Information Processing and Management 57 (2020) 102279

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Social media analytics and business intelligence research: A systematic review

Jaewoong Choi[a], Janghyeok Yoon [a,*], Jaemin Chung[a], Byoung-Youl Coh[b], Jae-Min Lee[b]

[a] *Department of Industrial Engineering, Konkuk University, 120 Neungdong-ro, Gwangjin-gu, Seoul 05029, Republic of Korea*
[b] *Future Technology Analysis Center, Korea Institute of Science and Technology Information, 66 Hoegiro, Dongdaemun-gu, Seoul, 02456, Korea*

ARTICLE INFO

Keywords:
Social media
Online VoC
Open data
Business intelligence
Systematic review
Sentiment analysis

ABSTRACT

Evidently, online voice of customers (VoC) expressed in social media has emerged as quality data for researchers who are willing to conduct customer-driven business intelligence (BI) research. Nevertheless, to the best of authors' knowledge, there is still a dearth of studies that deal with such remarkable research stream and address various open data (e.g., social media, intellectual property) from a BI research perspective. Therefore, this study has attempted to evaluate the applicability of social media data in BI research and provide a systematic review on the primary research articles in the domain. This study compared social media data with the other open data (e.g., gray literature, public government data) in terms of data content, collection, updatability and structure, which are determined through a thorough discussion with experts. Next, this study selected 57 social media-based BI research articles from the Web of Science (WoS) database and analyzed them with three research questions about the data, methodologies, and results to understand this research domain. Our findings are expected to inform the existing researchers in the research domain about the future research directions, enable newcomers to understand the overall process of analyzing social media data, and provide the practitioners with social media analysis approaches suitable for their environment.

1. Introduction

Easy access and flexibility have made social media the most powerful and effective window for customers to voice their opinions for nearly a decade (Alkhodair, Ding, et al., 2019; Chang, Ku, et al., 2017). According to recent statistics, about 79% of population in the United States had a social media profile, followed by East Asia (70%) and Northern Europe (67%) (Statistica, 2019a, Statistica 2019b). In addition, approximately 77% of customers read other customers' online product reviews, 75% of them trust the reviews rather than individual direct recommendations, and more than 80% of them obtain helpful devices or information through social media-based channels (Xiao, Wei, et al., 2016). Likewise, a majority of customers express their opinions about a product or service without any censorship (Balbi, Misuraca, et al., 2018), by simply posting or commenting on various word-of-mouth channels such as Twitter, Facebook, and Amazon. Specifically, most of the customers simply write about their satisfaction or feelings about a product or service on social media, whereas some influencers provide a detailed review on key functions of a product or service and/or frank criticisms of their shortcomings. Such subjective opinions of customers can determine whether the other potential customers will purchase a product or not, because the purchase decision of most of the customers is likely to depend on the experiences of those who

* Corresponding author.
E-mail address: janghyoon@konkuk.ac.kr (J. Yoon).

https://doi.org/10.1016/j.ipm.2020.102279
Received 24 September 2019; Received in revised form 8 April 2020; Accepted 21 April 2020
Available online 18 June 2020

have already used the product (Rose, Hair, et al., 2011). Further, commercial firms or entrepreneurs can utilize the online reviews of customers to identify customers' overall interest, satisfaction, or requirements of their products. Particularly, expressive posts with ample information can provide crucial hints about new product development, function improvement, or quality management (Bashir, Papamichail, et al., 2017). Herein, it is evident that the modern customers are not passive buyers of a product, but active participants who control the future development direction of the product and its survival either directly or indirectly, with their powerful voices that are embedded in social media.

Some pioneers, who recognized the advantages of social media by using it as a web platform for communicating with their customers, have launched their own social media channels and sought to understand the customers and the market that they engage in (He, Zha, et al., 2013). As a representative example, Samsung mobile opened its Facebook page named "Samsung Global" in 2010 to have approximately 50 million followers and to capture their candid responses on new products. By observing the contents of comments or the number of likes or shares, the company can identify its customers' responses or expectations from new products. As such, most of the commercial firms, whose main products are multifunctional such as smart phones, have launched their own channels on social media platforms such as Facebook, Instagram, and Twitter to listen to their customers' voices (Shirdastian, 2017). The underlying technologies of such high-tech products have already reached a mature state; therefore, product innovation through a technology-driven approach such as in-house R&D is rather difficult. The technology-intensive firms can adopt a market-driven approach where the future development direction of the product depends on the customer needs. For example, smart devices such as Samsung Galaxy Note are always confronted with customer needs for additional or new features, such as Samsung pay, touch pen, and applications (Jeong, Yoon, et al., 2018), which can be used as a starting point for establishing the future development direction of a technologically matured product. Moreover, customers' voices have a great influence on competitive markets. Indeed, Starbucks, one of the leading companies in the coffee chain market, opened an online platform named "My Starbucks Idea" in 2008 to crowd-source ideas and to interact with its customers, so that it can collect the customers' feedbacks or ideas about a prior product or service (Gallaugher & Ransbotham, 2010). These opinions, which may be trivial, can help the firm to strengthen the customer-firm relationship and to occupy a strong position in the competitive market, by providing the customers with crowd-customized services. Therefore, online VoC embedded in social media becomes a quality resource for both competitive advantage acquisition and product development.

Scholarly attention has been drawn toward understanding the importance of social media, one of the open data and the best virtual places for customers to express their views. Such academic interest in social media data has emerged as a scholarly attempt called social media analytics, which involves the process of gathering and analyzing various social media data and extracting valuable hidden information. Indeed, many researchers have collected a large amount of social media data generated by customers, and have derived statistical results, quantified satisfaction scores, or topics of text, which can help the commercial firms to establish strategic decisions in their business environment. Specifically, Jeong, Yoon, et al. (2018) collected the online review of Samsung Galaxy Note 5 from Reddit, one of the social media platforms, and captured customers' complaints about battery life, touch recognition, camera, etc. by applying topic modeling to the review data. In addition, they measured the scores of satisfaction and importance, and identified the opportunity score for each topic by using sentiment analysis and opportunity algorithm. Such information, which they derived, can support the related firm's decision making in the development of a product or service, by providing product features that have high importance scores but low satisfaction. Further, Trappey, Trappey, et al. (2018) collected the Amazon review data about three competing smart phones and quantified the interest of customers by using the major features that they defined. Their results aid the customer-centric product improvement by providing customers' expectations and the relative position of each product in the competitive market. Besides, there are numerous studies to analyze customer satisfaction quantitatively by applying sentiment analysis to the online data generated by customers. These studies attempted to provide quality information that can support the decision making in a business environment, and such approaches or analytical processes were defined as business intelligence (BI) by the modern scholars (Chang, Ku, et al., 2017; Ko, Jeong, et al., 2018; Wang, Feng, et al., 2018).

Here, it is evident that the online VoC expressed in social media has recently emerged as a good data source for researchers to conduct the BI research. Indeed, there are a number of valuable studies to derive quality information from the VoC available on social media; some of these can be developed to achieve an automated VoC monitoring system with a high reproducibility (Ko, Jeong, et al., 2018). Nonetheless, to the best of authors' knowledge, there is a dearth of studies that address the aforementioned remarkable research stream and the availability of social media in terms of BI research, although there are some social media-related review papers with different aspects (Olanrewaju, Hossain, et al., 2020; Salo, 2017). Accordingly, this study collects the social media-based BI literature from the Web of Science (WoS) and provides a systematic review to analyze the research trends, advantages, and limitations of each approach and categorize them. In addition, this study assesses the social media data in terms of the content, collection, updatability, and structure of the data, explaining its advantages and characteristics compared to the other open data in BI research. Finally, this study suggests a specific manual for collecting the social media data and performing high-quality BI research. In this review, we focus on the social media data used in prior studies and the results derived through their BI research with three pertinent research questions. (1) What are the promising social media platforms that are currently in use in this research domain? (2) What are the dominant methodologies or algorithms that researchers use to apply to social media data for BI? (3) What intelligent information does the social media-based BI research derive?

This paper is expected to make three contributions. First, this study can depict the suitability of social media data in BI research, and hence, can enrich the data source adopted by the existing researchers by explaining the advantages of social media that is used as a quality input data in BI research, unlike the traditional data such as patents and research articles. Second, our systematic review can provide specific social media platforms, latest methodologies, and algorithms, along with the advantages and limitations of prior studies to the researchers. Indeed, this study provides a detailed manual on the process of collecting, preprocessing, and analyzing the social media data, which can aid the prospective researchers or practitioners to solve business problems. Finally, this study can present current propositions and future research directions of the research stream, pinpointing the common limitations of current research and suggesting the possible research directions by using the latest algorithms.

The remainder of this paper proceeds as follows. Section 2 describes the theoretical background relevant to this study. Specifically, we review the definition of BI and depict various open data that can be used for BI research. Section 3 presents the evaluation results of the identified open data used to conduct BI. In Section 4, we categorize and compare the prior studies that use social media data to conduct quality BI research through a systematic review. Subsequently, we discuss the strengths and challenges of prior studies, and suggest the future direction of the social media-based BI research. Finally, we present the conclusions of this study.

2. Theoretical background

This section explicates the concept of BI research and introduces the classification of open data to understand the theoretical background of the study that addresses social media-based BI research.

2.1. Business intelligence

BI is an umbrella term that encompasses the entire process of acquiring, interpreting, collating, analyzing, and deriving quality information or knowledge in various business contexts (Davenport, 2006). Chaudhuri, Dayal, et al. (2011) remarked that a BI system enables the knowledge workers or entrepreneurs to understand their business or market better and make strategic decisions in a timely manner. Indeed, many excellent researchers have conducted the BI research by analyzing the critical business data available since the late 1990's when BI was conceptualized as a unified term (Chen, Chiang, et al., 2012). To the best of authors' knowledge, the earliest forms of BI involved a data-centric approach toward traditional data such as patents, research articles, or commercial data.

Particularly, because patents can represent the companies' underlying technologies and R&D activities, a number of researchers attempted to utilize this information in discovering the technology opportunity, monitoring the business environment or detecting the emerging technology. As a representative analytical approach toward patent data, Park and Yoon (2017) proposed a systematic approach that can recommend practical technology opportunities to a technology-intensive firm by capturing its technology capacity from its patents and identifying similar firms' technology portfolio through collaborative filtering. Their study is expected to be particularly pragmatic for small and medium enterprises with low human resource and initial capital, suggesting future directions in the development of technology verified by leading companies. Moreover, Yoon and Kim (2011)'s approach can help various kinds of stakeholders such as researchers and R&D policy makers to monitor the rapidly evolving R&D trends, by addressing the technical contents described in patents based on their subject-action-object structures and visualizing an overall trend based on the network structure. Wang et al. (2018) proposed a patent-based approach to identify the emerging topics in a target technology domain, by clustering patents based on their classification codes and measuring an emergence indicator for each cluster term. Likewise, most of the analytical results of the patent-based approaches can support the commercial firm's decision making on R&D activities or monitoring competitive firms. Furthermore, research articles reflect the recent trends in research and thus are often employed as input in the BI research to detect and track the emerging technology trends quantitatively, although they are not directly related to business opportunity as in the other data (Huang, Schuehle, et al., 2015; Kajikawa, Yoshikawa, et al., 2008).

