Textual sentiment and sector-specific reaction

Wolfgang Karl Härdle
Cathy Yi-Hsuan Chen
Elisabeth Bommes

Ladislaus von Bortkiewicz Chair of Statistics
Humboldt-Universität zu Berlin
http://lvb.wiwi.hu-berlin.de

News moves Markets

- Zhang et al. (2016): textual sentiment provides incremental information about future stock reactions
- Sectors react differently to sentiment
- Unsupervised vs. supervised approach in sentiment projection

But there is a lot of news...

Dimensions of News

- Source of news
  - Official channel: government, federal reserve bank/central bank, financial institutions
  - Internet: blog, social media, message board
- Content of news: signal vs. noise
  - Signal: nuance of context
  - Noise: increasing imprecision of deep parsing
- Arrangement of information
  - Bag of words
  - Sentence based

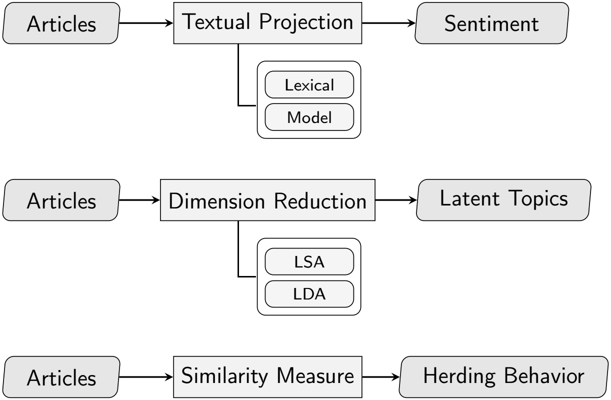
Dimensions of News ctd

- Type of news
  - Scheduled vs. non-scheduled
  - Expected vs. unexpected
  - Specific-event vs. continuous news flows

Challenge

- News are sector-specific
- How to distill sentiment across various sectors

The Power of Words: Textual Analytics



Sentiment Lexica

- *Opinion Lexicon* (BL)
  Hu and Liu (2004)
- *Financial Sentiment Dictionary* (LM)
  Loughran and McDonald (2011)
- *Multi-Perspective Question Answering Subjectivity Lexicon* (MPQA)
  Wilson et al. (2005)

Lexicon Correlation

Unsupervised Projection

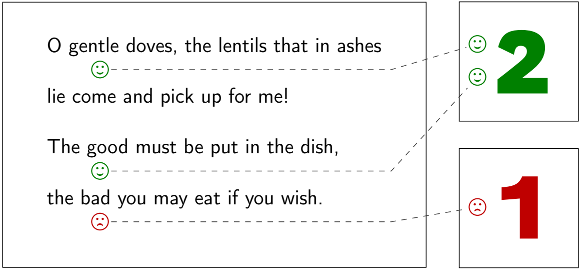


Figure: Example of Text Numerisization

- Many texts are numerisized via lexical projection
- Goal: Accurate values for positive and negative sentiment

Examples

Supervised Projection

- Training data: Financial Phrase Bank by Malo et al. (2014)
  - Sentence-level annotation of financial news
  - Manual annotation of 5,000 sentences by 16 annotators

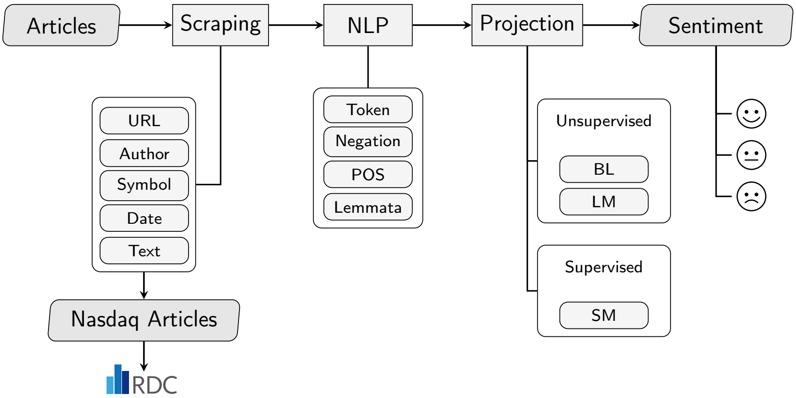
Research Questions

- Is the sentiment effect sector specific?
- Is supervised learning an effective approach in text classification?
- How well can one predict volatility or return?

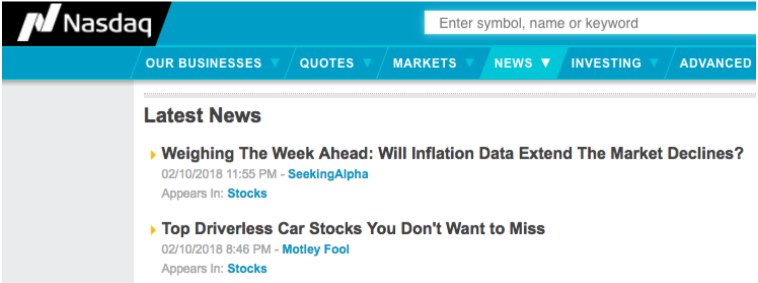
Outline

1. Motivation ✓
2. Data Collection
3. Sentiment Projection
4. Panel Regression
5. ARIMA-GARCH
6. Outlook

How to gather Sentiment Variables?



