Learning to Associate DBpedia Entities like Humans

Jörn Hees

2017-09-14
DBpedia Meetup Amsterdam
Jörn Hees

• Researcher at DFKI & TU Kaiserslautern
• Linked Data, Machine Learning, AI
• An RDFLib maintainer (Python)
• http://joernhees.de, @joernhees
• https://w3id.org/associations
What are Human Associations?

• Mental connections between concepts
• What's the first thing that comes to your mind when thinking about ... ?
• Example:
  • Dog
What are Human Associations?

• Mental connections between concepts

• What's the first thing that comes to your mind when thinking about ...?

• Example:
  • Dog: Cat, collar, leash, walk, fur, bark
What are Human Associations?

• Mental connections between concepts
• What's the first thing that comes to your mind when thinking about ...?
• Example:
  • Dog: Cat, collar, leash, walk, fur, bark
  • House
What are Human Associations?

• Mental connections between concepts
• What's the first thing that comes to your mind when thinking about ... ?
• Example:
  • Dog: Cat, collar, leash, walk, fur, bark
  • House: Roof, door, window, flat, live
Associations vs. Similarity

• Partially overlapping, but ≠
• Strongly Associated but not Similar:
  • Baby - Crying
• Similar but not Strongly Associated:
  • Dog - Terrier (100 ppl top answers: Cat (57 %), Collar (5 %), bark (2 %))
Outline

• Background
• My Research
• Evaluation
• Demo
Outline

• Background

• My Research

• Evaluation

• Demo
Motivation

• Associations are important for thinking:
  • Navigate from one thought to another
  • “Closeness of concepts in our mind”
    Chris Welty’s First Lady “Nixon” example

• Can we teach machines to do the same?
  • Using their Knowledge?
    • Linked Data
My Research

• Research Question:

• Is it possible to learn patterns for Human Associations from Linked Data?
My Research

- Research Question:
- Is it possible to learn patterns for Human Associations from Linked Data?
My Research

• Research Question:

• Is it possible to learn patterns for Human Associations from Linked Data?

https://w3id.org/associations
My Research

• Research Question:

• Is it possible to learn patterns for Human Associations from Linked Data?
My Research

- Research Question:
  - Is it possible to learn patterns for Human Associations from Linked Data?

- Goal:
  - Given an input node predict the output node(s) we would associate
My Research

• Research Question:
  • Is it possible to learn patterns for Human Associations from Linked Data?

• Dataset of "Semantic Associations" needed
Research Question:

- Is it possible to learn patterns for Human Associations from Linked Data?

- Dataset of "Semantic Associations" needed
Semantic Associations Dataset

- (Raw) Edinburgh Associative Thesaurus (EAT) as RDF (1.7 M triples)

- 727 verified distinct Semantic Associations

https://w3id.org/associations
Semantic Associations Dataset

- Edinburgh Associative Thesaurus (EAT) as RDF (1.7 M triples)
- 727 verified distinct Semantic Associations
Semantic Associations Dataset

- 727 verified distinct Semantic Associations

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbr:Cow</td>
<td>dbr:Milk</td>
</tr>
<tr>
<td>dbr:Camping</td>
<td>dbr:Tent</td>
</tr>
<tr>
<td>dbr:Expense</td>
<td>dbr:Money</td>
</tr>
<tr>
<td>dbr:Bed</td>
<td>dbr:Sleep</td>
</tr>
<tr>
<td>dbr:Pupil</td>
<td>dbr:Eye</td>
</tr>
</tbody>
</table>

- Not readily modelled in DBpedia!
- Not one property!
Machine Learning Outline

Training Data

<table>
<thead>
<tr>
<th>?source</th>
<th>?target</th>
</tr>
</thead>
<tbody>
<tr>
<td>db:Cow</td>
<td>db:Milk</td>
</tr>
<tr>
<td>db:Camping</td>
<td>db:Tent</td>
</tr>
<tr>
<td>db:Bed</td>
<td>db:Sleep</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Graph Pattern Learner

- Pattern Learner
- Fusion Training

SPARQL Endpoint

Training Phase
Machine Learning Outline

**Training Data**

<table>
<thead>
<tr>
<th>?source</th>
<th>?target</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbr:Cow</td>
<td>dbr:Milk</td>
</tr>
<tr>
<td>dbr:Camping</td>
<td>dbr:Tent</td>
</tr>
<tr>
<td>dbr:Bed</td>
<td>dbr:Sleep</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

