"Textual and Traditional Data in Auditing"

Khrystyna Bochkay
PhD Candidate
Rutgers Business School

Kyung Hee Yoon
PhD Candidate
Rutgers Business School

Kevin Moffitt
Assistant Professor
Rutgers Business School

What is text analytics?

- Existing analytical models rely heavily on structured numerical data.

- But within many domains, some of the most useful information is buried within unstructured texts (emails, customer reviews, tweets, etc.)

- Fortunately, text analytics can bridge the gap.

- **Text mining** (i.e., text data mining or text analytics), is the process of using technology to derive useful information from text.

What can be done?

- Sentiment analysis

- Document summarization, classification, and clustering

- Key-word search

- Association analysis

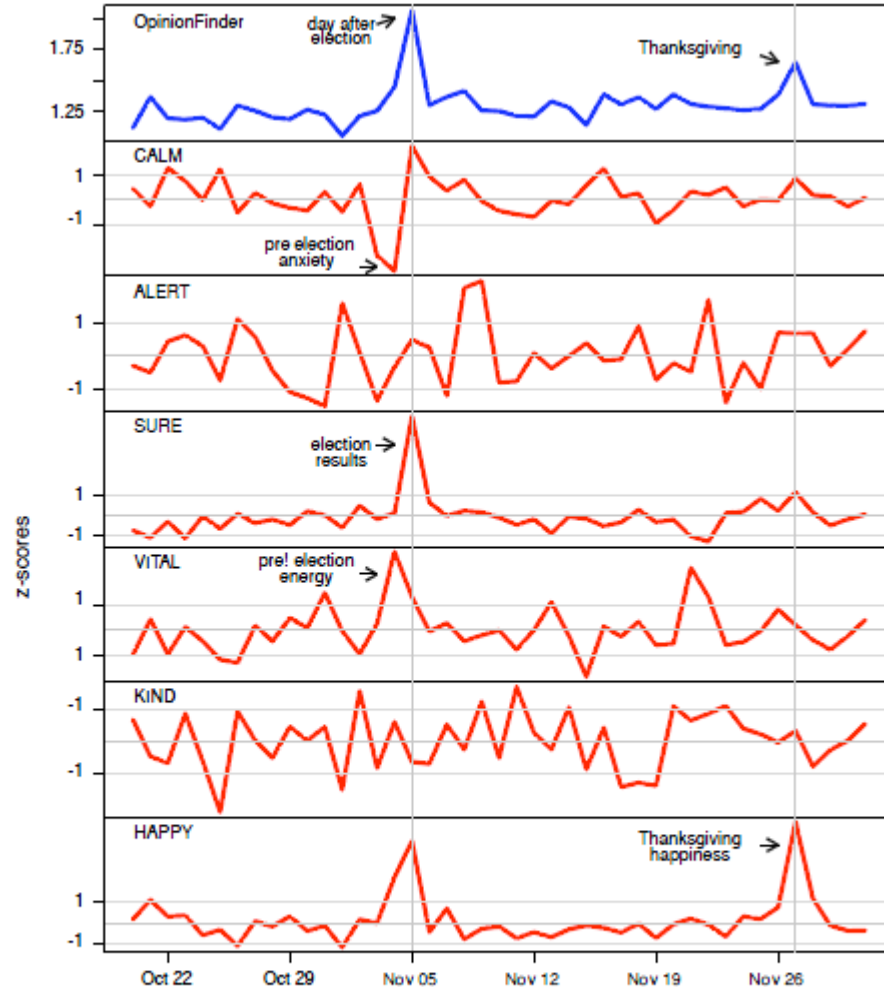- And many other things…

Examples

How people are feeling?
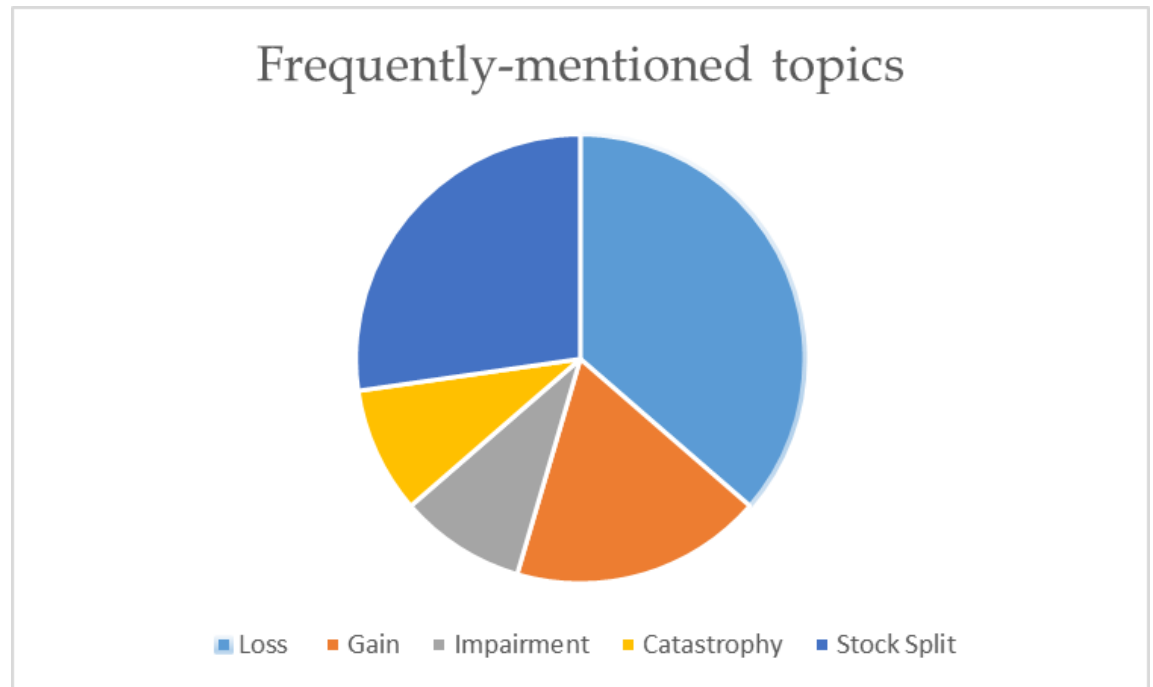

Fig. 2. Tracking public mood states from tweets posted between October 2008 to December 2008 shows public responses to presidential election and thanksgiving.

Examples

What is being discussed?

Frequently-mentioned topics

Loss · Gain · Impairment · Catastrophy · Stock Split

How it is being discussed?

Unexpected

Flood — Decrease

Loss

Large

Inventory

Damage

Text Mining Process

Interpretation /
Evaluation

Data Mining /
Pattern Discovery

Feature Selection

Text Transformation
(Feature Generation)

