Embeddings for Semantic Web ontologies

Robert Hoehndorf

Computational Bioscience Research Center
King Abdullah University of Science and Technology

Logic as foundation of AI

- We want our knowledge of the world to affect our decisions:
  - do action A if world believed to satisfy P
- We need a way to *represent* our knowledge of the world
  - Symbols are entities that *denote*, that we can *interpret*, and which we can *combine* into more complex symbol structures

Knowledge-based methods

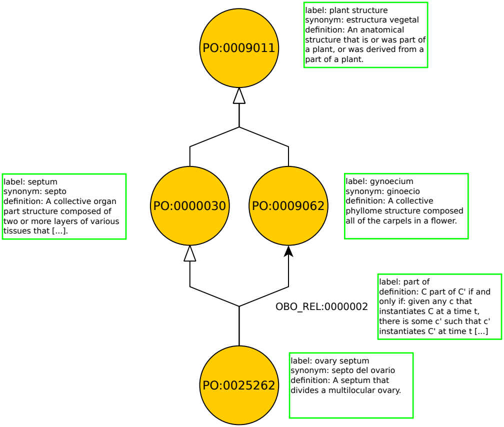
Newell & Simon, 1976

A physical symbol system has the necessary and sufficient means for general intelligent action.

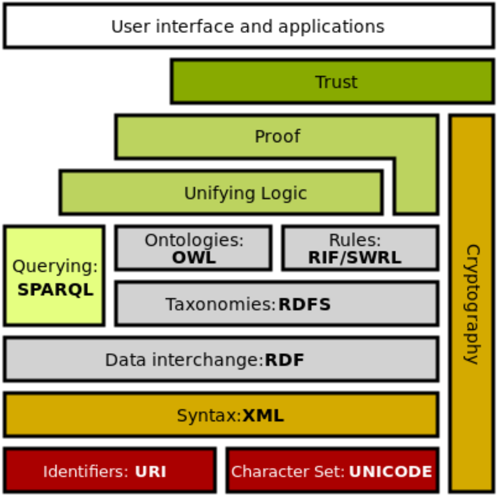
- Symbols are physical entities that can encode for our *knowledge* of the world (e.g., for a proposition P)
  - formal and natural languages
- P may not be *explicitly* represented $\Rightarrow$ need for symbol manipulation
  - reasoning, deduction
- main methods: mathematical logic, formal systems

Making sense of data... with ontologies

- ontology (philosophy) studies the nature of existence and categories of being
- an ontology (computer science) is the "explicit specification of a conceptualization of a domain" [Gruber, 1993]



Semantic Web ontologies



Web Ontology Language (OWL)

- OWL 2 is based on the Description Logic $\mathcal{SROIQ}(\mathcal{D})$

Description Logic $\mathcal{ALC}$: for A a concept name, C and D concepts, R role name:

$$
\begin{aligned}
C, D \rightarrow \quad & A \mid \\
& \top \mid \\
& \bot \mid \\
& \neg C \mid \\
& C \sqcup D \mid \\
& C \sqcap D \mid \\
& \forall R.C \mid \\
& \exists R.C \mid
\end{aligned}
$$

Web Ontology Language (OWL)

- OWL 2 is based on the Description Logic $\mathcal{SROIQ}(\mathcal{D})$
- $\mathcal{ALC}$ with
  - complex role inclusions: $r \circ s \subseteq r$
  - role hierarchy: $r \subseteq s$
  - role transitivity $r \circ r \subseteq r$
  - nominals: $\{a_1, ..., a_n\}$ as concept constructor
  - qualified number restrictions: $(\leq nr.Q)$
  - datatype properties: $\exists r.[\geq n(Integer)]$

Web Ontology Language (OWL)

