Sense Embeddings in Knowledge-Based Word Sense Disambiguation
Loïc Vial, Benjamin Lecouteux, Didier Schwab

To cite this version:
Loïc Vial, Benjamin Lecouteux, Didier Schwab. Sense Embeddings in Knowledge-Based Word Sense Disambiguation. IWCS, 2017, Montpellier, France. hal-02094762

HAL Id: hal-02094762
https://hal.archives-ouvertes.fr/hal-02094762
Submitted on 9 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Sense Embeddings in Knowledge-Based Word Sense Disambiguation
Loïc Vial, Benjamin Lecouteux, Didier Schwab
LIG - GETALP, Univ. Grenoble Alpes, France

Word Sense Disambiguation

Word Sense Disambiguation (WSD) consists in assigning the most appropriate sense to a word in a document, given a particular sense inventory. Similarity-based disambiguation systems are WSD systems composed of both:

1. A local algorithm, also called similarity measure, which computes a similarity score $\text{sim}(S_1, S_2)$ between two senses.
2. A global algorithm, which searches the best combination of senses at the document level, by using the local algorithm.

Common similarity measures

Among the most frequently used similarity measures based on definitions contained in a dictionary, there are:

- The Lesk algorithm [1], which computes the number of words in common between the two definitions:

  $$\text{Lesk}(S_1, S_2) = |D(S_1) \cap D(S_2)|, \quad \text{with } D(S) = \{w_0, w_1, \ldots, w_n\}$$

  \text{the definition of the sense } S \text{ in the dictionary.}

- The Extended Lesk algorithm [2], which takes into account not only the definitions of the senses, but also the definitions of all related senses in a semantic network:

  $$\text{ExtLesk}(S_1, S_2) = |(D(S_1) \cup D(\text{rel}(S_1))) \cap (D(S_2) \cup D(\text{rel}(S_2)))|, \quad \text{with } \text{rel}(S) \text{ the senses connected to } S \text{ through an explicit link in WordNet [3].}$$

Our similarity measure

The similarity measure that we propose, called VecLesk$(S_1, S_2)$, takes into account the closest senses in a sense embeddings model regarding their cosine similarity between the vector of the lemma of $S$, above a threshold $\delta_1$, and also between the vector of $S$, above a threshold $\delta_2$.

It is formally defined as:

$$\text{VecLesk}(S_1, S_2, \delta_1, \delta_2) = |\{D(S_1) \cup D(\text{rel}(S_1))\} \cap \{D(S_2) \cup D(\text{rel}(S_2))\}|$$

$$\text{rel}(S, \delta_1, \delta_2) = \{S' \mid \cosine(\phi(\text{lemma}(S)), \phi(S')) > \delta_1, \cosine(\phi(S), \phi(S')) > \delta_2\}$$

Therefore there is no more manually created semantic network used for extending the Lesk measure.

Some example results

Our sense vectors can be manipulated as word vectors.

For instance, close to bank (financial institution), we find the senses account, deposit and money; whereas close to bank (shore), we find the senses coast, sandbank and dip.

All sense vectors are available at the following URL: [https://github.com/getalp/WSD-IWCS2017-Vialetal](https://github.com/getalp/WSD-IWCS2017-Vialetal)

Evaluation

<table>
<thead>
<tr>
<th>System</th>
<th>SemEval 2007</th>
<th>SemEval 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>S2C [4]</td>
<td>75.80%*</td>
<td></td>
</tr>
<tr>
<td>Lesk</td>
<td>68.70%</td>
<td>50.65%</td>
</tr>
<tr>
<td>Extended Lesk</td>
<td>78.01%</td>
<td>61.42%</td>
</tr>
<tr>
<td>VecLesk (Baroni C [5])</td>
<td>75.29%</td>
<td>58.02%</td>
</tr>
<tr>
<td>VecLesk (Baroni P [5])</td>
<td>73.52%</td>
<td>53.46%</td>
</tr>
<tr>
<td>VecLesk (Dep [6])</td>
<td>73.02%</td>
<td>56.40%</td>
</tr>
<tr>
<td>VecLesk (GloVe [7])</td>
<td>73.00%</td>
<td>59.01%</td>
</tr>
<tr>
<td>VecLesk (Word2Vec [8])</td>
<td>73.30%</td>
<td>57.00%</td>
</tr>
</tbody>
</table>

Table 1: Results on SemEval 2007 and SemEval 2015 for each underlying word embeddings model.

*This system is comparable to our in terms of resources used, but it is biased: their threshold parameter $\delta$ is learned on the evaluation task. We would obtain 76.50% by doing the same.

Senses vectors

The definition of the vector of a sense $S$ is $\phi(S)$, and it corresponds to:

$$\phi(S) = \sum_{i=0}^{n} \phi(w_n) \times \text{weight}(\text{pos}(w_n)) \times \text{idf}(w_n)$$

- $\text{D}(S) = \{w_0, w_1, \ldots, w_n\}$ the definition of sense $S$ in the dictionary
- $\phi(w_n)$ the vector of the word $w_n$
- $\text{pos}(w_n) = \{n, v, a, r\}$ the part of speech of the word $w_n$: noun, verb, adjective or adverb
- $\text{idf}(w_n)$ the IDF value of $w_n$

$\phi(S)$ is then normalized so its length is the same as any word vector.

Table 2: Estimation of the parameters $\delta_1$ and $\delta_2$ on SemEval 2007 and SemEval 2015.

<table>
<thead>
<tr>
<th>Model</th>
<th>Baroni C [5] $\delta_1$</th>
<th>Baroni P [5] $\delta_2$</th>
<th>Deps [6] $\delta_1$</th>
<th>GloVe [7] $\delta_2$</th>
<th>Word2Vec [8] $\delta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SemEval 2007</td>
<td>0.6</td>
<td>0.6</td>
<td>0.5</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>SemEval 2015</td>
<td>0.5</td>
<td>0.8</td>
<td>0.5</td>
<td>0.5</td>
<td>0.6</td>
</tr>
</tbody>
</table>

- The Lesk measure is considerably improved with our extension.
- The scores almost reach the Extended Lesk.
- Our system allows to improve the word sense disambiguation of languages that have less resources than English.

References