HyperExpan: Taxonomy Expansion with Hyperbolic Representation Learning

Mingyu Derek Ma, Muhao Chen*, Te-Lin Wu*, Nanyun (Violet) Peng
What is a taxonomy

**Online catalog taxonomy**

**Scientific taxonomy**
Introduction

- Taxonomy curation is expensive and suffers from limited coverage
- Our task: taxonomy expansion
  - Attach new concept to an existing taxonomy
Introduction

- Taxonomy curation is expensive and suffers from limited coverage
- Our task: taxonomy expansion
  - Attach new concept to an existing taxonomy
- Taxonomy size grows exponentially
- Hyperbolic space can better capture lower-level concepts with better expressiveness
HyperExpan

A taxonomy expansion framework based on hyperbolic representation learning

1. Better preserves the taxonomical structure in a more expressive hyperbolic space
2. Characterizes concepts by exploiting sparse neighborhood information
3. Improves inference precision and generalizability by leveraging pretrained distributional features
Model design

Anchor concept

explode
change integrity
crumple
fry
roast
cook
griddle
frizzle
pan roast

Query concept

grill

Score

Anchor concept representation

Query concept representation
Model design

Query concept  grill  

Ego graph  

- change
- integrity
- explode
- fry
- crump erupt
- roast
- griddle
- frizzle
- pan roast

Anchor concept representation

Score

Query concept representation
Step 1: initial concept features

Ego graph
- explode
- crump
- erup
- grill
- frizzle
- pan roast

Initial concept features
- change
- integrity
- cook
- fry
- roast

Query concept
- grill

Initial concept feature

Score

Anchor concept representation

Query concept representation
Step 2: anchor concept representation

Initial concept feature

Hyperbolic GNN

Score

Anchor concept representation

Query concept representation
Step 2: anchor concept representation

- **Ego graph**
  - explode
  - crump erupt
  - change
  - integrity
  - cook
  - fry
  - roast
  - griddle
  - frizzle
  - pan roast

- **Query concept**
  - grill

- **Hyperbolic GNN**
  - Initial concept features
  - Hidden layers
  - Score
  - Anchor concept representation
  - Query concept representation
Step 2: anchor concept representation

Anchor concept representation

Hyperbolic GNN

Ego graph

Query concept

Initial concept feature

Initial concept features

Positional embeddings

Hidden layers

Score

Query concept representation

Anchor concept representation

Initial concept feature

Positional embeddings

Parent

Self

Children

Absolute

Depth
Step 2: anchor concept representation

---

**Ego graph**
- cook
- integrity
- change
- yell
- crump
- erupt
- pan
- roast
- frizzle
- griddle
- fry

**Query concept**
- grill

---

**Hyperbolic GNN**
- Initial concept features
- Hidden layers
- Logarithmic map
- Positional embeddings
  - Parent
  - Self
  - Children
  - Absolute
  - Depth

**Anchor concept representation**
- Score

---

**Initial concept feature**
Step 2: anchor concept representation
Step 2: anchor concept representation

Ego graph

Query concept

Initial concept feature

Anchor concept representation

Score

Hyperbolic GNN

Logarithmic map

Exponential map

Hidden layers

Initial concept features

Positional embeddings

Parent

Self

Children

Absolute Depth

Score
Step 2: anchor concept representation

Ego graph

Query concept: grill

Hyperbolic GNN

Initial concept feature

Positional embeddings

Graph Readout

Score

Anchor concept representation

Initial concept features

Hidden layers

Logarithmic map

Aggregation

Exponential map

Parent

Self

Children

Absolute

Depth
Step 3: matching module
Learning and inference

- **Training**
  - Self-supervision: positive + negative pairs

  *Training/seed graph*

  - change integrity

  - explode
  - cook

  - crump
  - fry
  - roast

  *Self-supervised data points*

  - roast, cook
  - change integrity, cook
  - explode, cook
  - crump, cook

- **Loss function**

  \[
  \mathcal{L}(\Theta) = -\frac{1}{|X|} \sum_{X_i \in X} \left[ \log \frac{f(n_p, n_c)}{\sum_{(n_j, n_c) \in X_i} f(n_j, n_c)} \right]
  \]
Learning and inference

