



武汉大学

Wuhan University



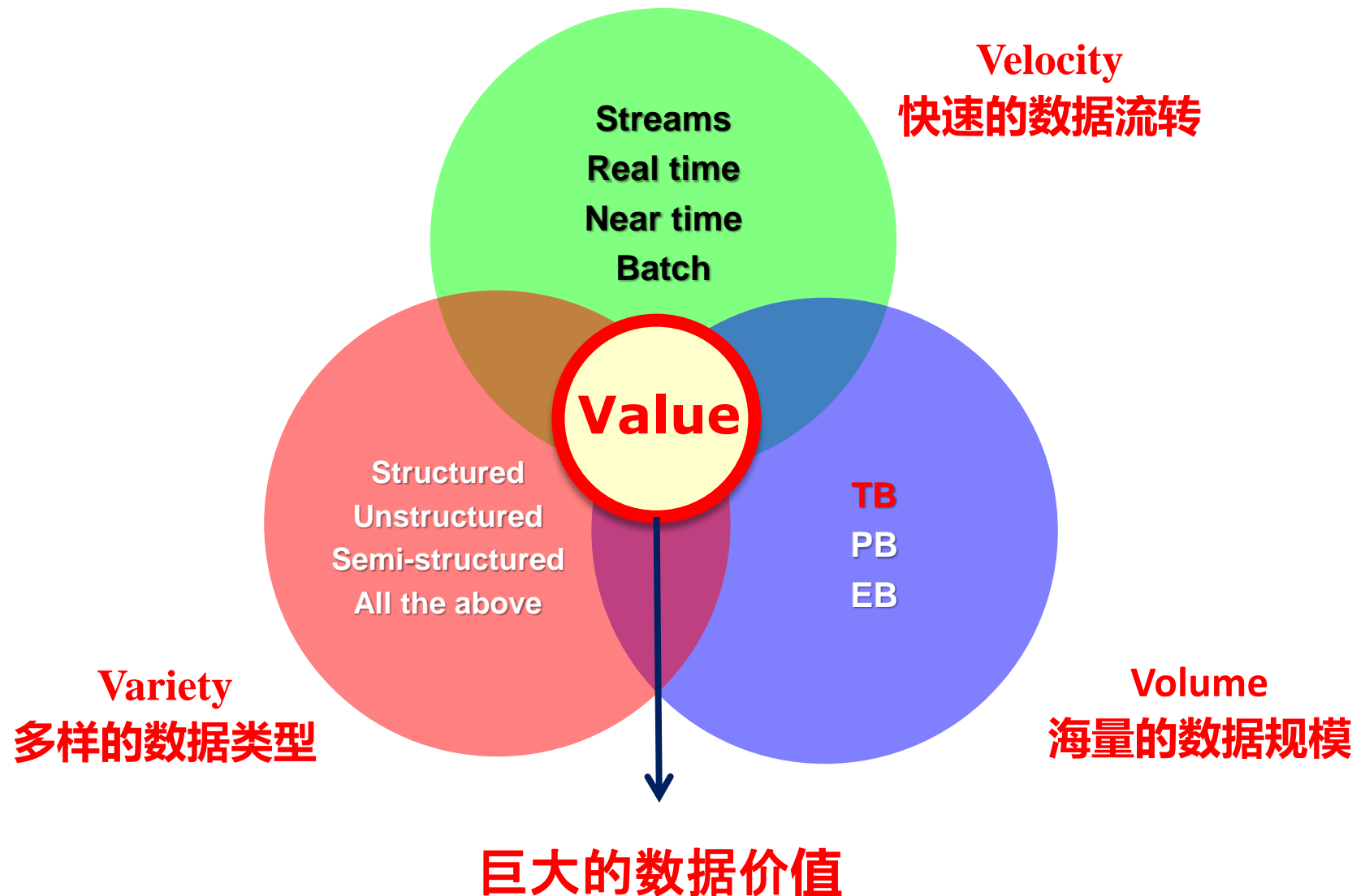
Text Analytics for Social Scientist

高宝俊 教授

武汉大学经济与管理学院

sinkgao@gmail.com

2022年6月



- **词频分析**

- **可读性**

- **情感分析**

- **研究项目**



Contents lists available at ScienceDirect

Journal of Corporate Finance

journal homepage: www.elsevier.com/locate/jcorpfin



Corporate culture and CEO turnover

Turnover ~ Culture X performance



Franco Fiordelisi^{a,b,*}, Ornella Ricci^a

^a University of Rome III, Italy

^b SDA Bocconi, Italy

ARTICLE INFO

Article history:

