◯ BRAIN

Brain Language Metrics on Earnings Call Transcripts

Product Summary

The exploitation of textual unstructured content (news, company filings, earnings calls etc) in financial analysis is quickly expanding across both quantitative and discretionary strategies as reflected in the growing number of academic papers and products in this domain.

The Brain Language Metrics on Earnings Calls Transcripts (BLMECT) dataset has the objective of monitoring several language metrics for the quarterly earnings call transcripts of 4500+ US stocks.

The dataset is composed of two parts. Part one includes several **language metrics** for the most recent earnings call transcript for each stock, namely:

1. Financial sentiment calculated using **Brain proprietary Large Language Model** approach

2. Percentage of words belonging to financial domain classified by language types:

    • *"Constraining"* language
    • *"Litigious"* language
    • *"Uncertainty"* language

3. Readability score

4. Lexical metrics such as lexical density and richness of text

5. Text statistics such as the transcript length

Part two includes the **differences** between the most recent earnings call transcript and the previous one:

1. Difference of the various language metrics (e.g. delta sentiment, delta readability score, delta percentage of a specific language type etc.)

2. Similarity metrics between documents, also with respect to a specific language type (for example similarity with respect to *"litigious"* language or *"uncertainty"* language)

The metrics calculation is reported separately for the following **sections** of the transcript:

a. Management Discussion
b. Analysts' Questions
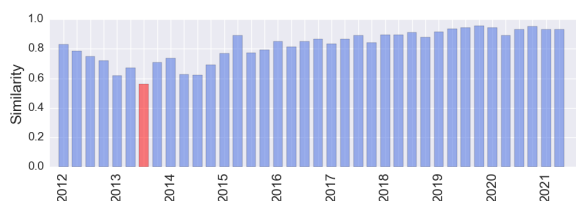c. Management Answers to Analysts' Questions

Dataset Frequency

The dataset is updated daily as new earnings call transcripts are published for some stocks. Data for each stock changes quarterly with new earnings calls. The historical dataset is available starting from 2012.
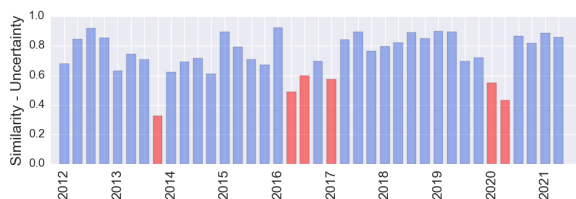
Examples of Metrics

One of the dataset metrics is the similarity between the most recent earnings call transcript and the previous one for various sections of the transcript.

Similarity with focus on generic financial domain language ("Management Discussion" section) - MSFT

Similarity with focus on "uncertainty" financial domain language ("Management Discussion" section) - MSFT

Another metric included in the dataset is the difference of sentiment score measuring the tone of the Analyst Questions as shown below.

Difference of Sentiment Score in "Analyst Questions" in AAPL Earnings Calls Transcripts.

Contacts

BRAIN is a research focused company that develops proprietary signals based on alternative data and algorithms for investment strategies on financial markets.

• EMAIL: contact@braincompany.co
• WEB: https://braincompany.co