4. Supervised Techniques I (flipped)

DS-GA 1015, Text as Data
Arthur Spirling

March 2, 2021

Housekeeping

Housekeeping

HW 1 out: coming in on March 9, 2021, at 11pm (NY time).

Housekeeping

HW 1 out: coming in on March 9, 2021, at 11pm (NY time).

TaD/NLP: Nanyun (Violet) Peng (UCLA), "Controllable Text Generation Beyond Auto-regressive Models."

Dictionaries

More Specifically

More Specifically

We have a set of key words, with attendant scores,

More Specifically

We have a set of key words, with attendant scores,

e.g. for movie reviews: '`terrible`' is scored as -1; '`fantastic`' as $+1$

More Specifically

We have a set of key words, with attendant scores,

e.g. for movie reviews: '`terrible`' is scored as -1; '`fantastic`' as $+1$

$\rightarrow$ the relative rate of occurrence of these terms tells us about the overall tone or category that the document should be placed in.

More Specifically

We have a set of key words, with attendant scores,

e.g. for movie reviews: '`terrible`' is scored as -1; '`fantastic`' as $+1$

$\rightarrow$ the relative rate of occurrence of these terms tells us about the overall tone or category that the document should be placed in.

i.e. for document i and words $m = 1, \ldots, M$ in the dictionary,

More Specifically

We have a set of key words, with attendant scores,

e.g. for movie reviews: '`terrible`' is scored as -1; '`fantastic`' as $+1$

$\rightarrow$ the relative rate of occurrence of these terms tells us about the overall tone or category that the document should be placed in.

i.e. for document i and words $m = 1, \ldots, M$ in the dictionary,

$$\boxed{\text{tone of document } i = \sum_{m=1}^{M} \frac{s_m w_{im}}{N_i}}$$

More Specifically

We have a set of key words, with attendant scores,

e.g. for movie reviews: '`terrible`' is scored as -1; '`fantastic`' as $+1$

$\rightarrow$ the relative rate of occurrence of these terms tells us about the overall tone or category that the document should be placed in.

i.e. for document i and words $m = 1, \ldots, M$ in the dictionary,

$$\boxed{\text{tone of document } i = \sum_{m=1}^{M} \frac{s_m w_{im}}{N_i}}$$

where s_m is the score of word m

More Specifically

We have a set of key words, with attendant scores,

e.g. for movie reviews: ‘`terrible`’ is scored as -1; ‘`fantastic`’ as $+1$

$\rightarrow$ the relative rate of occurrence of these terms tells us about the overall tone or category that the document should be placed in.

i.e. for document i and words $m = 1, \ldots, M$ in the dictionary,

$$\boxed{\text{tone of document } i = \sum_{m=1}^{M} \frac{s_m w_{im}}{N_i}}$$

where s_m is the score of word m

and w_{im} is the number of occurrences of the mth dictionary word in the document i

More Specifically

We have a set of key words, with attendant scores,

e.g. for movie reviews: ‘`terrible`’ is scored as -1; ‘`fantastic`’ as $+1$

$\rightarrow$ the relative rate of occurrence of these terms tells us about the overall tone or category that the document should be placed in.

i.e. for document i and words $m = 1, \ldots, M$ in the dictionary,

$$\boxed{\text{tone of document } i = \sum_{m=1}^{M} \frac{s_m w_{im}}{N_i}}$$

where s_m is the score of word m

and w_{im} is the number of occurrences of the mth dictionary word in the document i

and N_i is the total number of all dictionary words in the document.

More Specifically

We have a set of key words, with attendant scores,

e.g. for movie reviews: '`terrible`' is scored as -1; '`fantastic`' as $+1$

$\rightarrow$ the relative rate of occurrence of these terms tells us about the overall tone or category that the document should be placed in.

i.e. for document i and words $m = 1, \ldots, M$ in the dictionary,

$$\boxed{\text{tone of document } i = \sum_{m=1}^{M} \frac{s_m w_{im}}{N_i}}$$

where s_m is the score of word m

and w_{im} is the number of occurrences of the mth dictionary word in the document i

and N_i is the total number of all dictionary words in the document.

$\rightarrow$ just add up the number of times the words appear and multiply by the score (normalizing by doc dictionary presence)

(Simple) Example: Barnes' review of *The Big Short*

(Simple) Example: Barnes' review of *The Big Short*

Director and co-screenwriter Adam McKay (Step Brothers) bungles a great opportunity to savage the architects of the 2008 financial crisis in The Big Short, wasting an A-list ensemble cast in the process. Steve Carell, Brad Pitt, Christian Bale and Ryan Gosling play various tenuously related members of the finance industry, men who made made a killing by betting against the housing market, which at that point had superficially swelled to record highs. All of the elements are in place for a lacerating satire, but almost every aesthetic choice in the film is bad, from the U-Turn-era Oliver Stone visuals to Carell's sketch-comedy performance to the cheeky cutaways where Selena Gomez and Anthony Bourdain explain complex financial concepts. After a brutal opening half, it finally settles into a groove, and there's a queasy charge in watching a credit-drunk America walking towards that cliff's edge, but not enough to save the film.

