Part 2: Machine learning with biomedical ontologies

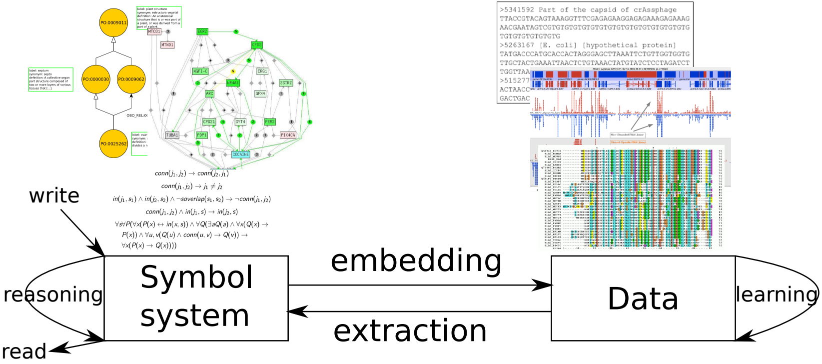
Robert Hoehndorf

Challenges with semantic similarity

- not data-driven but hand-crafted
- not task-specific
- usually outputs a single value
- hard to chose a similarity measure
- usually graph-based and losing some information

Next: machine learning methods for and with ontologies

Machine learning with ontologies: Overview



- phenotype
- function
- disease

- genotype
- protein sequence
- expression

Embedding formal knowledge

Embedding

An embedding is a map (morphism) from one mathematical structure X into another structure Y:
$f : X \hookrightarrow Y$
such that X is preserved in Y.

- Y may be more suitable than X for some operations/algorithms.
  - similarity
  - gradients, optimization

Embedding formal knowledge

Embedding

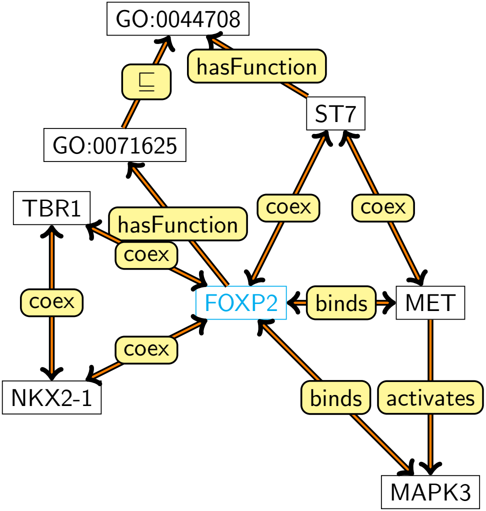
An embedding is a map (morphism) from one mathematical structure X into another structure Y:
$f : X \hookrightarrow Y$
such that X is preserved in Y.

- Y may be more suitable than X for some operations/algorithms.
  - similarity
  - gradients, optimization

We want to embed *ontologies* in $\Re^n$. Approaches:

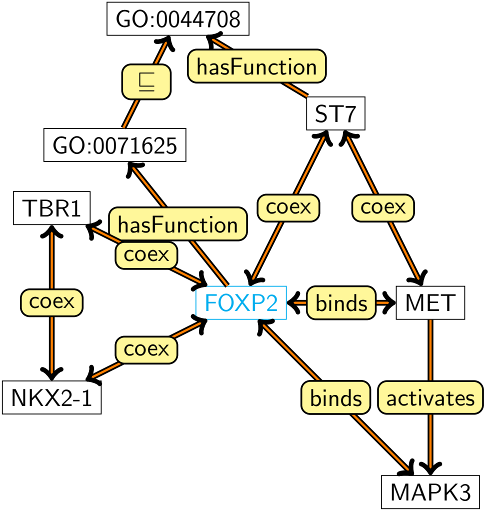
- graph-based
- syntactic
- model-theoretic

Random walks



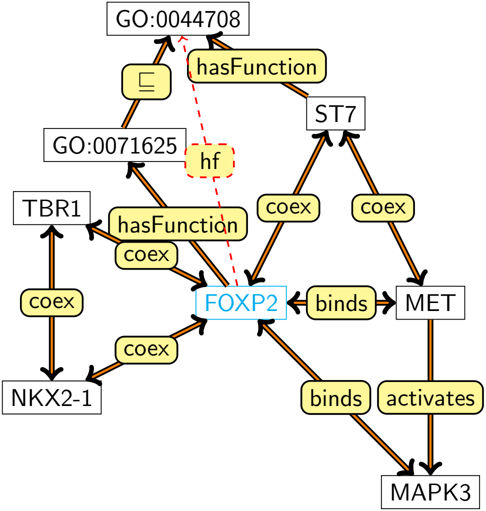
- FOXP2 is characterized by *adjacent* and close nodes and edges
- different edges may "transmit" information differently

Random walks



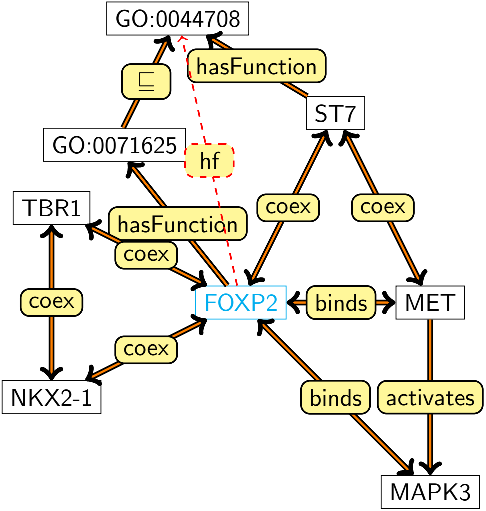
- precompute the deductive closure:
- for all ϕ: if $\mathcal{KG} \models \phi$, add ϕ to $\mathcal{KG}$

Random walks



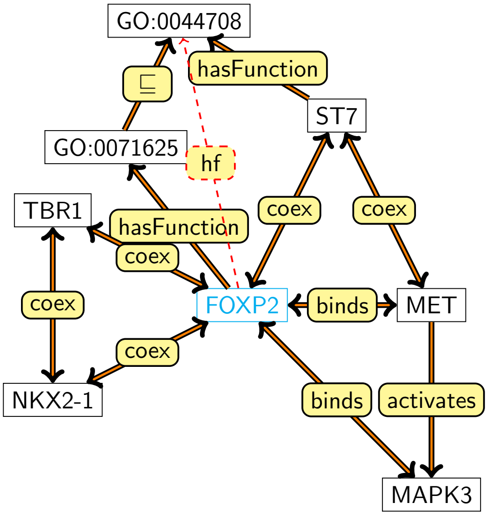
- precompute the deductive closure:
- for all ϕ: if $\mathcal{KG} \models \phi$, add ϕ to $\mathcal{KG}$

