huang_2017_analyst_information_discovery_and_interpretation_roles_a_topic_modeling_approach

Year

2017

Author(s)

Huang, Allen and Lehavy, Reuven and Zang, Amy and Zheng, Rong

Title

Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach

Venue

Management Science

Topic labeling

Manual

Focus

Secondary

Type of contribution

Established approach

Underlying technique

Manual labeling

Topic labeling parameters

\

Label generation

Method 1

We manually read the high-probability words in key topics and their respective sentences, to provide a short and intuitive label for each topic.

Table 1. Highest-Probability Words in the Top 10 Topics of Two Large Industries

Topic label	Top 20 words
Capital goods (GICS 2010)	
Comparing financial performance with expectation	Margin, estimate, guidance, EPS, expect, consensus, operating, revenue, lower, consensus, sales, expectation, below, per-share, segment, management, forecast, in-line, beat, outlook
Sales and revenue	Sales, improve, operating, margin, price, profit, revenue, estimate, decline, volume, share, segment, operating-income, improve, off-set, rise, lower, cost, currency, earnings
Growth	Growth, revenue-growth, organic, strong, sales, digit, acquisition, business, rate, expect, EPS-growth, strength, grow, line, margin, solid, core, single, segment, guidance
Business outlook	Business, expansion, good, term, margin, positive, looking, rate, big, difficult, customer, forward, guidance, market, pressure, area, line, opportunity, issue, new
Financial outlook	Revenue, growth, operating, margin, segment, increase, business, expect, year-over-year, forecast, result, acquisition, higher, estimate, decline, compare, income, report, strong, EPS
Valuation model	Multiple, price, stock, earnings, target-price, valuation, estimate, DCF (discount cash flow), cycle, risk, growth, EPS, current, price-to-earnings, group, relative, view, investor, peak, upside
Defense contracts	System, program, defense, contract, space, service, budget, electronic, aircraft, information, ship, missile, government, technology, international, sales, air, support, navy, DOD (department of defense)
Cash flows and financing	Cash, flow, free, share, capital, net, dividend, debt, balance, repurchase, increase, strong, sheet, margin, stock, free-cash-flow, earnings, growth, program, cash-flow
Raw materials and input price	Cost, price, increase, material, pricing, margin, higher, raw, volume, expect, incremental, inflation, impact, commodity, product, off-set, operating, steel, inventory, benefit
Geographic segments	Market, growth, China, Europe, global, emerging, America, demand, region, Asia, India, investment, country, north, economy, expect, middle, economic, European, east
Health care equipment and services (GICS 3510)	
Growth	Growth, margin, revenue, expect, operating, business, rate, gross, digit, market, expansion, improvement, EPS, organic, mix, increase, drive, single, grow, new
Earnings guidance and expectations	Estimate, EPS, guidance, share, expect, range, management, result, expectation, consensus, growth, earnings, impact, per-share, lower, new, in-line, revenue, report, stock
Geographic segments	Sales, division, currency, constant, growth, report, expect, divisional, product, FX (foreign exchange), gross, rate, Europe, business, impact, margin, international, foreign, tax, Japan
Income statement items	Income, net, revenue, expense, operating, after-tax, EPS, margin, gross, interest, share, cost, profit, rate, SG&A, dilute, pre-tax, amortization, item, adjust
Valuation	Estimate, EPS, target-price, multiple, price, share, risk, growth, valuation, price-to-earnings, stock, earnings, rating, base, trade, industry, group, forward, premium, peer
Medical cost	Enrollment, MLR (medical loss rate), commercial, cost, trend, medical, earnings, share, Medicare, expect, ratio, membership, higher, prior, SG&A, live, projection, increase, report, premium
Business outlook and opportunities	Business, positive, term, good, market, future, guidance, impact, looking, forward, rate, new, product, performance, opportunity, better, call, cost, issue, start
Cash flow and financing	Cash, debt, flow, share, net, asset, capital, cash-flow, liability, repurchase, balance, equity, note, investment, free, free-cash-flow, stock, dividend, sheet, expense
Medicare and Medicaid	Medicare, plan, commercial, member, Medicaid, advantage, health, premium, care, benefit, cost, membership, group, enrollment, business, contract, government, risk, Tricare, individual
Drug trial	Announce, disease, drug, product, category, treatment, trial, patient, update, system, new, agreement, Humira (a drug name), study, clinical, program, hub, pharmaceutical, administration, phase

Note. This table reports the top 20 words in each of the top 10 topics and our inferred topic labels for two of the five largest industries in terms of the total number of conference calls in our sample.

