Word Sense Disambiguation using a Bidirectional LSTM

**Kun Li**

Department of Computer Science, the University of Texas at Dallas

[Kxl162530@utdallas.edu](mailto:Kxl162530@utdallas.edu)

**Abstract**

In this paper I present a clean, yet effective, model for *word sense disambiguation*. My approach leverages a *bidirectional long short-term memory* network which is shared between all words. Central to this technology is a learned word vector embedding that has separate word vectors for each sense of a word. The model is trained end-to-end, directly from the raw text to sense labels, and makes effective use of word order. I evaluate my approach on two standard datasets, using identical hyperparameter settings, which are in turn tuned on a third set of held out data. I employ no external resources (e.g. knowledge graphs, part-of-speech tagging, etc.), language specific features, or hand-crafted rules, but still achieve statistically equivalent results to the best state-of-the-art systems.

**1 Introduction**

It’s common to see word vector embeddings as ways to represent words that are fed into machine learning models. These word vectors are often trained on large text corpora and convey some sort of relationship relative to surrounding context words. However, many common word vector embeddings would include the same vector representation of a word even though it might be used in different ways between contexts. Consider the following two examples for instance:

1. He sat down beside the Seine river **bank**.
2. He deposited the money at the Chase **bank**.

Although ‘bank’ is clearly being used in two different senses in these two examples, they would be represented by the same vector. Providing the sense in which a word is used could give machine learning systems valuable information that can contribute to how well they can perform a task. Word Sense Disambiguation (WSD) is considered an AI-Complete problem (Navigli 2009), meaning it is one of the centrally difficult problem in AI, so potential solutions and effective approaches towards this problem can have far-reaching implication for other key AI tasks. Consider the task of named entity recognition (NER), for instance. In NER, we attempt to classify different entities as organization, people, places, etc. In the first case, we should identify “bank” as a location instead of an organization, as is the case in the second sentence. By being able to predict word senses, we can help clarify these ambiguous situations. Furthermore, we can develop separate word vectors for each sense, which can ideally allow us to better represent words within their own context.

In this paper, I describe one WSD algorithm. It is based on a Long Short Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997). Since this model can take into account word order when classifying, it perform significantly better than a continuous bag of words model (Word2vec) (Mikolov et al., 2013; Iacobacci et al., 2016).

*Organization:* I review related work in Section 2. I introduce dataset in section 3, and the supervised WSD algorithm in Section 4. Experimental results are discussed in Section 5. We provide further discussion in Section 6 and future work in Section 7.