Received 16 September 2013

Received in revised form 13 November 2013

Accepted 18 November 2013

Available online 23 November 2013

JEL classification:

G14

G21

G34

M14

Keywords:

Corporate culture

Text analysis

Corporate governance

CEO


ABSTRACT

We study the effect of corporate culture on the relationship between firm performance and CEO turnover. Utilising a measure of cultural dimension developed in organisation behaviour research, we quantify corporate culture by assessing official documents using a text analysis approach. We employ this quantification to examine the impact of culture on CEO turnover, especially in the case of poor firm-specific performance. First, we find strong evidence of a negative relationship between firm-specific performance and CEO turnover. Second, we demonstrate that the probability of a CEO change, on average, is positively influenced by the competition- and creation-oriented cultures. The negative relationship between firm-specific performance and CEO turnover is reinforced by the control-oriented culture and reduced by the creation-oriented culture. Finally, we study the CEO insider or outsider succession and observe that the creation-oriented culture has a negative relationship with the probability of hiring an outsider. Moreover, the creation-oriented culture weakens the negative relationship existing between the firm-specific performance under the incumbent CEO and the probability of hiring an outsider.

© 2013 Elsevier B.V. All rights reserved.

Employee Flexibility, Exogenous Risk, and Firm Value

Published online by Cambridge University Press: 12 December 2019

Shiu-Yik Au, Ming Dong and Andreanne Tremblay 


Show author details ▾

Article


Supplementary materials

Metrics

Get access

 Share

 Cite

 Rights & Permissions

Abstract

We hypothesize that employee flexibility enhances firm value by helping firms respond to exogenous shocks. We estimate employee-flexibility scores through textual analysis of online job reviews, and we find that a high flexibility score leads to superior stock returns for firms exposed to external risk. During 2011–2017, the value-weighted hedge portfolio formed on employee flexibility earned a 5-factor annualized alpha of 9.5% during periods of high policy uncertainty. Earnings-announcement returns also suggest that investors do not fully value workforce flexibility. These results indicate that employee flexibility is a valuable corporate intangible that helps firms to manage risk during uncertain times.

| | |
|-------------|---|
| Type | Research Article |
| Information | <i>Journal of Financial and Quantitative Analysis</i> , Volume 56, Issue 3, May 2021, pp. 853 - |

Review of Financial Studeis, 2020, forthcoming

Measuring Corporate Culture Using Machine Learning*

Kai Li
Sauder School of Business
University of British Columbia
2053 Main Mall, Vancouver, BC V6T 1Z2
Canada
kai.li@sauder.ubc.ca

Feng Mai
School of Business
Stevens Institute of Technology
1 Castle Point Terrace
Hoboken, NJ 07030, USA
feng.mai@stevens.edu

Rui Shen
Nanyang Business School
Nanyang Technological University
50 Nanyang Avenue
Singapore 639798
shenrui@ntu.edu.sg

Xinyan Yan
School of Business Administration
University of Dayton
300 College Park
Dayton, OH 45469-2251
xyan02@udayton.edu

This version: September 2018

Abstract

Using one of the latest machine learning techniques—the word embedding model (Mikolov et al. 2013)—and 217,387 earnings call transcripts, we obtain five corporate cultural values: innovation, integrity, quality, respect, and teamwork (Guiso, Sapienza, and Zingales 2015) for 77,541 firm-year observations over the period 2001–2018. We conduct a large number of tests to validate our measures and demonstrate the advantages of our approach over several alternative methods. We show that corporate culture plays an important role in deal incidence and merger pairing, and that post-merger, acquirers' cultural values are positively associated with their target firms' cultural values pre-merger.

