(c) TransE-LE.
(d) TransE-LLE.

- Athlete
- Politicians
- Chemical
- City
- Clothing
- Country
- Sportsteam
- Journalist
- Televisionstation
- Room
Type-embodied knowledge representation learning (Xie et al., 2016b)

- **Key idea**
  - Translation after type-specific entity projection: \( \mathbf{M}_{rs}s + \mathbf{r} \approx \mathbf{M}_{ro}o \)

- **Modeling multiple type labels**
  - Projecting an entity with a linear combination of type matrices

\[
\mathbf{M}_{rs} = \frac{\sum_{i=1}^{n_s} \alpha_i \mathbf{M}_{c_i}}{\sum_{i=1}^{n_s} \alpha_i}, \quad \alpha_i = \begin{cases} 1, & c_i \in C_{rs} \\ 0, & c_i \notin C_{rs} \end{cases}
\]

- **Modeling hierarchical types**
  - Projection matrix of a type as a composition of projection matrices of its sub-types

\[
\text{addition: } \mathbf{M}_{c_i} = \beta_1 \mathbf{M}_{c_{i_1}} + \cdots + \beta_\ell \mathbf{M}_{c_{i_\ell}}
\]

\[
\text{multiplication: } \mathbf{M}_{c_i} = \mathbf{M}_{c_{i_1}} \odot \cdots \odot \mathbf{M}_{c_{i_\ell}}
\]
Part II: Models based on extra information

TOPICS:

INCORPORATING ENTITY TYPES
INCORPORATING TEXTUAL DESCRIPTIONS
INCORPORATING RELATION PATHS
INCORPORATING LOGIC RULES
Textual descriptions

- Concise descriptions of entities in knowledge graphs

(William Shakespeare, book/author/works_written, Romeo and Juliet)

- Other general textual information such as news releases and Wikipedia articles

Example from: Xie et al. (2016). Representation learning of knowledge graphs with entity descriptions. AAAI’16.
Initialization by word embeddings (Socher et al., 2013)

- **Key idea**
  - Initializing entity representations with pre-trained word embeddings

\[
\text{vec}(\text{Bengal tiger}) = \frac{1}{2} (\text{vec}(\text{Bengal}) + \text{vec}(\text{tiger}))
\]
Jointly embedding with text data (Wang et al., 2014; Zhong et al., 2015)

- **Key idea**
  - Jointly embedding relations, entities, and words into the same vector space so that one can make predictions between entities and words

- **Jointly embedding framework**
  - **Knowledge model**: Modeling triples in a knowledge graph
  - **Text model**: Modeling co-occurring word pairs in a text corpus
  - **Alignment model**: Aligning entity and word embedding spaces
    - Wikipedia anchors, entity names, entity descriptions
Jointly embedding with text data (cont.)

- **Knowledge model**

\[
\Pr(s|r, o) = \frac{\exp\{z(s, r, o)\}}{\sum_{s' \in \mathcal{E}} \exp\{z(s', r, o)\}} \quad z(s, r, o) = b - 0.5\|s + r - o\|
\]

\[
\mathcal{L}_K = - \sum_{(s, r, o) \in \mathcal{T}^+} \left[ \log \Pr(s|r, o) + \log \Pr(r|s, o) + \log \Pr(o|s, r) \right]
\]

- **Text model**

\[
\Pr(w|v) = \frac{\exp\{z(w, v)\}}{\sum_{w' \in \mathcal{V}} \exp\{z(w', v)\}} \quad z(w, v) = b - 0.5\|w - v\|
\]

\[
\mathcal{L}_T = - \sum_{(w, v) \in \mathcal{C}} \log \Pr(w|v)
\]

