Event-enhanced Learning for Knowledge Graph Completion

Ringsquandl M., Hildebrandt M., Lepratti R., Lamparter S. – Siemens
Kharlamov E., Horrocks I. – University of Oxford
Stepanova D. – Max-Planck Institut
Kröger P. – Ludwig-Maximilians Universität

Unrestricted © Siemens AG 2018
Outline

Industrial Knowledge Graphs
- Digital Twin of Manufacturing Systems

Missing Information in Engineering
- Motivating Example

Statistical Relational Learning
- Preliminaries
- Background Enhancement

Event Representation
- First-class citizens

Our Approach
- Combined objective function
- Link prediction results

Conclusion
- Major Take-Aways
Industrial Knowledge Graphs
Digital Twin of Manufacturing Systems

Bill of Material
Car Seat -> Base Part -> Headrest

Process Routing
Pre-Assembly -> Assembly -> Finishing

Process Operations
Line
RobotA
RobotB

Operational Data
Observations

Physical System

As Knowledge Graph

Assembly Line A
- hasPart Headrest
- hasPart Base Part

Assembly Robot
- connectedTo Finishing Station
- involvedIn Assembly
  - followedBy Finishing
- inputTo Base Part
  - consistsOf Headrest
- consistsOf Car Seat
Industrial Knowledge Graphs
Digital Twin of Manufacturing Systems

Bill of Material
- Car Seat
  - Base Part
    - Headrest

Process Routing
- Pre-Assembly
  - Assembly
    - Finishing
  - Line
    - Pick Headrest
    - Mount Headrest
    - Lock
  - RobotA
  - RobotB

Process Operations
- Observations

Operational Data

Physical Equipment
- Pre-Assembly
  - Assembly
  - Finishing

Physical System

As Knowledge Graph
- AssemblyLineA
  - hasPart
    - Assembly Robot
  - hasPart
    - Finishing Station
  - involvedIn
    - Assembly
  - involvedIn
    - Finishing
  - hasPart
    - Base Part
    - Headrest

As Event Time-series
- High Energy (Low Energy, Low Energy, Axis-5, Crash)
- High Energy (Low Energy, Low Energy, Axis-3, Crash)
Outline

Industrial Knowledge Graphs
• Digital Twin of Manufacturing System

Missing Information in Engineering
• Motivating Example

Statistical Relational Learning
• Preliminaries
• Background Enhancement

Event Representation
• First-class citizens

Our Approach
• Combined objective function
• Link prediction results

Conclusion
• Major Take-Aways
Missing Information in Engineering

Example

AssembleLineA

hasPart hasPart

Lock Robot connectedTo Finishing Station

involvedIn involvedIn

Prepare & Assembly

followedBy

Finishing

Before
Missing Information in Engineering Example

Before

New Assembly Process Entity

Prepare & Assembly

Finishing Station

Lock Robot connectedTo Finishing Station

hasPart hasPart

involvedIn involvedIn

Prepare & Assembly followedBy Finishing

AssembleLineA

Production Engineer
Missing Information in Engineering

Example

Before

AssemblyLineA

hasPart

Lock Robot

connectedTo

Finishing Station

involvedIn

Prepare & Assembly

followedBy

Finishing

New Assembly Process Entity

New Assembly Robot Entity

Production Engineer
Missing Information in Engineering

Example

Before

After
Missing Information in Engineering

Example

Can we infer these missing links?

