Making Sense of Word Embeddings

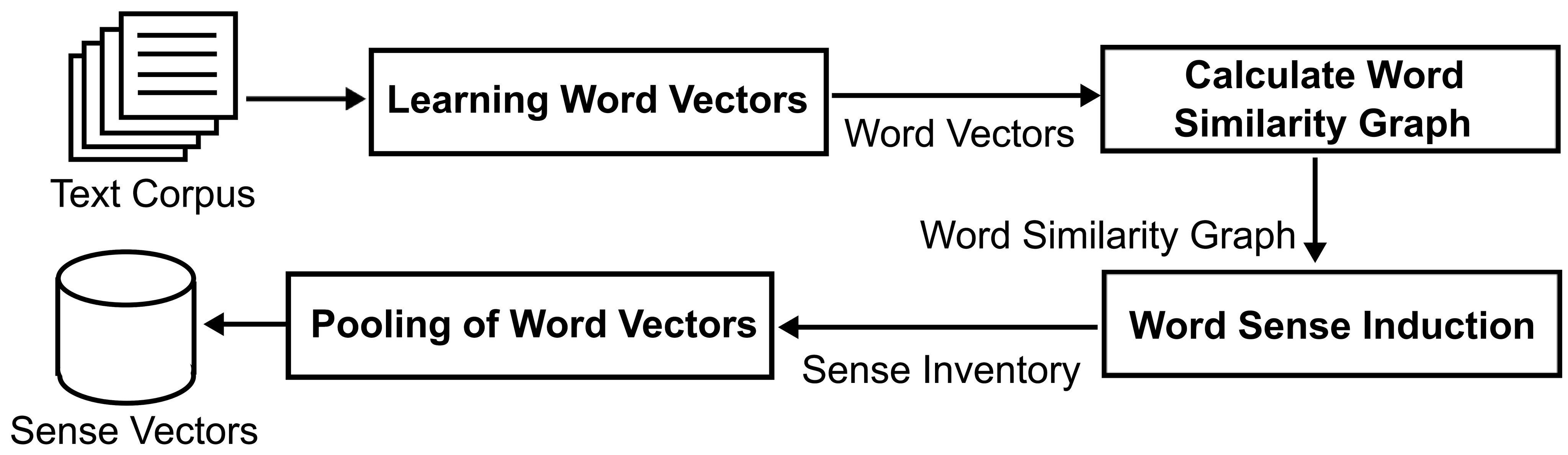
Maria Pelevina[1], Nikolay Arefiev[2], Chris Biemann[1], Alexander Panchenko[1]

[1] TU Darmstadt, Germany [2] Moscow State University, Russia

Introduction

We present a simple yet effective approach for learning word sense embeddings. In contrast to existing techniques, which either directly learn sense representations from corpora or rely on sense inventories from lexical resources, our approach can induce a sense inventory from existing word embeddings via clustering of ego-networks of related words. An integrated WSD mechanism enables labeling of words in context with learned sense vectors, which gives rise to downstream applications. Experiments show that the performance of our method is comparable to state-of-the-art unsupervised WSD systems.