**Graph Pattern Learner**

- Pattern Learner
- Fusion Training

**SPARQL Endpoint**

**Training Phase**

**Trained Model**

- Graph Patterns
- Fusion Model

Fusion Training
Machine Learning Outline

Training Data

<table>
<thead>
<tr>
<th>?source</th>
<th>?target</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbr:Cow</td>
<td>dbr:Milk</td>
</tr>
<tr>
<td>dbr:Camping</td>
<td>dbr:Tent</td>
</tr>
<tr>
<td>dbr:Bed</td>
<td>dbr:Sleep</td>
</tr>
</tbody>
</table>

Graph Pattern Learner

- Pattern Learner
- Fusion Training

Training Phase

SPARQL Endpoint

Application Phase

New Data

<table>
<thead>
<tr>
<th>?source</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbr:Fish</td>
</tr>
</tbody>
</table>

Trained Model

- Graph Patterns
- Fusion Model

https://w3id.org/associations
Machine Learning Outline

Training Data

<table>
<thead>
<tr>
<th>?source</th>
<th>?target</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbr:Cow</td>
<td>dbr:Milk</td>
</tr>
<tr>
<td>dbr:Camping</td>
<td>dbr:Tent</td>
</tr>
<tr>
<td>dbr:Bed</td>
<td>dbr:Sleep</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Graph Pattern Learner

Pattern Learner

Fusion Training

Training Phase

SPARQL Endpoint

New Data

<table>
<thead>
<tr>
<th>?source</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbr:Fish</td>
</tr>
</tbody>
</table>

Trained Model

Graph Patterns

Fusion Model

Application Phase

Prediction

<table>
<thead>
<tr>
<th>?target</th>
<th>?score</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbr:Fishing</td>
<td>4.2</td>
</tr>
<tr>
<td>dbr:Animal</td>
<td>2.1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

https://w3id.org/associations
Outline

- Background
- My Research
- Evaluation
- Demo
Outline

• Background
• My Research
• Evaluation
• Demo
Evaluation

- How good are the predictions?
- Training/Test set split
- Given a stimulus from the test set, what’s the rank of the true response in the prediction results?
Eval: Results
Eval: Results

our methods
Eval: Results

our methods
Eval: Results

our methods

neighbors

https://w3id.org/associations
Eval: Results

our methods

neighbors: bidi

https://w3id.org/associations
Eval: Results

our methods

Wiki Document Similarity
Eval: Results

our methods

Milne-Witten Similarity
Eval: Results

our methods

RDF2Vec

https://w3id.org/associations
Eval: Results

our methods

Word2Vec
## Evaluation Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Rec@1</th>
<th>Rec@2</th>
<th>Rec@3</th>
<th>Rec@5</th>
<th>Rec@10</th>
<th>MAP</th>
<th>NDCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>DocSim</td>
<td>4.2%</td>
<td>5.6%</td>
<td>5.6%</td>
<td>6.9%</td>
<td>12.5%</td>
<td>6.6%</td>
<td>12.5%</td>
</tr>
<tr>
<td>Word2Vec</td>
<td>4.2%</td>
<td>5.6%</td>
<td>8.3%</td>
<td>12.5%</td>
<td>15.3%</td>
<td>7.8%</td>
<td>12.5%</td>
</tr>
<tr>
<td>RDF2Vec</td>
<td>5.6%</td>
<td>12.5%</td>
<td>13.9%</td>
<td>15.3%</td>
<td>16.7%</td>
<td>10.3%</td>
<td>14.4%</td>
</tr>
<tr>
<td>MW</td>
<td>6.9%</td>
<td>9.7%</td>
<td>11.1%</td>
<td>13.9%</td>
<td>18.1%</td>
<td>11.0%</td>
<td>17.9%</td>
</tr>
<tr>
<td>NB Bidi WL PR</td>
<td>13.9%</td>
<td>20.8%</td>
<td>22.2%</td>
<td>29.2%</td>
<td>31.9%</td>
<td>20.2%</td>
<td>23.8%</td>
</tr>
<tr>
<td>NB Bidi WL InDeg</td>
<td>15.3%</td>
<td>20.8%</td>
<td>26.4%</td>
<td>31.9%</td>
<td>33.3%</td>
<td>21.4%</td>
<td>24.8%</td>
</tr>
<tr>
<td>gpl + precisions</td>
<td>25.0%</td>
<td>36.1%</td>
<td>44.4%</td>
<td>52.8%</td>
<td>62.5%</td>
<td>37.1%</td>
<td>46.0%</td>
</tr>
<tr>
<td>gpl + neural net</td>
<td>30.6%</td>
<td>36.1%</td>
<td>48.6%</td>
<td>51.4%</td>
<td>62.5%</td>
<td>40.3%</td>
<td>48.3%</td>
</tr>
</tbody>
</table>
### Evaluation Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Rec@1</th>
<th>Rec@2</th>
<th>Rec@3</th>
<th>Rec@5</th>
<th>Rec@10</th>
<th>MAP</th>
<th>NDCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>DocSim</td>
<td>4.2%</td>
<td>5.6%</td>
<td>5.6%</td>
<td>6.9%</td>
<td>12.5%</td>
<td>6.6%</td>
<td>12.5%</td>
</tr>
<tr>
<td>Word2Vec</td>
<td>4.2%</td>
<td>5.6%</td>
<td>8.3%</td>
<td>12.5%</td>
<td>15.3%</td>
<td>7.8%</td>
<td>12.5%</td>
</tr>
<tr>
<td>RDF2Vec</td>
<td>5.6%</td>
<td>12.5%</td>
<td>13.9%</td>
<td>15.3%</td>
<td>16.7%</td>
<td>10.3%</td>
<td>14.4%</td>
</tr>
<tr>
<td>MW</td>
<td>6.9%</td>
<td>9.7%</td>
<td>11.1%</td>
<td>13.9%</td>
<td>18.1%</td>
<td>11.0%</td>
<td>17.9%</td>
</tr>
<tr>
<td>NB Bidi WL PR</td>
<td>13.9%</td>
<td>20.8%</td>
<td>22.2%</td>
<td>29.2%</td>
<td>31.9%</td>
<td>20.2%</td>
<td>23.8%</td>
</tr>
<tr>
<td>NB Bidi WL InDeg</td>
<td>15.3%</td>
<td>20.8%</td>
<td>26.4%</td>
<td>31.9%</td>
<td>33.3%</td>
<td>21.4%</td>
<td>24.8%</td>
</tr>
<tr>
<td>gpl + precisions</td>
<td>25.0%</td>
<td>36.1%</td>
<td>44.4%</td>
<td>52.8%</td>
<td>62.5%</td>
<td>37.1%</td>
<td>46.0%</td>
</tr>
<tr>
<td>gpl + neural net</td>
<td>30.6%</td>
<td>36.1%</td>
<td>48.6%</td>
<td>51.4%</td>
<td>62.5%</td>
<td>40.3%</td>
<td>48.3%</td>
</tr>
</tbody>
</table>

