Business analytics using machine learning and large-scale textual data: Three essays

Shuang Zhai
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Recommended Citation
Zhai, Shuang, "Business analytics using machine learning and large-scale textual data: Three essays" (2020). Graduate Theses and Dissertations. 18255.
https://lib.dr.iastate.edu/etd/18255

This Thesis is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Business analytics using machine learning and large-scale textual data:

Three essays

by

Shuang Zhai

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Business and Technology (Information Systems)

Program of Study Committee:
Zhu Zhang, Major Professor
Arnold Cowan
Joey F. George
Zhengrui Jiang
Jin Tian

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University
Ames, Iowa
2020

Copyright © Shuang Zhai, 2020. All rights reserved.
DEDICATION

I would like to dedicate this dissertation to my family.
TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................ vi

LIST OF FIGURES ..................................................................................................... viii

ACKNOWLEDGMENTS ............................................................................................. xi

ABSTRACT ................................................................................................................... xii

CHAPTER 1. GENERAL INTRODUCTION ................................................................... 1

CHAPTER 2. EBB AND FLOW: THE EVOLUTION OF THE INTELLECTUAL CORE OF ACAD

EMIC INFORMATION SYSTEMS .................................................................................. 4
  2.1 Abstract ............................................................................................................... 4
  2.2 Introduction ......................................................................................................... 5
  2.3 Related Studies .................................................................................................. 8
  2.4 Methods ............................................................................................................. 22
    2.4.1 Latent Dirichlet Allocation (LDA) vs. Latent Semantic Analysis (LSA) .... 22
    2.4.2 Latent Dirichlet Allocation (LDA) .............................................................. 23
    2.4.3 Capturing Topic Dynamics by Dynamic Topic Models (DTM) ............... 24
    2.4.4 Measuring Topic Mass by Probability Mass .............................................. 25
    2.4.5 Measuring Inter-Corpus Difference by Kullback-Leibler (KL) Divergence . 26
  2.5 Data .................................................................................................................... 26
  2.6 Analysis .............................................................................................................. 27
    2.6.1 Topic Trend: Identity Level ........................................................................ 28
    2.6.2 Topic Trend: Granular Level .................................................................... 31
    2.6.3 Topic Trend: North American-based Journals vs. European-based Journals . 37
    2.6.4 Topic Morphing .......................................................................................... 43
  2.7 Discussion .......................................................................................................... 49
  2.8 Conclusion .......................................................................................................... 51
  2.9 References ........................................................................................................ 52
  2.10 Appendix: LDA Results ................................................................................... 57

CHAPTER 3. READ THE NEWS, NOT THE BOOKS: FORECASTING FIRMS’ LONG-TERM FINANCIAL PERFORMANCE VIA DEEP TEXT MINING .............................................. 58
  3.1 Abstract ............................................................................................................. 58
  3.2 Introduction ....................................................................................................... 58
  3.3 Related Work .................................................................................................... 61
    3.3.1 Finance Literature ................................................................................... 61
    3.3.2 Related Modeling Techniques ................................................................. 63
  3.4 Problem Formulation ....................................................................................... 67
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5</td>
<td>Deep Learning Models for Forecasting Firm Financial Ratios</td>
<td>70</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Single-Task Event MaxPooling Model (Single-EvMax)</td>
<td>71</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Single-Task Event Attention Model (Single-EvAttn)</td>
<td>72</td>
</tr>
<tr>
<td>3.5.3</td>
<td>Multi-Task Event Attention Model (Multi-EvAttn)</td>
<td>74</td>
</tr>
<tr>
<td>3.5.4</td>
<td>Ratio LSTM Model</td>
<td>75</td>
</tr>
<tr>
<td>3.5.5</td>
<td>Event Ratio Models</td>
<td>76</td>
</tr>
<tr>
<td>3.6</td>
<td>Empirical Results</td>
<td>77</td>
</tr>
<tr>
<td>3.6.1</td>
<td>Experimental Setup</td>
<td>77</td>
</tr>
<tr>
<td>3.6.2</td>
<td>Model Performance</td>
<td>83</td>
</tr>
<tr>
<td>3.7</td>
<td>Model Interpretability</td>
<td>86</td>
</tr>
<tr>
<td>3.7.1</td>
<td>Visualizing and Understanding Event Attention</td>
<td>86</td>
</tr>
<tr>
<td>3.7.2</td>
<td>Visualizing and Understanding Event and Ratio Interaction</td>
<td>88</td>
</tr>
<tr>
<td>3.7.3</td>
<td>Visualizing and Understanding Firm Embeddings</td>
<td>89</td>
</tr>
<tr>
<td>3.8</td>
<td>Conclusion, Implications, and Future Work</td>
<td>94</td>
</tr>
<tr>
<td>3.9</td>
<td>References</td>
<td>96</td>
</tr>
<tr>
<td>3.10</td>
<td>Appendix A: ARIMA Formulation</td>
<td>99</td>
</tr>
<tr>
<td>3.11</td>
<td>Appendix B: VAR Formulation</td>
<td>100</td>
</tr>
<tr>
<td>3.12</td>
<td>Appendix C: LSTM Formulation</td>
<td>101</td>
</tr>
<tr>
<td>3.13</td>
<td>Appendix D: Case Study: Attention Analysis on Expedia</td>
<td>102</td>
</tr>
</tbody>
</table>

CHAPTER 4. WHAT'S NEXT? HARNESSING AI TO FORECAST FIRM MATERIAL EVENT SEQUENCES

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Abstract</td>
<td>109</td>
</tr>
<tr>
<td>4.2</td>
<td>Introduction</td>
<td>109</td>
</tr>
<tr>
<td>4.3</td>
<td>Literature Review</td>
<td>111</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Corporate Reporting</td>
<td>111</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Corporate Strategy</td>
<td>113</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Event Study</td>
<td>114</td>
</tr>
<tr>
<td>4.3.4</td>
<td>Event Prediction</td>
<td>115</td>
</tr>
<tr>
<td>4.4</td>
<td>Related Modeling Techniques</td>
<td>116</td>
</tr>
<tr>
<td>4.4.1</td>
<td>AutoRegressive Integrated Moving Average (ARIMA) Model</td>
<td>116</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Recurrent Neural Networks (RNNs)</td>
<td>118</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Sequence-to-Sequence (Seq2Seq) Neural Network</td>
<td>119</td>
</tr>
<tr>
<td>4.4.4</td>
<td>Attention Mechanism</td>
<td>119</td>
</tr>
<tr>
<td>4.4.5</td>
<td>Transformer Model</td>
<td>119</td>
</tr>
<tr>
<td>4.5</td>
<td>Multiple Research Gaps</td>
<td>121</td>
</tr>
<tr>
<td>4.6</td>
<td>Corporate Event Sequence and Event Types</td>
<td>123</td>
</tr>
<tr>
<td>4.6.1</td>
<td>Corporate Event Sequence Prediction Task</td>
<td>123</td>
</tr>
<tr>
<td>4.6.2</td>
<td>Material Event Items by SEC</td>
<td>124</td>
</tr>
<tr>
<td>4.6.3</td>
<td>Corporate Event Types in Our Study</td>
<td>124</td>
</tr>
<tr>
<td>4.7</td>
<td>Task Definition</td>
<td>126</td>
</tr>
<tr>
<td>4.7.1</td>
<td>Task Framework</td>
<td>126</td>
</tr>
<tr>
<td>4.7.2</td>
<td>Task Formulation</td>
<td>127</td>
</tr>
<tr>
<td>4.8</td>
<td>Model Formulations</td>
<td>128</td>
</tr>
<tr>
<td>4.8.1</td>
<td>Event Attention Layer for Model Input</td>
<td>128</td>
</tr>
<tr>
<td>4.8.2</td>
<td>The GRU Model</td>
<td>129</td>
</tr>
</tbody>
</table>
CHAPTER 5. GENERAL CONCLUSION
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1</td>
<td>Related Key References</td>
<td>11</td>
</tr>
<tr>
<td>Table 2.2</td>
<td>Data Statistics</td>
<td>27</td>
</tr>
<tr>
<td>Table 2.3</td>
<td>LDA for MISQ, ISR and JMIS, Un-normalized Topic Mass, 1987-2019</td>
<td>57</td>
</tr>
<tr>
<td>Table 2.4</td>
<td>LDA for Basket of Eight Journals, Un-normalized Topic Mass, 1987-2019</td>
<td>57</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Stock Market: Theory Testing Literature</td>
<td>62</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>Stock Market: Prediction-related Literature</td>
<td>63</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>Firm Financial Ratios: Theory Testing Literature</td>
<td>64</td>
</tr>
<tr>
<td>Table 3.4</td>
<td>Definitions of Target Financial Ratios</td>
<td>68</td>
</tr>
<tr>
<td>Table 3.5</td>
<td>Data Statistics by Year</td>
<td>80</td>
</tr>
<tr>
<td>Table 3.6</td>
<td>ARIMA Besting-performing ((p, d, q)) for Final Testing</td>
<td>83</td>
</tr>
<tr>
<td>Table 3.7</td>
<td>Model Performance ((M=12, H=12, MAPE))</td>
<td>83</td>
</tr>
<tr>
<td>Table 3.8</td>
<td>Events Attention Example: sale_nwc, Apple Inc. Month 1</td>
<td>87</td>
</tr>
<tr>
<td>Table 3.9</td>
<td>Events Attention Visualization Example: Apple Inc. Month 11</td>
<td>88</td>
</tr>
<tr>
<td>Table 3.10</td>
<td>Fama-French 5-Industry Portfolios</td>
<td>90</td>
</tr>
<tr>
<td>Table 3.11</td>
<td>Events Attention Visualization Example: Expedia Month 1</td>
<td>103</td>
</tr>
<tr>
<td>Table 3.12</td>
<td>Events Attention Visualization Example: Expedia Month 2</td>
<td>104</td>
</tr>
<tr>
<td>Table 3.13</td>
<td>Events Attention Visualization Example: Expedia Month 4</td>
<td>105</td>
</tr>
<tr>
<td>Table 3.14</td>
<td>Events Attention Visualization Example: Expedia Month 5</td>
<td>105</td>
</tr>
<tr>
<td>Table 3.15</td>
<td>Events Attention Visualization Example: Expedia Month 7</td>
<td>106</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>3.16</td>
<td>Events Attention Visualization Example: Expedia Month 8</td>
<td>107</td>
</tr>
<tr>
<td>3.17</td>
<td>Events Attention Visualization Example: Expedia Month 11</td>
<td>107</td>
</tr>
<tr>
<td>4.1</td>
<td>Corporate Strategy Related Literature</td>
<td>114</td>
</tr>
<tr>
<td>4.2</td>
<td>Event Study Related Literature</td>
<td>116</td>
</tr>
<tr>
<td>4.3</td>
<td>Event Prediction Related Literature</td>
<td>117</td>
</tr>
<tr>
<td>4.4</td>
<td>Form 8-K Sections and Items by SEC</td>
<td>125</td>
</tr>
<tr>
<td>4.5</td>
<td>Proposed Event Types (Target Variables)</td>
<td>126</td>
</tr>
<tr>
<td>4.6</td>
<td>Event Type Distribution</td>
<td>138</td>
</tr>
<tr>
<td>4.7</td>
<td>Event Frequency Distribution</td>
<td>138</td>
</tr>
<tr>
<td>4.8</td>
<td>Event Matching Phrases</td>
<td>142</td>
</tr>
<tr>
<td>4.9</td>
<td>Model Performance on Time $t$ F1%</td>
<td>143</td>
</tr>
<tr>
<td>4.10</td>
<td>Categorical Cross Entropy and Perplexity</td>
<td>143</td>
</tr>
<tr>
<td>4.11</td>
<td>Model Performance on Time $t \pm 1$ F1%</td>
<td>144</td>
</tr>
<tr>
<td>4.12</td>
<td>Model Performance on Time $t \pm 2$ F1%</td>
<td>144</td>
</tr>
<tr>
<td>4.13</td>
<td>Model Performance on Time $t \pm 3$ F1%</td>
<td>145</td>
</tr>
<tr>
<td>4.14</td>
<td>Fred’s and Related Companies Financial Standing</td>
<td>147</td>
</tr>
<tr>
<td>4.15</td>
<td>Fama-French 5-Industry Portfolios</td>
<td>158</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.1</td>
<td>Number of Topics in Literature Over Time</td>
<td>22</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Dynamic Topic Model (DTM)</td>
<td>25</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>LDA vs. LSA for MISQ, ISR and JMIS, 1987-2006</td>
<td>28</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>LDA for MISQ, ISR and JMIS, 1987-2019</td>
<td>29</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>LDA for Basket of Eight Journals, 1987-2019</td>
<td>30</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>Eleven-topics LSA for ISR, MISQ, JAIS, JMIS, 2000-2016</td>
<td>31</td>
</tr>
<tr>
<td>Figure 2.7</td>
<td>Eleven-topics LDA for Basket of Eight Journals, 1987-2019 (Ours)</td>
<td>32</td>
</tr>
<tr>
<td>Figure 2.8</td>
<td>Eleven-Topics Distribution Comparison</td>
<td>33</td>
</tr>
<tr>
<td>Figure 2.9</td>
<td>Thirteen-topic Distribution of All Journals 1987-2006 vs. 2007-2019</td>
<td>36</td>
</tr>
<tr>
<td>Figure 2.10</td>
<td>Five-topic Analysis</td>
<td>38</td>
</tr>
<tr>
<td>Figure 2.11</td>
<td>North American-based Journals: Pre-Bologna vs. Post-Bologna</td>
<td>39</td>
</tr>
<tr>
<td>Figure 2.12</td>
<td>European-based Journals: Pre-Bologna vs. Post-Bologna Process</td>
<td>41</td>
</tr>
<tr>
<td>Figure 2.13</td>
<td>Topic Comparison</td>
<td>42</td>
</tr>
<tr>
<td>Figure 2.14</td>
<td>KL Divergences</td>
<td>43</td>
</tr>
<tr>
<td>Figure 2.15</td>
<td>Cross-topic Morphing Example 1</td>
<td>45</td>
</tr>
<tr>
<td>Figure 2.16</td>
<td>Cross-topic Morphing Example 2</td>
<td>46</td>
</tr>
<tr>
<td>Figure 2.17</td>
<td>Inter-topic Morphing Example 1</td>
<td>48</td>
</tr>
<tr>
<td>Figure 2.18</td>
<td>Inter-topic Morphing Example 2</td>
<td>49</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Deep Learning for Natural Language Processing: An Illustration</td>
<td>66</td>
</tr>
<tr>
<td>Figure 4.11</td>
<td>Horizontal Data Segmentation ............................................. 140</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>---------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Figure 4.12</td>
<td>Fred’s 8-K Event Sequence ................................................. 148</td>
<td></td>
</tr>
<tr>
<td>Figure 4.13</td>
<td>Encoder Attention .................................................................. 153</td>
<td></td>
</tr>
<tr>
<td>Figure 4.14</td>
<td>Encoder-Decoder Attention .................................................... 154</td>
<td></td>
</tr>
<tr>
<td>Figure 4.15</td>
<td>Decoder Attention .................................................................. 155</td>
<td></td>
</tr>
<tr>
<td>Figure 4.16</td>
<td>Firm Embedding ...................................................................... 157</td>
<td></td>
</tr>
<tr>
<td>Figure 4.17</td>
<td>Event Embedding .................................................................... 158</td>
<td></td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. Zhu Zhang, and my committee members, Dr. Joey F. George, Dr. Zhengrui Jiang, Dr. Jin Tian, and Dr. Arnold Cowan, for their guidance and support throughout my five years’ study at the Ivy College of Business, Iowa State University.

In addition, I would also like to thank my friends, colleagues, and department faculty and staff for making my time at Iowa State University a wonderful experience.
ABSTRACT

Natural Language Processing (NLP) techniques have been recognized as useful tools in various applications, such as language translation, stock price prediction, and semantic analysis. In recent years, the larger amount of data and faster computational hardware become available, NLP techniques are applied in a much broader application spectrum, and business applications are the frontier. My dissertation works at the intersection of business, Information Systems, and NLP. It extends NLP application scenarios in the business setting and employs NLP techniques in different real-world scenarios, including document topic categorization, company performance prediction, and corporate event sequence prediction. In particular, these problems are addressed by topic modeling techniques, recurrent neural networks, and sequence-to-sequence neural network architecture.

My dissertation comprises three independent essays. The first essay utilizes topic modeling techniques to demonstrate the scientific topic trends and topic morphing for top Management Information System (MIS) publications. The second essay proposes a recurrent neural network model to predict multiple company financial ratios from publicly available news articles. The third essay first formulates a unique business question, predicting corporate event sequences, based on the historical event sequences submitted in the U.S. Securities and Exchange Commission (SEC) filings. Then, it uses the Transformer architecture to address the problem.

In these essays, I apply and extend a variety of NLP tools to solve real-world business problems, including analyzing discipline publication trends and morphing, predicting company financial performance, and predicting corporate event sequences. NLP techniques employed in my dissertation include topic modeling, recurrent neural networks, and sequence-to-sequence networks. These essays contribute to discipline study and financial technology literature and develop tools for scholars, investors, and organizations to make informed decisions, from natural language.
CHAPTER 1. GENERAL INTRODUCTION

Natural language is an information vehicle used by human beings. It can carry and transfer information from one person to another. Humans use natural language to store, convey, and absorb information. Natural language is reflected in various forms, such as speaking language and written language. We also call written language as text. While speaking language can only have audiences who are at the same time and location as the speaker; texts, on the other hand, can benefit more audiences who are not at the same time and location as the speaker. Therefore, the text is viewed as an information carrier and can preserve information across temporal and spatial dimensions. Businesses have a lot of textual data. However, business participants have not fully utilized textual data to inform decisions. My dissertation extends the use cases for Natural Language Processing (NLP) in business analytics and develops tools to solve real-world business questions using natural language.

My dissertation comprises three logically independent essays. Each essay demonstrates one textual data application in business domains. The first essay utilizes the Latent Dirichlet Allocation (LDA) to analyze the discipline’s intellectual core trends for Management Information Systems (MIS) discipline. I compare research topics between North American based journals and European based journals, based on top MIS journals. I also employ Dynamic Topic Models (DTM) to identify two topic morphing behaviors: inter-topic morphing and cross-topic morphing. The second essay mines firm-related events in public news to predict various firm financial ratios. By exploiting neural architectures, including pseudo-event embeddings, Long Short-Term Memory Networks (LSTM), and attention mechanisms, my proposed news-powered neural network models which are shown to outperform standard econometric models operating on precise historical accounting data. The third essay formulates a unique business question, predicting corporate event sequences based on the firm’s historical event sequences. The problem was previously considered unapproachable. My
third essay proposes a novel solution that considers state-of-art techniques and real-world scenarios. In particular, it utilizes the Transformer model to forecast the firm’s material event series, based on the firm’s current 8-K reports submitted to the U.S. Securities and Exchange Commission (SEC). The proposed model demonstrates forecasting improvements over traditional sequence-to-sequence models and task-specific Markov Chain Monte Carlo simulations. The third essay also illustrates the inner-working of the proposed model to facilitate user understanding.

A variety of natural language processing techniques have been applied and extended in my dissertation. In the first essay, Latent Dirichlet Allocation (LDA) is applied to diagnose research topic trends, and Dynamic Topic Models (DTM) is adopted to identify topic morphing behaviors over time. In the second essay, I propose text-based Recurrent Neural Network (RNN) models to predict the firm’s financial ratios individually and simultaneously, from publicly available news articles. Furthermore, I propose text-ratio integration models to predict the firm’s financial ratios from both company news and its historical financial ratios. The proposed models are shown to outperform traditional time series models, such as the AutoRegressive Integrated Moving Average (ARIMA) models and Vector Autoregression (VAR) model. In the third essay, I propose a Sequence-to-Sequence (Seq2Seq) neural network model to predict corporate event sequences (CES) based on the firm’s historical event sequences. By employing the Transformer model, self-attention mechanism, and the encoder-decoder framework, my proposed CES model outperforms traditional recurrent neural network models such as Gated Recurrent Units (GRUs) and task-specific Markov Chain Monte Carlo (MCMC) simulation.

My dissertation contributes to the NLP literature in business domains by proposing various NLP analytical tools to facilitate business-related decision-making processes. The topic models demonstrate scholars can use texts, such as publications, to track discipline development trends, discover topic morphing, and better understand the dynamic of the field. The news empowered RNN model builds linkage between firm financial ratios and publicly available news articles, and provide decision supports for firm executives, main street investors, institutional investors, and third-party regulators and agencies. The corporate event series prediction model is to solve a unique
real-world business question that, by our knowledge, was not approachable before. The proposed model provides text-based data-driven organizational analysis and implications for business owners and executives, corporate internal and external investors, and third-party regulators and agencies. Furthermore, the demonstrated model inner-working illustrations can ease user understanding of the model and prompt NLP model usage in business domains.

In summary, my dissertation applies topic modeling techniques, including Latent Dirichlet Allocation and Dynamic Topic Models, on top MIS publications to analyze research topic trends and topic morphs in the field. Furthermore, my dissertation develops text-based business NLP models to facilitate decision-making processes and subsequently predict organization financial performance and event sequences, based on textual data such as news articles and SEC filings. The proposed NLP business models demonstrate their values and are shown to benefit various shareholders in their decision-making processes.
CHAPTER 2. EBB AND FLOW: THE EVOLUTION OF THE INTELLECTUAL CORE OF ACADEMIC INFORMATION SYSTEMS

Shuang Zhai, Joey F. George, and Zhu Zhang
Ivy College of Business, Iowa State University, Ames, IA, USA

Modified from a manuscript under review in Communication of the Association for Information Systems

2.1 Abstract

This paper explores the intellectual core of the academic study of information systems, based on our extended study on Sidorova, Evangelopoulos, Valacich, and Ramakrishnan (2008) and Goyal et al. (2018). We identified research topic trends and emergent topics using Latent Dirichlet Allocation and Dynamic Topic Model, for all Basket of Eight journals, North American based journals, and European based journals. We identified two types of topic morphing behaviors: cross-topic morphing and inter-topic morphing and demonstrated their examples. We found scholars from around the world have come to publish their work in the same narrow set of journals, regardless of where they work or what their research perspectives might be. Within this set of journals, there has been a growing movement toward homogeneity and convergence in research topics. The identified emergent topics include Virtual Collaboration, Security and Privacy, Social Media, Open Source Software, Online Review, and IT and Healthcare. The illustrated cross-topic morphing examples include Group Support System to Virtual Collaboration and Network Communication to Social Media. The illustrated inter-topic morphing examples include Value of IT and IT and Healthcare. We present a reinforcement loop between research topic convergence and topic morphing, with the empowerment by advanced technology, based on our analysis results.
Keywords: Information systems; Discipline; Research topics; Research trends; Latent Dirichlet Allocation (LDA); Dynamic Topic Model (DTM), European journals, North American journals, convergence, national journal lists.

2.2 Introduction

In 2008, Sidorova, Evangelopoulos, Valacich, and Ramakrisnan (2008) revealed intellectual cores of the Management Information Systems (MIS) discipline in MIS Quarterly. In the study, the authors categorized MIS researches published in three premier journals from 1985 through 2006 into five identity level topics: Information Technology (IT) and Organizations, IT and Individuals, IT and Groups, IT and Markets, and IS Development. Moreover, the authors separated all papers into four time periods (1987-1991, 1992-1996, 1997-2001, 2002-2006), and showed the most common research topic at the end of the 1980s, IS Development, had fallen to the least common by 2006. Conversely, the least common topic in the 1980s, IT and Markets, became the leading topic by 2006.

As time goes by, many questions related to the core IS research topics remain unsolved after the study. Two high-level questions we should ask: A. what are the research topic trends? B. how topics are morphed overtime?

AIS provides a Basket of Eight journal list on April 23, 2007 (revised on December 6, 2011). Both Sidorova et al. (2008) and Goyal et al. (2018) used part of the eight journals list. We believe a consideration of papers published in all journals in the Association of Information Systems Basket of Eight journal list would provide a more holistic picture of the field’s intellectual core. Also, both Sidorova et al. (2008) and Goyal et al. (2018) employed latent semantic analysis (LSA) in their studies.

Meanwhile, the AIS Basket of Eight journal list includes four North American-based IS journals and four European-based IS journals. When we study European journals, we realize the Bologna Process (COMM, 2015) officially began in 1998 with the Sorbonne Declaration. We wonder if there is any topic trend difference in pre-Bologna and post-Bologna Process.
Therefore, we address the following questions.

**A. Research topic trends:**

1. What are the research topic trends and emerged topics after Sidorova et al. (2008) study?
2. What are the research topic differences between Goyal et al. (2018) and our study?
3. What are research topic trends on both identity level and granular level, for all Basket of Eight journals?
4. What are the topic trend differences between North American-based IS journals and European-based IS journals? What are the research topic differences between pre-Bologna and post-Bologna processes?

**B. Research topic morphing:**

5. What are the types of topic morphing, and what are their examples?
6. What contributes to topic morphing? What can we learn from it?

For research topic trends, we compared MIS academic research trends on different levels from different journals, time periods, and topic granularity. We identified emergent research topics for all Basket of Eight journals, North American-based IS journals, and European-based IS journals. We have a common finding from these analyses. Scholars have come into the same narrow set of journals, regardless of their work locations or research perspectives. Within this set of journals, there has been a growing movement toward research topic homogeneity and convergence.

Interestingly, the more comprehensive MIS discipline content coverage and the broader individual level topic content coverage contribute to the field’s research topic convergence. While technologies are applied across business domains, successful advanced technology-empowered business applications are limited. Therefore, while MIS academic research opportunities have widely opened, research is squeezed into areas where technology sees success. Hence, technology advancements play a critical role in research topics homogeneity and convergence.

For research topic morphing, we identified two types of topic morphing behaviors, cross-topic morphing, and inter-topic morphing. The cross-topic morphing is one topic that has morphed into its new form and become a new topic, given the support from advanced technology. The examples
include Group Support System to Virtual Collaboration and Network Communication to Social Media. The inter-topic morphing describes the shift of topic concentration within one topic, such as the focus shift of IT value from cost to consumers, and the increased attention on hospitals, patients, and medicine in IT and Healthcare. Technology advancements promoted the focus shifts in these processes.

As a summary, topic convergence and topic morphing worked reciprocally and created a reinforcement loop supported by advanced technology. On the one hand, topics that can be benefited from technology will be morphed either into a new form or shift its topic focus to where technology empowerments play a higher weighted role. On the other hand, topics that can not be benefited from advanced technology will fade out popular research topics list and move the left research topics further towards homogeneity and convergence. Technology empowerment makes topic convergence and topic morphing as a mutual result.

In addition, Information and Communication Technologies (ICTs) and their usages have been marked significant changes since 2006. A glance at the 2002-2006 column in Table 1 of Sidorova et al. (2008) reveals few explicitly web-related topics, indicating academic research on the web only first appeared in the 2002-2006 MIS literature, even though the first web browser, Mosaic, was released in 1993. Missing completely, of course, are any research topics related to smartphones and mobile devices – the iPhone was first released on June 29, 2007. Therefore, the changes in ICTs since 2006 and their continuing high rates of penetration throughout the world imply revisiting the question of the intellectual core of the MIS field is in order, even if only to see how the MIS research of the past decade complements or expands the findings from Sidorova and colleagues (2008).

Recently, Goyal et al. (2018) provided an analysis of papers published in four ¹ MIS journals from the year 2000 to the first quarter of 2017 and increased the number of research themes from five to eleven in their study. Although Sidorova et al. (2008) also had thirteen topics analysis in their appendix, Goyal et al. (2018) proposed several topic labels that are different from the former. However, to better understand the trends of Sidorova et al. (2008) after 2006, an updated study

¹Management Information Systems Quarterly (MISQ), Information Systems Research (ISR), Journal of the Association for Information Systems (JAIS), and Journal of Management Information Systems (JMIS).
uses the same or similar terminologies and analyzes both the identity level and granular level are still in need.

In terms of methodology, we employ newer and arguably stronger computational models, including Latent Dirichlet Allocation (LDA) and Dynamic Topic Models (DTM), and employ them in the current study. These approaches serve to our analysis needs and are poised to depict the intellectual core of the field better because LDA provides a Bayesian generative view of document-topic mapping, and DTM can recognize the changing dynamic of topics over time.

We organize this study as the following. First, we present related literature and employed technologies. We then analyze topics to compare with Sidorova et al. (2008), Goyal et al. (2018), and between North American-based journals and European-journals and identify emergent topics. Next, we present two types of topic morphing, cross-topic morphing and inter-topic morphing, and their examples. At last, we discuss our reflection on topic convergence and topic morphing and provide our thoughts.

### 2.3 Related Studies

Management Information System (MIS) researchers have investigated the field’s research areas and topics since our discipline has started. We list the related studies in Table 2.1.

Across these studies, the number of topics identified by researchers grew from a small number, such as five topics, to a large number, such as 30 - 250 topics, to show diversity. In recent years, however, because of the maturity of our discipline, researches have shown a moderate number of topics, such as 11 - 13 topics, to represent our discipline’s identity and diversity. We show the number of topics by the previous studies by year in Figure 2.1. We also believe the moderate amount of topics can help researchers to understand our discipline topic dynamics while maintaining an open interest in different research areas. Therefore, in the current study, we employ the identity level - five topics and granular level – 11 topics and 13 topics to compare and present the topic dynamics.
For research topic categories, researchers have classified MIS publications from different facets. In the beginning, the research themes were viewed toward environmental interactions. Ives et al. (1980) used doctoral dissertations and evaluated five existing MIS research frameworks, proposing one comprehensive model which considers the environment, process, and information subsystem (ISS) variables.

By considering the environmental, technical, and management (Culnan, 1986, 1987; Farhoomand, 1987) aspects of the field, the nine categories proposed by Barki et al. (1988) were used in many studies (Barki et al., 1993; Farhoomand and Drury, 1999; Avison et al., 2008).

Organizational and computational areas are considered mainstream research topics for quite a while (Alavi and Carlson, 1992; Glass, 1992; Swanson and Ramiller, 1993; Claver et al., 2000), before the discussions of societal issues (Vessey et al., 2002) and economics of IS (Banker and Kauffman, 2004).

Although researchers had called for a hard and clear boundary for Information System research (Benbasat and Zmud, 2003), opposing views called for an embrace of the diversity of Information System (IS) research (DeSanctis, 2003). Communication in Association Information Systems (CAIS) had a special issue (vol 12, 2003) to discuss the definition of the core of Information System discipline.