Random walks



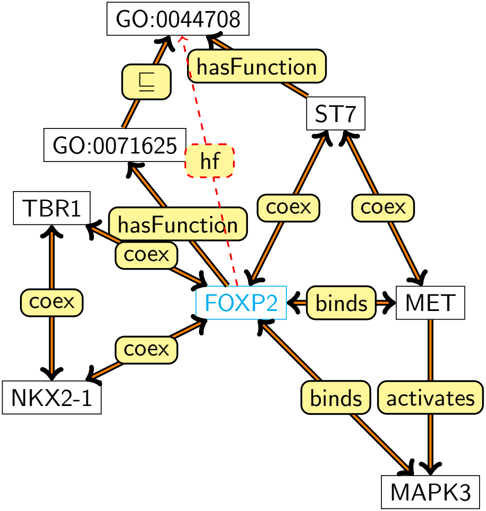
- Exploring the graph:

Random walks



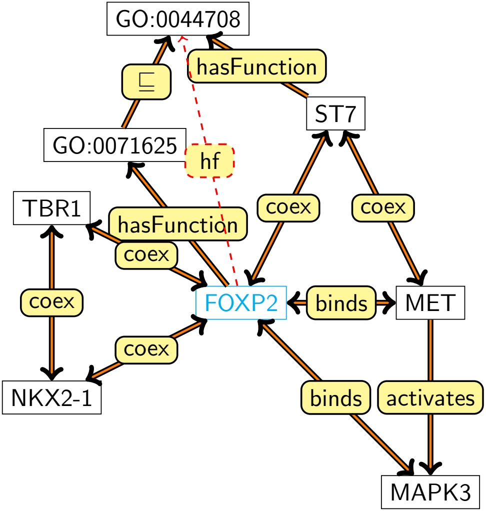
- Exploring the graph:
- :FOXP2 :binds :MET :coex :ST7 :hasFunction GO:0044708

Random walks



- Exploring the graph:
- :FOXP2 :binds :MET :coex :ST7 :hasFunction GO:0044708
- :FOXP2 :hasFunction GO:0071625 subClassOf GO:0044708

Random walks



- Exploring the graph:
- :FOXP2 :binds :MET :coex :ST7 :hasFunction GO:0044708
- :FOXP2 :hasFunction GO:0071625 subClassOf GO:0044708
- :FOXP2 :coex :TBR1 :coex :NKX2-1 :coex :TBR1 :coex ...

Word2Vec

Maximize:

$$\frac{1}{N}\sum_{n=1}^{N}\sum_{-c\leq j\leq c, j\neq 0}\log p(w_{n+j}|w_n) \tag{1}$$

with

$$p(w_o|w_i) = \frac{\exp({v'_{w_o}}^T v_{w_i})}{\sum_{w=1}^{W}\exp({v'_w}^T v_{w_i})} \tag{2}$$

(at least conceptually; different strategies are used to approximate Eqn. 2)

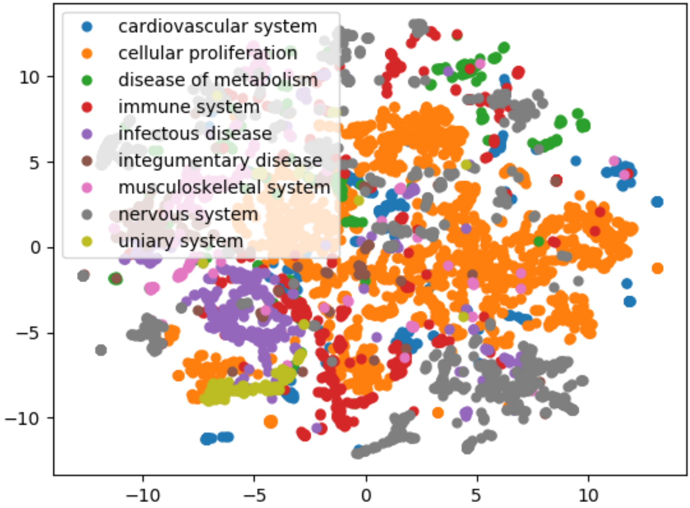
Word2Vec and Random Walks

- random walks "flatten" a graph
  - walks capture node neighborhood
  - and generate a "corpus"
- random walks capture graph "structure"
  - hub-nodes, communities, etc.
  - determine "importance" of nodes
- embeddings capture co-occurrence
  - similar graph neighborhood $\Rightarrow$ similar co-occurrence $\Rightarrow$ similar vector
- embeddings generate "feature" vectors
  - functions from symbols (words, labels) into $\Re^n$

What to do with embeddings?

- useful for edge prediction, similarity, clustering, as feature vectors
  - supervised: edge prediction (e.g., SVM, ANN)
    - e.g.: find a function $f : \Re^n \times \Re^n \mapsto [0, 1]$ s.t. $\sqrt{\frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}{T}}$ (RMSE) is minimized for a set of true labels y_k
  - unsupervised: clustering, similarity, visualization
    - cosine similarity (for L2-normalized features)
    - Word2Vec embeddings capture similarity between co-occurrence vectors

Visualizing embeddings



Supervised learning

- feature vectors represent graph neighborhood of nodes
  - adjacent nodes and edges
  - ontology classes (asserted & inferred)
- useful in supervised prediction tasks
- relation prediction:
  - input: two features vectors (from embedding function)
  - output: 0 or 1 (relation or not)
  - training data: positive and negative cases
    - $R(x, y)$ and $\neg R(x, y)$

