Machine learning with OWL 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

Logic as foundation of AI

- P may not be *explicitly* represented $\Rightarrow$ symbol manipulation
- Example:
  - Patient x is allergic to medication m: $all(x, m)$
  - Anybody allergic to m is also allergic to m':
    $\forall y(all(y, m) \rightarrow all(y, m'))$
  - It's not safe to prescribe medication if allergic:
    $\forall a, b(all(a, b) \rightarrow \neg safe(a, b))$
  - Is it safe to prescribe m' to x: $safe(x, m')$?

Knowledge-based methods

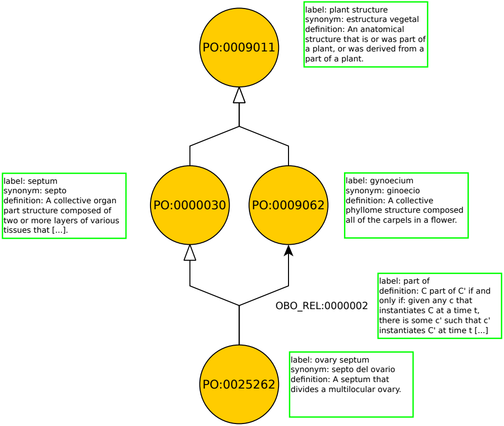
Newell & Simon, 1976

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]



Making sense of data... with ontologies

Ontologies provide

- standard identifiers — for integration of data
- terms — domain vocabulary
- definitions — for human understanding
- axioms — for machine understanding

Description Logic: ALC

Definition

An interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consists of a non-empty domain $\Delta^{\mathcal{I}}$ and an interpretation function $\cdot^{\mathcal{I}}$:

- $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ for all $A \in N_C$,
- $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ for all $r \in N_R$

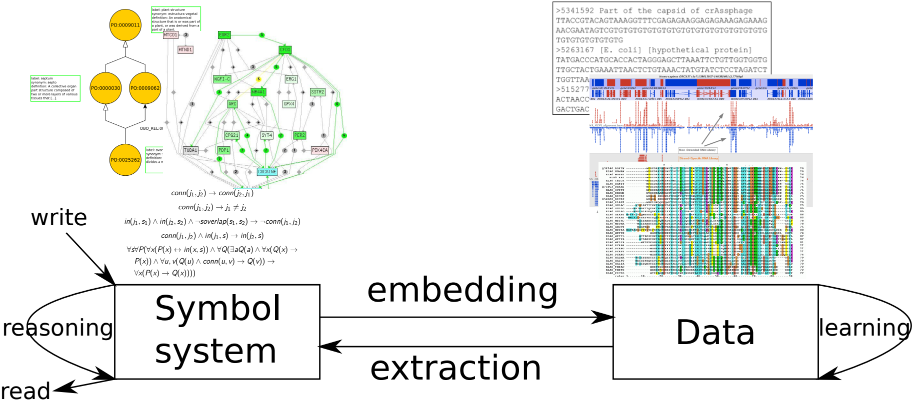
The interpretation function is extended to $\mathcal{ALC}$ concept descriptions as follows:

- $(C \sqcap D)^{\mathcal{I}} := C^{\mathcal{I}} \cap D^{\mathcal{I}}$
- $(C \sqcup D)^{\mathcal{I}} := C^{\mathcal{I}} \cup D^{\mathcal{I}}$
- $(\neg C)^{\mathcal{I}} := \Delta^{\mathcal{I}} - C^{\mathcal{I}}$
- $(\forall r.C)^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} | \text{for all } e \in \Delta^{\mathcal{I}} : (d, e) \in r^{\mathcal{I}} \text{ implies } e \in C^{\mathcal{I}}\}$
- $(\exists r.C)^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} | \text{there is } e \in \Delta^{\mathcal{I}} : (d, e) \in r^{\mathcal{I}} \text{ and } e \in C^{\mathcal{I}}\}$

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: $\exists r.[\geq n(Integer)]$

How do we combine logic and machine learning models?