Nasdaq Articles



- Terms of Service permit web scraping
- Data available at RDC
- Oct 2009 - Dec 2016: 580k articles
- S&P 500 companies: 240k articles

Article Timeline

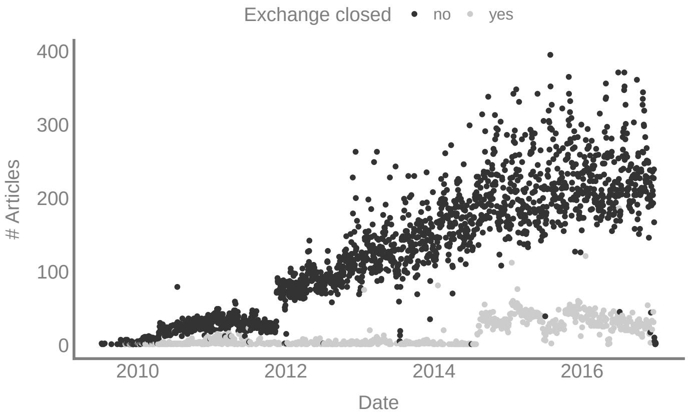


Figure: Number of Sector-specific Articles per Day

Attention Ratio

By Zhang et al (2016)

$$AR_i = T^{-1} \sum_{t=1}^{T} \mathrm{I}\left(c_{i,t} > 0\right) \qquad (1)$$

with $c_{i,t}$ as number of published articles for company i on day t.

Quantile	0%	20%	40%	60%	80%	100%
Attention Ratio	0.01	0.18	0.22	0.30	0.44	0.99

Table: Quantiles of Attention Ratio for all Nasdaq Companies

- Media coverage differs between companies
- Higher signal to noise ratio: select 100 companies More

Sector-specific articles

Sector	Abbr.	# Articles	# Comp.
Consumer Discretionary	CD	30,360	19
Consumer Staples	CS	12,210	10
Energy	EN	10,410	8
Financials	FI	34,570	13
Health Care	HC	16,950	13
Industrials	IN	16,440	13
Information Technology	IT	44,120	18
Materials	MA	3,820	3
~~Telecommunication Services~~	TE	5,880	2
~~Utilities~~	UT	780	1

Table: Number of Articles per Sector, Removal of TE and UT

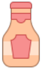
Lexical Sentiment

Project a sentence onto its polarity

$$S \in \{\textit{pos}\text{itive}, \textit{neut}\text{ral}, \textit{neg}\text{ative}\} = \{1, 0, -1\} \tag{2}$$

$$\begin{aligned} S &= \operatorname{sgn}\Big(\underbrace{\text{positive words}}_{w_{pos} \;-\; v_{pos} \;+\; v_{neg}} - \underbrace{\text{negative words}}_{w_{neg} \;-\; v_{neg} \;+\; v_{pos}} \Big) \\ &= \operatorname{sgn}\{ w_{pos} - w_{neg} - 2\,(v_{pos} - v_{neg}) \} \end{aligned} \tag{3}$$

by counting polarity words as w and negated polarity words as v.



Regularized Linear Models (RLM)

- Training data $(X_1, y_1) \dots (X_n, y_n)$ with $X_i \in \mathbb{R}^p$ and $y_i \in \{-1, 1\}$
- Linear scoring function $s(X) = \beta^\top X$ with $\beta \in \mathbb{R}^p$

Example

Regularized training error:

$$n^{-1} \sum_{i=1}^{n} \underbrace{L\{y_i, s(X)\}}_{\text{Loss Function}} + \lambda \underbrace{R(\beta)}_{\text{Regularization Term}} \tag{4}$$

with hyperparameter $\lambda \geq 0$.

RLM Estimation

- Optimize via Stochastic Gradient Descent More
- 5-fold cross validation More
- Oversampling More
- Choice of: $L(\cdot), R(\cdot), \lambda$, X (n-gram range, features) ...
- Three categories: one vs. all sub-models

Model Accuracy - Polarity

Supervised Learning

- Chosen model: Hinge loss, L1 norm, $\lambda = 0.0001$, ...
- Mean accuracy (oversampling): 0.80
- Mean accuracy (normal sample): 0.82

Lexicon-based

- Mean accuracy BL: 0.58
- Mean accuracy LM: 0.64

Evaluation BL

True \ Pred	-1	0	1	Total
-1	**214**	268	32	514
0	203	**1,786**	546	2,535
1	89	627	**452**	1,168
Total	506	2,681	1,030	4,217

