

Retraction

Retracted: Text Mining in Management Research: A Bibliometric Analysis

Security and Communication Networks

Received 26 December 2023; Accepted 26 December 2023; Published 29 December 2023

Copyright © 2023 Security and Communication Networks. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] G. Song, J. Wu, and S. Wang, "Text Mining in Management Research: A Bibliometric Analysis," *Security and Communication Networks*, vol. 2021, Article ID 2270276, 15 pages, 2021.

Research Article

Text Mining in Management Research: A Bibliometric Analysis

Guandong Song,¹ Jiying Wu ,¹ and Sihui Wang²

¹*School of Humanities and Law, Northeastern University, Shenyang 110167, Liaoning, China*

²*School of Music, Soochow University, Suzhou 215127, China*

Correspondence should be addressed to Jiying Wu; lilylove10@21cn.com

Received 22 October 2021; Revised 5 November 2021; Accepted 13 November 2021; Published 26 November 2021

Academic Editor: Jian Su

Copyright © 2021 Guandong Song et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The goal of this paper is to provide a bibliometric analysis of scientific publications that employ text mining in management. To accomplish this, the authors collected 1282 documents from the Web of Science and performed performance analysis and science mapping with the help of the Bibliometrix package in Rstudio. The performance analysis used a range of bibliometric indicators such as productivity, citations, *h*-index, and *m*-quotient, in order to identify research trends and the most influential journals, authors, countries, and literature in the study. Science mapping used author keywords co-occurrence, co-authorship, and co-citation analysis to reflect the conceptual, social, and intellectual structure of the research. Specifically, we have seen an exponential increase in the use of text mining in management in recent years. The United States is the dominant country for research, having the earliest studies and the highest number of literature and citations. Furthermore, the research themes showed that topic modeling is at the forefront of current text mining research about management. This study will help scholars and management practitioners interested in the intersection of text mining and management to quickly understand the latest advances in research.

1. Introduction

Data types can be structured, semi-structured, or even heterogeneous. The method of discovering knowledge can be mathematical, nonmathematical, or inductive. The discovered knowledge can be used for information management, query optimization, decision support, and data maintenance. Text mining is a knowledge-intensive process in which users interact with a set of documents by using a range of analysis tools to identify and explore the patterns of interest [1]. In contrast to generalized data mining, which deals with structured data, text mining focuses on the analysis and modeling of unstructured natural language text, such as online news, scientific research papers, and medical documents. Therefore, it is a comprehensive technology that exploits natural language processing, pattern classification, machine learning, statistics, and other techniques [2, 3].

The development of text mining began with the need to catalog text documents [4]. In 1958, Luhn incorporated the idea of word frequency statistics into document summarization. He then implemented it automatically on a

computer, setting a precedent for text mining [5]. In 1961, Doyle drew on Luhn's work to propose a new method for classifying library information in the form of word frequencies and associations [6]. With the development of information science and natural language processing, text mining activities have been extended from the early days of information retrieval and information summarization to information extraction. Until the 1990s, it was combined with many newly developed techniques to produce a wider range of analytical tasks, including document classification, clustering, document meaning extraction, association mining, trend analysis, and machine translation [7], which provides a variety of methods for discovering knowledge and patterns from massive amounts of unstructured text data. The steps of text mining process mainly include defining problems, establishing text mining database, text pre-processing, feature extraction and feature selection, and mining with algorithms.

With the advent of the era of big data, the information data in the network are growing explosively, which on the one hand enhances the efficiency of information generation

and delivery, and on the other hand, brings information overload [8], making it a very challenging task to extract high-quality information from the massive data [9]. Unstructured text is the most common format of web data [10], the knowledge and patterns implicit in it are valuable resources for management practices, for which many studies addressing the intersection of management and text mining have been generated. Obtaining the theoretical framework of these literatures can improve the understanding of scholars and management practitioners on the integration of text mining techniques in management operations. The purpose of analysis is to find the data fields that have the greatest impact on the forecast output and decide whether to define export fields. If the dataset contains hundreds of fields, browsing and analyzing these data will be a very time-consuming and tiring task. At this time, you need to choose a powerful method to help you complete these things.

Bibliometrics is a common approach to synthesize research results, which uses mathematical and statistical methods to systematically analyze books or other communication media in a given field [11], and is commonly used to uncover actors (researchers, institutions, countries), literature, research topics, and research trends in a research field. Currently, there are limited bibliometric studies in the academic community on the application of text mining techniques in different subject areas, and they are mainly focused on the field of biomedicine [12–14], and to our knowledge, there are no bibliometric studies on the application of text mining in management. For this reason, this paper collects relevant literature from the WOS database and performs bibliometrics to address the following research questions:

RQ1. What are the evolutionary trends in the application of text mining in management?

RQ2. Which literature, journals, authors, and countries are the most influential in the research of applying text mining in management?

RQ3. What are the structure characteristics of text mining literature about management?

2. Methods

2.1. Data Source and Search Strategy. The imported dataset for this study was extracted from the Web of Science (WOS), which covers 90 million documents from 15,000 journals [15], and the material covered is considered to have the highest quality standards [16]. In addition, considering that existing studies have demonstrated that the simultaneous use of multiple related databases cannot increase the number of documents due to duplication of literature between databases [17], the WOS was used as the only data source in this study. WOS is not only used as a document retrieval tool but also as a basis for scientific research evaluation. The total number of papers included by WOS in scientific research institutions reflects the scientific research level of the whole institution, especially the basic research level. The number of papers collected by WOS and the number of citations reflect its research ability and academic level.

In order to make the search terms more comprehensive, we first tried to use the keyword “text mining” as the topic for literature retrieval in WOS Core Collection (including SCIE, SSCI, CPCI-S, and CPCI-SSH). Furthermore, we identified 11 different keywords similar or related to text mining, which are commonly used in the field of management. Finally, in July 2021, we searched the WOS Core Collection database for articles with “text mining,” “sentiment analysis,” “natural language processing,” “text analytics,” “topic modeling,” “semantic network analysis,” “latent semantic analysis,” “latent Dirichlet allocation,” “document clustering,” “semantic relations,” “TF-IDF,” “lexical analysis” included in the title, abstract, author keywords, and keywords plus. 49209 articles in the field of text mining were obtained. The documents are preliminarily screened from the publication time (before December 30, 2020) and language (English), and then the documents are rescreened from the document types (articles, reviews, letters). Finally, 1282 articles in the field of management are retained by limiting the WOS categories to “operations research management science,” “management,” and “public administration.” Figure 1 shows the article extraction process.

2.2. Bibliometric Analysis. The eligible articles were analyzed by bibliometric analysis. In this regard, two methods were primarily used: performance analysis and science mapping [18].

Performance analysis is used to assess the citation impact of scientific results produced by different actors interacting in a research field; these actors include researchers as well as journals, countries, institutions, and departments. The traditional indicators of performance analysis are the number of articles and the number of citations, where the number of articles is used to characterize the quantity of research and the number of citations to characterize the quality of research. Hirsch combined these two indicators into a single- h -index, which refers to the number of papers that are cited more than h times [19]. h -index is widely used to evaluate the scientific output of individual researchers [20], research groups [21], research facilities [22], and countries [23] because of its many advantages. In his original proposal, Hirsch pointed out that the h -index, which considers a researcher’s research and can be easily calculated by ranking papers according to the “number of citations” in the scientific database of the Thomson ISI website [19]. Costas and Bordons point out that the h -index is an objective indicator that can play an important role in evaluating the performance of researchers [24], and Vanclay argues that an important advantage of the h -index is its robustness, as it is insensitive to a set of less cited papers [25]. Admittedly, the h -index also has many shortcomings, the most important of which is that it cannot compare researchers from different fields and at different career stages [19, 26]. In this study, only scholars who use text

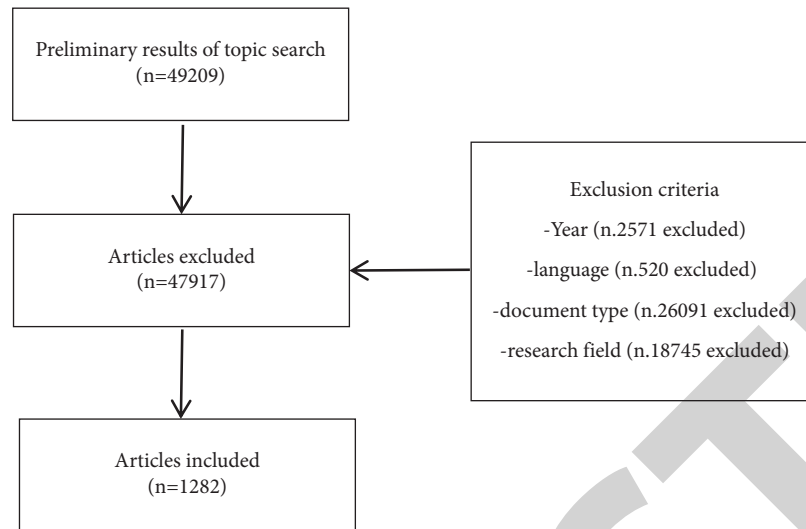


FIGURE 1: The flow chart for articles extraction on text mining in the field of management.

mining techniques to solve problems in the field of management are studied. In comparing scholars with different seniority, we further calculated another performance analysis metric, m -quotient [19], which can be defined as the ratio of h -index to the number of years since the researcher's first publication. In addition, to obtain more comprehensive performance analysis results, we also considered some variants of the above indicators, such as citation thresholds (≥ 500 , ≥ 300 , ≥ 200 , ≥ 100 , ≥ 50 , ≥ 10) and developmental stage dimensions of publications and citations (TP1, TP2, TP3, TC1, TC2, TC3), which we define after each table. Performance analysis is the first step in bibliometrics and is an important part of the overall performance improvement system. The purpose of performance analysis is to identify and measure the gap between expected performance and current performance. Without clarifying the problem and the performance gap, it is not possible to identify the cause and design or select a solution.