Keywords: machine learning; word embedding; unsupervised learning; corporate culture; cultural fit; acculturation; mergers and acquisitions

JEL classification: C45; G34

Text-Based Network Industries and Endogenous Product Differentiation

Gerard Hoberg

University of Southern California

Gordon Phillips

Dartmouth College and National Bureau of Economic Research

We study how firms differ from their competitors using new time-varying measures of product similarity based on text-based analysis of firm 10-K product descriptions. This year-by-year set of product similarity measures allows us to generate a new set of industries in which firms can have their own distinct set of competitors. Our new sets of competitors explain specific discussion of high competition, rivals identified by managers as peer firms, and changes to industry competitors following exogenous industry shocks. We also find evidence that firm R&D and advertising are associated with subsequent differentiation from competitors, consistent with theories of endogenous product differentiation.

文本相似性被用来重新定义行业边界和行业竞争程度。

该文的结论为：基于文本相似性的行业分类比SIC的行业分类更有解释力度。

Journal of Accounting Research

DOI: 10.1111/j.1475-679X.2010.00396.x
Journal of Accounting Research
Vol. 49 No. 2 May 2011
Printed in U.S.A.

CHICAGO BOOTH

Large-Sample Evidence on Firms' Year-over-Year MD&A Modifications

STEPHEN V. BROWN* AND JENNIFER WU TUCKER*

Received 4 January 2010; accepted 6 December 2010

ABSTRACT

The Securities and Exchange Commission (SEC) has expressed concern about the informativeness of firms' Management Discussion and Analysis (MD&A) disclosures. A firm's MD&A is potentially uninformative if it does not change appreciably from the previous year after significant economic changes at the firm. We introduce a measure for narrative disclosure—the degree to which the MD&A differs from the previous disclosure—and provide three findings on the usefulness of MD&A disclosure. First, firms with larger economic changes modify the MD&A more than those with smaller economic changes. Second, the magnitude of stock price responses to 10-K filings is positively associated with the MD&A modification score, but analyst earnings forecast revisions are unassociated with the score, suggesting that investors—but not analysts—use MD&A information. Finally, MD&A modification scores have declined in the past decade even as MD&A disclosures have become longer; the price reaction to MD&A modification scores has also weakened, suggesting a decline in MD&A usefulness.

文本相似性代表了披露内容的模板化，差异性代表了信息披露的增量内容。

结论：信息披露的增量内容与|CAR|显着正相关，表明增量信息被披露——投资者使用该信息。

但分析师在修订盈利预测时未使用MD&A的增量信息内容。

延伸阅读

Brown et al. (2016),
Peterson et al. (2015)
蒋艳辉(2014),
王雄元(2018)



BIG DATA & ANALYTICS IN NETWORKED BUSINESS

TOWARD A BETTER MEASURE OF BUSINESS PROXIMITY: TOPIC MODELING FOR INDUSTRY INTELLIGENCE¹

Zhan (Michael) Shi

Department of Information Systems, W. P. Carey School of Business, Arizona State University,
Tempe, AZ 85287-4606 U.S.A. {zmshi@asu.edu}

Gene Moo Lee

Department of Information Systems and Operations Management, College of Business, The University of Texas at Arlington,
Arlington, TX 76019 U.S.A. {gene.lee@uta.edu}

Andrew B. Whinston

Department of Information, Risk, and Operations Management, McCombs School of Business, The University of Texas at Austin,
Austin, TX 78712 U.S.A. {abw@uts.cc.utexas.edu}

In this article, we propose a new data-analytic approach to measure firms' dyadic business proximity. Specifically, our method analyzes the unstructured texts that describe firms' businesses using the statistical learning technique of topic modeling, and constructs a novel business proximity measure based on the output. When compared with existent methods, our approach is scalable for large datasets and provides finer granularity on quantifying firms' positions in the spaces of product, market, and technology. We then validate our business proximity measure in the context of industry intelligence and show the measure's effectiveness in an empirical application of analyzing mergers and acquisitions in the U.S. high technology industry. Based on the research, we also build a cloud-based information system to facilitate competitive intelligence on the high technology industry.

