

Predicting Changes to Moody's Credit Rating

Importance

The announcements of Moody's credit rating change have impact on the financial market. Predicting them in advance enables making more informed investment decisions. Vanguard is interested in predicting the next Moody's rating change to generate alpha.

Objective

Predict changes in Moody's credit rating for July and August 2018 for five companies: 21st Century Fox, AT&T, CBS, Comcast and Verizon. Predictions are made based on Security and Exchange Commission (SEC) filings and news headlines.

Moody's announcements are **rare**, highly **irregular** and **imbalanced**!

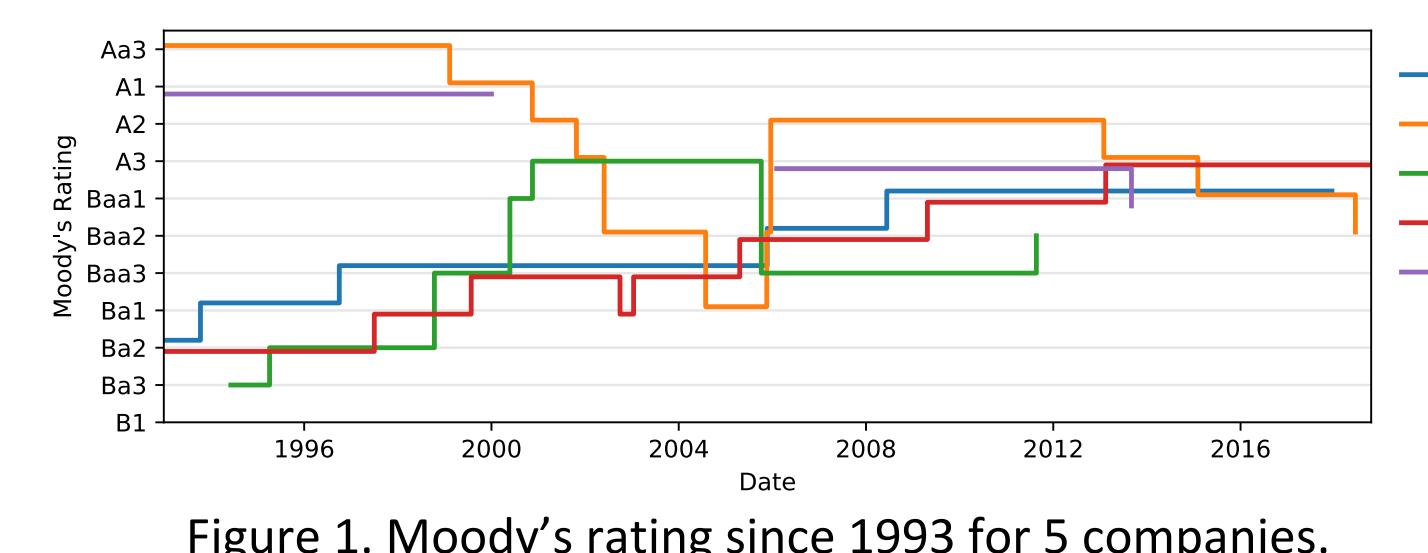


Figure 1. Moody's rating since 1993 for 5 companies.

