Embedding Complex Knowledge: From Geometric to Language Models

Jiaoyan Chen
Department of Computer Science
The University of Manchester

3rd March, 2026
KMI Seminar, The Open University

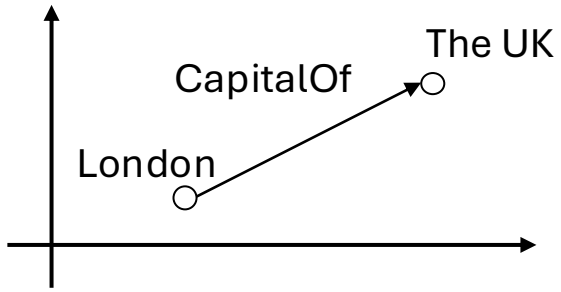
Embedding Symbolic Knowledge

- **Vector** or **parameter**-based representation of symbolic knowledge

- Why?
  - Knowledge inference with **uncertainty** (e.g., incompleteness, approximation & prediction, induction of schema & rule)
  - Similarity-based **matching** across modalities (e.g., retrieval, alignment and resolution)
  - **Inject knowledge** into parameter-based models (e.g., tuning LLM)
  - Kinds of **downstream applications** with machine learning and statistical models

Knowledge Graph Embedding

- Originate from **word embeddings**
  - Represent word cooccurrence and correlation by a neural network; e.g., similarity: (cat, kitten) > (cat, dog)
- Mostly aim at facts of RDF triples e.g., <London, CapitalOf, The UK>
  - Geometric modeling: TransE, TransR, TransH, ..



TransE: modeling relation as a translation mapping

- Sequence learning: RDF2Vec, ...
- Graph propagation: R-GCN, ...

Ontology (Description Logic) Embedding

- How to represent more complex ontologies of Description Logic (DL) in Euclidean space?

$\mathcal{T}$ = {Father $\sqsubseteq$ Parent $\sqcap$ Male, Mother $\sqsubseteq$ Parent $\sqcap$ Female,
Child $\sqsubseteq \exists$hasParent.Father, Child $\sqsubseteq \exists$hasParent.Mother,
hasParent $\sqsubseteq$ relatedTo}
$\mathcal{A}$ = {Father(Alex), Child(Bob), hasParent(Bob, Alex)}

A toy famil ontology in DL $\mathcal{EL}^{++}$ which allows complex concept construction:

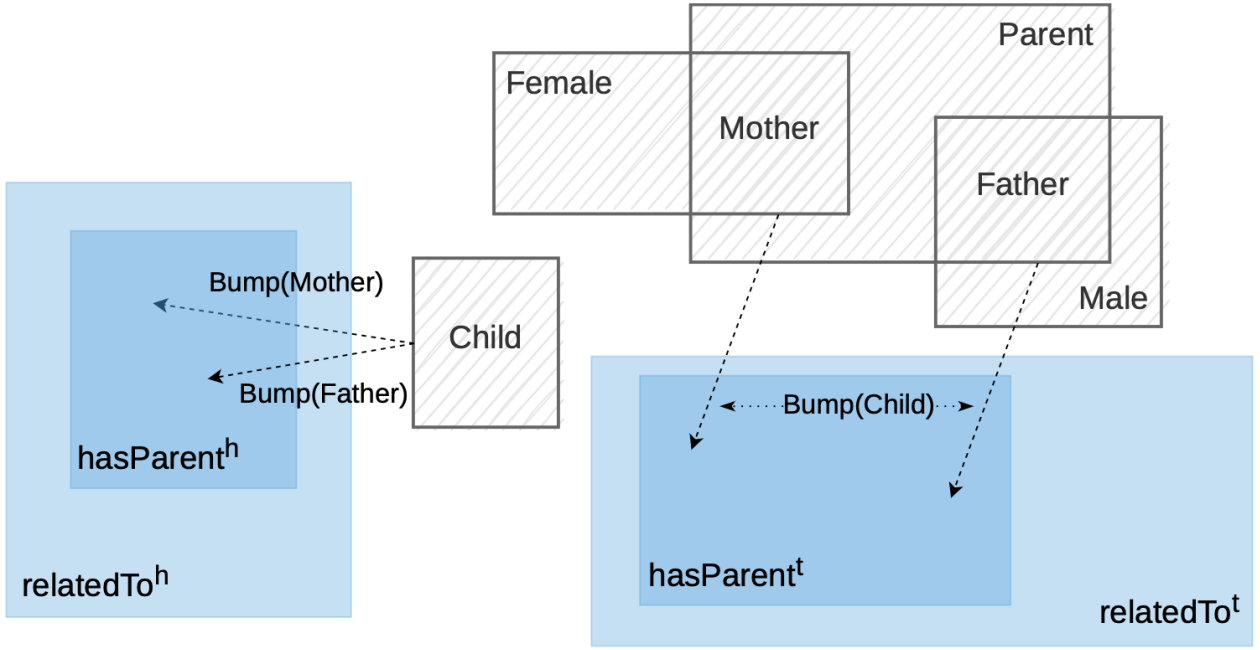
$$\bot \mid \top \mid A \mid C \sqcap D \mid \exists r.C \mid \{a\}$$

