# *Reading the Markets: Forecasting Prediction Markets by News Content Analysis*

***(or, How to Get Rich with Computational Linguistics)***


Kevin Lerman, Ari Gilder,
Mark Dredze, Fernando Pereira
UPenn Senior Design '06-'07


CS4701 Talk

# *What are Prediction Markets?*

- Buy and sell “shares” of future events
  - Sports, finance, legal rulings, **politics**, etc...
- Issuing company backs the shares
  - If the event happens, company pays $1/share
  - Otherwise, $0
- Investor-gamblers trade amongst themselves
  - Share price fluctuates according to likelihood of event
  - “Public thinks this event is X% likely to happen”

**2008 US Election - 2008 Presidential Election Winner (Individual) *See Specific Rules***

**Hillary Clinton to win 2008 US Presidential Election**

Last Price: 47.6

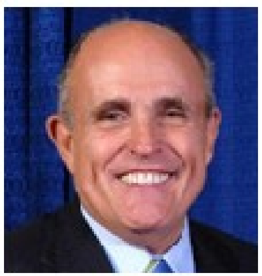
You can buy this at 47.8 — Buy

You can sell this at 47.4 — Sell

Explain | Trade Now


**2008 US Election - 2008 Presidential Election Winner (Individual) *See Specific Rules***

**Rudy Giuliani to win 2008 US Presidential Election**



Last Price: 16.8

You can buy this at 17.2 — Buy

You can sell this at 16.8 — Sell

Explain | Trade Now


| | | | | | |
|---|---|---|---|---|---|
| COURIC.DEPARTS.DEC07<br>Katie Couric departure from CBS Evening News announced on/before 31 December 2007 | **5.0** | **16.0** | 10.5 | 180 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| US.ACTION.NKOREA.DEC07<br>The United States to conduct overt military action against North Korea ON/BEFORE Dec 31st 2007 | **0.5** | **3.5** | 2.5 | 1538 | -0.5 |
| US.ACTION.NKOREA.JUN08<br>The United States to conduct overt military action against North Korea ON/BEFORE 30 June 2008 | **0.2** | **4.2** | 3.5 | 180 | 0 |
| US.ACTION.NKOREA.MAR08<br>The United States to conduct overt military action against North Korea ON/BEFORE Mar 31th 2008 | **1.1** | **4.6** | 4.0 | 1158 | 0 |

Source: Intrade (www.intrade.com)

# *The Goal*

- Value of a share ~ likelihood of underlying event
  - As perceived by the public
- How does the public figure out what's likely?
  - Following current events – reading the news
- Can we automatically recognize good/bad news?
  - Read the morning's news
  - Predict whether price will rise or fall today
    - ...well enough to trade and make a profit

# *Current Events Refresher: The 2004 US Election*

Ran for President

George Bush

John Kerry

Candidates for the DNC nomination

Hillary Clinton

Wesley Clark

Joe Lieberman

Howard Dean

John Kerry

Dick Gephardt

# *Predicting Sans News*

- Look **<u>only</u>** at the charts.
  - No news, no notion of what the security is about
- Wall Street sometimes uses this idea
  - "Technical Analysis"
- Pro: Easy to analyze
- Con: Not much info

# Features

- What was the price yesterday?
- What was the price movement yesterday?
- What was the price movement the day before?
- What was the direction of those price movements?
  - Strict binary.  Lets the model be more flexible

- How many shares were traded yesterday?
- Log2 of above
  - Again, let the model be more flexible

# Training and Testing Data

- One day = one datapoint
- Market runs for, say, 200 days
- Naïve thing to do would be cross-validation

- Can't do that here!
  - We'd be training our model on data from the future

- Every day, use all previous days as training data
  - Use the model to make exactly one prediction: today
  - Then throw it out and train a new model tomorrow
    - With today as a piece of training data

# Evaluation Method

- % days predicted correctly is a poor metric
  - Price movements are noisy labels
  - Ignores the magnitude of the movement
- Simulate investing according to the model
  - Every day, buy/short 1 share. Sell/cover it next day.
  - Normalize returns by "omniscience" figure
- Baselines for Comparison
  - 0 is reasonable (zero-sum market)
  - "Weather Forecasting"

