DEEP LEARNING APPLICATIONS IN AUDIT DECISION MAKING

Dissertation Defense

Ting Sun

Dissertation Committee
Chair: Dr. Miklos A. Vasarhelyi
Dr. Alexander Kogan
Dr. Helen Brown-Liburd
Dr. Rajendra P. Srivastava

April 16, 2018

Outline

- **Introduction**

- ***Essay One*: The Incremental Informativeness of Management Sentiment in Conference Calls for the Prediction of Internal Control Material Weaknesses**

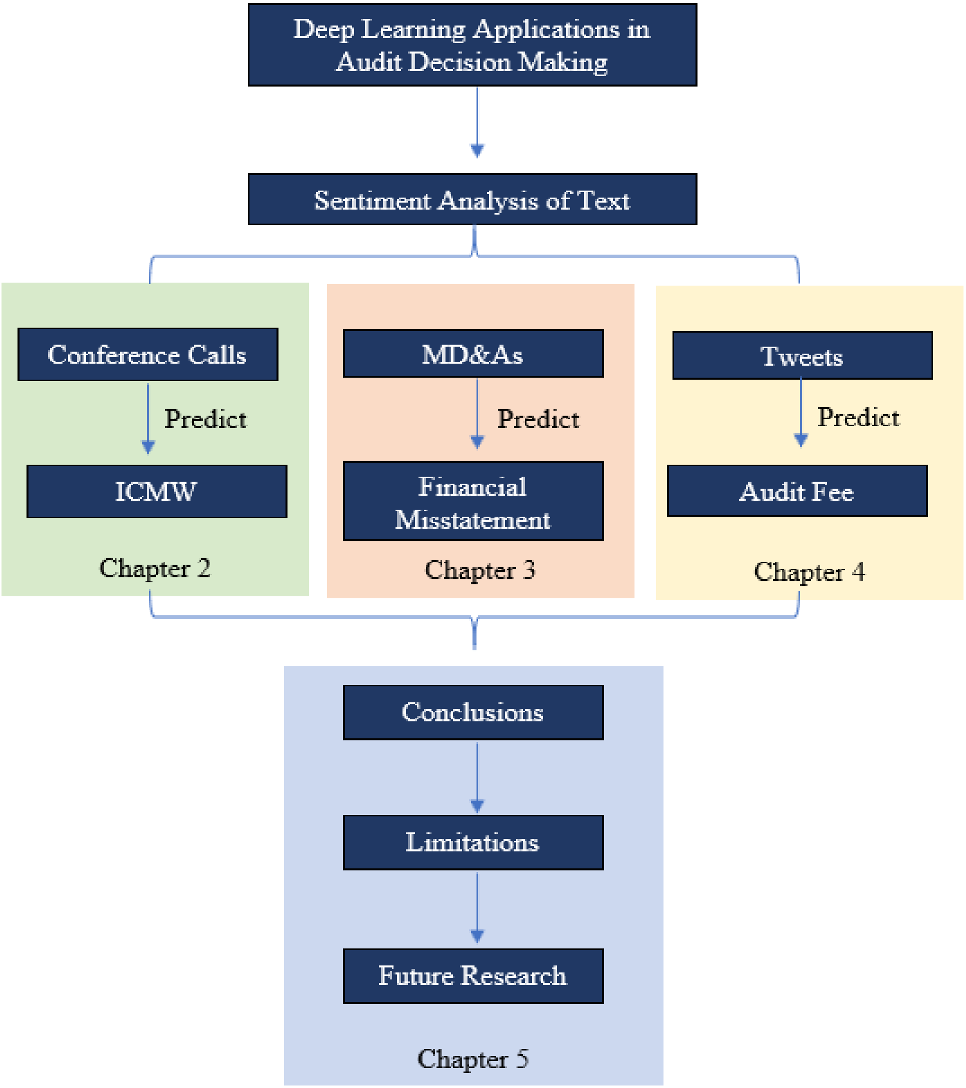
- ***Essay Two*: The Performance of Sentiment Features of MD&As for Financial Misstatements Prediction: A Comparison of Deep Learning and Bag of Words Approaches**

- ***Essay Three*: Predicting Audit Fees with Twitter: Do the 140 Characters reveal a company's audit risk?**