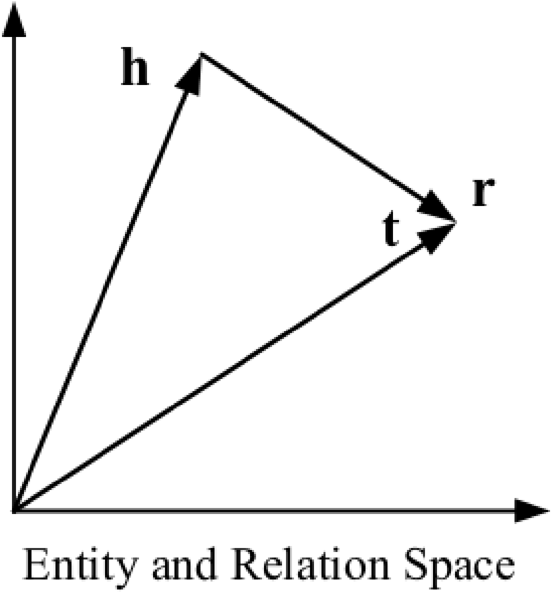
Features: supervised learning

Object property	Source type	Target type	Without reasoning		With reasoning	
			F-measure	AUC	F-measure	AUC
has target	Drug	Gene/Protein	0.94	0.97	0.94	0.98
has disease annotation	Gene/Protein	Disease	0.89	0.95	0.89	0.95
has side-effect*	Drug	Phenotype	0.86	0.93	0.87	0.94
has interaction	Gene/Protein	Gene/Protein	0.82	0.88	0.82	0.88
has function*	Gene/Protein	Function	0.85	0.95	0.83	0.91
has gene phenotype*	Gene/Protein	Phenotype	0.84	0.91	0.82	0.90
has indication	Drug	Disease	0.72	0.79	0.76	0.83
has disease phenotype*	Disease	Phenotype	0.72	0.78	0.70	0.77

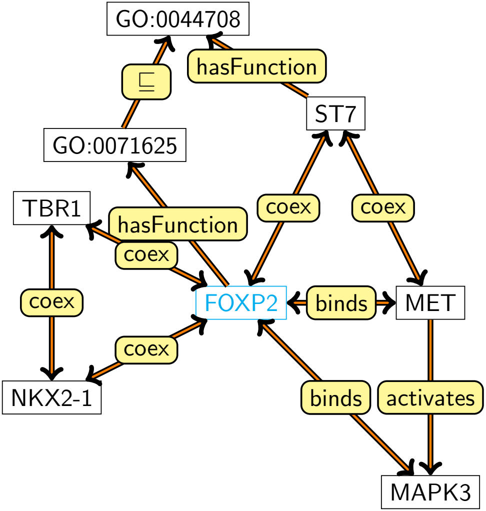
Some limitations

- “word”-based (Word2Vec):
  - semantics is reduced to co-occurrence in the random “walks”
  - “disjointWith” vs. “part-of” vs. “subClassOf”

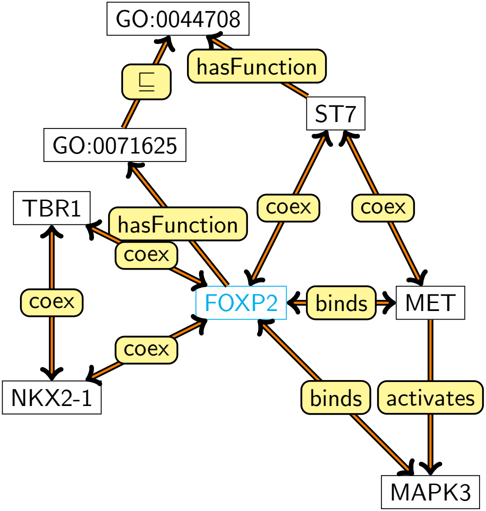
Translating embeddings



Translating embeddings

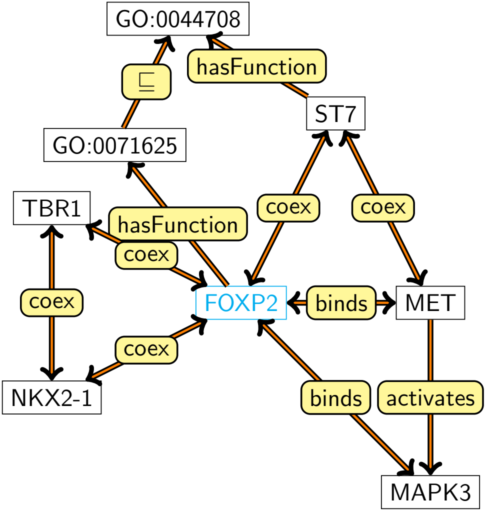


Translating embeddings



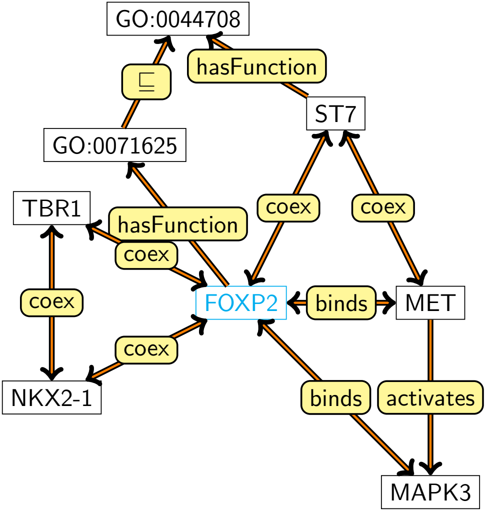
- FOXP2 + binds = MET

Translating embeddings



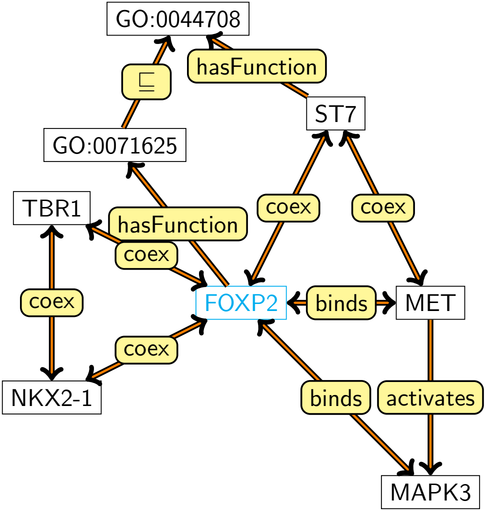
- FOXP2 + binds = MET
- MAP + activates = MAPK3

