



Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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

From technological development to social advance: A review of Industry 4.0 through machine learning

Changhun Lee¹, Chiehyeon Lim^{1,2,*}¹ Department of Industrial Engineering, Ulsan National Institute of Science and Technology, 50 Unist-gil, Eonyang-eup, Ulju-gun, Ulsan 44919, South Korea² Graduate School of Artificial Intelligence, Ulsan National Institute of Science and Technology, 50 Unist-gil, Eonyang-eup, Ulju-gun, Ulsan 44919, South Korea

ARTICLE INFO

Keywords:

Industry 4.0
Fourth industrial revolution
Review
Survey
Text mining

ABSTRACT

Industry 4.0 has attracted considerable interest from firms, governments, and individuals as the new concept of future computer, industrial, and social systems. However, the concept has yet to be fully explored in the scientific literature. Given the topic's broad scope, this work attempts to understand and clarify Industry 4.0 by analyzing 660 journal papers and 3,901 news articles through text mining with unsupervised machine learning algorithms. Based on the results, this work identifies 31 research and application issues related to Industry 4.0. These issues are categorized and described within a five-level hierarchy: 1) infrastructure development for connection, 2) artificial intelligence development for data-driven decision making, 3) system and process optimization, 4) industrial innovation, and 5) social advance. Further, a framework for convergence in Industry 4.0 is proposed, featuring six dimensions: connection, collection, communication, computation, control, and creation. The research outcomes are consistent with and complementary to existing relevant discussion and debate on Industry 4.0, which validates the utility and efficiency of the data-driven approach of this work to support experts' insights on Industry 4.0. This work helps establish a common ground for understanding Industry 4.0 across multiple disciplinary perspectives, enabling further research and development for industrial innovation and social advance.

1. Introduction

In January 2016, attendees of the 46th World Economic Forum in Davos discussed automation in industry under the theme "Industry 4.0." Researchers anticipate that connection and intelligence technologies will create a "fourth industrial revolution" and change industrial, economic, and social paradigms (Xu et al., 2018). However, despite its evident importance (Dalenogare et al., 2018; Li, 2018), the concept of "Industry 4.0" is not yet clear. What is it? What are the key issues? How can we validate its conceptualization? As of April 2020, a search for "Industry 4.0" on the Google Scholar web search engine generated more than 75,000 results across various fields including industrial engineering, computer science, electrical engineering, business, economics, and social science. Researchers such as Lasi et al. (2014), Lu (2017), Sung (2018), and Xu et al. (2018) have attempted to review and clarify the concept and scope of Industry 4.0. However, existing reviews are selective in the literature they use and are often limited in scope to accommodate the volume and variety of relevant documents.

Meanwhile, it is very difficult for human researchers themselves to review and integrate these documents.

This work develops an understanding of Industry 4.0 based on text mining of related documents. The journal papers, 660 in number, were identified as those having the keywords "Industry 4.0" or "fourth industrial revolution" from the Web of Science Core Collections, containing all available data published through September 8, 2018.¹ We applied an analytical method that incorporates metrics to measure the statistical and semantic significance of word features of data. Then, to identify patterns in the literature text corpus and to capture their essence, we used unsupervised machine learning algorithms such as spectral clustering (von Luxburg, 2007) and non-negative matrix factorization (NMF) algorithms (Lin, 2007). Spectral clustering is a graph partitioning method effective in exclusively identifying clusters that are nonlinearly separable in the original space, while NMF is a representation learning method for extracting latent factors of the original features; thus, we used the former to identify the distinct research topics in 660 articles and the latter for interpretation of the

* Corresponding author.

E-mail address: chlim@unist.ac.kr (C. Lim).

¹ This paper is based on a government project to clarify the concept of Industry 4.0 for national R&D planning that was conducted in September 2018.

<https://doi.org/10.1016/j.techfore.2021.120653>

Received 8 April 2020; Received in revised form 13 December 2020; Accepted 27 January 2021

Available online 26 February 2021

0040-1625/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

topics. Our analysis of the 660 papers identified basic statistics, significant keywords, and research topics of the literature. In addition, using the latent Dirichlet allocation (LDA) method (Blei et al., 2003), we conducted a similar analysis on 3901 news articles identified from the LexisNexis Business & Industry News Database (using the same keywords), containing all available data published through September 17, 2018. LDA is a generative probabilistic model for topic identification from collections of documents and words. The literature data include information on research topics and technology factors of Industry 4.0, whereas the news data generally describe areas of application and business aspects. Our intention was to integrate different perspectives on Industry 4.0 comprehensively by combining the two data sets.

Building on the findings quantitatively derived, we further reviewed the data manually to add human insight and develop an in-depth overview of Industry 4.0. As a result, we identified 31 research and application issues related to Industry 4.0, which we categorized into five levels: 1) infrastructure development for connection, 2) artificial intelligence development for data-driven decision making, 3) system and process optimization, 4) industrial innovation, and 5) social advance. Furthermore, we propose a framework describing the convergence in Industry 4.0 with six dimensions: connection, collection, communication, computation, control, and creation.

Establishing common ground for central concepts is essential for science (Fortunato et al., 2018). As a multidimensional agenda, Industry 4.0 continue to attract interest from practitioners and academics across a variety of disciplines. While many studies have reported technologies and applications about the agenda (e.g., Liao et al., 2017; Frank et al., 2019), theorizing and conceptualization efforts on Industry 4.0 *per se* are still relatively scarce. Our contribution is to provide a systematized view of dispersed knowledge about Industry 4.0 by integrating such knowledge into a unified conceptualization. In particular, our work is unique in that it uses a machine learning approach for comprehensive and reliable conceptualization. This helped us incorporate a broad range of studies on Industry 4.0 from technology and engineering to business and social science, thereby filling the gap between existing studies on technological development and studies on industrial innovation and social advance. This is a unique contribution of our work compared to other reviews on Industry 4.0 in manufacturing and engineering (e.g., Liao et al., 2017; Oztemel and Gursev, 2020). Fig. 1 summarizes this machine-learning-enabled contribution of our work, presenting a broader picture of research and application issues of Industry 4.0 from infrastructural technology development to social advance.

This paper is organized as follows. In Section 2, we review existing studies on Industry 4.0 to provide a foundation for this work. In Section 3, we describe the research method, which includes data collection and analysis. In Section 4, we describe the findings, showing an overview of

Industry 4.0. In Section 5, we discuss the implications of our work. Lastly, in Section 6 we conclude with some remarks.

2. Literature review

2.1. Diverse aspects of Industry 4.0

Researchers have studied and discussed diverse aspects of Industry 4.0. Fig. 2 shows the top 100 representative words of the Industry 4.0 literature from the 660 articles we collected. The degree of representation (i.e., the font size of a word) is calculated as the geometric mean of several word signification metric scores (Lim and Maglio, 2018). The figure indicates the diverse aspects of Industry 4.0. The top words associated with Industry 4.0 include “system,” “manufacturing,” “production,” “technology,” “smart,” “intelligent,” “data,” “process,” “Internet,” “network,” “sensor,” “cloud,” “security,” “factory,” “control,” “service,” “management,” “business,” “automation,” “communication,” “machine,” “innovation,” and “integration.” We did not include certain words with high scores in Fig. 2, such as “paper,” “proposed,” and “existing,” because their high scores are not attributable to their relevance to the topic but rather to the high frequency of their appearance in scientific documents in general.

The interdisciplinary nature of Industry 4.0 can also be illustrated by the publication journal information. The journals included *Technological Forecasting and Social Change*, *International Journal of Production Economics*, *International Journal of Production Research*, *International Journal of Advanced Manufacturing Technology*, *IEEE Transactions on Industrial Informatics*, *Computers in Industry*, *Applied Soft Computing*, *IEEE Internet of Things Journal*, *Future Generation Computer Systems*, *Journal of Manufacturing Systems*, and *Industrial Management & Data Systems*. See Section 3 for further details on the article information and data



Fig. 2. Top 100 representative words of the Industry 4.0 literature.

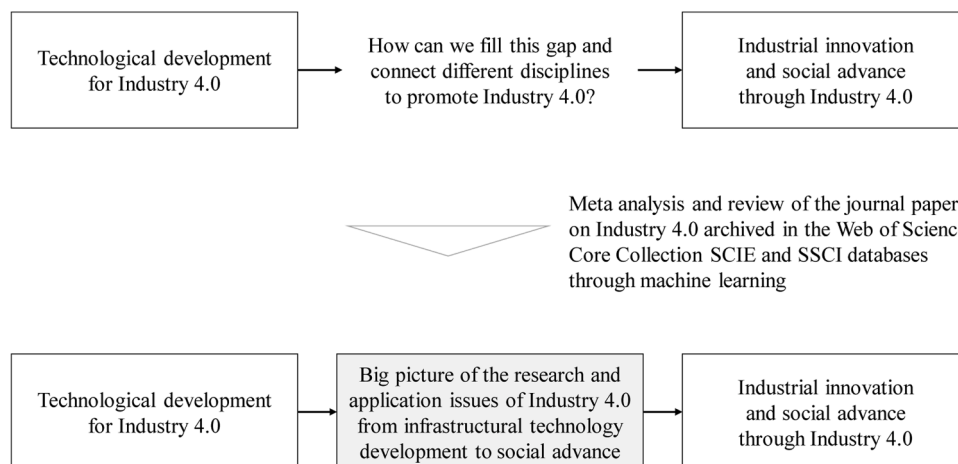


Fig. 1. Research overview.

preprocessing for creation of the word cloud.

