

Know Your Graph. State-of-the-Art Knowledge-Based WSD

Alexander Popov and Kiril Simov and Petya Osenova

IICT, Bulgarian Academy of Sciences

Sofia, Bulgaria

{alex.popov|kivs|petya}@bultreebank.org

Abstract

This paper introduces several improvements over the current state of the art in *knowledge-based* word sense disambiguation. Those innovations are the result of modifying and enriching a knowledge base created originally on the basis of WordNet. They reflect several separate but connected strategies: manipulating the shape and the content of the knowledge base, assigning weights over the relations in the knowledge base, and the addition of new relations to it. The main contribution of the paper is to demonstrate that the previously proposed knowledge bases organize linguistic and world knowledge suboptimally for the task of word sense disambiguation. In doing so, the paper also establishes a new state of the art for knowledge-based approaches. Its best models are competitive in the broader context of supervised systems as well.

1 Introduction

Word sense disambiguation (WSD) is a long-standing task in natural language processing (NLP), which has been approached through several broad families of computational techniques. While supervised learning models typically achieve the highest accuracy scores, problems related to the lack of gold corpora, data sparseness and suboptimal granularity of the computational lexicons have plagued the field and prevented significant breakthroughs. This paper presents results achieved with *knowledge-based word sense disambiguation* (KBWSD) algorithms as an alternative pathway. It builds on previous work in the subfield and demonstrates that KBWSD can achieve accuracy scores near and

in some cases even at the state-of-the-art, as a rule dominated by supervised approaches. Moreover, the experimental results indicate that in-depth analysis of knowledge representation and knowledge enrichment hold significant promises – both as an alternative to supervised WSD, able to sidestep data-related issues, and as a source of potentially powerful new training signals.

The contribution of the article is fourfold. Firstly, it presents a novel strategy for constructing knowledge bases used in KBWSD, showing that a structure centered on word senses rather than on synsets can be more effective for this particular task. Secondly, another novel approach to structuring the graph is explored in using word and synset embedding models in order to assign weights to relation arcs in the knowledge graph. Thirdly, a number of avenues for the enrichment of the semantic network used for KBWSD are pursued, thus linking WordNet to external resources. Finally, a new state of the art *for KBWSD* is established, which is competitive even when compared with supervised systems.

The paper is structured as follows: the next section presents related work; section 3 deals with some preliminary experiments aimed at replicating previous work described in the literature, but in a slightly modified setting; section 4 outlines several strategies for improving the structure of the knowledge base used for KBWSD; the penultimate section reports on our core experimental work, and the final section concludes the paper.

2 Related Work

State-of-the-art results in the broader field of WSD have been recently summarized in the Unified Evaluation Framework¹ (UEF) by Raganato et al. (2017a), which focuses on the *all-words*

¹<http://lcl.uniroma1.it/wsdeval>

disambiguation task based on WordNet 3.0 (Fellbaum, 2012). The top-performing systems in UEF, across supervised and knowledge-based ones, are an SVM model – an extension of the popular IMS system (Zhong and Ng, 2010), and a recurrent neural network model – context2vec (Melamud et al., 2016), with F1 scores on the concatenation of all evaluation corpora ranging between 69% and 69.7%. The supervised models are trained either on the SemCor corpus (Miller et al., 1993) or jointly on SemCor and the semi-automatically constructed OMSTI corpus (Taghipour and Ng, 2015). Very recently deep learning approaches have produced results above the 70% threshold (Luo et al., 2018a,b).

KBWSD has attracted a lot of interest from researchers over the years, since, at least notionally, it does not require any training data or additional resources beyond a computational lexicon and in some cases a knowledge base (KB). Furthermore, if the latter resources are designed in such a way so as to be domain-independent, this could provide a big advantage in dealing with data of heterogeneous origins. One of the earliest popular KBWSD methods is due to Lesk (1986); it takes the dictionary definitions of word senses for target words that occur together in a shared context and calculates the degree of overlap between them, seeking to maximise the latter metric. KBWSD systems, however, typically perform less well than supervised models, due to a number of hurdles, such as: the non-trivial issue of how to structure a KB and what to put inside it; how to explore a KB most effectively; how to integrate various pieces of knowledge into a holistic representation of meaning.