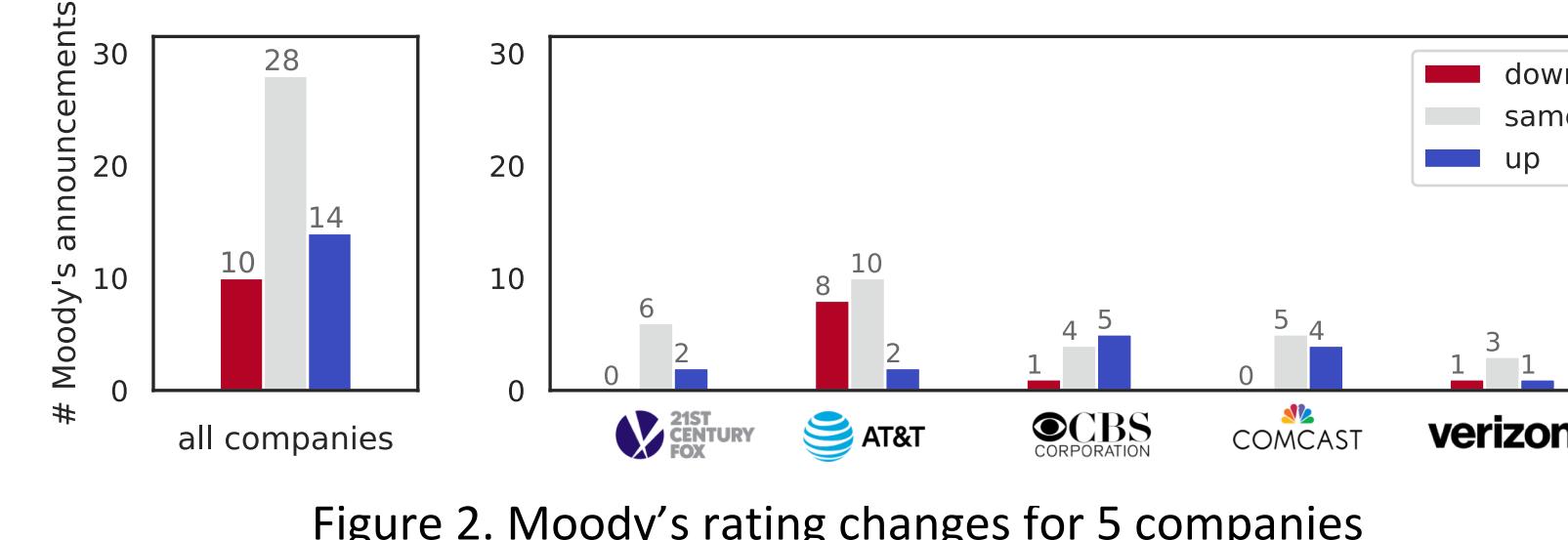


Figure 2. Moody's rating changes for 5 companies

Exploratory Data Analysis

- Moody's ratings
- SEC filings:
 - categorical and sentiment
- The New York Times news:
 - sentiment

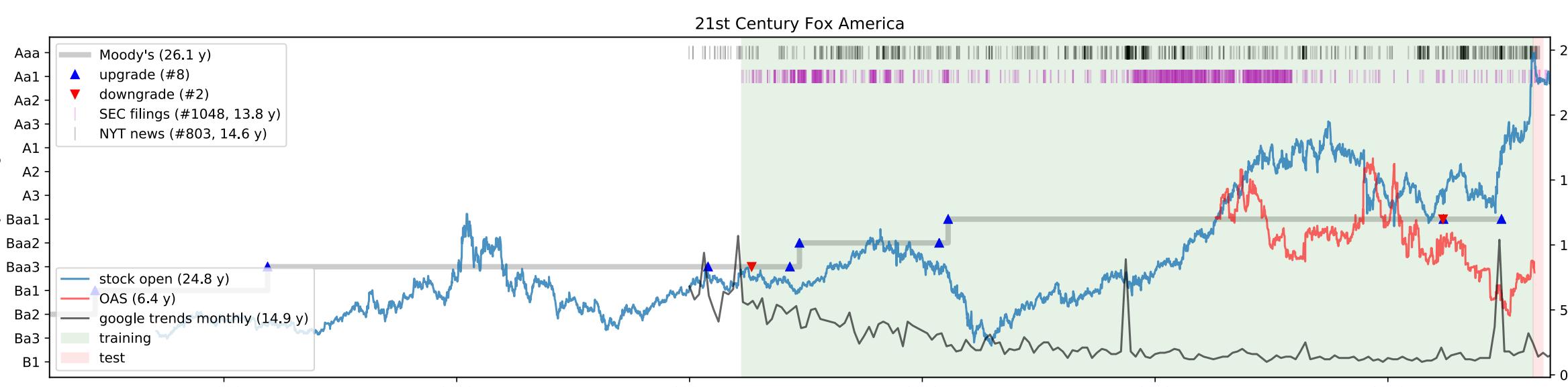


Figure 3. Data for 21st Century Fox: Moody's, SEC, NYT, OAS and Google trends.