Since then, to represent the diversity characteristics, the proposed number of research topics has been increased, and the level of granularity has become smaller (Palvia et al., 2003, 2004, 2007; Dwivedi and Kuljis, 2008; Avison et al., 2008; Sidorova et al., 2008; Goyal et al., 2018).

After more quantitative and automated computational methods were introduced, researchers were relived from manual work (Larsen et al., 2008), and the number of topics is back to a level that can highlight the discipline identity and maintain a reasonable granularity (Sidorova et al., 2008; Goyal et al., 2018). We applaud the moderate granularity level and continue this trend to conduct our analysis on this level.

Larsen et al. (2008) employed the Latent categorization method (LCM) and categorized the IS field into seven intellectual communities. Sidorova et al. (2008), Evangelopoulos et al. (2012), and
Goyal et al. (2018), employed the Latent Semantic Analysis (LSA) method. Sidorova et al. (2008) provided 2-, 3-, 4-, 5-, 8-, 12-, 13-, 100-factors analysis results. The authors suggested 5-factors result is the best among all the factor analysis in their study, and provided details about 13-factors analysis as a reasonable granular level result, without making it too granular (100-factors in the appendix) and distracting from the core in the main paper. Evangelopoulos et al. (2012) increased the number of topics to 30. Goyal et al. (2018) proposed a reasonable granular level 11-factors result, with a more recent sample.

Galliers and Whitley (2007) employed the categories from Banker and Kauffman (2004) and found electronic markets has gained a growing interest while the interest to decision support systems has declined. Palvia et al. (2007) manually assigned multiple topics to one article and found IS usage and IS management are the two topmost researched topics. Dwivedi and Kuljis (2008) surveyed IS academics in 18 European countries (Avgerou et al., 1999), and identified IS management and IS development are the most researched two topics.

IS development and IS management are the topmost two topics identified by Avison et al. (2008). Meanwhile, while IS development was on a decline, IS education, IS types, organizational environment, and IS usage were on the rise. IS development and IS usage are the most prominent topics in MCIS proceedings identified by Pouloudi et al. (2012), and these two topics account for 26.5% of the proceedings. Adding in IS implementation and IS research, the top four topics account for 40% of the accepted papers.

As a globally fast-growing discipline, MIS is at the intersection of many areas, such as business, technology, sociology, and many others (Glass, 1992); therefore, an updated study on MIS research trends, emergent and fade out topics, and research morphing, are needed. Furthermore, thoughts on the relationship between topic trends and topic morphing have a point. In addition, most of the previous studies focus on either North American-based or European-based articles alone; a study on analyzing the difference between the two groups is also in need. Therefore, we will bridge the gaps and provide our observations.
Table 2.1: Related Key References

<table>
<thead>
<tr>
<th>Article</th>
<th>Data Source</th>
<th>Geography</th>
<th>Years</th>
<th>Type</th>
<th>Method</th>
<th>Research Topics / Categories / Themes / Factors</th>
<th>Topic Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ives et al. (1980)</td>
<td>Doctoral dissertations</td>
<td>NA</td>
<td>1973-1979</td>
<td>abstract, key-words</td>
<td>manual</td>
<td>Five categories in the intersection of environment variables, process variables, and information subsystem.</td>
<td>5 types</td>
</tr>
<tr>
<td>Culnan (1986)</td>
<td>SSCI database, MISQ, Social Scisearch</td>
<td>Global</td>
<td>1972-1982</td>
<td>authors, title</td>
<td>co-citation</td>
<td>foundations / management theory, computing impacts, MIS / DSS implementation, individual differences, human factors, computer conferencing, and others.</td>
<td>9 factors</td>
</tr>
</tbody>
</table>
Table 2.1: continued

<table>
<thead>
<tr>
<th>Source</th>
<th>Database</th>
<th>Region</th>
<th>Dates</th>
<th>Search Keywords</th>
<th>Research Areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Farhoomand (1987)</td>
<td>MISQ, IM, SOS, CACM, MS, HBR</td>
<td>NA</td>
<td>1977-1985</td>
<td>keywords</td>
<td>Cost benefit analysis; database, systems, software, software design; data management, information resource management (IRM), ...</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>manual</td>
<td></td>
</tr>
</tbody>
</table>

5 clusters

16 themes
Table 2.1: continued

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Journals</th>
<th>Year</th>
<th>Keywords</th>
<th>Manual</th>
<th>Reference Disciplines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barki et al. (1988)</td>
<td>MISQ, IM, CACM, MS, DS</td>
<td>NA</td>
<td>1977-1987</td>
<td>keywords, abstract</td>
<td>MISQ, IM, CACM, MS, DS, NA</td>
</tr>
<tr>
<td>Glass (1992)</td>
<td>curriculum</td>
<td>NA</td>
<td>Pre 1992</td>
<td>article</td>
<td>manual</td>
</tr>
</tbody>
</table>

6 IS = Information Systems
<table>
<thead>
<tr>
<th>Study</th>
<th>Journals</th>
<th>Year</th>
<th>Keywords</th>
<th>Manual Method</th>
<th>Source</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swanson and Ramiller (1993)</td>
<td>Manuscript submission to ISR</td>
<td>NA</td>
<td>1987-1992</td>
<td>title, abstract, introduction, discussion, conclusion</td>
<td>manual</td>
<td>Computer-supported co-operative work, information and interface, decision support and knowledge-based systems, systems projects, evaluation and control, users, economics and strategy, introduction and impact, IS research</td>
</tr>
<tr>
<td>Claver et al. (2000)</td>
<td>MISQ, IM</td>
<td>NA</td>
<td>1981-1997</td>
<td>abstract, title, and keywords</td>
<td>manual</td>
<td>IS management, IS development/IS life cycle, information technologies, IS usage, and others</td>
</tr>
<tr>
<td>Author(s) and Year</td>
<td>Journal (s)</td>
<td>Region(s)</td>
<td>Year(s)</td>
<td>Feature(s)</td>
<td>Manual</td>
<td>Source</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------</td>
<td>-----------</td>
<td>---------</td>
<td>------------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Vessey et al. (2002)</td>
<td>ISR, JMIS, MISQ, MS, DS</td>
<td>NA</td>
<td>1995-1999</td>
<td>article</td>
<td>manual</td>
<td>Computer, system/software, data/information, problem domain-specific, systems/software management, organizational, societal, disciplinary issues concepts</td>
</tr>
<tr>
<td>Claver et al. (2000)</td>
<td>MISQ, IM</td>
<td>NA</td>
<td>1981-1997</td>
<td>abstract, title, and keywords</td>
<td>manual</td>
<td>IS management, IS development/IS life cycle, information technologies, IS usage, and others</td>
</tr>
</tbody>
</table>
Table 2.1: continued

<table>
<thead>
<tr>
<th>Authors</th>
<th>Journals/Collections</th>
<th>Timeframe</th>
<th>Data Source</th>
<th>Methodology</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vessey et al. (2002)</td>
<td>ISR, JMIS, MISQ, MS, DS</td>
<td>NA 1995-1999</td>
<td>article</td>
<td>manual</td>
<td>Computer, system/software, data/information, problem domain-specific, systems/software management, organizational, societal, disciplinary issues concepts 8 categories</td>
</tr>
<tr>
<td>Palvia et al. (2003)</td>
<td>CACM, DS, IM, ISR, JMIS, MS, MISQ</td>
<td>NA 1993-1997</td>
<td>keywords, abstract, article</td>
<td>manual</td>
<td>Details 8 are in the footnote 29 subjects</td>
</tr>
<tr>
<td>Banker and Kauffman (2004)</td>
<td>MS</td>
<td>NA</td>
<td>1954-2003</td>
<td>article</td>
<td>manual</td>
</tr>
<tr>
<td>---------------------------</td>
<td>----</td>
<td>----</td>
<td>------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>Oh et al. (2005)</td>
<td>ISR, JMIS, MISQ, MS</td>
<td>NA</td>
<td>1980-2002</td>
<td>authors</td>
<td>Co-authorship network analysis</td>
</tr>
<tr>
<td>Source</td>
<td>Database</td>
<td>Region</td>
<td>Years</td>
<td>Type</td>
<td>Details</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>------------</td>
<td>--------</td>
<td>---------</td>
<td>----------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Palvia et al. (2007)</td>
<td>IM</td>
<td>NA</td>
<td>1992-2005</td>
<td>keywords, abstract, article</td>
<td>Details are in the footnote</td>
</tr>
<tr>
<td>Galliers and Whitley (2007)</td>
<td>ECIS proceedings</td>
<td>EU</td>
<td>1993-2002</td>
<td>title, abstract, article</td>
<td>Society, IS organizational and strategy, system development, decision support systems, technology, research issues, economic, human, electronic markets</td>
</tr>
</tbody>
</table>
Table 2.1: continued

<table>
<thead>
<tr>
<th>Study</th>
<th>Source</th>
<th>Journals</th>
<th>Years</th>
<th>Lexical Method</th>
<th>Categories</th>
<th>Communities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Larsen et al. (2008)</td>
<td>65 information systems journals</td>
<td>Global</td>
<td>1990-2002</td>
<td>abstract latent categorization method (LCM)</td>
<td>MIS, global and societal, human-computer interaction, electronic commerce, systems and software engineering, information storage and retrieval, knowledge-based systems</td>
<td>7 communities</td>
</tr>
<tr>
<td>Sidorova et al. (2008)</td>
<td>MISQ, ISR, JMIS</td>
<td>NA</td>
<td>1985-2006</td>
<td>abstract latent semantic analysis (LSA)</td>
<td>IT and organizations, IS development, IT and individuals, IT and markets, IT and groups</td>
<td>2, 3, 4, 5, 8, 12, 13, 100 factors</td>
</tr>
<tr>
<td>Evangelopoulos et al. (2012)</td>
<td>EJIS</td>
<td>EU</td>
<td>1991-2008</td>
<td>abstract latent semantic analysis (LSA)</td>
<td>Clusters and topic factors</td>
<td>30 clusters / topic factors</td>
</tr>
<tr>
<td>Palvia et al. (2015)</td>
<td>MISQ, JAIS, JMIS, DS, EJIS, IM, ISR</td>
<td>NA + EU</td>
<td>2004-2013</td>
<td>title, keywords, abstract</td>
<td>manual</td>
<td>Details (^{14}) are in the footnote</td>
</tr>
<tr>
<td>---------------------</td>
<td>----------------------------------</td>
<td>---------</td>
<td>-------------</td>
<td>-------------------------</td>
<td>--------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Current Study</td>
<td>AIS basket of eight journals (MISQ, ISR, JAIS, JMIS, EJIS, ISJ, JIT, JSIS)</td>
<td>NA + EU</td>
<td>1987-2019</td>
<td>abstract</td>
<td>Latent Dirichlet Allocation (LDA), dynamic topic models (DTM)</td>
<td>Referred from Sidorova et al. (2008) and Goyal et al. (2018)</td>
</tr>
</tbody>
</table>

SSCI – Social Sciences Citation Index.


Decision support systems (DSS) decision theory; end-user computing (EUC); expert systems, artificial intelligence; human-computer interface; impact; implementation; information requirements analysis (IRA); interorganizational systems; management of and planning for IS; organizational design; strategic use of IS; technology transfer; other.


OS – Organizational Science.

Theory of MIS, artificial intelligence (AI) / expert system (ES) / neural networks (NN) / knowledge management (KM), global information technology (IT), hardware, software / programming languages, networks / telecommunications, Internet, electronic data interchange (EDI), electronic commerce (EC), multimedia, databases / DBMS, internal / external environment, organizational design / business process reengineering (BPR), innovation, resource management / IS management issues, IS planning, IS staffing, IS evaluation / control, security, IS development / methods and tools, IS implementation, IS usage, end user computing (EUC), executive information systems (EIS), decision support systems (DSS), group decision support systems (GDSS), IS function applications, IS education, IS research.

Resource management / IS management issues, IS development / methods and tools, organizational issues and design / BPR / workflow systems, electronic commerce / EDI, IS research, artificial intelligence / expert system / neural networks / knowledge management, IS usage, IS evaluation, software / programming languages, supply chain management (SCM) / ERP, IS implementation, IS planning, internal / external environment, Internet, IS staffing, global information technology, outsourcing, group decision support systems, innovation, theory of MIS, databases / DBMS, decision support systems, IS education, media and communications, security, IT value, multimedia, customer relationship management (CRM), end user computing, executive information systems, hardware, IS function application, networks / telecommunication.

IS development and implementation, knowledge management, IS and organizational issues, outsourcing and applications, IT architecture and performance, modeling, ERP, IT adoption, IS methodologies, critical success factors, IT innovation, social aspects of IS use, software products, IS publications, IT investment, job characteristics, public sector, soft systems methodology, programming, frames and references, project management, IS adoption, risk and trust issues, IS discipline, online customers, idea generation, group support systems (GSS), Web, social issues, claims and arguments.

Soft systems methodology, firm performance, IS publications, social perspectives, innovation, software, ERP, EDI, domain and data models, IS evaluation, e-business, outsourcing, mobile services, public sector, instrument development, project failure/success, online consumer, IS planning, team projects, DSS, agile software development, security, IT and groups, critical success factors, IS investment evaluation, IT jobs, imagination and knowledge, IS discipline, knowledge management, web site development.


2.4 Methods

While the previous computational studies assume one research paper can only have one research topic, we release this hard assumption and give one article the flexibility to be assigned to multiple research topics in this study. Furthermore, given the released assumption, we also study how the research topics are morphed over time.

2.4.1 Latent Dirichlet Allocation (LDA) vs. Latent Semantic Analysis (LSA)

The Latent Semantic Analysis (LSA) method (Deerwester et al., 1990) employed by Sidorova et al. (2008) and Goyal et al. (2018) used singular value decomposition (SVD) to reduce the dimensionality of the matrix. It is, in principle, a factor analysis technique and returns many topic groups. In the meantime, it returns high loading words for each topic. Users need to specify the number of topics and words in advance. One topic or theme can be assigned to one document. In other words, one document can only associate with one topic or theme.

Evangelopoulos (2016) illustrated the LSA process into three steps: pre-LSA textual data quantification step, including term filtering, stemming, lemmatization, and term weighting; core LSA
step, including SVD and dimensionality reduction; and post-LSA quantitative analysis step, including threshold selection, queries, clustering, and factor analysis.

In contrast, Latent Dirichlet Allocation (LDA) (Blei et al., 2003) can assign multiple topics or themes for each document, with the distribution for each topic, respectively. LDA is a parametric Bayesian algorithm that models the multi-stage data generation process in large text corpora. By following De Finetti’s representation theorem (De Finetti, 2017), LDA captures the intra-document statistics from a mixture distribution instead of a simplified version of the joint distribution that LSA assumes. Furthermore, LDA can be applied not only in unigram language models but also in n-gram models and paragraphs (Blei et al., 2003). With a solid probabilistic foundation, LDA outputs a set of document-topic distributions and a set of topic-term distributions, both provide compelling insights into the topical structure of a corpus. The Dirichlet distribution in LDA enables the high flexibility of determining the portion of words associated with a topic and the portion of topics associated with a document. Some topics can be associated with a large group of words, and some documents can be associated with a large number of topics.

2.4.2 Latent Dirichlet Allocation (LDA)

Given the disadvantage of LSA that one document can only associate with one topic, instead of using the Latent Semantic Analysis (LSA) (Deerwester et al., 1990) technique, we employ the Latent Dirichlet Allocation (LDA) (Blei et al., 2003) which represents the state of art in topic modeling research, to analyze the IS research landscape in our study. We take advantage of the probabilistic nature of LDA models to quantify topic strength for each article. Multiple topics with the topic weights will be assigned ”softly” to each paper.

In our analysis, each journal article abstract is considered one document, and all documents in the focal time-period are viewed as one corpus. The collection of all unique words in one corpus composes a dictionary. A corpus is formalized as a term-document matrix, in which each row represents one document, and each column represents one term from the vocabulary. If a word is exhibited in the document, the corresponding term-document cell in the matrix will be value 1;
otherwise, the cell value will be 0. The term-document matrix usually is large and sparse, given a real-world corpus. One family of techniques that aims at tackling the “curse of dimensionality” problem is topic-modeling methods, of which LSA and LDA are representative.

### 2.4.3 Capturing Topic Dynamics by Dynamic Topic Models (DTM)

As an extension of LDA, Blei and Lafferty (2006) proposed Dynamic Topic Models (DTM) to capture topic evolution over time. In the original LDA algorithm, documents are exchangeable in the entire corpus, i.e., the temporal ordering of documents is assumed to be irrelevant. After dividing the entire corpus into sequential slices (e.g., every five years), DTM only assumes documents in each time slice are exchangeable, and it models topic distribution progression from one time-slice to the following time-slice. A DTM model calculates keyword distributions for every topic across all time slices. Assuming the total number of topics \( K \) is known, we denote \( d \) as a document. DTM models a sequential generative process, illustrated by the graphical model representation in Figure 2.2. The model first draws the topic distribution \( \beta_{t,k} \) at the current time \( t \) conditional on the previous topic distribution \( \beta_{t-1,k} \sim N(\beta_{t-1,k}, \sigma^2 I) \). The chaining process quantifies word distribution by using \( \alpha_t | \alpha_{t-1} \sim N(\alpha_{t-1}, \delta^2 I) \) at each time period where \( \alpha \) is the mean of the logistic normal distribution that DTM model uses to draw per-document topic proportion. After defining \( \beta_t \) and \( \alpha_t \), the model draws the Dirichlet prior topic strength hyper-parameter \( \eta \sim N(\alpha_t, a^2 I) \) for each document. For each word in the document, the model draws the per-word topic assignment \( Z \sim Mult(\pi(\eta)) \) from a multinomial distribution, where \( \pi \) maps the multinomial natural parameter to the mean parameter, and then the word \( W_{t,d,n} \sim Mult(\beta_{t,z}) \). Model parameters are learned through variational methods such as variational Kalman filter and variational wavelet regression.

From the application point of view, DTM captures not only word distribution and topic structure at each time-slice, but also word distribution and topic dependency across time slices. By analyzing changes across the entire time period, topic evolution patterns can be revealed.
2.4.4 Measuring Topic Mass by Probability Mass

LDA computes a topic distribution over each document, and we use this property and apply the distribution statistics to formulate the topic mass score for each topic. Let us denote \( i \) as a topic, \( j \) as a document, \( w_{ij} \) as the posterior probability (which is effectively a weight in the topic representation) of topic \( i \) in document \( j \), \( m \) as the total number of topics, and \( n \) as the total number of documents. Each document has up to \( m \) number of topic weights. The topic mass for topic \( i \) is defined as its total probability mass across all documents:

\[
TopicMass_i = \sum_{j=1}^{n} w_{ij} \tag{2.1}
\]

Upon normalization, the topic mass for all topics should sum up to one.

\[
\sum_{i=1}^{m} NormalizedTopicMass_i = 1 \tag{2.2}
\]

In Sidorova et al. (2008), when a document has topic loading factors higher than a pre-determined threshold, an equal credit of 1 is assigned to all corresponding factors (topics). This approach has a significant disadvantage: equally assigned credit does not properly quantify the
typically non-uniform topic distribution within one document. Ideally, if a document exhibits different emphasis on different topics, the credit assignment mechanism should capture such subtlety. With our notion of topic mass grounded in probability, each document can assign different mass to different topics, and each topic can receive different credits from different documents.

2.4.5 Measuring Inter-Corpus Difference by Kullback-Leibler (KL) Divergence

In this study, we also investigate the different sub-communities within the MIS discipline by using Kullback-Leibler divergence (Kullback and Leibler, 1951). Such a problem is formulated as quantifying the topic distribution difference between two publication corpora produced by the sub-communities (e.g., the North American community and the European community).

Kullback-Leibler divergence looks at two probability distributions, and it measures to what extent one probability distribution diverges from another. Let us denote the first probability distribution as $P$, and the second probability distribution as $Q$. Kullback-Leibler divergence from $Q$ to $P$ is

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

(2.3)

, where $i$ is the element in the probability distribution.

If we indicate $P$ and $Q$ as word-level probability distributions from two corpora, we can measure the divergence between the two word probability distributions. If we indicate $P$ and $Q$ as topic-level probability distributions from two corpora, we can evaluate the topic divergence between two corpora. In this paper, we report the symmetrized version of KL divergence, which is

$$D_{KL}(P,Q) = D_{KL}(P||Q) + D_{KL}(Q||P)$$

(2.4)

2.5 Data

In this paper, we analyze abstracts of research articles published in the AIS Basket of Eight journals – MIS Quarterly (MISQ), Information Systems Research (ISR), Journal of the Management Information Systems (JMIS), Journal of AIS (JAIS), European Journal of Information Systems
(EJIS), Information Systems Journal (ISJ), Journal of Information Technology (JIT), and Journal of Strategic Information Systems (JSIS). MISQ, ISR, JMIS, and JAIS are North American-based IS journals, and EJIS, ISJ, JIT, and JSIS are European-based IS journals. The analysis period is between the years 1987 and 2019, as Sidorova et al. (2008) started their analysis with 1987 data, and the latest year for which we have complete data is 2019. There are 6,582 paper abstracts in our analysis (Table 2.2); 3,859 abstracts are from North American-based journals, and 2,723 paper abstracts are from European-based journals. Papers without an abstract are not included.

Table 2.2: Data Statistics

<table>
<thead>
<tr>
<th>Journal</th>
<th>Years</th>
<th>Articles</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>MISQ</td>
<td>1987-2019</td>
<td>1,105</td>
<td>3,859</td>
</tr>
<tr>
<td>JMIS</td>
<td>1987-2019</td>
<td>1,215</td>
<td></td>
</tr>
<tr>
<td>ISR</td>
<td>1990-2019</td>
<td>988</td>
<td></td>
</tr>
<tr>
<td>JAIS</td>
<td>2000-2019</td>
<td>551</td>
<td></td>
</tr>
<tr>
<td>JIT</td>
<td>1987-2019</td>
<td>753</td>
<td>2,723</td>
</tr>
<tr>
<td>EJIS</td>
<td>1991-2019</td>
<td>886</td>
<td></td>
</tr>
<tr>
<td>ISJ</td>
<td>1991-2019</td>
<td>586</td>
<td></td>
</tr>
<tr>
<td>JSIS</td>
<td>1991-2019</td>
<td>498</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td></td>
<td></td>
<td>6,582</td>
</tr>
</tbody>
</table>

2.6 Analysis

In preprocessing, we pruned the corpus vocabulary by removing stop words, stemming words by Porter stemmer (Porter et al., 1980), and removing terms that occurred only once. The final vocabulary size is 8,102. We use the topic coherence pipeline proposed by Röder et al. (2015) to assess the topics’ quality and interpretability. Given our goal to present meaningful topic comparisons, we analyze the identity level and moderate granular level. Particularly, we conduct our identity level analysis by setting the number of topics $K$ to 5 and granular level analysis by setting $K$ to 11 and 13. We employ the LDA technique for both levels analyses, and DTM method for topic evolution analysis, at the thirteen-topic level. Furthermore, for each analysis, we run our experiments for five times and take the averaged distribution value for each topic.
2.6.1 Topic Trend: Identity Level

The identity level is the most parsimonious approach to represent the macro-level topic structures. Therefore, we start with the five-topic analysis on the identity level using the LDA technique.

2.6.1.1 Comparison with Sidorova et al. (2008)

Realizing our dataset is much bigger, and Sidorova et al. (2008) conducted their analysis only on three journals (MISQ, ISR, JMIS), we run the LDA algorithm on a comparable dataset from the same three journals first. Summarized in Figure 2.3, the overall trends in our findings from LDA (on the left) are largely consistent with Sidorova et al. (2008) study from LSA (on the right) - IS Development is trending down while IT and Markets is trending up. However, because of the probabilistic nature of our topic mass calculation from LDA, one article can be assigned to multiple topics. The change in each topic is less harsh than the result of LSA.

![Figure 2.3: LDA (Left) vs. LSA (Right, Sidorova Data) for MISQ, ISR and JMIS, 1987-2006](image)

2.6.1.2 The Same Journals as Sidorova et al. (2008) until 2019

Next, we extend the Sidorova et al. (2008) analysis by adding data to 2019 (the full-year data available) for the same three journals and evaluate the topic trends.
In Figure 2.4, we normalize the topic mass and visualize the proportion of all topics. We see even clearer trends: IT and Markets has grown in popularity, IT and Groups has started to show its upward trend, IT and Individuals has went down, IS Development has descended steadily, and IT and Organization has remained stable. Our interpretation is IT and Groups has evolved from its founding meaning to a broader coverage. We will discuss this evolvement in detail in later sections. Many reasons can contribute to the upward trends for IT and Market and IT and Groups, such as the prevalent adoption of e-commerce and mobile devices. Additionally, these two categories can go very well together; we should not overlook the opportunities created by combining these two categories, such as live streaming businesses and the YouTube economy.

Figure 2.4: LDA for MISQ, ISR and JMIS, 1987-2019

2.6.1.3 Basket of Eight Journals until 2019

Furthermore, we extend our analysis by including all the Basket of Eight publications between the years 1987-2019. In Figure 2.5, we can see IS Development and IT and Organization have
attracted a large portion of focus from the beginning. The focus difference between IS Development and IT and Organization and the other three topics – IT and Groups, IT and Individuals, and IT and Markets – is quite significant.

However, IT and Individuals, IT and Markets, and IT and Groups have grown steadily over time, likely due to the increasing popularity of e-commerce and mobile commerce. Notably, it is interesting that the research interest in IT and Groups has increased steadily.

One way to explain this is to consider the changing definition of groups. While the group in IS research was once limited to workgroups, its context has extended from colleagues and business partners only to both real and virtual friends. Group interactions are now part of a much extensive network – the Internet. Another interesting finding when time gets closer to 2019 is the sign of convergence becomes much stronger and more apparent. One way to explain this is that MIS researchers have paid more attention to aspects with higher technological success. In the meantime, the interdisciplinary nature of our discipline has been acknowledged and appreciated.

Figure 2.5: LDA for Basket of Eight Journals, 1987-2019
2.6.2 Topic Trend: Granular Level

A reasonable level of granularity can avoid being buried in too many details and facilitate to make meaningful observations (Goyal et al., 2018). Therefore, for comparison purposes, we conduct the eleven-topic and thirteen-topic analysis on all Basket of Eight Journals to compare with Goyal et al. (2018) and Sidorova et al. (2008), respectively, in this section.

2.6.2.1 Longitudinal Analysis

We contribute to show an all basket of eight journals longitudinal topic study on the reasonable granularity level in this section. Goyal et al. (2018) demonstrated a longitudinal study of North American-based journals. We use their study as a benchmark (and use the same topic number for comparison) and show the difference between their result and ours. We show the adopted Goyal et al. (2018) graph in Figure 2.6, our analysis trend in Figure 2.7, and the topic distribution comparison in Figure 2.8.

Figure 2.6: Eleven-topics LSA for ISR, MISQ, JAIS, JMIS, 2000-2016 (adopted from Goyal et al. (2018))
Although we both analyze the 11-topics level, the difference between the two studies is quite striking. It also shows why our study is needed. The difference can come from several aspects: 1) the sample journals are different. While Goyal et al. (2018) conducted their analysis on four North American-based journals, we also consider four European-based journals. Therefore, the topics that European-based journals were interested in will be included in our study. 2) the sample periods are different. While Goyal et al. (2018) used a sample from 2000 – 2016, we include a longer period in our study from 1987 – 2019. 3) the techniques are different. As we mentioned before, LSA, which was employed by Goyal et al. (2018), restricts one topic to one article. And LDA, which was employed by our study, released this assumption, and assign multiple topics to one article. 4) intervals on graphs are different. Goyal et al. (2018) used a three-year rolling average, and we used a five-year interval (three years in the last period) distributions in our study.

In Goyal et al. (2018), the identified research themes (from the most number of articles to the least) are: knowledge management, technology adoption, e-commerce, recommender systems, security, virtual world, healthcare IT, trust, outsourcing, auctions, and privacy. Among these topics, we have two topics in common, two topics are similar, and the rest topics are different.
First, both Goyal et al. (2018) and our study have “technology adoption” and “IT outsourcing” in the returned topics. Second, based on our topic keywords, we use the term “online communication” instead of “virtual world” as one topic name. Third, also given our topic keywords, we recognize “security and privacy” as one topic while Goyal et al. (2018) recognize them as two separate topics. The rest of the topics are different.

While Knowledge Management is the topic with the highest number of articles in Goyal et al. (2018), IS Discipline Development has the most top topic distribution in our study. IS Discipline Development is not in the list of Goyal et al. (2018). The inclusion of earlier years publications and European-based journals can contribute to this difference. The closest topic to Knowledge Management in our study is Project and Risk Management, our 2nd highest distribution topic.
However, it dropped rapidly from top to bottom on our topic distribution list. The observation on IT Innovation and Adoption is largely consistent in both of our study and Goyal et al. (2018). The topic had a peak around 2006 and showed a slight decline afterward.

The rest of the topics in our study are IT and Individuals, IT and Organization, IT and Markets, IS Development, IT outsourcing, Decision Support System, Virtual Communication, and Security and Privacy. We can also see the sign of topic distribution convergence here, the same as the five-topic analysis in the previous section. The same as Project and Risk Management, Decision Support System, and IS Development are on the decline. Among the rest topics, IT and Organization, IT and Individuals and IT Outsourcing are relatively stable. The Outsourcing topic trend in Goyal et al. (2018) is largely consistent with ours.

We have three topics that have apparent upward trends: IT and Markets, Virtual Communication, and Security and Privacy. Comparing with Goyal et al. (2018), we believe their E-commerce, Recommender System, and Auctions can be considered as subareas of our IT and Markets. However, in Goyal et al. (2018), e-commerce shows a declining trend, which is opposite to our observation. Virtual Communication and Security and Privacy show a rising trend in both of our study and Goyal et al. (2018), while in our study, the trend is more visible.

As we pointed out earlier, the difference between our study and Goyal et al. (2018) can come from multiple sources. However, we can at least say that European-based journals contributed a lot to the discrepant topics to make them in the top eleven topics. They include IS Discipline Development, Project and Risk Management, IS Development, Decision Support System, IT and Individuals, IT and Organizations, and IT and Markets. Furthermore, because we included a longer time period, especially the beginning period of our discipline, our analysis results can better mark topics fading away and topics that are skyrocketing.

2.6.2.2 Emergent Topics

After conducting the longitudinal analysis, we shift gears to identify emergent topics in recent years. Sidorova et al. (2008) served as an excellent baseline to conduct such an analysis. Therefore,
we employ the analysis end year in Sidorova et al. (2008), 2006, as the cutoff time and discover what topics gained MIS researcher attention afterward. To balance meaningfulness and richness, we see through the lens of a thirteen-topic granular level consistent with Sidorova et al. (2008).