One of the most successful KBWSD approaches has been to use algorithms from the *Random Walks on Graph* family in order to obtain sense representations over particular textual contexts. For instance, Mihalcea (2005) constructs a subgraph with the possible word senses in a context and then runs *PageRank* (Page et al., 1999) over it in order to calculate the most prominent senses. Agirre and Soroa (2009) present an influential update on this method, within the UKB tool for WSD². In addition to the *static* version of *PageRank* (Spr; introduced in Agirre and Soroa (2008)), which also constructs a sub-graph from the WordNet semantic network as a preliminary

²<http://ixa2.si.ehu.es/ukb/>

step, they put forward *Personalized PageRank*, in which the whole KB is used, with the context words serving to inject probability mass into all candidate word senses. The final state of the *PageRank* vector over the graph indicates which are the most relevant concepts in the particular context. Two variants are described: *Ppr* and *Ppr_w2w*, where the second one is modified so as to put additional emphasis on connected concepts and away from the target senses themselves. The second strategy is used to prevent competing but related word senses from bolstering each other inordinately, but also makes the algorithm significantly slower. Another related KBWSD system is *Babelify*, which uses a *Random Walk with Restart* algorithm (Moro et al., 2014).

Two discussion points regarding Random Walks on Graph approaches are central in the context of this article: *knowledge base enrichment* and *algorithm parametrization*. Simov et al. (2016) have shown that the addition of new relation sets to the baseline WordNet 3.0 semantic network can have significant positive effects on the performance of the *PageRank* algorithm. Adding relations extracted from the manually disambiguated word sense glosses, for instance, is a major improvement; including dependency-based relations between manually disambiguated words from SemCor has also led to big error reductions. Therefore the enrichment and structural optimization of the KBs is clearly one possible avenue for the improvement of KBWSD accuracy. On the issue of optimal parameters selection, Agirre et al. (2018) have proposed an updated default parametrization for using the UKB system for *PageRank* KBWSD. They consider various possible changes: changing the length of the context under consideration, the number of iterations, the value of the damping factor, the use of word sense frequency information from WordNet, etc. This new default configuration yields much higher accuracy scores on the UEF evaluation data sets – in fact those are the best reported results for a KBWSD system that we are aware of. They are significantly above the challenging most frequent sense (MFS) baseline, and are not very far even from the accuracy scores reported for supervised systems.

3 Preliminary Experiments

Following the impressive improvements achieved via parameter optimization in Agirre et al. (2018)

and the observations by Simov et al. (2016) that KB extension can lead to significant improvements over using just the baseline WordNet relations, we undertook to combine the contributions of these two lines of research. In these preliminary experiments, version 3.2 of the UKB system has been used, which is set by default to the parameters described in Agirre et al. (2018). The immediate aim of the experimentation was to ascertain whether this optimized parametrization generates analogous positive effects when an extended KB is used in conjunction with the PageRank algorithm.

We have performed an exhaustive combination and evaluation of the relation sets presented in Simov et al. (2016). In several cases we have been able to obtain improvements over the results reported in Agirre et al. (2018). In addition to reusing these sets of relations, we have decided to further enrich the knowledge graph by assigning weights to the relation arcs, making use of a synset embedding model in order to calculate node similarities. We have constructed such a model following the methodology described in Goikoetxea et al. (2015): using the UKB system’s random walk function to produce an artificial corpus of node sequences, then training a Skip-gram model (Mikolov et al., 2013) over it. Thus, for each relation in the KB of the kind (*synset_1*, *synset_2*) its weight is set to the value of the cosine similarity function between the two synsets. The best results for the two types of setting – without and with weights – are reported in Table 1, in lines 9 and 10 respectively. Merely reusing the additional relation sets leads to an improvement of 0.9%. Setting the relation weights via the embedding model produces further gain of 0.4%.

4 Improving the Knowledge Base

In this section we present further changes to the knowledge base in order to improve accuracy. These modifications can be summed up as follows:

- Changing the shape of the graph by representing nodes in terms of word senses
- Defining weights on the knowledge graph arcs via embedding models
- Extraction of new relations from different external sources

We present each of these in turn.

