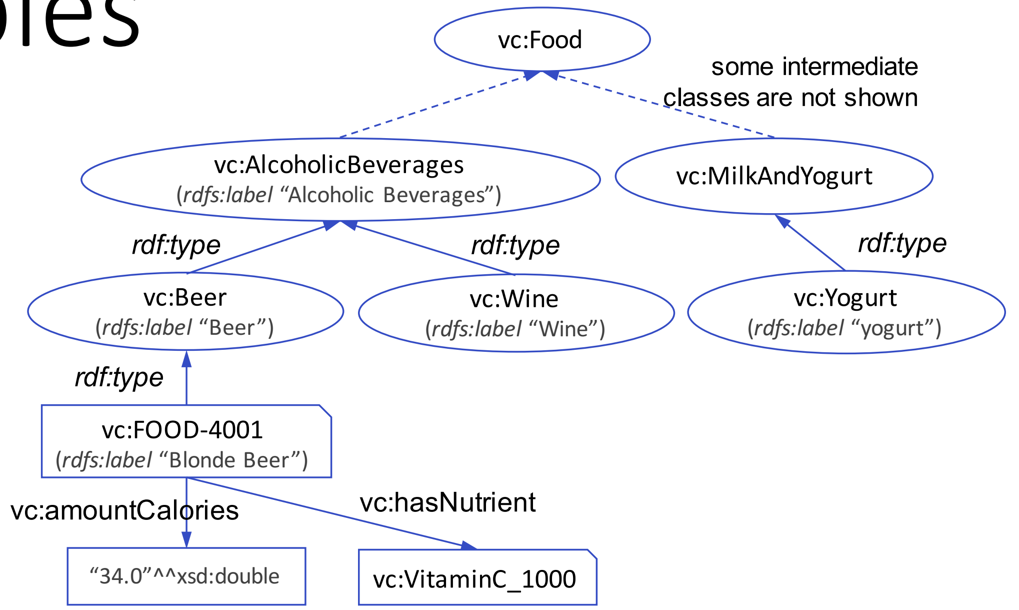
OWL2Vec*: Embedding of OWL Ontologies

Jiaoyan Chen

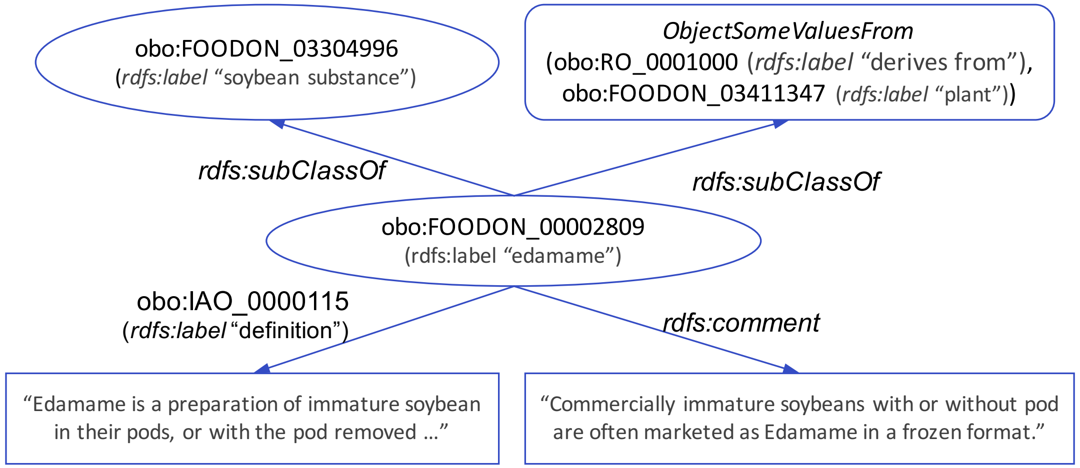
Department of Computer Science, University of Oxford

OWL Ontology and Examples

- TBox:
  - Atomic concepts (classes)
  - Atomic roles (object properties)
  - Complex classes and properties
    - by logical constructors (description logic)
  - Constraints, subsumption relationship, etc.
- ABox:
  - Individuals (instances)
  - Membership assertions, role assertions (facts)
- OWL/RDF/RDFS Built-in and bespoke properties
  - E.g., rdfs:subClassOf, vc:hasNutrient
- Others:
  - Meta data by annotation properties: label, synonym, class definition, comment, etc.
  - Entity, URI (prefix + name/id), Literal



The healthy lifestyle ontology named HeLis



The food ontology named FoodOn

KG/Ontology Embedding

- End-to-end paradigm
  - Score function, loss
  - Geometric learning (e.g., translation-based), decomposition, GNN, etc.

