All Words Are Not Made Equal

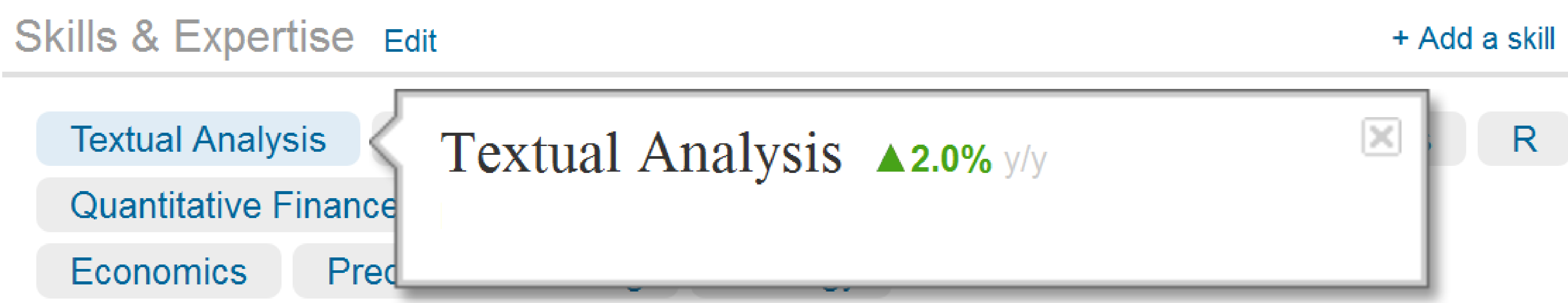
Sid Bhattacharyya
John Pendergrass
Nitish Ranjan Sinha

Presented by

Nitish Ranjan Sinha

Textual analysis is a popular skill among quants

- From Linked In



J.P. Morgan's $2 Billion Blunder

A massive trading bet boomeranged on J.P. Morgan Chase & Co., leaving the bank with at least $2 billion in trading losses and its chief executive, James Dimon, with a rare black eye following a long run as what some called the “King of Wall Street.”

The losses stemmed from wagers gone wrong in the bank's Chief Investment Office, which manages risk for the New York company. The Wall Street Journal reported early last month that large positions taken in that office by a trader nicknamed "the London whale" had roiled a sector of the debt markets.

A typical news article:
1130 words, 28 numbers!

UIC

Motivation

- Most financial information is in form of text
  - Aim to quantify textual information
  - Distinguish fact from opinion
- Better quantification of textual content lets us answer:
  - How does textual information come into prices?
    - Immediately? Grossman-Stiglitz Paradox
    - Slowly? Sinha(2011) Underreaction to News in the US Stock Market
  - What do people react to when they read news, after all?
    - All news articles report is old news
- Provide a simple yet robust way of identifying tone of news articles

UIC

Preview of Results

- Create a corpus of hand classified articles from financial press
- Almost 75% agreement among humans on “opinion” of the article
- Substantial disagreement between “Bag of Words”
  - Optimism and Positive facts: ND Dictionary is better
  - Pessimism and Negative facts: Harvard dictionary is better
- Easier to classify facts
  - Almost 80% agreement with humans using “Bag of Words”
- Harder to classify opinion
  - Optimism: 67% agreement from simple variations of Naïve Bayes classifier
  - Pessimism: 60% agreement from simple variations of Naïve Bayes classifier

Early Classifiers

- Tetlock (2007),
  - Harvard IV dictionary, from psychologists and sociologists. Developed by Philip Stone (1966)
- Loughran and McDonald (2010)
  - Chose almost 5000 most frequent words in 10-K
  - Hand tagged them as positive or negative
- Both widely used to measure tone of news articles
- Tone= (Positive Words- Negative Words)/Total Words

UIC

Tone has multiple sources

- Fact
  - “..at least $2 billion in trading losses”
- Opinion/Speculation
  - “.. with a rare black eye ”

An article can have at least two kinds of information: Present two different categorization problems

UIC

Articles have target audience

- News articles and 10K : two different audiences (and protections)
  - Newspapers often legally protected for speculation or providing opinion
  - Firms liable for opinion in 10K (some safe harbors)

Harvard Dictionary is from vernacular English
Notre Dame dictionary is from 10K

UIC

Can we quantify textual information more accurately?

