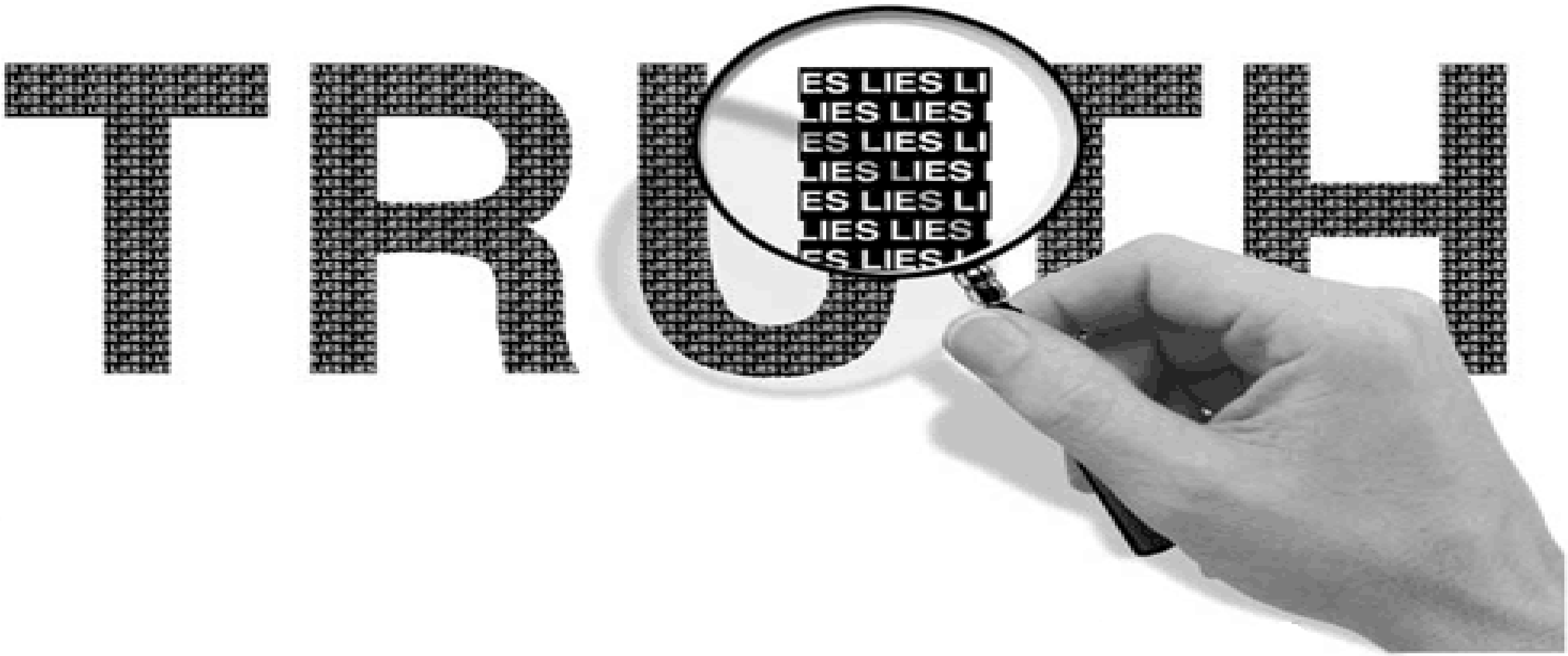


Between the Lines: Decipher the Firms' Fundamentals with Artificial Intelligence

Yanzhen Chen, Huaxia Rui, Andrew Whinston

Earnings Calls: Management Discussion

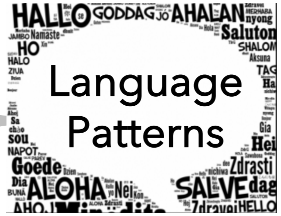
Earnings Calls: Q & A

Research Overview

Evasiveness: Relates Qs and As

Incoherence in responses

Emotional Inconsistencies: Presentations Vs. Answers



Firm Fundamentals

Stock Return

Evasive Answers and Firm Fundamentals

- **Q:** "Are there any issues related to recognizing revenues on these?" -*Analyst*
- **A:** Yes, with the backlog, the vast majority of the wireless backlog is clearly PAS (a product name). I think you saw the announcement at the end of June where we announced on the PAS infrastructure orders in China. And again, it's just the timing of deployment and achieving final acceptance. We've also got some CDMA to a lesser extent in the backlog. ... But Q3 is clearly a little more handset-oriented than we would typically run.

—*Michael Sophie(COO and EVP of UTStarcom)*

- **Results: Labeled as "detour statement", low earnings in the next quarter→ stock price shrunk to 2/3**

Our Big Data approach

- Our approach: a unified framework that integrates
- Machine learning
  - Topic Modeling and Deep Learning
- Big Data technologies
  - Cloud, NoSQL, Condor
- Validated using human ratings
- Automatic processing (vs. manual inspection)
- Finer granularity (vs. discrete)
- Highly consistent

Fintech

- Financial Innovation:
- New technologies, like *machine learning*, *predictive behavioral analytics* and *data-driven marketing*, will take the guess work out of financial decisions.
- Improved data analytics will help refine investment decisions.
- Our Paper=Empirical Asset Pricing + Machine Learning

Research Challenges

- Direction of Empirical research on Disclosure:
  - Validate the ideas of signaling game with natural language processing
- Lacking a proper measure
- Executives try to avoid the curse of

  "No News= Bad News", by tap-dancing around around the topic
- Our approach:

Announcement ➡ **Interactivity**

Solutions

- Open nature of earnings calls
  - The demand of info is observable
  - Directly analyze the language
  - Disclosure-nondisclosure intervals

    A refined semantic-syntactic level
- What they say

  Evasive answers, incoherent answers
- How they talk

  Inconsistency in emotion and cognition

Evasiveness Measure

- ***Approach:*** Topic Modeling
- ***Objectives:***

Consistency, scalability, finer granularity