Keywords: Big data analytics, business proximity, topic modeling, industry intelligence, information system

Spillover Effects of User-Generated Online Product Reviews

The Spillover Effects of User-Generated Online Product Reviews on Purchases: Evidence from Clickstream Data

Completed Research Paper

Young Kwark

University of Florida

young.kwark@warrington.ufl.edu

Gene Moo Lee

University of Texas at Arlington

gene.lee@uta.edu

Paul A. Pavlou

Temple University

pavlou@temple.edu

Liangfei Qiu

University of Florida

liangfei.qiu@warrington.ufl.edu

1. Introduction

EXAMINATION OF CEO-CFO SOCIAL INTERACTION THROUGH LANGUAGE STYLE MATCHING: OUTCOMES FOR THE CFO AND THE ORGANIZATION

ABSTRACT

This study proposes that CEO-CFO language style matching (LSM)—a form of unconscious verbal mimicry based on function words—can provide insights into social interaction processes between CEOs and CFOs. We argue and empirically verify that high CEO-CFO LSM reflects CFOs' strong attempts to ingratiate CEOs. Because ingratiation of superiors can lead to the superiors' positive evaluations of subordinates, CFOs who exhibit higher LSM with CEOs will receive higher compensation and are more likely to become board members of the associated firms. In addition, the proposed relationships will be stronger when CEOs are more powerful. Yet, in the presence of high CEO-CFO LSM, CFOs are less likely to voice different viewpoints and challenge CEOs in strategic decision processes. As a result, firms tend to undertake more mergers and acquisitions (M&As), and such M&As will be paid with a low percentage of cash (vs. stock) and realize lower announcement returns. Using a sample of over 1,000 U.S. firms in 2002-2013, we find empirical support for these predictions.

**Just DM Me (Politely):
Direct Messaging, Politeness, and Hiring Outcomes in Online Labor Markets**

Yili Hong¹, Jing Peng², Gord Burtch³, Ni Huang¹

¹ Bauer College of Business, University of Houston

² School of Business, University of Connecticut

³ Carlson School of Management, University of Minnesota

Abstract

This study examines the role of text-based direct messaging systems in online labor markets, which provide a communication channel between workers and employers, adding a personal touch to the exchange of online labor. We propose the effect of workers' use of the direct messaging system on employers' hiring decisions and conceptualize the information role of direct messaging. To empirically evaluate the information role of the direct messaging system, we leverage data on the direct messaging activities between workers and employers across more than 470,000 job applications on a leading online labor market. We report evidence that direct messaging with a prospective employer increases a worker's probability of being hired by 8.9%. However, the degree to which workers benefit from direct messaging is heterogeneous, and the effect amplifies for workers approaching employers from a position of disadvantage (lacking tenure or fit with the job) and attenuates as more workers attempt to message the same prospective employer. The effects also depend on message content. In particular, we find that the benefits of direct messaging for workers depend a great deal on the politeness of the workers, and this 'politeness effect' depends on several contextual factors. The beneficial effects are amplified for lower-status workers (i.e., workers lacking tenure and job fit), and workers who share a common language with the employer. At the same time, the beneficial effects weaken in the presence of typographic 截图(Alt + A) These findings provide important insights into when and what to message to achieve favorable hiring outcomes in online employment settings.

Keywords: direct messaging system, information value, politeness, online labor market, hiring outcome



Contents lists available at ScienceDirect

Information & Management

journal homepage: www.elsevier.com/locate/im



Recognizing CEO personality and its impact on business performance: Mining linguistic cues from social media

Shichao Wang, Xi Chen^{**}

Zhejiang University, China

ARTICLE INFO

Keywords:

Upper echelons theory
CEO personality
Social media
Business performance

ABSTRACT

Upper echelons theory suggests that CEO personality will influence organizational performance. However, difficulty in measuring CEO personality restrains related research. We capture linguistic cues CEOs leaving on social media and recognize their personality by text mining. To our knowledge, it is the first study introducing social media text mining approaches into the research stream that empirically inquires and extends upper echelons theory. Then, we investigate the CEO personality's impact on both operational and financial performance. Results show that CEO Extraversion, Emotional Stability, and Agreeableness improve Cost Efficiency and Profitability, while CEO Conscientiousness reduces them. CEO Openness to Experience negatively influences Profitability, and all facets of CEO personality improve Employee Productivity except for CEO Conscientiousness. The contribution of our research is multi-sided: (1). methodologically, we introduce a text mining approach to measure CEO personality; (2). theoretically, we provide empirical evidence for upper echelons theory; (3). practically, our results help companies evaluate CEO candidates from a personality perspective.