Method 2

Next, we invited an expert, who is an institutional investor / all-star analyst covering this industry, to label the intuition of the LDA-generated topics based on their keywords.

Motivation

Method 1

These labels are intended to validate that LDA is able to discern the underlying economic content of the topics.

(Notice also that: These subjective labels have no bearing on the empirical analyses, because the analyses treat each topic as a distinct cluster of words regardless of the label.)

Method 2

We provided the topic intuition generated by the all-star analyst to a human coder (a graduate student in an accounting master's degree program) and asked him to label the topics of each sentence in the call transcripts and analyst reports.

The coder was given only the intuition (i.e. label) of the 60 topics but not the keywords from the LDA outputs to avoid mimicking the LDA results. We asked the coder to assign each sentence to up to three topics due to the challenging nature of manual coding.

The LDA topic assignment is consistent with manual assignment in 69%, 66%, and 60% of the sentences for Campbell Soup, Coca-Cola, and Altria Group, respectively. These consistency rates are much higher than 5%, which is the consistency rate between a random assignment of topics and manual coding.

^ In other words, labels are used to support the evaluation of the LDA procedure, showing that the LDA topic assignment is fairly consistent with manual assignment.

Topic modeling

LDA

Topic modeling parameters

Nr of topics: 60
α: 0.1
β: 0.01

Nr. of topics

60

Label

Single or multi-word label manually assigned by the authors (Method 1) or by a domain expert (Method 2)

Label selection

\

Label quality evaluation

\

Assessors

\

Domain

Paper: Finance
Dataset: Finance

Problem statement

This study examines analyst information intermediary roles using a textual analysis of analyst reports and corporate disclosures.
We employ a topic modeling methodology from computational linguistic research to compare the thematic content of a large sample of analyst reports issued promptly after earnings conference calls with the content of the calls themselves.

Corpus

Origin: Thomson Reuters
Nr. of documents: 495.240
Details:
- 18,607 transcripts obtained from Thomson Reuters's StreetEvents Database from all available earnings conference calls
- 476,633 analyst reports obtained from Thomson Reuters' Investext Database) for S&P 500 firms during the period 2003–2012.

Document

Transcript and records content, together with company names, tickers, dates and time of the calls, participants and their titles.

Pre-processing

- Parsing XML transcripts and PDF reports to TXT
- First, we exclude venue-specific language or time- invariant information that is not associated with any eco-nomically relevant topics.
- For conference calls, we exclude narratives from operators, greeting words used by various speakers, and the safe harbor statements typically read by investor relations officers
- For analyst reports, we remove the tables, graphs, and "brokerage disclosures."
- all words into lower case
- remove all non-English characters (e.g., punctuation and numbers)
- convert all plural nouns into their singular forms
- remove stop words
- delete more contextual words that distinguish oral language from written language
- convert high-frequency phrases that constitute specific financial/ technical terms into one word (For example, "target price" is converted into "target-price,")
- remove S&P 500 company names and tickers

```
@article{huang_2017_analyst_information_discovery_and_interpretation_roles_a_to
pic_modeling_approach,
author = {Huang, Allen and Lehavy, Reuven and Zang, Amy and Zheng, Rong},
year = {2017},
month = {06},
pages = {},
title = {Analyst Information Discovery and Interpretation Roles: A Topic
Modeling Approach},
volume = {64},
journal = {Management Science},
doi = {10.1287/mnsc.2017.2751}
}
```