4.1 Sense-Based vs. Synset-Based WSD

When constructing the KB, the UKB system generalizes lexical meaning in a number of ways. One of these is the use of synset IDs as nodes in the semantic network. Here is an example of a relation as represented in the KB distributed together with the UKB system:

```
u:00007846-n v:04618781-n s:derivation
```

where *u:00007846-n* and *v:04618781-n* represent two nodes in the KB and where there exists a relation between them. The source of this connection is a derivational relation encoded in WordNet. *00007846-n* is the identifier³ for the synset including the following senses⁴:

```
person%1:03:00::, individual%1:03:00::,
someone%1:03:00::, somebody%1:03:00::,
mortal%1:03:00::, soul%1:03:00::
```

The second node corresponds to the following synset:

```
personhood%1:07:00::
```

As made apparent by this example, the derivational relation holds only between two of these senses⁵:

```
u:person%1:03:00::-n
v:personhood%1:07:00::-n
s:derivation
```

We have exploited this original representation in order to construct the new KB format. The biggest open question we have been faced with is how to represent the synonymy relation. The first option considered was representing it in the same way as in the WordNet distribution:

```
00007846-n person%1:03:00::-n
00007846-n individual%1:03:00::-n
00007846-n someone%1:03:00::-n
00007846-n somebody%1:03:00::-n
00007846-n mortal%1:03:00::-n
00007846-n soul%1:03:00::-n
```

This approach was our first choice. The relation representation in the KB follows the distinction between *lexical relations* holding among senses and *semantic relations* holding among synsets. After several experiments with this new representation of the KB it became clear that the novel format does not lead to improved accuracy.

Subsequently, we have excluded the synset identifiers from the representation of the KB. In

³This identifier is formed out of the original identifiers in WordNet. The original part-of-speech prefix is deleted and a part-of-speech suffix is added in its place.

⁴A sense in WordNet is defined as a combination of a lemma and a concept represented by a synset.

⁵This is how derivational and other kinds of lexical relations are represented in the original distribution of WordNet.

order to do this, a decision had to be made on how to represent the semantic and the synonymy relations. The option to represent them via calculating complete Cartesian products seems unrealistic. Hence our decision to utilize another source of information encoded in the distribution of WordNet – the ordering of the senses within each synset, which are supposed to represent their relative lexicographic importance. The main node for each synset is thus represented by the first sense in it. All remaining senses are mapped to this node, which is taken to be representative for the synset. All the semantic relations are mapped to the central nodes as well. The lexical relations are represented as before – between the corresponding word senses.

The above synset is now represented as follows:

```
person%1:03:00::-n individual%1:03:00::-n
person%1:03:00::-n someone%1:03:00::-n
person%1:03:00::-n somebody%1:03:00::-n
person%1:03:00::-n mortal%1:03:00::-n
person%1:03:00::-n soul%1:03:00::-n
```

The sense-centric experiments reported in the article have been performed using this format of the KB.

The next step in transforming the KB is the construction of a set of relations extracted from external sources, such as the ones described in Simov et al. (2016). Included in those is the set of external relations distributed together with the UKB system. This set is constructed on the basis of a WordNet gloss corpus, in which some of the words used in the gloss for each synset are annotated with the appropriate word senses from WordNet. Each constructed relation holds between the synset associated with the gloss and the synset of the corresponding sense in the annotation. It is relatively easy to reproduce this type of relation on the basis of the gloss corpus.

The relations represented in the **GraphRelSC** set are extracted from the SemCor corpus. This set of relations includes two types of nodes: for the semantically annotated words and for the nodes in the dependency tree of the corresponding sentence. The dependency analyses of the sentences are not part of the original annotation of SemCor. Thus, those cannot be reused for the construction of sense-based relations. The good news is that we do not need to have the actual dependency annotation, because in **GraphRelSC** the nodes corresponding to the dependency nodes are numbered with the numerical IDs of the documents, the sentences, and the words in the original cor-

pus. Thus, through a simple mapping we have been able to substitute the synset identifiers with the actual sense annotations in the corpus. In Table 1 this new set of relations is referred to as **SC**.

In addition to **GraphRelSC**, we have extracted a set of co-occurrence relations from each sentence in SemCor. This new set of relations is called **SCR**. We have not been able to reconstruct the **WN30gl** set of relations because the mapping from it to the original annotation is not straightforward. Similarly, we have decided not to use the third set of relations, **WN30glCon**.⁶

4.2 Relation Weighting

Each relation in the KB can be assigned an individual weight. These weights are exploited within the ranking algorithms implemented in the UKB system. The original sets of relations do not assign any weights to the arcs. In our preliminary experiments we have assigned each relation a weight determined by the similarity of its associated nodes. This similarity is derived by calculating the cosine similarity of the vector representations for the corresponding nodes.