■ stringr package

- **字符串匹配**: `str_match("The location is great; just walk across the street to the Galleria", "street")`
- `str_match_all("The location is great; just walk across the street to the Galleria", "street.+")`
- **分割字符串**: `str_split("The location is great; just walk across the street to the Galleria", ";")`
- **替换**: `str_replace_all("The location is great; just walk across the street to the Galleria", ";", "")`
-

■ 正则表达式 regular expression

- Corpus, Document, Term
- tm package Demo

■ Term Document Matrix (TDM)

- It is a table that provides the frequency count of every word (term) in each document.
- The number of rows in the TDM is equal to the number of unique terms
- The number of columns is equal to the number of documents.
- This allows us to store an entire library of text inside a single matrix

■ Zipf Law

■ TF-IDF

- This is a weighting scheme provided to sharpen the **importance of rare words** in a document, relative to the frequency of these words in the corpus.
- The TF-IDF is the importance of a word w in a document d in a corpus C
- Therefore it is a function of all these three, we write it as $\text{TF-IDF}(w, d, C)$

■ Term Frequency

$$f(w, d) = \frac{\#w \in d}{|d|}$$

- $|d|$ is the number of words in the document

■ log normalization

$$TF(w, d) = \ln[f(w, d)]$$

■ double normalization

$$TF(w, d) = \frac{1}{2} + \frac{1}{2} \frac{f(w, d)}{\max_{w \in d} f(w, d)}$$

■ Inverse Document Frequency

$$IDF(w, C) = \log_2 \frac{|C|}{|d_{t_i \in d}|}$$

- $|C|$ denotes the total number of documents
- $|d_{t_i \in d}|$ is the number of documents where the term t_i appears.

■ TF-IDF

$$TF\text{-}IDF(w, d, C) = TF(w, d) \times IDF(w, C)$$

- We may also directly use the **weightTfIdf** function in the **tm** package

■ Cosine Similarity

- we like to do is to find similar text or like sentences, eg. web search.
- Since documents are vectors in the TDM, we may want to find the closest vectors or compute the distance between vectors.

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \times \|B\|}$$

- $\|A\| = \sqrt{A \cdot A}$ is the dot product of A

■ Wordlcouds

- interesting ways in which to represent text. They give an instant visual summary. The **wordcloud** package in R may be used to create your own wordclouds.

8

Wang, Mai, and Chiang: *Market Dynamics and UGC About Tablet Computers*
Marketing Science, Articles in Advance, pp. 1–10, © 2013 INFORMS

Figure 4 Tag Cloud Visualization of Top 60 Noun Phrases in Tablet Reviews



of price dispersion. Therefore, all six price variables could be dependent variables for marketing research.

3.2. Translating Sales Ranks to Sales Quantities

The integration of structured product information and unstructured consumer-generated product reviews is another unique feature of our Tablet Computer data set that can support a research trend in textual analysis (e.g., Onishi and Manchanda 2012, Tirunillai and Tellis 2012). That is, research has shown that consumer reviews relate closely to sales (Chevalier and Mayzlin 2006, Liu 2006). However, Amazon does not provide product-level sales information, and data regarding sales volumes of tablet computers are not available in public sources.

Amazon provides sales ranks of products within a particular product category. We could use the log of sales rank as a dependent variable (Archak et al.

- Proclaimed Culture vs. Perceived Culture
- New Corpus + Text Analytics
- 前瞻性
- 战略

- 词频分析
- 可读性
- 情感分析
- 研究项目

- A metric of **how easy** it is to comprehend text.
- **Gunning-Fog Index**
 - Gunning (1952) developed the Fog index. The index estimates the years of formal education needed to understand text on a first reading.
 - A fog index of 12 requires the reading level of a U.S. high school senior (around 18 years old).
 - The index is based on the idea that poor readability is associated with **longer sentences and complex words**. Complex words are those that have more than two syllables. The formula for the Fog index is

$$0.4 \cdot \left[\frac{\backslash\#words}{\backslash\#sentences} + 100 \cdot \left(\frac{\backslash\#complex\ words}{\backslash\#words} \right) \right]$$

■ Coleman and Liau index

- Coleman and Liau (1975) index does not even require a count of syllables
- L is the average number of letters per hundred words and S the average number of sentences per hundred words.

$$CLI = 0.0588L - 0.296S - 15.8$$

- Standard readability metrics may not work well for financial text. Loughran and McDonald (2014) find that the Fog index is inferior to simply looking at 10-K file size.