Text Preprocessing
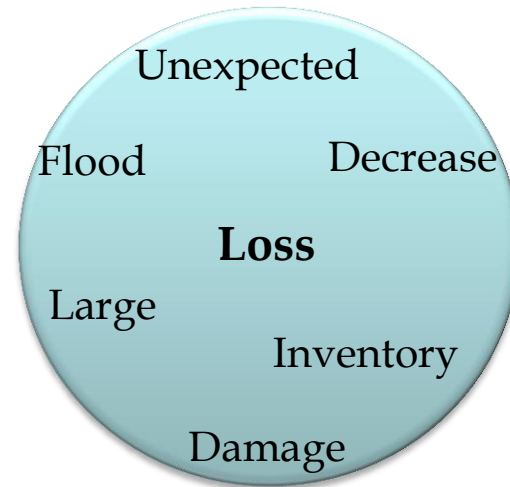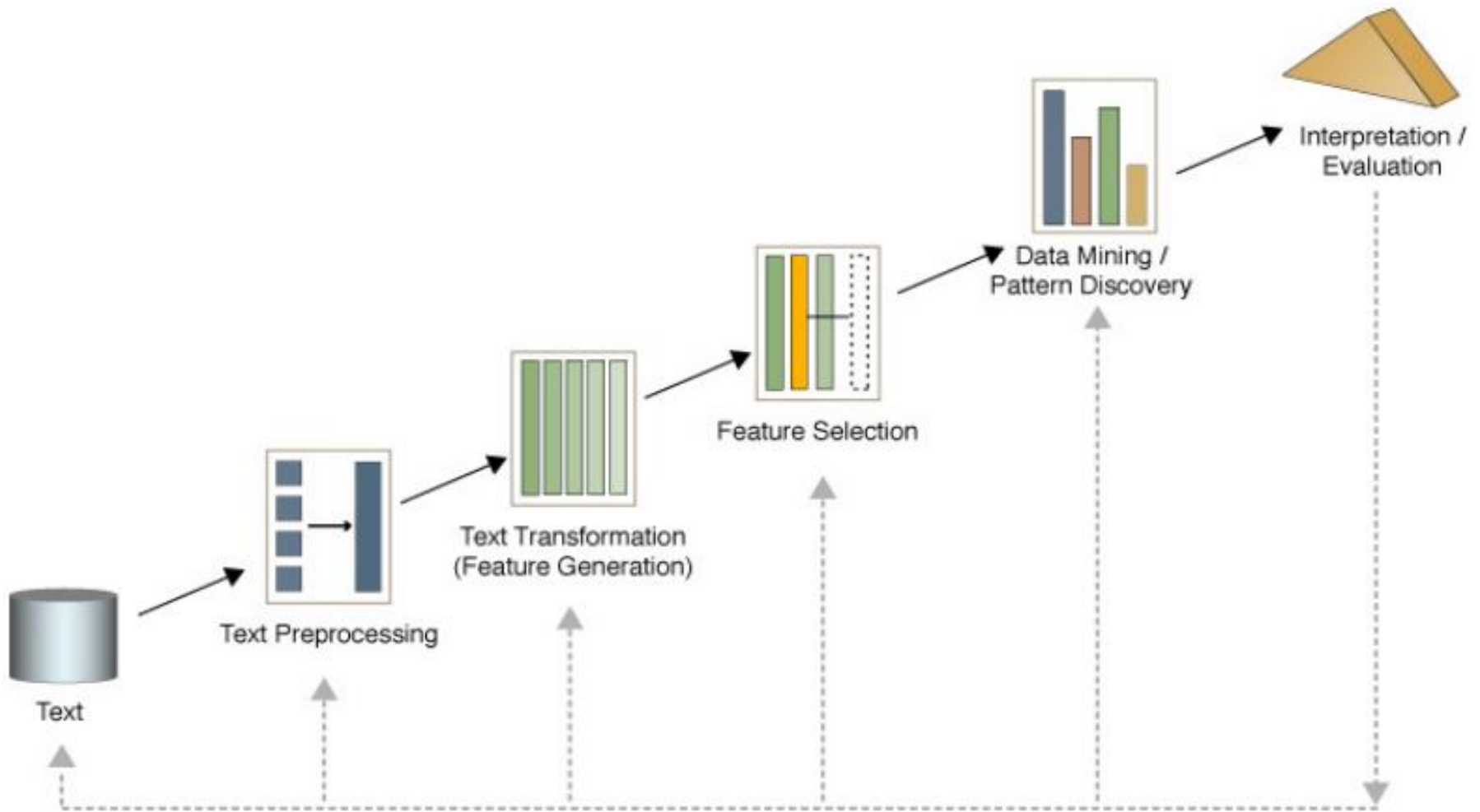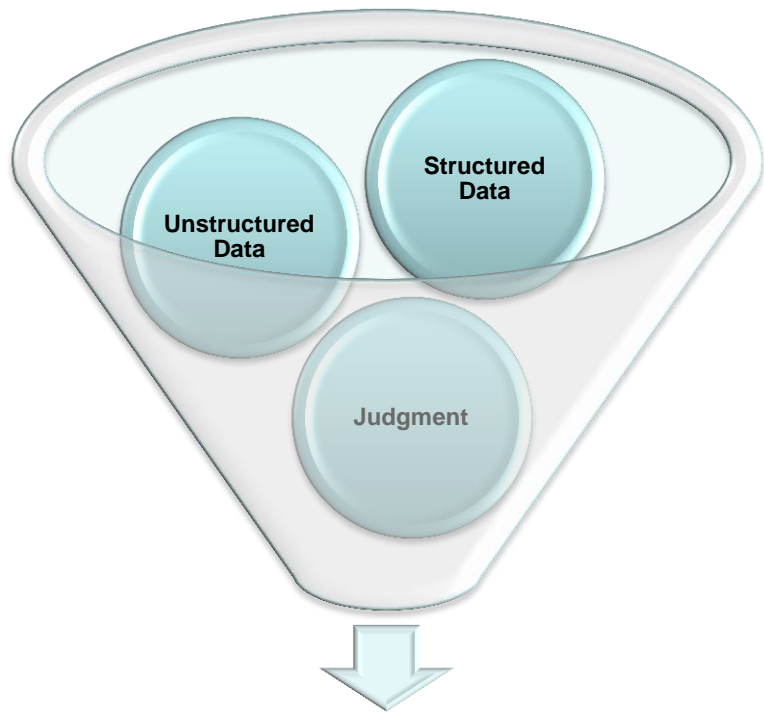
Text