  – Unsupervised learning to discover latent "topics" from a large collection of documents

- Inputs:
  – Transcripts of text: questions and answers
  – # of latent topics
- Outputs:
  – (1) keywords in each topic,
  – (2) distribution of topics for each question/answer

Topic Modeling

Carlos Kirjner
Analyst, Sanford C. Bernstein & Co. LLC
Q

Hi, two questions. If you run a website with proprietary high quality content today and had to choose a protocol to add metadata [ph] based (35:40) content, why wouldn't the clear choice be the Open Graph protocol instead of RDF or one of its variables? And if that happens and the semantic web arises on the back of Open Graph, doesn't it place Google at a fundamental disadvantage to achieve the vision that Larry laid out of the beginning of the call?

The second question is, what do you think of the future of vertical search? And why is that there are sites that specialize in several verticals such Travel, Local, that seem to do a better job than Google today? Is this going to change over time, and what happens with vertical search? Thank you.

Lawrence E. Page
Chief Executive Officer & Director, Google, Inc.
A

Carlos, I'll take those questions. I think if you – I'm not an expert on the protocols you're talking about. I think in general, we made a huge investment in Knowledge Graph and really understanding in detail about everything, and that's a major effort for us. We'd obviously love to have other people help us with that. I think it has been a little bit of a challenge in the past to get all the labeling aligned and all those things, so I think we have a big part to play in that. We're absolutely very excited about that and I think we're going to do a lot of work in that area. I think we're doing well in that space.

Vertical searches you asked about. I think our goal has always been to get you to the right place. But also to do that we need to really understand in detail your context, what you need, what's really going on with that information, if it's airline tickets, what are the flights between, what do they cost. It's products, some [indiscernible] (37:30). We need to know how much they cost again and what the shipping is and so on. So I think anyplace we can get that information accurately, we can present it to our users, we're very happy to do so. In general, we found that we've needed to really know more of that experience in order to provide a really high quality experience to our users. But again, we're always happy to also working with partners.

