Sentiment Analysis in Finance

Karan Uppal Vika Abrecht Gary Kazantsev
Bloomberg LP

DOES NEWS IMPACT FINANCIAL MARKETS?

NEWS SENTIMENT SYSTEM OVERVIEW

1. One million news stories ingested per day
2. Company tagging via Named Entity Recognition(NER)
3. Perform sentiment analysis on each story
4. Aggregate company sentiment in a time series

FINANCIAL SENTIMENT DEFINITION

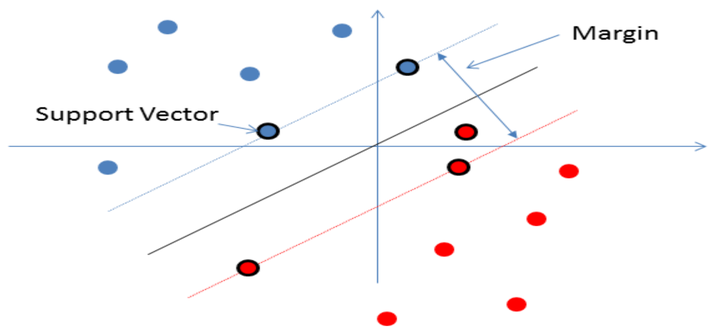
- How would a long-position investor feel about the story?
- 3 labels:

  POSITIVE - news may result in rise in stock price
  NEGATIVE - news may result in fall in stock price
  NEUTRAL – news has no effect on stock price
- Challenges:
  - How to interpret **job cuts**?
  - Are **acquisitions** positive or negative?
  - Are **higher earnings** always positive?

CLASSIFICATION FEATURES

- N-grams, some semantics
- Negations handled via state machine following McKeown & Biadsy
- Feature selection via TF/IDF (baseline), binomial (so-so), PMI (substantial gains)
- Parts of speech and lemmas aren't statistically significant
- Syntax gives small improvement, but is very expensive

CLASSIFIER



- Multi-class probabilistic non-linear SVM
- Separate models for short, long and multi-company stories
- Trained using top N by PMI features
- Thousands of training samples, imbalanced

CLASSIFICATION RESULTS

- Inter-annotator agreement 75%-88%
- Performance competitive over 3 classes
- Approaches inter-annotator agreement

SENTIMENT AGGREGATION

- For each company, aggregate individual story-level scores
- 8-hour sliding window
- Aggregates produced every 2 minutes
- Weighted confidence average in range [-1, 1]

TWITTER SENTIMENT

CHALLENGES AND DIFFERENCES

- Short text
  - ✓ 140 characters
  - ✓ Dealing with missing context
- Slang, errors, emoticons normalization
  - ✓ ROTFLMAO -> "rolling on the floor laughing…
  - ✓ Common misspellings: smaaart -> smart, goooood -> good
  - ✓ Emoticons to features
- Imbalanced dataset

FUTURE WORK

Extend to more content

– all content across all languages

Apply the same methods to other asset classes

– commodities, options, futures, fixed income, etc.

More powerful models

– Recursive Tensor Neural Networks

SCAN HERE TO DOWNLOAD THIS POSTER