Integrating Ontologies and Vector Space Embeddings Using Conceptual Spaces

Zied Bouraoui[1], Víctor Gutiérrez-Basulto[2], Steven Schockaert[2]

[1] CRIL Laboratory
Université d'Artois, France

[2] School of Computer Science & Informatics
Cardiff University, Cardiff, UK

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Knowledge representation using ontologies

Ontology

$$\begin{aligned}
\text{expertInAI}(X) &\leftarrow \text{authorOf}(X, Y), \text{hasTopic}(Y, \text{artificialIntelligence}) \\
\text{hasTopic}(X, \text{artificialIntelligence}) &\leftarrow \text{hasTopic}(X, \text{knowledgeRepresentation}) \\
\text{hasTopic}(X, \text{artificialIntelligence}) &\leftarrow \text{hasTopic}(X, \text{machineLearning}) \\
\text{hasTopic}(X, \text{artificialIntelligence}) &\leftarrow \text{hasTopic}(X, \text{multiAgentSystems}) \\
\text{hasTopic}(X, \text{artificialIntelligence}) &\leftarrow \text{hasTopic}(X, \text{naturalLanguageProcessing})
\end{aligned}$$

Facts

$$\{\text{authorOf}(\text{bob}, p), \text{hasTopic}(p, \text{knowledgeRepresentation})\}$$

Knowledge representation using ontologies

Ontology

$$\text{expertInAI}(X) \leftarrow \text{authorOf}(X, Y), \text{hasTopic}(Y, \text{artificialIntelligence})$$
$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{knowledgeRepresentation})$$
$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{machineLearning})$$
$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{multiAgentSystems})$$
$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{naturalLanguageProcessing})$$

Facts

$$\{\text{authorOf}(\text{bob}, p), \text{hasTopic}(p, \text{knowledgeRepresentation})\}$$

Consequences

$$\{\text{hasTopic}(p, \text{artificialIntelligence}), \text{expertInAI}(\text{bob})\}$$

Knowledge representation using ontologies

Ontology

$$\text{expertInAI}(X) \leftarrow \text{authorOf}(X, Y), \text{hasTopic}(Y, \text{artificialIntelligence})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{knowledgeRepresentation})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{machineLearning})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{multiAgentSystems})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{naturalLanguageProcessing})$$

Facts

$$\{\text{authorOf}(\text{bob}, p), \text{hasTopic}(p, \text{knowledgeRepresentation})\}$$

Consequences

$$\{\text{hasTopic}(p, \text{artificialIntelligence}), \text{expertInAI}(\text{bob})\}$$

Knowledge representation using ontologies

Ontology

$$\text{expertInAI}(X) \leftarrow \text{authorOf}(X, Y), \text{hasTopic}(Y, \text{artificialIntelligence})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{knowledgeRepresentation})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{machineLearning})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{multiAgentSystems})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{naturalLanguageProcessing})$$

Facts

$$\{\text{authorOf}(\text{bob}, p), \text{hasTopic}(p, \text{knowledgeRepresentation})\}$$

Consequences

$$\{\text{hasTopic}(p, \text{artificialIntelligence}), \text{expertInAI}(\text{bob})\}$$

Why do we need vector space embeddings?

Ontology

$$\text{expertInAI}(X) \leftarrow \text{authorOf}(X, Y), \text{hasTopic}(Y, \text{artificialIntelligence})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{knowledgeRepresentation})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{machineLearning})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{multiAgentSystems})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{naturalLanguageProcessing})$$

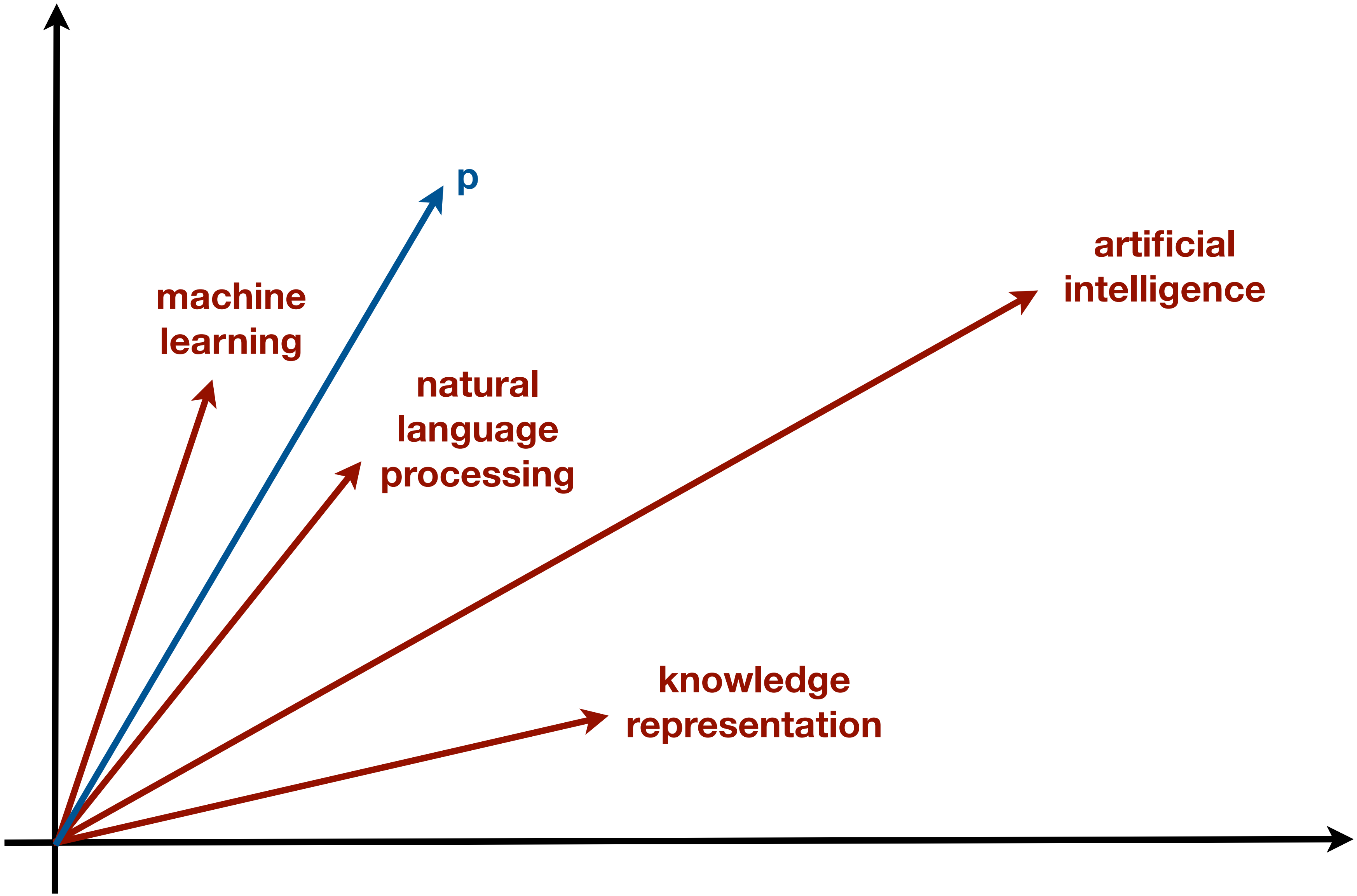
Facts

$$\{\text{authorOf}(\text{alice}, q), \text{hasTopic}(q, \text{planning})\}$$

Consequences

???

Why do we need vector space embeddings?



Neural theorem proving

Ontology

$$\text{expertInAI}(X) \leftarrow \text{authorOf}(X, Y), \text{hasTopic}(Y, \text{artificialIntelligence})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{knowledgeRepresentation})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{machineLearning})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{multiAgentSystems})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{naturalLanguageProcessing})$$

Facts

$$\{\text{authorOf}(\text{alice}, q), \text{hasTopic}(q, \text{planning})\}$$

Tim Rocktäschel, Sebastian Riedel: End-to-end Differentiable Proving. NIPS 2017: 3788-3800

Neural theorem proving

Ontology

$$\text{expertInAI}(X) \leftarrow \text{authorOf}(X, Y), \text{hasTopic}(Y, \text{artificialIntelligence})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{knowledgeRepresentation})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{machineLearning})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{multiAgentSystems})$$

$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{naturalLanguageProcessing})$$

Facts

$$\{\text{authorOf}(\text{alice}, q), \text{hasTopic}(q, \text{planning})\}$$

Consequences

$$\text{hasTopic}(p, \text{artificialIntelligence}) \quad : \quad \underbrace{\phi(\text{planning}) \cdot \phi(\text{knowledgeRepresentation})}_{\text{proof strength}}$$

Tim Rocktäschel, Sebastian Riedel: End-to-end Differentiable Proving. NIPS 2017: 3788-3800

Neural theorem proving

Ontology

$$\text{expertInAI}(X) \leftarrow \text{authorOf}(X, Y), \text{hasTopic}(Y, \text{artificialIntelligence})$$
$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{knowledgeRepresentation})$$
$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{machineLearning})$$
$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{multiAgentSystems})$$
$$\text{hasTopic}(X, \text{artificialIntelligence}) \leftarrow \text{hasTopic}(X, \text{naturalLanguageProcessing})$$

Facts

$$\{\text{authorOf}(\text{alice}, q), \text{hasTopic}(q, \text{planning})\}$$

Consequences

embedding of the term "knowledge representation"

$$\text{hasTopic}(p, \text{artificialIntelligence}) \quad : \quad \phi(\text{planning}) \cdot \phi(\text{knowledgeRepresentation})$$

Tim Rocktäschel, Sebastian Riedel: End-to-end Differentiable Proving. NIPS 2017: 3788-3800

Key issues

Where do the embeddings come from?

- Learned from the knowledge base itself (e.g. knowledge graph completion, neural theorem proving)
- Learned from text (e.g. word embeddings)
- But are these the right vectors for plausible reasoning?

Key issues

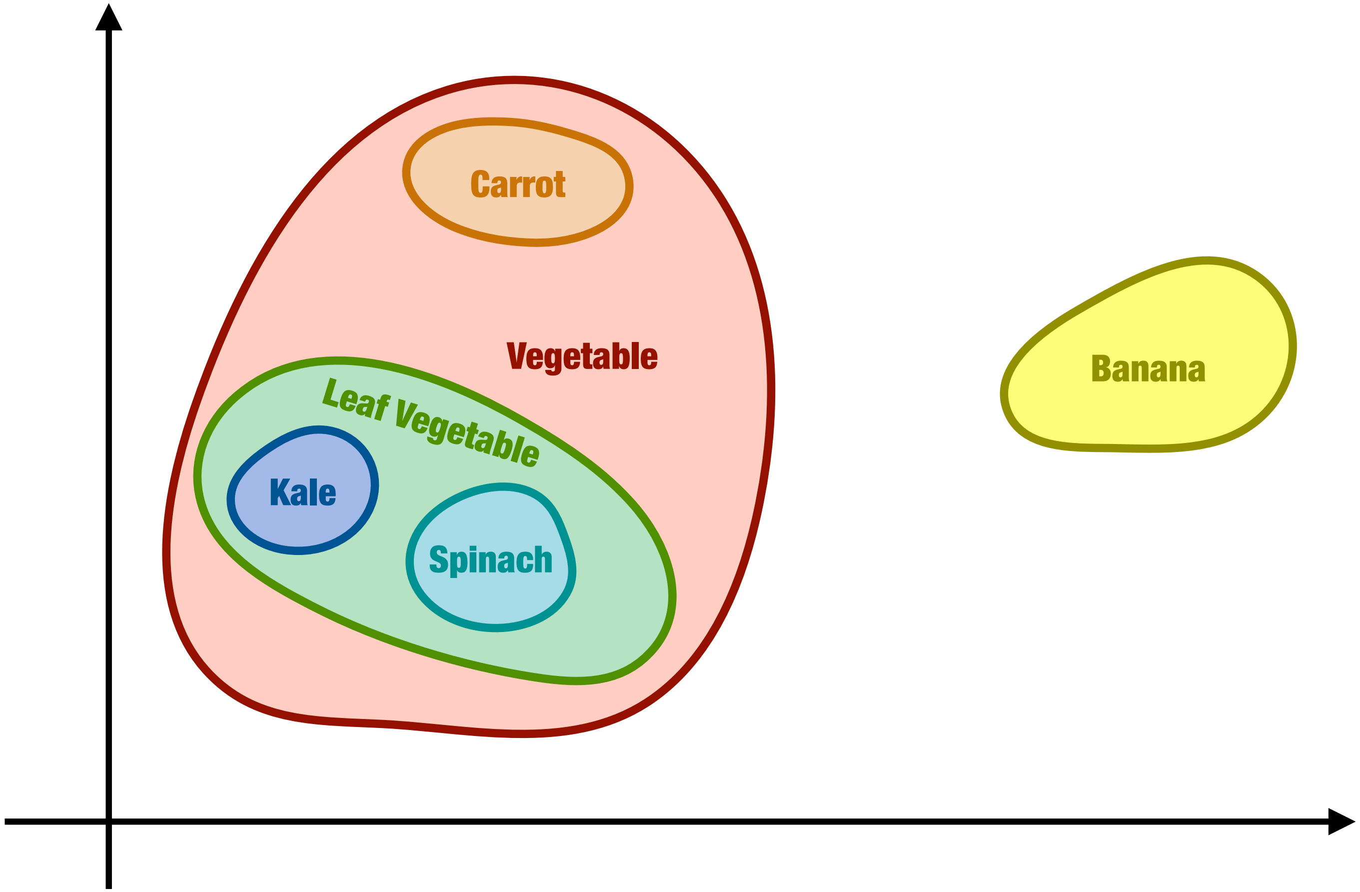
Where do the embeddings come from?

- Learned from the knowledge base itself (e.g. knowledge graph completion, neural theorem proving)
- Learned from text (e.g. word embeddings)
- But are these the right vectors for plausible reasoning?

What is the underlying principle?

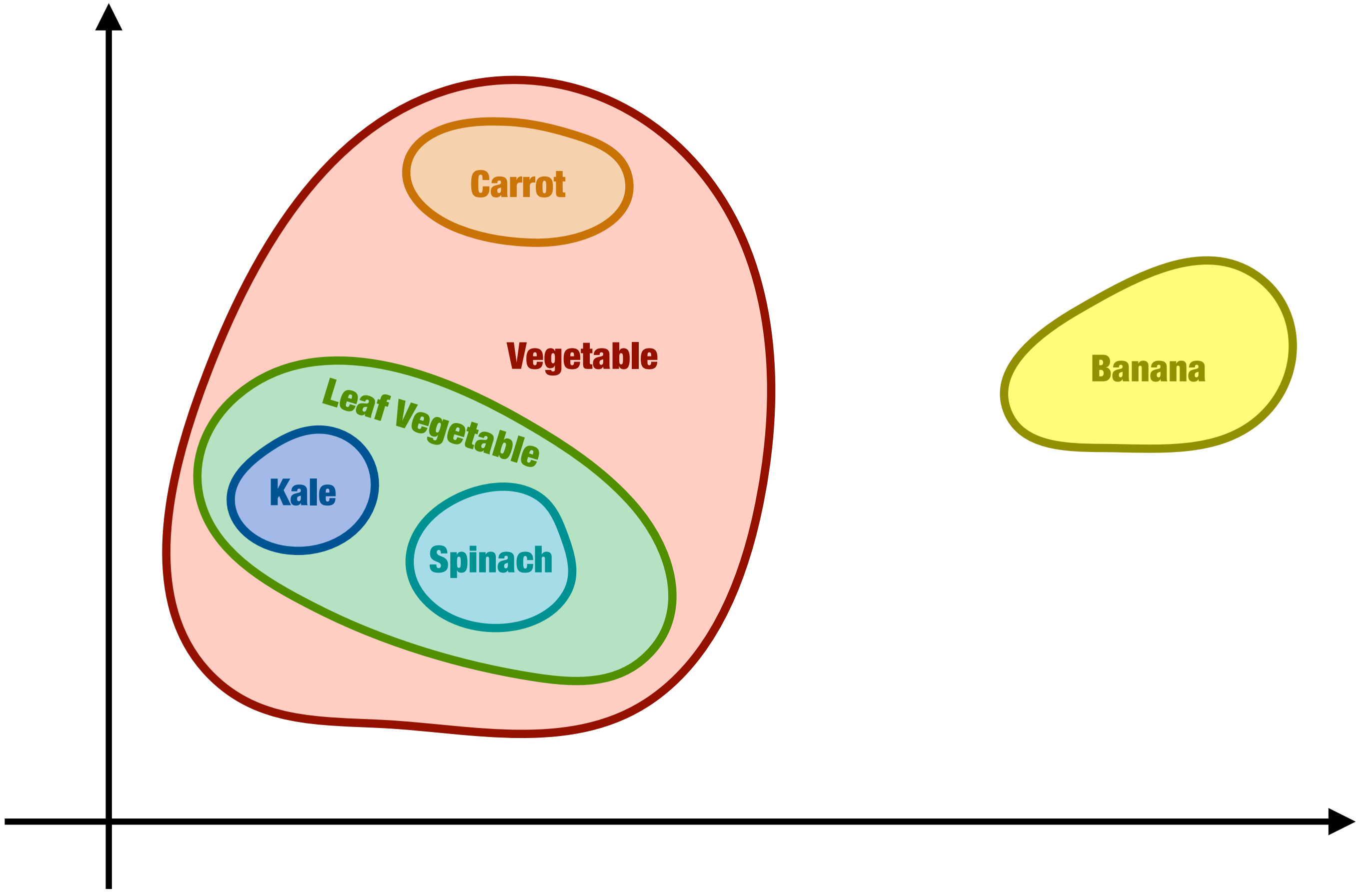
- Similarity-based reasoning is highly heuristic. No strong reason to believe that something is true just because it is true for a similar predicate or individual
- Is there a way to use embeddings to derive plausible consequences even if we don't have rules capturing "similar" situations?
- Can we find a single framework in which both rules and embeddings can be expressed?
- Can we formulate a model-theoretic semantics for inference methods that incorporate embeddings (e.g. to deal with inconsistency)?

Conceptual spaces



Peter Gärdenfors: Conceptual spaces - the geometry of thought. MIT Press 2000

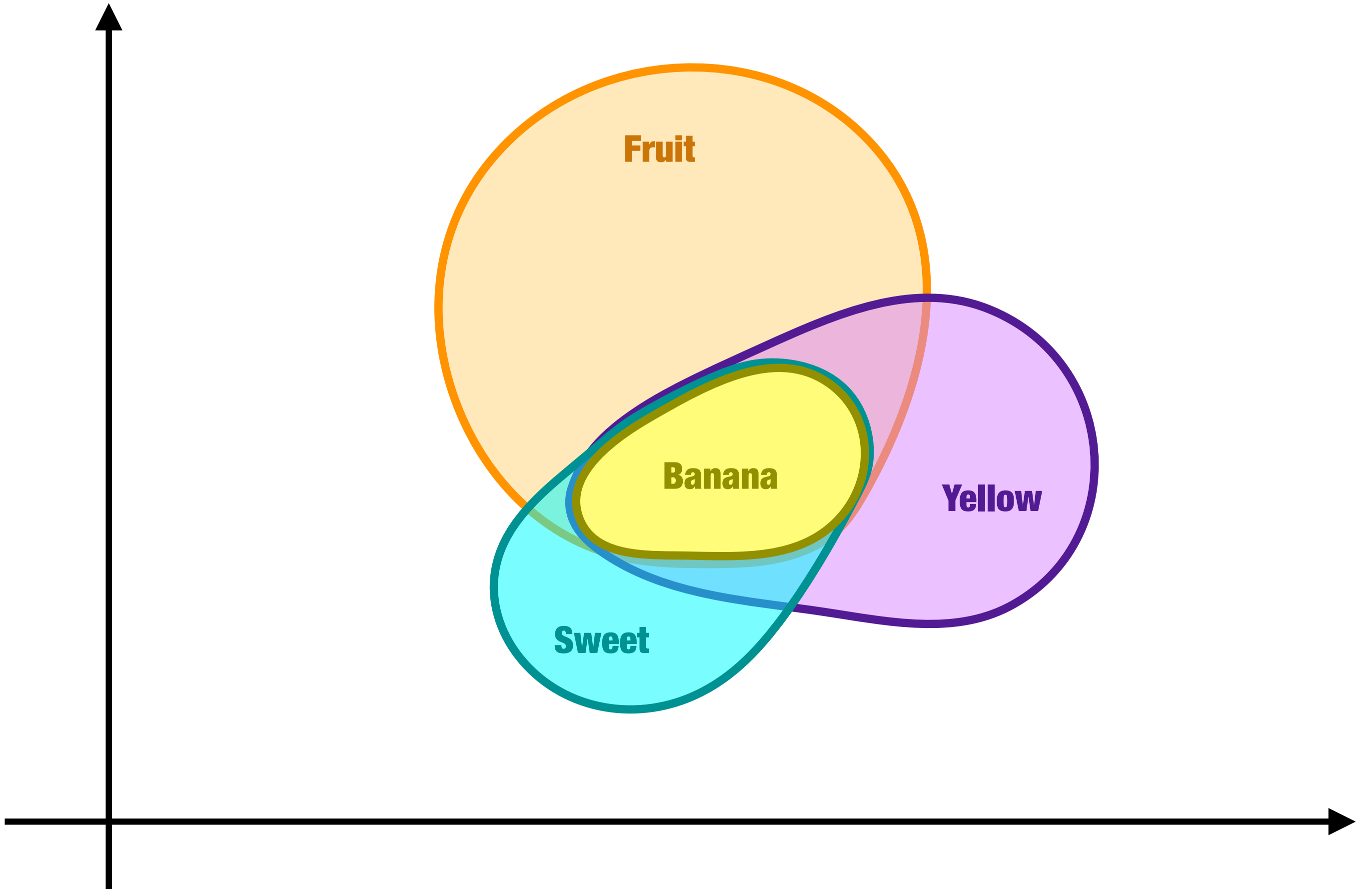
Conceptual spaces



$$\text{LeafVegetable}(X) \leftarrow \text{Spinach}(X)$$
$$\text{Vegetable}(X) \leftarrow \text{LeafVegetable}(X)$$
$$\bot \leftarrow \text{Banana}(X), \text{Vegetable}(X)$$

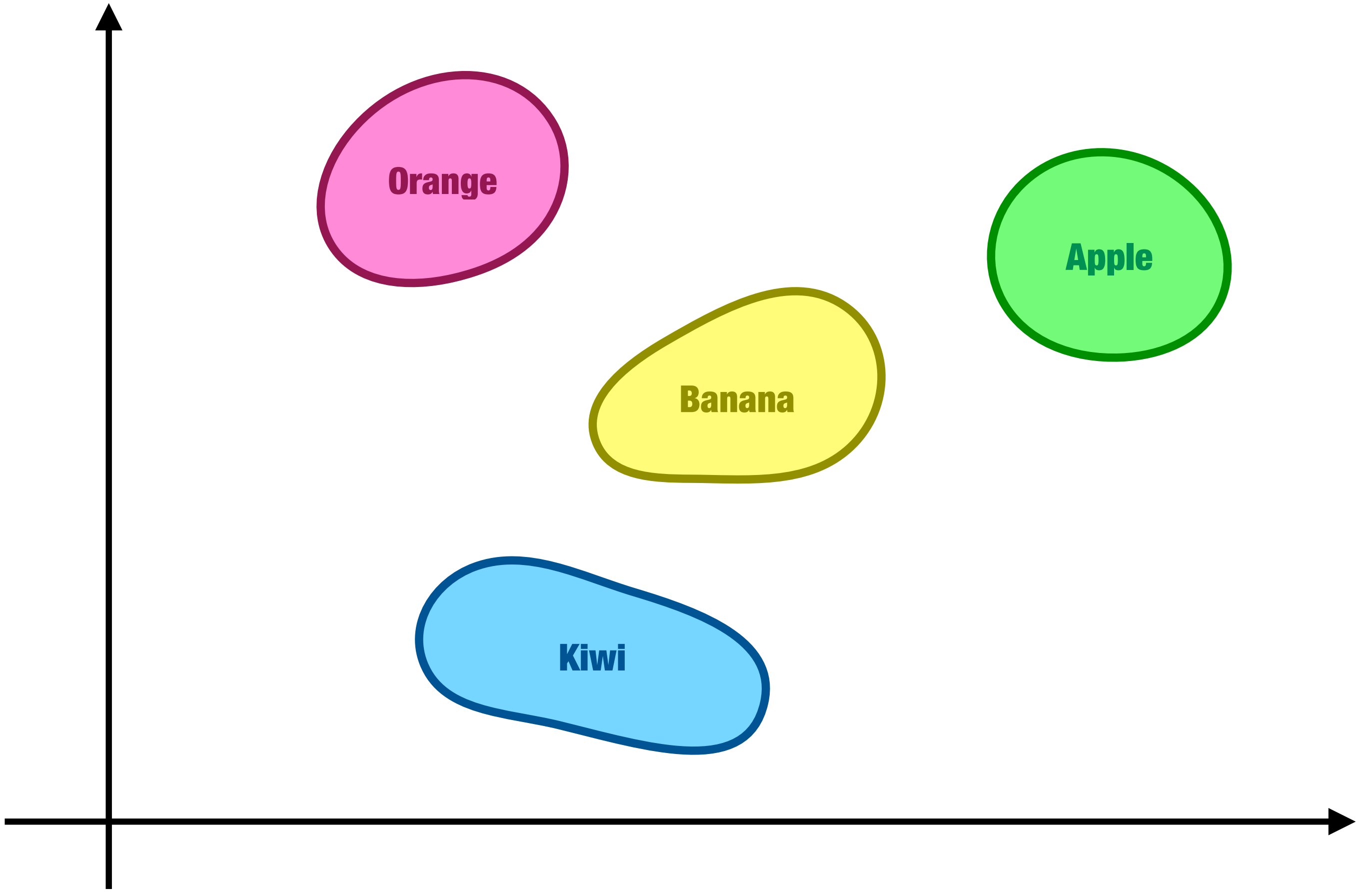
Peter Gärdenfors: Conceptual spaces - the geometry of thought. MIT Press 2000