Table: Confusion Matrix - BL Lexicon TXTfpblexical

Evaluation LM

True \ Pred	-1	0	1	Total
-1	**213**	289	12	514
0	200	**2,187**	148	2,535
1	111	772	**285**	1,168
Total	524	3,248	445	4,217

Table: Confusion Matrix - LM Lexicon TXTfpblexical

Evaluation SM

True \ Pred	-1	0	1	Total
-1	**389**	67	58	514
0	96	**2,134**	305	2,535
1	105	198	**916**	1,168
Total	539	2,399	1,279	4,217

Table: Confusion Matrix - Supervised Learning, estimated with Oversampling and evaluated on total Sample TXTfpbsupervised

Confusion Matrix with Oversampling | Choice of λ | Results Logistic Loss

Fractions

- Aggregation of sentence-level sentiment

$$\begin{aligned} PF &= n^{-1} \sum_{j=1}^{n} \mathbf{I}\left(Pol_j = 1\right) \\ NF &= n^{-1} \sum_{j=1}^{n} \mathbf{I}\left(Pol_j = -1\right) \end{aligned} \tag{5}$$

by Zhang et al (2016) with $j = 1, \ldots, n$ sentences in document.

- $PF_{i,t}$ and $NF_{i,t}$ account for fractions of company i on day t

Bullishness

$$B = \log\{(1 + PF)/(1 + NF)\} \quad (6)$$

by Antweiler and Frank (2004).

- $B_{i,t}$ accounts for bullishness of company i on day t
- Consider $|B_{i,t}|$ and $BN_{i,t} = \mathbf{I}\left(B_{i,t} < 0\right)B_{i,t}$

Sectors as Panels

Contemporaneous ($j = 0$) and lagged ($j = 1$) fixed effect panel regression

$$\log \sigma_{i,t} = \alpha + \beta_1 |B_{i,t-j}| + \beta_2 BN_{i,t-j} + \beta_3^\top X_{i,t-j} + \gamma_i + \varepsilon_{i,t} \quad (7)$$

$$R_{i,t} = \alpha + \beta_1 B_{i,t-j} + \beta_2^\top X_{i,t-j} + \gamma_i + \varepsilon_{i,t} \quad (8)$$

for stock i on day t with separate estimation of (7) and (8).

$X_{i,t}$ - control variables More Information

γ_i - company specific fixed effect satisfying $\sum_i \gamma_i = 0$

Stock Reaction Indicators

Range-based measure of volatility by Garman and Klass (1980)

- Notation: $\sigma_{i,t}$ Computation
- Based on open-high-low-close prices
- Equivalent results to realized volatility

Returns

$$R_{i,t} = \log(P^C_{i,t}) - \log(P^C_{i,t-1}) \tag{9}$$

with $P^C_{i,t}$ as closing price of stock i on day t

Contemporaneous - Volatility - Fractions

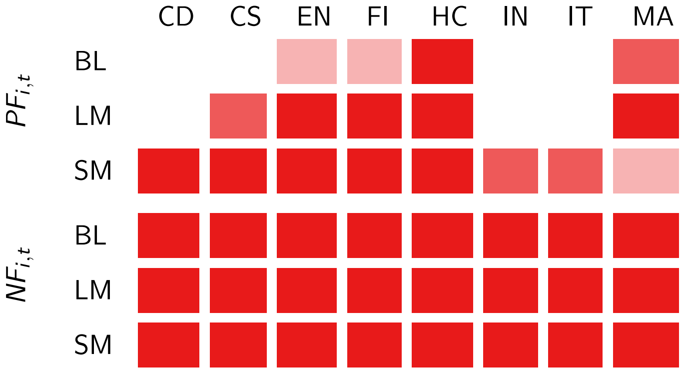


Table: Significance codes 0.01 0.05 0.1

Abbreviations

Contemporaneous - Volatility - Bullishness

		CD	CS	EN	FI	HC	IN	IT	MA
$\lvert B_{i,t} \rvert$	BL					0.01			0.01
	LM			0.05		0.01		0.01	
	SM	0.01	0.01	0.01	0.01	0.01	0.1		0.01
$BN_{i,t}$	BL	0.01	0.1	0.01			0.01	0.1	
	LM	0.01		0.01	0.05		0.01	0.01	0.05
	SM	0.05	0.01	0.01		0.01	0.05	0.01	0.05