- OWL 2 is based on the Description Logic $\mathcal{SROIQ}(\mathcal{D})$
- $\mathcal{ALC}$ with
  - complex role inclusions: $r \circ s \subseteq r$
  - role hierarchy: $r \subseteq s$
  - role transitivity $r \circ r \subseteq r$
  - nominals: $\{a_1, ..., a_n\}$ as concept constructor
  - qualified number restrictions: $(\leq nr.Q)$
  - datatype properties: $\exists r.[\geq n(Integer)]$
- two types of axioms:
  - TBox axioms: $C \sqsubseteq D$
  - ABox axioms: $C(a)$ or $R(a, b)$

Web Ontology Language (OWL)

- OWL 2 is based on the Description Logic $\mathcal{SROIQ}(\mathcal{D})$
- $\mathcal{ALC}$ with
  - complex role inclusions: $r \circ s \subseteq r$
  - role hierarchy: $r \subseteq s$
  - role transitivity $r \circ r \subseteq r$
  - nominals: $\{a_1, ..., a_n\}$ as concept constructor
  - qualified number restrictions: $(\leq nr.Q)$
  - datatype properties: $\exists r.[\geq n(Integer)]$
- two types of axioms:
  - TBox axioms: $C \sqsubseteq D$
  - ABox axioms: $C(a)$ or $R(a, b)$
- N.B.: ABox axioms $R(a, b)$ suffice to define "knowledge graphs"

Properties of symbolic methods

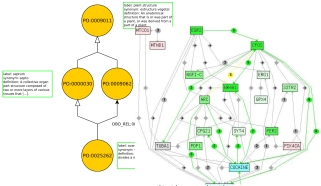
- well-studied for over 2,000 years
- model theory, proof theory
- soundness, completeness
- interpretable; explainable
- biases are "explicit"
  - in the axioms
  - in the deductive closure
- deductive inference $\rightarrow$ "zero-shot"

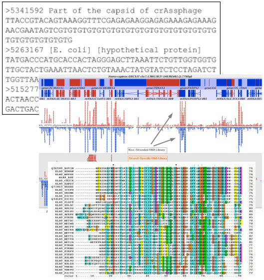
Properties of symbolic methods

- well-studied for over 2,000 years
- model theory, proof theory
- soundness, completeness
- interpretable; explainable
- biases are "explicit"
  - in the axioms
  - in the deductive closure
- deductive inference $\rightarrow$ "zero-shot"

Can we incorporate some of these properties in modern machine learning models?

How do we combine logic and machine learning models?





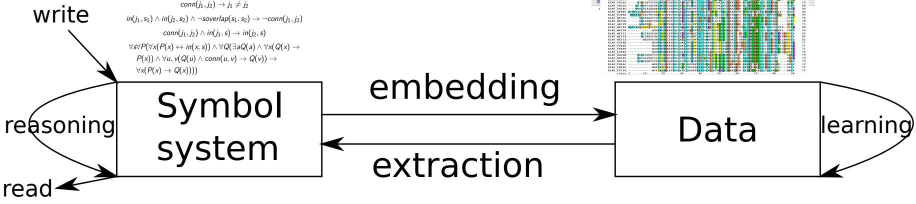
$$conn(j_1, j_2) \rightarrow conn(j_2, j_1)$$

$$conn(j_1, j_2) \rightarrow j_1 \neq j_2$$

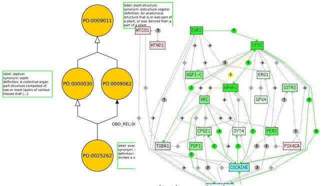
$$in(j_1, s_1) \wedge in(j_2, s_2) \wedge \neg soverlap(s_1, s_2) \rightarrow \neg conn(j_1, j_2)$$

$$conn(j_1, j_2) \wedge in(j_1, s) \rightarrow in(j_2, s)$$

$$\forall s \forall P(\forall x(P(x) \leftrightarrow in(x, s)) \wedge \forall Q(\exists a Q(a) \wedge \forall x(Q(x) \rightarrow P(x)) \wedge \forall u, v(Q(u) \wedge conn(u, v) \rightarrow Q(v)) \rightarrow \forall x(P(x) \rightarrow Q(x))))$$



How do we combine logic and machine learning models?