To summarize…

More Informed Decisions

Discovery of new patterns in texts leads to:

- ✓ Better prediction and association models
- ✓ More accurate anomaly detection techniques
- ✓ Faster document filtering, summarization,…

New Environment for Auditors

- A driver of change: Financial reporting scandal and SOX

- An enabler of change: Technology development

-> Holistic Business Risk Audit Approach (Trompeter & Wright, 2010)

Current Research

- Text information to understand firm's fundamental(i.e. Tetlock 2008, Li 2006)

- Auditors' reaction to textual information (Brazel et al. 2012)

My Research

- ## Audit fee and textual information

$$LogAuditFee = B_0 + B_1 Text + B_2 Size + B_3 InvRec + B_3 NumSeg + B_4 Foreign + B_5 Merger + B_6 Special$$
$$+ B_7 Loss + B_8 Btm + B_9 Volatility + B_{10} GC + B_{11} Big4$$

- ## Audit analytical procedure and non financial information

$$\text{Sales}_{at} = \text{Sales}_{bt-1} + \text{Sales}_{at-1} + \text{Inventory } a_t + \text{COGS}_{at-1} + \text{NFI}_t$$

RUTGERS

	Pred	Est.	t-stat
RiskFact	H(+)	0.13	(2.34)**
OptFact	H(-)	-0.13	(-2.67)***
Size	(+)	0.56	(12.35)***
InvRec	(+)	0.88	(3.27)***
NumSeg	(+)	-0.02	(-0.47)
Foreign	(+)	0.23	(2.30)**
Merger	(+)	0.05	(0.27)
Special	(+)	-0.05	(-0.47)
Loss	(+)	0.15	(1.31)*
Growth	(-)	0.07	(1.50)*
Btm	(-)	-0.10	(-1.49)*
LogMktReturn	(-)	-0.01	(-0.12)
Volatility	(+)	0.00	(-0.04)
Big4	(+)	0.28	(2.38)
GC	(+)	0.29	(1.70)
Intercept		7.88	(12.81)***
Industry Effects		Included	
Year Effects		Included	
R^2		0.73	
Model F (p-value)		11.69***	
# observation		260	

Text mining/analytics has many purposes

Understand ambiguous headlines

1. ASTRONAUT TAKES BLAME FOR GAS IN SPACECRAFT
2. IRAQI HEAD SEEKS ARMS
3. PANDA MATING FAILS; VETERINARIAN TAKES OVER
4. LACK OF BRAINS HINDERS RESEARCH

Detect sarcasm

1. **AUDITING IS REALLY, REALLY FUN**
2. **THE BEST MOVIE SINCE STAR WARS EPISODE I**

RUTGERS

Create a useful spam filter

1. **YOUR FREE CREDIT SCORE CAN HELP YOU!**
2. **DO YOU NEED AN URGENT LOAN?**

Help the external auditor

- I am interested in aiding external auditors through text analytic procedures

- Auditors can understand companies better by comparing them to their peers

- In this exercise, I compared risk disclosures between two technology firms (Adobe and McAfee) using a term-weighted (tfidf) matching algorithm