In order to assign weights, then, appropriate embedding models are necessary to provide vector representations. In the case of the preliminary experiments, we have used a synset embedding model constructed via random walks along the WordNet KB⁷. This is not the case with the new format of the KB where the nodes are represented as senses in WordNet. A direct construction of sense embeddings is also an option, which has not been realized for the purposes of this paper, and will be addressed in future research. Here we use pretrained embeddings.

Each word sense in WordNet is connected to a lemma and a synset. Thus, we could use either the synset embeddings, or the lemma (word) embeddings, or some combination between them. The first option has been disregarded since it does not distinguish between the various senses in the synonym sets. We have performed experiments with pretrained word embeddings such as GoogleNews⁸ and Glove⁹ (the 300-dimensional version),

⁶All new sets of relations, as well as the lexicon, available at <http://bultreebank.org/en/resources/>.

⁷All embedding models referred to here are also made accessible at <http://bultreebank.org/en/resources/>.

⁸<https://code.google.com/archive/p/word2vec/>

⁹<https://nlp.stanford.edu/projects/>

also with lemma embeddings trained analogously to the synset embeddings (see Goikoetxea et al. (2015) on how to generate pseudo corpora from WordNet and subsequently train embedding models on them).

For instance, for the sense `person%1:03:00::-n`, the associated list of synset and lemma IDs includes `00007846-n` and `person`. In the case when only a lemma embedding is available, the corresponding vector is used. In the case of multi-word expressions like `physical_object%1:03:00::-n`, the average of the vectors for the different component words is calculated. In the case when there are both a synset and a lemma embedding available, the concatenation of the two vectors is considered. The following embeddings have been used in the current work:

- **GoogleNews.** A word embedding model trained over 100 billion running words. The vectors are of size 300.
- **Glove.** Word embeddings trained over global contexts, as described in Pennington et al. (2014).
- **WN30WN30glConOne.** Synset and lemma embeddings trained by Simov et al. (2017).
- **WN30WN30glConOneWiki.** In this case a lemmatized Wikipedia corpus has been added to the pseudo corpus, in order to balance information from the knowledge graph with actual text data.

The weights associated with the different word senses in the new lexicon, which also play an important role for the optimal performance of the UKB system, are the same as those associated with synset identifiers in the original lexicon in the UKB distribution. This is so because the frequencies have been originally determined on the basis of sense occurrences in "various semantic concordance texts" (according to the documentation, quoted in Agirre et al. (2018)). Consequently, the conjunction of lemma and synset ID in the lexicon provides a unique mapping to a singular word sense, regardless of the structure of the KB.

4.3 Linking WordNet to VerbNet and FrameNet

VerbNet (Schuler, 2005) and FrameNet (Baker et al., 1998) are structured lexical resources which

[glove/](#)

provide information that is, from a theoretical point of view, complementary to that within WordNet. While the latter's semantic network is a rich representation of lexical semantics, the former give more insight into sentential semantics.

VerbNet classes bring together verbs that share the same syntactic subcategorization and semantic valency patterns. Membership in a VerbNet class does not necessarily indicate that the lexical items have a similar meaning (although that is often the case), but that they share some kind of structurally analogous behaviour, which is certainly a kind of information that is not present in WordNet. For instance, the verb *buy* is in the same class as *hire*, *lease*, *rent*, but in the same class are also the verbs *catch*, *choose*, *pluck*, *slaughter*¹⁰.

FrameNet organizes lexical knowledge around particular procedural scenarios called *frames* (Fillmore, 1968). An example of a FrameNet frame would be *Commerce_buy*¹¹, which specifies *frame elements*, i.e. participants and specifications of the situation, such as *Buyer*, *Goods*, *Seller*, *Means*, *Money*, etc. A frame is activated by certain *lexical units*, in this case *buy*, *buyer*, *client*, *purchase (noun)*, *purchase (verb)*, *purchaser*. Frames can be linked with one another, through relations like *Inherits from*, *Is used by*, etc.

As part of the effort to complement already existing relation sets for KBWSD, the WordNet semantic network has been partially connected to those of VerbNet and FrameNet. The Predicate Matrix resource¹² (De Lacalle et al., 2014, 2016), which automatically maps the WordNet, VerbNet, FrameNet, PropBank and MCR indices, has been used to obtain most of the cross-mappings; in the case of VerbNet, some of the verbs have already been mapped to WordNet senses, but Predicate Matrix can be used to extend the coverage of the mapping. In this way all WordNet sense identifiers that could be mapped to predicates in VerbNet classes and FrameNet frames have been organized in structures that reflect this kind of membership in the external resources. This has been done in a way that is similar to the one described earlier in relation to graphically connecting the various word senses in a synset. That is, the word sense

¹⁰<http://verbs.colorado.edu/verb-index/vn/get-13.5.1.php>

¹¹The frame index is available at <https://framenet.icsi.berkeley.edu/fndrupal/frameIndex>.