Learning Sense Embeddings from Word Embeddings

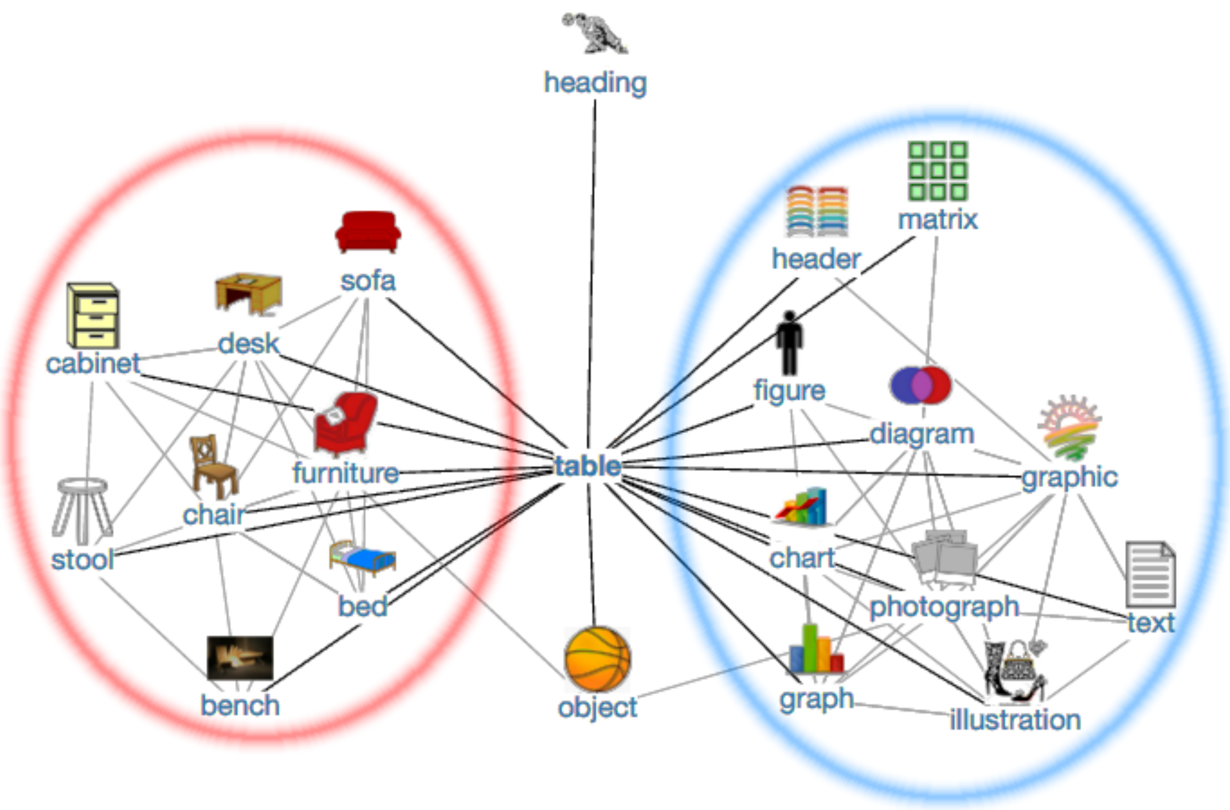


Schema of the word sense embeddings learning method.

	TWSI	JBT	w2v
table (furniture)	counter, console, bench, dinner table, dining table, desk, surface, bar, board	chair, room, desk, pulpit, couch, furniture, fireplace, bench, door, box, railing, tray	tray, bottom, bucket, basket, cup, pile, bracket, pot, cue, plate, jar, platter, ladder
table (data)	chart, list, index, graph, columned list, tabulation, standings, diagram, ranking	procedure, system, datum, process, mechanism, tool, method, database, calculation, scheme	diagram, brackets, stack, list, parenthesis, playfield, drop-down, cube, hash, results, tab
table (negotiations)	surface, counter, console, bargaining table, platform, negotiable, machine plate, level	—	—
table (geo)	level, plateau, plain, flatland, saturation level, water table, geographical level, water level	—	—

Word sense clusters from inventories derived from the Wikipedia corpus via crowdsourcing (TWSI), JoBimText (JBT) and word embeddings (w2v).

Word Sense Induction



Visualization of the ego-network of the word "table" with the "furniture" and the "data" sense clusters.

Algorithm 1: Word sense induction.

input : T – word similarity graph, N – ego-network size, n – ego-network connectivity, k – minimum cluster size

output: for each term $t \in T$, a clustering S_t of its N most similar terms

foreach $t \in T$ **do**
- $V \leftarrow N$ most similar terms of t from T
- $G \leftarrow$ graph with V as nodes and no edges E
- **foreach** $v \in V$ **do**
  - $V' \leftarrow n$ most similar terms of v from T
  - **foreach** $v' \in V'$ **do**
    - **if** $v' \in V$ **then** add edge (v, v') to E
  - **end**
- **end**
- $S_t \leftarrow \texttt{ChineseWhispers}(G)$
- $S_t \leftarrow \{s \in S_t : |s| \geq k\}$

end

Vector	Nearest Neighbours
table	tray, bottom, diagram, bucket, brackets, stack, basket, list, parenthesis, cup, trays, pile, playfield, bracket, pot, drop-down, cue, plate
table#0	leftmost#0, column#1, randomly#0, tableau#1, top-left0, indent#1, bracket#3, pointer#0, footer#1, cursor#1, diagram#0, grid#0
table#1	pile#1, stool#1, tray#0, basket#0, bowl#1, bucket#0, box#0, cage#0, saucer#3, mirror#1, birdcage#0, hole#0, pan#1, lid#0

Neighbours of the word “table” and its senses.

Word Sense Disambiguation with Sense Embeddings

Context representation based on k context or word vectors:

$$\bar{\mathbf{c}}_c = k^{-1} \sum_{i=1}^{k} vec_c(c_i)$$

$$\bar{\mathbf{c}}_w = k^{-1} \sum_{i=1}^{k} vec_w(c_i)$$

Similarity- and probability-based disambiguation in context:

$$s^* = \arg\max_i sim(\mathbf{s}_i, C) = \arg\max_i \frac{\bar{\mathbf{c}}_w \cdot \mathbf{s}_i}{\|\bar{\mathbf{c}}_w\| \cdot \|\mathbf{s}_i\|}$$

$$s^* = \arg\max_i P(C|\mathbf{s}_i) = \arg\max_i \frac{1}{1 + e^{-\bar{\mathbf{c}}_c \cdot \mathbf{s}_i}}$$