■ Flesch Reading Ease Score

- With a range of 90-100 easily accessible by a 11-year old, 60-70 being easy to understand for 13-15 year olds, and 0-30 for university graduates.

$$206.835 - 1.015 \left(\frac{\text{\#words}}{\text{\#sentences}} \right) - 84.6 \left(\frac{\text{\#syllables}}{\text{\#words}} \right)$$

■ Flesch-Kincaid Grade Level

- gives a number that corresponds to the grade level. As expected these two measures are negatively correlated.

$$0.39 \left(\frac{\text{\#words}}{\text{\#sentences}} \right) + 11.8 \left(\frac{\text{\#syllables}}{\text{\#words}} \right) - 15.59$$

■ Helpfulness

The image shows two TripAdvisor review profiles with various elements annotated for helpfulness analysis.

Top Profile (karenwl7705DO):

- Rating:** 5 stars (5 green circles)
- Rating date:** Reviewed 1 week ago
- Text comments:** "Fantastic location"
- Travel type:** Stayed February 2017, traveled as a couple
- Multi-dimensional ratings:**
 - Value: 5 stars
 - Location: 5 stars
 - Room Quality: 5 stars
 - Rooms: 5 stars
 - Cleanliness: 5 stars
 - Service: 5 stars
- Other info:** 1 review, 1 helpful vote

Bottom Profile (Hayleylouise87):

- Age & Women:** Level 1 Contributor, 25-34 woman from Birmingham, United Kingdom
- Location:** Birmingham, United Kingdom
- Review distribution (4):**
 - Excellent (0)
 - Very good (2)
 - Average (2)
 - Poor (0)
 - Terrible (0)
- Contributions:** 4 Contributions, 4 Helpful votes, 13 Cities visited
- Experience:** (indicated by a star icon)
- Other info:** 4 reviews, 3 hotel reviews

■ Applications in Finance and Accounting

- Given a goal of efficient markets, regulators want to foster transparency by making sure financial documents that are disseminated to the investing public are readable. Hence, metrics for readability are very important and are recently gaining traction.

■ Li (2008)

- 10-Ks with high Fog index and longer length have lower subsequent earnings.
- Thus managers with poor performance may try to hide this by increasing the complexity of their documents, mostly by increasing the size of their filings.

- **Lehavy et al (2011)**
 - a greater Fog index on 10-Ks is correlated with greater analyst following, more analyst dispersion, and lower accuracy of their forecasts.

■ Loughran and McDonald (2014)

- compare Fog index for 10-K to post filing measures of the **information environment** such as
- volatility of returns, dispersion of analyst recommendations
- When the text is readable, then there should be less dispersion in the information environment, i.e., lower volatility and lower dispersion of analysts expectations around the release of the 10-K.
- Whereas they find that :
- the Fog index does not seem to correlate well with these measures of the information environment,
- the **file size of the 10-K is a much better measure** and is significantly related to return volatility, earnings forecast errors, and earnings forecast dispersion

- DeFranco et al (2013)
 - Combine the Fog, Flesh-Kincaid, and Flesch scores to show that higher readability of **analyst's reports** is related to higher trading volume, suggesting that a better information environment induces people to trade more and not shy away from the market.