- **Alignment model** (by entity descriptions)

\[
\Pr(w|e) = \frac{\exp\{z(e, w)\}}{\sum_{w' \in \mathcal{V}} \exp\{z(e, w')\}} \quad z(e, w) = b - 0.5\|e - w\|
\]

\[
\mathcal{L}_A = - \sum_{e \in \mathcal{E}} \sum_{w \in \mathcal{D}_e} \left[ \log \Pr(w|e) + \log \Pr(e|w) \right]
\]
Key idea

- Entity: structure-based embedding + description-based embedding
- Description-based embedding as composition of word embeddings

Triple scoring function

\[ f(s, r, o) = - \|s_K + r - o_K\| \quad \text{① score for structure-based embeddings} \]
\[ - \|s_T + r - o_T\| - \|s_K + r - o_T\| - \|s_T + r - o_K\| \]

② score for description-based embeddings

- \(s_K, o_K\): structure-based entity embeddings
- \(s_T, o_T\): description-based entity embeddings, modeled as compositions of word embeddings
Modeling description-based entity embeddings

- **Continuous bag-of-words encoder**: Composition by addition, ignoring word orders

\[
\text{subject} + \text{relation} = \text{object}
\]

- Keywords of subject + Keywords of relation = Keywords of object
Modeling description-based entity embeddings

- **Convolutional neural network encoder:** Composition by CNN, taking word orders into account

subject + relation = object

description of subject

2nd Pooling & nonlinear

2nd Convolution

1st Pooling & nonlinear

1st Convolution

description of object
Part II: Models based on extra information

TOPICS:

- INCORPORATING ENTITY TYPES
- INCORPORATING TEXTUAL DESCRIPTIONS
- INCORPORATING RELATION PATHS
- INCORPORATING LOGIC RULES
Relation paths

- Multi-hop relationships between entities, extremely useful for predicting missing links

Modeling relation paths

- Relation path: A sequence of relations linking two entities

\[ p = (r_1, r_2, \cdots, r_\ell) \iff s \xrightarrow{r_1} e_1 \xrightarrow{r_2} \cdots \xrightarrow{r_\ell} o \]

- Path representation: **Composition** of relation representations

\[ p = r_1 \circ r_2 \circ \cdots \circ r_\ell \quad (\circ \text{ is a composition operation}) \]
Path-based TransE (Lin et al., 2015)

- Key idea
  - Taking relation paths as translations between long distance entities

subject + path = object

- Semantic composition

addition: \( p = r_1 + r_2 + \cdots + r_\ell \)

multiplication: \( p = r_1 \odot r_2 \odot \cdots \odot r_\ell \)

RNN: \( c_i = f(W[c_{i-1}; r_i]) \)
Path-based TransE (cont.)

- **Optimization problem**

  - **Modeling relation-connected triple \((s, r, o)\)**
    
    \[
    \mathcal{L}(s, r, o) = \sum_{(s', r', o') \in \mathcal{T}^-} \max(0, \gamma + \|s + r - o\| - \|s' + r' - o'\|)
    \]

    loss of triple

  - **Modeling path-connected triple \((s, p, o)\)**
    
    \[
    f(s, p, o) = -\|s + p - o\| = -\|p - (o - s)\| \approx -\|p - r\|
    \]

    \[
    \mathcal{L}(p, r) = \sum_{(s, r', o) \in \mathcal{T}^-} \max(0, \gamma + \|p - r\| - \|p - r'\|)
    \]

    loss of path

  - **Combining the two parts**
    
    \[
    \min \sum_{(s, r, o) \in \mathcal{T}^+} \left[ \mathcal{L}(s, r, o) + \frac{1}{Z} \sum_{p \in \mathcal{P}(s, o)} R(p|s, o) \mathcal{L}(p, r) \right]
    \]

    ① loss of \((s, r, o)\)

    ② loss of all paths linking \(s\) and \(o\)

- **Reliability of path**

  \[
  R(p|s, o) = \frac{1}{Z} \sum_{p \in \mathcal{P}(s, o)} \mathcal{L}(p, r)
  \]
Path-based TransE (cont.)