¹²<http://adimen.si.ehu.es/web/PredicateMatrix>

N	Knowledge base	SNE-2	SNE-3	SME-07	SME-13	SME-15	ALL
1	WN _{synsets}	64.6	62.8	51.6	67.9	66.6	64.3
2	WN _{senses}	66.5	61.2	51.6	66.5	70.5	64.8
3	WN+VNM _{senses+w}	67.6	61.4	53.2	65.9	71.6	65.3
4	WN+VNM+FNM _{senses+w}	67.4	62.4	53.6	66.0	71.4	65.5
5	WN+VNM+FNM+FNR _{senses+w}	67.1	62.7	54.3	65.8	71.1	65.4
6	WN+GL _{synsets} (Agirre et al., 2018)	68.8	66.1	53.0	68.8	70.3	67.3
7	WN+GL _{senses}	69.0	65.7	55.4	69.2	71.8	67.7
8	WN+GL _{senses+w}	69.3	66.0	55.2	69.4	71.6	67.9
9	WN+GL+SC _{synsets}	70.1	67.0	53.0	69.5	70.9	68.2
10	WN+GL+SC _{synsets+w}	70.4	67.6	53.4	69.5	71.9	68.6
11	WN+GL+SC _{senses}	70.1	67.8	57.4	69.0	72.2	68.8
12	WN+GL+SC _{senses+w+ctx=20,35}	70.2	67.8	58.2	69.1	72.4	68.9
13	WN+GL+SC+VNM+FNM _{senses+w}	69.7	67.6	57.4	68.6	72.5	68.5
14	WN+GL+SC _{senses+w+ctx=10,15,25,30}	70.3	67.9	57.8	69.8	71.8	69.0

Table 1: Accuracy scores on the UEF data sets with different KBs. Only the results for the Ppr_w2w mode of the PageRank algorithm are reported. The synset-based models use KBs where the nodes are represented by synset IDs in WordNet; sense-based models use the new KB configurations described in the paper; the +w subscript means that a model takes into account relation weights. **WN** stands for the original WordNet relations; **GL** – the relations from the annotated gloss corpus; **SC** – the relations from the automatically parsed SemCor corpus; **VNM** – the sense groupings from VerbNet; **FNM** – the sense groupings from FrameNet; **FNR** – the links between FrameNet predicate senses and role-type senses. All experiments use the default parametrization from Agirre et al. (2018), with the exception of the cases marked with a subscript $ctx=num$, where the context windows have been changed to include num words. The best result (line 14) is achieved for contexts with 10, 15, 25 and 30 words. **SNE** stands for *Senseval* and **SME** stands for *SemEval*.

WSD system	SNE-2	SNE-3	SME-07	SME-13	SME-15	ALL
Luo et al. (2018a)	72.8	70.3	—*	68.5	72.8	71.1
Luo et al. (2018b)	72.2	70.5	—*	67.2	72.6	70.6
Raganato et al. (2017b)	72.0	69.1	64.8*	66.9	71.5	69.9
Iacobacci et al. (2016)†	73.3	69.6	61.1	66.7	70.4	69.7
Melamud et al. (2016)†	72.3	68.2	61.5	67.2	71.7	69.4
Agirre et al. (2018)†	68.8	66.1	53.0	68.8	70.3	67.3
Moro et al. (2014)†	67.0	63.5	51.6	66.4	70.3	65.5
WN 1st sense†	66.8	66.2	55.2	63.0	67.8	65.2
This work _{best}	70.3	67.9	57.8	69.8	71.8	69.0

Table 2: The state of the art across WSD systems. The dagger symbol indicates that the result is reported in the UEF (Raganato et al., 2017a). * Luo et al. (2018a,b) do not report accuracy on SME07 since it is used as a development set; this is also true for Raganato et al. (2017b), which however does report the result. The KBWSD models in the table (ours and those by Agirre and Moro) are deterministic, i.e. they would always produce the same results with particular KBs, as no actual training is involved.

in a class/frame with the highest associated frequency of use is promoted to main node status that stands in for the whole structure, then all the rest of the senses are connected to it by setting a connection weight according to their respective frequencies (including a +1-count smoothing). The

verb class hierarchy and inter-frame connection relations are also included in the new subgraph.

