



Sentiment Analysis: Stock Prediction

Misuk Kim

mskim@dm.snu.ac.kr

Content

I. Introduction

II. 연구 동향

III. To Do..

Reference

Introduction

Thomson Reuters Taps Into Twitter For Big Data Sentiment Analysis

Posted Feb 3, 2014 by Ingrid Lunden (@ingridlunden)

6 Like 338 Tweak



Twitter bases the major river, but a new service up its position as a big now incorporating sentiment and trading platform. To be clear it is not a deal "combining a number of

You can think of this as Bloomberg last year, in Tweets related to speci

Here, the Thomson Reuters

<http://techcrunch.com/2014/02/03/twitter-raises/>

TRENDING ON TWITTER: SOCIAL SENTIMENT ANALYTICS

February 20, 2014

If you've spent any time on Twitter, you know that tweets can be snarky, informative, nonsensical, hilarious, and sometimes provocative. This is the "sentiment" of the Twittersphere. What Twitter does for the reputation of Lady Gaga or Bill Gates also applies to companies and stocks. Knowing the fast-moving social sentiment of a particular investment opportunity could provide a trader with that extra edge. That's why Bloomberg provides subscribers with social sentiment tools, which categorize opinion and provide charts and visualizations based on this data.

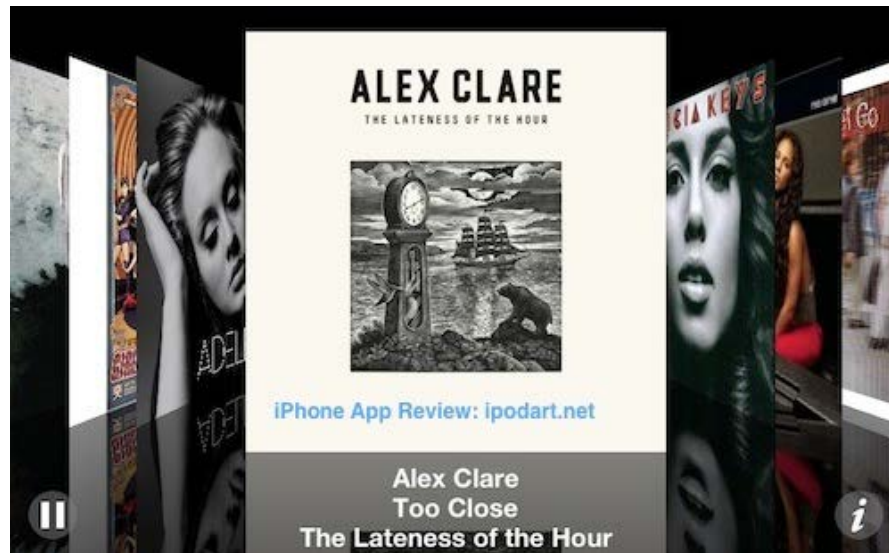


The blue line illustrates Regeneron Pharmaceuticals' stock price, while the green, amber and red bars indicate social sentiment and volume