Figure 2.9 shows the top thirteen-topic distributions pre- and post- 2006 for all basket of eight journals. The darker color represents the top thirteen-topic before 2006, and the lighter gray color represents the top thirteen-topic after 2006. First, we see dominant roles for IS Development, IS Management, IT and Individuals, and IS Discipline Development before 2006 and IT and Organization, IT Adoption and Use, Group Support System, and Project and Risk Management follow after. Decision Support System, IT Outsourcing, IT and Markets, Online Auction, and Online Market are down the list.

The new topics we can see in the lighter gray color are Social Media, Security and Privacy, Open Source Software, and Virtual Collaboration. Note that, even in the thirteen-topic solution, some aspects of the original five-topic solution are still represented in both time periods: IS Development, IT and Individuals, IT and Organization, and IT and Markets. Group Support System and Decision Support System are not presented in the thirteen-topics after 2006. We briefly introduce the new topics emerging since 2006 (non-overlapping lighter color).

Social Media. Social media is now in all aspects of our personal lives. The term “word-of-mouth” can be applied in both personal and business contexts. Companies use social media to improve their business services (Zhang et al., 2012; Lu et al., 2013) and make recommendations (Jabr and Zheng, 2013). Social media has become a vehicle that is used by both customers and companies. The sharing nature of social media made the information flow became more manageable, and it can be created anywhere. Given its ubiquitous characteristic, it is not surprising that MIS researchers have increased their research focus on social media.

Security and Privacy. The Web provides two sides for users. On the one hand, it provides convenience, but on the other hand, it brings security concerns and issues. The virtual threats on the Web can have significant harm to individuals and companies. Therefore, MIS researchers started to pay more attention to these security and privacy issues and factors (Wang et al., 2013).
Figure 2.9: Thirteen-topic Distribution of All Journals 1987-2006 vs. 2007-2019
Open Source Software. The open-source community has been growing, not only in terms of the number of open-source projects but also the collaboration among contributors (Faraj et al., 2015). The network effect (Singh et al., 2011) is identified in the open-source community. As a form of web-based crowd-sourcing technology, MIS researchers have evaluated open-source software services and innovations (Orlikowski and Scott, 2015; Barrett et al., 2015; Lusch and Nambisan, 2015).

Virtual Collaboration. Because of the rapid spread of the World Wide Web (WWW), traditional face-to-face communication methods no longer satisfy business and individual needs. Virtual collaboration (George et al., 2008) and communications have become widely adopted. In business circumstances, inter-organizational and cross-organizational collaborations started to rely more on virtual technologies, and knowledge coordination has an impact on virtual team performance (Kanawattanachai and Yoo, 2007). Individuals obtained greater freedom to collaborate with people virtually. The contribution behavior in virtual communities (Tsai and Baggozzi, 2014) is different than in the real world. In the virtual world, barriers between people are diminishing.

2.6.3 Topic Trend: North American-based Journals vs. European-based Journals

Based on our observation in the previous section, the inclusion of five new journals (four of which are predominantly European) into our data has resulted in some unobservable discoveries than previous studies. Publications from all basket of eight journals allow us to investigate the potential differences between sub-communities in MIS discipline. To illustrate this opportunity, we examine the differences in topic focus between the four North American-based and the four European-based IS journals in this section. We expand the comparison by looking at the differences pre-and post-Bologna processes. The Bologna process refers to a decades’ long effort to standardize higher education across the European Union (COMM, 2015). The implementation of these efforts has also affected scholarship: more European scholars have moved to publish in English and to target highly ranked international journals for publishing their research (Borghans and Cörvers, 2009). Although we aware many time stamps and reasons can be used to demonstrate the comparison, we
use the Bologna process, a well-reasoned event, to illustrate the pre- and post- topic distribution comparison. The process officially began in 1998 with the Sorbonne Declaration, so we chose the end of 2002 and the beginning of 2003 as our dividing line for pre- and post- the process, reasoning that any effects of the Bologna process would not be noticeable for at least four years.

2.6.3.1 Identity Level

We first investigate the identity level difference between the two MIS communities and show it in Figure 2.10. Using the identity level categories, we find all five topics within North American-based journal publications. However, we can only see two out of five topics within European-based publications in pre-Bologna process periods: IT Organization and IS Development.

However, we can identify a significant change afterward. We found the two MIS communities share all the five identity level research topics after 2003. IT and Individuals, IT and Markets, and IT and Groups became part of mainstream research topics in the European-based journals, and the research topics interested by North American-based journals and European-based journals on the identity level become the same.

![Figure 2.10: Five-topic Analysis for North American-based vs. European-based Journals](image)

2.6.3.2 Granular Level

The research topic convergence on identity level led to an exploration of the result on the granular level. We continue to use thirteen-topics to identify the effects of the Bologna process on North American-based journals and European-based journals.
Figure 2.11: North American-based Journals: Pre-Bologna vs. Post-Bologna Process
Figure 2.11 summarizes the shift of thirteen-topic in North American-based journals, from the pre-Bologna period to the post-Bologna period. We witness the rise of emerging topics such as Online Review, Social Media, Virtual Collaboration, IT and Healthcare, IT Outsourcing, and Security and Privacy, in North American-based journals.

Online Review. Online review as an information pool has been used in MIS research. Facilitated by online reviews, questions such as whether online markets will overtake brick-and-mortar stores and how online and offline platforms can be integrated generated a lot of research attention (Fang et al., 2015; Ghose et al., 2013; Hoehle and Venkatesh, 2015; Steinbart et al., 2016). On the consumer side, information from the web has also relied heavily. Businesses, therefore, seek to acquire more customers using online platforms. Examples such as one-on-one marketing and search engine optimization (SEO) (Dou et al., 2010) are widely used. Another platform that relies heavily on online reviews is the sharing economy. Researchers studied customer safety concerns from Uber (Greenwood and Wattal, 2017), competition between Airbnb providers and local hotels (Zervas et al., 2017), and a relationship of HIV trends and the usage of the online dating section on Craigslist (Chan and Ghose, 2014).

IT and Healthcare. IT and Healthcare did not receive much attention ten years ago. However, researchers have realized that a comprehensive and efficient Information Management System can make exceptional contributions to human lives and society (Yan and Tan, 2014; Bardhan et al., 2015). From our analysis result, it has the same scale as Online Review, Social Media, and Virtual Collaboration. We have also observed businesses have combined online platforms with healthcare, such as answering patient questions and distance appointments. Powered by more advanced technologies for IT and Healthcare, we will keep an eye on this topic as well.

Figure 2.12 summarizes the shift of thirteen-topic structure in European-based journals, for pre-Bologna and post-Bologna processes. Like previous results, we observe the topic surges on IT and Healthcare, Social Media, Virtual Collaboration, and Security and Privacy, which are also observed in the North American-based journals. Among the emergent topics for the European-based journals, IS Discipline Development, IT and Individuals, IS Management, and IT Innovation
Figure 2.12: European-based Journals: Pre-Bologna vs. Post-Bologna Process
and Adoption, are topics that existed in North American-based journals pre-Bologna Process. It signals an increasing share of similar research interests between the two communities.

Figure 2.13: Topic Comparison: European-based vs. North American-based Journals, and Pre-Bologna Process vs. Post-Bologna Process

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Group Support System</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Knowledge Management</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Project and Risk Management</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>EDI and Interorganizational Systems</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT Outsourcing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IS Case Study</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT and Markets</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Business Process Development</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Value of IT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IS Planning</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT and Organization</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Decision Support System</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IS Development</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IS Discipline Development</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT and Healthcare</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Social Media</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT and Individuals</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IS Management</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Virtual Collaboration</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT Innovation and Adoption</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Security and Privacy</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HR Issues in IS</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measurement and Instrument</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Online Review</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2.13 integrated the topics of North American-based and European-based journals on pre-Bologna and post-Bologna periods. Between 1987 and 2002, we found an overlapping interest for seven topics between the two groups of journals, while the North American community portfolio of topics included six topics not considered by the European community. Similarly, for this period, the European community worked on six topics not reported in the North American journals. Beginning
in 2003, the two communities had 12 topics in common, while each community had one topic that unique to its journals. The difference between the two communities after 2002 is much lessened.

2.6.3.3 Kullback-Leibler (KL) Divergence

To further quantify the difference between North American-based journals and European-based journals for pre-Bologna Process and post-Bologna Process, we computed the KL divergence between the two corpora, considering both topic distribution (computed on thirteen-topic) and word distribution. Figure 2.14 shows the shrinkage of both topic distribution divergence and linguistic (word) distribution divergence. It appears that the two communities are experiencing intellectual convergence.

![Figure 2.14: KL Divergences: Pre-Bologna Process vs. Post-Bologna Process, for North American-based Journals and European-based Journals](image)

2.6.4 Topic Morphing

As of now, our analysis and the previous studies (Sidorova et al., 2008; Goyal et al., 2018) are premised on the assumption that despite the topics’ ebb and flow, each topic has a coherent and stable internal semantic. However, the semantic of one topic can change, given technology
advancement. Therefore, it is reasonable to offer two types of topic morphing processes: cross-topic morphing and inter-topic morphing. The cross-topic morphing is one topic that has transformed into a new form and become a new topic empowered by advanced technology. Inter-topic morphing describes the topic concentration shift process of a particular topic, also supported by advanced technology.

In this section, we share our observations of topic-morphing phenomena using Dynamic Topic Models (DTM) on all Basket of Eight journals. In principle, the DTM algorithm is equipped to model such intertwined processes in large corpora. Though our dataset may not have a sufficient size on par with the capacity of DTM, running the algorithm with thirteen-topic still revealed some very interesting patterns. For instance, we can demonstrate how the focus of some topics has shifted, as exhibited in the change of word distributions. Since morphing does not reveal itself for each topic in our dataset, we picked four morphing examples with the most observable patterns. Moreover, each topic can have both generic terms whose probabilistic trends are stable and specialized terms have probability trends change dramatically over time. We choose the terms for elucidation are from the top ten keywords and have the most observable trends. Unlike the previous longitudinal analysis graphs that each line represents one topic, in here, each line represents a term (a word stem, to be precise) temporal trend.

2.6.4.1 Cross-topic Morphing Example 1: Group Support System → Virtual Collaboration

Figure 2.15 visualizes the cross-topic morphing of Group Support System to Virtual Collaboration, from their keywords probability shifts. We can see that representable words for Group Support System, such as group and support, fell out of favor. The representable terms for Virtual Collaboration, such as social and communication, have gained popularity. The completely faded keywords group and support made Group Support System morphed into a completely new topic, Virtual Collaboration. Virtual Collaboration started with sharing similar keywords with Group Support System. However, the context has completely changed and needs to be adapted to societal
and technological advancements. The easy accessibilities of the Internet, the Web, online platforms and mobile applications enable different groups of people to work together effortlessly. Rather than being limited by physical constraints and existing business relationships, Virtual Collaboration extended the boundaries to various types of environments and groups, with goals that range from social to political to affiliations based on shared interests and activities.

![Figure 2.15: Cross-topic Morphing Example 1: Group Support System → Virtual Collaboration](image)

Traditional technology can only support group communications within limited business partners. However, because of the availability and adoption of personal computers, group communications are not constrained by limited business partners, and the capacity for group communications has enlarged tremendously. The coverage of IT and Group has shifted entirely from a narrowed form of business groups to a much broader population. The audience of group communications has been enlarged to anyone with Internet access. Therefore, IT and Groups topic has to morph, given technology advancements. Meantime, the advancement of the Internet makes Virtual Communication in shape. Virtual Communication is a product of IT and Groups, digital device adoption, and Internet technology.
2.6.4.2 Cross-topic Morphing Example 2: Network Communication → Social Media

Figure 2.16 visualizes the cross-topic morphing of Network Communication to Social Media. The fast development of digital and mobile technologies has changed how people communicate. Online channels, such as social media, are vastly utilized by companies and individuals. The keyword network had a leading weight when network facilities were the main concerns for companies. As more companies have stable physical networks in place, how to use it to communicate better became more relevant. Companies realize that networks can conduct businesses and acquire new customers. Individuals understand networks can be used to interact with others. The touching points of networks are on the social front. Therefore, the rising of keyword social is not surprising anymore. Traditional company-network communications, which are limited to a small, reachable range, have naturally disappeared because of the maturity of much broader scale networking technologies. The social interaction approaches it created also contributed to society and research interest shifts.

![Figure 2.16: Cross-topic Morphing Example 2: Network Communication → Social Media](image)

With a constrained network bandwidth, network communications can only happen from limited resources. Therefore, it will be used first on the most critical societal factors, such as business and
security. However, after the advancement in mobile devices and networks, such as 1G to 5G, lower tiers of needs, such as social activity and entertainment, are in line to be served. Social media can be used in both business and personal contexts and fulfill both critical business needs and casual personal requirements. Given the mobile and network technology advancement, Network Communication has to morph. Because of the empowerment from technology, Social Network became its next form. In our example, we see the weight of the word social increased significantly as it has become the first touching point between people. We also observe the weight of the word network declined a lot because it is not the main barrier for most people anymore. Social Network is a product of Network Communication, mobile device adoption, and advanced network technology.

2.6.4.3 Inter-topic Morphing Example 1: Value of IT

Figure 2.17 presents the inter-topic keywords trajectory of the Value of IT. We can observe the keywords price and cost show descending trends, especially in the 21st century. The facility costs and expenses were the primary concern when computers were first introduced to businesses and individuals. However, after machines and IT facilities were adopted, the focus has moved away from their costs and prices. The topic focus has progressed beyond the initial anxiety over the essential business value of IT, with a focus on IT costs and return on investment, to a focus on the firm’s customers and the interactions with consumers to improve their products and customer satisfaction. Therefore, the trending up keywords consumer and product reinforced this change. The narrow primary focus of the Value of IT has shifted to a broader focus on the economics of IT and how to outperform competitors by maximizing the firm’s IT capabilities.

The inter-topic morphing examples also show the critical role of technology advancement. The value of IT is not shown from its cost and price, rather its product. We can see it from the declining weights of words cost and price and the increasing weight of the word product. Moreover, while consumers are aware of the change, their products’ expectations became higher with the available technologies. Therefore, we observe the context for the Value of IT has morphed within the topic.
Although the topic name remains the same, the focus has shifted. The new focus of IT’s Value is a product of its previous form and focus and advanced technologies and adoptions.

### 2.6.4.4 Inter-topic Morphing Example 2: IT and Healthcare

Figure 2.18 presents the keywords trajectory for IT and Healthcare. Healthcare awareness and technology innovation have made the healthcare industry and technology industry good partners in creating premium societal and business benefits. The keywords trajectory for health, care, and healthcare reflect the consideration shift. Furthermore, the participation of MIS research in IT and Healthcare has not limited to healthcare in general or on the conceptual level; it has penetrated to those practical practices, such as hospital applications, patient treatments, and medicine. This sign of increasing practical relevance from IT and Healthcare research is revealed here.

We see the increased practical relevance of MIS research in this example. While IT and Healthcare has existed as an MIS research topic for many years, empowered by recent technology advances, such as machine learning and artificial intelligence, healthcare industry shows a high demand for advanced system and technology to support their care needs. We see the increasing word weights on hospital, patient, and medicine. They are critical components in the practitioner’s daily practice.
Therefore, IT and Healthcare started to find its stand on both practical relevance and business, even on medicine, while maintaining its healthcare core stream. The role of technology advancement is inevitable for the morphing of IT and Healthcare. The new focus of IT and Healthcare is a product of its old focus, technology advancement, and practical relevance.

### 2.7 Discussion

MIS research positioning has been a topic since the beginning of our discipline (Ives et al., 1980). It has gone through the debate on having a hard boundary (Benbasat and Zmud, 2003) or seeing the “thousand flowers” to bloom (DeSanctis, 2003). As time goes to the 2nd ten years of the 21st century, computers are widely available, and computer systems are widely adopted. Management Information Systems (MIS) is recognized as an essential component in modern organizations. MIS serves as the technical foundation and supports the entire firm’s operation. It is not surprising that MIS research has intersections with other domains, such as marketing, finance, security, and healthcare.

With widely accessible computers, information systems are everywhere and support everyone’s daily activities. MIS has penetrated to every detail of modern society. MIS research topics can
cover all business areas supported by information systems, obviously or inevitably. Today, it is not hard to find a business scenario that has an MIS component in it. The increased number of business applications that employ MIS makes academic MIS research have a much bigger playground. However, we observe research topics distribution became more balanced, in both the identity level analysis in Section 2.6.1 and the granular level analysis in Section 2.6.2. There is a shift between the previous unpopular topics group and the popular topics group.

Technology advancement played a critical role. On the identity level, the weights of technology in IT and Groups, IT and Individuals, and IT and Markets increased dramatically in the last two decades. On the contrary, while primary information systems are implemented in most organizations, the research interest and space for IS Development and IT and Organization declined. On the granular level, Virtual Communication, Security and Privacy, and IT and Markets are areas in need of reliable technology supports. Their signs of progress are highly dependent on technology advancement. Without advanced technologies, their movements will be constrained. Therefore, we observe their growth from technology empowerment. In contrast, Project Management, as a topic that does not require highly dependent on advanced technology, has fallen out of interest by many MIS researchers.

In addition, we identified emergent topics in Section 2.6.2.2 and Section 2.6.3.2. The emergent topics, such as Social Media and Online Review, also show the critical role of advanced technology. Without systematic support from technologies, they will not be invented. Without continuous support from advanced technologies, they will not become widely available and further become new norms of modern activities. In addition, as advanced technologies become more reliable, the uses and interests of IT and Healthcare grow significantly. Without technology, hospitals, patients, and medicine are only the vocabulary limited within a small group of people. With the empowerment from advanced technology, virtually accessible physicians and distancing operations are possible for people in need.

With the supports from advanced technology, domains can be integrated with MIS becomes broader, and the content coverage of MIS discipline becomes wider. In the meantime, the prosperous
application areas supported by advanced technology are still limited. Therefore, research topics are constrained by technology advancement boundaries. As a result, we observe a topic convergence trend.

In addition, the advancement of technology and topic convergence together push MIS research to evolve. We provide one plausible answer to Goyal et al. (2018) on why specific topics are enduring, and certain topics had a limited lifespan. The peaked and troughed topics only related to a specific context, rather than the core MIS phenomena, as mentioned by Goyal et al. (2018). More importantly and one step further, technology advancements contribute the most in their lifespans. We identified two types of topic morphing behaviors, cross-topic morphing and inter-topic morphing, and their examples in Section 2.6.4.

With the empowerment of advanced technology, cross-topic morphing describes one topic’s essential change, and inter-topic morphing demonstrates the topic focus change. Without technology advancement, Group Support System to Virtual Collaboration and Network Communication to Social Media would not be possible. Without technology advancement, the Value of IT would remain emphasis on cost, and IT and Healthcare usage would continue to be limited.

Therefore, topic convergence and topic morphing created a reinforcement loop over each other, with advanced technology empowerment. Technology advancements will accelerate topic morphing processes when the integration of technology and the topic is plausible. Meanwhile, technological progress will expedite topic convergence if successful use cases of advanced technologies are still limited in business domains.

2.8 Conclusion

Over the past 30 years, MIS research topics have ebbed and flowed over time, along with technology advancement. The emergence of new research topics in the last decade and a half, most of which have been focused on the internet- and mobility-based systems, illustrate how the research interests of the field have reflected the changing role of information systems in business and society. At one time in the field’s history, academic information systems researchers were limited
to studying computing in business organizations, because that is where the computers were. With the ubiquity of computing, our research territory has expanded considerably, and our analyses show that researchers have reacted accordingly. Analyses such as ours demonstrate that our field is not stagnant, that it is growing and shifting to encompass new realities, and that it has done so throughout its history.

In this work, we identified that our discipline has shown maturity from topic convergence and topic morphing and described the reinforcement loop between the two. We analyzed research topic trends from different levels and demonstrated emergent topics. Finally, we identified two topic morphing behaviors, cross-topic morphing, and inter-topic morphing, and illustrated their corresponding examples. On the one hand, topic convergence is a product of topic morphing and technology advancement. On the other hand, topic convergence pushes topics to morph, with the empowerment from advanced technologies.

Like all studies, this one has its limitations. First, although seemingly large, our dataset is only modest in size for the analysis techniques we used. Therefore, the data set may not fully exploit the power of these high-capacity topic modeling algorithms. One should be cautious in extrapolating our findings. Second, the labeling of topics, a manual process, is considerably influenced by Sidorova et al. (2008)’s schema and possible annotator bias. Third, following Sidorova et al. (2008), our dataset is comprised of article abstracts alone. However, it is arguable that including the full text of research papers would enrich the data and could lead to more interesting and nuanced discoveries.

2.9 References


### 2.10 Appendix: LDA Results

#### Table 2.3: LDA for Three North American-based Journals (MISQ, ISR and JMIS), Un-normalized Topic Mass, 1987-2019

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IT and Groups</td>
<td>31.54</td>
<td>41.37</td>
<td>46.24</td>
<td>55.68</td>
<td>79.14</td>
<td>127.02</td>
<td>94.70</td>
</tr>
<tr>
<td>IT and Market</td>
<td>15.41</td>
<td>21.85</td>
<td>36.12</td>
<td>64.58</td>
<td>104.43</td>
<td>164.90</td>
<td>120.76</td>
</tr>
<tr>
<td>IT and Organization</td>
<td>60.89</td>
<td>78.63</td>
<td>90.89</td>
<td>93.95</td>
<td>125.57</td>
<td>157.19</td>
<td>87.72</td>
</tr>
<tr>
<td>IT and Individual</td>
<td>86.59</td>
<td>113.01</td>
<td>94.55</td>
<td>104.27</td>
<td>123.66</td>
<td>148.22</td>
<td>74.66</td>
</tr>
<tr>
<td>IS Development</td>
<td>107.46</td>
<td>120.99</td>
<td>117.14</td>
<td>120.44</td>
<td>133.62</td>
<td>159.54</td>
<td>88.19</td>
</tr>
</tbody>
</table>

#### Table 2.4: LDA for Basket of Eight Journals, Un-normalized Topic Mass, 1987-2019

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IT and Groups</td>
<td>71.05</td>
<td>127.68</td>
<td>160.35</td>
<td>195.29</td>
<td>310.44</td>
<td>364.58</td>
<td>255.40</td>
</tr>
<tr>
<td>IT and Market</td>
<td>46.70</td>
<td>96.81</td>
<td>139.61</td>
<td>192.96</td>
<td>290.12</td>
<td>365.08</td>
<td>268.03</td>
</tr>
<tr>
<td>IT and Organization</td>
<td>162.66</td>
<td>354.02</td>
<td>364.14</td>
<td>371.54</td>
<td>470.48</td>
<td>436.92</td>
<td>245.23</td>
</tr>
<tr>
<td>IT and Individual</td>
<td>59.39</td>
<td>114.40</td>
<td>122.64</td>
<td>203.77</td>
<td>318.00</td>
<td>386.23</td>
<td>267.91</td>
</tr>
<tr>
<td>IS Development</td>
<td>246.82</td>
<td>397.84</td>
<td>317.53</td>
<td>347.28</td>
<td>434.97</td>
<td>405.72</td>
<td>255.83</td>
</tr>
</tbody>
</table>
3.1 Abstract

In this paper, we show that mining firm-related events in public news can effectively predict various firm financial ratios. By exploiting state-of-the-art neural architectures including pseudo-event embeddings, Long Short-Term Memory Networks, and attention mechanisms, our news-powered deep learning models are shown to outperform standard econometric models operating on precise historical accounting data. We also observe forecasting quality improvement in multi-task learning settings, i.e., when multiple financial ratios are predicted simultaneously. Our forecasting models (and byproducts such as attention maps and firm embeddings) benefit various stakeholders with not only quality predictions but also explainable insights. Our proposed models can be applied when accounting data is not available or available.

Keywords: natural language processing, deep learning, forecasting

3.2 Introduction

While public companies’ financial performance is anchored in their stock returns, it is vital to evaluate a firm’s financial status through a wide spectrum of quantitative measures, e.g., profitability, efficiency, and solvency, many of them available in the form of financial ratios. Why? Such ratios are concrete measures of firm fundamentals, hence tangible basis of decisions for many mar-
ket participants, particularly the long-term-minded investors with low trading frequency. There are other important stakeholders, e.g., credit rating agencies and insurance companies, eyeing firms’ financial pictures much broader than stock returns. Outside the stock market, private companies also participate in the economy and undergo financial scrutiny.

A company’s financial standing is complex and dynamic, and the sheer amount of information for stakeholders to acquire digest is overwhelming. Although the information overload problem is commonly acknowledged in finance and much broader settings (Hemp, 2009), few techniques are readily available to help less sophisticated Main Street stakeholders overcome this challenge. Given the information asymmetry, many people and agencies seek insights from financial reports that disclose various accounting and financial ratios.

However, significant disadvantages exist of heavy reliance on (numerical) accounting data in financial reports:

- Public companies’ financial reports are released in a retrospective fashion, i.e., they are released after the fact and usually have months of delay. Investors can only “postprocess” the data after the quarter or year end. Hence, it is impossible to gain timely information, or more preferably, forward-looking insight, from accounting data in financial reports. While institutional investors and researchers are used to building various econometric models based on quantitative accounting and financial data in financial disclosures, Main Street investors often do not have the sophistication to either pursue such modeling effort or make intuitive sense of the model output. Even when they rely on professional help, the decision process lacks transparency and accountability.

- Private companies are not legally required to disclose their financial data. Investors of those companies have to seek other information sources, for example, insider or public news, to gain insight.

Hence we ask these questions: 1) Can stakeholders acquire intelligence before the release of financial reports, or when accounting data is not available at all? 2) If accounting data is available, can another source of data, such as news, bring in additional value?
We propose to mine such intelligence from public news. There are many possible advantages. Firstly, news articles are frequently published online, and mostly free. Main Street investors have easy access to the same information as other advanced investors do. Secondly, news are often published immediately after an event has taken place, so investors can capture their needed information in a timely manner. Thirdly, readers can entertain alternative views on the same event, as news coverage is often accompanied by expert opinions or community reactions. This is considered important for Main Street investors since they don’t often have extensive expert support as advanced investors do. Fourthly, if investors systematically follow a company’s news, trends are likely to emerge and lead to more educated decisions. Lastly, when accounting data is not available, news is naturally considered an essential source of up-to-date information for stakeholders constantly looking for financial opportunities.

Given the advantages of public news as a powerful source of financial intelligence, we therefore aim to automate the news-powered, full-fledged forecasting of firm long-term financial performance, by exploiting state-of-the-art natural language processing and machine learning capabilities. Specifically, we expect to make contributions on both methodological and application fronts:

- Novel, long-horizon (e.g. a year) models that help not only institutional investors but also Main Street investors forecast firms’ long-term financial performance.
- Low-barrier models that only exploit easily accessible information in public news to provide insights for all.
- Enhanced transparency and accountability in financial decision making, through explainability provided by model artifacts.
- Low-latency, forward-looking intelligence to help investors make almost-real-time investment decisions, rooted in timeliness of news data.
- Comprehensive financial overview of firms through full-fledged, simultaneous forecasting of multiple financial ratios.
3.3 Related Work

In this section, we review related literature in finance (stock market and firm financial ratios) and modeling techniques, respectively.

3.3.1 Finance Literature

**Stock market**  A public firm’s stock market performance is often the first indicator of its overall financial health. Therefore, numerous studies have been dedicated to the understanding or prediction of stock markets. Due to the vast body of literature in this area, we only review those directly relevant to our research, i.e., studies based on publicly available non-numerical information. A representative set of *theory testing* studies in this area are summarized in Table 3.1. Fueled by recent advancements in Artificial Intelligence (AI), researchers have extensively studied how to *predict* stock market behavior by using *machine learning models*. Table 3.2 summarizes a representative set of such studies.

**Firm financial ratios**  Besides stock returns, there exist other quantitative measures of a company’s financial health, many of which form various financial ratios. A comprehensive taxonomy of firm financial ratios is defined by Wharton Research Data Services (WRDS)\(^1\), and it constitutes the focal phenomenon of our study. A set of representative *theory testing* studies that involve financial ratios are summarized in Table 3.3\(^2\).

**Textual data and machine learning**  Textual data have gained popularity in economics research (Loughran and McDonald, 2016; Gentzkow et al., 2019). Several recent finance studies exploited textual data to investigate firm organization (Hoberg and Phillips, 2018), corporate culture (Li et al., 2018), corporate innovation (Bellstam et al., 2017), and merger and acquisition (Routledge et al., 2016), accompanied by machine learning techniques of varying capacity.

