CORE: Context-Aware Open Relation Extraction with Factorization Machines

Fabio Petroni

Luciano Del Corro

Rainer Gemulla
Open relation extraction

- Open relation extraction is the task of extracting new facts for a potentially unbounded set of relations from various sources.

natural language text + knowledge bases
Enrico Fermi was a professor in theoretical physics at Sapienza University of Rome.

Input data: facts from natural language text

**Surface relation**

**Open information extractor**

**Surface fact**

"professor at"(Fermi,Sapienza)

**Tuple**
Input data: facts from knowledge bases

- **KB fact**: employee(Fermi, Sapienza)
- **KB relation**: "professor at"(Fermi, Sapienza)
- **Entity link**: e.g., string match heuristic

Relation extraction techniques taxonomy

**Relation extraction**

- **Close in-KB**
  - **Distant supervision**
    - set of predefined relations
  - **Tensor completion**
    - RESCAL (Nickel et al., 2011)
    - limited scalability with the number of relations; large prediction space
  - **Relation clustering**
    - "black and white" approach

- **Open in-KB out of-KB**
  - **Latent factors models**
    - PITF (Drumond et al., 2012)
  - **Matrix completion**
    - NFE (Riedel et al., 2013)
    - CORE (Petroni et al., 2015)
    - restricted prediction space
## Matrix completion for open relation extraction

<table>
<thead>
<tr>
<th>Tuple</th>
<th>Surface Relation</th>
<th>KB Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Caesar, Rome)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(Fermi, Rome)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(Fermi, Sapienza)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(de Blasio, NY)</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

*Employee, born in, professor at, mayor of*

*Example: tuples x relations*
Matrix completion for open relation extraction

| (Caesar, Rome) | ? | ? | ? | ? |
| (Fermi, Rome) | 1 | ? | ? | ? |
| (Fermi, Sapienza) | ? | 1 | ? | 1 |
| (de Blasio, NY) | ? | ? | 1 | ? |

- **KB relation**
- **Surface relation**
- **Employee**
- **Professor at**
- **Mayor of**

**tuples x relations**
Matrix factorization

- learn latent semantic representations of tuples and relations

\[
\begin{align*}
&\text{tuple latent factor vector} \\
&\text{relation latent factor vector}
\end{align*}
\]

- leverage latent representations to predict new facts

(\text{Fermi, Sapienza})

\[
\begin{pmatrix}
0.8 \\
-0.5
\end{pmatrix}
\begin{pmatrix}
0.9 \\
-0.3
\end{pmatrix}
\]

- in real applications latent factors are uninterpretable

\text{related with science}

\text{related with sport}
Matrix factorization

- learn latent semantic representations of tuples and relations

\[ \begin{array}{c|c|c}
\text{(Term, Sapienza)} & \text{professor at} & 0.8 \\
\hline
\text{related with science} & -0.5 & 0.9 \\
\text{related with sport} & -0.3 &
\end{array} \]

CORE integrates contextual information into such models to improve prediction performance

- in real applications latent factors are uninterpretable
Tom Peloso joined Modest Mouse to record their fifth studio album.
Contextual information

How to incorporate contextual information within the model?
CORE - latent representations of variables

- associates latent representations $f_v$ with each variable $v \in V$

<table>
<thead>
<tr>
<th>tuple</th>
<th>(Peloso, Modest Mouse)</th>
</tr>
</thead>
<tbody>
<tr>
<td>relation</td>
<td>join</td>
</tr>
<tr>
<td>entities</td>
<td>Peloso</td>
</tr>
<tr>
<td></td>
<td>Modest Mouse</td>
</tr>
<tr>
<td>context</td>
<td>Music</td>
</tr>
<tr>
<td></td>
<td>record</td>
</tr>
<tr>
<td></td>
<td>album</td>
</tr>
</tbody>
</table>

latent factor vectors
CORE - modeling facts

- models the input data in terms of a matrix in which each row corresponds to a fact \( x \) and each column to a variable \( v \)
- groups columns according to the type of the variables
- in each row the values of each column group sum up to unity
CORE - modeling context

- aggregates and normalizes contextual information by tuple
  - a fact can be observed multiple times with different context
  - there is no context for new facts (never observed in input)
- this approach allows us to provide comprehensive contextual information for both observed and unobserved facts
CORE - factorization model

- uses factorization machines as underlying framework
- associates a score $s(x)$ with each fact $x$

$$s(x) = \sum_{v_1 \in V} \sum_{v_2 \in V \setminus \{v_1\}} x_{v_1} x_{v_2} f_{v_1}^T f_{v_2}$$