Translating embeddings



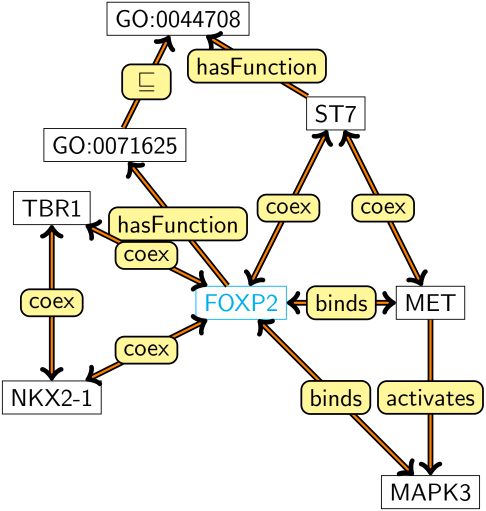
- FOXP2 + binds = MET
- MAP + activates = MAPK3
- MET + binds = FOXP2

Translating embeddings



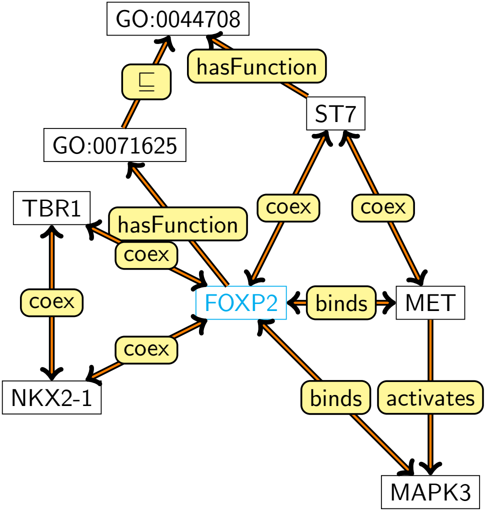
- FOXP2 + binds = MET
- MAP + activates = MAPK3
- MET + binds = FOXP2
- ST7 + hasFunction = `GO:0044708`

Translating embeddings



- FOXP2 + binds = MET
- MAP + activates = MAPK3
- MET + binds = FOXP2
- ST7 + hasFunction = `GO:0044708`
- ...

Translating embeddings



- FOXP2 + binds - MET = 0
- MAP + activates - MAPK3 = 0
- MET + binds - FOXP2 = 0
- ST7 + hasFunction - `GO:0044708` = 0
- ...

Translating embeddings

Algorithm 1 Learning TransE

input Training set $S = \{(h, \ell, t)\}$, entities and rel. sets E and L, margin γ, embeddings dim. k.

initialize $\boldsymbol{\ell} \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each $\ell \in L$

$\boldsymbol{\ell} \leftarrow \boldsymbol{\ell} / \|\boldsymbol{\ell}\|$ for each $\ell \in L$

$\mathbf{e} \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each entity $e \in E$

loop

$\mathbf{e} \leftarrow \mathbf{e} / \|\mathbf{e}\|$ for each entity $e \in E$

$S_{batch} \leftarrow \text{sample}(S, b)$ // sample a minibatch of size b

$T_{batch} \leftarrow \emptyset$ // initialize the set of pairs of triplets

for $(h, \ell, t) \in S_{batch}$ **do**

$(h', \ell, t') \leftarrow \text{sample}(S'_{(h,\ell,t)})$ // sample a corrupted triplet

$T_{batch} \leftarrow T_{batch} \cup \{((h, \ell, t), (h', \ell, t'))\}$

end for

Update embeddings w.r.t. $\displaystyle\sum_{((h,\ell,t),(h',\ell,t')) \in T_{batch}} \nabla \left[\gamma + d(\boldsymbol{h} + \boldsymbol{\ell}, \boldsymbol{t}) - d(\boldsymbol{h'} + \boldsymbol{\ell}, \boldsymbol{t'})\right]_{+}$

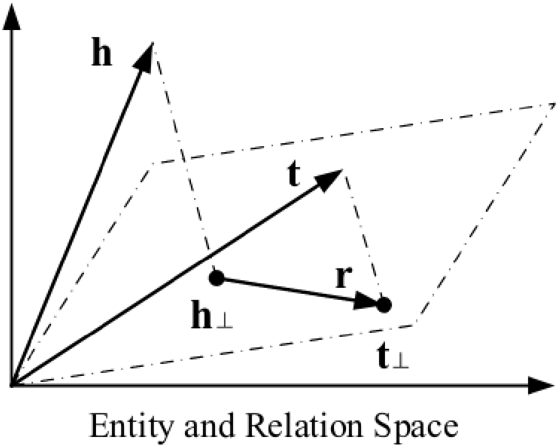
end loop

Bordes et al. (2013). Translating Embeddings for ModelingMulti-relational Data.

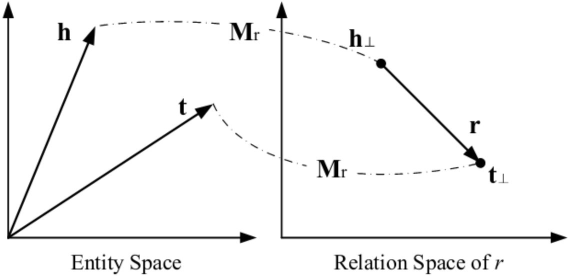
Some properties of TransE

- graph-based
  - works well on RDF graphs
  - and ontology graphs
- 1:1 relations only
  - not suitable for hierarchies (1-N relations)
  - not suitable for N-N relations
  - no transitive, symmetric, reflexive relations

Translating embeddings



Translating embeddings



(c) TransR.