---


\(^2\)We italicized the dependent variable or independent variable when it is either a financial ratio or some variations (e.g., the nominator or denominator).
Table 3.1: Stock Market: Theory Testing Literature

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Input genre</th>
<th>Primary independent variables</th>
<th>Methodology</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock return</td>
<td>public news</td>
<td>anomaly variables,</td>
<td>linear regression</td>
<td>(Engelberg et al., 2018)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>information-day dummy variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stock return,</td>
<td>public news</td>
<td>news sentiment</td>
<td>vector auto-regressive (VAR) model</td>
<td>(Tetlock, 2007)</td>
</tr>
<tr>
<td>market volume</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earning, stock</td>
<td>public news</td>
<td>news sentiment</td>
<td>linear regression</td>
<td>(Tetlock et al., 2008)</td>
</tr>
<tr>
<td>return</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stock return</td>
<td>public news</td>
<td>media coverage</td>
<td>CAPM, F-F 3-factor model, Carhart 4-factor model, Pastor-Stambaugh liquidity model</td>
<td>(Fang and Peress, 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stock return</td>
<td>public news</td>
<td>news sentiment (fraction of positive and negative words)</td>
<td>time series regression, GARCH model</td>
<td>(Garcia, 2013)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal trading</td>
<td>earning announcement</td>
<td>capital gain overhang, unexpected earnings, idiosyncratic return volatility</td>
<td>Cox proportional hazard rate model, Fama-French-momentum 4-factor model</td>
<td>(Weisbrod, 2018)</td>
</tr>
<tr>
<td>volume, abnormal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>return</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stock price,</td>
<td>message board</td>
<td>message sentiment, disagreement, message volume, trading volume</td>
<td>classifier ensembles, linear regression</td>
<td>(Das and Chen, 2007)</td>
</tr>
<tr>
<td>return, volatility</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trading volume,</td>
<td>message board</td>
<td>number of messages, message sentiment and the agreement index</td>
<td>naive bayes classifier, linear regression</td>
<td>(Antweiler and Frank, 2004)</td>
</tr>
<tr>
<td>volatility</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notably, economics-rooted studies are still mostly driven by theory testing, viewing AI-based tools merely as modeling aids instead of centerpieces. We believe the community should diversify its methodology portfolio by embracing forecasting frameworks based on machine learning backbones. This study aims at fully unleashing the power of big textual data through state-of-the-art deep learning models.
Table 3.2: Stock Market: Prediction-related Literature

<table>
<thead>
<tr>
<th>Output variable</th>
<th>Input genre</th>
<th>Input variables</th>
<th>Method</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>stock price, price movement, return</td>
<td>public news</td>
<td>bag-of-words features, noun phrases, named entities</td>
<td>support vector machine (SVM)</td>
<td>(Schumaker and Chen, 2009)</td>
</tr>
<tr>
<td>stock price movement</td>
<td>public news</td>
<td>event extraction, event embedding</td>
<td>convolutional neural network (CNN)</td>
<td>(Ding et al., 2015)</td>
</tr>
<tr>
<td>stock price movement</td>
<td>press release</td>
<td>bag-of-words features</td>
<td>Rocchio, k Nearest Neighbors (kNN), linear SVM, non-linear SVM</td>
<td>(Mittermayer and Knolmayer, 2006)</td>
</tr>
<tr>
<td>stock price movement</td>
<td>press release</td>
<td>bag-of-words features</td>
<td>support vector machine (SVM)</td>
<td>(Luss and Aspremont, 2015)</td>
</tr>
<tr>
<td>stock price volatility</td>
<td>earnings calls</td>
<td>unigrams, bigrams, named entity, part-of-speech features</td>
<td>semiparametric Gaussian copula regression</td>
<td>(Wang and Hua, 2014)</td>
</tr>
<tr>
<td>stock price movement</td>
<td>social media</td>
<td>word embedding</td>
<td>generative recurrent neural network (RNN)</td>
<td>(Xu and Cohen, 2018)</td>
</tr>
<tr>
<td>stock price volatility</td>
<td>time series</td>
<td>historical stock price</td>
<td>generative neural stochastic model</td>
<td>(Luo et al., 2018)</td>
</tr>
</tbody>
</table>

3.3.2 Related Modeling Techniques

3.3.2.1 Time Series Models

Since corporate events have temporal component, time series model is the first related model family.

**AutoRegressive Integrated Moving Average (ARIMA) Models** The ARIMA model\(^3\) (Pankratz, 1983) is the *de facto* standard for univariate time series analysis in econometrics, hence a natural baseline model for our study. ARIMA model has three components, an autoregression (AR) component on the variable itself, a differencing (I) component if the time series is not stationary, and a moving average (MA) component on error terms.

\(^{3}\)Formally defined in Appendix 3.10.
Table 3.3: Firm Financial Ratios: Theory Testing Literature

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Input format</th>
<th>Primary independent variables</th>
<th>Methodology</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net cash flow as percentage of committed capital, net cash flow</td>
<td>numerical</td>
<td>price to dividend ratio, yield spread</td>
<td>linear regression</td>
<td>(Robinson and Sensoy, 2016)</td>
</tr>
<tr>
<td>Cash holding, Tobin’s Q</td>
<td>numerical</td>
<td>capital expenditures, working capital, cash flow, debt</td>
<td>linear regression</td>
<td>(Anderson and Hamadi, 2016)</td>
</tr>
<tr>
<td>Cost of equity</td>
<td>numerical</td>
<td>cash flow to capital ratio</td>
<td>linear regression, CAPM, F-F 3-factor model, Carhart 4-factor model</td>
<td>(Frank and Shen, 2016)</td>
</tr>
<tr>
<td>Tobin’s Q, EBITA</td>
<td>numerical</td>
<td>cash flow, R&amp;D expenses, capital expenditure, debt to asset ratio</td>
<td>linear regression</td>
<td>(Kang et al., 2017)</td>
</tr>
<tr>
<td>Daily stock return</td>
<td>numerical</td>
<td>earning smoothing</td>
<td>Jones model, linear regression</td>
<td>(Chen et al., 2017)</td>
</tr>
</tbody>
</table>
**Vector Autoregression (VAR) Model** The VAR method is a multivariate time series method which can describe the inter-dependency between the participated time series when they are influencing each other. Each time series can be described as its own lag values, other time series’ lag values, and an error term. Lag values are represented by autoregression (AR) terms. VAR model was used to identify relationship among firm employment growth, sales growth, and profits growth (Coad, 2010), and relationship among corporate social responsibility, corporate social irresponsibility, and firm performance (Kang et al., 2016).

### 3.3.2.2 Deep Learning Models

In recent years, deep learning models are acknowledged to fit well on sequential inputs. Therefore, we recognize deep learning model as another model family to handle time series data.

**Deep Learning (DL) for Natural Language Processing (NLP)** Deep learning (LeCun et al., 2015) is a family of machine learning models that are composed of multiple linear and nonlinear processing layers to learn representations of data with multiple levels of abstraction. It discovers intricate structures in large data sets by using the backpropagation algorithm (Rumelhart et al., 1986) to learn model parameters that are used to compute the representation in each layer from the representation in the previous layer. DL methods have dramatically improved the state-of-the-art in many domains, such as speech recognition, visual object detection, and genomics.

A centerpiece of natural language processing (NLP) is semantic representations of words and sentences. DL methods model them as word and sentence embeddings. An embedding is a mapping from a discrete object, such as a word or sentence, to a dense vector of real numbers. Glove (Pennington et al., 2014) and Word2Vec (Mikolov et al., 2013) are widely-used word embedding techniques.

While one can trivially combine word embeddings to produce a sentence embedding, researchers have studied techniques to generate more robust sentence representations. A building block of our models, Smooth Inverse Frequency (SIF) (Arora et al., 2017) is an unsupervised, PCA-based
technique to produce sentence embeddings for downstream NLP applications. An illustration of
the conceptual process can be found in Figure 3.1.

In the context of economic forecasting with textual data, it is important to notice that such
model architectures maintain distributed (vector) representations of text all the way through and
progressively build up predictive features, without prematurely collapsing high-dimensional textual
data into simplistic measures (e.g., sentiment, or event type).

![Figure 3.1: Deep Learning for Natural Language Processing: An Illustration](image)

**Long Short-Term Memory Networks (LSTMs)** Long Short-Term Memory (LSTM) net-
works\(^4\), first proposed by Hochreiter and Schmidhuber (1997), are a representative family of deep
learning architectures for sequence data. They model nonlinear temporal dynamics, and are capa-
bile of digesting vector input and emitting vector output at every time stamp. Numerous studies
have been conducted using LSTMs, such as information extraction (Miwa and Bansal, 2016), syn-
tactic parsing (Dyer et al., 2015), speech recognition (Graves et al., 2013), machine translation
(Bahdanau et al., 2015), and question answering (Wang and Nyberg, 2015). In this study, we

\(^4\)Formally defined in Appendix 3.12.
use LSTMs as backbones (an instance of “downstream models” in Figure 3.1) to model series of firm-specific pseudo events, all of which represented by embeddings.

### 3.4 Problem Formulation

Based on the broad context, related literature, and potential practical implications, we observe the following gaps in existing research:

- Unlike financial market studies, the corporate finance literature seems to be dominated by theory testing research, and lacks visible study tackling problems related to prediction of firm financial ratios. Given their indispensable value for a wide range of stakeholders, forecasting capabilities are highly desirable.

- Almost all financial-ratio-related studies (see Table 3.3) use numerical variables as input. For reasons we articulated in Section 1, power of textual data awaits to be exploited.

- Textual data as model input are inherently of very high dimensionality, the Achilles’ heel of classical econometric models such as ARIMA. On the other hand, deep learning represents state-of-the-art advancement in machine learning research, and is perfectly fitting for learning implicit high-level representations in text (by forgoing explicit feature engineering) and tackling dimensionality issues (by operating on fixed-size dense vectors) at the same time. As natural as it is, the integration between textual data and deep learning models is yet to emerge in leading economics literature.

We propose to bridge the research gaps by designing and evaluating a text-based, end-to-end forecasting framework that:

- forecasts firms’ long-term financial performance, as embodied in the financial ratios summarized in Table 3.4.\(^5\)

\(^5\)It is impractical (and not necessary) to work with all 70+ financial ratios defined in WRDS document. Hence we choose a representative measure from each category, without loss of generality.
Table 3.4: Definitions of Target Financial Ratios

<table>
<thead>
<tr>
<th>Category</th>
<th>Variable Name</th>
<th>Code</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valuation</td>
<td>Price to Earning (diluted, excl.EI)</td>
<td>pe_exi</td>
<td>price/earning</td>
</tr>
<tr>
<td>Efficiency</td>
<td>Sales to Working Capital</td>
<td>sale_nwc</td>
<td>sales/working capital</td>
</tr>
<tr>
<td>Solvency</td>
<td>Debt Ratio</td>
<td>de_ratio</td>
<td>total liability/shareholders' equity</td>
</tr>
<tr>
<td>Soundness</td>
<td>Cash Flow Margin</td>
<td>cfm</td>
<td>income before EI and depre./sales</td>
</tr>
<tr>
<td>Profitability</td>
<td>Net Profit Margin</td>
<td>npm</td>
<td>net income/sales</td>
</tr>
<tr>
<td>Capitalization</td>
<td>Common Equity to Invested Capital</td>
<td>equity_invcap</td>
<td>common equity/invested capital</td>
</tr>
<tr>
<td>Liquidity</td>
<td>Cash Ratio</td>
<td>cash_ratio</td>
<td>cash and short-term invest./current liabilities</td>
</tr>
</tbody>
</table>

- achieves its modeling competence by natural integration of state-of-the-art deep learning model architectures and textual data from public news.
- distinguishes itself by its unique flexibility to model multiple financial ratios simultaneously.
- exploits the integration of textual data and numerical data from a deep learning model perspective.
- most importantly, has general applicability in a broad array of economic forecasting problems with similar configurations and requiring high-dimensional input.

The conceptual structure of our problem formulation, forecasting future values of target financial ratios based on historical event series, is illustrated in Figure 3.2. We define a firm-specific pseudo event as a sentence that mentions a firm. The intuition is that semantic signals encoded in such pseudo events have predictive power for the focal firm’s future financial status. Formally

\[
\text{single - task : } y_{(i)}^{H} = f_{j=0}^{M}(E_{(i)j}, R_{(i)j}) \quad (y \in Y),
\]  

(3.1)
Figure 3.2: Conceptual Structure of the Problem Formulation

\[ \text{multi-task : } Y_{(i)}^H = f_{j=0}^M (E_{(i)j}, R_{(i)j}), \quad (3.2) \]

\[ E_{(i)j} = \sum_{k=0}^{S_{(i)jk}} (S_{(i)jk}) \quad (3.3) \]

where,

- \( Y \) is all target financial ratios summarized in Table 3.4. \( y \) is one of the financial ratios. \( y \in Y \).
- \( i \) indexes companies; \( j \) indexes time windows; \( k \) indexes sentences.
- \( H \) denotes the prediction horizon, and \( M \) denotes the size of memory, both measured in terms of number of time windows.
- \( E_{(i)j} \) is the aggregate event embedding for company \( C_{(i)} \) in time window \( m_j \); \( S_{(i)jk} \) is the encoding of a sentence \( k \) that mentions company \( C_{(i)} \) in time window \( m_j \).
- \( R_{(i)j} \) denotes the ratio of company \( C_{(i)} \) in time window \( m_j \).
- \( f \) is a learned model (function) that maps all pseudo event embeddings in memory to target values.
• $g$ is a function that aggregates multiple sentence embeddings into one pseudo event embedding.

In principle, $f$ and $g$ can be parameterized as any function approximator. Broadly speaking, there are two types of pseudo events. The following are several sample sentences, which we treat as pseudo events (the italicized entities represent the focal companies under consideration).

(a) To describe what has happened:

- *Amazon.com Inc.* invested $175 million in LivingSocial in December.
- Chicago-based *Boeing Co.*, the world’s largest aerospace company, said it received 127 orders in August, up from 115 in July.

(b) To describe what is going to happen:\(^6\):

- *AT&T* said that it would expand the rollout of its high-speed wireless technology, called Long-Term Evolution, or LTE, under the T-Mobile agreement.
- Now, as rulings start coming in, it might be time for a détente that helps *Apple* maximize the value of its patents, said Kevin Rivette, a managing partner at 3LP Advisors LLC, a firm that advises on intellectual property.

### 3.5 Deep Learning Models for Forecasting Firm Financial Ratios

Given the conceptual formulation in Figure 3.2, the problem has a natural time series setup. Therefore we choose Long Short-Term Memory (LSTMs) networks, state-of-the-art deep learning models for sequential data, as the backbone of our model. In other words, the functions $f$ in Equation 3.1 and Equation 3.2 are parameterized as LSTM layers (In practice, we use *BiLSTM* layers).

In the following subsections, we discuss several model variations motivated by different parameterization of function $g$ in Equation 3.3 and the effort to simultaneously model multiple target ratios.

\(^6\)Notice that such forward-looking information or speculation, though not explicitly differentiated by our model, does become a subset of the input and supports our model’s design rationale.
3.5.1 Single-Task Event MaxPooling Model (Single-EvMax)

We illustrate our Single-EvMax model in Figure 3.3. The model only predicts a single target ratio, hence the “single-task” part of the name.

![Figure 3.3: Single-EvMax Model Architecture](image)

The first layer of Single-EvMax model is the *pseudo events input* layer. The challenge here is that while the LSTM layers only accept a single-vector input per time window, there typically exist multiple pseudo events per window, hence requiring multiple event embeddings be aggregated into one. In the literature, such operations are often termed "pooling", and represent the essence of the \( g \) function in Equation 3.3. Pooling broadly refers to some form of combining "nearby" values of a signal (e.g., through averaging) or picking one representative value (e.g., through maximization or sub-sampling). In our case, within every time window, we max-pool on every dimension of all
pseudo event embedding vectors in the \textit{MaxPooling} layer. Formally, we define the max-pooling component as:

\[ \forall d, v^{<d>}_{(i)j} = \max_k E^{<d>}_{(i)jk} \quad (3.4) \]

Where \( d \) indexes individual dimensions in the embedding space, and \( v \) is the aggregated vector for all events in the same time window \( m_j \) after the pooling operation.

After the \textit{BiLSTM} layers, the model will generate a dense vector representation of the company in the \textit{Dense} layer. Finally we predict the target financial ratio with a linear function of the dense vector.

3.5.2 Single-Task Event Attention Model (Single-EvAttn)

Notice that the max-pooling trick described in the previous subsection is quite naive, and produces a crude aggregate of multiple event embeddings. A potentially more powerful approach is to preserve the entirety of each event embedding and adaptively assign weights to them. In other words, the model should pay more “attention” to the more informative pseudo events.

Interestingly enough, the deep learning community has extensively explored \textit{attention mechanisms} recently. Essentially an imitation of human sight mechanism, they were used primarily in computer vision in the early days. They learn to focus on particular regions, instead of scanning the entire scene end to end.

The intuition for attention mechanisms in natural language processing is to assign higher attention to text units (typically words or tokens) that contain more information for the task on hand. \cite{Bahdanau2015} was the first to apply attention mechanisms in NLP, and more specifically, in machine translation.

More recently, \cite{Vaswani2017} used multi-head attentions alone to solve sequence prediction problems which are traditionally handled by other neural networks such as LSTMs and Convolution Neural Networks (CNNs).
Figure 3.4 illustrates our improved model by integrating an attention mechanism. The first layer of our model is pseudo events input layer. Within every time window, we select the top $K$ event embedding vectors (with the largest L2 norms) as the representation of the company in that time window.

The selected pseudo event embeddings are fed into the event attention layer, a self-attention mechanism inspired by previous work. We institute the attention mechanism to distribute weights among pseudo events for more effective prediction of the target variable. Formally, we define our event attention mechanism as:

$$ u_{ijk} = \text{sigmoid}(W_{ij}E_{ijk} + b_{ij}) $$

(3.5)
\[ \alpha_{ijk} = \frac{\exp(u_{ijk})}{\sum_k \exp(u_{ijk})} \quad (3.6) \]

\[ v_{ij} = \sum_{k=1}^{K} \alpha_{ijk} E_{ijk} \quad (3.7) \]

where \( u_{ijk} \) is a pseudo event’s raw attention score, \( \alpha_{ijk} \) is the normalized attention weight, and \( v_{ij} \) is the weighted sum vector to represent company \( i \) at time window \( j \).

After the BiLSTM layers, the model will again generate a dense vector representation of the company in the Dense layer and use it to compute the value of the target ratio.

### 3.5.3 Multi-Task Event Attention Model (Multi-EvAttn)

While we can certainly perform learning and inference on each individual financial ratio, it is conceivable that multiple ratios may share some common underlying characteristics rooted in the common business operations and dynamics of the firm. As it turns out, there is a family of machine learning techniques called Multi-Task Learning (MTL), which improves generalization by leveraging the domain-specific information contained in the training signals of related tasks (Caruana, 1997). The key rationale of MTL is shared representation among related tasks, and it can also be viewed as a regularization technique. The layered architecture of deep learning models makes it quite feasible to practice MTL.

We illustrate our multi-task model in Figure 3.5, which still relies on the event attention mechanism to produce input for BiLSTM layers. The multiple financial ratios naturally form the multiple learning tasks that leverage common representations among each other. The model is learned by optimizing an aggregate loss function that is a linear combination of the individual task loss.

Notice that in all three model variations, the final Dense layer before output is labeled as “firm embedding”, because it is effectively a semantic representation of the firm learned by the current model. See more discussion in Section 3.7.3.
3.5.4 Ratio LSTM Model

We use financial ratio series as input and LSTM model as the backbone in this model. We can write the single-task model as

\[ y^H_{(i)} = LSTM(r_{(i)j=0}^M) \]  

(3.8)

and the multi-task model as

\[ Y^H_{(i)} = LSTM(r_{(i)j=0}^M) \]  

(3.9)

where \( r \) is the financial ratio of company \( C_{(i)} \) from time window \( j = 0 \) to \( j = M \).
3.5.5 Event Ratio Models

3.5.5.1 Event Single-Ratio Model

On top of the Single-EvAttn model, we integrate the financial ratio series in this Event Ratio model. We illustrate our Event Ratio model in Figure 3.6.

For event stream, the \textit{Pseudo Events Input Layer}, \textit{Event Attention Layer}, and \textit{BiLSTM} layers are the same for Single-EvAttn model and Event Ratio model.

For the financial ratio stream, we take the ratio series as the input in the \textit{Ratio Input Layer}, and feed them into a \textit{Dense Layer}.

\begin{equation}
l_t = \text{Dense}(k_t) \tag{3.10}
\end{equation}

Then the \textit{Dense Layer} results are used as inputs in the \textit{Ratio LSTM Layer}. Next, we integrate two streams of data, i.e. textual data stream and numerical data stream. Specifically, we take the hidden states from the \textit{Events LSTM Layers} and \textit{Ratio LSTM Layer} and merge (multiply) them in the \textit{Merge (Multiply) Layer}.

\begin{equation}
m_t = \text{Multiply}(l_t, h_t) \tag{3.11}
\end{equation}

Inspired by Luong et al. (2015), we first conduct a temporal self-attention on the event and ratio integrated vector $m_1$ to $m_M$, and their attention scores $b_1$ to $b_M$ are computed as:

\begin{equation}
z_i = \text{ELU}(W_i m_i + b_i) \tag{3.12}
\end{equation}

\begin{equation}
b_i = \frac{\exp(z_i)}{\sum_k \exp(z_k)} \tag{3.13}
\end{equation}

Since the ratio LSTM architecture fits our task well, we apply the attention weights to the ratio LSTM hidden states $l_1$ to $l_M$ to calculate the event ratio weighted vector $j_t$ as:

\begin{equation}
j_t = \sum_{k=1}^{K} b_i l_i \tag{3.14}
\end{equation}
Next, we concatenate the event ratio weighted vector $j_t$ with the last hidden state of the Events LSTM Layer, and compute the context vector $C_t$ as:

$$C_t = \text{Concatenate}(j_t, h_M)$$

(3.15)

Then, the context vector $C_t$ goes thru a Dense Layer to produce $d_t$, and $d_t$ is used to predict the target financial ratio $Y_{(i)}^H$.

$$d_t = \text{Dense}(C_t)$$

(3.16)

$$y_{(i)}^H = \text{Dense}(d_t)$$

(3.17)

3.5.5.2 Event Multi-Ratio Model

In the Event Multi-Ratio model, all target variables are predicted simultaneously (the same as the Multi-EvAttn model).

$$Y_{(i)}^H = \text{Dense}(d_t)$$

(3.18)

3.6 Empirical Results

3.6.1 Experimental Setup

Candidate models We evaluate and compare the following models in our experiments:

1. Time Series Models:
   1a) ARIMA: the baseline time series model for single target variable, based on historical values of the target financial ratio.
   1b) VAR: the baseline time series model for multiple target variables, based on historical values of the financial ratios.

2. Text-based DL Models:
   2a) Single-EvMax: single-task event maxpooling model.
   2b) Single-EvAttn: single task event attention model.
Figure 3.6: Event Ratio Model Architecture
2c) Multi-EvAttn (unweighted): multi-task event attention model (each individual task loss receiving equal weight in the aggregate loss).

2d) Multi-EvAttn (weighted): multi-task event attention model (each individual task loss receiving unequal weight in the aggregate loss)

3. Ratio-based DL Models:

3a) Single-Ratio LSTM: single-ratio LSTM model.

3b) Multi-Ratio LSTM: multi-ratio LSTM model.

4. Text-Ratio Integration DL Models:

4a) Event Single-Ratio: event attention and single-ratio prediction model.

4b) Event Multi-Ratio: event attention and multi-ratio prediction model.

Models 1a), 2a), 2b), 3a), and 4a) work with individual target variables, and models 1b), 2c), 2d), and 4b) predict multiple target variables simultaneously. We employ Mean Absolute Percentage Error (MAPE) as the loss function and performance measure.

\[
\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^{N} \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|
\]  

(3.19)

Data  Although our framework is applicable to public and private firms, we only have access to public firms’ financial data. Hence we focus on Fortune 1,000 companies in our study. Each company’s ticker ID and financial ratios are collected from Wharton Research Data Services (WRDS). We use news articles from a major business news service published between the year 2011 and 2015. The descriptive statistics of the dataset are listed in Table 3.5. We collected each company’s financial ratios also from Wharton Research Data Services (WRDS).

Preprocessing text  We use Python NLTK as our text processing tool. For each news article, we extract its content and publication time, and segment each article into sentences.

By consulting a pre-compiled gazetteer of company names, name variants, and abbreviations, we extract all sentences that contain mentions of each focal company, followed by simple noise-

---

\footnote{Here is another practical constraint we have to work with. Though the models are capable of handling much larger text corpora, this is the maximum amount of news articles we were able to collect.}
Table 3.5: Data Statistics by Year

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of Articles</th>
<th>Number of Pseudo Events</th>
<th>Number of Companies Covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>18,615</td>
<td>92,707</td>
<td>707</td>
</tr>
<tr>
<td>2012</td>
<td>14,840</td>
<td>77,043</td>
<td>697</td>
</tr>
<tr>
<td>2013</td>
<td>58,211</td>
<td>289,007</td>
<td>819</td>
</tr>
<tr>
<td>2014</td>
<td>49,711</td>
<td>183,976</td>
<td>817</td>
</tr>
<tr>
<td>2015</td>
<td>39,705</td>
<td>135,936</td>
<td>810</td>
</tr>
<tr>
<td>Total</td>
<td>181,082</td>
<td>778,669</td>
<td>927 (union)</td>
</tr>
</tbody>
</table>

reduction heuristics. Each such sentence is treated as a pseudo event. In case of multiple in-list companies in one sentence, the corresponding pseudo event will participate multiple companies’ event sequence. After dropping no-news-coverage firms, 927 companies remain in our data. We then replace all company names and name variants with a special token ‘FOCOMP’ (meaning 'focal company'),

in order to help the model generalize across firms by sharing statistical strength of similar event embeddings.

**Sentence (pseudo event) encoding** We use the SIF sentence embedding method (Arora et al., 2017) to encode every extracted pseudo event into a 300-dimension dense vector.

**Preprocessing target ratios** To tackle the problem of missing values in the firm financial ratios downloaded from WRDS, we first exclude companies that have more than two missing values in the study period, then fill in the small number of missing values by linear interpolation between neighboring values. After eliminating those with too many missing values in financial ratios, finally 707 companies remain in our data.

**Model instantiation, training, and evaluation setup** In principle, one can stack a large number of BiLSTM layers to build a deep neural network. For practical reasons of computational complexity, we only stack 2 BiLSTM layers as our model backbone.
In our experiments, we define each time window $m_j$ as a month, and set memory size $M = 12$. In line with the vision of long-term forecasting illustrated in Figure 3.2, we set the prediction horizon $H = 12$.

We also set $K=5$ (see Section 3.5.2 for the definition of $K$), the median number of pseudo events per company per month.

We conduct model evaluation in the most stringent temporal out-of-sample fashion. The two-phase process is illustrated in Figure 3.7. In the validation phase, we use event data from the year 2011 to 2014 as training data and predict target ratios at the end of the year 2015. We tune our model hyper-parameters (e.g., number of epochs) by validating against known true values of target ratios. Once model parameters are determined based on validation, we retrain the same model by using all data from the year 2011 to 2015 and evaluate the final prediction at the end of 2016.

Last but not least, in all deep learning models, we use exponential linear units (ELUs) (Clevert et al., 2015) as the activation function, and deploy the Adam optimizer (Kingma and Ba, 2014).

**ARIMA model instantiation** ARIMA($p,d,q$) model is well-known to be suited for small memory sizes, so we look for our baseline model among a set of small-memory-size ARIMA models. To be able to find the best performing ARIMA model parameters, we experiment combinations of the following parameter settings: autoregressive term $p$ in $\{1, 2, 3\}$, differencing term $d$ in $\{0, 1\}$, and the moving average term $q$ in $\{0, 1, 2\}$. Among these ARIMA model variations, we find the best-performing configuration $(p,d,q)$ in the validation phase and then use it in final testing. Table 3.6 lists such configuration for each target ratio.

**VAR model instantiation** Similar to the instantiation of the ARIMA model, we also need to find the best performing VAR($p$) model parameters. In particular, we experiment autoregressive term $p$ in $\{1, 2, 3, 4, 5\}$ on the differencing $\{0, 1, 2\}$. Among these VAR model variations, we find the best-performing configuration $p$ and the differencing value in the validation phase and then use it in final testing. We choose a model configuration from the lowest average of variable
Figure 3.7: Model Validation and Testing
Table 3.6: ARIMA Besting-performing \((p,d,q)\) for Final Testing

<table>
<thead>
<tr>
<th>target variable</th>
<th>(p)</th>
<th>(d)</th>
<th>(q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sale_nwc</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>pe_exi</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>de_ratio</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>cfm</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>npm</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>equity_invcap</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>cash_ratio</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>eps</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

performance in the validation period. The configuration we use for the final testing period is VAR(1) on differencing = 2, and \(p = 1\).

### 3.6.2 Model Performance

Our model performance is summarized in Table 3.7. Each MAPE number is an average over 5 runs of the same model configuration.

<table>
<thead>
<tr>
<th>Target Ratio</th>
<th>1a) ARIMA</th>
<th>1b) VAR</th>
<th>2a) Single Ev-Max</th>
<th>2b) Single EvAttn</th>
<th>2c) Multi EvAttn (un-weighted)</th>
<th>2d) Multi EvAttn (weighted)</th>
<th>3a) Multi-Ratio LSTM</th>
<th>3b) Multi-Ratio LSTM</th>
<th>4a) Event Single-Ratio</th>
<th>4b) Event Multi-Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>sale_nwc</td>
<td>100.55%</td>
<td>2885.92%</td>
<td>57.64%</td>
<td>55.95%</td>
<td>52.10%</td>
<td>52.61%</td>
<td>25.79%</td>
<td>30.11%</td>
<td>25.57%</td>
<td>32.51%</td>
</tr>
<tr>
<td>pe_exi</td>
<td>100.28%</td>
<td>1381.58%</td>
<td>96.83%</td>
<td>96.64%</td>
<td>70.84%</td>
<td>69.88%</td>
<td>50.87%</td>
<td>53.46%</td>
<td>48.68%</td>
<td>54.76%</td>
</tr>
<tr>
<td>de_ratio</td>
<td>66.92%</td>
<td>246.32%</td>
<td>64.07%</td>
<td>55.44%</td>
<td>51.07%</td>
<td>51.59%</td>
<td>20.71%</td>
<td>22.53%</td>
<td>20.65%</td>
<td>27.76%</td>
</tr>
<tr>
<td>cfm</td>
<td>187.38%</td>
<td>205.38%</td>
<td>94.70%</td>
<td>89.10%</td>
<td>74.33%</td>
<td>72.20%</td>
<td>43.08%</td>
<td>45.65%</td>
<td>41.51%</td>
<td>48.17%</td>
</tr>
<tr>
<td>npm</td>
<td>163.54%</td>
<td>232.25%</td>
<td>112.55%</td>
<td>95.85%</td>
<td>104.89%</td>
<td>98.67%</td>
<td>69.59%</td>
<td>73.07%</td>
<td>65.66%</td>
<td>85.31%</td>
</tr>
<tr>
<td>equity_invcap</td>
<td>33.15%</td>
<td>77.10%</td>
<td>107.69%</td>
<td>90.76%</td>
<td>77.67%</td>
<td>73.16%</td>
<td>19.77%</td>
<td>24.79%</td>
<td>18.80%</td>
<td>29.17%</td>
</tr>
<tr>
<td>cash_ratio</td>
<td>58.65%</td>
<td>141.58%</td>
<td>102.11%</td>
<td>85.33%</td>
<td>90.50%</td>
<td>81.61%</td>
<td>32.73%</td>
<td>42.22%</td>
<td>32.45%</td>
<td>51.46%</td>
</tr>
</tbody>
</table>

Model 1a) ARIMA and 1b) VAR are conducted on ratio time series only. We can see the ARIMA model performs better than VAR on individual ratios. It indicates the interactions among all variables can negatively influence the target variable’s individual prediction. In other words, lumping sum all variables together is not the right solution direction for this task. Furthermore, from the higher differencing value 2 in the VAR(1) model, we can also see it is harder for the VAR
model to be stationary than the ARIMA model. On the one hand, traditional time series models with stationary assumptions are not the best appropriate tools for our task. On the other hand, lumping sum all variables together is also not the right solution path for this task.

