Exploring Semantic Capacity of Terms

Jie Huang, Zilong Wang, Kevin Chen-Chuan Chang, Wen-mei Hwu, Jinjun Xiong

University of Illinois at Urbana-Champaign
University of California at San Diego
IBM Thomas J. Watson Research Center
IBM-Illinois Center for Cognitive Computing Systems Research (C3SR)

Semantic Capacity

Artificial Intelligence

What is the **Semantic Capacity** SC(·) of "Artificial Intelligence"?

- SC(Artificial Intelligence) < SC(Computer Science)
- SC(Artificial Intelligence) > SC(Machine Learning)
- SC(Artificial Intelligence) > SC(Greedy Algorithm)

Semantic Capacity

Research Profiling

Engineering & Materials Science

Neural networks · Deep learning · Object recognition · Unsupervised learning · Labels · Character recognition · Pixels · Classifiers

Backpropagation · Invariance · Convolution · Network architecture · Color · Learning algorithms · Robots · Machine learning

Mathematics

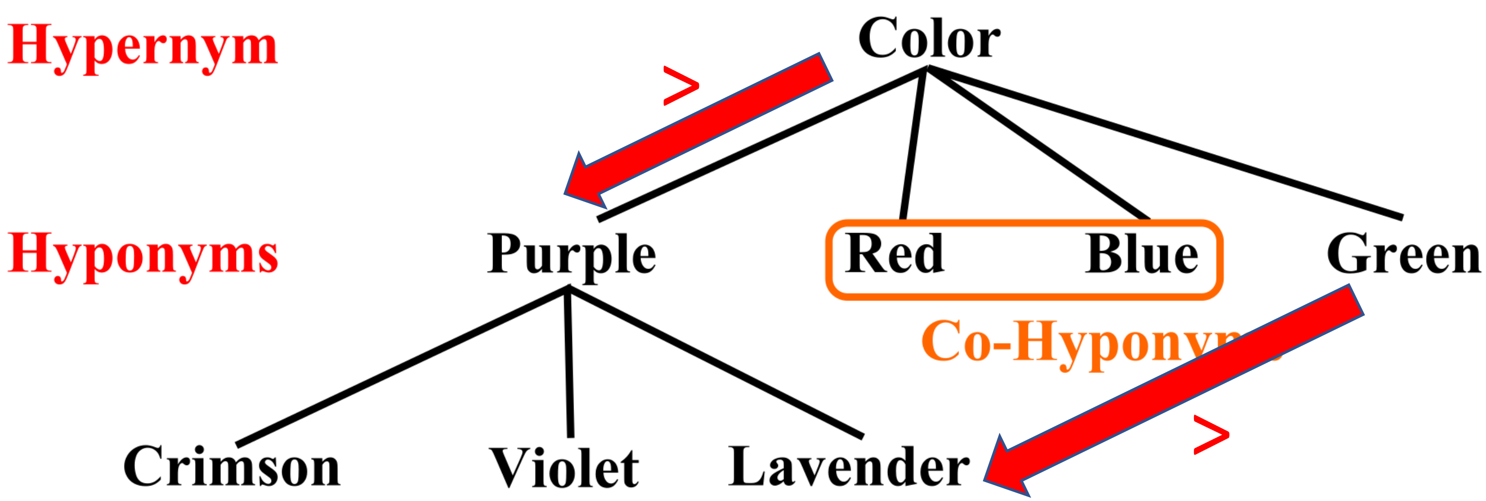
Learning · Segmentation · Training · Neural Networks · Pose Estimation · Loss Function · Multilayer · Energy

Discriminative Training · Object Classification · Machine Learning · Model · Labeling · Scale Invariant Feature Transform

https://www.elsevier.com/solutions/elsevier-fingerprint-engine

Semantic Capacity

Hypernym-Hyponym Discovery



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

Semantic Capacity?

