Noun Compound Interpretation

Girishkumar Ponkiya          Pushpak Bhattacharyya
girishp@cse.iitb.ac.in       pb@cse.iitb.ac.in

Girish K. Palshikar
gk.palshikar@tcs.com

Abstract

Noun compounds (or noun sequences) are a productive, continuous sequence of more than one nouns. Most of the noun compounds appear only once in a large corpus. These characteristics of noun compounds make them a special case, and demand special treatment. Here, the problem is to find noun compounds from text, parse them if required, and extract semantic relation between components of the noun compound. A task of extracting an abstract relation between components of the noun compound (e.g., apple pie: Made_Of), or paraphrasing noun compound using verb and prepositions (apple pie: “a pie made of apple” or “a pie with apple flavor”), is known as interpretation of noun compound (or noun compound interpretation). For our work, we use a set of predefined abstract labels as semantic relations.

Following are major bottlenecks in current system, and our approaches to solve the same:

1. There is no acceptable inventory of semantic relations. We have analyzed a inventory of semantic relations[1], and we trying to refine the inventory. We are also planning to use a data driven approach which will help in refining the current inventories.

2. Inspite of millions of noun compounds in large corpora, there is no sufficiently large annotated dataset for supervised training. We are planning to use semi-supervised approach to tackle this.

3. Context influences semantic relations. But, the present datasets have annotation for each noun compound without context. We are planning to study how context can be used for the task. We are planning to use web-extracted information to bring the context in play.

References