Natural language processing (NLP) in management research: A literature review

Yue Kang, Zhao Cai, Chee-Wee Tan, Qian Huang & Hefu Liu

Taylor & Francis
Taylor & Francis Group

Check for updates

REVIEW

Natural language processing (NLP) in management research: A literature review

Yue Kang[a], Zhao Cai [b*], Chee-Wee Tan [c], Qian Huang[a] and Hefu Liu [a]

[a]*School of Management, University of Science and Technology of China, Hefei, People's Republic of China;* [b]*Nottingham University Business School China, University of Nottingham Ningbo China, Ningbo, People's Republic of China;* [c]*Department of Digitalization, Copenhagen Business School, Copenhagen, Denmark*

Natural language processing (NLP) is gaining momentum in management research for its ability to automatically analyze and comprehend human language. Yet, despite its extensive application in management research, there is neither a comprehensive review of extant literature on such applications, nor is there a detailed walkthrough on how it can be employed as an analytical technique. To this end, we review articles in the UT Dallas List of 24 Leading Business Journals that employ NLP as their focal analytical technique to elucidate how textual data can be harnessed for advancing management theories across multiple disciplines. We describe the available toolkits and procedural steps for employing NLP as an analytical technique as well as its advantages and disadvantages. In so doing, we highlight the managerial and technological challenges associated with the application of NLP in management research in order to guide future inquires.

Keywords: natural language processing; literature review; sentiment analysis; machine learning

1. Introduction

Digitalization has endowed researchers with unprecedented opportunities for harvesting a wealth of textual data for investigating contemporary phenomena. Natural Language Processing (NLP), a computer-assisted analytical technique aimed at automatically analyzing and comprehending human language (Manning, Manning, & Schütze, 1999), allows scholars to easily extract beneficial insights contained in textual datasets while avoiding burdensome computational work (Collobert et al., 2011; Green, 2012). In recent years, NLP has experienced significant breakthroughs with the emergence of Artificial Intelligence (AI) (Barr & Feigenbaum, 2014; Hirschberg & Manning, 2015; Lu, 2019). Advances in machine translation (Och, 2003), pattern matching (Califf & Mooney, 2003), sentiment analysis (Liu, 2012), and speech recognition (Weber, 2002) has not only brought about radical changes in people's daily lives, but they have also revolutionized business practices. For instance, the Norwegian News Agency[1] (NTB) documented that robotic journalists can produce a news article within 30 s upon the conclusion of an event. Likewise, NLP

---

*Corresponding author. Email: zhao.cai@nottingham.edu.cn

has given rise to smart devices (Siri, Cortana), simultaneous translation (Baidu's STACL[2]), speech-to-text recognition (iFLYTEK's iflyrec[3]), sentiment-based market predictions (e.g. financial news and Twitter)[4], as well as intelligent shopping guides and robotic customer service attendants (Alibaba[5]), which together, have reshaped the business landscape. To gain an in-depth appreciation into consumer behavior (Chi-Hsien & Nagasawa, 2019; Hou, Zhao, Zhao, & Zhang, 2016), corporate governance (Chanda & Goyal, 2019; Law & Chung, 2020), and market dynamics (Verma, Malhotra, & Singh, 2020), scholars have turned to NLP as a means of deriving insights from textual datasets in annual reports and press releases published by firms, online reviews and social media postings generated by consumers, as well as legislative and policy documents promulgated by governments (Bansal, Sharma, & Singh, 2019). Harnessing NLP, researchers can capture concepts objectively and precisely, thereby eliminating the tedious work associated with manual textual coding.

To date, the vast development of NLP has substantially contributed to advances in management theories. By the end of 2019, there were 72 articles published in the UT Dallas List of 24 Leading Business Journals (UTD List) that draw on NLP as their focal analytical technique. Yet, despite the burgeoning number of studies centered on NLP, there is neither a comprehensive review of extant literature on how it has been applied to management research, nor is there a detailed walkthrough on how it can be employed as an analytical technique. We hence argue that a systematic review of extant literature is long overdue to elucidate how NLP has been applied to management research across disciplines and provide scholars with plausible, innovative approaches for applying this analytical technique toward the enrichment of management theories. In the same vein, we contend that it is essential to develop a step-by-step tutorial illuminating how NLP can be employed for data analysis to render it more accessible to scholars dealing with textual data. To this end, this study addresses the abovementioned knowledge gaps by conducting a review of extant literature on NLP to chart out a roadmap for how the analytical technique can be applied to inform management research. The scope of our literature review encompasses the UTD List of 24 Leading Business Journals because articles published in these journals generally possess a solid theoretical foundation, rigorous empirical design, meticulous analytical procedures, and rich datasets. By constraining the scope of our literature review to this list, we adhere to Ricks, Toyne, and Martinez's (1990) advocated principle that reviewed literature must be the "recent literature on the basic issues so that the reader can be brought up to date and guided towards what can be read in order to obtain the depth of understanding desired" (p. 220).

Given this, we aim to contribute to the application of NLP in management research on three fronts. First, we attempt to illustrate how textual data can be harnessed for management research across multiple disciplines, concentrating specifically on where textual datasets can be acquired and how they can be leveraged to solve ongoing management problems. Second, we endeavor to offer a methodical guide for scholars intending to employ NLP in their analytical work, outlining the procedural steps and toolkits available as well as their advantages and disadvantages. Finally, we sought to expound the managerial and technological challenges confronting scholars in the application of NLP, which in turn point to opportunities for further innovation on how textual data can be ingeniously applied to advance management theories.

The remaining manuscript is structured as follows. We first offer an overview of NLP, including its history. We then describe the methodological steps taken in conducting our review of extant literature on the application of NLP in articles published in the UTD List of 24 Leading Business Journals. Following this, we elaborate on how NLP has been applied in management research across a variety of disciplines, highlighting discipline-specific considerations with regards to its application and deriving opportunities for future research through interdisciplinary comparisons. Thereafter, we present a step-by-step tutorial of basic processing procedures for NLP and related algorithms. We conclude by identifying potential challenges and future directions for the development of NLP from both managerial and technological perspectives.

2. Natural language processing: an overview

NLP constitutes a core interest in the field of artificial intelligence and computer science. NLP studies comprise theories and methods that enable effective communication between humans and computers in natural language. As a scientific field of study, NLP assimilates computer science, linguistics, and mathematics with a primary goal of translating human (or natural) language into commands that can be executed by computers. NLP consists of two research directions: *Natural Language Understanding* (NLU) and *Natural Language Generation* (NLG). The principal mission of NLU is to comprehend the natural language (human language) (Schank, 1972) by deciphering documents and extracting valuable information for downstream tasks. In contrast, NLG is the production of text in natural languages that are understandable by humans based on the provision of structured data, text, graphics, audio, and video, (McDonald, 2010). NLG can be further divided into three categories: text-to-text (Genest & Lapalme, 2011), such as translation and abstract; text-to-other, such as text-generated images (Xu et al., 2018); and other to text (other-to-Text), such as video-generated text (Rohrbach et al., 2013).

Thus far, the development of NLP has gone through four periods: the germination period before 1956, the rapid development period from 1957 to 1970, the low-speed development period from 1971 to 1993, and the recovery period from 1994 to the present.

The germination period encompasses the beginnings of NLP research. Alan Turing first proposed the concept of the 'Turing Machine' in 1936. The 'Turing Machine', as the theoretical origin for the modern computer, prompted the birth of the electronic computer in 1946, providing a material basis for machine translation and, subsequently, NLP. Given the needs of machine translation, foundational research in NLP was carried out during this period. In 1948, Shannon applied the probabilistic model of discrete Markov processes to the automation of the description language. He then applied the concept of entropy in thermodynamics to the probabilistic algorithm of language processing to measure the amount of information contained in human language. In the early 1950s, Kleene investigated finite automata and regular expressions and, in 1956, Chomsky proposed a context-free grammar and applied it to NLP. Such work directly led to the generation of two NLP techniques based on rules and probability. These two methods have sparked decades of disputes regarding rule-based methods and probabilistic methods in NLP. The birth of artificial intelligence in 1956 further opened a new chapter in NLP as AI gradually combined with other NLP technologies over the next few decades to enrich the technical means of NLU and NLG, thereby broadening the social application of NLP.

The first major development of NLP occurred from 1957 to 1970 when NLP was rapidly integrated into the field of artificial intelligence. During this period, research on both rule-based methods and probabilistic methods made great progress. From the mid-1950s to the mid-1960s, symbolist scholars like Chomsky began the study of formal language theory and generative syntax. Concurrently, scholars focusing on probabilistic methods adopted the Bayesian method based on statistical research methods and also made great progress during this period. However, in the field of artificial intelligence, most scholars only paid attention to the study of reasoning and logic. Very few scholars from statistics and electronics majors studied probability-based statistical methods and neural networks. Important achievements in research during this period included the successful Transformations and Discourse Analysis Project (TDAP) developed by the University of Pennsylvania in 1959 and the establishment of the Brown American English Corpus. In 1967, American psychologist Neisser proposed the concept of cognitive psychology, directly linking NLP with human cognition.

The second major development of NLP started in 1971 and ended in 1993. By then, it was apparent that applications based on NLP could not be solved in a short amount of time and new problems regarding the use of statistical approach and the establishment of corpus were constantly emerging. Many people lost confidence in the study of NLP as a result. The 1970s thus mark a low point in the study of NLP. Despite this, researchers in developed countries continued their research and, in the 1970s, successfully developed statistical methods based on the Hidden Markov Model (HMM) that greatly benefited the field of speech recognition. Furthermore, the development of discourse analysis in the early 1980s also made significant progress in human-machine dialog. Few years later, as NLP researchers reflected on past research, finite state models and statistical methods also began to revive.

After the mid-1990s, two events fundamentally promoted the recovery and development of NLP research. The first was the rapid increase of speed and storage in computers, which improved the material foundation for NLP and made the commercial development of speech and language processing possible. The second event was the commercialization of the Internet in 1994. Overall, the development of network technology during this period has made the demand for natural language-based information retrieval and information extraction more prominent. In 2001, Yoshua Bengio proposed the first neural language model, the feed-forward neural network. In 2008, Ronan Collobert was the first to apply multitasking to NLP's neural network. In 2013, Tomas Mikolov developed Word2Vec, a statistical method that can effectively learn independent word embedding from a text corpus based on neural networks in Google. In 2014, Ilya Sutskever proposed the sequence-to-sequence learning model, a general framework for mapping a sequence to another sequence using a neural network. Based on these statistical models, people make machines better understand and produce human language.

