

# Chapter 29

## Rule-Based Polarity Aggregation Using Rhetorical Structures for Aspect- Based Sentiment Analysis

**Nuttapong Sanglerdsinlapachai**

*Sirindhorn International Institute of Technology, Thammasat University, Pathumthani, Thailand &  
Japan Advanced Institute of Science and Technology, Ishikawa, Japan*

**Anon Plangprasopchok**

*National Electronics and Computer Technology Center, Pathumthani, Thailand*

**Tu Bao Ho**

*John von Neumann Institute, Vietnam National University, Ho Chi Minh City, Vietnam & Japan  
Advanced Institute of Science and Technology, Ishikawa, Japan*

**Ekawit Nantajeewarawat**

*Sirindhorn International Institute of Technology, Thammasat University, Pathumthani, Thailand*

### **ABSTRACT**

*The segments of a document that are relevant to a given aspect can be identified by using discourse relations of the rhetorical structure theory (RST). Different segments may contribute to the overall sentiment differently, and the sentiment of one segment may affect the contribution of another segment. This work exploits the RST structures of relevant segments to infer the sentiment of a given aspect. An input document is first parsed into an RST tree. For each aspect, relevant segments with their relations in the resulting tree are localized and transformed into a set of features. A set of classification rules is subsequently induced and evaluated on data. The proposed framework performs well in several experimental settings, with the accuracy values ranging from 74.0% to 77.1% being achieved. With proper strategies for removing conflicting rules and tuning the confidence threshold, f-measure values for the negative polarity class can be improved.*

DOI: 10.4018/978-1-6684-6303-1.ch029

## INTRODUCTION

Sentiment analysis or opinion mining is an interesting research topic in recent years (Pang & Lee, 2008). It is widely applied to textual information in various domains, i.e., reviews of products and services by customers (Hu & Liu, 2004; Ding et al., 2008; Jo & Oh, 2011), financial criticism on microblogs (Si et al., 2013), medical records (Denecke & Deng, 2015), and online news (Chen & Li, 2017). Instead of determining the sentiment of the whole text, several research works, e.g., Jo and Oh (2011) and Moghadam and Ester (2012), addressed sentiments at smaller levels such as parts, components, attributes, or aspects of an entity of interest. This kind of analysis is referred to as aspect-based sentiment analysis. It basically involves two main tasks, i.e., aspect extraction and sentiment classification of extracted aspects (Liu & Zhang, 2012).

For aspect extraction, descriptive statistics and topic modelling techniques have been used; for example, term frequencies and Latent Dirichlet Allocation (LDA) were used in the works of Hu and Liu (2004) and Jo and Oh (2011), respectively. For determining aspect sentiments, two major approaches, machine-learning-based and lexicon-based, have been applied. These two approaches, however, still perform poorly when they are applied to complicated text with rich linguistic structures. Consider, for example, the sentence “The new phone is fine, but its battery still lacks capacity for one-day use”, which should be classified as negative with respect to the “power” aspect. Since the terms “fine” and “lacks” indicate positive and negative sentiments, respectively, this sentence may be misclassified as neutral by using a lexicon-based method. For training a machine-learning classification model, when this sentence is used as a negative instance, terms in the first clause “The new phone is fine” may be incorrectly taken as features for the negative class. The use of linguistic structures makes it possible to consider only the second clause “but its battery still lacks capacity for one-day use” to be relevant to the “power” aspect, and the undesired effects of terms occurring in the first clause can be eliminated.

Recently, attempts to utilise linguistic structures to the sentiment analysis have been reported. Polanyi and van den Berg (2011) discussed the application of the Linguistic Discourse Model to sentiment analysis and observed that a discourse structure affects sentiments at the discourse level; however, no experimental result was reported. Zirn, Niepert, Stuckenschmidt, and Strube (2011) applied the Rhetorical Structure Theory (RST) (Mann & Thompson, 1988) to indicate parts of text that are relevant to a sentiment in product reviews, and an overall sentiment score was calculated from the indicated parts, with the sentiment classification accuracy of 69% being achieved. Sanglerdsinlapachai, Plangprasopchok, and Nantajeewarawat (2016) applied RST to identify text portions relevant to specific aspects in a dataset containing mobile phone reviews, and the sentiment classification accuracy of 72.4% was obtained by averaging the polarity scores of the relevant parts.

In this study, the authors step further by investigating the dependence of the overall sentiment of a given group of related text portions and the polarity scores of its main part (nucleus) and complement parts (satellites) and examining the effects of RST relation types on polarity score aggregation. Sentiment classification rules are induced from RST components, i.e., nuclei, satellites, and/or relation types. The induced rules are evaluated at the level of local aspect segments, each of which consists of related text portions with a well-defined boundary and can be easily annotated with a sentiment. Comparisons with methods that do not employ information about RST components are conducted.

The rest of this paper is organized as follows: The second section describes how RST can be potentially useful for sentiment analysis. The third section presents data preparation and rule induction methods. The fourth section describes experiments on rule-based sentiment analysis and their results. The fifth

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:  
[www.igi-global.com/chapter/rule-based-polarity-aggregation-using-rhetorical-structures-for-aspect-based-sentiment-analysis/308506](http://www.igi-global.com/chapter/rule-based-polarity-aggregation-using-rhetorical-structures-for-aspect-based-sentiment-analysis/308506)

## Related Content

---

### Data Mining in Public Administration

John Wang, Xiaohua Huand Dan Zhu (2009). *Social and Political Implications of Data Mining: Knowledge Management in E-Government* (pp. 302-314).

[www.irma-international.org/chapter/data-mining-public-administration/29078](http://www.irma-international.org/chapter/data-mining-public-administration/29078)

### Semi-Automatic Design of Spatial Data Cubes from Simulation Model Results

Hadj Mahboubi, Sandro Bimonte, Guillaume Defuant, Jean-Pierre Chanetand François Pinet (2013). *International Journal of Data Warehousing and Mining* (pp. 70-95).

[www.irma-international.org/article/semi-automatic-design-spatial-data/75616](http://www.irma-international.org/article/semi-automatic-design-spatial-data/75616)

### XML Tree Classification on Evolving Data Streams

Albert Bifetand Ricard Gavaldà (2012). *XML Data Mining: Models, Methods, and Applications* (pp. 199-218).

[www.irma-international.org/chapter/xml-tree-classification-evolving-data/60910](http://www.irma-international.org/chapter/xml-tree-classification-evolving-data/60910)

### An Experimental Approach and Monitoring Tools for Evaluating a Dynamic Cubing System

Anne Tchounikine, Maryvonne Miqueland Usman Ahmed (2016). *International Journal of Data Warehousing and Mining* (pp. 1-19).

[www.irma-international.org/article/an-experimental-approach-and-monitoring-tools-for-evaluating-a-dynamic-cubing-system/171117](http://www.irma-international.org/article/an-experimental-approach-and-monitoring-tools-for-evaluating-a-dynamic-cubing-system/171117)

### Algebraic and Graphic Languages for OLAP Manipulations

Franck Ravat, Olivier Teste, Ronan Tournierand Gilles Zurfluh (2008). *International Journal of Data Warehousing and Mining* (pp. 17-46).

[www.irma-international.org/article/algebraic-graphic-languages-olap-manipulations/1798](http://www.irma-international.org/article/algebraic-graphic-languages-olap-manipulations/1798)