In addition to creating Fig. 2 (quantitative analysis), we conducted a detailed qualitative review of the literature and found that the scope of Industry 4.0 involves diverse aspects (Schwab, 2017; Xu et al., 2018) such as big data (Opresnik and Taisch, 2015), optimization (Ivanov et al., 2016), artificial intelligence (Li et al., 2017), automation (Bahrin et al., 2016), wireless networks (Lin et al., 2015), the Internet of things (IoT) (Wan et al., 2016), business information systems (Lasi et al., 2014), and human-machine interaction (Gorecky et al., 2017). Related applications include cyber-physical systems (Jazdi, 2014), intelligent manufacturing (Zhong et al., 2017), sustainable manufacturing (de Sousa Jabbour et al., 2018), smart factory (Büchi et al., 2020), and smart services (Lim and Maglio, 2018; Lim et al., 2018c). Implications surrounding Industry 4.0 include organizational learning (Tortorella et al., 2020), value creation (Kagermann, 2015), innovation for small and medium-sized enterprises (Müller et al., 2018), and social innovation (Morrar et al., 2017). These are just a few examples of the trans-disciplinary nature of Industry 4.0. The Industry 4.0 concept lies at an intersection of technology development and application, firm management, industrial innovation, and social advance, among other disciplines.

2.2. Pertinent reviews of Industry 4.0

Although each study introduced in the previous section represents the specific perspectives of the respective authors, a multi-perspective study is required to understand and describe Industry 4.0. This is because Industry 4.0 is a broad agenda rather than a specific issue. As such, its technologies and applications have been surveyed by several researchers, who discuss challenges and research issues. For example, Liao et al. (2017) conducted a systematic literature review through the process of paper identification, screening, eligibility assessment, and quantitative and qualitative analyses of selected papers. Their analysis outcomes include the categorization of journal and conference papers on Industry 4.0, identification of frequent words related to Industry 4.0, identification of main research issues and application fields of Industry 4.0, and suggestion of a research agenda. Similarly, Lu (2017) surveyed technology and application issues of Industry 4.0 through a systematic literature review. This work particularly focused on the cyber-physical systems and interoperability of Industry 4.0, as well as the enabling technologies such as IoT, big data, cloud computing, and mobile computing. The three main application areas of Industry 4.0 discussed in this work were smart factories, smart products, and smart cities. Other similar reviews published later include Dohale and Kumar (2018), who introduced various studies on Industry 4.0 under the categories of conceptual, empirical, review, and case study articles, and Oztemel and Gursev (2020), who introduced national initiatives related to Industry 4.0 and worldwide Industry 4.0 projects, as well as to the components of Industry 4.0 including machine-to-machine communication, enterprise resource planning systems, cloud systems, data mining, and augmented reality.

Although the aforementioned reviews cover broad contexts of Industry 4.0, several other reviews mainly focus on specific contexts. For example, Alcácer and Cruz-Machado (2019) focus on the key technologies for manufacturing systems in the context of Industry 4.0. The key technologies they categorize include industrial IoT, cloud computing, big data, simulation, augmented reality, additive manufacturing (three-dimensional [3D] printing), systems integration, autonomous robots, and cybersecurity. The authors state that these technologies are integrated for use in developing smart factories that include cyber-physical systems and internet of services. Similarly, Adamson et al. (2017) focused on the concept of cloud manufacturing, while Zhong et al. (2017) reviewed studies on intelligent manufacturing in the context of Industry 4.0 and investigated related efforts worldwide. Meanwhile, Damiani et al. (2018) specifically focused on the augmented and virtual reality technologies and their applications in Industry 4.0.

The key functions of these technologies include tracking, positioning, object detection, recognition, and display interaction. The key application areas include management of production systems and virtual training with the advantages of production error diagnostics, safety management of production systems, and improvement of cooperation between humans and machine. Mourtzis (2020) show the state of the art and new trends in the simulation for the design and operation of manufacturing systems. The author conducted a systematic literature review with the process of paper search from multiple databases, identification of relevant papers, and full-text reading and grouping into research topics. The analysis findings include the historical evolution of simulation, categorization of product and production lifecycle simulation tools, and identification of topics regarding the digitalization in simulation toward Industry 4.0. The author also presents future challenges posed by the new simulation approaches based on the findings from main analyses. Finally, Aulbur et al. (2016) focus on the national level of Industry 4.0 and analyze the skill development for Industry 4.0 in BRICS and other developed nations such as USA, Germany, Korea, and Japan.

Although existing reviews provide a variety of key information and insights on Industry 4.0, most reviews are selective in the literature they use and may be limited in scope to accommodate the volume and variety of sources. Specifically, existing reviews mainly focus on manufacturing or engineering in general, and there is no study yet that provides a full-range overview of the research and application issues of Industry 4.0 from manufacturing and engineering to business and social science, despite the merit of understanding the whole spectrum of Industry 4.0 as a basis for understanding and promoting it from technological development to social advance. In addition, given the variety and volume of documents on Industry 4.0, a quantitative method is necessary to understand Industry 4.0, as no human researcher may be able to review and integrate all these documents. However, as yet, there is no such work. Therefore, as described in Section 1, our objective is to develop a unified understanding of Industry 4.0 from thousands of scientific and news articles in the fields of industrial engineering, computer science, electrical engineering, business, economics, social science, among others, based on a data-driven machine learning approach to reduce subjectivity and increase inclusiveness. Furthermore, we also conduct a qualitative review of the key papers exemplified in this section and combine it with the quantitatively derived findings to identify future challenges in Industry 4.0, similar to the approach of Mourtzis (2020).

3. Method

Developing a unified conceptualization of Industry 4.0 is not an easy task given the variety and volume of relevant studies. Text mining is a process by which previously unknown knowledge is discovered from a large quantity of textual data (Bird et al., 2009). Text mining methods and machine learning algorithms have been used for many purposes on a wide range of data types, such as for analyzing technology trends with patent data (Yoon and Park, 2004), for understanding customers using hotel review data (Mankad et al., 2016), and for describing specific research fields with scientific documents (Lim and Maglio, 2018). Text mining is an appropriate method for achieving our review objective because we aim to explore aspects and areas of Industry 4.0 in a comprehensive manner, and such work is difficult to perform manually. In addition, despite insights from experts, subjective descriptions of Industry 4.0 can be difficult to evaluate. A data-driven approach, such as text mining with metrics and machine learning algorithms, can be an excellent alternative (Blei et al., 2003). Moreover, an expert's analysis built upon a machine's findings from massive quantities of data often generates rich insights and uncovers new implications (Ordenes et al., 2014). Therefore, we collected and analyzed a comprehensive set of 660 journal articles as shown in Fig. 3. We used the Python numpy, scipy, pandas, nltk, scikit-learn, and matplotlib libraries in this process.

We identified and downloaded journal papers related to Industry 4.0

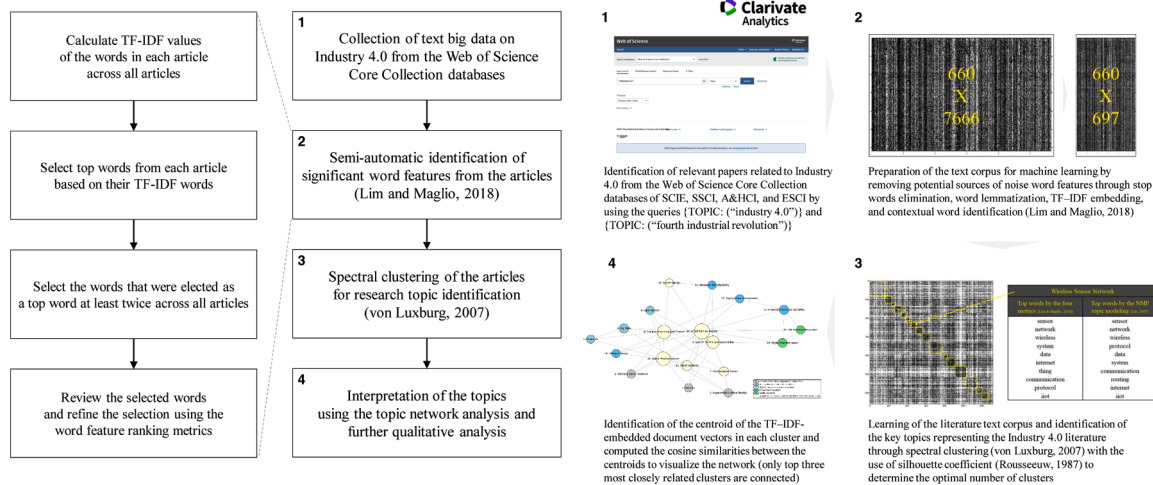


Fig. 3. Overview of the process for text mining of the Industry 4.0 literature.

from the Web of Science Core Collection databases of Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI), Arts & Humanities Citation Index (A&HCI), and Emerging Sources Citation Index (ESCI) by using the queries {TOPIC: (Industry 4.0)} and {TOPIC: (“fourth industrial revolution”)}. The Web of Science Core Collection indexes high-quality papers, and all of the papers found involved features of Industry 4.0. We collected only the “research article” type of paper because other types of papers, such as reviews and editorials, may be too broad and may generate noise in learning specific topics and characteristics of the literature. Unlike existing reviews of selected sample papers, our review involved collecting and analyzing full population data from the databases in a semiautomatic manner. We downloaded 684 articles initially but removed 24 as being duplicates or incomplete, thus resulting in a text corpus of 660 articles for analysis.