Interestingly, the traditional data including patents are still employed as the main data, but some researchers remarked that the latest trends in BI research are gradually focusing toward analyzing customer-generated data. In addition, Chung and Tseng (2012) proposed another facet to the prior concept of BI, by highlighting that the online product reviews are a major data source for understanding the customers and market. Indeed, various social media-based approaches have provided the commercial firms or stakeholders with abundant and unique business opportunities by monitoring the customers' opinions and identifying the competitive position in the market they engage in (Pang & Lee, 2008). Thus, a consensus now exists that the social media data is appropriate for BI researchers to understand the customers and market, although there is a debate over which is the best open data for BI research. Accordingly, this study investigates various existing data sources and compares their analytic suitability in the next subsection, before conducting a systematic review on the social media-based BI research.

2.2. Open data classification

As mentioned previously, prior studies employed many kinds of data to conduct BI research from their own unique perspectives. Therefore, we tried to identify various existing open data and classify them in this section. According to the open source intelligence, open data is defined as unclassified open sources of information, which is widely available, but differs from an open source software (Bradbury, 2011). Based on the relevant literature (Best Jr & Cumming, 2007; Steele, 2007), we classify the open data into 1) mass media, 2) commercial material, 3) grey literature, 4) public government data, 5) professional/academic publication data, and 6) web-based data. The actual examples of each category are presented in the following table.

Firstly, mass media includes traditional information derived from printed newspapers, magazines, radio, or television (Best Jr & Cumming, 2007). The traditional mass media data has been used in BI research to serve targeted advertisements effectively, to predict customer churn, or to monitor audience interests (Chan-Olmsted, 2002; Rust, Moorman, et al., 2010). Secondly, commercial material involves geospatial information such as commercial imagery, or an enterprise's financial/industrial assessment report, which is closely coupled with the enterprise's business opportunity; however, it is a relatively rare data source (Gibson, 2004; Steele, 2007). Grey literature publications include materials or research reports that are available to the public, but are not published commercially;

for example, various reports (technical reports, statistical reports, memoranda, etc.), technical documentation, commercial documentation, and bibliographies (Alberani, Pietrangeli, et al., 1990; Conn, Valentine, et al., 2003). Particularly, intellectual property such as patents and trademarks are the descriptive grey literature of the technology and business fields, and thus, they are frequently used in BI research (Lee & Lee, 2017; Yoon & Kim, 2012). Public government data refers to government-issued official reports or data (Yang & Lee, 2012) and specific datasets or statistics that are available on the government website (e.g. US: www.data.gov, UK: www.data.gov.uk, France: www.data.gouv.fr, Singapore: www.data.gov.sg, India: www.data.gov.in, Korea: www.data.gov.kr) (Attard, Orlandi, et al., 2015; Gurstein, 2011). Recently, various kinds of datasets including agriculture, climate, consumer, and education are being actively posted on the websites because of the political trends of open government and the advent of big data era (Krishnamurthy & Awazu, 2016), but the doubt whether these public government data or the other open data have the appropriate information to help solve the business problems remains. However, professional/academic publication data are rather frequently utilized in BI research (Porter, Garner, et al., 2018; Shiau, Dwivedi, et al., 2018). This is because the information obtained from journals, conferences, and symposia contains the latest trends and the best methodologies, algorithms or techniques (Wallace, Van Fleet, et al., 2011), which can be applied to firms' R&D activities, despite its intensely academic and scientific nature. Lastly, web-based data includes social media (e.g. Facebook, Twitter, and Instagram), online discussion groups, and citizen media such as YouTube, which have easy accessibility, and can therefore outpace other data sources. Particularly, social media has emerged as a representative data source for BI research, because it provides voluminous, real-time, and customer-generated data of potential value to commercial firms, stakeholders, or researchers (Jimenez-Marquez, Gonzalez-Carrasco, et al., 2019). Similarly, online discussion groups refer to deliberative social media where a group of individuals formally or informally exchange informative comments on specific topics, products, or services, whereas Reddit, Facebook group, and Google group are representative.

3. Open data evaluation

3.1. Evaluation dimension

Some researchers insisted that social media data is sufficiently valuable as an input in social media-based BI research, and further in VoC monitoring system (Jeong, Yoon, et al., 2018; Xu, Wang, et al., 2017). In this study, we evaluate the analytical suitability of social media data, and compare it with the other data sources in terms of four evaluation dimensions. Specifically, we assessed the analytical suitability of each open data source in terms of the content, collection, updatability, and structure of the data (Table 2). Each dimension was selected to reflect common attribute of each open data in the process of collecting and analyzing the data, as well as continuously operating BI system. The four dimensions were selected by 10 experts, who are currently engaged in BI research. The experts' pool involves four professors and six graduate students. The professors have conducted BI research by using various open data such as social media, research proposals, and patents to discover business opportunities in terms of product or technology for approximately 10 years. In addition, the graduate students have also been involved in similar research domains such as social media mining, customer understanding, and data-driven business planning for approximately four years.

Firstly, the data content represents the intrinsic properties of each open data such as contents, professionalism, and users' intellectual level, addressing the degree to which the commercial firms or researchers can obtain helpful information in business decision making from the open data (Jimenez-Marquez, Gonzalez-Carrasco, et al., 2019). Secondly, the data collection identifies the existence of an open application programming interface (API) service to collect an immense amount of data. Specifically, this dimension indicates whether the open data can be continuously collected, accumulated, and used as an input in BI research (Stieglitz et al., 2018). Thirdly, the data updatability indicates whether the open data contains the latest data, and whether its database can be updated continually (Cvijikj & Michahelles, 2011); this, as well as the data collection are closely related to the systematization of BI research. Lastly, the data structure assesses whether the open data can be processed in a structured format in the BI research or system (Huang and Nguyen, 2015).

3.2. Evaluation results

In this section, the integrated evaluation results obtained after consulting the experts, are reported (Table 3). The experts evaluated each open data type for its content, collection, updatability, and structure, and ranked them as high, medium, and low. We summarized the experts' evaluation of each open data and specified them in the following table.

Firstly, a majority of experts regarded that it is difficult to utilize mass media and commercial data for BI research in terms of data collection and management, because they are physical materials and therefore, excessive elaboration is required to convert them into

Table 1
Open data classification.

Category	Examples
Mass media	Printed newspaper, magazines, radio, television
Commercial material	Commercial imagery, financial/industrial assessment
Grey literature	Intellectual property, newsletters, technical reports, bibliographies
Public government data	Public government reports, government's open data
Professional/Academic publication data	Information from journal, conferences and symposia
Web-based data	Social media, online discussion group, online publication

Table 2
Open data evaluation.

Dimension	Description
Data content	Contents covered by each open data as a data source, professionality, and users' intellectual level; the extent to which contents relates to BI research
Data collection	Whether there is an open API service for collecting a large amount of open data
Data updatability	Whether each open data accumulates the latest data and its database can be updated continuously
Data structure	Whether each open data can be processed easily in a structured (analytical) format

Table 3
Evaluation results.

Category	Data content	Data collection	Data updatability	Data structure
Mass media	Low	Low	Low	Low
Commercial data	High	Low	Low	Low
Grey literature	High	High	High	High
Public government data	Medium	Medium	Low	High
Professional/Academic publication data	Medium	High	High	High
Web-based data	High	High	High	Medium

an analytical format. On the contrary, the grey literature, typified by patents or trademarks, describes the technology or business areas where commercial firms are involved (Lee & Lee, 2017; Yoon & Kim, 2012), and are therefore considered to be closely coupled with BI research. In addition, most of the grey literature is published periodically and accumulated constantly by the official authorities, enabling a systematic database construction. Hence, the grey literature was highly appreciated by a majority of our experts in terms of data content, collection, updatability, and structure. However, public government data received a lower evaluation than the grey literature for the same indicators. Taking Data.gov as an example, it can be realized that a representative public government data website provides over 300,000 datasets derived from the government departments of various areas such as agriculture, climate, ecosystem, and education (Krishnamurthy & Awazu, 2016); however, there are a few datasets that are directly or potentially related to business opportunities or customers' voices. In addition, although the datasets are provided in analytical formats such as html, bin, xml, pdf, zip, originator data format, and csv, they are published non-periodically and in different formats, making it difficult to collect and accumulate datasets continuously. Lastly, web-based data received high evaluation along with the grey literature. Web-based data classified largely by social media, online discussion groups, and citizen media, is generated by the common users or customers in real time, ensuring its high data updatability. In addition, most of the web-based data has a unique open API service, otherwise which web crawling can be applied, facilitating a systematic collection. Particularly, on social media or online discussion group page, several customers express their opinions on specific products, services, or topics without any censorship by simply posting or commenting, and these can be used in BI research for purposes such as VoC monitoring (Jeong, Yoon, et al., 2018), product opportunity identification (Ko, Jeong, et al., 2018), and customer churn prediction (Shirazi & Mohammadi, 2018). However, the customer-generated data has a major downside because it requires preprocessing, owing to its unstructured data including typing errors, emoticons, abbreviations, or inconsistent data formats such as text, video, or audio (ur Rehman, Chang, et al., 2016).

4. Systematic review and results

4.1. Systematic process

In this paper, we conducted a systematic review on the articles that analyze social media data and derive quality information that is useful for business decision making, by following the guidelines developed by Kitchenham (2004). We adopted a systematic review approach (Kitchenham, 2004) rather than a concept-centric approach for several reasons (Webster & Watson, 2002). First, social media-based BI research came from decision-making issues in real-life business environment rather than a theoretical background. Accordingly, a systematic review approach is employed to select social media-based BI research articles and analyze their approaches for solving business problem. Secondly, we attempted to analyze the recent trends in the social media-based BI research in terms of methodology, algorithm and data used, rather than the development path of theoretical background by tracking the citation relationship in the research domain. Consequently, this review can provide a detailed manual on the overall process of collecting, preprocessing, and analyzing the social media data. Furthermore, we expect this systematic review to encourage the prospective researchers to devise new research items in this research domain, by suggesting new methodologies, data, or problems. To accomplish these aims, we defined the following questions:

- RQ1. What are the promising social media platforms that are currently in use in this research domain?
- RQ2. What are the dominant methodologies or algorithms that researchers use to apply to social media data for BI?
- RQ3. What intelligent information does the social media-based BI research derive?