- Link prediction performance on FB15k

<table>
<thead>
<tr>
<th>Metric</th>
<th>Mean Rank Raw</th>
<th>Mean Rank Filter</th>
<th>Hits@10 (%) Raw</th>
<th>Hits@10 (%) Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESCAL</td>
<td>828</td>
<td>683</td>
<td>28.4</td>
<td>44.1</td>
</tr>
<tr>
<td>SE</td>
<td>273</td>
<td>162</td>
<td>28.8</td>
<td>39.8</td>
</tr>
<tr>
<td>SME (linear)</td>
<td>274</td>
<td>154</td>
<td>30.7</td>
<td>40.8</td>
</tr>
<tr>
<td>SME (bilinear)</td>
<td>284</td>
<td>158</td>
<td>31.3</td>
<td>41.3</td>
</tr>
<tr>
<td>LFM</td>
<td>283</td>
<td>164</td>
<td>26.0</td>
<td>33.1</td>
</tr>
<tr>
<td>TransE</td>
<td>243</td>
<td>125</td>
<td>34.9</td>
<td>47.1</td>
</tr>
<tr>
<td>TransH</td>
<td>212</td>
<td>87</td>
<td>45.7</td>
<td>64.4</td>
</tr>
<tr>
<td>TransR</td>
<td>198</td>
<td>77</td>
<td>48.2</td>
<td>68.7</td>
</tr>
<tr>
<td>TransE (Our)</td>
<td>205</td>
<td>63</td>
<td>47.9</td>
<td>70.2</td>
</tr>
<tr>
<td>PTransE (ADD, 2-step)</td>
<td>200</td>
<td>54</td>
<td>51.8</td>
<td>83.4</td>
</tr>
<tr>
<td>PTransE (MUL, 2-step)</td>
<td>216</td>
<td>67</td>
<td>47.4</td>
<td>77.7</td>
</tr>
<tr>
<td>PTransE (RNN, 2-step)</td>
<td>242</td>
<td>92</td>
<td>50.6</td>
<td>82.2</td>
</tr>
<tr>
<td>PTransE (ADD, 3-step)</td>
<td>207</td>
<td>58</td>
<td>51.4</td>
<td>84.6</td>
</tr>
</tbody>
</table>

Addition composition operation performs best

\[ s \xrightarrow{r_1} e_1 \xrightarrow{r_2} o \]

\[ s + r_1 = e_1 \]

\[ e_1 + r_2 = o \]

\[ s + (r_1 + r_2) = o \]
Compositionalizing other models (Guu et al., 2015)

- **TransE** (composition by addition)
  \[
  f(s, r, o) = -\|s + r - o\| \Rightarrow f(s, p, o) = -\|s + (r_1 + \cdots + r_\ell) - o\|
  \]

- **RESCAL** (composition by multiplication)
  \[
  f(s, r, o) = s^\top M_r o \Rightarrow f(s, p, o) = s^\top (M_{r_1} \odot \cdots \odot M_{r_\ell}) o
  \]

- **DistMult** (composition by multiplication)
  \[
  f(s, r, o) = s^\top \text{diag}(r) o \Rightarrow f(s, p, o) = s^\top \text{diag}(r_1 \odot \cdots \odot r_\ell) o
  \]
Advantage of compositional learning (Guu et al., 2015)

- Modeling triples separately introduces **cascading errors**

- Modeling paths compositionally reduces cascading errors
Path-RNN (Neelakantan et al., 2015)

- Modeling paths (sequences of relations) with RNN

\[ p \approx r \]
Path-RNN with entity information (Das et al., 2017)

- Entity information is useful in path modeling
  - Without entity information
    
    ![Diagram without entity information](image1)

  - With entity information
    
    ![Diagram with entity information](image2)
Taking paths as sequences of entities and relations, modeled with RNN
Part II: Models based on extra information

TOPICS:

