Semantic Web and Machine Learning

Time to re-sync

Filip Železný, Czech Technical University in Prague
Bottom Line of the Talk

“Ask not what machine learning can do for you, ask what you can do for machine learning.”

I will show how people try to make the semantic web understandable to machine learning algorithms.
Machine Learning

Data: ground observations, facts

Knowledge: patterns allowing inference
## How data is (usually) encoded in Machine Learning

<table>
<thead>
<tr>
<th>Name</th>
<th>Thread pitch (mm)</th>
<th>Minor diameter tolerance</th>
<th>Nominal diameter (mm)</th>
<th>Head shape</th>
<th>Price for 50 screws</th>
<th>Available at factory outlet?</th>
<th>Number in stock</th>
<th>Flat or Phillips head?</th>
</tr>
</thead>
<tbody>
<tr>
<td>M4</td>
<td>0.7</td>
<td>4g</td>
<td>4</td>
<td>Pan</td>
<td>$10.08</td>
<td>Yes</td>
<td>276</td>
<td>Flat</td>
</tr>
<tr>
<td>M5</td>
<td>0.8</td>
<td>4g</td>
<td>5</td>
<td>Round</td>
<td>$13.89</td>
<td>Yes</td>
<td>183</td>
<td>Both</td>
</tr>
<tr>
<td>M6</td>
<td>1</td>
<td>5g</td>
<td>6</td>
<td>Button</td>
<td>$10.42</td>
<td>Yes</td>
<td>1043</td>
<td>Flat</td>
</tr>
<tr>
<td>M8</td>
<td>1.25</td>
<td>5g</td>
<td>8</td>
<td>Pan</td>
<td>$11.98</td>
<td>No</td>
<td>298</td>
<td>Phillips</td>
</tr>
<tr>
<td>M10</td>
<td>1.5</td>
<td>6g</td>
<td>10</td>
<td>Round</td>
<td>$16.74</td>
<td>Yes</td>
<td>488</td>
<td>Phillips</td>
</tr>
<tr>
<td>M12</td>
<td>1.75</td>
<td>7g</td>
<td>12</td>
<td>Pan</td>
<td>$18.26</td>
<td>No</td>
<td>998</td>
<td>Flat</td>
</tr>
<tr>
<td>M14</td>
<td>2</td>
<td>7g</td>
<td>14</td>
<td>Round</td>
<td>$21.19</td>
<td>No</td>
<td>235</td>
<td>Phillips</td>
</tr>
<tr>
<td>M16</td>
<td>2</td>
<td>8g</td>
<td>16</td>
<td>Button</td>
<td>$23.57</td>
<td>Yes</td>
<td>292</td>
<td>Both</td>
</tr>
<tr>
<td>M18</td>
<td>2.1</td>
<td>8g</td>
<td>18</td>
<td>Button</td>
<td>$25.87</td>
<td>No</td>
<td>664</td>
<td>Both</td>
</tr>
<tr>
<td>M20</td>
<td>2.4</td>
<td>8g</td>
<td>20</td>
<td>Pan</td>
<td>$29.09</td>
<td>Yes</td>
<td>486</td>
<td>Both</td>
</tr>
<tr>
<td>M24</td>
<td>2.55</td>
<td>9g</td>
<td>24</td>
<td>Round</td>
<td>$33.01</td>
<td>Yes</td>
<td>982</td>
<td>Phillips</td>
</tr>
<tr>
<td>M28</td>
<td>2.7</td>
<td>10g</td>
<td>28</td>
<td>Button</td>
<td>$35.66</td>
<td>No</td>
<td>1067</td>
<td>Phillips</td>
</tr>
<tr>
<td>M36</td>
<td>3.2</td>
<td>12g</td>
<td>36</td>
<td>Pan</td>
<td>$41.32</td>
<td>No</td>
<td>434</td>
<td>Both</td>
</tr>
<tr>
<td>M50</td>
<td>4.5</td>
<td>15g</td>
<td>50</td>
<td>Pan</td>
<td>$44.72</td>
<td>No</td>
<td>740</td>
<td>Flat</td>
</tr>
</tbody>
</table>
How knowledge is encoded in Machine Learning

- **IF** (humidity = high) and (outlook = sunny) THEN play=no (3.0/0.0)
- **IF** (outlook = rainy) and (windy = TRUE) THEN play=no (2.0/0.0)
- **OTHERWISE** play=yes (9.0/0.0)
Semantic Web and Machine Learning (1)

- Learning **concept taxonomies** through **hierarchical clustering**
- Learning **deep annotation** rules, concept/relation population
- Ontology **alignment**
- **Duplicate** detection
- Etc.
- Not the scope of this talk
Lots of motivation: FreeBase, DBpedia, Google Knowledge Graph, ...