Text preprocessing (Step 2 in Fig. 3) aimed to prepare the text corpus for subsequent analysis by removing potential sources of noise. This step involved eliminating “stop words” (e.g., “it,” “and,” and “for”), any words containing letters not found in the alphabet, changing all text to lowercase (e.g., from “Industry” to “industry”), lemmatizing all words (e.g., from “processes” to “process”), as well as applying customized rules (e.g., we deleted words that are commonly used in journal articles but that are non-contextual, such as “paper” and “result”). We then focused on word feature selection, after creating the document-term matrix with term frequency (TF)-inverse document frequency (IDF) (or collectively, TF-IDF) embedding. Typically, a word feature can be either one singular word or a set of multiple words. This study focuses on the one-word feature case that measures the feature value using TF or a TF-IDF measure. It should be noted that TF alone cannot reflect the feature value of a datum meaningfully in cases that the feature value should represent a relative position in a datum compared with the values in other data. Thus, many text mining-based studies employ a TF-IDF measure instead of simple frequency (Wu et al., 2008). The basic TF-IDF value is calculated by multiplying TF and IDF, a logarithmically scaled fraction of the documents that contain the word. In short, the value of TF-IDF increases as the appearance of a word in a datum increases and as the word appears occasionally in other data. Variations of the TF-IDF calculation method are often used to reflect the intention of the analysts (e.g., to smoothen the effect of the document frequency value).

Words in a corpus of journal articles can be categorized into three types: 1) specialized words developed by the authors or specific companies, such as acronyms and application names; 2) contextual words relevant to the topic, such as “sensor,” “network,” and “smart”; and 3) general words frequently found in scientific publications, such as “method” and “approach,” and in English documents, such as “within” and “over.” Word-feature selection requires the inclusion of Type 2

words and the exclusion of Type 1 and Type 3 words. Type 1 words tend to have very high TF-IDF values because they appear several times in an item yet are not present in other items. Type 3 words have low values for the opposite reason. Type 2 words, which appear several times in many data items, generally have high TF-IDF values. The process of identifying several meaningful keywords (Parameter 1) from each datum using the TF-IDF measure is useful in filtering out many Type 3 words (i. e., overly general words). Given the definition of the TF-IDF measure, such identified keywords distinguish the datum from other data in the same corpus. Selecting the keywords that were elected as a top word at least several times (Parameter 2) across the data set is useful in excluding Type 1 words (i.e., extremely case-sensitive words). An issue of this word feature reduction process is deciding the parameter values. Using the algorithm and word significance metrics proposed by Lim and Maglio (2018), we found the optimal values of Parameters 1 and 2 to be six and two, respectively. Accordingly, we eliminated Type 1 and Type 3 words from the original data set and identified Type 2 words that represent the Industry 4.0 literature. As the final result, 697 word features were selected from the original 7666 features.

For learning the literature text corpus and identifying key topics representing the Industry 4.0 literature (Step 3 in Fig. 3), we applied spectral clustering (von Luxburg, 2007), which is based on graph partitioning and uses the Laplacian matrix derived from a similarity matrix of data. This algorithm was chosen based on the results of pilot studies that involved testing the performance of various clustering algorithms on our data, including spectral clustering, affinity propagation clustering (Frey and Dueck, 2007), density-based spatial clustering of applications with noise (Birant and Kut, 2007), and k-means clustering (Hartigan and Wong, 1979) on the reduced data after principal component analysis (Jolliffe, 2002). The mean of the silhouette coefficients (Rousseeuw, 1987) of the entire data set was our performance metric for algorithm testing. The graph-partitioning problem in spectral clustering is an NP-hard problem, requiring the use of a heuristic algorithm, meaning that the clustering result changes with each run and it is difficult to determine an optimal number of clusters. We used the mean of the silhouette coefficient values to determine the number of clusters in the text corpus. We checked the average of 100 iterations for each of 40 cases (i.e., testing for the right number of clusters from 1 cluster to 40 clusters) based on the observation that the clustering quality decreases monotonously from the number 31. We also reviewed the result of each case manually.

Finally, we concluded that the optimal number of clusters was 19. In interpreting each cluster, we used the NMF algorithm (Lin, 2007) and the word significance metrics (Lim and Maglio, 2018) to identify words representative of each cluster. NMF is a representation learning method

for extracting the latent representation of data with nonnegative values, which are often more interpretable in describing the original features (Lee and Seung, 1999). Thus, we used NMF in our work to interpret the topic across article data and word features in a specific cluster. The metrics we adopted from Lim and Maglio (2018) include the mean of the TF-IDF scores of a word feature across data, where high value indicates that the word is generally important across many data, as well as, the mean of the dot product scores of a word feature to other features, where high value indicates that the word is close to many other words and important across many data at the same time. We also interpreted the results using the visualization method of Longabaugh's (2012) as follows. Step 3 in Fig. 3 shows a binary adjacency matrix. Each cluster is highlighted by a yellow square. The density of each cluster indicates the degree of homogeneity (i.e., text similarity), whereas its size indicates the quantity of data. We subsequently validated the outcome from spectral clustering (i.e., hard clustering) using the NMF topic modeling algorithm (i.e., soft clustering) for different topic number scenarios to check whether any topics were missed by the analyses. We found that this was not the case.

Finally, we performed a network analysis and a further qualitative analysis of the clustering result for interpretation (Step 4 in Fig. 3). For the network analysis, we identified the centroids of the clusters (i.e., the centroid of the TF-IDF-embedded document vectors in each cluster) and computed the cosine similarities between the centroids. Given that all of the clusters were highly related, the similarity scores were generally high and all nodes (clusters) were connected; thus, we identified the top three most closely related clusters for each cluster, and connected only these to observe the most significant relationships between the research topics. The size of a node represents its network degree (i.e., relationship strength). For the qualitative analysis, we reviewed the mining outcomes and details in the papers to add depth to our findings.

Although the literature data set shows the research topics and technology factors of Industry 4.0, it does not include much information regarding the application areas and business aspects. Instead, news articles such as introductions, compliments, and critiques of Industry 4.0 have such information. Hence, these two types of data complement each other for gaining an understanding of Industry 4.0 and enable theoretical and managerial insights to be derived. For news data collection, we searched for the same keywords in the LexisNexis Business & Industry News Database, obtaining 3901 news articles from the database.

We analyzed the news data following a process similar to that portrayed in Fig. 3. The difference between the literature and news data analysis processes lies in Step 3 in the figure. Whereas a journal article typically focuses on one specific topic, a news article often discusses multiple topics. Thus, we applied a hard-clustering method (spectral clustering; von Luxburg, 2007) for the literature data analysis, as hard clustering assigns each article to one cluster. On the other hand, we applied a soft clustering method (LDA; Blei et al., 2003) for the news data analysis, as soft clustering assigns each article to multiple clusters. LDA is a generative probabilistic model for collections of documents and words; it assumes that every document contains every topic latently and that every word contributes to every topic. This method has been applied to a variety of topic extraction problems (e.g., Griffiths and Mark, 2004; Chang and Blei, 2009; Mankad et al., 2016). Following the same feature algorithm as that used in the literature data analysis case (Step 2 in Fig. 3), we identified 1156 word features as significant in the news article corpus. Using the three well-known metrics cosine similarity, Jaccard similarity, and Kullback-Leibler divergence (Huang, 2008), we concluded the optimal number of topics to be 51. We interpreted each topic using its representative words and articles.

4. Findings

This section describes the findings from our data analysis. Section 4.1 describes the basic statistics of the literature and the statistically and semantically significant keywords for Industry 4.0. Section 4.2 identifies

the key research topics of Industry 4.0 and their five-level hierarchy. Section 4.3 describes the application-oriented findings from news articles. Finally, Section 4.4 integrates all these findings to list 31 research and application issues related to Industry 4.0.

4.1. Statistics on the Industry 4.0 literature

Table 1 shows basic statistics on the Industry 4.0 literature, including the year of publication, country of the corresponding author, authors' keywords, and Web of Science research areas for the 660 scientific articles. The statistics on the year of publication show the rapid expansion of the literature and the current relevance of our work. In this context, it is not surprising that as of September 8, 2018, 1934 researchers had contributed to Industry 4.0 research through SCIE, SSCI, A&HCI, and ESCI journals. The country statistics show the relevance of manufacturing and Industry 4.0 as many of the countries with the most published work are traditional manufacturing powerhouses, such as Germany, Italy, and South Korea. Table 1 also shows the topic statistics for the literature. Note that the table excludes "Industry 4.0" and "fourth industrial revolution" as these were the search keywords used to identify these works. The paper keywords designated by the authors highlight enabling technologies (e.g., IoT), applications (e.g., cyber-physical systems), and objectives (e.g., digitalization) of Industry 4.0. The Web of Science research areas show the interdisciplinary nature of the literature, led by manufacturing, computer science, electrical engineering, and industrial engineering.

