Interactive Medical Word Sense Disambiguation with Instance and Feature Labeling

Yue Wang, MS[1], Kai Zheng, PhD[2], Hua Xu, PhD[3], Qiaozhu Mei, PhD[4,1]
[1]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI; [2]Department of Informatics, University of California, Irvine, CA; [3]School of Biomedical Informatics, University of Texas Health Science Center, Houston, TX; [4]School of Information, University of Michigan, Ann Arbor, MI

Introduction

Medical documents contain many ambiguous words. Word sense disambiguation (WSD), i.e. assigning the appropriate meaning to an ambiguous word in context, is a critical step for many medical natural language processing applications, such as named entity extraction and computer-assisted review. Previous works have proposed many approaches for medical WSD, including unsupervised learning, supervised learning, semi-supervised learning, active learning, and interactive search and classification. However, these approaches still lack the ability to learn from domain knowledge, and take relatively long time to reach a reasonable performance, which is known as the "cold-start" problem. In this study, we design a novel interactive learning algorithm that directly incorporates expert knowledge into the WSD model training process. We consider two types of expert knowledge in the WSD task: prior knowledge and reasoning process. The expanded form of many ambiguous abbreviations are documented in medical ontologies, and the expert may know contextual words of a particular sense before looking at any instance, both of which are prior knowledge. Upon seeing an instance containing an ambiguous word, the expert can pinpoint the words and phrases that support his decision on the word sense, which is the reasoning process hidden behind the label. In the new interactive learning process, an expert can express knowledge as the association between senses and textual patterns, and the machine learning algorithm can directly learn from such association. Experiments on one biomedical literature corpus and two clinical notes corpora show that the proposed algorithm makes better use of human efforts in training WSD models than all previous approaches, achieving the state-of-the-art performance with the least effort from domain experts.

Methods

The proposed algorithm has an interactive learning component and two computational components (Figure 1A).

(1) Labeling instances and features. In the labeling process, a domain expert can come up with informative contextual words, label instances, and highlight informative text snippets as "rationales" behind the label decision. As a result, the expert provides two types of supervision: instance labels and feature labels. An instance label is assigned to an individual instance containing an ambiguous word, as in traditional machine learning. A *feature label* is assigned to a word, a phrase, or a textual pattern, all of which can be viewed as features. When the expert *provides* informative words for a label (sense), these words are explicitly associated with the label. When the expert *highlights* words in an instance, the highlighted words are implicitly associated with the label of that instance.

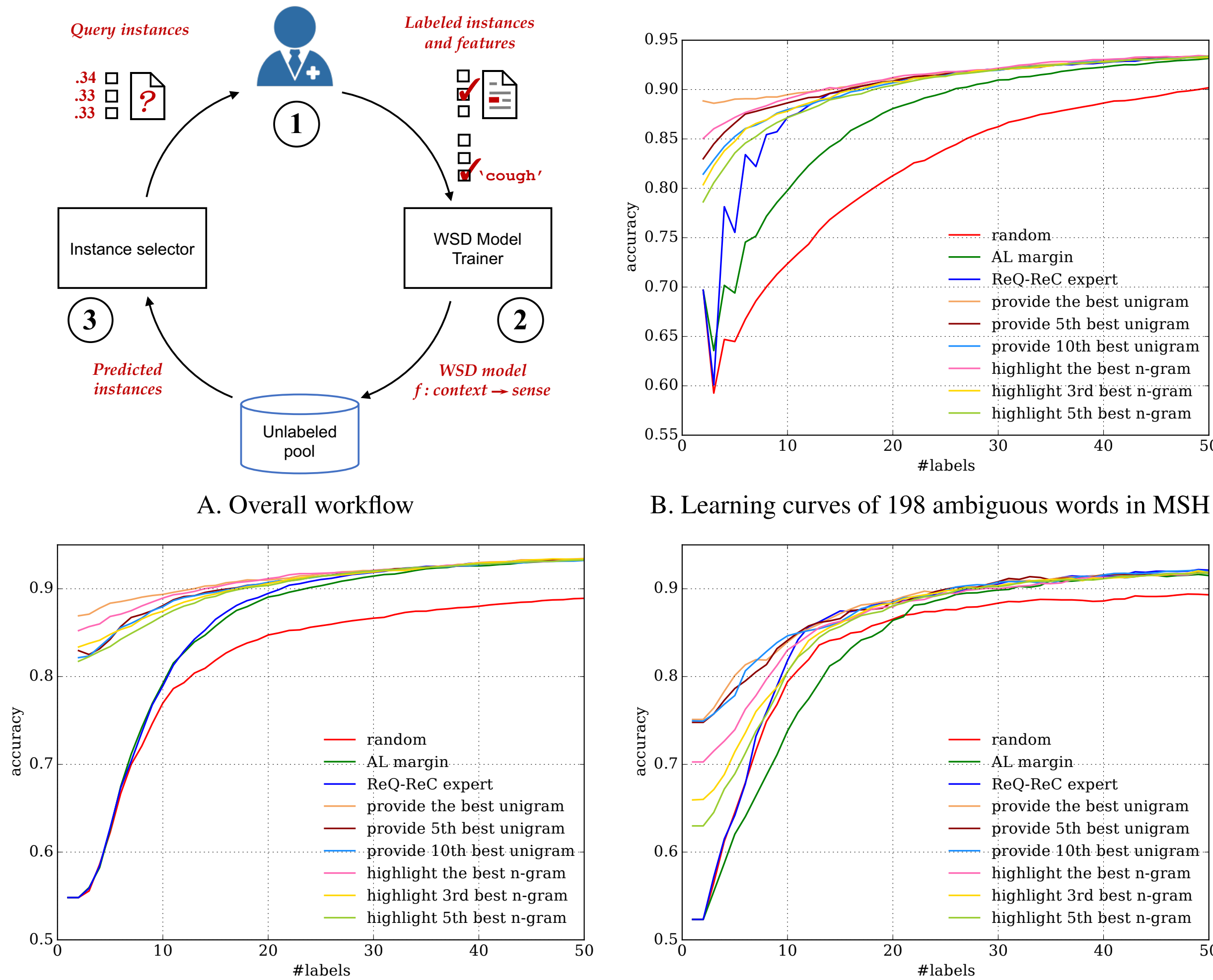
(2) Learning from labeled instances and features. The machine learning module takes in both instance labels and feature labels from (1) to train a WSD classifier. At each iteration, labeled features may contain arbitrary textual patterns (e.g. n-grams) that are hard to know beforehand. We start with a set of base features (unigrams) and dynamically expand the feature set as new features are provided/highlighted. We use logistic regression with a linear kernel as the WSD classifier. We train the classifier by adding a new term to the original logistic regression loss function, such that any instance containing a labeled feature is more likely to bear that label. A labeled feature may carry substantially more information than a labeled instance if the feature appears in many instances.