- INCORPORATING ENTITY TYPES
- INCORPORATING TEXTUAL DESCRIPTIONS
- INCORPORATING RELATION PATHS
- INCORPORATING LOGIC RULES
Logic rules

- First-order Horn clauses
  \[
  \forall x, y: (x, \text{capitalOf}, y) \implies (x, \text{locatedIn}, y)
  \]
  The capital city of a country must be located in that country

- Having a close relationship to relation paths
  \[
  \forall x, y, z: (x, \text{Director}, y) \land (y, \text{Language}, z) \\
  \implies (x, \text{Language of Film}, z)
  \]
  \[
  \forall x, y, z: (x, \text{Release Region}, y) \land (y, \text{Official Language}, z) \\
  \implies (x, \text{Language of Film}, z)
  \]

Hard rules vs. soft rules

- **Hard rules**
  - Rules that always hold with no exception
    \[ \forall x, y: (x, \text{capitalOf}, y) \Rightarrow (x, \text{locatedIn}, y) \]
    *The capital city of a country must be located in that country*
  - Usually requiring extensive manual effort to create or validate

- **Soft rules**
  - Rules with different confidence levels that can handle uncertainty
    \[ \forall x, y: (x, \text{bornIn}, y) \Rightarrow (x, \text{nationality}, y) \quad (\text{confidence} = 0.8) \]
    *A person is very likely (but not necessarily) to have a nationality of the country where he/she was born*
  - Automatically extracted via modern rule mining systems
Automatically extracted soft rules

- Soft rules mined from YAGO by AMIE+ (Galárraga et al. 2015)

