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.

collected financial reports from SEC and do data preprocessing

collecting metadata from both website and parsing from reports

construct risk-sentiment dataset

adopt different models on risk-sentiment dataset to obtain classifier

use above classifier and 10-K corpus to build RiskFinder

Figure. System Construction Flow

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

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

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.

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.

### Risk-Sentiment Dataset

| Model          | Accuracy | Precision | Recall | F1-measure |
|----------------|----------|-----------|--------|------------|
| Siamese-CBOW   | 0.656    | 0.776     | 0.438  | 0.558      |
| FastText       | 0.813    | 0.774     | 0.621  | 0.684      |

* 4-fold cross-validation

### RiskFinder

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.

### 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.

### 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.