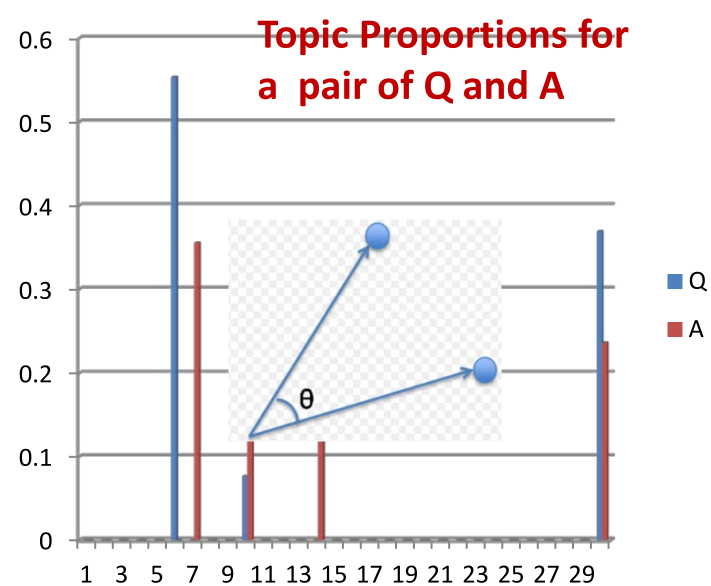
Latent Topics of Qs or As

Table 1: Sample of Latent Topics

Topics	Most Freq Key Words							
1	customers	capacity	production	business	fleet	shipment	fuel	new
2	europe	growth	market	china	brazil	international	u.s	asia
3	million	data	question	launch	clinical	study	fda	going
4	million	year	quarter	guidance	rate	tax	number	impact
5	projects	going	that's	rigs	coast	crude	gulf	oil

Only the top eight words are displayed due to space limit.

LDA



Topic Modeling Output

Table 1: Sample of Latent Topics

Topics	Most Freq Key Words							
1	customers	capacity	production	business	fleet	shipment	fuel	new
2	europe	growth	market	china	brazil	international	u.s	asia
3	million	data	question	launch	clinical	study	fda	going
4	million	year	quarter	guidance	rate	tax	number	impact
5	projects	going	that’s	rigs	coast	crude	gulf	oil

Only the top eight words are displayed due to space limit.

