Hierarchical Embeddings for Hypernymy Detection and Directionality

Kim Anh Nguyen, Maximilian Köper, Sabine Schulte im Walde, Ngoc Thang Vu
Institute for Natural Language Processing - University of Stuttgart, Germany
{nguyenkh, koepermn, schulte, thangvu}@ims.uni-stuttgart.de

Motivation

- **Hypothesis**: each common context of a hyponym–hypernym relation is an indicator to determine which of two words is semantically more general.
- **Goal**: learn the hierarchical embeddings for hypernymy detection and directionality.
- **Procedure**: strengthen the distributional similarity of hypernym pairs and generate a distributional hierarchy between hyponyms and hypernyms.

Contributions

1. Propose a novel neural model HyperVec to learn hierarchical embeddings for hypernymy addressing detection and directionality tasks.
2. Present an unsupervised measure to score hypernym relations based on HyperVec.
3. The HyperVec is able to generalize over unseen hypernymy pairs.
4. The HyperVec outperforms other state-of-the-art unsupervised measures and embedding models.

Models

1. **Hierarchical Hypernymy**:
   - Learn the hierarchical embeddings in a specific order. The similarity score for hypernymy is higher than the similarity score for other relations:
     \[ L(u,v) = \frac{1}{\#(u,v)} \sum_{u \in H^+(v)} \partial(\vec{u}, \vec{v}) \]
   - Learn the distributional hierarchy between hypernyms and hyponyms, as an indicator to differentiate between hypernym and hyponym:
     \[ L(v,u) = \sum_{v \in H^-(u)} \partial(\vec{v}, \vec{u}) \]
   - Incorporate the Skip-gram with negative sampling model:
     \[ J(u,v) = \#(u,v) \log \sigma(\vec{u}, \vec{v}) + k \cdot E_{c/N} \log \sigma(-\vec{u}, \vec{v}) \]
   - The final objective function is defined as follows:
     \[ J = \sum_{u \in V_W} \sum_{v \in V_C} J(u,v) + L(u,v) + L(v,u) \]

2. **Unsupervised Hypernymy Measure**:
   - **HyperVec** shows the following properties:
     1. high similarity between hypernyms and hyponyms.
     2. hierarchy between hypernyms and their hyponyms.
   - The measure is defined as follows:
     \[ \text{HyperScore}(u,v) = \cos(\vec{u}, \vec{v}) + \frac{\mu(\vec{u})}{\mu(\vec{v})} \]

Experimental Settings

- **Benchmark**: EnCoWo14A corpus ≈14.5 billion tokens.
- **Baseline**: default SGNS (word2vec).
- **100 dimensions, window 5, negative samples 15, learning rate 0.025**.
- **Learn HyperVec for nouns and verbs**.

Graded Entailment

- **HyperLex**: dataset of graded lexical entailment.
- Provides flexible entailment on a continuous scale, e.g., duck-animal is 5.6 out of 6.0 but reversed animal-duck is only 1.0.
- 2,616 word pairs, seven semantic relations, and two word classes (nouns and verbs).
- We compared HyperScore against the most prominent state-of-the-art models.

Generalizing Hypernymy I

Motivation: explore HyperVecs potential for generalization

1. Rely on a small seed set only, rather than using a large set of training data.
2. Learn only based on the 200 concepts (and their hyponyms) from the BLESS dataset.
3. Performance measured using Average Precision (AP) ranking measure.

Generalizing Hypernymy II

- **Project (default) representations from any arbitrary language into our modified English HyperVec space**.
- Mapping function between source and target space using least-squares error method.
- **DE → EN** and **IT → EN** word translations based on Europarl.
- Compare the original vs. mapped representation on hypernymy ranking retrieval task.

---

**Table:**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Baseline</th>
<th>HyperScore</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVALution</td>
<td>0.353</td>
<td>0.390</td>
</tr>
<tr>
<td>Lenci/Benotto</td>
<td>0.382</td>
<td>0.448</td>
</tr>
<tr>
<td>Weeds</td>
<td>0.441</td>
<td>0.585</td>
</tr>
</tbody>
</table>

**Table:**

<table>
<thead>
<tr>
<th>Model</th>
<th>Embeddings</th>
</tr>
</thead>
<tbody>
<tr>
<td>FR</td>
<td>SGNS</td>
</tr>
<tr>
<td>DEM</td>
<td>PARAGRAM</td>
</tr>
<tr>
<td>SLQ5</td>
<td>OrderEmb</td>
</tr>
<tr>
<td>WN</td>
<td>Word2Gauss</td>
</tr>
<tr>
<td>VIS</td>
<td>HyperGauss</td>
</tr>
</tbody>
</table>

**Table:**

<table>
<thead>
<tr>
<th>Model</th>
<th>Embeddings</th>
</tr>
</thead>
<tbody>
<tr>
<td>German</td>
<td>Hypl/All</td>
</tr>
<tr>
<td>DE → SGNS</td>
<td>0.28</td>
</tr>
<tr>
<td>DE → EN</td>
<td>0.68</td>
</tr>
<tr>
<td>Italian</td>
<td>Hypl/All</td>
</tr>
<tr>
<td>IT → SGNS</td>
<td>0.38</td>
</tr>
<tr>
<td>IT → EN</td>
<td>0.44</td>
</tr>
</tbody>
</table>