Recently, in addition to research focused on improving existing algorithms or proposing new methods for NLU and NLG in the field of computer science, scholars in other fields have begun to realize the value of NLP and apply it to research. For example, many scholars in management have proposed new methods and improved algorithms according to different management scenarios. In the field of classification, for instance, the Deidentification and Anonymization for Sharing medical Texts (DAST) framework was proposed as a means for clustering medical text records (Li & Qin, 2017). Xu, Liao, Lau, and Zhao (2014) tried to overcome the difficulty of

achieving accurate annotation in a mass of manually labeled data by introducing a novel active learning method for large-margin classifiers. Gabel, Guhl, and Klapper (2019) combined a neural network language model and dimensionality reduction to propose a new approach for analyzing market structure. Similarly, T. Y. Lee and Bradlow (2011) focused on market structure analysis and presented a method for identifying product attributes and a brand's relative position based on online customer reviews. Das and Chen (2007) developed a methodology for extracting small investor sentiment from stock message boards. In terms of sentiment analysis, Fang, Dutta, and Datta (2014) proposed a hybrid approach that uses sentiment analysis to resolve the high cost of acquiring labeled data. To improve topic modeling, Toubia, Iyengar, Bunnell, and Lemaire (2018) proposed guided latent Dirichlet allocation (LDA) to extract features of entertainment products. Bao and Datta (2014) also improved topic modeling by taking sentence structure into account and developing a sent-LDA model for measuring risk types from 10-K filings. Ansari, Li, and Zhang (2018) developed a novel covariate-guided, heterogeneous, supervised topic model to succinctly characterize products in terms of latent topics and specify consumer preferences via topics. J. Liu and Toubia (2018) took into account the semantic relation between two different types of documents and accordingly proposed a hierarchically dual latent Dirichlet allocation (HDLDA). Grounded in the Correlated Topic Model (CTM), Trusov, Ma, and Jamal (2016) incorporated visitation intensity, heterogeneity, and dynamics to propose an approach that uncovers individual user profiles from online surfing data. Bansal et al. (2019) proposed a novel Fuzzy Analytical Hierarchical process (FAHP) based on sentence structure, term-frequency, thematic word, and sentence proximity to evaluate sentences in the production of an efficient and effective summary of legal judgement. To annotate electronic medical records (EMR) accurately and effectively, Xu et al. (2016) suggested an indirect annotation method. This method splits medical terms into words and, applies phrase sense disambiguation (PSD) to 'compound terms' in order to refine candidate words for annotation. Taken together, these studies indicate that NLP has been increasingly applied to a variety of topics in management research. Our paper thus scrutinizes different disciplines' application of NLP to tease out how it has been used and identify research opportunities based on interdisciplinary comparison.

3. Literature review methodology

To obtain a comprehensive analysis on the preeminent literature in management (Webster & Watson, 2002), we searched articles in 24 leading business journals identified by the University of Texas at Dallas: *Academy of Management Journal*, *Academy of Management Review*, *Administrative Science Quarterly*, *Management Science*, *Strategic Management Journal*, *Journal of International Business Studies*, *Journal of Finance*, *Journal of Financial Economics*, *Review of Financial Studies*, *Journal of Accounting and Economics*, *Journal of Accounting Research*, *Accounting Review*, *Manufacturing and Service Operations Management*, *Operations Research*, *Journal of Operations Management*, *Production and Operations Management*, *Journal of Consumer Research*, *Journal of Marketing*, *Marketing Science*, *Journal of Marketing Research*, *Organization Science*, *MIS Quarterly*, *Journal on Computing*, and *Information Systems Research*. The keywords "natural language processing" were used in searching a total of 123 articles.

We located sentences that mentioned NLP to determine whether papers were directly related to this technique. 50 articles were deleted as a result because some articles proposed the application of NLP in future research, some discarded this technique due to stated drawbacks, and some only mentioned NLP in the author biography. 73 articles were thus selected as the paper pool for further analysis. *Management Science* had the greatest number of articles (16), *Marketing Science* had the second (12), and the *Journal of Marketing Research* and *Information Systems Research* had the third (8) (Figure 1).

After identifying the articles that substantively addressed NLP, we then identified the articles' distribution. As depicted in Figure 2, the overall distribution shows that the past two years have witnessed a rapid growth in applications of NLP to management research. Furthermore, according to Figure 3, the growth of research on the
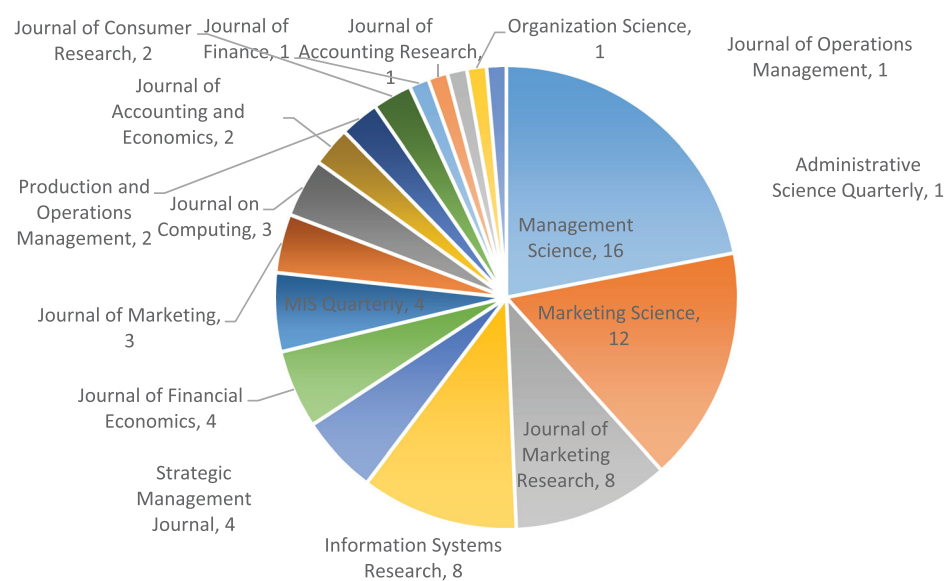

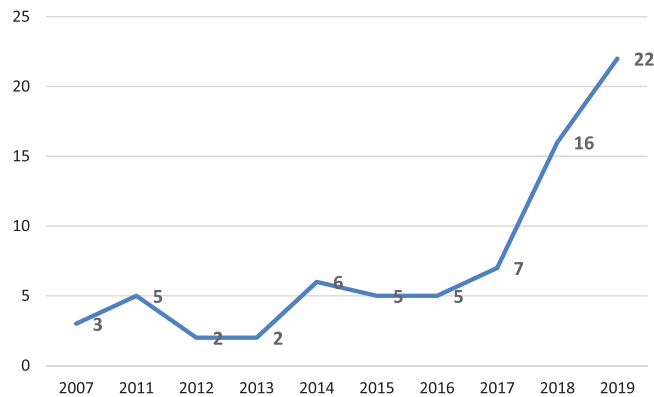Figure 1.   Number of Articles from UTD24.

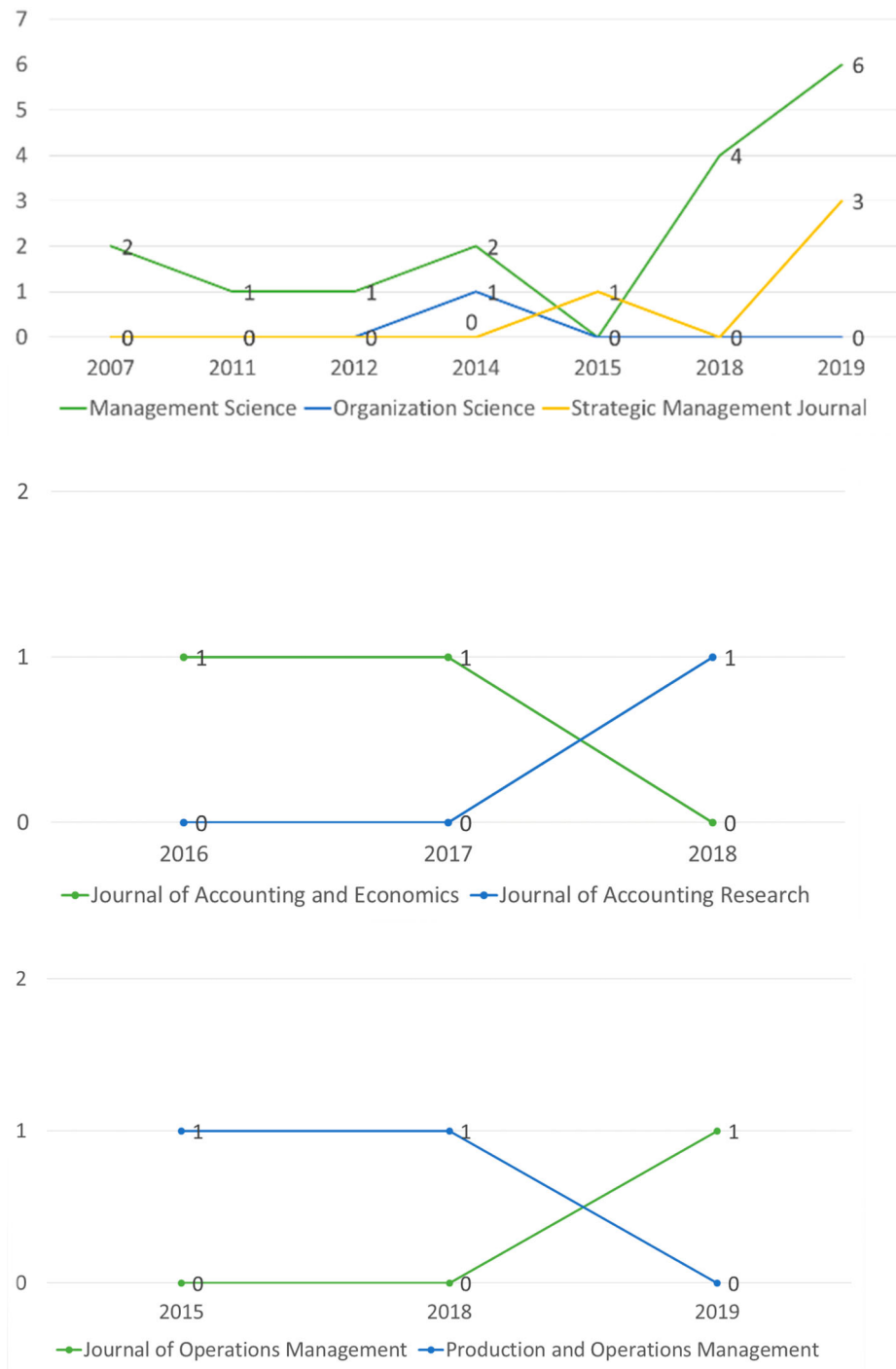Figure 2.   Article Distribution between 2007 and 2019.