Conceptual spaces



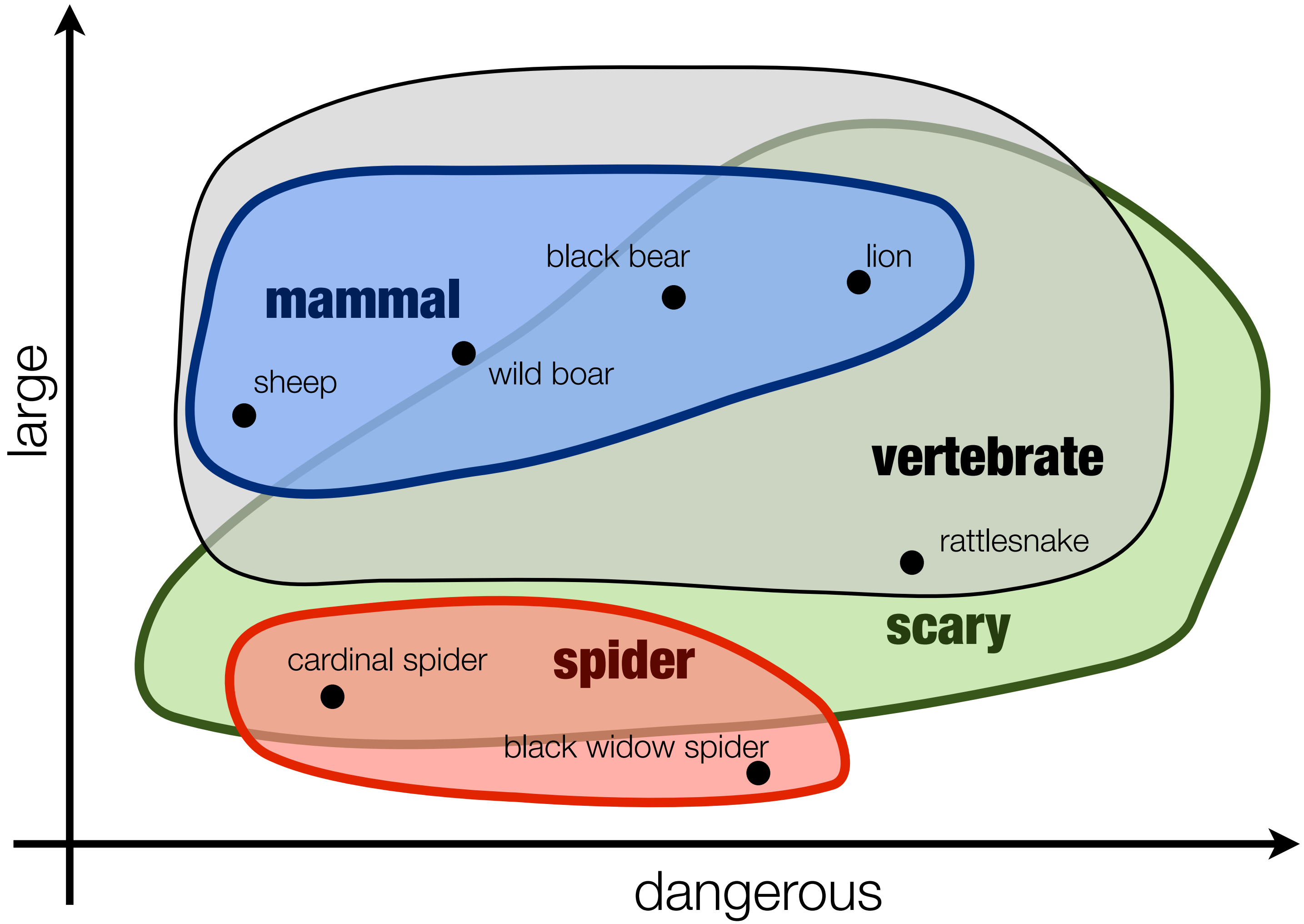
$$\text{Banana}(X) \leftarrow \text{Fruit}(X), \text{Yellow}(X), \text{Sweet}(X)$$

Conceptual spaces



Banana is **between** Orange, Apple and Kiwi

Conceptual spaces: quality dimensions



"Sufficiently large spiders are always scary"

Conceptual spaces: domains

hue
saturation
intensity

colour domain

size

size domain

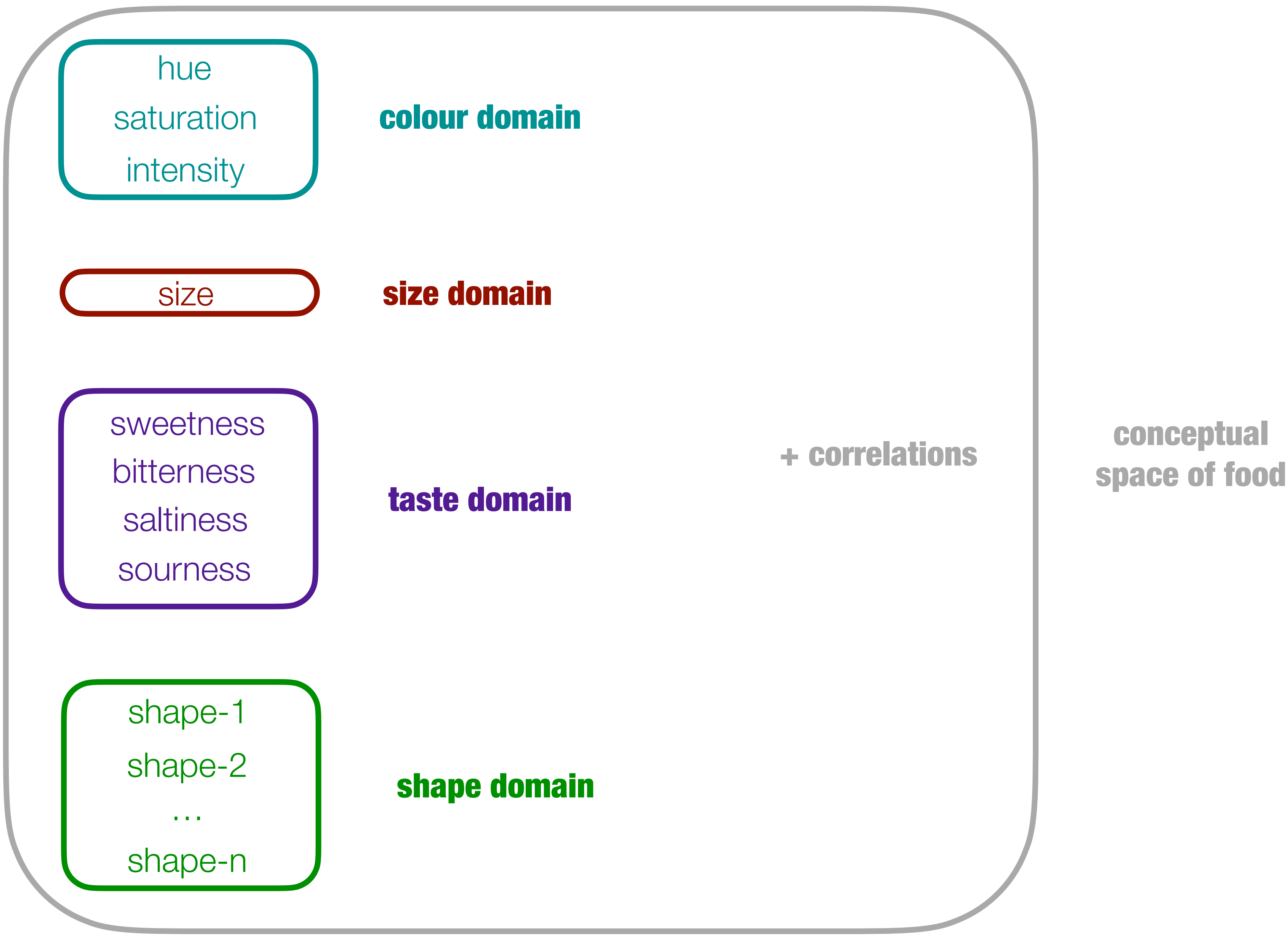
sweetness
bitterness
saltiness
sourness

taste domain

shape-1
shape-2
. . .
shape-n

shape domain

Conceptual spaces: domains



Conceptual spaces

How can we obtain conceptual space representations in practice?

Can we find a generalisation of conceptual spaces for capturing relational rules (e.g. ontologies, existential rules)?

Can we take inspiration from conceptual spaces for developing plausible reasoning strategies, even in cases where we only have partial knowledge about the conceptual space representations?

Overview

Learning conceptual space representations

Relational conceptual spaces

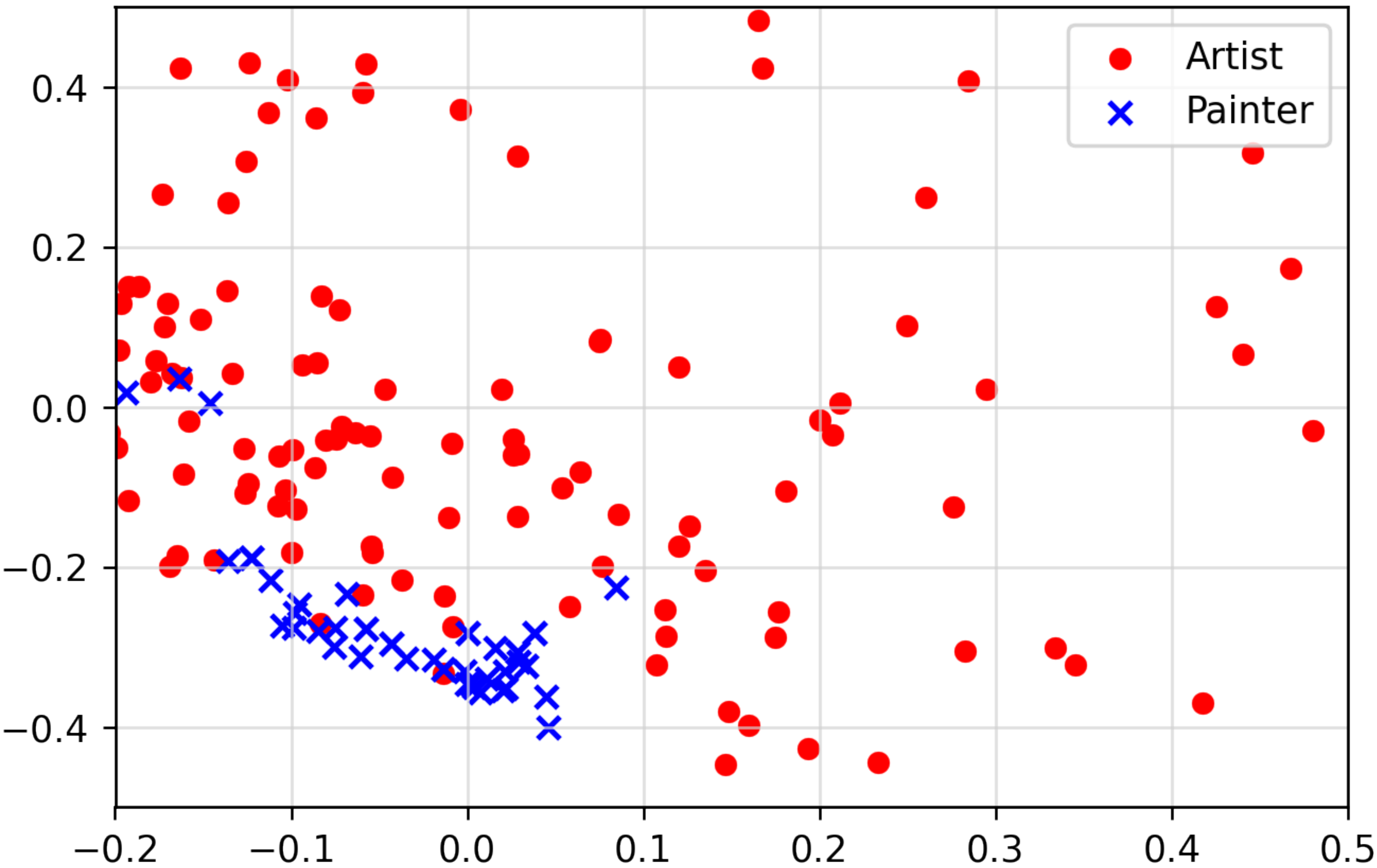
Using vectors for plausible reasoning over symbolic knowledge

Learning conceptual space representations

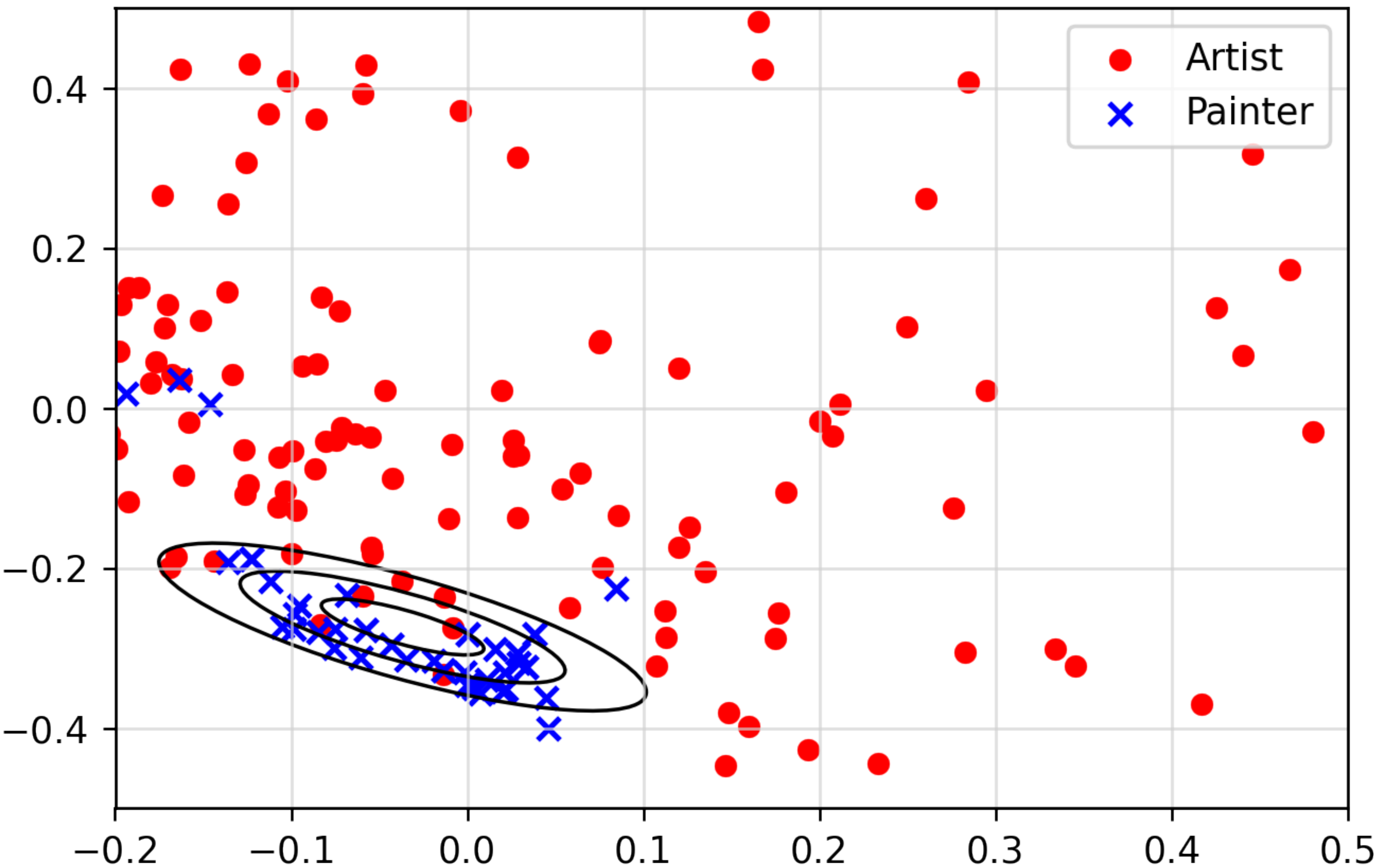
Relational conceptual spaces

Using vectors for plausible reasoning over symbolic knowledge

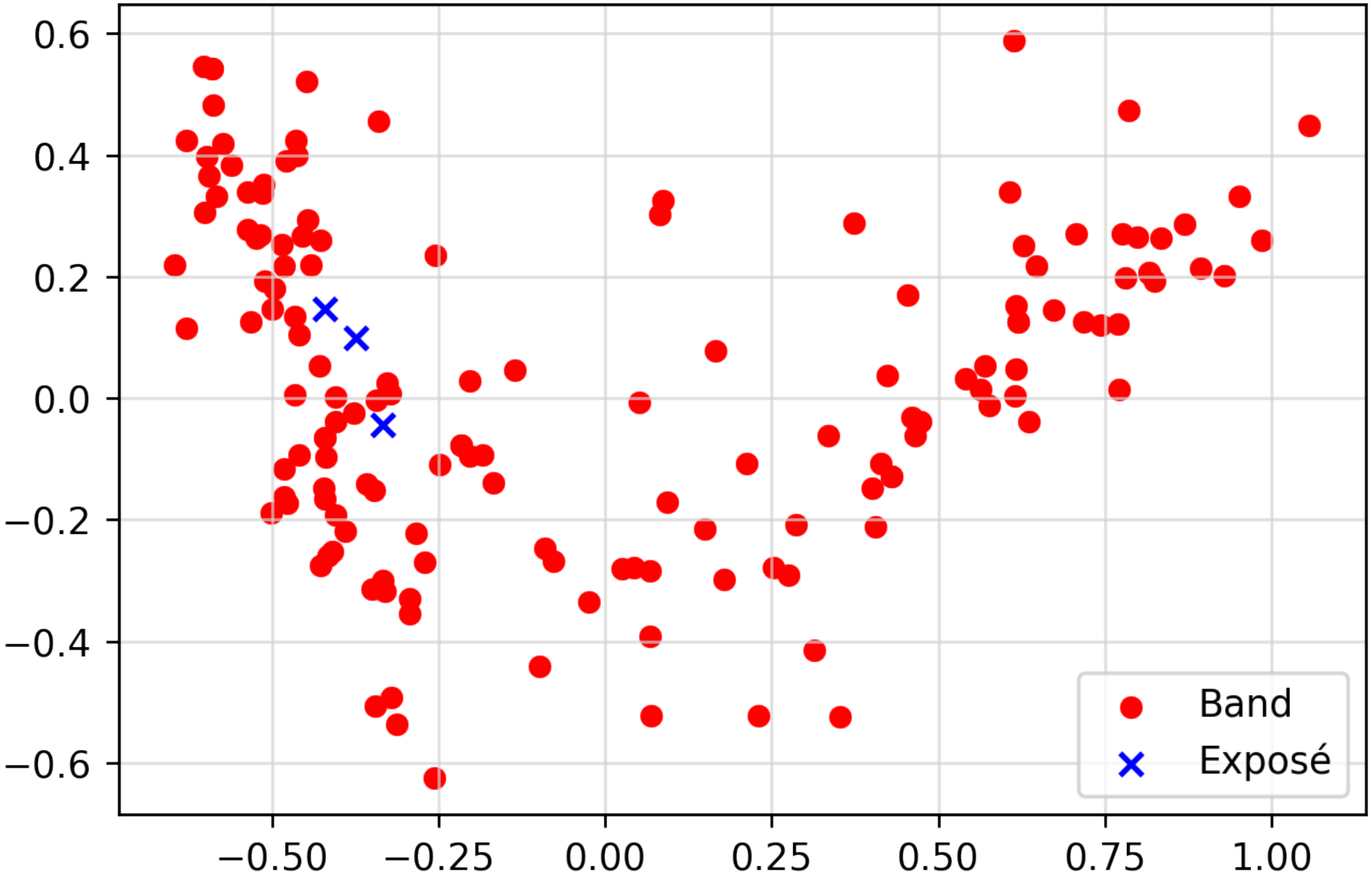
Modelling Concepts as Regions



Learning Gaussian Representations



Bayesian learning with prior knowledge



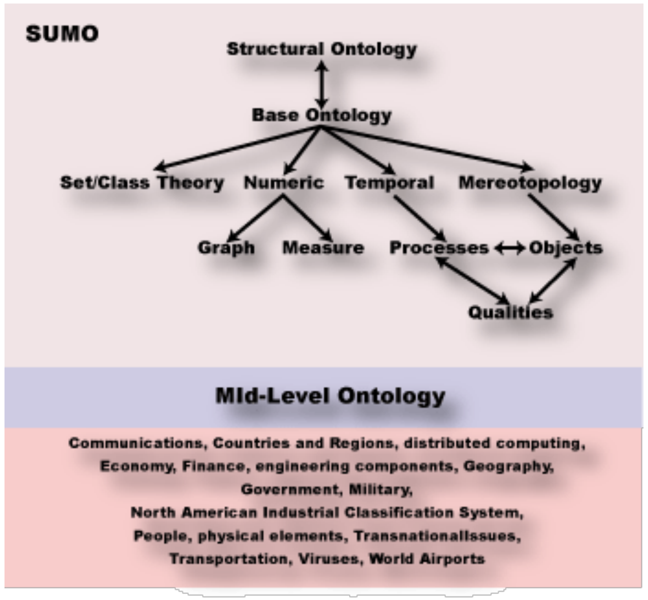
Bayesian learning with prior knowledge

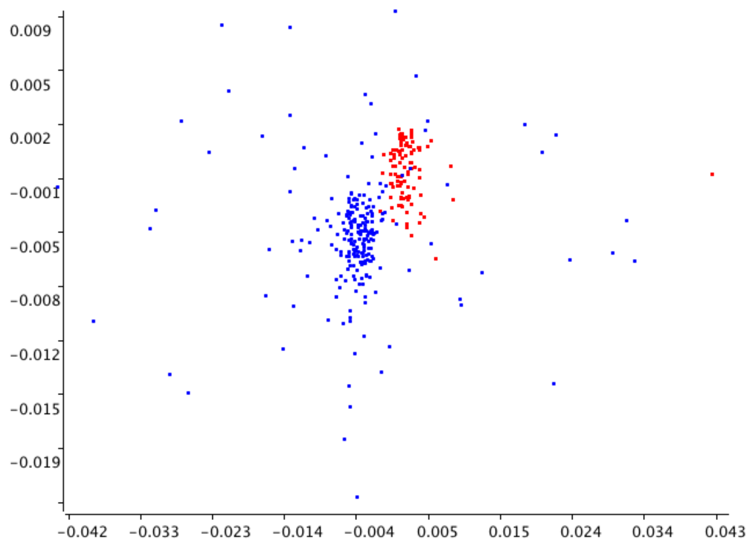
Control how common the instances are

The variance of this Gaussian encodes how much the instances are dispersed across the space

$$P(C|v_a) = \lambda_C \cdot G_C(v_a)$$

Prior knowledge





Prior on mean and variance

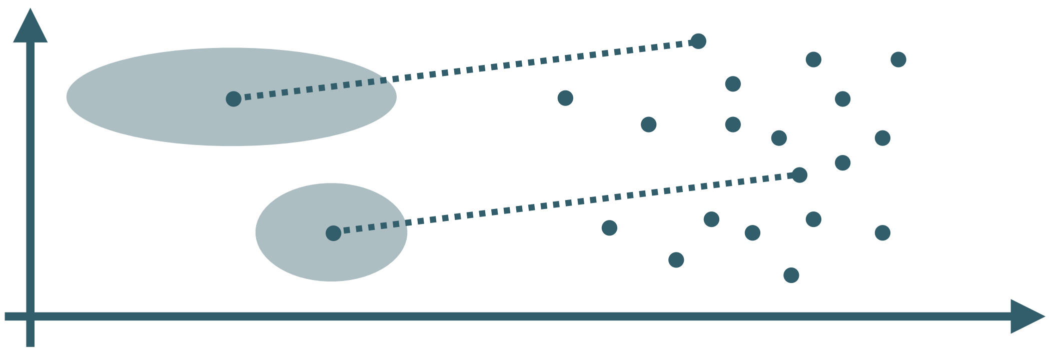
Using taxonomic parents as priors

$$A \sqsubseteq C_1, \ldots, A \sqsubseteq C_k$$

Mean of the Gaussian representing A should be probable according to the Gaussians representing $C_1,..,C_k$

Variance of the Gaussian representing A should be similar to the variance of the Gaussians representing its taxonomic siblings