- 词频分析
- 可读性
- 情感分析
- 研究项目

- Identify the **orientation of opinion** in a piece of text
- Motivation
 - Movie: is this review positive or negative?
 - Products: what do people think about the new iPhone?
 - Public sentiment: how is consumer confidence? Is despair increasing?
 - Politics: what do people think about this candidate or issue?
 - Prediction: predict election outcomes or market trends from sentiment
- Three Levels: Document, Sentence, Feature level

■ Unsupervised Methods – Lexicon based

- R Code and Packages
- Problems: Amplifier/De-amplifiers, Negations, The "but" Conjunctions, Sarcasm
- Domain dependent

■ Sentiment Lexicons

- Minqing Hu and Bing Liu - OPINION LEXICON @ <http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>
- SentiWordNet @ <http://sentiwordnet.isti.cnr.it/>
- The General Inquirer @ <http://www.wjh.harvard.edu/~inquirer>
- Fiance: Loughran McDonald Master Dictionary_2014 @ http://www3.nd.edu/~mcdonald/Word_Lists.html

- **Scores vs. Polarity**
- **LIWC, Very Common in Business Literature**
- **Supervised Methods**
 - Machine learning
 - Naïve Bayes, SVM, Maximum Entropy Classifier
- **APIs: Stanford CoreNLP**

■ Polarity and Subjective

$$\text{polarity} = \frac{n_{pos} - n_{neg}}{n_{pos} + n_{neg}}$$

$$\text{subjectivity} = \frac{n_{pos} + n_{neg}}{N}$$

- **Numeric Rating vs. Text Sentiment**
- **Aspect Sentiment**

EXPERT BLOGS AND CONSUMER PERCEPTIONS
OF COMPETING BRANDS¹

Xueming Luo

Department of Marketing, Fox School of Business, Temple University,
Philadelphia, PA 19122 U.S.A. {luoxm@temple.edu}

Bin Gu

Department of Information Systems, W. P. Carey School of Business, Arizona State University,
Tempe, AZ 85287 U.S.A. {bin.gu@asu.edu}

Jie Zhang

Department of Information Systems and Operations Management, College of Business, University of Texas at Arlington,
Arlington, TX 76019 U.S.A. {jiezhang@uta.edu}

Chee Wei Phang

Department of Information Management and Information Systems, School of Management, Fudan University,
Shanghai 200433 CHINA {phangcw@fudan.edu.cn}

Social media-based expert blogs are a crucial and credible online information source for consumers, yet little is known about the role of expert blogs for general consumer brand perceptions that bestow long-term value for the firm. On the basis of a novel dataset with 7,871 brand-day observations of over 131,000 expert blogs on major brands in the PC industry, this study reveals how expert blogs may act as a quality signal leading general consumers' perceptions of rival brands. Our analyses reveal three key insights: (1) Expert blog sentiments and volume on a focal brand have a positive relationship with consumer perceptions of the focal brand but a negative relationship with those of its competing brands. (2) Focal brand's blogs have a reinforcing relationship with its subsequent blogs and a cannibalistic relationship with competitors' subsequent blogs. (3) Brand position (leading versus non-leading in the market) plays a nontrivial role in these competitive, dynamic relationships. For example, a 1 percent unexpected increase in Acer's expert blog sentiment would be associated with a 1.5 percent lift in its own brand quality as perceived by consumers, thus winning more consumer hearts and minds. Also, a 1 percent unexpected increase in Acer's expert blog sentiment would be associated with a .7 percent (Toshiba) to 1.5 percent (Sony Vaio) drop in its rival brand quality, thus undermining consumer hearts and minds of competing brands in the same industry. Also, the effects are asymmetrical across leading versus non-leading brands as a 1 percent unexpected increase in expert blog sentiment is associated with a 1.6 percent drop in rival-brand quality perceptions among consumers for non-leading brands, vis-à-vis only a 0.88 percent drop for leading brands. For researchers and managers, these results unravel some neglected benefits of expert blogs.

■ Amazon Mechanical Turk vs. 文本挖掘

Table 4 Top 20 Most Frequent Product Opinions Identified in “Digital Camera” and “Camcorder” Product Categories