Bagshaw (1986) also compared the univariate ARIMA model with static VAR models on seven economic variables. The authors showed the same result, i.e., the univariate ARIMA performs better than static VAR. Given related literature (Litterman, 1986), while the VAR model is a general approach in multivariate time series, it has a serious problem, which is overparameterization. A large number of variables can consequently make VAR parameters a large number. Moreover, it is very easy for the number of parameters to reach or exceed the number of observations. A good forecasting model needs to capture both the systematic relationships and the random variation in the data. In economic data, however, the systematic relationship is usually only a small percentage of the total variations. The goal of the desired model is to filter out the weak systematic variation and work heavily on random variation. Therefore, the VAR model does not have this advantage. Moreover, from our data specifics, the dynamics among all the companies created another layer of challenge and became another constraint.

Next, we look at deep learning (DL) model families. First, we look at models based on textual input only. They are model 2a), 2b), 2c) and 2d). The text-based DL models, having no access to historical time series, show very competitive forecasting performance. We observe the baseline ARIMA model produces high MAPE on 5 out of 7 target ratios (sale, nwc, pe, exi, de, ratio, cfm, and npm), and textual DL models consistently outperform in these variables.

Diversely, when ARIMA produces reasonably low MAPE (on equity, invcap and cash_ratio), deep learning models lose the race. It suggests the existence of different dynamics underlying different financial ratios. Capitalization and liquidity ratios are more tied to the firm’s internal operations. Hence, they are more stable, receive less news coverage, and consequently more amenable to an inherently linear model (ARIMA). In contrast, the other five ratios all involve sales and the stock market. Hence, they are more volatile, receive more news coverage, and consequently see more success with news-based non-linear DL models.
Within the text-based deep learning models, model 2b) Single-EvAttn model consistently outperforms 2a) Single-EvMax model, which underscores the superiority of the attention mechanism over the relatively naive max-pooling trick. In addition, the 2c) Multi-EvAttn model (unweighted, i.e., each task-specific loss gets an equal weight of 1), by sharing representations among multiple ratios, yields further performance gain over 2b) Single-EvAttn model, on 5 out of 7 target ratios. In our attempt to further boost the MTL performance, we assign higher weights to the loss of the two “harder” target ratios, npm and cash_ratio. When moderate over-weighting is in place (roughly between 2 and 5), we observe another wave of error reduction on model 2d) Multi-EvAttn (weighted).

As of now, we can see that even when we do not have a firm’s accounting information, we can still use our text-based DL model to forecast the firm’s future financial ratios.

Next, we integrate ratio series in the input and use them in the proposed deep learning architectures. We use LSTM as the backbone and financial ratio time series as input in the 3a) Single-Ratio LSTM model and 4a) Multi-Ratio LSTM model. First, we observe that deep learning models can take the ratio series as input and predict future multi-steps ahead ratios better than the previous models when there is no stationary assumption. This observation suggests using time series as input and applies them in deep learning architecture is a plausible approach for our task. Second, the same as the comparison between ARIMA and VAR, we also observe that multi-task learning on ratios alone did not bring in additional performance improvement. This observation further confirms our previous suggestion that lumping sum all variable series together is not the best approach.

So far, we have known that 1) textual inputs can provide prediction improvements, and 2) ratio series work well with deep learning architecture. Given these observations, we can work on the integration of these two data streams and make them both as inputs. We use both events and ratio series in model 4a) Event Single-Ratio model and 4b) Event Multi-Ratio model as inputs. From 4a) model performance, we can see all variables’ performance is lifted by integrating these two streams, while some variables gain more, and some variables gain less. The most benefited variables are
price to earnings ratio (\texttt{pe-exi}), cash flow margin (\texttt{cfm}), and net profit margin (\texttt{npm}). Interestingly, they are also in the five winning variables from the text-based DL models. We can further confirm our intuition that the prediction performance is highly relevant to the text’s content. Even these targeted ratios are all about the same company, if textual inputs are limited to the particular aspect, the model improvements from text data will also be limited.

The model performance in 4a) Event Single-Ratio model makes this model the best for our task, within all of our proposed models. Furthermore, the comparison between 4a) and 4b) further confirms that lumping sum all ratio series together is not the best approach for our task.

Additionally, when only textual data is available (model 2a, 2b, 2c, and 2d), multi-task learning can bring in the overview look to the existing text. However, when ratio series are available, they have dominant impacts. Therefore, textual input is better to serve each ratio individually and solve each ratio’s prediction one at a time. In other words, when accounting information is not available, multi-task learning on textual inputs is the best approach. And when accounting information is available, combining textual input and ratio history in a single-task learning framework is the best approach.

3.7 Model Interpretability

Deep learning models are often criticized for being big black boxes. While such criticism is not completely unreasonable considering the nonparametric nature of DL models, this section represents our effort to open up the boxes and gain insights into their inner workings. More specifically, we extract pseudo event attention weights and company embeddings at the end of testing process, then visualize and analyze them.

3.7.1 Visualizing and Understanding Event Attention

The attention mechanism introduced in Section 3.5.2 provides a rather nice form of model interpretability, in the sense that the normalized attention scores quantify the relative importance of pseudo events within each time window, as determined by the learned model. We can extract these
scores from the Event Attention Layer in Figure 3.4 and Figure 3.5, as well as their corresponding pseudo events.

In this example, we focus on Apple Inc., and work with the Single-EvAttn model for target ratio sale_nwc. We visualize the overall attention map in Figure 3.8, and zoom into the specifics in Table 3.8 and Table 3.9.

In Table 3.8 (for year 2015 month 1), the most-attended pseudo event is #4, which discusses Apple Inc.’s patent filing two years ago. Intuitively, the event likely creates a positive long-term impact on the firm’s financial well-being. The model correctly assigns it a higher weight.

Table 3.8 actually suggests a very interesting mistake of the model. Pseudo event #1 was deemed relevant to Apple Inc. by the event extraction algorithm due to company name matching, 8A much more detailed example on Expedia is presented in Appendix 3.13.
and then assigned higher weights by the attention mechanism due to semantic matching (it is certainly sales-related). The only problem is that the sentence actually talks about Samsung (at the time, it was a supplier of chips used in Apple devices), the correct understanding of which can only arise from a larger context, and requires NLP capabilities arguably beyond state of the art (e.g., anaphora resolution across sentence boundaries).

Table 3.9: Events Attention Visualization Example: Apple Inc. Month 11

<table>
<thead>
<tr>
<th>Event</th>
<th>Attention weight</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.29</td>
<td>In addition to sales from its own handsets, it gains revenue from each iPhone sold because its chip unit manufactures the main processor used in Apple’s phones and tablets.</td>
</tr>
<tr>
<td>2</td>
<td>0.21</td>
<td>These stock grants are meant to reward them down the road for their hard work in helping to keep Apple the most innovative company in the world.</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
<td>The company sought an order for Apple to produce the documents.</td>
</tr>
<tr>
<td>4</td>
<td>0.12</td>
<td>Apple, ..., added 2.5 percent to $388.83.</td>
</tr>
<tr>
<td>5</td>
<td>0.12</td>
<td>Apple, meanwhile, opened its iTunes store in 2003.</td>
</tr>
</tbody>
</table>

3.7.2 Visualizing and Understanding Event and Ratio Interaction

In Figure 3.9, we illustrate an example for our integration model: we use both events information and ratio series of as inputs, and predict its 12 months ahead financial ratios.

Specific topics from news coverage are observed to have the capability to facilitate ratio predictions. For example, patent filing can affect a company’s future cash flow, profits, and earnings. Therefore, given the news coverage, our integration model can gain additional information on the top of its historical financial ratio series. The prediction performance can be improved. Additionally, market-related contents are also easy to be found in news articles. Therefore, earnings-related ratios, such as the price-to-earnings ratio, can benefit from this type of news. Additionally, a firm’s new products or services is another popular topic in the news. This type of news can generate additional sales to the firm and further increase its cash flow and profits. Thus, given a firm’s
Figure 3.9: Events and Ratio Integration Example: Apple Inc. Overview

historical ratio series, its future ratio predictions can be further promoted by textual, i.e., news, inputs.

In this example’s event stream, we can see the higher attention events, such as "Apple doesn’t sell the cheapest phones," "Sell Apple," and "Apple, the No." are negative information to the firm. Therefore, they could contribute to the drop in cash flow margin (cfm), net profit margin (npm), and sales to net working capital (sales\_nwc) of the firm one year later.

From the model performance, we can confirm our intuition that when news coverage and a firm’s historical ratio are combined, the integrated model gain from both textual inputs and the historical ratio input.

3.7.3 Visualizing and Understanding Firm Embeddings

An extremely useful byproduct of our model is firm embeddings, which map firms to vectors of real numbers. Technically, each firm embedding is a semantic representation of the firm learned
in the context of a model. Understanding of these embeddings is usually derived from the relative positioning of firms in a high-dimensional space, typically visualized in a 2-D space. The notion of proximity can intuitively guide peer benchmarking for internal management and portfolio-building for outside investors.

A traditional way to understand firm relationships is industry segmentation. For reference, we list the Fama-French 5-industry portfolios (Fama and French, 2019) in Table 3.10.

Table 3.10: Fama-French 5-Industry Portfolios

<table>
<thead>
<tr>
<th>Industry</th>
<th>Portfolio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumer</td>
<td>Consumer Durables, NonDurables, Wholesale, Retail, and Some Services</td>
</tr>
<tr>
<td></td>
<td>(Laundries, Repair Shops)</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>Manufacturing, Energy, and Utilities</td>
</tr>
<tr>
<td>HiTech</td>
<td>Business Equipment, Telephone and Television Transmission</td>
</tr>
<tr>
<td>Healthcare</td>
<td>Healthcare, Medical Equipment, and Drugs</td>
</tr>
<tr>
<td>Others</td>
<td>Mines, Construction, Building Maintenance, Transportation, Hotels,</td>
</tr>
<tr>
<td></td>
<td>Business Services, Entertainment, and Finance</td>
</tr>
</tbody>
</table>

In contrast, our models generate firm embeddings without knowing the company (since we replaced all company names with the special token ‘FOCOMP’), and certainly not knowing any industry membership. We extract every firm’s embedding by capturing the activations of the Dense Layer in Figure 3.4 and Figure 3.5, then use t-SNE (van der Maaten and Hinton, 2008) to visualize the high-dimensional firm embeddings in a 2-D space. Below we briefly discuss firm embeddings encoded by two different ratios.

**Sale\_nwc-encoded firm embeddings** Figure 3.10 visualizes the firm embeddings encoded by the Single-EvAttn model for sale\_nwc. Firstly, our firm embedding space overall agrees with industry segmentation quite well. Most consumer companies (in blue) are clustered in the top portion of the graph: one big cluster is in the right upper corner (region 1) and a smaller cluster is in the mid upper region (region 2). The healthcare industry (in gold color) and the HiTech industry (in green) are clustered in the left bottom corner (region 3). The manufacturing companies are somewhat scattered in the entire space.
When there is disagreement, the learned embedding structure seems to pick up interesting semantics. Although Netflix (in purple color) is Fama-French classified into the ‘Others’ category, its firm embedding is situated very close to Walt Disney and CBS (region 4), likely due to similar product offerings. Walt Disney and CBS, on the other hand, though officially classified as HiTech, are located much closer to companies in the consumer industry than typical HiTech ones (e.g., Google and Microsoft in region 3). Another example is Nike (region 2), which, though officially belongs to the manufacturing industry, actually falls into the neighborhood of the consumer industry. Its “persona” in public perceptions does seem to be much more similar to Macy’s and Foot Locker than to Caterpillar.

**De_ratio-encoded firm embeddings**  Figure 3.11 visualizes the firm embeddings encoded by the Single-EvAttn model for de_ratio (de_ratio values are shown in the figure when applicable). We have the following observations:

- Retail companies (region 3), such as Wal-Mart, Costco, Best Buy, the Home Depot, and Target, have similar debt ratios, and they are indeed recognized together by our firm embedding.

- The solvency situation of HiTech companies (region 2) seems fairly similar, hence they are clustered together, with IBM being an exception (our models didn’t succeed in forecasting its relatively high debt ratio).

- Berkshire Hathaway (region 1), surprisingly, is quite far away from other financial institutions. Does it make sense? As it turns out, its debt_ratio (1.18) is much lower than companies such as Citigroup (7.44), JPMorgan Chase (9.49), and Morgan Stanley (10.12). Why? Berkshire Hathaway is known for its conservative debt use, as its chairman and CEO Warren Buffett believes in running businesses with as little debt as possible.

All examples above show that our news-powered DL models learned sophisticated semantic representations of firms far richer than hard industry segmentation and otherwise requiring non-trivial knowledge.
Figure 3.10: Firm Embedding Encoded by `sale_nwc`
Figure 3.11: Firm Embedding Encoded by $\text{de}_{\text{ratio}}$
3.8 Conclusion, Implications, and Future Work

In this paper, we show the effectiveness of deep text mining models in forecasting firms’ long-term financial performance, driven by firm-specific pseudo-events. The news-powered deep learning models yield impressively competitive forecasting performance compared to standard econometric models based on historical accounting data. We also illustrate the potentiality of further forecasting improvement via multi-task learning. Subsequently, we propose an integration model to combine the news data and financial ratio streams into an end-to-end model. Additionally, we provide two more model artifacts, such as attention maps and firm embeddings, which offering meaningful insights for model interpretation and decision making, especially when accounting data is not available.

Our work has considerable implications for firm stakeholders, the financial industry, and the research community:

- Internally, the forecasting models and their artifacts provide decision support for firm executives, when performing various tasks such as strategic planning, financial forecasting, and peer benchmarking.

- Externally, institutional investors will benefit from the models’ predictive outcome and explanatory insight. Such enhanced transparency and accountability are even more critical when dealing with private firms who have no duty to disclose their accounting books.

- Main Street investors are typically not equipped or inclined to conduct sophisticated quantitative analysis based on (arguably clean) numerical accounting data. Our models and artifacts provide previously non-existent, and interpretable decision aid\(^9\), without compromising on quality of prediction at all.

- Third-party regulators and agencies represent another family of beneficiaries. For example, S&P Global Ratings’ analysis can potentially leverage enriched information channels and the

\(^9\)The linear form of ARIMA models and their coefficients are deemed interpretable by many researchers, but not necessarily the best genre of interpretability for mathematically unsophisticated laypeople. Though we do not intend to engage in a philosophical debate around model interpretability, we believe our model artifacts provide alternative avenues of this important notion.
interplay of multi-faceted financial pictures, both of which are essential components in our framework.

- On the methodological front, our problem formulation and model architecture are largely problem independent, hence have general applicability in a broad spectrum of economic forecasting problems.

- Last but not least, artifacts derived from our models can help members of the business research community, in particular theory-minded researchers, generate new concepts and hypotheses, and even conduct preliminary theoretical explorations.

An early attempt at bringing large-scale textual data and state-of-the-art deep learning models into financial forecasting, our work is not without its limitations. We are poised to extend it in multiple directions:

- The success of our models heavily relies on the quality of input. Room for improvement exists in both event extraction and event encoding techniques.

- Other types of non-numerical data can participate in the modeling and forecasting process, e.g., textual portions of SEC filings, sell-side reports, etc.

- The potential of MTL is yet to be fully realized by designing more educated information-sharing model architectures.

- The generalizability of our model architecture is yet to be fully explored in broader economic forecasting settings.

One final note: the purpose of this research is not to establish a horse race and declare the winner between econometric models and AI-based models. While ARIMA is certainly not the apex of the former, our deep learning models have great potential for improvement as well. What we strive, and what we hope to have achieved, is to illustrate the empirical power of textual data in economic forecasting, and demonstrate corresponding modeling capabilities. In reality, sophisticated researchers and agencies will likely exploit both families of models.
3.9 References


### 3.10 Appendix A: ARIMA Formulation

In ARIMA($p, d, q$) model, $p$ denotes the order of the autoregression (AR), $d$ denotes the order of differencing (I), and $q$ denotes the order of the moving average (MA).

\[
(1 - \sum_{i=1}^{p} \phi_i L^i)(1 - L)^d Y_t = (1 - \sum_{i=1}^{q} \theta_i L^i) \epsilon_t \tag{3.20}
\]

\[
\epsilon_t \sim N(0, \sigma^2)
\]

where $L$ is the lag operator, and $\epsilon_t$ is a white noise process.
If we use Box-Jenkins backshift operator, Equation 3.20 can be written as

\[(1 - B)^p(1 - B)^dY_t = (1 - B)^q\epsilon_t\] (3.21)

\[\epsilon_t \sim N(0, \sigma^2)\]

where \(B\) is the backshift operator, \((1 - B)^p = 1 - \phi_1B - \phi_2B^2 - \cdots - \phi_pB^p\), and \((1 - B)^q = 1 - \theta_1B - \theta_2B^2 - \cdots - \theta_pB^p\).

Equation 3.20 and Equation 3.21 can be also written as

\[\phi_p(B)\nabla^dY_t = \theta_q(B)\epsilon_t\] (3.22)

\[\epsilon_t \sim N(0, \sigma^2)\]

or

\[\phi(B)\nabla^dY_t = \theta(B)\epsilon_t\] (3.23)

\[\epsilon_t \sim N(0, \sigma^2)\]

where \(\nabla\) is the difference operator, \(\phi(B) = \phi_p(B) = 1 - \sum_{i=1}^{p} \phi_iB^i\) is the \(p\)-order AR operator, \(\theta(B) = \theta_j(B) = 1 - \sum_{j=1}^{q} \theta_jB^j\) is the \(q\)-order MA operator, and \(\epsilon_t\) is a white noise process.

### 3.11 Appendix B: VAR Formulation

In VAR model, \(p\) denotes the order of the autoregression (AR), and \(k\) denotes the number of variables.

\[Y_t = A_1Y_{t-1} + A_2Y_{t-2} + \ldots + A_pY_{t-p} + c + \epsilon_t\] (3.24)

where \(A\) is a time-invariant \((k \times k)\) matrix, \(c\) is a constant, and \(\epsilon_t\) is an error term which satisfying that \(E(\epsilon_t) = 0\) and no correlation across times.
We use $k = 2$ and $p = 1$ as an example. The model VAR(1) can be presented as:

$$Y_{1,t} = a_{1,1}Y_{1,t-1} + a_{1,2}Y_{2,t-1} + c_1 + \epsilon_{1,t}$$

(3.25)

$$Y_{2,t} = a_{2,1}Y_{1,t-1} + a_{2,2}Y_{2,t-1} + c_2 + \epsilon_{2,t}$$

(3.26)

### 3.12 Appendix C: LSTM Formulation

Each LSTM cell has an input gate $i_t$, an output gate $o_t$, and a forget gate $f_t$, at each time step $t$. And output vector of cell at time $t$, $h_t$, is based on the previous cell state $h_{t-1}$, the current state input $x_t$, and the current cell state $c_t$. By including the sigmoid activation function $\sigma_g$ and hyperbolic tangent activation function $\sigma_c$, we can write the LSTM with forget gate as:

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f)$$

(3.27)

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i)$$

(3.28)

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o)$$

(3.29)

$$c_t = f_t \odot c_{t-1} + i_t \odot \sigma_c(W_c x_t + U_c h_{t-1} + b_c)$$

(3.30)

$$h_t = o_t \odot \sigma_h(c_t)$$

(3.31)

where $W_f$, $W_i$, $W_o$, $U_f$, $U_i$, and $U_o$ are weight matrices, and $b_f$, $b_i$, and $b_o$ are biases.

A natural extension of LSTM network is a bidirectional LSTM (BiLSTM), which lets the sequence pass through the architecture in both directions and aggregate the information at each time step.

$$h_t = \sigma(h_t^+, h_t^-)$$

(3.32)
3.13 Appendix D: Case Study: Attention Analysis on Expedia

In this Expedia example here, we work with two target ratios - $\text{cfm}$ (cash flow margin) and $\text{npm}$ (net profit margin), and hence examine two Single-EvAttn models in parallel. We choose these two ratios in this case study because they share some common impact factors, and we’d like to see if and how our event attention model captures the semantic subtlety. Figure 3.12 and Figure 3.13 provide the bird’s-eye view of overall attention maps.

Table 3.11 shows pseudo events in year 2015 month 1. We can see pseudo event #2 and #3 got higher attentions from the model. After looking at these two sentences carefully, we found they are both focused on company internal activities that lead to change in cash flow and profit. For example, event #2 is talking about Expedia expanded its business via a partnership with
Travelocity. And under the partnership, Expedia’s responsibility is to provide customer services for Travelocity and support Travelocity’s website. Both responsibilities imply more revenue hence more cash flow and net profit for Expedia. If we move to another high-attention pseudo event #3, we can see this sentence is talking about how Expedia updates its strategies, i.e. bring down room prices and take commissions from hotels, to generate more revenues. It means our event attention model pays higher attention to the pseudo events that are talking about the firm’s profit generation (1a) activities and internal cost control (1b) activities. Consistent with human intuition, those activities have larger impacts when forecasting a firm’s future cash flow margin and net profit margin.

Table 3.11: Events Attention Visualization Example: Expedia Month 1

<table>
<thead>
<tr>
<th>Event</th>
<th>cfm attention weight</th>
<th>npm attention weight</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.18</td>
<td>0.19</td>
<td>The company, which develops software for firms such as Barclays Plc and Expedia, is headquartered in the U.S. and has about a third of its 11,000 employees in Russia and Ukraine. ...</td>
</tr>
<tr>
<td>2</td>
<td>0.24</td>
<td>0.20</td>
<td>Under that partnership, Expedia provided customer service for Travelocity and supported its websites in the U.S. and Canada. (1a)</td>
</tr>
<tr>
<td>3</td>
<td>0.23</td>
<td>0.24</td>
<td>Online travel agents and intermediaries such as Priceline Group and Expedia are pushing down room prices and taking commissions from hotels, while enabling travelers to organize trips individually, poaching customers from tour operators. (1a, 1b)</td>
</tr>
<tr>
<td>4</td>
<td>0.19</td>
<td>0.19</td>
<td>At least 30 people will be on hand to lead workshops and provide advice, including Di-Ann Eisnor, ...; John Malloy, ...; Jerry Engel, ...; and Sam Friend, ...; now owned by Expedia.</td>
</tr>
<tr>
<td>5</td>
<td>0.16</td>
<td>0.19</td>
<td>You will now receive the Game Plan newsletter Last year, Google licensed hotel-booking software from Room 77, a startup backed by Expedia.</td>
</tr>
</tbody>
</table>

In Table 3.12, we can see pseudo event #3 talks about an acquisition activity (1c) of Expedia. Obviously, merger and acquisition activities will impact on company’s cash flow, hence this pseudo event got higher attention. In event #4, the pseudo event talks about an Expedia strategy: dis-

---

10 Code used in Figure 3.14. To be discussed later.
playing travel agents by packaging and selling discounted airfares and hotel rooms. This strategy is likely to increase company sales and lead to higher cash flow and profits.

Table 3.12: Events Attention Visualization Example: Expedia Month 2

<table>
<thead>
<tr>
<th>Event</th>
<th>cfm attention weight</th>
<th>npm attention weight</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.20</td>
<td>0.18</td>
<td>Expedia intends to keep the Orbitz brand intact, Khosrowshahi said on the call.</td>
</tr>
<tr>
<td>2</td>
<td>0.20</td>
<td>0.14</td>
<td>Today, Expedia landed one for its shareholders.</td>
</tr>
<tr>
<td>3</td>
<td>0.23</td>
<td>0.27</td>
<td>TripAdvisor surged 24 percent after Expedia agreed to acquire Orbitz Worldwide. (1c)</td>
</tr>
<tr>
<td>4</td>
<td>0.25</td>
<td>0.28</td>
<td>The online booking company Expedia displaced travel agents by packaging and selling discounted airfares and hotel rooms. (1a, 1b)</td>
</tr>
<tr>
<td>5</td>
<td>0.11</td>
<td>0.13</td>
<td>Tarran Vaillancourt, an Expedia spokeswoman, declined to comment.</td>
</tr>
</tbody>
</table>

In Table 3.13 we can see several external impact factors mentioned in month 4. Sentence 1 talks about Google’s search engine impact on Expedia and its competitors (2b1, 2b2). More visibility in searched results on Google will likely generate more sales for Expedia. Sentence 5 talks about a newcomer, the world’s biggest online retailer, is trying to compete with Expedia to gain some market shares from Expedia. This pseudo event implies the impact is not only applied to Expedia itself but also to the entire industry.

In Table 3.14 we can see that the models corresponding to different target variables, cfm and npm, have different attention weights on different pseudo events. Pseudo event #1 talks about another company bought a large amount of Elong’s stock shares from Expedia. This market activity is likely to have a larger impact on Expedia’s cash flow. Pseudo event #2 talks about Goldman’s announcement about Expedia. Pseudo event #4 talks about Expedia’s stock share’s sell in biotechnology industry and small-caps. Output variable cfm focus more on event 1 and 2 while npm focus more on event 2 and 4. All pseudo events 1, 2, and 4 are external impacts (2c) to Expedia, and they are all events in stock market.
Table 3.13: Events Attention Visualization Example: Expedia Month 4

<table>
<thead>
<tr>
<th>Event</th>
<th>cfm attention weight</th>
<th>npm attention weight</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.26</td>
<td>0.26</td>
<td>Microsoft, Expedia, publishers and others have asked the EU to examine complaints that Google favors its own services over competitors and hinders specialized search engines that compete with it. (2b1, 2b2)</td>
</tr>
<tr>
<td>2</td>
<td>0.18</td>
<td>0.21</td>
<td>Expedia is joining U.S. issuers from Coca-Cola to Berkshire Hathaway, which sold notes in the shared currency this year as European Central Bank stimulus drives down funding costs in the region.</td>
</tr>
<tr>
<td>3</td>
<td>0.20</td>
<td>0.20</td>
<td>With Google commanding almost all of the search market in some European countries, critics including Microsoft and Expedia are fed up with the company, which they say highlights its own Web services in query results at the expense of rivals.</td>
</tr>
<tr>
<td>4</td>
<td>0.10</td>
<td>0.12</td>
<td>Expedia climbed after quarterly revenue exceeded estimates.</td>
</tr>
<tr>
<td>5</td>
<td>0.25</td>
<td>0.22</td>
<td>The world’s biggest online retailer by revenue will be competing with Priceline Group, Expedia, startup Airbnb and others for a piece of the online hotel booking market. (2b1, 2b2)</td>
</tr>
</tbody>
</table>

Table 3.14: Events Attention Visualization Example: Expedia Month 5

<table>
<thead>
<tr>
<th>Event</th>
<th>cfm attention weight</th>
<th>npm attention weight</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.20</td>
<td>0.15</td>
<td>Ctrip, which operates the country’s biggest travel-booking website, bought the Elong stake from U.S.-based Expedia, becoming its biggest shareholder. (2c)</td>
</tr>
<tr>
<td>2</td>
<td>0.30</td>
<td>0.27</td>
<td>Goldman talks about here include Caterpillar, Coca Cola, Phillips 66, United Technologies, Automatic Data Processing, and Expedia. (2c)</td>
</tr>
<tr>
<td>3</td>
<td>0.17</td>
<td>0.21</td>
<td>Stocks rose, with the Standard &amp; Poor’s 500 Index paring a weekly loss, as Gilead Sciences and Expedia rallied after Thursday’s selloff in biotechnology and small-cap shares. (2c)</td>
</tr>
<tr>
<td>4</td>
<td>0.16</td>
<td>0.22</td>
<td>Stocks pared a weekly loss on Friday, as Gilead Sciences and Expedia rallied after Thursday’s selloff in biotechnology and small-cap shares. (2c)</td>
</tr>
<tr>
<td>5</td>
<td>0.17</td>
<td>0.15</td>
<td>Expedia reached a record, rising 6.7 percent for the fifth straight gain and the longest streak since January.</td>
</tr>
</tbody>
</table>
In Table 3.15 we can see the previously mentioned Elong stock share sell event #2, which relates to stock market (2c), gets more attention. Event #5 receives high attention as well, due to the discussion of business environment (2b1, 2b2).

Table 3.15: Events Attention Visualization Example: Expedia Month 7

<table>
<thead>
<tr>
<th>Event</th>
<th>cfm attention weight</th>
<th>npm attention weight</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.13</td>
<td>0.17</td>
<td>Expedia jumped 13 percent and Amgen added 2.9 percent as results beat estimates.</td>
</tr>
<tr>
<td>2</td>
<td>0.26</td>
<td>0.24</td>
<td>In May, Expedia sold its stake in ELong, a Chinese online travel company. (2c)</td>
</tr>
<tr>
<td>3</td>
<td>0.12</td>
<td>0.17</td>
<td>Amgen rallied 2.9 percent and Expedia jumped 13 percent on better-than-estimated earnings.</td>
</tr>
<tr>
<td>4</td>
<td>0.16</td>
<td>0.14</td>
<td>Expedia also topped a record, leading consumer-discretionary shares higher after second-quarter sales and profit topped analysts’ estimates.</td>
</tr>
<tr>
<td>5</td>
<td>0.33</td>
<td>0.29</td>
<td>Other homegrown technology companies include Zillow Group, Expedia and Zulily. (2b1, 2b2)</td>
</tr>
</tbody>
</table>

Table 3.16 is another example of different target variable allocating different attention weights to pseudo events. The output variable cfm pays more attention to event 1 and 2 while the output variable npm pays more attention to event 1 and 3. Pseudo event 1 and 2 are talking about external impact factors to the entire business environment (2b1, 2b2), and those impacts will lead to an impact on Expedia’s cash flow. Meanwhile, pseudo event 1 and 3 are talking about the competition and direct external impact on Expedia, hence lead to the impact on Expedia’s net profit. The attention models capture the important and related pseudo events precisely.

In Table 3.17, the highlighted pseudo events focus on external impact factors. Pesudo event 2 and 3 are talking about external partnership (2a) of Expedia. The business partnership is likely to impact Expedia’s revenue and cash flow. Pseudo event 5 is talking about stock market movement (2c) for Expedia, which may have impact on Expedia’s net profits.