Using the embedding of the concept name



Gibbs Sampling

Generate sequences of parameters μ_{C0}, μ_{C1},… and Σ_{C0}, Σ_{C1},… for each concept

Steps:

- Init parameters μ_{C0} and Σ_{C0}
- repeatedly iterate over all concepts and in the i^{th} iteration, choose the next samples μ_{Ci} and Σ_{Ci} for each concept C

Use known dependencies between concepts to construct informative priors on μ_{Ci} and Σ_{Ci}

Making prediction

$$P(C|v) = \frac{\lambda_C}{N} \sum_{i=1}^{N} p(v; \mu_C^i, \Sigma_C^i)$$

Average over the Gibbs samples

$$\sum_{i=1}^{s} \log(\lambda_C P(v_i|C)) + \sum_{i=1}^{r} \log(1 - \lambda_C P(u_i|C))$$

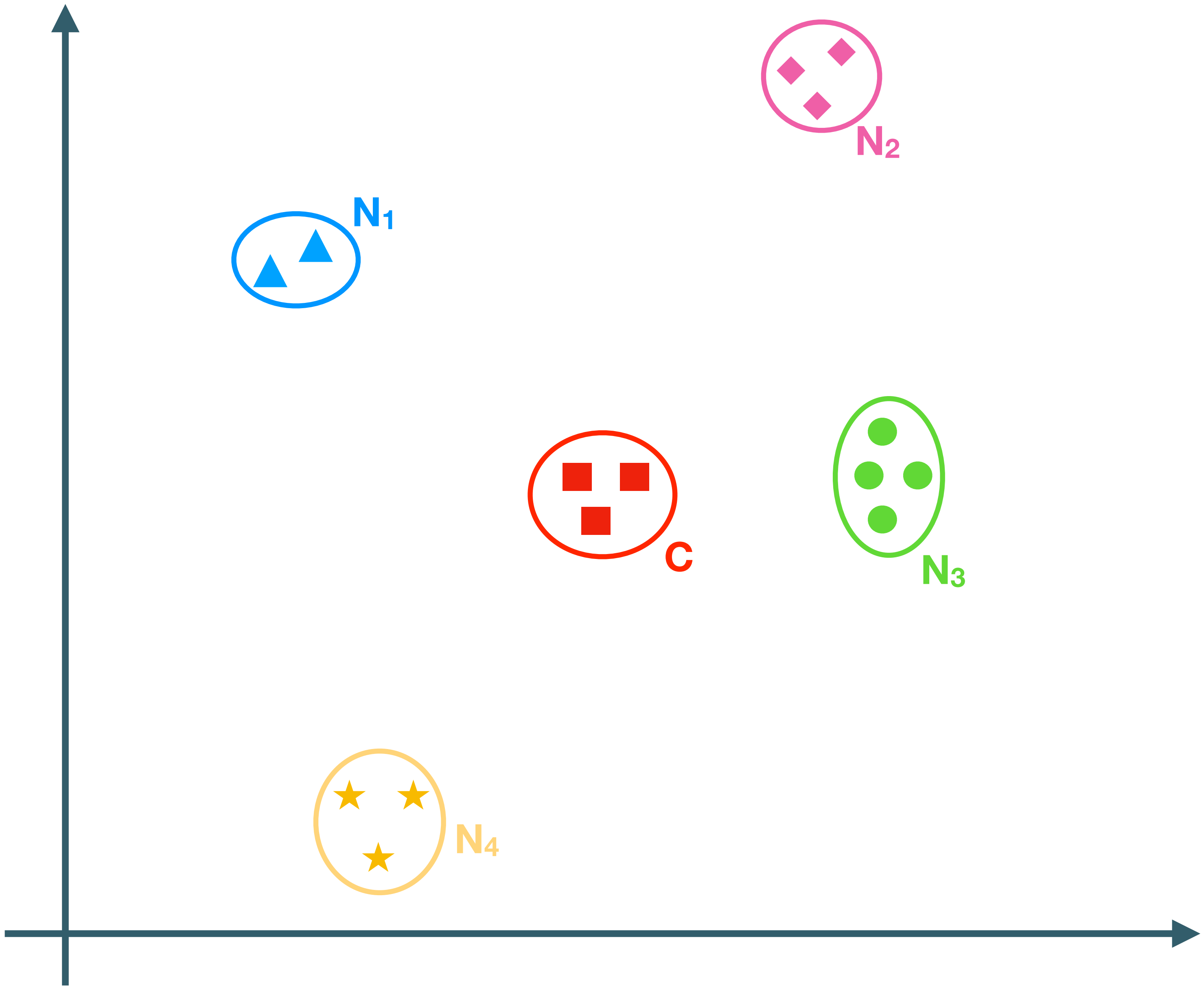
maximizing the likelihood to obtain estimates of the scaling parameters λ

Bayesian learning with prior knowledge

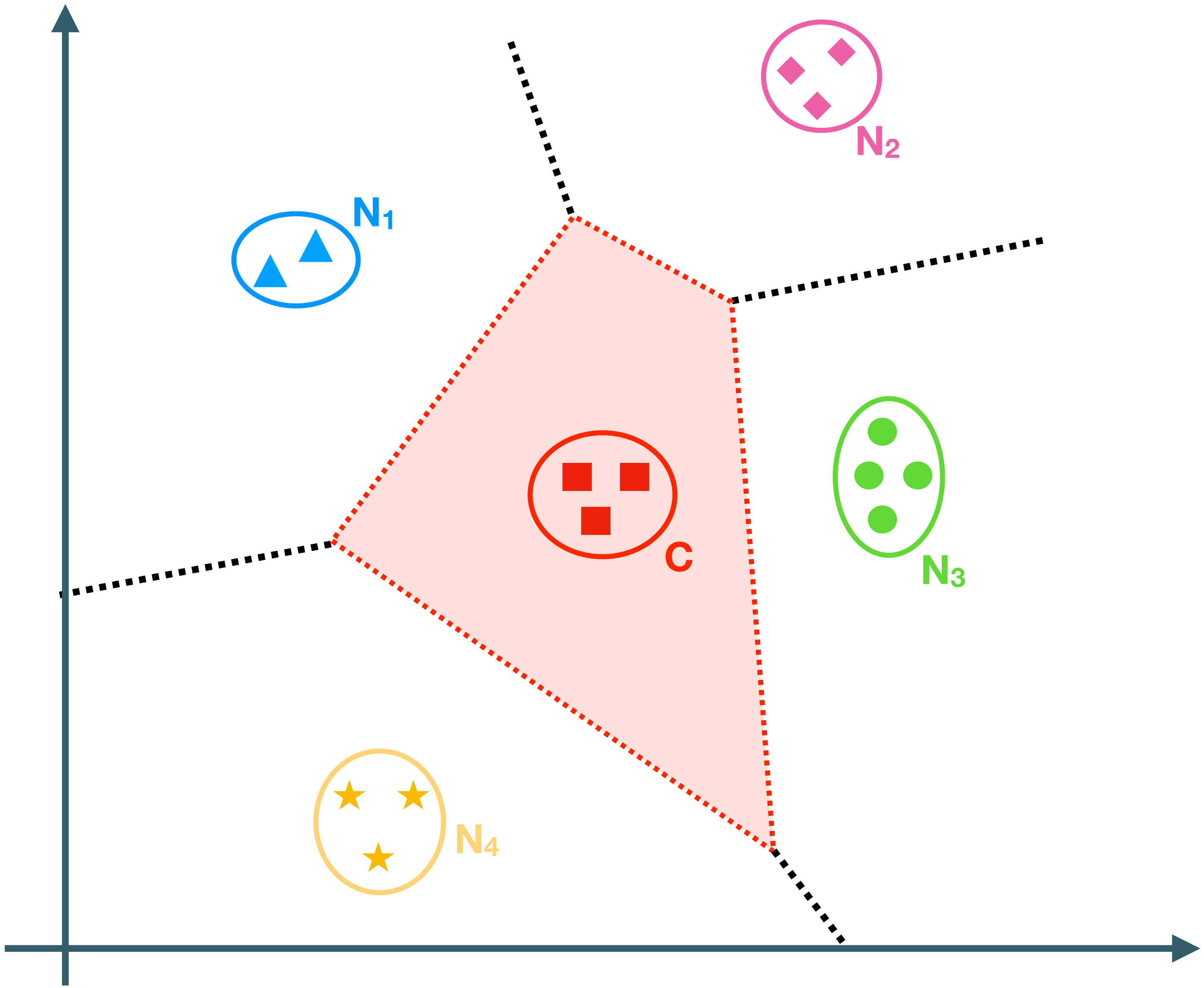
	SVM linear				SVM quadratic			
	Pr	Rec	F1	AP	Pr	Rec	F1	AP
$1 \leq \lvert X \rvert \leq 5$	0.033	0.509	0.062	0.055	0.086	0.046	0.060	0.144
$5 < \lvert X \rvert \leq 10$	0.084	0.922	0.154	0.067	0.116	0.404	0.180	0.163
$10 < \lvert X \rvert \leq 50$	0.111	0.948	0.199	0.081	0.151	0.382	0.216	0.247
$\lvert X \rvert > 50$	0.153	0.217	0.180	0.230	0.224	0.721	0.342	0.260

	Flat prior				Informed prior			
	Pr	Rec	F1	AP	Pr	Rec	F1	AP
$1 \leq \lvert X \rvert \leq 5$	0.212	0.416	0.281	0.290	0.258	0.508	0.343	0.328
$5 < \lvert X \rvert \leq 10$	0.186	0.368	0.247	0.273	0.202	0.474	0.283	0.340
$10 < \lvert X \rvert \leq 50$	0.199	0.496	0.284	0.210	0.242	0.886	0.380	0.276
$\lvert X \rvert > 50$	0.316	0.312	0.314	0.328	0.361	0.678	0.471	0.404

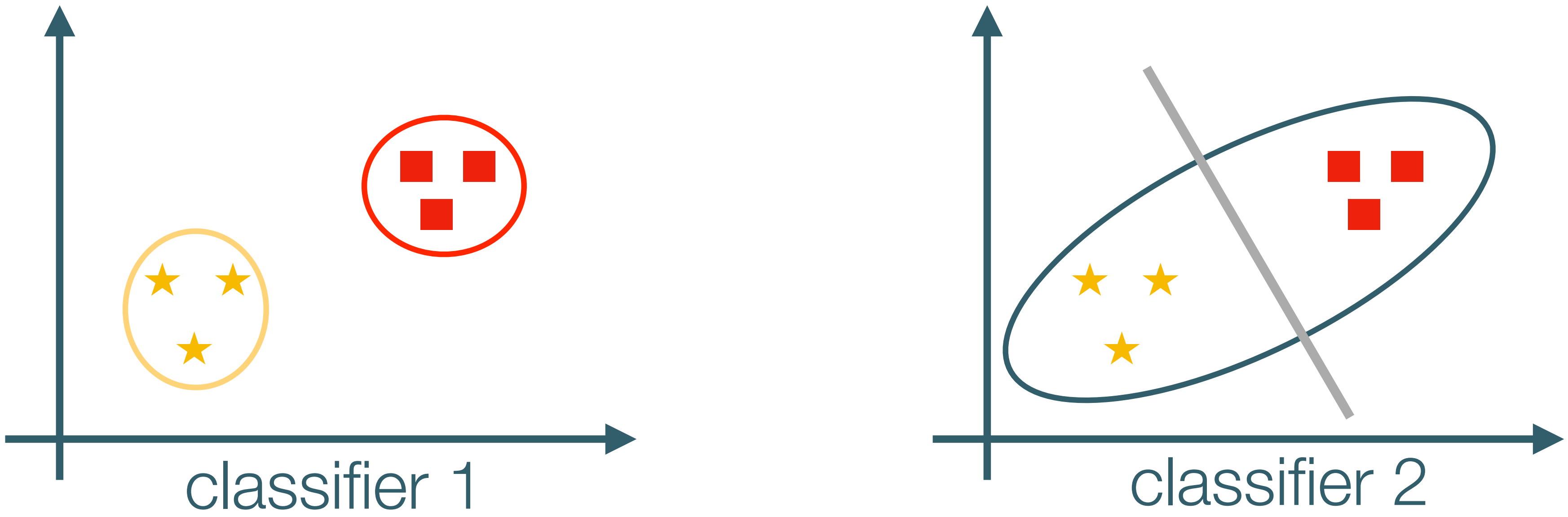
Underestimate the coverage of a category



Conceptual neighbours



How to find conceptual neighbours?



Classifier 2 much better than classifier 1

⇒ A and B are likely conceptual neighbours

How to find conceptual neighbours?

High confidence	**Medium confidence**
Actor – Comedian	Cruise ship – Ocean liner
Journal – Newspaper	Synagogue – Temple
Club – Company	Mountain range – Ridge
Novel – Short story	Child – Man
Tutor – Professor	Monastery – Palace
Museum – Public aquarium	Fairy tale – Short story
Lake – River	Guitarist – Harpsichordist

Predicting conceptual neighbourhood from text

In British geography, a hamlet is considered smaller than a village and …

Find likely conceptual neighbours from large BabelNet concepts

↓

Use these to train a text classifier that can predict conceptual neighbourhood

↓

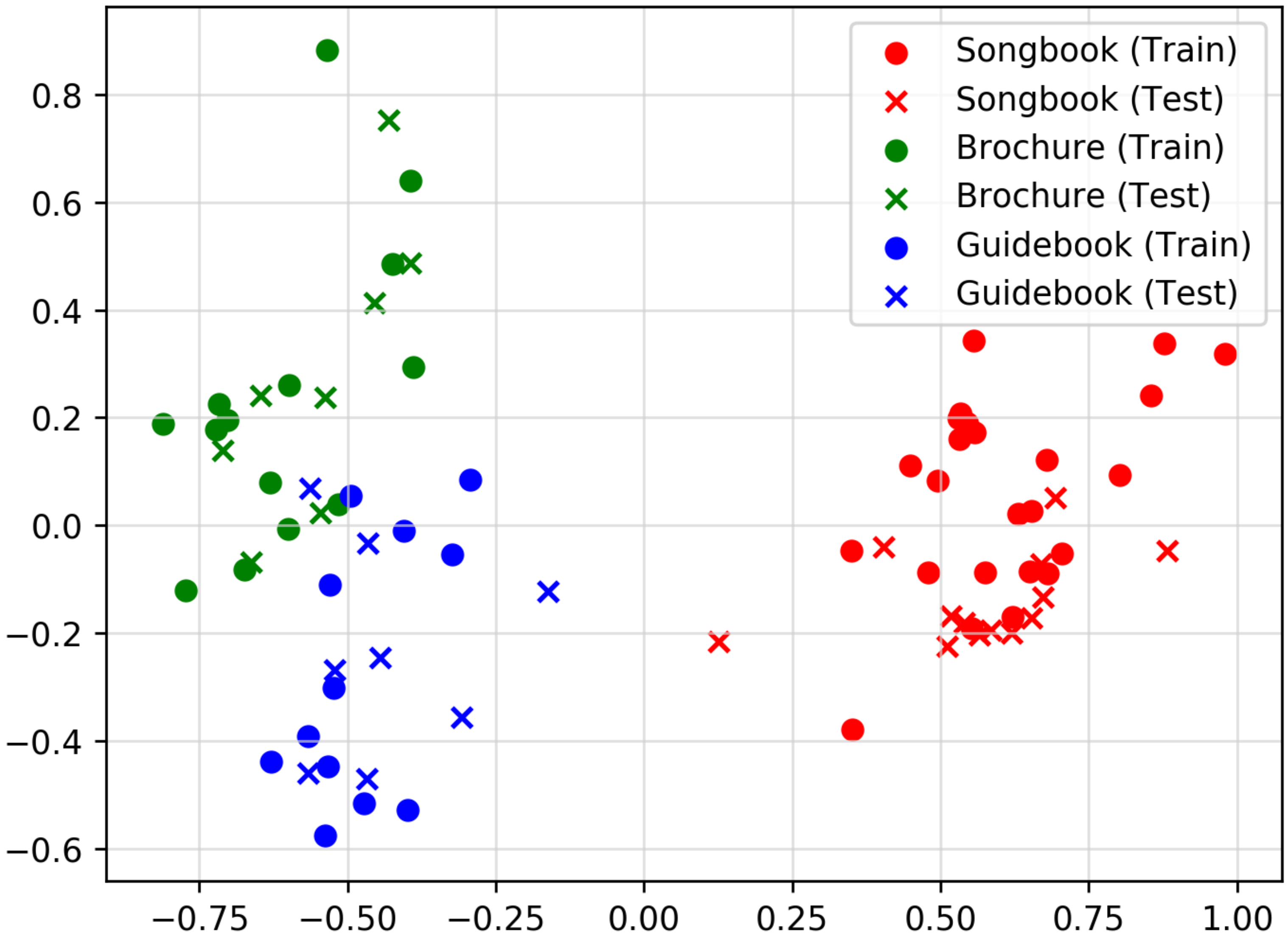
Use the text classifier to identify conceptual neighbours among small BabelNet concepts

Results

	Pr	Rec	F1
Gauss	23.0	27.4	22.3
Multi	37.7	75.2	44.2
Similarity₁	28.7	69.2	33.8
$Similarity_2$	30.0	68.1	34.0
$Similarity_3$	31.6	67.2	34.3
$Similarity_4$	32.8	78.5	38.2
$Similarity_5$	37.2	80.6	42.8
$SECOND\text{-}WEA_1$	32.7	**90.1**	41.9
$SECOND\text{-}WEA_2$	42.2	82.6	49.3
$SECOND\text{-}WEA_3$	43.4	83.1	50.4
$SECOND\text{-}WEA_4$	**47.7**	84.2	**54.2**
$SECOND\text{-}WEA_5$	44.0	82.6	51.1
$SECOND\text{-}BERT_1$	38.5	87.1	47.0
$SECOND\text{-}BERT_2$	43.9	84.1	50.8
$SECOND\text{-}BERT_3$	44.9	84.4	52.2
$SECOND\text{-}BERT_4$	46.2	85.4	53.3
$SECOND\text{-}BERT_5$	43.8	84.7	51.3

	Acc	F1	Pr	Rec
Avg.	**70.6**	**69.0**	**69.4**	**69.0**
BERT	66.9	65.8	65.9	66.2
#sents	61.6	46.6	43.3	54.3

Results


Songbook (Train)
Songbook (Test)
Brochure (Train)
Brochure (Test)
Guidebook (Train)
Guidebook (Test)
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

Results

Concept	Top neighbour	F1
amphitheater	velodrome	0.67
proxy server	application server	0.61
ketch	cutter	0.74
quintet	brass band	0.67
sand dune	drumlin	0.71

Concept	Top neighbour	F1
bachelor's degree	undergraduate degree	0.34
episodic video game	multiplayer game	0.34
501(c) organization	not-for-profit arts organization	0.29
heavy bomber	triplane	0.41
ministry	United States government	0.33

Similarity in Entity Embeddings

Castle

https://en.wikipedia.org/wiki/Gamlehaugen

https://fr.wikipedia.org/wiki/Château_de_Versailles

Museum

https://en.wikipedia.org/wiki/University_Museum_of_Bergen

https://fr.wikipedia.org/wiki/Musée_du_Louvre

Similarity in Entity Embeddings

Paris

https://fr.wikipedia.org/wiki/Musée_du_Louvre

https://fr.wikipedia.org/wiki/Château_de_Versailles

Bergen

https://en.wikipedia.org/wiki/University_Museum_of_Bergen

https://en.wikipedia.org/wiki/Gamlehaugen

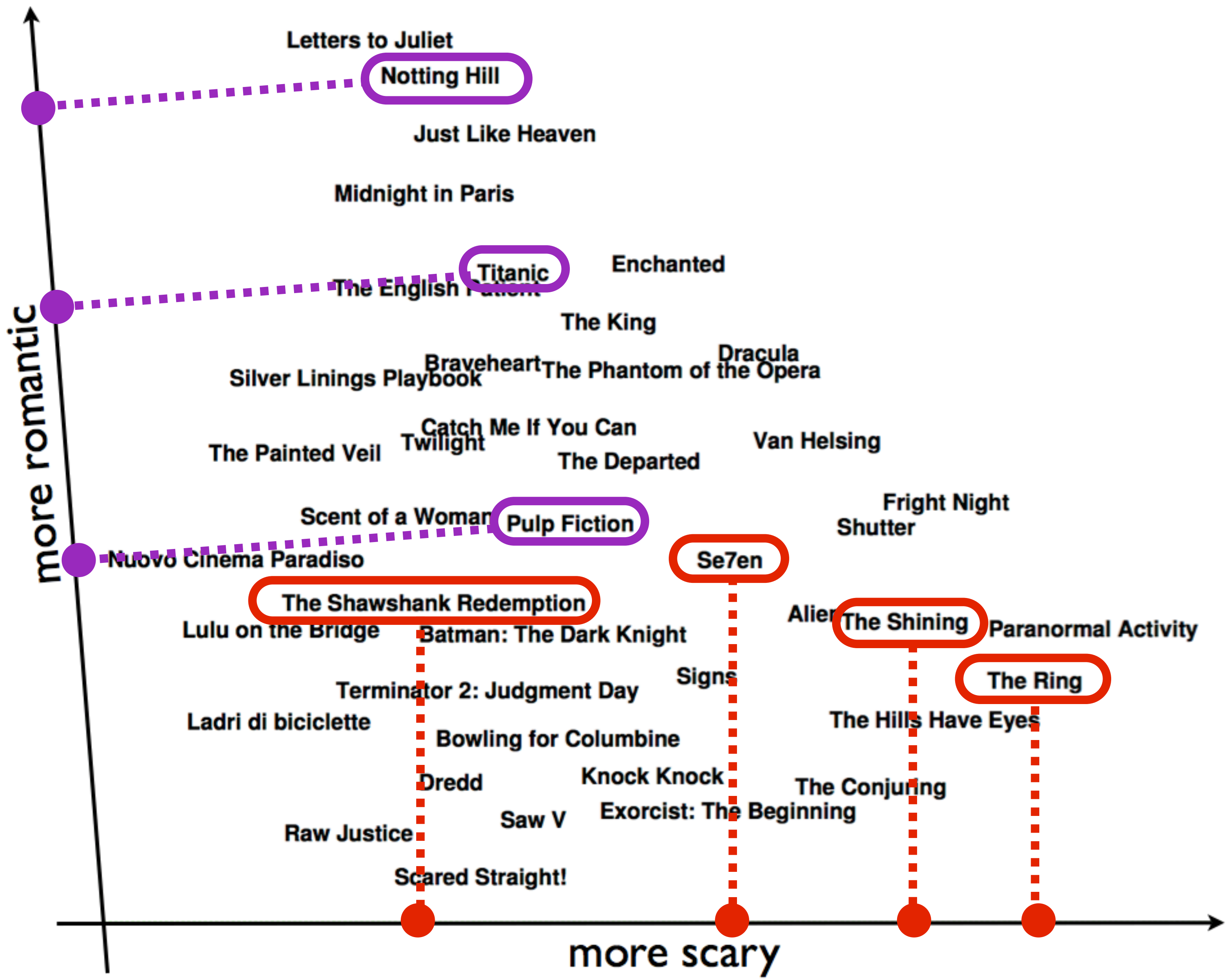
Similarity in Entity Embeddings

The similarity is **inherently multi-faceted**, however standard entity embeddings do not reflect those facets

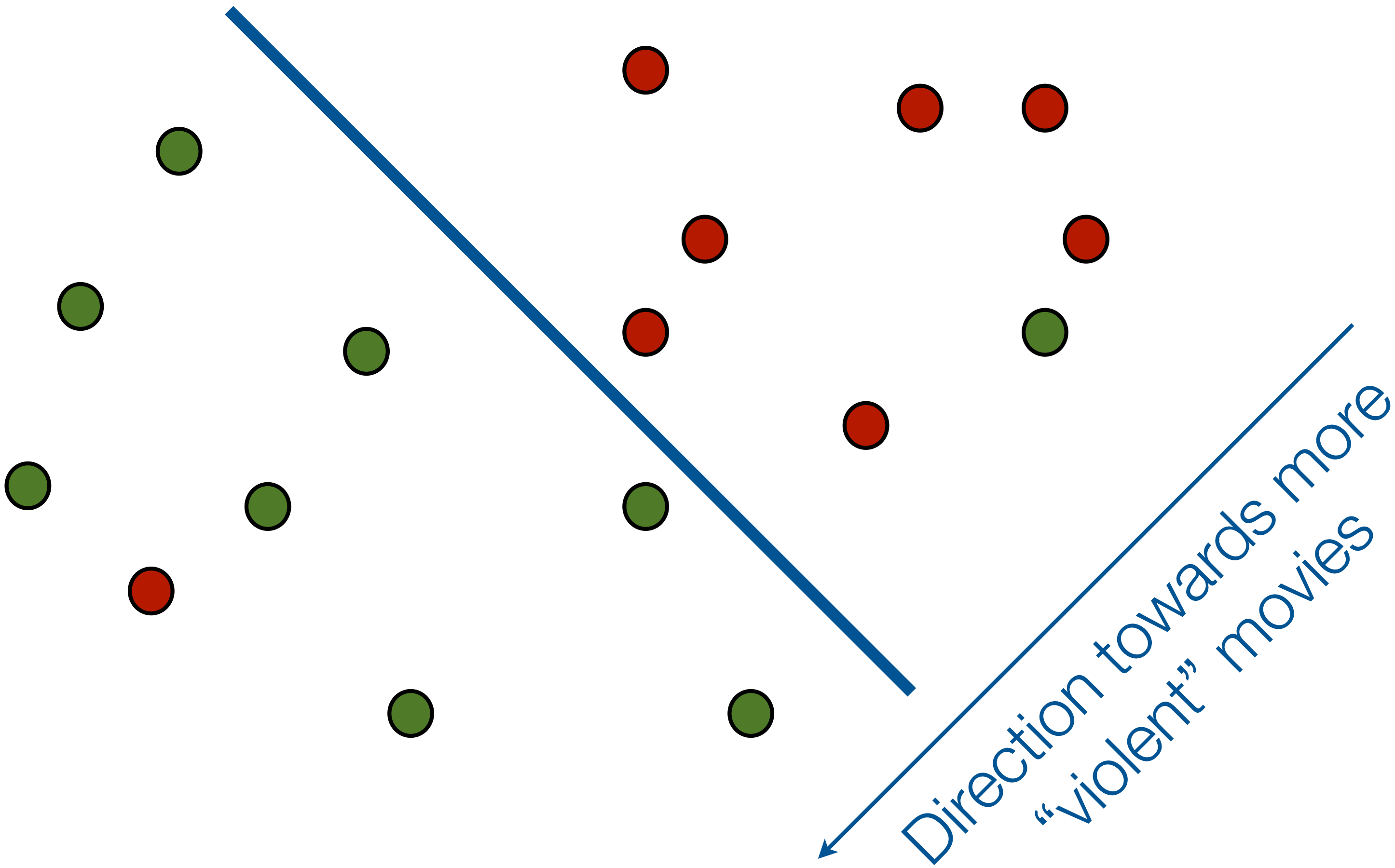
Instead of learning one embedding for a giving domain, we learn **several low dimensional embeddings**, each of which capture different aspect of similarity.