<table>
<thead>
<tr>
<th>Rule</th>
<th>Precision in the unknown region</th>
<th>Std. Confidence</th>
<th>PCA Confidence</th>
<th>New Predictions</th>
<th>Total predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>?a &lt;isMarriedTo&gt; ?a ?b &lt;hasChild&gt; ?b =&gt; ?a &lt;hasChild&gt; ?b</td>
<td>34.48%</td>
<td>57.57%</td>
<td>57.57%</td>
<td>1865</td>
<td>1865</td>
</tr>
<tr>
<td>?a &lt;isMarriedTo&gt; ?a ?b ?c &lt;hasChild&gt; ?b =&gt; ?a &lt;hasChild&gt; ?b</td>
<td>33.33%</td>
<td>56.11%</td>
<td>56.11%</td>
<td>1068</td>
<td>1586</td>
</tr>
<tr>
<td>?a &lt;isMarriedTo&gt; ?a ?b ?c &lt;hasChild&gt; ?b =&gt; ?a &lt;hasMarriedTo&gt; ?b</td>
<td>100.00%</td>
<td>53.01%</td>
<td>91.79%</td>
<td>5158</td>
<td>5158</td>
</tr>
<tr>
<td>?a &lt;created&gt; ?b ?a &lt;produced&gt; ?b =&gt; ?a &lt;created&gt; ?b</td>
<td>3.33%</td>
<td>49.83%</td>
<td>58.68%</td>
<td>976</td>
<td>976</td>
</tr>
<tr>
<td>?a &lt;actedIn&gt; ?b ?a &lt;created&gt; ?b =&gt; ?a &lt;directed&gt; ?b</td>
<td>3.45%</td>
<td>38.29%</td>
<td>45.23%</td>
<td>518</td>
<td>608</td>
</tr>
<tr>
<td>?a &lt;isMarriedTo&gt; ?c ?c &lt;livesIn&gt; ?b =&gt; ?a &lt;livesIn&gt; ?b</td>
<td>75.00%</td>
<td>33.71%</td>
<td>69.88%</td>
<td>326</td>
<td>326</td>
</tr>
<tr>
<td>?a &lt;directed&gt; ?b =&gt; ?a &lt;created&gt; ?b</td>
<td>20.00%</td>
<td>28.19%</td>
<td>28.19%</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>?a &lt;isMarriedTo&gt; ?a ?b ?c &lt;livesIn&gt; ?b =&gt; ?a &lt;livesIn&gt; ?b</td>
<td>100.00%</td>
<td>22.61%</td>
<td>68.42%</td>
<td>299</td>
<td>542</td>
</tr>
<tr>
<td>?a &lt;actedIn&gt; ?b ?a &lt;created&gt; ?b =&gt; ?a &lt;produced&gt; ?b</td>
<td>15.52%</td>
<td>20.64%</td>
<td>38.98%</td>
<td>750</td>
<td>790</td>
</tr>
<tr>
<td>?a &lt;created&gt; ?b =&gt; ?a &lt;created&gt; ?b</td>
<td>24.14%</td>
<td>18.75%</td>
<td>32.46%</td>
<td>26380</td>
<td>26380</td>
</tr>
<tr>
<td>?a &lt;livesIn&gt; ?f ?f &lt;hasCapital&gt; ?b =&gt; ?a &lt;livesIn&gt; ?b</td>
<td>16.67%</td>
<td>18.53%</td>
<td>18.53%</td>
<td>2851</td>
<td>2852</td>
</tr>
<tr>
<td>?a &lt;livesIn&gt; ?f ?f &lt;isLocatedIn&gt; ?b =&gt; ?a &lt;livesIn&gt; ?b</td>
<td>89.66%</td>
<td>17.80%</td>
<td>17.80%</td>
<td>4225</td>
<td>4227</td>
</tr>
<tr>
<td>?a &lt;hasChild&gt; ?c ?b &lt;hasChild&gt; ?c =&gt; ?a &lt;isMarriedTo&gt; ?b</td>
<td>76.67%</td>
<td>17.32%</td>
<td>40.99%</td>
<td>1170</td>
<td>1720</td>
</tr>
<tr>
<td>?a &lt;created&gt; ?b</td>
<td>0.00%</td>
<td>16.11%</td>
<td>37.47%</td>
<td>4973</td>
<td>5268</td>
</tr>
<tr>
<td>?c &lt;directed&gt; ?b ?a &lt;directed&gt; ?b =&gt; ?a &lt;produced&gt; ?b</td>
<td>42.86%</td>
<td>15.44%</td>
<td>21.78%</td>
<td>471</td>
<td>473</td>
</tr>
<tr>
<td>?c &lt;hasOfficialLanguage&gt; ?b ?c &lt;livesIn&gt; ?a =&gt; ?a &lt;hasOfficialLanguage&gt; ?b</td>
<td>33.33%</td>
<td>14.29%</td>
<td>64.29%</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>?b &lt;directed&gt; ?b =&gt; ?a &lt;directed&gt; ?b</td>
<td>100.00%</td>
<td>13.64%</td>
<td>18.32%</td>
<td>152</td>
<td>219</td>
</tr>
<tr>
<td>?b &lt;hasCapital&gt; ?f ?f &lt;livesIn&gt; ?b =&gt; ?a &lt;livesIn&gt; ?b</td>
<td>48.28%</td>
<td>13.49%</td>
<td>13.49%</td>
<td>2856</td>
<td>3209</td>
</tr>
<tr>
<td>?c &lt;hasAcademicAdvisor&gt; ?a ?c &lt;livesIn&gt; ?b =&gt; ?a &lt;worksAt&gt; ?b</td>
<td>6.87%</td>
<td>13.44%</td>
<td>39.04%</td>
<td>552</td>
<td>552</td>
</tr>
<tr>
<td>?a &lt;produced&gt; ?b =&gt; ?a &lt;directed&gt; ?b</td>
<td>0.00%</td>
<td>11.26%</td>
<td>13.40%</td>
<td>16262</td>
<td>17188</td>
</tr>
<tr>
<td>?a &lt;livesIn&gt; ?c &lt;livesIn&gt; ?b =&gt; ?a &lt;livesIn&gt; ?b</td>
<td>46.43%</td>
<td>10.58%</td>
<td>14.88%</td>
<td>438</td>
<td>438</td>
</tr>
<tr>
<td>?c &lt;isCitizenOf&gt; ?b ?a &lt;influences&gt; ?c =&gt; ?a &lt;isCitizenOf&gt; ?b</td>
<td>3.33%</td>
<td>10.56%</td>
<td>51.96%</td>
<td>1199</td>
<td>1199</td>
</tr>
<tr>
<td>?a &lt;produced&gt; ?b =&gt; ?a &lt;created&gt; ?b</td>
<td>10.34%</td>
<td>10.06%</td>
<td>15.63%</td>
<td>16272</td>
<td>17463</td>
</tr>
<tr>
<td>?a &lt;worksAt&gt; ?b =&gt; ?a &lt;graduatedFrom&gt; ?b</td>
<td>3.57%</td>
<td>8.99%</td>
<td>16.25%</td>
<td>2217</td>
<td>2217</td>
</tr>
<tr>
<td>?a &lt;livesIn&gt; ?b =&gt; ?a &lt;wasBomln&gt; ?b</td>
<td>0.00%</td>
<td>8.85%</td>
<td>22.53%</td>
<td>20196</td>
<td>20196</td>
</tr>
<tr>
<td>?a &lt;exports&gt; ?b =&gt; ?a &lt;imports&gt; ?b</td>
<td>3.45%</td>
<td>8.68%</td>
<td>10.33%</td>
<td>380</td>
<td>333</td>
</tr>
</tbody>
</table>