Science mapping is another bibliometric analysis method, which is mainly used to reveal the conceptual, social, and intellectual structure of scientific research, as well as aspects of their dynamic evolution [27]. Keyword co-occurrence [28], co-authorship [29], and co-citation analysis [30] techniques were used to scientifically map the articles on text mining in management. Among them, keyword co-occurrence uses the authors' keywords as the unit of analysis to study the conceptual structure of scientific research. Co-authorship was performed on the 100 influential authors to highlight the social structure. And, the co-citation results were presented by a historiograph [31], which depicts the citation evolution of the 20 most influential documents, and aims to obtain the intellectual structure of the research.

To perform performance analysis and scientific mapping, we used the Bibliometrix Rstudio package, a unique open source tool for quantitative research in scientometrics and bibliometrics, designed by Aria and Cuccurullo [32]. The choice of bibliometrix was based on a comparative study of software tools for conducting bibliometric analysis, which

concluded that bibliometrix "contains the more extensive set of technologies implemented, and together with the ease of its interface" [33].

3. Results

3.1. Performance Analysis. In this section, we conduct a performance analysis of articles using the bibliometric analysis indicators introduced earlier, such as number of publications, number of citations, h -index, m -quotient, and some variable metrics based on the number of publications and citations, to answer RQ1 and RQ2, which are the trends and main actors in the application of text mining in management.

3.1.1. Evolution of Articles. A total of 1282 articles that met the research criteria were published from 1995 to 2020. Figure 2 shows the annual distribution of these articles and the fitted curves constructed by the logistic regression model. Regression model is a predictive modeling technology, which studies the relationship between dependent variables and independent variables. This technique is usually used for predictive analysis, time series models, and discovering causal relationships between variables. From two articles (1.56%) in 1995 to 266 articles (20.75%) in 2020, the use of text mining in management shows an exponential growth ($R^2 = 0.9689$). Based on the number of articles published per year, the research development can be divided into three phases. The first phase was from 1995 to 2005 and can be called the initial phase of the research, in which less than 10 articles were published per year, with an average of 3 articles per year. The second phase is 2006–2016, which can be called the development phase of the study, in which the number of articles published per year ranged from 12 to 87, with an average of 47 articles per year. The third phase is 2017–2020, which can be called the expansion phase of the study, in which the number of published articles per year ranges between 128 and 266, with an average of 182 articles per year.

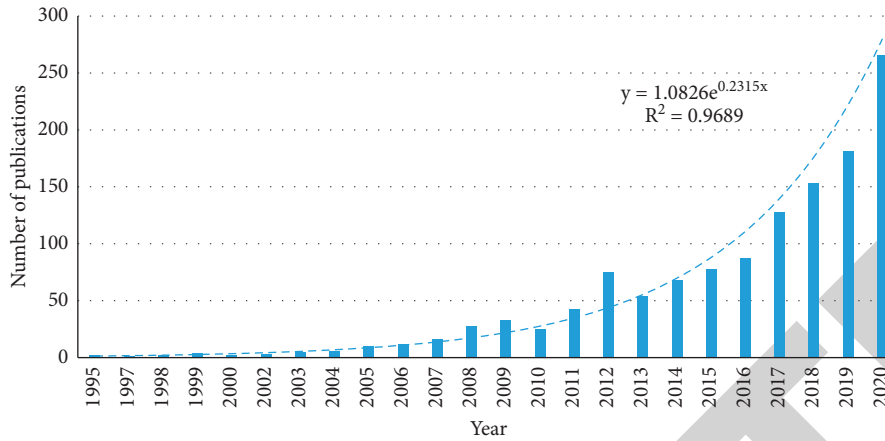


FIGURE 2: Distribution of articles over time on text mining in management.

In the remaining analysis of this study, we will use the three phases of productivity division as the temporal dimension to observe the development of the research actors and the research themes over time. We use T_1 , T_2 , and T_3 to denote the initial phase (1995–2005), the development phase (1996–2016), and the expansion phase (2017–2020), respectively.

3.1.2. Analysis of Important Journals. 1282 articles were published in 182 journals in total. Table 1 shows the top 20 journals in terms of the number of publications. The ranking of journals in the table is based on the total number of publications (TP). When the total number of publications is the same, the total number of citations (TC) is used to rank journals.

As can be seen from Table 1, the 20 journals published 67.60% of the articles in the field. Two journals of operations research and management science, Applied Expert Systems (ESA) and Decision Support Systems (DSS), published the most articles, with 427 and 140 articles, respectively, accounting for 33.31% and 10.92% of the total number of articles. Only four journals published more than 2% of the number of articles. In addition to these two journals, Tourism Management (TMG) and Journal of Management Information systems (JMIS) accounted for 2.26% and 2.03% of the total number of publications, respectively.

Another important indicator for analysis is the number of citations of the journals. Citation statistics is a basic method of bibliometrics, which refers to the counting of literature in the same discipline cited in published papers. The citations and references at the end of the paper are representative of the literature most needed by the user. The citations basically reflect the main characteristics of the information obtained by the user using formal channels. The top five most cited journals are: Applied Expert Systems (ESA), Decision Support Systems (DSS), Tourism Management (TMG), Information Systems Research (ISR), and Miss Quarterly (MISQ). Among them, Expert Application Systems (ESA) and Decision Support Systems (DSS) were the most cited. The former was cited 12,731 times, about twice as many as the latter (6,224 times). It should be noted that although Information Systems Research (ISR) and Miss Quarterly (MISQ) have average publication levels, they rank

high in terms of citations, which may be related to the fact that they have published some important papers in the field. An analysis of the distribution of journal citations confirms this conclusion, with Information Systems Research (ISR) and Expert Systems with Applications (ESA) being the only two journals to have published a paper with more than 500 citations. Miss Quarterly (MISQ) is one of the few journals to have published a paper with more than 200 citations.

The results of the h -index show that when combined with the number and impact of published articles, Expert Systems and Applications (ESA), Decision Support Systems (DSS), Tourism Management (TMG), Journal of Management Information Systems (JMIS), and Miss Quarterly (MISQ) are the most valuable journals in the field of text mining for management.

The analysis of the journals' development over time (Table 2) also yielded some interesting results. The earliest journals to publish articles on the application of text mining in management were Applied Expert Systems (ESA), Decision Support Systems (DSS), Management Information Systems (JMIS), and International Journal of Information Technology Decision Making (IJITDM). More than half (55%) of the journals started publishing relevant articles in 2006–2016, and three of them are noteworthy. One is Tourism Management (TMG), which published two articles with 475 citations. The second is Information Systems Research (ISR) with six published articles and 883 citations. Third, the International Journal of Contemporary Hospitality Management (IJCHM) published only one article with 59 citations. 2017–2020, Tourism Management Perspectives (TMP), IEEE Systems Journal (ISJ), and International Journal of Forecasting (IJF) started to publish relevant articles. The calculations of journal m -quotient validate the above results, that is, in addition to ESA (Expert Systems with Applications) and DSS (Decision Support Systems), which are two stable performers, some journals that joined the research later also performed well.

3.1.3. Analysis of Important Articles. Over the decades, many influential management articles applying text mining have been published. One way to find these influential

TABLE 1: Ranking of the most productive journals on text mining in management.