# So, Does It Work?

- Quite well, actually

| | Weather | Tech Analysis |
|---|---|---|
| Bush (Election) | 20% | 21% |
| Kerry (Election) | 15% | 2% |
| Clark (DNC) | 13% | 20% |
| Clinton (DNC) | -8% | 38% |
| Dean (DNC) | 24% | 23% |
| Gephardt (DNC) | 1% | 8% |
| Kerry (DNC) | 6% | -6% |
| Lieberman (DNC) | 2% | 3% |
| **Average** | **9.1%** | **13.6%** |

# On to News

- We'll build the news system separately
  - No price/volume history information
- Looks like a sentiment-classification problem
  - Is this product review saying good or bad things about the product?
  - Is this news article saying good or bad things about the candidate?
- Bag-of-words techniques work well for sentiment problems, maybe we can adapt them

# Bag of Words?

- Turn each word into a feature: How many times do we see it?

  <u>“We are learning machine learning!”</u>
  we:1 are:1 learning:2 machine:1

  - Implicitly, literature:0 history:0 dinosaur:0 ...
  - Typically, exclude “stopwords” (“are”, maybe “we”)
  - Normalize counts to document length:

    we:0.25 learning:0.5 machine:0.25

# News Stories are not Product Reviews

- They don't discuss one candidate exclusively

**Howard Dean**, the former governor of Vermont, and Senator **John Kerry** of Massachusetts squabbled so intensely over their differences on the war in Iraq and on each other's credentials that the Rev. **Al Sharpton** of New York finally stepped in and urged an end to disputes that he said could hurt the Democrats in their attempt to win the White House.

(source: New York Times, 5/4/2003)

# News Stories are not Product Reviews

- They don't discuss one candidate exclusively
- So, predefine a list of entities for each security
  - eg. {Bush, Kerry, Iraq}
  - Look for sentences that mention an entity
  - Associate each token in that sentence with that entity
  - Produces features like “said:Bush”, “casualties:Iraq”

# *Price Movements are not Product Ratings*

- They reflect not public perception, but **<u>change</u>** in public perception

Nine Democratic presidential candidates battled tonight over the war in Iraq and over how to provide health care insurance for all Americans...

(source: USA Today, 5/3/2003)

The nine Democrats vying for the White House clashed over the U.S.-led war against Iraq, health insurance and President Bush's tax cut...

(source: New York Times, 5/4/2003)

The nine candidates debated for the first time on Saturday in South Carolina, an early primary state...

(source: WPTZ, 5/5/2003)

# Price Movements are not Product Ratings

- They reflect not public perception, but **change** in public perception

- So, look at changes in feature counts
  - Compare today's prominence of a feature to its prominence over the past three days
  - Learn from the change, not the raw count

# So, Does It Work?

- Not really.

| | Weather | Tech Analysis | News |
|---|---|---|---|
| Bush (Election) | 20% | 21% | 21% |
| Kerry (Election) | 15% | 2% | 12% |
| Clark (DNC) | 13% | 20% | 14% |
| Clinton (DNC) | -8% | 38% | -12% |
| Dean (DNC) | 24% | 23% | 41% |
| Gephardt (DNC) | 1% | 8% | 1% |
| Kerry (DNC) | 6% | -6% | 8% |
| Lieberman (DNC) | 2% | 3% | -13% |
| **Average** | **9.1%** | **13.6%** | **9.0%** |

# Justification for "Interesting" campaign designation

- Ask Wikipedia, knower of all things:

- Article on 2004 Presidential election discusses:
  - Bush
  - Kerry
  - Dean
  - Clark (a little, mostly to say he was too late and had no well-articulated positions)

- Other candidates only mentioned

  - Page visited 10/22/2007. http://en.wikipedia.org/wiki/United_Stated_presidential_election

# *Improving the Entities*

- We throw out sentences that contain 2+ entities
  - About 25% of the sentences that contain 1+ entities
- These sentences can be critical!

**President Bush ultimately defeated Senator Kerry in the debate.**

- But we need to understand their structure
  - "Senator Kerry ultimately defeated President Bush in the debate" ?