| Camera feature | Evaluation | Freq. | Camcorder feature | Evaluation | Freq. |
|-----------------|-------------|-------|--------------------------|-------------|-------|
| Ease of use | Easy | 405 | User friendliness | Easy | 112 |
| Picture quality | Great | 354 | Size/weight | Small | 72 |
| Size | Small | 321 | Video quality | Great | 48 |
| Ease of use | Easy to use | 160 | Video quality | Excellent | 43 |
| Picture quality | Good | 151 | Size/weight | Compact | 38 |
| Picture quality | Excellent | 120 | Video quality | Good | 38 |
| Size | Compact | 115 | Picture/image quality | Good | 36 |
| Picture quality | Clear | 92 | Picture/image quality | Excellent | 35 |
| LCD | Large | 85 | User friendliness | Easy to use | 30 |
| Size | Light | 78 | User friendliness | Great | 23 |
| Picture quality | Sharp | 71 | Different output formats | Minidv | 22 |
| Picture quality | Blurry | 70 | Audio quality | Good | 21 |
| Design | Great | 65 | Size/weight | Light | 20 |
| Ease of use | Very easy | 65 | Size/weight | Lightweight | 19 |
| Picture quality | Amazing | 63 | Picture/image quality | Nice | 18 |
| Video quality | Great | 59 | Picture/image quality | Clear | 17 |
| Ease of use | Simple | 57 | User friendliness | Very easy | 17 |
| Size | Little | 54 | Picture/image quality | Very good | 17 |
| Picture quality | Crisp | 54 | Video quality | Very good | 17 |
| Battery life | Good | 48 | Picture/image quality | Great | 15 |

SOCIAL NETWORK INTEGRATION AND USER CONTENT GENERATION: EVIDENCE FROM NATURAL EXPERIMENTS¹

Ni Huang

W. P. Carey School of Business, Arizona State University, 300 E. Lemon Street,
Tempe, AZ 85287 U.S.A. {ni.huang@asu.edu}

Yili Hong

W. P. Carey School of Business, Arizona State University, 300 E. Lemon Street,
Tempe, AZ 85287 U.S.A. {hong@asu.edu}

Gordon Burtch

Carlson School of Management, University of Minnesota, 321 19th Avenue South,
Minneapolis, MN 55455 U.S.A. {gburtch@umn.edu}

This study examines how social network integration (i.e., integration of online platforms with other social media services, for example, with Facebook or Twitter) can affect the characteristics of user-generated content (volume and linguistic features) in the context of online reviews. Building on the social presence theory, we propose a number of hypotheses on how social network integration affects review volume and linguistic features of review text. We consider two natural experiments at leading online review platforms (Yelp.com and TripAdvisor.com), wherein each implemented a social network integration with Facebook. Constructing a unique panel dataset of online reviews for a matched set of restaurants across the two review sites, we estimate a difference-in-differences (DID) model to assess the impact of social network integration. We find that integration with Facebook increased the production of user-generated content and positive emotion in review text, while simultaneously decreasing cognitive language, negative emotion, and expressions of disagreement (negations) in review text. Our findings demonstrate that social network integration works as a double-edged sword. On the one hand, integration provides benefits in terms of increased review quantity. On the other hand, these benefits appear to come at the cost of reduced review quality, given past research which has found that positive, emotional reviews are perceived by users to be less helpful. We discuss the implications of these results as they relate to the creation of sustainable online social platforms for user content generation.

Keywords: Social network integration, online reviews, natural experiment, difference-in-differences, text analytics

- 词频分析
- 可读性
- 情感分析
- 研究项目

The Helpfulness of Online Reviews

The image shows two TripAdvisor review snippets. The top snippet is for a hotel review by 'karenwl7705DO' with a 5-star rating and a 'Fantastic location' title. The bottom snippet is for a user profile for 'Hayleylouise87' with a Level 1 Contributor status and a 'Review distribution' chart. Red boxes and lines highlight specific elements related to the helpfulness of online reviews.

Top Review Snippet:

- Rating:** 5 stars (indicated by 5 green circles)
- Rating date:** Reviewed 1 week ago
- Text comments:** "Fantastic location"
- Travel type:** Stayed February 2017, traveled as a couple
- Multi-dimensional ratings:**
 - Value: 5 stars
 - Location: 5 stars
 - Sleep Quality: 5 stars
 - Rooms: 5 stars
 - Cleanliness: 5 stars
 - Service: 5 stars

Bottom Review Snippet:

- Age & Women:** Hayleylouise87, Level 1 Contributor
- Location:** Birmingham, United Kingdom
- Contributions:** 4 Contributions, 4 Helpful votes, 13 Cities visited
- Experience:** (indicated by a green checkmark)
- Review distribution (4):**
 - Excellent (0)
 - Very good (2)
 - Average (2)
 - Poor (0)
 - Terrible (0)