4.2. Key research topics of Industry 4.0 and their five-level hierarchy

Fig. 4 shows 19 key research topics of Industry 4.0 (i.e., 19 clusters in the literature data). Based on the interpretation method explained in Section 3 (Step 3 in Fig. 3), we named each topic by considering the keywords suggested by various metrics and topic modeling algorithms as well as the titles and abstracts of the associated papers. In addition, we assessed the homogeneity of each cluster by reviewing the contents of the top and bottom items in each cluster during interpretation (as identified based on the cosine similarity between each item and the centroid of its cluster) as well as the publication journals in each cluster; a cluster can be considered homogeneous if the top and bottom items discuss a similar topic and the journals' subject areas are strongly correlated. Fig. 4 shows the topic interpretation and naming outcomes.

The 19 topics can be categorized into five levels. The first level is "infrastructure development for connection," including the topics of wireless sensor networks, security, and augmented and virtual reality. The second level is "AI development for data-driven decision making," including the topics of big data, prognosis, and optimization. The third level is "system and process optimization," including the topics of collaborative robots, process monitoring and control, system control and automation, system design, cyber-physical systems, smart systems, and smart service models. The fourth level is "industrial innovation," including the topics of advanced manufacturing, supply chain management, the energy industry, and industrial innovation and SMEs. The fifth level is "social advance," including the topics of the fourth industrial revolution and digital transformation. Further discussion on the individual topics is given in Section 4.4.

The fourth and fifth levels, "industrial innovation" and "social advance," represent the objectives of Industry 4.0; the third level, "system and process optimization," concerns the method for Industry 4.0; and the first and second levels, "infrastructure development for connection" and "AI development for data-driven decision making," show the basis for Industry 4.0. Fig. 5 displays a network of the relationships among the 19 research topics (see Section 3 for a description of the network visualization method). The nodes in the center have strong relationship values (i.e., they are connected to many other nodes), whereas the nodes in the boundaries are weaker and are connected through the strong nodes. The third-level topics (light yellow

Table 1
Statistics on the Industry 4.0 literature.

Year of publication	Frequency*	Country of the corresponding author	Frequency	Authors' keywords	Frequency	Web of Science research areas	Frequency
2013	1	Germany	128	"Internet of things"	86	Engineering, Manufacturing	91
2014	1	China	68	"cyber-physical systems"	72	Computer Science, Information Systems	85
2015	35	Italy	38	"smart factory"	32	Automation & Control Systems	82
2016	95	Spain	36	"big data"	30	Engineering, Electrical & Electronic	80
2017	244	Taiwan	34	"cloud computing"	22	Engineering, Industrial	79
2018	284	South Korea	33	"digitalization"	12	Telecommunications	66

*Note that this paper is based on a government project to clarify the concept of Industry 4.0 for national R&D planning that was conducted in September 2018. The frequency for the year 2018 includes papers published through September 8. As of January 2021, the Web of Science Core Collection databases indicate that 601 papers were published in 2018, 1,045 papers in 2019, and 1,687 papers in 2020.

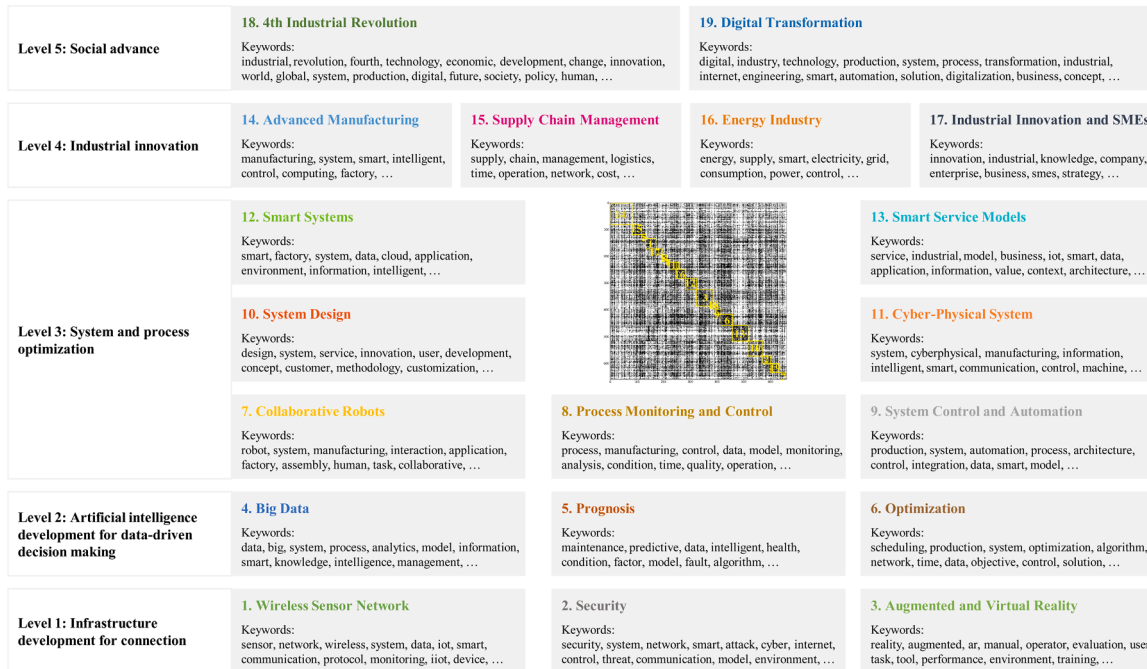


Fig. 4. Key research topics of Industry 4.0 and their five-level hierarchy.

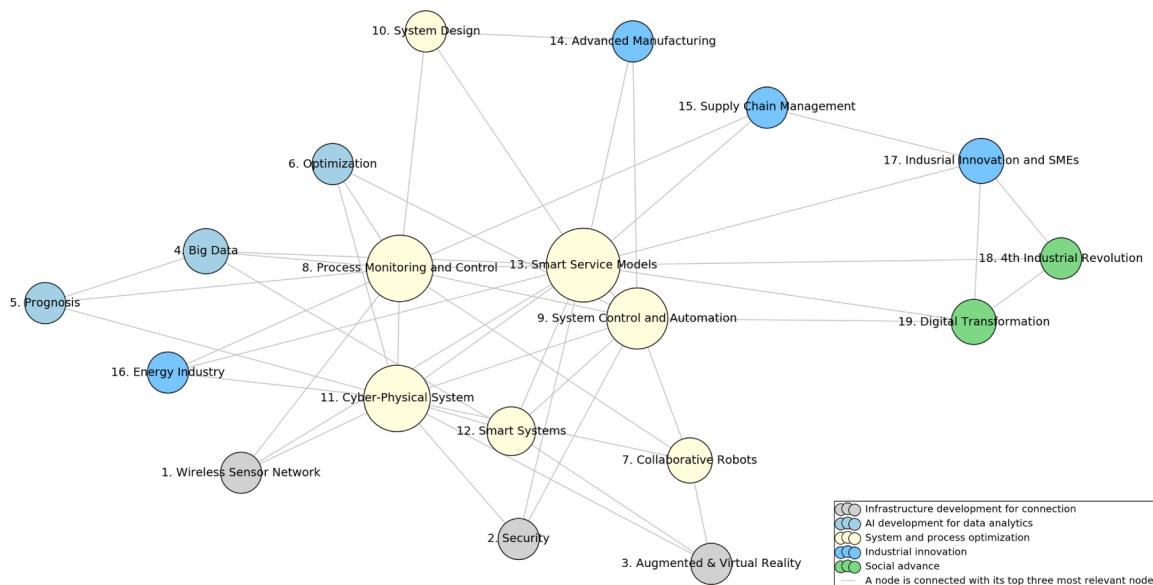


Fig. 5. Relationships among the key research topics of Industry 4.0.

nodes) connect the lower-, base-level topics (light gray and gray-blue nodes) with the higher-level topics (blue and green nodes), showing the story of how the Industry 4.0 objectives can be achieved by the methods enabled with the bases. For this reason, this work defines Industry 4.0 as “the paradigm toward industrial innovation and social advance through the optimization of systems and processes based on the development of connection infrastructure and artificial intelligence.”

4.3. Further information found from news articles

This section describes the results of our analysis of the news data. As described in Section 3, 51 topics were identified from the news data. Topic interpretation was performed using the same approach as was applied to the literature data. Most topics in the 3901 news articles can be categorized as technology issues, industrial application issues, national issues, or social issues related to Industry 4.0. Few topics were related to market reports, events, or executive perspectives of Industry 4.0, and we excluded these topics from our analysis because of their low relevance to our research objective. In summary, the news articles showed how firms and countries deal with technologies and applications of Industry 4.0.

The first topic category, technology issues, includes sensing, the IoT, big data analytics, AI, cybersecurity, robotics, cloud solutions, and blockchain. As expected, this result is consistent with the finding from the literature data. A notable point is the emergence of blockchain technology, which was difficult to observe clearly in the literature data. The second category, application issues, includes advanced manufacturing, fashion supply chains, health care, energy, and transportation. As expected, news articles contained more information on applications related to Industry 4.0 than did journal papers. In addition, we observed that numerous startups are working in various industries, such as finance, fashion, and energy. This kind of firm-related information could be observed only in the news data. The third category, national issues, includes interests of Germany, China, India, South Korea, the United Arab Emirates, and other countries, showing the importance of Industry 4.0 in many countries. We also found that policies and regulations are essential for Industry 4.0 but that these vary by country. This information could be ascertained only from the news data. The fourth category, social issues, includes the fourth industrial revolution, digital transformation, the smart city, and education. The findings for the first and second topic categories are consistent with those from the literature data analysis, whereas those for the third and fourth categories were identified from the news data. The new findings also indicated that Industry 4.0 movements should ultimately contribute to improving people’s quality of life and to the development of appropriate education methods for future generations.

