

DEPARTMENT OF COMPUTER SCIENCE

MPhil Degree Oral Presentation

MPhil Candidate:	Mr Tianjie ZHAN
Supervisor:	Dr Chun Hung LI
External Examiner:	Prof Jian Huang LAI
Time:	29 October 2010 (Friday) 4:00 pm – 6:00 pm (35 mins presentation and 15 mins Q & A)
Venue:	RRS 702A, Sir Run Run Shaw Building, HSH Campus

“Semantic Analysis for Extracting Finer-grained Opinion Aspects”

Abstract

Web documents of user-generated content such as forum discussion and product reviews, attracted research interests in opinion mining. Previous works on extraction of opinion aspects and words obtain lexical features based on grammatical rules and WordNet-like lexical resources together with machine learning analyses. Rule based approaches involve in using syntactical and semantic rules to learn the co-occurrence patterns of relevant lexicons for identifying opinion and its aspect. To enhance identification of words' semantic similarity, the WordNet-like lexical database provides semantic relations, e.g. synonyms and hypernyms, for distinguishing distinct opinion aspects and sentiments. But such approaches often represents sentence as “bag of word” or “sequence of word” without fully utilizing semantic dependency relations in a sentence which is essential for extracting semantic relevant words to represent opinion aspects or opinions. To address above problems, this dissertation considers two issues on the topic of extracting fine-grained opinion aspects in sentence level.

The first issue is how to take full advantage of semantic information within a sentence. We introduce a syntactic dependency parser as baseline to develop our sentential semantic parser for learning the composition structure of a sentence. Starting with the atomic sentence which contains only one verb predicate comprising of a subject, main verb and object, compound sentences and complex sentences are parsed into several atomic sentences within a hierarchical layered structure. For example, a compound sentence could be parsed as an atomic sentence with its subject or object as an atomic sentence. Based on the assumption that parts of a atomic sentence can be divided into core parts (subject-verb-object or subject-verb) and other parts for modifying the components of core parts, atomic sentence can be parsed

into representation comprising of a <subject, verb, object> triple and a set of head-modifier dependency relations.

The second issue is how to identify distinct finer-grain opinion aspects in discriminative semantic orientation. To address the problem, we focus on identifying discriminative clusters of semantic relevant lexicons to indicate distinct opinion aspects, among which noun is most representative and essential for describing specific opinion aspect. To develop related noun features and their semantic dependent context features, we extract noun fragments and their semantic neighbor features which are determined by their adjacency relationship with noun fragment within the sentence structure. Given the obtained noun fragments and their semantic neighbors, we propose two approaches to extract finer-grained opinion aspects in an unsupervised manner. The first approach is to employ non-negative matrix factorization (NMF) method on the co-occurrence frequency matrix between noun fragments and their semantic neighbor features. To combine the two views of noun-sentence co-occurrence and association relationship between noun fragment and its semantic neighbors, the second approach takes adjectives as the representative of semantic neighbor features, and a tri-factorization generative model is developed to factorize the corpus along three views of noun fragments, adjectives and sentences to generate semantic relevant clusters for opinion aspect representation.

Finally, the experimental results demonstrate the advantage of our NMF-based approach over the conventional approaches based on noun-sentence co-occurrence frequency matrix. In the second approach, experimental results indicate that our tri-factorization model outperforms the baseline models based on latent Dirichlet allocation. Effectiveness of the two approaches is shown by improved numerical measurement as well as recognition of semantically cohesive fine-grained opinion aspects.

***** ALL INTERESTED ARE WELCOME*****