- If we can find all hypernym-hyponym pairs -> **tree**
  => semantic capacity can be solved to some extent

- However…
  - Hearst patterns (Hearst, COLING'1992, with extended patterns) can only find **2.5%** (35/1393) pairs
    => impossible to measure semantic capacity of terms

Observation

Artificial Intelligence associates with**:**

1) many terms, e.g., AI terms
2) broad terms, e.g., CS, CV, ML, …

   r(AI, ML) > r(AI, SVM)

Semantic Capacity Association Hypothesis:

Terms with higher semantic capacity associate with

1) ***more*** terms
2) terms with ***higher*** semantic capacity than lower ones

Normalized Pointwise Mutual Information

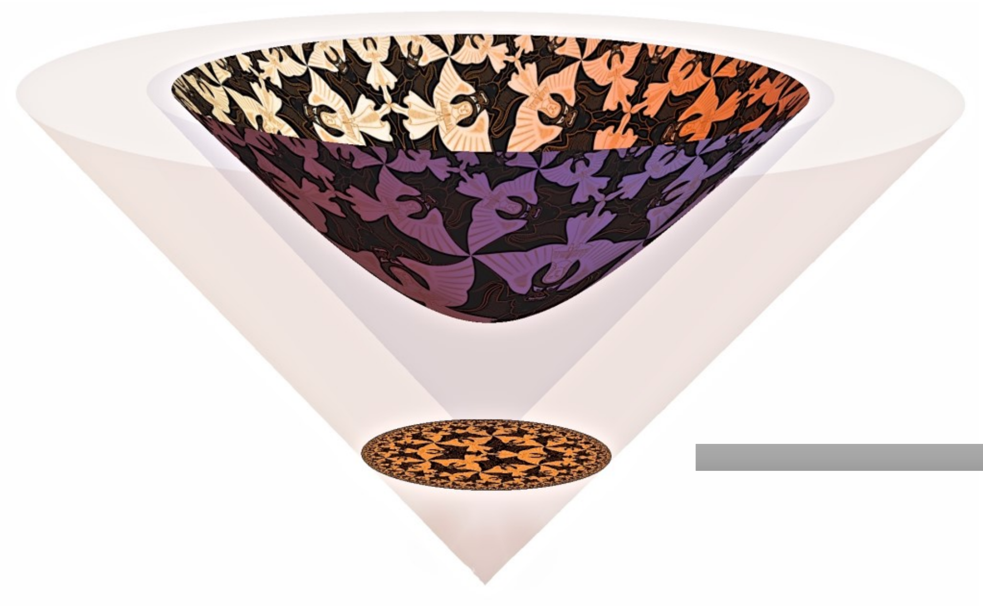
$$npmi(x,y) = -\log \frac{p(x,y)}{p(x)p(y)} / \log p(x,y)$$

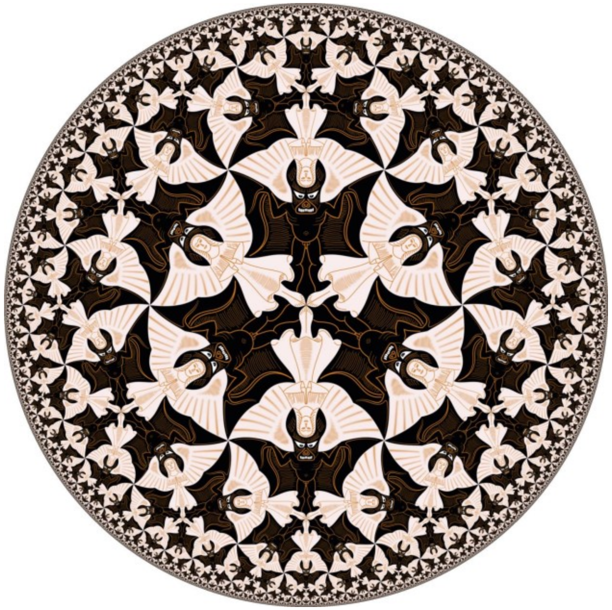
Range from -1 to 1:

