ACL 2020 reviews for the submission "Don't Neglect the Obvious: On the Role of Unambiguous Words in Word Sense Disambiguation"

Overall recommendation scores: 4-4-3.5 (Scale 1-5)

Review #1

What is this paper about, what contributions does it make, what are the main strengths and weaknesses?

This paper presents a method of extending the data set that allows propagation methods to utilize more data to estimate sense vectors to do word sense disambiguation with. The method involves a very simple, but apparently so-far overlooked idea, namely to sense-tag all unambiguous words in a very large corpus (Unambiguous Word Annotation; UWA). Simple as it may seem, extending the training set with such data was found to lead to considerable improvement. I think this is a clever contribution, and a good case of making progress in NLP without tinkering on increasingly complex learning models. Such work should, to my mind, be encouraged and I highly recommend this paper to be accepted. The experiments are worked out well, and the paper is well written.

Reasons to accept

I think this is a clever contribution, and a good case of making progress in NLP without tinkering on increasingly complex learning models. Such work should, to my mind, be encouraged.

Reasons to reject

As a smart use of data, it is perhaps a little on the light side as a free-standing contribution (even for a short paper), but to be honest, I'm not to bothered by that.

Overall Recommendation: 4

Questions for the Authors(s)

One point I would have liked to see more elaboration is the discussion of the results. To my mind, the discussion of the number of examples needed can be compressed quite a bit and instead some attention can be given to two questions that came up: (1) why is there such a striking difference between SemEval data sets in the amount of improvement the UWA approach yields. Comparing it to OMSTI: multiple percentage points in SE2, SE3, SE07, but hardly any improvement in SE13 and SE15. Relatedly, more in-depth analysis into the kinds of words, senses, contexts for which the UWA approach was found to improve over other approaches would substantiate the insights of this paper.

Missing References

-

Typos, Grammar, and Style

l124: their method further require[s] from [?] a Wikipedia corpus

Additional Suggestions for the Author(s)

Bender Rule: this paper does not mention on which language it is tested (aside from noticing that English Wikipedia was used; line 161). The degree and nature of the polysemy of words in a language varies across languages, and not observing that the scope of this work is English gives the (in my opinion) false suggestion that this work can be succesfully applied regardless of language.

A comment on the clarity of the writing is that it took me until page 3 to understand how the approach would benefit WSD. This likely has to do with me not being up to date with the latest developments in WSD (who can keep up with the state of the art in every task these days?), but it would benefit the paper's appeal to a wider audience if the high-level intuition were conveyed early on in the paper. Unpacking the phrase 'standard propagation algorithms' (which presupposes knowledge of those algorithms) into something along the lines of 'these unambiguous words allow standard propagation algorithms to estimate sense representations for unseen meanings of ambiguous words more accurately' would help the reader who doesn't know those algorithms before reading the paper. Otherwise, the paper is very well written and clear.

Review #2

What is this paper about, what contributions does it make, what are the main strengths and weaknesses?

The paper introduces a method for constructing a corpus of unambiguous word annotations and show the influence of such a resource for word sense disambiguation (WSD). To construct such a corpus they use two corpora (wikipedia and web text), perform some pre-processing (lemmatization, POS, tokenization) and keep only sentences where a lemma has only one meaning in WordNet. In total thez cover 98494 senses and have 10 sentences for 81.2% of these senses. To show the performance using this resource they use an existing LMMS that is based on BERT and uses WordNet to infere senses. Based on an existing WS evaluation framework they observe the best scores when using their corpus in combination with SemCor.

Furthermore, they investigate how many sentences of the single-sense-words they require to obtain good results. In their experiments they reveal that already 3 sentences is enough and using more sentences does not yield to any further improvement.

The paper is a great contribution to the conference: it introduces a new dataset and also shows its impact. Furthermore, the authors also show some analysis about how many sentences of unambiguous word annotations are required to observe improvements.

Reasons to accept

- new resource for WSD
- presenting results using existing method
- state-of-the-art performance
- analysis of various parameters (number of sentences used)

Reasons to reject

- the contribution is a minor one, however, I think this should not be considered as a drawback as the paper is also a short paper

Overall Recommendation: 4

Questions for the Authors(s)

- How many of the UWA sentences are already covered in SC?
- is there some correlation between the frequency of the words and the number of words needed to improve the WSD performance?
- what is the influence of the NER filtering? (also adding some example in the paper would be nice)

- will the resource be made publicly available?