R	Journal	TP	P%	TC	<i>h</i>	≥10	≥50	≥100	≥200	≥300	≥500
1	ESA	427	33.31	12731	57	268	65	26	10	1	1
2	DSS	140	10.92	6224	46	115	71	17	4		
3	TMG	29	2.26	1819	18	20	8	5	3	1	
4	JMIS	26	2.03	540	16	19	1				
5	IMG	24	1.87	492	12	24	19	2			
6	ECR	23	1.79	171	8	7					
7	IJITDM	21	1.64	311	4	3	1	1	1		
8	ISR	20	1.56	979	10	10	5	2	1	1	1
9	ISEBM	20	1.56	174	8	6					
10	MISQ	18	1.4	656	13	13	3	1	1		
11	CEUS	18	1.4	408	11	11	2				
12	TASM	18	1.4	179	8	6					
13	IJCHM	17	1.33	376	10	11	3				
14	IJPR	13	1.01	182	8	7	1				
15	SSC	11	0.86	69	6	2					
16	TMP	9	0.7	123	6	4					
17	ISJ	9	0.7	53	4	1					
18	JKM	8	0.62	98	4	8	8	2			
19	AOR	8	0.62	73	3	2					
20	IJF	8	0.62	54	5	3					

R, rank; TP, total publications; P%, percentage of articles in the journal; TC, total citations; *h*, *h*-index of the journal; ≥10, ≥50, ≥100, ≥200, ≥300, ≥500, number of articles with more than 10, 50, 100, 200, 300, and 500 citations; ESA, Expert Systems with Applications; DSS, Decision Support Systems; TMG, Tourism Management; JMIS, Journal of Management Information Systems; IMG, Information Management; ECR, Electronic Commerce Research; IJITDM, International Journal of Information Technology Decision-Making; ISR, Information Systems Research; ISEBM, Information Systems and E-Business Management; MISQ, Miss Quarterly; CEUS, Computers Environment and Urban Systems; TASM, Technology Analysis Strategic Management; IJCHM, International Journal of Contemporary Hospitality Management; IJPR, International Journal of Production Research; SSC, Safety Science; TMP, Tourism Management Perspectives; ISJ, IEEE Systems Journal; JKM, Journal of Knowledge Management; AOR, Annals of Operations Research; IJF, International Journal of Forecasting.

TABLE 2: Evolution of the most productive journals on text mining in management.

R	Journal	<i>m</i>	Year	TP1	TP2	TP3	TC1	TC2	TC3	TC1/TP1	TC2/TP2	TC3/TP3
1	ESA	2.375	1998	12	238	177	211	9592	2928	17.58	40.30	16.54
2	DSS	2.000	1999	7	82	51	465	4763	996	66.43	58.09	19.53
3	IJCHM	1.667	2016		1	16		59	317		59.00	19.81
4	IJF	1.250	2018			8			54			6.75
5	TMG	1.200	2007		2	27		475	1344		237.50	49.78
6	TMP	1.200	2017			9			123			13.67
7	ISR	1.111	2013		6	14		883	96		147.17	6.86
8	MISQ	0.867	2007		9	9		560	96		62.22	10.67
9	IMG	0.857	2008		7	17		287	205		41.00	12.06
10	ECR	0.800	2012		1	22		7	164		7.00	7.45
11	ISJ	0.800	2017		0	9		0	53			5.89
12	JMIS	0.696	1999	4	12	10	98	266	176	24.50	22.17	17.60
13	CEUS	0.688	2006		6	12		260	148		43.33	12.33
14	JKM	0.667	2016		2	6		10	88		5.00	14.67
15	ISEBM	0.571	2008		9	11		128	46		14.22	4.18
16	TASM	0.500	2006		10	8		161	18		16.10	2.25
17	SSC	0.429	2008		3	8		37	32		12.33	4.00
18	IJPR	0.348	1999	1	2	10	11	67	104	11.00	33.50	10.40
19	IJITDM	0.235	2005	3	7	11	12	274	25	4.00	39.14	2.27
20	AOR	0.111	1995	1	2	5	1	31	41	1.00	15.50	8.20

R, rank; *m*, *m* quotient; Year, the publishing year of the first relevant article; TP1, TP2, TP3, total publications of T1, T2, T3; TC1, TC2, TC3, total citations of T1, T2, T3; TC1/TP1, TC2/TP2, TC3/TP3, average number of citations per article; abbreviation of journal names are found in Table 1.

articles is to analyze the number of citations received [16]. Table 3 shows the 20 most frequently cited articles, and the articles in the table are ranked based on the number of citations received.

According to Table 3, there are two articles with a total of more than 500 threshold citations. Among them, Romero and Ventura’s publication was the most cited with a total of 595 citations. This paper reviews the application of text

TABLE 3: The most cited articles on text mining in management.

R	Author (s)	Title	Year	Journal	TC	TC/Y
1	Romero and Ventura	Educational data mining: a survey from 1995 to 2005	2007	ESA	595	42.50
2	Goh et al.	Social media brand community and consumer behavior: quantifying the relative impact of user- and marketer-generated content	2013	ISR	513	64.13
3	Choi et al.	Destination image representation on the web: content analysis of Macau travel-related websites	2007	TMG	391	27.93
4	Archak et al.	Deriving the pricing power of product features by mining consumer reviews	2011	MSC	374	37.40
5	Cao et al.	Exploring determinants of voting for helping online user reviews: a text mining approach	2011	DSS	286	28.60
6	Moraes et al.	Document-level sentiment classification: an empirical comparison between SVM and ANN	2013	ESA	265	33.13
7	Mostafa et al.	More than words: social networks' text mining for consumer brand sentiments	2013	ESA	259	32.38
8	Xiang et al.	A comparative analysis of major online review platforms: implications for social media analytics in hospitality and tourism	2017	TMG	257	64.25
9	Randhawa et al.	A bibliometric review of open innovation: setting a research agenda	2016	JPIMG	255	51.00
10	Gerber et al.	Predicting crime using twitter and kernel density estimation	2014	DSS	252	36.00
11	Li et al.	Using text mining and sentiment analysis for online forums hotspot detection and forecast	2010	DSS	250	22.73
12	Guo et al.	Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent Dirichlet allocation	2017	TMG	248	62.00
13	Sidorova et al.	Uncovering the intellectual core of the information systems discipline	2008	MISQ	242	18.62
14	Lee et al.	An approach to discovering new technology opportunities: keyword-based patent map approach	2009	TECH	240	20.00
15	Bennasa et al.	Feature selection using joint mutual information maximisation	2015	ESA	237	39.50
16	Peng et al.	A descriptive framework for the field of data mining and knowledge discovery	2008	IJITDM	231	17.77
17	Zhang et al.	A comparative study of TF*IDF, LSI, and multi-words for text classification	2011	ESA	230	23.00
18	Nassirtoussi et al.	Text mining for market prediction: a systematic review	2014	ESA	224	32.00
19	Ghobakhloo et al.	The future of manufacturing industry: a Strategic road map toward industry 4.0	2018	JMTM	212	70.67
20	Ghiassi et al.	Twitter brand sentiment analysis: a hybrid system using n -gram analysis and dynamic artificial neural network	2013	ESA	210	26.25

R, rank; Year, the year of the article being published; Journal, the journal of the article published; TC, total citations of the article; TC/Y, the average number of citations for this article per year; abbreviation of journal names are found in Table 1, except for MSC, Management Science; JPIMG, Journal of Product Innovation Management; TECH, Technovation; JMTM, Journal of Manufacturing Technology Management.

mining in different types of web-based educational systems [34]. The articles published by Xiang et al. (2017), Goh et al. (2013), and Guo et al. (2017) are the three most frequently cited articles in terms of annual citations. All three articles are text mining studies on social media comments. Among them, the article published by Goh et al. (2013) is earlier and has the top two total citations. Therefore, it can serve as an important reference for the application of text mining in management.

3.1.4. Analysis of Important Authors. Compared with structured data in databases, text has limited or no structure at all, and the content of text is a natural language used by humans, making it difficult for computers to process its semantics. The peculiarities of textual data sources make it impossible to apply existing data mining techniques directly. For this reason, it is necessary to analyze the text and extract metadata representing its features. These features can be saved in a structured form as an intermediate representation of the document. It is aimed at scanning and extracting the desired facts from the text. As text mining arises and develops, more and more scholars apply it to the management field, promoting the smooth and rapid development of

related research. Based on the article data downloaded from the WOS, a total of 3247 authors participated in the research. Like journals, we used the total number of published articles as a benchmark to select the top 20 authors in the field (Table 4). Authors in the table are sorted by the total number of publications (TP) and total citations (TC) when the same.

As seen in Table 4, Park is the author with the highest number of published and cited articles, with 16 relevant articles and an h -index of 12. He proposed a keyword-based patent mapping approach that combines text mining with patent blank domain discovery to guide new technology creation activities, and received more than 240 citations [35]. Van den Poel, Chen, and Feuerriegel are the three authors behind Park. Chen was the first (1999) to participate in the research among the top 20 authors, and Feuerriegel is the latest (2016) to participate in the research among those authors. However, Feuerriegel has published nine papers in the five years of participation. Another author of note is Fan, who received 630 citations for nine articles, an average of 70 per article. He is the author with the highest number of citations per article. A similar author to him is Lee et al., who published nine papers with 568 citations at an average of 63.11 citations per article. The main reasons for the low number of citations in the papers published by other authors

TABLE 4: Ranking of the most productive authors on text mining in management.