Retain words in Hu & Liu Dictionary...

Director and co-screenwriter Adam McKay (Step Brothers) bungles a great opportunity to savage the architects of the 2008 financial crisis in The Big Short, wasting an A-list ensemble cast in the process. Steve Carell, Brad Pitt, Christian Bale and Ryan Gosling play various tenuously related members of the finance industry, men who made made a killing by betting against the housing market, which at that point had superficially swelled to record highs. All of the elements are in place for a lacerating satire, but almost every aesthetic choice in the film is bad, from the U-Turn-era Oliver Stone visuals to Carell's sketch-comedy performance to the cheeky cutaways where Selena Gomez and Anthony Bourdain explain complex financial concepts. After a brutal opening half, it finally settles into a groove, and there's a queasy charge in watching a credit-drunk America walking towards that cliff's edge, but not enough to save the film.

Retain words in Hu & Liu Dictionary...

great *savage*
crisis *wasting*

tenuously
killing
superficially *swelled*

bad

complex *brutal*

drunk
enough

Simple math. . .

Simple math. . .

negative 11

Simple math...

negative	11
positive	2

Simple math...

negative	11
positive	2
total	13

Simple math...

negative 11

positive 2

total 13

tone $= \frac{2-11}{13} = \frac{-9}{13}$

Simple math...

negative 11

positive 2

total 13

tone $= \frac{2-11}{13} = \frac{-9}{13}$

Harrison's review of *The Addams Family*

The best gags in this new version update classic jokes: the furious villagers are now controlled via social media fake news, while Wednesday appears identically wan in every one of her friend's Instagram filters. The family predilection for using dangerous weapons survives intact, even if the darkly Gothic spirit of the Barry Sonnenfeld films is largely missing.

The standard of the animation, however, is less than lush, and doesn't sit well in comparison with most recent studio releases, such as the similar Hotel Transylvania franchise. But, with a decent voice cast and a style that pays homage to the original drawings, The Addams Family passes muster as a disconcertingly cheerful, family-friendly romp.

Harrison's review of *The Addams Family*

The best gags in this new version update classic jokes: the furious villagers are now controlled via social media fake news, while Wednesday appears identically wan in every one of her friend's Instagram filters. The family predilection for using dangerous weapons survives intact, even if the darkly Gothic spirit of the Barry Sonnenfeld films is largely missing.

The standard of the animation, however, is less than lush, and doesn't sit well in comparison with most recent studio releases, such as the similar Hotel Transylvania franchise. But, with a decent voice cast and a style that pays homage to the original drawings, The Addams Family passes muster as a disconcertingly cheerful, family-friendly romp.

What rating do you think this received?

Harrison's review of *The Addams Family*

The best gags in this new version update classic jokes: the furious villagers are now controlled via social media fake news, while Wednesday appears identically wan in every one of her friend's Instagram filters. The family predilection for using dangerous weapons survives intact, even if the darkly Gothic spirit of the Barry Sonnenfeld films is largely missing.

The standard of the animation, however, is less than lush, and doesn't sit well in comparison with most recent studio releases, such as the similar Hotel Transylvania franchise. But, with a decent voice cast and a style that pays homage to the original drawings, The Addams Family passes muster as a disconcertingly cheerful, family-friendly romp.

What rating do you think this received?

3 stars (0.27)

Harrison's review of *6 Underground*

Harrison's review of *6 Underground*

The writers of Deadpool, Paul Wernick and Rhett Reese, were employed here, and the aim is clearly to replicate that brand of post-modern snark, but the jokes don't stretch further than having One's crew tediously quote lines from other films while pulling off a heist. The explosions and locations are flashy but, without any opportunities for engagement with the narrative, the two-hour-plus runtime is truly punishing.

If Netflix have funded their fair share of lame comedies and stodgy action, this at least provides fans of the low-brow with a two-for-one. Otherwise, it's a laboured Mission Impossible knock-off that offers little more than watching money unenjoyably wasted. In the era of fake news, 6 Underground feels like a fake film: a disposable, unmemorable product that even those seeking Bay's usual high-octane idiocy will find exhausting.

Harrison's review of *6 Underground*

The writers of Deadpool, Paul Wernick and Rhett Reese, were employed here, and the aim is clearly to replicate that brand of post-modern snark, but the jokes don't stretch further than having One's crew tediously quote lines from other films while pulling off a heist. The explosions and locations are flashy but, without any opportunities for engagement with the narrative, the two-hour-plus runtime is truly punishing.