- weighted pairwise interactions of latent factor vectors
CORE - prediction

produce a ranked list of tuples for each relation

given a relation

- rank reflects the likelihood that the corresponding fact is true
- to generate this ranked list:
  - fix a relation $r$
  - retrieve all tuples $t$, s.t. the fact $r(t)$ is not observed
  - add tuple context
  - rank unobserved facts by their scores
CORE - parameter estimation

- parameters: $\Theta = \{ b_v, f_v \mid v \in V \}$
- all our observations are positive, no negative training data
- Bayesian personalized ranking, open-world assumption

- pairwise approach, $x$ is more likely to be true than $x-$

$$\maximize \sum_x f(s(x) - s(x-))$$

- stochastic gradient ascent

$$\Theta \leftarrow \Theta + \eta \nabla_{\Theta} (\cdot)$$
Experiments - dataset

440k facts extracted from The New York Times corpus

15k facts from Freebase

Contextual information

- article metadata
  - news desk (e.g., foreign desk)
  - descriptors (e.g., finances)
  - online section (e.g., sports)
  - section (e.g., a, d)
  - publication year

- entity type
  - person
  - organization
  - location
  - miscellaneous

- bag-of-word sentences where the fact has been extracted

letters to indicate contextual information considered
Experiments - methodology

- we consider (to keep experiments feasible):
  - 10k tuples
  - 19 Freebase relations
  - 10 surface relations

- for each relation and method:
  - we rank the tuples subsample
  - we consider the top-100 predictions and label them manually

- evaluation metrics:
  - number of true facts
  - MAP (quality of the ranking)

- methods:
  - **PITF**, tensor factorization method
  - **NFE**, matrix completion method (context-agnostic)
  - **CORE**, uses relations, tuples and entities as variables
  - **CORE+m, +t, +w, +mt, +mtw**
### Results - Freebase relations

<table>
<thead>
<tr>
<th>Relation</th>
<th>#</th>
<th>PITF (0.47)</th>
<th>NFE (0.81)</th>
<th>CORE (0.83)</th>
<th>CORE+m (0.84)</th>
<th>CORE+t (0.87)</th>
<th>CORE+w (0.87)</th>
<th>CORE+mt (0.93)</th>
<th>CORE+mtw (0.94)</th>
</tr>
</thead>
<tbody>
<tr>
<td>person/company</td>
<td>208</td>
<td>70</td>
<td>92</td>
<td>91</td>
<td>90</td>
<td>91</td>
<td>92</td>
<td>95</td>
<td>96</td>
</tr>
<tr>
<td>person/place_of_birth</td>
<td>117</td>
<td>1</td>
<td>92</td>
<td>90</td>
<td>92</td>
<td>92</td>
<td>89</td>
<td>93</td>
<td>92</td>
</tr>
<tr>
<td>location/containedby</td>
<td>102</td>
<td>7</td>
<td>63</td>
<td>62</td>
<td>63</td>
<td>61</td>
<td>61</td>
<td>62</td>
<td>68</td>
</tr>
<tr>
<td>parent/child</td>
<td>88</td>
<td>9 (0.01)</td>
<td>64 (0.6)</td>
<td>64 (0.56)</td>
<td>64 (0.59)</td>
<td>64 (0.62)</td>
<td>64 (0.57)</td>
<td>67 (0.67)</td>
<td>68 (0.63)</td>
</tr>
<tr>
<td>person/place_of_death</td>
<td>71</td>
<td>1 (0.0)</td>
<td>67 (0.93)</td>
<td>67 (0.92)</td>
<td>69 (0.94)</td>
<td>67 (0.93)</td>
<td>67 (0.92)</td>
<td>69 (0.94)</td>
<td>67 (0.92)</td>
</tr>
<tr>
<td>person/parts</td>
<td>67</td>
<td>20 (0.1)</td>
<td>51 (0.64)</td>
<td>52 (0.62)</td>
<td>51 (0.61)</td>
<td>49 (0.64)</td>
<td>47 (0.6)</td>
<td>53 (0.67)</td>
<td>53 (0.65)</td>
</tr>
<tr>
<td>author/works_written</td>
<td>65</td>
<td>24 (0.08)</td>
<td>45 (0.59)</td>
<td>49 (0.62)</td>
<td>51 (0.69)</td>
<td>50 (0.68)</td>
<td>50 (0.68)</td>
<td>51 (0.7)</td>
<td>52 (0.67)</td>
</tr>
<tr>
<td>person/nationality</td>
<td>61</td>
<td>21 (0.08)</td>
<td>25 (0.19)</td>
<td>27 (0.17)</td>
<td>28 (0.2)</td>
<td>26 (0.2)</td>
<td>29 (0.19)</td>
<td>27 (0.18)</td>
<td>27 (0.21)</td>
</tr>
<tr>
<td>neighbor/neighborhood_of</td>
<td>39</td>
<td>3 (0.0)</td>
<td>24 (0.44)</td>
<td>23 (0.45)</td>
<td>26 (0.5)</td>
<td>27 (0.47)</td>
<td>27 (0.49)</td>
<td>30 (0.51)</td>
<td>30 (0.52)</td>
</tr>
</tbody>
</table>