- **Conclusion, Limitation, and Future Research**

Introduction

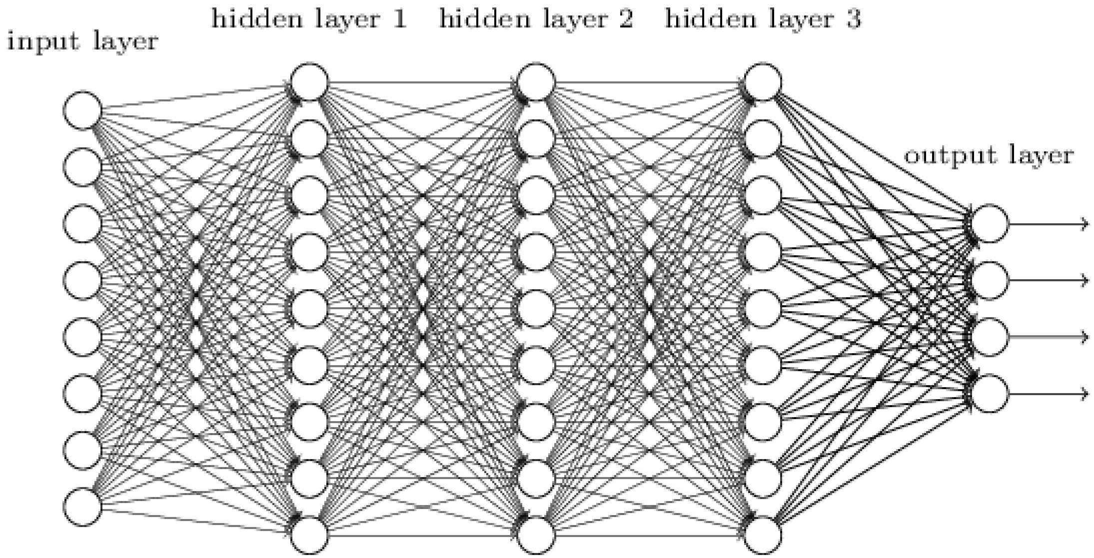
- **Motivation**
  - Deep Learning has been widely applied to computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation and etc. However, the application of deep learning in auditing is just emerging.
  - Big Four accounting firms are exploring the value of deep learning for auditing (e.g., KPMG and Deloitte), but they have not disclosed the details about their experiences.
  - Limited research has demonstrated the use of big data as additional audit evidence or integrated Deep Learning into an auditor's decision-making process
- **Objective**
  - Demonstrate the application of Deep Learning in auditing by using deep learning-based sentiment analysis of three types of textual data to help auditors assess risks.
  - Examine the effectiveness of deep learning-based approach of sentiment analysis as compared to traditional approaches


Deep Learning Applications in Audit Decision Making
Sentiment Analysis of Text
Conference Calls
Predict
ICMW
Chapter 2
MD&As
Predict
Financial Misstatement
Chapter 3
Tweets
Predict
Audit Fee
Chapter 4
Conclusions
Limitations
Future Research
Chapter 5

The Road Map

❖ Deep Learning

Use deep neural networks to extract high-level and abstract features from raw data by building multiple layers of representations that are expressed in terms of other, simpler representations (Goodfellow et al. 2016).



Multilayer Perceptrons (MLP)

- ❖ **The superiority of Deep Learning**
- ❖ **Why does Deep Learning perform well relative to other approaches**
  - Integration of feature extraction within the training process
  - Collection of very large data sets
  - Technology development
- ❖ **The need of Deep Learning for audit decision making**
  - Big data as supplementary audit evidence
  - The difficulties in big data analysis
  - The need of audit procedure automation

Essay One (Chapter 2)

The Incremental Informativeness of Management Sentiment in Conference Calls for the Prediction of Internal Control Material Weaknesses (ICMW)

❖Motivation

- ➢ The quality of ICFR audit is unsatisfactory
- ➢ Conference calls contain incremental information beyond mandated disclosures for the situation of the company (Allee and Deangelis, 2015; Sedor, 2002)
- ➢ The effectiveness of ICFR concerns both investors and managers
  - Investors: perceives higher information asymmetry, lower transparency, higher risk premium, lower sustainability of earnings etc. (Lopez, Vandervelde and Wu, 2009). The market negatively reacts to ICMW (reduced share prices and shareholder dissatisfaction) (Hammersley, Myers, and Shakespeare, 2008; Ye and Krishnan, 2008).
  - Managers: are hold accountable for ICMW (increased turnover, reduced compensation, etc.) (Johnstone, Li, and Rupley, 2011; Hoitash, Hoitash, and Johnstone, 2012)

❖Objective

- ➢ Examine the relationship between sentiment features of conference call transcripts and ICMWs;
- ➢ Investigate whether the sentiment features contain incremental information for the prediction of ICMW