- 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:

$f\colon X \hookrightarrow Y$

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

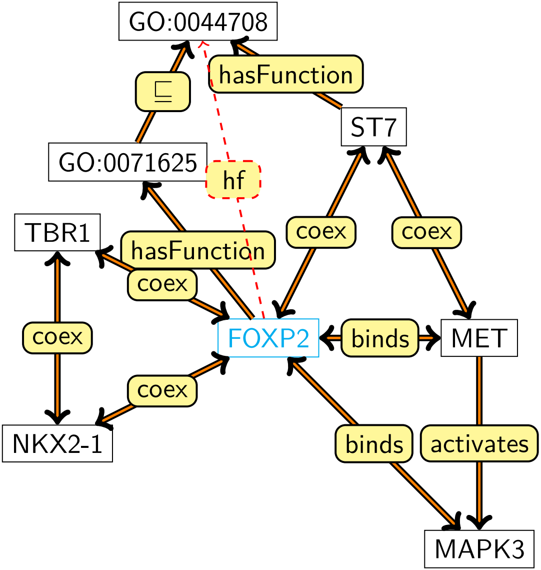
An embedding is a map (morphism) from one mathematical structure X into another structure Y:
$f\colon X \hookrightarrow Y$
such that X is preserved in Y.

- 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

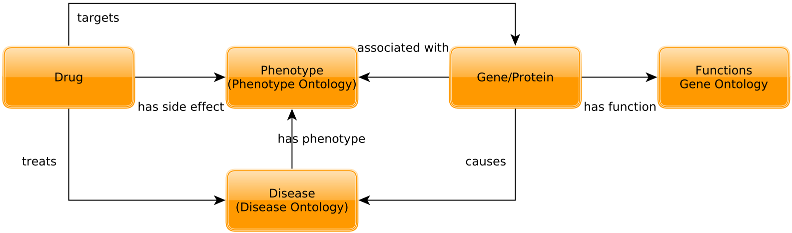
Knowledge graph embeddings



graph embedding, e.g.:

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

Relation prediction



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

Alshahrani et al., Bioinformatics, 2017

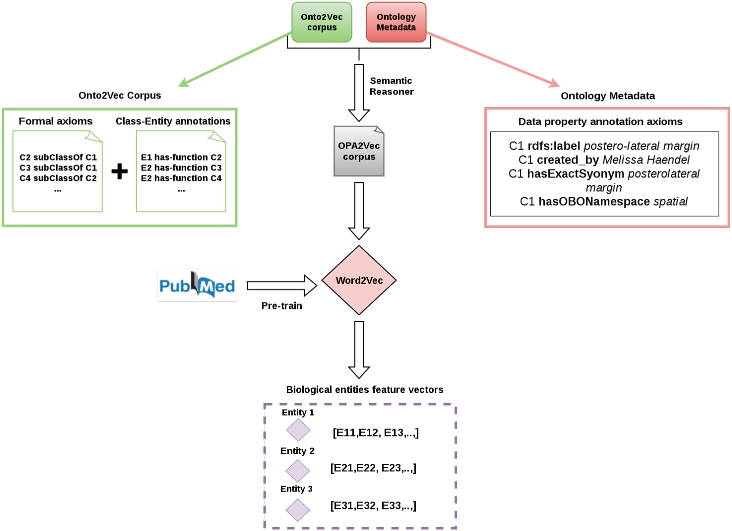
Onto2Vec: Exploiting axioms

Gene Ontology:

- behavior $\sqsubseteq \forall$in-taxon.metazoa
- behavior $\sqcap$ 'developmental process' $\sqsubseteq \bot$
- 'cell proliferation' $\sqcap$ $\exists$in-taxon.fungi $\sqsubseteq \bot$
- 'cell growth' $\equiv$ growth $\sqcap$ $\exists$'results in growth of'.cell
- ...

Can we embed these axioms *directly*?

OPA2Vec: multimodality and transfer learning



Smaili et al., Bioinformatics, 2018, 2019

Onto2Vec/OPA2Vec: multimodality and transfer learning

Protein-protein interaction benchmark (ROCAUC)

	Human	Yeast
Onto2Vec (no reasoner)	0.7385	0.7439
Onto2Vec	0.7614	0.7701
Onto2Vec (NN)	0.8779	0.8731
OPA2Vec	0.8727	0.8622
OPA2Vec (NN)	**0.9033**	**0.9011**
Resnik (baseline)	0.7891	0.7924

- “trainable” semantic similarity measures over ontologies
  - applicable to a large class of problems in bioinformatics
- based on co-occurrence patterns in Description Logic axioms *and* their deductive closure

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

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 $\mathbb{R}^n$ (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

