Summary of Uncovering Information: Can AI Tell Us Where to Look?

Lixuan Feng 20250418

1 What are the research questions?

- Can LLMs detect and measure newsworthy(new or surprising) textual information within financial disclosures?

2 Why are the research questions interesting?

- Relevant information is important but often buried in a sea of text, hard to find.
- Prior Study measuring information in narrative disclosures
  - like measuring properties of text is straightforward, but limited in assessing surprising disclosures due to their reliance on subjective word classification.
  - like comparing a firm's new 10K to the prior year's is insightful, but the large volume of yearly disclosures makes such direct comparisons challenging.
- Use generative AI technology to build a measure of information that is context-specific
  - help stakeholders quickly sift through unstructured information to find what's important.
  - accounts for every type of disclosure released by the firm and filed in EDGAR.

3 What is the paper's contribution?

- Development of a method for identifying information within unstructured data
- Literature on use the output of LLMs to predict returns
  - Prior Study: LLM is used to transform a source document into another representation.
  - Extend: preserves the source document and uses an LLM to overlay a measure of information
- Literature on costly information processing
  - Prior Study: study annual and quarterly reports.
  - Extend: forming priors over all disclosures.

4 What hypotheses are tested in the paper?

- LLMs can identify new information from company communicates.
  - measurement explains the majority of the market reaction

5 Sample: appropriateness of the sample selection procedures

- Every type of disclosure released by the firm and filed in EDGAR.

6 Comment on the appropriateness of variable definition and measurement

- Market Reaction; Information Content(LLMs)

7 Comment on the appropriateness of the regress/predict model

- LLMs accurately estimate word probabilities based on context.

8 What difficulties arise in drawing inferences from the empirical work?

- Limited Attention: Selective Focus on Specific Document Types(this paper only consider this?) or Selective Focus within Documents

9 Describe at least one publishable and feasible extension of this research.

- Use LLMs to extract new ESG report info by training on past disclosures and identifying surprising content.

10 Describe the relationship between papers.

- They all investigate whether LLMs can effectively analyze and extract information from corporate or news texts.