



**HAL**  
open science

# 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.science/hal-02094762v1>**

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

## 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  $sim(S_1, S_2)$  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:  
 $Lesk(S_1, S_2) = |D(S_1) \cap D(S_2)|$ , with  $D(S) = \{w_0, w_1, \dots, w_n\}$  the definition of the sense  $S$  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:  
 $ExtLesk(S_1, S_2) = |(D(S_1) \cup D(rel(S_1))) \cap (D(S_2) \cup D(rel(S_2)))|$ , with  $rel(S)$  the senses connected to  $S$  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:

$$VecLesk(S_1, S_2, \delta_1, \delta_2) = |(D(S_1) \cup D(rel(S_1, \delta_1, \delta_2))) \cap (D(S_2) \cup D(rel(S_2, \delta_1, \delta_2)))|$$
$$rel(S, \delta_1, \delta_2) = \{S' \mid cosine(\phi(lemma(S)), \phi(S')) > \delta_1, cosine(\phi(S), \phi(S')) > \delta_2\}$$

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>



## Sense vectors

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

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

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

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

## Evaluation

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

**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.

Model Parameters	Baroni C [5]		Baroni P [5]		Deps [6]		GloVe [7]		Word2Vec [8]	
	$\delta_1$	$\delta_2$	$\delta_1$	$\delta_2$	$\delta_1$	$\delta_2$	$\delta_1$	$\delta_2$	$\delta_1$	$\delta_2$
SemEval 2007	0.6	0.6	0.5	0.5	0.6	0.8	0.5	0.6	0.5	0.6
SemEval 2015	0.5	0.8	0.5	0.6	0.6	0.8	0.5	0.7	0.5	0.6

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

- The **Lesk** measure is **considerably improved** with our extension.
- The scores **almost** reach the **Extended Lesk**, **without** the need of a semantic network like the one from **WordNet**.
- Our system allows to **improve the word sense disambiguation of languages that have less resources** than English.

## References

- [1] Michael Lesk. Automatic sense disambiguation using mrd: how to tell a pine cone from an ice cream cone. In *Proceedings of SIGDOC '86*, pages 24–26, 1986.
- [2] Satantjee Banerjee and Ted Pedersen. An adapted lesk algorithm for word sense disambiguation using wordnet. In *CICLing 2002*, Mexico City, February 2002.
- [3] George A. Miller. Wordnet: A lexical database. *ACM*, Vol. 38(No. 11):p. 1–41, 1995.
- [4] Xinxiong Chen *et al.* A unified model for word sense representation and disambiguation. In *EMNLP 2014*, pages 1025–1035.
- [5] Marco Baroni *et al.* Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In *ACL 2014*, pages 238–247.
- [6] Omer Levy and Yoav Goldberg. Dependency-based word embeddings. In *ACL 2014*, pages 302–308.
- [7] Jeffrey Pennington *et al.* Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, 2014.
- [8] Tomas Mikolov *et al.* Distributed representations of words and phrases and their compositionality. In *NIPS 2013*, pages 3111–3119.