R	Author	TP	P%	TC	TC/TP	h	m	Year
1	Park	16	1.25	794	49.63	12	0.800	2007
2	Abrahams	10	0.78	373	37.30	6	0.600	2012
3	Van Den Poel	10	0.78	309	30.90	9	0.643	2008
4	Fan	9	0.70	630	70.00	6	0.353	2005
5	Lee	9	0.70	568	63.11	8	0.571	2008
6	Yoon	9	0.70	534	59.33	7	0.467	2007
7	Chen	9	0.70	443	49.22	9	0.391	1999
8	Feuerriegel	9	0.70	267	29.67	9	1.500	2016
9	Lau	8	0.62	281	35.13	7	0.700	2012
10	Thorleuchter	8	0.62	195	24.38	7	0.583	2010
11	Wei	8	0.62	185	23.13	6	0.375	2006
12	Lee	8	0.62	178	22.25	7	0.389	2004
13	Rita	7	0.55	325	46.43	6	0.857	2015
14	Valencia-Garcia	7	0.55	207	29.57	7	0.389	2004
15	He	7	0.55	187	26.71	6	0.667	2013
16	Yang	7	0.55	115	16.43	6	0.333	2004
17	Cortez	6	0.47	360	60.00	5	0.714	2015
18	Bose	6	0.47	327	54.50	5	0.500	2012
19	Kim	6	0.47	306	51.00	6	0.545	2011
20	Lee	6	0.47	79	13.17	5	0.500	2012

R, rank; TP, total publications; P%, percentage of articles written by the author; TC, total citations; TC/TP, average number of citations per article by this author; h, h-index of the author; m, m quotient of the author; Year, the year the author published his first paper.

are: there is no large amount of reading literature, lack of understanding of the dynamics at home and abroad, and lack of innovation; in order to increase the number of papers, divide one paper into two or more; ignoring or neglecting references; unintentionally or intentionally not quoting the literature of domestic peers; citing second-hand references, I did not read the original text, so the accuracy of references is poor.

For a more detailed analysis of the authors, this paper plots the annual distribution of the number of publications and citations of the authors (Figure 3), showing the research trends of the 20 authors with the highest number of publications. The size of the circles in the graph indicates the number of papers published by the authors. The larger the circle, the higher the number of papers published by that author in the corresponding year. The color of the circle represents the number of times the author has been cited. The darker the color of the circle, the higher the number of citations received by that author in that year.

Prior to 2005, only Chen, Fan, Lee, Valencia-Garcia, Yang and Lee published articles, indicating that they were pioneers in introducing text mining into the management field. However, from the downloaded publications, the citations of these foundation articles are not high. From 2006 to 2016, the remaining 15 authors started publishing articles. In other words, in this period, all authors published articles. Park and Van den Poel performed best in terms of the number of published articles and Bose performed best in terms of the number of citations. From 2017 to 2020, the study enters an expansion period with an average of more than 100 articles published per year, but eight (40%) authors

did not publish. Of the remaining 12 authors, Feuerriegel and Rita published the most articles and made the fastest progress, as in the previous period they had only 2 and 1 publications, respectively.

3.1.5. Analysis of Important Countries. Science and technology innovation refers to innovation in the field of science and technology, including both new discoveries of natural science knowledge and technological innovation. In modern society, universities and scientific and technological research institutions are the basic disciplines for basic science and technology innovation, and enterprises are the basic disciplines for technological innovation in applied engineering technologies and processes. Since science and technology are the most important factors contributing to the advancement of knowledge and economic growth, countries are paying more and more attention to investment in scientific research [36]. The purpose of this section is to analyze the country distribution of articles. It should be emphasized that authors who publish multiple articles may publish in different countries, since authors are usually mobile. The analysis of authors' countries in this paper is based on the country to which the authors belonged at the time of publication [37]. Similar to the journal analysis and author analysis, the top 20 countries with the highest number of publications are counted in this paper (Table 5), and the ranking of countries in the table is based on the total number of published articles. When the countries have the same number of total publications, they are sorted by the number of total citations.

The USA was the most published and influential country with 291 papers and an h-index of 47. On the one hand, this may be related to the size of the country, the investments in R & D, the number of researchers, and the language facilities [38]. On the other hand, it may be due to the fact that research in text mining in management started in the USA, generating many highly cited articles that ended up receiving almost twice as many citations as the second-ranked country. The second country is China, with 208 publications and an h-index of 35. Although the number of publications and impact indicators are not as high as those of the USA, China's indicators are much higher than those of South Korea, which ranks third, and Spain, which ranks fourth. 45% of the countries in the table belong to Europe and 30% to Asia. Only two Latin American countries (Mexico and Brazil) perform average in terms of number of articles and impact. Singapore is another notable country. Despite the limited number of published articles, it has the highest average number of citations per article of all countries at 53.56.

The results of the analysis of the temporal distribution of national publications (Table 6) show that China is the most influential country considering the time of participation in the study with an m-quotient of 2.33, followed by Korea and the United States with m-quotients of 2.14 and 1.81, respectively. The analysis of the number of articles and citations at different stages shows that seven countries conducted studies from 1995 to 2005, with the United States and Spain publishing the most articles, with 14 and 4, respectively. The other countries started their studies in 2006–2016. Portugal, a new country involved in the study,

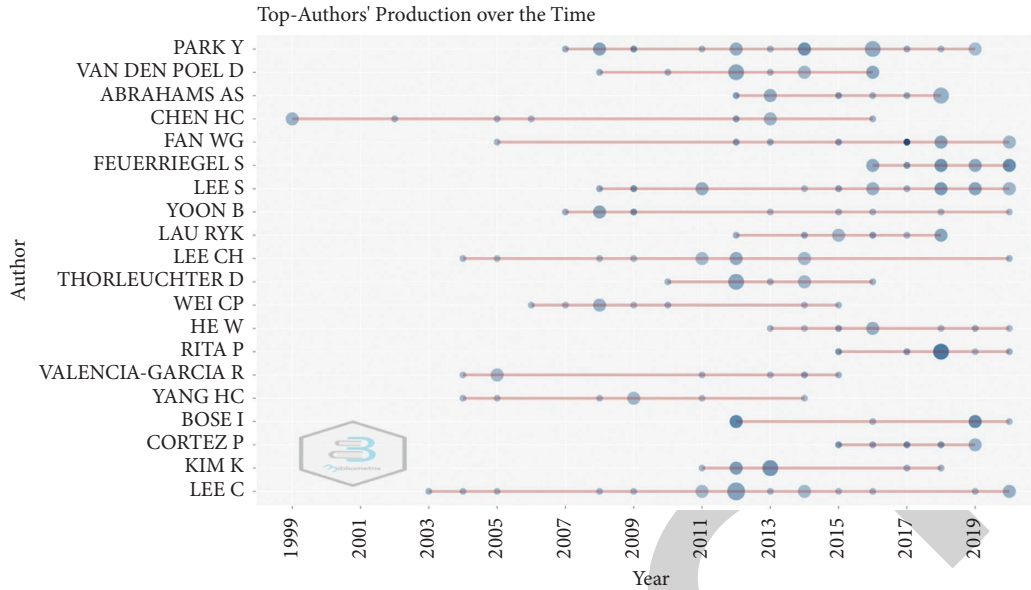


FIGURE 3: Top authors' production over time.

TABLE 5: Ranking of the most productive counties on text mining in management.

R	Country	TP	TC	TC/TP	<i>h</i>	>200	>100	>50	>10
1	USA	291	8718	29.96	47	7	21	44	164
2	China	208	4680	22.50	35	5	9	24	102
3	South Korea	88	2483	28.22	30	1	4	16	52
4	Spain	62	2345	37.82	23	2	5	14	33
5	UK	55	1476	26.84	19	1	3	10	27
6	India	55	1037	18.85	17	0	2	3	29
7	Germany	36	841	23.36	15	0	1	5	19
8	Australia	32	833	26.03	12	1	3	4	16
9	Netherlands	30	609	20.30	14	0	0	2	19
10	Italy	28	402	14.36	10	0	0	1	11
11	Canada	23	891	38.74	12	1	2	5	15
12	Turkey	22	540	24.55	12	0	1	3	12
13	France	21	160	7.62	8	0	0	0	7
14	Belgium	20	642	32.10	12	0	1	5	16
15	Brazil	20	586	29.30	10	1	2	2	11
16	Singapore	18	964	53.56	12	1	1	4	14
17	Portugal	18	871	48.39	12	1	3	7	14
18	Japan	16	336	21.00	7	0	1	1	4
19	Mexico	14	282	20.14	6	0	1	2	6
20	Greece	13	447	34.38	8	0	2	2	7

R, rank; TP, total publications; TC, total citations; TC/TP, the average number of citations per article in the country; *h*, *h*-index of the country; ≥ 10 , ≥ 50 , ≥ 100 , ≥ 200 , number of articles with more than 10, 50, 100, and 200 citations.

TABLE 6: Evolution of the most productive countries on text mining in management.

R	Country	<i>m</i>	Year	T1	T2	T3	CI	C2	C3	TC1/TP1	TC2/TP2	TC3/TP3
1	China	2.33	2006		63	145		2692	1988		42.73	13.71
2	South Korea	2.14	2007		46	42		2079	404		45.20	9.62
3	USA	1.81	1995	14	108	169	702	5705	2311	50.14	52.82	13.67
4	Portugal	1.50	2013		3	15		441	430		147.00	28.67
5	UK	1.27	2006		15	40		679	797		45.27	19.93
6	Germany	1.15	2008		18	18		673	168		37.39	9.33
7	Italy	1.11	2012		8	20		217	185		27.13	9.25
8	Brazil	1.11	2012		11	9		517	69		47.00	7.67
9	Spain	1.10	2000	4	38	20	47	1990	308	5.00	52.37	15.40

TABLE 6: Continued.