4.4. Thirty-one research and application issues related to Industry 4.0

Building on the quantitatively derived findings, we further reviewed the journal papers and news articles manually to add human insight and develop a balanced overview of Industry 4.0. In this process, we integrated several industry- or nation-specific topics identified from news data into a single generic topic, such as integrating health care system optimization, manufacturing process optimization, and other similar topics into “system and process optimization in different industries” and integrating German, South Korean, and other country topics into “understanding and promotion of social advance in different countries.” We list the results of this process in Table 2, which synthesizes all topics found in the two data sets into 31 research and application issues related to Industry 4.0. We categorize these issues according to the five levels of Industry 4.0 and identify the main source for each.

Some of the findings related to infrastructure development for connection (Level 1) are as follows: 1.1. To enable Industry 4.0, it is fundamental to develop and expand wireless sensor and telecommunications networks for data collection and exchange. Examples of specific

Table 2
The 31 research and application issues identified for Industry 4.0.

Level	Research and/or application issue	Main source	
		Literature	News
Level 1: Infrastructure development for connection	1.1. Wireless sensor and telecommunications network development and expansion	v	v
	1.2. Cybersecurity reinforcement	v	v
	1.3. Augmented and virtual reality technology development	v	v
	1.4. Blockchain use for connecting people, objects, and organizations		v
	1.5. Cloud solution infrastructure development		v
	1.6. Integrated infrastructure development customized for different industries		v
Level 2: AI development for data-driven decision making	2.1. Big data collection and integration	v	v
	2.2. Prediction and prognosis technology development	v	v
	2.3. Optimization technology development	v	
	2.4. Data-learning AI development customized for different industries		v
Level 3: System and process optimization	3.1. Human-robot collaboration system development	v	v
	3.2. Process monitoring and control case development	v	v
	3.3. System control and automation case development	v	v
	3.4. New system type design	v	
	3.5. Cyber-physical system development	v	
	3.6. Smart system development and implementation	v	v
	3.7. New smart service model design	v	v
	3.8. System and process optimization customized for different industries		v
Level 4: Industrial innovation	4.1. Manufacturing industry innovation case development	v	
	4.2. Supply chain management innovation case development	v	v
	4.3. Energy industry innovation case development	v	v
	4.4. Large-, medium-, and small-enterprise innovation case development		v
	4.5. Support for open innovation of technology startups		v
	4.6. Understanding and promotion of innovation mechanisms in different industries	v	v
Level 5: Social advance	5.1. Smart city case development		v
	5.2. Understanding of technology-driven social advance mechanisms	v	v
	5.3. Promotion of digital transformation	v	v
	5.4. Policy development for innovation promotion and side-effect prevention		v
	5.5. Management of labor and employment markets		v
	5.6. Education for next generations of Industry 4.0		v
	5.7. Understanding and promotion of social advance in different countries		v

topics in the literature include new sensor development, network routing protocols, and network data-gathering protocols. 1.2. Meanwhile, the network should be secure for authentication, control, filtering, and other functions. 1.3. Augmented and virtual reality technologies are also essential as infrastructure for interactions between humans and machines. Although these technologies are for physical connection, organizational connection is also required to collect data on business interactions and processes. However, conventional centralized networks are limited to managing the trust and security issues in the interactions and processes. 1.4. Blockchain is useful for this purpose as this technology enables direct and secure interactions in and across organizations and minimizes the roles and authorities of traditional intermediaries. News articles showed that this technology is particularly useful in the financial industry, where numerous types of stakeholders interact by relying on the trust of intermediaries. These components of Industry 4.0 infrastructure are limited in an organization's local environment. 1.5. Cloud solution infrastructure development is essential today as managing and processing the interaction data require high-level hardware, software, and human resources. We found many news articles describing how professional IT companies do this job for their clients by providing cloud solutions. 1.6. Finally, news articles showed many cases of integrated infrastructure development customized for different industries, such as integration of telecommunications and security technologies.

Some of the findings related to AI development for data-driven decision making (Level 2) are as follows: 2.1. Big data collection and integration constitute one of the most important tasks for enabling Industry 4.0 because data are resources for training AI for jobs in industries. Many of the journal articles dealt with studies on integrating and using data from multiple sources for analysis purposes rather than collecting data from one particular source (e.g., integrating data from multiple machines in a factory or from social networks of different people). 2.2. Related to this, there were various types of journal and news articles on the development of prediction and prognosis technology for specific targets, such as developing predictive maintenance systems for factories and developing human health prediction systems for precision medicine. Once we can predict future conditions, prescriptions will be needed. 2.3. Thus, there were also various articles on the development of optimization technology, including optimization modeling and algorithms for efficient production management and network operations. 2.4. Finally, news articles showed many cases of AI development processes customized for different industries, such as AI in homes and in the fields of logistics, transportation, agriculture, and chemical engineering.

Some of the findings related to system and process optimization (Level 3) are as follows: 3.1. Human-robot collaboration system development is the essential concept in Industry 4.0. Automation should be a complement to existing human jobs, where robots can take the form of both hardware (e.g., manufacturing robots) and software (e.g., AI built into manufacturing robots). Examples of specific topics in the literature include the development of human-robot cooperation models for assisted shop-floor tasks in factories, safety and security issues, and use of robots in industry. 3.2. Process monitoring and control case development was a research issue frequently mentioned in the literature because Industry 4.0 began as industrial process improvement with technology. Here, process management includes that for both manufacturing and business processes. Examples of related research include monitoring manufacturing process conditions and improving process outcome quality with data. 3.3. Another research issue frequently arising in the Industry 4.0 literature was system control and automation case development, which includes design of automation architectures for system control and development of agent-based methods in cyber-physical production systems. 3.4. Although most of these studies focused on the improvement of existing industry systems, some of the recent studies focused on the design of novel systems such as designing crowdsourcing systems for data collection, designing 3D

printing systems for device manufacturing, and designing products for mass customization. 3.5. Cyber-physical system development was another key issue in system and process optimization as this concept involves many of the aforementioned technologies and concepts, such as telecommunications, augmented reality, process monitoring, prediction, and control. 3.6. The issue of smart system development and implementation refers to the integration of technology components for real system development implementation with smart functions. Examples of related research include smart factory development and smart ship development. 3.7. Although most of these studies focused on production contexts, some of the recent studies focused on the design of service models for business, including cloud service design, blockchain-based service design, smart living service architecture design, and smart logistics business model design. 3.7. Finally, news articles reported many cases of system and process optimization customized for different industries, such as energy operations management and logistics optimization.

Some of the findings related to industrial innovation (Level 4) are as follows: 4.1. Manufacturing industry innovation case development includes industry-level case studies (i.e., not system-level studies) on the application of new technologies in the manufacturing industry. These studies discuss the state of the art and future trends in manufacturing (e.g., how the scope of automation will continue to expand beyond specific mechanical systems). 4.2. Supply chain management innovation case development includes studies on the development of a new model (e.g., a scheduling model) and the impact of a new technology (e.g., IoT technology) for increasing the efficiency of a supply chain. 4.3. Energy industry innovation case development includes case studies on the development of connection infrastructure and AI for the optimization of systems and processes in the energy industry, such as cybersecurity management for power plants and energy demand forecasting for supply. 4.4. Whereas the Industry 4.0 literature discussed these industry-level case studies, news articles introduced enterprise-level innovation activities of particular large, medium, and small enterprises. 4.5. Meanwhile, we found various news articles that discuss the attempts of technology startups and their collaboration regarding open innovation. As Industry 4.0 involves new technologies, multiple countries and investment bodies have tried to develop an ecosystem that supports the open innovation of technology startups. 4.6. Finally, researchers have studied how specific industries (e.g., manufacturing, energy) have been advanced with technologies, human resources, and organizational capabilities for understanding the mechanisms and challenges of innovation. As Industry 4.0 is not yet established, research on this issue will contribute to preparation for and improvement of the future.

Some of the findings related to social advance (Level 5) are as follows: 5.1. Smart city case development is an issue derived from news articles that discussed continuous development of smart city cases emerging from all over the world. The cases cover not only the use of specific technologies for prediction and automation but also the humanities, sociology, and the arts from the perspectives of convenience and human psychology. Such studies on the ways a society develops based on technologies and application cases and addressing the accompanying challenges will contribute to the preparation for and creation of the future fourth industrial revolution. 5.2. Researchers have tried to develop an understanding of technology-driven mechanisms of social advance by studying the changes in economic and social indices with regard to Industry 4.0. 5.3. The issue of the promotion of digital transformation was distributed throughout the SCIE and SSCI journals according to the level of the transformation target (i.e., whether the target is an industry-specific system, an industry, or society as a whole). 5.4. The development of policies to promote innovation toward Industry 4.0 and to prevent side effects is one of the issues derived from the news articles. These articles discussed how policies should be pursued according to a country's particular socioeconomic environment. 5.5. Management of labor and employment markets is another issue derived from news articles as all large, medium, and small enterprises require

human resources for the skills related to new technologies and applications in Industry 4.0. 5.6. Consequently, there were other news articles addressing how education from graduate school to elementary school should be delivered in the rapidly changing industrial, social, and economic environment as technology develops. 5.7. Finally, there were other news articles reporting cases of technology-enabled social advance in various countries and discussing the implications for promoting it.