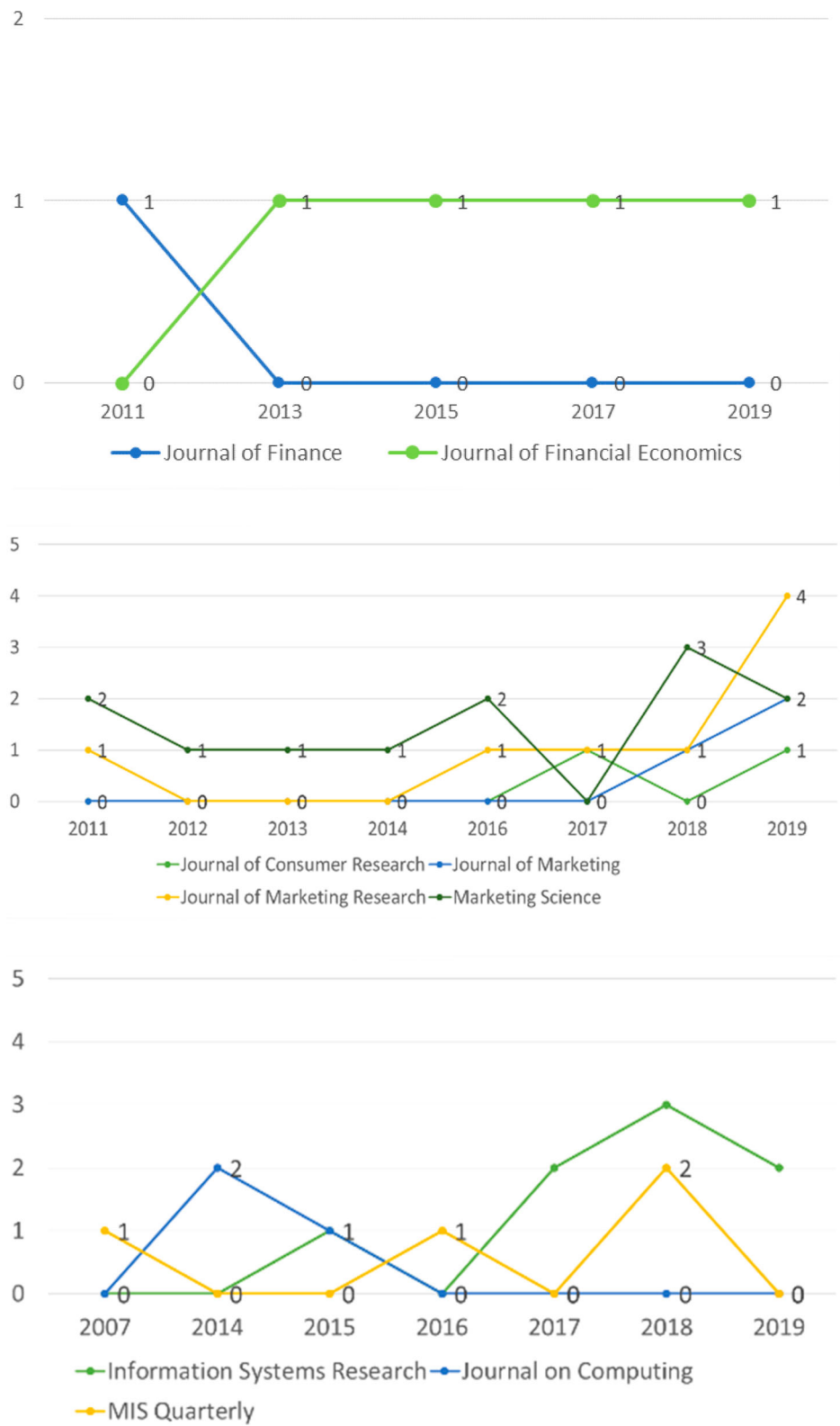Figure 3.   Article Distribution among UTD24 between 2007 and 2019.
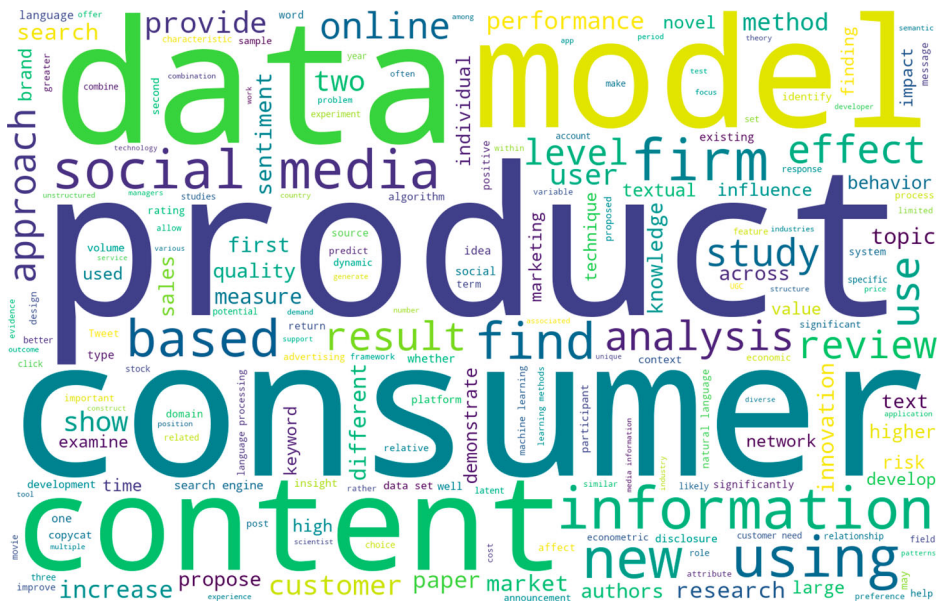
Figure 3. Continued

Figure 4. Word Cloud of NLP Applications in Management.

application of NLP in different management disciplines is accelerating, especially in marketing, management science, and strategic management topics.

To further understand the topic of these papers, a word cloud was used to analyze their abstracts. We implemented a Python project to text mine and synthesize the contents of each of the 72 papers. As shown in Figure 4, common words related to NLP were collated and constructed as a word cloud, from which a holistic picture of the main application of NLP in management came to be illustrated. Based on text mining results and the selected articles, we found NLP has been applied to several areas of interest within the management research field and highlight in this study the areas of consumer research, product management, social media, and information management.

4. Natural language processing in management

As textual data are ubiquitously generated with business operations, NLP research in different management disciplines focuses on a variety of research topics and analytical methods. Specifically, studies in information systems and marketing fields tend to utilize NLP to capture customers' opinions and behaviors. Other disciplines, such as accounting, finance, and operations management, tend focus on extracting nuanced information from firms' filings, such as annual reports and patent data. For analytical methods, information systems, marketing, and operations management scholars have adopted deep learning algorithms that take the context of words into account, which means that computers can express words according to their semantic value to "understand" the text more accurately. In other fields, traditional NLP methods like calculating word frequency remain the mainstream in terms of analytical methods. These differences make it necessary to determine each discipline's research topics and

analytical methods to identify research opportunities based on comparisons among disciplines. As such, this section analyzes these differences to shed light on the current progress of NLP research and research opportunities in management field.

4.1. *Natural language processing in accounting*

4.1.1. *Application*

Accounting scholars mainly explore different factors that affect firm performance such as firm value. With the rapid development of Internet and social media, customers use on online platforms to express their opinions about companies or brands (e.g. Twitter). This information and accompanying sentiments can affect other customers' purchase behavior (Tang, 2018) and investors' investment behavior (Nguyen, Calantone, & Krishnan, 2019), which can influence a firm's value. In this process, investor types and tweet senders or commenters present differentiated influencing mechanisms. For example, dedicated and quasi-index institutional investors increase their stock-holdings in response to increases in positive sentiment in tweet, but transient investors decrease their stockholdings in response to increases in negative sentiment. Further, the filings published by companies themselves, such as 10-K filings, contain information on the business activities of enterprises (Frankel, Jennings, & Lee, 2016), which impacts market reaction and can thus predict firms' future performance. Analysts playing information intermediary roles have confirmed and clarified company filings (e.g. conference calls), driving investors to place a greater value on new information in analyst reports when managers face greater incentives to withhold value-relevant information (Huang, Lehavy, Zang, & Zheng, 2018).

4.1.2. *Approach*

Classification algorithms, topic modeling, and sentiment analysis are the most commonly used methods in accounting research. In text classification, supervised learning methods (e.g. SVR) enable researchers to explore the latent relationship among filings via examinations of how specific words and phrases in Management Discussion and Analysis (MD&A) explain accrual levels (Frankel et al., 2016). In topic modeling, LDA, or the classification of texts into a specified number of topics, is the most frequently applied learning method. Accounting researchers set the topic number based on NLP standards, such as perplexity score (Huang et al., 2018) or management interpretation of filings, i.e. hiring several individuals with accounting or financial backgrounds and manually grouping the LDA topics into broad categories (Dyer, Lang, & Stice-Lawrence, 2017). The interpretation of LDA results has verified topic changes over time (Dyer et al., 2017) some of which have served expand and revise certain concepts. For example, to capture the discovery and interpretation roles of analyst reporting, Huang et al. (2018) operationalized the discovery role as cases in which analysts discussed topics that received little or no mention by managers during conference calls and the interpretation role as cases in which analysts discussed conference call topics in their prompt reports. For sentiment analysis, the most common application is examining customers' attitudes and emotions towards a brand or the market as well as the sharing of these sentiments on social media platforms (Tang, 2018). Additionally, researchers have further classified emotions based on sentiment lexicon or algorithms to investigate the different mechanisms of each emotion. For

example, Nguyen et al. (2019) captured eight types of emotion—joy, trust, surprise, anticipation, sadness, fear, disgust, and anger—and identified their influences on institutional investors' investment decisions and firm value. Further, the Infegy Atlas social media analysis platform and the National Research Council Canada (NRC)'s Word-Emotion Association Lexicon are relatively more suitable for accounting research given the specific meaning of proper nouns in accounting contexts.

In conclusion, accounting scholars have mainly focused on the explaining function of the textual features of annual reports, announcements, and customers' opinions posted on social media by attending to topics, tone, and sentiment. The most frequent algorithms applied by these scholars includes LDA and sentiment analysis methods.

4.2. *Natural language processing in finance*

4.2.1. *Application*

Researchers in finance have generally discussed the prediction effect of news' and announcements' linguistics features in terms of tone, sentiment, frequency, and unusualness (entropy) of word flow on market reaction and corporate sales or opposite causality (Calomiris & Mamaysky, 2019; Froot, Kang, Ozik, & Sadka, 2017; Hendershott, Livdan, & Schürhoff, 2015). Further, they have investigated the differences of this prediction effect at the country-level via the division of developed and emerging markets (Calomiris & Mamaysky, 2019).

4.2.2. *Approach*

Some platforms or databases offer released news and analytical tools to assist financing researchers' analysis, such as Thomson Reuters News Analytics (TRNA), the Thomson Reuters Machine Readable News archive, and Reuters Data Feed (RDF). The RDF offers a Reuters algorithm that identifies the degree of positivity, neutrality, and negativity of words used in articles by firms' to determine tone. However, though there are emerging tools that assist in sentiment analysis, many financing researchers assert that a major challenge in management research is that the word list developed for psychology and sociology is inappropriate for such a context. In an effort to amend this, T. Loughran and Mcdonald (2011) used Harvard-IV-4 TagNeg's (H4N) word list to reclassify a selection of negative words and develop six new word lists that better reflect tone in financial texts.

In conclusion, the research topics in the financial field tend to be more focused than those of accounting, and existing databases and platforms have assisted scholars in the application of NLP to the calculation of related constructs.

4.3. *Natural language processing in information systems*

4.3.1. *Application*

Within the discipline of information systems, the most investigated topics regarding NLP application are organization research and consumer behavior. Studies on organization have encompassed issues regarding R&D investment, innovation, and more. New entry threats in the fast-changing information technology industry have been shown to have significant impacts on firm decision making, such as in the case of R&D investment and the characteristics of industry that differentiate the mechanism

(Pan, Huang, & Gopal, 2019). Moreover, researchers have proven that the capabilities of data analytics have contributed to firms' innovation, a finding that has been classified as process improvement and new technology improvement, or centralized innovation and decentralized innovation (Wu, Hitt, & Lou, 2019; Wu, Lou, & Hitt, 2019). In online communities, interpersonal communication reflects a user's personality and how it can influence other members' identification and how similarities between two users can accentuate purchasing behavior (Adamopoulos, Ghose, & Todri, 2018; Hong & Pavlou, 2017). Interpersonal interactions among community members enable them to share insights and spur innovative ideas (Hwang, Singh, & Argote, 2019). Such interactions first require searching for specific information. For instance, consumers' preferences and search habits can be ascertained via keyword, which demonstrates how search engine selection and design is crucial in consumers' subsequent purchasing and search behavior and crucial to understanding search performance (e.g. click-through rate) and advertising performance (e.g. ad position) (Ghose, Ipeirotis, & Li, 2019; Gong, Abhisek, & Li, 2018).

4.3.2. *Approach*

Information systems scholars have mainly applied five categories of NLP: methods in text preprocessing, text representation, classification, topic modeling, and deep learning. In the preprocessing method, parts-of-speech are applied to the identification of nouns, verbs, or adjectives and the resulting word frequency is used for further analysis. Researchers choose different parts-of-speech according to their research contexts, e.g. evaluation phrases (adjectives and adverbs) to measure the service features (Ghose, Ipeirotis, & Li, 2012). This method is often combined with clustering or classification algorithms and manual coding for information extraction and accuracy improvement.