R	Country	m	Year	T1	T2	T3	CI	C2	C3	TC1/TP1	TC2/TP2	TC3/TP3
10	Netherlands	1.08	2008		11	19		310	299		28.18	15.74
11	Australia	0.92	2008		11	21		578	255		52.55	12.14
12	Belgium	0.92	2008		14	6		542	100		38.71	16.67
13	Greece	0.89	2012		5	8		253	194		50.60	24.25
14	Turkey	0.86	2007		8	14		371	169		46.38	12.07
15	India	0.77	1999	1	8	46	11	401	625	11.00	50.13	13.59
16	Canada	0.55	1999	1	12	10	55	716	120	55.00	59.67	12.00
17	Singapore	0.50	1997	1	12	5		909	55		75.75	11.00
18	Japan	0.44	2005	1	9	6	6	265	65	6.00	29.44	10.83
19	France	0.35	1998	1	7	13	3	47	110	3.00	6.71	8.46
20	Mexico	0.26	1998	2	5	7	8	231	43	4.00	46.20	6.14

R, rank; m , m -quotient; Year, the year the country published its first article; TP1, TP2, TP3, total publications of T1, T2, T3; TC1, TC2, TC3, total citations of T1, T2, T3; TC1/TP1, TC2/TP2, TC3/TP3, average number of citations per article.

published four articles in this period, with an average of 147 citations per article, the highest of all countries and about twice as many as Singapore (75.75). In the most recent period, 2017–2020, most countries showed almost the same or increasing research trends as in the previous period. Only Spain, Belgium, and Singapore reported a significant decrease in the number of articles. The number of citations is much lower than in the previous period due to the short period of publication of the literature in this phase, but Spain still ranks first with an average of 28.67 citations per article.

3.2. Science Mapping. As a complement to the results of the performance analysis, this section will perform a scientific mapping analysis of the text mining literature on management to answer RQ3, that is, to examine how the conceptual, social, and intellectual structure of the literature is characterized.

3.2.1. Conceptual Structure. The thesis is a vehicle for creative thinking in scientific research. Its main task is to convey scientific information. At the same time, it also has the significance of cultural storage and cultural accumulation. Whether from the perspective of transmitting information or storing information, the citation of subject terms or keywords will bring great convenience to the storage and retrieval of literature. Keywords are nouns or phrases that are used to express the thematic content of a document [39], not only for scientific and technical papers but also for scientific and technical reports and academic papers. Author keyword co-occurrence analysis can find keywords that appear frequently in an article, as well as keywords that appear in the same article [40]. Therefore, this paper identifies the research hotspots of text mining in management through co-occurrence analysis of author keywords. This was done by retaining the top 50 author keywords and clustering these keywords using the Louvain clustering algorithm [41]. The results are shown in Figure 4. The size of the nodes in the figure indicates the number of times the authors used the keywords, and the colors in the figure indicate the different keyword clusters. Thus, four keyword

clusters were identified with representative keywords for text mining (purple circles), sentiment analysis (blue circles), natural language processing and machine learning (pink circles), and topic modeling (green circles).

The first clustering theme is sentiment analysis, and the main keywords include sentiment analysis, social media, twitter, opinion mining, social network analysis, etc. Articles on this topic mainly focus on how to quickly and effectively perform sentiment analysis and opinion mining on posts and comments of social media users such as Facebook [42], Twitter [43], and YouTube [44]. The second clustering theme is text mining. The main keywords include text mining, classification, clustering, data mining, business intelligence, etc. Related papers mainly study the application of text mining in the field of management by developing personalized text classification and text clustering technology [45, 46], or designing reasonable text classification and text clustering process [47]. The third clustering theme is natural language processing and machine learning. The main keywords include natural language processing, machine learning, deep learning, information retrieval, information extraction, etc. the papers in this theme mainly discuss the automation of text mining technology in the field of management from the perspective of artificial intelligence, For example, sentiment detection out of textual snippets based on machine learning [48], fake news detection system based on deep learning model [49], multimodal sentiment analysis (MSA) based on deep learning [44], traffic accident text analysis based on machine learning [50]. The fourth clustering theme is topic modeling, and the main keywords include topic modeling, latent Dirichlet allocation, big data, text analytics, online reviews, etc. Related papers focus on the application of using the Latent Dirichlet Allocation (LDA) topic model in mining hidden topics in some common texts, such as online reviews [51], patent data [52], and research papers [53]. The hierarchical method decomposes a given dataset hierarchically until some conditions are met. Specifically, it can be divided into “bottom-up” and “top-down” schemes. Initially, each data record forms a separate group. In the next iteration, it combines those adjacent to each other into a group until all records form a group or a condition is met. In order to clarify the

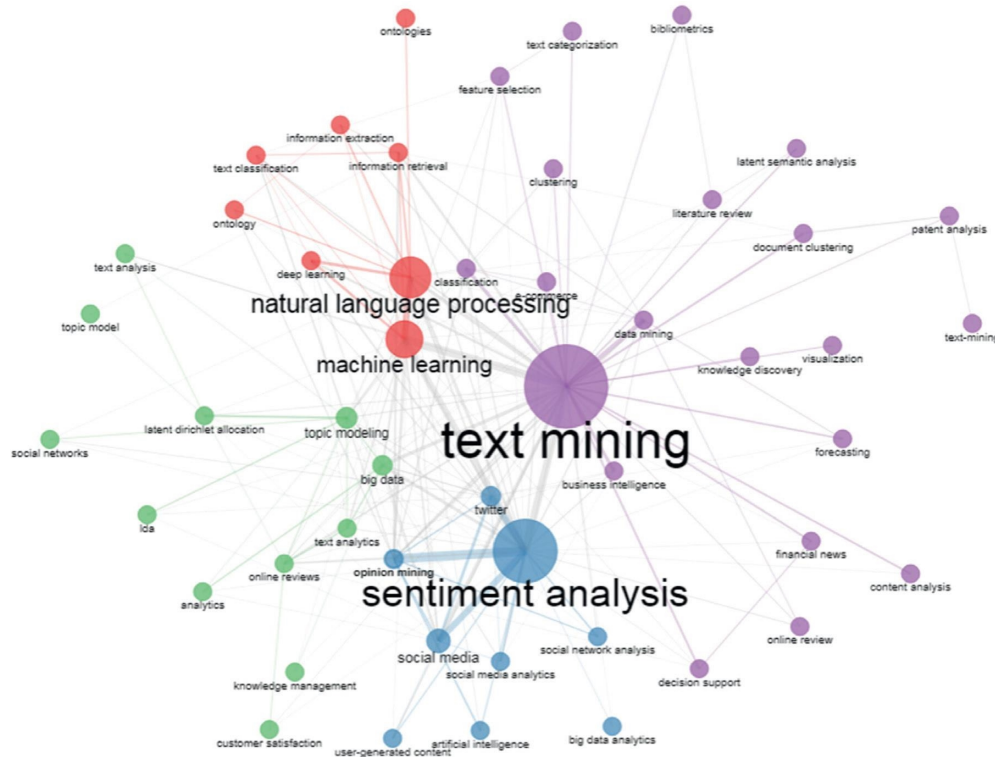


FIGURE 4: The conceptual structure of the dataset.

association and hierarchy between topics, Professor Blei, the proponent of LDA, proposed the hierarchical Latent Dirichlet Allocation (hLDA) based on LDA and hierarchical method, but no relevant application is found in management papers at present.

In order to obtain more in-depth information about the different topics, this paper further provides the evolution of the topics in the temporal dimension. As mentioned earlier, the application of text mining in the management field is divided into three phases, 1995–2005, 2006–2016, and 2017–2020, according to the research process. The co-occurrence of each stage considers 250 author keywords, and the results are presented in the form of strategic diagram, as shown in Figures 5–7.

Each research theme in the strategic diagram has two parameters. The first one is the parameter represented by the horizontal axis, which is called “centrality,” that is, the external connection strength between the theme and other themes. We can understand it as a parameter to measure the importance of the theme in the development of the whole research field. The second one is the parameter represented by the vertical axis, called “density,” that is, the connection strength between keywords within the topic, which can be understood as a parameter to measure the degree of the theme [54]. In this sense, the upper right quadrant of the strategic diagram shows themes with high density and centrality, indicating that these themes are well developed and important to the construction of the research field. The lower right quadrant of the strategic diagram contains themes with high centrality, but low density, and these themes are important for the development of the research

field, but they are not well developed, and are generally basic themes in the research field. The upper left quadrant of the strategic diagram includes themes with high density, but low centrality; it shows that these themes have developed well but have limited impact on the research field. Most of these themes are peripheral themes or highly specialized themes. The lower left quadrant of the strategic diagram highlights themes’ low density and centrality, the importance and development of these themes are both weakly, and they may be the emerging or disappearing themes in the field [55].