Translating embeddings

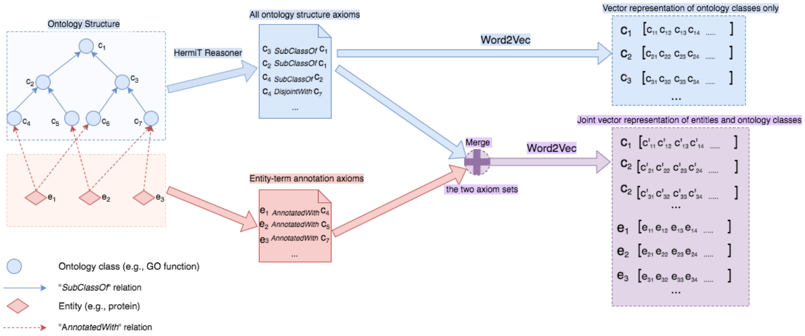
Method	Ent. embedding	Rel. embedding	Scoring function $f_r(h,t)$	Constraints/Regularization
TransE [14]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$-\Vert\mathbf{h}+\mathbf{r}-\mathbf{t}\Vert_{1/2}$	$\Vert\mathbf{h}\Vert_2=1, \Vert\mathbf{t}\Vert_2=1$
TransH [15]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r},\mathbf{w}_r\in\mathbb{R}^d$	$-\Vert(\mathbf{h}-\mathbf{w}_r^\top\mathbf{h}\mathbf{w}_r)+\mathbf{r}-(\mathbf{t}-\mathbf{w}_r^\top\mathbf{t}\mathbf{w}_r)\Vert_2^2$	$\Vert\mathbf{h}\Vert_2\le 1, \Vert\mathbf{t}\Vert_2\le 1$ $\vert\mathbf{w}_r^\top\mathbf{r}\vert/\Vert\mathbf{r}\Vert_2\le\epsilon, \Vert\mathbf{w}_r\Vert_2=1$
TransR [16]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^k,\mathbf{M}_r\in\mathbb{R}^{k\times d}$	$-\Vert\mathbf{M}_r\mathbf{h}+\mathbf{r}-\mathbf{M}_r\mathbf{t}\Vert_2^2$	$\Vert\mathbf{h}\Vert_2\le 1, \Vert\mathbf{t}\Vert_2\le 1, \Vert\mathbf{r}\Vert_2\le 1$ $\Vert\mathbf{M}_r\mathbf{h}\Vert_2\le 1, \Vert\mathbf{M}_r\mathbf{t}\Vert_2\le 1$
TransD [50]	$\mathbf{h},\mathbf{w}_h\in\mathbb{R}^d$ $\mathbf{t},\mathbf{w}_t\in\mathbb{R}^d$	$\mathbf{r},\mathbf{w}_r\in\mathbb{R}^k$	$-\Vert(\mathbf{w}_r\mathbf{w}_h^\top+\mathbf{I})\mathbf{h}+\mathbf{r}-(\mathbf{w}_r\mathbf{w}_t^\top+\mathbf{I})\mathbf{t}\Vert_2^2$	$\Vert\mathbf{h}\Vert_2\le 1, \Vert\mathbf{t}\Vert_2\le 1, \Vert\mathbf{r}\Vert_2\le 1$ $\Vert(\mathbf{w}_r\mathbf{w}_h^\top+\mathbf{I})\mathbf{h}\Vert_2\le 1$ $\Vert(\mathbf{w}_r\mathbf{w}_t^\top+\mathbf{I})\mathbf{t}\Vert_2\le 1$
TranSparse [51]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^k,\mathbf{M}_r(\theta_r)\in\mathbb{R}^{k\times d}$ $\mathbf{M}_r^1(\theta_r^1),\mathbf{M}_r^2(\theta_r^2)\in\mathbb{R}^{k\times d}$	$-\Vert\mathbf{M}_r(\theta_r)\mathbf{h}+\mathbf{r}-\mathbf{M}_r(\theta_r)\mathbf{t}\Vert_{1/2}^2$ $-\Vert\mathbf{M}_r^1(\theta_r^1)\mathbf{h}+\mathbf{r}-\mathbf{M}_r^2(\theta_r^2)\mathbf{t}\Vert_{1/2}^2$	$\Vert\mathbf{h}\Vert_2\le 1, \Vert\mathbf{t}\Vert_2\le 1, \Vert\mathbf{r}\Vert_2\le 1$ $\Vert\mathbf{M}_r(\theta_r)\mathbf{h}\Vert_2\le 1, \Vert\mathbf{M}_r(\theta_r)\mathbf{t}\Vert_2\le 1$ $\Vert\mathbf{M}_r^1(\theta_r^1)\mathbf{h}\Vert_2\le 1, \Vert\mathbf{M}_r^2(\theta_r^2)\mathbf{t}\Vert_2\le 1$
TransM [52]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$-\theta_r\Vert\mathbf{h}+\mathbf{r}-\mathbf{t}\Vert_{1/2}$	$\Vert\mathbf{h}\Vert_2=1, \Vert\mathbf{t}\Vert_2=1$
ManifoldE [53]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$-(\Vert\mathbf{h}+\mathbf{r}-\mathbf{t}\Vert_2^2-\theta_r^2)^2$	$\Vert\mathbf{h}\Vert_2\le 1, \Vert\mathbf{t}\Vert_2\le 1, \Vert\mathbf{r}\Vert_2\le 1$
TransF [54]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$(\mathbf{h}+\mathbf{r})^\top\mathbf{t}+(\mathbf{t}-\mathbf{r})^\top\mathbf{h}$	$\Vert\mathbf{h}\Vert_2\le 1, \Vert\mathbf{t}\Vert_2\le 1, \Vert\mathbf{r}\Vert_2\le 1$
TransA [55]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d,\mathbf{M}_r\in\mathbb{R}^{d\times d}$	$-(\vert\mathbf{h}+\mathbf{r}-\mathbf{t}\vert)^\top\mathbf{M}_r(\vert\mathbf{h}+\mathbf{r}-\mathbf{t}\vert)$	$\Vert\mathbf{h}\Vert_2\le 1, \Vert\mathbf{t}\Vert_2\le 1, \Vert\mathbf{r}\Vert_2\le 1$ $\Vert\mathbf{M}_r\Vert_F\le 1, [\mathbf{M}_r]_{ij}=[\mathbf{M}_r]_{ji}\ge 0$
KG2E [45]	$\mathbf{h}\sim\mathcal{N}(\boldsymbol{\mu}_h,\boldsymbol{\Sigma}_h)$ $\mathbf{t}\sim\mathcal{N}(\boldsymbol{\mu}_t,\boldsymbol{\Sigma}_t)$ $\boldsymbol{\mu}_h,\boldsymbol{\mu}_t\in\mathbb{R}^d$ $\boldsymbol{\Sigma}_h,\boldsymbol{\Sigma}_t\in\mathbb{R}^{d\times d}$	$\mathbf{r}\sim\mathcal{N}(\boldsymbol{\mu}_r,\boldsymbol{\Sigma}_r)$ $\boldsymbol{\mu}_r\in\mathbb{R}^d,\boldsymbol{\Sigma}_r\in\mathbb{R}^{d\times d}$	$-\mathrm{tr}(\boldsymbol{\Sigma}_r^{-1}(\boldsymbol{\Sigma}_h+\boldsymbol{\Sigma}_t))-\boldsymbol{\mu}^\top\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\mu}-\ln\frac{\det(\boldsymbol{\Sigma}_r)}{\det(\boldsymbol{\Sigma}_h+\boldsymbol{\Sigma}_t)}$ $-\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}-\ln(\det(\boldsymbol{\Sigma}))$ $\boldsymbol{\mu}=\boldsymbol{\mu}_h+\boldsymbol{\mu}_r-\boldsymbol{\mu}_t$ $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_h+\boldsymbol{\Sigma}_r+\boldsymbol{\Sigma}_t$	$\Vert\boldsymbol{\mu}_h\Vert_2\le 1, \Vert\boldsymbol{\mu}_t\Vert_2\le 1, \Vert\boldsymbol{\mu}_r\Vert_2\le 1$ $c_{min}\mathbf{I}\le\boldsymbol{\Sigma}_h\le c_{max}\mathbf{I}$ $c_{min}\mathbf{I}\le\boldsymbol{\Sigma}_t\le c_{max}\mathbf{I}$ $c_{min}\mathbf{I}\le\boldsymbol{\Sigma}_r\le c_{max}\mathbf{I}$
TransG [46]	$\mathbf{h}\sim\mathcal{N}(\boldsymbol{\mu}_h,\sigma_h^2\mathbf{I})$ $\mathbf{t}\sim\mathcal{N}(\boldsymbol{\mu}_t,\sigma_t^2\mathbf{I})$ $\boldsymbol{\mu}_h,\boldsymbol{\mu}_t\in\mathbb{R}^d$	$\boldsymbol{\mu}_r^i\sim\mathcal{N}(\boldsymbol{\mu}_t-\boldsymbol{\mu}_h,(\sigma_h^2+\sigma_t^2)\mathbf{I})$ $\mathbf{r}=\sum_i\pi_r^i\boldsymbol{\mu}_r^i\in\mathbb{R}^d$	$\sum_i\pi_r^i\exp\left(-\frac{\Vert\boldsymbol{\mu}_h+\boldsymbol{\mu}_r^i-\boldsymbol{\mu}_t\Vert_2^2}{\sigma_h^2+\sigma_t^2}\right)$	$\Vert\boldsymbol{\mu}_h\Vert_2\le 1, \Vert\boldsymbol{\mu}_t\Vert_2\le 1, \Vert\boldsymbol{\mu}_r^i\Vert_2\le 1$
UM [56]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	——	$-\Vert\mathbf{h}-\mathbf{t}\Vert_2^2$	$\Vert\mathbf{h}\Vert_2=1, \Vert\mathbf{t}\Vert_2=1$
SE [57]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{M}_r^1,\mathbf{M}_r^2\in\mathbb{R}^{d\times d}$	$-\Vert\mathbf{M}_r^1\mathbf{h}-\mathbf{M}_r^2\mathbf{t}\Vert_1$	$\Vert\mathbf{h}\Vert_2=1, \Vert\mathbf{t}\Vert_2=1$