There are some prominent terms that often appear in the social media-based BI research, e.g., “social media” or “business.” In addition, terms such as "product review" or "service review" were also used, since they have similar concepts as the core terms and represent one of the social media categories However, instead of these core terms, some research articles often refer to specific social media platforms such as Twitter, Facebook, or Amazon that they used in their titles, abstracts, or keywords, clarifying the analysis target or input data. Similarly, researchers often mention the specific methodology, algorithm, or framework such as the sentiment analysis, network analysis, or topic modeling that they utilized to address social media data in their title or abstract, whereas some of them use common terms such as "social media mining" or "social media analysis" instead. As such, researchers that are engaged in this research domain describe their social media-based BI research by using different keywords or expressions. Therefore, we found it difficult to cover all the articles in the domain of social media-based BI research in this review, and therefore created a search query with the highest search purity. Alternatively, we used the 'or' operator than 'and' operator in searching query with the main keywords, for discovering as many research articles as possible. The main keywords that are mentioned previously are featured in the following search query (SQ):

- SQ. "TITLE: ((social and media or ((product or service) near (review))) and (mining or analy* or sentiment or business)) NOT TITLE: (systematic and review)".

This review aims to address the analytical social media-based BI research; therefore, we excluded the review papers and literature surveys that contain the term "systematic review" in their titles. Conversely, the keyword "sentiment" is included in the search query, because sentiment analysis is one of the effective methodologies that deals with the subjective content embedded in social media data and is thus frequently used in this research domain. This search query was used on the WoS, one of the popular academic search engines, on July 4th, 2018. By doing so, we obtained a set of potential research articles that analyze social media to conduct BI research. Here, we limited the search period to the latest five years and narrowed the document type to peer-reviewed journal articles, considering that the social media has recently emerged as a data source for BI research and the latest research trends are our major concern. The WoS search for articles by using the aforementioned search query returned 633 records. These retrieved articles included not only social media-based BI research articles, but also research articles to improve algorithm performance or literature study on social media. This initial list of articles (Group 1) was utilized as the subject in exclusion criteria and trend analysis.

Subsequently, we examined the title and abstract of Group 1 articles and selected the primary articles by considering several exclusion criteria, to improve the validity of the review. Specific exclusion criteria (EC) are presented as follows:

- EC1. Only the peer-reviewed articles of a journal are considered.
- EC2. Book review, meeting abstract, and editorial material are excluded.
- EC3. Title or abstract that does not provide a definite and sufficient description of the research is excluded.
- EC4. The articles that are not accessible are excluded.
- EC5. Only the articles that are written in English are considered.
- EC6. The articles should be unique; the ones that have repetitive content are excluded.
- EC7. Only the research articles are included in this review; literature survey articles are excluded.
- EC8. Articles that do not analyze social media or product review in BI research are excluded; e.g., articles that use social media as an input in new computing algorithm to describe performance improvement are excluded.
- EC9. Articles that do not derive quality information that can support business decision making from the social media are excluded.

After filtering the Group 1 articles (633 articles) by applying the exclusion criteria, 57 articles were shortlisted for a systematic review. Specifically, 10 articles were found to be deduplicated and 566 articles were considered to be irrelevant after examining their abstracts. Above all, we excluded the research articles that do not correspond with the aim of this review; e.g., review papers, surveys, bibliometrics, and theoretical conceptualization. Also, conference papers or reports were excluded since their referred journal article would represent state-of-the-art results with high impact. Such approach is consistent with approaches from previous review papers on social media, where only peer-review journal articles have been analyzed (Chan & Ngai, 2011; Olanrewaju, Hossain, et al., 2020). In addition, we excluded the research articles that are faintly related to BI research in terms of the research purpose, based on their abstracts. Particularly, we excluded a number of research articles that aim to propose new models or improve computing algorithms and use social media data as input data only to show their superior performance. Finally, the primary articles (Group 2) that are considered to derive the analytical results or propose a novel approach to address the social media data in terms of BI research are shortlisted. We considered each paper of Group 2 as a representative example of social media-based BI research and conducted a thorough analysis on each of them. The list of Group 2 articles is presented in Table A.1 in the appendix.

4.2. Review results

In this subsection, the general trend analysis results of Group 1 articles are presented, and the systematic review results of Group 2 articles are described.

4.2.1. General trends

First, we plotted a graph displaying the change in the number of Group 1 articles according to the publication year (Fig. 1). From the figure, it is evident that the number of research articles published has been increasing since 2014 (83, 118, 150, and 180 articles in order). Particularly, nearly 100 research articles had already been published by the month of July in 2018. Accordingly, the surge

Table 4
Distribution of Group 2 articles based on the social media type and the specific platform.

Type	number of articles	Specific platform	number of articles
Commercial reviews	29	Amazon	6
		Yelp	5
		TripAdvisor	4
		Jindong	2
		Booking.com	2
		iPhone Apps Plus	2
		Etc.	8
SNS	24	Twitter	11
		Flickr	4
		Yahoo Finance	3
		Facebook	2
		Weibo	2
		Qzone, StockTwits	2
Online discussion group	2	Reddit	2
Multi-data	2	Amazon-Patent	1
		Twitter-Patent	1

in social media-based BI research articles seemed to continue. Likewise, a similar upward trend was observed for Group 2 articles. The increase in the number of Group 2 articles in 2012 and 2015 was fourfold, and that of 2018 was also high despite the short period. Therefore, we could confirm that for BI research, the social media is recognized as an appropriate data source, and this research domain can be a promising one. Here in, we suspected that this growing trend in social media-based BI research was driven by the proliferation of particular authors or journals.

Not surprisingly, the articles were mainly published in the peer-review journals focusing on the data-driven approaches toward information processing and management, intelligent system, and expert decision support system. Throughout the analysis period, the relevant journals have consistently published social media-based BI research papers. In addition to the journals in the field of information science, some of the Group 2 articles had been published in the journals of various fields such as marketing, service business, electronic commerce, tobacco control, and industrial safety, indicating that social media analysis is being used for different purposes in various business environments.

4.2.2. What are the promising social media platforms that are currently in use in this research domain?

To address RQ1, this systematic review on Group 2 articles investigated the different social media platforms that are used to conduct BI research. Eventually, we identified multiple platforms and grouped their data into four types to analyze the articles effectively: 1) commercial review, 2) social networking service (SNS), 3) online discussion group, and 4) multi-data. Table 4 shows the distribution of Group 2 articles based on the social media platforms that they used. This indicates that the majority of the platforms used are of type commercial review or SNS data. Specifically, from the perspective of individual platforms, Twitter, Amazon, Yelp, and TripAdvisor were frequently employed and can be considered promising social media platforms.

The benefits, difficulties, and rationale for using each platform's data are presented in the following paragraphs. In addition, our findings revealed different features of platforms, such as main contents, data delivery method, degree of personalization, sharing, and collectivity. For example, in group-centered platforms such as Reddit, the common users form a community on a specific topic first, and then share the post and discuss the topic; however, social media platforms also include individual-centered platforms such as Facebook, Instagram, or Twitter. In the latter platforms, individual users usually build their own private network, and share posts within a more limited scope. Thus, we could investigate the promising platforms that were used in the Group 2 articles.

4.2.2.1. Online reviews on commercial products/services. Online reviews refer to the empirical data generated by customers who purchased or experienced the products/services. Similar to the other types of social media platforms, online reviews' platforms also have a voluminous amount of data, and a majority of them provide open API services, which facilitate the data collection. The main strength of online review data lies in their unified topics. Here, the customers write about the products or services that they experienced, and thus, the online review data are accumulated based on the product/service, thereby minimizing the data selection procedures after collection. Because of the descriptive features of commercial review data, we could divide them into two types: 1) product reviews on e-commerce sites (e.g. Amazon and Jindong) and 2) service reviews on crowd-sourced social media platforms (e.g. Yelp and TripAdvisor). In summary, out of 29 articles using commercial reviews, 14 articles of them used online product reviews and the remaining 15 articles used online service reviews.

Firstly, in online product reviews, the customers who purchase the products through e-commerce sites usually write reviews and share their experiences. As such, online product review data contain customers' experiential knowledge and opinions on functions, appearance, and specifications of a product. Indeed, several analytical approaches that are proposed in 14 articles (Group 2) have employed these review data to understand customer requirements or analyze the product's competition in market (Wang, Feng, et al., 2018). Qi, Zhang, et al. (2016) extended the use of online product review toward supporting customer-oriented product improvement strategies. Specifically, they applied the conjoint analysis and sentiment analysis to approximately 700,000 Jindong online reviews and identified the customer requirements from them, positing that online product reviews can provide new and innovative ideas for new product development or improvement similar to the other studies (Chong, Li, et al., 2016). Considering that online product reviews have a significant impact on customers' purchase decisions, not surprisingly, a number of studies have been conducted (Govindaraj & Gopalakrishnan, 2016). Interestingly, a majority of them adopt sentiment analysis to investigate benefits of online reviews (Salehan & Kim, 2016), or to rank competitive products (Liu, Bi, et al., 2017).

Furthermore, service reviews, where customers who have experienced online reservation services or offline services (e.g., accommodation business and restaurant business) share their experiences, are a type of commercial review. Such shared experiences on services can also influence other customers' purchase decisions (Xu, Wang, et al., 2017), and can have positive or negative effects on brand reputation (Filieri, McLeay, et al., 2018). In addition, service review data can convey fruitful information such as complaints, opinions, and experiences, which can be utilized to derive prior problems, identify customer needs, and measure a service quality. Indeed, 15 articles in Group 2 paid attention to the benefits of service review data and their ripple effects on the customer, by extending an application of the data to BI research. Especially, Chang, Ku, et al. (2017), who proposed an integrated framework including data collection, preprocessing, and data analysis, have empirically remarked the intricate relationships of customers' ratings, sentiments, and the type of customer by using TripAdvisor data and Google Trends data. Likewise, some profound studies that focused on complex correlation between customers and their review data were conducted (Gui, Zhou, et al., 2017; Micu, Micu, et al., 2017). Their results can help commercial firms to identify customers' individual characteristics from a large-scale review data and provide crowd-customized services. Further, by emphasizing the rich contents of service review data, several studies proposed qualitative approaches to measure customer satisfaction or service quality to utilize the VoC in the direction of service improvement (Song, Lee, et al., 2016). Accordingly, here, it can be confirmed that a large amount of service review data not only reflects customers' characteristics, but also indicates the quality of the service they experienced.

In summary, the key research questions of commercial review-based studies can be summarized as customer understanding, satisfaction of brand/product/service, and the analysis of relationship between customers and reviews. Lastly, a majority of commercial review-based studies indicate that the preprocessing process is an essential step in their approach because of the data quality. However, the online review data often contains abbreviations, emoticons, or typos in its text, similar to the other social media data. In addition, there are quite a number of autocomplete texts in online product reviews, and thus, the data cleansing process is necessary before analysis.