In text representation, term frequency-inverse document frequency (TF-IDF) is the most common method for getting word vectors and is often combined with cosine similarity to measure constructs. For example, to capture whether a service provider's skills match project requirements, Hong and Pavlou (2017) applied TF-IDF to analyze user profiles and project descriptions on an online outsourcing platform. They then computed the cosine similarity between the two documents (service provider's profile and project description) for each project-service provider pair. To capture the degree of new entry threat to an incumbent, Pan et al. (2019) calculated the cosine similarity score between companies' business descriptions using TF-IDF weighted words vector. Wang, Li, and Singh (2018) detected functional similarity based on app descriptions and consumer reviews of each app on the iOS App Store using bag-of-words, TF-IDF, and singular value decomposition (SVD), an algorithm for solving the problem of data sparsity in high-dimensional data. Bag-of-words is another frequently used method for measuring word features in documents. In terms of measuring innovation and given the fact that novel ideas can be detected through shifting vocabulary, Wu, Hitt, et al. (2019) applied bag-of-words to the identification of patent terms in patent abstracts collected from NBER Patent Citation Data File. The authors then calculated the age of each word based on when it first appeared in similar technological domains to get the "novelty score."

There are two mainstream methods for classifying texts: supervised learning and unsupervised learning. Ghose et al. (2012) used a supervised learning method to measure two text features—readability (e.g. textual complexity, syllables, spelling

errors) and subjectivity—that were likely to affect consumers' intellectual efforts in internalizing review content. They trained a 4-gram dynamic language model classifier provided by the LingPipe toolkit to derive the probability of subjectivity in customers' reviews. Similarly, Wang et al. (2018) clustered each app's word vector based on TF-IDF using the Markov clustering algorithm to identify original from copycat apps.

Furthermore, it is the general consensus that LDA is well suited to information systems studies and, as such, is commonly used. For example, Adamopoulos et al. (2018) applied several different NLP methods to the analysis of tweets in order to capture both consumers' personalities (i.e. similarities between sender and recipient of word of mouth messages) and characteristics of tweets. Some researchers have tried to construct individuals' information networks on crowdsourcing platforms based on individuals' communication (Hwang et al., 2019). Ghose et al. (2019) also calculated the "Topic Entropy" based on LDA to measure textual topic complexity. To quantify keyword ambiguity, Gong et al. (2018) applied LDA to a corpus constructed using keyword-specific search results and subsequently computed the entropy of topic distribution for each keyword.

Sentiment analysis was sometimes found to be combined with deep learning because the sentiment (positive, negative, or neutral) of the corpus can be understood as a classification problem. Additionally, publicly available commercial sentiment analysis mechanisms based on deep learning (e.g. AlchemyApi) has been shown to provide richer metrics for capturing textual sentiment compared to other naïve metrics (e.g. lexicon-based scores) (Adamopoulos et al., 2018).

Beyond text preprocessing, text representation, classification, and sentiment analysis, the literature in information systems also features other applications of NLP. For instance, Adamopoulos et al. (2018) first preprocessed tweets and other user-generated content, then matched these word vectors with the Linguistic Inquiry and Word Count (LIWC) psycholinguistic dictionary to compute relative scores reflecting the percentile score for specific personality categories. In an effort to better understand leaders' language usage, Johnson, Safadi, and Faraj (2015) explored online community leaders' posts in three communities—Blender Artists, Gearbox Software, and Northern Sounds—from the perspective of morphology, lexical, syntactic, and semantic analysis. Furthermore, Wu, Lou, et al. (2019) turned to online resumes written by current and past employees to measure a firm's data analytics capabilities.

In sum, topics in information systems research using NLP as analytical method cover both organizational and individual levels. The most common methods in information systems studies are bag-of-words, TF-IDF, and LDA, which are generally applied in information extraction to construct measurements, such as customer's personality, UGC's sentiment, firms' data analytics capability, and innovation.

4.4. *Natural language processing in marketing*

4.4.1. *Application*

Marketing research relevant to NLP aims to refine customer experience, improve sales performance, and maintain brand reputation. Customers' behavior regarding searching and browsing online, communicating with other users, and expressing opinions on social media are all activities closely followed by firms. The invention of smartphones enables users to quickly post brief lines of text and express emotion at anytime compared PCs (Melumad, Inman, & Pham, 2019). These technologies shorten the distance

among people, even strangers, and allow for the immediate exchange of ideas, which can drive innovation and allows individual customers to impact product/service ideas (Stephen, Zubcsek, & Goldenberg, 2016). Beyond innovation, social media provides opportunities for customers to share their shopping or usage experience, which contains valuable information about product quality, price, promotion, and service (Archak, Ghose, & Ipeirotis, 2011; Caldieraro, Zhang, Cunha, & Shulman, 2018; Dotzel & Shankar, 2019; Liu, Lee, & Srinivasan, 2019; Liu, Singh, & Srinivasan, 2016; Xiong & Bharadwaj, 2014). Customers' patterns and opinions help firms forecast sales and adjust strategies to positively stimulate demand, though it has been noted that negative customer feedback on social media can turn disastrous if a company does not detect and reduce adverse impacts (Herhausen, Ludwig, Grewal, Wulf, & Schoegel, 2019).

4.4.2. *Approach*

Diversity of research methods is one of the most important concerns for marketing researchers. As such, all six major NLP methods—text preprocessing, text representation, classification, topic modeling, sentiment analysis, and deep learning—are applied in marketing studies. In text preprocessing, part-of-speech is still the most frequently adopted method for identifying characteristics, e.g. product features (Archak et al., 2011). In combing with the clustering method, researchers can separate sparse words and phrases into a set of similar phrases then extract the "context" in which a particular word appears.

The most frequently used method in text representation is bag-of-words which is well-established and easy to use. For instance, to provide implications for the design of international products, Hermosilla, Gutiérrez-Navratil, and Prieto-Rodríguez (2018) relied on bag-of-words and the "cosine similarity" metric to process IMDb's "summary plot" and "synopsis" to measure storyline similarity. In addition, Liu et al. (2016) used n-gram to examine the variety of content in Tweets.

For classification, labeling is a great challenge. Scholars have either hired research assistants (Villarroel Ordenes et al., 2019) or outsourced these tasks online (Lee, Hosanagar, & Nair, 2018), via crowdsourcing through sites like Amazon Mechanical Turk. Moreover, to improve the accuracy of learning methods, the ensemble learning model —which includes logistic regression, Naive Bayes, and support vector machines—is frequently applied to identify large numbers of text features, such as brand mention, product mention, readability index, and more (Lee et al., 2018). Barring regular algorithms, advanced algorithms have sprung up in the field of computer science and have been actively adopted by marketing scholars. For example, Liu et al. (2016) used DynamicLMClassifier to measure sentiment in tweets. However, some studies have focused more on the clustering algorithm without labeled data. This hierarchical agglomerative clustering algorithm is the most popular method in marketing for extracting firm and product features (Archak et al., 2011; Ghose et al., 2019). Nowadays, the NLP module is embedded in most business software. For instance, Caldieraro et al. (2018) used classification and linguistic processing algorithms provided by SPSS software to identify the top concepts that appear in the loan description on Lending Club. Based on these concepts, the authors used the top six categories as dummy variables to control for variations across the loan description content.

Meanwhile, LDA is still the basic algorithm in topic modeling. Rutz, Sonnier, and Trusov (2017) adopted bag-of-words and LDA to capture the content of text ads to get a better understanding of the perceptions and performance of paid search ads. Beyond

studies of online message management, Dotzel and Shankar (2019) applied two topic modeling methods—LDA and the correlated topic model (CTM)—to measure the quality of innovations.

There are two mainstream methods in sentiment analysis: the lexicon-based method and algorithm. In the lexicon method field, based on Linguistic Inquiry and Word Count (LIWC) text-mining dictionaries and the Stanford Sentence and Grammatical Dependency Parser. Herhausen et al. (2019) calculated the proportion of words associated with negative emotions in Facebook posts to identify and measure the intensity of high and low arousal for each negative post. To do so, they used existing LIWC dictionaries for fear/anxiety, anger, and sadness, and developed a new dictionary for locating disgust in order to calculate an intensity score per aforementioned emotion word category. And, for algorithms, Dotzel and Shankar (2019) applied a sentiment analysis algorithm—tidytext—for the sake of capturing the overall sentiment of innovation announcements. To combine the advantages of both approaches, several researchers conducted comprehensive sentiment analyses of product reviews (Ordenes, Ludwig, De Ruyter, Grewal, & Wetzels, 2017), social media messages (Melumad et al., 2019), and online buzz for video games (Xiong & Bharadwaj, 2014). The methods they used contained: Stanford Sentence and Grammatical Dependency Parse, SentiStrength, Linguistic Inquiry and Word Count (LIWC), and Scoutlab/Lithium social media sentiment monitoring systems.

With the rapid development of deep learning to ensure high accuracy, many researchers have introduced these algorithms in NLP. For example, Liu et al. (2019) combined Amazon MTurkers' labeled data and deep learning—Convolutional Neural Networks (CNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM)—to capture the quality and price information of products. Further, Hu, Dang, and Chintagunta (2019) analyzed the quality of Groupon deals based on MTurkers' labeled data and a CNN model.

Finally, some studies did not specifically highlight the algorithms they used. For instance, based on customers' word-of-mouth on Facebook and Twitter, Kumar, Bhaskaran, Mirchandani, and Shah (2013) measured the stickiness index, defined as an array of the degree to which a user or an instance of word of mouth is specific to each category of topics. They did so to identify how closely other keywords, such as "dessert," "sundae," and "food" are associated with a focal keyword like "ice cream." Similarly, to examine inspiration among customers regarding the generation of new ideas, Stephen et al. (2016) quantified the similarity of pairs of ideas at the concept level based on the preprocessing step in NLP.

To conclude, marketing scholars have applied a wide variety of algorithms to their studies, such as part-of-speech, bag-of-words, classification methods, and deep learning algorithms.

4.5. *Natural language processing in strategy*

4.5.1. *Application*

Strategic management studies have mainly focused on startups and mature enterprises. For new entrants, it is critical to strategize how to optimally position themselves within product markets. Barlow, Verhaal, and Angus (2019) proposed that the optimally distinct entry point is at a high level of exemplar similarity (the most salient category member or a clear market leader) and a low level of prototype similarity (the most

representative member in a market category). Lacking management experience is another prominent feature of novice entrepreneurs. The advice offered by their peers have been found to influence their startup's performance (Chatterji, Delecourt, Hasan, & Koning, 2019). For mature companies, it is urgent to achieve innovation for the sake of gaining competitive advantage. To this effect, researchers have analyzed two key factors of innovation, i.e. how far entrepreneurial organizations should search to improve their chances for success (Angus, 2019) and how organizations should combine the distant or diverse knowledge they gather in creativity (Kaplan & Vakili, 2015).

4.5.2. *Approach*

Researchers in strategy usually apply the calculation index in NLP to concept measurement. For example, Barlow et al. (2019) applied NLP to clean app descriptions and calculate the measurement of two concepts—prototype similarity and exemplar similarity. Angus (2019) used basic cosine similarity in NLP to measure the distance between the text descriptions of a developer's first and second apps on the Google Play store. Chatterji et al. (2019) counted the proportion of management words relative to words describing the product, technology or customer after being assigned to managers with formal or informal styles to capture the advice approach. Kaplan and Vakili (2015) used topic-modeling to extract 100 topics from the USPTO document abstract. To identify whether patents originated as novel ideas, they selected all patents over a threshold, which was determined by three expert coders and weighted for topic, as well as the requirement that they appeared in the first 12 months of topic formation (based on application date).

   In short, one of the issues that strategic management scholars are most concerned about is how to make an appropriate strategy to gain competitive advantage. The strategy is captured mainly via topic modeling and sentiment analysis.

4.6.   *Natural language processing in operations management*

4.6.1.   *Application*

Researchers in the operations management field have applied several NLP methods in a wide range of topics, such as product management, operational efficiency, inventory management, and sales forecast.