If Netflix have funded their fair share of lame comedies and stodgy action, this at least provides fans of the low-brow with a two-for-one. Otherwise, it's a laboured Mission Impossible knock-off that offers little more than watching money unenjoyably wasted. In the era of fake news, 6 Underground feels like a fake film: a disposable, unmemorable product that even those seeking Bay's usual high-octane idiocy will find exhausting.

What rating do you think this received?

Harrison's review of *6 Underground*

The writers of Deadpool, Paul Wernick and Rhett Reese, were employed here, and the aim is clearly to replicate that brand of post-modern snark, but the jokes don't stretch further than having One's crew tediously quote lines from other films while pulling off a heist. The explosions and locations are flashy but, without any opportunities for engagement with the narrative, the two-hour-plus runtime is truly punishing.

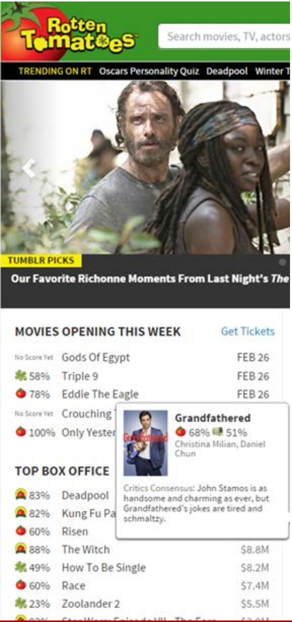
If Netflix have funded their fair share of lame comedies and stodgy action, this at least provides fans of the low-brow with a two-for-one. Otherwise, it's a laboured Mission Impossible knock-off that offers little more than watching money unenjoyably wasted. In the era of fake news, 6 Underground feels like a fake film: a disposable, unmemorable product that even those seeking Bay's usual high-octane idiocy will find exhausting.

What rating do you think this received?

1 star (-0.167)

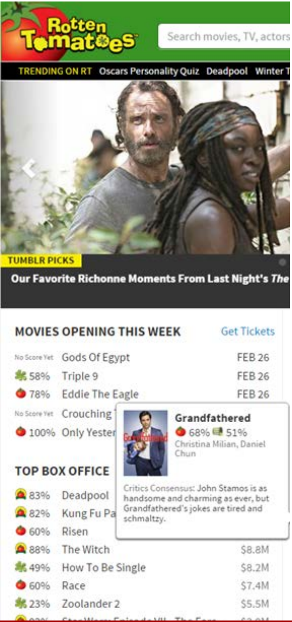
Exercise

Exercise



You are working for `rottentomatoes.com`, and want to automatically code (written) movie reviews as being between 1 and 5 stars.

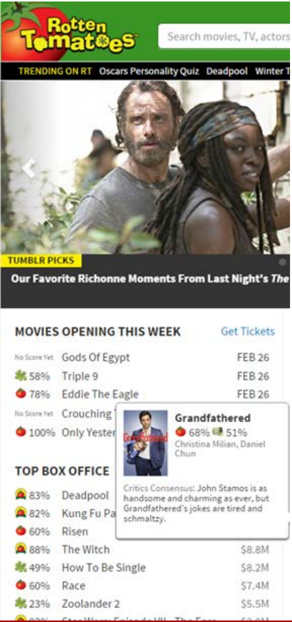
Exercise



You are working for `rottentomatoes.com`, and want to automatically code (written) movie reviews as being between 1 and 5 stars.

1. Would the Hu & Liu approach work better for distinguishing a 1 star review from a 5 star review, or a 4 from a 5 star review? Why? How could you improve upon this?

Exercise

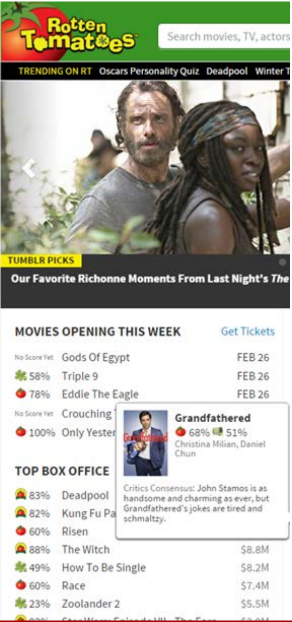


You are working for `rottentomatoes.com`, and want to automatically code (written) movie reviews as being between 1 and 5 stars.

1. Would the Hu & Liu approach work better for distinguishing a 1 star review from a 5 star review, or a 4 from a 5 star review? Why? How could you improve upon this?

2. Why does sarcasm cause problems, and what should we do about it?

Exercise



You are working for `rottentomatoes.com`, and want to automatically code (written) movie reviews as being between 1 and 5 stars.

1 Would the Hu & Liu approach work better for distinguishing a 1 star review from a 5 star review, or a 4 from a 5 star review? Why? How could you improve upon this?