Table 2 Heat map similarity between Adobe (X-axis) and McAfee (Y-axis) for year 2007

Adobe 2007

McAfee 2007	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
01	0	18	2	5	0	3	0	0	3	0	0	0	0	0	0	0	0	3	0	0
02	28	15	4	9	0	5	0	0	6	7	0	0	23	0	0	0	0	5	18	5
03	24	12	0	0	0	16	0	0	0	7	0	0	0	0	0	0	0	0	0	0
04	0	3	19	18	0	4	0	0	5	0	0	0	54	0	0	0	8	4	0	0
05	0	0	0	0	0	0	0	0	0	34	0	0	9	0	0	0	0	0	0	0
06	0	0	0	0	0	0	0	0	49	0	0	16	17	0	0	0	0	0	0	0
07	0	0	0	0	41	0	0	0	0	0	0	12	0	0	0	0	0	54	0	0
08	0	0	0	21	25	0	0	14	0	0	0	28	0	0	16	0	17	0	0	0
09	0	3	3	87	0	4	0	0	5	0	0	0	0	0	0	0	18	4	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	2	2	6	0	2	58	54	3	11	0	2	4	0	3	0	0	2	0	9
12	0	7	8	21	0	9	0	0	11	7	0	0	8	0	0	0	0	10	0	5
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	70	0	0	0
14	0	0	20	22	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0	0
16	0	0	24	0	0	29	0	0	0	0	0	0	0	35	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	4	38	3	0	5	0	0	6	0	0	0	38	0	0	0	0	5	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	0	67	0
20	0	2	2	4	0	2	4	30	3	3	0	10	0	0	37	0	36	2	0	2
21	0	0	0	10	0	0	24	0	25	7	0	8	0	0	0	0	0	0	0	5
22	0	0	0	0	0	0	0	0	0	0	0	0	31	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	64	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	21	0	0	0	0	19	0	0	0	0	0	0	0	0

McAfee risk not in Adobe

- We have experienced, and may continue to experience, material weaknesses and significant deficiencies in our internal control and financial reporting environment, which impacts the accuracy, completeness and timeliness of our external financial reporting.

Adobe risks not in McAfee

- Catastrophic events may disrupt our business.

- If our goodwill or amortizable intangible assets become impaired we may be required to record a significant charge to earnings.

Matched risks

Adobe

- Adverse changes in general economic or political conditions in any of the major countries in which we do business could adversely affect our operating results.

0.87

McAfee

- Conditions and changes in the national and global economic and political environments may adversely affect our business and financial results.

Matched risks

Adobe

- We may incur substantial costs enforcing or acquiring intellectual property rights and defending against third-party claims as a result of litigation or other proceedings.

0.58

- We may not be able to protect our intellectual property rights, including our source code, from third-party infringers, or unauthorized copying, use, disclosure or malicious attack.

0.53

McAfee

- We face numerous risks relating to the enforceability of our intellectual property rights and our use of third party intellectual property, many of which could result in the loss of our intellectual property rights as well as other material adverse impacts on our business and financial results and condition.

Matched risks

Adobe

- We may not realize the anticipated benefits of past or future acquisitions, and integration of these acquisitions may disrupt our business and management.

0.49

McAfee

- We face risks associated with past and future acquisitions.

Matched risks

Adobe

- We rely on distributors to sell our products and any adverse change in our relationship with our distributors could result in a loss of revenue and harm our business.

McAfee

0.34

- We face a number of risks related to our product sales through distributors and other third parties.

Matched risks

Adobe

- Delays in development or shipment of new products or upgrades to existing products could cause a decline in our revenue.

0.28

McAfee

- We face product development risks due to rapid changes in our industry. Failure to keep pace with these changes could harm our business and financial results.

0.24

- If our products do not work properly, we could experience negative publicity, damage to our reputation, legal liability, declining sales and increased expenses.