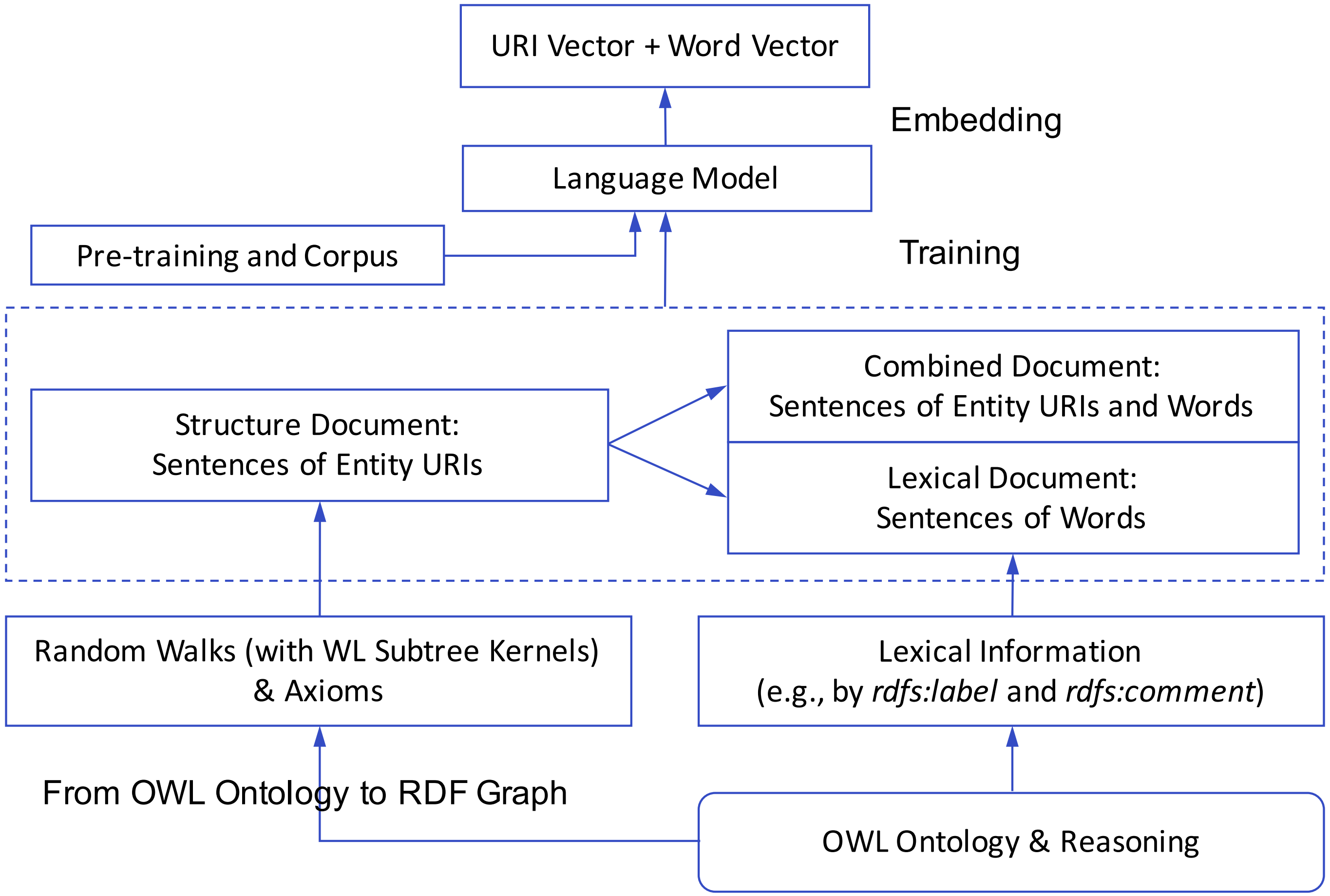
- Pipeline paradigm
  - Transformed into sequences (document) with semantics “kept”
  - Train neural language models
  - E.g., node2vec, RDF2Vec