Learning interpretable dimensions

Learning interpretable dimensions



Learning interpretable dimensions



movies whose associated text contains the word "violent"
movies whose associated text does not contain the word "violent"

Semantic attributes

horror movies	zombie, much gore, slashers, vampires, scary monsters, ...
supernatural	a witch, ghost stories, mysticism, a demon, the afterlife, ...
scientist	experiment, the virus, radiation, the mad scientist, ...
criminal	the mafia, robbers, parole, the thieves, the mastermind, ...
the animation	the voices, drawings, the artwork, the cartoons, anime, ...
touching	inspirational, warmth, dignity, sadness, heartwarming, ...
budget	a low budget film, b movies, independent films, ...
political	socialism, idealism, terrorism, leaders, protests, equality, corruption, ...
clever	schemes, satire, smart, witty dialogue, ingenious, ...
bizarre	odd, twisted, peculiar, lunacy, surrealism, obscure, ...
predictable	forgettable, unoriginal, formulaic, implausible, contrived, ...
twists	unpredictable, betrayals, many twists and turns, deceit, ...
romantic	lovers, romance, the chemistry, kisses, true love, ...
eerie	paranoid, spooky, impending doom, dread, ominous, ...
scary	shivers, chills, creeps, frightening, the dark, goosebumps, ...
cheesy	camp, corny, tacky, laughable, a guilty pleasure, ...
she's	her apartment, her sister, her death, her family, the heroine, actress, ...
his life	his son, his quest, his guilt, a man, his voice, his fate, his anger, ...
hilarious	humorous, really funny, a very funny movie, amusing, ...
vhs	laserdisc, videotape, this dvd version, first released, this classic, ...
violence	violent, cold blood, knives, bad people, brotherhood, ...
adaptation	the stage version, the source material, the novel, ...
sequel	the trilogy, the first film, the same formula, this franchise, ...
era	the fifties, the sixties, the seventies, a period piece, the depression, ...

Thematic properties

horror movies	zombie, much gore, slashers, vampires, scary monsters, ...
supernatural	a witch, ghost stories, mysticism, a demon, the afterlife, ...
scientist	experiment, the virus, radiation, the mad scientist, ...
criminal	the mafia, robbers, parole, the thieves, the mastermind, ...
the animation	the voices, drawings, the artwork, the cartoons, anime, ...
touching	inspirational, warmth, dignity, sadness, heartwarming, ...
budget	a low budget film, b movies, independent films, ...
political	socialism, idealism, terrorism, leaders, protests, equality, corruption, ...
clever	schemes, satire, smart, witty dialogue, ingenious, ...
bizarre	odd, twisted, peculiar, lunacy, surrealism, obscure, ...
predictable	forgettable, unoriginal, formulaic, implausible, contrived, ...
twists	unpredictable, betrayals, many twists and turns, deceit, ...
romantic	lovers, romance, the chemistry, kisses, true love, ...
eerie	paranoid, spooky, impending doom, dread, ominous, ...
scary	shivers, chills, creeps, frightening, the dark, goosebumps, ...
cheesy	camp, corny, tacky, laughable, a guilty pleasure, ...
she's	her apartment, her sister, her death, her family, the heroine, actress, ...
his life	his son, his quest, his guilt, a man, his voice, his fate, his anger, ...
hilarious	humorous, really funny, a very funny movie, amusing, ...
vhs	laserdisc, videotape, this dvd version, first released, this classic, ...
violence	violent, cold blood, knives, bad people, brotherhood, ...
adaptation	the stage version, the source material, the novel, ...
sequel	the trilogy, the first film, the same formula, this franchise, ...
era	the fifties, the sixties, the seventies, a period piece, the depression, ...

Results

		Place types			Movies			Organisations		Buildings	
		Fours.	Geo.	OpenC.	KeyW.	Genre	Rating	Country	HL.	Country	AL.
DT-D1	MDS	0.34	0.26	0.26	**0.26**	0.38	0.43	0.67	0.24	0.47	0.47
DT-D3	MDS	0.52	0.27	0.32	**0.27**	**0.43**	0.47	0.70	0.27	0.47	0.46
SVM	MDS	0.65	0.31	0.35	0.25	**0.54**	0.54	0.71	**0.26**	0.38	0.39
Gaussian	MDS	0.81	0.45	**0.46**	0.26	0.58	0.48	0.74	0.27	0.53	0.51

Organising quality dimensions into domains

Select the words that can best be represented as directions in the 100-dimensional space

Organising quality dimensions into domains

Select the words that can best be represented as directions in the 100-dimensional space

↓ ↓ ↓

Cluster these words

$$d(a,b) = \begin{cases} 1 - \cos(\mathbf{w}_a, \mathbf{w}_b) & \text{if } o(a,b) \leq \lambda \\ 1 & \text{otherwise} \end{cases}$$

$$o(a,b) = \min\left(\frac{|pos_a \cap pos_b|}{|pos_a|}, \frac{|pos_a \cap pos_b|}{|pos_b|}\right)$$

Organising quality dimensions into domains

Select the words that can best be represented as directions in the 100-dimensional space

↓

Cluster these words

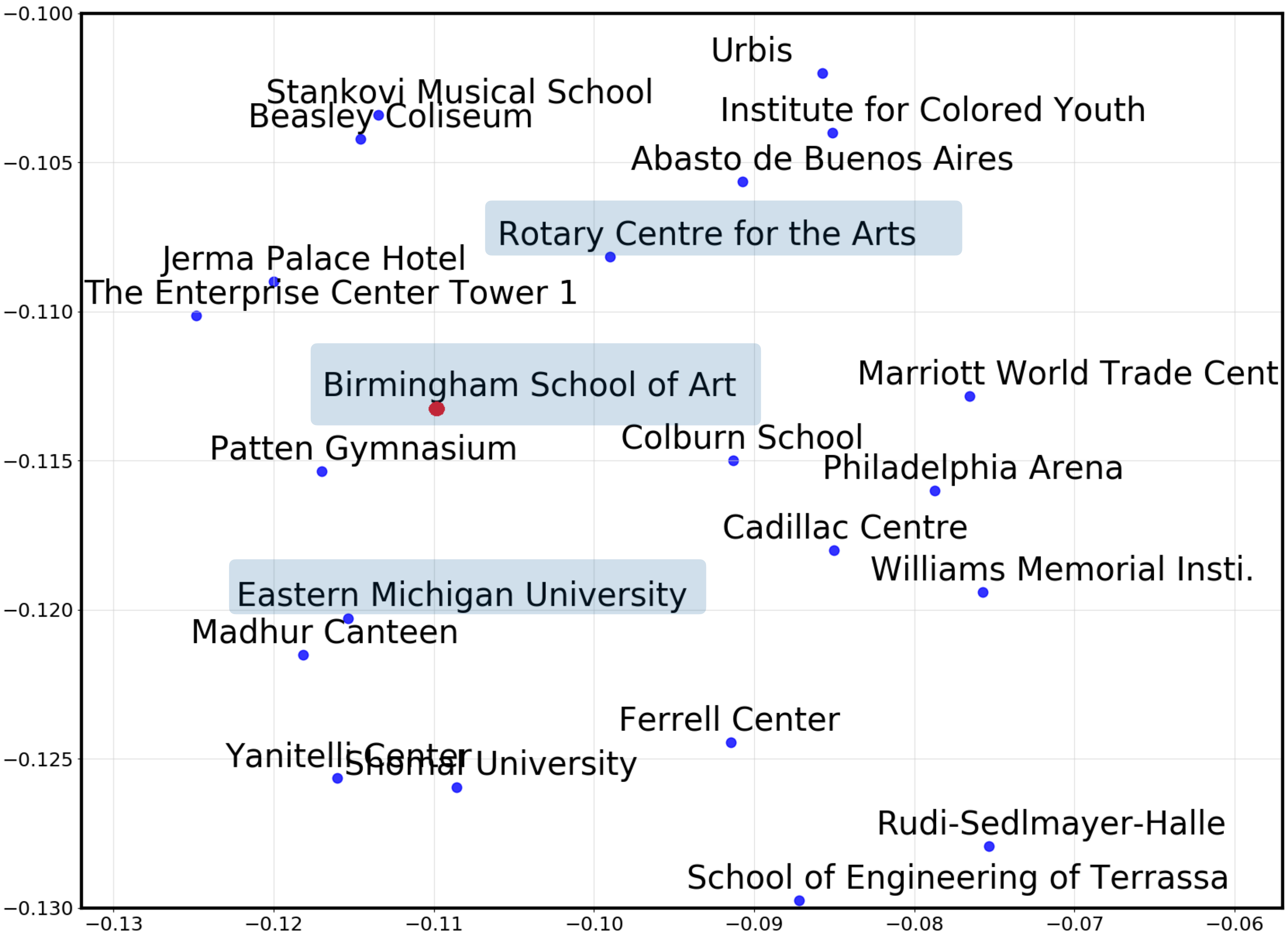
↓

Find the 10-dimensional subspace that can best model the words from the top cluster as directions

↓

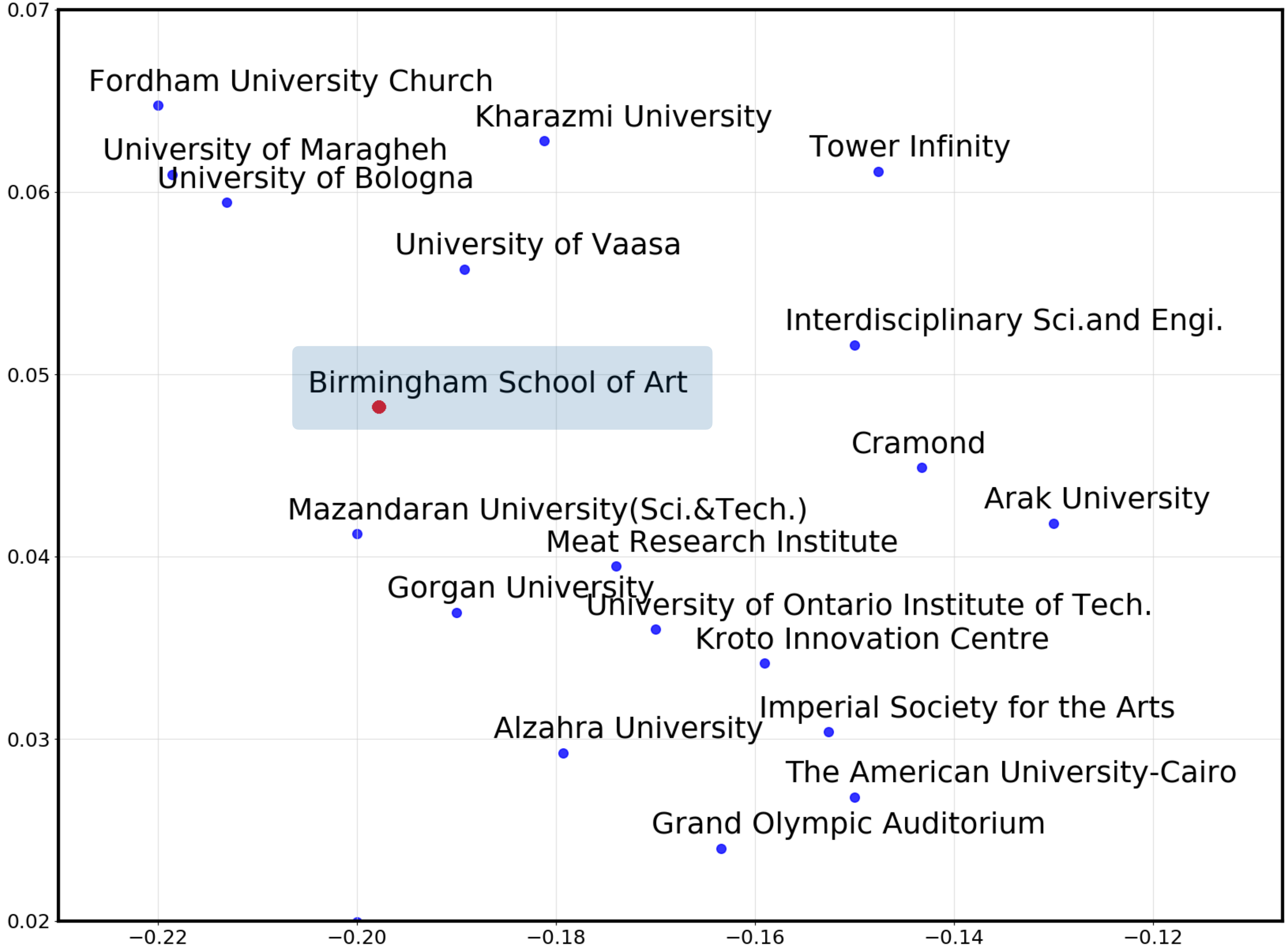
Repeat the same process on the 90-dimensional remainder space, disregarding words that are already modelled in the subspace

Organising quality dimensions into domains



First two principal components of the full space

Organising quality dimensions into domains

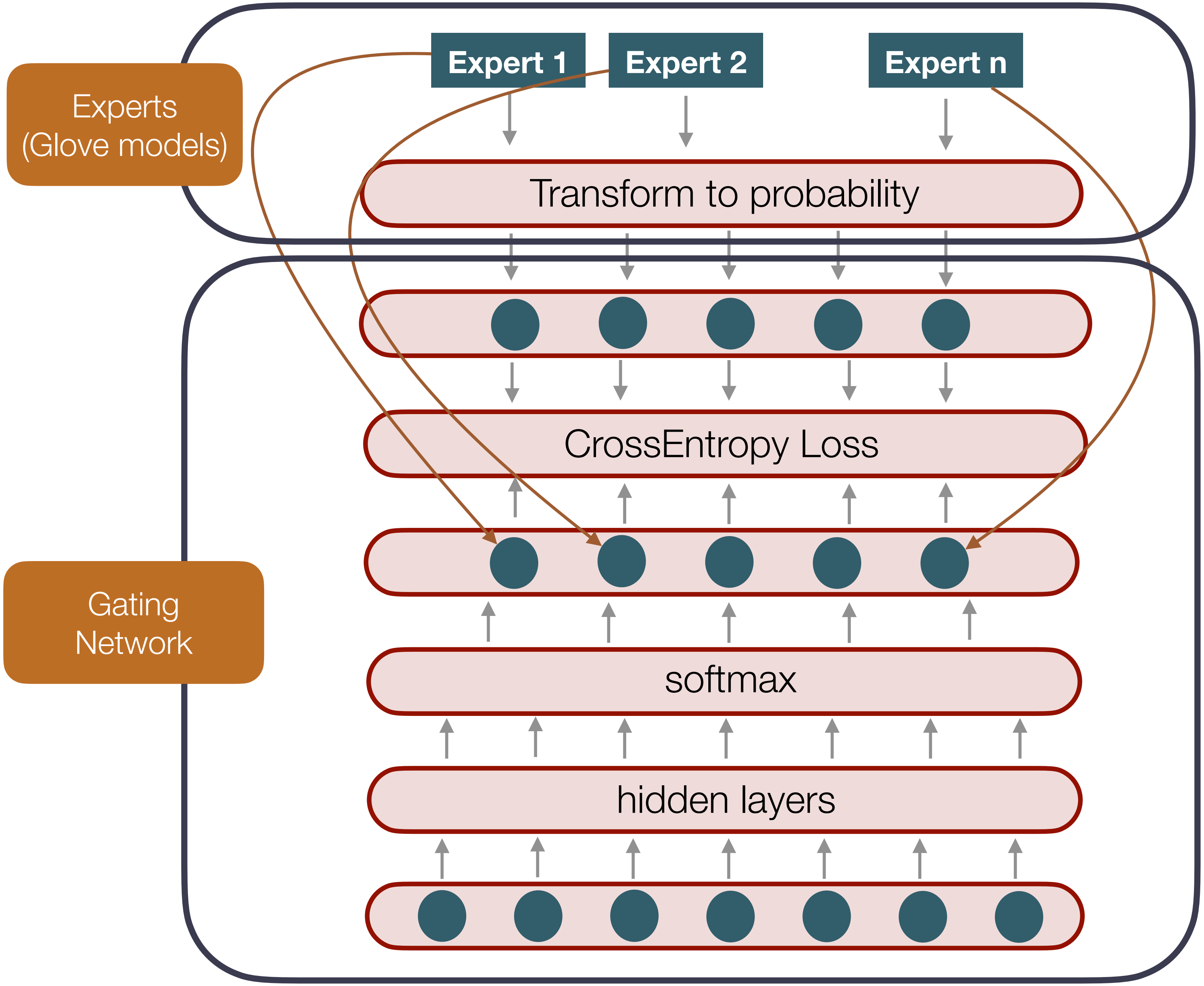


First two principal components of the “place type” subspace

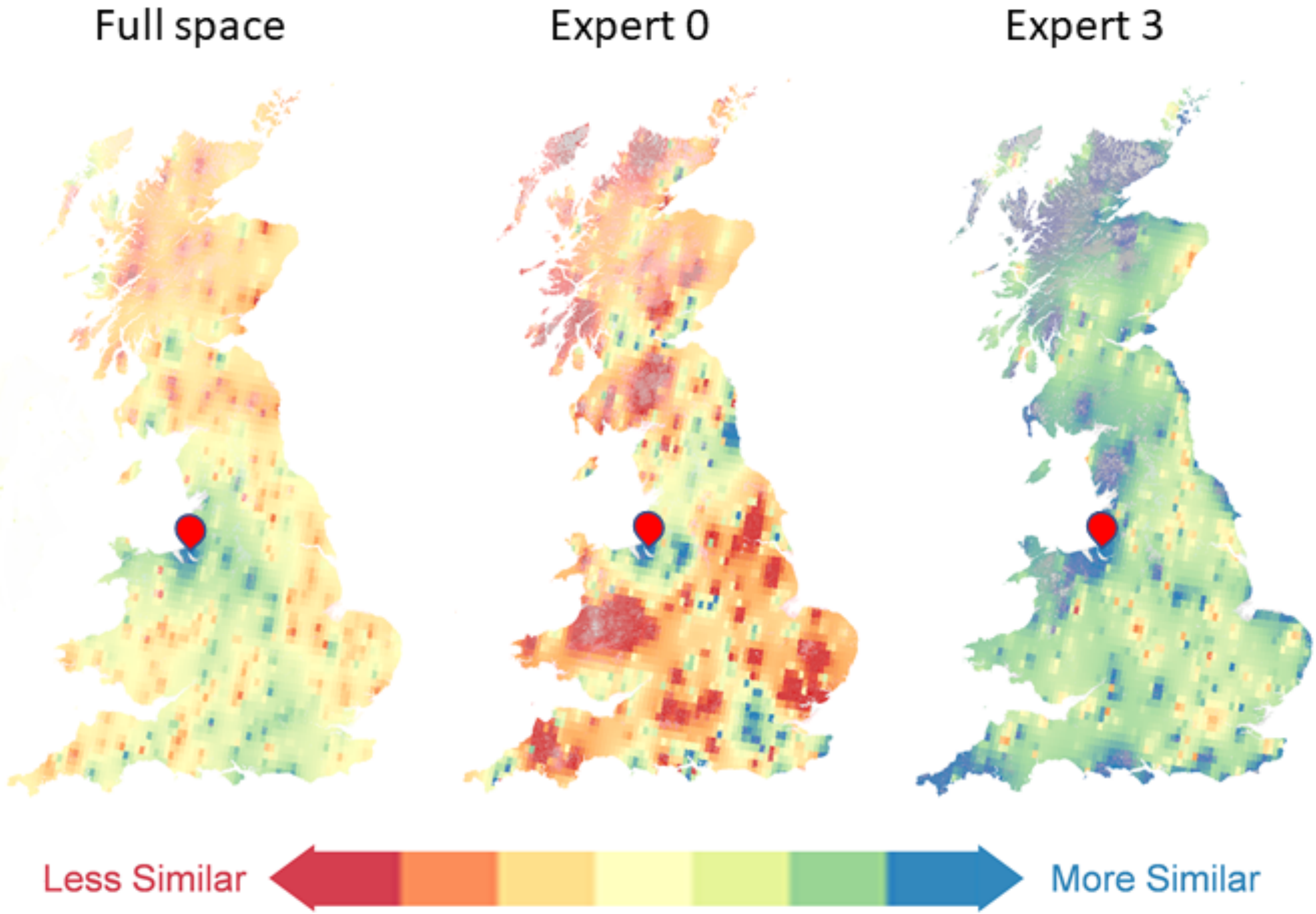
Results

		Place types			Movies			Organisations		Buildings	
		Fours.	Geo.	OpenC.	KeyW.	Genre	Rating	Country	HL.	Country	AL.
DT-D1	MDS	0.34	0.26	0.26	**0.26**	0.38	0.43	0.67	0.24	0.47	0.47
	IncAgg	**0.45**	**0.30**	**0.30**	0.25	**0.40**	**0.47**	**0.76**	**0.26**	**0.50**	0.50
	CosIncAgg	0.45	0.26	**0.30**	0.24	0.38	0.43	0.75	0.23	0.43	0.42
	IncHDB	0.43	0.26	0.28	0.25	0.38	0.40	0.50	0.22	0.46	0.46
	NonIncHDB	0.30	0.20	0.27	0.23	0.34	0.40	0.50	0.20	0.46	0.47
	NonIncAgg	0.33	0.24	0.27	0.23	0.33	0.42	0.40	0.21	0.48	0.47
DT-D3	MDS	0.52	0.27	0.32	**0.27**	**0.43**	0.47	0.70	0.27	0.47	0.46
	IncAgg	**0.58**	**0.34**	**0.34**	**0.27**	0.41	**0.47**	0.77	**0.30**	**0.54**	**0.52**
	CosIncAgg	0.54	0.28	**0.34**	0.25	0.40	0.45	**0.78**	0.26	0.47	0.45
	IncHDB	0.57	0.26	0.31	**0.27**	0.41	0.45	0.70	0.27	0.49	0.50
	NonIncHDB	0.43	0.24	0.27	0.26	0.38	0.44	0.60	0.21	0.48	0.49
	NonIncAgg	0.36	0.30	0.29	0.24	0.38	0.45	0.65	0.22	0.51	0.50
SVM	MDS	0.65	0.31	0.35	0.25	**0.54**	0.54	0.71	**0.26**	0.38	0.39
	IncAgg	**0.73**	0.33	**0.37**	**0.26**	**0.54**	**0.55**	0.76	**0.26**	**0.52**	**0.51**
	CosIncAgg	0.62	0.33	0.34	0.25	0.52	0.53	**0.80**	0.12	0.50	0.50
	IncHDB	0.65	0.30	0.36	0.23	0.50	0.51	0.70	0.20	0.51	**0.51**
	NonIncHDB	0.60	**0.35**	**0.37**	0.24	0.46	0.52	0.68	0.24	**0.52**	**0.51**
	NonIncAgg	0.58	**0.35**	0.35	0.24	0.48	0.51	0.72	0.26	0.50	**0.51**
Gaussian	MDS	0.81	0.45	**0.46**	0.26	0.58	0.48	0.74	0.27	0.53	0.51
	IncAgg	**0.87**	**0.48**	0.45	**0.28**	**0.60**	**0.51**	**0.81**	0.27	0.54	**0.55**
	CosIncAgg	0.81	0.45	**0.46**	**0.28**	**0.60**	**0.51**	**0.81**	**0.28**	0.53	0.53
	IncHDB	0.84	0.43	0.43	0.27	**0.60**	**0.51**	0.80	**0.28**	0.54	0.53
	NonIncHDB	0.75	0.41	0.40	0.23	0.51	0.47	0.75	0.27	**0.59**	0.53
	NonIncAgg	0.71	0.46	0.45	0.22	0.52	0.46	0.77	0.27	0.58	0.53

Learning Multi-Facet Entity Embeddings



Summary on relational conceptual spaces



Summary on learning conceptual space representations

We can **model concept as convex regions**, using Gaussian representations with prior knowledge

The role of **conceptual neighborhood**, for modelling categories, focusing especially on categories with a relatively **small number of instances**

Learning **multi-facets embeddings**, characterised as **quality dimensions** in the embedding using heuristic methods and MoE model

Open questions

Can we learn **conceptual spaces from data**?