- Embedding: Region-based
  - Individual – Point
  - Concept – Ball, Box, …

Ontology (Description Logic) Embedding

- Example: Box^{2}EL
  - Individual: one n-point
  - Concept: one n-box
    - Conjunction, subsumption, membership
  - Relation: two n-boxes (head & tail)
    - Composition, subsumption
  - Concept interaction: bumping vector
    - Existential quantification

$Child \sqsubseteq \exists hasParent.Father$



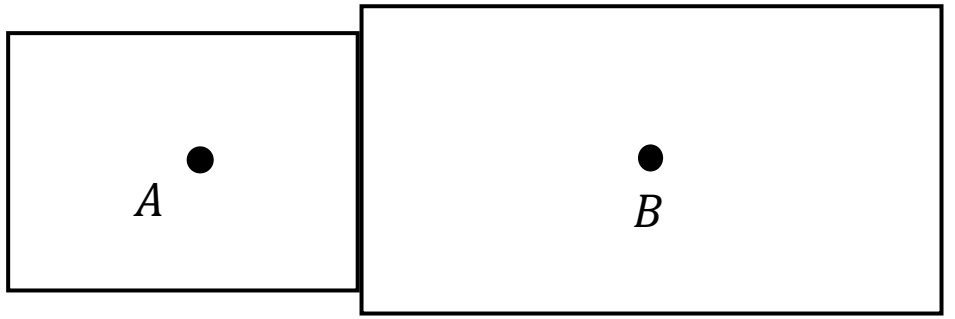
Representation of the family ontology in Box^{2}EL

Jackermeier, Mathias, Jiaoyan Chen, and Ian Horrocks. "Dual box embeddings for the description logic EL++." *Proceedings of the ACM Web Conference 2024*. 2024.

Ontology (Description Logic) Embedding

- Box^{2}EL training:
  - Element-wise distance of two boxes:

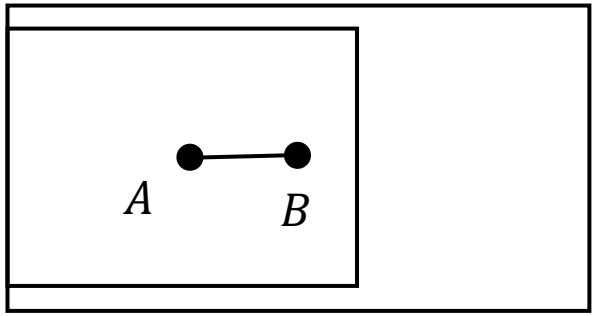
$$\boldsymbol{d}(A,B) = |\boldsymbol{c}(A) - \boldsymbol{c}(B)| - \boldsymbol{o}(A) - \boldsymbol{o}(B)$$



The horizontal dimension distance is 0

  - Score/loss of concept subsumption (inclusion of two boxes):

$$\mathcal{L}_{\subseteq}(A,B) = \begin{cases} \|\max\{\mathbf{0},\ \boldsymbol{d}(A,B) + 2\boldsymbol{o}(A) - \gamma\}\| & \text{if } B \neq \emptyset \\ \max\{0, \boldsymbol{o}(A)_1 + 1\} & \text{otherwise,} \end{cases}$$



The horizontal dimension loss is 0

Box A moves left, the horizontal dimension loss > 0, penalize the embeddings by training;