- Inference

**Query node pending to attach**

```
+-----------------+          +-----------------+          +-----------------+
|                 |          |                 |          |                 |
|   change integrity   |          |   change integrity   |          |   change integrity   |
|-----------------+          +-----------------+          +-----------------+          +-----------------+          +-----------------+
|                 |          |                 |          |                 |          |                 |          |                 |
|   explode       |          |   cook          |          |   explode       |          |   cook          |          |   change integrity |
|                 |          |                 |          |                 |          |                 |          |                 |
|   crump         |          |   fry           |          |   crump         |          |   fry           |          |   roast         |
|                 |          |                 |          |                 |          |                 |          |                 |
|                 |          | grill           |          |                 |          | grill           |          |
```

**Calculate matching scores with candidates**

```
+-----------------+          +-----------------+          +-----------------+
|                 |          |                 |          |                 |
|   change integrity   |          |   change integrity   |          |   change integrity   |
|-----------------+          +-----------------+          +-----------------+          +-----------------+          +-----------------+
|                 |          |                 |          |                 |          |                 |          |                 |
|   explode       |          |   cook          |          |   explode       |          |   cook          |          |   change integrity |
|                 |          |                 |          |                 |          |                 |          |                 |
|   crump         |          |   fry           |          |   crump         |          |   fry           |          |   roast         |
|                 |          |                 |          |                 |          |                 |          |                 |
|                 |          | grill           |          |                 |          | grill           |          |
```

**Ranking list**

- **cook**
- roast
- fry
- change integrity
- ...
We hypothesize that the absolute position embedding (line 9) caused if we remove relative position embeddings means "without". "i/o" means "instead of", "w/o" means "without". The metrics. MRR metrics are scaled by 10 to amplify the performance difference.

<table>
<thead>
<tr>
<th>Model</th>
<th>MR ↓ MRR ↑</th>
<th>Recall % ↑</th>
<th>Precision % ↑</th>
<th>MR ↓ MRR ↑</th>
<th>Recall % ↑</th>
<th>Precision % ↑</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>@1 @5 @10</td>
<td>@1 @5 @10</td>
<td>@1 @5 @10</td>
<td>@1 @5 @10</td>
<td>@1 @5 @10</td>
<td>@1 @5 @10</td>
</tr>
<tr>
<td>WordNet-Verb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Candidates #: 11,936)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARBORIST</td>
<td>608.7 0.280</td>
<td>10.8 24.0 27.7</td>
<td>6.7 4.8 3.2</td>
<td>1095.1 0.435</td>
<td>16.5 28.4 34.1</td>
<td>16.8 5.8 3.5</td>
</tr>
<tr>
<td>TaxoExpan</td>
<td>502.8 0.439</td>
<td>12.4 28.2 35.2</td>
<td>12.4 5.6 3.5</td>
<td>649.6 0.562</td>
<td>19.7 38.2 47.4</td>
<td>20.1 7.8 4.8</td>
</tr>
<tr>
<td>TMN</td>
<td>465.0 0.479</td>
<td>14.9 31.6 37.9</td>
<td>13.2 6.4 4.0</td>
<td>501.0 0.595</td>
<td>20.7 40.5 50.1</td>
<td>21.1 8.3 5.1</td>
</tr>
<tr>
<td>GCN</td>
<td>456.9 0.445</td>
<td>10.9 27.2 34.5</td>
<td>10.9 5.4 3.5</td>
<td>684.1 0.563</td>
<td>20.9 39.8 47.3</td>
<td>21.3 8.1 4.8</td>
</tr>
<tr>
<td>GAT</td>
<td>471.7 0.449</td>
<td>11.6 28.7 35.6</td>
<td>11.6 5.7 3.6</td>
<td>640.7 0.585</td>
<td>22.3 40.9 49.7</td>
<td>22.7 8.3 5.1</td>
</tr>
<tr>
<td>HyperExpan</td>
<td>400.8 0.517</td>
<td>15.0 32.8 42.7</td>
<td>15.0 6.6 4.3</td>
<td>573.6 0.607</td>
<td>23.9 42.1 52.5</td>
<td>24.4 8.6 5.4</td>
</tr>
<tr>
<td>MAG-PSY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Candidates #: 21,187)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARBORIST</td>
<td>119.9 0.722</td>
<td>21.0 48.4 62.9</td>
<td>25.8 12.5 7.7</td>
<td>284.7 0.602</td>
<td>15.1 38.9 49.4</td>
<td>24.6 12.6 8.0</td>
</tr>
<tr>
<td>TaxoExpan</td>
<td>68.5 0.775</td>
<td>26.1 56.9 69.5</td>
<td>33.8 14.7 9.0</td>
<td>189.8 0.661</td>
<td>15.9 42.9 55.4</td>
<td>25.8 13.9 9.0</td>
</tr>
<tr>
<td>TMN</td>
<td>73.0 0.781</td>
<td>25.8 58.7 70.5</td>
<td>33.4 15.2 9.1</td>
<td>160.5 0.667</td>
<td>16.0 43.1 56.3</td>
<td>26.0 14.0 9.1</td>
</tr>
<tr>
<td>GCN</td>
<td>51.4 0.742</td>
<td>23.8 52.5 64.3</td>
<td>30.8 13.6 7.4</td>
<td>90.3 0.653</td>
<td>14.5 39.6 53.3</td>
<td>23.6 12.9 8.7</td>
</tr>
<tr>
<td>GAT</td>
<td>48.6 0.751</td>
<td>23.6 52.4 65.8</td>
<td>30.5 13.5 8.5</td>
<td>92.2 0.676</td>
<td>15.9 41.9 56.0</td>
<td>25.9 13.6 9.1</td>
</tr>
<tr>
<td>HyperExpan</td>
<td>38.4 0.827</td>
<td>28.8 63.0 75.3</td>
<td>37.2 16.3 9.7</td>
<td>74.4 0.689</td>
<td>16.1 44.6 58.0</td>
<td>26.1 14.5 9.4</td>
</tr>
<tr>
<td>MAG-CS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Candidates #: 22,754)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Experiments

