

Abstract:

Awareness of others' opinions plays a crucial role in the decision making process performed by simple customers to top-level executives of manufacturing companies and various organizations. Today, with the advent of Web 2.0 and the expansion of social networks, a vast number of texts related to people's opinions have been created. However, exploring the enormous amount of documents, various opinion sources and opposing opinions about an entity have made the process of extracting and analyzing opinions very difficult. Hence, there is a need for methods to explore and summarize the existing opinions. Accordingly, there has recently been a new trend in natural language processing science called "opinion mining". Primary researches in this area were limited to extracting and analyzing the positive or negative sentiments of individuals from textual information. In recent years, new opinion mining researches have progressed more towards being feature-based. Our main goal in this study is to provide a feature-level opinion mining and a structured summarization of opinions (the users' sentiments separated by text features). Current methods of sentiment detection have many weaknesses in dealing with linguistic and conceptual complexities.

Therefore, considering the complexities of the structured summarizer of opinions, the semantic approach has been employed in this study in order to extract features, detect sentiment related to each feature and finally express the relations and display the results. The main idea of this research is to consider the meaning (knowledge) to detect the sentiment of a text by applying a semantic framework in order to use it integratedly in all stages of opinion summarization. Thus, a semi-automatic method is proposed for creating the opinion ontology. The purpose of this framework is to convert the vast number of opinions from the unstructured texts to the form of structured semantic data in a reasonable time. Applying the proposed ontology helps to fully identify various features and their relations with the main entity as well as the expressed sentiment for each feature.

Prior to the construction of opinion ontology, a conceptual model of opinions (an ontology schema), independent of the language and the opinions domain, is proposed in order to keep them in a structured format. In the following step, the target language is selected and a semi-automatic method (using the basic knowledge) is proposed to construct the opinion ontology from documentation (in a particular domain). Then, the ontology of the constructed opinions is applied to detect features, analyze their sentiments and group them. Finally, a summary of the opinions is presented in a structured form and in format of semantic data. The results of the evaluations have indicated that the proposed framework provides a good quality for the structured opinion summarization.

In addition, in line with the main purpose of the thesis, a corpus of Persian opinions, a glossary of general and specific sentiment words (for commercial goods), a set of various Persian text-processing tools required for opinion mining and also FerdowsNet for Persian have been provided.