Box A moves right, the horizontal dimension loss is still 0

Ontology (Description Logic) Embedding

- Box^{2}EL training: loess/scores for axioms of each normal form (NF)
  - NF1: $C \sqsubseteq D$ $$\mathcal{L}_1(C, D) = \mathcal{L}_{\subseteq}(\mathrm{Box}(C), \mathrm{Box}(D))$$
  - NF2: $C \sqcap D \sqsubseteq E$ $$\mathcal{L}_2(C, D, E) = \mathcal{L}_{\subseteq}\left(\mathrm{Box}(C) \cap \mathrm{Box}(D), \mathrm{Box}(E)\right)$$
  - NF3: $C \sqsubseteq \exists r.D$ $$\mathcal{L}_3(C, r, D) = \frac{1}{2}\Big(\mathcal{L}_{\subseteq}(\mathrm{Box}(C) + \mathrm{Bump}(D), \mathrm{Head}(r)) + \mathcal{L}_{\subseteq}(\mathrm{Box}(D) + \mathrm{Bump}(C), \mathrm{Tail}(r))\Big).$$
  - NF4: $\exists r.C \sqsubseteq D$ $$\mathcal{L}_4(r, C, D) = \mathcal{L}_{\subseteq}(\mathrm{Head}(r) - \mathrm{Bump}(C), \mathrm{Box}(D))$$
  - NF5: $C \sqcap D \sqsubseteq \bot$ $$\mathcal{L}_5(C, D) = \|\max\{\mathbf{0}, -(\boldsymbol{d}(\mathrm{Box}(C), \mathrm{Box}(D)) + \gamma)\}\|$$

Ontology (Description Logic) Embedding

- Box^{2}EL training: Loess for axioms of each normal form (NF)

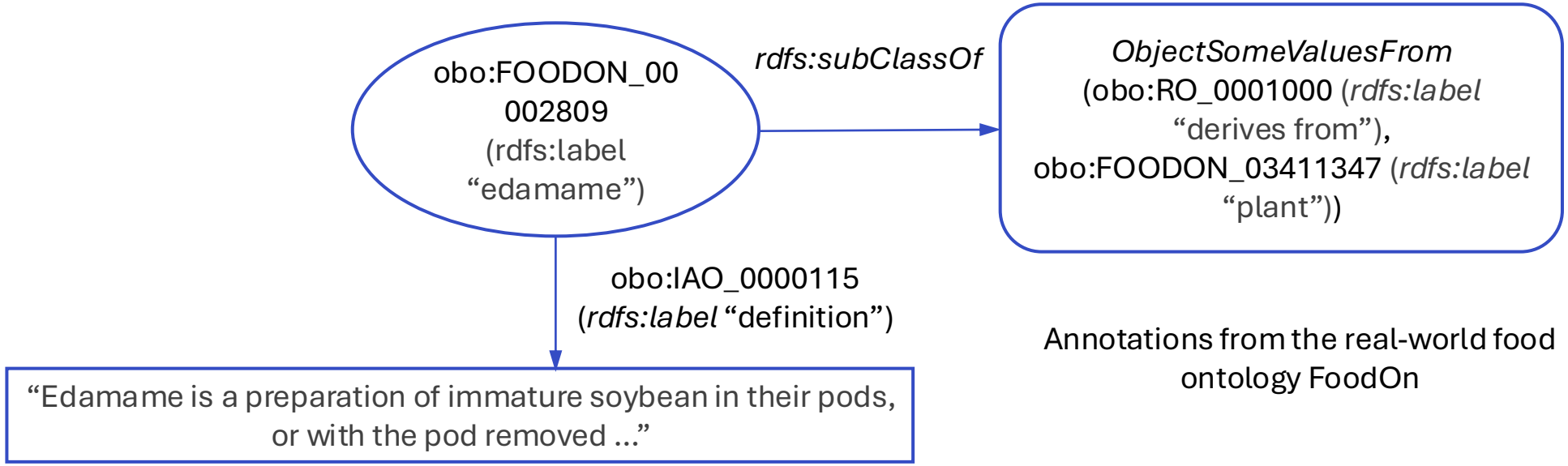
  - NF6: $r \sqsubseteq s$ $\quad \mathcal{L}_6(r,s) = \frac{1}{2}\left(\mathcal{L}_{\subseteq}(\text{Head}(r), \text{Head}(s)) + \mathcal{L}_{\subseteq}(\text{Tail}(r), \text{Tail}(s))\right)$

  - NF7: $r_1 \circ r_2 \sqsubseteq s$ $\quad \mathcal{L}_7(r_1, r_2, s) = \frac{1}{2}\left(\mathcal{L}_{\subseteq}(\text{Head}(r_1), \text{Head}(s)) + \mathcal{L}_{\subseteq}(\text{Tail}(r_2), \text{Tail}(s))\right)$

ABox can be transformed into TBox: $C(a) \rightsquigarrow \{a\} \sqsubseteq C$, $r(a,b) \rightsquigarrow \{a\} \sqsubseteq \exists r.\{b\}$

Text-aware Ontology Embedding

- OWL ontology includes more than formal semantics (e.g., labels, textual definitions)



- How to jointly embed the informal textual knowledge and the formally defined knowledge?