- HyperExpan get large performance increase compared with GCN and GAT due to expressiveness of the hyperbolic space
- HyperExpan outperforms previous SOTA TMN
Ablation study

<table>
<thead>
<tr>
<th>Model</th>
<th>MRR ↑</th>
<th>Rec ↑ @10</th>
<th>Prec ↑ @1</th>
</tr>
</thead>
<tbody>
<tr>
<td>w/o trainable curvature</td>
<td>0.490</td>
<td>40.8</td>
<td>14.4</td>
</tr>
<tr>
<td>anchor + parent + children</td>
<td>0.506</td>
<td>42.2</td>
<td>15.0</td>
</tr>
<tr>
<td>#4 + anchor’s ancestors</td>
<td>0.505</td>
<td>42.5</td>
<td>15.5</td>
</tr>
<tr>
<td>#5 + anchor’s descendants</td>
<td>0.517</td>
<td>42.7</td>
<td>15.0</td>
</tr>
<tr>
<td>#6 + anchor’s siblings</td>
<td>0.502</td>
<td>41.7</td>
<td>14.5</td>
</tr>
<tr>
<td>w/o Relative Pos Emb</td>
<td>0.497</td>
<td>40.8</td>
<td>13.0</td>
</tr>
<tr>
<td>w/o Absolute Pos Emb</td>
<td>0.503</td>
<td>41.2</td>
<td>14.3</td>
</tr>
<tr>
<td>w/o both Positional Emb</td>
<td>0.482</td>
<td>38.8</td>
<td>12.5</td>
</tr>
<tr>
<td>HyperExpan</td>
<td>0.517</td>
<td>42.7</td>
<td>15.0</td>
</tr>
</tbody>
</table>

- Trainable curvature leads fine-grained manifold setting
- Adding descendant or ancestors of the anchor node is helpful, anchor’s sibling nodes are not
- Positional embeddings are helpful

*Recall % P @1 @5 @10

MRR is scaled by 10, i/o means instead of, w/o means without
Conclusion

- HyperExpan: a taxonomy expansion model which better preserves the taxonomical structure in an expressive hyperbolic space
- Use HGNN to incorporate neighborhood information and positional features of concepts
- Experimental results show that HyperExpan performs better than its Euclidean counterparts and achieves the state-of-the-art
Thank You

HyperExpan: Taxonomy Expansion with Hyperbolic Representation Learning

Mingyu Derek Ma, Muhao Chen*, Te-Lin Wu*, Nanyun (Violet) Peng

Code available at:
github.com/PlusLabNLP/HyperExpan