Results taken from: [http://resources.mpi-inf.mpg.de/yago-naga/amie/data/yago2/amie_yago2.html](http://resources.mpi-inf.mpg.de/yago-naga/amie/data/yago2/amie_yago2.html)
Jointly embedding with logic rules (Guo et al., 2016)

- Key idea
  - Jointly embedding subject-relation-object triples and **hard** rules

- Jointly embedding framework
  - Triples: **Atomic formulae** modeled by translation assumption
  - Rules: **Complex formulae** modeled by t-norm fuzzy logics

\[(\text{Paris, Capital-Of, France}) \Rightarrow (\text{Paris, Located-In, France})\]
Jointly embedding with logic rules (cont.)

- **Modeling triples**
  - Translation assumption: \( s + r \approx o \)
    \[
    I(s, r, o) = 1 - \frac{1}{3\sqrt{d}} \| s + r - o \|_{\ell_1} \in [0, 1]
    \]

- **Modeling rules**
  - **T-norm fuzzy logics**: Truth value of a complex formulae is a composition of truth values of its constituents
    \[
    I(f_1 \Rightarrow f_2) = I(f_1) \cdot I(f_2) - I(f_1) + 1
    \]
    \[
    I(f_1 \land f_2 \Rightarrow f_3) = I(f_1) \cdot I(f_2) \cdot I(f_3) - I(f_1) \cdot I(f_2) + 1
    \]

- **Joint learning**
  - Minimizing a global loss over both triples and rules
    \[
    \min \sum_{f^+ \in \mathcal{F}^+} \sum_{f^- \in \mathcal{F}^-} \max(0, \gamma - I(f^+) + I(f^-))
    \]
    \( f^+ \) and \( f^- \) can be either atomic or complex formulae
Rule-guided embedding (Guo et al., 2018)

- **Key idea**
  - Knowledge graph embedding with iterative guidance from *soft* rules

- **Iterative learning framework**
  - **Soft label prediction**: Use current embeddings and soft rules to predict soft labels for unlabeled triples
  - **Embedding rectification**: Integrate both labeled and unlabeled triples to update current embeddings

---

**Learning resources**
- Labeled triples $\mathcal{L} = \{(x_\ell, y_\ell)\}$
- Unlabeled triples $\mathcal{U} = \{x_u\}$
- Soft rules $\mathcal{F} = \{(f_p, \lambda_p)\}$ and groundings $\mathcal{G} = \{g_{pq}\}$
Rule-guided embedding (cont.)