Finally, what’s extremely exciting about this somewhat lengthy case study is that we realize event attention maps in our models (and attention maps in general) not only can serve as sense-making tools, but also can facilitate concept and hypothesis generation. Figure 3.14 illustrates
Table 3.16: Events Attention Visualization Example: Expedia Month 8

<table>
<thead>
<tr>
<th>Event</th>
<th>cfm attention weight</th>
<th>npm attention weight</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.26</td>
<td>0.26</td>
<td>With established tour operators facing competition from low-cost airlines and online booking sites such as Expedia.com, Thomas Cook and TUI are under pressure to invest in new offerings. (2b1, 2b2)</td>
</tr>
<tr>
<td>2</td>
<td>0.24</td>
<td>0.19</td>
<td>Google’s arguments were countered by Thomas Vinje, a lawyer with Clifford Chance who represents FairSearch Europe, whose members include Microsoft, Expedia and Nokia Oyj. (2b1, 2b2)</td>
</tr>
<tr>
<td>3</td>
<td>0.22</td>
<td>0.22</td>
<td>Fighting Competition Priceline, whose sites include Booking.com and Kayak, has struck partnerships and made purchases to fend off competition from Google and Expedia. (2a)</td>
</tr>
<tr>
<td>4</td>
<td>0.12</td>
<td>0.13</td>
<td>Expedia’s shares rose to a record on July 31 after second-quarter sales and profit topped analysts’ estimates.</td>
</tr>
<tr>
<td>5</td>
<td>0.17</td>
<td>0.20</td>
<td>When asked whether Priceline has seen mounting competition from Google or Expedia, Huston said there has not been a tangible change.</td>
</tr>
</tbody>
</table>

Table 3.17: Events Attention Visualization Example: Expedia Month 11

<table>
<thead>
<tr>
<th>Event</th>
<th>cfm attention weight</th>
<th>npm attention weight</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.15</td>
<td>0.19</td>
<td>Expedia lost 2.6 percent to $122.01, paring an earlier decline of 3.6 percent.</td>
</tr>
<tr>
<td>2</td>
<td>0.22</td>
<td>0.18</td>
<td>Expedia has had a partnership with HomeAway for two years, Khosrowshahi said in Wednesday’s statement. (2a)</td>
</tr>
<tr>
<td>3</td>
<td>0.23</td>
<td>0.21</td>
<td>Bringing HomeAway into Expedia’s portfolio of brands ’is a logical next step,’ Khosrowshahi said. (2a)</td>
</tr>
<tr>
<td>4</td>
<td>0.22</td>
<td>0.18</td>
<td>Last week, travel-booking site Expedia agreed to acquire vacation-rental company HomeAway for $3.9 billion.</td>
</tr>
<tr>
<td>5</td>
<td>0.18</td>
<td>0.24</td>
<td>Online travel company Expedia rebounded 1.9 percent after sliding 2.9 percent Tuesday, while Marriott International and Carnival each added 1.2 percent. (2c)</td>
</tr>
</tbody>
</table>
a tentative concept hierarchy that emerged from our case study of Expedia, upon consolidating insights gained from examining various attention maps (the codes map to previously discussed pseudo event examples).

We can see two high-level impact factors can impact a firm's cash flow margin and net profit margin: internal and external factors. Three kinds of firm internal activities can affect firm's cfm and npm.

External factors can include three facets as well. One facet is companies’ partnership that can integrate companies’ strength to gain more market share. Larger market share likely lead to larger profits. The second external factor is business environment change, which can impact the entire industry or the focal company alone. The third external impact factor is the firm’s stock market activities.
CHAPTER 4. WHAT’S NEXT? HARNESSING AI TO FORECAST FIRM MATERIAL EVENT SEQUENCES

Shuang Zhai and Zhu Zhang

Ivy College of Business, Iowa State University, Ames, IA, USA

Modified from a manuscript to be submitted to Information Systems Research

4.1 Abstract

In this paper, we formulate a unique business problem previously considered unapproachable, and we propose a novel solution that considers the real-business scenario for the task. In particular, we propose a Transformer model to forecast the firm’s material event series, based on its SEC 8-K current reports. Our proposed model demonstrates forecasting improvements over traditional sequence-to-sequence models and task-specific Markov Chain Monte Carlo simulations.

Our paper extends the literature on financial AI models and corporate management strategy. We also demonstrate the inner working of the proposed model to increase user perception and understanding of the model. It is a significant step forward on how to utilize and decode AI models for real business scenarios.

Keywords: Artificial Intelligence, Natural Language Processing, Corporate Event Sequence

4.2 Introduction

“The medium is the message.”

- Marshall McLuhan

How to quickly understand a company’s underlying dynamics are essential to investors. We leverage a firm’s 8-K financial reports - a textual based medium - to decode messages the firm sends to its investors. Particularly, we propose a deep learning approach to identify a firm’s important
material events and forecast its subsequent material event sequence, from 8-K reports. A material
event is defined as a matter if there is a substantial likelihood that a reasonable person would
consider it important, and a "rule of thumb" impact scale of a material event is five to ten percent
of net income. ¹

Investors can use a company’s historical Corporate Event Sequence (CES) to project its future
CES. For instance, an acquisition is a sign of financing, and insufficient funding is a vane for
refinancing. Similarly, a failing operational decision can bring an executive personnel change,
and a new senior-level appointment can be expected after that. CES embodies information that
consistency to corporate strategy and continuity to time; therefore, it is equipped to illuminate a
pathway for organizational strategy maze for the outsiders.

Publicly traded companies need to file reports to the U.S. Security and Exchange Commission
(SEC) regularly. The 8-K report, also called 'material event report' or 'current report,' should be
submitted by companies when certain types of corporate events occur. In other words, an 8-K
report should be filed when a public company has an event that its shareholders should be aware
of.

Corporate events embody firm strategies, short-term plans, long-term plans, and management
decision-making processes. Therefore, CES can be viewed as a tool to probe organizational blueprint
and managerial decision-making behaviors. However, the barrier for investors in understanding
corporate events well never go away. Instead, the content expansion in financial reports makes
investors’ costs to grasp a firm’s underlying dynamics increased dramatically, especially for non-
professional investors.

Meanwhile, artificial intelligence (AI) has drawn attention to businesses and became the new
frontier of business development and innovation. AI models are also applied in today’s real com-
panies for analytic and prediction. However, how to interpret an AI model remains a challenge.

of Financial Accounting Concepts No. 2, Qualitative Characteristics of Accounting Information ("Concepts Statement
As a result, we provide a tool to connect investors and financial reports. Specifically, we develop a model to automatically read corporate reports, understand its semantics, reveal essential corporate events, unleash company decision-making process logic, and forecast future event series, based on the AI identified covert patterns. More importantly, we demonstrate the inner-working of our model attentions, and it can be used to disclose influential events and corporate events relationships. We illustrate the function flow of our model in Figure 4.1.

In summary, we use a state-of-art AI technique to solve a real-world problem that was not formulated in the proposed way before and provide interpretation capability to facilitate the model understanding.

4.3 Literature Review

4.3.1 Corporate Reporting

For an 8-K corporate report, various items\(^2\) are required to be filed in it. Therefore, studies tried to categorize 8-Ks into different categories. Zhao (2016) classified Form 8-K into seven categories. Among those categories, information about business and operations, financial information, corporate governance and management, events related to Regulation Fair Disclosure, and other events, are considered the top five 8-K categories which cover more than 95% of all the reports in their study.

Feuerriegel and Pröllochs (2018) used latent Dirichlet allocation (LDA) method and categorized 8-K reports into topics: energy sector, insurance sector, change of trustee, real estate, corporate structure, loan payment, amendment of shareholder rights, earnings results, securities sales, stock option award, credit rating, income statements, business strategy, securities lending, management change, health care sector, tax report, stock dilution, mergers and acquisitions, and public relations. Earnings results and public relations are the top two topics in their study.

8-K reports include voluntary items (Item 2.02, 7.01, and 8.01) and mandatory items. He and Plumlee (2019) categorized voluntary items into a business combination, conference presentations,\(^2\)https://www.sec.gov/fast-answers/answersform8khtm.html
Figure 4.1: Function Architecture of Our Model

dividend announcements, litigation, patents, restructuring, security offerings, share repurchase, and shareholder agreement.

SEC has set up rules for companies to report their material events \(^3\). While we assume firms all follow SEC’s rules to file reports, Bird et al. (2018) show firms are more likely to strategically misclassify events into 'other events' category (Item 8.01) when the events are semantically negative, or the market intention is high. As a consequence, the misclassification will generate lower search

\(^3\)https://www.sec.gov/fast-answers/answersform8khtm.html
traffic online and smaller market reaction. This study provides a hint that when an event is categorized as ‘other events,’ it has a chance to be better categorized into another category.

Another way to categorize 8-K reports is by sentiments. Goldstein and Wu (2015) categorized events in 8-K reports into good events and bad event. The authors found a firm’s stock return is affected by its report’s sentiment and timeliness of the disclosure. Particularly, good news will lower the firm’s abnormal return, and bad news will increase the firm’s abnormal return.

Public firms also need to submit their 10-K report every year. Dyer et al. (2017) used Latent Dirichlet Allocation (LDA) method to categorize 10-K reports into topics. Given the increasingly longer 10-K reports, authors found governmental requirements from the Financial Accounting Standards Board (FASB) and SEC can explain most of the length increase in 10-K reports. Particularly, fair value, internal control, and risk factor disclosures can account for almost all of the length increase.

However, 10-K annual reports and 10-Q quarterly reports have apparent and significant drawbacks compared to 8-Ks. First, 10-K/Qs are designed to cover a mixed category of information. It is easy for them to plunge lower readability and create higher barriers for amateur readers. While the length of 10-K/Qs gets longer and longer (Cazier and Pfeiffer, 2015; Dyer et al., 2017), not all investors have the skill to decipher the insightful message from the lengthy 10-K/Qs. Second, when understanding difficulties are encountered in 10K/Qs, most retail investors do not have enough resources as advanced institutional investors do. Third, 10-K/Qs are released a long time after the event and get considerably prolonged-release intervals. It means investors have to wait one quarter or even longer to see the updated release from the company.

4.3.2 Corporate Strategy

Corporate events can reveal a company’s short-term and long-term strategies, and corporate strategy alters a company’s complication level and has an impact on the company’s inter-working (Tanriverdi and Du, 2020). Bowman and Helfat (2001) show corporate strategy does matter for corporate profitability. Capron and Pistre (2002) found in a merger and acquisition (M&A) event,
the acquirer is expected to earn abnormal return only when it transfers its resources to the target company, instead of receiving resources entirely from the target. Because jointly using resources indicates the uniqueness of the acquirer’s resources and the worth of the added value from the synergy. Berchicci et al. (2012) show companies with superior environmental capacities are significantly likely to acquire companies with similar physical facilitates but inferior environmental capacities.

Furthermore, Ozilgin and Penno (2005) found market leader firms’ operation success disclosures can impact their follower’s operation decision. Because follower companies can make a competitive advantage over the leading company based on the leader company’s disclosure, it is a strategic decision for leading companies to handle their operational choice and financial disclosures.

We summarize the related corporate strategy literature in Table 4.1.

<table>
<thead>
<tr>
<th>Study</th>
<th>Event Type</th>
<th>Trigger</th>
<th>Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bowman and Helfat (2001)</td>
<td>profitability</td>
<td>corporate strategy</td>
<td>variance decomposition</td>
</tr>
<tr>
<td>Capron and Pistre (2002)</td>
<td>profitability</td>
<td>merger company transfers its resources to the target company</td>
<td>ordinary least squares (OLS) regression</td>
</tr>
<tr>
<td>Berchicci et al. (2012)</td>
<td>acquisition</td>
<td>superior environmental capacities</td>
<td>conditional logistic regression</td>
</tr>
<tr>
<td>Ozilgin and Penno (2005)</td>
<td>operation decision</td>
<td>market leader firm’s operation success disclosure</td>
<td>game theory</td>
</tr>
</tbody>
</table>

4.3.3 Event Study

Event study, particularly for the company’s stock return, has been studied extensively. Grewal et al. (2018) found mandatory non-financial disclosure, environmental, social, and governance (ESG) disclosures especially, will lead to negative (positive) five-day abnormal cumulative stock return (centered on the event day) for weak (strong) pre-regulation ESG performance companies. Their study demonstrated non-financial disclosure could be used by equity holders to infer their future costs and benefits. Their study is consistent with the equity market perceives cost (benefit) for weak (strong) non-financial performance firms when they file non-financial mandatory disclosure.
Brandt et al. (2008) proposed a portfolio forming strategy to form portfolio based on firm’s return reaction on its earnings announcement. Authors show the strategy formed portfolio can generate Earning Announcement Return (EAR) drift, since EAR can capture broader unexpected information of the company. Moreover, the EAR strategy has no reversal effect after the 3rd quarter. We summarize the related event study literature in Table 4.2.

Furthermore, studies show 8-K report filing as an event will affect the firm’s abnormal return, volume, and volatility. Lerman and Livnat (2010) exam the effect of 8-K report regime update on August 23, 2004. The authors found that required disclosure items can generate a firm’s abnormal volume and return volatility. It validates the value of 8-K reports. Zhao (2016) used 8-Ks to show higher information intensity will reduce the company’s future return and future volatility. For mandatory disclosures, McMullin et al. (2015) showed the increased frequency of mandatory 8-Ks would impact the efficiency of price formation and increase the speed of information impounded into the stock price.

Bozanic et al. (2018) studies forward-looking statements (FLS) in earning releases, particularly contents in Item 2.02 - Results of Operations and Financial Conditions from 8-Ks. Authors found forecast-like FLS can lead to stronger investor response and more accurate analyst forecasts. Authors also suggest incorporating FLS measures into studies for a more comprehensive view of the firm’s voluntary disclosures.

4.3.4 Event Prediction

Studies exam how certain events can be predicted. Agarwal and Taffler (2008) used both market-based and accounting-ratio-based models to predict non-finance industry UK firms’ bankruptcy. Tian et al. (2015) used least absolute shrinkage and selection method (LASSO) to select important predictors and forecast corporate bankruptcy. Lyandres and Zhdanov (2013) showed investment opportunities are highly related to the likelihood of firm bankruptcy. Dietrich and Sorensen (1984) used Logit estimation to predict the probability of a firm to be a merger and acquisition target. We summarize the related single event prediction literature in Table 4.3.
Table 4.2: Event Study Related Literature

<table>
<thead>
<tr>
<th>Study</th>
<th>Dependent Variable</th>
<th>Event</th>
<th>Independent Variables</th>
<th>Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grewal et al. (2018)</td>
<td>cumulative</td>
<td>mandated</td>
<td>firm energy and climate change, firm human capital, firm governance quality, quantity of ESG disclosures, research and development expenditures, return on assets</td>
<td>cross section analysis</td>
</tr>
<tr>
<td></td>
<td>abnormal</td>
<td>non-financial disclosures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brandt et al. (2008)</td>
<td>earnings</td>
<td>announcement return</td>
<td>return on benchmark size, book-to-market Fama-French portfolio</td>
<td>random walk with drift</td>
</tr>
<tr>
<td></td>
<td>announcement</td>
<td>returns</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lerman and Livnat (2010)</td>
<td>abnormal</td>
<td>volume and return volatility</td>
<td>trading volume, stock price return volatility</td>
<td>regression</td>
</tr>
<tr>
<td>Zhao (2016)</td>
<td>return and</td>
<td>volatility</td>
<td>information intensity, book-to-market ratio, past month return, and past year return</td>
<td>cross-sectional regression</td>
</tr>
<tr>
<td></td>
<td>volatility</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>disclosure</td>
<td>increase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bozanic et al. (2018)</td>
<td>stock return</td>
<td>forward-looking statements</td>
<td>market value, book-to-market, sentence count, forward-looking sentences</td>
<td>OLS regressions</td>
</tr>
<tr>
<td></td>
<td>disclosure</td>
<td>disclosure</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.4 Related Modeling Techniques

4.4.1 AutoRegressive Integrated Moving Average (ARIMA) Model

ARIMA model (Pankratz, 1983) is the de facto standard model for univariate time series analysis in econometric. It has three components: an autoregression (AR) component on the variable itself, a differencing (I) component if the time series is not stationary, and a moving average (MA) component on error terms.

In ARIMA\((p, d, q)\) model, \(p\) denotes the order of the autoregression (AR), \(d\) denotes the order of differencing (I), and \(q\) denotes the order of the moving average (MA).

\[
(1 - \sum_{i=1}^{p} \phi_i L^i)(1 - L)^d Y_t = (1 - \sum_{i=1}^{q} \theta_i L^i) \epsilon_t
\]

\(\epsilon_t \sim N(0, \sigma^2)\)

, where \(L\) is the lag operator, and \(\epsilon_t\) is a white noise process.
### Table 4.3: Event Prediction Related Literature

<table>
<thead>
<tr>
<th>Study</th>
<th>Target Event Type</th>
<th>Input Data</th>
<th>Method/Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agarwal and Taffler (2008)</td>
<td>bankruptcy</td>
<td>non-finance industry UK firms listed on London Stock Exchange (LSE)</td>
<td>contingent claims model, z-score model</td>
</tr>
<tr>
<td>Tian et al. (2015)</td>
<td>bankruptcy</td>
<td>CRSP equity data, COMPUSTAT accounting data</td>
<td>discrete hazard mode and least absolute shrinkage and selection method (LASSO)</td>
</tr>
<tr>
<td>Dietrich and Sorensen (1984)</td>
<td>merger and acquisition</td>
<td>COMPUSTAT data</td>
<td>logistic model</td>
</tr>
</tbody>
</table>

If we use Box-Jenkins backshift operator, Equation 4.1 can be written as

\[(1 - B)^p (1 - B)^d Y_t = (1 - B)^q \epsilon_t \quad (4.2)\]

\[\epsilon_t \sim N(0, \sigma^2)\]

, where \(B\) is the backshift operator, \((1 - B)^p = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\), and \((1 - B)^q = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\). Equation 4.1 and Equation 4.2 can be also written as

\[\phi_p(B) \nabla^d Y_t = \theta_q(B) \epsilon_t \quad (4.3)\]

\[\epsilon_t \sim N(0, \sigma^2)\]

or

\[\phi(B) \nabla^d Y_t = \theta(B) \epsilon_t \quad (4.4)\]

\[\epsilon_t \sim N(0, \sigma^2)\]

, where \(\nabla\) is the difference operator, \(\phi(B) = \phi_p(B) = 1 - \sum_{i=1}^p \phi_i B^i\) is the \(p\)-order AR operator, \(\theta(B) = \theta_j(B) = 1 - \sum_{j=1}^q \theta_j B^j\) is the \(q\)-order MA operator, and \(\epsilon_t\) is a white noise process.
4.4.2 Recurrent Neural Networks (RNNs)

The time series model, such as ARIMA, is not designed to tackle sequence prediction problems. Even one can force the ARIMA model to conduct sequential predictions, such as predicting values one at a time in a row, the model is not well capable of handling the sequential dynamics, especially in the prediction phase.

Therefore, deep learning models, especially recurrent neural networks (RNNs), naturally render themselves to sequence prediction problem.

Time series models decompose the time series into trend, seasonal, and irregular components. Deep learning models take time series as a whole, compute representation, and infer patterns from an integrated and holistic nonlinear approach.

Gated Recurrent Units (GRUs) is a representative deep learning architecture in RNNs, and it was first proposed by Cho et al. (2014). Numerous works have used GRUs for natural language processing tasks, such as part of speech (POS) tagging, information extraction, syntactic parsing, speech recognition, machine translation (Bahdanau et al., 2015; Luong et al., 2015), and question answering (McCann et al., 2018).

Each GRU cell has an input vector $x_t$, output vector $h_t$, reset gate vector $r_t$, and update gate vector $z_t$, at each time step $t$. By including the sigmoid activation function $\sigma_g$ and the hadamard product $\odot$, we can write GRU as the following:

\[
r_t = \sigma_g(W_{ir} x_t + b_{ir} + W_{hr} h_{t-1} + b_{hr}) \tag{4.5}
\]

\[
z_t = \sigma_g(W_{iz} x_t + b_{iz} + W_{hz} h_{t-1} + b_{hz}) \tag{4.6}
\]

\[
n = \tanh(W_{in} x_t + b_{in} + r_t \odot (W_{hn} h_{t-1} + b_{hn})) \tag{4.7}
\]

\[
h_t = (1 - z_t) \odot n + z_t \odot h_{t-1} \tag{4.8}
\]
where $W_{ir}$, $W_{hr}$, $W_{iz}$, $W_{in}$, and $W_{hn}$ are weight matrices and $b_{ir}$, $b_{hr}$, $b_{iz}$, $b_{in}$, and $b_{hn}$ are biases.

### 4.4.3 Sequence-to-Sequence (Seq2Seq) Neural Network

The sequence-to-sequence model was introduced by Sutskever et al. (2014). It is widely used in machine translation tasks (Bahdanau et al., 2015; Luong et al., 2015), i.e., translating sentences from one language to another language, such as French to English. The Sequence-to-sequence model follows an encoder-decoder framework. There are two basic components in the encoder-decoder framework, the encoder, and the decoder. The encoder will encode the information embedded in the input sequence, convey the inferred information to the decoder, and the decoder will generate the output sequence. The framework can also be viewed as a completion task, i.e., giving the first half of the task and generating the second half of the task.

### 4.4.4 Attention Mechanism

The attention mechanism has been employed broadly in recent natural language processing (NLP) studies. The intuition for attention is to assign higher weights to sections where contain important information about the task. Attention will highlight the segments where the model should not overlook and should pay more attention to. Attentions can be applied to the word-level, sentence-level (Yang et al., 2016), aspect-level (Wang et al., 2016), or in an interactive approach (Ma et al., 2017).

### 4.4.5 Transformer Model

The attention mechanism can be used by itself, and it is called self-attention. Vaswani et al. (2017) used multi-head attention alone, i.e., the Transformer model, to solve the sequence prediction task, which is traditionally handled by other neural network techniques such as Recurrent Neural Networks (RNNs) and Convolution Neural Networks (CNNs). In the Transformer model, both encoders and decoders contain many self-attentions.
The transformer model follows the encoder-decoder framework. The sequential information from the input sequence is handled by positional encoding. A series of encoders form the full encoder, and a series of decoders stacked into the full decoder. Information is processed from the encoder and carried to all of the decoders in the decoder component. The output is generated from the last stacked decoder. We illustrated the information flow in a Transformer model in Figure 4.2.

![Figure 4.2: The Encoder and Decoder in Transformer Model](image)

In each encoder of the Transformer model, multiple self-attention heads are processed simultaneously on information from the input sequence. After going through the stacked encoders, learned information is brought to the encoder-decoder attention heads. Furthermore, the decoders also contain self-attention heads and process information after the encoder-decoder attention heads. Multiple decoders are stacked before making the output. We further illustrate the insides of encoders and decoders of a Transformer model in Figure 4.3.

Within each attention head in the Transformer model, the input is first projected by a linear transformation before processed by the attention head. Therefore, every attention head looks at information from the entire input sequence. After processing within each attention head, learned
information is concatenated before feeding into another linear projection. We illustrate the inner-working of the multi-head self-attention in Figure 4.4.

4.5 Multiple Research Gaps

There are several research gaps we can identify from previous literature.

First, researches increasingly reveal the merit of textual data in financial studies. However, most of them are centered on the financial market, such as stock price, return, and volatility (Fang and Peress, 2009; Tetlock, 2010; Edmans, 2011). Although market-related values are important to study, other indicators can also reflect a company’s inner-working and performance, such as actual corporate events.

Second, corporate events are reported in the form of written reports. Instead of treating binary to having an event or not, the textual contents in reports have a lot of information. However, the text has very high dimensionality, and it presents challenges to traditional econometric models.
Third, many textual data can be used to investigate firm performance. Scholars tried to render themselves to machine learning models, given the text’s high dimensionality. Deep learning models have applied to market prediction using news articles (Ding et al., 2014, 2015). However, they only considered a single event and didn’t recognize contents in financial reports.

Fourth, none of the previously mentioned studies focus on the nature of the events. Moreover, they did not model the inter-relationships among the corporate events and event sequences explicitly.

Why studying corporate event series is needed? Various stakeholders, not only investors but also management teams and regulators, can find CES useful. CES provides not only corporate strategy hints but also operation patterns. Given the continuous characteristic of CES, it notches up contents for stakeholders to achieve higher profits and a better company market position in a timely manner.

We contribute to the literature on five-folds.

• (1) we formulate a real-world problem that was not expressed in the proposed way before.
• (2) we enrich the understanding of Form 8-K Current Report.

• (3) we expand the use case of the state-of-art Transformer Model on a real-business problem and provide the state-of-art result.

• (4) we demonstrate the interpretations of the proposed deep learning model and contribute to the understanding of AI models.

• (5) we provide managerial and governance implications based on the state-of-art AI model to facilitate practical AI-based management in reality.

4.6 Corporate Event Sequence and Event Types

4.6.1 Corporate Event Sequence Prediction Task

Corporate event sequence is a series of corporate events that represent corporate decisions. From the order of corporate events, one can foresee what events are likely to happen within the company. For instance, if a firm announced a merger and acquisition, one can expect the firm to request a loan or other financial supports. Furthermore, a firm’s operation situation will impact the company’s decision on director and senior personnel appointments. From a firm’s historical event series, one can infer the company’s decision-making behavior and forecast their next steps. Corporate event series prediction is to use a firm’s historical events to predict what events are likely to happen in the future. In other words, we can consider corporate event series prediction as a story completion task. We view the historical event sequence as the first half of the story, and the prediction task is to forecast future events’ happening sequence as the second half of the story. The entire prediction task is to complete the firm’s events stories. We introduce corporate event types in the next two sections.
4.6.2 Material Event Items by SEC

SEC requires public traded companies to report certain corporate events in their 8-K reports. SEC describes the types of events in Form 8-K by two levels: Section and Item 4. We summarize SEC Sections (higher level category) and Items (lower level category) in Table 4.4.

Although SEC has its Sections (higher level category) and Items (lower level category), we can see from the table that they are not classified by the actual activity or event nature. For instance, finance-related events are in many sections. Also, the Other Events section is ambiguous. As Bird et al. (2018) pointed out that companies can strategically misclassify events into Other Events Section and purposely manipulate market reaction. Therefore, we propose our event types in Table 4.5 based on the activity types and nature of the events, such as business activities, financial activities, personnel change activities, etc., to reflect the actual meaning and purpose of the events.

4.6.3 Corporate Event Types in Our Study

Since the categories used by SEC reporting purpose did not satisfy users’ needs to grasp the essence of the events quickly, and they give the misclassification opportunity for companies, we propose our event types that are categorized based on company activities and events nature. The proposed event types will be used as the target variables in our study.

We first read thousands of 8-Ks by ourselves and designed taxonomies. We characterize 8-Ks into multiple event types, based on the understanding of the report contents and the event activity nature. Then, we map every 8-K to one of our proposed event types. We list our event types in Table 4.5.

Meanwhile, there are several implementation details we want to mention. First, since Item 9.01 is designated for Financial Statements and Exhibits, and it typically serves as the attachment one event type, we did not treat this Item as a standalone event. Second, since some reports can be filed under different Items, the mapping between Items and our Event Types is many to many. In other words, one Item can be seen in multiple Event Types. The bold Items in Table 4.5 are Items that

---

4 https://www.sec.gov/fast-answers/answersform8khtm.html
<table>
<thead>
<tr>
<th>Sec</th>
<th>Section Description</th>
<th>Item</th>
<th>Item Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Registrant’s Business and Operations</td>
<td>1.01</td>
<td>Entry into a Material Definitive Agreement</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.02</td>
<td>Termination of a Material Definitive Agreement</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.03</td>
<td>Bankruptcy or Receivership</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.04</td>
<td>Mine Safety</td>
</tr>
<tr>
<td>2</td>
<td>Financial Information</td>
<td>2.01</td>
<td>Completion of Acquisition or Disposition of Assets</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.02</td>
<td>Results of Operations and Financial Condition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.03</td>
<td>Direct Financial Obligation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.04</td>
<td>Increase a Direct Financial Obligation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.05</td>
<td>Exit or Disposal Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.06</td>
<td>Material Impairments</td>
</tr>
<tr>
<td>3</td>
<td>Securities and Trading Markets</td>
<td>3.01</td>
<td>Delisting, Transfer of Listing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.02</td>
<td>Unregistered Sales of Equity Securities</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.03</td>
<td>Material Modification to Rights of Security Holders</td>
</tr>
<tr>
<td>4</td>
<td>Accountants and Financial Statements</td>
<td>4.01</td>
<td>Changes in Registrant’s Certifying Accountant</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.02</td>
<td>Non-Reliance</td>
</tr>
<tr>
<td>5</td>
<td>Corporate Governance and Management</td>
<td>5.01</td>
<td>Changes in Control of Registrant</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.02</td>
<td>Senior Personnel Change</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.03</td>
<td>Articles, Bylaws, and Fiscal Year Amendments</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.04</td>
<td>Temporary Suspension of Trading</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.05</td>
<td>Code of Ethics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.06</td>
<td>Change in Shell Company Status</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.07</td>
<td>Vote of Security Holders</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.08</td>
<td>Shareholder Director Nominations</td>
</tr>
<tr>
<td>6</td>
<td>Asset-Backed Securities (ABS)</td>
<td>6.01</td>
<td>ABS Informational Material</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.02</td>
<td>Change of Servicer or Trustee</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.03</td>
<td>Change in Credit Enhancement or Support</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.04</td>
<td>Failure to Make a Required Distribution</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.05</td>
<td>Securities Act Updating Disclosure</td>
</tr>
<tr>
<td>7</td>
<td>Regulation FD</td>
<td>7.01</td>
<td>Regulation FD Disclosure</td>
</tr>
<tr>
<td>8</td>
<td>Other Events</td>
<td>8.01</td>
<td>Other Events</td>
</tr>
<tr>
<td>9</td>
<td>Financial Statements</td>
<td>9.01</td>
<td>Financial Statements and Exhibits</td>
</tr>
</tbody>
</table>
can be seen in more than one Event Types. For instance, Item 7.01 Regulation FD Disclosure and Item 8.01 Other Events can be used from companies to report a senior personnel change. Therefore, both Item 7.01 and Item 8.01 are considered in our PC: Personnel Change Event Type.