4.2.2.2. Social networking service platforms. BI studies using SNS data accounted for 33% of Group 2 articles because SNS is the most commercially available and accessible social media platforms. Indeed, SNS, often used as a narrow concept of social media, is based on an individual user's private network (e.g., Facebook and LinkedIn) or micro blog (e.g., Twitter, Weibo, and Flickr). Regardless of the specific features of the SNS platforms, the main usage layer consists of a voluminous amount of common users and they generate content on various topics in real-time; thus, a large amount of SNS data is produced. In addition, a majority of the SNS platforms offer their unique open API service. However, SNS data has a relatively large number of typos, emoticons, and so on, which require preprocessing. For example, Twitter, which was used in more than 22% of Group 2 articles, generates more than 500 million tweets on average per day. This platform limits the number of characters to 140; this compels the user to express their thoughts, opinions, emotions, and so on in abbreviated words and emoticons. Therefore, although the Twitter data contain candid opinions about various topics including new products, services, or brand, it requires a systematic preprocessing process to extract and refine available keywords from short texts and to derive their contextual meanings.

These stressed features are also revealed in the collected BI research papers. Above all, social media-based BI research of this type makes good use of the immensity of SNS data. Indeed, most of the studies that use SNS data gathered a large amount of textual data as many as hundreds of thousands, and refined their dataset based on language and keywords. In addition, the studies that analyzed the photo data posted on Flickr collected over 100,000 datasets (Miah, Vu, et al., 2017; Sun & Lee, 2017). These SNS data-based studies collected their target data relatively easily through open API services or web archives (e.g., Cardiff Online Social Media Observatory) (Burnap, Rana, et al., 2015). Such systematic collection method is an important cornerstone in automating the approach or analytical process proposed in each study. In other words, given the goal of developing a BI system rather than conducting a one-off study, a systematic approach to gathering, such as open API, should be employed.

Information obtained from SNS data such as Twitter, Flickr, and Facebook includes text, photos, locations, and user information. The textual data contains the information about users' intentions, opinions, sentiments and so on, similar to other social media, and therefore, a number of studies have attempted to monitor the topics that users are primarily interested in and their reactions by using text mining, sentiment analysis, or sometimes machine learning (ML) algorithms (Xu, Qi, et al., 2018; Yoo, Song, et al., 2018). Further, there are several delicate studies that focus on the textual information itself by identifying terms or topics with a high analytical value or deriving a correlation between predefined product attributes and comment length (Lo, Chiong, et al., 2017; Wong and Lacka, 2017). By contrast, some studies concentrated on the relationship between the users, who are the generators of SNS data (Fang, Sang, et al., 2014; Xie, Li, et al., 2014). SNS users build their own private networks in the SNS platforms and interact with each other, by exchanging opinions about posts, events, topics and so on. Therefore, in the social network, user communities or influencers (opinion leaders) are generated naturally. Several researchers addressed the user ID information to understand the user needs or opinions better by identifying potential user communities or analyzing influencer opinions.

4.2.2.3. Online discussion group. Online discussion groups refer to platforms where user communities are formed to discuss various topics such as specific products, technologies, or services, and they exchange the opinions and information about the topic by generating posts and comments. Of the various social media platforms, Reddit is the closest to the online discussion group because it has over 1.2 million subreddits to address various topics including news, sciences, movies, books, music, and so on. Users of a subreddit are more likely to have a specialized knowledge or interest in the topic than the other social media users, and thus, their posts have more specific and ample content (Park, Conway, et al., 2018). By taking a subreddit of a specific smartphone as an example, it is found that its users not only express satisfaction on the product, but also describe specific features, pros, and cons, and compare them with other competitive products. In addition, according to a study that analyzed the text of Samsung Galaxy Note 5 subreddit, various topics such as calling, software update, Samsung pay, camera, and battery life were covered (Jeong, Yoon, et al., 2018). Likewise, several promising researchers proposed a quantitative approach to discover product/service opportunities by using Reddit's text data (Ko, Jeong, et al., 2018), but this data is expected to be applied in BI research owing to its ample textual contents. Reddit, like the other social media platforms, sometimes includes inscriptions, typos, and emoticons, which require preprocessing, but it is expected to be acceptable owing to its ample textual information.

4.2.2.4. Multi data. As mentioned in Section 2.2, there are various types of open data available for BI research, but the reference information they can provide for business decision making is different. For example, patents are known to reflect the outcomes of firms' R&D activities, and therefore, patent-based BI research studies tried to understand firms' business activities by using the firms' patents (Yoon, Park, et al., 2015), whereas social media contains VoC, and social media-based BI research studies analyzed the firms' product or service in terms of the market (Trappey, Trappey, et al., 2018). Nevertheless, the majority of prior BI research studies tend to utilize a single type of open data to derive quality information that can support experts in a business environment. In addition, most of the social media-based BI research studies in Group 2 articles employed only social media data to understand firms' business in terms of the products/services/brands evaluated by customers in the market. In fact, there are some studies where two or more social media platforms are used for analysis (Akay, Dragomir, et al., 2015a; Nikfarjam, Sarker, et al., 2015), but they are nothing more than a general social media-based BI research because the multiple social media data have a similar perspective of firms' business.

However, there are multi data-based BI research articles where both social media and patents are analyzed at the same time (Li, Xie, et al., 2018; Trappey, Trappey, et al., 2018). These studies tried to address both firms' technical competence and customers' opinion on the firms' products, services, or brands. In other words, attempts were made to understand the integrated process where commercial firms develop their technologies by using R&D activities, launch new products, and the customers experience the products. Such attempts should be made and a systematic BI system should be formed to implement a market-driven approach by relating customers' needs to actual technologies. However, this type of research articles made a good contribution by analyzing multi-data, but they require further enhancement. For instance, research, where social media and patents were linked indirectly with the same analysis period, did not consider the difference between the patent issuance date and the product release data (Li, Xie, et al., 2018). Further, empirical studies revealed that there is a clear time lag from three months to three years between the patent disclosure and market launch (Gerken, Moehrle, et al., 2015; Tietze, Reul, et al., 2009). An ontology-based approach that matches the patents and commercial review data directly made a substantial contribution to the academic (Trappey, Trappey, et al., 2018). However, the knowledge ontology schema construction requires a laborious expert intervention and thus, the approach could be developed into a systematic BI system. Likewise, attempts to match two or more heterogeneous open data can be advantageous in analyzing the commercial firms' business activities from an integrated perspective, but more systematic and accurate approach is required than the currently proposed ones.

4.2.3. What are the dominant methodologies or algorithms that researchers use to apply to social media data for business intelligence?

In this section, we focused on the approaches adopted by Group 2 articles, to identify the efficient ways to deal with social media data. Hence, we sought to understand the purpose and content of each approach adopted, before analyzing the algorithms used. Here, we investigated the analytical procedure of Group 2 articles and defined an integrated framework for social media-based BI research as shown in Fig. 2. This framework involves four steps: data collection, preprocessing, analysis through multiple algorithms, and result validation and interpretation.

First, data collection refers to gathering of text, image, or video data through open API service or web crawling, and configuring a database, the cornerstone for analysis. Considering the security and privacy of social media data that is most important in analyzing social media data but often ignored as in the Cambridge Analytica affair (Cadwalladr & Graham-Harrison, 2018), the use of open API services would be best. Therefore, we investigated all available open API services and websites where social media data are shared, to provide specific manuals for social media-based BI research. The list of open API services of various social media platforms, including the platforms utilized in Group 2 articles, is presented in Table A.2 in the appendix. Next, data preprocessing or cleansing should be done to obtain a quality input dataset from raw data. Particularly, data preprocessing of text data consists of a keyword extraction process from the texts through natural language processing and a keyword set refinement process by removing stopwords. As mentioned previously, because social media users often write relatively short texts using emoticons, abbreviations, and typos, it is necessary to refine the extracted keyword set. Additionally, to obtain more significant keywords, some researchers used the frequency of a keyword in a document (TF; term frequency) or the number of documents (DF; document frequency) where the keyword appears. Emoticons ("^^", ":-D", etc.), web link ("www"), onomatopoeic words ("haha," "blah," etc.), etc. are often excluded, and further, general terms ("product," "process," "device," etc.) or irrelevant words ("day," "month," etc.) are also excluded. After such thorough removal process, the unstructured text data is usually converted into a structured format (e.g., document-term matrix) with the selected keywords. Alternatively, some researchers used word embedding results for analyzing short-text data without complicated preprocessing.

During the data analysis and validation steps, multiple algorithms and techniques were applied; however, the represented depths of the algorithms varied. By considering the depth and intention behind using the algorithms, we classified them into five analytical

Table 5
Analytical approaches of Group 2 articles and purposes.

Analytical approach	Purpose	Frequency	Subsection
Sentiment analysis	Detecting sentiment polarity of customers' opinions or response to a specific topic (product, service, brand)	27	Lexicon-based Text classification-based Deep learning-based
Topic modeling	Identifying the topics mentioned by customers, general users or the public.	13	LDA (Latent Dirichlet Allocation) LSA (Latent Semantic Analysis)
ML-based approach	Classifying customers' sentiment polarity or predicting stock prices, product sales or customers' sentiment	27	Naïve Bayesian Random forest Decision tree Self Organizing Map Neural Network Random walk Logistic Regression Support Vector Machine (SVM) Gaussian Mixture Model Hidden Markov Model Clustering
Network-based approach	Identifying customers' interaction network or discovering influential users.	7	Social network analysis Network clustering Chance discovery Hyper graph
Theoretical approach	Quantifying customer satisfaction, service lifecycle, or service quality.	6	Fuzzy theory Game theory Opportunity algorithm Kano model Life cycle analysis Multi criteria decision analysis

approaches: 1) sentiment analysis, 2) topic modeling, 3) ML-based approach, 4) network-based approach, and 5) theoretical approach. Table 5 represents the number of Group 2 articles and the purpose of each category, where the articles of each category can be overlapped. Not surprisingly, the most widely used approaches were textual data-based algorithms such as sentiment analysis and topic modeling.

Particularly, the overwhelming use of sentiment analysis indicates that most of the social media-based BI researches focus on emotional response or subjective opinion expressed by the customers (Burnap, Rana, et al., 2015; Sun, Wang, et al., 2015). Because sentiment analysis computes people's polarity of sentiment embedded in their texts as a score, the majority of Group 2 articles have adopted it to analyze customers' response or satisfaction on a product, brand, or topic quantitatively (Duan, Yu, et al., 2016; Kang & Park, 2014). Specifically, the computed sentiment score enables the researcher to identify qualitative information such as whether customers' response is positive or negative; however, prior studies used the number of documents as the quantitative index of customer interest. In addition, sentiment analysis can detect the polarity of sentiment from short and low quality text (Guzman & Maalej, 2014), and thus, social media data, which consists of short phrases with few keywords or mostly involves stopwords including abbreviations and emoticons, can be analyzed relatively accurately and effectively through this prominent method. Therefore, the majority of Group 2 articles have applied sentiment analysis to social media data with relatively low quality to identify customer's satisfaction, service quality, or brand image (Chang, Ku, et al., 2017; Xu, Wang, et al., 2017). For example, in a real-life business environment, sentiment analysis results enable the entrepreneurs or experts to specifically identify the features their customers are disappointed with or satisfied with for their products or services. Therefore, sentiment analysis would be an important foundation to improve customer understanding through social media mining. However, to achieve better results, additional analysis such as ML algorithms, network analysis, or various theories in addition to sentiment analysis is required in social media-based BI research.