Table: Significance codes 0.01 0.05 0.1

Abbreviations

Contemporaneous - Returns - Fractions

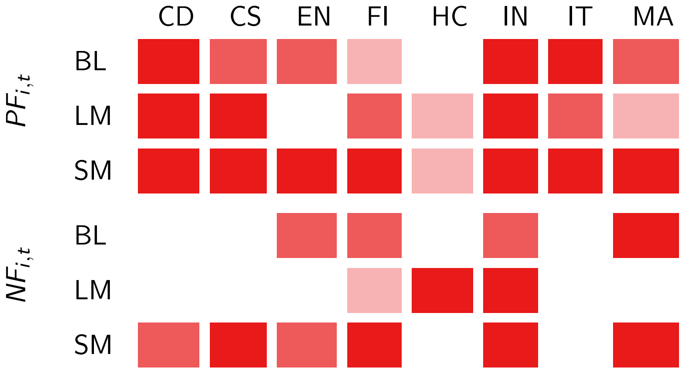


Table: Significance codes 0.01 0.05 0.1

Abbreviations

Contemporaneous - Returns - Bullishness

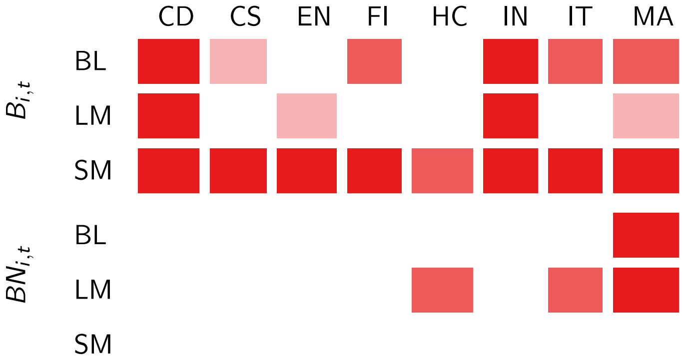


Table: Significance codes 0.01 0.05 0.1

Abbreviations

Lagged - Volatility - Fractions

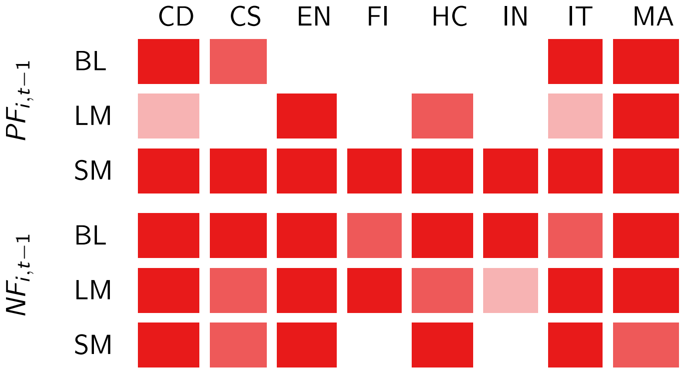


Table: Significance codes ■ 0.01 ■ 0.05 ■ 0.1

Abbreviations

Lagged - Volatility - Bullishness

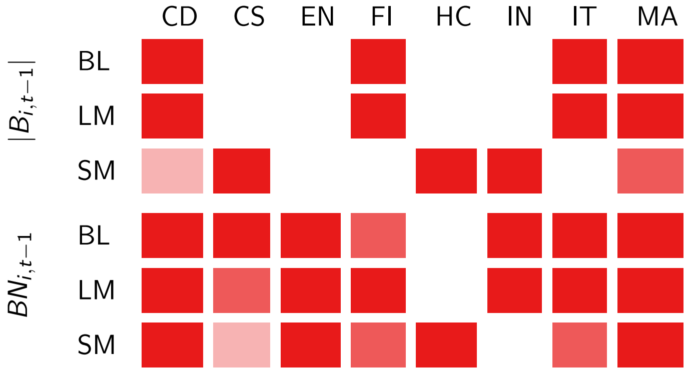


Table: Significance codes ■ 0.01 ■ 0.05 ■ 0.1

Abbreviations

S&P 500 Sector Indices

AR(1)-GARCH(1, 1) model with control variables

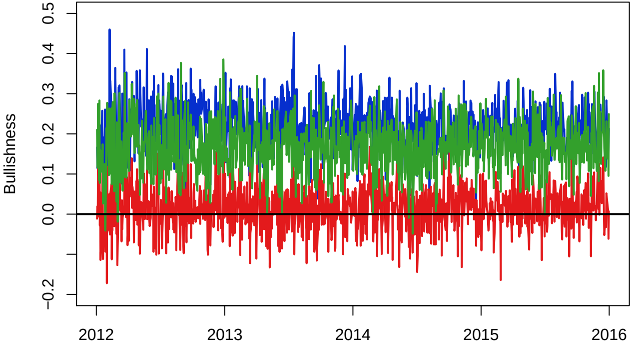
$$R_{i,t} = c_i + \varphi R_{i,t-1} + \varepsilon_{i,t} \quad (10)$$

$$\sigma^2_{i,t} = \omega_i + \alpha_i \varepsilon^2_{i,t-1} + \beta_i \sigma^2_{i,t-1} + \theta_i PF_{i,t-1} + \gamma_i NF_{i,t-1} \quad (11)$$

for sector index i on day t.

$PF_{i,t}$ - Fraction of positive words
$NF_{i,t}$ - Fraction of negative words

Why not Bullishness?



- Financial sector, BL (green), LM (red), SM (blue)
- Aggregated news for markets are very bullish
- Potential news bias?