<table>
<thead>
<tr>
<th>ID</th>
<th>Event Type</th>
<th>Code</th>
<th>SEC Items</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Business Combination and Restructuring</td>
<td>BC</td>
<td>1.01, 1.02, 2.01, 7.01, 8.01</td>
<td>merger, acquisition, join venture, separation, spin-off</td>
</tr>
<tr>
<td>2</td>
<td>Financial Activities</td>
<td>FN</td>
<td>1.01, 1.02, 2.03, 2.04, 2.05, 2.06, 3.02, 6.01, 6.02, 6.03, 6.04, 6.05, 7.01, 8.01</td>
<td>lend, borrow, loan, Notes, payment, debt, stock, repurchase, dividend, asset-backed securities (ABS)</td>
</tr>
<tr>
<td>3</td>
<td>Operation Activities</td>
<td>OA</td>
<td>1.01, 1.02, 7.01, 8.01</td>
<td>operation, contract, consulting, service, product, supply</td>
</tr>
<tr>
<td>4</td>
<td>Senior Personnel Change</td>
<td>PC</td>
<td>1.01, 1.02, 5.02, 7.01, 8.01</td>
<td>executive officer/director, retire, leave, appointment</td>
</tr>
<tr>
<td>5</td>
<td>Information Disclosure</td>
<td>ID</td>
<td>2.02, 4.01, 4.02, 5.07, 5.08, 7.01, 8.01</td>
<td>conference, presentation, statement, exhibit</td>
</tr>
<tr>
<td>6</td>
<td>Document Update</td>
<td>DU</td>
<td>3.03, 5.01, 5.03, 5.05, 5.06, 7.01, 8.01</td>
<td>by-laws, code of ethics</td>
</tr>
<tr>
<td>7</td>
<td>Intellectual Property Activities</td>
<td>IP</td>
<td>1.01, 1.02, 7.01, 8.01</td>
<td>intellectual property, patent approval</td>
</tr>
<tr>
<td>8</td>
<td>Litigation and Lawsuit</td>
<td>LL</td>
<td>1.01, 1.02, 7.01, 8.01</td>
<td>settlement, litigation, lawsuit</td>
</tr>
<tr>
<td>9</td>
<td>Delisting, Trading suspension</td>
<td>DL</td>
<td>3.01, 5.04, 7.01, 8.01</td>
<td>delisting, trading suspension</td>
</tr>
<tr>
<td>10</td>
<td>Bankruptcy</td>
<td>BK</td>
<td>1.03, 7.01, 8.01</td>
<td>bankruptcy</td>
</tr>
<tr>
<td>11</td>
<td>None</td>
<td>NA</td>
<td></td>
<td>no material event happened</td>
</tr>
</tbody>
</table>

### 4.7 Task Definition

#### 4.7.1 Task Framework

Our task is based on a company’s historical event sequence to predict its future event sequence. It is a sequence-to-sequence prediction problem. Company events formed its historical event series, and we will use them to infer the company’s hidden decision-making rationale and predict its future event series. We illustrate the framework of our model in Figure 4.5.
In particular, we use corporate event sequences in history/memory $M$ to predict event sequences in forecasting horizon $H$. Memory $M$ and forecasting horizon $H$ are formed by smaller time intervals.

We can also view the CES prediction as a story completion task. Once we know what events have happened at the company in the past, we aim to complete the story by predicting what events are likely to occur in the future.

Figure 4.5: Time Structure of the Model

4.7.2 Task Formulation

Now, let’s formally define the task as,

$$y^{(i)}_q \in Y^H_{(i)}$$

(4.9)

$$y^{(i)}_q = \sum_{j=0}^{M} (E^{(i)}_{(i)j})$$

(4.10)

$$E^{(i)}_{(i)j} = g_{k=0}^{\mid K \mid} (S_{(i)jk})$$

(4.11)

where,

- $Y^H$ denotes the event sequence, $y$ denotes the event types in Table 4.5, and the sequence $Y^H$ is a series of events: $Y^H = [y_1, ..., y_H]$. 
• $M$ denotes the size of memory, and $H$ denotes the size of the forecasting horizon, both are measured in terms of the number of time windows.

• $ev$ denotes event index in Table 4.5, and $|Ev|$ denotes total number of event types.

• $i$ indexes companies.

• $j$ indexes time windows in $M$, and $j = \{1, ..., M\}$.

• $q$ indexes time windows in $H$, and $q = \{1, ..., H\}$.

• $|K|$ denotes the number of events per time window.

• $S_{(i)jk}$ is the event embedding (dense vector) of the $k$th event of company $C_i$ in time window $j$.

• $E_{(i)j}$ is the aggregate event embedding (dense vector) of company $C_i$ in time window $j$.

• $g$ is a function that aggregates multiple event embeddings into one embedding.

• $f$ is a learned function (our proposed model) that maps all event embeddings in Memory $M$ to forecasting horizon $H$.

In principle, $g$ and $f$ can be parameterized as any function approximator.

4.8 Model Formulations

4.8.1 Event Attention Layer for Model Input

Since multiple 8-Ks can be filed within the same time window, we select the top $|K|$ event embeddings for each company within each time window, based on the event embedding’s L2 norm value. We institute various treatments of function $g$ in Equation 4.11, such as attention mechanism. In the Event Attention Layer, we implement attention to the top $|K|$ event embeddings and obtain the weighted embedding at each time window.
We define our Event Attention Layer as,

$$\tilde{h}_{(i)j} = sigmoid(W_{(i)jk}S_{(i)jk} + b_{(i)j})$$  \hspace{1cm} (4.12)$$

$$\alpha_{(i)jk} = softmax(W_{(i)j}\tilde{h}_{(i)j})$$  \hspace{1cm} (4.13)$$

$$E_{(i)j} = \sum_{k=1}^{[K]} \alpha_{(i)jk}S_{(i)jk}$$  \hspace{1cm} (4.14)$$

The weighted sum context vector $E_{(i)j}$ is used as the aggregated semantic representation of the company $C_i$’s events in time window $j$. Next, $E_{(i)j}$ is used as model input and fed into the following models.

4.8.2 The GRU Model

Our task works on time sequences and is formed as a sequence-to-sequence problem. Recurrent Neural Networks (RNNs) are the typical deep learning models to handle sequential tasks. Therefore, we start the comparison models by using one of the RNNs - Gated Recurrent Units (GRUs) as the backbone. In particular, we use GRUs as the processing units in the encoder-decoder framework. We illustrate the GRU model in Figure 4.6.

In the model, the last hidden state of the encoder is directly connected to the decoder. In the decoder, the hidden state $h_{st}$ is used to predict the event type for each timestamp. We employ $sigmoid$ function to compute $y_t$ as,

$$y_t = sigmoid(W_{sth_{st}})$$  \hspace{1cm} (4.15)$$

4.8.3 The GRU Attention Model

We implement an Alignment Attention Layer in the GRU Attention model to align decoder and encoder. Particularly, information from the encoder is carried over to the decoder. To be able to capture what events happened in history play more roles in the prediction horizon, we employ
attention mechanisms (Bahdanau et al., 2015; Luong et al., 2015; Kadlec et al., 2016; Cui et al., 2017) to capture the dynamics. The decoder of the GRU Attention model looks at every hidden state in the encoder. The Alignment Attention Layer assigns attention scores to every hidden state in the encoder and aggregates them. We illustrate the GRU Attention model in Figure 4.7. We are inspired by Bahdanau et al. (2015), Luong et al. (2015), Kadlec et al. (2016) and Cui et al. (2017), and employed attention and self-attention mechanisms to compute event attentions for our corporate event sequence (CES) prediction task.

We follow the "general" approach in Luong et al. (2015) and obtain attention scores as,

$$score(h_{tr}, h_{sr}) = h_{tr}^T W_{h} h_{sr}$$

(4.16)

, where $h_{tr}$ is the hidden state of target sequence and $h_{sr}$ is the hidden state of the source sequence. The context vector $c_t$ is the weighted sum of the product of attention scores and the hidden states in the encoder.

$$c_t = \sum_{j=0}^{M} score(h_{tr}, h_{sr})_j h_{sr_j}$$

(4.17)
The updated decoder hidden state $h_{st}$ uses information from both context vector $c_t$ and the target sequence hidden state $h_{tr}$ for final prediction.

$$h_{st} = \tanh(W_c[c_t; h_{tr}])$$ (4.18)

, where $[; ;]$ denotes concatenation along the sequence dimension.

4.8.4 The Proposed CES Transformer Model

We use Transformer (Vaswani et al., 2017) as the backbone of our proposed model and train our event type embeddings. We illustrate our proposed model in Figure 4.8.

Before we feed data to the Transformer, we create our model inputs in Event Attention Layer, as illustrated in Section 4.8.1. Then, for the model to learn temporal relationships, we implement the positional encoding layer as,

$$PE_{pos,2i} = \sin(pos / 10000^{2i/d_{model}})$$ (4.19)

$$PE_{pos,2i+1} = \cos(pos / 10000^{2i/d_{model}})$$ (4.20)
where $pos$ is the position of the input in sequence, $i$ is the dimension of event embedding, and $d_{model}$ is the dimensionality of the event embedding.

In the encoder, $E_{(i)j}$ is going through the multi-head attention blocks (with layer normalization, dropout, and residual connection (He et al., 2016)) and feed-forward neural networks multiple times. We follow the Transformer scaled dot-product attention approach and assign the attention scores as,

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$  \hspace{1cm} (4.21)
where $Q$, $K$, $V$ are inputs to attention, i.e., $E_{(i)j}$, $h$ is number of parallel attention heads, and $d_k = d_{\text{model}}/h$.

The position-wise feed-forward network is defined as,

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2$$

(4.22)

where $W_1$ and $W_2$ are parameter matrices and $b_1$ and $b_2$ are biases.

Outputs from different attention heads are concatenated as,

$$MultiHead(Q, K, V) = \text{Concat}(\text{head}_1, ..., \text{head}_h)$$

(4.23)

$$\text{head}_i = Attention(QW_{Qi}, KW_{Ki}, VW_{Vi})$$

(4.24)

where $W_{Qi}$, $W_{Ki}$, and $W_{Vi}$ are parameter matrices.

The information from the encoder is carried over to the decoder. Besides self-attentions from the decoder inputs, every decoder looks at the output from the encoder and assigns attention scores to encoder outputs. The encoder-decoder attention is implemented in the second multi-head attention in the decoder.

After the decoder, we employ sigmoid function to compute $y(i)_q$ as,

$$y(i)_q = \text{sigmoid}(W_d h_{dec})$$

(4.25)

where $W_d$ is parameter matrix and $h_{dec}$ is the output from $N$ stacked decoder blocks.

### 4.9 Experiments

#### 4.9.1 Evaluation Models

We evaluate the following models in our experiments:

- **MCMC**: Markov Chain Monte Carlo (MCMC) simulation, the baseline model. (Details are discussed in 4.9.1.1.)
• **GRU Model**: event sequence as input, and no attention.

• **GRU Attention Model**: event sequence as input, and with attention.

• **CES Transformer Model (CEST)**: event sequence as input, Transformer model.

### 4.9.1.1 Markov Chain Monte Carlo (MCMC)

For every company, we gathered its event sequence through the training period, and construct an event transition matrix.

Given the obtained transition matrix, we can implement the Markov Chain Monte Carlo (MCMC) simulations. In particular, we view each row of the transition matrix as the event type probability distribution. During the simulation, we first recognize event types at the last training timestamp as the current event type $E_t$. Second, based on event frequency distribution $Freq$, we draw the number of events $freq_t$ for this timestamp. Third, we draw the next event type $E_{t+1}$ based on $E_t$’s probability distribution, for $freq_t$ times. Fourth, if there is more than one event, we average the draw event distributions and use the result as the next timestamp’s event distribution. By making this approach repetitively, we can sample $E_{t+2}$ based on $E_{t+1}$’s probability distribution, and so on. Finally, we reach $E_{t+H}$ and complete the sampling process.

In the experiment, for each event type, we sample its sequences 50 times, and we use the averaged performance as our model baseline.

\[
Pr(E_t) = f(E_t|E_{t-1})
\]  
(4.26)

### 4.9.2 Evaluation Methods

#### 4.9.2.1 Per Type Evaluation

Based on the sigmoid results, we use a threshold number to convert the predicted value to its binary format at each time $t$ in $H$ as,
Algorithm 1: Markov Chain Monte Carlo (MCMC) Simulation

Data: 1. transition matrix \( \text{Dist} \) created by company sequences from training period
2. event frequency distribution \( \text{Freq} \) created based on event frequency in training period

Input : the event types in the last time step in \( M \)
Output: simulated event sequences

repeat
   while data sequence is in testing data set do
      potential_first_input ← true_event_types_{t-1} ;
      if potential_first_input is null then
         initial_event_type ← 10
      else
         while initial_event_type in potential_first_input do
            event_input ← initial_event_type ;
            for \( t ← 0 \) to target_sequence_length do
               freq_t ← Freq ;
               for \( i ← 0 \) to freq_t do
                  averagedDist ← Dist(event_input) ;
               
               event_input ← averagedDist ;
               for each event type do
                  compute \( tp, tn, fp, \) and \( fn \) ;
            
         end
         until simulated times complete;
      
      precision ← total_tp, total.fp ;
      recall ← total_tp, total_fn ;
      fmeasure ← precision, recall ;

   end

\[
\text{Eval}_{ev,t} = \begin{cases} 
1 & \text{if } y_{ev,t} > \text{threshold} \\
0 & \text{otherwise} 
\end{cases}
\]

In experiments, we use the natural threshold value of 0.5. We evaluate classification performance on each event type. In particular, we compute classification criteria, i.e., precision, recall, and F1 score, for every event type.

Furthermore, predicting event types correctly within a reasonable temporal approximate period is also important and meaningful for the event series prediction task in reality. Therefore, we use two approaches, i.e., *precise evaluation* and *fuzzy evaluation* to evaluate our models.
For every event type, we compute its true positive (TP), false positive (FP), false negative (FN), and true negative (TN) for both approaches, and evaluate our model’s precision (Pr), recall (Re), and F-measure (F1).

**Precise Evaluation**: we compute the confusion matrix based on $Eval_{ev,t}$ for each event type as,

$$
Precision_{ev,t}(Pr_{ev,t}) = \frac{TP_{ev,t}}{TP_{ev,t} + FP_{ev,t}}
$$

(4.27)

$$
Recall_{ev,t}(Re_{ev,t}) = \frac{TP_{ev,t}}{TP_{ev,t} + FN_{ev,t}}
$$

(4.28)

$$
F1_{ev,t} = \frac{2 \times Pr_{ev,t} \times Re_{ev,t}}{Pr_{ev,t} + Re_{ev,t}}
$$

(4.29)

**Fuzzy Evaluation**: because predicting event correctly close to exact time $t$ also has practical implications, i.e., forecasting accurately of the event type close to time $t$ is also useful, we compute the confusion matrix within $[t - z, t + z]$ time window, and $z \in [1, 2, \ldots]$. We update the true positive for precision and recall for $t \in [t - z, t + z]$, and re-compute the precision, recall, and F1 measures for $t \in [t - z, t + z]$ as,

$$
Pr_{ev,[t-z,t+z]} = \frac{TP_{ev,[t-z,t+z]}}{TP_{ev,[t-z,t+z]} + FP_{ev,t}}
$$

(4.30)

$$
Re_{ev,[t-z,t+z]} = \frac{TP_{ev,[t-z,t+z]}}{TP_{ev,[t-z,t+z]} + FN_{ev,t}}
$$

(4.31)

$$
F1_{ev,[t-z,t+z]} = \frac{2 \times Pr_{ev,[t-z,t+z]} \times Re_{ev,[t-z,t+z]}}{Pr_{ev,[t-z,t+z]} + Re_{ev,[t-z,t+z]}}
$$

(4.32)

Precise evaluation is a special case of fuzzy evaluation when $z=0$. We report all precision evaluation and fuzzy evaluation $z \in \{1, 2, 3\}$ results in the following sections.
4.9.2.2 Binary Cross Entropy and Perplexity

We also care about per event type’s performance at every timestamp. Therefore, we use Binary Cross Entropy (BCE) to measure the bits needed to identify one event type at every time $t$.

$$BCE = -\frac{1}{H} \sum_{t=1}^{H} \sum_{ev=1}^{|Ev|} y_{ev,t} \log \hat{P}(y_{ev,t}) + (1 - y_{ev,t}) \log \hat{P}(1 - y_{ev,t})$$  \hspace{1cm} (4.33)

Then, we use perplexity to measure how well the event sequences are predicted by the model, as the following,

$$Perplexity = 2^{BCE}$$  \hspace{1cm} (4.34)

Both BCE and perplexity can be used to further assess the model performance.

4.10 Model Implementation

4.10.1 Data

4.10.1.1 Data Description

There was an item schema update in Aug 2004. To be consistent, we use 8-K current reports filed to SEC’s EDGAR system (the Electronic Data Gathering, Analysis, and Retrieval system) between August of the year 2004 and the year 2018 as our data. Our study focuses on the Fortune 1,000 companies, and we use them as our focal companies.

4.10.1.2 Data Distribution

We show the event type distribution and event sequence distribution in Table 4.6 and Table 4.7, respectively. Both tables show the event type distribution and event frequency distribution for the training set, validation set, and testing set are similar. Therefore, the learning model trained from data in the training set is meaningful to be applied to data in the validation set and testing set.
Table 4.6: Event Type Distribution

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Event Type</th>
<th>BC</th>
<th>FN</th>
<th>OA</th>
<th>PC</th>
<th>ID</th>
<th>DU</th>
<th>LL</th>
<th>DL</th>
<th>BK</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>Count</td>
<td>65789</td>
<td>159395</td>
<td>13611</td>
<td>17778</td>
<td>303445</td>
<td>38151</td>
<td>4688</td>
<td>10516</td>
<td>4083</td>
<td>372</td>
</tr>
<tr>
<td>Train</td>
<td>Count%</td>
<td>6.47%</td>
<td>15.67%</td>
<td>1.34%</td>
<td>17.48%</td>
<td>29.84%</td>
<td>3.75%</td>
<td>0.46%</td>
<td>1.03%</td>
<td>0.40%</td>
<td>0.04%</td>
</tr>
<tr>
<td>Valid</td>
<td>Count</td>
<td>18307</td>
<td>44277</td>
<td>3161</td>
<td>49457</td>
<td>93965</td>
<td>9640</td>
<td>911</td>
<td>2306</td>
<td>621</td>
<td>14</td>
</tr>
<tr>
<td>Valid</td>
<td>Count%</td>
<td>6.18%</td>
<td>14.96%</td>
<td>1.07%</td>
<td>16.70%</td>
<td>31.74%</td>
<td>3.26%</td>
<td>0.32%</td>
<td>0.78%</td>
<td>0.21%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Test</td>
<td>Count</td>
<td>18085</td>
<td>43604</td>
<td>2778</td>
<td>40519</td>
<td>92835</td>
<td>9863</td>
<td>911</td>
<td>2236</td>
<td>674</td>
<td>24</td>
</tr>
<tr>
<td>Test</td>
<td>Count%</td>
<td>6.13%</td>
<td>14.78%</td>
<td>0.94%</td>
<td>16.79%</td>
<td>31.47%</td>
<td>3.32%</td>
<td>0.31%</td>
<td>0.76%</td>
<td>0.23%</td>
<td>0.01%</td>
</tr>
</tbody>
</table>

Figure 4.9: Event Type Distribution

Table 4.7: Event Frequency Distribution

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Event Frequency</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>Count</td>
<td>239218</td>
<td>254990</td>
<td>150757</td>
<td>53087</td>
<td>12003</td>
<td>2446</td>
<td>262</td>
<td>37</td>
</tr>
<tr>
<td>Train</td>
<td>Count%</td>
<td>33.56%</td>
<td>35.77%</td>
<td>21.15%</td>
<td>7.45%</td>
<td>1.68%</td>
<td>0.34%</td>
<td>0.04%</td>
<td>0.01%</td>
</tr>
<tr>
<td>Valid</td>
<td>Count</td>
<td>73323</td>
<td>75799</td>
<td>43624</td>
<td>14818</td>
<td>3068</td>
<td>511</td>
<td>48</td>
<td>9</td>
</tr>
<tr>
<td>Valid</td>
<td>Count%</td>
<td>34.72%</td>
<td>35.89%</td>
<td>20.66%</td>
<td>7.92%</td>
<td>1.45%</td>
<td>0.24%</td>
<td>0.02%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Test</td>
<td>Count</td>
<td>74496</td>
<td>75170</td>
<td>43580</td>
<td>14206</td>
<td>3263</td>
<td>442</td>
<td>42</td>
<td>1</td>
</tr>
<tr>
<td>Test</td>
<td>Count%</td>
<td>35.27%</td>
<td>35.59%</td>
<td>20.63%</td>
<td>6.73%</td>
<td>1.54%</td>
<td>0.21%</td>
<td>0.02%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>
4.10.1.3 Data Segmentation

We split companies by time in the horizontal data segmentation. We split the training, validation, and testing datasets by time period, and we use data from 08/2004-12/2014 as training, 01/2015-12/2016 as validation, 01/2017-12/2018 as testing. After preprocessing, we have 550 companies within the dataset, and 14,300 sequences in training, 4,400 sequences in validation, and 4,400 sequences in testing. We use a rolling window by quarter to generate sequences. We illustrate the horizontal data segmentation method in Figure 4.11.

4.10.1.4 Data Preprocessing

We extracted the company name, report content, and published date for each focal company from their 8-K current reports in EDGAR. We use Python for reports downloading and content extraction. Since there was an item schema update in Aug 2004, our event matching starts from the first report that follows the updated item schema.

![Event Frequency Distribution](image)
Figure 4.11: Horizontal Data Segmentation
4.10.2 Map Reports to Event Types

For event types contain no ambiguity (non-bold event types in Table 4.5), we match the report to our event categories from the report item number directly, followed by the matching shown in Table 4.5. For event types that could be matched to multiple types (bolded event types in Table 4.5), we use Spacy phrase matcher \(^5\) to map the report content to one of the event types in Table 4.5. We list our matching phrases in Table 4.8. The matcher is case insensitive and uses exact match on all matching patterns, using OR logic.

4.10.3 Training Details

We train our own event embeddings in all deep learning models. The same set of event embedding is shared across the encoder and decoder. We initialize event embeddings by Xavier uniform distribution (Glorot and Bengio, 2010), also known as Glorot initialization. The initialized event embeddings values are sampled from \(\mathcal{U}(-a, a)\), where \(a = \text{gain} \times \sqrt{\frac{6}{\text{fan}_{\text{in}} + \text{fan}_{\text{out}}}}\), \(\text{fan}_{\text{in}}\) and \(\text{fan}_{\text{out}}\) are computed based on input and embedding dimensions and the default value for gain, a scaling factor, is 1.0. Model parameters are initialized by Glorot initialization as well. We define time window \(j\) to be a month and memory size \(M = 36\). We define forecasting time window \(q\) to be a month and the forecasting horizon \(H = 12\). We use sigmoid (Goodfellow et al., 2016) as the activation function, Adam (Kingma and Ba, 2014) with \(\beta_1 = 0.9\) and \(\beta_2 = 0.98\) as the optimizer, and Binary Cross Entropy as the loss function. For Transformer attentions, we use ELUs (Clevert et al., 2015) as the activation function.

The batch size for all models is 512. The number of head blocks (layers) for Transformer models is 4, and the number of heads in each block is 5. Event embedding size is \(d_{\text{model}} = 50\), and \(d_k = d_{\text{model}}/h = 50/5 = 10\). The learning rate is 0.001. The feed-forward layer size for Transformer models is 70.

<table>
<thead>
<tr>
<th>Event Type</th>
<th>Matching Phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business combination and restructuring</td>
<td>merge, merger, acquire, acquirer, acquisition, joint venture, separate, separation, spin-off, spin off, offtake, sell, sold, realignment, realign, offer, organizational, reorganization, purchase, business, restruct, reconstruct, limited liability, combination, combine, prospectus, sale, purchaser, seller, mergers, sellers, integrating, integration, integrate</td>
</tr>
<tr>
<td>Financial activities</td>
<td>principal amount, senior note, lend, borrow, loan, financial obligation, underwrite, credit, pay, debt, securities, repurchase, impairment, loss, note, net worth, dividend, deficiencies, net, capital, cost, redeem, redemption, price, trust, pension, stock split, bond, stock, option, tax, gain, revenue, rate of return, return, rate, charge-off, finance, financial, asset-backed, funding, earnings per share, earning per share, earnings-per-share, earning-per-share, EPS, dilute, return on equity, equity ratio, bank, investment, waive, paper, mortgage, liquidity, compensation, incentive, award, attrition, benefit, derivative, notes, underwriting, underwriters</td>
</tr>
<tr>
<td>Operation activities</td>
<td>consulting, service, product, operate, supply, system, procedure, strategy, lease, migrate, contract, equipment, assemble, inspect, procurement, decision, resource, develop, constructe, site, accident, occur, process, phase, study, determine, efficacy, safe, carrier, collaborate, manufacture, subscribe, recall, distributor, participated, participation, drilling, wells</td>
</tr>
<tr>
<td>Senior personnel change</td>
<td>director, directors, executive, vice, manager, officer, retire, leave, appoint, step down, board, committee, resign, serve, severance, not to compete, employ, hire, position, CEO</td>
</tr>
<tr>
<td>Information disclosure</td>
<td>release, presentation, statement, exhibit, reference, report, announce, announcement, meeting, furnish, date, result, outlook, conference, call, webcast, signatures, table of content, filing</td>
</tr>
<tr>
<td>Document update</td>
<td>by-laws, amendment, modification, article, code of ethics, bylaw, rights, error, rule, order, notice</td>
</tr>
<tr>
<td>Intellectual property activities</td>
<td>license, certificate, intellectual property, patent, grant, approve, permission, approval, approvable, FTC, FDA</td>
</tr>
<tr>
<td>Litigation and lawsuit</td>
<td>settlement, litigation, lawsuit, investigation, settle, suit, court, complaint, jury, hearing, subpoena, attorney, justic, violate</td>
</tr>
<tr>
<td>Delisting, trading suspension</td>
<td>delist, trading suspension, trade suspension, listing, cancel, cancellation</td>
</tr>
<tr>
<td>Bankruptcy</td>
<td>bankrupt, bankruptcy</td>
</tr>
</tbody>
</table>
At each time $t$, the top 4 L2 norm embeddings are selected for training. At the generation or testing period, event types that higher than 0.5 are selected and used as input for the next timestamp.

### 4.11 Model Results

We show our models precise (time $t$) evaluation F1% performance in Table 4.9, and cross entropy and perplexity in Table 4.10.