Wang et al. Knowledge Graph Embedding: A Survey ofApproaches and Applications.

Beyond graphs

- so far, all the methods were based on graphs
  - ontologies are not graphs!
  - converting ontologies to graphs loses information
  - no axioms, no definitions
- maybe we won't need the graph?

Onto2Vec

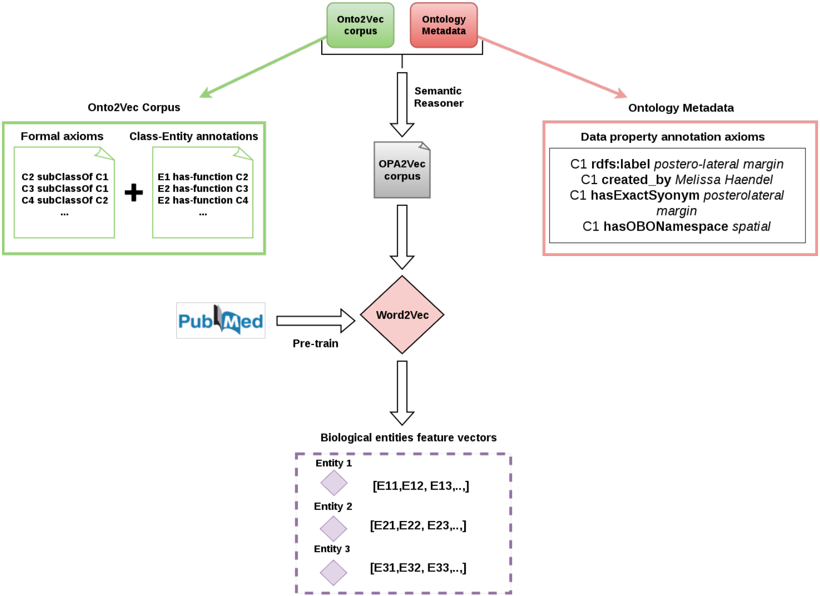


Ontology Structure
HermiT Reasoner
All ontology structure axioms
c3 SubClassOf c1
c2 SubClassOf c1
c4 SubClassOf c2
c4 DisjointWith c7
...
Word2Vec
Vector representation of ontology classes only
Merge
the two axiom sets
Word2Vec
Joint vector representation of entities and ontology classes
Entity-term annotation axioms
e1 AnnotatedWith c4
e2 AnnotatedWith c5
e3 AnnotatedWith c7
...
Ontology class (e.g., GO function)
"SubClassOf" relation
Entity (e.g., protein)
"AnnotatedWith" relation

Combination with text

- ontologies contain more than axioms:
  - labels, synonyms, definitions, authors, etc.
- Description Logic axioms != natural language
- transfer learning: learn on one domain/task, apply to another
  - e.g.: learn on literature, apply to ontologies
  - words have "meaning" in literature, Description Logic symbols have "meaning" in ontology axioms
- Ontologies Plus Annotations 2 Vec (OPA2Vec) combines both

Ontologies Plus Annotations 2 Vec



Axioms contribute to prediction tasks: GO and GO-PLUS

	Human	Yeast	Arabidopsis
GO_Onto2Vec	0.7660	0.7701	0.7559
GO_Onto2Vec_NN	0.8779	0.8711	0.8364
GO_plus_Onto2Vec	0.7880	0.7943	0.7889
GO_plus_Onto2Vec_NN	**0.9021**	**0.8937**	**0.8834**

How to overcome the semantic gap?