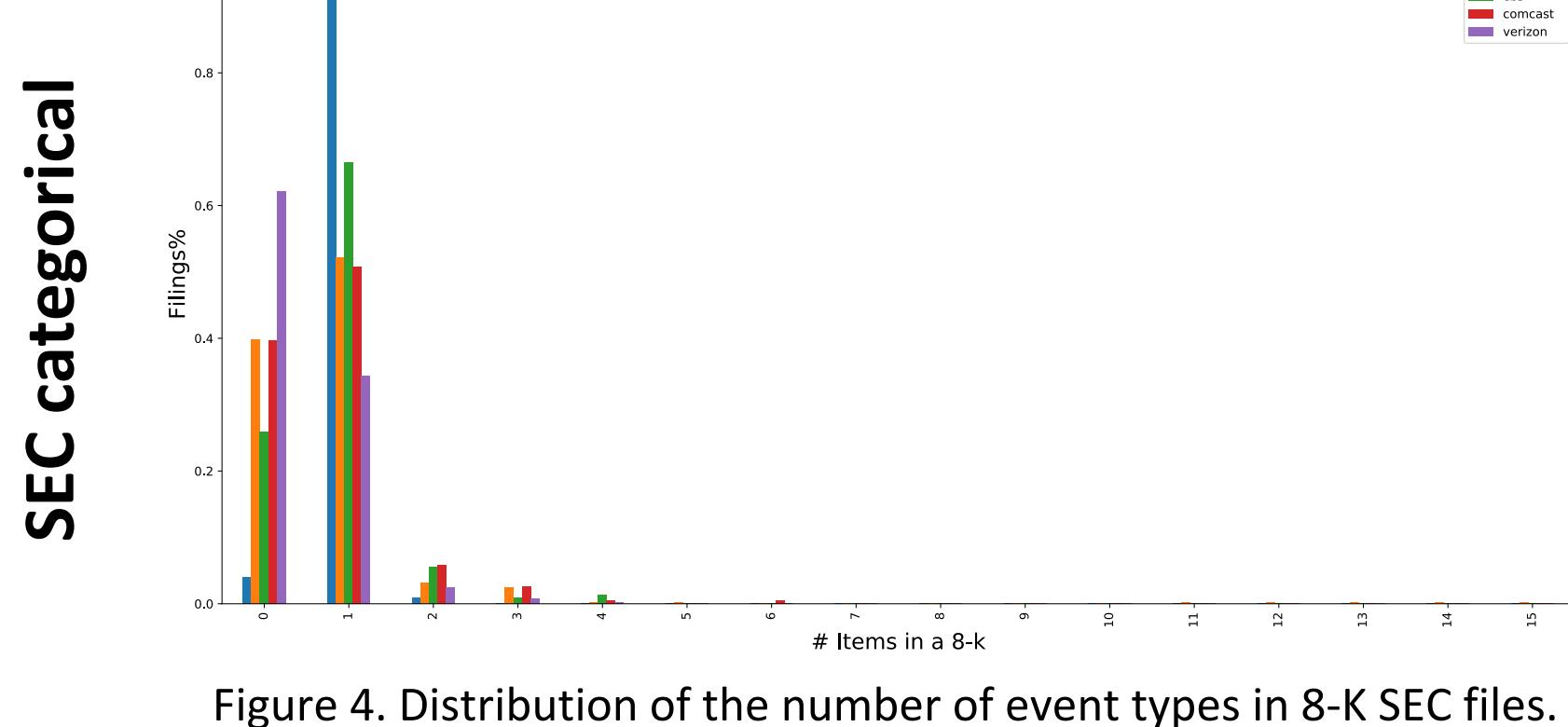


Figure 4. Distribution of the number of event types in 8-K SEC files.