Before

After
Outline

**Industrial Knowledge Graphs**
- Digital Twin of Manufacturing System

**Missing Information in Engineering**
- Motivating Example

**Statistical Relational Learning**
- Preliminaries
- Background Enhancement

**Event Representation**
- First-class citizens

**Our Approach**
- Combined objective function
- Link prediction results

**Conclusion**
- Major Take-Aways
Statistical Relational Learning
Preliminaries

As Probabilistic Knowledge Graph

\[ T_{e_i, r_k, e_j} = \begin{cases} 1, & \text{if } (e_i, r_k, e_j) \text{ is true} \\ 0, & \text{otherwise} \end{cases} \]

→ Each triple is a binary random variable

\[ T_{e_i, r_k, e_j} \sim \text{Ber}(p) \]

Probability \( p \) is modeled as scoring function \( f \) over latent features of entities and relations

\[ p \approx f(\theta_{e_i, r_k, e_j}) \]

→ Optimize \( f \) by learning these latent features
Latent features $\theta_{e_i,r_k,e_j}$:

Each entity is modeled as d-dimensional vector $\vec{e_i} \in \mathbb{R}^d$

Each relation is modeled as d-dimensional vector $\vec{r_k} \in \mathbb{R}^d$ (or matrix $R_k \in \mathbb{R}^d \times d$)

TransE model:

$$f \left( \theta_{e_i,r_k,e_j} \right) = -\| \vec{e_i} + \vec{r_k} - \vec{e_j} \|$$
Latent features $\theta_{e_i, r_k, e_j}$:

Each entity is modeled as a $d$-dimensional vector $\vec{e}_i \in \mathbb{R}^d$

Each relation is modeled as a $d$-dimensional vector $\vec{r}_k \in \mathbb{R}^d$ (or matrix $R_k \in \mathbb{R}^{d \times d}$)

TransE model:

$$f\left(\theta_{e_i, r_k, e_j}\right) = -\| \vec{e}_i + \vec{r}_k - \vec{e}_j \|$$
Statistical Relational Learning
Representation Learning

Latent features $\theta_{e_i,r,k,e_j}$:

Each entity is modeled as $d$-dimensional vector $\overrightarrow{e_i} \in \mathbb{R}^d$

Each relation is modeled as $d$-dimensional vector $\overrightarrow{r_k} \in \mathbb{R}^d$ (or matrix $R_k \in \mathbb{R}^{d \times d}$)

TransE model:

$$f\left(\theta_{e_i,r_k,e_j}\right) = -\| \overrightarrow{e_i} + \overrightarrow{r_k} - \overrightarrow{e_j} \|$$
Statistical Relational Learning
Representation Learning

Latent features $\theta_{e_i, r_k, e_j}$:

Each entity is modeled as $d$-dimensional vector $\vec{e}_i \in \mathbb{R}^d$

Each relation is modeled as $d$-dimensional vector $\vec{r}_k \in \mathbb{R}^d$ (or matrix $R_k \in \mathbb{R}^{d \times d}$)

TransE model:

$$f\left(\theta_{e_i, r_k, e_j}\right) = -\| \vec{e}_i + \vec{r}_k - \vec{e}_j \|$$
Learning latent features:

Additional background information can be used to enhance the latent features, most notably:

- Text, e.g. Wikipedia entries of entities / relations

Barack Obama

From Wikipedia, the free encyclopedia

"Barack" and "Obama" redirect here. For other uses, see Barack (disambiguation) and Obama (disambiguation).

Learning latent features:

Additional background information can be used to enhance the latent features, most notably:

- Text, e.g. Wikipedia entries of entities / relations
- Rules, e.g. Transitivity $r_a(x, z) \leftarrow r_b(x, y), r_c(y, z)$
Learning latent features:
Additional background information can be used to enhance the latent features, most notably:

- Text, e.g. Wikipedia entries of entities / relations
- Rules, e.g. Transitivity $r_a(x, z) \leftarrow r_b(x, y), r_c(y, z)$
- Type-constraints, e.g. domain of bornIn is Person
Learning latent features:
Additional background information can be used to enhance the latent features, most notably:

- Text, e.g. Wikipedia entries of entities / relations
- Rules, e.g. Transitivity $r_a(x, z) \leftarrow r_b(x, y), r_c(y, z)$
- Type-constraints, e.g. domain of $bornIn$ is $Person$
- ... and Events, e.g. alarms of production equipment