Additionally, the automatically created FrameNet extension¹³ by Bryl et al. (2012), which

¹³<https://dh.fbk.eu/technologies/wordnet-sense-repository-framenet-extension>

maps FrameNet roles to WordNet synsets, has been used to add links between predicate word senses grouped in frames and role-filling word senses. These new relations are weighted in accordance with the numbers provided by the generated sense repository, which correspond to frequencies of use in an automatically tagged corpus. The procedures described above should provide higher inter-predicate connection density (the class/frame membership relations) in the KB and also more syntagmatically oriented relations (the FrameNet role senses mapped to frame lexical unit senses).

5 Experiments

In this section we provide several empirical points of view to the currently presented project of extending and optimizing the knowledge graph used for WSD. First we examine how the reconfiguration of the graph in terms of word senses compares with the synset-based graph and show that the new structure outperforms the previous one. In parallel with that we demonstrate that the introduced KB extensions lead to significant improvements over the baseline graphs. We also provide our own contribution to the optimal parametrization of the UKB system.

KB permutations Table 1 shows the various combinations of KB relations, KB structuring and parametrization of the UKB system. Several noteworthy observations come to the fore:

1. By comparing lines 1&2, 6&7, 9&11, it becomes clear that the KBs structured around word senses perform better than those where nodes are represented by synset IDs.
2. The addition of relations extracted from the VerbNet-WordNet-FrameNet (gold and automatic) mappings does improve the baseline results over the WN relations (lines 3-5). VNM builds over WN accuracy, FNM builds over WN+VNM, while FNR does not seem to decisively improve results (though on some data sets it does help – SNE-3 and SME-07).
3. The addition of the gloss and SemCor relations has a very significant effect on accuracy when compared to using just the baseline WN relations (lines 6-14).
4. The VN-FN relations do not seem to reliably improve accuracy when added to the gloss and SemCor relations, in fact they improve

- accuracy only on one data set (SME-15) and in the rest of the cases bring it down (line 13).
5. The default parameters from Agirre et al. (2018) are indeed a good optimization of the UKB system. We have been able to improve the result with the best KB only in one case, which we report here, and the improvement is not very big (0.1%; line 14). The result nevertheless indicates that there is space for optimizing the interaction between KB and algorithm.

Comparison with state-of-the-art models Table 2 situates our best result in the context of the state-of-the-art results in WSD at large. Again, several observations are worth pointing out:

1. The combination of UKB with our best-performing graph comfortably beats the WN 1st sense heuristic, which is not the case for many WSD systems.
2. The model significantly outperforms all KB-WSD models reported in the UEF, including the improved parametrization of Agirre et al. (2018), whose results we have improved upon with 1.7%.
3. Our result is at this point very close to the top-performing WSD systems, regardless of whether they are supervised or not. This has typically not been the case for KBWSD systems. For one data set (SME-13) the present model achieves the highest result of all¹⁴ and for another one (SME-15) it is better than all but the two leading supervised systems.

Improvements with static PageRank Finally, in table 3 we show the performance of three different versions of the PageRank algorithm, as implemented in the UKB system and described in Agirre and Soroa (2009). The results with the static version have been very significantly improved in comparison to the results reported in Agirre et al. (2018). The static mode performs better than the WN 1st sense heuristic when used with the specified graphs. The difference between the $w1$ and $w2$ models is as follows: with $w1$ weights are set via a combination of **WN30WN30glConOne** embeddings for synsets and lemmas, and with $w2$ – via a combination of

¹⁴SME-13 contains sense annotations of nouns only; with its extensive taxonomic network, WordNet is a powerful tool for representing nominal meaning. More detailed analysis is required to ascertain the real reasons for the high accuracy scores.

Knowledge base	Spr	Ppr	Ppr_w2w
WN+GL _{synsets} (Agirre et al., 2018)	57.7	65.6	67.3
WN+GL+SC _{senses+w1}	65.5	65.2	68.9
WN+GL+SC _{senses+w2}	66.5	65.1	68.0
WN+GL+SC+SCR _{senses+w2}	66.6	65.1	68.1

Table 3: Comparison between different PageRank versions (accuracy measured on all UEF test data sets). **Spr** stands for *static* mode, **Ppr** is *Personalized PageRank*, **Ppr_w2r** is *Ppr with emphasis on neighbouring concepts*. *w1* denotes weights set via the combination of **WN30WN30glConOne** embeddings for synsets and lemmas; *w2* denotes the combination of **WN30WN30glConOneWiki** embeddings.