- Avg. Inter-Human Agreement: ~32%
Outline

• Background
• My Research
• Evaluation
• Demo
Outline

- Background
- My Research
- Evaluation
- Demo
Machine Learning Outline

**Training Data**
- DBpedia: Cow, Milk, Camping, Tent, Bed, Sleep, ...

**Graph Pattern Learner**
- Pattern Learner
- Fusion Training

**SPARQL Endpoint**

**New Data**
- DBpedia: Fish

**Trained Model**
- Graph Patterns
- Fusion Model

**Prediction**
- DBpedia: Fishing, Score 4.2
- DBpedia: Animal, Score 2.1
- ...

---

https://w3id.org/associations
Machine Learning Outline

Demo

SPARQL Endpoint

New Data
- ?source
- dbr:Fish

Application Phase

Trained Model
- Graph Patterns
- Fusion Model

Prediction
<table>
<thead>
<tr>
<th>?target</th>
<th>?score</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbr:Fishing</td>
<td>4.2</td>
</tr>
<tr>
<td>dbr:Animal</td>
<td>2.1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
HUMAN ASSOCIATION PREDICTION DEMO

This page demonstrates how human associations can be simulated with Linked Data.

For this demo, we used the Graph Pattern Learner to train a machine learning model on a training dataset of human associations (e.g., Dog – Cat).

Click continue to try the trained model out yourself by entering a source node and have it predict target nodes that humans are likely to associate. As a fallback you can also watch a short video of the demo (YouTube).

https://w3id.org/associations#demos
This page demonstrates how human associations can be simulated with Linked Data.

For this demo, we used the Graph Pattern Learner to train a machine learning model on a training dataset of human associations (e.g., Dog – Cat).

Click continue to try the trained model out yourself by entering a source node and have it predict target nodes that humans are likely to associate. As a fallback you can also watch a short video of the demo (YouTube).

https://w3id.org/associations#demos
Other Applications

• TasteDive (Recommendation Engine) Books
  • $\sim 50 \%$ Recall@$10$

• DBpediaNYD
  • $\sim 63 \%$ Recall@$10$
Summary

• **Goal**
  Learning Graph Patterns for Associations

• **Semantic Association Dataset**

• **Graph Pattern Learner**
  Learns SPARQL Patterns for Source-Target-Pairs
  Demo

https://w3id.org/associations
Future Work

• Apply Evolutionary Algorithm
  • to other datasets
  • to other problems

• Extensions:
  • Literals
  • LOD-a-lot & #LD
Discussion

Thanks for your attention

Questions?

https://w3id.org/associations