2 Why does sarcasm cause problems, and what should we do about it?

3 Why might be generally nervous about BOW approaches?

Sarcasm Detection

Sarcasm Detection

Notoriously difficult problem (see Pang & Lee, 2008)

Sarcasm Detection

Notoriously difficult problem (see Pang & Lee, 2008)

e.g. "Great for insomniacs!" (book review)

Sarcasm Detection

Notoriously difficult problem (see Pang & Lee, 2008)

e.g. “Great for insomniacs!” (book review)

Early approaches simply used presence of tags e.g. `[laughter]` in proceedings to make inferences about what was sarcastic.

Sarcasm Detection

Notoriously difficult problem (see Pang & Lee, 2008)

e.g. "Great for insomniacs!" (book review)

Early approaches simply used presence of tags e.g. `[laughter]` in proceedings to make inferences about what was sarcastic. Or prosody in spoken word (e.g. lower pitch).

Sarcasm Detection

Notoriously difficult problem (see Pang & Lee, 2008)

e.g. "Great for insomniacs!" (book review)

Early approaches simply used presence of tags e.g. `[laughter]` in proceedings to make inferences about what was sarcastic. Or prosody in spoken word (e.g. lower pitch).

Davidov et al ('Semi-Supervised Recognition of Sarcastic Sentences in Twitter and Amazon') get good results by looking for particular pattern:

Sarcasm Detection

Notoriously difficult problem (see Pang & Lee, 2008)

e.g. "Great for insomniacs!" (book review)

Early approaches simply used presence of tags e.g. `[laughter]` in proceedings to make inferences about what was sarcastic. Or prosody in spoken word (e.g. lower pitch).

Davidov et al ('Semi-Supervised Recognition of Sarcastic Sentences in Twitter and Amazon') get good results by looking for particular pattern: [high frequency word] [content word] [high frequency word] in reviews

Sarcasm Detection

Notoriously difficult problem (see Pang & Lee, 2008)

e.g. "Great for insomniacs!" (book review)

Early approaches simply used presence of tags e.g. `[laughter]` in proceedings to make inferences about what was sarcastic. Or prosody in spoken word (e.g. lower pitch).

Davidov et al ('Semi-Supervised Recognition of Sarcastic Sentences in Twitter and Amazon') get good results by looking for particular pattern: [high frequency word] [content word] [high frequency word] in reviews

btw punctuation adds relatively little to accuracy.

Using LIWC

Using LIWC

RESEARCH ARTICLE

Examining long-term trends in politics and culture through language of political leaders and cultural institutions

Kayla N. Jordan, Joanna Sterling, James W. Pennebaker, and Ryan L. Boyd

PNAS February 26, 2019 116 (9) 3476-3481; first published February 11, 2019; https://doi.org/10.1073/pnas.1811987116

Edited by Steven Pinker, Harvard University, Cambridge, MA, and approved December 28, 2018 (received for review July 11, 2018)

Using LIWC

RESEARCH ARTICLE

Examining long-term trends in politics and culture through language of political leaders and cultural institutions

Kayla N. Jordan, Joanna Sterling, James W. Pennebaker, and Ryan L. Boyd

PNAS February 26, 2019 116 (9) 3476-3481; first published February 11, 2019; https://doi.org/10.1073/pnas.1811987116

Edited by Steven Pinker, Harvard University, Cambridge, MA, and approved December 28, 2018 (received for review July 11, 2018)

Authors use LIWC to study 'analytic' thinking and 'clout' (confidence) of Trump, US Presidents and other country leaders.

Using LIWC

RESEARCH ARTICLE

Examining long-term trends in politics and culture through language of political leaders and cultural institutions

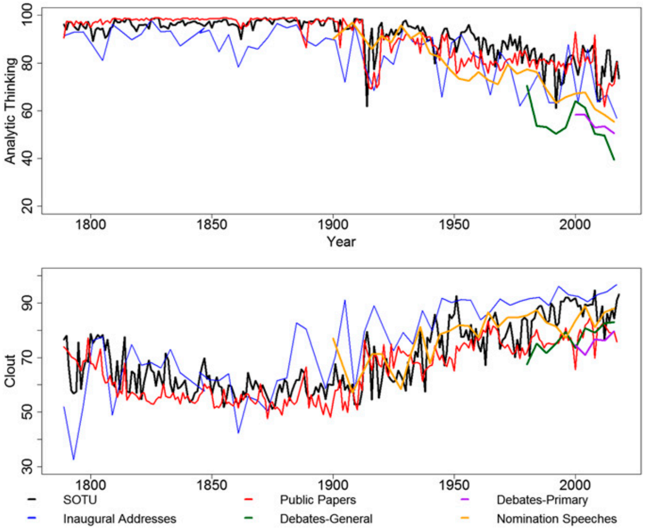
Kayla N. Jordan, Joanna Sterling, James W. Pennebaker, and Ryan L. Boyd

PNAS February 26, 2019 116 (9) 3476-3481; first published February 11, 2019; https://doi.org/10.1073/pnas.1811987116

Edited by Steven Pinker, Harvard University, Cambridge, MA, and approved December 28, 2018 (received for review July 11, 2018)

Authors use LIWC to study 'analytic' thinking and 'clout' (confidence) of Trump, US Presidents and other country leaders.

