Risk Analysis Based on 10-K Item 1a

Kevin Moffitt

Yue Liu

Background

- Firms report "Risk Factors" in their 10-K item 1A (beginning in 2005)

  -Information about the most significant risks that apply to the company or to its securities

  -listed in order of their importance. Some risks may be true for the entire economy, some may apply only to the company's industry sector or geographic region, and some may be unique to the company. (SEC)

  -Smaller reporting companies are not required to report (they may do so voluntarily)

  -No formal terms for the risks (described in sentences and paragraphs)→similar risks may be expressed differently

Literature

Table 1. Summary of studies that involve identification of risk categories

Author-year	Unit of analysis	# of risk categories	Method of defining risk categories
Campbell et al. (2014)	item 1A	5	subjectively define risk categories
Huang and Li (2011)	individual risk factors	25	manually identify risk categories by reading 10-Ks
Mirakur (2011)	individual risk factors	29	manually identify risk categories by reading 10-Ks
Miihkinen (2013)	risk disclosures in Finland	5	manually identify risk categories
Bao and Datta (2014)	individual risk factors	30	sent-LDA topic model with a predefined number of clusters

Objectives of this paper

- Automatically identify similar risks from individual risk factors in item 1A to generate risk categories

  - Does not need the arbitrary set of number of risk categories beforehand

  -Evaluate our method

  -Further develop a risk measure for risk analysis

  -Develop a system that can help auditors with decision making from the aspect of risks (unique risks, risk supposed to have but actually not, etc.)

- Most prior studies that involve risk categorization based on Item 1A either manually generate risk categories, and no studies specifically develop a system to assist audit

Methodology

- Data preprocessing

1. Download 10-Ks for the retailing industry (sic starting with 52-59) from 2007 to 2015, and extract item 1a

2. Split item 1a into risk factors and extract each into a separate text file (we call them risk factor files)

- Pilot study

- Top 10 companies in the retailing industry with the largest average total assets in 2007-2015

Specifically: Walmart, CVS, Target, Home Depot, McDonald's, Lowe's, Amazon, Kroger, Costco, Macy's

- 1,420 risk factors for the 10 company in 2007-2015 (raw file)

- Short version of risk factors: keep the first 35 words for each risk factor (short file)

- Extract important words for each short file

- Extract important phrases for each short file (weighted by importance)

- Use Microsoft Bing Search API to search each important phrase, and keep the top 25 hits (weighted by rank)

- Use Wikipedia to search each important phrase, and keep the top 10 hits (weighted by rank)

- Calculate similarity of risk factors for:
1. The same company over different years
2. Different companies in the same year
- 6 different similarity scores:
1. similarity_word_weak
2. similarity_word_strong
3. similarity_phrase_match_weak
4. similarity_phrase_match_strong
5. similarity_bing
6. similarity_wikipedia

- similarity_word_weak/strong: based on important words for each file

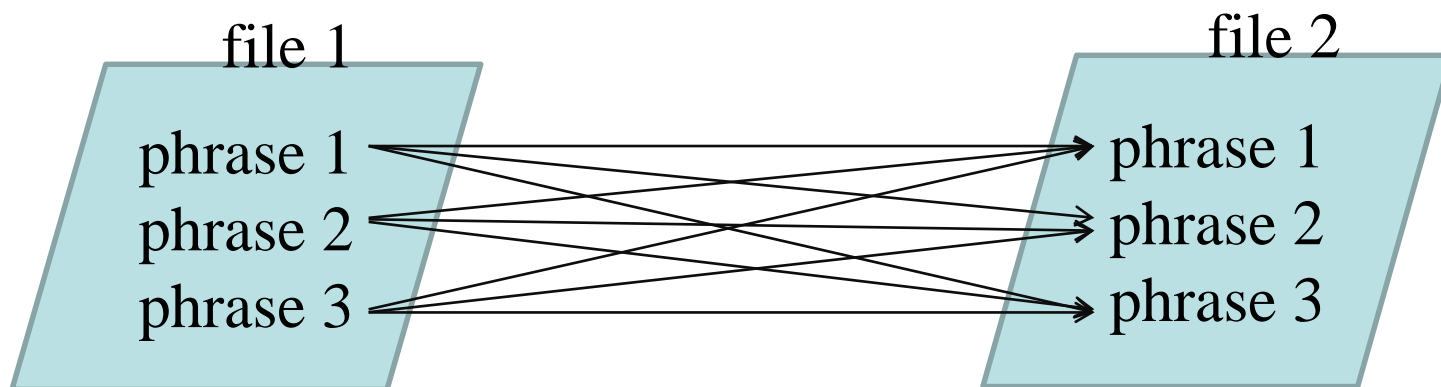$$\text{similarity\_word\_weak}_{12} = \frac{\text{\# of words same in both files}}{\text{total \# of words in file } 1}$$

$$\text{similarity\_word\_}strong_{12} = \frac{\text{\# of words same in both files}}{\sqrt{\text{total \# of words in file } 1}}$$