- -1: never co-occur
- 0: occur independently
- 1: co-occur completely

→ **Association**

Hyperbolic Geometry



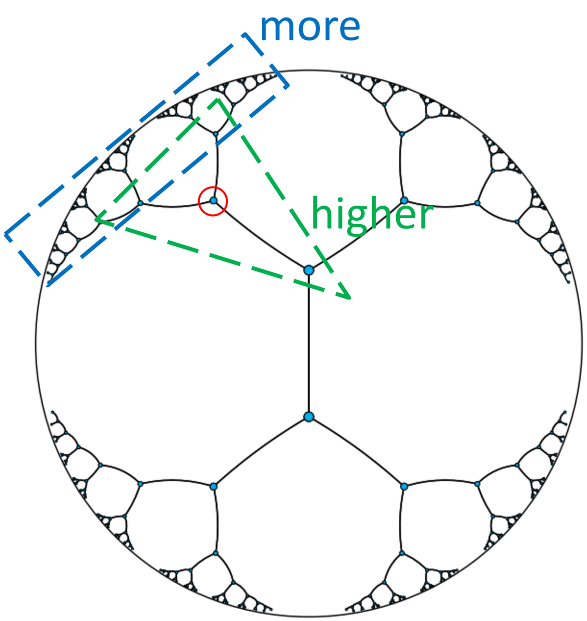


Poincaré disk

http://inspirehep.net/record/1355197/plots

Poincaré Embedding

$$d_p(\mathbf{x}, \mathbf{y}) = \operatorname{arcosh}\left(1 + 2\frac{\|\mathbf{x} - \mathbf{y}\|^2}{(1 - \|\mathbf{x}\|^2)(1 - \|\mathbf{y}\|^2)}\right)$$



- The distance of points increases exponentially as they are closer to the boundary

Why Hyperbolic Space?

- Volumes grow exponentially with radius
- Number of terms grows exponentially as semantic capacity gets lower

Maximillian Nickel and Douwe Kiela. 2017. Poincaré embeddings for learning hierarchical representations. In NIPS. 6338–6347.

Lorentz Model

- An equivalent model for hyperbolic space:
  - Perform Riemannian optimization more efficiently
  - Distance function avoids numerical instabilities

- Poincaré -> Lorentz

$$\ell(x_1, \ldots, x_n) = \frac{(1 + \|\mathbf{x}\|^2, 2x_1, \ldots, 2x_n)}{1 - \|\mathbf{x}\|^2}$$

- Lorentz -> Poincaré

$$\ell^{-1}(x_0, x_1, \ldots, x_n) = \frac{(x_1, \ldots, x_n)}{x_0 + 1}$$

Maximillian Nickel and Douwe Kiela. 2018. Learning Continuous Hierarchies in the Lorentz Model of Hyperbolic Geometry. In ICML. 3776–3785.

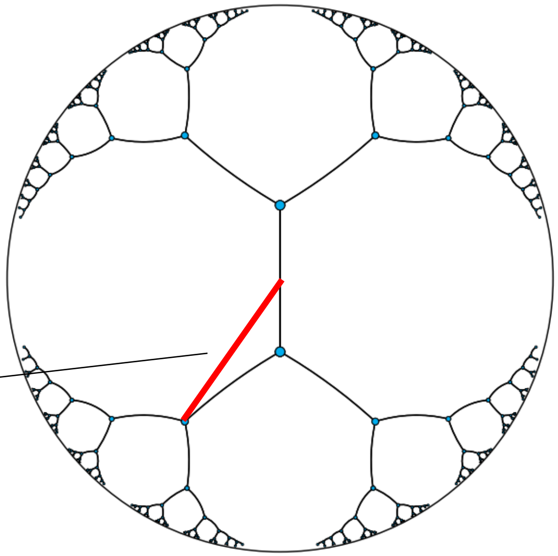
Lorentz Model with NPMI

$$\mathcal{L}(\Theta) = -\sum_{(x,y)\in\mathcal{D}} npmi(x,y) \cdot \log s(x,y)$$