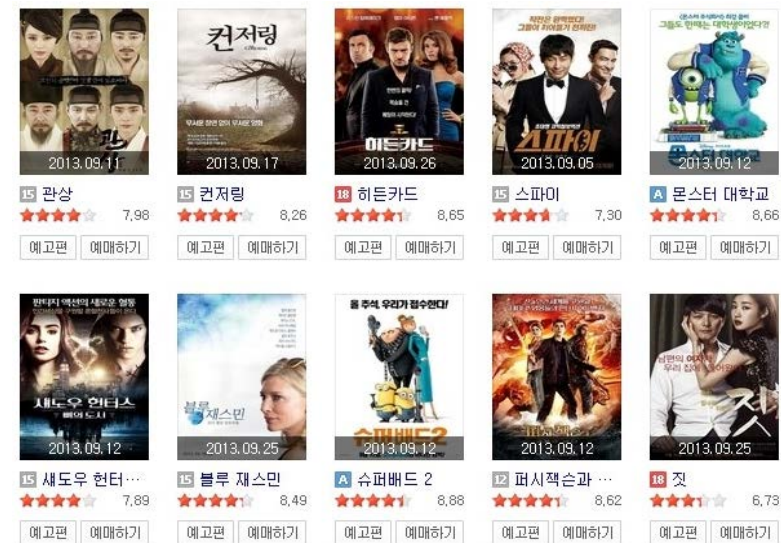
- 감성 분석은 다양한 분야에 활용되고 있음
- 가장 활발하게 활용되는 분야: FINANCE
ex) Thomson Reuters, Bloomberg
- 자연어처리 분야에서 화자의 감정이나 의견을 자동으로 인식하고 추출하려는 연구가 활발히 이루어지고 있음
- 문장 내에 주관적인 표현을 인식하고, 이를 통해 주관적인 문장이라고 판단하는 것이 중요함
- 주관적은 표현을 컴퓨터가 인식할 수 있게 주식언어로 표현해야 함
- 이슈
 - ✓ NEWS → DATA
 - ✓ NEWS 시각화
- 적용분야 예
 - ✓ 상품평 분석(영화, 전자제품 등)
 - ✓ 정치인의 담화 특성 분석
 - ✓ 주식 및 금융 시장 예측

Introduction

- 기존의 텍스트 분석은 **주제 분류**(교육, 오락, 등), **장르 분류**(소설, 실화, 등)에 중점을 둠
 - ✓ 주제 분류 및 장르 분류에는 Machine Learning 기법의 성과가 좋음
- 최근에 다양한 온라인 리뷰들의 감성분석이 활발하게 연구되고 있음
 - ✓ 감성 분석에는 Machine Learning이 좋은 성과를 나타내지 않음
- 전체적인 감성 분석이 아닌 **상세한(In-depth) 요소 분석**이 필요



[음악 앨범] 목소리, 노래, 녹음 질, 창의력 등



[영화 리뷰] 대본, 연기, 영화예술, 감독, 장면, 속도감 등

Introduction

- 관심 분야
 - News의 sentiment가 주가에 미치는 영향 파악
 - 현재 주가를 예측하는 numerical한 데이터로는 한계가 있음
 - 비정형 데이터 중 News를 이용하여 주가의 방향 예측

연구동향

Foundations and Trends in Information Retrieval
Vol. 2, No. 1-2 (2008) 1-131
© 2008 Bo Pang and Lillian Lee. This is a pre-publication version; there are formatting and potentially small wording differences from the final version.
DOI: 10.1392/ftr.v2i1.131

now
the essence of knowledge

Opinion mining and sentiment analysis

Bo Pang¹ and Lillian Lee²

¹ Yahoo! Research, 701 First Ave, Sunnyvale, CA 94089, U.S.A., bpang@yahoo-inc.com
² Computer Science Department, Cornell University, Ithaca, NY 14853, U.S.A., llee@cs.cornell.edu

Abstract

An important part of our information-gathering behavior has always been to find out what other people think. With the growing availability and popularity of opinion-rich resources such as online review sites and personal blogs, new opportunities and challenges arise as people now can, and do, actively use information technologies to seek out and understand the opinions of others. The sudden eruption of activity in the area of opinion mining and sentiment analysis, which deals with the computational treatment of opinion, sentiment, and subjectivity in text, has thus occurred at least in part as a direct response to the surge of interest in new systems that deal directly with opinions as a first-class object.

This survey covers techniques and approaches that promise to directly enable opinion-oriented information-seeking systems. Our focus is on methods that seek to address the new challenges raised by sentiment-aware applications, as compared to those that are already present in more traditional fact-based analysis. We include material on summarization of evaluative text and on broader issues regarding privacy, manipulation, and economic impact that the development of opinion-oriented information-access services gives rise to. To facilitate future work, a discussion of available resources, benchmark datasets, and evaluation campaigns is also provided.

● Opinion mining and sentiment analysis(Bo Pang and Lillian Lee)(2008)

- Computer scientists and computational linguists
- 텍스트의 감성을 분류하는 자동화 기법 개발

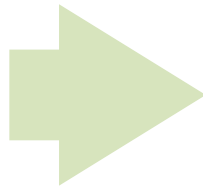
Ex) Document를 positive or negative 감성으로 분류

상품 리뷰, 영화 리뷰, 주식 및 금융 시장, 선거 결과 예측 등에 적용

그러나, 자동화된 감성 분석은 한계에 부딪힘 → 체계적이고 근거가 충분한 이론에 뒷받침된 정교한 감성 분석이 필요함

● Self Design framework 사용

- Positive/ negative
- Positive/neutral/negative
- Critical/supportive
- Emotional/critical



- ✓ 연구자의 직관적으로 이루어짐
- ✓ 연구들이 다 제각각임
- ✓ 다른 연구들과 비교하기 어려움

연구동향

- **Mining and Summarizing Customer Reviews(Minqing Hu and Bing Liu)(2004)**
 - 고객들의 리뷰를 요약
 - 전자기기의 상품 리뷰들 중 특정한 **상품 특징**에 대해서 감성 분석 수행
 - Ex) Picture quality and screen size
- **Sentiment Analysis of Movie Reviews on Discussion Boards using a Linguistic Approach(Thet *et al.*)(2010)**
 - **Clause level**로 감성분석 수행
 - 영화 리뷰에 대해 다양한 측면으로 의견을 분석함
 - Ex) cast, director, story, music
- **Appraisal of Opinion Expressions in Discourse(Asher *et al.*)(2009)**
 - Positive/negative 표현을 넘어서 **다양한 의견**을 분석
 - Ex) motivation, recommendation, speculation

연구동향

- **On the Importance of Text Analysis for Stock Price Prediction(Lee, Heeyoung, et al) (2014)**
 - Relationship between Textual Information and Stock Price
 - Contributions
 - Demonstrate that text information impact the stock price of the corresponding company for several days
 - Demonstrate that the model which includes textual information performs significantly better than a model with financial information alone
 - Release a corpus that aligns these financial events with the corresponding stock prices

To do

1. 관련 논문 조사
2. 추가 적인 데이터 확보
3. NLP 알고리즘 공부
4. Text data와 숫자 데이터의 효과적인 visualization

Reference

- [1] Wiebe, Janyce, Theresa Wilson, and Claire Cardie. Annotating Expressions of Opinions and Emotions in Language. Language Resources and Evaluation (formerly Computers and the Humanities), 39(2/3):164–210, 2005.
- [2] Wiebe, Janyce. Instructions for annotating opinions in newspaper articles. Department of Computer Science Technical Report TR-02- 101, University of Pittsburgh. 2002.
- [3] Wilson, Theresa Ann. Fine-Grained Subjectivity and Sentiment Analysis: Recognizing the Intensity, Polarity, and Attitudes of Private States. Ph.D Dissertation, University of Pittsburgh. 2008.
- [4] Shin, Hyopil, Munhyong Kim, Yu-Mi Jo, Hayeon Jang, and Andrew Cattle. Annotation Scheme for Constructing Sentiment Corpus in Korean. PACLIC 26. 181-190. 2012.
- [5] Cattle, Andrew, Munhyong Kim, and Hyopil Shin. Morphemebased Annotation Tool for Korean Text. AAC-2013. San Diego State University. 2013.
- [6] Hayeon Jang, Munhyong Kim, Hyopil Shin. KOSAC: A full-fledged Korean Sentiment Analysis Corpus, PACLIC 27. 366-373. 2013.
- [7] Hayeon Jang, Hyopil Shin. Effective Use of Linguistic Feature for Sentiment Analysis of Korean. PACLIC 24. 2010.

Thank you