

RiskFinder: A Sentence-level Risk Detector for Financial Reports



Yu-Wen Liu and Ming-Feng Tsai
Department of Computer Science
National Chengchi University, Taiwan



Chuan-Ju Wang
Research Center of Information Technology Innovation
Academia Sinica, Taiwan



Liang-Chih Liu
Department of Information and Finance Management
National Taipei University of Technology, Taiwan

Abstract

This paper presents a web-based information system, RiskFinder, for facilitating the analyses of soft (textual information) and hard information (numerical measures) in financial reports and broadens the analyses from word level to sentence level.

System Description



Metadata

metadata, including basic information on each company as well as several relevant financial measures.

1. Company name
2. Company's CUSIP number
3. Report release date
4. Annualized post-event return volatilities
5. Number and percentage of highlighted risk-related sentences in selected report

Visualization

Our system displays time-aligned quantitative information (e.g. **historical prices** and **trading volumes**) to facilitate the analysis of textual and numeric information and capture more insight in the financial risk.

- Red Vertical Line: release date of the report
- Users can adjust the window (6) to show the corresponding quantitative information for a certain period.

Risky Sentences Detector

RiskFinder focuses on the sentence-level investigation. To obtain better sentence representations, we adopt two sentence embedding techniques, FastText and Siamese-CBOW, to conduct sentence-level textual analyses. Due to the significant superior performance of FastText, we adopt it to highlight high-risk sentences in the reports in the proposed system.

	Siamese-CBOW	FastText
Accuracy	0.656	0.813
Precision	0.776	0.774
Recall	0.438	0.621
F1-measure	0.558	0.648

* 4-fold cross-validation

Risk-Sentiment Dataset

Form 10-K risk-sentiment dataset provides the financial risk-sentiment dataset consisting of two types of data:

1. A set of risk-labeled financial sentences
2. The pre-trained sentence embeddings.

	High-Risk	Total
# of Sentence	138	432

Cronbach's alpha: 0.784

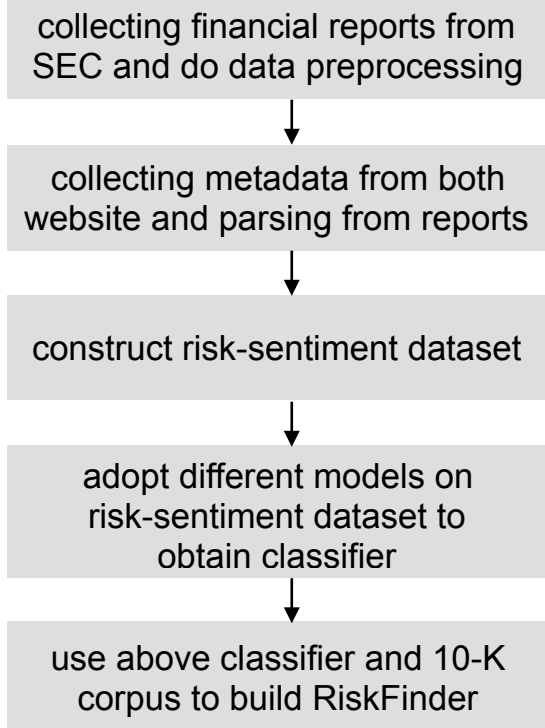


Figure. System Construction Flow

* RiskFinder is now online:
<https://cfda.csie.org/RiskFinder/>

