**Verb Sense Induction using Phrase Embeddings**

**Dan Prendergast and Sam Young**

**Introduction**

One of the quintessential problems of computational linguistics is the ambiguity of words in a language. Consider the word “run.” In some contexts, “run” can mean movement under an agent’s own direction, as in “to run a marathon.” In another context, the word can mean to exercise leadership over a group or entity, as in “to run a company.” This type of ambiguity makes automated processing of text and speech difficult. Without the ability to make the distinction between senses, an automated assistant may mistake your intent to “run a marathon this March” as an intent to plan and coordinate a sporting event in March.

**Problem**

The task of identifying the proper meaning of these different uses of words is called word sense disambiguation (WSD). WSD can largely be divided into supervised learning, unsupervised learning, and knowledge-based approaches (Navigli, 2009). The knowledge-based approaches, such as WordNet (Miller, 1995), are often used as a gold standard for evaluating other WSD methods. The disadvantage of supervised learning and knowledge-based approaches is the considerable effort required to create manually annotated datasets. On the other hand, unsupervised methods do not require labeled data. The disadvantage of unsupervised learning is that they cannot assign sense labels. Instead they perform word sense induction (or discrimination). That is, they can identify when occurrences of a word fall into different sense classes, but they cannot apply semantic labels to the classes.

Klapaftis and Manandhar (2013) provide overview an of word sense induction (WSI) methods. These methods can be categorized as Bayesian, graph-based, or vector-based approaches. Bayesian approaches such as Brody and Lapata (2009) and Blei et al. (2003) use a generative approach to WSI using probability distributions of words and senses. The graph-based methods (Dorow and Widdows, 2003; Véronis, 2004; Agirre et al., 2006b) first construct a graph with context words and target words as vertices and co-occurrences of those words as edges. After the corpus is ingested and the graph built, graph reduction algorithms are used to find the significant clusters of vertices (Klapaftis and Manandhar, 2013). Lastly, vector-based WSI uses a co-occurrence matrix to develop a vector representation for a target word. Various clustering techniques can then be applied to the vectors to induce the separate word senses as in Pedersen (2007), Niu et al. (2007), and Pinto et al. (2007).

**Approach**

We intend to induce verb sense distinctions using verb-object phrase embeddings. Our approach is based on the assumption that the object of a verb adds context that assists in disambiguating the verb sense. (Dligach and Palmer, 2008) shows that by utilizing the semantic context of verb objects using a dependency-parsed corpus, error rates of verb disambiguation tasks can be decreased by as much as 15%.

One method of creating phrase embeddings is through simple composition of individual word embeddings. Mitchell and Lapata (2008) formalize additive and multiplicative composition models. Their work shows that compositions involving multiplication of subject and verb embeddings performed better than simple addition on a sentence similarity task (Mitchell and Lapata, 2008).

In this project, we will create phrase embeddings through composition of verb and object word embeddings. Then we will evaluate the effectiveness of discriminating different verb senses by clustering these phrase embeddings. We hypothesize that verb senses can be effectively induced using verb-object phrase embeddings. The advantage of this approach is that it can be performed locally on a local sentence-by-sentence basis, unlike the graph-based or Bayesian WSI approaches described above. Such a system could be useful in speech recognition or translation systems where disambiguation must be performed in real-time.

**Evaluation**

In this project we will evaluate the effectiveness of multiple word embedding sets and composition methods in the task of disambiguating verb senses. We will use two pre-trained word embedding sets -- one using the word2vec algorithm (Mikolov, 2013) available from <https://github.com/idio/wiki2vec/> and the other using the GloVe algorithm (Pennington, Socher, and Manning, 2014) available from <https://nlp.stanford.edu/projects/glove/>. We will select 20 verbs from the PropBank labeled dataset that demonstrate multiple senses. Each occurrence of the verb will be extracted from the PropBank dataset along with its object and tagged sense. For each verb-object pair, we will calculate phrase embeddings using two different composition methods -- cross product and addition. Using a 2 x 2 factorial design we will vary the word embedding set and composition method to evaluate all 4 configurations.

