Model-theoretic embeddings

Maxat Kulmanov

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"

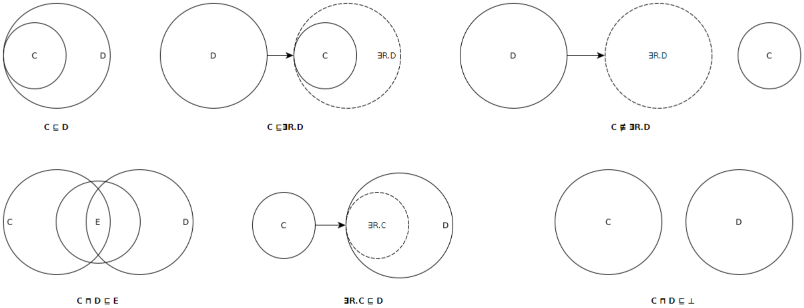
How to overcome the semantic gap?

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

ELEmbeddings



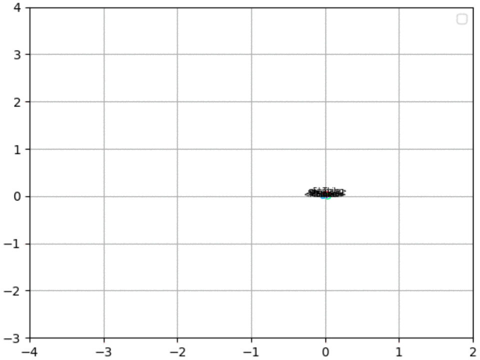
Algorithm

- 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:
  - $C \sqsubseteq D$
  - $C \sqcap D \sqsubseteq E$
  - $C \sqsubseteq \exists R.D$
  - $\exists R.C \sqsubseteq D$

Family Domain

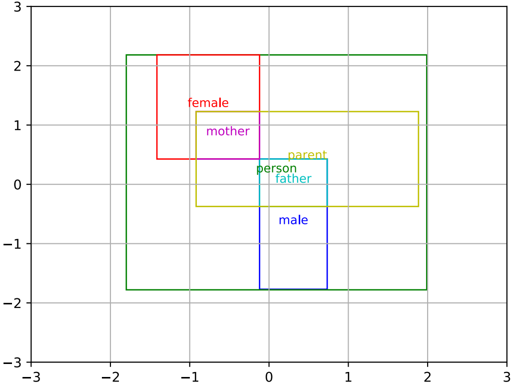
$$
\begin{aligned}
Male &\sqsubseteq Person && (1)\\
Female &\sqsubseteq Person && (2)\\
Father &\sqsubseteq Male && (3)\\
Mother &\sqsubseteq Female && (4)\\
Father &\sqsubseteq Parent && (5)\\
Mother &\sqsubseteq Parent && (6)\\
Female \sqcap Male &\sqsubseteq \bot && (7)\\
Female \sqcap Parent &\sqsubseteq Mother && (8)\\
Male \sqcap Parent &\sqsubseteq Father && (9)\\
\exists hasChild.Person &\sqsubseteq Parent && (10)\\
Parent &\sqsubseteq Person && (11)\\
Parent &\sqsubseteq \exists hasChild.\top && (12)
\end{aligned}
$$

Family domain

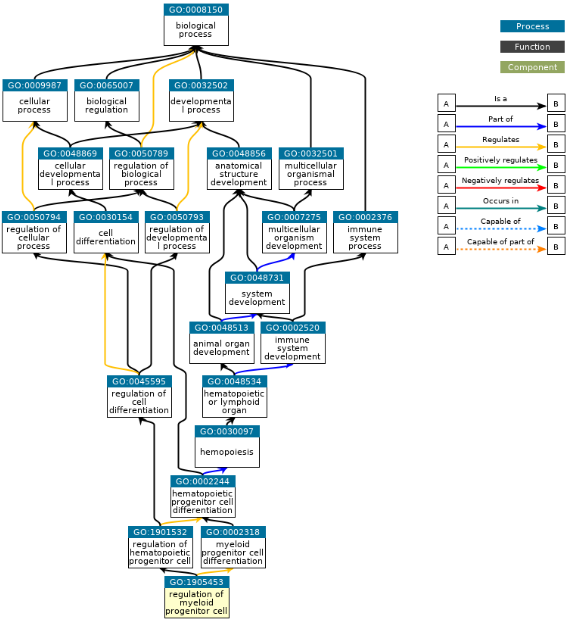


- model with $\Delta = R^n$
- support quantifiers, negation, conjunction,...

EL BOX Embeddings



Gene Ontology



QuickGO - https://www.ebi.ac.uk/QuickGO

DeepGOZero - function prediction