Thus, the evolution of the four categories of topics mentioned above can be observed in the figures. The basic theory of text mining, as a prerequisite for its application in the management field, has been central to the entire research process. Although the density fluctuates slightly, it has been at a low to medium level, indicating that its own theoretical system needs to be improved. The centrality and density of natural language processing and machine learning topics were low at the beginning of the study and have increased to varying degrees during the development phase, indicating that their own development and importance to field research is increasing. Sentiment analysis is a new topic with high centrality and low density in the development phase, and further decreases in density in the research expansion phase, indicating that although the topic of sentiment analysis is important, in recent years, there has been a decrease in research on this topic by relevant researchers, and topic modeling is a new topic in the research expansion phase. It has high density and centrality, promoting research on the application of text mining in the management field. It is currently the most cutting-edge research topic.

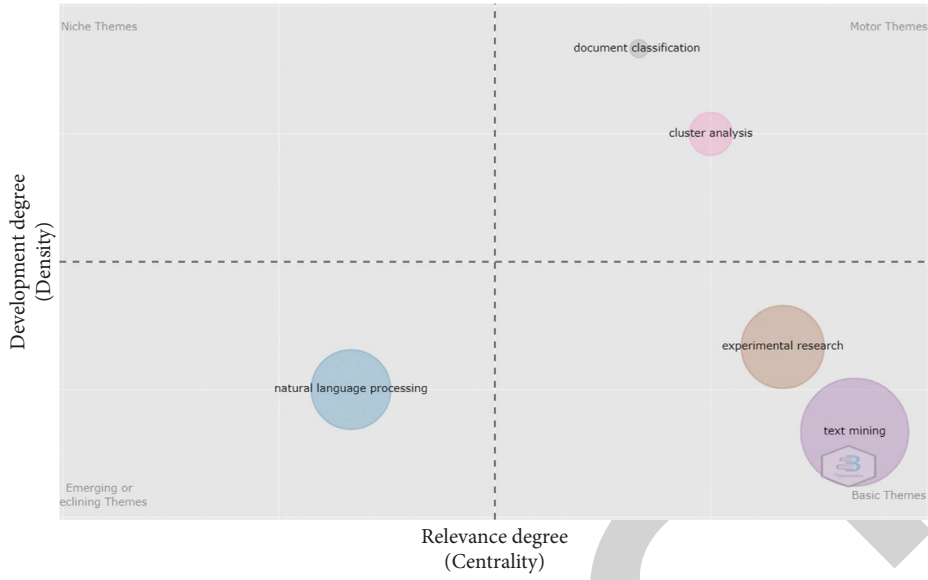


FIGURE 5: Strategic diagram of the initial stage (1995–2005).

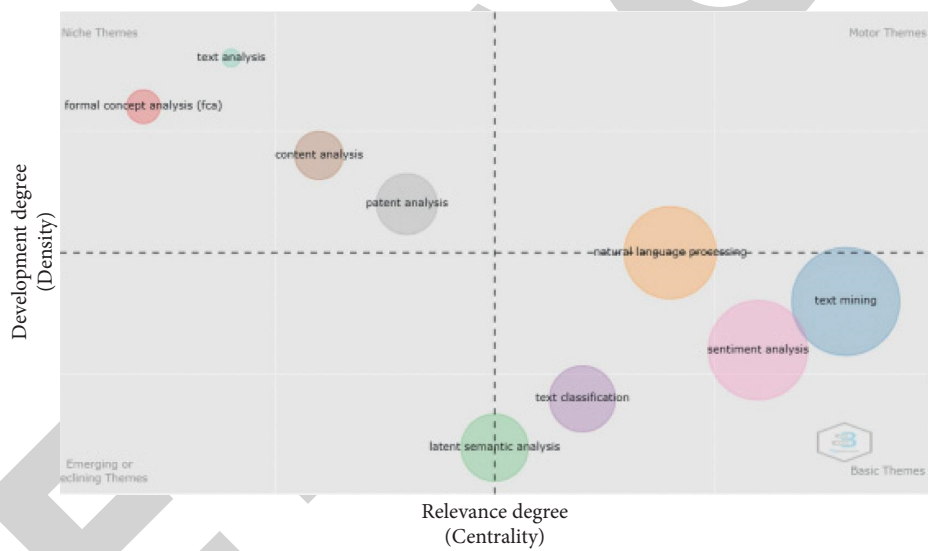


FIGURE 6: Strategic diagram of the development stage (2006–2016).

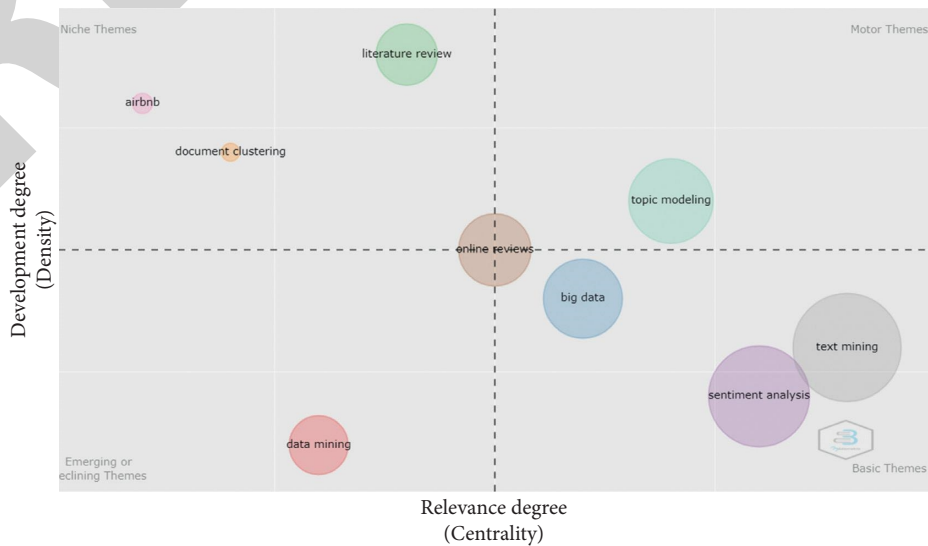


FIGURE 7: Strategic diagram of the expansion stage (2017–2020).

3.2.2. Social Structure. We further visualized the co-authorship relationships in the dataset through co-occurrence techniques. As with the co-occurrence analysis of author keywords, the Louvain clustering algorithm was again used. The co-authorship of the 100 most influential authors is shown in Figure 8. The size of the circles indicates the number of articles authored by the authors and the cross-coverage between the circles indicates the number of co-authored articles between the authors. It should be noted that isolated nodes do not mean that these authors do not collaborate with others, but only prove that they do not collaborate with the top 100 most influential authors. Therefore, to avoid misinterpretation, these isolated nodes are omitted from the figure. As can be seen from the figure, there are 21 collaborative groups of the top 100 influential authors, divided into 14 categories. The collaborative groups led by Abraham AS, Van den poel, and Park have the closest collaborative relationships.

Figure 9 presents the results for the authors' countries. The color of the country represents the number of national papers and the thickness of the inter-country linkage represents the frequency of collaboration. The USA and China are the darkest colored countries, and researchers from these countries have authored the most articles, which is consistent with the results of the analysis in Section 3.1.5. The coarsest link between the USA and China indicates that authors from both countries collaborate and communicate more frequently. In addition, the USA and South Korea, as well as China and the UK collaborate more frequently. A total of 17 pairs of countries collaborated more than five times in this area, as shown in Table 7.

3.2.3. Intellectual Structure. In order to explore the intellectual structure of the research, a historiograph was plotted in Figure 10, showing the historical development of the 20 most influential documents in chronological order [31]. From Tan's article published in 2008 to Xiang's article published in 2017, these 20 documents constitute a complete citation network. Several key nodes in the network need attention. First, Li's article published in 2010 proposed an unsupervised text mining method to detect and forecast hotspots in online forums [10]. This paper leads to several branches of research. Second, Yu published an article in 2013, quoting Li's article. Yu used sentiment analysis technology to explore the impact of social media and conventional media on the short-term performance of the company's stock market [56]. This article is one of the first relevant studies to study the impact of social media sources and traditional media. The third one is Nassirtussi's article in 2014. This paper is a literature review that systematically reviews studies related to market prediction based on online text mining and points out possibilities for future research [57]. Finally, Abrahams' article published in 2015. Building on the existing text mining research, Abrahams proposes an integrated text analytics framework for enterprise product defect discovery [58]. Table 8 shows the details of these four articles.

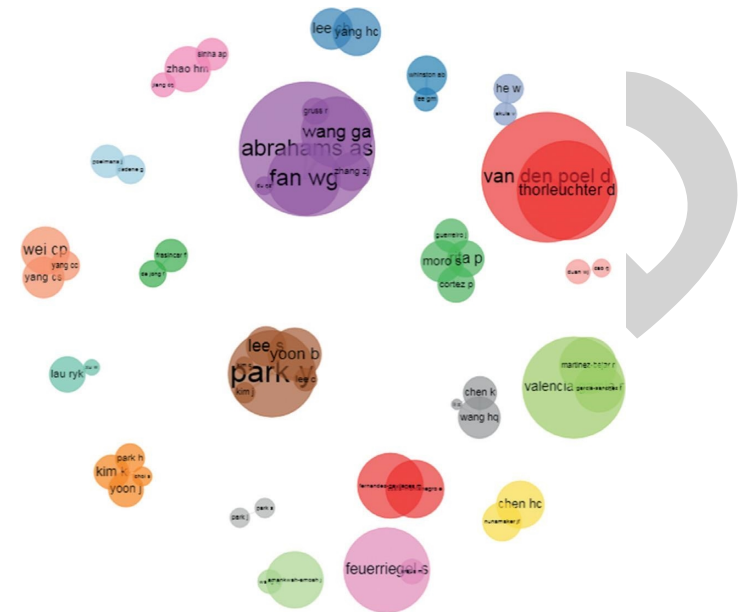


FIGURE 8: The social structure of the dataset.