Regression Results

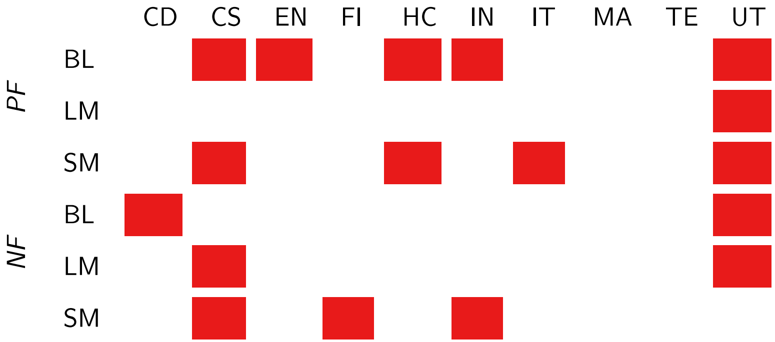


Table: Significance codes 0.01 0.05 0.1

Financials Lags

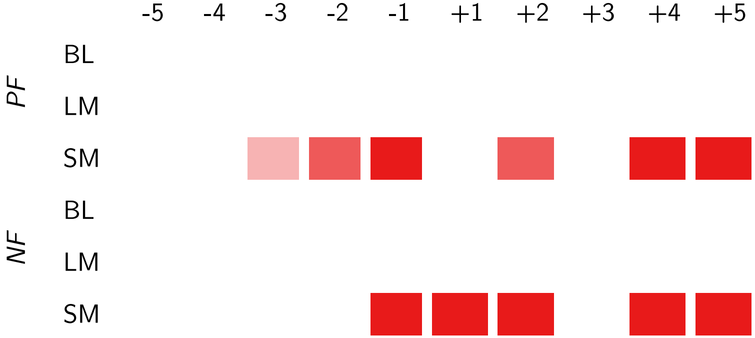


Table: Significance codes 0.01 0.05 0.1

What's next?

- Closer look at sectors : sectoral attributes, concentration, competition...
- Textual sentiment spillover : network modelling

Textual sentiment and sector-specific reaction

Wolfgang Karl Härdle
Cathy Yi-Hsuan Chen
Elisabeth Bommes

Ladislaus von Bortkiewicz Chair of Statistics
Humboldt-Universität zu Berlin
http://lvb.wiwi.hu-berlin.de

Bibliography

Antweiler, W. and Frank, M. Z.
Is All That Talk Just Noise?
J. Fin., 2004

Garman, M. and Klass, M.
On the Estimation of Security Price Volatilities from Historical Data
J. Bus., 1980

Härdle, W. K. and Lee, Y. J. and Schäfer D. and Yeh Y. R.
Variable Selection and Oversampling in the Use of Smooth Support Vector Machines for Predicting the Default Risk of Companies
J. Forecast., 2009

Hu, M. and Liu, B.
Mining and Summarizing Customer Reviews
10th ACM SIGKDD, 2004

Loughran, T. and McDonald, B.
When is a liability not a liability?
J. Financ., 2011

Malo, Pekka and Sinha, Ankur and Korhonen, Pekka and Wallenius, Jyrki and Takala, Pyry
Good debt or bad debt
Journal of the Association for Information Science and Technology, 2014

Wilson, T. and Wiebe, J. and Hoffmann, P.
Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis
HLT-EMNLP, 2005

Zhang, J., Chen C. Y., Härdle, W. K. and Bommes, E.
Distillation of News into Analysis of Stock Reactions
JBES, 2016

Zhang, X., Yichao, W., Wang, L. and Runze, L.
A Consistent Information Criterion for Support Vector Machines in Diverging Model Spaces
J. Mach. Learn. Res., 2016

Appendix

Tagging Example - BL

... McDonald's has an obesity **problem** that continues to get **worse**. And that's nothing to do with the food itself, but rather the huge menus that can now double as medieval fortification. For perspective, the chain's menu has grown 70% since 2007. And while more offerings might seem **like** a **good** thing, large menus result in **slower** service and more flare-ups between franchisees and the corporation.
Bloated menus raise inventory costs for smaller franchisees and **lead** to lower profit margins. The McDonald's corporate franchise fee is based upon sales instead of profits, making it a smaller **concern** for the company overall. ...

3 **positive words** and 5 **negative words**

TXTMcDbm
Article source

Tagging Example - LM

... McDonald's has an obesity **problem** that continues to get **worse**. And that's nothing to do with the food itself, but rather the huge menus that can now double as medieval fortification. For perspective, the chain's menu has grown 70% since 2007. And while more offerings might seem like a **good** thing, large menus result in **slower** service and more flare-ups between franchisees and the corporation.
Bloated menus raise inventory costs for smaller franchisees and lead to lower profit margins. The McDonald's corporate franchise fee is based upon sales instead of profits, making it a smaller **concern** for the company overall. ...