Topic modeling is also frequently used in Group 2 articles to reduce various contents mentioned by customers into *n*-topics. Indeed, they have applied various topic modeling algorithms such as latent Dirichlet allocation (LDA) and latent Semantic Analysis (LSA) to the text expressed by the keyword vector, rendering the identification of key topics from a large amount of customer-generated data. For example, by applying LDA to the document-keyword matrix, the probabilistic distribution of the topic by each document and that of the keywords by each topic can be obtained (Blei, Ng, et al., 2003). Accordingly, based on the results of LDA, each topic can be interpreted primarily by using keywords with high probabilities, and the key topic covered by each document can be identified. Specifically, Group 2 articles utilized topic modeling to identify all the topics or events that general users often refer to or to analyze customers' candid complaints or opinions on a specific topic, product, service, or brand (Brandt, Bendler, et al., 2017; Trappey, Trappey, et al., 2018; Xu, Qi, et al., 2018).

Numerous studies have applied well-developed ML algorithms to classify, predict, or cluster problems in social media data. Indeed, pioneering researchers have attempted to predict stock prices (Ho, Damien, et al., 2017; Jiang, Liang, et al., 2014), product sales (Chong, Li, et al., 2016), or sentiment score (Yoo, Song, et al., 2018), by employing multiple models such as artificial neural networks, support vector machine, regression analysis, and Naïve Bayesian model. These studies have a performance verification process in common, where their prediction performance is compared with existing prediction models in terms of precision, recall, accuracy, and F-measure. First, researchers trying to predict quantitative figures such as stock prices or product sales at certain point of time, have shown an interesting tendency to use sentiment scores in social media text as independent variables (Jiang, Liang, et al., 2014; Nguyen, Shirai, et al., 2015; Sul, Dennis, et al., 2017). Challenging researchers attempting to address the polarity of sentiment in the social media text, have taken an additional step to compare their performance with those of the existing models (Gui, Zhou, et al., 2017). Although their well-established models outperformed the traditional ML-based models such as Naïve Bayesian model and random forest in terms of accuracy, recall, and F-measure, the researchers should validate their models against various social media datasets and improve them so that they can be widely used in future researches. In addition, most of the studies using the results of sentiment analysis in ML-based models require a qualitative judgment by analysts who specify the threshold for classifying sentiment scores for large learning datasets.

Meanwhile, some studies have suggested creative approaches to identify new business opportunities by analyzing commercial review data on products or services, SNS, and online discussion group data. Specifically, they have tried to detect possible defects or side effects when using the product considering the customers' opinions (Akay and Erlandsson, 2015a; Mummalaneni, Gruss, et al., 2018), to recommend the appropriate products by customer type based on the text mining results (Xu et al., 2017) or to predict the life cycle phase of a service by applying review data with time information to the hidden Markov model (Kim & Lee, 2017). Of the traditional ML-based classifiers, support vector machine (SVM) is most consistently used throughout the analysis period from 2013 to 2017 (Govindaraj & Gopalakrishnan, 2016; Nikfarjam, Sarker, et al., 2015). This is because this dominant algorithm shows good performance on high dimensional sparse data, such as text data of social media which are written in unique words of each customer. Of course, in Group 2 articles, relatively new ML algorithms such as convolutional neural networks (CNNs) and long short-term memory (LSTM) were employed. CNNs have shown superior performance in image classification (Kim, 2014), but their use is being extended to social media mining (Gui, Zhou, et al., 2017). Indeed, a study has adopted CNNs for classifying the sentiment of users for the occurred events, where all sentences are represented in two-dimensional vectors as images, based on the words vectorized by Word2Vec algorithm (Yoo, Song, et al., 2018). Meanwhile, LSTM, a special kind of recurrent neural network, shows excellent performance in solving the problems dealing with sequential data such as text or time series data, because it can remember inputs of the past steps and address long sequence with long time lags (Lample, Ballesteros, et al., 2016). Leveraging such feature of the LSTM, some researchers have applied bidirectional LSTM to text data of social media data, to identify safety issues with a product (Xie and Zeng, 2017).

Social media mining can identify various entities such as customers, tags, keywords, and topics. However, their relationships are complex and numerous, which makes it difficult to grasp the data structure. To deal with such complex and massive social media data, some researchers have used social network analysis and expressed it as a network or a graph structure (Miah, Vu, et al., 2017; Xie, Li, et al., 2014). Social network analysis can identify the overall structure by representing the relationships of nodes connected in various directions visually and investigating important nodes by using well-developed quantitative indicators such as degree centrality, closeness centrality, and betweenness centrality (Kim & Hastak, 2018). Indeed, some studies in Group 2 articles have calculated the similarities between users using their tags, keywords, etc. and then constructed the network (Akay, Dragomir, et al., 2015b). On the created network, they have applied several algorithms including clustering or self organization map theory and tried to identify the latent user community (Xie, Li, et al., 2014) or influential users (Fang, Sang, et al., 2014). In addition, a study has expressed the co-occurrence relationships of topics that were derived through topic modeling in the form of a graph, otherwise which are difficult to grasp from a large volume of text data (Ko, Jeong, et al., 2018). Based on the constructed graphs, the study was able to lay the groundwork for visualizing their results and adopting the chance discovery theory. Thus, network structure or network analysis can be considered as the basis for effective analysis of large amount of data, and many researchers have conducted additional analysis based on the network indeed. In addition to the chance discovery theory utilized in the aforementioned study, researchers have employed various theories such as opportunity algorithm to prioritize unmet needs (Jeong, Yoon, et al., 2018), Kano model to represent the relationship between customer expectation and satisfaction (Qi, Zhang, et al., 2016), life cycle model to address life quantitatively (Kim & Lee, 2017), multi criteria decision analysis to deal with various criteria in the decision making (Kang & Park, 2014), game theory (Wang, Chen, et al., 2018) or fuzzy theory (Liu, Bi, et al., 2017). Likewise, abstract theories have been addressed quantitatively in social media-based BI research.

4.2.4. What intelligent information does the social media-based BI research derive?

With the social media data and methodologies of Group 2 articles summarized in Sections 4.2.2 and 4.2.3, we categorized their analytical subject and application method to derive their own intelligent information (Table 6). Social media-based BI research represented by Group 2 articles is of interest to customers who mainly use social media, their relationship, the content they generated, or the characteristics of social media data itself. Each researcher attempted to understand the customer or social media with their own perspective by identifying customers' opinions, latent relationships among the customers, or high-interest topics.

There have been a number of studies to discover product opportunities, or even business opportunities, by analyzing the content generated by general users semantically. These studies can provide business opportunities in a relatively intuitive form, such as a set

Table 6
Analytical subjects and specific results of Group 2 articles.

Subject of analysis	Application method
Contents generated by customer	Identifying Product & Business opportunities Analyzing competitive position in a target market Analyzing market performance of a product Surveilling product quality & Detecting safety hazard Identifying customer requirement and satisfaction Investigating service quality & gap Discovering customer-driven product technology and functions …
Relationship among customers	Detecting latent customer communities Identifying influential users in a network Mining product adopter information Analyzing tourists' behavior …
Customer reaction	Predicting stock market movement Monitoring customer reaction & Detecting burst in sentiment-aware topics Detecting anomalous emotion & adverse event Predicting online product sales Forecasting dynamic customer needs Monitoring the development trend of emerging technologies …
Characteristics of social media data	Examining the theoretical foundation of review helpfulness Investigating the predictors of readership and helpfulness of online customer reviews

of keywords or labeled topics, which enable the commercial firms to reflect customers' opinions or experiences on product development directly (Jeong, Yoon, et al., 2018; Ko, Jeong, et al., 2018). Rather than providing specific insights to firms with discovered opportunities directly, several researchers have attempted to provide more analytical results by analyzing competitive intelligence or identifying market performance or market share (He, Zha, et al., 2013; Jin, Ji, et al., 2016; Wang, Feng, et al., 2018). In a similar way, several researchers, who are interested in quantifying customer reactions on a product, service, or brand offered by commercial firms, measured customer satisfaction, service quality, and product risk quantitatively, which enables the firms to reflect customers' needs or opinions in their outputs (Kang & Park, 2014; Mummalaneni, Gruss, et al., 2018; (Shirdastian, 2017); Song, Lee et al., 2016). In such a fashion, several researchers have proposed systematic methodologies such as recommendation system and experts support system (Jeong, Yoon, et al., 2018; Stavrianou & Brun, 2015), whereas most of the studies use one-off analysis to grasp customers satisfaction, loyalty, or perception. To summarize, there have been proposed methodologies or frameworks for understanding customers by mining customers' intrinsic information from the text they write, which can provide the commercial firms with customer-driven insights. Further, several leading researchers have attempted to discover new functions or emerging technologies of products associated with the identified customer needs by analyzing social media and patents together (Li, Xie, et al., 2018; Trappey, Trappey, et al., 2018).

Several researchers have focused on the interaction and relationship between customers who use social media (Fang, Sang, et al., 2014; Miah, Vu, et al., 2017). They have attempted to improve the understanding of customers by detecting potential community with the user pool or identifying influential users or product adopters (Xie, Li, et al., 2014; Zhao, Wang, et al., 2016). Rather than analyzing each customer in such a fashion, some researchers addressed the entire customer group by monitoring their emotional change and reaction to new technologies, topics, products, or brands (Burnap, Rana, et al., 2015; Li, Xie, et al., 2018; Yoo, Song, et al., 2018). Unlike the aforementioned research flows, there were some challenging researchers who attempted to predict stock prices, product sales, etc. based on the customer information identified from the social media; this indicates that the influence of customers is not just on products, but also on the status of commercial firms including stock prices (Chong et al., 2016; Ho, Damien, et al., 2017; Jiang, Liang, et al., 2014; Nguyen and Velcin, 2015). Similarly, there are a few studies where social media analytics were used to analyze service lifecycle (Kim & Lee, 2017). Yet, although interest in such research fields is growing, there seems to be room for further development. Researchers interested in the nature of social media itself attempted to identify the correlation between the commercial review data on social media and the price of product or analyze the benefits of each review comment (Filieri, McLeay, et al., 2018; Hu and Lee, 2017; Salehan & Kim, 2016). While such studies may seem weak in the nature of BI research, their results can help the commercial firms or other researchers to understand the social media data better.

5. Discussions and concluding remarks

This paper has attempted to analyze the applicability of social media in BI research and the recent trend of the research domain. This was done by evaluating the open data of social media as input data and by a systematic review on articles that conducted social media-based BI research. The evaluation results obtained after sufficient consultation with experts stated that the social media has high relevance to BI research, because of the high scores of data collection, updatability, and structure. Accordingly, we concluded that the analytical suitability of social media in BI research would be higher than any other open data. In addition, such applicability of social media was identified in the systematic review results, because general trends showed an increasing number of social media-based BI research papers, which were collected from WoS database.