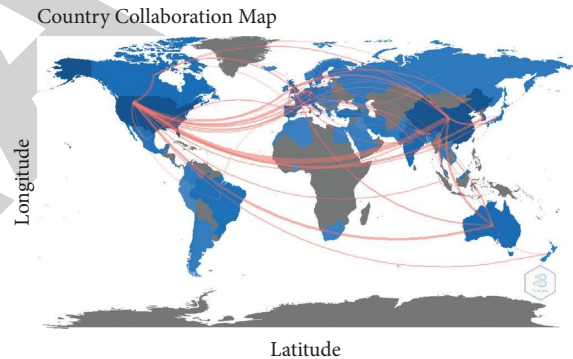


FIGURE 9: The rate of collaboration between countries based on authors' affiliations.

TABLE 7: Countries with more than 5 cooperation frequencies.

R	Cooperative country1	Cooperative country2	Frequency
1	USA	China	73
2	USA	Korea	17
3	China	United Kingdom	16
4	China	Australia	14
5	USA	Australia	14
6	USA	Canada	14
7	USA	Netherlands	10
8	Germany	Belgium	8
9	Netherlands	Belgium	8
10	USA	United Kingdom	8
11	USA	Germany	7
12	China	Japan	6
13	China	Singapore	6
14	Korea	United Kingdom	6
15	USA	Mexico	6
16	China	Korea	5
17	United Kingdom	Australia	5

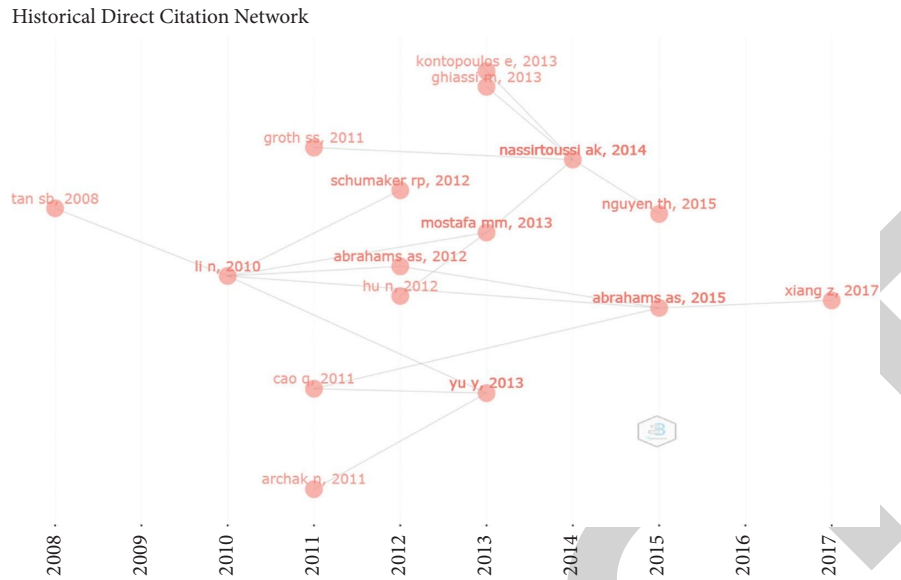


FIGURE 10: The historic evolution of co-citations amongst the 20 most relevant articles.

TABLE 8: Document information table of important nodes of citation network.

Id	Paper	LCS	GCS
1	Li, 2010, DECIS SUPPORT SYST DOI 10.1016/J.DSS.2009.09.003. Using text mining and sentiment analysis for online forums hotspot detection and forecast	26	250
2	Yu, 2013, DECIS SUPPORT SYST DOI 10.1016/J.DSS.2012.12.028. The impact of social and conventional media on firm equity value: a sentiment analysis approach	25	198
3	Nassirtoussi, 2014, EXPERT SYST APPL DOI 10.1016/J.ESWA.2014.06.009. Text mining for market prediction: a systematic review	20	226
4	Abrahams, 2015, PROD OPER MANAG DOI 10.1111/POMS.12303. An integrated text analytic framework for product defect discovery	22	94

R, rank; LCS, local citation score; GCS, global citation score.

4. Conclusions

The purpose of this work was to obtain an overview of text mining in the area of management. Performance analysis and scientific mapping in bibliometrics are used to assess the presence and structure of scientific publications. Performance analysis assesses the productivity and impact of a given document through several bibliometric indicators. The scientific mapping, as a complement to the performance analysis, reflects the conceptual, social, and intellectual structure of the literature in the field through co-occurrence of author keywords, co-authorship analysis, and historical co-citation analysis. These analyses were performed through Rstudio’s Bibliometrix package. In addition, different dimensions were analyzed in order to broaden the research horizon, including journals, articles, authors, and countries. The present study is useful for having a comprehensive overview of the state of text mining in the area of management. However, there are some limitations that must be mentioned. Using the WOS as the unique database for literature collection is considered the first limitation, which limits the number of analyzable literature. Furthermore, some exclusion criteria were used to refine the literatures collected (e.g., language, publication year, type of documents, and research fields). Future studies could be

conducted by expanding the literature pool in order to obtain more comprehensive findings.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

The authors confirm that the content of the manuscript has not been published or submitted for publication elsewhere. All authors have seen the manuscript and approved to submit to the journal.

Conflicts of Interest

The authors declare that there are no conflicts of interest in this paper.

Acknowledgments

This work was supported by project of China National Academy of Innovation Strategy “Analysis of Scientific

Research Competency Elements of Outstanding Scientific and Technological Innovation Talents” (2020020200015).