- similarity_phrase_match_weak/strong: based on important phrases for each file

$$\text{similarity\_phrase\_match\_weak}_{12} = \sum_{i=1}^{3} phrase\_match_{1i} \cdot weight_{1i}$$

$$\text{similarity\_phrase\_match\_}strong_{12} = \sum_{i=1}^{3} phrase\_match_{1i} \cdot 1.5weight_{1i}$$

- similarity_bing/wikipedia:

-Use Bing/Wikipedia search results to add similarity score for phrases share the same links.

Preliminary Results-examples

Matched Risks

MACY'S INC
2012
794367_2012-03-28-fs14.txt

the company's sales and operating results could be adversely affected by product safety concerns. if the company's merchandise offerin do not meet applicable safety standards or our consumers' expectations regarding safety, the company could experience decreased sale experience increased costs and/or be exposed to legal and reputational risk. events that give rise to actual, potential or perceived produ safety concerns could expose the company to government enforcement action and/or private litigation. reputational damage caused by or perceived product safety concerns could have a negative impact on the company's sales and operating results.

the company's sales and operating results could be adversely affected by product safety concerns. if the company's merchandise offerin do not meet applicable safety standards or our consumers' expectations regarding safety, the company could experience

word_weak	word_strong	phrase_match_weak	phrase_match_strong	phrase_bing	phrase_wiki
0.636363636	2.110579412	1	1.5	1	1.138457542

TARGET CORP
2012
27419_2012-03-15-b10.txt

failure to address product safety concerns could adversely affect our sales and results of operations. if our merchandise offerings, including food, drug and children's products, do not meet applicable safety standards or our guests' expectations regarding safety, we could experience lost sales and increased costs and be exposed to legal and reputational risk. all of our vendors must comply with applicable product safety laws, and we are dependent on them to ensure that the products we buy comply with all safety standards. events that give rise to actual, potential or perceived product safety concerns, including food or drug contamination, could expose us to government enforcement action or private litigation and result in costly product recalls and other liabilities. in addition, negative guest perceptions regarding the safety of the products we sell could cause our guests to seek alternative sources for their needs, resulting in sales in those circumstances it may be difficult and costly for us to regain the confidence of our quests.

第 1 项(共 2194 无筛选器 搜索

- MACY-2012:

- the company's sales and operating results could be adversely affected by product safety concerns. if the company's merchandise offerings do not meet applicable safety standards or our consumers' expectations regarding safety, the company could experience

- TARGET-2012

- failure to address product safety concerns could adversely affect our sales and results of operations. if our merchandise offerings, including food, drug and children's products, do not meet applicable safety standards or our guests'

- Similarity scores:

word_weak	word_strong	phrase_match_weak	phrase_match_strong	phrase_bing	phrase_wiki
0.636363636	2.110579412	1	1.5	1	1.138457542

- HOME DEPOT-2015
- if we are unable to effectively manage and expand our alliances and relationships with selected suppliers of both brand name and proprietary products, we may be unable to effectively execute our strategy to differentiate ourselves

- LOWE'S-2015
- our inability to effectively manage our relationships with selected suppliers of brand name products could negatively impact our business plan and financial results. we form strategic relationships with selected suppliers to market and develop products

- Similarity scores:

word_weak	word_strong	phrase_match_weak	phrase_match_strong	phrase_bing	phrase_wiki
0.4	0.894427191	0.75494187	1.132412805	0.779112131	0.793160397

- LOWE'S-2014

- future litigation or governmental proceedings could result in material adverse consequences, including judgments or settlements. we are, and in the future will become, involved in lawsuits, regulatory inquiries, and governmental and other legal proceedings arising

- WALMART-2014

- we are subject to certain legal proceedings that may adversely affect our results of operations, financial condition and liquidity. we are involved in a number of legal proceedings, which include consumer, employment, tort and other

- Similarity scores:

word_weak	word_strong	phrase_match_weak	phrase_match_strong	phrase_bing	phrase_wiki
0.166666667	0.577350269	0	0	0	0

Future steps

- Add some other measures of similarity (e.g., based on other dictionaries/websites/compare the sentiment of the risks)
- Manually code whether two files are matched or not to validate the similarity measures
- Cluster files based on file similarity and label the clusters
- Further risk analysis