- **Soft label prediction**
  - Soft label prediction
  - Soft label to be predicted
  - Truth value computed by current embeddings
  - Soft label should stay close to truth value
  - Confidence level of soft rule
  - Slackness to handle uncertainty

  \[
  \min \frac{1}{2} \sum_{x_u \in U} (s(x_u) - \phi(x_u))^2 + C \sum_{p,q} \xi_{pq}
  \]
  \[
  \text{s.t. } \lambda_p (1 - \pi(g_{pq}|S)) \leq \xi_{pq}, \forall g_{pq} \in G
  \]

- **Embedding rectification**
  - Embedding rectification
  - Loss of labeled triples with their hard labels
  - Loss of unlabeled triples with their soft labels

  \[
  \min \frac{1}{|L|} \sum_{\ell} \ell(\phi(x_\ell), y_\ell) + \frac{1}{|U|} \sum_{u} \ell(\phi(x_u), s(x_u))
  \]
Influence of confidence levels of soft rules on link prediction

Soft rules (even those with moderate confidence levels) are highly beneficial
Main obstacle: Propositionalization

- First-order rules have to be propositionalized using entities in knowledge graphs (grounding)
  - First-order rule
    \[
    \forall x, y: (x, \text{capitalOf}, y) \Rightarrow (x, \text{locatedIn}, y)
    \]
  - Proposition rules
    (Paris, capitalOf, France) ⇒ (Paris, locatedIn, France)
    (Rome, capitalOf, Italy) ⇒ (Rome, locatedIn, Italy)
    (Berlin, capitalOf, Germany) ⇒ (Berlin, locatedIn, Germany)
    (Beijing, capitalOf, China) ⇒ (Beijing, locatedIn, China)
    (Moscow, capitalOf, Russia) ⇒ (Moscow, locatedIn, Russia)
    ⋮

Scales exponentially with graph size (number of entities)
To avoid grounding: Regularizing relation embeddings

- Key idea
  - Modeling first-order rules by regularizing relation embeddings (using no entity embeddings)

First-order rule

\[ \forall x, y: (x, r_p, y) \Rightarrow (x, r_q, y) \]

Any two entities linked by relation \( r_p \) should also be linked by \( r_q \)

Equivalent statement

\[ \forall s, o \in \mathcal{E}: f(s, r_p, o) \leq f(s, r_q, o) \]

For any two entities \( s \) and \( o \), if \( (s, r_p, o) \) is a valid triple with a high score, then \( (s, r_q, o) \) with an even higher score will also be predicted as valid by the embedding model.
Applying to entity pair model (Demeester et al., 2016)

- Triple scoring function: \( f(s, r, o) = v_{so}^T v_r \)
- **Non-negative** entity pair representation: \( v_{so} \geq 0, \forall s, o \in \mathcal{E} \)

\[ v_{rp} \leq v_{rq} \quad \Rightarrow \quad \forall s, o \in \mathcal{E}: f(s, r_p, o) \leq f(s, r_q, o) \]

Applying to ComplEx (Ding et al., 2018)

- Triple scoring function: \( f(s, r, o) = \text{Re}(s^T \text{diag}(r) o) \)
- **Non-negative** entity representation: \( \text{Re}(e) \geq 0, \text{Im}(e) \geq 0, \forall e \in \mathcal{E} \)

\[ \text{Re}(r_p) = \text{Re}(r_q) \quad \text{and} \quad \text{Im}(r_p) \leq \text{Im}(r_q) \quad \Rightarrow \quad \forall s, o \in \mathcal{E}: f(s, r_p, o) \leq f(s, r_q, o) \]

**Pros:** Complexity independent of graph size

**Cons:** Can only handle rules in the simplest form \( \forall x, y: (x, r_p, y) \Rightarrow (x, r_q, y) \)
To avoid grounding: Regularizing relation embeddings (cont.)

