**Automated WordNet Construction Using Word Embeddings**

Mikhail Khodak, Andrej Risteski, Christiane Fellbaum, Sanjeev Arora
Computer Science Department, Princeton University

---

**Contributions:**
- Method to represent and match synsets using machine translation and word embeddings.
- Unsupervised procedure to induce and represent word-senses as clusters of related words.
- Test sets for French and Russian WordNets.
- Automated WordNets for French and Russian.

**Synset Representation:**
- Find a word embedding $\nu_w$ (e.g., GloVe) for each word $w$ in the target language.
- Represent a candidate synset $S$ of word $w$ as a combination $u_S$ of vector representations of lemmas and glosses provided by WordNet. [1]
  - Represent a set of lemmas $L$ by a sum: $\sum_{w \in L} v_w$.
  - Represent a gloss $G$ via SIF embeddings: $\sum_{w \in G} \frac{a}{a + P(w)} v_w$ for $a = 10^{-4}$. [2]
- Match $S$ to $w$ if $u_S \cdot v_w \geq \alpha$ (a learned cutoff).

**Isometric mapping of sense-clusters found for "fox"**

**Sense Clusters:**
- Approximate $v_w$ as a sparse linear representation of atoms $a_i$ via dictionary learning. [3]
- Represent each sense of $w$ as a cluster of words close to both $v_w$ and one of its atoms $a_i$.
- Improve performance by using these sense-clusters to prune irrelevant synsets matched to $w$.

---

**Translations**

French word 'dalle'
French word 'лук' ("bow", "onion")

**Threshold Matching**

return all synsets with score above a threshold $\alpha$

**Score = 0.521**

**Correct synsets:** flag.n.06 - slab.n.01

**Retrieved synsets:** slab.n.01

---

**Contributions:**
- Wordnet Libre du Français (WOLF) [4]
- Universal Wordnet [5]
- Extended Open Multilingual Wordnet [6]

---

**F-Score of Synset Matching for French and Russian**

<table>
<thead>
<tr>
<th>Synset Representation</th>
<th>Synset Representation + Sense Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---