- none of the models discussed above are truly “semantic”
  - all syntactic
  - graph-based or based on axioms

How to overcome the semantic gap?

- none of the models discussed above are truly “semantic”
  - all syntactic
  - graph-based or based on axioms
- what do we actually mean by “semantics”?

How to overcome the semantic gap?

- none of the models discussed above are truly “semantic”
  - all syntactic
  - graph-based or based on axioms
- what do we actually mean by “semantics”?
  - formal definition of “truth” relies on “models”

How to overcome the semantic gap?

- none of the models discussed above are truly "semantic"
  - all syntactic
  - graph-based or based on axioms
- what do we actually mean by "semantics"?
  - formal definition of "truth" relies on "models"
  - universal algebra over formal languages (with signature Σ)

Description Logic EL++

Name	Syntax	Semantics
top	$\top$	$\Delta^{\mathcal{I}}$
bottom	$\bot$	$\emptyset$
nominal	$\{a\}$	$\{a^{\mathcal{I}}\}$
conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
existential restriction	$\exists r.C$	$\{x \in \Delta^{\mathcal{I}} \mid \exists y \in \Delta^{\mathcal{I}} : (x, y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$
generalized concept inclusion	$C \sqsubseteq D$	$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
role inclusion	$r_1 \circ ... \circ r_n \sqsubseteq r$	$r_1^{\mathcal{I}} \circ ... \circ r_n^{\mathcal{I}} \subseteq r^{\mathcal{I}}$

Models

- Interpretations and Σ-structures
- Model $\mathfrak{A}$ of a formula ϕ: ϕ is true in $\mathfrak{A}$ ($\mathfrak{A} \models \phi$)
- Theory T: set of formulas
- $\mathfrak{A}$ is a model of T if $\mathfrak{A}$ is a model of all formulas in T
- Ontologies are (special kinds of) theories

EL Embeddings

- given a theory/ontology T with signature $\Sigma(T)$
- aim: find $f_e : \Sigma(T) \mapsto \Re^n$ s.t. $f_e(\Sigma(T))$ is a model of T ($f_e(\Sigma(T)) \models T$)

EL Embeddings

- given a theory/ontology T with signature $\Sigma(T)$
- aim: find $f_e : \Sigma(T) \mapsto \Re^n$ s.t. $f_e(\Sigma(T))$ is a model of T ($f_e(\Sigma(T)) \models T$)
- more general: find an algorithm that maps symbols (signatures) into $\Re^n$ so that the *semantics* of the symbol (expressed through axioms and explicit in model structures) is preserved

EL Embeddings

- given a theory/ontology T with signature $\Sigma(T)$
- aim: find $f_e : \Sigma(T) \mapsto \Re^n$ s.t. $f_e(\Sigma(T))$ is a model of T ($f_e(\Sigma(T)) \models T$)
- more general: find an algorithm that maps symbols (signatures) into $\Re^n$ so that the *semantics* of the symbol (expressed through axioms and explicit in model structures) is preserved
- any consistent EL++ theory has infinite models

EL Embeddings

- given a theory/ontology T with signature $\Sigma(T)$
- aim: find $f_e : \Sigma(T) \mapsto \Re^n$ s.t. $f_e(\Sigma(T))$ is a model of T ($f_e(\Sigma(T)) \models T$)
- more general: find an algorithm that maps symbols (signatures) into $\Re^n$ so that the *semantics* of the symbol (expressed through axioms and explicit in model structures) is preserved
- any consistent EL++ theory has infinite models
- any consistent EL++ theory has models in $\mathbb{R}^n$ (Loewenheim-Skolem, upwards)

Key idea

- for all $r \in \Sigma(T)$ and $C \in \Sigma(T)$, define $f_e(r)$ and $f_e(C)$
- $f_e(C)$ maps to points in an open n-ball such that $f_e(C) = C^{\mathcal{I}}$:
  $C^{\mathcal{I}} = \{x \in \mathbb{R}^n | \, \|f_e(C) - x\| < r_e(C)\}$
  - these are the *extension* of a class in $\Re^n$
- $f_e(r)$ maps a binary relation r to a vector such that
  $r^{\mathcal{I}} = \{(x, y) | x + f_e(r) = y\}$
  - that's the TransE property for *individuals*
- use the axioms in T as constraints

Algorithm

- normalize the theory:
  - every $\mathcal{EL}^{++}$theory can be expressed using four normal forms (Baader et al., 2005)
- eliminate the ABox: replace each individual symbol with a singleton class: a becomes $\{a\}$
- rewrite relation assertions $r(a, b)$ and class assertions $C(a)$ as $\{a\} \sqsubseteq \exists r.\{b\}$ and $\{a\} \sqsubseteq C$
  - something to remember for the next class-vs-instance discussion?
- normalization rules to generate:
  - $C \sqsubseteq D$
  - $C \sqcap D \sqsubseteq E$
  - $C \sqsubseteq \exists R.D$
  - $\exists R.C \sqsubseteq D$

Algorithm: loss functions

$$
\begin{aligned}
&loss_{C \sqsubseteq D}(c, d) = \\
&\max(0, \|f_\eta(c) - f_\eta(d)\| + r_\eta(c) - r_\eta(d) - \gamma) \\
&+ |\,\|f_\eta(c)\| - 1| + |\,\|f_\eta(d)\| - 1|
\end{aligned} \tag{3}
$$

Algorithm: loss functions

Let $h = \frac{r_\eta(c)^2 - r_\eta(d)^2 + \|f_\eta(c) - f_\eta(d)\|^2}{2\|f_\eta(c) - f_\eta(d)\|}$, then the center and radius of the smallest n-ball containing the intersection of $\eta(C)$ and $\eta(D)$ are $f_\eta(c) + \frac{h}{\|f_\eta(c) - f_\eta(d)\|}(f_\eta(d) - f_\eta(c))$ and $\sqrt{r_\eta(c)^2 - h^2}$.

