

Word Sense Disambiguation by Semi-Supervised Learning

Zheng-Yu Niu¹, Dong-Hong Ji¹, Chew-Lim Tan², and
Ling-Peng Yang¹

¹ Institute for Infocomm Research
21 Heng Mui Keng Terrace, 119613 Singapore
{zniu, dhji, lpyang}@i2r.a-star.edu.sg

² Department of Computer Science, National University of Singapore
3 Science Drive 2, 117543 Singapore
tanc1@comp.nus.edu.sg

Abstract. In this paper we propose to use a semi-supervised learning algorithm to deal with word sense disambiguation problem. We evaluated a semi-supervised learning algorithm, local and global consistency algorithm, on widely used benchmark corpus for word sense disambiguation. This algorithm yields encouraging experimental results. It achieves better performance than orthodox supervised learning algorithm, such as kNN, and its performance on monolingual benchmark corpus is comparable to a state of the art bootstrapping algorithm (bilingual bootstrapping) for word sense disambiguation.

1 Introduction

In this paper, we address the problem of word sense disambiguation (WSD), which is to assign an appropriate sense to an occurrence of a word in a given context. Many learning algorithms have been proposed or investigated to deal with this problem, including knowledge or dictionary based algorithms, and corpus based algorithms. Corpus based algorithms can be categorized as supervised learning algorithms, weakly supervised learning algorithms [1, 3, 5–8], and unsupervised learning algorithms. In WSD task, we often face a shortage of labeled training data, but there is a large amount of unlabelled data which can be cheaply acquired. As a result, a great deal of work [1, 3, 5–8] have been devoted to effective usage of unlabeled data for improving the performance of WSD systems.

Here we use a semi-supervised learning algorithm [9] to perform WSD. Compared with other weakly supervised learning based WSD algorithms, such as bootstrapping or co-training, semi-supervised learning algorithm explores the manifold structure to determine the labels of unlabeled points. Secondly, bootstrapping and co-training require that the class distribution should be fixed during the iteration procedure to avoid degenerate solutions.

This paper is organized as follows. In section 2 we will define feature vector and distance measure for WSD. In section 3 we will describe the semi-supervised