Prior research

- **Internal control over financial reporting**
  - traditional firm-level fundamentals: size, age, financial performance, business complexity, growing speed, restructuring experiences (Doyle, et al., 2007a; Ashbaugh-Skaife, et al., 2007), accruals (Doyle, et al., 2007b)
- **Sentiment features of conference calls**
  - stock trading volume and return variance (Frankel, Johnson, and Skinner, 1999; Price, et al., 2012).
  - future performance ,analyst responses (i.e., Mayew and Venkatachalam, 2012; Druz, et al., 2015; Davis, et al., 2015).
  - Financial misstatement (Hobson, Mayew, and Venkatachalam, 2012; Larker and Zakolyukina,2012; Burgoon et al. 2016)
- **Investors' concern of ICMW**
  - investors perceive higher information asymmetry, lower financial statement transparency, higher risk premium, lower sustainability of earnings, and lower earnings predictability (Lopez, Vandervelde and Wu 2009).
  - the market negatively reacts to the disclosure of internal control weakness, in terms of reduced share prices (Hammersley, Myers, and Shakespeare 2008) and higher cost of capital (Ashbaugh-Skaife et al. 2009).
- **Managers' concern of ICMW**
  - it is primarily the management's responsibility to design and maintain the internal control system (PCAOB, 2007)
  - Johnstone, Li, and Rupley (2011) : an adverse ICFR opinion leads to increased management turnover
  - Hoitash, Hoitash, and Johnstone (2012) : ICMW disclosures are negatively related to the change in CFO total compensation, bonus compensation, and equity compensation, especially for firms with stronger governance oversight
- **Social psychology research**
  - Leakage hypothesis (Ekman and Friesen, 1969), the act of deception will make a single person feel guilty, stressful, and fear of detection.
  - DePaulo, Rosenthal, Rosenkrantz, and Green (1982) and Kraut (1980) suggest that a person may experience relatively heightened cognitive processing when telling a lie than telling the truth.

Hypotheses

- H1: The sentiment features of conference calls are significantly associated with the likelihood of internal control material weaknesses.
- H2: The explanatory ability of the model that incorporates sentiment features of conference calls along with major financial determinants is superior to that of the model that merely uses the financial determinants.
- H3: The predictive ability of the model that incorporates sentiment features of conference calls along with major financial determinants is superior to that of the model that merely uses the financial determinants.

Data

Source: SeekiNF

Sample selection procedure

Initial conference call transcript samples from Seek iNF	6379
Less: Missing fiscal year or CIK information	(1582)
use the conference call in the last quarter if a company has multiple conference calls	
Remaining:	2408
Less:	
Missing internal control information	(15)
Missing Compustat data	(731)
Missing Audit Analytics data	(11)
Final sample	1651

corresponding to fiscal year from 2004 to 2014, among which, 189 firm-years are related to ICMW.

Tool:

Alchemy Language API, a deep learning based text analysis cloud services of IBM Watson

Sentiment features:

- Overall sentiment score:

measures the sentiment strength of the entire document, ranged from -1 to 1

- Joy score:

the probability that an emotion of joy is implied by any part (e.g., sentence, paragraph) of the text.

Logistic regression

The Baseline Model

$$ICMW = \beta_0 + \beta_1 Marketvalue + \beta_2 Aggregateloss + \beta_3 Zscore + \beta_4 Segments + \beta_5 Foreign + \beta_6 Inventory + \beta_7 Restructure + \beta_8 Acquisition + \beta_9 Resign + \beta_{10} Big4 + \beta_{11} Litigation + \sum IndustryFE + \varepsilon$$

The Sentiment Model

$$ICMW = \beta_0 + \alpha_1 Sentiment + \alpha_2 Joy + \beta_1 Marketvalue + \beta_2 Aggregateloss + \beta_3 Zscore + \beta_4 Segments + \beta_5 Foreign + \beta_6 Inventory + \beta_7 Restructure + \beta_8 Acquisition + \beta_9 Resign + \beta_{10} Big4 + \beta_{11} Litigation + \sum IndustryFE + \varepsilon$$

Results of primary analysis

- **Logistic regression of the determinants of ICMW**

	Predicted sign	Estimate coefficients of group A		Estimate coefficients of group B	
		Baseline model A	Sentiment model A	Baseline model B	Sentiment model B
Intercept	+/-	-1.6784**	-1.6211*	-2.1248*	-2.0097*
Sentiment	-		0.6243		0.2979
Joy	-		-1.3762***		-1.5264**
Marketvalue	-	-0.2551***	-0.2495***	-0.2591***	-0.2537***
Aggregateloss	+	-0.3105	-0.3137	-0.1360	-0.1379
Zscore	-	-0.0040	-0.0008	-0.0047	-0.0035
Segments	+	0.3424***	0.3547***	0.2512	0.2559
Foreign	+	0.3927	0.4047	0.5328	0.5575
Inventory	+	0.1535	0.1585	0.5008	0.5555
Growth	+			-0.0193	-0.0286
Restructure	+	-0.1366	-0.1330	-0.1187	-0.1420
Acquisition	+	0.0601	0.0935	0.2901	0.3429
Resign	+	2.2631***	2.2322***	2.3476***	2.3188***
Big4	-	-0.1079	-0.0984	-0.0007	0.0260
Litigation	+	0.1908	0.2211	0.2760	0.3119
Industry indicator		Included	Included	Included	Included
# total observations		1651	1651	1228	1228
Likelihood ratio, χ^2 (p-value)		89.85 (0.0001)	98.17 (0.0001)	63.42 (0.0001)	70.44 (0.0001)
Pseudo R^2		0.0757	0.0827	0.0785	0.0872
Likelihood-ratio test: Likelihood ratio (p-value)		8.32** (0.0156)		7.02** (0.0300)	