# Beyond Bag of Words

- Parsing: Which words are related to which others?
- Semantic Role Labelling:  How are they related?
- Danger: These both use their own machine learning models, so they introduce errors

# Bag of Branches

accused (verb)
Kerry (subject)
(good for Kerry)

plans (verb)
Bush (subject)
(good for Bush)

accused (verb)
Kerry (object)
(good for Bush)

plans (noun)
Bush (modifier)
(good for Kerry)

# So, Does It Work?

- Depends which benchmark we're trying to beat
  - Technical Analysis is surprisingly resilient!

| | Weather | Tech Analysis | BOW | Parsing |
|---|---|---|---|---|
| Bush (Election) | 20% | 21% | 21% | 39% |
| Kerry (Election) | 15% | 2% | 12% | 20% |
| Clark (DNC) | 13% | 20% | 14% | 6% |
| Clinton (DNC) | -8% | 38% | -12% | -10% |
| Dean (DNC) | 24% | 23% | 41% | 35% |
| Gephardt (DNC) | 1% | 8% | 1% | 3% |
| Kerry (DNC) | 6% | -6% | 8% | 8% |
| Lieberman (DNC) | 2% | 3% | -13% | -19% |
| **Average** | **9.1%** | **13.6%** | **9.0%** | **10.3%** |

# A closer look

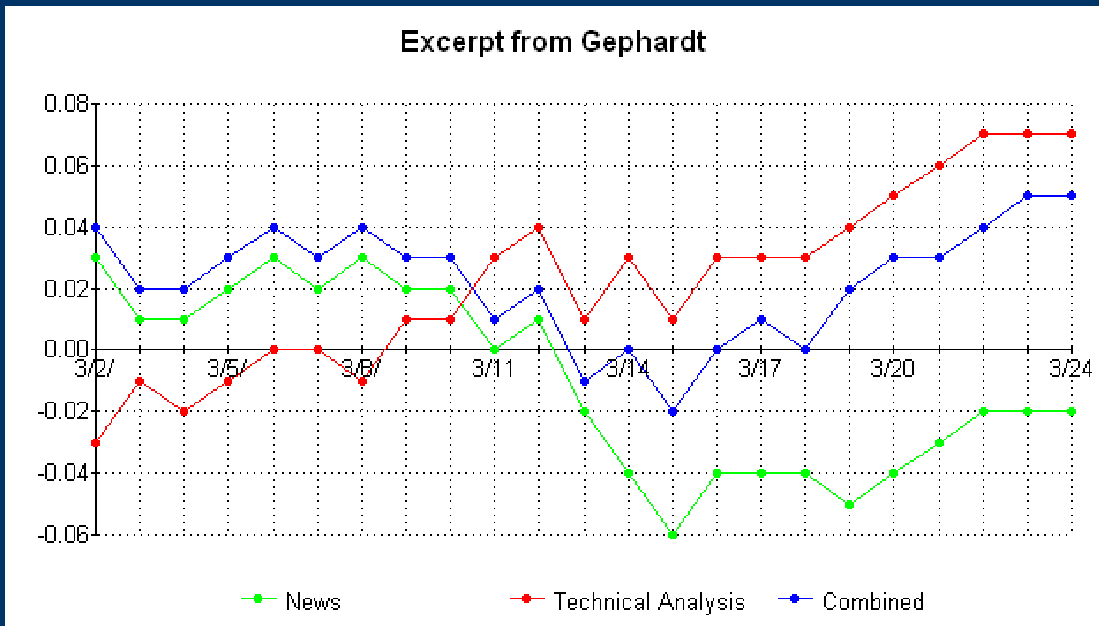
- Using news, we beat the tech analysis half the time
- The methods disagree on which markets are hard
- They capture *non-redundant* information

| | Tech Analysis | News | Max() |
|---|---|---|---|
| Bush (Election) | 21% | 39% | 39% |
| Kerry (Election) | 2% | 20% | 20% |
| Clark (DNC) | 20% | 6% | 20% |
| Clinton (DNC) | 38% | -10% | 38% |
| Dean (DNC) | 23% | 35% | 35% |
| Gephardt (DNC) | 8% | 3% | 8% |
| Kerry (DNC) | -6% | 8% | 8% |
| Lieberman (DNC) | 3% | -19% | 3% |
| **Average** | **13.6%** | **10.3%** | **21.4%** |