1 **positive word** and 4 **negative words**

TXTMcDlm

Back

Correlation - Positive Sentiment

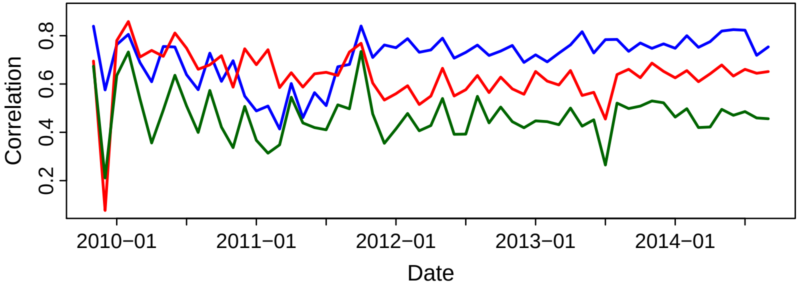


Figure: Monthly correlation between positive sentiment: BL and LM , BL and MPQA, LM and MPQA. Source: Zhang et al. (2016)

Correlation - Negative Sentiment

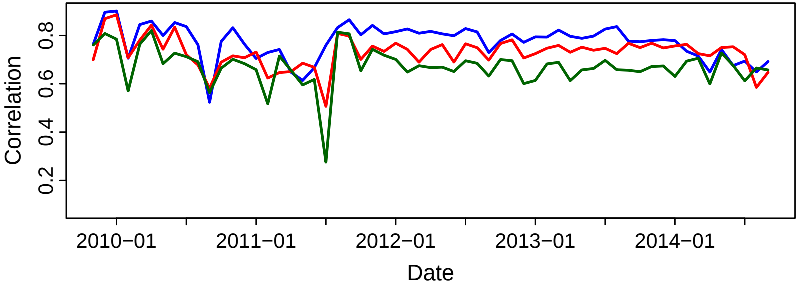


Figure: Monthly correlation between negative sentiment: BL and LM, BL and MPQA, LM and MPQA. Source: Zhang et al. (2016) Back

Natural Language Processing (NLP)

- Text is unstructured data with implicit structure
  - Text, sentences, words, characters
  - Nouns, verbs, adjectives, ..
  - Grammar
- Transform implicit text structure into explicit structure
- Reduce text variation for further analysis
- Python Natural Language Toolkit (NLTK)
- TXTnlp

Tokenization

- String

  ''McDonald's has its work cut out for it. Not only are sales falling in the U.S., but the company is now experiencing problems abroad.''

- Sentences

  ''McDonald's has its work cut out for it.'',
  ''Not only are sales falling in the U.S., but the company is now experiencing problems abroad.''

- Words

  ''McDonald'', '''s'', ''has'', ''its'', ''work'', ''cut'', ''out'' ...

Negation Handling

- "not good" $\neq$ "good"
- Reverse polarity of word if negation word is nearby
- Negation words
  `"n’t", "not", "never", "no", "neither", "nor", "none"`

Part of Speech Tagging (POS)

- Grammatical tagging of words
  - `dogs - noun, plural (NNS)`
  - `saw - verb, past tense (VBD) or noun, singular (NN)`
- Penn Treebank POS tags
- Stochastic model or rule-based

Lemmatization

- Determine canonical form of word
  - `dogs - dog`
  - `saw (verb) - see and saw (noun) - saw`
- Reduces dimension of text
- Takes POS into account
  - Porter stemmer: `saw (verb and noun) - saw`

Back

Attention Ratio II

Sector	Attention Ratio				
	Min	Q1	Q2	Q3	Max
Consumer Discretionary	0.448	0.523	0.630	0.737	0.929
Consumer Staples	0.443	0.500	0.521	0.622	0.871
Energy	0.448	0.512	0.534	0.697	0.854
Financials	0.464	0.616	0.686	0.891	0.979
Health Care	0.443	0.512	0.583	0.636	0.841
Industrials	0.458	0.522	0.577	0.661	0.857
Information Technology	0.444	0.528	0.655	0.848	0.991
Materials	0.533	0.585	0.637	0.640	0.643

Table: Attention Ratio of 100 Companies by Sector. Q1, Q2 and Q3 represent 25%, 50% and 75% quantile, respectively.

