

ADDING ARTIFICIAL INTELLIGENCE THROUGH OPINION MINING

¹ M A KHADEER

² K RAKESH

³ D JAMUNA

ABSTRACT

The advent of Web 2.0 and social media content has stirred much excitement and created abundant opportunities for understanding the opinions of the general public and consumers toward social events, political movements, company strategies, marketing campaigns, and product references.

Many new and exciting social, geo political, and business-related research questions can be answered by analyzing the thousands, even millions, of comments and responses expressed in various blogs (such as the blogosphere), forums (such as Yahoo Forums), social media and social network sites (including YouTube, Facebook, and Flickr), virtual worlds (such as Second Life), and tweets (Twitter). *Opinion mining*, a sub discipline within data mining and computational linguistics, refers to the computational techniques for extracting, classifying, understanding, and assessing the opinions expressed in various online news sources, social media comments, and other user-generated content. Sentiment analysis is often used in opinion mining to identify

sentiment, affect, subjectivity, and other emotional states in online text.

1. INTRODUCTION

In spite of recent advances, there are still several promising new directions for developing and advancing new opinion mining research. For example, much past and current opinion mining research has focused on English, Chinese, Arabic, and several European languages. Advanced and mature techniques have been developed especially for English.

However, in light of the large amount of public opinions expressed by citizens in different parts of the world, new, scalable opinion mining and sentiment analysis resources and techniques need to be developed for various languages.

Future work in multilingual opinion mining will require bootstrapping techniques for analyzing obscure and lesser-known languages for quick situation assessment. Frameworks and methods for integrating sentiments and opinions expressed with other computational representations—such as interesting topics or product features

extracted from user-generated text, participant reply networks, spikes and outbreaks of ideas or events—are also critically needed. Sentiment and opinion alone will not allow researchers to sufficiently understand the complex dynamics in opinions expressed by a large group of participants.

KEY WORDS: opinion, mining, data-warehouse.

2. DEFINITIONS

In general, people can express opinions on any *target entity*—products, services, individuals, organizations, or events. In this context, the term object is used to denote the target entity that has been commented on.

An object can have a set of components (or parts) and a set of attributes (or properties), which we collectively call the features of the object. For example, a particular brand of cellular phone is an object. It has a set of components (such as battery and screen) and a set of attributes (such as voice quality and size), which are all called *features* (or *aspects*).

An opinion can be expressed on any feature of the object and also on the object itself.

3. DISCUSSION

To address the broad questions posed in this research, and guided by the literature reviewed, we developed a framework for analysis with four major stages: stakeholder analysis, topical analysis, sentiment analysis, and stock modeling.

During the stakeholder analysis stage, we identified the stakeholder groups participating in Web forum discussions. In the topical analysis stage, the major topics of discussion driving communication in the Web forum are determined. The sentiment analysis stage consists of assessing the opinions expressed by the Web forum participants in their discussions. Finally, in the stock modeling stage, we examine the relationships between various attributes of Web forum discussions and the firm's stock behavior.

The stock models we developed to test the hypotheses regarding the relationship between the Web forum discussions and the Wal-Mart stock behavior directly followed those proposed in previous research. The models examined the correlation and developed both contemporaneous and predictive regression models using the variables presented in the research design. In the contemporaneous regressions, the stock behavior variables are

regressed on the measures of the forum discussions occurring on the same day. In the predictive regressions, measures of the forum discussions occurring on a specific day are utilized to predict the stock behavior on the following trading day. We summarize selected results from predictive regression using overall forum data.

Disagreement and subjectivity also held significant relationships with volatility, where less disagreement and high levels of subjectivity predicted periods of high stock volatility. When the forum consolidated or intensified its sentiment position and utilized highly subjective language, perhaps in efforts to influence other participants, increases in stock volatility were likely to follow in the subsequent trading day.

4. CONCLUSION

Finally, despite these difficulties and challenges, the field has made significant progress over the past few years. This is evident from the large number of start-up companies that provide sentiment-analysis and opinion mining services. A real, substantial need exists in industry for such services. This practical need and the technical challenges will keep the field vibrant and lively for years to come.

5. ACKNOWLEDGEMENT

The authors express their deep gratitude to the Principal and the Management members of JPNCE for their encouragement and extensive support in preparing and publishing of this paper.

6. REFERENCES

1. B. Liu, "Sentiment Analysis and subjectivity," *Handbook of Natural Language Processing*, 2nd ed., N. Indurkha and F.J. Damerau, eds., Chapman & Hall, 2010, pp. 627–666.
2. B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, vol. 2, nos. 1–2, 2008, pp. 1–135.
3. J. Wiebe et al., "Learning Subjective Language," *Computational Linguistics*, vol. 30, Sept. 2004, pp. 277–308.
4. M. Hu and B. Liu, "Mining and summarizing Customer Reviews," *Proc. ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD)*, ACM Press, 2004, pp. 168–177.
5. N. Jindal and B. Liu, "Opinion Spam and Analysis," *Proc. Conf. Web Search and Web Data Mining (WSDM)*, ACM Press, 2008, pp. 219–230.

AUTHORS



1. M.A. Khadeer, II Year-
M.Tech (CSE) Research
Scholar at CSE Dept.
Jayaprakash Narayan

College of Engineering, Mahabubnagar. His areas of Interest are in Data warehousing and mining, Networking and Web Technologies.



2. Mr. K. Rakesh, Working
as Assistant Professor in CSE
Dept. Jayaprakash Narayan
College of Engineering,

Mahabub Nagar. His areas of Interest are in Advanced compiler designing, Data Mining, Network security, guided M. Tech and B. Tech Students IEEE Projects.



3. Prof. D. Jamuna, Working
as Professor & Head of CSE
Dept. Jayaprakash Narayan
College of Engineering,

Mahabubnagar, M.Tech(SE) from School of Information Technology, JNTUH, Hyderabad. BE(CSE) from Vijayanagara Engineering College, Bellary. Experience 15 Years in Teaching Profession. Her areas of Interest are in Wireless Sensor Networks, Data Mining, Networking, guided M. Tech

and B. Tech Students IEEE Projects. She is Member of CSI.