Semantics Embedded

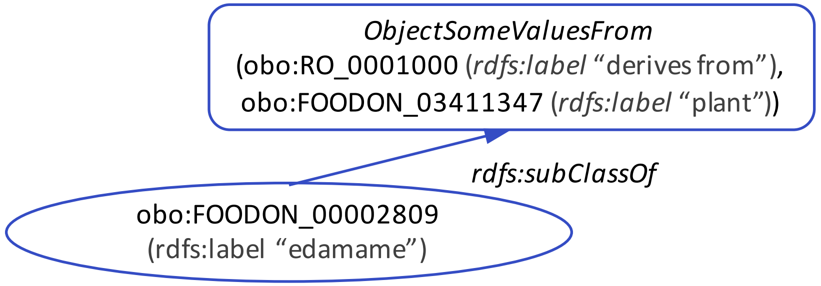
	Graph Structure	Axioms	(Text) Literals	Logical Constructors
RDF2Vec, TransE, HAKE, etc.	Yes	No	Possible	No
Onto2Vec	No	Yes	No	No
Opa2Vec	No	Yes	Yes	No
EL Embedding	Partial	No	No	Partial ($\mathcal{EL}^{++}$)
Quantum Embedding	Partial	No	No	Partial ($\mathcal{ALC}$)
OWL2Vec	Yes	No	No	Yes
OWL2Vec*	Yes	Yes	Yes	Yes

- Ontological schema + large scale facts (KG) e.g., DBpedia
  - E.g., JOIE (KDD'19)
  - Rule injection in embedding