$\rightarrow$ find analytic thinking has decreased; confidence has increased


Analytic Thinking
20
40
60
80
100
1800
1850
1900
1950
2000
Year
Clout
30
50
70
90
1800
1850
1900
1950
2000
SOTU
Inaugural Addresses
Public Papers
Debates-General
Debates-Primary
Nomination Speeches

Under The Hood

Under The Hood

"Analytical thinking– a high number reflects formal, logical, and hierarchical thinking; lower numbers reflect more informal, personal, here-and-now, and narrative thinking."

Under The Hood

"Analytical thinking– a high number reflects formal, logical, and hierarchical thinking; lower numbers reflect more informal, personal, here-and-now, and narrative thinking."

hmm "Due to prior commercial agreements, the precise algorithms are not available."

Under The Hood

"Analytical thinking– a high number reflects formal, logical, and hierarchical thinking; lower numbers reflect more informal, personal, here-and-now, and narrative thinking."

hmm "Due to prior commercial agreements, the precise algorithms are not available."

How do you think this variable is formed?

Being Careful...

Being Careful…

In principle, it is straightforward to extend dictionary from one domain to another

Being Careful...

In principle, it is straightforward to extend dictionary from one domain to another

$\rightarrow$ matter of adding extra words in the various categories.

Being Careful...

In principle, it is straightforward to extend dictionary from one domain to another

$\rightarrow$ matter of adding extra words in the various categories.

But much care is needed when a dictionary designed for one context is applied to another.

Being Careful...

In principle, it is straightforward to extend dictionary from one domain to another

→ matter of adding extra words in the various categories.

But much care is needed when a dictionary designed for one context is applied to another.

e.g. Loughran & MacDonald, 2011: common dictionaries fail badly when applied to financial texts

Being Careful...

In principle, it is straightforward to extend dictionary from one domain to another

→ matter of adding extra words in the various categories.

But much care is needed when a dictionary designed for one context is applied to another.

e.g. Loughran & MacDonald, 2011: common dictionaries fail badly when applied to financial texts—e.g. `cost` is a neutral term in reports, but negative in Harvard IV

Being Careful...

In principle, it is straightforward to extend dictionary from one domain to another

→ matter of adding extra words in the various categories.

But much care is needed when a dictionary designed for one context is applied to another.

e.g. Loughran & MacDonald, 2011: common dictionaries fail badly when applied to financial texts—e.g. `cost` is a neutral term in reports, but negative in Harvard IV

Q in your field, is there an example of a word(s) that is used in a fundamentally different way, in terms of sentiment, relative to more common uses?

Followup: how bad could it be?

Followup: how bad could it be?

Loughran and McDonald (2011) show that *Harvard General Inquirer* is “inappropriate” for business applications.

Followup: how bad could it be?

Loughran and McDonald (2011) show that *Harvard General Inquirer* is "inappropriate" for business applications.

L&M (2015) "The Use of Word Lists in Textual Analysis" compares (proprietary) `Diction` software word lists for tone with their own lists, which are trained on finance documents.

Followup: how bad could it be?

Loughran and McDonald (2011) show that *Harvard General Inquirer* is "inappropriate" for business applications.

L&M (2015) "The Use of Word Lists in Textual Analysis" compares (proprietary) `Diction` software word lists for tone with their own lists, which are trained on finance documents.

83% of freq counts of Diction 'optimistic' words don't appear on L&M list. For 'pessimistic' words, 70% of Diction word frequencies don't appear on L&M.

Followup: how bad could it be?

Loughran and McDonald (2011) show that *Harvard General Inquirer* is "inappropriate" for business applications.

L&M (2015) "The Use of Word Lists in Textual Analysis" compares (proprietary) `Diction` software word lists for tone with their own lists, which are trained on finance documents.

83% of freq counts of Diction 'optimistic' words don't appear on L&M list. For 'pessimistic' words, 70% of Diction word frequencies don't appear on L&M. Also show that L&M word lists (from company filings) are statistically significant predictor of volatility and direction makes sense (not so for Diction).

Events

Events, dear boy...

Events, dear boy…

Scholars of International Relations need access to events

Events, dear boy. . .

Scholars of International Relations need access to events

Real time media reports are obvious source. . .

Events, dear boy...

Scholars of International Relations need access to events

Real time media reports are obvious source...


ASIA PACIFIC

Leaders of South Korea and Japan Meet in Effort to Mend Ties

点击查看本文中文版 | Read in Chinese

By CHOE SANG-HUN NOV. 1, 2015

Events, dear boy...

Scholars of International Relations need access to events

Real time media reports are obvious source...

Yet need to be coded automatically to be helpful.

Events, dear boy...

Scholars of International Relations need access to events

Real time media reports are obvious source...

