Pei Lun Hung
CS 410
Professor ChengXiang Zhai
November 15th, 2020

Technology Review

1. Introduction. Accurately representing words for use in models is a challenging task, bearing on at least two different issues. The first is with regards to word usage—the syntactical and semantical intricacies of how a word functions within its surrounding context. The second is with respect to meaning: a given word may carry a host of different connotations across distinct contexts. For example, "leaf" can mean one thing when the surrounding context is about nature, and another when the document is about binary trees. ELMo (Embeddings from Language Models) representations seek to directly combat these two challenges.

In another vein, ULMFiT (Universal Language Model Fine-tuning) addresses additional issues in natural language processing. Namely, fine-tuning language models is often difficult, requiring task-specific documents or training from scratch. ULMFiT tackles these concerns and allows for vigorous inductive transfer learning for any natural language processing task.

2. Applications.

2.1 *ELMo* ELMo word representations are "functions of the entire input sequence" (Peters 2018). That is, they aim to capture contextualized word embeddings[1], using vectors resulting from a

[1] Thus, ELMo word representations differ, crucially, from most common word embeddings—such as in GloVe (Global Vectors for Word Representation), where the word "leaf" is represented by the same vector cross-contextually.

bidirectional LSTM trained on a precise job. Peters et al. found that ELMo is performative across six target natural language processing tasks[2].

Because ELMo uses vectors derived from a bidirectional LSTM, ELMo-enhanced models have both a sense of the next word in a sequence of words *and* a sense of the preceding word. As a result, incorporating ELMo into a model results in more efficient training: Peters et al. found that models enhanced with ELMo need much less training data to reach a given target performance level. For a semantic role labeling task that they simulated, an ELMo model was able to hit the same F_1 measure as the non-ELMo baseline model by using just 1 percent of the training data; the baseline model, however, required a full 10 percent of the training data to achieve the same results. ELMo, then, can be used to enhance existing models, serving as a steppingstone for natural language processing tasks and facilitating faster convergence on training data.

2.2 ULMFiT ULMFiT, proposed by Howard and Ruder, is an "extremely sample-efficient transfer learning method that can be applied to any NLP task" (Howard and Ruder 2018). It drastically surpasses preceding transfer learning techniques in terms of performance by allowing for new fine-tuning techniques[3]—techniques that enable better learning while, at the same time, mitigating "catastrophic forgetting."

The three stages of ULMFiT—language model pre-training, language model fine-tuning, and classifier fine-tuning—incorporate these techniques and quicken the learning process. Making use of ULMFiT on future models may allow researchers to discover new state-of-the-art models,

[2] Namely: question answering, textual entailment, semantic role labeling, coreference resolution, named entity extraction, and sentiment analysis.

[3] Howard and Ruder describe discriminative fine-tuning, slanted triangular learning rates, and gradual unfreezing as techniques to retain knowledge and mitigate forgetting during the fine-tuning process.

and, because it works well with small datasets, ULMFiT might also be useful in performing tasks that have a significant amount of unlabeled data.

3. Conclusion. Converting a large corpus of text data into vector representations is an important tool for computational linguists and natural language processing researchers, first introduced by the vector space model. Developments in the last few years have introduced the concept of representing words in a *contextualized* manner, wherein the vector representation of a word is contingent not solely on the vocabulary word's distribution in the training data set, but also on the particular context surrounding the word. ELMo and ULMFiT are powerful tools that can encode semantic and syntactic information about a word, capture contextual information surrounding a word occurrence, and enable fine-tuning techniques.

Works Cited

Howard, Jeremy, and Sebastian Ruder. “Universal Language Model Fine-Tuning for Text Classification.” *ACL Anthology,* Association for Computational Linguistics, July 2018, www.aclweb.org/anthology/P18-1031/.

Peters, Matthew, et al. “Deep Contextualized Word Representations.” *ACL Anthology*, Association for Computational Linguistics, June 2018, www.aclweb.org/anthology/N18-1202/.