Chapter 20

Word-Sense Disambiguation by Examples

Taijiro Tsutsumi

Abstract

This chapter describes a method of disambiguating multi-sense words in a sentence by using example sentences in which such words are already disambiguated, and by using taxonyn and synonym hierarchies. As a knowledge base, we developed a small-scale text database containing 730 example sentences in English that include the verb “take,” and prototyped a program that resolves 12 senses of the verb “take” in the input sentences. Our test results show the feasibility of our approach. The advantages of the approach are: (1) it does not require special semantic categorization; (2) the knowledge base is easy to create and maintain; (3) closely related senses, as in polysemous words, can be disambiguated; and (4) the approach is robust.

This chapter is reprinted from Tsutsumi 1991.
Much effort has been devoted to research on resolving word-sense ambiguity, since this is one of the most fundamental and important issues in natural language understanding. There seem to be two major approaches to the issue (Fass 1988). One is the approach based on selectional restrictions, Preference Semantics, the case structure of verbs, and so on. It usually involves the use of scripts, a semantic network, or frames as semantic representations of world knowledge. The other includes mechanisms for spreading activation and marker passing, and seeks the links between two nodes that represent word-senses in a semantic network. The idea of association between words, in particular, is becoming much more important in this area (Allen 1987, Hirst 1987).

However, most approaches are not powerful enough for natural language applications that by their nature require wide-coverage and high-quality processing, such as machine translation and automatic text summarizing. We often see two kinds of compromise: systems that disambiguate only easy cases, and systems that disambiguate hard cases as well as easy cases, but only in a limited domain. The bottleneck is the difficulty of building a large-scale and accurate knowledge base, usually in the form of a semantic network.

Recent computer technology has made it easy to analyze a large volume of texts, create a large text database, store it on high-speed access devices, and retrieve the data fast. Our approach, described in this chapter, does not use a popular type of knowledge base, such as a semantic network, but instead uses a text database containing example sentences, and taxonym and synonym hierarchies. We describe the background of our approach in section 20.2. Section 20.3 gives an approximate definition of closeness between words and between sentences, which is a basis for evaluating the plausibility of an interpretation of an input sentence. Section 20.4 describes the procedure for disambiguating word-senses. Sections 20.5 and 20.6 describe experiments and conclusions, respectively.

### 20.2 Background of our approach

#### 20.2.1 Relations between sentences

If two sentences are significantly analogous in terms of words and syntax, they may share some common meaning. We would expect lexical and syntactic ambiguities in an input sentence to be resolved to some extent by using analogies with sentences that relate to the input sentence. Let us consider the following sentences, in which only the word “take” is ambiguous: