

Predicting Market Movements: From Breaking News to Emerging Social Media

Hsinchun Chen

Department of Management Information Systems
University of Arizona
1130 E. Helen St.
Tucson, AZ 85721
hchen@eller.arizona.edu

Abstract. In this talk I will present several studies conducted at the AI Lab of the University of Arizona that aim to understand and predict market movements using text mining, breaking news, and social media.

In “User-Generated Content on Social Media: Predicting New Product Market Success from Online Word-of-Mouth,” we explore the predictive validity of various text and sentiment measures of online WOM for the market success of new products. The context of our study is the Hollywood movie industry where the forecast of movie sales is highly challenging and has started to incorporate online WOM. We first examine the evolution patterns of online WOM over time, followed by correlation analysis of how various sentiment measures are related to the metrics of new product success. Overall, the number of WOM messages was found to be the most useful predictor of the five new product metrics.

In “AZ SmartStock: Stock Prediction with Targeted Sentiment and Life Support,” we develop a text-based stock prediction engine with targeted sentiment and life support considerations in a real world financial setting. We focus on inter-day trading experiments, with the 5-, 10-, 20-, and 40-day trading windows. We focus on S&P 500 firms in order to minimize the potential illiquid problem associated with thinly traded stocks. News articles from major newswires were extracted from Yahoo! Finance. Life support of a company is extracted from aggregated energy (novelty) of terms used in the news articles where the company is mentioned. The combined Life-Support model was shown to out-perform other models in the 10-day trading window setting.

In “A Stakeholder Approach to Stock Prediction using Finance Social Media,” we utilize firm-related finance web forum discussions for the prediction of stock return and trading of firm stock. Considering forum participants uniformly as shareholders of the firm, suggested by prior studies, and extracting forum-level measures provided little improvement over the baseline set of fundamental and technician variables. Recognizing the true diversity among forum participants, segmenting them into stakeholder groups based upon their interactions in the forum social network and assessing them independently, refined the measures extracted from the forum and improved stock return prediction. The superior performance of the stakeholder-level model represented a statistically significant improvement over the baseline in directional accuracy, and provided an annual return of 44% in simulated trading of firm stock.