How to learn meaningful region representations for concept that do not have instances?

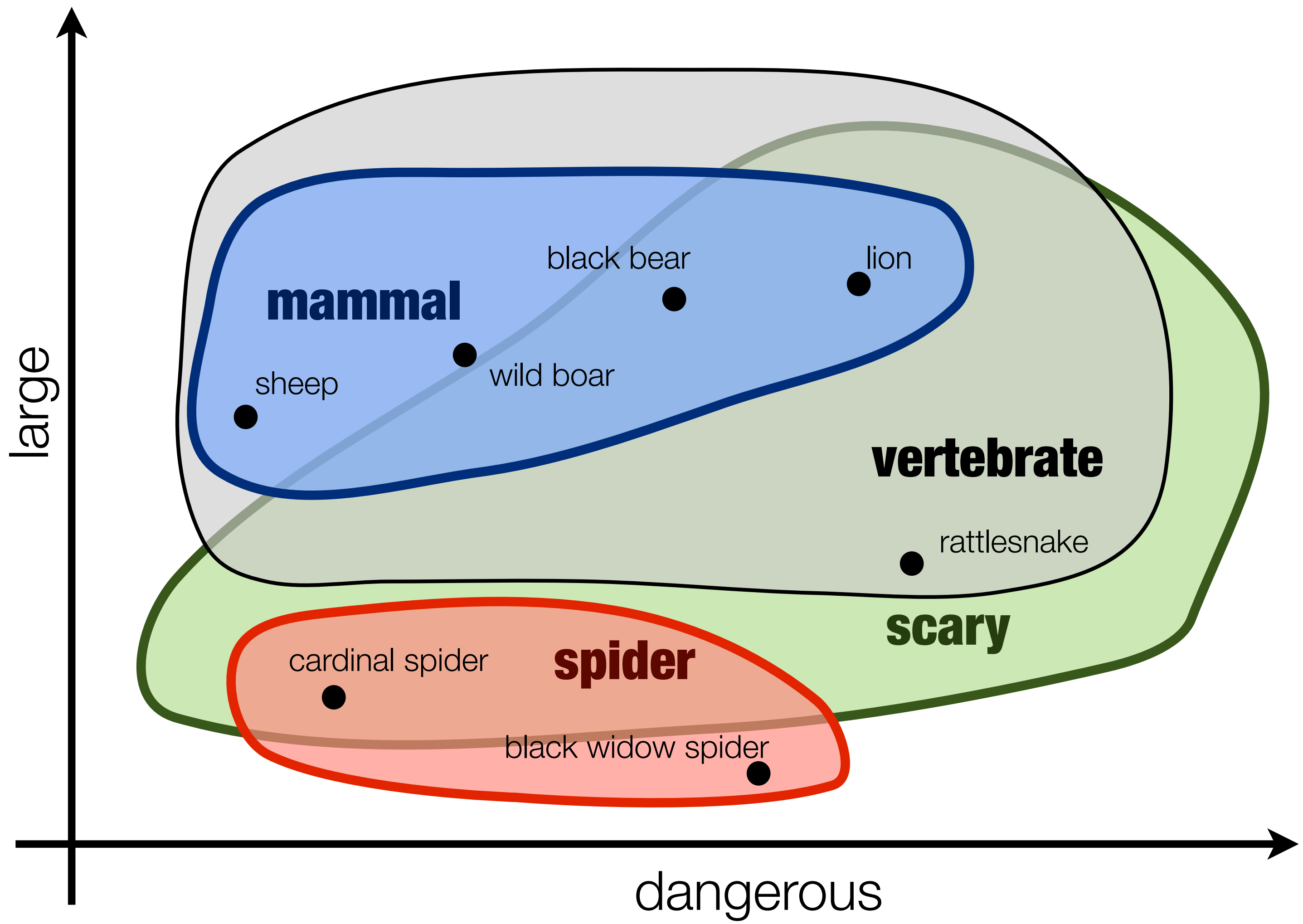
How to learning disentangled representations from contextualised word embeddings?

Learning conceptual space representations

Relational conceptual spaces

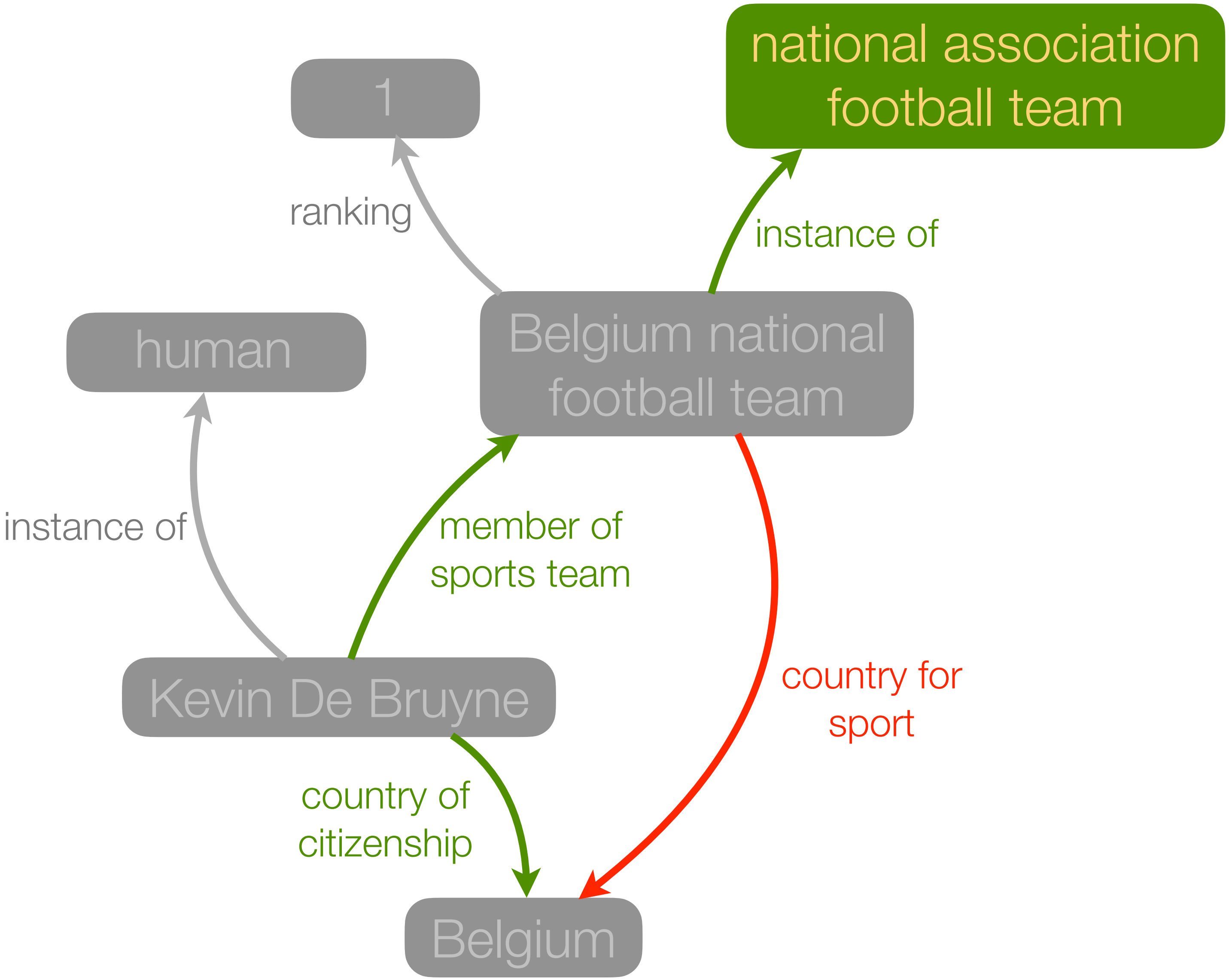
Using vectors for plausible reasoning over symbolic knowledge

What is the relational counterpart of a conceptual space?



Knowledge graphs

Knowledge graphs



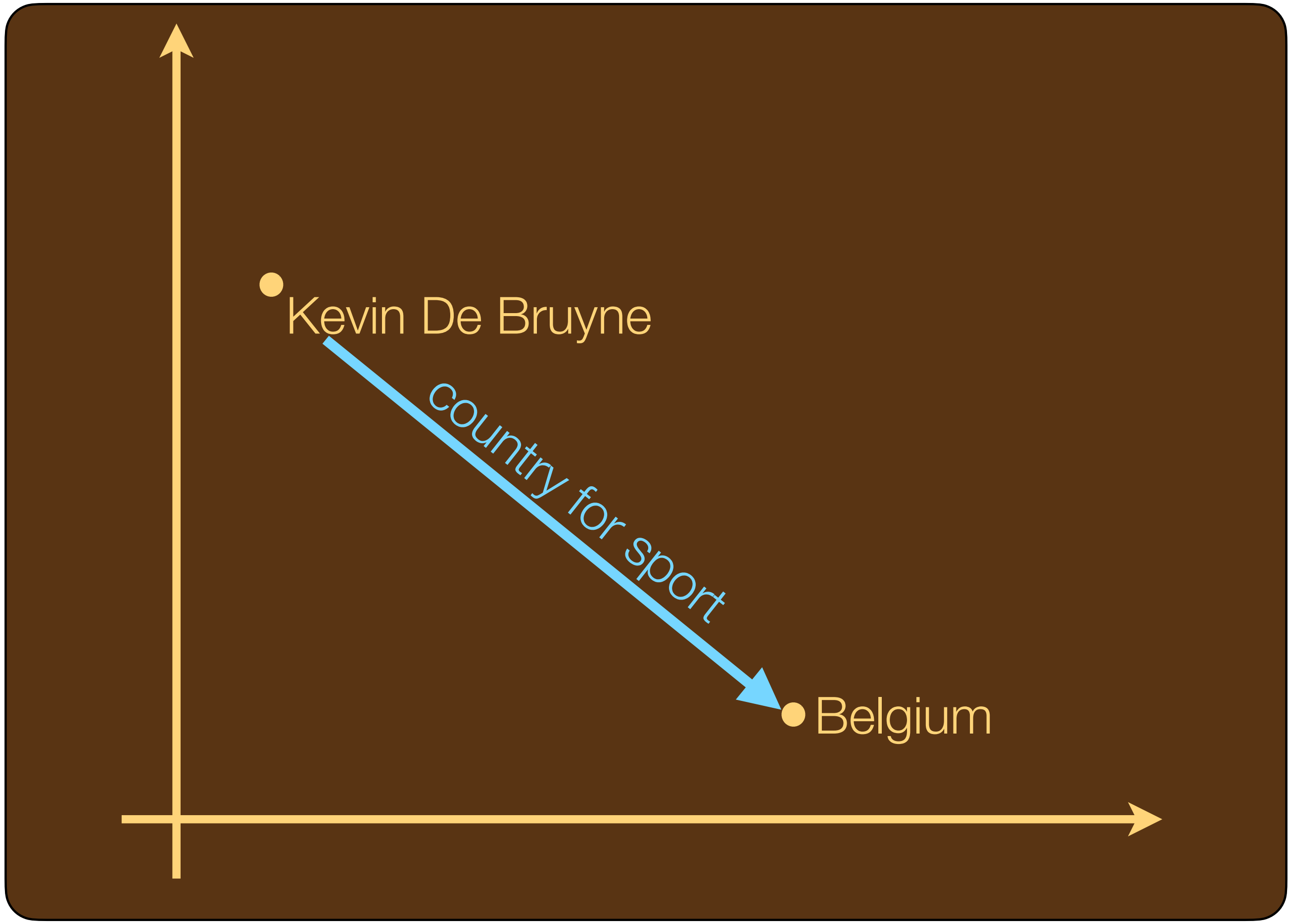
Neural Link Prediction

$$\begin{cases} f_r(\mathbf{h}, \mathbf{t}) > \lambda & \text{if } (h, r, t) \text{ is a valid triple} \\ f_r(\mathbf{h}, \mathbf{t}) < \lambda & \text{otherwise} \end{cases}$$

TransE (Bordes et al 2013)

Translation Intuition: For a triple (h, r, t), $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ if the given fact is true, else $\mathbf{h} + \mathbf{r} \neq \mathbf{t}$

Scoring function: $f_r(\mathbf{h}, \mathbf{t}) = -d(\mathbf{h} + \mathbf{r}, \mathbf{t})$

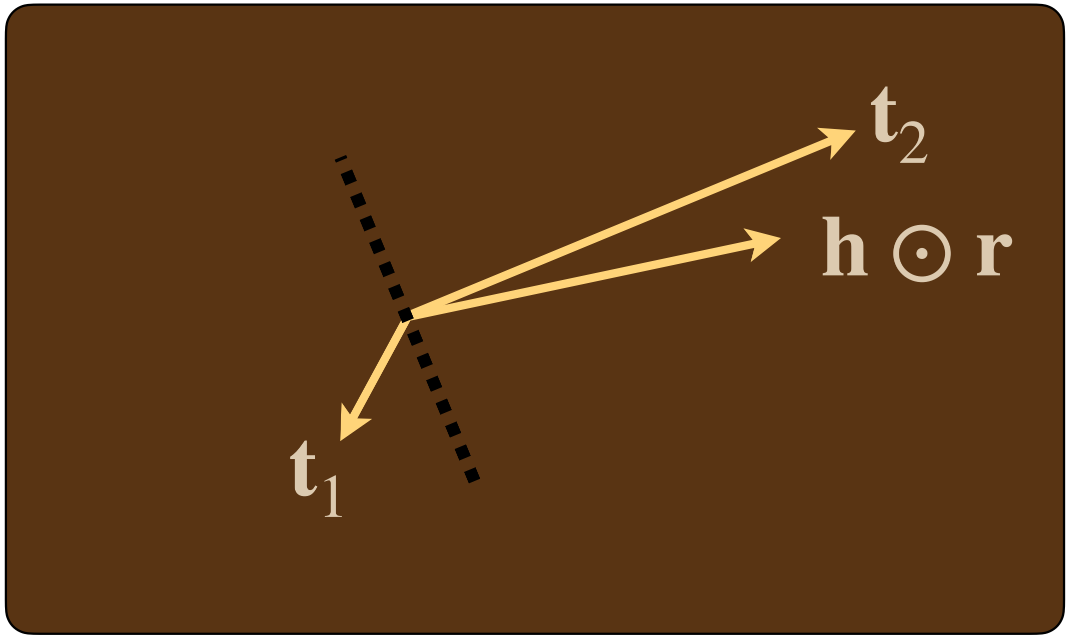


DistMult (Yang et al 2015)

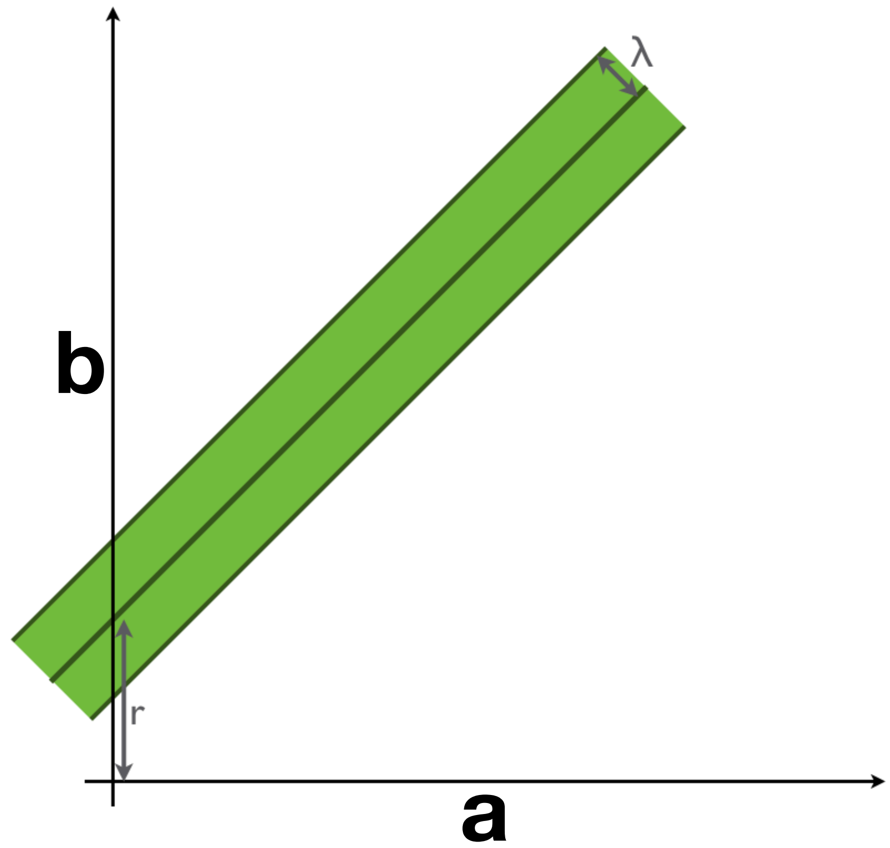
DistMult adopts **bilinear modeling**

$$f_r(\mathbf{h}, \mathbf{t}) = (\mathbf{h} \odot \mathbf{r}) \cdot \mathbf{t} = \sum_i h_i \cdot r_i \cdot t_i$$

Intuition: The score function can be seen as the similarity between $\mathbf{h} \odot \mathbf{r}$ and $\mathbf{t}$