Through a systematic review on Group 2 articles, which were considered closest to our analytical purpose, we were able to categorize various social media platforms that are currently used in this research domain and identify the most promising social media type (e.g., online reviews, SNS, online discussion group, and multi-data) and social media platforms (e.g., Twitter, Amazon, Yelp, TripAdvisor). Next, we revealed the dominant methodologies or algorithms that researchers have applied to social media data for BI research and analyzed their frequency, considering their analytical purposes. Consequently, we could investigate the intelligent content the articles have derived from social media such as emerging product opportunities, influential users in a network, detecting customer reaction on new technology or product, and the correlation between review data and its benefits. With the aforementioned analytical results, this paper is expected to be the first step towards understanding the current research trends in social media-based BI research and pinpointing the advantages and limitations of prior studies in this research domain, encouraging future researchers to develop existing analytical approach to social media data.

Our detailed analysis on Group 2 articles with three critical RQs provided a sufficient description of social media-based BI research in terms of methodology, data and intelligent results. However, we have found several development points. First, social media-based BI research is likely to rely heavily on sentiment analysis to enhance its understanding of customer. Sentiment analysis is a well-established method because it enables the researchers to interpret customers' language accurately and its results are intuitive. However, to provide intelligent information that commercial firm can refer to in a real-life business environment, social media-based BI studies should not depend only on sentiment analysis itself. In other words, rather than applying sentiment analysis to social media data to investigate whether the customers are good or bad, prominent researchers in this research domain need to continually devise and suggest quantitative approaches to track specific reason why a product has received negative reviews from its customers and to discover new features to satisfy existing customers or to lead new ones. Because a large amount of social media data includes a variety of information such as texts, users, location, photographs, or video in addition to the sentiment score, there are hidden hints about product opportunities, emerging technologies, or customer-driven functions. In a similar way, social media-based BI studies need to evolve into a systematic methodology or framework beyond a one-off analysis. For example, some pioneering researchers have applied sophisticated algorithms to social media data, which are relatively difficult, to reproduce similar results for various datasets in a real-life business environment. Accordingly, such excellent researchers would be able to develop better expert support systems, considering the reproducibility or practical applicability of their approaches.

Next, social media-based BI researcher can consider combining the social media data with other open data such as patents, to conduct BI research from a broader perspective. As mentioned previously, social media data reflects customer-driven information without any censorship. However, because not only customers, but also technology, firms, or competitive environment should be considered in decision making in a real-life business environment, social media-based BI researcher can also attempt to analyze multiple open data such as patents, research papers, government open data with social media data to reflect various perspectives. Our findings of Group 2 articles showed that there are a few attempts to combine technical data such as patents with social media data such as Twitter and Amazon directly or indirectly. However, such attempts need to be more active and evolved. Indeed, several challenging researchers addressed the two different data by introducing a qualitative method such as ontology or comparing time-series trends at the same analysis period, which were relatively difficult to apply in a volatile business environment because of its low reproducibility. In analyzing the two disparate data together, it would be possible to utilize trademark data as a mediator or apply well-established text mining techniques. Product name or applicant information registered in trademarks can provide researchers with hints on linking technical data and customers' needs. In addition, by using word embedding results accurately, common words used by customers are expected to be quantitatively compared to technical terms in patent documents. Further, an application of word embedding to text preprocessing is expected to develop existing methods that are not robust to typos or abbreviations in social media. With such advanced techniques, social media-based BI researchers would be able to understand and deal with customers' languages more accurately.

So far, we have identified the feasibility of using social media in BI research and discussed the future directions with the experts. With our findings, this research is expected to encourage and motivate prominent researchers to participate in and further develop social media-based BI research. Indeed, we described how the social media data can be utilized in BI research and explained the data collection and analysis process, which can help new researchers who are not familiar with the social media analytics. Answers to the three RQs can extend the research scope of existing researchers in this research domain by suggesting new methodologies or data platforms. In addition, this paper can help the experts or practitioners in a real-life business environment to grasp the latest trend and seek scientific approaches in this domain. The detailed comparison of the methodologies or frameworks allows experts to find and adopt the appropriate ones for their business environment. This paper is expected to provide many academic and practical contributions, but there are some limitations. First, although we selected the main academic search engine for collecting the related papers, we cannot guarantee that the entire social media-based BI research papers were included and analyzed for a systematic review. In addition, we excluded the conference papers since their referred journal article would represent state-of-the-art results with high impact. However, since there are some conference papers with significant impacts (Asur & Huberman, 2010), various search engines and conference papers could be used in future works for identifying more research articles. Also, a concept-centric approach could be applied to social media-based BI research articles by analyzing their cited papers with a larger analysis period and associating them with various business theories. In addition, more appropriate research questions need to be added to analyze Group

2 articles from a more diverse perspective. If these points are addressed well in future studies, more accurate understanding of this research domain can be achieved. Lastly, there are some aspects of social media-based BI research not covered in this paper, such as the language of social media and related business domains. The collected articles analyzed customers' words on various business domains such as tourism, medicine, hotel, movie, electronics, technology, and food service, which were not mainly covered in this paper. In addition, since the level of business domain associated with social media data varies by social media type or platform, a quantitative approach for identifying the business domain that each data refers to from various social media platforms should be developed in future studies. As such, analyzing customer characteristics or identifying social media platforms by business domain would be possible. Also, the language of social media needs to be analyzed since it could provide quality information from a national, regional or individual user perspective. Most of the articles collected in this paper have used social media written in English or Chinese and future works could collect research articles in various languages and analyze the characteristics of social media data by language.

CRediT authorship contribution statement

Jaewoong Choi: Investigation, Writing - original draft. **Janghyeok Yoon:** Writing - review & editing. **Jaemin Chung:** Investigation, Resources. **Byoung-Youl Coh:** Project administration. **Jae-Min Lee:** Project administration.

Acknowledgement

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07045768); and the Korea Institute of Science and Technology Information (KISTI) in 2018.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ipm.2020.102279.

Appendix A

Table A1
The list of the primary articles in this systematic review.

Reference	Title
(Akay and Erlandsson, 2015a)	Network-Based Modeling and Intelligent Data Mining of Social Media for Improving Care
(Akay, Dragomir, et al., 2015b)	A Novel Data-Mining Approach Leveraging Social Media to Monitor Consumer Opinion of Sitagliptin
(Brandt, Bendler, et al., 2017)	Social media analytics and value creation in urban smart tourism ecosystems
(Burnap, Rana, et al., 2015)	Detecting tension in online communities with computational Twitter analysis
(Chang, Ku, et al., 2017)	Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor
(Chong, Li, et al., 2016)	Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach
(Dai & Hao, 2017)	Mining social media data for opinion polarities about electronic cigarettes
(Duan, Yu, et al., 2016)	Exploring the Impact of Social Media on Hotel Service Performance: A Sentimental Analysis Approach
(Fang, Sang, et al., 2014)	Topic-Sensitive Influencer Mining in Interest-Based Social Media Networks via Hypergraph Learning
(Filieri, McLeay, et al., 2018)	Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services
(Govindaraj & Gopalakrishnan, 2016)	Intensified Sentiment Analysis of Customer Product Reviews Using Acoustic and Textual Features
(Gui, Zhou, et al., 2017)	Learning representations from heterogeneous network for sentiment classification of product reviews
(He, Tian, et al., 2017)	Application of social media analytics: a case of analyzing online hotel reviews
(Ho, Damien, et al., 2017)	The time-varying nature of social media sentiments in modeling stock returns
(Hu, Chen, et al., 2017)	The effect of user-controllable filters on the prediction of online hotel reviews
(Jeong, Yoon, et al., 2018)	Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis
(Jiang, Liang, et al., 2014)	Analyzing market performance via social media: a case study of a banking industry crisis
(Jin, Ji, et al., 2016)	Identifying comparative customer requirements from product online reviews for competitor analysis
(Kang & Park, 2014)	Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach
(Kangale, Kumar, et al., 2016)	Mining consumer reviews to generate ratings of different product attributes while producing feature-based review-summary
(Kim & Lee, 2017)	Stochastic service life cycle analysis using customer reviews
(Ko, Jeong, et al., 2018)	Identifying Product Opportunities Using Social Media Mining: Application of Topic Modeling and Chance Discovery Theory
(Li, Xie, et al., 2018)	Identifying and monitoring the development trends of emerging technologies using patent analysis and Twitter data mining: The case of perovskite solar cell technology

(continued on next page)

Table A1 (*continued*)

Reference	Title
(Liu, Bi, et al., 2017)	Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory
(Lo, Chiong, et al., 2017)	An unsupervised multilingual approach for online social media topic identification
(Miah, Vu, et al., 2017)	A Big Data Analytics Method for Tourist Behaviour Analysis
(Micu, Micu, et al., 2017)	Analyzing user sentiment in social media: Implications for online marketing strategy
(Mummalaneni, Gruss, et al., 2018)	Social media analytics for quality surveillance and safety hazard detection in baby cribs
(Nave, Rita, et al., 2018)	A decision support system framework to track consumer sentiments in social media
(Nguyen, Shirai, et al., 2015)	Sentiment analysis on social media for stock movement prediction
(Nikfarjam et al., 2015)	Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features
(Ouyang, Guo, et al., 2017)	SentiStory: multi-grained sentiment analysis and event summarization with crowdsourced social media data
(Qi, Zhang, et al., 2016)	Mining customer requirements from online reviews: A product improvement perspective
(Reddick, Chatfield, et al., 2017)	A social media text analytics framework for double-loop learning for citizen-centric public services: A case study of a local government Facebook use
(Salehan & Kim, 2016)	Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics
(Sánchez-Franco, Navarro-García, et al., 2016)	Online Customer Service Reviews in Urban Hotels: A Data Mining Approach
(Shirdastian, 2017)	Using big data analytics to study brand authenticity sentiments: The case of Starbucks on Twitter
(Song, Lee, et al., 2016)	Diagnosing service quality using customer reviews: an index approach based on sentiment and gap analyses
(Stavrianou & Brun, 2015)	Expert recommendations based on opinion mining of user-generated product reviews
(Sul, Dennis, et al., 2017)	Trading on Twitter: Using Social Media Sentiment to Predict Stock Returns
(Sun, Lachanski, et al., 2016)	Trade the tweet: Social media text mining and sparse matrix factorization for stock market prediction
(Sun & Lee, 2017)	Tour recommendations by mining photo sharing social media
(Sun, Zhang, et al., 2018)	Detecting users' anomalous emotion using social media for business intelligence
(Sussman, Garcia, et al., 2014)	Consumers' perceptions of vape shops in Southern California: an analysis of online Yelp reviews
(Trappey, Trappey, et al., 2018)	Consumer driven product technology function deployment using social media and patent mining
(Tuarob & Tucker, 2015)	Automated Discovery of Lead Users and Latent Product Features by Mining Large Scale Social Media Networks
(Wang, Chen, et al., 2018)	Game-Theoretic Cross Social Media Analytic: How Yelp Ratings Affect Deal Selection on Groupon?
(Wang, Wang, et al., 2017)	Ranking product aspects through sentiment analysis of online reviews
(Wang, Feng, et al., 2018)	Topic analysis of online reviews for two competitive products using latent Dirichlet allocation
(Wong, Chan, et al., 2017)	An ANN-based approach of interpreting user-generated comments from social media
(Xie, Li, et al., 2014)	Mining Latent User Community for Tag-Based and Content-Based Search in Social Media
(Xie and Zeng, 2017)	Mining e-cigarette adverse events in social media using Bi-LSTM recurrent neural network with word embedding representation
(Xu, Wang, et al., 2017)	Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors
(Xu et al., 2018)	Detecting bursts in sentiment-aware topics from social media
(Yang, Geng, et al., 2016)	A web sentiment analysis method on fuzzy clustering for mobile social media users
(Yoo, Song, et al., 2018)	Social media contents based sentiment analysis and prediction system
(Zhao, Wang, et al., 2016)	Mining Product Adopter Information from Online Reviews for Improving Product Recommendation

Table A.2
Open API services of each social media platform.