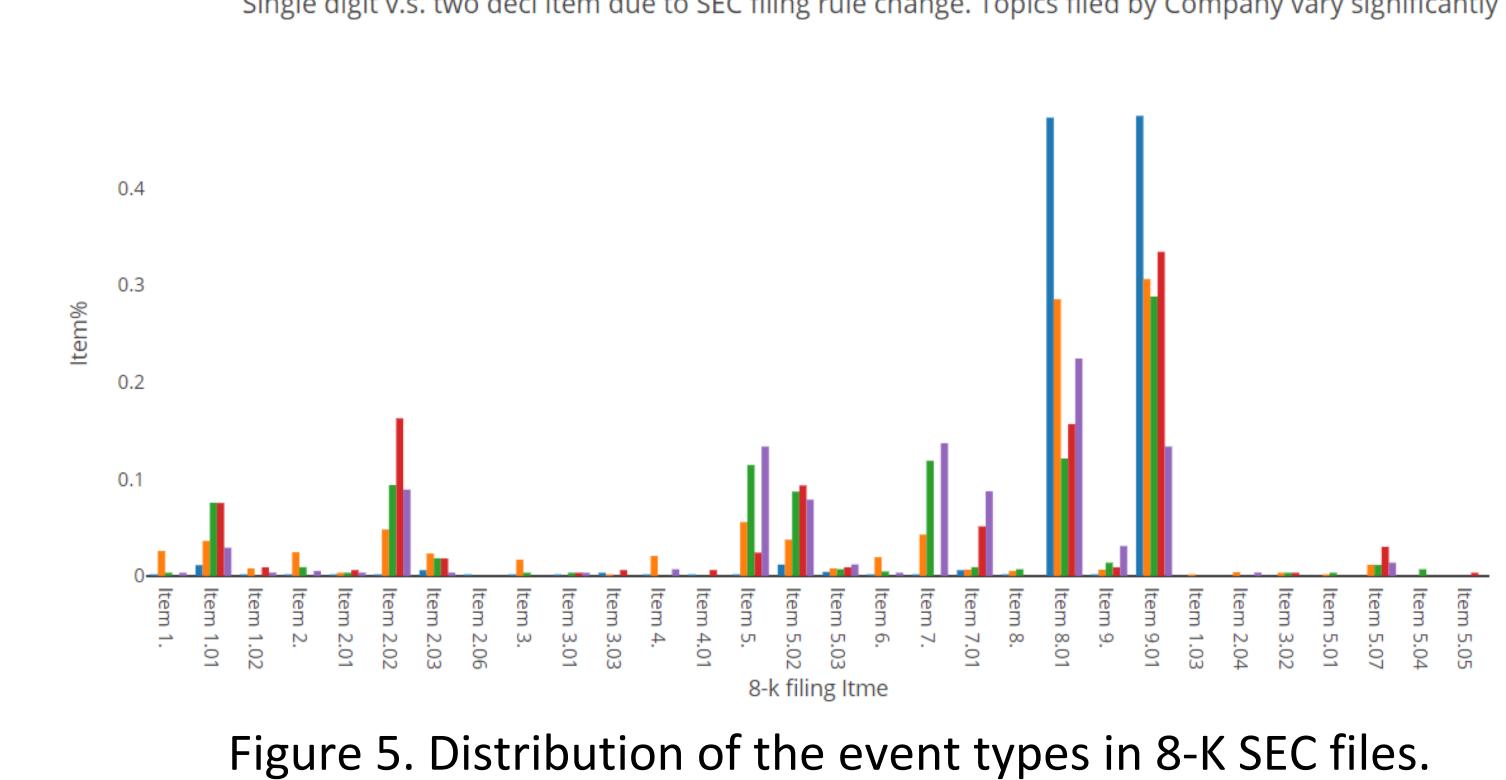


Figure 5. Distribution of the event types in 8-K SEC files.

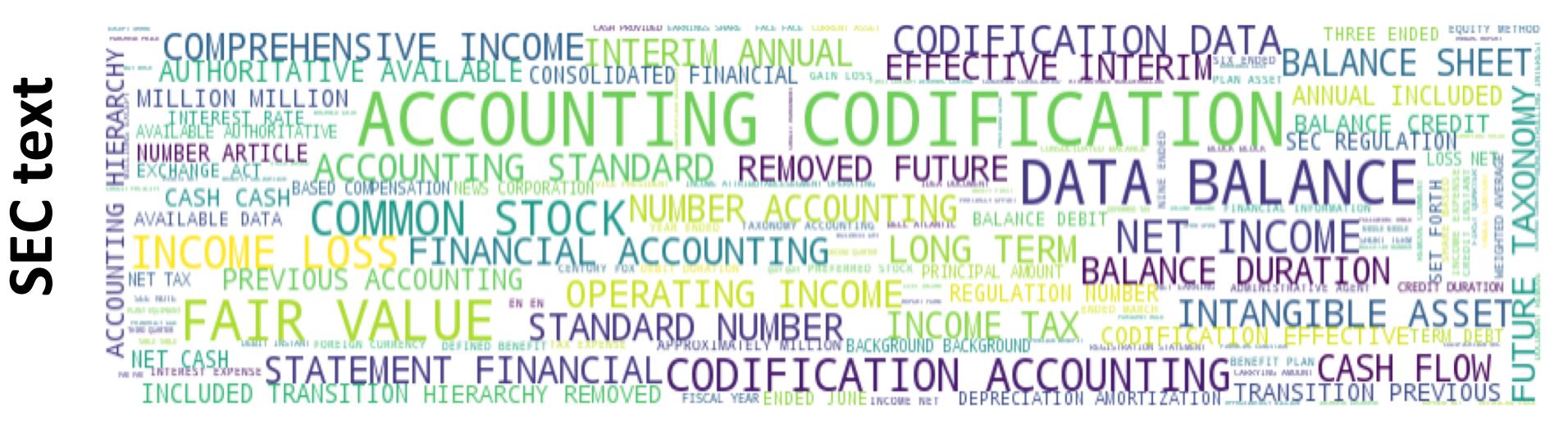


Figure 6. Word cloud on all SEC filings for 21st Century Fox.