# *Exploiting Parallel Data Streams*

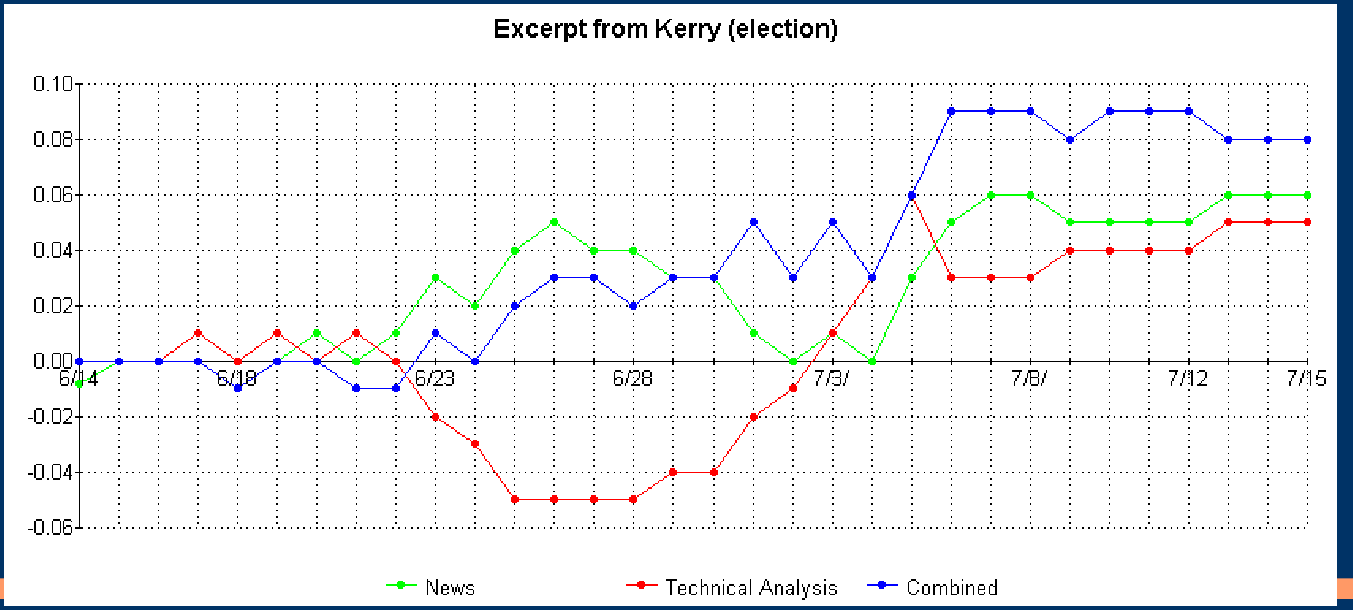
- Ideally, use news-based when news is “interesting,” else use technical analysis
- It's hard to detect when news is “interesting”
- But, easy to detect when news *was* “interesting”
  - News was interesting when news predictions were good
- So, use whichever system has been doing better lately

# System Switching at Work




(graphs of cumulative money made/lost over time by each system, as % of omniscience)

Combined stays with the better performing system




It can even beat both systems by mixing them correctly

# So, Does It Work?
## *(Last one, I promise)*

- Yes!

| | Weather | Tech Analysis | Parsing | Combined |
|---|---|---|---|---|
| Bush (Election) | 20% | 21% | 39% | 29% |
| Kerry (Election) | 15% | 2% | 20% | 13% |
| Clark (DNC) | 13% | 20% | 6% | 15% |
| Clinton (DNC) | -8% | 38% | -10% | 23% |
| Dean (DNC) | 24% | 23% | 35% | 30% |
| Gephardt (DNC) | 1% | 8% | 3% | 8% |
| Kerry (DNC) | 6% | -6% | 8% | 21% |
| Lieberman (DNC) | 2% | 3% | -19% | 2% |
| **Average** | **9.1%** | **13.6%** | **10.3%** | **17.6%** |

# Conclusions

- Prediction markets are inefficient
  - At least the Iowa Electronic Markets
- But they are in fact correlated to developing news
  - ...well enough to build a predictive model
  - ...if the news is interesting enough
- Errors introduced by statistical parsing/role labelling are more than offset by the higher quality features you can extract from the output