Algorithm: loss functions

$$
\begin{aligned}
&loss_{C \sqsubseteq \exists R.D}(c, d, r) = \\
&\max(0, \|f_\eta(c) + f_\eta(r) - f_\eta(d)\| + r_\eta(c) - r_\eta(d) - \gamma) \\
&+ |\,\|f_\eta(c)\| - 1| + |\,\|f_\eta(d)\| - 1|
\end{aligned} \tag{4}
$$

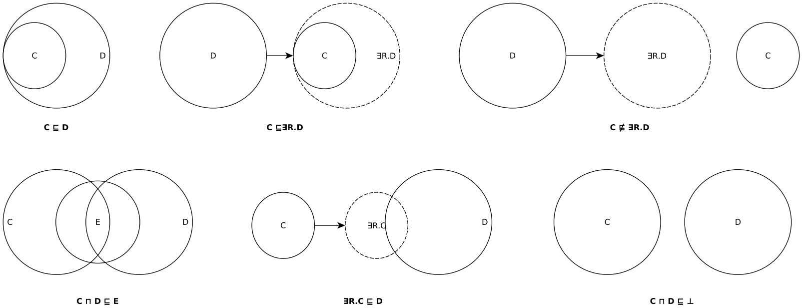
Algorithm: loss functions

$$
\begin{aligned}
&loss_{\exists R.C \sqsubseteq D}(c, d, r) = \\
&\max(0, \|f_\eta(c) - f_\eta(r) - f_\eta(d)\| - r_\eta(c) - r_\eta(d) - \gamma) \\
&+ |\,\|f_\eta(c)\| - 1| + |\,\|f_\eta(d)\| - 1|
\end{aligned} \tag{5}
$$

Algorithm: loss functions

$$
\begin{aligned}
& loss_{C \sqcap D \sqsubseteq \bot}(c, d, e) = \\
& \max(0, r_\eta(c) + r_\eta(d) - \|f_\eta(c) - f_\eta(d)\| + \gamma) \\
& + |\,\|f_\eta(c)\| - 1| + |\,\|f_\eta(d)\| - 1|
\end{aligned} \tag{6}
$$

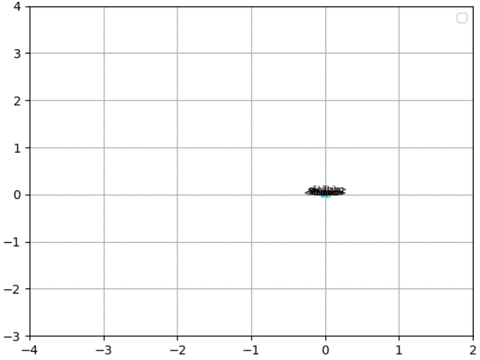
Algorithm: loss functions



EL Embeddings

$$
\begin{align}
Male &\sqsubseteq Person \tag{7}\\
Female &\sqsubseteq Person \tag{8}\\
Father &\sqsubseteq Male \tag{9}\\
Mother &\sqsubseteq Female \tag{10}\\
Father &\sqsubseteq Parent \tag{11}\\
Mother &\sqsubseteq Parent \tag{12}\\
Female \sqcap Male &\sqsubseteq \bot \tag{13}\\
Female \sqcap Parent &\sqsubseteq Mother \tag{14}\\
Male \sqcap Parent &\sqsubseteq Father \tag{15}\\
\exists hasChild.Person &\sqsubseteq Parent \tag{16}\\
Parent &\sqsubseteq Person \tag{17}\\
Parent &\sqsubseteq \exists hasChild.\top \tag{18}
\end{align}
$$

EL Embeddings



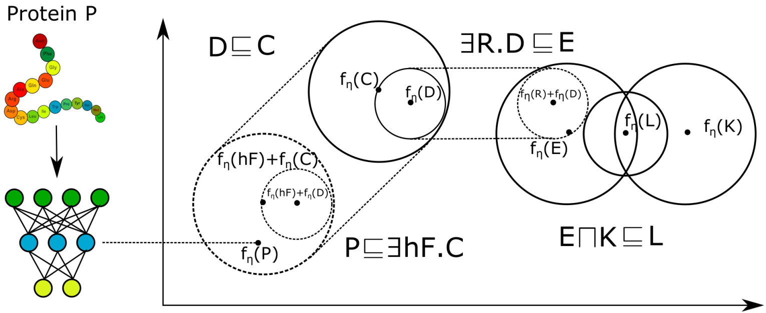
- model with $\Delta = R^n$
- support quantifiers, negation, conjunction,...

IJCAI 2019

Zero-shot prediction with EL Embeddings

- zero-shot prediction:
  - no instance of class C is observed
    - e.g., no protein has *ever* been observed with function C
    - e.g., no individual has *ever* been observed with disease C
  - zero-shot prediction: predict instances of C
    - assumption: C is used in axioms, e.g., $C \sqsubseteq \exists R.D$ or $C \equiv A \sqcap B$
- no training data
  - can we exploit axioms and entailment within the embedding space?

Zero-shot prediction with EL Embeddings



Kulmanov & Hoehndorf, DeepGOZero: Improving protein function prediction from sequence and zero-shot learning based on ontology axioms. ISMB, 2022.

Zero-shot prediction: more general

Embedding

An embedding is a map (morphism) from one mathematical structure X into another structure Y:
$f : X \hookrightarrow Y$
such that X is preserved in Y.

Zero-shot prediction: more general

Embedding

An embedding is a map (morphism) from one mathematical structure X into another structure Y:
$f : X \hookrightarrow Y$
such that X is preserved in Y.

- existence of f^{-1}
  - inverse embedding
  - allows "extraction" of X from Y
- formulation of the prediction problem in X
  - e.g., classification: $X \sqsubseteq \exists R.Y$
  - loss expressed in terms of f^{-1}
- joint embedding of X and Z in Y
  - Z is a third structure providing external information

mOWL

- high-performance software library for machine learning with Semantic Web (OWL) ontologies
- ontology embeddings, zero-shot, constrained optimization
- contains
  - graph generation (DL2Vec, OWL2Vec*, Taxonomy)
  - graph embedding (random walk + word2vec, node2vec, various knowledge graph embedding methods from PyKEEN)
  - model-based embeddings (ELEm, EMEl, ELBE)
  - category-theoretical embeddings
  - FuzzFuzzFuzz ALC embeddings
- Algorithms written in Python + Scala (OWLAPI), tuned for performance

`https://github.com/bio-ontology-research-group/mowl`

Hands-on part

There is a notebook available that shows how to use mOWL with all the methods introduced here. Run the code in the Notebook. When complete, change the task from prediction of PPIs to the prediction of gene–disease associations as in the Semantic Similarity notebook.