Measure Validation

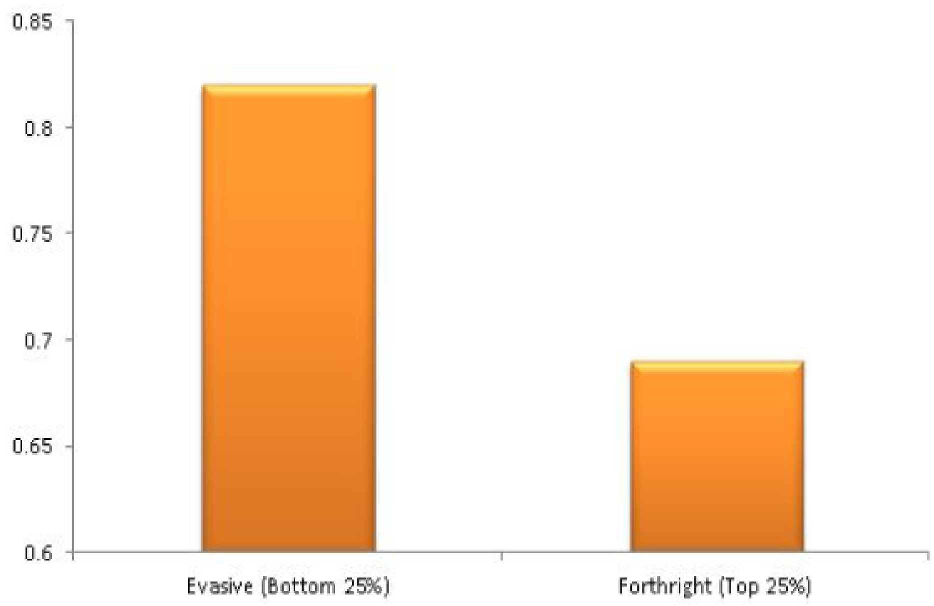


Fig. 1. Fractions of documents overlapped with different measures of evasiveness

- Cross valid the top/bottom 0.25 quartile according to ratings by subjects and LDA algorithm.
- Subjects: 67 business major undergrads

Coherence Measure

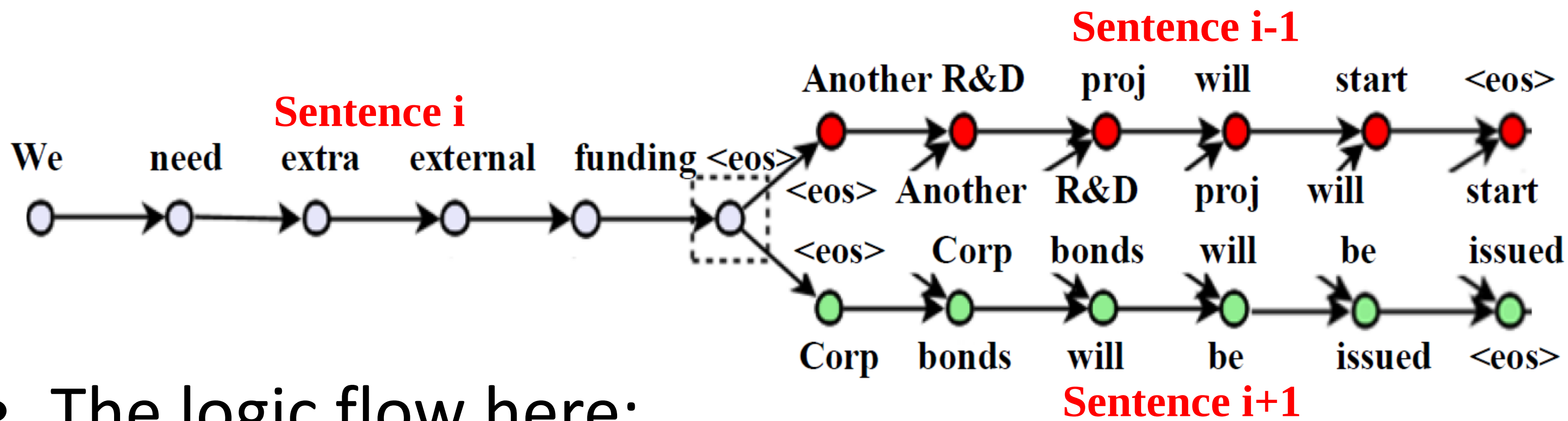
Approach: Deep Learning Model

Objectives: Leverage the *expert knowledge* of the industry and firm from analysts and executives, guide the model to extract important factors rather than engineering them.

Skip Thoughts Model

- Thoughts:
  - A chain of reasoning, a logical path of sorts
- Our Target
  - Exploit information between the lines to represent sentences by thoughts.
- Solution: Skip Thought Models
- Used to predict sentences, inferring missing sentences, etc.

Thoughts on Transit



- The logic flow here:
  - Reason---Consequence---Solution
- Other possible logic flows:
  - A reason---Another reason---Consequence
  - An action---Its impact---Evaluation of the Action
  - and many more…

Skip Thoughts

- Incoherence Measure
  - How smooth are the underlying thoughts?
  - How well does the sentences fits into each others based on the context?