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

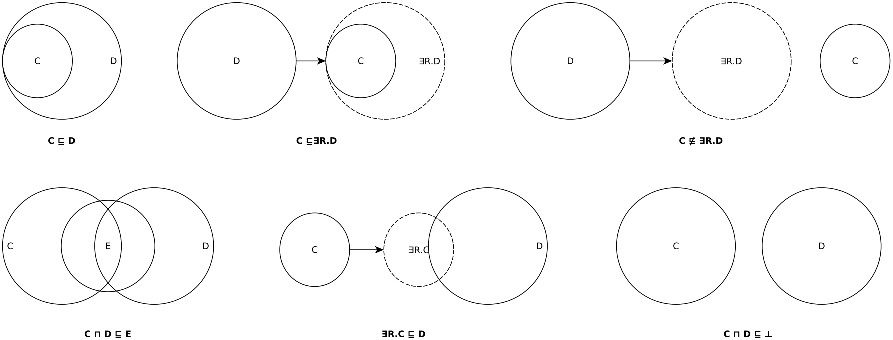
- for all $i \in C^{\mathcal{I}}$, $i + r^{\mathcal{I}} \in D^{\mathcal{I}}$
- $C^{\mathcal{I}} + r^{\mathcal{I}} \subseteq D^{\mathcal{I}}$

Algorithm: loss functions

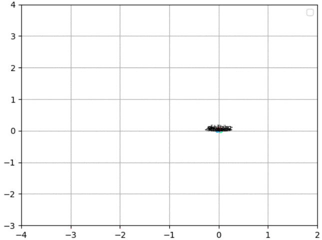
$$
\begin{aligned}
& loss_{C \sqsubseteq \exists R.D}(c, d, r) = \\
& \max(0, \|f_e(c) + f_e(r) - f_e(d)\| + r_e(c) - r_e(d) - \gamma) \\
& + |\,\|f_e(c)\| - 1| + |\,\|f_e(d)\| - 1|
\end{aligned} \tag{1}
$$

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

EL Box Embeddings

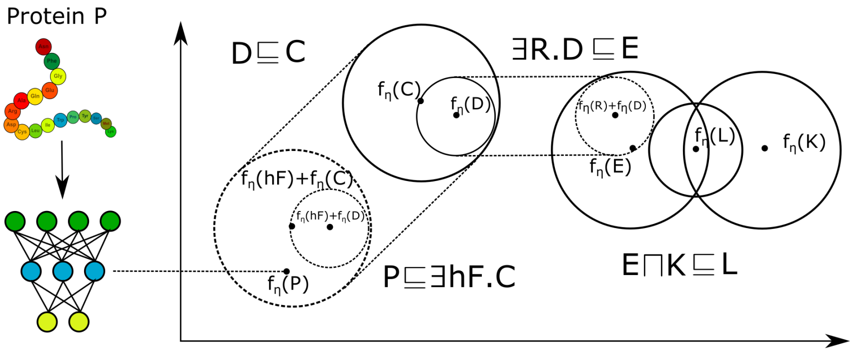
Limitations of EL Embeddings:

- intersection of two *n*-balls is no *n*-ball
  - solution: axis-parallel rectangles (boxes) instead of *n*-balls (ELBE)
- relation model (TransE) allows only 1:1 relations
  - solution: more complex relation models
- expressivity limited to EL$++$
  - how about (unrestricted) negation, union, universal quantifiers?
- imbalance of axioms in normal forms
  - weights
- missing completeness result!

Zero-shot prediction with EL Embeddings

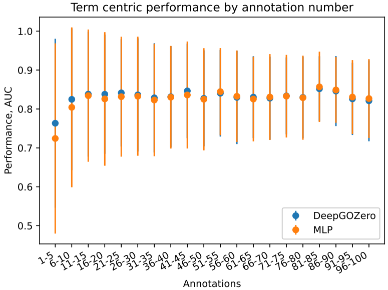
- zero-shot phenotype prediction:
  - no instance of class C is observed
    - no protein has *ever* been observed with function C
  - zero-shot prediction: predict instances of C
    - assumption: C is used in axioms, e.g., $C \sqsubseteq \exists R.D$ or $A \sqcap B \equiv C$
- no training data
  -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

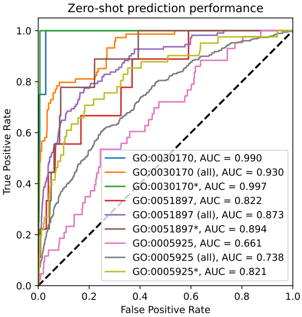
Zero-shot prediction with EL Embeddings



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

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

What's next?

- sound and complete ontology embeddings:
  - "model-generating" geometric embeddings (the embedding *is* the model)
  - neural networks, fuzzy logic, differentiable t-norms
  - category theory and (homotopy) type theory
    - categorical semantics to generate graphs

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
- AI models in biology need to rely on biological background knowledge
  - from > 100 years of experiments and observation
  - needs appropriate "embedding" algorithms

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