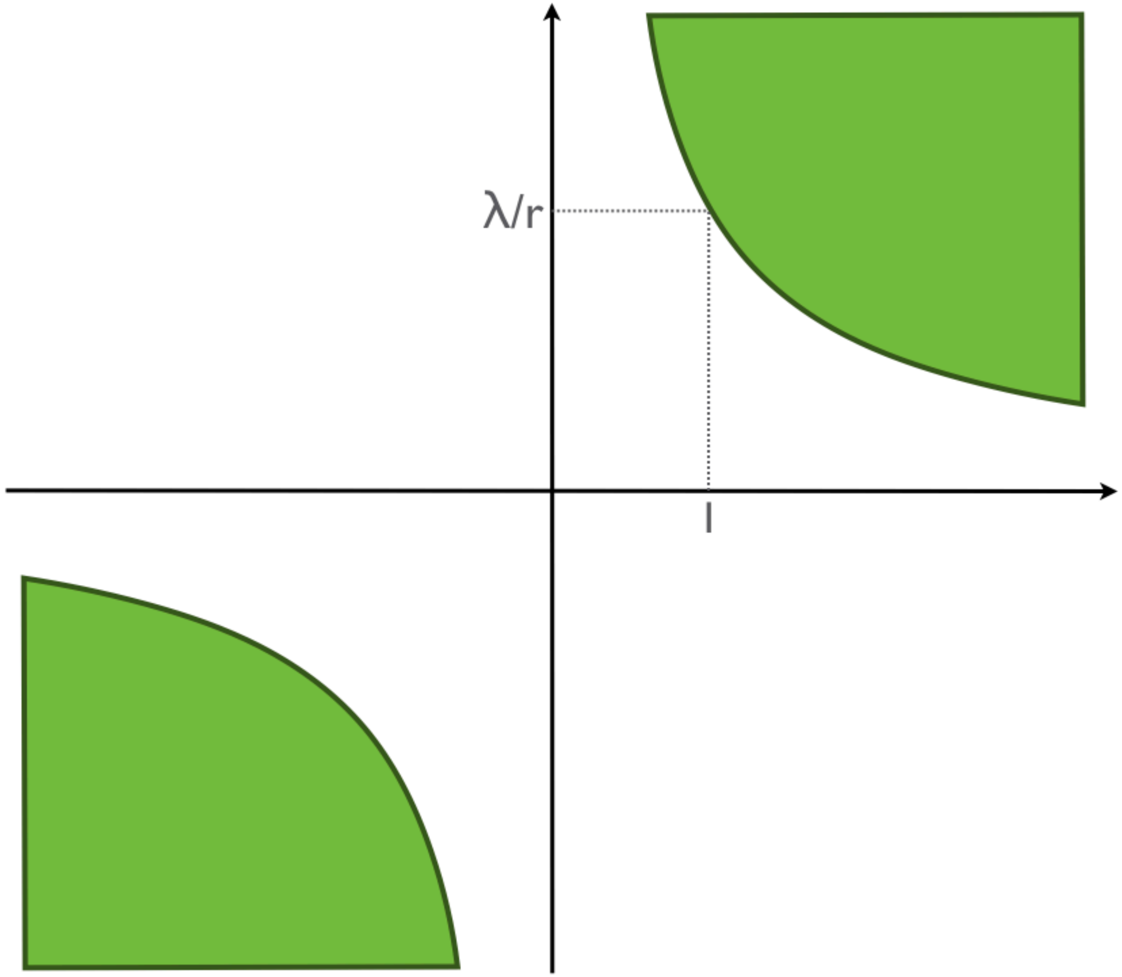


Region based view of relations: TransE



$$f_r(\mathbf{a}, \mathbf{b}) = -\, d(\mathbf{a} + \mathbf{r}, \mathbf{b})$$

Region based view of relations: DistMult

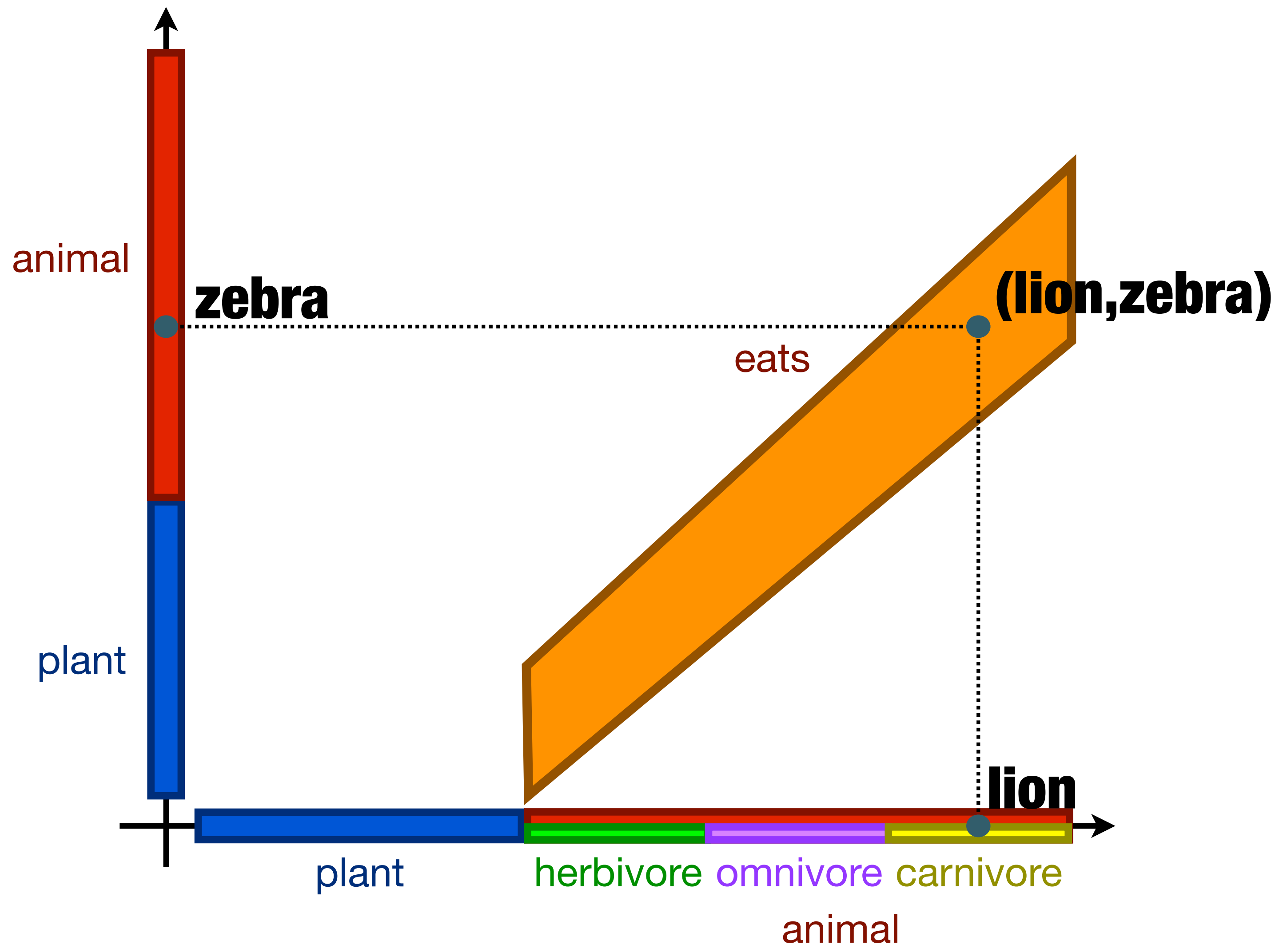


$$f_r(\mathbf{a}, \mathbf{b}) = \mathbf{a} \odot \mathbf{r} \odot \mathbf{b}$$

Modelling rules as spatial constraints

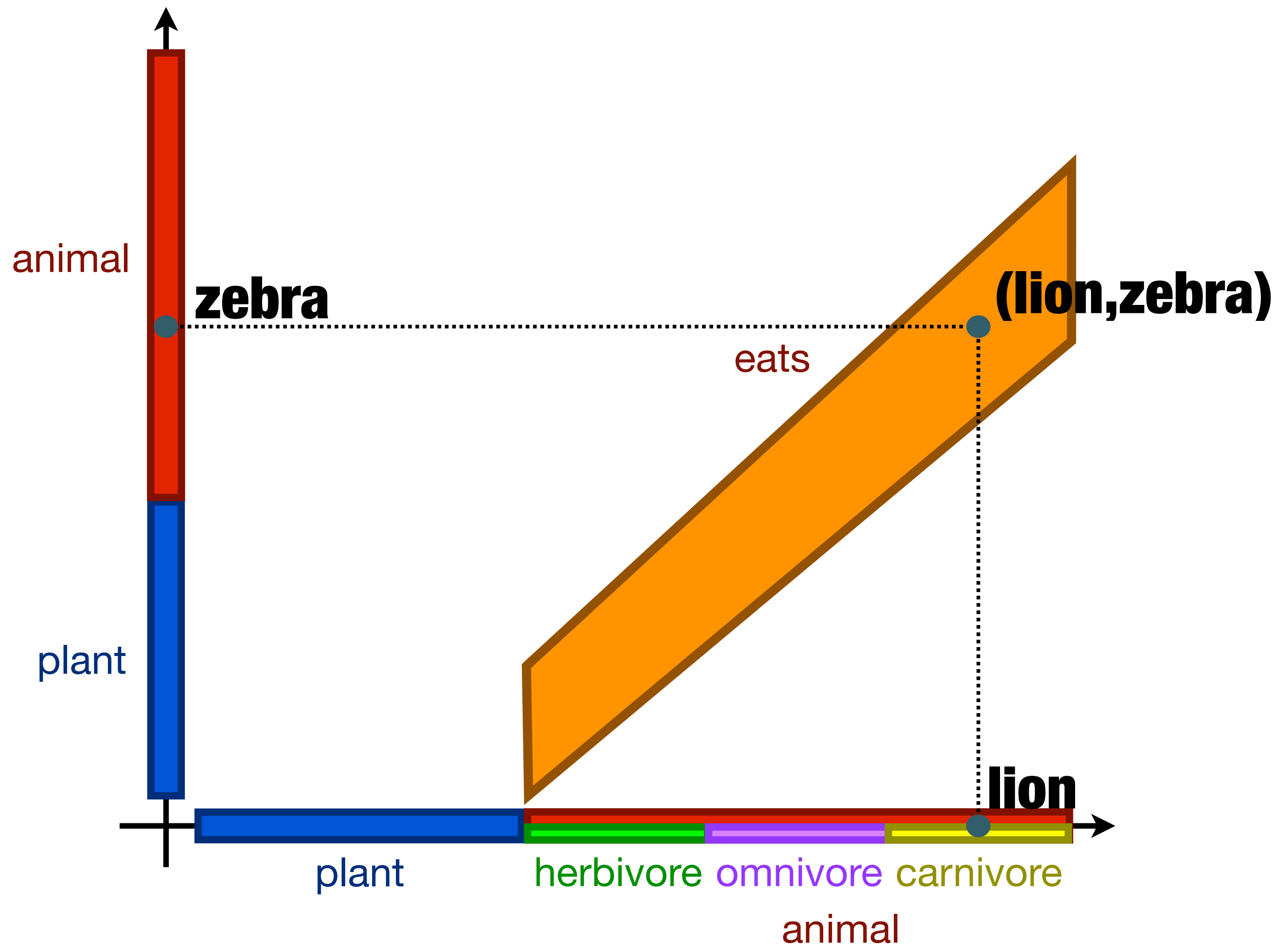
Victor Gutierrez Basulto, Steven Schockaert: From Knowledge Graph Embedding to Ontology Embedding? An Analysis of the Compatibility between Vector Space Representations and Rules. KR 2018

Modelling rules as spatial constraints



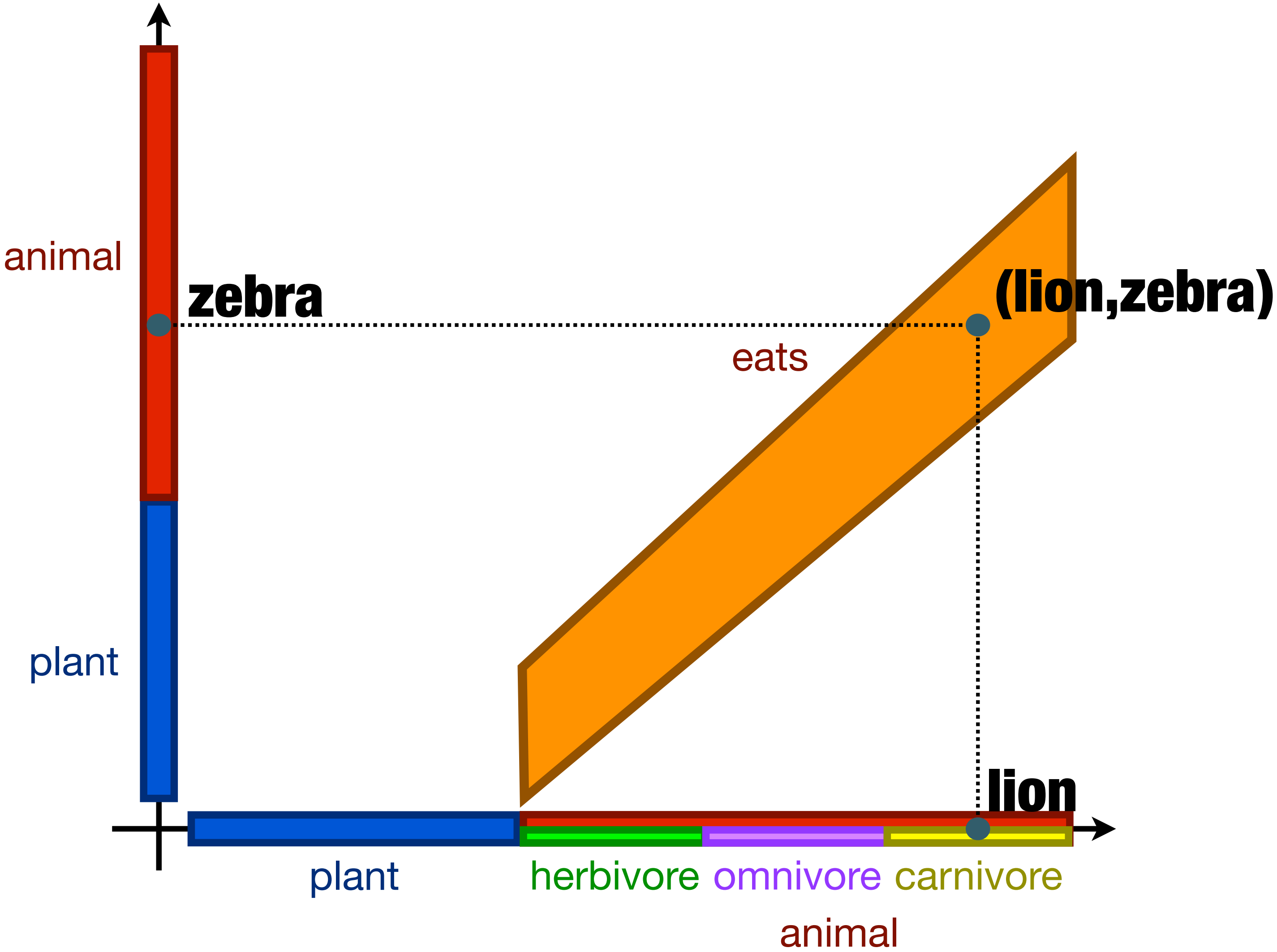
Victor Gutierrez Basulto, Steven Schockaert: From Knowledge Graph Embedding to Ontology Embedding? An Analysis of the Compatibility between Vector Space Representations and Rules. KR 2018

Modelling rules as spatial constraints



$$\text{Animal}(Y) \leftarrow \text{Carnivore}(X), \text{Eats}(X, Y)$$

Modelling rules as spatial constraints



$$\exists Y . \text{Eats}(X, Y) \wedge \text{Animal}(Y) \leftarrow \text{Carnivore}(X)$$

Modelling rules as spatial constraints

Each individual a is represented by a point $\eta(a) \in \mathbb{R}^n$

Each k-ary relation r is represented by a convex region $\eta(r) \subseteq \mathbb{R}^{k \cdot n}$

We refer to the mapping η as a **geometric interpretation**

Modelling rules as spatial constraints

Each individual a is represented by a point $\eta(a) \in \mathbb{R}^n$

Each k-ary relation r is represented by a convex region $\eta(r) \subseteq \mathbb{R}^{k \cdot n}$

We refer to the mapping η as a **geometric interpretation**

The relational fact $r(a_1, \ldots, a_k)$ is satisfied in a geometric interpretation η if:

$$\eta(a_1) \oplus \ldots \oplus \eta(a_k) \in \eta(r)$$

Modelling rules as spatial constraints

Now consider a rule of the following form

$$r(X_1, \ldots, X_k) \leftarrow s(X_1, \ldots, X_k)$$

This rule is satisfied by a geometric interpretation η if

$$\eta(s) \subseteq \eta(r)$$

Modelling rules as spatial constraints

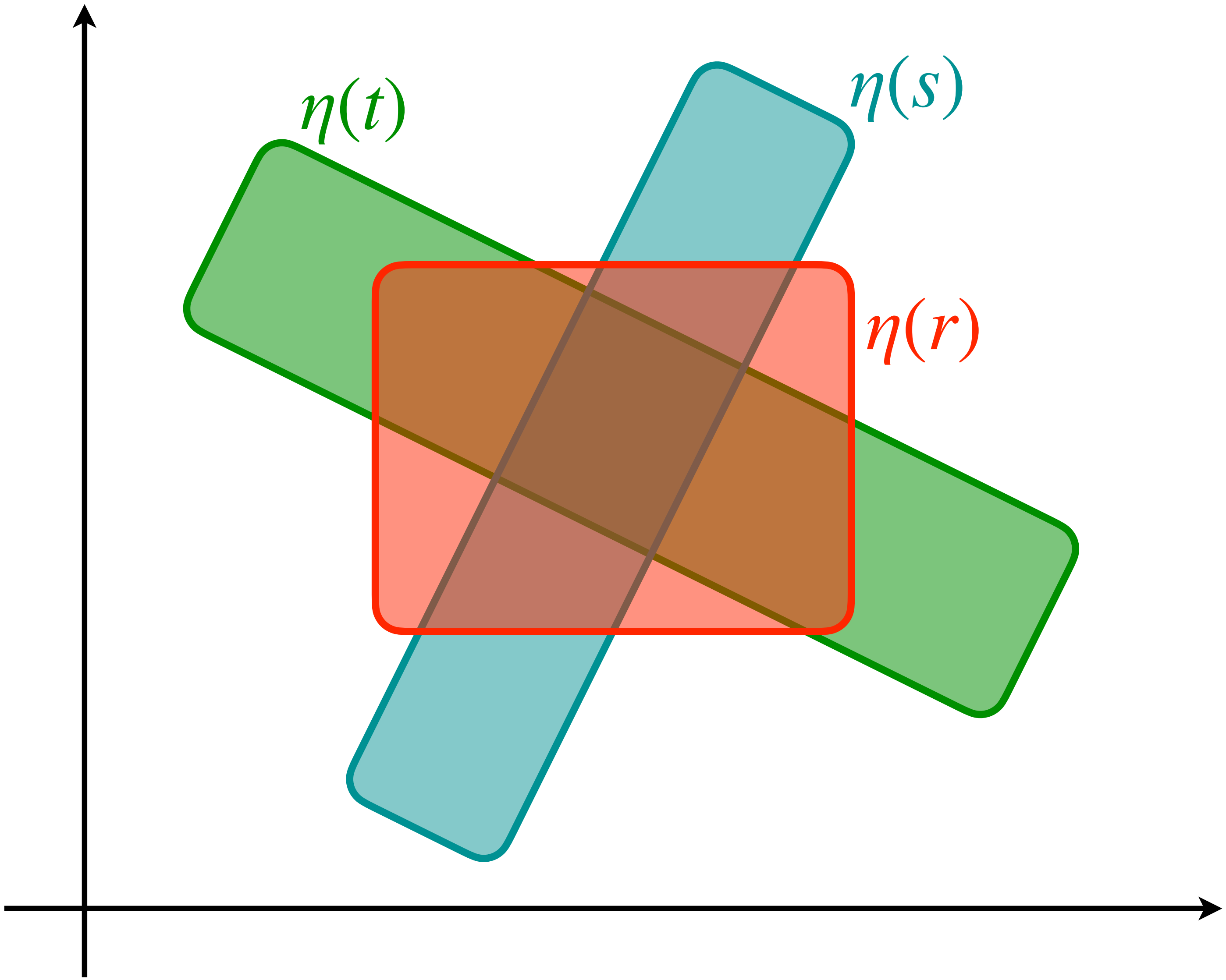
Now consider a rule of the following form

$$r(X_1, \ldots, X_k) \leftarrow s(X_1, \ldots, X_k), t(X_1, \ldots, X_k)$$

This rule is satisfied by a geometric interpretation η if

$$\eta(s) \cap \eta(t) \subseteq \eta(r)$$

Modelling rules as spatial constraints



$$r(X_1, X_2) \leftarrow s(X_1, X_2), t(X_1, X_2)$$

Modelling rules as spatial constraints

Now consider a rule of the following form

$$r(X_1, X_3) \leftarrow s(X_1, X_2), t(X_2, X_3)$$

Modelling rules as spatial constraints

Now consider a rule of the following form

$$r(X_1, X_3) \leftarrow s(X_1, X_2), t(X_2, X_3)$$

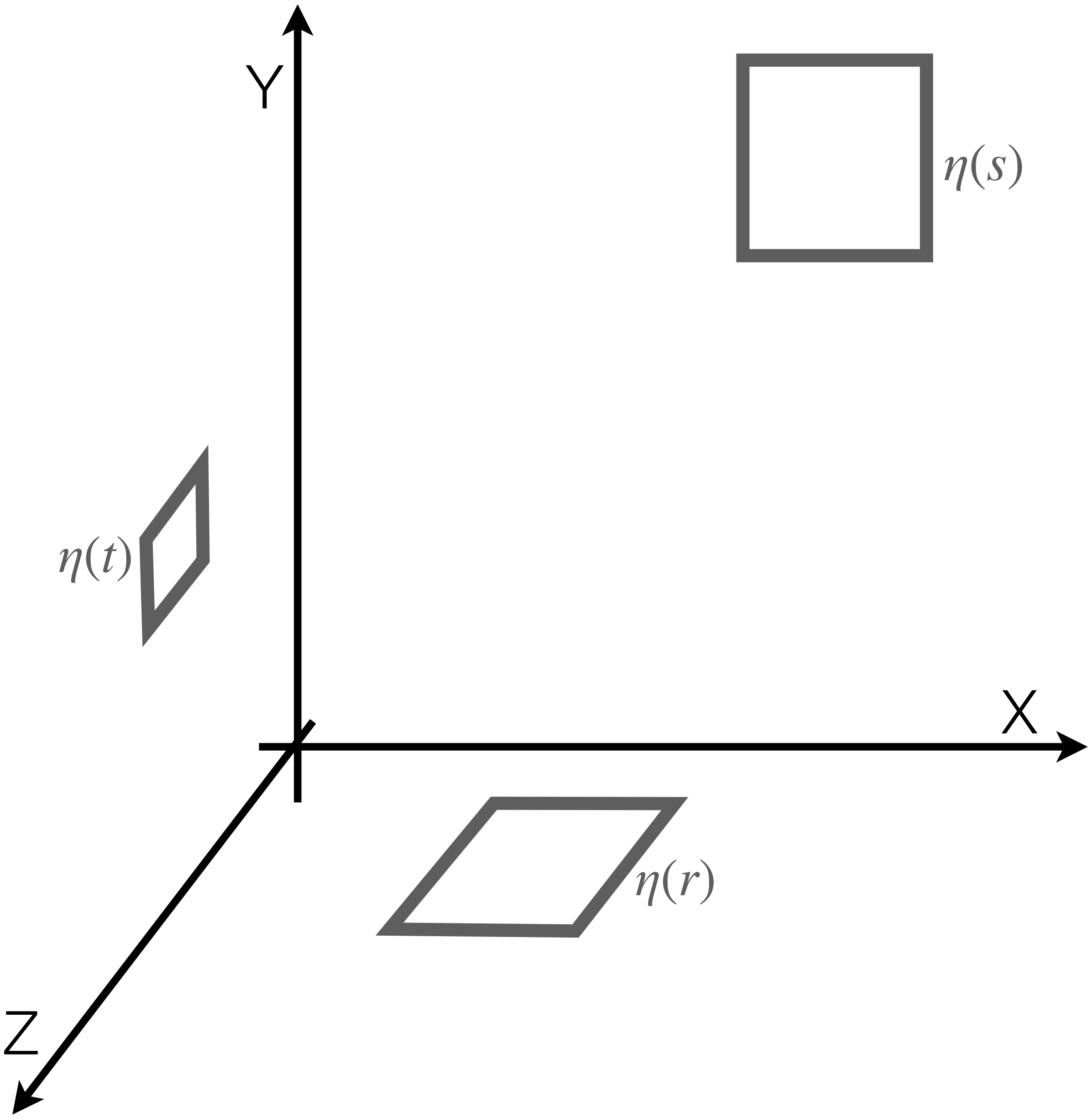
We can always view binary relations as ternary relations in which one argument is ignored

$$r^*(X, Y, Z) \equiv r(X, Z)$$
$$s^*(X, Y, Z) \equiv s(X, Y)$$
$$t^*(X, Y, Z) \equiv s(Y, Z)$$

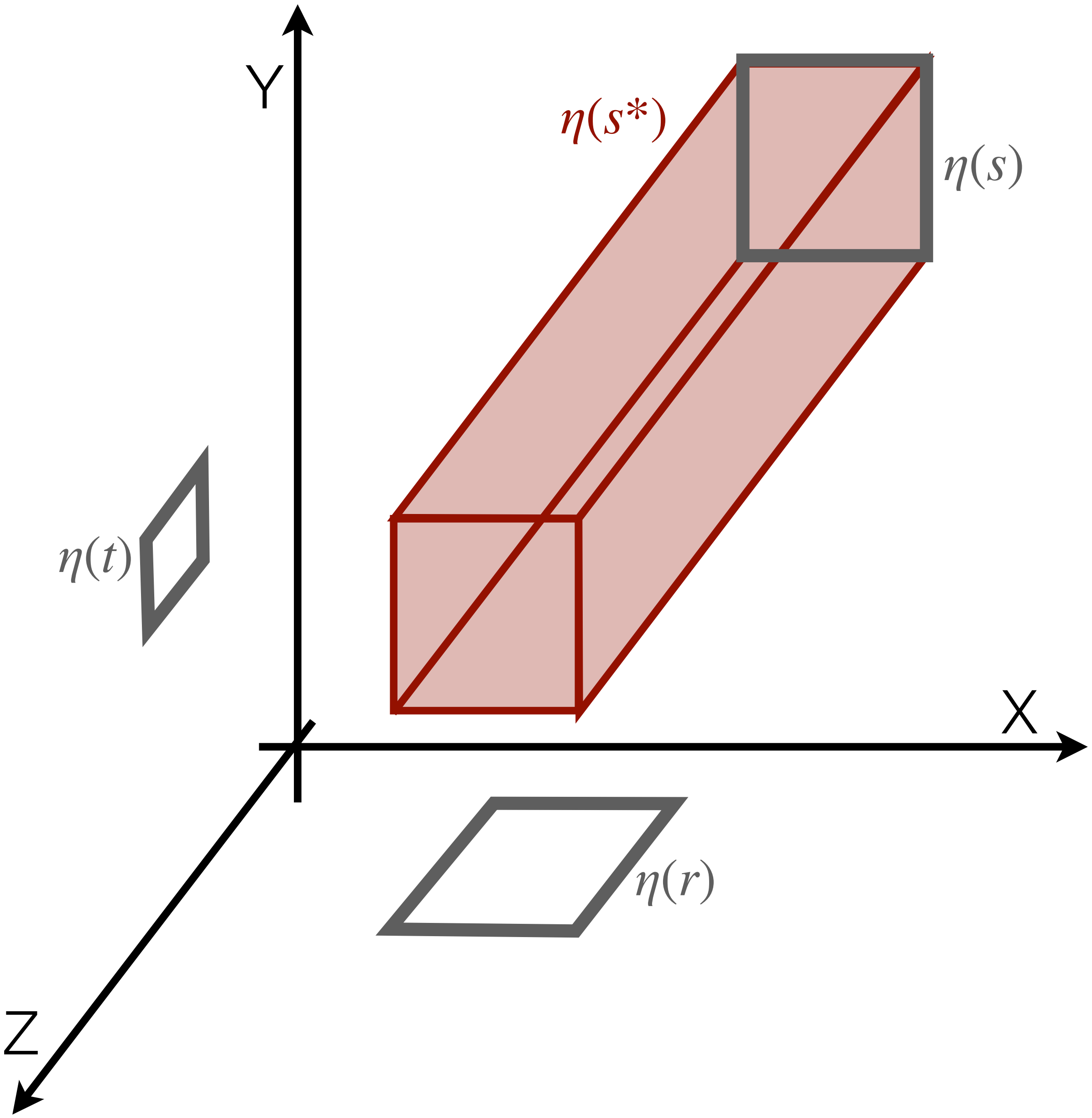
Leading to the following constraint:

$$\eta(s^*) \cap \eta(t^*) \subseteq \eta(r^*)$$

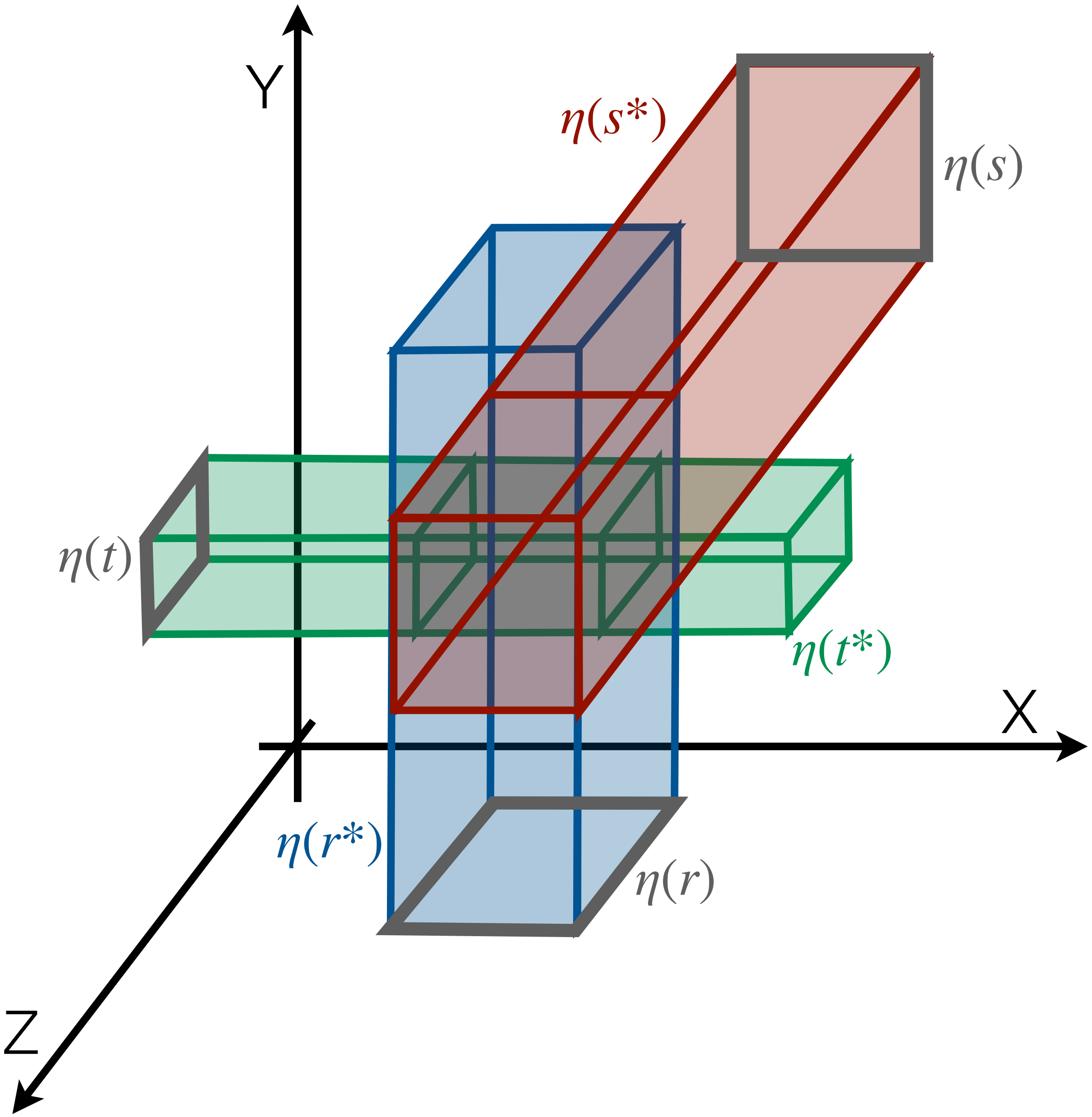
Modelling rules as spatial constraints



Modelling rules as spatial constraints



Modelling rules as spatial constraints



Modelling rules as spatial constraints

Let us now formally define the relationship between $\eta(r)$ and its extension $\eta(r^*)$, for a given relation r

Let $I \subseteq \{1,...,k\}$, then we define the restriction of a vector $(x_1, \ldots, x_{k \cdot n}) \in \mathbb{R}^{k \cdot n}$ to I as follows:

$$(x_1, \ldots, x_{k \cdot n}) \downarrow I = \bigoplus_{i \in I} (x_{n \cdot i+1}, \ldots, x_{n \cdot i+n})$$

Modelling rules as spatial constraints

Let us now formally define the relationship between $\eta(r)$ and its extension $\eta(r^*)$, for a given relation r

Let $I \subseteq \{1,...,k\}$, then we define the restriction of a vector $(x_1, \ldots, x_{k \cdot n}) \in \mathbb{R}^{k \cdot n}$ to I as follows:

$$(x_1, \ldots, x_{k \cdot n}) \downarrow I = \bigoplus_{i \in I} (x_{n \cdot i+1}, \ldots, x_{n \cdot i+n})$$

For instance, for $n = 2$, $k = 4$ and $I = \{1,4\}$ we have

$$(\underline{x_1, x_2}, \underline{x_3, x_4}, \underline{x_5, x_6}, \underline{x_7, x_8}) \downarrow \{1,4\} = (\underline{x_1, x_2}, \underline{x_7, x_8})$$

vector representing the first argument of a 4-ary relation

vector representing the last argument of a 4-ary relation

Modelling rules as spatial constraints

Intuitively, if $(x_1, \ldots, x_{k \cdot n})$ is the representation of a tuple $(a_1, \ldots, a_k)$ then $(x_1, \ldots, x_{k \cdot n}) \downarrow I$ is the representation of the tuple we obtain if we only keep the arguments at the positions that belong to I

Modelling rules as spatial constraints

Intuitively, if $(x_1, \ldots, x_{k\cdot n})$ is the representation of a tuple $(a_1, \ldots, a_k)$ then $(x_1, \ldots, x_{k\cdot n}) \downarrow I$ is the representation of the tuple we obtain if we only keep the arguments at the positions that belong to I

Let $R \subseteq \mathbb{R}^{l\cdot n}$ be a region, corresponding to the representation of some l-ary relation r. The cylindrical extension of R is given by:

$$\textsf{ext}_I^k(R) = \{\mathbf{x} \in \mathbb{R}^{k\cdot n} \,|\, \mathbf{x} \downarrow I \in R\}$$

Note how this cylindrical extension corresponds to the representation of a k-ary relation, which is defined in terms of the l-ary relation r, with the remaining arguments being ignored. The set of indices I determines which of the k arguments are non-trivial.