10-Fold Cross Validation Result

		AUC	Overall Accuracy	False Positive Rate	False Negative Rate
Logistic Regression	Baseline model	0.6931	0.7288	0.2426	0.4621
	Sentiment model	0.6955	0.6915	0.2994	0.3690
Random Forest	Baseline Model	0.7228	0.7256	0.2545	0.4069
	Sentiment Model	0.7274	0.8357	0.1033	0.5724
ANN	Baseline model	0.6726	0.7171	0.2530	0.4828
	Sentiment model	0.6838	0.7081	0.2664	0.4621

Additional analysis: the Number of ICMW

Multinomial Logistic Regression of the Probability of ICMW by the

		Oneweak vs. Noweak		Moreweak vs. Noweak		(3)-(1)
Independent Variable	Expected Sign	Coefficient	P-value	Coefficient	P-value	
		(1)	(2)	(3)	(4)	(5)
Intercept	+/-	-17.3622	0.996	-31.5101	0.986	-14.1479
Sentiment	-	-0.8239	0.421	1.8955*	0.061	2.7194*
Joy	-	-0.2783	0.679	-2.4116***	0.001	-2.1333**
Marketvalue	-	-0.1967**	0.016	-0.2944***	0.001	-0.0977
Aggregateloss	+	-0.5376	0.118	-0.1183	0.695	0.4193
Zscore	-	-0.0116	0.122	0.0158	0.111	0.0274**
Segments	+	0.3123*	0.097	0.3994**	0.023	0.0871
Foreign	+	-0.4587	0.551	14.7832	0.991	15.2419
Inventory	+	0.1788	0.883	0.3791	0.746	0.2003
Restructure	+	0.0677	0.790	-0.2939	0.240	-0.3616
Acquisition	+	0.0342	0.893	0.1367	0.559	0.1025
Resign	+	1.6018***	0.004	2.6901***	0.001	1.0883*
Big4	-	-0.2373	0.397	0.0069	0.979	0.2442
Litigation	+	-0.1963	0.620	0.5673	0.141	0.7636
Industry indicator variables		Included		Included		
Number of total observations		1651 (107 Moreweak, 85 oneweak, and 1459 Noweak)				
Likelihood ratio, χ^2		149.07*** (0.0001)				

The Persistency of ICMW

Logistic Regression of the Probability of ICMW by the

		Dependent variable: Firstyrmw		Dependent variable: Persistmw	
Independent variable	Expected Sign	Coefficient	P-value	Coefficient	P-value
		(1)	(2)	(3)	(4)
Intercept	+/-	-2.2647***	0.007	-2.6044**	0.027
Sentiment	-	-0.5544	0.722	0.3843	0.719
Joy	-	-0.6643	0.530	-1.7097***	0.009
Marketvalue	-	-0.3090***	0.009	-0.2614***	0.001
Aggregateloss	+	0.6380	0.123	-0.5148	0.150
Zscore	-	0.0086	0.575	0.0004	0.964
Segments	+	0.2556	0.375	0.3628**	0.050
Foreign	+	0.1887	0.441	0.3161	0.765
Inventory	+	-0.3979	0.835	0.9401	0.395
Restructure	+	0.0986	0.802	-0.1742	0.512
Acquisition	+	0.6709*	0.082	0.0506	0.842
Resign	+	0.4467	0.705	2.6591***	0.001
Big4	-	-0.3437	0.398	-0.2630	0.339
Litigation	+	0.5390	0.348	0.5584	0.176
Industry indicator variables		Included		Included	
Number of total observations		1499		1554	
Number of observations with no MW		1462		1462	
Number of MW observations		37		92	
Likelihood ratio, χ^2		29.24		92.11***	
(p-value)		0.1726		0.0001	

Conclusion

- ❖ The sentiment features, especially the joy score, improves both explanatory and predictive ability of the model for ICMW prediction
- ❖ Compared to companies with one material weakness, companies with more than one material weakness has higher overall sentiment score, lower joy score, and more likely to have auditor resignation.
- ❖ The Joy score determinates persistent ICMW effectively

Contribution

- ❖ Documents that the sentiment and emotion of conference calls provide additional information for the detection of ICMW. Auditors can use conference call transcripts as supplementary evidence source to support ICFR audit

Essay Two

The Performance of Sentiment Feature of MD&As for Financial Misstatements Prediction: A Comparison of Deep Learning and Bag of Words Approaches

❖ Motivation

- Quantitative information disclosed by financial statements may contain misleading information that does not fairly present the financial position and the performance of the company.
- Other business communication documents provide incremental qualitative evidence of sentiment and other linguistic features that can be used to uncover financial misstatements (e.g., Larcker and Zakolyukina, 2012; Lee, Lusk, and Halperin, 2014; Czerney, Schmidt, and Thompson, 2014).
- Researchers argue that "bag of words" approach is too simplistic to obtain the accurate meaning of the text (Salton and McGill, 1983).

