Graph-based embeddings in mOWL

Fernando Zhapa-Camacho

<2023-02-13 Mon>

Learning objectives

- Learn about different graph representation of ontologies
- Learn how to use graphs from ontologies in machine learning

Introduction

Ontologies axioms are divided into: ABox
assertions of the world

- `Father(John)`
- `hasChild(John, Mary)`

Introduction

Ontologies axioms are divided into: RBox

- relationships between roles:
- `is_part_of` inverseOf `has_part`
- `negatively_regulates subpPropertyOf regulates`

Introduction

Ontologies axioms are divided into: TBox

- Concept descriptions
- `Mother subClassOf Person`

Graphs from ontologies

- ABox can be easily transformed into a Knowledge Graph
  - `Father(John)`: (John, instanceof, Father)
  - `hasChild(John, Mary)`: (John, hasChild, Mary)

Graphs from ontologies

- ABox can be easily transformed into a Knowledge Graph
    - `Father(John)`: (John, instanceof, Father)
    - `hasChild(John, Mary)`: (John, hasChild, Mary)
- What about the TBox?

Graphs from ontologies

- ABox can be easily transformed into a Knowledge Graph
    - `Father(John)`: (John, instanceof, Father)
    - `hasChild(John, Mary)`: (John, hasChild, Mary)
- What about the TBox?
- Multiple methods to *project* an ontology into a graph

Graphs from ontologies

- ABox can be easily transformed into a Knowledge Graph
  - `Father(John)`: (John, instanceof, Father)
  - `hasChild(John, Mary)`: (John, hasChild, Mary)
- What about the TBox?
- Multiple methods to *project* an ontology into a graph
- Some methods could undergo loss of information

Methods implemented in mOWL

Taxonomy

- Only parses axioms of the form $C \sqsubseteq D$
- C, D are atomic concepts
- Graphs represent the hierarchy of concepts

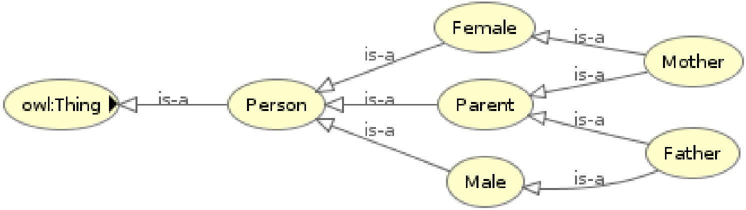

Figure: Family ontology representation

Methods implemented in mOWL

DL2Vec

Condition 1	Condition 2	Triple(s)
$A \sqsubseteq QR_0 \ldots QR_m D$ $A \equiv QR_0 \ldots QR_m D$	$D := B_1 \sqcup \ldots \sqcup B_n \vert B_1 \sqcap \ldots \sqcap B_n$	$\langle A, (R_0 \ldots R_m), B_i \rangle$ for $i \in 1 \ldots n$
$A \sqsubseteq B$		$\langle A, SubClassOf, B \rangle$
$A \equiv B$		$\langle A, EquivalentTo, B \rangle$

Figure: DL2Vec projection rules

Methods implemented in mOWL

OWL2Vec*

Axiom of condition 1	Axiom or triple(s) of condition 2	Projected triple(s)
$A \sqsubseteq \square r.\ D$	$D \equiv B \mid B_1 \sqcup \ldots \sqcup B_n \mid B_1 \sqcap \ldots \sqcap B_n$	$\langle A, r, B\rangle$ or
or		
$\square r.\ D \sqsubseteq A$		
$\exists r. \top \sqsubseteq A$ (domain)	$\top \sqsubseteq \forall r.\ B$ (range)	$\langle A, r, B_i\rangle$ for $i \in 1, \ldots, n$
$A \sqsubseteq \exists r.\ \{b\}$	$B(b)$	
$r \sqsubseteq r'$	$\langle A, r', B\rangle$ has been projected	
$r' \equiv r^-$	$\langle B, r', A\rangle$ has been projected	
$s_1 \circ \ldots \circ s_n \sqsubseteq r$	$\langle A, s_1, C_1\rangle \ldots \langle C_n, s_n, B\rangle$ have been projected	
$B \sqsubseteq A$	–	$\langle B, rdfs{:}subClassOf, A\rangle$
		$\langle A, rdfs{:}subClassOf^-, B\rangle$
$A(a)$	–	$\langle a, rdf{:}type, A\rangle$
		$\langle A, rdf{:}type^-, a\rangle$
$r(a, b)$	–	$\langle a, r, b\rangle$

$\square$ is one of: $\geq, \leq, =, \exists, \forall$. A, B, B_i and C_i are atomic concepts (classes), s_i, r and r' are roles (object properties), r^- is the inverse of a relation r, a and b are individuals (instances), $\top$ is the top concept (defined by *owl:Thing*)

Figure: OWL2Vec* projection rules

Limitations of projection methods

- Some are not complete (Taxonomical projection):
  - $(C \sqcap D) \sqsubseteq E \rightarrow \textit{undefined}$

Limitations of projection methods

- Some are not complete (Taxonomical projection):
  - $(C \sqcap D) \sqsubseteq E \rightarrow$ *undefined*
- Some are not invertible (OWL2Vec*):
  - $(C \sqsubseteq \exists R.D) \rightarrow (C, R, D)$
  - $(C \sqsubseteq \forall R.D) \rightarrow (C, R, D)$
  - Inverse of $(C, R, D) \rightarrow$??