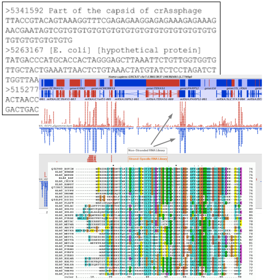
$conn(j_1, j_2) \rightarrow conn(j_2, j_1)$

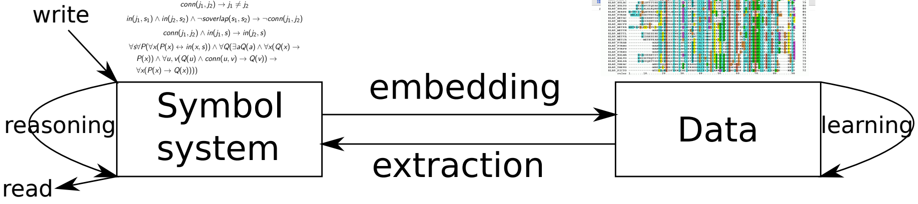
$conn(j_1, j_2) \rightarrow j_1 \neq j_2$

$in(j_1, s_1) \wedge in(j_2, s_2) \wedge \neg soverlap(s_1, s_2) \rightarrow \neg conn(j_1, j_2)$

$conn(j_1, j_2) \wedge in(j_1, s) \rightarrow in(j_2, s)$

$\forall s \forall P(\forall x(P(x) \leftrightarrow in(x, s)) \wedge \forall Q(\exists a Q(a) \wedge \forall x(Q(x) \rightarrow P(x)) \wedge \forall u, v(Q(u) \wedge conn(u, v) \rightarrow Q(v)) \rightarrow \forall x(P(x) \rightarrow Q(x))))$





- 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

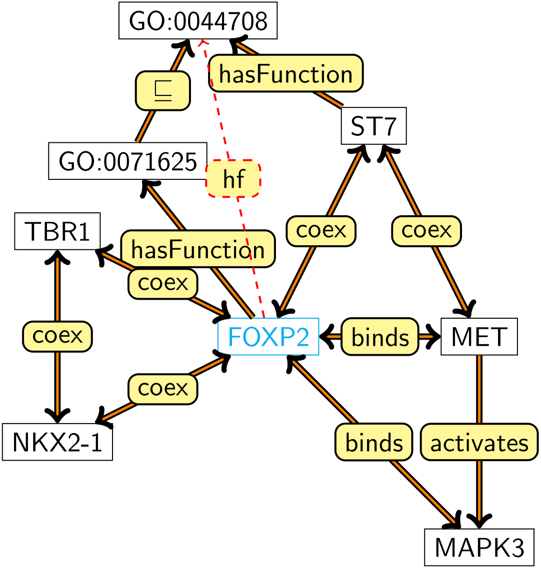
An embedding is a map (morphism) from one mathematical structure X into another structure Y:
$f : X \hookrightarrow Y$
such that X is preserved in Y.

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

We want to embed formalized *knowledge bases* in $\mathbb{R}^n$. Approaches:

- graph-based, syntactic
- model-theoretic

Knowledge graph embeddings



graph embedding, e.g.:

- iterated random walks generate "sentences"
- embedding through language model
- use embedding for clustering, edge prediction, etc.

