ABSTRACT

Opinions of people are an important source of information in our everyday life. Nowadays, online customer reviews have become an invaluable source of opinions. Besides helping customers to make a more informed purchase decision, online reviews are also beneficial for service providers to obtain free feedback. For trendy products or services, there often exists an overwhelming amount of reviews in which manual human analysis is burdensome. Hence, automated opinion mining is used to extract important discussed features (aspect) and their related comments (sentiment). As extraction of correct aspect-sentiment pairs is critical for the outcome of overall opinion mining, in this thesis, we are motivated to improve the extraction in review sentences. Current works have limitations in terms of identifying non-noun-noun compound noun and parent-child relationship aspects in the extraction. The non-noun-noun compound noun aspect consists of compound noun in which either one or both of words are not noun such as hard (adjective) disk (noun) and boot (verb) up (particle). The parent-child aspect consists of general aspect (parent) and specific aspect (child). In the sentence, "The screen of laptop is big", the parent-child aspect is "Screen of laptop" in which parent is "laptop" and the child is "screen". To address this problem, an aspect-sentiment pair extraction using rules and compound noun lexicons (ASPERC) model is proposed. The rules and compound noun lexicons enable to extract more aspects by identifying non-noun-noun compound noun and parent-child aspects which eventually contribute to more aspect-sentiment pair extraction. The experiment is conducted with SemEval dataset consisting of 250 laptop training sentences as well as 63 laptop and 63 restaurant testing sentences for BE baseline (9 rules), ASPERC (63 rules and compound noun lexicons) and ASPER (63 rules only) models to compare their performance in extraction. The experimental result confirms that ASPERC and ASPER have higher recall and slightly lower precision than BE

baseline. The recalls have incremented by 28.58% (ASPERC) and by 22.55% (ASPER) in laptop domain and 11.2% in restaurant domain compared to BE baseline. The precisions have declined by 2% (ASPERC) and 6% (ASPER) in laptop domain and declined by 1.09% (ASPERC) and by 2.12% (ASPER) in restaurant domain compared to baseline BE. Based on the results, ASPERC and ASPER models are able to extract more aspect sentiment pairs with satisfactory correct pairs. The reasonable outcome of ASPER also indicates that the rules are applicable across various domain.