We have graphs. Next step?

- Ontology has been transformed into a graph
- Graphs as input for a machine learning model
- mOWL supports two ways to *embed* a graph:
  - Random-walk based embeddings
  - Knowledge Graph Embedding (KGE) models

Embedding with random walks

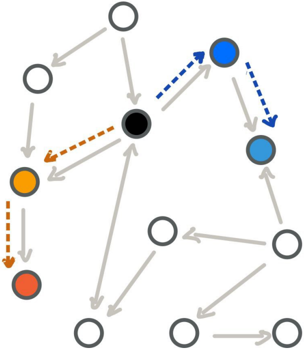

Figure: Random walks

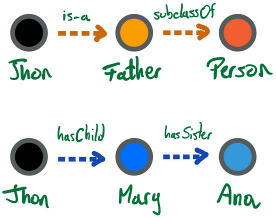

Figure: Sequences generated

Embedding with random walks

- Generated sentences become input in a language processing model:
    - Word2Vec
    - Transformers

Embedding with random walks

- Word2Vec embeddings preserve co-occurrence
- Generated embeddings can be used to compute similarity between entities E_1 and E_2
- Similarity can be computed using cosine similarity.

$$sim(v(E_1) \cdot v(E_2)) = 1 \rightarrow similar$$

$$sim(v(E_1) \cdot v(E_2)) = 0 \rightarrow dissimilar$$

Embedding with random walks

- Word2Vec embeddings preserve co-occurrence
- Generated embeddings can be used to compute similarity between entities E_1 and E_2
- Similarity can be computed using cosine similarity.

$$sim(v(E_1) \cdot v(E_2)) = 1 \rightarrow similar$$

$$sim(v(E_1) \cdot v(E_2)) = 0 \rightarrow dissimilar$$

- Protein-protein interactions
- Gene-disease associations

KGE models

- Graphs are composed by triples *head*, *relation*, *tail*
- TransE is an example of a translational KGE model.
- Translational models consider *relation* to be a translation operation between *head* and tail

$$||h + r - t|| \approx 0$$

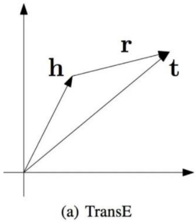

Figure: TransE

KGE models

- The *score* of a triple is given by $d(h, r, t) = ||h + r - t||$, where the lower d is, the more plausible the triple to hold true.
- TransE is the most representative model.

$$\mathcal{L} = \left[\gamma + d(\boldsymbol{h} + \boldsymbol{r}, \boldsymbol{t}) - d\left(\boldsymbol{h} + \boldsymbol{r}, \boldsymbol{t}'\right)\right]_+$$

- $d(\boldsymbol{h} + \boldsymbol{r}, \boldsymbol{t})$ is the score of a *positive* triple
- $d\left(\boldsymbol{h} + \boldsymbol{r}, \boldsymbol{t}'\right)$ is the score of a *negative* triple
- γ is a margin parameter