5. Discussion

5.1. Comparison with relevant discussion and debate on Industry 4.0

The World Manufacturing Forum (WMF) was held in Cernobbio, Italy, on September 28–29, 2018, where more than 900 people from 40 countries participated in a variety of discussions on Industry 4.0. The forum emphasized that the manufacturing industry is the driving force for the global economy and that a big-data-based society is coming. Several messages in particular were highlighted in the forums, many of which relate to the issues listed in Table 2. For example, “strengthen and expand infrastructures to enable future-oriented manufacturing” refers to Level 1 issues, “explore real value of data-driven decision making” to Level 2 issues, “develop effective policies to support global business initiatives” to Level 3 issues, “cultivate a positive perception of manufacturing” and “assist SMEs with digital transformation” to Level 4 issues, and “promote education skills development for societal well-being” to Level 5 issues. Thus, the findings of this work are consistent with existing discussion and debate on Industry 4.0. This consistency matches with other reviews of Industry 4.0 as well; for example, our findings include all the enabling technologies of Industry 4.0 (IoT, big data, and cloud computing) and application areas of Industry 4.0 (smart factory, smart product, and smart city) that are discussed in Lu (2017), as well as all the components of Industry 4.0 (e.g., cloud systems, data mining, and augmented reality) that are discussed in Oztemel and Gursev (2020). In the context of existing work, our clarification is significant because it confirms, complements, aggregates, and simplifies insightful but subjective expert perspectives on Industry 4.0 based on data.

5.2. Convergence in Industry 4.0

As shown in Section 4, Industry 4.0 is broad and interdisciplinary in nature. As such, the *convergence* of different capabilities is key for industrial innovation and social advance in Industry 4.0. For example, Issue 3.3 (system control and automation case development) involves the convergence of technologies for sensors, telecommunications, and analytics, and Issue 5.1 (smart city case development) requires social science capabilities that go beyond particular technologies. Building on

the previous discussion, we propose a convergence framework for Industry 4.0 (Fig. 6). In Industry 4.0, people, objects, and organizations are *connected* to *collect* data from specific systems and processes and to *communicate* with each other. The underlying information and knowledge in data are *computed* with AI and are used to optimally *control* the systems and processes and ultimately to *create* value in industries and societies. These six “C” dimensions are interrelated. For example, data collection and communication are inherently related (e.g., in wireless sensor networks), and computation using process data is a prerequisite to process control.

An integrated interpretation of Table 2 and Fig. 6 can be formulated as follows: As shown in Table 2, Level 1 issues are primarily related to connection, collection, and communication technologies. Accordingly, the relevant papers were published in SCIE and ESCI journals in engineering and science. In other words, these issues relate to the convergence within engineering and science. Level 2 issues are closely related to collection, communication, and computation as well as control technologies, thereby requiring the convergence of engineering and science. Meanwhile, Level 3 issues deal with the creation of value with AI-based control, using information from collected data analytics. In other words, these issues are related to the convergence of engineering and science as well as to the convergence of the humanities and sociology. Likewise, Level 4 issues relate to the convergence of engineering, science, the humanities, and sociology for solving practical problems of industry and management. For Level 5 issues, most of the associated papers were from SSCI and ESCI journals in the humanities and sociology. That is, these issues deal with the convergence of the humanities and sociology. Taken together, issues from Level 1 through Level 5 form a spectrum of convergences from engineering and science to the humanities and sociology.

5.3. Challenges in the convergence for Industry 4.0

As with any large-scale initiative for change, the move to Industry 4.0 is not easy. Given its multidimensional nature, the challenges of Industry 4.0 are inherently relevant to the convergence in Industry 4.0 discussed in the previous section. Therefore, this section focuses on the challenges that we identified from both the machine learning (i.e., quantitative analysis) and manual literature review (i.e., qualitative analysis) outcomes. Fig. 7 illustrates an overview of the challenges in the convergence for Industry 4.0 that should be addressed in future research.

For the infrastructure development (level 1), we observed convergence in the development of technologies of sensors, communications, security, augmented reality, blockchain, and others for the connectivity, collection, and communications capabilities (e.g., Lee et al., 2019; Egger and Masood, 2020). Meanwhile, many of these studies focused on the technology development itself. However, news articles in the context of Industry 4.0 indicate that it is important to develop the application cases of infrastructural technologies for eventual system and process optimization as well as for industrial innovation or social advance. For the AI development (level 2), we found multiple studies on the development of technologies of big data, machine learning (learning intelligence), and computational intelligence for the collection, communications, and computation capabilities (e.g., Yu et al., 2018; Jha et al., 2019). However, many of these studies were limited to laboratory-restricted AI development without ultimate service functionality. Recent studies show that a service-orientation in AI development is useful in enhancing the utility and acceptability of AI, such as AI-based monitoring and improvement of user’s behaviors (Lim et al., 2019) and informatics-based services in manufacturing industries (Lim et al., 2018b). Meanwhile, news articles and practical journal articles the context of Industry 4.0 indicate the huge demand of developing reliable and serviceable artificial intelligence for real-world applications. For the system and process optimization (level 3), most relevant studies and news articles deal with a specific case of optimization and

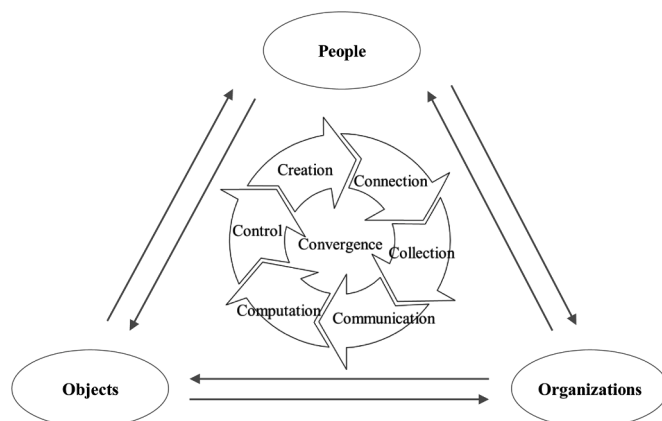


Fig. 6. Convergence in Industry 4.0.

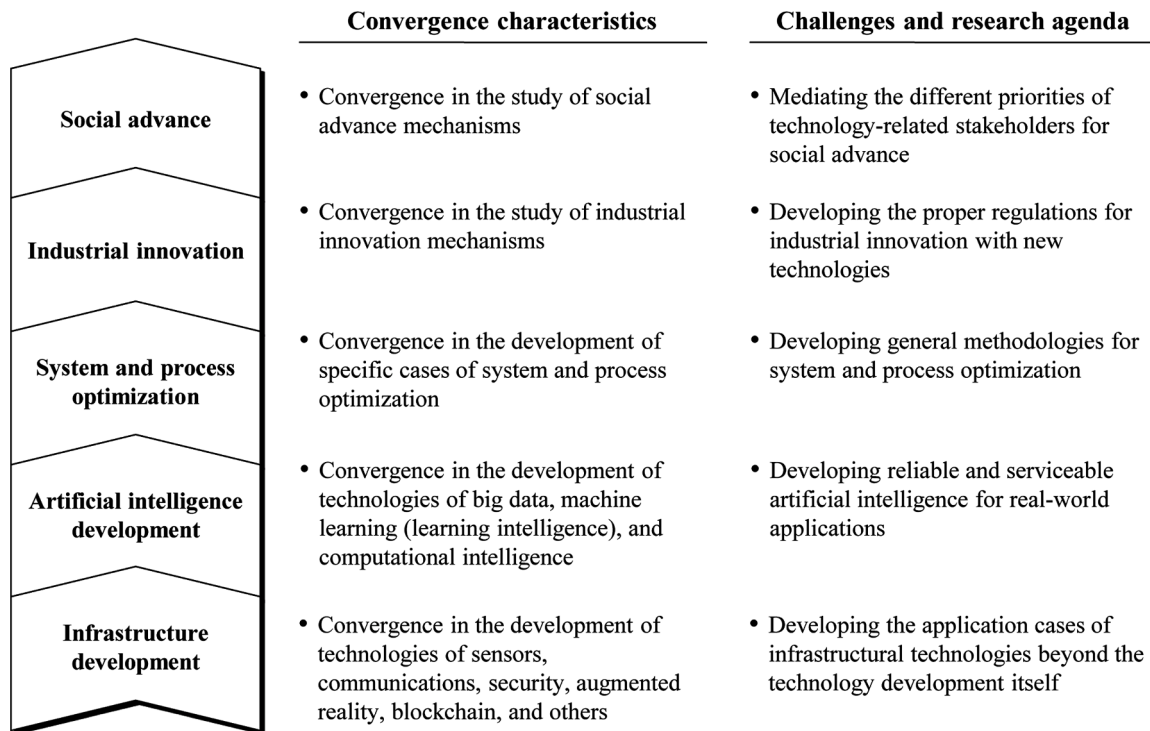


Fig. 7. Challenges and future research agenda in the convergence for Industry 4.0.

automation (e.g., [K.H. Kim et al., 2018](#); [Wang et al., 2020](#)). However, the case-sensitive reference is often not applicable to other cases. As such, we believe the development of a general methodology applicable for multiple cases is necessary to promote Industry 4.0, such as the six sigma methodology of 20 years ago ([Linderman et al., 2003](#)).