Modelling rules as spatial constraints

Consider again the following rule:

$$r(X_1, X_3) \leftarrow s(X_1, X_2), t(X_2, X_3)$$

This rule is satisfied in a geometric interpretation η if

$$\mathsf{ext}^3_{\{1,2\}}\big(\eta(s)\big) \cap \mathsf{ext}^3_{\{2,3\}}\big(\eta(t)\big) \subseteq \mathsf{ext}^3_{\{1,3\}}\big(\eta(r)\big)$$

Modelling rules as spatial constraints

We can similarly model existential rules:

$$\exists X_2 \,.\, r(X_1, X_2) \wedge s(X_2, X_3) \leftarrow t(X_1, X_3)$$

This rule is satisfied in a geometric interpretation η if

$$\eta(t) \subseteq \left(\mathsf{ext}^3_{\{1,2\}}\big(\eta(r)\big) \cap \mathsf{ext}^3_{\{2,3\}}\big(\eta(s)\big)\right) \downarrow \{1,3\}$$

Modelling rules as spatial constraints

We can similarly model existential rules:

$$\exists X_2 \,.\, r(X_1, X_2) \wedge s(X_2, X_3) \leftarrow t(X_1, X_3)$$

This rule is satisfied in a geometric interpretation η if

$$\eta(t) \subseteq \left(\mathsf{ext}^3_{\{1,2\}}\big(\eta(r)\big) \cap \mathsf{ext}^3_{\{2,3\}}\big(\eta(s)\big) \right) \downarrow \{1,3\}$$

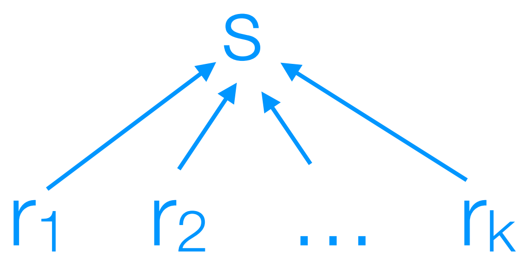
Can KG embeddings model arbitrary rules?

Consider a **bilinear model**, i.e.:

$$f_r(a, b) = \mathbf{a}^T \mathbf{M_r} \mathbf{b}$$

Suppose the following rules are modelled:

$$r_1(X, Y) \rightarrow s(X, Y)$$
$$\cdots$$
$$r_k(X, Y) \rightarrow s(X, Y)$$



Victor Gutierrez Basulto, Steven Schockaert: From Knowledge Graph Embedding to Ontology Embedding? An Analysis of the Compatibility between Vector Space Representations and Rules. KR 2018

Can KG embeddings model arbitrary rules?

Consider a **bilinear model**, i.e.:

$$f_r(a, b) = \mathbf{a}^T \mathbf{M_r} \mathbf{b}$$

Suppose the following rules are modelled:

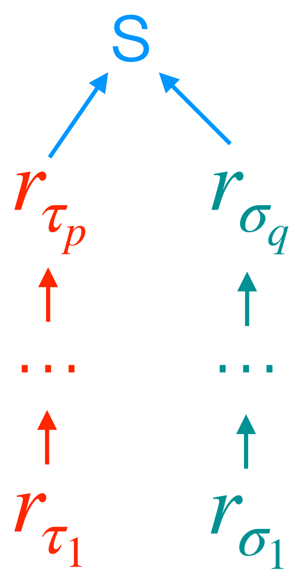
$$r_1(X, Y) \rightarrow s(X, Y)$$
$$\ldots$$
$$r_k(X, Y) \rightarrow s(X, Y)$$

Then there exists a permutation of the predicates:

$$\{r_1, \ldots, r_k\} = \{\underline{r_{\tau_1}, \ldots, r_{\tau_p}}, \underline{r_{\sigma_1}, \ldots, r_{\sigma_q}}\}$$

Such that: $\forall 1 \leq i < p \,.\, r_{\tau_i}(X, Y) \rightarrow r_{\tau_{i+1}}(X, Y)$

$\forall 1 \leq i < q \,.\, r_{\sigma_i}(X, Y) \rightarrow r_{\sigma_{i+1}}(X, Y)$



Victor Gutierrez Basulto, Steven Schockaert: From Knowledge Graph Embedding to Ontology Embedding? An Analysis of the Compatibility between Vector Space Representations and Rules. KR 2018

Can KG embeddings model arbitrary rules?

Consider a model in which relations can be modelled by arbitrary **convex polytopes**

Then all (sets of) rules of the following form (called **quasi-chained**) can be modelled

$$B_1 \wedge \ldots \wedge B_i \wedge \ldots \wedge B_n \rightarrow \exists X_1, \ldots, X_j . H_1 \wedge \ldots \wedge H_k$$

First-order atom which shares at most one variable with $B_1, \ldots, B_{i-1}$

Victor Gutierrez Basulto, Steven Schockaert: From Knowledge Graph Embedding to Ontology Embedding? An Analysis of the Compatibility between Vector Space Representations and Rules. KR 2018

Can KG embeddings model arbitrary rules?

Consider a model in which relations can be modelled by arbitrary **convex polytopes**

Such a model cannot model the following rule

$$\perp \leftarrow r_1(X, Y), r_2(X, Y)$$

together with the following facts:

$$\{r_1(a, a), r_1(b, b), r_2(a, b), r_2(b, a)\}$$

Victor Gutierrez Basulto, Steven Schockaert: From Knowledge Graph Embedding to Ontology Embedding? An Analysis of the Compatibility between Vector Space Representations and Rules. KR 2018

Can KG embeddings model arbitrary rules?

Indeed, if $\eta(r_1)$ and $\eta(r_2)$ are convex, and we have

$$\eta(a) \oplus \eta(a) \in \eta(r_1)$$

$$\eta(b) \oplus \eta(b) \in \eta(r_1)$$

$$\eta(a) \oplus \eta(b) \in \eta(r_2)$$

$$\eta(b) \oplus \eta(a) \in \eta(r_2)$$

Then we also have

$$\frac{(\eta(a) + \eta(b))}{2} \oplus \frac{(\eta(a) + \eta(b))}{2} \in \eta(r_1) \cap \eta(r_2)$$

Victor Gutierrez Basulto, Steven Schockaert: From Knowledge Graph Embedding to Ontology Embedding? An Analysis of the Compatibility between Vector Space Representations and Rules. KR 2018

Can KG embeddings model arbitrary rules?

Indeed, if $\eta(r_1)$ and $\eta(r_2)$ are convex, and we have

$$\eta(a) \oplus \eta(a) \in \eta(r_1)$$

$$\eta(b) \oplus \eta(b) \in \eta(r_1)$$

$$\eta(a) \oplus \eta(b) \in \eta(r_2)$$

$$\eta(b) \oplus \eta(a) \in \eta(r_2)$$

Then we also have

$$\frac{(\eta(a) + \eta(b))}{2} \oplus \frac{(\eta(a) + \eta(b))}{2} \in \eta(r_1) \cap \eta(r_2)$$

Victor Gutierrez Basulto, Steven Schockaert: From Knowledge Graph Embedding to Ontology Embedding? An Analysis of the Compatibility between Vector Space Representations and Rules. KR 2018

Summary on relational conceptual spaces

We can **model relational knowledge using convex regions**, similarly to conceptual spaces, by considering the Cartesian product of "standard" conceptual spaces

Existential rules can be viewed as **spatial constraints** over such representations

This makes it possible, in principle, to exploit given relational knowledge when learning an entity embedding, allowing us to generalise from a given ontology and knowledge graph in a principled way.

Open questions

Is there a **larger fragment of existential rules** that can be faithfully modelled in terms of geometric interpretations with convex regions?

Is there a way to **relax the convexity assumption** such that arbitrary existential rules can be captured, while keeping the representations simple enough to be learnable?

In practice, it is difficult to learn good representations when allowing arbitrary convex polytopes. Is it possible to find **interesting special cases** that can still capture a non-trivial fragment of existential rules, while being **easier to learn**?

Embeddings essentially correspond to a **single interpretation**. There is no obvious counterpart of a "knowledge base", as a set of possible interpretations.

Alternative approach

Consider the following propositional rules:

$$\text{mother} \leftarrow \text{female}, \text{parent}$$
$$\text{female} \leftarrow \text{mother}$$
$$\text{parent} \leftarrow \text{mother}$$

Under the conceptual spaces view, these rules correspond to the following constraint

$$\eta(\text{mother}) = \eta(\text{female}) \cap \eta(\text{parent})$$

Alternative approach

Consider the following propositional rules:

$$\text{mother} \leftarrow \text{female}, \text{parent}$$

$$\text{female} \leftarrow \text{mother}$$

$$\text{parent} \leftarrow \text{mother}$$

Under the conceptual spaces view, these rules correspond to the following constraint

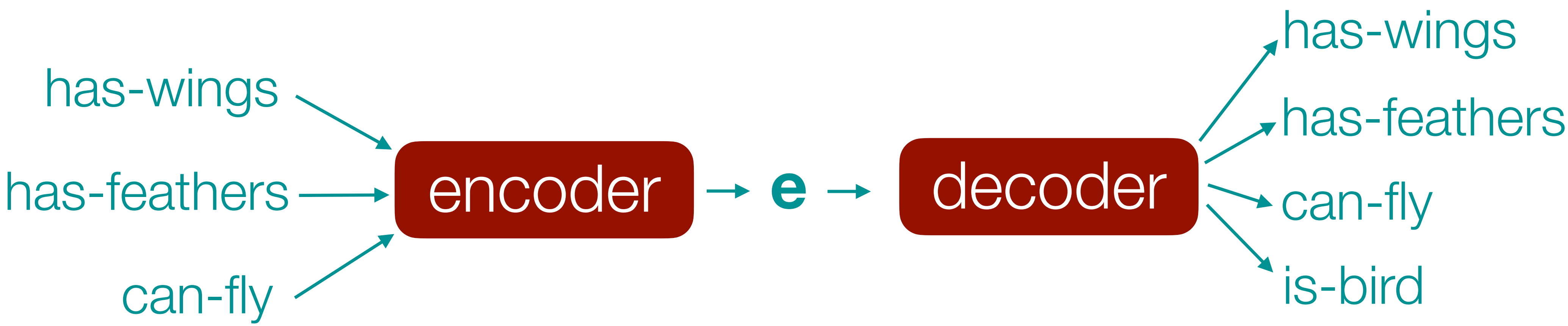
$$\eta(\text{mother}) = \eta(\text{female}) \cap \eta(\text{parent})$$

In practice, labels are usually predicted using vector dot products, e.g. we may assume

$$\eta(\text{mother}) = \{\mathbf{x} \in \mathbb{R}^n : \sigma(\mathbf{x} \cdot \mathbf{v}_{mother}) \geq 0.5\}$$

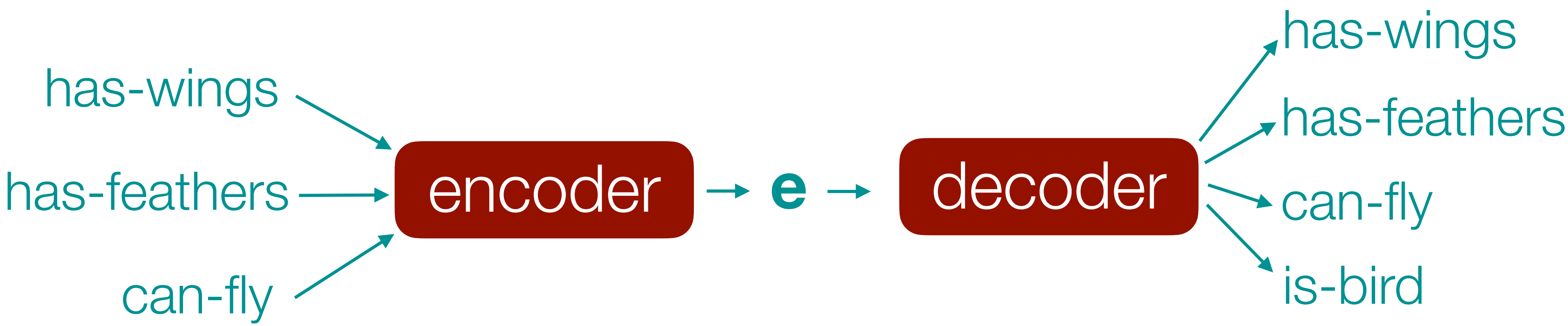
The above constraint cannot be modelled using such regions

An encoder-decoder view



Steven Schockaert: Modelling Monotonic and Non-Monotonic Attribute Dependencies with Embeddings: A Theoretical Analysis. AKBC 2021

An encoder-decoder view



has-wings ∧ has-feathers ∧ can-fly → is-bird

Steven Schockaert: Modelling Monotonic and Non-Monotonic Attribute Dependencies with Embeddings: A Theoretical Analysis. AKBC 2021

An encoder-decoder view

Assumption: entities are encodes by aggregating attribute vectors

$$\text{Emb}(a_1, \ldots, a_n) = \frac{1}{n}(\mathbf{a_1} + \ldots + \mathbf{a_n})$$

the attributes which entity e is known to satisfy

embedding of attribute a_n

Steven Schockaert: Modelling Monotonic and Non-Monotonic Attribute Dependencies with Embeddings: A Theoretical Analysis. AKBC 2021

An encoder-decoder view

Assumption: entities are encodes by aggregating attribute vectors

$$\text{Emb}(a_1, \ldots, a_n) = \frac{1}{n}(\mathbf{a_1} + \ldots + \mathbf{a_n})$$

$$\text{Emb}(a_1, \ldots, a_n) = \frac{\mathbf{a_1} + \ldots + \mathbf{a_n}}{\|\mathbf{a_1} + \ldots + \mathbf{a_n}\|}$$

$$\text{Emb}(a_1, \ldots, a_n) = \text{argmax}_{\mathbf{e}} \sum_{i=1}^{n} \log \sigma(\mathbf{e} \cdot \mathbf{a_i}) + \kappa \|\mathbf{e}\|^2$$

$$\text{Emb}(a_1, \ldots, a_n) = \max(\mathbf{a_1}, \ldots, \mathbf{a_n})$$

Steven Schockaert: Modelling Monotonic and Non-Monotonic Attribute Dependencies with Embeddings: A Theoretical Analysis. AKBC 2021

An encoder-decoder view

Assumption: labelling function depends on (possibly different) attribute embeddings

$$\text{Lab}(\mathbf{e}) = \{b \in \mathscr{A} \mid \mathbf{e} \cdot \tilde{\mathbf{b}} \geq \lambda_b\}$$

An encoder-decoder view

Assumption: labelling function depends on (possibly different) attribute embeddings

$$\text{Lab}(\mathbf{e}) = \{b \in \mathscr{A} \mid \mathbf{e} \cdot \tilde{\mathbf{b}} \geq \lambda_b\}$$

$$\text{Lab}(\mathbf{e}) = \{b \in \mathscr{A} \mid d(\mathbf{e}, \tilde{\mathbf{b}}) \leq \theta_b\}$$

$$\text{Lab}(\mathbf{e}) = \{b \in \mathscr{A} \mid \text{ReLU}(\mathbf{e}) \cdot \mathbf{b} \geq 0\}$$

$$\text{Lab}(\mathbf{e}) = \{b \in \mathscr{A} \mid \mathbf{b} \preceq \mathbf{e}\}$$

An encoder-decoder view

$Emb(a_1, ..., a_n)$	$Lab(\mathbf{e})$	**Monotonic**	**Non-mon.**
$\frac{1}{n}\sum_i \mathbf{a_i}$	$\{b \mid \mathbf{e} \cdot \tilde{\mathbf{b}} \geq \lambda_b\}$	✗	✗
$\frac{1}{n}\sum_i \mathbf{a_i}$	$\{b \mid d(\mathbf{e}, \tilde{\mathbf{b}}) \leq \theta_b\}$	✗	✗
$\frac{\sum_i \mathbf{a_i}}{\lVert\sum_i \mathbf{a_i}\rVert}$	$\{b \mid \mathbf{e} \cdot \tilde{\mathbf{b}} \geq \lambda_b\}$	✗	✗
$\frac{\sum_i \mathbf{a_i}}{\lVert\sum_i \mathbf{a_i}\rVert}$	$\{b \mid d(\mathbf{e}, \tilde{\mathbf{b}}) \leq \theta_b\}$	✗	✗
$\arg\max_{\mathbf{e}} \sum_i \log \sigma(\mathbf{e} \cdot \mathbf{a_i}) + \kappa \lVert\mathbf{e}\rVert^2$	$\{b \mid \mathbf{e} \cdot \tilde{\mathbf{b}} \geq \lambda_b\}$	✗	✗
$\arg\max_{\mathbf{e}} \sum_i \log \sigma(\mathbf{e} \cdot \mathbf{a_i}) + \kappa \lVert\mathbf{e}\rVert^2$	$\{b \mid d(\mathbf{e}, \tilde{\mathbf{b}}) \leq \theta_b\}$	✗	✗
$\frac{1}{n}\sum_i \mathbf{a_i}$	$\{b \mid \mathrm{ReLU}(\mathbf{e}) \cdot \mathbf{b} \geq 0\}$	✓	✓
$\mathbf{a_1} \odot ... \odot \mathbf{a_n}$	$\{b \mid \mathbf{e} \cdot \tilde{\mathbf{b}} \geq 0\}$	✓	✓
$\mathbf{a_1} \odot ... \odot \mathbf{a_n}$	$\{b \mid \mathbf{e} \cdot \mathbf{b} \geq 0\}$	✗	✗
$\max(\mathbf{a_1}, ..., \mathbf{a_n})$	$\{b \mid \mathbf{b} \preceq \mathbf{e}\}$	✓	✗

Steven Schockaert: Modelling Monotonic and Non-Monotonic Attribute Dependencies with Embeddings: A Theoretical Analysis. AKBC 2021

An encoder-decoder view

$Emb(a_1, ..., a_n)$	$Lab(\mathbf{e})$	**Monotonic**	**Non-mon.**
$\frac{1}{n}\sum_i \mathbf{a_i}$	$\{b \mid \mathbf{e} \cdot \tilde{\mathbf{b}} \geq \lambda_b\}$	✗	✗
$\frac{1}{n}\sum_i \mathbf{a_i}$	$\{b \mid d(\mathbf{e}, \tilde{\mathbf{b}}) \leq \theta_b\}$	✗	✗
$\frac{\sum_i \mathbf{a_i}}{\lVert\sum_i \mathbf{a_i}\rVert}$	$\{b \mid \mathbf{e} \cdot \tilde{\mathbf{b}} \geq \lambda_b\}$	✗	✗
$\frac{\sum_i \mathbf{a_i}}{\lVert\sum_i \mathbf{a_i}\rVert}$	$\{b \mid d(\mathbf{e}, \tilde{\mathbf{b}}) \leq \theta_b\}$	✗	✗
$\arg\max_{\mathbf{e}} \sum_i \log \sigma(\mathbf{e} \cdot \mathbf{a_i}) + \kappa\lVert\mathbf{e}\rVert^2$	$\{b \mid \mathbf{e} \cdot \tilde{\mathbf{b}} \geq \lambda_b\}$	✗	✗
$\arg\max_{\mathbf{e}} \sum_i \log \sigma(\mathbf{e} \cdot \mathbf{a_i}) + \kappa\lVert\mathbf{e}\rVert^2$	$\{b \mid d(\mathbf{e}, \tilde{\mathbf{b}}) \leq \theta_b\}$	✗	✗
$\frac{1}{n}\sum_i \mathbf{a_i}$	$\{b \mid \text{ReLU}(\mathbf{e}) \cdot \mathbf{b} \geq 0\}$	✓	✓
$\mathbf{a_1} \odot ... \odot \mathbf{a_n}$	$\{b \mid \mathbf{e} \cdot \tilde{\mathbf{b}} \geq 0\}$	✓	✓
$\mathbf{a_1} \odot ... \odot \mathbf{a_n}$	$\{b \mid \mathbf{e} \cdot \mathbf{b} \geq 0\}$	✗	✗
$\max(\mathbf{a_1}, ..., \mathbf{a_n})$	$\{b \mid \mathbf{b} \preceq \mathbf{e}\}$	✓	✗

Steven Schockaert: Modelling Monotonic and Non-Monotonic Attribute Dependencies with Embeddings: A Theoretical Analysis. AKBC 2021

Summary on encoder-decoder view

The encoder-decoder view offers an alternative way to integrate rules and vectors, which is less demanding than conceptual spaces:

- **Conceptual spaces**: every point in the space corresponds to a model of the rule base
- **Encoder-decoder model**: every point that can be generated using the encoder corresponds to a model of the rule base

The limitations identified for the encoder-decoder view also apply to Graph Neural Network based approaches to inductive knowledge graph completion

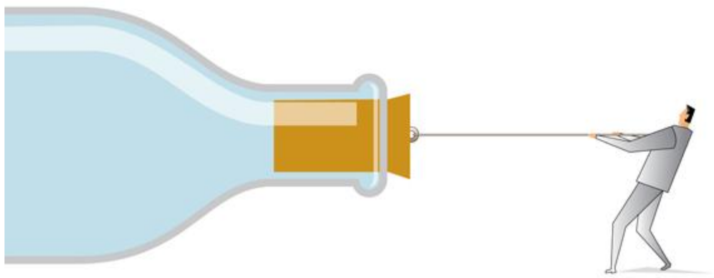
Learning conceptual space representations

Relational conceptual spaces

Using vectors for plausible reasoning over symbolic knowledge

Plausible Reasoning?