❖ Research Questions

(1) Do sentiment features in MD&As add information for financial misreporting prediction?

(2) If they do, are they effective only for fraud prediction or for misstatement including both fraud and error?

(3) How effective the model using deep learning based sentiment features is as compared to the model using sentiment feature obtained by bag of words approach?

Prior Literature

❖ Financial misstatement prediction

- Dechow et al. (2011) : several measures of accrual quality, gross profit, “soft” assets, and other financial factors are highly associated with misstatements.
- Cecchini et al. (2010) : based on Support Vector Machines (SVM) to detect frauds with financial data. The power of the learning machine is increased to be able to correctly labeled 80% of the fraudulent companies.
- Perols,et al., (2017): developed models to detect financial statement fraud using a dataset with 51 fraud firms, 15,934 non-fraud firm-years, and 109 explanatory variables from prior research.
- Larcker and Zakolyukina (2012): A total of 29,663 CC transcripts were analyzed to extract sentiment attributes to detect financial misstatements.
- Perols (2011): compares the performance of six popular statistical and machine learning models in detecting financial statement fraud .The results show that logistic regression and support vector machines perform well relative to an artificial neural network, bagging, C4.5, and stacking.

❖ Sentiment features of MD&A and financial misstatement

- Churyk, Lee, and Clinton 2009: more words, less terms with positive emotions like optimism and energy, and more terms with negative emotion like anxiety
- Humpherys et al. (2011): active language than those of non-fraudulent firms.
- Loughran and McDonald (2011): The appearance of a list of 13 problematic

Method: Data

- 31,466 MD&As of 10-K filings for fiscal years from 2006 to 2015 are processed using deep learning and "bag of words" based sentiment analysis separately.
- deep learning approach: Sentiment_DL and Joy
- bag of words approach: Sentiment_TM
- 82 other predictors for financial frauds and misstatements: based on the research (Perols, Bowen, Zimmermann, and Samba, 2017; Dechow et al. , 2011; Perols, 2011; Cecchini et al. , 2010; Beneish, 1999; Huang et al. , 2012; Churyk et al. , 2009).

RUTGERS

The Structure of Model Sets

Panel A: Misstatement Prediction				
		Baseline model	Model 1	Model 2
DV		MISSTATEMENT	MISSTATEMENT	MISSTATEMENT
IV	Sentiment Predictors	N/A	SENTIMENT_DL JOY	SENTIMENT_TM
	Other predictors	82 variables related to misstatement	82 variables related to misstatement	82 variables related to misstatement
Panel B: Fraud Prediction				
		Baseline model	Model 1	Model 2
DV		FRAUD	FRAUD	FRAUD
IV	Sentiment Predictors	N/A	SENTIMENT_DL JOY	SENTIMENT_TM
	Other predictors	82 variables related to misstatement	82 variables related to misstatement	82 variables related to misstatement
Panel C: Error Prediction				
		Baseline model	Model 1	Model 2
DV		ERROR	ERROR	ERROR
IV	Sentiment Predictors	N/A	SENTIMENT_DL JOY	SENTIMENT_TM
	Other predictors	82 variables related to misstatement	82 variables related to misstatement	82 variables related to misstatement

Classification Algorithms

5 machine learning algorithms:

- Random Forest
- Logistic Regression
- Naïve Bayes
- Deep Neural Network: three hidden layers(175,350,150)
- Traditional Neural Network: one hidden layer consisting of 100 neurons

With each algorithm, we analyze 9 models

Totally, **45** models are developed

Random Forest:
prediction results of 10-fold cross validation

	Baseline Model			Model 1 (deep learning)			Model 2 (bag of words)		
Metrics	(A) MIS	(B) FRAUD	(C) ERROR	(D) MIS	(E) FRAUD	(F) ERROR	(G) MIS	(H) FRAUD	(I) ERROR
Accuracy	66.69%	75.36%	61.91%	66.88%	77.28%	61.77%	65.23%	75.76%	63.73%
Type one error rate	33.31%	24.64%	38.06%	33.11%	22.72%	38.23%	34.77%	24.24%	36.27%
Type two error rate	33.33%	24.04%	38.53%	33.33%	22.12%	37.61%	34.89%	24.02%	36.70%
AUC	0.7232	0.8288	0.6673	0.7325	0.8524	0.6683	0.7224	0.8506	0.6786

Predictor Importance: Random Forest

Top 10 important predictors of fraud detection models

Baseline Model		Model 1		Model 2	
Predictor	**Scaled Importance**	**Predictor**	**Scaled Importance**	**Predictor**	**Scaled Importance**
SOFT	1	SOFT	1	SOFT	1
RECAT	0.7189	RECAT	0.9222	RECAT	0.6731
PPENTAT	0.4805	PPENTAT	0.5454	PPENTAT	0.5061
PENSION	0.3816	MVE	0.4480	SENTIMENT_TM	0.4223
MVE	0.3004	SENTIMENT_DL	0.4160	MVE	0.3474
LEASE	0.2775	FAAT	0.3788	PENSION	0.3313
AT	0.2557	PENSION	0.3613	AT	0.2906
FAAT	0.2512	AT	0.3042	LEASE	0.2458
LTXINT	0.2198	SALEAT	0.2372	FAAT	0.2447
CLEASE	0.1795	LTXINT	0.2152	SALEAT	0.2089

Answers to RQs:

(1)Do sentiment features add information for financial misreporting prediction?