For the industrial innovation (level 4), we observed convergence in the study of industrial innovation mechanisms and its promotion for the creation capabilities (e.g., [Lim et al., 2018a](#); [Wirtz et al., 2018](#); [Horváth and Szabó, 2019](#)). Further, we found that there are relatively few studies on developing the proper regulations for industrial innovation with new technologies, despite its due importance. Regulations are important in any industry, in particular, industries such as the financial and health-care industries where stakeholders in these industries are concerned with money and human life. However, such regulations are often incomplete and sometimes must be altered when new technologies and paradigms emerge. For example, from news articles, we found that the regulations for blockchain-based ecosystems without certified intermediaries have not been fully prepared. [OECD \(2019\)](#) also discusses the importance of regulations for blockchain applications, as a blockchain inherently involves multiple stakeholders and is a tool to inherit and develop existing systems. Regulations should guide the proper use of such new technologies with specific criteria in order to prevent potential problems and failures from inappropriate use. For the social advance (level 5), we observed convergence in the study of social advance mechanisms and its promotion for the creation capabilities (e.g., [Bolton et al., 2018](#); [Subramony et al., 2018](#)). Given the broadness of the concept, there exists diverse priorities of technology-related stakeholders in any context of Industry 4.0. In this context, mediating the different priorities would be critical. For example, [K.J. Kim et al. \(2018\)](#) show that the health industry involves patients, family of patients, healthy people, doctors, nurses, hospital administrative staff, and government employees, and that the conflict between stakeholders is one of the most critical factors hindering technological development and industrial innovation in the health industry. We believe that understanding and mediating the different priorities of technology-related stakeholders is crucial in promoting system and process optimization,

industrial innovation, and eventual social change.

6. Concluding remarks

Our work advances the understanding of Industry 4.0 by mapping dispersed knowledge from the scientific literature and news articles to construct a more systematized and integrated conceptualization. Our work identifies key research and application issues related to Industry 4.0, from the issues on technological development to social change. Furthermore, it recognizes the multidimensional nature of Industry 4.0. Empirical findings are consistent with the information we can glean by reading existing studies and news items, and this implies that the findings of our semi-automated analysis of textual big data make sense, naturally and clearly reflecting the existing structure of Industry 4.0 research and application issues. Thus, our contribution is to aggregate and confirm the key concepts and areas of broad study and applications of Industry 4.0 based on data. We believe our work is important because an intensive discussion is ongoing among researchers and practitioners regarding the importance of Industry 4.0 in industry and society, and our work can provide clarification and establish a foundation for this discussion. In terms of methodological contribution, to the best of our knowledge, there are no machine-learning-based reviews similar to our work, except certain few examples (e.g., [Fortunato et al., 2018](#); [Lim and Maglio, 2018](#)), and we concur with [Chiarello et al. \(2020\)](#) that this approach is useful for researchers to monitor the evolution of a research agenda and to define its boundaries. As such, we call for more reviews with the machine-learning-based approach as a new review method for interdisciplinary fields. The method presented in this paper can be used in future studies to understand other interdisciplinary topics in science, engineering, the humanities, and sociology.

Nonetheless, we see several limitations to our work, which can be addressed in future research. First, our findings are data-dependent, which means the results can change if the data source or time frame for data collection changes. For the literature data, although our strategy was to use all available data from the Web of Science Core Collection databases as the Web of Science indexes high-quality journals, the use of

one database may present a limitation in terms of the comprehensiveness. To address this, the scope and source of data could be expanded to other databases (e.g., Scopus) or reduced to specific domains (e.g., production area journals) to suit the purpose. Similarly, additional various search keywords could be used for collecting the literature and news data. For example, there are certain similar terms related to Industry 4.0 that could return useful data, such as Industrie 4.0 (Germany), Intelligent factories (Italy), and Made in China 2025 (China) (Aulbur and Bigghe, 2016). We collected the literature and news data up to September 2018; yet in five years' time, after further technological development and social change has progressed, data by 2023 will undoubtedly show an extended list of research topics, application areas, and factors for Industry 4.0. Nonetheless, within our collected data, the unique characteristic of our contribution is demonstrated in the use of a machine-learning-based approach to develop a full range overview of the research and application issues of Industry 4.0. Although this study analyzed the broad aspects of Industry 4.0, an analysis of user perspectives on Industry 4.0 applications is missing. Surveys or interviews with users could be conducted in the future to further develop our understanding of Industry 4.0. Patent data could be analyzed to perform an in-depth analysis of technological aspects of Industry 4.0. Other types of data on Industry 4.0, such as company profiles, could be used to investigate enterprise aspects of Industry 4.0.

Second, our work did not provide detailed information on each research topic, application area, or factor of Industry 4.0. A systematic review of existing literature for a specific topic may be valuable for describing a particular aspect of Industry 4.0 in more detail. The analysis of abstracts, titles, and keywords was appropriate for achieving our research objective (i.e., examining a wide range of studies on Industry 4.0 and developing a high-level overview of the entire literature), but full-text analysis may be appropriate if the research objective focuses on a specific topic. Furthermore, our findings obtained from text mining should be integrated with real research and development projects related to Industry 4.0 through empirical studies (e.g., Frank et al., 2019; Tortorella et al., 2020). We have conducted such projects with industry and government on matters such as smart cars and transportation, health, and building systems. An integration of the current study with such projects will further facilitate the in-depth understanding and promotion of Industry 4.0.

Acknowledgements

The authors would like to sincerely thank the associate editor and anonymous reviewers for their helpful and insightful comments and suggestions that have resulted in a much improved version of this manuscript. This work was supported by the National Research Foundation of Korea (NRF-2018S1A5A2A03032736 funded from the Ministry of Education of the Republic of Korea, NRF-2020R1F1A1076067 funded from the Ministry of Science and ICT), as well as the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2020-0-01336, Artificial Intelligence Graduate School Program - UNIST).