Knowledge to (knowledge) graphs

- $X \sqsubseteq Y$: $X \xrightarrow{\text{is-a}} Y$
- $X \sqsubseteq \exists\text{part-of}.Y$: $X \xrightarrow{\text{part-of}} Y$
- $X \sqsubseteq \exists\text{regulates}.Y$: $X \xrightarrow{\text{regulates}} Y$
- $X \sqcap Y \sqsubseteq \bot$: $X \xleftrightarrow{\text{disjoint}} Y$
- $X \equiv Y$: $X \overset{\equiv}{\leftrightarrow} Y$, $\{X, Y\}$

Asserted and inferred:

- $X \sqsubseteq \exists\text{part-of}.Y$
- $Y \sqsubseteq \exists\text{part-of}.Z$
- $\text{part-of} \circ \text{part-of} \subseteq \text{part-of}$
- entails: $X \sqsubseteq \exists\text{part-of}.Z$

More semantics

- walk- (and word-)based methods are not truly "semantic"
- semantics relies on interpretation (Σ-algebras)
- an interpretation is a *model* of T if all $C \sqsubseteq D$ are satisfied

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	$\exists r.C$	$\{x \in \Delta^{\mathcal{I}} \mid \exists y \in \Delta^{\mathcal{I}} : (x, y) \in r^{\mathcal{I}} \land y \in C^{\mathcal{I}}\}$
generalized concept inclusion	$C \sqsubseteq D$	$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
role inclusion	$r_1 \circ ... \circ r_n \sqsubseteq r$	$r_1^{\mathcal{I}} \circ ... \circ r_n^{\mathcal{I}} \subseteq r^{\mathcal{I}}$

EL Embeddings

- given a knowledge base T with signature $\Sigma(T)$
- aim: find $f_e : \Sigma(T) \mapsto \mathbb{R}^n$ s.t. $f_e(\Sigma(T))$ is a model of T ($f_e(\Sigma(T)) \models T$)
  - maps symbols into $\mathbb{R}^n$ while preserving their model-theoretic semantics

EL Embeddings

- given a knowledge base T with signature $\Sigma(T)$
- aim: find $f_e : \Sigma(T) \mapsto \mathbb{R}^n$ s.t. $f_e(\Sigma(T))$ is a model of T ($f_e(\Sigma(T)) \models T$)
  - maps symbols into $\mathbb{R}^n$ while preserving their model-theoretic semantics
- preliminaries:
  - any consistent $\mathcal{EL}^{++}$theory has infinite models

EL Embeddings

- given a knowledge base T with signature $\Sigma(T)$
- aim: find $f_e : \Sigma(T) \mapsto \mathbb{R}^n$ s.t. $f_e(\Sigma(T))$ is a model of T ($f_e(\Sigma(T)) \models T$)
  - maps symbols into $\mathbb{R}^n$ while preserving their model-theoretic semantics
- preliminaries:
  - any consistent $\mathcal{EL}^{++}$theory has infinite models
  - any consistent $\mathcal{EL}^{++}$theory has models in $\mathbb{R}^n$ (theorem of Löwenheim-Skolem, upwards)

EL Embeddings

- for all $r \in \Sigma(T)$ and $C \in \Sigma(T)$, define $f_e(r)$, $f_e(C)$, and $r_e(C)$

EL Embeddings

- for all $r \in \Sigma(T)$ and $C \in \Sigma(T)$, define $f_e(r)$, $f_e(C)$, and $r_e(C)$
- $f_e(C)$ and $r_e(C)$ map to points in an open n-ball with the aim that $C^{\mathcal{I}} = \{x \in \mathbb{R}^n | \, \|f_e(C) - x\| < r_e(C)\}$
  - these will be the *extension* of a unary predicate in $\mathbb{R}^n$

EL Embeddings

- for all $r \in \Sigma(T)$ and $C \in \Sigma(T)$, define $f_e(r)$, $f_e(C)$, and $r_e(C)$
- $f_e(C)$ and $r_e(C)$ map to points in an open n-ball with the aim that $C^{\mathcal{I}} = \{x \in \mathbb{R}^n | \, \|f_e(C) - x\| < r_e(C)\}$
  - these will be the *extension* of a unary predicate in $\mathbb{R}^n$
- $f_e(r)$ maps a binary predicate r to a vector with the aim that $r^{\mathcal{I}} = \{(x, y) | x + f_e(r) = y\}$