- Problems with “Bag of Words” method
  - Language in Financial Newspapers is not similar to language in 10-Ks
  - “Bag of Words” methods do not distinguish facts from opinions
  - Opinions are subjective but humans reliably understand them
  - All words are not made equal (for classifying an article)
    - “bad” and “horrible” have the same weight
- Our solution:
  - Use a classifier that assigns different weights for different words
  - Distinguish fact from opinion

UIC

We use text data from two sources

- “Abreast of the market” column from WSJ, 1984-2000
- All the unique words in 10K, provided by Bill McDonald
  - Includes word, word count, # of documents in which the word occurs

UIC

Language in Financial Newspapers is not similar to language in 10-Ks

- Among top 7000 words in “Abreast of the Market” and 10Ks, 3534 not unique to the source
- 3466 top words in “Abreast of the Market” do not appear in 10Ks
  - Market Condition: “rally”, “yesterday”, “jumped”, “eased”, “slid”
  - Opinion: “think”, “rumors”, “speculation”,
  - State of mind: “worried”, ”anxiety”
- 3466 top words in 10 K do not occur in the “Abreast of the Market”
  - Accounting: “deferred”, “goodwill”, ”expenditures”, “receivable”
  - Legal: “accordance”, “reference”, “impairment”, “materially”

UIC

We propose using naïve Bayesian classification for quantifying textual information

Tag → **Train** → **Classify**

Tag: Tag news articles as positive/negative

Train: Learns probability associated with a positive document given the word/feature

Classify: Obtains probability of document being positive given the word/feature

How to tag news articles?

Two possibilities

- Market Reaction to individual articles
  - Nature of these articles poses a problem
  - We are interested in whether news affects prices, using prices to tag news is a joint test of market efficiency and classifier efficacy
  - Market may not incorporate information right away
    - Use minute by minute return, hourly return, weekly return…
- Human tagged articles

Tagging: Continued

- We obtained tags for articles in two stages
  - First Stage
    - A finance professor with management consulting and equity research experience classified each article on a scale of 1 to 5 for the following questions
      - Opinion
        - » After reading this article I feel optimistic about the future.
        - » After reading this article I feel pessimistic about the future.
      - Facts
        - » I think this article <u>presents</u> positive developments in the market.
        - » I think this article <u>presents</u> negative developments in the market.
    - The five point scale corresponds to
      - Strongly agree, Agree, Hard to say, Disagree, Strongly disagree

UIC

Tagging: Continued

- We obtained tags for articles in two stages
  - First Stage

  - Second Stage
    - Another professor read the same articles

Hand coded data: Summary Statistics

	Optimism	Pessimism	Positive	Negative
No	102	88	28	36
Yes	50	64	124	116

Relatively rare to have opinion of either kind.

Can we distinguish between opinion and fact

	Opinion		Fact	
	Optimism	Pessimism	Positive	Negative
Optimism	1			
Pessimism	-0.6	1		
Positive	0.05	-0.15	1	
Negative	-0.17	0.26	-0.19	1

- Is optimism different from pessimism?
- Is optimism different from positive?

UIC

How often do humans agree with each other?

- Another coder read all 152 articles
  - Similar level of agreement on opinion as reported in other sentiment analysis studies

Sample of 152 human tagged articles	Optimism	Pessimism	Positive	Negative
Agreement	73.3%	79.3%	57.3%	63.3%

UIC

At the training stage, the classifier learns informative features

- Naïve Bayes assigns different weights to individual words/features
- We defined features four different ways
  - Most frequent 3000 words among tagged articles
  - Words from the Harvard IV dictionary
  - Words from the Notre Dame dictionary
  - Most frequent 10,000 Character N-Grams

Example of Ngrams

- “good news”
  - N=4: “good”, “ood ”, “od n”, “d ne”, “ new”, “news”
  - N=3: “goo”, “ood”, “od ”, “d n”, “ ne”, “new”, “ews”
  - ..
  - .
- Do not need to model sequence of words
  - “od n”, “d n” etc.
- Do not need to model inflections
  - “has” and “have” – “ha”

