

Title	電子カルテからのアスペクトベースの感情分析
Author(s)	Sanglerdsinlapachai, Nuttapong
Citation	
Issue Date	2019-09
Type	Thesis or Dissertation
Text version	ETD
URL	<a href="http://hdl.handle.net/10119/16182">http://hdl.handle.net/10119/16182</a>
Rights	
Description	Supervisor: Dam Hieu Chi, 知識科学研究科, 博士

# Abstract

Sentiment analysis is a process of understanding an opinion in a written or spoken language. It may be applied at different scales, ranging from phrases to a whole document. Instead of determining the sentiment of an entire text portion, aspect-based sentiment analysis addresses sentiments corresponding to parts, components, attributes, or aspects of an entity of interest, which are mentioned in the given text portion. This dissertation studies how a linguistic structure is used to improve aspect-based sentiment analysis and how to apply sentiment analysis to a document in medical domain, especially, a clinical narrative.

In our study, an aspect mentioned in a text portion is first detected, and elementary discourse units (EDUs) relevant to the aspect are then localized by using the linguistic structure, i.e., the rhetorical structure theory (RST). Using lexicon-based approaches, the polarity scores of terms occurring in an EDU are combined into the polarity score of the EDU. We propose a new score aggregation strategy that utilizes RST to aggregate scores from all EDUs relevant to the aspect. Experimental results on online product reviews demonstrate that our new score aggregation method improves sentiment classification at the level of local aspect segments.

To apply the proposed method to clinical text in electronic medical records (EMRs), some extensions are required. The medical-domain-knowledge corpus, i.e., the Unified Medical Language System (UMLS), is employed to detect aspects mentioned in a clinical narrative. Local aspect segments are then formed by using RST. However, occurrences of medicine-technical terms, e.g., disease names or treatment processes, make the sentiment on a clinical narrative hard to analyse. For example, the sentiment of the text portion “Appears to have premature atrial contraction with bundle showing” depends greatly on the meaning of the term “premature atrial contraction”. Semantic types of technical terms, provided by UMLS, are incorporated into lexicon-based sentiment classification methods of two types, i.e., methods using a generic sentiment lexicon and those using a trained sentiment lexicon. Preliminary results show that different classification methods are appropriate for text portions containing different semantic types. Classifier combination is then employed to select a classification method that is most suitable for an input text portion.

*Keywords:* Aspect-Based Sentiment Analysis, Lexicon-Based Sentiment Classification, Rhetorical Structure Theory, Clinical Narrative, Unified Medical Language System