Yes

(2) If they do, are they effective only for fraud prediction or for misstatement including both fraud and error?

Fraud prediction

intentional misstatements and unintentional misstatements are two different types of events and that distinction between the two is important to increase the power of the tests (Kim, Baik, and Cho, 2016).

(3) How effective the model using deep learning based sentiment features is as compared to the model using sentiment feature obtained by bag of words approach?

Significant improvement of effectiveness in terms of Accuracy, AUC, false positive rates

Conclusion

Considering its effectiveness and efficiency, deep learning-based sentiment analysis is a promising technique for audit analytics

Contribution

- Provide evidence for the informativeness of the sentiment features of MD&A for Financial fraud detection
- Demonstrate the superiority of Deep Learning to Bag-of-words for the task of fraud prediction

Essay Three

Predicting Audit Fee with Twitter: Do the 140 Characters Reveal the Company's Audit Risk?

❖Motivation

- Auditors devote substantial time to understand as much as possible the company and its management to mitigate the audit risk of engaging with or continuing to serve a client.
- The efficiency of social media makes more information regarding the company's operational and financial situation available to us at high speed.
- Little research explores the powerful insights provided by social media and how auditors could leverage them to support risk assessment in planning (Debreceny, 2015).

❖Research Question

- Does Tweets provide audit risk information that influences audit pricing?

❖Objective

- Investigate whether Tweets can serve as a predictor of the audit fee, which reflects the audit risk perceived by the auditor
- Demonstrate the effectiveness of deep learning based sentiment analysis of social media data for audit fee decision support

Prior Literature

❖ Audit Fees

- Audit fee model (Simunic 1980)
- Risky clients are likely to pay high audit fees (O'Keefe et al. 1994; Lyon and Maher 2005; Venkataraman, Weber, and Willenborg, 2008)
- Audit fee is negotiated and determined in the engagement letter. Once engaged, the negotiated fee will not change except in response to unexpected significant changes (Hackenbrack, Jenkins, and Pevzner, 2014)

❖ Twitter

- company's adoption of Twitter and its market effect (Blankespoor, Miller, and White 2013; Prokofieva 2015; Lee, Hutton and Shu 2015; Debreceny, Rahman, and Wang 2016)
- Sentiment of tweets (Bonner, 2008; Mian and Sankaraguruswamy, 2012; Prokofieva, 2015; Debreceny, Rahman, and Wang, 2016)
- Retweets (Wu and Shen, 2015; Blankespoor et al. 2014; Prokofieva, 2015)
  - Measure the popularity of the disseminated information. It is an indicator that Twitter disclosure has been read and disseminated further.

Hypotheses

H1: The audit fee of a company is positively associated with the negativity of the Tweets mentioned the company.

H2: The positive association between the audit fee and the negativity of Tweets is stronger for companies with more Retweets.

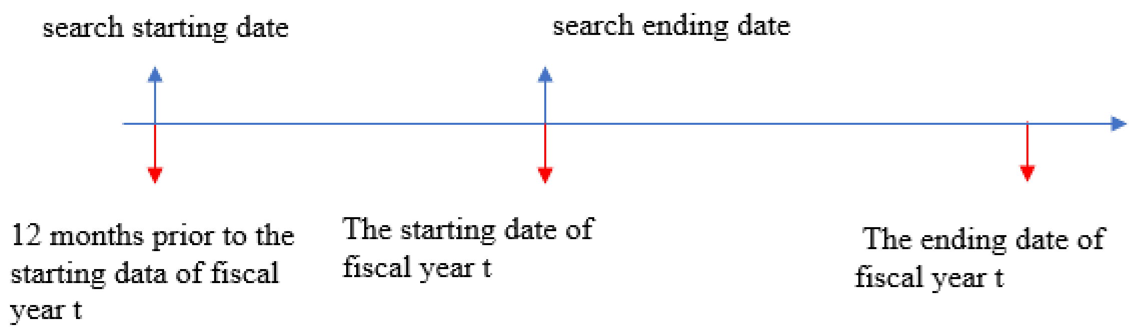
H3: The information of Tweets improves the predictive ability of the audit fee model

Method

❖ **Tool: IBM® Twitter Insight:**

- this tool includes APIs that allow searches for Twitter content based on keywords, timeframes, and other query parameters, provides real-time analysis of Tweets, and returns Tweets with related properties, such as the number of retweet and the overall sentiment (e.g., positive, negative, ambivalent, or neutral) .