UIC

Optimism: Results from classification

CATEGORY: OPTIMISM

Sample of 152 human tagged articles (120 for training, 32 for held out)	Harvard IV word count	Notre Dame dictionary word count	Naïve Bayes (Harvard IV)	Naïve Bayes (Notre Dame dictionary)	Naïve Bayes (Most frequent 3000 words)
Accuracy	50%	60.53%	67.48%	65.94%	67.10%

Sample of 152 human tagged articles	Optimism	Pessimism	Positive	Negative
Agreement	**73.3%**	79.3%	57.3%	63.3%

UIC

Pessimism: Results from classification

CATEGORY: PESSIMISM

Sample of 152 human tagged articles (120 for training, 32 for held out)	Harvard IV word count	Notre Dame dictionary word count	Naïve Bayes (Harvard IV)	Naïve Bayes (Notre Dame dictionary)	Naïve Bayes (Most frequent 3000 words)
Accuracy	59.87%	31.58%	56.87%	57.13%	59.45%

Sample of 152 human tagged articles	Optimism	Pessimism	Positive	Negative
Agreement	73.3%	**79.3%**	57.3%	63.3%

UIC

Positive: Results from classification

CATEGORY: POSITIVE FACTS

Sample of 152 human tagged articles (120 for training, 32 for held out)	Harvard IV word count	Notre Dame dictionary word count	Naïve Bayes (Harvard IV)	Naïve Bayes (Notre Dame dictionary)	Naïve Bayes (Most frequent 3000 words)
Accuracy	56.58%	59.21%	81.16%	81.16%	82.00%

Sample of 152 human tagged articles	Optimism	Pessimism	Positive	Negative
Agreement	73.3%	79.3%	**57.3%**	63.3%

UIC

Negative: Results from classification

CATEGORY: NEGATIVE FACTS

Sample of 152 human tagged articles (120 for training, 32 for held out)	Harvard IV word count	Notre Dame dictionary word count	Naïve Bayes (Harvard IV)	Naïve Bayes (Notre Dame dictionary)	Naïve Bayes (Most frequent 3000 words)
Accuracy	53.29%	35.53%	76.87%	75.74%	75.13%

Sample of 152 human tagged articles	Optimism	Pessimism	Positive	Negative
Agreement	73.3%	79.3%	57.3%	**63.3%**

UIC

Overall comments

- The classifier is only as good as
  - Tagged data
  - Feature set
    - Ngrams perform better than word or word N-grams
- Be sensitive to the vocabulary of your text source
  - 10K, News articles, and Twitter have different vocabularies

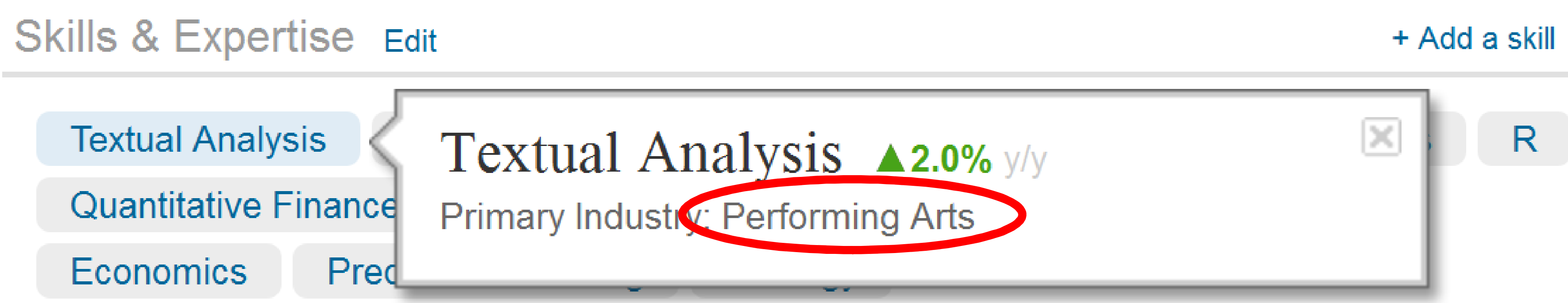
Software: State of Progress

- Word Count: Python
- Naïve Bayes Classifier: Python+NLTK
- Plan to release estimated classifier
  - Even a better plan: Tag your own articles, will get better results for your application
- R-package coming, stay tuned

UIC

Some textual analysis of textual analysis

- Linked In thinks



UIC