(3) Selecting instances. An instance selector picks up new instances from the unlabeled pool, presents them to the expert, and asks for labels. We use the margin active learning algorithm to select uncertain instances.

To evaluate the new algorithm, we compare it with several baseline methods, including random sampling, margin active learning[1], and ReQ-ReC expert[2]. We use two settings of the proposed algorithm: providing labeled features and highlighting informative features in labeled instances. To make head-to-head comparison with a previous work[2], we use the same evaluation settings. We simulate an expert by leveraging benchmark medical WSD corpora, including MSH (biomedical literature), UMN, and VUH (clinical notes). Labeled instances are sense-tagged examples in these corpora, and labeled features are n-grams ($n = 1, 2, 3$) with high information gain for an ambiguous word. In all experiments, the expert provides only **one** labeled feature at the beginning; all subsequent labels are instance labels. To make realistic assumption on feature labeling, we simulate experts that provide or highlight the k-th best feature.

Results

Figure 1B, 1C, and 1D show the learning curves of different algorithms on three WSD corpora. The new interactive learning algorithm outperforms strong baseline methods. On biomedical literature WSD corpus (MSH), the new algorithm achieves 90% accuracy with 15 labels, saving 40% of labeling effort compared to active learning. WSD in clinical notes is more difficult. On UMN corpus, the new algorithm achieves 90% accuracy with 15 labels, saving 35% of labeling effort compared to active learning. On VUH corpus, the new algorithm saves labels at the beginning.



A. Overall workflow

B. Learning curves of 198 ambiguous words in MSH

C. Learning curves of 75 ambiguous words in UMN

D. Learning curves of 24 ambiguous words in VUH

Figure 1: A: Diagram of the proposed interactive learning process; B,C,D: Learning curves on three WSD corpora.

Discussion

The proposed approach effectively handles the "cold-start" problem in active learning. Active learning works best when the model has a reasonably good "understanding" of the problem space so that the selected instances are the most informative. At the beginning, the model trained on very few labeled instances can perform poorly. In the new learning process, human experts can kick off training by providing domain knowledge, giving the model a warm start.

References

1. Chen Y, Cao H, Mei Q, Zheng K, Xu H. Applying active learning to supervised word sense disambiguation in MEDLINE. Journal of the American Medical Informatics Association. 2013;20(5):1001–1006.

2. Wang Y, Zheng K, Xu H, Mei Q. Clinical word sense disambiguation with interactive search and classification. In: AMIA Annual Symposium Proceedings. vol. 2016. American Medical Informatics Association; 2016. .