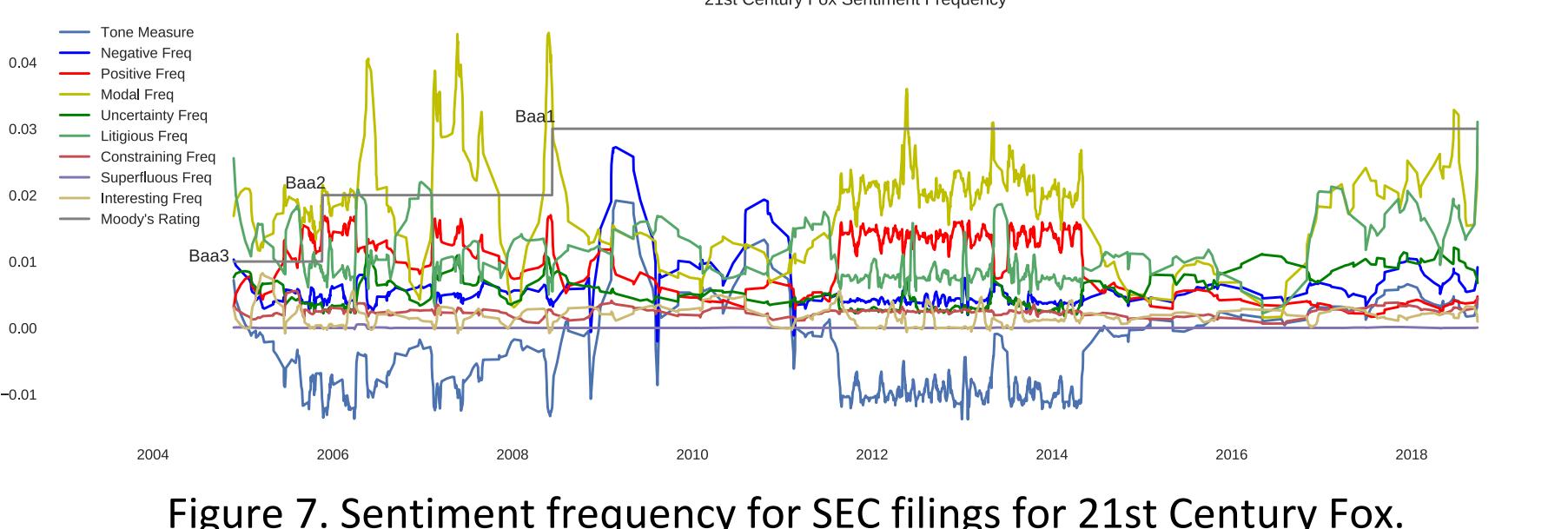


Figure 7. Sentiment frequency for SEC filings for 21st Century Fox.

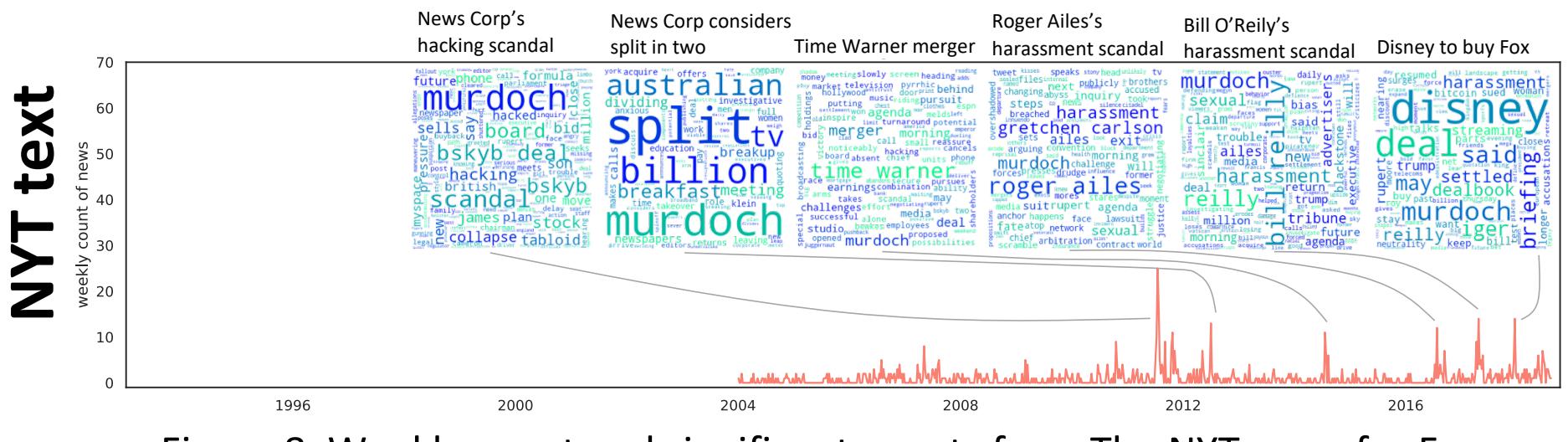


Figure 8. Weekly count and significant events from The NYT news for Fox.