Back

Loss Functions for Classification

- Logistic: Logit

$$L\{y, s(X)\} = \log(2)^{-1} \log[1 + \exp\{-s(X)y\}] \qquad (12)$$

- Hinge: Support Vector Machines

$$L\{y, s(X)\} = \max\{0,\, 1 - s(X)y\} \qquad (13)$$

Back

Regularization Term

- L2 norm

$$R(\beta) = 2^{-1} \sum_{i=1}^{p} \beta_i^2 \tag{14}$$

- L1 norm

$$R(\beta) = \sum_{i=1}^{p} |\beta_i| \tag{15}$$

Back

RLM Example

Sentence 1: "The profit of Apple increased."
Sentence 2: "The profit of the company decreased."

$$y = (1, -1) \quad (16)$$

$$X = \begin{array}{c} \\ \textit{the} \\ \textit{profit} \\ \textit{of} \\ \textit{Apple} \\ \textit{increased} \\ \textit{company} \\ \textit{decreased} \end{array} \begin{array}{c} \begin{array}{cc} X_1 & X_2 \end{array} \\ \begin{pmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \end{array} \quad (17)$$

Back

k-fold Cross Validation (CV)

- Partition data into k complementary subsets
- No loss of information as in conventional validation
- Stratified CV: equally distributed response variable in each fold

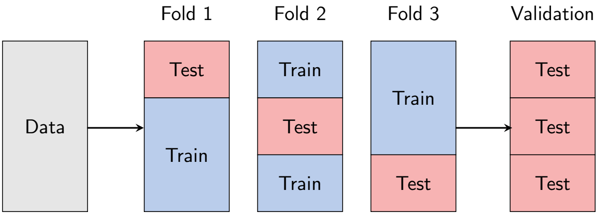


Figure: 3-fold Cross Validation

Back

Oversampling

- Härdle et al. (2009) Trade-off between Type I and Type 2 error in classification Error types
- Balance size of neutral sentences and ones with polarity in sample
- Duplicate sentences within folds of stratified cross validation until the sample is balanced

Back

Classification Error Rates

- Type I error rate $= \text{FP}/(\text{FP} + \text{TP})$
- Type II error rate $= \text{FN}/(\text{FN} + \text{TN})$
- Total error rate $= (\text{FN} + \text{FP})/(\text{TP} + \text{TN} + \text{FP} + \text{FN})$

with TP as true positive, TN as true negative, FP as false positive and FN as false negative.

Back

Stochastic Gradient Descent (SGD)

- Approximately minimize loss function

$$L(\theta) = \sum_{i=1}^{n} L_i(\theta) \tag{18}$$

- Iteratively update

$$\theta_i = \theta_{i-1} - \eta \, \frac{\partial L_i(\theta)}{\partial \theta} \tag{19}$$

SGD Algorithm

1. Choose learning rate η
2. Shuffle data
3. For $i = 1, \ldots, n$, do:

$$\theta_i = \theta_{i-1} - \eta \, \frac{\partial L_i(\theta)}{\partial \theta}$$

Repeat 2 and 3 until approximate minimum obtained.

SGD Example

$X \sim \mathrm{N}(\mu, \sigma)$ and $x_1, \ldots, x_n$ as randomly drawn sample

$$\min_{\theta} \; n^{-1} \sum_{i=1}^{n} (\theta - x_i)^2$$

Update step

$$\theta_i = \theta_{i-1} - 2\eta(\theta_{i-1} - x_i)$$

Optimal gain

Set $2\eta = 1/i$ and obtain $\theta_n = \bar{x}$ with $\bar{x}$ as sample mean.

SGD Example ctd

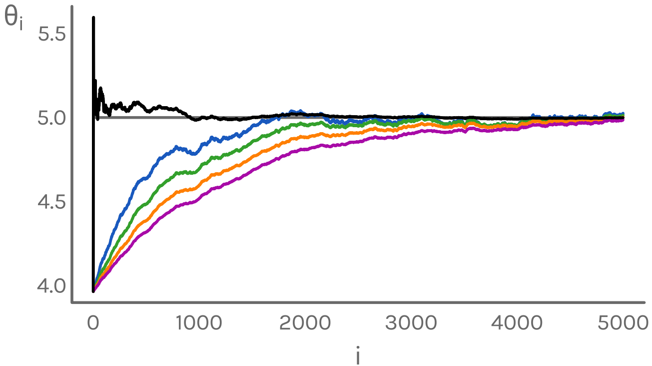


Figure: Estimate Mean via SGD, $x_t \sim \mathrm{N}(5, 1)$

$\eta \in \{1/t,\ 1/1000,\ 1/1500,\ 1/2000,\ 1/2500\}$ TXTSGD

Back

Garman and Klass range-based Measure of Volatility

$$\sigma^2_{i,t} = 0.511(u-d)^2 - 0.019\{c(u+d) - 2ud\} - 0.383c^2 \quad (20)$$

with $u = \log(P^H_{i,t}) - \log(P^O_{i,t}), \quad d = \log(P^L_{i,t}) - \log(P^O_{i,t}),$

$c = \log(P^C_{i,t}) - \log(P^O_{i,t})$

for company i on day t with $P^H_{i,t}$, $P^L_{i,t}$, $P^O_{i,t}$, $P^C_{i,t}$ as highest, lowest, opening and closing stock prices, respectively.

