Restaurant Review Problem Statement

Jason D. M. Rennie
jrennie@csail.mit.edu

July 7, 2004

We are intererested in the general problem of extracting and summarizing information contained in informal communication. Restaurant discussion bulletin-boards provide an interesting domain for extracting and summarizing information. There are a number of boards where people discuss and review restaurants. These boards tend to be much more up-to-date, extensive and opinionated than other venues. But, it can be difficult to locate all the reviews of a particular restaurant and summarize the opinions. Search provides an effective solution for finding threads that discuss a particular restaurant. But, search does a poor job of resolving references. Many discussions mention the restaurant name sparingly, making full use of the context of previous posts; many posts that review a restaurant never mention the restaurant by name. We are interested in doing the job that search cannot: automatically determining context so that each sentence (or part of sentence) may be linked to the appropriate restaurant(s). Once context is resolved, one can present review information in a more accessible and readable form, listing only the sentences that are relevant to the restaurant in question. Given the ability to determine sentiment, one can also compile a summary sentiment, indicating the number of people who like/dislike the restaurant.

In summary, we set forth the following task:

- For each restaurant,
  - Find all discussion of that restaurant, and
  - For each person who discusses the restaurant, determine his/her sentiment (positive or negative) about the restaurant.

This work has implications that go beyond extracting information from restaurant reviews. Identifying named entities and resolving referneces are two important natural language problems. Most work on these problems have been done on more structured domains such as newspaper text. Our extraction and resolution techniques will apply to other informal communication domains, such as non-restaurant product reviews, personal e-mail, mailing lists and other discussions.