ASIA PACIFIC

Leaders of South Korea and Japan Meet in Effort to Mend Ties

点击查看本文中文版 | Read in Chinese

By CHOE SANG-HUN NOV. 1, 2015

Yet need to be coded automatically to be helpful.

Exercise

Exercise

Consider compiling an events data set (by hand or by machine) from press reports.

Exercise

Consider compiling an events data set (by hand or by machine) from press reports.

1 Would you prefer to have the universe of press reports from the United States, or the United Kingdom or North Korea? Why, and what does that tell you about journalistic agendas?

Exercise

Consider compiling an events data set (by hand or by machine) from press reports.

1. Would you prefer to have the universe of press reports from the United States, or the United Kingdom or North Korea? Why, and what does that tell you about journalistic agendas?
2. Sports reports cause problems for automatic event extraction. Why?

Exercise

Consider compiling an events data set (by hand or by machine) from press reports.

1. Would you prefer to have the universe of press reports from the United States, or the United Kingdom or North Korea? Why, and what does that tell you about journalistic agendas?
2. Sports reports cause problems for automatic event extraction. Why?

England bounce back to defeat Italy in the Six Nations

England managed to secure victory against Italy despite going behind early on. Photo by David Rogers/Getty Images

Why are these stories hard to parse?

Why are these stories hard to parse?


HEALTH • COVID-19

Exclusive: New York Tells Hospitals In Leaked Email That Staff 'Must' Come Before the Elderly In COVID-19 Vaccination Drive

Why are these stories hard to parse?

HEALTH • COVID-19

Exclusive: New York Tells Hospitals In Leaked Email That Staff 'Must' Come Before the Elderly In COVID-19 Vaccination Drive

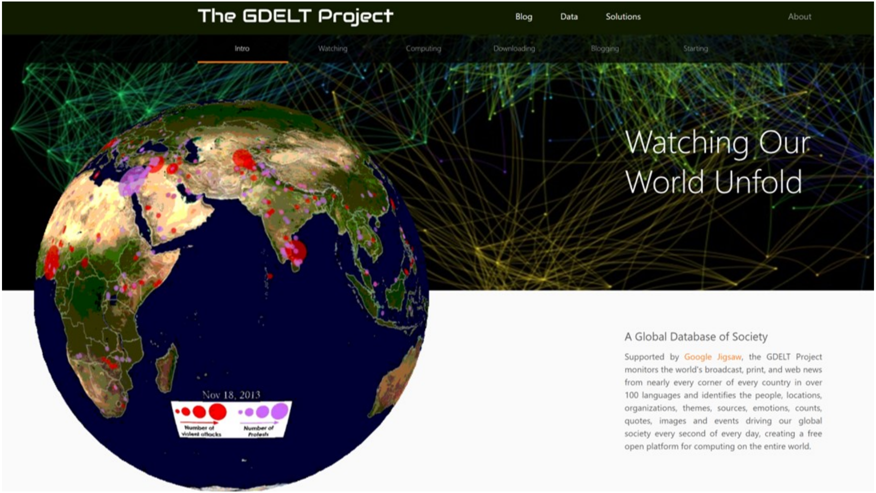
Prime Minister warns tier restrictions could come back in England

The road map will start as a national approach

SHARE

By Neil Shaw
16:21, 22 FEB 2021

NEWS


The GDELT Project
Blog
Data
Solutions
About
Intro
Watching
Computing
Downloading
Blogging
Starting
Watching Our World Unfold
Nov 18, 2013
Number of violent attacks
Number of Protests
A Global Database of Society
Supported by Google Jigsaw, the GDELT Project monitors the world's broadcast, print, and web news from nearly every corner of every country in over 100 languages and identifies the people, locations, organizations, themes, sources, emotions, counts, quotes, images and events driving our global society every second of every day, creating a free open platform for computing on the entire world.

FOMC

FOMC

FOMC

Federal Reserve Board oversees federal reserve system,

FOMC

Federal Reserve Board oversees federal reserve system, and sets monetary policy via Federal Open Market Committee decisions on interest rates.

FOMC

Federal Reserve Board oversees federal reserve system, and sets monetary policy via Federal Open Market Committee decisions on interest rates.

Meet at least four times a year,

FOMC

Federal Reserve Board oversees federal reserve system, and sets monetary policy via Federal Open Market Committee decisions on interest rates.

Meet at least four times a year, and release statement after meeting (followed by minutes, also release 'Beige book' of economic indicators 8 times a year)

FOMC

Federal Reserve Board oversees federal reserve system, and sets monetary policy via Federal Open Market Committee decisions on interest rates.

Meet at least four times a year, and release statement after meeting (followed by minutes, also release 'Beige book' of economic indicators 8 times a year)

FOMC perception of economic situation has extremely influential effects on financial markets,

FOMC

Federal Reserve Board oversees federal reserve system, and sets monetary policy via Federal Open Market Committee decisions on interest rates.