Text-aware Ontology Embedding

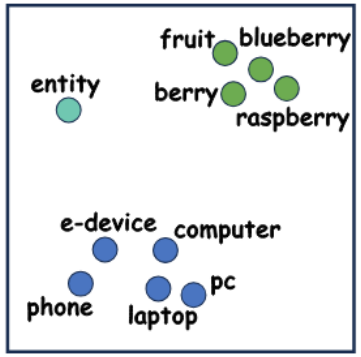
- Non-contextual: word2vec
  - **OWL2Vec***: Train a Word2Vec model from an OWL ontology, following RDF2Vec
  - Corpus (sentences) extraction via seriation (Manchester OWL Syntax), walking on the graph, OWL to RDF projection, etc.
- Contextual: Transformer-based language models
  - Common task-specific paradigm: pre-train then fine-tune
  - E.g., **BERTMap** which fine-tunes a BERT alike language model for ontology alignment

They both lose formally defined semantics!

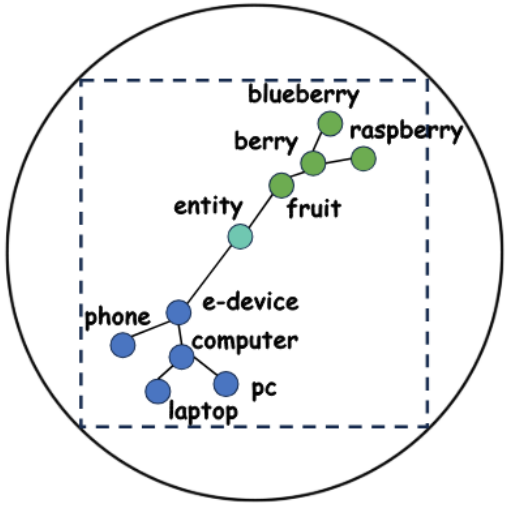
Chen, Jiaoyan, et al. "OWL2Vec*: embedding of OWL ontologies." *Machine Learning* 110.7 (2021): 1813-1845.
He, Yuan, et al. "BERTMap: a BERT-based ontology alignment system." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 36. No. 5. 2022.

Text-aware Ontology Embedding

- LM as hierarchy encoder (**HiT**)
  - Re-train a BERT alike LM by an ontology
  - Force the LM's concept encodings to a hierarchy in a hyperbolic space (Poincare ball)
    - Motived by its efficiency for representing hierarchies



Concept's text embedding in Euclidean Space by the last layer (tanh activation) of an LM, which is in a d-dimensional hyper-cube



Concept's text embedding by an ontology retrained LM, which is in a Poincare Ball of radius $\sqrt{d}$ that circumscribe the hyper cube

He, Yuan, et al. "Language models as hierarchy encoders." *Advances in Neural Information Processing Systems* 37 (2024): 14690-14711.

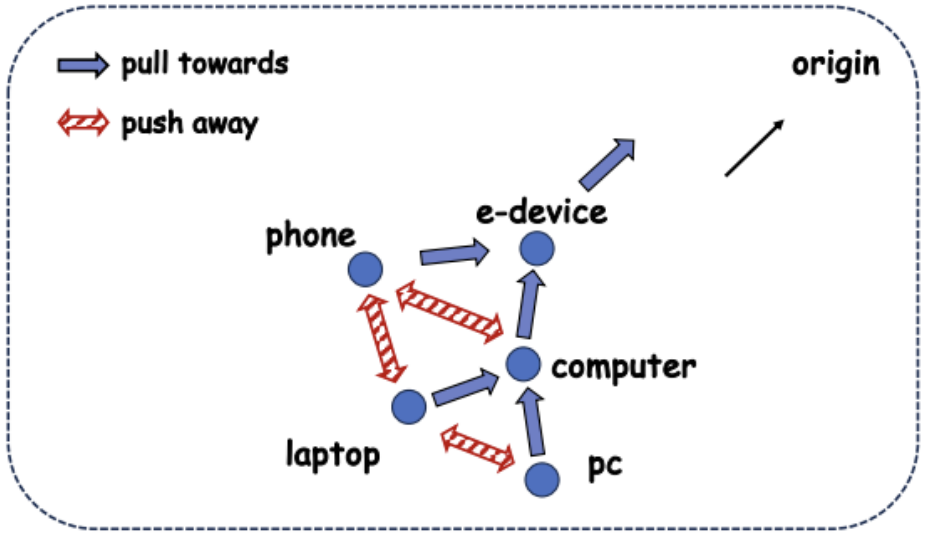
Text-aware Ontology Embedding

- Training of **HiT**
  - Hyperbolic clustering loss: clustering related concepts and distancing unrelated concepts (a concept is close to its positive parent and distanced from its negative parent)