- Visualization of relation embeddings learned by regularizing ComplEx (Ding et al., 2018)

<table>
<thead>
<tr>
<th>Relation</th>
<th>Real Component</th>
<th>Imaginary Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>country</td>
<td>-0.57</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>-0.52</td>
<td>-0.81</td>
</tr>
<tr>
<td>location_country</td>
<td>-0.57</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>-0.52</td>
<td>-0.81</td>
</tr>
<tr>
<td>owning_company</td>
<td>-0.06</td>
<td>-0.42</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>-0.68</td>
</tr>
<tr>
<td>owner</td>
<td>-0.06</td>
<td>-0.42</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>-0.68</td>
</tr>
<tr>
<td>spouse</td>
<td>0.15</td>
<td>1.39</td>
</tr>
<tr>
<td></td>
<td>-0.87</td>
<td>-0.63</td>
</tr>
<tr>
<td>spouse^{-1}</td>
<td>-0.83</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>0.47</td>
<td>-0.63</td>
</tr>
<tr>
<td>child^{-1}</td>
<td>0.38</td>
<td>-0.29</td>
</tr>
<tr>
<td></td>
<td>0.47</td>
<td>-0.64</td>
</tr>
<tr>
<td>parent</td>
<td>0.38</td>
<td>-0.29</td>
</tr>
<tr>
<td></td>
<td>0.47</td>
<td>-0.64</td>
</tr>
<tr>
<td>position</td>
<td>0.81</td>
<td>-0.11</td>
</tr>
<tr>
<td></td>
<td>0.39</td>
<td>1.01</td>
</tr>
<tr>
<td>honours</td>
<td>-0.81</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>0.73</td>
<td>1.01</td>
</tr>
<tr>
<td>official_language</td>
<td>0.84</td>
<td>-0.44</td>
</tr>
<tr>
<td></td>
<td>0.61</td>
<td>-0.86</td>
</tr>
<tr>
<td>language</td>
<td>0.84</td>
<td>-0.41</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>-0.80</td>
</tr>
</tbody>
</table>

Equivalence
\[
\text{Re}(r_p) = \text{Re}(r_q) \\
\text{Im}(r_p) = \text{Im}(r_q)
\]

Inversion
\[
\begin{align*}
\text{Re}(r_p) &= \text{Re}(r_q) \\
\text{Im}(r_p) &= -\text{Im}(r_q)
\end{align*}
\]

Implication
\[
\begin{align*}
\text{Re}(r_p) &\leq \text{Re}(r_q) \\
\text{Im}(r_p) &= \text{Im}(r_q)
\end{align*}
\]
To avoid grounding: Adversarial training (Minervini et al., 2017)

- **Key idea**
  - Modeling first-order rules with *adversarially generated* entities rather than real entities

- **Adversarial training architecture**
  - **Adversary**: Generate a set of adversarial entity embeddings on which the rules are violated most
  - **Discriminator**: Learn an embedding model compatible with real input (triples) while satisfying the rules on the adversarial set

Generating rather than traversing entities, with complexity independent of graph size
Review

Problem
• To learn distributed representations of entities and relations from extra information beyond subject-relation-object triples

Incorporating entity types
• Difficulty: hierarchical types and multiple type labels

Incorporating textual descriptions
• Jointly embedding knowledge graphs and words
• Modeling entity embedding as a composition of word embeddings

Incorporating relation paths
• Taking path as a sequence of relations, sequence modeling by addition, multiplication, or RNN
• Intermediate entities may also be included during sequence modeling
Incorporating logic rules

- Hard rules versus soft rules
- Difficulty: propositionalization (grounding)
- Avoiding grounding by regularizing entity representations
- Avoiding grounding by adversarial training
References


