A Mapping between Classifiers and Training Conditions for WSD

Aarón Pancardo-Rodríguez,[1] Manuel Montes-y-Gómez,[1,2]
Luis Villaseñor-Pineda,[1] Paolo Rosso [2]

[1] National Institute of Astrophysics, Optics and Electronics, Mexico
{aaron_cyberman, mmontesg, villasen}@inaoep.mx
[2] Polytechnic University of Valencia, Spain.
{mmontes,prosso}@dsic.upv.es

Abstract. This paper studies performance of various classifiers for Word Sense Disambiguation considering different training conditions. Our preliminary results indicate that the number and distribution of training examples has a great impact on the resulting precision. The Naïve Bayes method emerged as the most adequate classifier for disambiguating words having few examples.

1 Introduction

The objetive of Word Sense Disambiguation (WSD) is to distinguish between the different senses of a word, that is, to identify the correct sense of a word in a context. The state of the art of WSD [1] shows that the supervised paradigm is the most efficient. Under this approach, the disambiguation process is carried out using information that is estimated from data. Several statistical and machine learning techniques have been applied to learn classifiers from disambiguated corpora. For instance, statistical classifiers, decision trees, decision lists, memory-based learners, and kernel methods such as Support Vector Machines (SVM).

The comparison among the different approaches to WSD is difficult. The last edition of the Senseval competition showed that the SVM is emerging as one of the most powerful supervised techniques for WSD [3]. Although important, this comparison focuses on the entire systems as black boxes, and does not consider the details about the individual classifiers and the fine tunning of their paramethers.

Some researchers have attempted to compare the performace of classifiers under equal training conditions. For instance, Paliouras et al [2] disambiguated all content words from Semcor using various classifiers (e.g., J48, Naïve Bayes, PART, k-nn and a decision table). Their results indicated that the decision tree induction outperforms other algorithms. Zavrel et al [4] investigated the performance of some classifiers (neuronal networks, memory-based leraning, rule induction, decision trees, maximum entropy, winnow perceptrons, Naïve-Bayes, and SVM) and some ensembles on a diverse set of natural language processing tasks. Their results showed that the SVM algorithm is the most prommising for WSD.

In the study of the global execution of some classifiers, we focus our attention on providing information about the behaviour of the classifiers under different training