4.6.2.   *Approach*

Jain, Girotra, and Netessine (2014) combined text-pattern recognition algorithms and manual inspection to extract supplied information from the raw, unprocessed bill of lading forms for each import transaction. Ko, Mai, Shan, and Zhang (2019) crawled physician review data from Vitals (http://www.vitals.com) to measure the importance of technical and interpersonal quality dimensions to patients. Cui, Gallino, Moreno, and Zhang (2018) used valence as well as Facebook comments' informativeness and sentiment to measure endorsement. Informativeness was determined by the number of words, sentences, and unique words in a comment. Because comments on social media are usually short and written by heterogeneous users, simple sentiment analysis cannot accurately classify them. As such, the authors

adopted the recursive neural tensor network (RNTN) on top of the Stanford Sentiment Treebank corpus to classify each comment into categories of positive, negative, or neutral. Hoberg and Phillips (2018) analyzed the business descriptions in 10-K filings to capture how product language overlaps at different levels (across-industry, within-industry, between industry). They formed word vectors for each firm based on the text in each firm's product description and omitted common words that are used by more than 25% of all firms. Cosine similarity for two word vectors was then calculated to measure similarities in across-industry and within-industry language.

Essentially, the issues and applications of NLP in OM research are diverse. In addition to traditional word frequency analysis, OM research has also used deep learning algorithms for text classification.

4.7. *Natural language processing in other management disciplines*

NLP enables scholars to analyze political events. For example, Berman, Melumad, Humphrey, and Meyer (2019) examined tweets posted during and shortly after four key debates leading up to the 2016 U.S. presidential election to understand how tweeting evolves, namely in terms of changes in linguistic style, topical content, and robotic detection during significant political events. In another study, Srivastava, Goldberg, Manian, and Potts (2018) analyzed electronic messages exchanged among full-time employees to measure cultural fit. They used LIWC to measure linguistic style and combined these linguistic categories with Jensen-Shannon (JS) divergence to calculate cultural fit. Oettl (2012) used acknowledgment in immunology papers to measure scientists' helpfulness to others. The author applied name identification algorithms proposed by Councill, Giles, Han, and Manavoglu (2005) to discern those who achieved what they called the "helpfulness star.[6]" In addition, latent semantic analysis enabled Tuertscher, Garud, and Kumaraswamy (2014) to explore the prevalence of justifications in documents produced by A Toroidal LHC Apparatus (ATLAS) scientists. They tested the relationship between justification and knowledge sharing at ATLAS, a complex technological system developed at CERN (Conseil Européenn pour la Recherche Nucléaire), Geneva. To capture the evolution of the perception of expertise over time, Croidieu and Kim (2018) adopted LDA to extract 98 topics and manually labeled these raw topics into descriptive codes, first-order concepts, then second-order themes, and developed four aggregate dimensions for expertise legitimation.

4.8. *Inter-disciplinary comparisons and research opportunities*

By reviewing relevant literature in all areas of management, we found that information systems and marketing studies analyze data from multiple sources, such as social media, online reviews, and company announcements. Data are collected at different levels, ranging from customer opinion to enterprise behavior. Datasets used by studies in accounting, strategy, and operations management are relatively limited, focusing mostly on documents issued by companies.

Market reaction is the concentration of finance studies using the NLP technique. We found that analytical methods, classification, topic modeling, and sentiment analysis are frequently used in each research field. These traditional algorithms enable scholars to explore themes, similarities and differences between documents, and the sentiment orientation of texts. However, deep learning, an advanced NLP method

for mining text thoroughly, has been applied only in information systems and marketing studies. In terms of theoretical framework, almost every study in the information systems, operations management, and marketing fields adopt extant management theories to explain the research model and justify theoretical contributions. However, for the fields of accounting, finance, and strategy, there remains a lack of attention to utilizing and contributing to existing theories.

Analyzing past literature across disciplines also provides a pathway for conducting NLP research in the management field. For accounting, a salient problem is the inaccurate processing of specific terms in accounting filings, such as "taxes" or "liabilities." In previous studies, these words have usually been treated with sentiment orientation in sentiment analysis or lexicon, both of which do not conform to the research context. Therefore, accounting researchers may use deep learning to identify these terms in improving algorithmic accuracy. Rather than relying on existing databases or platforms, financing researchers might turn to newly proposed methods or lexicons that are more suitable for their research scenarios. Similarly, information systems and marketing scholars might take the role of forerunner to explore NLP as a way of enriching traditional management theories. For research in strategy, the key to adopting NLP for the extraction of accurate information is to build association rules between texts and strategies. Deep learning is a feasible way of doing this as it allows for the extraction and analysis of relevant texts to capture strategies. Finally, it is worth mentioning here that, due to a high variety of research topics in operations management studies, the success of NLP application depends on method choice. Finding relevant and effective analytical methods in operations management research requires scholars to refer to the literature in other areas with relevant research topics.

5. Natural language processing methods and toolkits

Given the abundant research opportunities for management scholars to utilize NLP to generate findings, a step-by-step tutorial on how to perform NLP is imperative. This section addresses this need by reviewing the NLP methodologies most often used in research, including techniques needed to convert text into machine-readable representations as well as procedures for extracting the textual information incorporated into subsequent modeling and analyses. The objective of this section is to introduce available techniques, frequently used packages, and advantages and drawbacks of each method.

In general, the process of NLP can be divided into four parts: text preprocessing, text representation, model training, and model evaluation. Text preprocessing aims to get a "clean" text in order to improve the efficiency and accuracy of the following analysis by removing meaningless symbols, checking spelling mistakes, tokenization, POS tagging, removing stop words, and stemming. After preprocessing the text, researchers must choose how best to represent the words in type that can be effectively and accurately calculated by computers. Text representation converts words into numbers, which will always be vector or matrix in this step. Based on these word vectors, it is possible to apply algorithms—such as classification, sentiment analysis, and topic extraction, etc.—to train a model for the sake of solving the real problem. After training the model, researchers then need to evaluate the model to ensure it has optimal generalizability to other corpora. Figure 5 demonstrates a flow chart depicting the steps of conducting NLP.
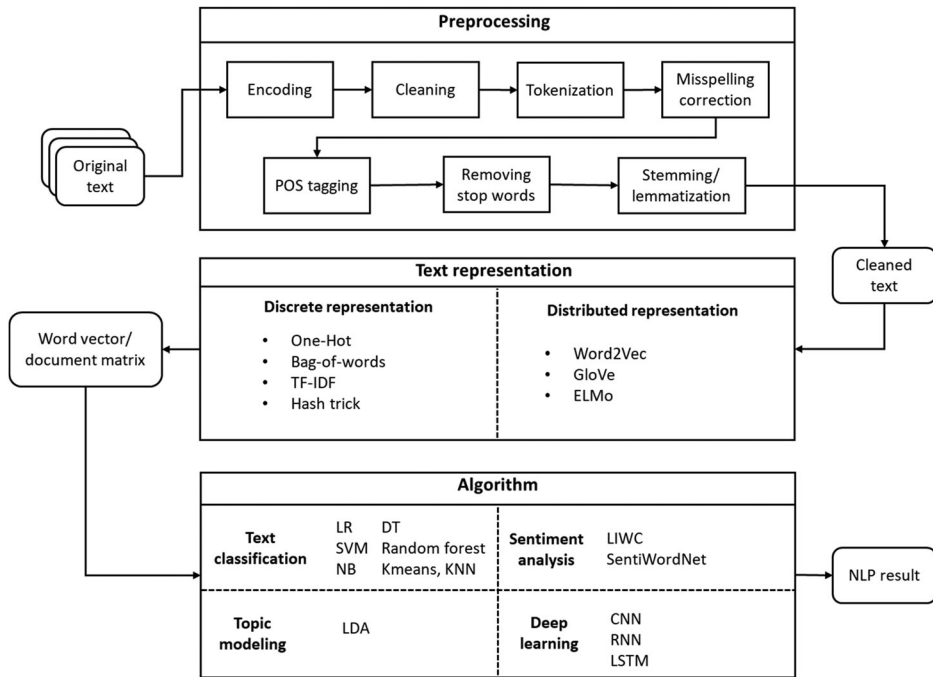
Figure 5.   Flow Chart of NLP.

5.1.   *Preprocessing*

5.1.1.   *Encoding*

The first step in NLP is reading data for further analysis. However, even in English, a recurring challenge is the handling of character encoding. To address this issue, researchers can first write the script in UTF-8. If encoding settings are ignored, loading data directly will fail due to the default encoding for Windows. It is important to make sure every file is in the "correct" encoding format. Encoding settings can easily be changed in txt files before reading them.

5.1.2.   *Cleaning*

Recently, many researchers have crawled text from a variety of websites (e.g. Twitter, Facebook, TripAdvisor.com, etc.). This data usually contains HTML tags and non-textual information, such as images and emojis, which should be cleaned or removed from the data set. Given the abundance of online data and the irregularity of information to be removed, cleaning is a very involved process. Some packages in Python, such as pattern.web, can be applied to remove HTML tags. Additionally, the combination of beautifulsoup and Regular Expression in Python can also be used to recognize and remove tags. However, cleaning needs may depend on the purpose of the analysis, images, and other nontextual information that must be retained. For example, some researchers have paid explicit attention to emoticons to enrich sentiment dictionaries (Lee et al., 2018; Ordenes et al., 2017). In this step of the process, researchers should be mindful of and remove phrases that are

automatically generated by computers that may occur within the text (e.g. "HTML"). Doing so researchers may keep the exact text they want to use for analysis in research.

5.1.3. *Tokenization*

Tokenization is the process of breaking the text into units, often in the form of words and sentences. When tokenizing, the delimiters that define a token (space, period, semicolon, etc.) need to be determined. For example, in English, space is used to determine a word. There are many tokenization packages in Python that deliver smart tokenization procedures, e.g. NLTK and StanfordCoreNLP. Regardless, researchers are advised to pay close attention to instances specific to the textual corpora. For example, when analyzing corporations' 10-K filings, it is usually necessary to encode corporation name before tokenization.

5.1.4. *Misspelling correction*

The misspelling detection process usually consists of a check on, whether it be via an input string as a valid index or a dictionary word and the correction of spelling errors. N-gram and dictionary lookup are two well-known misspelling detection techniques. For spelling correction algorithms, the most studied are edit distance, similarity keys, rule-based techniques, n-gram-based techniques, probabilistic techniques, and neural networks. Most text-mining packages have prepackaged spellers that can help correct spelling mistakes (e.g. the PyEnchant). In using these spellers, researchers should be aware of the domain-specific language that might not appear in the speller orthat the speller might incorrectly "fix."

5.1.5. *POS tagging*

POS tagging is an acronym for part-of-speech tagging, a process of attaching each word in a sentence with a suitable tag from a given set of tags (e.g,. N represents noun, ADJ represents adjective, ADV represents verb, etc.) according to the definition, context, and part of speech. An example could be the relationship between adjacent or related words in a phrase, sentence, or paragraph. Some marked corpora can be easily applied in management research, such as NLTK and Pattern in Python. In management research, some researchers have only analyzed nouns and verbs with the help of this technique (Pan et al., 2019; Wang et al., 2018).

5.1.6. *Removing stop words*

Stop words are frequently occurring words in a natural language that carry very little or no significant semantic context in a sentence and are unnecessary for certain NLP applications, such as "a" and "the" in English. Common NLP tools (e.g. NLTK in Python) that contain predefined lists of such stop words can be applied to remove these words. Furthermore, researchers can download online lists of stop words and amend them according to their research purposes. For instance, it is sensible to add common, domain-specific words to these lists, e.g. brand names in a corpus of tweets about the brand.

5.1.7. *Stemming/lemmatization*

Lemmatization is the process of reducing the words in inflectional forms—usually meaningful words—to general forms. Stemming refers to extracted word stem or root forms that cannot necessarily express complete semantic thought. Stemming is often applied in information retrieval for extended retrieval. In application, there are also different emphases. Meanwhile, lemmatization is mainly applied in text mining and NLP and is used for finer-grained and more accurate text analysis and expression, e.g. machine translation. Several prepackaged stemmers exist in most text-mining tools, e.g. WordNet, SpaCy, and Snowball in Python.