$$\mathcal{L}_{cluster} = \sum_{(e,e^+,e^-)\in\mathcal{D}} \max(d_c(\mathbf{e},\mathbf{e}^+) - d_c(\mathbf{e},\mathbf{e}^-) + \alpha, 0)$$

  - Hyperbolic centripetal loss: make parent closer to origin

$$\mathcal{L}_{centri} = \sum_{(e,e^+,e^-)\in\mathcal{D}} \max(\|\mathbf{e}^+\|_c - \|\mathbf{e}\|_c + \beta, 0)$$

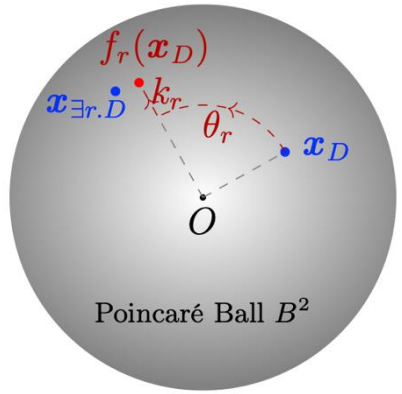


- Subsumption inference with HiT embeddings
  - Consider both contrastive and centripetal losses

$$s(e_1 \sqsubseteq e_2) = -(d_c(\mathbf{e}_1,\mathbf{e}_2) + \lambda(\|\mathbf{e}_2\|_c - \|\mathbf{e}_1\|_c))$$

Text-aware Ontology Embedding

- LM as $\mathcal{EL}^{++}$ ontology encoder (**OnT**)
  - Extend HiT to complex concepts E.g., $\exists isParentOf.Person$
- Solution #1: verbalization
  - $\exists isParentOf.Person$ → "something that is parent of some person"
  - Embedding notion: $\exists r.D$ → $\boldsymbol{x}_{\exists r.D}$
  - Does not represent roles' logics e.g., if $\mathrm{A} \sqsubseteq \mathrm{B}$ then $\exists r.\mathrm{A} \sqsubseteq \exists r.\mathrm{A}$
- Solution #2: Relation by rotation
  - A rotation function to represent a relation: $\exists r.D$ → $f_r(\boldsymbol{x}_D)$
  - Learning: $\boldsymbol{x}_{\exists r.D} \prec f_r(\boldsymbol{x}_D), f_r(\boldsymbol{x}_D) \prec \boldsymbol{x}_{\exists r.D}$



Yang, Hui, et al. "Language Models as Ontology Encoders." *The 24th International Semantic Web Conference (ISWC)*. 2025.

Summary

- Geometric models
  - **Box^{2}EL** (concept as a box and a bump vector, relation as two boxes)
- Language models
  - Non-contextual (OWL2Vec*)
  - Transformer-based
    - **BERTMap**: fine-tune a pre-trained encoder model with an additional classifier for concept equivalence matching
    - **HiT**: Continuously train a pre-trained encoder model for representing hierarchical concepts in a Poincare ball
    - **OnT**: Extend HiT for conserving relationship between concepts

Discussion on Application

- Link prediction in knowledge intensive domains
  - E.g., **protein-protein interaction prediction** with the Gene Ontology as evaluated in the Box^2EL paper
- Ontology construction and curation
  - **Subsumption completion**, including named concepts and complex concepts in either formal representation or natural language (evaluated in Box^2EL, HiT and OnT papers)
  - **Subsumption matching and new concept placement**; can be combined with LLM and agentic AI, where such embeddings can efficiently get candidates and reduce space space via indexing
- Hierarchical retrieval
  - E.g., most queries to SNOMED CT are **out-of-vocabulary**, which require to return most close concepts with a subsumption relationship instead of equivalence

- Acknowledgement
  - Key contributors: Yuan He (Oxford, now in Amazon), Hui Yang (Manchester), Mathias Jackermeier (Oxford), Ian Horrocks (Oxford)
  - Main funders: EPSRC (OntoEm & ConCur), Samsung Research UK

- Feel free to contact me
  - jiaoyan.chen@manchester.ac.uk

- Q&A