Infer sentence i-1

$$\sum_i \left(\sum_{t_{i+1}} logP(w^t_{i+1}|w^{<t}_{i+1}, h_i) + \sum_{t_{i-1}} logP(w^t_{i-1}|w^{<t}_{i-1}, h_i)\right)$$

Info from sentence i

Infer sentence i+1

How They Talk: Emotion Inconsistencies

Approach: Bag of Words (Dictionary Approach)

Objectives: quantifying emotional levels based on a psych-linguistic literature.

Tools: LIWC

Psychological Processes	Abbrev	Examples
Affective processes	affect	happy, cried
Positive emotion	posemo	love, nice, sweet
Negative emotion	negemo	hurt, ugly, nasty
Anxiety	anx	worried, fearful
Anger	anger	hate, kill, annoyed
Sadness	sad	crying, grief, sad

How They Talk: Emotion Inconsistencies

- Capture nuanced emotional change between MD presentation and Q&A ***<u>on the same topic.</u>***
- Deep Structure of Two Layers:
  - What topics are in the executive's mind;
  - How they feel towards each topic
- Two topic proportions: $\mathrm{T}_{s,t}^{pres}, \mathrm{T}_{s,t}^{A}$
- Topic specific psychological reactions :

Emotion Inconsistencies

- LASSO regression: Calculate topic specific emotional response matrices

$$\min_{L_{s,t}^{pres}} \frac{1}{2} \| P_{s,t}^{pres} - L_{s,t}^{pres} \cdot T_{i,j,t}^{pres} \|_2^2 + \lambda \| L_{s,t}^{pres} \|_1$$

$$\min_{L_{s,t}^{A}} \frac{1}{2} \| P_{s,t}^{A} - L_{i,t}^{A} \cdot T_{s,t}^{A} \|_2^2 + \lambda \| L_{s,t}^{A} \|_1$$

- Emotional Inconsistencies:

$$DP_{s,t}^2 = \| L_{s,t}^{A} - L_{s,t}^{pres} \|_2$$

for a given speaker on the same topic

Data

- Unique Dataset courtesy of Goldman Sachs
  - S&P 500 Companies
  - From 2010 to June-2015
- Large scale textual data:
  - 1.4 Million Q&A Conversations
- Analysts’ forecast data: I/B/E/S
- Financial data: CRSP/Compustat

Results: Next Qtr.'s Earning

DV: SUE	Relationship	Significance
Evasiveness	-0.0363	**
Coherence	0.0371	**
Emotional Inconsistency	-0.0290	*
DV: SAFE	**Relationship**	**Significance**
Evasiveness	-0.0373	**
Coherence	0.0178	*
Emotional Inconsistency	-0.0142	

Results: Next Day's Stock Ret

DV: AR_1	Relationship	Significance
Evasiveness	-0.0246	**
Coherence	0.0051	*
Emotional Inconsistency	-0.0148	
DV: Ret_1	**Relationship**	**Significance**
Evasiveness	-0.0287	**
Coherence	0.0391	
Emotional Inconsistency	-0.0137	

Results: Trading Strategy

	CAPM	Three-Factor	Four-Factor
Alpha	0.2719**	0.2631**	0.2657**
	(0.1134)	(0.1134)	(0.1164)
Market	-0.0001	0.0004*	0.0003
	(0.0011)	(0.0002)	(0.0011)
SMB		-0.0022	-0.0021
		(0.0028)	(0.0028)
HML		-0.0020**	-0.0018*
		(0.0009)	(0.0010)
UMD			0.0003
			(0.0018)
Earnings Call Days	158	158	158
Adjusted R^2	0.0002	0.0049	0.0052

Summary

1. Proposed new machine learning-based measures for investment intelligence.
2. Understand the motivation of answering questions evasively and less coherently.
3. Quantify market reaction of such language patterns.
3. Plan to develop a prototype platform for Fintech real-time recommendation.

Thank you!

"The art of reading between the lines is as old as manipulated information."

– Serge Schmemann