OWL2Vec*: A Language Model based Framework



1. From OWL Ontology to RDF Graph
    - Reasoning by E.g. HermiT

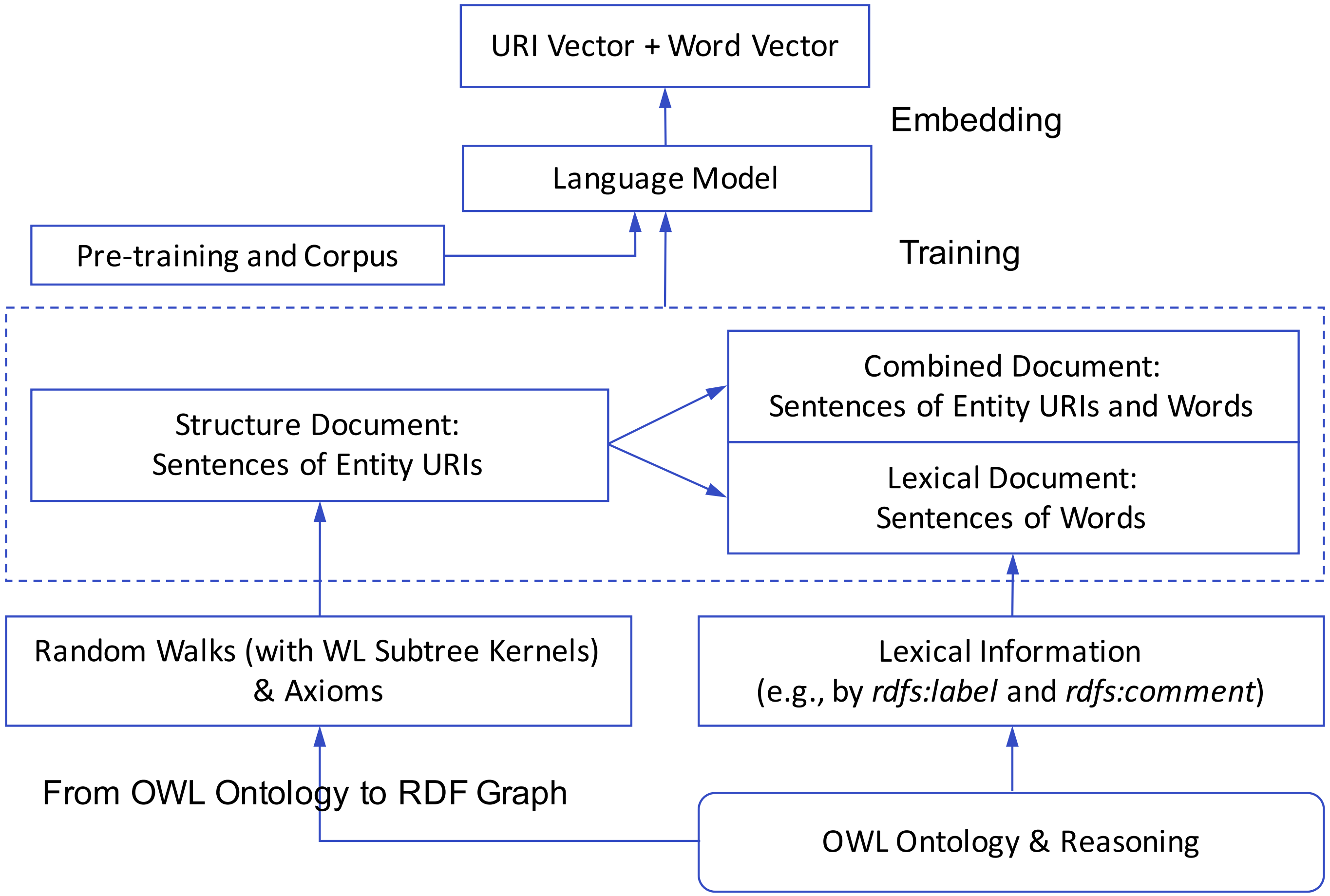


Solution #1: W3C OWL to RDF Graph Mapping
e.g.,
<obo:FOODON_00002809, rdfs:subClassOf, _:x>
<_:x, rdf:type, owl:Restriction>
<_:x, owl:OnProperty, obo:RO_0001000>
<_:x, owl:SomeValueFrom, obo:FOODON_03411347>

Solution #2: Projection Rules
e.g., <obo:FOODON_00002809, rdfs:subClassOf, obo:FOODON_03411347

Axiom of Condition 1	Axiom or Triple(s) of Condition 2	Projected Triple(s)
$A \sqsubseteq \square r.D$ or $\square r.D \sqsubseteq A$	$D \equiv B \mid B_1 \sqcup \ldots \sqcup B_n \mid B_1 \sqcap \ldots \sqcap B_n$	$\langle A, r, B \rangle$ or $\langle A, r, B_i \rangle$ for $i \in 1, \ldots, n$
$\exists r.\top \sqsubseteq A$ (domain)	$\top \sqsubseteq \forall r.B$ (range)	
$A \sqsubseteq \exists r.\{b\}$	$B(b)$	
$r \sqsubseteq r'$	$\langle A, r', B \rangle$ has been projected	
$r' \equiv r^-$	$\langle B, r', A \rangle$ has been projected	
$s_1 \circ \ldots \circ s_n \sqsubseteq r$	$\langle A, s_1, C_1 \rangle \ldots \langle C_n, s_n, B \rangle$ have been projected	
$B \sqsubseteq A$	–	$\langle B, rdfs{:}subClassOf, A \rangle$ $\langle A, rdfs{:}subClassOf^-, B \rangle$
$A(a)$	–	$\langle a, rdf{:}type, A \rangle$ $\langle A, rdf{:}type^-, a \rangle$
$r(a, b)$	–	$\langle a, r, b \rangle$

OWL2Vec*: A Language Model based Framework



1. From OWL Ontology to RDF Graph
2. Structure Document -- Sequences of entities (URIs)
   - Random walk, Weisfeiler-Lehman (WL) subtree kernel
   - Axioms (OWL Manchester Syntax)
3. Lexical Document -- Sequences of words
   - Text from annotation properties
   - From structure document
4. Combined Document -- Sequences of words and URIs
   - Replace partial entities in every entity sequence by their words
5. Language Model (CBOW architecture)
   - Pre-training
   - URI vector vs average word vector

Case Studies

- Ontology completion
  - Class membership and subsumption prediction
    - Input: the OWL2Vec* embeddings of two entities (as learned features)
    - Classifiers e.g., Random Forest and Logistic Regression
      - Learn from the known axioms
    - Output: a score in [0, 1]

- Others
  - Clustering, alignment, neural-symbolic AI, domain applications, etc.