Meet at least four times a year, and release statement after meeting (followed by minutes, also release 'Beige book' of economic indicators 8 times a year)

FOMC perception of economic situation has extremely influential effects on financial markets, so are made with great care via 'Fedspeak'—

Federal Reserve Board oversees federal reserve system, and sets monetary policy via Federal Open Market Committee decisions on interest rates.

Meet at least four times a year, and release statement after meeting (followed by minutes, also release 'Beige book' of economic indicators 8 times a year)

FOMC perception of economic situation has extremely influential effects on financial markets, so are made with great care via 'Fedspeak'—which is ambiguous in tone.

Example from Sept 2015

Example from Sept 2015

Information received since the Federal Open Market Committee met in July suggests that economic activity is expanding at a moderate pace. Household spending and business fixed investment have been increasing moderately, and the housing sector has improved further; however, net exports have been soft.

Example from Sept 2015

Information received since the Federal Open Market Committee met in July suggests that economic activity is expanding at a moderate pace. Household spending and business fixed investment have been increasing moderately, and the housing sector has improved further; however, net exports have been soft.

The Committee continues to see the risks to the outlook for economic activity and the labor market as nearly balanced but is monitoring developments abroad. Inflation is anticipated to remain near its recent low level in the near term but the Committee expects inflation to rise gradually toward 2 percent over the medium term as the labor market improves further and the transitory effects of declines in energy and import prices dissipate.

Problem and Approach

Problem and Approach

Can we predict interest rate decisions $\{-1, 0, +1\}$ at next meeting from prior FOMC statements, minutes, books since last meeting?

Problem and Approach

Can we predict interest rate decisions $\{-1, 0, +1\}$ at next meeting from prior FOMC statements, minutes, books since last meeting? (range is 1994–2013)

Problem and Approach

Can we predict interest rate decisions $\{-1, 0, +1\}$ at next meeting from prior FOMC statements, minutes, books since last meeting? (range is 1994–2013)

$\rightarrow$ How?

Problem and Approach

Can we predict interest rate decisions $\{-1, 0, +1\}$ at next meeting from prior FOMC statements, minutes, books since last meeting? (range is 1994–2013)

$\rightarrow$ How?

Use 'hawkish' and 'dovish' dictionary,

Problem and Approach

Can we predict interest rate decisions $\{-1, 0, +1\}$ at next meeting from prior FOMC statements, minutes, books since last meeting? (range is 1994–2013)

$\rightarrow$ How?

Use 'hawkish' and 'dovish' dictionary, with key measure being ratio of those terms.

Problem and Approach

Can we predict interest rate decisions $\{-1, 0, +1\}$ at next meeting from prior FOMC statements, minutes, books since last meeting? (range is 1994–2013)

$\rightarrow$ How?

Use 'hawkish' and 'dovish' dictionary, with key measure being ratio of those terms.

$\rightarrow$ nouns (`strength` vs `recession`) and adjectives (`healthy`, vs `weak`)

Problem and Approach

Can we predict interest rate decisions $\{-1, 0, +1\}$ at next meeting from prior FOMC statements, minutes, books since last meeting? (range is 1994–2013)

→ How?

Use 'hawkish' and 'dovish' dictionary, with key measure being ratio of those terms.

→ nouns (`strength` vs `recession`) and adjectives (`healthy`, vs `weak`)

NB statements generally most marginally informative (as expected, since they come first),

Problem and Approach

Can we predict interest rate decisions $\{-1, 0, +1\}$ at next meeting from prior FOMC statements, minutes, books since last meeting? (range is 1994–2013)

→ How?

Use 'hawkish' and 'dovish' dictionary, with key measure being ratio of those terms.

→ nouns ('`strength`' vs '`recession`') and adjectives ('`healthy`', vs '`weak`')

NB statements generally most marginally informative (as expected, since they come first), with pseudo-R^2 (from ordered logit/probit?) ~ 0.15

Problem and Approach

Can we predict interest rate decisions $\{-1, 0, +1\}$ at next meeting from prior FOMC statements, minutes, books since last meeting? (range is 1994–2013)

→ How?

Use 'hawkish' and 'dovish' dictionary, with key measure being ratio of those terms.

→ nouns ('`strength`' vs '`recession`') and adjectives ('`healthy`', vs '`weak`')

NB statements generally most marginally informative (as expected, since they come first), with pseudo-R^2 (from ordered logit/probit?) ~ 0.15

→ rising to ~ 0.25 when all sources included (NB: speeches generally uninformative)

Questions

Questions

Is this a 'good' system (in terms of prediction)?

Questions

Is this a ‘good’ system (in terms of prediction)? What should we compare it to?

Questions

Is this a ‘good’ system (in terms of prediction)? What should we compare it to?

What *exactly* is being modeled here?

Questions

Is this a ‘good’ system (in terms of prediction)? What should we compare it to?