Platforms	Open API service
Twitter	https://developer.twitter.com/en/docs.html
Facebook	https://developers.facebook.com/
Amazon	https://developer.amazon.com/
Reddit	https://www.reddit.com/dev/api
Yelp	https://www.yelp.com/developers
TripAdvisor	https://developer-tripadvisor.com/content-api/
YouTube	https://developers.google.com/youtube/v3/
Flickr	https://www.flickr.com/services/api/
Booking.com	https://developers.booking.com/api/index.html
Glassdoor	https://www.glassdoor.com/developer/index.htm
Pinterest	https://developers.pinterest.com/
Instagram	https://www.instagram.com/developer/
Trustpilot	https://developers.trustpilot.com/
Tumblr	https://www.tumblr.com/docs/en/api/v2
WordPress	https://developer.wordpress.org/rest-api/
Blogger	https://developers.google.com/blogger/
Snapchat	https://developers.snapchat.com/

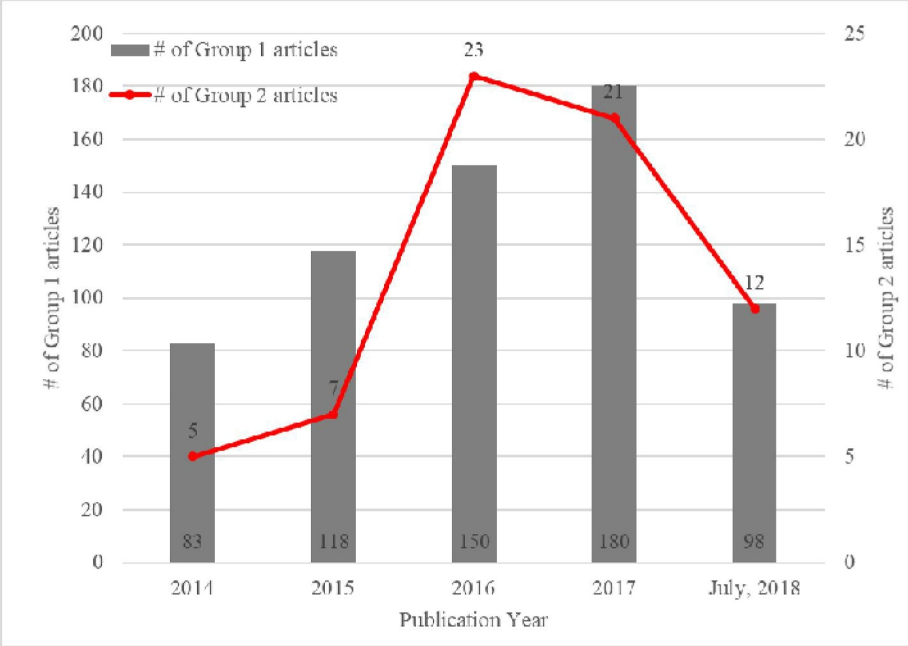

Fig. 1. Annual trends of social media-based BI research article (Group 1)

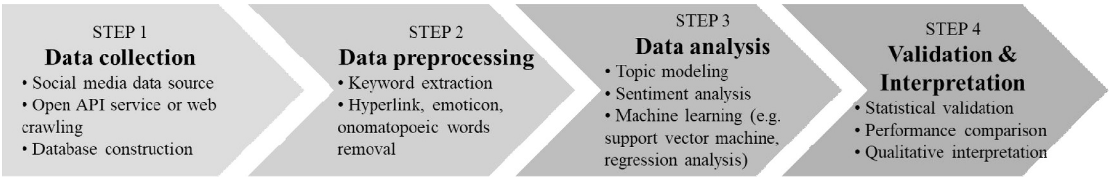

Fig. 2. Framework of social media-based BI research

References

Akay, Dragomir, & Erlandsson (2015a). Network-based modeling and intelligent data mining of social media for improving care. *IEEE journal of biomedical and health informatics, 19*, 210–218.

Akay, Dragomir, & Erlandsson (2015b). A novel data-mining approach leveraging social media to monitor consumer opinion of sitagliptin. *IEEE journal of biomedical and health informatics, 19*, 389–396.

Alberani, Pietrangeli, & Mazza (1990). The use of grey literature in health sciences: a preliminary survey. *Bulletin of the Medical Library Association, 78*, 358.

Alkhodair, Ding, Fung, & Liu (2019). Detecting breaking news rumors of emerging topics in social media. *Information Processing&Management*.

Asur, & Huberman (2010). *Predicting the future with social media. 2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology, 1*, IEEE492–499.

Attard, Orlandi, Scerri, & Auer. (2015). A systematic review of open government data initiatives. *Government Information Quarterly, 32*, 399–418.

Balbi, Misuraca, & Scepi (2018). Combining different evaluation systems on social media for measuring user satisfaction. *Information Processing&Management, 54*, 674–685.

Bashir, Papamichail, & Malik (2017). Use of Social Media Applications for Supporting New Product Development Processes in Multinational Corporations. *Technological Forecasting and Social Change, 120*, 176–183.

Blei, Ng, & Jordan (2003). Latent dirichlet allocation. *Journal of machine Learning research, 3*, 993–1022.

Bradbury (2011). In plain view: open source intelligence. *Computer Fraud & Security, 2011*, 5–9.

Brandt, Bendler, & Neumann (2017). Social media analytics and value creation in urban smart tourism ecosystems. *Information & Management, 54*, 703–713.

Burnap, Rana, Avis, Williams, Housley, Edwards, Morgan, & Sloan (2015). Detecting tension in online communities with computational Twitter analysis. *Technological Forecasting and Social Change, 95*, 96–108.

Cadwalladr, & Graham-Harrison (2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The guardian, 17*, 22.

Chan, & Ngai (2011). Conceptualising electronic word of mouth activity: An input-process-output perspective. *Marketing Intelligence and Planning, 29*, 488–516.

Chang, Ku, & Chen (2017). Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. *International Journal of Information Management*.

Chan-Olmsted (2002). Branding and Internet marketing in the age of digital media. *Journal of Broadcasting & Electronic Media, 46*, 641–645.

Chaudhuri, Dayal, & Narasayya (2011). An overview of business intelligence technology. *Communications of the ACM, 54*, 88–98.

Chen, Chiang, & Storey (2012). Business intelligence and analytics: from big data to big impact. *MIS quarterly*, 1165–1188.

Chong, Li, Ngai, Ch'ng, & Lee (2016). Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach. *International Journal of Operations & Production Management, 36*, 358–383.

Chung, & Tseng (2012). Discovering business intelligence from online product reviews: A rule-induction framework. *Expert systems with applications, 39*, 11870–11879.

Conn, Valentine, Cooper, & Rantz (2003). Grey literature in meta-analyses. *Nursing research, 52*, 256–261.