Knowledge acquisition bottleneck



- Rules and ontologies manually curated
- Experts are not good programmers
- Costs crowdsourcing

Knowledge bases are inevitably incomplete

Objective

Equip KR-symbolic systems with **inductive capabilities** using vectors

Develop formalisms incorporating **knowledge from vectors** to infer plausible **concept inclusions (rules)**

In a **principled way!**

Inductive reasoning

Tomatoes contain vitamin B6

Mushrooms contain vitamin B6

Carrots contain vitamin B6

Inductive reasoning

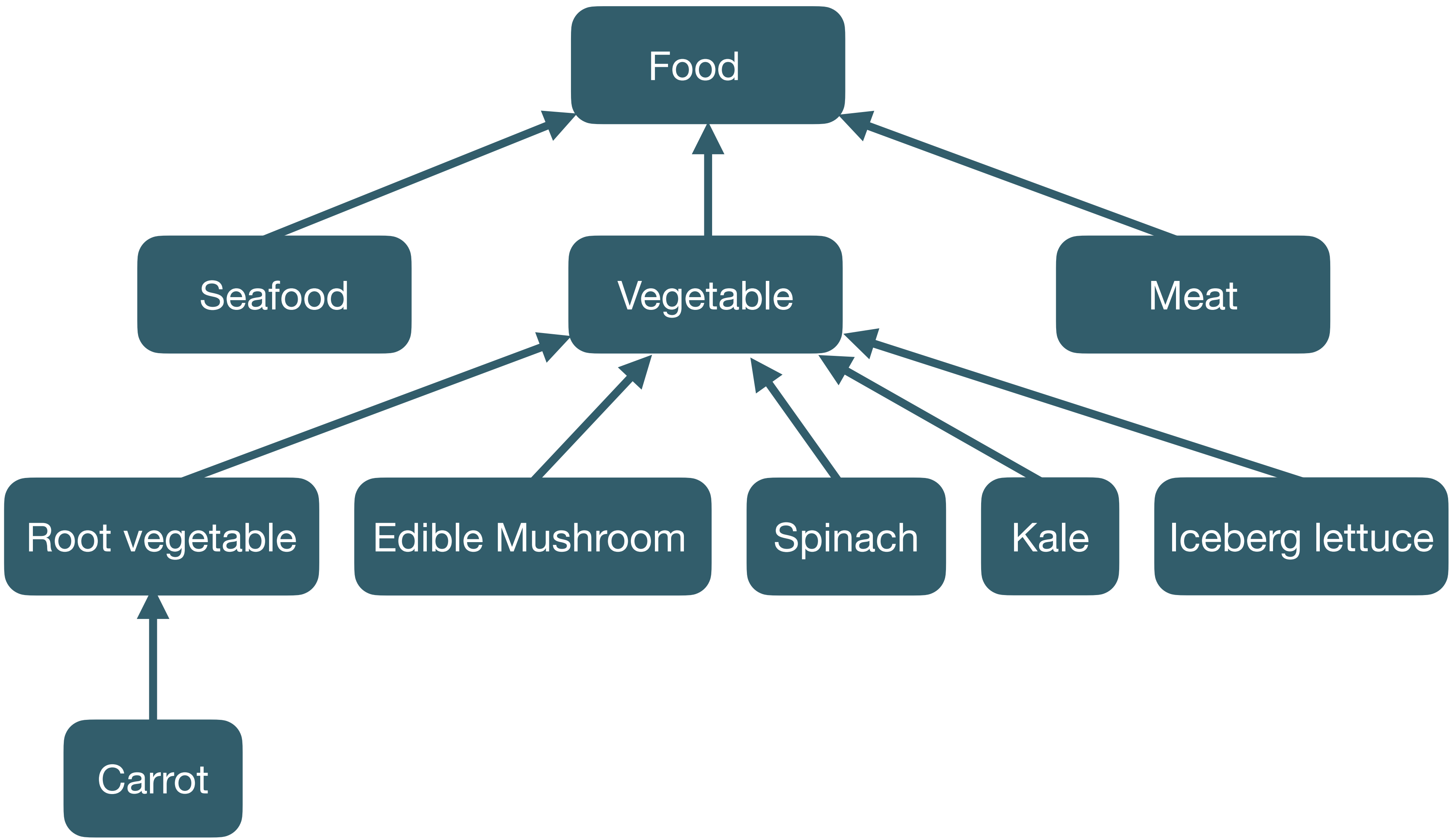
Tomatoes contain vitamin B6
Mushrooms contain vitamin B6

Carrots contain vitamin B6

Kale contains vitamin B6
Spinach contains vitamin B6

Carrots contain vitamin B6 ???

Taxonomies are too coarse-grained

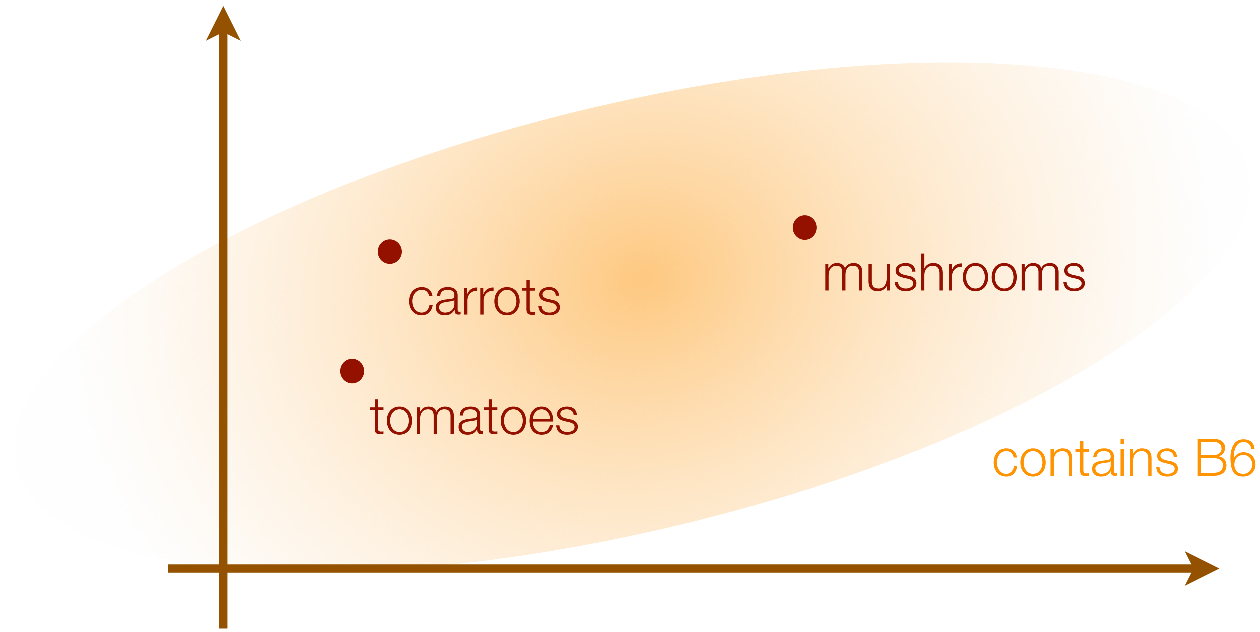


source: wikidata

Inductive reasoning

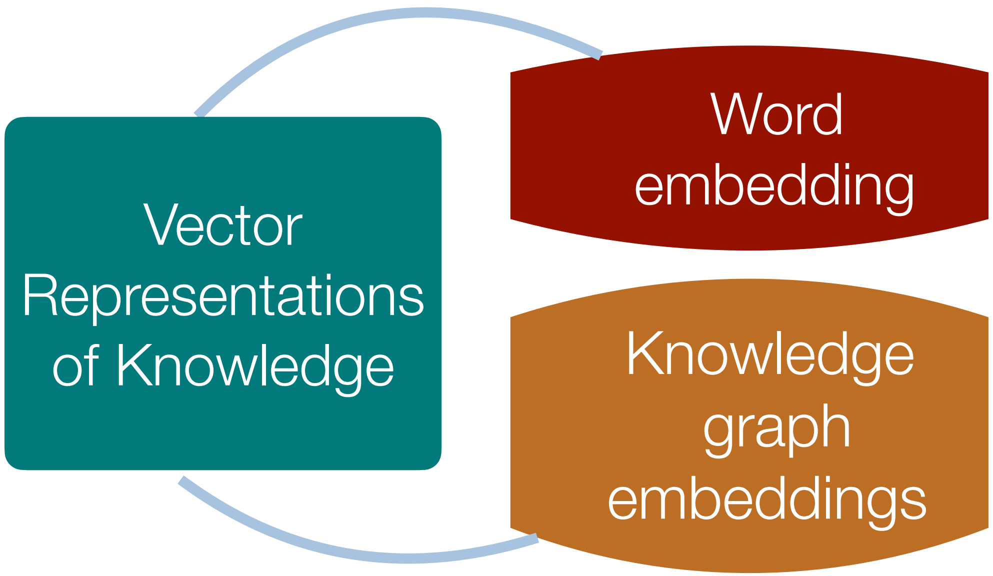
Tomatoes contain vitamin B6
Mushrooms contain vitamin B6

Carrots contain vitamin B6



§

- Several data-driven approaches have been proposed for automatically extending ontologies.



- exploit Wikidata, Freebase, BabelNet
- encode information about the similarity between different concepts
- but no other dependencies: subsumption, existential restrictions (unlike e.g some ontology languages)

Deductive and Inductive Reasoning

Exploit **rules and other symbolic KR approaches** for learning higher quality vector space representations

Use vector representations to **infer missing knowledge**

- knowledge graph triples (Neelakantan et al. 2015; Xie et al. 2016)
- ABox assertions (Rizzo et al. 2013; Bouraoui and Schockaert 2018)
- **concept inclusions** (Li, Bouraoui, and Schockaert 2019)

Our proposal

An inference mechanism based on a clear **model-theoretic semantics**

- Inference of plausible concepts inclusions

Formalisation of some form of **inductive reasoning** in description logic ontologies

- integration between the deductive and inductive inferences

Computational complexity bounds for reasoning (subsumption) in the description logic EL

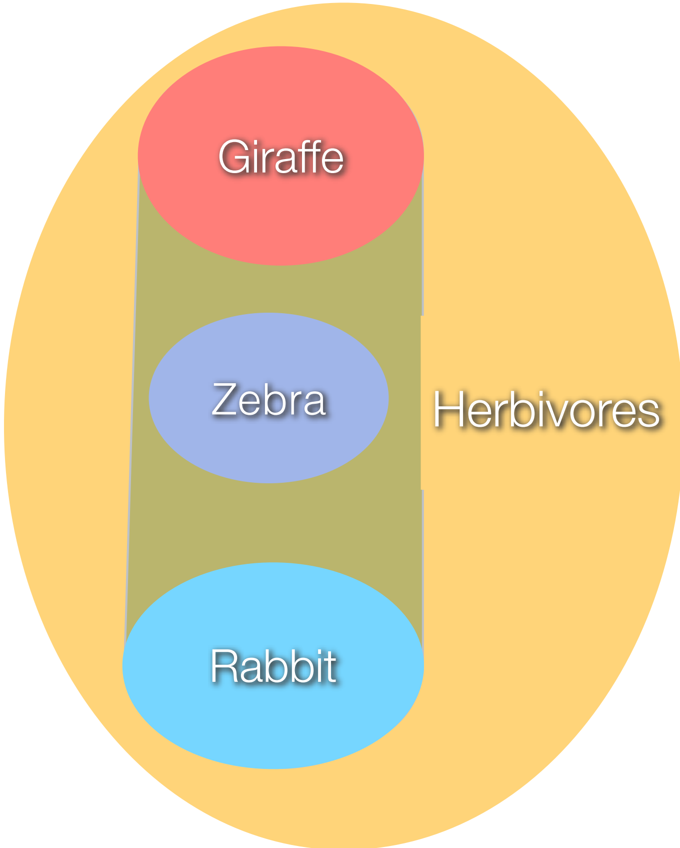
What kind of Inductive Reasoning?

INTERPOLATION: cognitive models of **category-based induction.**

Natural properties (concepts)

- correspond to **convex regions** in a suitable vector space

Conceptual betweenness: C is conceptually between A and B if it has all the natural properties of A and B



EL Description Logic

Horn description logic

$$C, D := \top \mid A \mid C \sqcap D \mid \exists r.C \mid \bot$$

Young $\sqcap$ Cat $\sqsubseteq$ Cute
Adult $\sqcap$ WildCat $\sqsubseteq$ Dangerous
Young $\sqcap$ Dog $\sqsubseteq$ Cute
WildCat $\sqsubseteq$ $\exists$eats.Meat

Core of **huge medical ontologies**, e.g. SNOMED CT

Reasoning is **tractable (PTIME)**, e.g. concept subsumption

Extending EL description logic

$$C, D := \top \mid A \mid C \sqcap D \mid \exists r\,.\,C \mid N$$

$$N, N' := A' \mid N \sqcap N' \mid N \bowtie N'$$

- A' belongs to a distinguished set of natural concept names
- Knowledge about conceptual betweenness $N \bowtie N'$ can be **obtained from vector representations**

Extending EL description logic

$$\frac{\begin{array}{r}\text{Rabbit} \sqsubseteq \text{Herbivore}\\ \text{Giraffe} \sqsubseteq \text{Herbivore}\\ \text{Zebra} \sqsubseteq \text{Rabbit} \bowtie \text{Giraffe}\\ \text{Herbivore} \sqsubseteq \exists \text{eats} \,.\, \text{Plant}\end{array}}{\text{Zebra} \sqsubseteq \text{Hervbivore}}$$

$$\frac{\begin{array}{l}A \sqcap C \sqsubseteq B\\ A \sqcap D \sqsubseteq B\end{array}}{A \sqcap (C \bowtie D) \sqsubseteq B}$$

Herbivore is a natural concept name

Non-interference

From data we can **only** learn betweenness information about **concept names**

$$\frac{\text{Nectarine} \sqsubseteq \text{Plum} \bowtie \text{Peach}}{\text{Sweet} \sqcap \text{Nectarine} \sqsubseteq (\text{Sweet} \sqcap \text{Plum}) \bowtie (\text{Sweet} \sqcap \text{Nectarine})}$$

$\text{Sweet} \boxdot (\text{Plump}, \text{Peach})$

Formalisation of underlying semantics

Which concepts are natural?

How to interpret conceptual betweenness?

- **Feature-enriched interpretations** (related to Formal Concept Analysis)
- **Geometric interpretations** (related to Conceptual Spaces)

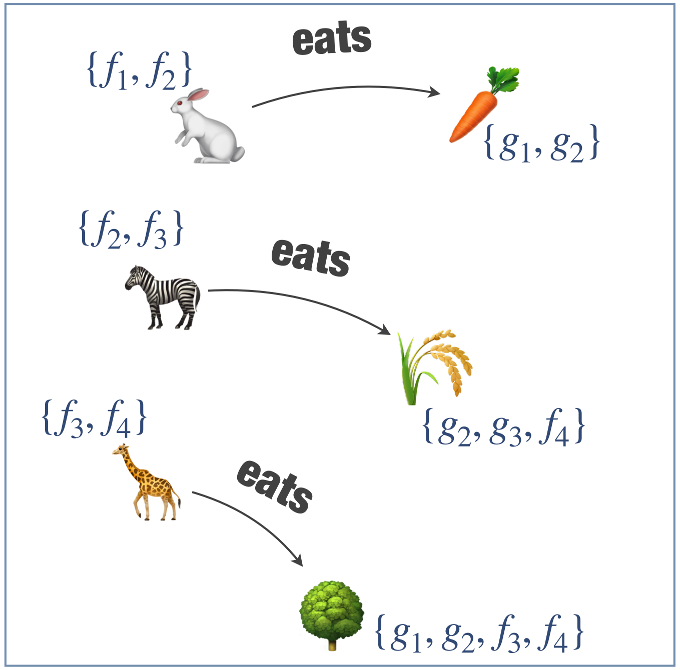
Formalisation of underlying semantics

Feature-enriched interpretation

- A classical DL interpretation + finite set of features

- C is **natural** if it is completely characterised by its set of features
  i.e., d is an instance of C iff the features of C are contained in the features of d

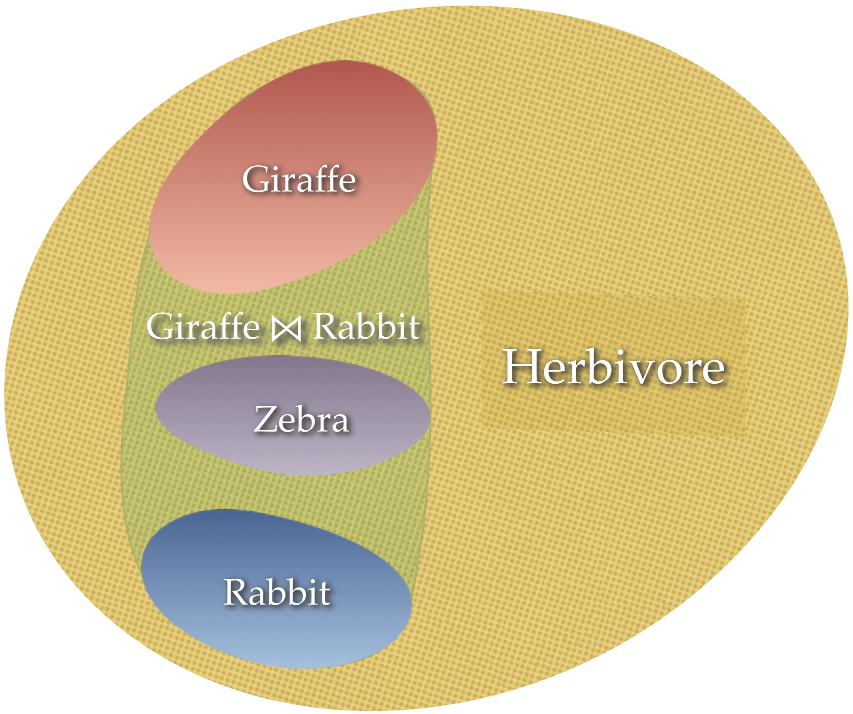
- $A \bowtie B$ is the concept characterised by the intersection of the features of A and B
  - $A \bowtie B$ is a natural concept



Formalisation of underlying semantics

Geometric Interpretation

- Concepts are interpreted as **regions from a vector space**
- A concept is **natural** if the region interpreting it is convex
- $A \bowtie B$ is interpreted as the **convex hull of the** union the regions interpreting A and B



Complexity

We look at the problem of concept subsumption

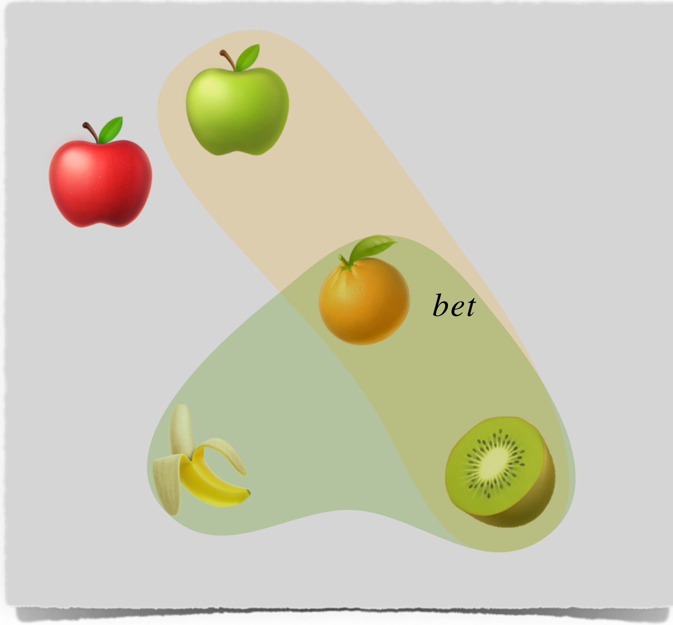
$$\mathcal{T} \models C \sqsubseteq D$$

- coNP-complete under the feature semantics
- PSpace-hard under the geometric semantics

Harder than in pure EL! :(

A unifying approach

Betweenness Semantics
An interpretation consists of a classical DL interpretation $\mathscr{I}$ and a **ternary relation** $bet \subseteq \Delta^{\mathscr{I}} \times \Delta^{\mathscr{I}} \times \Delta^{\mathscr{I}}$



- C ⋈ D contains **all individuals that between instances** of C and D

- If we impose certain properties to bet (e.g. symmetry), we can see the feature-based semantics as a special case.

Defeasible Reasoning

We focused on entailment, but conceptual betweenness is **learnt from data, so noisy**

Carrot is between Lettuce and Courgette

$$0.5 : \text{Carrot} \sqsubseteq \text{Lettuce} \bowtie \text{Courgette}$$
$$0.75 : \text{Carrot} \sqsubseteq \text{Lettuce} \bowtie \text{Courgette} \bowtie \text{Tomato}$$
$$1 : \text{Lettuce} \sqsubseteq \text{Green}$$
$$1 : \text{Courgette} \sqsubseteq \text{Green}$$
$$1 : \text{Carrot} \sqsubseteq \text{Orange}$$

It is straightforward to add a **defeasible mechanism**, e.g. a possibilistic extension

Interpolation vs Similarity

Adding similarity (e.g. probabilities, weights) to DLs seems straightforward. Why concept betweenness?

Challenge: How do we relate **similarity of instances** to **plausible inferences.**

Given that $\mathcal{T} \models C \sqsubseteq D$, **how similar** E needs to be to infer that $E \sqsubseteq D$?

Analogical Reasoning

In AI analogical reasoning builds on **analogical proportions:**

A is to B what C is to D

Man is to King what Woman is to Queen

Note: Analogical knowledge can be learnt from data GPT3 or matrix factorization

Develop a formalism that uses (learnt) analogies allowing to **extrapolate and translate**

Analogical Reasoning

Rule extrapolation

$$\frac{\begin{array}{c}\text{Young} \sqcap \text{Cat} \sqsubseteq \text{Cute} \\ \text{Adult} \sqcap \text{WildCat} \sqsubseteq \text{Dangerous} \\ \text{Young} \sqcap \text{Dog} \sqsubseteq \text{Cute} \\ \text{Cat} : \text{WildCat} :: \text{Dog} : \text{Wolf}\end{array}}{\text{Adult} \sqcap \text{Wolf} \sqsubseteq \text{Dangerous}}$$

Analogical Reasoning

Rule translation

$$\frac{\text{Program} \sqsubseteq \exists \text{specifies} . \text{Software} \quad \text{Program} : \text{Plan} :: \text{Software} : \text{Building}}{\text{Plan} \sqsubseteq \exists \text{specifies} . \text{Building}}$$

Transfer knowledge from one domain to another

Extending EL with Analogies

$EL^{\bowtie}$ + **analogical assertions** of the following form

$$A \triangleright B :: C \triangleright D$$

Feature-based semantics

A classical DL interpretation $\mathcal{I}=(\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$

A mapping π from from individuals to features from a set $\mathcal{F} = \mathcal{F}_1 \cup \ldots \cup \mathcal{F}_k$, where $\mathcal{F}_i$ is viewed as a domain

A equivalence relation ~ on {1,…,k} indicating which domains are equivalent

For each pair $(i,j) \in \sim$, a bijection $\sigma_{i,j}$ between $\mathcal{F}_i$ and $\mathcal{F}_j$, satisfying $\sigma_{i,j}^{-1}=\sigma_{j,i}$ and $\sigma_{i,j} \circ \sigma_{j,k} = \sigma_{i,k}$

Notable Properties

Lifting analogies

$$\frac{A_1 \triangleright B_1 :: C_1 \triangleright D_1 \qquad A_2 \triangleright B_2 :: C_2 \triangleright D_2}{(A_1 \sqcap A_2) \triangleright (B_1 \sqcap B_2) :: (C_1 \sqcap C_2) \triangleright (D_1 \sqcap D_2)}$$

$$\frac{A \triangleright B :: C \triangleright D}{(\exists r . A) \triangleright (\exists r . B) :: (\exists r . C) \triangleright (\exists r . D)}$$

Extrapolation and translation patterns

Open questions: interpolation

- Beyond conceptual betweenness

if burglary $(L, T-2)$, burglary $(L, T-1)$ then burglary (L, T)

if burglary $(L, T-1)$, burglary $(L_1, T-1)$, burglary $(L_2, T-1)$, n(L, L_1), $n(L, L_2)$, $L_1 \neq L_2$ then burglary (L, T)

if burglary$(L, T-2)$, burglary $(L_1, T-1)$, burglary $(L_2, T-1)$, n(L, L_1), n(L, L_2), $L_1 \neq L_2$ then burglary (L, T)

if burglary $(L_1, T-2)$, burglary $(L_2, T-2)$, burglary $(L, T-1)$, n(L, L_1), n(L, L_2), $L_1 \neq L_2$ then burglary(L, T)

- Better understanding of the complexity of the formalisms under the **bet** semantics

Open questions: analogies

What is the **exact computational complexity** of reasoning with analogies?

Is there a **simpler semantics**?

A **semantics driven by an application**?

Other **rule-like formalisms**?

Summary on plausible reasoning in DLs

We **provided model-theoretical semantics**, that allows to capture conceptual betweenness on concepts such that the interpolation pattern is sound

Explored different alternatives varying in complexity

Initiated the study of **analogical reasoning**

We believe these are valuable first steps towards effectively using knowledge from vector embeddings to enable ontology-based systems with inductive capabilities

Conclusions

There has been a lot of work on identifying plausible missing triples in knowledge graphs using embeddings

Similarly, we may try to **identify plausible missing generic knowledge** in ontologies

However, if we want to **tightly integrate** "knowledge completion" with deductive reasoning, we need a principled mechanism

One possible answer is to **extend conceptual spaces**, modelling relations as regions in high-dimensional spaces and viewing rules as qualitative spatial constraints between these regions

Another possibility is to abstract away from actual conceptual space representations and develop a calculus for **reasoning about incomplete qualitative constraints on conceptual space representations** (e.g. rules and betweenness assertions)