Calls for learning algorithms which, apart from data, understand formal knowledge

Taxonomies, rules, etc.
Inductive Logic Programming

- Background knowledge
- Inductive Logic Programming
- New Knowledge
- Data
Inductive Logic Programming

- Background knowledge
  - First-order Logic Theory

- Data
  - Logical Ground Facts

- Inductive Logic Programming

- New Knowledge
  - First-order Logic Theory
ILP: example

Inductive Logic Programming

Chemical structures and molecular formulas are shown, with labels for various compounds.
atom(drug1, c1, carbon).
dbond(drug1, c1, c2).

anthracene(Drug, [Ring1,Ring2])
IF
benzene(Drug,Ring1),
benezene(Drug,Ring2).

mutagenic(drug1).
mutagenic(drug2).
....
not mutagenic(drug3)
not mutagenic(drug4)
....

Srinivasan et al., Artif. Intell. 1996
ILP: basic principle

Search through a subsumption lattice

mutagenic(Drug)

mutagenic(Drug) IF benzene(Drug, Ring)

mutagenic(Drug) IF bond(double, Ring, Z).

mutagenic(Drug) IF benzene(Drug, Ring), bond(double, Ring, Z).
ILP: discussion

- Energetic research since the 90’s
- Mainly in Europe
- Annual ILP conferences (cca 50 people)
  - ILP’08: Semantic Web Keynote by Frank Harmelen
- Current focus: combining with probabilistic inference
  - Probabilistic ILP, Statistical Relational Learning

- Main issues of ILP:
  - **Scalability**, handling **uncertainty**, **numerical** reasoning
  - **First-order** logic (mainly Datalog) not a semantic web standard
ILP with Description Logics

**DL-Learner**: learns concept descriptions from examples (concept instances) and counterexamples, and an OWL-DL ontology

Principle: search through a subsumption lattice as in ILP

Hellman et al. ISWC 2008
Taxonomies as Guidance in ILP

- Rule search co-guided by the Gene Ontology

- General mechanism implemented in ILP system Aleph

M. Zakova et al, ECML 2007

A Full Hybrid Approach

Learning views (rules) from a Datalog base AND an ontology

Example e.g. happy(Mary), counter-example happy(Paul)

\[
\text{RICH}(X) \leftarrow \text{famous}(X), \neg \text{scientist}(X)
\]

\[
\text{RICH\(\cup\)} \text{UNMARRIED} \subseteq \exists \text{WANTS-TO-MARRY}^-.T
\]

\[
\text{WANTS-TO-MARRY} \sqcap \text{LOVES}
\]

\[
\text{UNMARRIED}(\text{Mary})
\]

\[
\text{UNMARRIED}(\text{Joe})
\]

\[
\text{happy}(X) \leftarrow \text{famous}(X), \text{WANTS-TO-MARRY}(Y,X)
\]

F. Lisi: TPLP, 2010
Semantic Web and Machine Learning: Workaround

Knowledge base (+data) → Machine Learning → New Knowledge

Produce a data table which also accounts for background knowledge. Then use your favorite, fantastic machine learning algorithm.
Propositionalization

- An ILP-inspired technique to automate the workaround
- First-order features constructed via combinatorial search
- Each is true or false for a particular sample

<table>
<thead>
<tr>
<th></th>
<th>Feature 1</th>
<th>Feature 2</th>
<th>Feature 3</th>
<th>class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Sample 2</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Propositionalization

- A technique to automate the workaround
- First-order features constructed via *combinatorial search*
- Each is true or false for a particular sample

```
<table>
<thead>
<tr>
<th>Feature 1</th>
<th>Feature 2</th>
<th>Feature 3</th>
<th>mutagenic?</th>
</tr>
</thead>
<tbody>
<tr>
<td>drug 1</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>drug 2</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>
```

`benzene(Drug, Ring1), benzene(Drug, Ring2), not Ring1=Ring2`
**Propositionalization: state of the art**

- **Treeliker**: a fast algorithm for propositionalization
- Open SW, comes with documentation and example data, google it
- Produces (up to) very complex features efficiently, from very complex data/knowledge
- Prolog/Datalog representation, no support for SW standards (yet)

Kuzelka et al., Machine Learning 2011; ECML 2011
Szaboova et al., Proteome Sci. 2012
Ontopropositionalization (\(?)\)

SELECT ?e ?r
WHERE { (m ?r ?e) UNION (?e ?r m) }

\[
\begin{align*}
\phi(\text{Nikita}) &= 1; \\
\phi(\text{Léon}) &= 1; \\
\phi(\text{Black Swan}) &= 0
\end{align*}
\]

- European Movie, subclass American Movie
- Action Movie, hasGenre
- Thriller Movie, hasGenre
- French Movie, type
- Hollywood Movie, type
- Jean Reno, actedIn
- Anne Parillaud, actedIn
- Patrice Ledoux, produced
- Luc Besson, directed
- Scott Franklin, produced
- Mark Heyman, directed
- Natalie Portman, actedIn

error in predicting viewer feedback

W. Cheng, CKMI 2011
Kernels for RDF Data: another Workaround

Knowledge base (+data) → Machine Learning → New Knowledge

\[ \kappa(x, y) = \langle \phi(x), \phi(y) \rangle \]
Kernels for RDF Data

Knowledge base (+data)

Any RDF graph

Machine Learning

$\kappa(x, y) = \langle \phi(x), \phi(y) \rangle$

New Knowledge

Used successfully for property-value and link prediction.