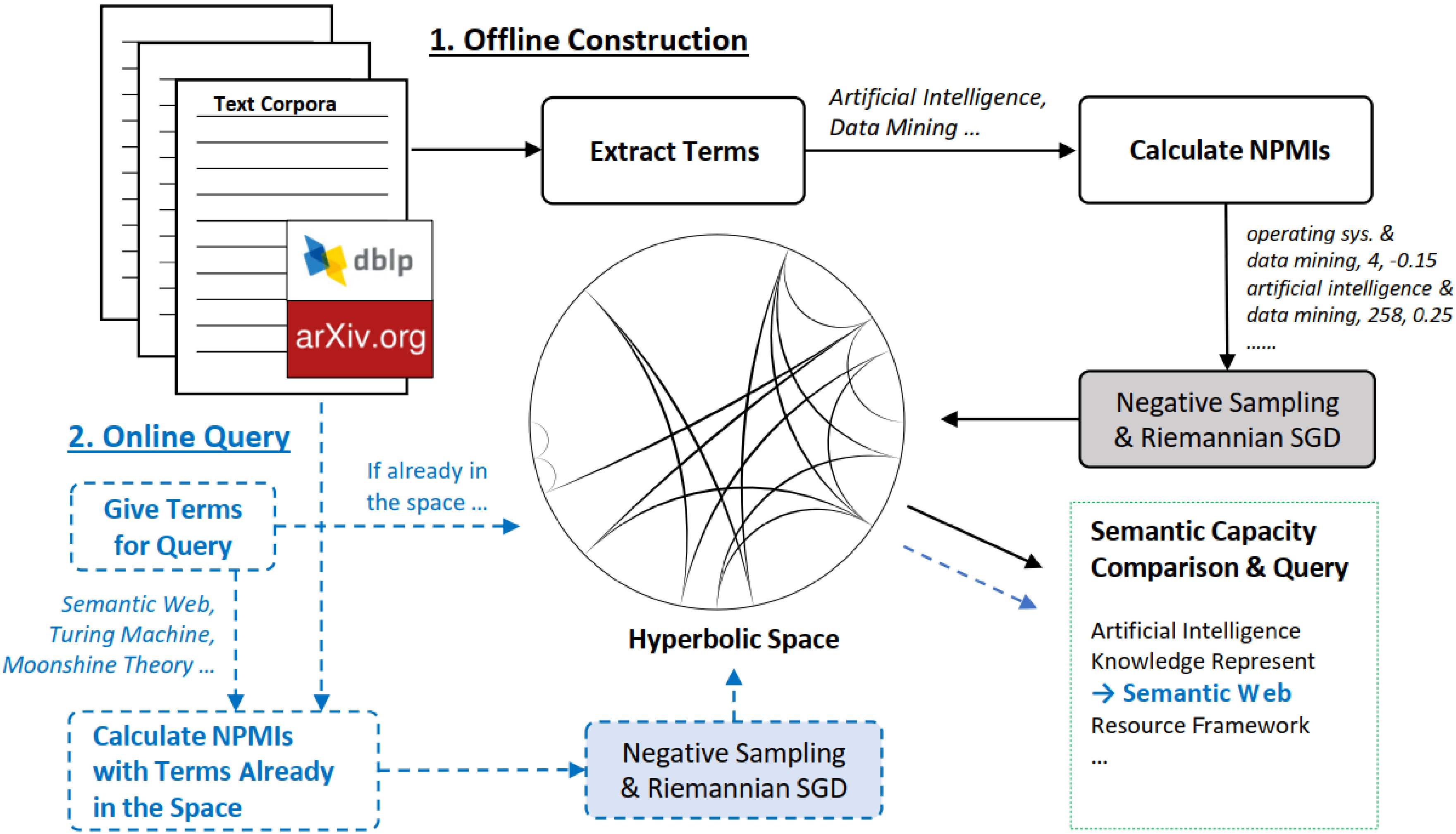
$$\mathcal{D} = \{(x,y) : npmi(x,y) > \delta\} \qquad s(x,y) = \frac{\exp(-d_\ell(\mathbf{x},\mathbf{y}))}{\sum_{y'\in\mathcal{N}(x)} \exp(-d_\ell(\mathbf{x},\mathbf{y}'))}$$