5.2. *Text representation*

Although preprocessing enables text data—unstructured data that is converted into structured data—it is still necessary to have effective methods for document representation. These methods are required for further advanced text analysis, such as information extraction, sentiment analysis, text classification, and more. Only in this way can computers "understand" the text.

The two most common representations of text include discrete representation and distributed representation. One-Hot encoding is a traditional and basic feature representation of a word. This method represents a word as a vector whose dimension is the length of a text generated through a corpus in which the position of the current word has a value of 1 and the position of the rest is 0. The DictVectorizer of scikit-learn (sklearn) in Python can be applied to get this kind of vectorized text. However, One-Hot cannot measure the relationship between different words and the importance of words. Moreover, the matrix of the document is sparse, which wastes a large amount of computing and storage resources. Alternately, bag-of-words (BOW) is used to form a vector representing a document using the frequency count of each term in the document. Sklearn in Python also offers CountVectorizer to rapidly generate document matrices. Unfortunately, the BOW representation scheme has its own limitations: (1) high dimensionality of the representation; (2) loss of the semantic relationship that exists among terms in a document; and (3) interference of common words with high word frequency. Term frequency-inverse document frequency (TF-IDF) addresses BOW's third weakness. TF is the frequency of a word appearing in the document, and IDF represents the proportion of the text containing a certain word in the corpus. Equation (1) is the classical formula of TF-IDF used for text representation. $w_{i,j}$ is the weight for word i in document j, $tf_{i,j}$ is the word frequency of word i in document j, df_i is the document frequency of term i in the whole corpus, and N is the number of documents in the corpus. TfidfVectorizer and the combination of CountVectorizer and TfidfTransformer in sklearn can be used to generate TF-IDF document matrices. A major problem of TF-IDF similar to that of One-Hot and BOW is the high dimensionality of document matrices, or the size of the whole vocabulary across the entire dataset, which results in huge computations.

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right) \tag{1}$$

Hash Trick, the most common text dimension reduction algorithm, can be used to solve the high dimensionality problem mentioned above. The resulting text after hash trick dimension reduction can no longer be considered a representative of the original text in terms of meaning. Therefore, as long as the vector of the corpus is not too large, it is more optimal to use the above three methods. As these methods are highly explanatory, we will know which word corresponds to each one-dimensional feature and from there can modify the weight of each word to further improve text representation. If the dataset is large-scale, Hash Trick can improve vector generation speed so the vectors after dimensionality reduction can still aid in completing classification and clustering. In the HashingVectorizer of sklearn, the algorithm based on a signed hash trick is implemented.

The key idea of distributed representation is that a word's meaning is given according to the words that frequently appear nearby. Word2Vec makes use of the context information of the word, thus the semantic information is richer and the trained word vector is low-dimensional and dense. Gensim in Python offers the word2vec method for training the model. Researchers can choose appropriate corpora to train word embedding according to their research contexts. Word2Vec is based on Cbow/Skip-Gram, a local context window method that cannot retain the overall relationship between words, which often results in too much weigh in high exposure vocabulary. However, Global Vector (GloVe) combines the global statistics of the Latent Semantic Analysis (LSA) and the local context window, incorporating global a priori statistical information that can speed up the training of the model and control the relative weight of the words. ELMo is a new type of deep contextualized word representation that models the complex features of words (such as syntax and semantics) and the changes in words in the context of language (i.e. modeling polysemous words). Unlike Word2Vec, ELMo's main purpose is to train a complete language model and use that model to process text that needs to be trained. From there, it generates a corresponding word vector. In this way, different word vectors can be generated in different sentences for the same word in ELMo.

5.3. *Modeling*

In this section, we review common algorithms and models that have been applied in management studies in order to inspire management scholars to integrate these models into management research appropriately.

5.3.1. *Text classification*

Most machine learning methods have been applied in the field of text classification as opposed to other fields. These methods are fall into two categories: supervised learning and unsupervised learning. Data labeling is a major challenge in supervised learning. Researchers can either download labeled data sets or conduct a survey on Amazon MTurk to outsource labeling of data (Ghose et al., 2019; Wang et al., 2018). The most popular supervised learning algorithms are Logistic Regression (LR), Support Vector Machine (SVM), Naïve Bayes (NB), Decision Tree (DT), and Random Forest. Precision, recall, F1, AUC, and ROC are commonly used to compare the performance of different algorithms so researchers can choose the algorithm that works best for their purposes. Kmeans and K-Nearest-Neighbour (KNN) are the most

frequently used unsupervised clustering algorithms. Although there is no need to label data before training, interpreting clustering results is often difficult. Because clustering algorithms are usually calculated based on the distance between vectors, results calculated from these algorithms differ from the ones generated from researchers' logical analysis. This discrepancy leads to a situation that irrelevant words in logical analysis appear in the same category. Researchers must be patient in identifying patterns in the results. Sklearn in Python provides various kinds of classification and clustering algorithms.

5.3.2. *Sentiment analysis*

The above-mentioned classification methods can also be applied in sentiment analysis if the data set contains sentiment labels. Another way to capture document or sentence sentiment is the lexicon. Each word is assigned a sentiment score in the dictionary and the sum of all words in a sentence or document is the final sentiment score for the sentence or document in question. The most popular lexicons are Linguistic Inquiry and Word Count (LIWC) psycholinguistic dictionary, SentiWordNet, MPQA Subjectivity Cues Lexicon, General Inquirer, and Bing Liu Opinion Lexicons.

5.3.3. *Topic modeling*

Topic modeling mimics the data generation process in that the writer chooses a topic to write about and then chooses words to express these topics. Topics are defined as word distributions that commonly co-occur and thus have a certain probability of appearing in a topic. A document, then, is then described as a probabilistic mixture of topics. The most frequently used tool for topic modeling is latent Dirichlet allocation (LDA), as it can be easily realized with the help of gensim in Python. The output of LDA are the topics and corresponding keywords. The number of topics and keywords is up to users.

5.3.4. *Deep learning*

With the rapid development of deep learning and big data, some deep learning algorithms are applied in text classification, such as CNN, RNN, BP neural network, LSTM, and more.

5.4. *Model evaluation*

The trained model needs to be evaluated before being implemented in management analytics to ensure the model is adequately generalizable to the corpus. These evaluation indexes assist researchers in choosing the most appropriate model for their research contexts. What follows is a brief breakdown of the most frequently used indicators Table 1.

For classification algorithms, the error rate, accuracy, precision, recall, F1 score, ROC and AUC are all useful for measuring performance. *Error rate* is the proportion of the number of samples misclassified to the total number of samples. *Accuracy* is the proportion of the number of correctly classified samples to the total number of samples. For binary classification missions, precision, recall and F1 score can be

Table 1.   Summary of Management Research Applying Natural Language Processing (NLP).

	Data	Model			Theory	Opportunities
		Algorithm Category	Algorithm	Paper		
Accounting	• **10-K filings** Dyer et al. (2017); Frankel et al. (2016) • **Conference call** Huang et al. (2018) • **Social media** Nguyen et al. (2019); Tang, (2018)	Classification Topic Modeling Sentiment Analysis	SVR LDA Lexicon	Frankel et al. (2016) Dyer et al. (2017); Huang et al. (2018) Nguyen et al. (2019); Loughran and Mcdonald (2011)	• **Psychology theory** Nguyen et al. (2019)	• Adopt deep learning methods to "understand" terms in accounting filings
Finance	• **News** Calomiris & Mamaysky (2019); Hendershott et al. (2015) • **Announcement** Froot et al. (2017)	Classification Sentiment Analysis	Louvain community detection algorithm Lexicon Reuters algorithm	Calomiris and Mamaysky (2019) Calomiris and Mamaysky (2019); Froot et al. (2017) Hendershott et al. (2015)	**N/A**	• Extend the analytical method
Information Systems	• **Online platform** Adamopoulos et al., (2018); Hong & Pavlou, (2017); Hwang et al. (2019); Johnson et al. (2015) • **App store** Wang et al. (2018) • **Patent** Wu, Hitt, et al. (2019) • **Resume**	Preprocessing Text representation Classification	Part-of-speech Bag-of-words TF-IDF Classifier Clustering	Ghose et al. (2019) Wu, Hitt, et al. (2019); Wang et al. (2018) Hong and Pavlou (2017); Pan et al. (2019); Wang et al. (2018) Ghose et al. (2019) Ghose et al. (2019); Wang et al. (2018)	• **Theory of interpersonal similarity** Adamopoulos et al. (2018) • **Signaling theory** Hong & Pavlou (2017) • **Analogical reasoning theory** Hwang et al. (2019) • **Functional leadership theory, leader–member exchange theory, shared leadership theory,**	• Combine the NLP with management theories

	Wu, Hitt, et al. (2019); Wu, Lou, et al. (2019) • **Search engine** Gong et al. (2018)	Topic Modeling	LDA	Adamopoulos et al. (2018); Hwang et al. (2019); Ghose et al. (2019); Gong et al. (2018)	**communication as constitutive of organizing theory** Johnson et al. (2015) • **Optimal sequential search framework** Ghose et al. (2019) • **Economic search theory** Gong et al. (2018)	
		Deep learning	Alchemyapi	Adamopoulos et al. (2018)		
Marketing	• **Social media** Lee et al. (2018); Herhausen et al. (2019); Hermosilla et al. (2018); Kumar et al. (2013); Melumad et al. (2019); Villarroel Ordenes et al. (2019); Liu et al. (2016); Xiong & Bharadwaj (2014) • **Online review** Archak et al. (2011); Ghose et al. (2012); Ordenes et al. (2017); Liu et al. (2019) • **P2P** Caldieraro et al., (2018) • **Announcement** Dotzel & Shankar (2019)	Preprocessing	Part-of-speech	Archak et al. (2011); Ghose et al. (2012)	• **Countersignaling theory** Caldieraro et al. (2018) • **Resource-based view** Dotzel & Shankar (2019) • **Emotion regulation theory** Herhausen et al. (2019) • **Fuzzy-trace theory** Melumad et al. (2019) • **Random utility theory** Rutz et al. (2017) • **Wisdom of crowds, organizational sociology, and networks, brainstorming perspective** Stephen et al. (2016) • **Efficient market theory** Xiong & Bharadwaj (2014)	• Combine the NLP with management theories
		Text representation	Bag-of-words	Hermosilla et al. (2018); Rutz et al. (2017)		
			N-gram	Liu et al. (2016)		
		Classification	Classifier	Villarroel Ordenes et al. (2019); Lee et al. (2018); Liu et al. (2016); Caldieraro et al. (2018)		
			Clustering	Archak et al. (2011); Ghose et al. (2012)		
		Topic Modeling	LDA	Rutz et al. (2017); Dotzel and Shankar (2019)		
			CTM	Dotzel and Shankar (2019)		
		Sentiment Analysis	Lexicon	Herhausen et al. (2019)		
			Algorithm	Dotzel and Shankar (2019)		
		Deep Learning	CNN, RNN, LSTM	Liu et al. (2019); Hu et al. (2019)		

Table 1.   Continued.

	Data	Model			Theory	Opportunities
		Algorithm Category	Algorithm	Paper		
Strategy	• **App store** Angus (2019); Barlow et al. (2019) • **Patent** Kaplan & Vakili (2015)	Topic Modeling Sentiment Analysis		Angus (2019) Kaplan and Vakili (2015); Barlow et al. (2019)	• **Problemistic search theory** Angus (2019) • **Tension theory of creativity** Kaplan & Vakili (2015)	• Explore appropriate corpus • Adopt deep learning to analyze company filings
Operations Management	• **Online platform** Cui et al. (2018); Ko et al. (2019) • **10-K filings** Hoberg & Phillips (2018)	Classification Topic Modeling Sentiment Analysis	Text-pattern recognition algorithm LDA RNTN	Jain et al. (2014) Ko et al. (2019) Cui et al. (2018)	• **Goal orientation, anticipatory, and information economic theories** Ko et al. (2019) • **Operations management theory** Jain et al. (2014)	• Select appropriate model

Table 2. Confusion Matrix.