U. Loesch: ESWC 2012
Knowledge Graph as a Tensor

Probabilistic Model

- The RESCAL bilinear model

\[
\begin{align*}
\mathbf{Y}_k & \approx \mathbf{E} \mathbf{W}_k \mathbf{E}^T \\
\mathbf{f}_{ijk}^{\text{RESCAL}} & := \mathbf{e}_i^T \mathbf{W}_k \mathbf{e}_j = \sum_{a=1}^{H_e} \sum_{b=1}^{H_e} w_{abk} e_{ia} e_{jb}
\end{align*}
\]

Learned vector embeddings (latent variables)

Result: Similar vectors - similar semantics

Nickel et al., ICML 2011
Probabilistic Model

- The ER-MLP model (Google’s “Knowledge Vault”)

\[ f_{ijk}^{ER-MLP} := w^\top g(h_{ijk}^c) \]
\[ h_{ijk}^c := C^\top \phi_{ijk}^{ER-MLP} \]
\[ \phi_{ijk}^{ER-MLP} := [e_i; e_j; r_k]. \]

Dong et al., KDD 2014
Probabilistic Model

\[ f_{ijk}^{\text{ER-MLP}} := w^T g(h_{ijk}^c) \]
\[ h_{ijk}^c := C^T \phi_{ijk}^{\text{ER-MLP}} \]
\[ \phi_{ijk}^{\text{ER-MLP}} := [e_i; e_j; r_k]. \]

Dong et al., KDD 2014
Embeddings Capture Semantics

\[ f_{ijk}^{\text{ER-MLP}} := w^\top g(h_{ijk}^c) \]
\[ h_{ijk}^c := C^\top \phi_{ijk}^{\text{ER-MLP}} \]
\[ \phi_{ijk}^{\text{ER-MLP}} := [e_i; e_j; r_k]. \]

<table>
<thead>
<tr>
<th>Relation</th>
<th>Nearest Neighbors</th>
</tr>
</thead>
<tbody>
<tr>
<td>children</td>
<td>parents (0.4)</td>
</tr>
<tr>
<td>birth-date</td>
<td>children (1.24)</td>
</tr>
<tr>
<td>edu-end</td>
<td>job-start (1.41)</td>
</tr>
<tr>
<td></td>
<td>spouse (0.5)</td>
</tr>
<tr>
<td></td>
<td>gender (1.25)</td>
</tr>
<tr>
<td></td>
<td>edu-start (1.61)</td>
</tr>
<tr>
<td></td>
<td>birth-place (0.8)</td>
</tr>
<tr>
<td></td>
<td>parents (1.29)</td>
</tr>
<tr>
<td></td>
<td>job-end (1.74)</td>
</tr>
</tbody>
</table>

Dong et al., KDD 2014
Other Recent Interesting Approaches

- Gaussian embeddings
  - Dos Santos et al, ECML 2016
  - Embed entities as normal random vars
  - Conflicting “gradient forces” result in greater variance

- “Injecting” logical knowledge into conventional learning paradigms
  - Diligenti et al, MLJ 2012: Horn rules to regularize SVM’s
  - Rocktaeschel et al. NAACL 2015: Tensor factorization guided by logical rules

- General, encompassing SRL frameworks
  - Such as Markov Logic Networks [Richardson, Domingos - MLJ 2006]
  - Scalability issues
Other Recent Interesting Approaches (2)

- **Lifted Relational Neural Networks**
  - Hybrid FOL-Neural paradigm
  - A weighted FOL theory is a template to derive ground neural networks
  - Sourek et al., NIPS Neural Symb Integr. Workshop 2015, ILP 2016

- **Differentiable databases**
  - Compiles a deductive database s.t. inference is a differentiable function
  - Cohen, NIPS 2016

- **Machine Learning ontologies**
  - Used to plan learning processes
  - Zakova et al., IEEE T-ASE, 2007; Vavpetic, PhD thesis 2016
Conclusions

● Exciting recent research in ML triggered by the growing SW
● Some great novel concepts include
  ○ Vector and other embeddings of knowledge graphs
  ○ Tensor and neural learning approaches
  ○ Hybrid symbolic-neural (or, vector space) approaches
● Some standing challenges
  ○ E.g. embedding non-ground knowledge