$$\min_{\Theta} \mathcal{L}(\Theta) \quad \text{s.t. } \forall \boldsymbol{\theta}_i \in \Theta : \boldsymbol{\theta}_i \in \mathbf{H}^n$$

$$\Rightarrow \qquad SC(x) = \frac{1}{||\ell^{-1}(\mathbf{x})||}$$



Framework



1. Offline Construction
Text Corpora
dblp
arXiv.org
Extract Terms
Artificial Intelligence,
Data Mining ...
Calculate NPMIs
operating sys. &
data mining, 4, -0.15
artificial intelligence &
data mining, 258, 0.25
......
Negative Sampling
& Riemannian SGD
2. Online Query
If already in
the space ...
Give Terms
for Query
Semantic Web,
Turing Machine,
Moonshine Theory ...
Hyperbolic Space
Semantic Capacity
Comparison & Query
Artificial Intelligence
Knowledge Represent
→ Semantic Web
Resource Framework
...
Calculate NPMIs
with Terms Already
in the Space
Negative Sampling
& Riemannian SGD

Experiments

- Hypernym-hyponym pairs in three scientific domains
- Abstracts of papers are used to find the co-occurrences between terms

	number of pairs			number of terms		
	all	top 1	top 2	all	top 1	top 2
Computer Science	782	93	325	651	11	109
Physics	1393	105	452	1090	14	127
Mathematics	1070	158	399	826	18	153

Baselines

- Popularity: $SC(x) \propto freq(x)$
- Poincaré GloVe (Tifrea et al., ICLR'2019)

Variants:

- Euclidean Model (Co-occurrence)
- Euclidean Model (NPMI)
- Lorentz Model (Co-occurrences)
- Lorentz Model (NPMI)

Human Annotation by Layman, Professional, Expert

Evaluation on Offline Construction

	Computer Science			Physics			Mathematics		
	all	top 1	top 2	all	top 1	top 2	all	top 1	top 2
Popularity	65.47	64.52	65.54	62.67	55.24	54.42	66.45	68.99	62.66
Poincaré GloVe	65.47	70.97	67.38	61.45	56.19	54.87	63.27	68.35	64.41
Euclidean Model (Co-occurrences)	69.44	71.69	70.77	67.77	54.29	60.40	68.82	78.06	69.42
Euclidean Model (NPMI)	71.00	73.92	75.46	58.15	47.62	53.76	64.95	65.19	65.79
Lorentz Model (Co-occurrences)	69.57	73.12	72.00	67.34	70.48	62.39	68.66	75.95	68.92
Lorentz Model (NPMI)	**74.25**	**88.39**	**77.11**	**72.52**	**82.48**	**74.07**	**72.34**	**80.76**	**73.86**

The Lorentz model with NPMI outperforms all the baselines significantly

Hearst patterns (with extended patterns) can only find **2.5%** (35/1393) pairs

Evaluation on Online Query

	Computer Science			Physics			Mathematics		
	all	top 1	top 2	all	top 1	top 2	all	top 1	top 2
Human Annotation (Layman)	64.33	75.31	68.27	58.67	56.14	58.82	62.00	67.62	64.26
Human Annotation (Professional)	78.33	82.72	80.32	79.67	91.23	81.96	80.00	**91.43**	83.53
Human Annotation (Expert)	**79.33**	86.42	82.73	**83.00**	**94.74**	**87.06**	**82.33**	83.81	**84.34**
Lorentz Model (NPMI)	77.40	**92.59**	**84.09**	78.20	91.58	79.29	76.20	80.00	79.28

The Lorentz model with NPMI can achieve performance comparable to professionals, with a small margin to experts, and much better than laymen

Conclusion

- **Semantic capacity:** a value that measures the semantic scope of terms
- **Semantic capacity association hypothesis** => **the Lorentz model with NPMI**
- **Two-step model:** offline construction and online query
- Experiments on three scientific domains

Thanks!