KGE models

- Other models have appeared by changing

$$h + r \approx t$$

into

$$f(h) + f(r) \approx f(t)$$

- Some variations of TransE are:
  - TransH
  - TransR
  - TransD

KGE models

Method	Ent. embedding	Rel. embedding	Scoring function $f_r(h,t)$
TransE [14]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$-\|\mathbf{h}+\mathbf{r}-\mathbf{t}\|_{1/2}$
TransH [15]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r},\mathbf{w}_r\in\mathbb{R}^d$	$-\|(\mathbf{h}-\mathbf{w}_r^\top\mathbf{h}\mathbf{w}_r)+\mathbf{r}-(\mathbf{t}-\mathbf{w}_r^\top\mathbf{t}\mathbf{w}_r)\|_2^2$
TransR [16]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^k,\mathbf{M}_r\in\mathbb{R}^{k\times d}$	$-\|\mathbf{M}_r\mathbf{h}+\mathbf{r}-\mathbf{M}_r\mathbf{t}\|_2^2$
TransD [50]	$\mathbf{h},\mathbf{w}_h\in\mathbb{R}^d$ $\mathbf{t},\mathbf{w}_t\in\mathbb{R}^d$	$\mathbf{r},\mathbf{w}_r\in\mathbb{R}^k$	$-\|(\mathbf{w}_r\mathbf{w}_h^\top+\mathbf{I})\mathbf{h}+\mathbf{r}-(\mathbf{w}_r\mathbf{w}_t^\top+\mathbf{I})\mathbf{t}\|_2^2$
TranSparse [51]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^k,\mathbf{M}_r(\theta_r)\in\mathbb{R}^{k\times d}$ $\mathbf{M}_r^1(\theta_r^1),\mathbf{M}_r^2(\theta_r^2)\in\mathbb{R}^{k\times d}$	$-\|\mathbf{M}_r(\theta_r)\mathbf{h}+\mathbf{r}-\mathbf{M}_r(\theta_r)\mathbf{t}\|_{1/2}^2$ $-\|\mathbf{M}_r^1(\theta_r^1)\mathbf{h}+\mathbf{r}-\mathbf{M}_r^2(\theta_r^2)\mathbf{t}\|_{1/2}^2$
TransM [52]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$-\theta_r\|\mathbf{h}+\mathbf{r}-\mathbf{t}\|_{1/2}$
ManifoldE [53]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$-(\|\mathbf{h}+\mathbf{r}-\mathbf{t}\|_2^2-\theta_r^2)^2$
TransF [54]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$(\mathbf{h}+\mathbf{r})^\top\mathbf{t}+(\mathbf{t}-\mathbf{r})^\top\mathbf{h}$
TransA [55]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d,\mathbf{M}_r\in\mathbb{R}^{d\times d}$	$-(\|\mathbf{h}+\mathbf{r}-\mathbf{t}\|)^\top\mathbf{M}_r(\|\mathbf{h}+\mathbf{r}-\mathbf{t}\|)$

Figure: KGE methods

KGE models

Method	Ent. embedding	Rel. embedding	Scoring function $f_r(h,t)$
RESCAL [13]	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{M}_r \in \mathbb{R}^{d \times d}$	$\mathbf{h}^\top \mathbf{M}_r \mathbf{t}$
TATEC [64]	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d, \mathbf{M}_r \in \mathbb{R}^{d \times d}$	$\mathbf{h}^\top \mathbf{M}_r \mathbf{t} + \mathbf{h}^\top \mathbf{r} + \mathbf{t}^\top \mathbf{r} + \mathbf{h}^\top \mathbf{D} \mathbf{t}$
DistMult [65]	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$\mathbf{h}^\top \mathrm{diag}(\mathbf{r}) \mathbf{t}$
HolE [62]	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$\mathbf{r}^\top (\mathbf{h} \star \mathbf{t})$
ComplEx [66]	$\mathbf{h}, \mathbf{t} \in \mathbb{C}^d$	$\mathbf{r} \in \mathbb{C}^d$	$\mathrm{Re}(\mathbf{h}^\top \mathrm{diag}(\mathbf{r}) \bar{\mathbf{t}})$
ANALOGY [68]	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{M}_r \in \mathbb{R}^{d \times d}$	$\mathbf{h}^\top \mathbf{M}_r \mathbf{t}$
SME [18]	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$(\mathbf{M}_u^1 \mathbf{h} + \mathbf{M}_u^2 \mathbf{r} + \mathbf{b}_u)^\top (\mathbf{M}_v^1 \mathbf{t} + \mathbf{M}_v^2 \mathbf{r} + \mathbf{b}_v)$ $((\mathbf{M}_u^1 \mathbf{h}) \circ (\mathbf{M}_u^2 \mathbf{r}) + \mathbf{b}_u)^\top ((\mathbf{M}_v^1 \mathbf{t}) \circ (\mathbf{M}_v^2 \mathbf{r}) + \mathbf{b}_v)$
NTN [19]	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r}, \mathbf{b}_r \in \mathbb{R}^k, \underline{\mathbf{M}}_r \in \mathbb{R}^{d \times d \times k}$ $\mathbf{M}_r^1, \mathbf{M}_r^2 \in \mathbb{R}^{k \times d}$	$\mathbf{r}^\top \tanh(\mathbf{h}^\top \underline{\mathbf{M}}_r \mathbf{t} + \mathbf{M}_r^1 \mathbf{h} + \mathbf{M}_r^2 \mathbf{t} + \mathbf{b}_r)$
SLM [19]	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^k, \mathbf{M}_r^1, \mathbf{M}_r^2 \in \mathbb{R}^{k \times d}$	$\mathbf{r}^\top \tanh(\mathbf{M}_r^1 \mathbf{h} + \mathbf{M}_r^2 \mathbf{t})$
MLP [69]	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$\mathbf{w}^\top \tanh(\mathbf{M}^1 \mathbf{h} + \mathbf{M}^2 \mathbf{r} + \mathbf{M}^3 \mathbf{t})$

Figure: KGE methods

KGE Embeddings:

We can use KGE models to predict new triples. For example:

1. Axiom to query: $a = C \sqsubseteq \exists R.D$
2. Graph representation: (C, R, D)
3. TransE score: $s = ||h + r - t||$
4. Score s gives us the plausability of the axiom a to hold true.