Evaluation

- Experiment Settings
  - Randomly split the explicitly declared membership/subsumption axioms into train (70%), valid (10%), and test (20%)
  - Remove the valid and test axioms from the ontology
  - Given the head entity, rank all the candidate tail entities, and calculate Mean Reciprocal Rank (MRR), Hits@K

	DL Expressivity	Instances #	Classes #	Axioms #	Membership #	Subsumption #
HeLis	$\mathcal{ALCHIQ(D)}$	20, 318	277	172, 213	20, 318	261
FoodOn	$\mathcal{SRIQ}$	359	28, 182	241, 581	0	29, 778
GO	$\mathcal{SRI}$	0	44, 244	513, 306	0	72, 601

Evaluated ontologies and their statistics

To reproduce: https://github.com/KRR-Oxford/OWL2Vec-Star

Overall Results (Comparison with Baselines)

Method	HeLis			
	MRR	Hits@1	Hits@5	Hits@10
RDF2Vec	0.345	0.219	0.460	0.655
TransE	0.181	0.09	0.232	0.355
TransR	0.298	0.184	0.391	0.559
DistMult	0.253	0.166	0.304	0.437
Quantum Embeding	0.159	0.132	0.163	0.190
Onto2Vec	0.211	0.108	0.268	0.397
OPA2Vec	0.237	0.146	0.286	0.408
OWL2Vec	0.335	0.215	0.397	0.601
Pre-trained *Word2Vec*	0.899	0.877	0.923	0.933
OWL2Vec*	**0.953**	**0.932**	**0.978**	**0.987**

(a) Membership Prediction

Method	FoodOn				GO			
	MRR	Hits@1	Hits@5	Hits@10	MRR	Hits@1	Hits@5	Hits@10
RDF2Vec	0.078	0.053	0.097	0.119	0.043	0.017	0.057	0.087
TransE	0.029	0.011	0.044	0.065	0.015	0.005	0.018	0.030
TransR	0.072	0.044	0.093	0.130	0.048	0.016	0.076	0.113
DistMult	0.076	0.045	0.099	0.134	0.046	0.018	0.68	0.097
EL Embeding	0.040	0.014	0.067	0.099	0.018	0.005	0.021	0.036
Onto2Vec	0.034	0.014	0.047	0.064	0.016	0.004	0.021	0.036
OPA2Vec	0.093	0.058	0.117	0.156	0.075	0.032	0.106	0.157
OWL2Vec	0.091	0.052	0.121	0.152	0.031	0.012	0.040	0.067
Pre-trained *Word2Vec*	0.136	0.089	0.175	0.227	0.123	0.055	0.177	0.260
OWL2Vec*	**0.213**	**0.143**	**0.287**	**0.357**	**0.170**	**0.076**	**0.258**	**0.376**

(b) Subsumption Prediction

- Settings for OWL2Vec*:
  - no reasoning
  - W3C OWL to RDF Graph Mapping
  - other settings are set via the valid set and will be discussed next.
- Quantum Embedding, EL Embedding, Onto2Vec, TransE **<** RDF2Vec, TransR, DistMult, OWL2Vec, OPA2Vec **<** Pre-trained Word2vec **< OWL2Vec***

Ablation Study

Setting	HeLis			
	MRR	Hits@1	Hits@5	Hits@10
$D_s + V_{uri}$	0.353	0.226	0.470	0.668
$D_{s,l} + V_{uri}$	0.448	0.295	0.623	0.814
$D_{s,l} + V_{word}$	0.938	0.920	0.961	0.974
$D_{s,l} + V_{uri,word}$	0.952	**0.934**	0.974	0.984
$D_{s,l,rc} + V_{uri}$	0.446	0.299	0.618	0.799
$D_{s,l,rc} + V_{word}$	0.945	0.926	0.970	0.979
$D_{s,l,rc} + V_{uri,word}$	0.951	0.932	**0.975**	**0.987**
$D_{s,l,tc} + V_{uri}$	0.505	0.355	0.695	0.854
$D_{s,l,tc} + V_{word}$	0.943	0.923	0.969	0.976
$D_{s,l,tc} + V_{uri,word}$	**0.953**	0.932	**0.975**	**0.987**