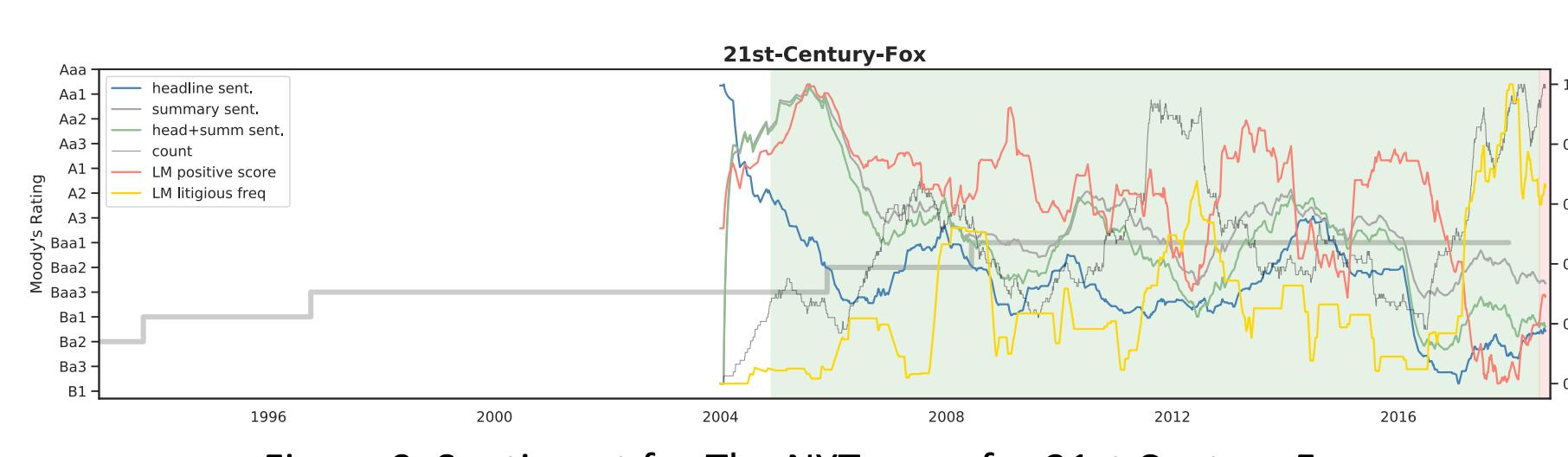


Figure 9. Sentiment for The NYT news for 21st Century Fox.

Results

We employed classification models based on two approaches:

1. Pairing each SEC filing with Moody's rating at the time: 2619 data points.
 - Random Forest model using sentiment scores of 8-K files as input.
 - Bag-of-words model using event text as input.
 - Long short-term memory network (LSTM) using event text as input.
 - Wide and deep network using sentiment scores and event text as input.