EL Embeddings

- for all $r \in \Sigma(T)$ and $C \in \Sigma(T)$, define $f_e(r)$, $f_e(C)$, and $r_e(C)$
- $f_e(C)$ and $r_e(C)$ map to points in an open n-ball with the aim that $C^{\mathcal{I}} = \{x \in \mathbb{R}^n | \, \|f_e(C) - x\| < r_e(C)\}$
  - these will be the *extension* of a unary predicate in $\mathbb{R}^n$
- $f_e(r)$ maps a binary predicate r to a vector with the aim that $r^{\mathcal{I}} = \{(x, y) | x + f_e(r) = y\}$
- use axioms in T to find f_e and r_e
- sufficient to focus on certain normal forms of axioms in T

EL Embeddings

- eliminate the ABox:
  - rewrite role assertions $r(a, b)$ as $\{a\} \sqsubseteq \exists r.\{b\}$
  - rewrite class assertions $C(a)$ as $\{a\} \sqsubseteq C$
  - replace $\{a\}$ with N_a
- normalize the TBox (Baader et al., 2005):
  - $C \sqsubseteq D$ and $C \sqsubseteq \bot$
  - $C \sqcap D \sqsubseteq E$ and $C \sqcap D \sqsubseteq \bot$
  - $C \sqsubseteq \exists R.D$ and $C \sqsubseteq \exists R.\bot$
  - $\exists R.C \sqsubseteq D$

EL Embeddings

- eliminate the ABox:
  - rewrite role assertions $r(a, b)$ as $\{a\} \sqsubseteq \exists r.\{b\}$
  - rewrite class assertions $C(a)$ as $\{a\} \sqsubseteq C$
  - replace $\{a\}$ with N_a
- normalize the TBox (Baader et al., 2005):
  - $C \sqsubseteq D$ and $C \sqsubseteq \bot$
  - $C \sqcap D \sqsubseteq E$ and $C \sqcap D \sqsubseteq \bot$
  - $C \sqsubseteq \exists R.D$ and $C \sqsubseteq \exists R.\bot$
  - $\exists R.C \sqsubseteq D$
- minimize loss
  - randomly initialize f_e and r_e
  - one loss function for each normal form

Algorithm: loss functions

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

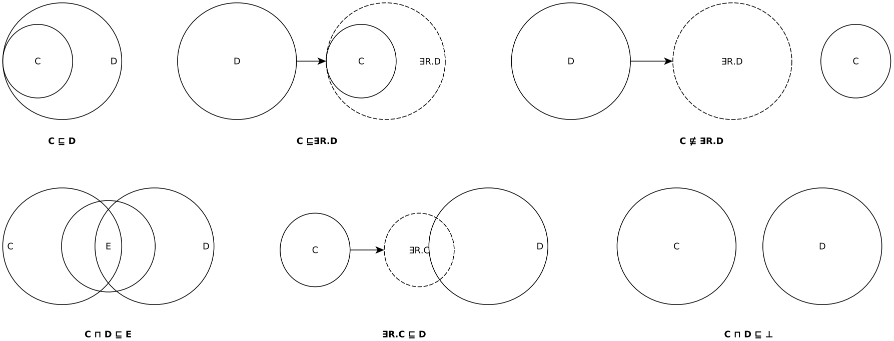
- for all $i \in C^{\mathcal{I}}$, $i + r^{\mathcal{I}} \in D^{\mathcal{I}}$
- $C^{\mathcal{I}} + r^{\mathcal{I}} \subseteq D^{\mathcal{I}}$