(a) Membership Prediction

Setting	FoodOn				GO			
	MRR	Hits@1	Hits@5	Hits@10	MRR	Hits@1	Hits@5	Hits@10
$D_s + V_{uri}$	0.083	0.047	0.116	0.150	0.041	0.017	0.055	0.084
$D_{s,l} + V_{uri}$	0.146	0.095	0.188	0.247	0.098	0.040	0.146	0.206
$D_{s,l} + V_{word}$	**0.213**	**0.143**	**0.287**	**0.357**	**0.170**	**0.076**	**0.258**	**0.376**
$D_{s,l} + V_{uri,word}$	0.175	0.113	0.235	0.298	0.137	0.068	0.204	0.319
$D_{s,l,rc} + V_{uri}$	0.174	0.115	0.230	0.295	0.103	0.045	0.149	0.219
$D_{s,l,rc} + V_{word}$	0.195	0.126	0.265	0.338	0.155	0.066	0.237	0.348
$D_{s,l,rc} + V_{uri,word}$	0.195	0.132	0.254	0.316	0.139	0.060	0.206	0.297
$D_{s,l,tc} + V_{uri}$	0.147	0.086	0.199	0.270	0.097	0.042	0.139	0.202
$D_{s,l,tc} + V_{word}$	0.171	0.102	0.240	0.303	0.150	0.061	0.230	0.343
$D_{s,l,tc} + V_{uri,word}$	0.165	0.097	0.231	0.305	0.139	0.061	0.204	0.300

(b) Subsumption Prediction

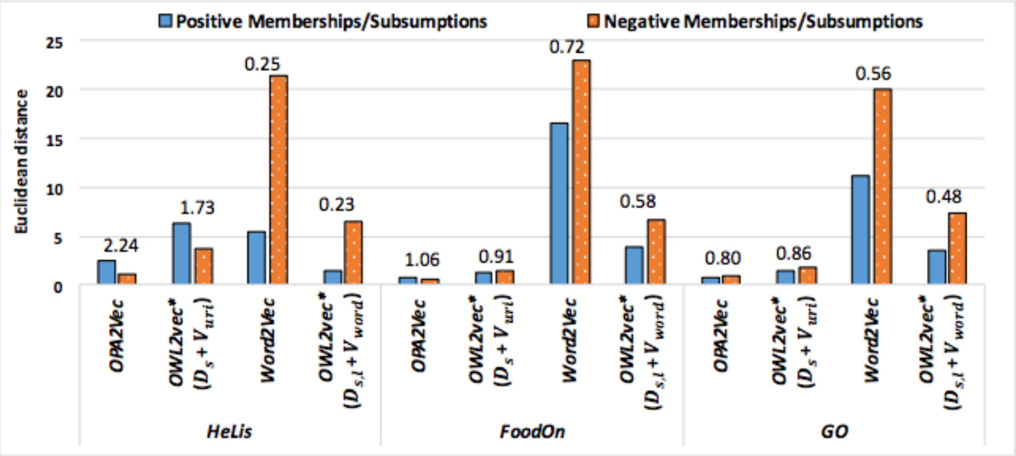
- Evaluation on documents, URI vector and word vector of **OWL2Vec***
- HeLis:
  - All three documents ≈ structural + literal documents
  - URI vector + word vector > word vector alone >> URI vector alone
- FoodOn and GO
  - Structural document + literal document + word vector is the best