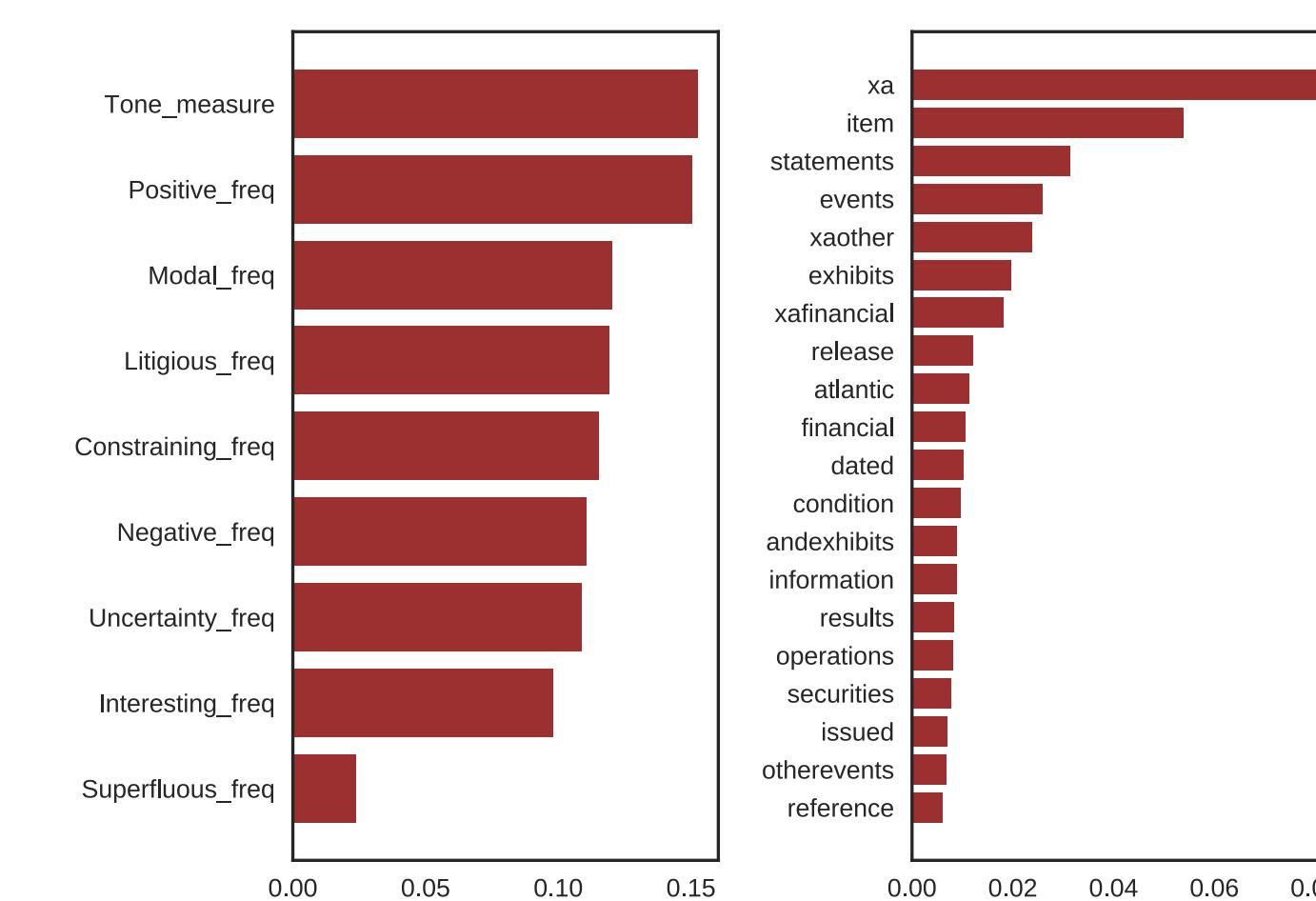


Figure 10. Top importance features
Random Forest (left) and bag-of-words (right).

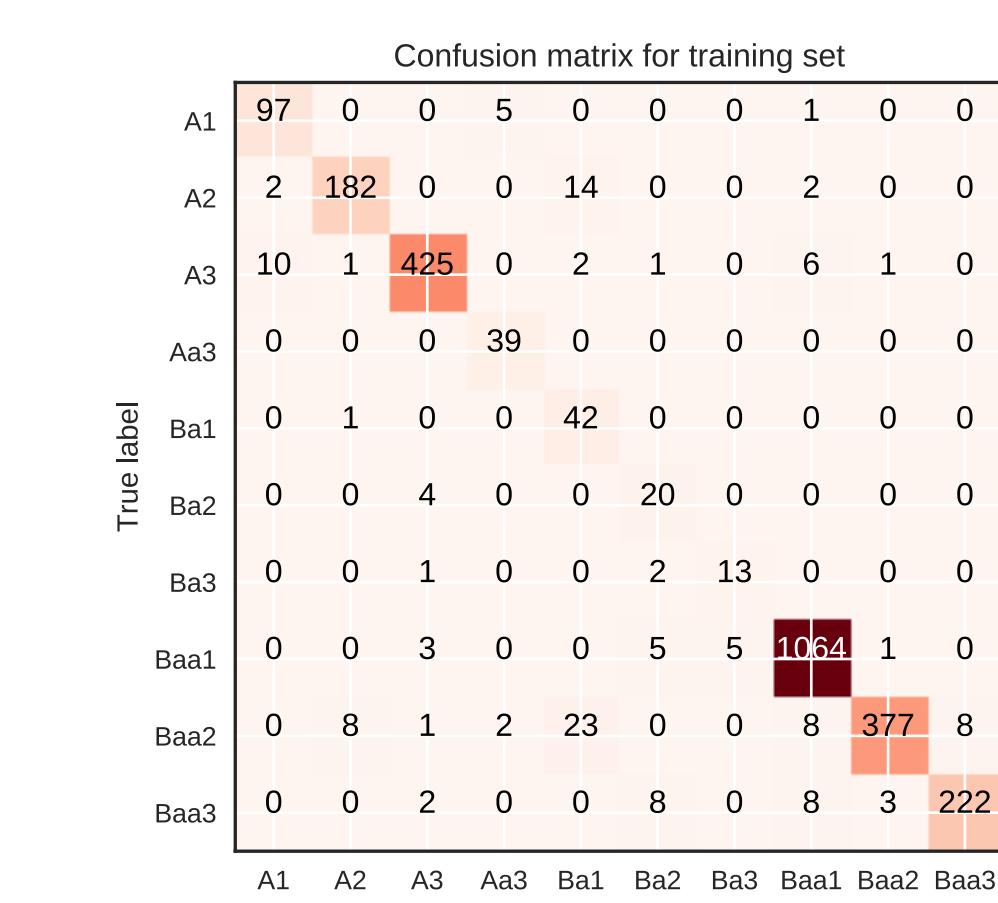
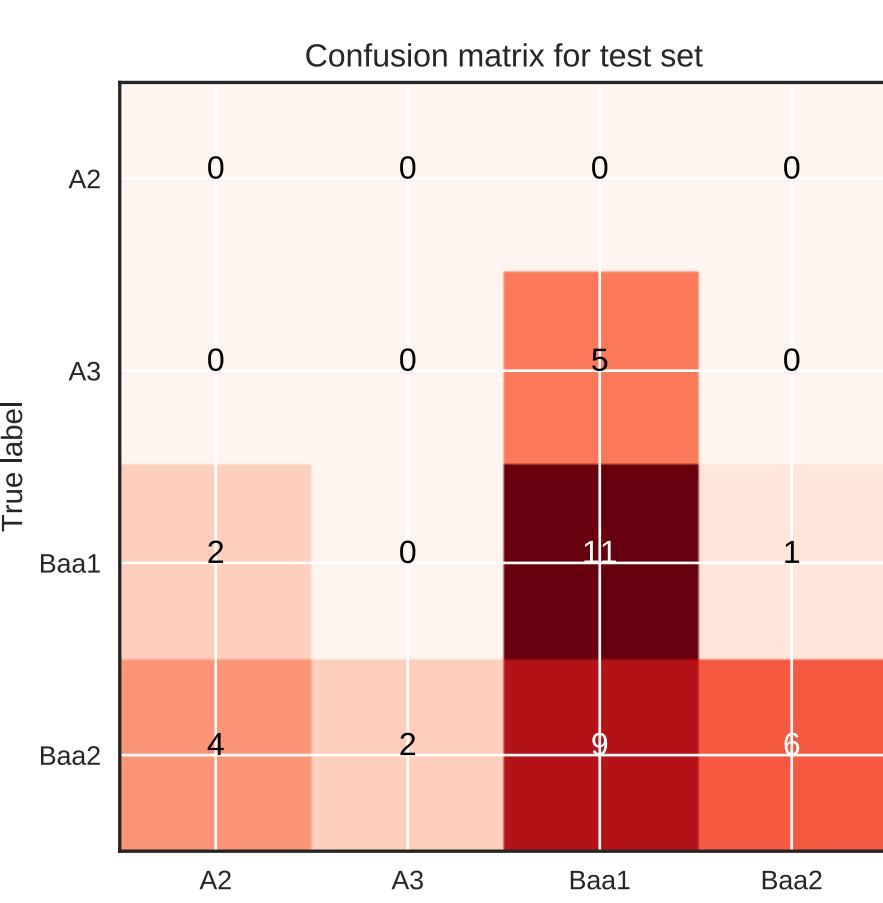