❖ Sample



	Number of Observations
U.S. listed companies in 2015	6130
Less: financial, insurance, and real estate firms	(235)
Less: observations with financial variable data missing in Compustat or Compustat Segments	(1,869)
Less: observations with audit data missing in AuditAnalytics	(1,215)
Less: observations with missing Twitter data	(479)
Final sample	2332

Audit Fee Model

$$Lnauditfee = \beta_0 + \beta_1 Negativity + \beta_2 Retweets + \beta_3 Negativity * Retweets + \beta_4 Tweets + \beta_5 Roaearnings + \beta_6 Size + \beta_7 Invrec + \beta_8 Leverage + \beta_9 Currentratio + \beta_{10} BTM + \beta_{11} Growth + \beta_{12} Loss + \beta_{13} Segments + \beta_{14} Foreign + \beta_{15} Merger + \beta_{16} Special + \beta_{17} Firstyear + \beta_{18} Big4 + \beta_{19} IC + \beta_{20} GC + \sum IndustryFE + \varepsilon$$

Where:

- Negativity: the percentage of tweets with negative sentiment among all tweets mentioned the company minus the percentage of tweets with positive sentiment among all tweets mentioned the company
- Retweets: the maximum number of retweets for each tweet mentioned the company.
- Tweets: the count of all tweets mentioned the company

RUTGERS

Results: full sample

		Model 1		Model 2		Model 3		Model 4	
Variable	Expected Sign	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	+/-	9.8690***	0.001	9.8707***	0.001	9.8701***	0.001	9.8577***	0.001
Negativity	+	0.0411	0.573	0.0453	0.534	0.0745	0.314	0.0790	0.285
Retweets	+	$8.85e^{-8}$	0.834	$8.61e^{-8}$	0.839	$1.39e^{-6}$**	0.036	$1.41e^{-6}$**	0.033
Negativity×Retweets	+					$7.14e^{-6}$**	0.011	$7.26e^{-6}$***	0.010
Tweets	+/-	-$7.24e^{-9}$	0.365	-$7.14e^{-9}$	0.371	-$7.96e^{-9}$	0.318	-$7.89e^{-9}$	0.323
Roaearnings	-	-0.0005***	0.001	-0.0005***	0.001	-0.0005***	0.001	-0.0005***	0.001
Size	+	0.4433***	0.001	0.4428***	0.001	0.4430***	0.001	0.4431***	0.001
Invrec	+	0.3891***	0.001	0.3887***	0.001	0.3894***	0.001	0.3855***	0.001
Leverage	+	0.1010***	0.001	0.1013**	0.001	0.1017***	0.001	0.1013***	0.001
Currentratio	-	-0.0012**	0.014	-0.0012**	0.015	-0.0012**	0.014	-0.0012**	0.015
BTM	-	-0.0005	0.461	-0.0006	0.433	-0.0005	0.469	-0.0006	0.439
Growth	-	-0.0009**	0.019	-0.0009**	0.020	-0.0009**	0.017	-0.0008**	0.021
Loss	+	0.1179***	0.001	0.1182***	0.001	0.1188***	0.000	0.1153***	0.001
Foreign	+	0.3588***	0.001	0.3593***	0.001	0.3591***	0.001	0.3586***	0.001
Merger	+	0.1272***	0.001	0.1280***	0.001	0.1272***	0.001	0.1267***	0.001
Special	+	0.1596***	0.001	0.1592***	0.001	0.1596***	0.001	0.1542***	0.001
Firstyear	-	-0.0782*	0.061			-0.0771*	0.064		
Resignation	+			-0.1769*	0.065			-0.1767*	0.065
Dismissal	-			0.0136	0.791			0.0203	0.694
GC	+	0.3067***	0.001	0.2900***	0.001	0.2892***	0.001	0.2840***	0.001
Big4	+	0.4669***	0.001	0.4669***	0.001	0.4665***	0.001	0.4745***	0.001
IC	+	0.0634***	0.001	0.0635***	0.001	0.0634***	0.001	0.0633***	0.001
Industry effect		Included		Included		Included		Included	
Observations		2332		2332		2332		2332	
Adjusted R^2		0.8612		0.8611		0.8615		0.8615	