Algorithm: loss functions

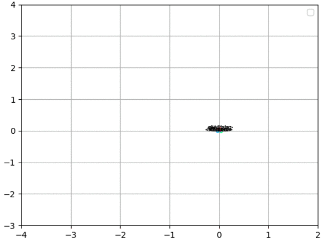
$$
\begin{aligned}
&loss_{C \sqsubseteq \exists R.D}(c, d, r) = \\
&\max(0, \|f_e(c) + f_e(r) - f_e(d)\| + r_e(c) - r_e(d) - \gamma) \qquad (1)\\
&+ |\,\|f_e(c)\| - 1| + |\,\|f_e(d)\| - 1|
\end{aligned}
$$

- for all $i \in C^{\mathcal{I}}$, $i + r^{\mathcal{I}} \in D^{\mathcal{I}}$
- $C^{\mathcal{I}} + r^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
- margin γ
- regularization

Algorithm: loss functions



EL Embeddings



Theorem (Soundness)

Let T be a theory in $\mathcal{EL}^{++}$. If $\gamma \leq 0$ and $loss_n(\eta(T)) = 0$ then T has a model.

$$
\begin{aligned}
Male &\sqsubseteq Person \\
Female &\sqsubseteq Person \\
Father &\sqsubseteq Male \\
Mother &\sqsubseteq Female \\
Father &\sqsubseteq Parent \\
Mother &\sqsubseteq Parent \\
Female \sqcap Male &\sqsubseteq \bot \\
Female \sqcap Parent &\sqsubseteq Mother \\
Male \sqcap Parent &\sqsubseteq Father \\
\exists hasChild.Person &\sqsubseteq Parent \\
Parent &\sqsubseteq Person \\
Parent &\sqsubseteq \exists hasChild.\top
\end{aligned}
$$

Predicting protein–protein interactions based on protein functions

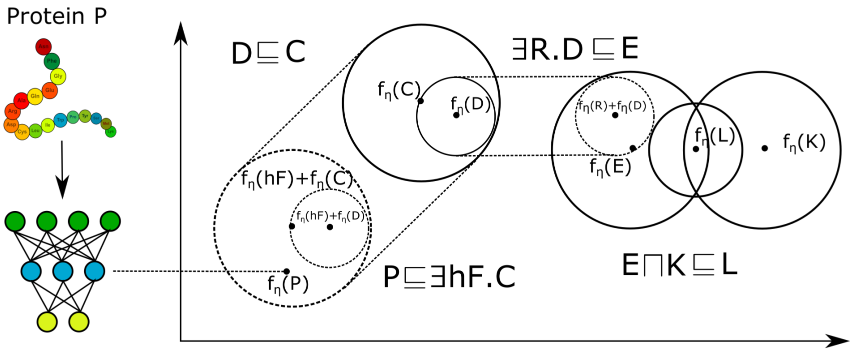
Predicting PPIs in yeast using functional relatedness of proteins:

Method	Hits@10 (R)	Hits@10 (F)	Hits@100 (R)	Hits@100 (F)	Mean Rank (R)	Mean Rank (F)	AUC (R)	AUC (F)
TransE (RDF)	0.03	0.05	0.22	0.27	855	809	0.84	0.85
TransE (plain)	0.06	0.13	0.41	0.54	378	330	0.93	0.94
SimResnik	0.08	0.18	0.38	0.49	713	663	0.87	0.88
SimLin	0.08	0.17	0.34	0.45	807	756	0.85	0.86
EmEL ++	0.07	0.17	0.48	0.65	336	291	0.94	0.95
EL Embeddings	0.10	0.23	0.50	0.75	247	187	0.96	0.97
ELBE	**0.11**	**0.26**	**0.57**	**0.77**	**201**	**154**	**0.96**	**0.97**