References

- [1] R. A. Poldrack, J. A. Mumford, and T. E. Nichols, *The Text Mining Handbook*, Cambridge University Press, United Kingdom, 2007.
- [2] C. Zong, R. Xia, and J. Zhang, *Text Data Mining*, Tsinghua University Press, Beijing, China, 2021.
- [3] M. Grobelnik, D. Mladenic, and M. Jermol, “Exploiting text mining in publishing and education,” in *Proceedings of the ICML-2002 Workshop on Data Mining Lessons Learned*, pp. 34–39, Ljubljana, Slovenia, April 2002.
- [4] G. Miner, J. Elder IV, A. Fast, T. Hill, R. Nisbet, and D. Delen, *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*, Academic Press, Massachusetts, 2012.
- [5] H. P. Luhn, “The automatic creation of literature abstracts,” *IBM Journal of Research and Development*, vol. 2, no. 2, pp. 159–165, 1958.
- [6] L. B. Doyle, “Semantic road maps for literature searchers,” *Journal of the ACM*, vol. 8, p. 553e578, 1961.
- [7] J. Žižka, F. Dařena, and A. Svoboda, *Text Mining with Machine Learning: Principles and Techniques*, CRC Press, Boca Raton, Florida, United States, 2019.
- [8] Q. Jones, G. Ravid, and S. Rafaei, “Information overload and the message dynamics of online interaction spaces: a theoretical model and empirical exploration,” *Information Systems Research*, vol. 15, no. 2, pp. 194–210, 2004.
- [9] C. Li and Y. Zhu, “The challenges of data quality and data quality assessment in the big data era,” *Data Science Journal*, vol. 14, no. 1, pp. 21–23, 2015.
- [10] N. Li and D. D. Wu, “Using text mining and sentiment analysis for online forums hotspot detection and forecast,” *Decision Support Systems*, vol. 48, no. 2, pp. 354–368, 2010.
- [11] A. Pritchard, “Statistical bibliography or bibliometrics?” *Journal of Documentation*, vol. 25, no. 4, pp. 348–349, 1969.
- [12] X. Zhai, Z. Li, K. Gao, Y. Huang, L. Lin, and L. Wang, “Research status and trend analysis of global biomedical text mining studies in recent 10 years,” *Scientometrics*, vol. 105, no. 1, pp. 509–523, 2015.
- [13] X. Chen, H. Xie, G. Cheng, L. K. M. Poon, M. Leng, and F. L. Wang, “Trends and features of the applications of natural language processing techniques for clinical trials text analysis,” *Applied Sciences*, vol. 10, no. 6, p. 2157, 2020.
- [14] T. Hao, X. Chen, G. Li, and J. Yan, “A bibliometric analysis of text mining in medical research,” *Soft Computing*, vol. 22, no. 23, pp. 7875–7892, 2018.
- [15] C. Forliano, P. D. Bernardi, and D. Yahiaoui, “Entrepreneurial universities: a bibliometric analysis within the business and management domains,” *Technological Forecasting and Social Change*, vol. 165, no. 1, Article ID 120522, 2021.
- [16] J. M. Merigó, A. Mas-Tur, N. Roig-Tierno, and D. Ribeiro-Soriano, “A bibliometric overview of the journal of business research between 1973 and 2014,” *Journal of Business Research*, vol. 68, no. 12, pp. 2645–2653, 2015.
- [17] A. W. Harzing and S. Alakangas, “Google scholar, scopus and the web of science: a longitudinal and cross-disciplinary comparison,” *Scientometrics*, vol. 106, no. 2, pp. 787–804, 2016.
- [18] E. C. Noyons, H. F. Moed, and M. Luwel, “Combining mapping and citation analysis for evaluative bibliometric purposes: a bibliometric study,” *Journal of the American Society for Information Science*, vol. 50, no. 2, pp. 115–131, 1999.
- [19] J. E. Hirsch, “An index to quantify an individual’s scientific research output,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 46, pp. 16569–16572, 2005.
- [20] W. Glänzel, “On the opportunities and limitations of the H-index,” *Science Focus*, vol. 1, no. 1, pp. 10–11, 2006.
- [21] A. F. Van Raan, “Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups,” *Scientometrics*, vol. 67, no. 3, pp. 491–502, 2006.
- [22] A. L. Kinney, “National scientific facilities and their science impact on nonbiomedical research,” *Proceedings of the National Academy of Sciences*, vol. 104, no. 46, pp. 17943–17947, 2007.
- [23] E. Csajbók, A. Berhidi, L. Vasas, and A. Schubert, “Hirsch-index for countries based on essential science indicators data,” *Scientometrics*, vol. 73, no. 1, pp. 91–117, 2007.
- [24] R. Costas and M. Bordons, “Advantages, limitations and its relation with other bibliometric indicators at the micro level,” *Journal of Informetrics*, vol. 1, no. 3, pp. 193–203, 2007.
- [25] J. Vanclay, “On the robustness of the h-index,” *Journal of the American Society for Information Science and Technology*, vol. 58, no. 10, pp. 1547–1550, 2007.
- [26] C. Kelly and M. Jennions, “The h-index and career assessment by numbers,” *Trends in Ecology & Evolution*, vol. 21, no. 4, pp. 167–170, 2006.
- [27] M. Gutiérrez-Salcedo, M. Á. Martínez, J. A. Moral-Muñoz, E. Herrera-Viedma, and M. J. Cobo, “Some bibliometric procedures for analyzing and evaluating research fields,” *Applied Intelligence*, vol. 48, no. 5, pp. 1275–1287, 2018.
- [28] M. Callon, J. P. Courtial, W. A. Turner, and S. Bauin, “From translations to problematic networks: an introduction to co-word analysis,” *Social Science Information*, vol. 22, no. 2, pp. 191–235, 1983.
- [29] H. P. F. Peters and A. F. van Raan, “Structuring scientific activities by co-author analysis an exercise on a university faculty level,” *Scientometrics*, vol. 20, no. 1, pp. 235–255, 1991.
- [30] H. Small, “Co-citation in the scientific literature: a new measure of the relationship between two documents,” *Journal of the American Society for Information Science*, vol. 24, no. 4, pp. 265–269, 1973.
- [31] E. Garfield, “Historiographic mapping of knowledge domains literature,” *Journal of Information Science*, vol. 30, no. 2, pp. 119–145, 2004.
- [32] M. Aria and C. Cuccurullo, “Bibliometrix: an R-tool for comprehensive science mapping analysis,” *Journal of Informetrics*, vol. 11, no. 4, pp. 959–975, 2017.
- [33] J. A. Moral Muñoz, E. Herrera Viedma, A. Santisteban Espejo, and M. J. Cobo, “Software tools for conducting bibliometric analysis in science: an up-to-date review,” *El Profesional de la Información*, vol. 29, no. 1, pp. 1–20, 2020.
- [34] C. Romero and S. Ventura, “Educational data mining: a survey from 1995 to 2005,” *Expert Systems with Applications*, vol. 33, no. 1, pp. 135–146, 2007.
- [35] S. Lee, B. Yoon, and Y. Park, “An approach to discovering new technology opportunities: keyword-based patent map approach,” *Technovation*, vol. 29, no. 6–7, pp. 481–497, 2009.
- [36] B. Becker, “Public R&D policies and private R&D investment: a survey of the empirical evidence,” *Journal of Economic Surveys*, vol. 29, no. 5, pp. 917–942, 2015.
- [37] J. M. Merigó, A. M. Gil-Lafuente, and R. R. Yager, “An overview of fuzzy research with bibliometric indicators,” *Applied Soft Computing*, vol. 27, pp. 420–433, 2015.

- [38] M. Gaviria-Marin, J. M. Merigó, and H. Baier-Fuentes, "Knowledge management: a global examination based on bibliometric analysis," *Technological Forecasting and Social Change*, vol. 140, pp. 194–220, 2019.
- [39] C. Xiang, W. Yuan, and H. Liu, "A scientometrics review on nonpoint source pollution research," *Ecological Engineering*, vol. 99, pp. 400–408, 2017.
- [40] F. J. Martínez-López, J. M. Merigó, L. Valenzuela-Fernández, and C. Nicolás, "Fifty years of the european journal of marketing: a bibliometric analysis," *European Journal of Marketing*, vol. 52, no. 1, pp. 439–468, 2018.
- [41] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, 2008.
- [42] H. N. Tran and E. Cambria, "Ensemble application of ELM and GPU for real-time multimodal sentiment analysis," *Memetic Computing*, vol. 10, no. 1, pp. 3–13, 2018.
- [43] E. Kontopoulos, C. Berberidis, T. Dergiades, and N. Bassiliades, "Ontology-based sentiment analysis of twitter posts," *Expert Systems with Applications*, vol. 40, no. 10, pp. 4065–4074, 2013.
- [44] P. D. Mahendhiran and S. Kannimuthu, "Deep learning techniques for polarity classification in multimodal sentiment analysis," *International Journal of Information Technology and Decision Making*, vol. 17, no. 03, pp. 883–910, 2018.
- [45] S. Lo, "Web service quality control based on text mining using support vector machine," *Expert Systems with Applications*, vol. 34, no. 1, pp. 603–610, 2008.
- [46] C. P. Wei, C. S. Yang, and H. W. Hsiao, "A collaborative filtering-based approach to personalized document clustering," *Decision Support Systems*, vol. 45, no. 3, pp. 413–428, 2008.
- [47] J. H. Suh, C. H. Park, and H. J. Si, "Applying text and data mining techniques to forecasting the trend of petitions filed to e-people," *Expert Systems with Applications*, vol. 37, no. 10, pp. 7255–7268, 2010.
- [48] M. Giatsoglou, M. G. Vozalis, K. Diamantaras, A. Vakali, G. Sarigiannidis, and K. C. Chatzisavvas, "Sentiment analysis leveraging emotions and word embeddings," *Expert Systems with Applications*, vol. 69, pp. 214–224, 2017.
- [49] Y. F. Huang and P. H. Chen, "Fake news detection using an ensemble learning model based on self-adaptive harmony search algorithms," *Expert Systems with Applications*, vol. 159, Article ID 113584, 2020.
- [50] C. Arteaga, A. Paz, and J. Park, "Injury severity on traffic crashes: a text mining with an interpretable machine-learning approach," *Safety Science*, vol. 132, Article ID 104988, 2020.
- [51] L. Y. Dong, S. J. Ji, C. J. Zhang et al., "An unsupervised topic-sentiment joint probabilistic model for detecting deceptive reviews," *Expert Systems with Applications*, vol. 114, pp. 210–223, 2018.
- [52] J. Wang and C. C. Hsu, "A topic-based patent analytics approach for exploring technological trends in smart manufacturing," *Journal of Manufacturing Technology Management*, vol. 32, no. 1, pp. 110–135, 2020.
- [53] L. Zhou, L. Zhang, Y. Zhao, R. Zheng, and K. Song, "A scientometric review of blockchain research," *Information Systems and E-Business Management*, pp. 1–31, 2020.
- [54] T. Cahlik, "Comparison of the maps of science," *Scientometrics*, vol. 49, no. 3, pp. 373–387, 2000.
- [55] H. K. Zhai, Q. Li, and X. W. Wei, "Research topics and evolution paths in the field of international mental health literacy from 2009 to 2018," *Journal of Southwest University for Nationalities*, vol. 42, no. 3, p. 7, 2021.
- [56] Y. Yu, W. Duan, and Q. Cao, "The impact of social and conventional media on firm equity value: a sentiment analysis approach," *Decision Support Systems*, vol. 55, no. 4, pp. 919–926, 2013.
- [57] A. K. Nassirtoussi, S. Aghabozorgi, T. Y. Wah, and D. Ngo, "Text mining for market prediction: a systematic review," *Expert Systems with Applications*, vol. 41, no. 16, pp. 7653–7670, 2014.
- [58] A. S. Abrahams, W. Fan, G. A. Wang, Z. J. Zhang, and J. Jiao, "An integrated text analytic framework for product defect discovery," *Production and Operations Management*, vol. 24, no. 6, pp. 975–990, 2015.