Ablation Study

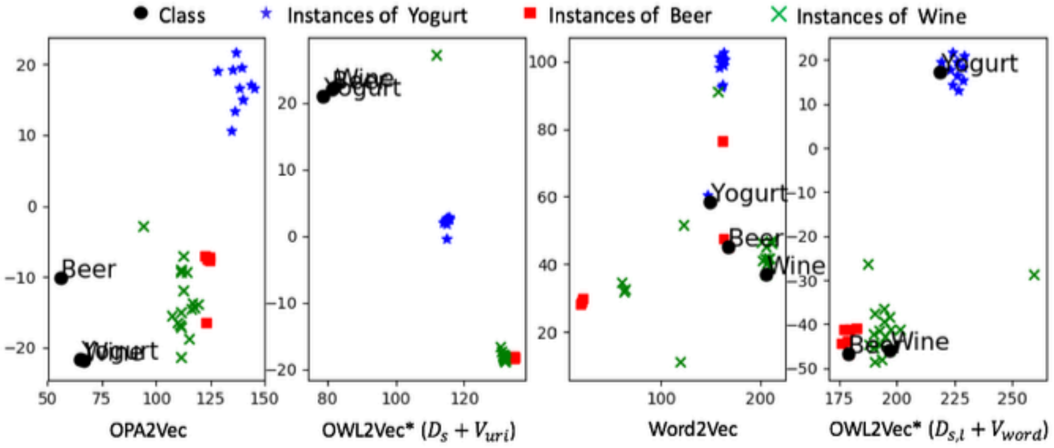
- Structure Document
  - **OWL to RDF Mapping** vs **Project Rules**
    - Varies from different walking strategies and walking depths
    - **The former has a bit higher optimum performance**
  - Walking strategies
    - Using WL subtree kernel achieves a bit higher performance, and needs a smaller depth
  - Reasoning in advance
    - Very limited impact
- Pre-training
  - Negative impact on membership/subsumption prediction
  - Would have positive impact on other case studies

Euclidean distance & visualization

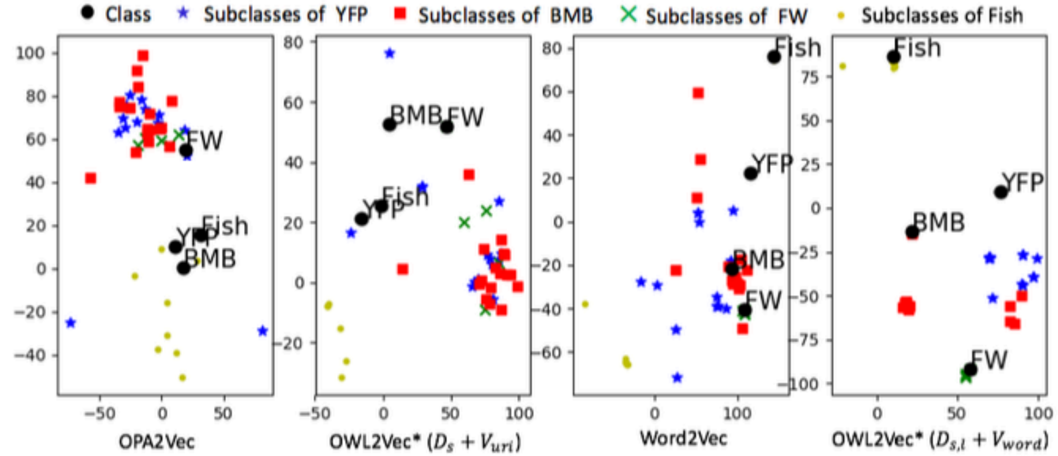
The average Euclidean distance between the head entity and the tail entity, of positive axioms and negative axioms



The more discriminative Euclidean distance, the better performance



(a) Three classes of HeLis – "Yogurt", "Beer" and "Wine", and their instances (10, 6 and 17 respectively).



(b) Four classes of FoodOn – "Yogurt Food Product" (YFP), "Barley Malt Beverage" (BMB), "Fruit Wine" (FW) and "Fish", and their subclasses (14, 17, 5 and 8 respectively).

Conclusion and Outlook

- A language model based embedding framework for OWL ontologies
  - Graph structure
  - Textual literals
  - Axioms
  - Logical constructors

- Future work
  - Software (standalone application, compatible for different components)
  - Sub-symbolic reasoning
  - Ontology + large scale facts (KG)
  - Applications

Application: Ontology Quality Assurance

- Alignment
  - Food ontologies, biomedical ontologies
  - How to deal with two embedding spaces?
  - How to map entities with none equivalent relationships?
    - e.g., Chemical effects [Myklebust et al. ISWC’19]
- Error detection
  - E.g., Wrong subsumption
- Completion

Application: Neural-Symbolic AI

- Semantic composition for zero-shot learning
  - Using description logic to describe those “unseen classes” ([Chen et al. KR 2020])

- Domain modeling for transfer learning
  - E.g., explain or augment transfer learning by domain knowledge ([Chen et al. KR 2018])

Ontology embedding to bridge the symbolic KR and sub-symbolic ML!

Thanks for your attentions!

Contact: jiaoyan.chen@cs.ox.ac.uk