To evaluate the results, we will perform k-means clustering for each of the phrases with the same verb, where k is the number of senses for the verb found in the PropBank dataset. We will determine the best cluster-to-sense alignment using overall accuracy of sense assignment to the PropBank gold standard. To make this more precise, let be the sense provided to the verb-object phrase by the label in PropBank, let be the assignment of a sense to a particular cluster, and let N be the number of unique verb-object phrases for a particular verb. The accuracy for a particular assignment is given by…

This accuracy will indicate overall success in verb sense disambiguation. We will compare this to the accuracy we could expect to achieve from random assignment of the verb-object phrases to one of the senses found in the PropBank dataset.

**Paper Presentations:**

Dan will present… Jeff Mitchell and Mirella Lapata. 2008. Vector-based models of semantic composition. In Association for Computational Linguistics (ACL).

Sam will present… Mandelbaum, Amit, and Adi Shalev. "Word embeddings and their use in sentence classification tasks." *arXiv preprint arXiv:1610.08229* (2016).

**References**

Agirre, E., Martı´nez, D., Lo´pez de Lacalle, O., & Soroa, A. (2006b). Two graph-based algorithms for state-of-the-art wsd. In Proceedings of the conference on empirical methods in natural language processing (pp. 585–593). Sydney, Australia: ACL.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. J. Mach. Learn. Res., 3, 993–1022.

Brody, S., & Lapata, M. (2009). Bayesian word sense induction. In Proceedings of the 12th conference of the european chapter of the association for computational linguistics, EACL ’09 (pp. 103–111). Stroudsburg, PA, USA: Association for Computational Linguistics.

Dligach, Dmitriy, and Martha Palmer. "Novel semantic features for verb sense disambiguation." *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*. Association for Computational Linguistics, 2008.

Dorow, B., & Widdows, D. (2003). Discovering corpus-specific word senses. In Proceedings of the 10th conference of the European chapter of the ACL (pp. 79–82). Budapest, Hungary: ACL.

Kim, Joo-Kyung, Marie-Catherine de Marneffe, and Eric Fosler-Lussier. "Neural word embeddings with multiplicative feature interactions for tensor-based compositions." *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*. 2015.

Kingsbury, Paul, and Martha Palmer. "From TreeBank to PropBank." *LREC*. 2002.

Klapaftis, Ioannis P., and Suresh Manandhar. "Evaluating word sense induction and disambiguation methods." *Language resources and evaluation* 47.3 (2013): 579-605.

Mandelbaum, Amit, and Adi Shalev. "Word embeddings and their use in sentence classification tasks." *arXiv preprint arXiv:1610.08229* (2016).

Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781* (2013).

Miller, George A. "WordNet: a lexical database for English." *Communications of the ACM* 38.11 (1995): 39-41.

Jeff Mitchell and Mirella Lapata. 2008. Vector-based models of semantic composition. In Association for Computational Linguistics (ACL), pages 236–244.

Navigli, Roberto. "Word sense disambiguation: A survey." *ACM computing surveys (CSUR)* 41.2 (2009): 10.

Niu, Z.-Y., Ji, D.-H., & Tan, C.-L. (2007). I2r: Three systems for word sense discrimination, chinese word sense disambiguation, and english word sense disambiguation. In Proceedings of the fourth international workshop on semantic evaluations (SemEval-2007) (pp. 177–182). Prague, Czech Republic: Association for Computational Linguistics.

Pedersen, T. (2007). Umnd2: Senseclusters applied to the sense induction task of senseval-4. In Proceedings of the fourth international workshop on semantic evaluations (pp. 394–397). Prague, Czech Republic: ACL.

Pennington, Jeffrey, Richard Socher, and Christopher Manning. "Glove: Global vectors for word representation." *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 2014.

Pinto, D., Rosso, P., & Jime´nez-Salazar, H. (2007). Upv-si: Word sense induction using self term expansion. In Proceedings of the fourth international workshop on semantic evaluations (SemEval2007) (pp. 430–433). Prague, Czech Republic: Association for Computational Linguistics.

Véronis, J. (2004). Hyperlex: lexical cartography for information retrieval. Computer Speech & Language, 18(3), 223–252.