**Average MAP**

<table>
<thead>
<tr>
<th>Relation</th>
<th>#</th>
<th>NFE</th>
<th>CORE</th>
<th>CORE+m</th>
<th>CORE+t</th>
<th>CORE+w</th>
<th>CORE+mt</th>
<th>CORE+mtw</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.09</td>
<td>0.46</td>
<td>0.47</td>
<td>0.49</td>
<td>0.47</td>
<td>0.49</td>
<td>0.51</td>
</tr>
<tr>
<td>Weighted Average MAP</td>
<td>#</td>
<td>0.14</td>
<td>0.64</td>
<td>0.64</td>
<td>0.66</td>
<td>0.67</td>
<td>0.66</td>
<td>0.70</td>
</tr>
</tbody>
</table>

---

## Results - surface relations

<table>
<thead>
<tr>
<th>Relation</th>
<th>#</th>
<th>PITF</th>
<th>NFE</th>
<th>CORE</th>
<th>CORE+m</th>
<th>CORE+t</th>
<th>CORE+w</th>
<th>CORE+mt</th>
<th>CORE+mtw</th>
</tr>
</thead>
<tbody>
<tr>
<td>head</td>
<td>162</td>
<td>34 (0.18)</td>
<td>80 (0.66)</td>
<td>83 (0.66)</td>
<td>82 (0.63)</td>
<td>76 (0.57)</td>
<td>77 (0.57)</td>
<td>83 (0.69)</td>
<td><strong>88 (0.73)</strong></td>
</tr>
<tr>
<td>scientist</td>
<td>144</td>
<td>44 (0.17)</td>
<td>76 (0.6)</td>
<td>74 (0.55)</td>
<td>73 (0.56)</td>
<td>74 (0.6)</td>
<td>73 (0.59)</td>
<td>78 (0.66)</td>
<td><strong>78 (0.69)</strong></td>
</tr>
<tr>
<td>base</td>
<td>133</td>
<td>10 (0.01)</td>
<td>85 (0.71)</td>
<td>86 (0.71)</td>
<td>86 (0.78)</td>
<td>88 (0.79)</td>
<td>85 (0.75)</td>
<td>83 (0.76)</td>
<td><strong>89 (0.8)</strong></td>
</tr>
<tr>
<td>visit</td>
<td>118</td>
<td>4 (0.0)</td>
<td>73 (0.6)</td>
<td>75 (0.61)</td>
<td>76 (0.64)</td>
<td>80 (0.68)</td>
<td>74 (0.64)</td>
<td>75 (0.66)</td>
<td><strong>82 (0.74)</strong></td>
</tr>
<tr>
<td>attend</td>
<td>92</td>
<td>11 (0.02)</td>
<td>65 (0.58)</td>
<td>64 (0.59)</td>
<td>65 (0.63)</td>
<td>62 (0.6)</td>
<td>66 (0.63)</td>
<td>62 (0.58)</td>
<td><strong>69 (0.64)</strong></td>
</tr>
<tr>
<td>adviser</td>
<td>56</td>
<td>2 (0.0)</td>
<td>42 (0.56)</td>
<td><strong>47 (0.58)</strong></td>
<td>44 (0.58)</td>
<td>43 (0.59)</td>
<td>45 (0.63)</td>
<td>43 (0.53)</td>
<td>44 (0.63)</td>
</tr>
<tr>
<td>criticize</td>
<td>40</td>
<td>5 (0.0)</td>
<td>31 (0.66)</td>
<td>33 (0.62)</td>
<td>33 (0.7)</td>
<td>33 (0.67)</td>
<td>33 (0.61)</td>
<td>35 (0.69)</td>
<td><strong>37 (0.69)</strong></td>
</tr>
<tr>
<td>support</td>
<td>33</td>
<td>3 (0.0)</td>
<td>19 (0.27)</td>
<td>22 (0.28)</td>
<td>18 (0.21)</td>
<td>19 (0.28)</td>
<td>22 (0.27)</td>
<td><strong>23 (0.27)</strong></td>
<td>21 (0.27)</td>
</tr>
<tr>
<td>praise</td>
<td>5</td>
<td>0 (0.0)</td>
<td>2 (0.0)</td>
<td>2 (0.01)</td>
<td>4 (0.03)</td>
<td>3 (0.01)</td>
<td>3 (0.02)</td>
<td><strong>5 (0.03)</strong></td>
<td>2 (0.01)</td>
</tr>
<tr>
<td>vote</td>
<td>3</td>
<td>2 (0.01)</td>
<td>3 (0.63)</td>
<td>3 (0.63)</td>
<td>3 (0.32)</td>
<td>3 (0.49)</td>
<td>3 (0.51)</td>
<td>3 (0.59)</td>
<td><strong>3 (0.64)</strong></td>
</tr>
</tbody>
</table>

