ON THE USAGE OF THE NAÏVE BAYES MODEL IN AUTOMATIC WORD SENSE DISAMBIGUATION

IULIAN DANIEL BUZDUGAN, ANCA CIOATES

The present survey paper analyzes the usage of the Naïve Bayes model in automatic word sense disambiguation, with special focus on supervised and unsupervised techniques. The most recent results in the field are discussed, especially with respect to unsupervised word sense disambiguation (Hristea and Popescu 2009), (Hristea, Popescu and Dumitrescu 2009). With regard to supervised disambiguation the authors study the possibility of enriching the discussed mathematical model with Markov chains and come to the conclusion that, despite common sense expectations, this enrichment does not improve disambiguation results.

Keywords: Word sense disambiguation, Supervised disambiguation, Unsupervised disambiguation, Bayesian classification, The EM algorithm, WordNet.

1. Introduction

Word Sense Disambiguation (WSD) consists in determining the meaning of a word in the specific context where it occurs. WSD is a core research problem in computational linguistics[1] and natural language processing and was recognized as such since the beginning of the scientific interest in machine translation[2] and in artificial intelligence, in general. WSD has many important applications not only in natural language understanding, but also in: machine translation, speech processing, text processing etc. Its most important application is nowadays considered to be information retrieval (web search engines).

The problem of WSD is defined as that of computationally determining the sense of a word in a specific context and represents a classification problem.

---

[1] The initiator of mathematical and computational linguistics in Romania is Solomon Marcus [see, for instance, (MARCUS, 1966)]. Professor Marcus is equally well known as one of the most prominent initiators of this field worldwide.

[2] Machine translation has been anticipated since 1933 by Smirnov-Troianski (MARCUS, NICOLAU and STATI, 1966). In 1957, in New-York, I.B.M. performed the first public demonstration of automatic translation from one language to another. In Romania the first automatic translations are performed in 1962, in Timisoara, under the supervision of Gr. C. Moisil. The algorithm was elaborated by E. Domonkoş, being implemented on MECIPT (DOMONKOŞ, 1962).