Cvijikj, & Michahelles (2011). *Monitoring trends on facebook. 2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing*. IEEE895–902.
Dai, & Hao (2017). Mining social media data for opinion polarities about electronic cigarettes. *Tobacco control, 26*, 175–180.
Davenport (2006). Competing on analytics. *harvard business review, 84*, 98.
Duan, Yu, Cao, & Levy (2016). Exploring the impact of social media on hotel service performance: A sentimental analysis approach. *Cornell Hospitality Quarterly, 57*, 282–296.
Fang, Sang, Xu, & Rui (2014). Topic-sensitive influencer mining in interest-based social media networks via hypergraph learning. *IEEE Trans. Multimedia, 16*, 796–812.
Filieri, McLeay, Tsui, & Lin (2018). Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services. *Information & Management*.
Gallaugher, & Ransbotham (2010). Social media and customer dialog management at Starbucks. *MIS Quarterly Executive, 9*.
Gerken, Moehrle, & Walter (2015). One year ahead! Investigating the time lag between patent publication and market launch: insights from a longitudinal study in the automotive industry. *R&D Management, 45*, 287–303.
Gibson (2004). Open source intelligence: An intelligence lifeline. *The RUSI Journal, 149*, 16–22.
Govindaraj, & Gopalakrishnan (2016). Intensified Sentiment Analysis of Customer Product Reviews Using Acoustic and Textual Features. *ETRI Journal, 38*, 494–501.
Gui, Zhou, Xu, He, & Lu (2017). Learning representations from heterogeneous network for sentiment classification of product reviews. *Knowledge-Based Systems, 124*, 34–45.
Gurstein (2011). Open data: Empowering the empowered or effective data use for everyone. *First Monday, 16*.
Guzman, & Maalej (2014). *How do users like this feature? a fine grained sentiment analysis of app reviews. 2014 IEEE 22nd international requirements engineering conference (RE)*. IEEE153–162.
He, Tian, Tao, Zhang, Yan, & Akula (2017). Application of social media analytics: a case of analyzing online hotel reviews. *Online Information Review, 41*, 921–935.
He, Zha, & Li (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management, 33*, 464–472.
Ho, Damien, Gu, & Konana (2017). The time-varying nature of social media sentiments in modeling stock returns. *Decision Support Systems, 101*, 69–81.
Hu, Chen, & Lee (2017). The effect of user-controllable filters on the prediction of online hotel reviews. *Information & Management, 54*, 728–744.
Huang, Liu, & Nguyen (2015). *Location-based event search in social texts. 2015 International Conference on Computing, Networking and Communications (ICNC)*. IEEE668–672.
Huang, Schuehle, Porter, & Youtie (2015). A systematic method to create search strategies for emerging technologies based on the Web of Science: illustrated for 'Big Data'. *Scientometrics, 105*, 2005–2022.
Jeong, Yoon, & Lee (2018). Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *International Journal of Information Management*.
Jiang, Liang, Chen, & Ding (2014). Analyzing market performance via social media: a case study of a banking industry crisis. *Science China Information Sciences, 57*, 1–18.
Jimenez-Marquez, Gonzalez-Carrasco, Lopez-Cuadrado, & Ruiz-Mezcua (2019). Towards a big data framework for analyzing social media content. *International Journal of Information Management, 44*, 1–12.
Jin, Ji, & Gu (2016). Identifying comparative customer requirements from product online reviews for competitor analysis. *Engineering Applications of Artificial Intelligence, 49*, 61–73.
Best Jr, & Cumming. (2007). *Open source intelligence (OSINT): issues for congress*: Dec.
Kajikawa, Yoshikawa, Takeda, & Matsushima (2008). Tracking emerging technologies in energy research: Toward a roadmap for sustainable energy. *Technological Forecasting and Social Change, 75*, 771–782.
Kang, & Park (2014). Review based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. *Expert systems with applications, 41*, 1041–1050.
Kangale, Kumar, Naeem, Williams, & Tiwari (2016). Mining consumer reviews to generate ratings of different product attributes while producing feature-based review-summary. *International Journal of Systems Science, 47*, 3272–3286.
Kim (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.
Kim, & Hastak (2018). Social network analysis: Characteristics of online social networks after a disaster. *International Journal of Information Management, 38*, 86–96.
Kim, & Lee (2017). Stochastic service life cycle analysis using customer reviews. *The Service Industries Journal, 37*, 296–316.
Kitchenham (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University, 33*, 1–26.
Ko, Jeong, Choi, & Yoon (2018). Identifying Product Opportunities Using Social Media Mining: Application of Topic Modeling and Chance Discovery Theory. *IEEE Access, 6*, 1680–1693.
Krishnamurthy, & Awazu (2016). Liberating data for public value: The case of Data. gov. *International Journal of Information Management, 36*, 668–672.
Lample, Ballesteros, Subramanian, Kawakami, & Dyer. (2016). Neural architectures for named entity recognition. *arXiv preprint arXiv:1603.01360*.
Lee, & Lee (2017). Identifying new business opportunities from competitor intelligence: An integrated use of patent and trademark databases. *Technological Forecasting and Social Change, 119*, 170–183.
Li, Xie, Jiang, Zhou, & Huang (2018). Identifying and monitoring the development trends of emerging technologies using patent analysis and Twitter data mining: The case of perovskite solar cell technology. *Technological Forecasting and Social Change*.
Liu, Bi, & Fan. (2017). Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory. *Information Fusion, 36*, 149-161.
Lo, Chiong, & Cornforth (2017). An unsupervised multilingual approach for online social media topic identification. *Expert systems with applications, 81*, 282–298.
Miah, Vu, Gammack, & McGrath (2017). A big data analytics method for tourist behaviour analysis. *Information & Management, 54*, 771–785.
Micu, Micu, Geru, & Lixandroiu (2017). Analyzing user sentiment in social media: Implications for online marketing strategy. *Psychology & Marketing, 34*, 1094–1100.
Mummalaneni, Gruss, Goldberg, Ehsani, & Abrahams (2018). Social media analytics for quality surveillance and safety hazard detection in baby cribs. *Safety science, 104*, 260–268.
Nave, Rita, & Guerreiro (2018). A decision support system framework to track consumer sentiments in social media. *Journal of Hospitality Marketing & Management*, 1–18.
Nguyen, Shirai, & Velcin (2015). Sentiment analysis on social media for stock movement prediction. *Expert systems with applications, 42*, 9603–9611.
Nikfarjam, Sarker, O'connor, Ginn, & Gonzalez (2015). Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *Journal of the American Medical Informatics Association, 22*, 671–681.
Olanrewaju, Hossain, Whiteside, & Mercieca (2020). Social media and entrepreneurship research: A literature review. *International Journal of Information Management, 50*, 90–110.
Ouyang, Guo, Zhang, Yu, & Zhou (2017). SentiStory: multi-grained sentiment analysis and event summarization with crowdsourced social media data. *Personal and Ubiquitous Computing, 21*, 97–111.
Pang, & Lee (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval,, 2*, 1–135.
Park, Conway, & Chen (2018). Examining thematic similarity, difference, and membership in three online mental health communities from Reddit: a text mining and visualization approach. *Computers in human behavior, 78*, 98–112.
Park, & Yoon (2017). Application technology opportunity discovery from technology portfolios: Use of patent classification and collaborative filtering. *Technological Forecasting and Social Change, 118*, 170–183.
Porter, Garner, Carley, & Newman (2018). Emergence scoring to identify frontier R&D topics and key players. *Technological Forecasting and Social Change*.
Qi, Zhang, Jeon, & Zhou (2016). Mining customer requirements from online reviews: A product improvement perspective. *Information & Management, 53*, 951–963.
Reddick, Chatfield, & Ojo (2017). A social media text analytics framework for double-loop learning for citizen-centric public services: A case study of a local government Facebook use. *Government Information Quarterly, 34*, 110–125.

Rehman, ur, Chang, Batool, & Wah (2016). Big data reduction framework for value creation in sustainable enterprises. *International Journal of Information Management, 36*, 917–928.
Rose, Hair, & Clark (2011). Online customer experience: A review of the business-to-consumer online purchase context. *International Journal of Management Reviews, 13*, 24–39.
Rust, Moorman, & Bhalla (2010). Rethinking marketing. *harvard business review, 88*, 94–101.
Salehan, & Kim (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems, 81,* 30–40.
Salo (2017). Social media research in the industrial marketing field: Review of literature and future research directions. *Industrial Marketing Management, 66*, 115–129.
Sánchez-Franco, Navarro-García, & Rondán-Cataluña (2016). Online customer service reviews in urban hotels: A data mining approach. *Psychology & Marketing, 33*, 1174–1186.
Shiau, Dwivedi, & Lai (2018). Examining the core knowledge on facebook. *International Journal of Information Management, 43,* 52–63.
Shirazi, & Mohammadi (2018). A big data analytics model for customer churn prediction in the retiree segment. *International Journal of Information Management*.
Shirdastian, Laroche, & Richard (2017). Using big data analytics to study brand authenticity sentiments: The case of Starbucks on Twitter. *International Journal of Information Management*.
Song, Lee, Yoon, & Park (2016). Diagnosing service quality using customer reviews: an index approach based on sentiment and gap analyses. *Service Business, 10,* 775–798.
Statistica. (2019a). Social media - Statistics & Facts. In.
Statistica. (2019b). Social media usage in the United States - Statistics & Facts.
Stavrianou, & Brun (2015). Expert recommendations based on opinion mining of user-generated product reviews. *Computational Intelligence, 31*, 165–183.
Steele (2007). Open source intelligence. *Handbook of intelligence studies,* 129–147.
Stieglitz, Mirbabaie, Ross, & Neuberger (2018). Social media analytics–Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management, 39*, 156–168.
Sul, Dennis, & Yuan (2017). Trading on twitter: Using social media sentiment to predict stock returns. *Decision Sciences, 48,* 454–488.
Sun, Lachanski, & Fabozzi (2016). Trade the tweet: Social media text mining and sparse matrix factorization for stock market prediction. *International Review of Financial Analysis, 48*, 272–281.
Sun, & Lee (2017). Tour recommendations by mining photo sharing social media. *Decision Support Systems, 101*, 28–39.
Sun, Wang, Cheng, & Fu (2015). Mining affective text to improve social media item recommendation. *Information Processing&Management, 51*, 444–457.
Sun, Zhang, Li, Sun, Ren, Zomaya, & Ranjan (2018). Detecting users' anomalous emotion using social media for business intelligence. *Journal of Computational Science, 25*, 193–200.
Sussman, Garcia, Cruz, Baezconde-Garbanati, Pentz, & Unger (2014). Consumers' perceptions of vape shops in Southern California: an analysis of online Yelp reviews. *Tobacco induced diseases, 12*, 22.
Tietze, Reul, & Herstatt (2009). The relation of patent ownership and firm success cases from the LCD Flat-Panel-Display industry. *International Journal of Technology Intelligence and Planning, 5*, 90–109.
Trappey, Trappey, Fan, & Lee (2018). *Consumer driven product technology function deployment using social media and patent mining, 36*, 120–129.
Tuarob, & Tucker (2015). Automated discovery of lead users and latent product features by mining large scale social media networks. *Journal of Mechanical Design, 137*, Article 071402.
Wallace, Van Fleet, & Downs (2011). The research core of the knowledge management literature. *International Journal of Information Management, 31*, 14–20.
Wang, Chen, & Liu (2018). Game-Theoretic Cross Social Media Analytic: How Yelp Ratings Affect Deal Selection on Groupon. *IEEE Transactions on Knowledge and Data Engineering, 30*, 908–921.
Wang, Feng, & Dai (2018). Topic analysis of online reviews for two competitive products using latent Dirichlet allocation. *Electronic Commerce Research and Applications, 29*, 142–156.
Wang, Porter, Wang, & Carley (2018). An approach to identify emergent topics of technological convergence: A case study for 3D printing. *Technological Forecasting and Social Change*.
Wang, Wang, & Song (2017). Ranking product aspects through sentiment analysis of online reviews. *Journal of Experimental & Theoretical Artificial Intelligence, 29*, 227–246.
Webster, & Watson (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly* xiii-xxiii.
Wong, Chan, & Lacka (2017). An ANN-based approach of interpreting user-generated comments from social media. *Applied Soft Computing, 52*, 1169–1180.
Xiao, Wei, & Dong (2016). Crowd intelligence: Analyzing online product reviews for preference measurement. *Information & Management, 53*, 169–182.
Xie, Li, Mao, Li, Cai, & Zheng (2014). Mining latent user community for tag-based and content-based search in social media. *The Computer Journal, 57*, 1415–1430.
Xie, Liu, & Zeng, Dajun (2017). Mining e-cigarette adverse events in social media using Bi-LSTM recurrent neural network with word embedding representation. *Journal of the American Medical Informatics Association, 25*, 72–80.
Xu, Qi, Huang, Wu, & Fu (2018). Detecting bursts in sentiment-aware topics from social media. *Knowledge-Based Systems, 141*, 44–54.
Xu, Wang, Li, & Haghighi (2017). Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors. *International Journal of Information Management, 37*, 673–683.
Yang, Geng, & Liao (2016). A web sentiment analysis method on fuzzy clustering for mobile social media users. *EURASIP Journal on Wireless Communications and Networking, 2016,* 128.
Yang, & Lee (2012). Mining open source text documents for intelligence gathering. *Information Technology in Medicine and Education (ITME), 2012, 2*, 969–973 *International Symposium on*.
Yoo, Song, & Jeong (2018). Social media contents based sentiment analysis and prediction system. *Expert systems with applications,, 105*, 102–111.
Yoon, & Kim (2011). Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks. *Scientometrics, 88,* 213–228.
Yoon, & Kim (2012). Detecting signals of new technological opportunities using semantic patent analysis and outlier detection. *Scientometrics, 90,* 445–461.
Yoon, Park, Seo, Lee, Coh, & Kim (2015). Technology opportunity discovery (TOD) from existing technologies and products: A function-based TOD framework. *Technological Forecasting and Social Change, 100,* 153–167.
Zhao, Wang, He, Wen, Chang, & Li (2016). Mining product adopter information from online reviews for improving product recommendation. *ACM Transactions on Knowledge Discovery from Data (TKDD), 10,* 29.