Actual Value	Predicted Value	
	Positives	Negatives
Positives	True Positive (TP)	False Negative (FN)
Negatives	False Positive (FP)	True Negative (TN)

calculated based on the confusion matrix as shown in Table 2. *Precision* $\left(\frac{TP}{TP + FP}\right)$ represents how many of the predicted positive samples are true positive samples. *Recall* $\left(\frac{TP}{TP + FN}\right)$ represents how many positive examples in the sample were predicted to be correct. *F1 score* $\left(\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times TP}{\text{sample number} + TP - TN}\right)$ measures different preferences for precision/recall. Many algorithms generate a real-valued or probabilistic prediction for the test sample and then compare this predicted value with a classification threshold. If this prediction is greater than the threshold, it is classified as a positive class; otherwise, it is a negative class. Samples are sorted according to the prediction results of the algorithm. They are predicted one by one as positive examples in this order. Two important values are calculated each time and plotted with the horizontal and vertical coordinates respectively to obtain the *ROC curve*. The vertical axis is the True Positive Rate $\left(\frac{TP}{TP + FN}\right)$ and the horizontal axis is the False Positive Rate $\left(\frac{FP}{TN + FP}\right)$. *AUC* is the area under the ROC curve. If the ROC curve of one algorithm can completely cover the other, it can be asserted that the former is better than the latter; but if the two ROC curves overlap, the AUC can be used for judgment.

6. Discussion

6.1. *Managerial challenges and future developments*

6.1.1. *Data*

Given abundances in data, many researchers crawl data from websites such as Twitter, Facebook, and Amazon using API or self-built programs. The cleaning of such massive data sets requires enormous amounts of time, energy, and money as it requires researchers to retool and learn new set of skills. Furthermore, in classification tasks, labeled data is a prerequisite for analysis. However, there exist a few mature labeled datasets in management. To improve the quality and accuracy of algorithms, researchers must code data manually, which is time-consuming and cannot guarantee quality. Fortunately, the combination of manual coding and training algorithms can ensure quality and efficiency (Song, Huang, Tan, & Yu, 2019). While crawling public data on some platforms may be permissible under the law, it may conflict with platforms' terms of service. Through terms of service, platforms develop rules to control access to content that users have chosen to make publicly available (Berger et al., 2019).

6.1.2.   *Model*

Due to a lack of NLP models applicable in management discipline, scholars have turned to improving algorithms (Abbasi, Zhou, Deng, & Zhang, 2018; Rutz, Trusov, & Bucklin, 2011) or developed new dictionaries (Loughran & Mcdonald, 2011) in order to capture specific features of management scenarios. Conventional algorithms are not accurate enough to meet the requirements of specific scenarios, such as small samples. The meaning of some indexes or measures in NLP, such as readability, is still not well defined in management research (Loughran & McDonald, 2016). In the study of management, algorithm accuracy is not the only way to measure algorithms; scholars also need to interpret the results of the algorithm. However, the results of some clustering algorithms can fall short of clearly identifying the specific meaning of each class. Some scholars will again manually code these results (Dyer et al., 2017), which greatly increases costs. Finally, it is also necessary to choose a model appropriate to the research context, as task that also takes time and effort.

6.1.3.   *Theory*

It is difficult to make theoretical contributions purely through NLP. For example, while topic modeling and clustering algorithms are generally applied in management research to study a concept's dimensions, design and interpretation of these dimensions based on extant literature and theory are still necessary. Additionally, the core of many algorithms in NLP is a mathematical calculation that is difficult to understand from a theoretical perspective. Furthermore, because deep learning is a highly nonlinear and very complex model, especially now that it is end-to-end, output corresponds to input. The whole training process is completely modeled as a result, leaving few places for manual intervention. For management, it is often both an accurate result and a process of reasoning. For example, to evaluate customers' satisfaction, it is necessary to evaluate the score correctly. Enterprises, on the other hand, need to know why they got the score so they can determine which part failed to satisfy customers, and how to improve. For the all-black box model, even deep learning ultimately revolutionizes NLP by greatly improving the accuracy of NLP tasks. Yet, it is still difficult for companies and researchers to comprehend for business insights. It has thus been noted that the results of specific NLP models might not be generalizable to other research fields considering the special character of measurement or methods (Hong & Pavlou, 2017).

6.2.   ***Technological challenges and future developments***

6.2.1.   *Natural language understanding*

The core issue of natural language understanding (NLU) in NLP is that it is a prerequisite for many tasks, such as natural language generation. *Rules.* The current model of NLU has yet to obtain a "real" understanding of natural language. Although there are already some basic rules for NLU, it is difficult to understand natural language and for machine learning to comprehend the inherent characteristics only based on these rules given that the same meaning can be expressed in many ways. It's important to note here that human language does not always follow the rules that govern it. It is possible that omitted words and substandard grammar might not interfere with communication since humans can speculate the meaning of sentences. For computers,

however, these are obstacles. *Metaphor*. Through metaphors, multiple unrelated concepts can be combined to generate new concepts. For example, "this room is like an oven refers to the high temperature of a room, but a computer cannot understand the meaning of these words using word representation and syntactic analysis. *Creativity*. There are a lot of new words and expressions emerging on the Internet every day. Some are created based on important events or particular scenarios, which enhances the difficulty of discovering and equipping computers to understand these words. Given that human language was created by thousands of people over a long period of time, Gouws, Van Rooyen, and Engelbrecht (2010) believe that researchers can use information from structured data sources and knowledge bases. The authors claim that human beings learn language by applying their surroundings to the object through experience and interaction. However, the computation is enormous in this situation. Finally, because neuroscience and cognitive science investigates the underlying mechanism of how human brain processes language, knowledge in neuroscience and cognitive science can be used to build models.

6.2.2. *Construction of large-scale corpora*

Mainstream NLP tasks are based on statistical machine learning, which requires a large-scale corpus. In many missions, these corpora need to be built manually, which is a very laborious task. Data sharing is thus an indispensable practice for promoting research and development. Moreover, any corpus, regardless of size or type, cannot comprehensively cover all cases in a given field. There is also no uniform standard for corpus' labeling systems. If the category is too broad, it is impossible to describe the language comprehensively and meticulously. If it is too narrow, the efficiency of labeling will be reduced and result in a serious problem of sparsity and poor robustness of the trained model. As manually annotating a corpus is a time-consuming and laborious task, it is necessary to study how to use large scale data without manual or partially labeled data. Semi-supervised learning is particularly immature for use in this regard.

6.3. *Conclusion*

NLP is an important branch of artificial intelligence that adopts conventional machine learning algorithms, unique NLP methods, and deep learning to build models based on a variety of research backgrounds to extract document information and understand documents. With the development of big data, cloud computing, and deep learning, NLP is widely used in various disciplines, including management research. NLP facilitates people's life and effectively improve efficiency and accuracy of work.

Recently, NLP has experienced explosive growth in management studies. With the continuous improvement and breakthrough of algorithms and peculiar datasets, NLP is providing scholars with more accurate analytical methods for textual data. The measurement of concepts and the assembly of methods for specific problem-solving in the business world have attracted a great deal of attention as a result. In the future, NLP will undoubtedly assist researchers to make theoretical and practical contributions to management literature. We hope this review offers a roadmap for how NLP can be applied as an analytical technique for management research to generate theoretical and practical insights.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Notes

1. https://sg.news.yahoo.com/robo-journalism-gains-traction-shifting-media-landscape-014451181.html
2. https://www.aclweb.org/anthology/P19-1289
3. https://www.iflyrec.com/
4. https://patents.google.com/patent/US8285619B2/en
5. https://damo.alibaba.com/air/1235
6. Helpfulness star scientists benefit others through influencing the formation and quality of new ideas by discussion, feedback, encouragement, and criticism with peers.

ORCID

Zhao Cai http://orcid.org/0000-0001-8705-4164
Chee-Wee Tan http://orcid.org/0000-0003-4245-541X
Hefu Liu http://orcid.org/0000-0003-3854-7821

References

Abbasi, A., Zhou, Y., Deng, S., & Zhang, P. (2018). Text analytics to support sense-making in social media: A language-action perspective. *MIS Quarterly*, *42*(2), 427–438.
Adamopoulos, P., Ghose, A., & Todri, V. (2018). The impact of user personality traits on word of mouth: Text-mining social media platforms. *Information Systems Research*, *29*(3), 612–640.
Angus, R. W. (2019). Problemistic search distance and entrepreneurial performance. *Strategic Management Journal*, *40*(12), 2011–2023.
Ansari, A., Li, Y., & Zhang, J. Z. (2018). Probabilistic topic model for hybrid recommender systems: A stochastic variational Bayesian approach. *Marketing Science*, *37*(6), 987–1008.
Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, *57*(8), 1485–1509.
Bansal, N., Sharma, A., & Singh, R. K. (2019). Fuzzy AHP approach for legal judgement summarization. *Journal of Management Analytics*, *6*(3), 323–340. doi:10.1080/23270012.2019.1655672
Bao, Y., & Datta, A. (2014). Simultaneously discovering and quantifying risk types from textual risk disclosures. *Management Science*, *60*(6), 1371–1391.
Barlow, M. A., Verhaal, J. C., & Angus, R. W. (2019). Optimal distinctiveness, strategic categorization, and product market entry on the Google Play app platform. *Strategic Management Journal*, *40*(8), 1219–1242.
Barr, A., & Feigenbaum, E. A. (2014). *The handbook of artificial intelligence*. Butterworth-Heinemann.
Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2019). Uniting the tribes: Using text for marketing insight. *Journal of Marketing*, *84*(1), 1–25.
Berman, R., Melumad, S., Humphrey, C., & Meyer, R. (2019). A tale of two Twitterspheres: Political microblogging during and after the 2016 primary and presidential debates. *Journal of Marketing Research*, *56*(6), 895–917.
Caldieraro, F., Zhang, J. Z., Cunha, M., & Shulman, J. D. (2018). Strategic information transmission in peer-to-peer lending markets. *Journal of Marketing*, *82*(2), 42–63.

Califf, M. E., & Mooney, R. J. (2003). Bottom-up relational learning of pattern matching rules for information extraction. *Journal of Machine Learning Research*, *4*(Jun), 177–210.

Calomiris, C. W., & Mamaysky, H. (2019). How news and its context drive risk and returns around the world. *Journal of Financial Economics*, *133*(2), 299–336.

Chanda, U., & Goyal, P. (2019). A Bayesian network model on the interlinkage between Socially Responsible HRM, employee satisfaction, employee commitment and organizational performance. *Journal of Management Analytics*, *7*(1), 105–138. doi:10.1080/23270012.2019.1650670

Chatterji, A., Delecourt, S., Hasan, S., & Koning, R. (2019). When does advice impact startup performance? *Strategic Management Journal*, *40*(3), 331–356.

Chi-Hsien, K., & Nagasawa, S. (2019). Applying machine learning to market analysis: Knowing your luxury consumer. *Journal of Management Analytics*, *6*(4), 404–419. doi:10.1080/23270012.2019.1692254

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, *12*(Aug), 2493–2537.

Councill, I. G., Giles, C. L., Han, H., & Manavoglu, E. (2005). *Automatic acknowledgement indexing: Expanding the semantics of contribution in the CiteSeer digital library.* Proceedings of the 3rd international conference on Knowledge capture.

Croidieu, G., & Kim, P. H. (2018). Labor of love: Amateurs and lay-expertise legitimation in the early US radio field. *Administrative Science Quarterly*, *63*(1), 1–42.