Events are fundamentally different!
Outline

Industrial Knowledge Graphs
  • Digital Twin of Manufacturing System

Missing Information in Engineering
  • Motivating Example

Statistical Relational Learning
  • Preliminaries
  • Representation Learning

Event Representation
  • First-class citizens

Our Approach
  • Combined objective function
  • Link prediction results

Conclusion
  • Major Take-Aways
Event Representation
First-class citizens

Event as entity in the Knowledge Graph

Event can have multiple occurrences as token

\[
S_1 = (\text{High Energy}, \text{Low Energy}, \text{Low Energy}, \text{Axis-5 Crash})
\]

\[
S_2 = (\text{High Energy}, \text{Low Energy}, \text{Low Energy}, \text{Axis-3 Crash})
\]

- Exact order is important
- Only small part of the KG is affected by Events
Event Representation
Shared latent features

Event as entity in the Knowledge Graph

Robot

hasSource

Low Energy

belongsToGroup

Energy Events

Process

effects

Combined vector representation

\( e_i \)

\( S_1 \) (High Energy, Low Energy, Low Energy, Axis-5 Crash)

\( S_2 \) (High Energy, Low Energy, Low Energy, Axis-3 Crash)

Event can have multiple occurrences as token

- Exact order is important
- Only small part of the KG is affected by Events
Event Representation
Shared latent features

Event as entity in the Knowledge Graph

- Event can have multiple occurrences as token

**Combined vector representation**

\[ e_i \]

- Exact order is important
- Only small part of the KG is affected by Events
- Can also have missing links for event entities

**Diagram:**
- Robot
  - hasSource
- Energy Events
  - belongsToGroup
- Low Energy
  - effects
  - Process
- High Energy
  - Low Energy
  - Axis-5 Crash

**Equations:**

\[ S_1 = (\text{High Energy, Low Energy, Low Energy, Axis-5 Crash}) \]
\[ S_2 = (\text{High Energy, Low Energy, Low Energy, Axis-3 Crash}) \]
Outline

Industrial Knowledge Graphs
• Digital Twin of Manufacturing System

Missing Information in Engineering
• Motivating Example

Statistical Relational Learning
• Preliminaries
• Background Enhancement

Event Representation
• First-class citizens

Our Approach
• Combined objective function
• Link prediction results

Conclusion
• Major Take-Aways
Our Approach (1)
Shared entity representations

\[ L_{\text{joint}} = L_K + \alpha L_S \]

\[ S = \{ (\text{Low Energy}, \text{Low Energy}, \text{High Energy}) \} \]

\[ K = \{ (\text{Low Energy}, \text{hasSource}, \text{Robot}) \} \]

Joint Optimization:
- Mini-batches of event sequences and triples
- Grid search-based hyperparameter tuning \( d \in \{40, 60, 80\}, m \in \{3, 5, 7\}, \eta, \alpha, \ldots \)
- Max. 100 Epochs with early stopping and Bernoulli-based negative sampling
Our Approach (2)

Weighted combination of entity representations

\[
L_{joint} = L_K + \alpha L_S
\]

\[
S = \{(\text{Low Energy, Low Energy, High Energy})\}
\]

\[
K = \{(\text{Low Energy, hasSource, Robot})\}
\]

\[
\tilde{e}_1 = e_{12} \otimes e_{11} := e_{12} \circ a_r + e_{11} \circ b_r
\]

Relation specific weighting
Our Approach – Event Context
Learning Latent Features of Events

- log σ (e₃⁺ c) - log σ (e₃⁻ c)

Full context concatenation - $EKL_{full}$

$\mathbf{c}_i = \bigoplus_{j=1}^{\lfloor \frac{m}{2} \rfloor} \mathbf{e}_{i-j} \bigoplus_{j=1}^{\lfloor \frac{m}{2} \rfloor} \mathbf{e}_{i+j}$

Causation context concatenation - $EKL_{cause}$

$\mathbf{c}_i = \bigoplus_{j=1}^{m-1} \mathbf{e}_{i-j}$
Our Approach – Event Context
Learning Latent Features of Events