Figure 11. Confusion matrices for Random Forest
training (left) and test (right).



2. Paring Moody's announcements with time series data from the preceding 6 months.

- Inputs: SEC categorical event type, SEC sentiment, NYT sentiment.
- Used oversampling (SMOTE) to balance classes.
- Training set cross-validation accuracy: 61%.
- Our model assumes a rating change and predicts its type for each day in July-August.

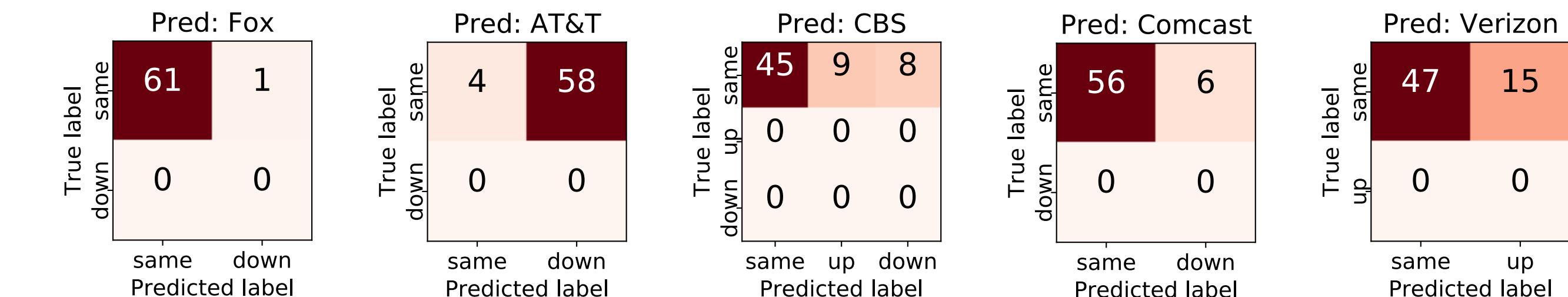


Figure 12. Confusion matrices for July-August 2018 prediction using XGBoost model.

Conclusions

Modeling of rare events is extremely challenging (if not futile), especially with limited unstructured data. We recommend to get more data (Moody's ratings for more companies) and perform time series classification with a deep model.

Acknowledgments

We would like to express our great appreciation to Professor Smaranda Muresan as well as our mentor Eunice Hameyie from Vanguard, who provided insights and NLP expertise to this project.