What *exactly* is being modeled here? That is, suppose you were building a predictor of interest rate changes from scratch, what would you include?

Using Twitter to Predict Financial Markets

scientific reports

Explore Content ⌄ Journal Information ⌄ Publish With Us ⌄

nature › scientific reports › articles › article

Open Access | Published: 27 February 2014

When Can Social Media Lead Financial Markets?

Ilya Zheludev, Robert Smith & Tomaso Aste

Scientific Reports **4**, Article number: 4213 (2014) | Cite this article

3641 Accesses | **44** Citations | **31** Altmetric | Metrics

Using Twitter to Predict Financial Markets

scientific reports

Explore Content ⌄ Journal Information ⌄ Publish With Us ⌄

nature › scientific reports › articles › article

Open Access | Published: 27 February 2014

When Can Social Media Lead Financial Markets?

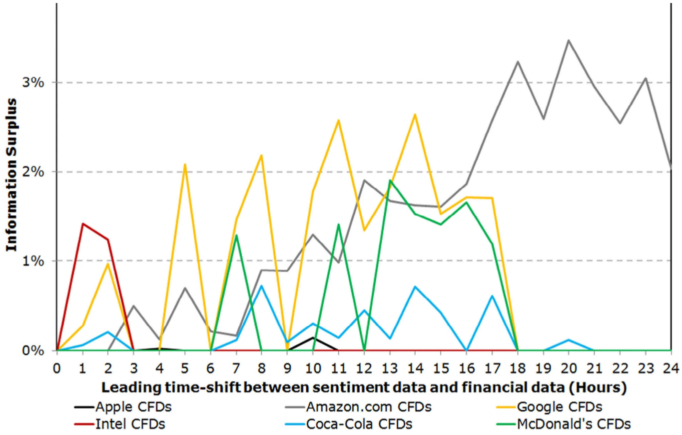
Ilya Zheludev, Robert Smith & Tomaso Aste

Scientific Reports **4**, Article number: 4213 (2014) | Cite this article

3641 Accesses | **44** Citations | **31** Altmetric | Metrics

Authors want to predict (future) stock price movements from sentiment of tweets about those stocks (e.g. "`$TSLA` looks great!")

Find: some information in sentiments for 12 of 28 assets they investigate.



Information Surplus
3%
2%
1%
0%
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Leading time-shift between sentiment data and financial data (Hours)
Apple CFDs
Amazon.com CFDs
Google CFDs
Intel CFDs
Coca-Cola CFDs
McDonald's CFDs

Questions

Questions

Q Authors contrast sentiment approaches to previous efforts that look at (mere) volume —

Questions

Q Authors contrast sentiment approaches to previous efforts that look at (mere) volume —why? That is, why would you expect volume of tweets to be informative?

Questions

- Q Authors contrast sentiment approaches to previous efforts that look at (mere) volume —why? That is, why would you expect volume of tweets to be informative?

- Q Also contrast approach to more NLP-heavy attempts to automatically 'read'/understand tweets on direction of stock price.

Questions

Q Authors contrast sentiment approaches to previous efforts that look at (mere) volume —why? That is, why would you expect volume of tweets to be informative?

Q Also contrast approach to more NLP-heavy attempts to automatically ‘read’/understand tweets on direction of stock price. What knowledge model are such approaches using?

Questions

- Q Authors contrast sentiment approaches to previous efforts that look at (mere) volume —why? That is, why would you expect volume of tweets to be informative?

- Q Also contrast approach to more NLP-heavy attempts to automatically 'read'/understand tweets on direction of stock price. What knowledge model are such approaches using?

- Q Tweet sentiment shows heavy autocorrelation.

Questions

Q Authors contrast sentiment approaches to previous efforts that look at (mere) volume —why? That is, why would you expect volume of tweets to be informative?

Q Also contrast approach to more NLP-heavy attempts to automatically 'read'/understand tweets on direction of stock price. What knowledge model are such approaches using?

Q Tweet sentiment shows heavy autocorrelation. Why? What does this mean?

Questions

- Q Authors contrast sentiment approaches to previous efforts that look at (mere) volume —why? That is, why would you expect volume of tweets to be informative?

- Q Also contrast approach to more NLP-heavy attempts to automatically 'read'/understand tweets on direction of stock price. What knowledge model are such approaches using?

- Q Tweet sentiment shows heavy autocorrelation. Why? What does this mean?

- Q Authors find that sentiment is informative about price time series.

Questions

Q Authors contrast sentiment approaches to previous efforts that look at (mere) volume —why? That is, why would you expect volume of tweets to be informative?

Q Also contrast approach to more NLP-heavy attempts to automatically 'read'/understand tweets on direction of stock price. What knowledge model are such approaches using?

Q Tweet sentiment shows heavy autocorrelation. Why? What does this mean?

Q Authors find that sentiment is informative about price time series. Does that mean it causes it?