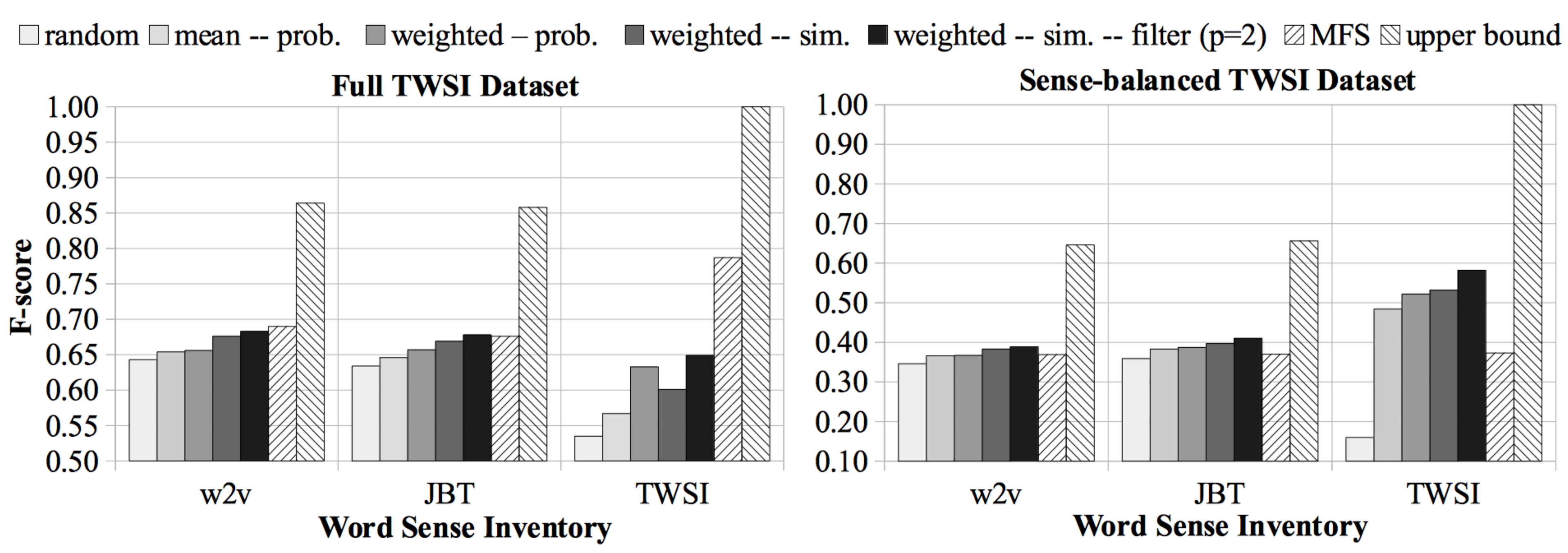
Sense vector:

$$\mathbf{s}_i = \frac{\sum_{k=1}^{n} \gamma_i(w_k) vec_w(w_k)}{\sum_{k=1}^{n} \gamma_i(w_k)}$$

Filtering the k context words:

$$\max_i f(\mathbf{s}_i, c_j) - \min_i f(\mathbf{s}_i, c_j)$$

Results: WSD Evaluation on the TWSI and the SemEval 2013 Task 13 Datasets



Performance of our method trained on the Wikipedia corpus on the full (on the left) and on the sense-balanced (on the right) TWSI dataset.

Model		Supervised Evaluation: Jacc. Ind.	Supervised Evaluation: Tau	Supervised Evaluation: WNDCG	Clustering Evaluation: F.NMI	Clustering Evaluation: F.B-Cubed
Baselines	One sense for all	0.171	0.627	0.302	0.000	0.631
	One sense per instance	0.000	0.953	0.000	0.072	0.000
	Most Frequent Sense (MFS)	0.579	0.583	0.431	–	–
SemEval	AI-KU (add1000)	0.176	0.609	0.205	0.033	0.317
	AI-KU	0.176	0.619	0.393	0.066	0.382
	AI-KU (remove5-add1000)	0.228	0.654	0.330	0.040	0.463
	Unimelb (5p)	0.198	0.623	0.374	0.056	0.475
	Unimelb (50k)	0.198	0.633	0.384	0.060	0.494
	UoS (#WN senses)	0.171	0.600	0.298	0.046	0.186
	UoS (top-3)	0.220	0.637	0.370	0.044	0.451
	La Sapienza (1)	0.131	0.544	0.332	–	–
	La Sapienza (2)	0.131	0.535	0.394	–	–
Sense emb.	AdaGram, $\alpha = 0.05$, 100 dim. vectors	0.274	0.644	0.318	0.058	0.470
Our models	w2v – weighted – sim. – filter ($p = 2$)	0.197	0.615	0.291	0.011	0.615
	w2v – weighted – sim. – filter ($p = 2$): nouns	0.179	0.626	0.304	0.011	0.623
	JBT – weighted – sim. – filter ($p = 2$)	0.205	0.624	0.291	0.017	0.598
	JBT – weighted – sim. – filter ($p = 2$): nouns	0.198	0.643	0.310	0.031	0.595
	TWSI – weighted – sim. – filter ($p = 2$): nouns	0.215	0.651	0.318	0.030	0.573

The best configurations of our method on the SemEval 2013 Task 13 dataset. All systems were trained on the ukWaC corpus

Deutsche Forschungsgemeinschaft

Code & Data: https://github.com/tudarmstadt-lt/sensegram