Zero-shot prediction with EL Embeddings

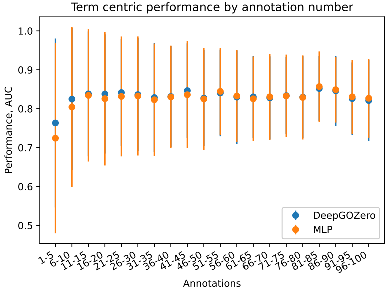
- zero-shot prediction:
  - no instance of class C is observed
    - e.g., no protein has *ever* been observed with function C
  - zero-shot prediction: predict instances of C
    - assumption: C is used in axioms, e.g., $C \sqsubseteq \exists R.D$ or $C \equiv A \sqcap B$
- no training data
  - can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

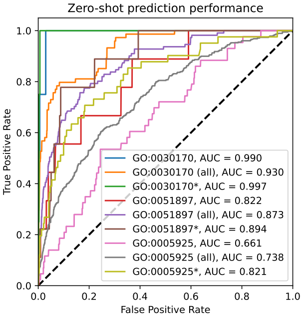
Zero-shot prediction with EL Embeddings



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

Zero shot protein function prediction





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, EMEI, ELBE)
- Algorithms written in Python + Scala (OWLAPI), tuned for performance

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

Limitations and open research questions

- intersection of two n-balls is no n-ball
  - solution: axis-parallel rectangles (boxes) instead of n-balls (ELBE – EL Box Embeddings)

Limitations and open research questions

- intersection of two *n*-balls is no *n*-ball
  - solution: axis-parallel rectangles (boxes) instead of *n*-balls (ELBE – EL Box Embeddings)
- missing completeness, degenerated models
  - no "entailment": $T \models \phi$ iff $Mod(T \cup \{\neg\phi\}) = \emptyset$

Limitations and open research questions

- intersection of two *n*-balls is no *n*-ball
  - solution: axis-parallel rectangles (boxes) instead of *n*-balls (ELBE – EL Box Embeddings)
- missing completeness, degenerated models
  - no "entailment": $T \models \phi$ iff $Mod(T \cup \{\neg\phi\}) = \emptyset$
- relation model (TransE) allows only 1:1 relations
  - more complex relation models?

Limitations and open research questions

- intersection of two *n*-balls is no *n*-ball
  - solution: axis-parallel rectangles (boxes) instead of *n*-balls (ELBE – EL Box Embeddings)
- missing completeness, degenerated models
  - no "entailment": $T \models \phi$ iff $Mod(T \cup \{\neg\phi\}) = \emptyset$
- relation model (TransE) allows only 1:1 relations
  - more complex relation models?
- expressivity limited to EL$++$
  - (unrestricted) negation, union, universal quantifiers?

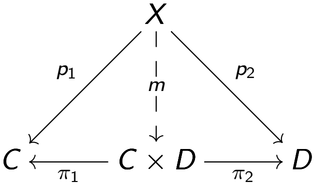
Categorical semantics

- category theory provides an alternative way to represent semantics of Description Logic theories
  - Curry-Howard-Lambek correspondence
  - intuitionistic (non-classical, constructive) logic
- relation to proof theory (Gentzen-style sequent calculus)

Categorical semantics

- category theory provides an alternative way to represent semantics of Description Logic theories
  - Curry-Howard-Lambek correspondence
  - intuitionistic (non-classical, constructive) logic
- relation to proof theory (Gentzen-style sequent calculus)

$C \sqcap D$:



(2)

Categorical semantics

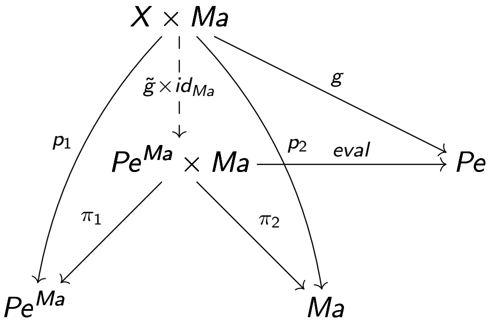
- category theory provides an alternative way to represent semantics of Description Logic theories
  - Curry-Howard-Lambek correspondence
  - intuitionistic (non-classical, constructive) logic
- relation to proof theory (Gentzen-style sequent calculus)