Back

Evaluation Supervised Learning

True \ Pred	-1	0	1	Total
-1	**1,983**	298	254	2,535
0	96	**2,134**	305	2,535
1	105	469	**1,961**	2,535
Total	2,184	2,901	2,520	7,605

Table: Confusion Matrix - Supervised Learning with Oversampling

Back

Choice of λ

- Fine grid with $\lambda_i \in [5 \cdot 10^{-6}, 0.05]$, $i = 1, \ldots, 9999$
- Estimate penalized SVM model
- Results remain stable
  - $\hat{\lambda}_{CV} = 0.000155$
  - Accuracy: 0.8

Choice of λ also possible via information criterion, e.g. Zhang et al. (2016)

Back

Evaluation Logistic Loss Function

True \ Pred	-1	0	1	Total
-1	**397**	55	62	514
0	103	**2,115**	317	2,535
1	58	193	**917**	1,168
Total	558	2,363	1,296	4,217

Table: Confusion Matrix - Supervised Learning, estimated with Oversampling and evaluated on total Sample, Accuracy: 0.80

Back

Abbreviations

Sector	Abbreviation
Consumer Discretionary	CD
Consumer Staples	CS
Energy	EN
Financials	FI
Health Care	HC
Industrials	IN
Information Technology	IT
Materials	MA
Telecommunication	TE
Utilities	UT

Table: Sector Abbreviations

Volatility Regression Returns Regression

Control Variables

$R_{M,t}$ - S&P 500 index return

$\log VIX_t$ - CBOE VIX More Information

$\log \sigma_{i,t}$ - Range-based volatility

$R_{i,t}$ - Return

Back

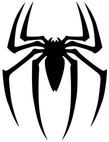
VIX

- Implied volatility
- Measures market expectation of S&P 500
- Calculated by Chicago Board Options Exchange (CBOE)
- Measures 30-day expected volatility
- Calculated with put and call options with more than 23 days and less than 37 days to expiration

Back

Crawling and Scraping

- Automatically extract information from web pages
- Crawling
  - Any information
  - Follows links
  - General information extraction



- Scraper
  - Specific information
  - Specific web pages
  - Easy to obtain high quality data

Legality of Web Scraping

- It is public / Google does it
  - Search engines add value
  - Log in systems, paywalls, ...?
- Highly context specific
  - Commerical v non-commercial
  - Internal v third party use
- Technicalities
  - Bandwidth usage
  - Denial-of-service (DoS) attack

European Union

- Ryanair Ltd v PR Aviation BV (2015)
  - PR Aviation: price comparison of flights
  - Copyright and database right infringement?
  - ToS prohibited data extraction for commercial purposes
- Decision by Court of Justice of the European Union
  - No infringement of intellectual property, no creative input
  - ToS still apply, liability in terms of breach of contract
- In contrast NLA v Meltwater (2013)
  - Scraping of news headlines and links to articles
  - Intellectual property is infringed because of creative input

United States

Pro

- Web data is public, should be accessible
- Unfair market power of Facebook, Google, LinkedIn, ...
- First Amendment protects information gathering

Contra

- Copyright infringement
- Breach of contract
- Violation of the Computer Fraud and Abuse Act (CFAA), 1986
- Trespass to chattels

LinkedIn v hiQ and vice versa

If you exclude someone from sites like LinkedIn, Facebook and Twitter, you are excluding them from the modern version of the town square.

Laurence Tribe, Harvard law professor

- hiQ predicts who is when quitting their job
- LinkedIn: CFAA violation, hiQ: blocked
- LinkedIn ordered to give access to public profiles

Academia is save, right?

Aaron Swartz

- Harvard research fellow
- Automatic download of JSTOR articles
- Laptop in restricted closet at MIT
- No civil law suit by MIT and JSTOR
- Federal charges: wire fraud, CFAA violations
- Possible penalty of $1 million and 35 years in prison

Unclear outcome, suicide on January 11, 2013

Bright Side

Cap Verde is beautiful and does not extradite

Ethical Scraping for Academia

- Technical
  - Use API if provided
  - Appear as a bot, not as a human
  - Provide user agent string with contact data
  - Decreased rate of requests
  - Check robots.txt Google's robots.txt

- Usage
  - Strictly non-commercial
  - Restrict further access to academia

- Ask for permission, not for forgiveness!

Scraping How To

python™

- Complete framework: Scrapy
- Fast and easy: Beautiful Soup
- Low level: lxml

R

- Complete framework: RCrawler
- Fast and easy: rvest
- Low level: XML

Back to Flowchart Nasdaq Articles

Google's robots.txt

```
User-agent: *
Disallow: /search
Allow: /search/about
Allow: /search/howsearchworks
Disallow: /sdch
Disallow: /groups
Disallow: /index.html?
Disallow: /?
Allow: /?hl=
Disallow: /?hl=*&
Allow: /?hl=*&gws_rd=ssl$
Disallow: /?hl=*&*&gws_rd=ssl
Allow: /?gws_rd=ssl$
Allow: /?pt1=true$
Disallow: /imgres
Disallow: /u/
Disallow: /preferences
Disallow: /setprefs
Disallow: /default
Disallow: /m?
Disallow: /m/
Allow:    /m/finance
```

Back