WN30WN30glConOneWiki embeddings, again for synsets and for lemmas. The **SCR** relation set, which is built on the basis of co-occurrences of word senses in SemCor, contributes positively to the best static model. Compared with previous state-of-the-art results when using static PageRank, here we can see an improvement of nearly 9%.

Bearing in mind that the static version of the algorithm is much faster than the personalized ones, these results should also be interpreted as important, as they demonstrate that KB improvement might be even more beneficial for less sophisticated methods that nevertheless offer a good trade-off in terms of speed of execution.

6 Conclusion

We have presented results from a series of experiments with a KBWSD system with state-of-the-art default parametrization and have shown that accuracy can be further improved through the manipulation and extension of the KB. The present models achieve the highest reported accuracy scores for a KBWSD system that we are aware of; they also enter in the close orbit of the highest-scoring supervised systems, achieving a new state of the art on the Semeval-13 data set.

Further improvement to the content of the KB is certainly possible. Based on the reported experiments, an intuition emerges that the relation weighting schema has to have a dynamic character. This would correspond to the promotion and demotion of semantic features within context. The current implementation relies on the PageRank algorithms to maintain this kind of dynamics. The experiments also demonstrate that the different weighting schemata improve the performance of different algorithms, which suggests complex patterns of interaction between algorithm and graph structure. The current schemata do not distinguish

between different kinds of relations. The difference between paradigmatic and syntagmatic relations necessitates different weighting approaches. Using the cosine similarity measure over the entire embedding space seems to be a suboptimal blanket strategy. In our future research we plan to train relation embeddings, following the approach for generating pseudo corpora, as in the case of training synset and lemma embeddings.

Additionally, we plan on further investigating strategies for generalizing knowledge from external resources such as VerbNet and FrameNet, as well as other ones which can be mapped to WordNet, such as PropBank (Kingsbury and Palmer, 2002) and the OntoNotes sense groupings (Snow et al., 2007). As has been demonstrated, structuring a lexico-semantic graph in an optimal way can make a big difference. Detailed error analysis and sophisticated linguistic theory should be employed in order to capture the principles underlying a good knowledge base.

Acknowledgements

This research has been funded by the Bulgarian National Science Fund grant number 02/12/2016 — *Deep Models of Semantic Knowledge (DemoSem)*. Alexander Popov’s contribution has been also supported by the Bulgarian Ministry of Education and Science under the National Research Programme “Young scientists and postdoctoral students” approved by DCM # 577 / 17.08.2018.

References

- Eneko Agirre, Oier Lopez de Lacalle, and Aitor Soroa. 2018. The risk of sub-optimal use of open source NLP software: UKB is inadvertently state-of-the-art in knowledge-based WSD. In *Proceedings of Workshop for NLP Open Source Software (NLP-OSS)*, pages 29–33. Association for Computational Linguistics.