$C \sqcup D$:

$$\begin{array}{ccccc} C & \xrightarrow{\iota_1} & C+D & \xleftarrow{\iota_2} & D \\ & \searrow^{i_1} & \downarrow^{m} & \swarrow_{i_2} & \\ & & X & & \end{array} \tag{3}$$

Categorical semantics

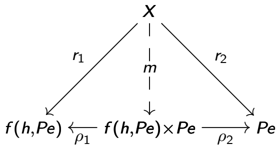
$$Male \sqsubseteq Person \tag{4}$$

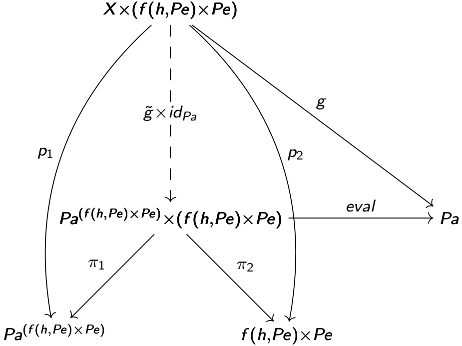


(5)

Categorical semantics

$$\exists hasChild.Person \sqsubseteq Parent$$





Categorical embeddings

- semantics induced by commutativity in diagrams
- morphisms: linear transformations
- commutativity of “triangles” as optimization objective
- computation of entailments = testing commutativity in diagrams
- implemented in mOWL
- soundness and completeness results still missing

Categorical embeddings

Method	Hits10		Hits@100		Mean Rank		AUC	
	R	F	R	F	R	F	R	F
TransE (RDF)	0.03	0.05	0.22	0.27	855	809	0.84	0.85
TransE (plain)	0.06	0.13	0.41	0.54	378	330	0.93	0.94
SimResnik	0.08	0.18	0.38	0.49	713	663	0.87	0.88
SimLin	0.08	0.17	0.34	0.45	807	756	0.85	0.86
EmEL ++	0.07	0.17	0.48	0.65	336	291	0.94	0.95
EL Embeddings	**0.10**	**0.23**	0.50	**0.75**	247	187	**0.96**	**0.97**
CatEmbeddings (n=50)	0.06	0.14	0.48	0.68	246	198	**0.96**	0.96
CatEmbeddings (=1024)	0.09	0.17	**0.52**	0.71	**231**	**185**	**0.96**	**0.97**

Table: Prediction performance for yeast protein-protein interactions. (R=Raw, F=Filtered)

Fuzzy logic

- crisp sets defined by membership
- fuzzy sets defined by “degree” of membership for all domain entities: $m(e, S)$
- (differentiable) fuzzy operators (t-norm, t-conorm)
- quantifiers: finite models, domain closure
  - sub-sampling infinite domains
- fuzzy entailments (question answering)

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

What's next?

- sound and complete ontology embeddings:
  - "model-generating" embeddings (the embedding *is* the model)
  - neural networks, fuzzy logic, differentiable t-norms
  - crucial for deductive inference
- expanding mOWL
  - library for **all** embedding methods for Semantic Web ontologies
  - algorithms for neuro-symbolic systems

Summary

Feigenbaum, 1977

[The domain-specific knowledge] plays a critical role in organizing and constraining search. The theme is that in the knowledge is the power.

- traditionally: for discrete search
- more recently: for continuous search and optimization (machine learning)
- hallmark of a knowledge-based system: gain new knowledge and change behavior accordingly
- encoding background knowledge
- reduce biases (making biases explicit)
- provide formal guarantees

Acknowledgements

Thank you

Amyloid beta
Protein classified with blood
coagulation.

A Semantic Haiku

generated from the UniProt Knowledgebase

`http://borg.kaust.edu.sa`
robert.hoehndorf@kaust.edu.sa