Additional Suggestions for the Author(s)

- not sure if CoreNLP can be considered as state-of-the-art for lemmatization and pos tagging

Review #3

What is this paper about, what contributions does it make, what are the main strengths and weaknesses?

The first contribution of the paper is a method to construct a corpus of Unambiguous Words Annotations, the second contribution is to show how existing WSD models can benefit from this corpus. Overall, I liked the idea to bootstrap the unambiguous words to generate more data. Along these lines probably more things also can be done, i.e. to align monosemous words in different languages, etc. The setup of the experiment is sound to me. The results are encouraging.

Main strengths:

- A new automatically constructed corpus of Unambiguous Word Annotations (UWA), that extends the coverage of SemCor Evaluation of WSD model LMMS on a combined corpus of SemCor and UWA.

Main weakness

- Methodologically the paper is very simple and may not be innovative enough in this respect.

Reasons to accept

A new resource for WSD has been constructed with a wide coverage of unambiguous words.

Reasons to reject

A very simple idea and technical approach (yet it works).

Overall Recommendation: 3.5

Author Response

General Response to Reviewers:

We thank the reviewers for their supportive comments and for bringing to our attention a few aspects that require further clarification.

--R1--

[(1) why is there such a striking difference between SemEval data sets in the amount of improvement the UWA approach yields. Comparing it to OMSTI: multiple percentage points in SE2, SE3, SE07, but hardly any improvement in SE13 and SE15.]

There are two major factors at play in this comparison: 1) the test sets vary significantly in the senses they cover; 2) the related corpora include annotations for ambiguous words which overlap with these test sets in varying proportions. For example, SC+OMSTI increases the number of examples for ambiguous senses by 8.7x for SE13 and 4.3x for the other test sets, when compared with SC. Since we compared UWA w/SC, where all annotations for ambiguous words come from SC, this helps explain the difference observed in the case of SE13, which is the only dataset where OMSTI provides improvements with respect to SemCor. This is indeed a very relevant question that will be clarified and further analyzed in the revised version of our paper.

[Relatedly, more in-depth analysis into the kinds of words, senses, contexts for which the UWA approach was found to improve over other approaches would substantiate the insights of this paper.]

We agree that such an analysis would be helpful and we're planning to use the extra-page in the revised version for this purpose. We'll report on improvements when using UWA by part-of-speech and other factors.

[Bender Rule: this paper does not mention on which language it is tested (aside from noticing that English Wikipedia was used; line 161). The degree and nature of the polysemy of words in a language varies across languages, and not observing that the scope of this work is English gives the (in my opinion) false suggestion that this work can be succesfully applied regardless of language.]

We understand the concern and will make this clear in the revised version.

[A comment on the clarity of the writing is that it took me until page 3 to understand how the approach would benefit WSD. [...] it would benefit the paper's appeal to a wider audience if the high-level intuition were conveyed early on in the paper.]

We thank the reviewer for this suggestion. We'll make adjustments to the early sections so that the main contributions are discussed earlier in the paper.

--R2--

[How many of the UWA sentences are already covered in SC?]

The overlap between the unique senses covered by UWA and SemCor is of 8.9%. This means that 91.1% of the unambiguous words in UWA are not covered in SemCor.

[Is there some correlation between the frequency of the words and the number of words needed to improve the WSD performance?]

We are not sure if we fully understood the question. In this paper, focusing just on unambiguous words, we show that after a small number of examples the WSD performance stops improving, irrespective of their overall frequency. We believe the effect of word frequency would be more impactful on the number of required examples for ambiguous words, as frequency and polysemy have been shown to correlate (e.g. Yaghoobzadeh et al. ACL 2019).

[What is the influence of the NER filtering? (also adding some example in the paper would be nice)]

The main motivation for adding this NER filtering is to avoid mistaking named entities and concepts. For instance, in the following example the NER filter helped remove a false positive: "*Inception* was a box-office hit." In this case 'Inception' made reference to a movie and not to the unambiguous word 'inception' from WordNet. We'll add the example to the paper. This reduction in false positives (discarding sentences where a word is an unknown named entity) also reduced the number of senses present in UWA by about 5% (from 103K to 98K, approximately), due to senses that were left without any example sentences.

[Will the resource be made publicly available?]

Yes, we'll freely release the UWA dataset (and improved sense embeddings) upon acceptance.

--R3--

We thank the reviewer for their supportive comments and for acknowledging the simplicity of our approach. We also appreciate their interesting suggestion for future work of aligning monosemous words in different languages.