**Table 4.9: Model Performance on Time $t$ F1%**

<table>
<thead>
<tr>
<th>Event Type</th>
<th>MCMC</th>
<th>GRU</th>
<th>GRU Attention</th>
<th>CES Transformer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Restructuring</td>
<td>8.54%</td>
<td>3.89%</td>
<td>4.64%</td>
<td>7.35%</td>
</tr>
<tr>
<td>Financial Activities</td>
<td>19.66%</td>
<td>36.23%</td>
<td>36.88%</td>
<td>39.56%</td>
</tr>
<tr>
<td>Operation Activities</td>
<td>1.42%</td>
<td>17.91%</td>
<td>25.56%</td>
<td>25.56%</td>
</tr>
<tr>
<td>Senior Personnel Change</td>
<td>20.86%</td>
<td>20.37%</td>
<td>7.17%</td>
<td>26.07%</td>
</tr>
<tr>
<td>Information Disclosure</td>
<td>34.59%</td>
<td>74.77%</td>
<td>80.73%</td>
<td>82.35%</td>
</tr>
<tr>
<td>Document Updates</td>
<td>4.30%</td>
<td>0.21%</td>
<td>0.10%</td>
<td>1.72%</td>
</tr>
<tr>
<td>Intellectual Property</td>
<td>0.46%</td>
<td>13.66%</td>
<td>14.66%</td>
<td>22.66%</td>
</tr>
<tr>
<td>Litigation and Lawsuit</td>
<td>1.26%</td>
<td>0.00%</td>
<td>4.10%</td>
<td>5.22%</td>
</tr>
<tr>
<td>Delisting</td>
<td>0.48%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Bankruptcy</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>No Event</td>
<td>44.16%</td>
<td>53.90%</td>
<td>64.63%</td>
<td>64.27%</td>
</tr>
<tr>
<td>Micro-Averaging</td>
<td>30.40%</td>
<td>51.22%</td>
<td>57.25%</td>
<td>58.68%</td>
</tr>
</tbody>
</table>

**Table 4.10: Categorical Cross Entropy and Perplexity**

<table>
<thead>
<tr>
<th></th>
<th>MCMC</th>
<th>GRU</th>
<th>GRU Attention</th>
<th>CES Transformer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross Entropy</td>
<td>4.72</td>
<td>3.97</td>
<td>3.85</td>
<td>3.65</td>
</tr>
<tr>
<td>Perplexity</td>
<td>26.35</td>
<td>15.67</td>
<td>14.42</td>
<td>12.55</td>
</tr>
</tbody>
</table>

In Table 4.9, we can see although the proposed Transformer-based CEST model does not outperform the baseline model and two comparison models on every event type, it beats on six out of ten event types, and no event type is very close the highest performance. Additionally, the micro
Table 4.11: Model Performance on Time $t \pm 1$ F1%

<table>
<thead>
<tr>
<th>Event Type</th>
<th>MCMC</th>
<th>GRU</th>
<th>GRU attention</th>
<th>CES Transformer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Restructuring</td>
<td>20.87%</td>
<td>11.18%</td>
<td>12.02%</td>
<td>16.58%</td>
</tr>
<tr>
<td>Financial Activities</td>
<td>40.85%</td>
<td>57.08%</td>
<td>55.28%</td>
<td>56.91%</td>
</tr>
<tr>
<td><strong>Operation Activities</strong></td>
<td>4.09%</td>
<td>36.46%</td>
<td>41.56%</td>
<td><strong>43.39%</strong></td>
</tr>
<tr>
<td>Senior Personnel Change</td>
<td><strong>42.65%</strong></td>
<td>38.38%</td>
<td>30.40%</td>
<td>41.50%</td>
</tr>
<tr>
<td>Information Disclosure</td>
<td>59.68%</td>
<td>83.90%</td>
<td>87.23%</td>
<td><strong>88.45%</strong></td>
</tr>
<tr>
<td>Document Updates</td>
<td><strong>11.30%</strong></td>
<td>0.92%</td>
<td>0.41%</td>
<td>4.13%</td>
</tr>
<tr>
<td>Intellectual Property</td>
<td>1.29%</td>
<td><strong>27.16%</strong></td>
<td>16.41%</td>
<td>26.29%</td>
</tr>
<tr>
<td><strong>Litigation and Lawsuit</strong></td>
<td>3.50%</td>
<td>0.00%</td>
<td>5.26%</td>
<td><strong>8.88%</strong></td>
</tr>
<tr>
<td>Delisting</td>
<td>1.16%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Bankruptcy</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>No Event</td>
<td>69.37%</td>
<td>72.33%</td>
<td><strong>79.43%</strong></td>
<td>79.01%</td>
</tr>
<tr>
<td><strong>Micro-Averaging</strong></td>
<td>55.33%</td>
<td>67.73%</td>
<td>71.45%</td>
<td><strong>72.49%</strong></td>
</tr>
</tbody>
</table>

Table 4.12: Model Performance on Time $t \pm 2$ F1%

<table>
<thead>
<tr>
<th>Event Type</th>
<th>Baseline</th>
<th>GRU</th>
<th>GRU attention</th>
<th>CES Transformer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Restructuring</td>
<td><strong>29.49%</strong></td>
<td>16.77%</td>
<td>17.75%</td>
<td>23.39%</td>
</tr>
<tr>
<td>Financial Activities</td>
<td>52.28%</td>
<td><strong>66.99%</strong></td>
<td>64.87%</td>
<td>66.36%</td>
</tr>
<tr>
<td><strong>Operation Activities</strong></td>
<td>6.39%</td>
<td>47.40%</td>
<td>51.91%</td>
<td><strong>54.08%</strong></td>
</tr>
<tr>
<td>Senior Personnel Change</td>
<td><strong>54.14%</strong></td>
<td>48.51%</td>
<td>39.22%</td>
<td>51.17%</td>
</tr>
<tr>
<td>Information Disclosure</td>
<td>70.42%</td>
<td>88.39%</td>
<td>90.83%</td>
<td><strong>91.78%</strong></td>
</tr>
<tr>
<td>Document Updates</td>
<td><strong>16.76%</strong></td>
<td>1.22%</td>
<td>0.52%</td>
<td>5.34%</td>
</tr>
<tr>
<td>Intellectual Property</td>
<td>2.04%</td>
<td>33.21%</td>
<td>25.57%</td>
<td><strong>34.85%</strong></td>
</tr>
<tr>
<td>Litigation and Lawsuit</td>
<td>5.41%</td>
<td>0.00%</td>
<td>8.95%</td>
<td><strong>10.94%</strong></td>
</tr>
<tr>
<td>Delisting</td>
<td><strong>1.69%</strong></td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Bankruptcy</td>
<td>0.14%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>No Event</td>
<td>78.08%</td>
<td>79.56%</td>
<td><strong>85.08%</strong></td>
<td>84.68%</td>
</tr>
<tr>
<td><strong>Micro-Averaging</strong></td>
<td>66.26%</td>
<td>75.66%</td>
<td>78.59%</td>
<td><strong>79.45%</strong></td>
</tr>
</tbody>
</table>
Table 4.13: Model Performance on Time $t \pm 3$ F1%

<table>
<thead>
<tr>
<th>Event Type</th>
<th>Baseline</th>
<th>GRU</th>
<th>GRU attention</th>
<th>CES Transformer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Restructuring</td>
<td>35.76%</td>
<td>21.03%</td>
<td>22.54%</td>
<td>29.79%</td>
</tr>
<tr>
<td>Financial Activities</td>
<td>59.37%</td>
<td>74.29%</td>
<td>72.91%</td>
<td>74.29%</td>
</tr>
<tr>
<td>Operation Activities</td>
<td>8.22%</td>
<td>55.28%</td>
<td>61.76%</td>
<td>63.13%</td>
</tr>
<tr>
<td>Senior Personnel Change</td>
<td>61.20%</td>
<td>56.46%</td>
<td>47.96%</td>
<td>59.28%</td>
</tr>
<tr>
<td>Information Disclosure</td>
<td>76.03%</td>
<td>92.26%</td>
<td>93.90%</td>
<td>94.49%</td>
</tr>
<tr>
<td>Document Updates</td>
<td>21.22%</td>
<td>1.22%</td>
<td>0.62%</td>
<td>6.79%</td>
</tr>
<tr>
<td>Intellectual Property</td>
<td>2.65%</td>
<td>45.87%</td>
<td>31.22%</td>
<td>45.67%</td>
</tr>
<tr>
<td>Litigation and Lawsuit</td>
<td>7.01%</td>
<td>0.00%</td>
<td>12.36%</td>
<td>15.41%</td>
</tr>
<tr>
<td>Delisting</td>
<td>2.44%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Bankruptcy</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>No Event</td>
<td>82.60%</td>
<td>85.09%</td>
<td>89.20%</td>
<td>88.88%</td>
</tr>
<tr>
<td>Micro-Averaging</td>
<td>72.37%</td>
<td>82.36%</td>
<td>84.58%</td>
<td>85.15%</td>
</tr>
</tbody>
</table>

averaging F1% for all event types outperforms the baseline MCMC model and the two sequence-to-sequence comparison models.

We should point out that several event types have very small amounts of data compared to regular event types based on the data statistics. It means the model is trained based on very unbalanced data. As a result, we can observe several event types, such as document updates, delisting, and bankruptcy have low predictions, because of their small data shares.

Table 4.10 demonstrates our proposed model generates consistent lower cross entropy and perplexity than the baseline model and the two comparison models. Experimental results verify the effectiveness of our proposed model on predicting business material event sequences.

We can see the model results are consistent with our intuition that SEC 8-K reports embody useful information to predict firm future events. Furthermore, they confirm our proposed model architecture, i.e., the multi-head attention Transformer model can be used to tackle the event sequence forecasting task.

We show fuzzy evaluation $z=1$ F1% performance in Table 4.11, $z=2$ F1% performance in Table 4.12, and $z=3$ F1% performance in Table 4.13.
From the model results, we can see if we look for one quarter variation forecasting horizon \((t\pm3)\), our model’s outperforming event types can produce pretty good prediction results. Furthermore, the overall model’s micro averaging for all event types are consistently better than the baseline and comparison models in all evaluation variations.

The multiple evaluation variations also endorse our model architecture and confirm our intuition that we can use historical firm material event sequences to predict future event series, using the CES Transformer model.

### 4.12 Example: Fred’s Inc. - The Fading Discount Retailer

#### 4.12.1 Company Description

Fred’s, a Tennessee based discount merchandise retailer, carried both packaged products and pharmacy businesses and had most of their stores opened in southeastern states of the U.S. The company was founded in Coldwater, Mississippi, in 1947. On the discount storefront, Fred’s business is similar to Family Dollar and Dollar General. However, their product price range is higher. For instance, they also offer grill and patio furniture during the summertime while Family Dollar and Dollar General do not carry. On the pharmacy business front, Fred’s is similar to Rite Aid, Walgreens, and CVS. They offer services for patients’ prescriptions.

We list the focal company Fred’s and its related companies financial standing from Fortune.com in Table 4.14. The market value is as of the available year March 31.

#### 4.12.2 Example Time Period

We use Fred’s Inc., a company that ended with bankruptcy, as an example to illustrate our model. In this example, the memory period is 01/2015-12/2017 \((M=36)\), and the prediction period is 01/2018-12/2018 \((H=12)\).

---

Table 4.14: Fred’s and Related Companies Financial Standing

<table>
<thead>
<tr>
<th>Company</th>
<th>Year</th>
<th>Fortune 500 Rank</th>
<th>Revenues ($M)</th>
<th>Profits ($M)</th>
<th>Market Value($M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fred’s</td>
<td>2016</td>
<td>928</td>
<td>2,125</td>
<td>-7</td>
<td>557</td>
</tr>
<tr>
<td>CVS</td>
<td>2016</td>
<td>7</td>
<td>153,290</td>
<td>5,237</td>
<td>113,947</td>
</tr>
<tr>
<td>Walgreens</td>
<td>2016</td>
<td>19</td>
<td>103,444</td>
<td>4,220</td>
<td>90,874</td>
</tr>
<tr>
<td>Rite Aid</td>
<td>2016</td>
<td>107</td>
<td>26,528</td>
<td>2,109</td>
<td>8,529</td>
</tr>
<tr>
<td>Dollar General</td>
<td>2016</td>
<td>139</td>
<td>20,369</td>
<td>1,165</td>
<td>24,522</td>
</tr>
<tr>
<td>Family Dollar</td>
<td>2015</td>
<td>281</td>
<td>10,489</td>
<td>285</td>
<td>9,069</td>
</tr>
</tbody>
</table>

While we should learn successful stories from leading companies, the sad stories from fading companies should not be ignored. They should be studied even harder, especially when a firm faces a similar situation. If companies follow what successful companies do, they may not lead themselves to the same success. However, if companies follow what the fading companies did, they have a higher chance of facing the same result as the fading company. Therefore, we use a fading company as an example.

Moreover, our model can facilitate the process of identifying influential events. Hence, our model can ease this learning process. If companies learn what the sequence of decisions the fading company made, they can know what they are likely to experience if they took similar choices and what the result will look like at the end. Therefore, studying a fading company’s end story can prevent companies from trying those costly operational or managerial paths themselves.

Furthermore, fading companies are more intended to try different operational approaches; therefore, they can provide a series of amplified event relationships. Hence, these companies are also in line with our demonstration purpose.

### 4.12.3 Corporate Event Sequence Stories

We list the company’s 8-K events in Figure 4.12, and we will illustrate event sequences in this company during the sample period in detail in this section.
Figure 4.12: Fred’s 8-K Event Sequence
The company has tried to enlarge its pharmacy business before our sample period. Therefore, we can see the company named a new President and Chief Operating Officer (COO) who had working experience from Family Dollar and CVS. The combined expertise from both small-box general merchandising and pharmacy store operation made the new President and COO a good fit for Fred’s combined business. Right after the new President was named, the company acquired a drug store to expand its pharmacy business at the beginning of our memory period (03/2015). And in the next two years, pharmacy business expansion is the most important strategy for the company. It has prepared several talented high-level management team members for the development, such as new Board Chairman (03/2016), new CFO (04/2016), new EVP and CMO (08/2016), and even a new CEO (09/2016). Finally, at the end of 2016 (12/2016), Fred’s entered an agreement with Rite Aid to purchase 865 of its stores. In the proposed deal, Walgreens and Fred’s will both be the acquirers, while Walgreens was going to acquire even more Rite Aid stores than Fred’s do.

The high-level personnel changes (event type PC) have worked on preludes for “success” of the merger agreement. It is a common strategy for companies to appoint new management members, especially chief officers, when they are preparing on a big project, such as acquisition, and require their business connections and experience. From this chain of actions, we can summarize the sequence of the key events as:

\[
\text{New President} \xRightarrow{1} \text{New CEO} \xRightarrow{2} \text{Purchase Agreement}.
\]

After the purchase agreement, it is common for companies to require more money to support the merge. Therefore, one should expect loan activities afterward. Indeed, we can see the company increased its debt (01/2017) and loan (02/2017) in the next two months after the purchase agreement (12/2016). Often, it is not enough to have one loan, especially for large acquisition activities. Therefore, we see multiple debt and loan activities from Fred’s after their purchase agreement announcement. Follow this logic, we can summarize this event sequence as:

\[
\text{Purchase Agreement} \xRightarrow{4} \text{Debt(s)} \xRightarrow{5} \text{Loan(s)}.
\]
However, the Federal Trade Commission (FTC) rejection changed the subsequent stories entirely. The FTC rejected Walgreens’ bid for Rite-Aid in 01/2017 because of antitrust concerns. Fred’s released a merger extension 8-K report to reflect this FTC decision (01/2017). In the report, Fred’s stated ”the Federal Trade Commission (“FTC”) requests that additional stores be sold, and Walgreens agrees to sell such stores, Fred’s Pharmacy has agreed to buy those stores”. While still preparing the merger by obtaining more loans (02/2017, 06/2017), the company had worked on its board structure and tried to salvage the deal. Fred’s increased six board members in total within the next three months and its number of board members increased from seven to ten (03/2017), ten to eleven (04/2017), and eleven to thirteen (04/2017). However, they did not work things out, and the purchase agreement was terminated. Eventually, Walgreens ended with acquiring less Rite-Aid stores and cut Fred’s out of the deal (06/2017). We can summarize these event sequences as:

\[
\text{The FTC Rejection} \stackrel{7}{\Rightarrow} \text{Personnel Change}
\]

, and \[
\text{The FTC Rejection} \stackrel{8}{\Rightarrow} \text{Merger Termination}
\]

Since previous loans were intended to prepare the merger, after the merger termination, the loans were terminated afterwards. We can illustrate this event sequence as:

, and \[
\text{Merger Termination} \stackrel{9}{\Rightarrow} \text{Loan Termination}
\]

In addition, because of the merger deal’s termination, we can expect some personnel change. Furthermore, Fred’s loss had widened than the previous year. The personnel changes turned out happened more than expected, including the departure of Executive Vice President, Chief Financial Officer and Secretary (07/2017), the retirement of Board Member (08/2017), the appointment of new Board of Director Chairman (08/2017), and the resigns of the previous Board Director Chairman (08/2017), Chief Operating Officer (09/2017), Senior Vice President (02/2018), Board of Directors (02/2018, 04/2018, 05/2018), Executive Vice President, Chief Operating Officer (05/2018), and Executive Vice President, Chief Merchandising and Marketing Officer (05/2018). More importantly, we notice the Board of Directors Chairman was changed to a person from the investment
company. One senior personnel change can lead to many subsequent personnel changes. We can summarize these event sequences as:

\[
\text{Merger Termination} \xrightarrow{10} \text{Personnel Change} \xrightarrow{11} \text{Personnel Change(s)}.\]

Given the merger termination and the overall company performance, to avoid to be acquired, Fred’s amended its shareholder rights (09/2017) and stock purchase program (12/2017). We can summarize this event sequence as:

\[
\text{Merger Termination} \xrightarrow{12} \text{Stock Purchase Amendment}.\]

Because of the terminated merger deal and the underperformed store sales, Fred’s made a credit and security amendment with its investment company. They decided to open up the list for selling company properties, including selling its real estate (08/2018). Sometimes, one financial obligation is not enough to satisfy the immediate company’s needs. Therefore, other financial obligations might follow after the initial one as well. Meanwhile, if an asset sale is started, we can see when the asset sale is completed from the company’s reporting. In fact, we observe Fred’s had to sell its prescription files and related data and records, retail pharmaceutical inventory, and certain other assets to purchasers in multiple transactions. We can summarize the entire event sequence as:

\[
\text{Merger Termination} \xrightarrow{13} \text{Asset Sale} \xrightarrow{14} \text{Asset Sale(s)} \xrightarrow{15} \text{Asset Sale(s) Completion}.\]

Although the company appointed a new CEO early on, the series of decisions made by senior management team members lead the company to an asset sale at a later time. We can summarize the event sequence as:

\[
\text{Personnel Change} \xrightarrow{16} \text{Asset Sale}.\]

Nine months after our example time period, although Fred’s tried different ways to turn the company around, it filed for bankruptcy protection and planned to shut down all of its stores. The termination of the merger and acquisition activity indicated the end of the story.

\[
\text{Merger Termination} \xrightarrow{17} \text{Filing Bankruptcy}.\]
In this example, we show examples of corporate event series, and within every event series, we explained its rationale. In the next section, we will demonstrate our model’s attention weights to alleviate human works in this process.

4.12.4 Model Insights

What Our Model Attention Maps Can Identify For Users? This section will illustrate what our model can identify for users, from its attention weights. There are three main features the model attention can identify for us:

- identify influential events in memory period.
- identify important event relationships between memory and forecasting periods.
- identify influential events in forecasting period.

We will illustrate them in turn.

4.12.4.1 Identify Influential Events in Memory Period

We can identify influential events that happened during the memory time period, using the encoder attention map. We illustrate an encoder attention map in Figure 4.13.

In Figure 4.13, we can pinpoint an important time period - 11/2016 to 10/2017 - which has higher attention weights (darker color on the attention map) during the company’s historical period. True enough, the most influential event for the company - the merger activity with Rite Aid and Walgreens - happened within these highlighted high attention time stamps. The encoder attention map also highlights months after the merger activity to show the prolonged influence from the merger event.

Furthermore, although the attention map can not directly demonstrate the event relations, we can combine our knowledge from Section 4.12.3 and link the stories to our attention map.

For instance, our attention map assigns higher weights to both events in the Purchase Agreement Debt(s) event sequence. In addition, we can identify other important events and event sequences
happened after the merger activity, including the FTC Rejection $\xRightarrow{7}$ Personnel Change and Asset Sale $\xRightarrow{14}$ Asset Sale(s) from the encoder’s attention map.

Meanwhile, although not all events in the event sequences will be highlighted by the attention map, we can still use them to validate the importance of those identified influential events. For example, the merger activity in multiple event sequences can be streamed out from one influential event, although some events are not highlighted by the attention map. For example, Purchase Agreement $\xRightarrow{6}$ Debt(s) $\xRightarrow{6}$ Loan(s) and the FTC Rejection $\xRightarrow{8}$ Merger Termination are both streamed out from the merger activity. In addition, both Merger Termination $\xRightarrow{9}$ Loan Termination and Merger Termination $\xRightarrow{10}$ Personnel Change start from the merger termination event.

Figure 4.13: Encoder Attention: Identify Influential Events in Memory Period
4.12.4.2 Identify Important Event Relationships Between Memory and Forecasting Periods

Figure 4.14 is an attention map that describes the relationship between the encoder and the decoder. From the highlighted timestamps, we can identify the personnel change event in August 2016 - a new CEO appointment, which played an important and prolonged role for the company. Specifically, 23 months after the CEO was appointed, the company has undertaken a series of events and reached a point to sell its assets. The relationship between the senior personnel change and the financial obligation can be identified from our encoder-decoder attention map. Interestingly, our model can identify those assets sale events were lead by a series of CEO related decisions, even after the CEO had resigned.

![Encoder-Decoder Attention](image)

Figure 4.14: Encoder-Decoder Attention: Identify Important Event Relationships Between Memory and Forecasting Periods

4.12.4.3 Identify Influential Events in Forecasting Period

Like the encoder attention map, the decoder attention map can emphasize influential events within the forecasting period. From Figure 4.15, we can identify the high self-attention weights event: the asset sales completion event. It is result event in the event sequences Merger Termination Asset Sale and Personnel Change Asset Sale.
Furthermore, we can also identify other influential events, such as personnel change event in Personnel Change $\Rightarrow$ Personnel Change(s) and asset sales event in Asset Sale(s) $\Rightarrow$ Asset Sale(s) Completion.

Figure 4.15: Decoder Attention: Identify Influential Events in Forecasting Period

4.12.5 Model Goals

What Our Model Can Do? There are five functions that our model can offer:

- (1) Formulate a useful business scenario - from a series of corporate events to forecast its consecutive event sequence, while this setup is not considered solvable by most time series models.
- (2) Forecast Corporate Event Sequence (CES).
- (3) Identify influential events in the memory period.
- (4) Identify important event relationships between memory and forecasting periods.
(5) Identify influential events in the forecasting period.

Therefore, when we need to forecast event series from previous event series, we can use our CEST model. Furthermore, if we want to understand the inner working of the model further, our model’s attention maps can be used to identify influential company events during the memory period and forecasting period, respectively. In addition, if we want to understand event relationships, we can use the model’s encoder-decoder attention maps to observe the desired event relationships.

4.13 Model Byproducts

We have two model byproducts: firm embedding and event embedding from our model. Embedding is a dense vector that carries specific semantic meaning. For instance, we can use firm embedding to represent an individual company’s decision-making behavior, based on their historical event sequences. Similarly, we can use event embedding to describe the meaning of each event type.

4.13.1 Firm Embedding

We use vectors before the last timestamp’s sigmoid prediction as our potential firm embedding. For each firm, we average all potential firm embeddings in the test set to obtain the firm embedding for each specific company.

Then, we use T-SNE (Maaten and Hinton, 2008) to plot the relative positions for each company. For demonstrate purpose, we illustrate selected firm embedding in Figure 4.16. We color code companies by the Fama-French 5-industry portfolios (Fama and French, 2019) in Table 4.15.

As a byproduct of the model, firm embeddings show each company’s relative positions compared to other companies. First, we observe close distance for companies in the same industry, such as financial services companies (in purple) and HiTech companies (in green). We can see companies with the same color tend to be closer. It means they share certain similar decision-making behavior, based on their industry specifics. Second, we can observe examples of companies not in the same industry but having a close distance. For instance, we can see an example of the relative position
for amazon.com and Netflix. This example tells us that although companies could be categorized into different industries, they can share a similar decision-making process or behavior if some of their businesses are competing with each other. Use amazon.com and Netflix as an example. We are aware that amazon.com has broad business services, including Amazon prime video service for its members. It can be viewed as a competitor to Netflix because they might target the same or similar group of customers. Therefore, their decision-making process and strategies might be similar.

Figure 4.16: Firm Embedding
Table 4.15: Fama-French 5-Industry Portfolios

<table>
<thead>
<tr>
<th>Industry</th>
<th>Portfolio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumer</td>
<td>Consumer Durables, NonDurables, Wholesale, Retail, and Some Services (Laundries, Repair Shops)</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>Manufacturing, Energy, and Utilities</td>
</tr>
<tr>
<td>HiTech</td>
<td>Business Equipment, Telephone and Television Transmission</td>
</tr>
<tr>
<td>Healthcare</td>
<td>Healthcare, Medical Equipment, and Drugs</td>
</tr>
<tr>
<td>Others</td>
<td>Mines, Construction, Building Maintenance, Transportation, Hotels, Business Services, Entertainment, and Finance</td>
</tr>
</tbody>
</table>

4.13.2 Event Embedding

Another byproduct of our model is event embedding. We illustrate it in Figure 4.17. We train the event embeddings in our model. And we also use T-SNE (Maaten and Hinton, 2008) to plot the relative semantic position for each event.

First, we can see intellectual property activities (IP) is very far from other event types. Second, litigation and lawsuit (LL), delisting and trading suspension (DL), and bankruptcy (BK), these
irregular events, are relatively closer to each other. Third, business operation activities, such as information disclosure (ID), document update (DU), and operation activities (OA) as relatively closer. Fourth, personnel change (PC), financial activities (FN), and business combination (BC) are considered normal but infrequent (such as every quarter) business events; therefore, they stay closer to each other.

4.14 Business Implication

In this section, we demonstrate how our model can be used to help understanding company material events. To address our model’s practicality, we use aphorism and quotes from Mark Twain, Charlie Munger (Clark, 2017), Warren Buffett (Buffett and Clark, 2006) and Jeff Bezos to demonstrate the underlying thinking and use cases of our model.

“History doesn’t repeat itself but it often rhymes.” - Mark Twain

“I very frequently get the question: ‘What’s going to change in the next 10 years?’ ... I almost never get the question: ‘What’s not going to change in the next 10 years?’ And I submit to you that that second question is actually the more important of the two – because you can build a business strategy around the things that are stable in time.” - Jeff Bezos

In general, our deep learning model provides a computational approach for investors to exam a firm’s history and find out things are important for the company while providing timely variant flexibility. For corporate strategy, it is important to have a computational tool to forecast long-term events for a company.

“Favorable surprises are easy to handle. It’s the unfavorable surprises that cause the trouble.” - Charlie Munger

Although our model does not distinguish if an event is favorable or not for an investor, our model’s fundamental goal is to help limit the focal company’s future surprising events and get companies and investors prepared ahead of time.
“The difference between a good business and a bad business is that good businesses throw up one easy decision after another. The bad businesses throw up painful decisions time after time.” - Charlie Munger

“We don’t go into companies with the thought of effecting a lot of changes. That doesn’t work any better in investments than it does in marriages.” - Warren Buffett

“A good managerial record is far more a function of what business boat you get into than it is of how effectively you row. Should you find yourself in a chronically leaking boat, energy devoted to changing vessels is likely to be more productive than energy devoted to patching leaks.” - Warren Buffett

A company must go through diligent processes and many careful thoughts on any given project. From an outsider’s eyes, it is just an announcement on their 8-K report. On the contrary, if a firm’s management team is unstable, such as changing their higher management team members frequently, the outsider can scratch the possibility of shaky company status.

“What we learn from history is that people don’t learn from history.” - Warren Buffett

Our model provides a tool to grasp a firm’s material event history, highlight the company’s behaviors, see the firm’s unique character, and compare companies across industries or even the entire market.

” 'One solution fits all’ is not the way to go .... The right culture for the Mayo Clinic is different from the right culture at a Hollywood movie studio. You can’t run all these places with a cookie-cutter solution.” - Charlie Munger

We truly believe it. Therefore, our model offers companies and investors the opportunity to closely examine their focal companies from a one-on-one fashion while maintaining an overarching framework. On the one hand, our model learns statistical patterns from all Fortune companies, no matter what size of the company and industry. On the other hand, when we open up the model and look
at one company at once, we can leverage the entire model and examine the particular company individually. Our model provides tailored and customized insights for each company instead of a cookie-cutter solution.

4.15 Discussion: From AI Models to Managing AI to AI-based Management

In this paper, we demonstrate a flow of solving a real business problem using AI models and dive into the deep learning model to make interpretations. Because deep learning models are often criticized as black boxes, we present a way to open up the box and probe the inner working of the proposed neural network models. Given the computational and logical complexities of neural network deep learning models, it is critical to managing AI models well to serve business needs. The serving approach we took is to use interpretation as a tool to manage AI.

However, we also want to take managing AI a step further and project where AI models and results probing can lead to us. In business reality, we need a powerful AI model to work on a real-world task and provide details on how it works out. From the user’s perspective, we need to manage AI models by understanding the model results and model interpretation. Furthermore, when users are used to managing AI models, more AI models will be implemented and used in reality. At that time, AI models are also objects that need to be included in the management system, good or bad. Therefore, AI-based management is approaching, and managers should be aware of it.

How to govern the AI-based management systems? On the one hand, model results should not be blindly taken. By helping from tools, such as the interpretations that we demonstrated, managers should understand each model’s rationale and make sense of the model results by incorporating knowledge from reality. On the other hand, managers should also be open-minded on new observations, but with a grind of salt.

Other than the AI model results, computational complexity should also be considered by management teams in the new AI-based management era. Although this is not the focus of our demonstration in this paper, it certainly has a say on the firm’s AI strategy and success.
The combination of AI models and hardware supports will make a good starting point for a firm’s AI implementation. On top of that, managers’ skills in understanding AI models and interpreting model results also become required. Furthermore, after mastering the skills in interpreting AI models, managers can start to know the extent to which the state-of-art AI models can achieve and what pitfalls AI models can potentially make. In this case, AI models can be used to serve business needs better.

4.16 Conclusion and Future Work

In this paper, we demonstrated a flow of AI models to manage AI to AI-based management. We illustrate a real business task that by our knowledge is not well solved in real business and introduce a state-of-art neural network model to tackle the problem. In particular, we use Transformer models to predict corporate event series from the company’s SEC 8-K Current Reports. Our proposed Transformer model outperforms its Markovian and traditional seq2seq counterparts.

We dive into the Transformer model and demonstrate the inner working of the model. Furthermore, we provide interpretations for each component of the model. We also demonstrated the effectiveness of our model by comparing the model predictions and the actual truth. We also provide our thoughts on how to manage AI by interpreting AI models and how to be prepared for the AI-based management era at the end of the work.

Our work has considerable implications for firm stakeholders, the financial industry, and the research community:

- Internally, the forecasting models and their artifacts provide decision support for firm executives, when performing various tasks such as strategic planning, financial forecasting, and peer benchmarking.

- Externally, institutional investors will benefit from the models’ predictive outcome and explanatory insight. Such enhanced transparency and accountability are even more critical when dealing with private firms with no duty to disclose their accounting books.
• Our models and artifacts provide previously non-existent, and interpretable decision aid, without compromising on the quality of prediction at all.

• Third-party regulators and agencies represent another family of beneficiaries. For example, S&P Global Ratings’ analysis can potentially leverage enriched information channels and the interplay of multi-faceted financial pictures, both of which are essential components in our framework.

• On the methodological front, our unique problem formulation, i.e., event sequence prediction, casts new light into the area of financial forecasting powered by AI technologies. The deep-learning model architecture also enables the natural integration of numerical and textual data, previously considered extremely challenging due to a lack of methodological aids. The formulation and models are largely problem independent. Hence, they have general applicability in a broad spectrum of economic forecasting problems.

An early attempt at bringing large-scale textual data and state-of-the-art deep learning models into financial forecasting, our work is not without its limitations. We are poised to extend it in multiple directions:

• The success of our models heavily relies on the quality of input. Room for improvement exists in both event extraction and event encoding techniques.

• Other types of non-numerical data can participate in the modeling and forecasting process, e.g., business news, sell-side reports, etc.

• The generalizability of our model architecture is yet to be fully explored in broader economic forecasting settings.

4.17 References


CHAPTER 5. GENERAL CONCLUSION

My dissertation develops and extends natural language processing (NLP) tools for business analytics, organization performance prediction, and corporate events sequences prediction.

Findings in my dissertation show Management Information Systems (MIS) scholars from around the world have come to publish their work in the same narrow set of journals, regardless of where they work or what their research perspectives might be. Within this set of journals, there has been a growing movement toward homogeneity and convergence in research topics. The analyses provide support for MIS scholars to quantify the discipline research trend and better assess field dynamics. My first essay also identified two topic morphing behaviors, cross-topic morphing and inter-topic morphing, and illustrated their examples. My second essay describes when a firm’s financial performance is not available, the proposed single-task and multi-task text-based recurrent neural network models can predict firm financial ratios individually and simultaneously, using news articles. The proposed models outperform traditional time series models. Additionally, when a firm’s historical financial ratios are available, the proposed integration model can consolidate historical ratios and news articles to predict future firm financial ratios. The news empowered integration model outperforms the ratio’s deep learning model baseline. My third essay first formulated a unique and real business question, predicting corporate event sequences based on the firm historical event sequences. It then proposed a model to complete the company’s event sequence by predicting the subsequently corporate event series, based on its historical report history. The proposed models can benefit scholars, firm executives, main street investors, institutional investors, and third-party regulators and agencies. In addition, the proposed problem formulation and model architectures are largely problem independent. Therefore, they have general applicability in a broad spectrum of scholarly and organizational scenarios.
Based on the three essays in this dissertation, a few interesting research directions are emerging. First of all, real-world business problems are highly dynamic, and business environment and firm-level knowledge are important to be considered. Therefore, it would be appealing to integrate these aspects into proposed models. Secondly, leveraging larger pretrained generation models and transfer knowledge to business domains is another direction. Last but not least, other NLP tasks and frameworks, such as question answering and multi-modality architectures, can be integrated business analytics to solve real-world business problems. Meanwhile, various business fields, such as health IT, marketing, and finance, have the data supports and can be benefited from advanced NLP models. The integration of NLP techniques and real-world business scenarios is an open and exciting direction for my future research.