Cui, R., Gallino, S., Moreno, A., & Zhang, D. J. (2018). The operational value of social media information. *Production and Operations Management*, *27*(10), 1749–1769.

Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science*, *53*(9), 1375–1388.

Dotzel, T., & Shankar, V. (2019). The relative effects of business-to-business (vs. Business-to-consumer) service innovations on firm value and firm risk: An empirical analysis. *Journal of Marketing*, *83*(5), 133–152.

Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The evolution of 10-K textual disclosure: Evidence from latent Dirichlet allocation. *Journal of Accounting and Economics*, *64*(2–3), 221–245.

Fang, F., Dutta, K., & Datta, A. (2014). Domain adaptation for sentiment classification in light of multiple sources. *INFORMS Journal on Computing*, *26*(3), 586–598.

Frankel, R., Jennings, J., & Lee, J. (2016). Using unstructured and qualitative disclosures to explain accruals. *Journal of Accounting and Economics*, *62*(2–3), 209–227.

Froot, K., Kang, N., Ozik, G., & Sadka, R. (2017). What do measures of real-time corporate sales say about earnings surprises and post-announcement returns? *Journal of Financial Economics*, *125*(1), 143–162.

Gabel, S., Guhl, D., & Klapper, D. (2019). P2V-MAP: Mapping market structures for large retail assortments. *Journal of Marketing Research*, *56*(4), 557–580.

Genest, P.-E., & Lapalme, G. (2011). *Framework for abstractive summarization using text-to-text generation.* Proceedings of the Workshop on Monolingual Text-To-Text Generation.

Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science*, *31*(3), 493–520.

Ghose, A., Ipeirotis, P. G., & Li, B. (2019). Modeling consumer footprints on search engines: An interplay with social media. *Management Science*, *65*(3), 1363–1385.

Gong, J., Abhisek, V., & Li, B. (2018). Examining the impact of keyword ambiguity on search advertising performance: A topic model approach. *MIS Quarterly*, *42*(3), 805–829.

Gouws, S., Van Rooyen, G., & Engelbrecht, H. A. (2010). *Measuring conceptual similarity by spreading activation over Wikipedia's hyperlink structure.* Proceedings of the 2nd Workshop on The People's Web Meets NLP: Collaboratively Constructed Semantic Resources.

Green, G. M. (2012). *Pragmatics and natural language understanding.* Routledge.

Hendershott, T., Livdan, D., & Schürhoff, N. (2015). Are institutions informed about news? *Journal of Financial Economics*, *117*(2), 249–287.

Herhausen, D., Ludwig, S., Grewal, D., Wulf, J., & Schoegel, M. (2019). Detecting, preventing, and mitigating online firestorms in brand communities. *Journal of Marketing*, *83*(3), 1–21.

Hermosilla, M., Gutiérrez-Navratil, F., & Prieto-Rodríguez, J. (2018). Can emerging markets tilt global product design? Impacts of Chinese colorism on Hollywood castings. *Marketing Science*, *37*(3), 356–381.

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, *349* (6245), 261–266.

Hoberg, G., & Phillips, G. (2018). Conglomerate industry choice and product language. *Management Science*, *64*(8), 3735–3755.

Hong, Y., & Pavlou, P. A. (2017). On buyer selection of service providers in online outsourcing platforms for IT services. *Information Systems Research*, *28*(3), 547–562.

Hou, J., Zhao, H., Zhao, X., & Zhang, J. (2016). Predicting mobile users' behaviors and locations using dynamic Bayesian networks. *Journal of Management Analytics*, *3*(3), 191–205. doi:10.1080/23270012.2016.1198242

Hu, M., Dang, C., & Chintagunta, P. K. (2019). Search and learning at a daily deals website. *Marketing Science*, *38*(4), 609–642.

Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. (2018). Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science*, *64*(6), 2833–2855.

Hwang, E. H., Singh, P. V., & Argote, L. (2019). Jack of all, master of some: Information network and innovation in crowdsourcing communities. *Information Systems Research*, *30* (2), 389–410.

Jain, N., Girotra, K., & Netessine, S. (2014). Managing global sourcing: Inventory performance. *Management Science*, *60*(5), 1202–1222.

Johnson, S. L., Safadi, H., & Faraj, S. (2015). The emergence of online community leadership. *Information Systems Research*, *26*(1), 165–187.

Kaplan, S., & Vakili, K. (2015). The double-edged sword of recombination in breakthrough innovation. *Strategic Management Journal*, *36*(10), 1435–1457.

Ko, D. G., Mai, F., Shan, Z., & Zhang, D. (2019). Operational efficiency and patient-centered health care: A view from online physician reviews. *Journal of Operations Management*, *65*(4), 353–379.

Kumar, V., Bhaskaran, V., Mirchandani, R., & Shah, M. (2013). Practice prize winner—creating a measurable social media marketing strategy: Increasing the value and ROI of intangibles and tangibles for Hokey Pokey. *Marketing Science*, *32*(2), 194–212.

Law, K. S., & Chung, F.-L. (2020). Knowledge-driven decision analytics for commercial banking. *Journal of Management Analytics*, 1–22. doi:10.1080/23270012.2020.1734879

Lee, D., Hosanagar, K., & Nair, H. S. (2018). Advertising content and consumer engagement on social media: Evidence from Facebook. *Management Science*, *64*(11), 5105–5131.

Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research*, *48*(5), 881–894.

Li, X. B., & Qin, J. (2017). Anonymizing and sharing medical text records. *Information Systems Research*, *28*(2), 332–352.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167.

Liu, J., & Toubia, O. (2018). A semantic approach for estimating consumer content preferences from online search queries. *Marketing Science*, *37*(6), 930–952.

Liu, X., Lee, D., & Srinivasan, K. (2019). Large-scale cross-category analysis of consumer review content on sales conversion leveraging deep learning. *Journal of Marketing Research*, *56*(6), 918–943.

Liu, X., Singh, P. V., & Srinivasan, K. (2016). A structured analysis of unstructured big data by leveraging cloud computing. *Marketing Science*, *35*(3), 363–388.

Loughran, T. I. M., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, *54*(4), 1187–1230.

Loughran, T., & Mcdonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65.

Lu, Y. (2019). Artificial intelligence: A survey on evolution, models, applications and future trends. *Journal of Management Analytics*, *6*(1), 1–29. doi:10.1080/23270012.2019.1570365

Manning, C. D., Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT press.

McDonald, D. D. (2010). Natural language generation. *Handbook of Natural Language Processing*, *2*, 121–144.

Melumad, S., Inman, J. J., & Pham, M. T. (2019). Selectively emotional: How smartphone use changes user-generated content. *Journal of Marketing Research*, *56*(2), 259–275.

Nguyen, H., Calantone, R., & Krishnan, R. (2019). Influence of social media emotional word of mouth on institutional investors' decisions and firm value. *Management Science*, *66*(2), 887–910.

Och, F. J. (2003). *Minimum error rate training in statistical machine translation*. Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1.

Oettl, A. (2012). Reconceptualizing stars: Scientist helpfulness and peer performance. *Management Science*, *58*(6), 1122–1140.

Ordenes, F. V., Ludwig, S., De Ruyter, K., Grewal, D., & Wetzels, M. (2017). Unveiling what is written in the stars: Analyzing explicit, implicit and discourse patterns of sentiment in social media. *Journal of Consumer Research*, *43*(6), 875–894.

Pan, Y., Huang, P., & Gopal, A. (2019). Storm clouds on the horizon? *New Entry Threats and R&D Investments in the U.S. IT Industry. Information Systems Research*, *30*(2), 540–562.

Ricks, D. A., Toyne, B., & Martinez, Z. (1990). Recent developments in international management research. *Journal of Management*, *16*(2), 219–253.

Rohrbach, M., Qiu, W., Titov, I., Thater, S., Pinkal, M., & Schiele, B. (2013). *Translating video content to natural language descriptions*. Proceedings of the IEEE International Conference on Computer Vision.

Rutz, O. J., Sonnier, G. P., & Trusov, M. (2017). A new method to aid copy testing of paid search text advertisements. *Journal of Marketing Research*, *54*(6), 885–900.

Rutz, O. J., Trusov, M., & Bucklin, R. E. (2011). Modeling indirect effects of paid search advertising: Which keywords lead to more future visits? *Marketing Science*, *30*(4), 646–665.

Schank, R. C. (1972). Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology*, *3*(4), 552–631.

Song, T., Huang, J., Tan, Y., & Yu, Y. (2019). Using user- and marketer-generated content for box office revenue prediction: Differences between microblogging and third-party platforms. *Information Systems Research*, *30*(1), 191–203.

Srivastava, S. B., Goldberg, A., Manian, V. G., & Potts, C. (2018). Enculturation trajectories: Language, cultural adaptation, and individual outcomes in organizations. *Management Science*, *64*(3), 1348–1364.

Stephen, A. T., Zubcsek, P. P., & Goldenberg, J. (2016). Lower connectivity is better: The effects of network structure on redundancy of ideas and customer innovativeness in interdependent ideation tasks. *Journal of Marketing Research*, *53*(2), 263–279.

Tang, V. W. (2018). Wisdom of crowds: Cross-sectional variation in the informativeness of third-party-generated product information on Twitter. *Journal of Accounting Research*, *56*(3), 989–1034.

Toubia, O., Iyengar, G., Bunnell, R., & Lemaire, A. (2018). Extracting features of entertainment products: A guided latent Dirichlet allocation approach informed by the psychology of media consumption. *Journal of Marketing Research*, *56*(1), 18–36.

Trusov, M., Ma, L., & Jamal, Z. (2016). Crumbs of the cookie: User profiling in customer-base analysis and behavioral targeting. *Marketing Science*, *35*(3), 405–426.

Tuertscher, P., Garud, R., & Kumaraswamy, A. (2014). Justification and interlaced knowledge at ATLAS, CERN. *Organization Science*, *25*(6), 1579–1608.

Verma, N., Malhotra, D., & Singh, J. (2020). Big data analytics for retail industry using MapReduce-Apriori framework. *Journal of Management Analytics*, 1–19. doi:10.1080/23270012.2020.1728403

Villarroel Ordenes, F., Grewal, D., Ludwig, S., Ruyter, K. D., Mahr, D., Wetzels, M., … Kopalle, P. (2019). Cutting through content clutter: How speech and image acts drive consumer sharing of social media brand messages. *Journal of Consumer Research*, *45*(5), 988–1012.

Wang, Q., Li, B., & Singh, P. V. (2018). Copycats vs. Original mobile apps: A machine learning copycat-detection method and empirical analysis. *Information Systems Research*, *29*(2), 273–291.

Weber, D. (2002). *Object interactive user interface using speech recognition and natural language processing*. Google Patents. US patent number 6434524.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*(2), 13–23.

Wu, L., Hitt, L., & Lou, B. (2019). Data analytics, innovation, and firm productivity. *Management Science*.

Wu, L., Lou, B., & Hitt, L. (2019). Data analytics supports decentralized innovation. *Management Science*, *65*(10), 4863–4877.

Xiong, G., & Bharadwaj, S. (2014). Prerelease buzz evolution patterns and new product performance. *Marketing Science*, *33*(3), 401–421.

Xu, B., Xu, K., Fu, L., Li, L., Xin, W., & Cai, H. (2016). Healthcare data analytics: Using a metadata annotation approach for integrating electronic hospital records. *Journal of Management Analytics*, *3*(2), 136–151. doi:10.1080/23270012.2016.1141331

Xu, K., Liao, S. S., Lau, R. Y. K., & Zhao, J. L. (2014). Effective active learning strategies for the use of large-margin classifiers in semantic annotation: An optimal parameter discovery perspective. *INFORMS Journal on Computing*, *26*(3), 461–483.

Xu, T., Zhang, P., Huang, Q., Zhang, H., Gan, Z., Huang, X., & He, X. (2018). *Attngan: Fine-grained text to image generation with attentional generative adversarial networks*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.