RUTGERS

Going-concern Opinion

		GC companies		Non-GC companies	
Variable	Expected Sign	Coefficient	p-value	Coefficient	p-value
		(1)	(2)	(3)	(4)
Intercept	+/-	11.8946***	0.001	9.2993***	0.001
Negativity	+	-0.1755	0.387	0.1703**	0.032
Retweets	+	-7.28e^{-6}*	0.095	1.88e^{-6}***	0.004
Negativity× Retweets	+	-3.17e^{-5}	0.201	8.90e^{-6}***	0.001
Tweets	+/-	2.88e^{-9}	0.994	-1.18e^{-8}	0.121
Roaearnings	-	-0.0005***	0.002	0.0159***	0.001
Size	+	0.3330***	0.001	0.4680***	0.001
Invrec	+	0.1294	0.587	0.5596***	0.001
Leverage	+	0.0446	0.177	0.1813***	0.001
Currentratio	-	0.0021	0.906	-0.0008*	0.093
BTM	-	-0.0012	0.357	-0.0129***	0.000
Growth	-	-0.0006	0.180	-0.0045*	0.425
Loss	+	-0.3048	0.219	0.1511***	0.001
Foreign	+	0.4588***	0.009	0.3256***	0.001
Merger	+	0.0814	0.610	0.1104***	0.001
Special	+	0.0758	0.446	0.1502***	0.001
Resignation	+	-0.4833*	0.053	-0.1539	0.137
Dismissal	+	0.0022	0.986	0.0468	0.408
Big4	+	0.6886***	0.001	0.4318***	0.001
IC	+	0.0058	0.849	0.0960***	0.001
Industry effect		Included		Included	

Restatement Risk

		Restatement risk (Top Quintile)		Restatement risk (Middle Quintile)		Restatement risk (Bottom Quintile)	
Variable	Expected Sign	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
		(1)	(2)	(3)	(4)	(5)	(6)
Intercept	+/-	12.3178***	0.001	9.7059***	0.001	11.0437***	0.001
Negativity	+	0.3887	0.303	0.1976**	0.034	-0.1123	0.787
Retweets	+	-6.10e^{-6}	0.593	1.36e^{-6}*	0.058	-3.13e^{-6}	0.636
Negativity × Retweets	+	-6.10e^{-5}	0.391	6.13e^{-6}**	0.047	5.50e^{-6}	0.866
Tweets	+/-	4.49e^{-7}	0.385	-1.23e^{-8}	0.114	4.56e^{-8}	0.741
Roaearnings	-	-0.2972**	0.050	-0.3312***	0.001	-0.1006	0.727
Size	+	0.4234***	0.001	0.4814***	0.001	0.3070***	0.001
Invrec	+	0.9707**	0.050	0.5887***	0.001	1.3936*	0.058
Leverage	+	0.4758*	0.074	-0.0026	0.963	-0.0989	0.709
Currentratio	-	-0.0602**	0.038	-0.0160***	0.001	-0.0143	0.365
BTM	-	-0.0079	0.891	-0.0467***	0.001	-0.2754**	0.045
Growth	-	-0.1223	0.187	-0.0037	0.682	-0.0385*	0.088
Loss	+	0.2706*	0.051	0.0995***	0.002	0.0386	0.863
Foreign	+	0.2077	0.241	0.2909***	0.001	0.1768	0.278
Merger	+	0.0658	0.663	0.0582**	0.044	0.0560	0.794
Special	+	0.0820	0.738	0.1490***	0.001	0.2699*	0.094
Pscore	+	-10.8065**	0.035	-0.8217	0.384	0.8056	0.932
Firstyear	-	0.0310	0.873	-0.0732	0.158	-0.0509	0.843
GC	+	0.1367	0.648	0.2059**	0.012	0.7562	0.134
Big4	+	0.7411***	0.001	0.4124***	0.001	0.8107***	0.001
IC	+	0.1234**	0.046	0.1122***	0.001	Omitted	
Industry effect		Included		Included		Included	
Observations		90		2172		70	

Predictive Performance

	Linear Regression		RF		ANN	
	Baseline model	Sentiment Model	Baseline model	Sentiment model	Baseline model	Sentiment model
	(1)	(2)	(3)	(4)	(5)	(6)
RMSE	0.5984	0.4720	0.6902	0.6879	0.6261	0.6248
MAE	0.4297	0.3671	0.4170	0.4269	0.4617	0.4619

$$MAE = \frac{1}{n}\sum_{j=1}^{n} |y_j - \hat{y}_j|$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

Conclusion

- Twitter provides qualitative information regarding the risk of the prospective client. It can be used as a technology shortcut to improve the quality of audit decision making (Western Intergovernmental Audit Forum, 2013).
- Deep learning can be leveraged to identify the sentiment of social media data to offer efficient and effective evidence with limited human bias
- **Contribution**
  - First research examining the information content of social media for audit decision making
  - Provides supporting evidence that social media can be used as additional information source to for audit risk assessment

RUTGERS

Conclusion, Limitation, and Future Research

- Conclusion

  Deep learning is a promising technology that can be used to improve the effectiveness of auditors' risk assessment

- Contribution

  Findings in this dissertation contribute to auditing research by investigating how deep learning can be implemented to extract sentiment features from business communication documents and utterances in social media to help auditors improve the quality of risk assessment

- Limitation
  - The deep learning algorithm applied in this study is not exclusively trained with finance-specific data
  - Conference calls (Q&A, manager vs. analysts)
  - Data availability issue of Tweets (only one year)

- Future Research
  - Framework
  - Audit-specific data
  - More business communication data sources
  - how the auditor without programming skills can use the deep learning tools in practice
  - a comparison between AI and human auditors can be conducted as a behavior research

Thank you!