

Learning Industry-Specific Selective Disclosures from SEC 10-K Regulatory Filings

The recent policy change permitting foreign firms to submit regulatory filings to the U.S. Securities and Exchange Commission (SEC) using International Financial Reporting Standards (IFRS) rather than U.S. Generally Accepted Accounting Principles (GAAP), combined with the SEC's on-going XBRL Voluntary Filing Plan (VFP), hints at a future where all firms will support continuous, XBRL-based financial reporting. Discussions surrounding XBRL have focused on automated querying and analysis of financial facts and figures, largely overlooking the significance of textual narratives such as Management's Discussion and Analysis (MD&A). Prior research in the finance and accounting literature has established the significance of selective disclosures revealed in narratives like MD&A. In this paper, we complement XBRL-based automated reporting by developing automated measures for analyzing textual narratives in SEC 10-K regulatory filings. We introduce a structured extension to Bayesian topic modeling that parallels structured information retrieval; where terms appear within the document alter their significance. A genetic algorithm automatically identifies textual narratives within the filing and is robust in the face of mislabeled and incorrectly ordered elements. A language model, weighed by document structure and conditioned on industry sector learns topics within text passages. Selective disclosures are evaluated by comparing automatically learned topics to Annual Reviews of Corporate Reporting Practices from the Association for Investment Management and Research.

Current status of the paper:

While the genetic algorithm for identifying passages has been implemented and evaluated, the model for structure-weighted topic discovery is still being implemented. Completion of the paper will depend upon the success of the algorithm as we try and complete it over the holidays. I would very much like to present this work, but there is still implementation work to be done.