Notable mentions:
- Recurrent Neural Networks
- Sequence Auto-Encoders
- Skipgram / Co-occurrence (TEKE)
### Results

#### Link Prediction

| Dataset      | $|\mathcal{V}|$ | $|\mathcal{S}|$ | $|\mathcal{E}|$     | $|\mathcal{R}|$ |
|--------------|-------------|-------------|-----------------|-------------|
| Manufacturing | 6,791       | 56,000      | 3,180 (728)     | 10          |
| Traffic      | 11,000      | 157,000     | 13,113 (4,000)  | 5           |

**Shared / combined architecture**

### Table: Model Performance

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean Rank</th>
<th>Hits@10</th>
<th>Hits@3</th>
<th>Hits@1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dataset: Manufacturing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TransE</td>
<td>317</td>
<td>36.1</td>
<td>23.2</td>
<td>7.5</td>
</tr>
<tr>
<td>TEK_E</td>
<td>596</td>
<td>24.5</td>
<td>10.8</td>
<td>3.6</td>
</tr>
<tr>
<td>EKL$_{Full}$</td>
<td>285 / 663</td>
<td>37.9 / 23.5</td>
<td>25.0 / 12.3</td>
<td>8.0 / 4.8</td>
</tr>
<tr>
<td>EKL$_{Cause}$</td>
<td>280 / 691</td>
<td>38.1 / 21.4</td>
<td>25.8 / 11.5</td>
<td>7.4 / 5.2</td>
</tr>
<tr>
<td>EKL$_{Auto}$</td>
<td>302 / 692</td>
<td>34.5 / 22.5</td>
<td>23.6 / 10.1</td>
<td><strong>9.6 / 2.7</strong></td>
</tr>
<tr>
<td><strong>Dataset: Traffic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TransE</td>
<td>4126</td>
<td>26.8</td>
<td>24.6</td>
<td>9.5</td>
</tr>
<tr>
<td>TEK_E</td>
<td>897</td>
<td>25.3</td>
<td>22.6</td>
<td>18.9</td>
</tr>
<tr>
<td>EKL$_{Full}$</td>
<td>1118 / 758</td>
<td>27.0 / 27.3</td>
<td><strong>25.3 / 24.5</strong></td>
<td>21.1 / 20.6</td>
</tr>
<tr>
<td>EKL$_{Cause}$</td>
<td>999 / 783</td>
<td>27.2 / 27.0</td>
<td>24.7 / 24.4</td>
<td>20.0 / 20.5</td>
</tr>
<tr>
<td>EKL$_{Auto}$</td>
<td>944 / 840</td>
<td>27.5 / <strong>27.7</strong></td>
<td>24.8 / 24.8</td>
<td><strong>22.2 / 20.6</strong></td>
</tr>
</tbody>
</table>

Results
Link Prediction of „Zero-Shot“ Event Entities

- TEKE
- EKL\textsubscript{Cause}
- EKL\textsubscript{Full}
- EKL\textsubscript{Auto}
Outline

Industrial Knowledge Graphs
• Digital Twin of Manufacturing System

Missing Information in Engineering
• Motivating Example

Statistical Relational Learning
• Preliminaries
• Background Enhancement

Event Representation
• First-class citizens

Our Approach
• Combined objective function
• Link prediction results

Conclusion
• Major Take-Aways
Conclusion
Major Take-Aways

• Synchronization of Digital Twins can be formulated as Knowledge Graph completion
• Events as background enhancement need specialized embedding models
• Event embeddings can improve KG completion for some parts
• Future challenge: Detect and introduce missing entities

https://github.com/NetherNova/event-kge

Trans(E,H), RESCAL, TEKE
Auto-Encoders, Concatenation

Martin Ringsquandl
Siemens AG, Corporate Technology
Semantics and Reasoning
martin.ringsquandl@siemens.com