- Eneko Agirre and Aitor Soroa. 2008. Using the multilingual central repository for Graph-based Word Sense Disambiguation. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. European Language Resources Association (ELRA). [Http://www.lrec-conf.org/proceedings/lrec2008/](http://www.lrec-conf.org/proceedings/lrec2008/).
- Eneko Agirre and Aitor Soroa. 2009. Personalizing PageRank for Word Sense Disambiguation. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 33–41, Athens, Greece. Association for Computational Linguistics.
- Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley Framenet Project. In *Proceedings of the 17th International Conference on Computational Linguistics - Volume 1, COLING '98*, pages 86–90, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Volha Bryl, Sara Tonelli, Claudio Giuliano, and Luciano Serafini. 2012. A novel framenet-based resource for the semantic web. In *Proceedings of the 27th Annual ACM Symposium on Applied Computing, SAC '12*, pages 360–365, New York, NY, USA. ACM.
- Maddalen Lopez De Lacalle, Egoitz Laparra, Itziar Aldabe, and German Rigau. 2016. Predicate Matrix: automatically extending the semantic interoperability between predicate resources. *Language Resources and Evaluation*, 50(2):263–289.
- Maddalen Lopez De Lacalle, Egoitz Laparra, and German Rigau. 2014. Predicate Matrix: extending SemLink through WordNet mappings. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland. European Language Resources Association (ELRA).
- Christiane Fellbaum. 2012. Wordnet. *The Encyclopedia of Applied Linguistics*.
- Charles J. Fillmore. 1968. The case for case. In E. Bach and R. T. Harms, editors, *Universals in Linguistic Theory*, pages 1–88. Holt, Rinehart, and Winston.
- Josu Goikoetxea, Aitor Soroa, and Eneko Agirre. 2015. Random walks and neural network language models on knowledge bases. In *HLT-NAACL*, pages 1434–1439. The Association for Computational Linguistics.
- Ignacio Iacobacci, Mohammad Taher Pilehvar, and Roberto Navigli. 2016. Embeddings for Word Sense Disambiguation: An Evaluation Study. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 897–907, Berlin, Germany. Association for Computational Linguistics.
- Paul Kingsbury and Martha Palmer. 2002. From treebank to propbank. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*. European Language Resources Association (ELRA).
- Michael Lesk. 1986. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *Proceedings of the 5th Annual International Conference on Systems Documentation, SIGDOC '86*, pages 24–26, New York, NY, USA. ACM.
- Fuli Luo, Tianyu Liu, Zexue He, Qiaolin Xia, Zhifang Sui, and Baobao Chang. 2018a. Leveraging gloss knowledge in neural word sense disambiguation by hierarchical co-attention. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1402–1411. Association for Computational Linguistics.
- Fuli Luo, Tianyu Liu, Qiaolin Xia, Baobao Chang, and Zhifang Sui. 2018b. Incorporating Glosses into Neural Word Sense Disambiguation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2473–2482. Association for Computational Linguistics.
- Oren Melamud, Jacob Goldberger, and Ido Dagan. 2016. context2vec: Learning Generic Context Embedding with Bidirectional LSTM. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, pages 51–61, Berlin, Germany. Association for Computational Linguistics.
- Rada Mihalcea. 2005. Unsupervised large-vocabulary word sense disambiguation with graph-based algorithms for sequence data labeling. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 411–418, Vancouver, British Columbia, Canada. Association for Computational Linguistics.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- George A. Miller, Claudia Leacock, Randee Tengi, and Ross T. Bunker. 1993. A Semantic Concordance. In *Proceedings of the Workshop on Human Language Technology, HLT '93*, pages 303–308, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Andrea Moro, Alessandro Raganato, and Roberto Navigli. 2014. Entity linking meets word sense disambiguation: a unified approach. *Transactions of the Association for Computational Linguistics*, 2:231–244.

- Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The PageRank Citation Ranking: Bringing Order to the Web. Technical Report 1999-66, Stanford InfoLab. Previous number = SIDL-WP-1999-0120.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543. Association for Computational Linguistics.
- Alessandro Raganato, Jose Camacho-Collados, and Roberto Navigli. 2017a. Word Sense Disambiguation: A unified evaluation framework and empirical comparison. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, volume 1, pages 99–110.
- Alessandro Raganato, Claudio Delli Bovi, and Roberto Navigli. 2017b. Neural Sequence Learning Models for Word Sense Disambiguation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1156–1167, Copenhagen, Denmark. Association for Computational Linguistics.
- Karin Kipper Schuler. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, University of Pennsylvania.
- Kiril Simov, Petya Osenova, and Alexander Popov. 2016. Using Context Information for Knowledge-Based Word Sense Disambiguation. In *Artificial Intelligence: Methodology, Systems, and Applications*, pages 130–139, Cham. Springer International Publishing.
- Kiril Simov, Petya Osenova, and Alexander Popov. 2017. Comparison of word embeddings from different knowledge graphs. In *Language, Data, and Knowledge*, pages 213–221, Cham. Springer International Publishing.
- Rion Snow, Sushant Prakash, Daniel Jurafsky, and Andrew Y. Ng. 2007. Learning to merge word senses. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 1005–1014, Prague, Czech Republic. Association for Computational Linguistics.
- Kaveh Taghipour and Hwee Tou Ng. 2015. One million sense-tagged instances for word sense disambiguation and induction. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, pages 338–344, Beijing, China. Association for Computational Linguistics.
- Zhi Zhong and Hwee Tou Ng. 2010. It makes sense: A wide-coverage word sense disambiguation system for free text. In *Proceedings of the ACL 2010 System Demonstrations*, pages 78–83, Uppsala, Sweden. Association for Computational Linguistics.