**Average MAP**

<table>
<thead>
<tr>
<th>Relation</th>
<th>#</th>
<th>PITF</th>
<th>NFE</th>
<th>CORE</th>
<th>CORE+m</th>
<th>CORE+t</th>
<th>CORE+w</th>
<th>CORE+mt</th>
<th>CORE+mtw</th>
</tr>
</thead>
<tbody>
<tr>
<td>head</td>
<td>100</td>
<td>0.04</td>
<td>0.53</td>
<td>0.53</td>
<td>0.51</td>
<td>0.53</td>
<td>0.53</td>
<td>0.55</td>
<td>0.59</td>
</tr>
<tr>
<td>scientist</td>
<td>100</td>
<td>0.08</td>
<td>0.62</td>
<td>0.61</td>
<td>0.63</td>
<td>0.63</td>
<td>0.61</td>
<td>0.65</td>
<td>0.70</td>
</tr>
</tbody>
</table>

Weighted Average MAP

- NFE: 0.62
- CORE: 0.61
- CORE+m: 0.63
- CORE+t: 0.63
- CORE+w: 0.61
- CORE+mt: 0.65
- CORE+mtw: 0.70
Anecdotal results

**author(x,y)**

<table>
<thead>
<tr>
<th>ranked list of tuples</th>
<th>similar relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Winston Groom, Forrest Gump)</td>
<td>0.98 “reviews x by y”(x,y)</td>
</tr>
<tr>
<td>2 (D. M. Thomas, White Hotel)</td>
<td>0.97 “book by”(x,y)</td>
</tr>
<tr>
<td>3 (Roger Rosenblatt, Life Itself)</td>
<td>0.95 “author of”(x,y)</td>
</tr>
<tr>
<td>4 (Edmund White, Skinned Alive)</td>
<td>0.95 “’s novel”(x,y)</td>
</tr>
<tr>
<td>5 (Peter Manso, Brando: The Biography)</td>
<td>0.95 “’s book”(x,y)</td>
</tr>
</tbody>
</table>

**“scientist at”(x,y)**

<table>
<thead>
<tr>
<th>ranked list of tuples</th>
<th>similar relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Riordan Roett, Johns Hopkins University)</td>
<td>0.87 “scientist”(x,y)</td>
</tr>
<tr>
<td>2 (Dr. R. M. Roberts, University of Missouri)</td>
<td>0.84 “scientist with”(x,y)</td>
</tr>
<tr>
<td>3 (Linda Mayes, Yale University)</td>
<td>0.80 “professor at”(x,y)</td>
</tr>
<tr>
<td>4 (Daniel T. Jones, Cardiff Business School)</td>
<td>0.79 “scientist for”(x,y)</td>
</tr>
<tr>
<td>5 (Russell Ross, University of Iowa)</td>
<td>0.78 “neuroscientist at”(x,y)</td>
</tr>
</tbody>
</table>

- semantic similarity of relations is one aspect of our model
- similar relations treated differently in different contexts
Conclusion

- CORE, a matrix factorization model for open relation extraction that incorporates contextual information
- based on factorization machines and open-world assumption
- extensible model, additional contextual information can be integrated when available
- experimental study suggests that exploiting context can significantly improve prediction performance

- Source code, datasets, and supporting material are available at https://github.com/fabiopetroni/CORE
Thank you!

Questions?

Fabio Petroni
Sapienza University of Rome, Italy

Current position:
PhD Student in Engineering in Computer Science

Research Interests:
data mining, machine learning, big data

petroni@dis.uniroma1.it