LEDIR: An Unsupervised Algorithm for Learning Directionality of Inference Rules

Rahul Bhagat, Patrick Pantel, Eduard Hovy
Information Sciences Institute
University of Southern California
Marina del Rey, CA
{rahul,pantel,hovy}@isi.edu

Abstract

Semantic inference is a core component of many natural language applications. In response, several researchers have developed algorithms for automatically learning inference rules from textual corpora. However, these rules are often either imprecise or underspecified in directionality. In this paper we propose an algorithm called LEDIR that filters incorrect inference rules and identifies the directionality of correct ones. Based on an extension to Harris's distributional hypothesis, we use selectional preferences to gather evidence of inference directionality and plausibility. Experiments show empirical evidence that our approach can classify inference rules significantly better than several baselines.

1 Introduction

Paraphrases are textual expressions that convey the same meaning using different surface forms. Textual entailment is a similar phenomenon, in which the presence of one expression licenses the validity of another. Paraphrases and inference rules are known to improve performance in various NLP applications like Question Answering (Harabagiu and Hickl 2006), summarization (Barzilay et al. 1999) and Information Retrieval (Anick and Tipirneni 1999).

Paraphrase and entailment involve inference rules that license a conclusion when a premise is given. Deciding whether a proposed inference rule is fully valid is difficult, however, and most NL systems instead focus on plausible inference. In this case, one statement has some likelihood of being identical in meaning to, or derivable from, the other. In the rest of this paper we discuss plausible inference only.

Given the importance of inference, several researchers have developed inference rule collections. While manually built resources like WordNet (Fellbaum 1998) and Cyc (Lenat 1995) have been around for years, for coverage and domain adaptability reasons many recent approaches have focused on automatic acquisition of paraphrases (Barzilay and McKeown 2001) and inference rules (Lin and Pantel 2001; Szpektor et al. 2004). The downside of these approaches is that they often result in incorrect inference rules or in inference rules that are underspecified in directionality (i.e. asymmetric but are wrongly considered symmetric). For example, consider an inference rule from DIRT (Lin and Pantel 2001):

$$X \text{ eats } Y \Leftrightarrow X \text{ likes } Y \quad (1)$$

All rules in DIRT are considered symmetric. Though here, one is most likely to infer that "*X eats Y*" ⇒ "*X likes Y*", because if someone eats something, he most probably likes it[1], but if he likes something he might not necessarily be able to eat it. So for example, given the sentence "*I eat spicy food*", one is mostly likely to infer that "*I like spicy food*". On the other hand, given the sentence "*I like rollerblading*", one cannot infer that "*I eat rollerblading*".

In this paper, we propose an algorithm called **LEDIR** (pronounced "leader") for **LE**arning **D**irectionality of **I**nference **R**ules. Our algorithm filters incorrect inference rules and identifies the directionality of the correct ones. Our algorithm

[1] There could be certain usages of "*X eats Y*" where, one might not be able to infer "*X likes Y*" (for example metaphorical). But, in most cases, this inference holds.

Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 161–170, Prague, June 2007.