References

- Adamson, G., Wang, L., Holm, M., Moore, P., 2017. Cloud manufacturing—a critical review of recent development and future trends. *Int J Computer Integr Manuf* 30 (4–5), 347–380.
- Alcácer, V., Cruz-Machado, V., 2019. Scanning the industry 4.0: a literature review on technologies for manufacturing systems. *Engineering Science and Technology, an International Journal* 22 (3), 899–919.
- Aulbur, W., CJ, A., Bigghe, R., 2016. Skill development for Industry 4.0. Roland Berger GmbH, Munique.
- Bahrin, M.A.K., Othman, M.F., Azli, N.N., Talib, M.F., 2016. Industry 4.0: a review on industrial automation and robotic. *Jurnal Teknologi* 78 (6–13), 137–143.
- Birant, D., Kut, A., 2007. ST-DBSCAN: an algorithm for clustering spatial-temporal data. *Data Knowl Eng* 60 (1), 208–221.
- Bird, S., Klein, E., Loper, E., 2009. *Natural Language Processing With Python*. O'Reilly Media, Inc.
- Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3 (Jan), 993–1022.
- Bolton, R.N., McColl-Kennedy, J.R., Cheung, L., Gallan, A., Orsingher, C., Witell, L., Zaki, M., 2018. Customer experience challenges: bringing together digital, physical and social realms. *Journal of Service Management*.
- Büchi, G., Cugno, M., Castagnoli, R., 2020. Smart factory performance and Industry 4.0. *Technol Forecast Soc Change* 150, 119790.
- Chang, J., Blei, D., 2009. Relational topic models for document networks. *Artificial Intelligence and Statistics* 81–88.
- Chiarello, F., Trivelli, L., Bonaccorsi, A., Fantoni, G., 2018. Extracting and mapping industry 4.0 technologies using wikipedia. *Comput Ind* 100, 244–257.
- Dalenogare, L.S., Benitez, G.B., Ayala, N.F., Frank, A.G., 2018. The expected contribution of Industry 4.0 technologies for industrial performance. *Int J Prod Econ* 204, 383–394.
- Damiani, L., Demartini, M., Guizzi, G., Revetria, R., Tonelli, F., 2018. Augmented and virtual reality applications in industrial systems: a qualitative review towards the industry 4.0 era. *IFAC-PapersOnLine* 51 (11), 624–630.
- de Sousa Jabbour, A.B.L., Jabbour, C.J.C., Foropon, C., Godinho Filho, M., 2018. When titans meet—Can industry 4.0 revolutionise the environmentally-sustainable manufacturing wave? The role of critical success factors. *Technol Forecast Soc Change* 132, 18–25.
- Dohale, V., Kumar, S., 2018. A Review of Literature on Industry 4.0. National Convention of IIIE and International Conference.
- Egger, J., Masood, T., 2020. Augmented reality in support of intelligent manufacturing—a systematic literature review. *Comput Ind Eng* 140, 106195.
- Fortunato, S., Bergstrom, C.T., Borner, K., Evans, J.A., Helbing, D., Milojević, S., Vespignani, A., 2018. Science of science. *Science* 359 (6379), eaa0185.
- Frank, A.G., Dalenogare, L.S., Ayala, N.F., 2019. Industry 4.0 technologies: implementation patterns in manufacturing companies. *Int J Prod Econ* 210, 15–26.
- Frey, B.J., Dueck, D., 2007. Clustering by passing messages between data points. *Science* 315 (5814), 972–976.
- Gorecky, D., Khamis, M., Mura, K., 2017. Introduction and establishment of virtual training in the factory of the future. *Int J Computer Integr Manuf* 30 (1), 182–190.
- Griffiths, T.L., Steyvers, M., 2004. Finding scientific topics. *Proc Natl Acad Sci* 101 (1), 5228–5235.
- Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: a k-means clustering algorithm. *Journal of the Royal Statistical Society. Series c (applied statistics)* 28 (1), 100–108.
- Horváth, D., Szabó, R.Z., 2019. Driving forces and barriers of Industry 4.0: do multinational and small and medium-sized companies have equal opportunities? *Technol Forecast Soc Change* 146, 119–132.
- Huang, A., 2008. Similarity measures for text document clustering. *Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008)* 4, 9–56.
- Ivanov, D., Dolgui, A., Sokolov, B., Werner, F., Ivanova, M., 2016. A dynamic model and an algorithm for short-term supply chain scheduling in the smart factory industry 4.0. *Int J Prod Res* 54 (2), 386–402.
- Jazdi, N., 2014. Cyber physical systems in the context of Industry 4.0. In *2014 IEEE International Conference On automation, Quality and testing, Robotics* (pp. 1–4). IEEE.
- Jha, K., Doshi, A., Patel, P., Shah, M., 2019. A comprehensive review on automation in agriculture using artificial intelligence. *Artificial Intelligence in Agriculture* 2, 1–12.
- Jolliffe, I., 2002. *Principal Component Analysis*. John Wiley & Sons, Ltd.
- Kagermann, H., 2015. Change through digitization? Value creation in the age of Industry 4.0. *Management of Permanent Change*. Springer Gabler, Wiesbaden, pp. 23–45.
- Kim, K.H., Kim, K.J., Lim, C., Heo, J.Y., 2018. Development of a lifelogs-based daily wellness score to advance a smart wellness service. *Serv Sci* 10 (4), 408–422.
- Kim, K.J., Lim, C., Kim, K.H., Kang, S.H., You, H., Jun, C.H., Kim, J.H., 2018b. Development of service concepts that utilize health-related data: a case study with the National Health Insurance Service of South Korea. *IIEE Transactions on Healthcare Systems Engineering* 8 (4), 237–249.
- Lasi, H., Fetteke, P., Kemper, H.G., Feld, T., Hoffmann, M., 2014. Industry 4.0. *Business & information systems engineering* 6 (4), 239–242.
- Lee, D.D., Seung, H.S., 1999. Learning the parts of objects by non-negative matrix factorization. *Nature* 401 (6755), 788–791.
- Lee, J., Azamfar, M., Singh, J., 2019. A blockchain enabled Cyber-Physical System architecture for Industry 4.0 manufacturing systems. *Manufacturing Letters* 20, 34–39.
- Li, B.H., Hou, B.C., Yu, W.T., Lu, X.B., Yang, C.W., 2017. Applications of artificial intelligence in intelligent manufacturing: a review. *Frontiers of Information Technology & Electronic Engineering* 18 (1), 86–96.
- Li, L., 2018. China's manufacturing locus in 2025: with a comparison of "Made-in-China 2025" and "Industry 4.0". *Technol Forecast Soc Change* 135, 66–74.
- Liao, Y., Deschamps, F., Loures, E.D.F.R., Ramos, L.F.P., 2017. Past, present and future of Industry 4.0—a systematic literature review and research agenda proposal. *Int J Prod Res* 55 (12), 3609–3629.
- Lim, C., Kim, M.J., Kim, K.H., Kim, K.J., Maglio, P.P., 2018a. Using data to advance service: managerial issues and theoretical implications from action research. *Journal of Service Theory and Practice* 28 (1), 99–128.
- Lim, C., Maglio, P.P., 2018. Data-driven understanding of smart service systems through text mining. *Serv Sci* 10 (2), 154–180.
- Lim, C., Kim, M.J., Kim, K.H., Kim, K.J., Maglio, P.P., 2018. Customer process management: A framework for using customer-related data to create customer value. *Journal of Service Management* 30 (1), 105–131.
- Lim, C., Kim, M.J., Heo, J.Y., Kim, K.J., 2018b. Design of informatics-based services in manufacturing industries: case studies using large vehicle-related databases. *Journal of Intelligent Manufacturing* 29 (3), 497–508.

- Lim, C., Kim, K.H., Kim, M.J., Heo, J.Y., Kim, K.J., Maglio, P.P., 2018c. From data to value: A nine-factor framework for data-based value creation in information-intensive services. *International Journal of Information Management* 39, 121–135.
- Lin, C.J., 2007. Projected gradient methods for nonnegative matrix factorization. *Neural Comput* 19 (10), 2756–2779.
- Lin, F., Chen, C., Zhang, N., Guan, X., Shen, X., 2015. Autonomous channel switching: towards efficient spectrum sharing for industrial wireless sensor networks. *IEEE Internet of Things Journal* 3 (2), 231–243.
- Linderman, K., Schroeder, R.G., Zaheer, S., Choo, A.S., 2003. Six Sigma: a goal-theoretic perspective. *J Oper Manage* 21 (2), 193–203.
- Longabaugh, B., 2012. *Visualizing Adjacency Matrices in Python*. <http://sociograph.blogspot.com/2012/11/visualizing-adjacency-matrices-in-python.html>.
- Lu, Y., 2017. Industry 4.0: a survey on technologies, applications and open research issues. *Journal of Industrial Information Integration* 6, 1–10.
- Mankad, S., Han, H.S., Goh, J., Gavrimeni, S., 2016. Understanding online hotel reviews through automated text analysis. *Serv Sci* 8 (2), 124–138.
- Morrar, R., Arman, H., Mousa, S., 2017. The fourth industrial revolution (Industry 4.0): a social innovation perspective. *Technology Innovation Management Review* 7 (11), 12–20.
- Mourtzis, D., 2020. Simulation in the design and operation of manufacturing systems: state of the art and new trends. *Int J Prod Res* 58 (7), 1927–1949.
- Müller, J.M., Buliga, O., Voigt, K.I., 2018. Fortune favors the prepared: how SMEs approach business model innovations in Industry 4.0. *Technol Forecast Soc Change* 132, 2–17.
- OECD, 2019. *The Policy Environment For Blockchain Innovation and Adoption 2019 OECD Global Blockchain Policy Forum Summary Report*. OECD BLOCKCHAIN POLICY SERIES.
- Opresnik, D., Taisch, M., 2015. The value of big data in servitization. *Int J Prod Econ* 165, 174–184.
- Ordenes, F.V., Theodoulidis, B., Burton, J., Gruber, T., Zaki, M., 2014. Analyzing customer experience feedback using text mining: a linguistics-based approach. *Journal of Service Research* 17 (3), 278–295.
- Oztemel, E., Gursev, S., 2020. Literature review of Industry 4.0 and related technologies. *J Intell Manuf* 31 (1), 127–182.
- Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20, 53–65.
- Schwab, K., 2017. *The fourth industrial revolution*. Currency.
- Subramony, M., Solnet, D., Groth, M., Yagil, D., Hartley, N., Kim, P.B., Golubovskaya, M., 2018. Service work in 2050: toward a work ecosystems perspective. *Journal of Service Management*.
- Sung, T.K., 2018. Industry 4.0: a Korea perspective. *Technol Forecast Soc Change* 132, 40–45.
- Tortorella, G.L., Vergara, A.M.C., Garza-Reyes, J.A., Sawhney, R., 2020. Organizational learning paths based upon industry 4.0 adoption: an empirical study with Brazilian manufacturers. *Int J Prod Econ* 219, 284–294.
- Von Luxburg, U., 2007. A tutorial on spectral clustering. *Stat Comput* 17 (4), 395–416.
- Wan, J., Tang, S., Shu, Z., Li, D., Wang, S., Imran, M., Vasilakos, A.V., 2016. Software-defined industrial internet of things in the context of industry 4.0. *IEEE Sens J* 16 (20), 7373–7380.
- Wang, Y., Pan, Z., Yuan, X., Yang, C., Gui, W., 2020. A novel deep learning based fault diagnosis approach for chemical process with extended deep belief network. *ISA Trans* 96, 457–467.
- Wirtz, J., Patterson, P.G., Kunz, W.H., Gruber, T., Lu, V.N., Paluch, S., Martins, A., 2018. Brave new world: service robots in the frontline. *Journal of Service Management*.
- Wu, H.C., Luk, R.W.P., Wong, K.F., Kwok, K.L., 2008. Interpreting tf-idf term weights as making relevance decisions. *ACM Transactions on Information Systems (TOIS)* 26 (3), 1–37.
- Xu, L.D., Xu, E.L., Li, L., 2018. Industry 4.0: state of the art and future trends. *Int J Prod Res* 56 (8), 2941–2962.
- Yoon, B., Park, Y., 2004. A text-mining-based patent network: analytical tool for high-technology trend. *The Journal of High Technology Management Research* 15 (1), 37–50.
- Yu, K.H., Beam, A.L., Kohane, I.S., 2018. Artificial intelligence in healthcare. *Nature Biomedical Engineering* 2 (10), 719–731.
- Zhong, R.Y., Xu, X., Klotz, E., Newman, S.T., 2017. Intelligent manufacturing in the context of industry 4.0: a review. *Engineering* 3 (5), 616–630.

Changhun Lee is currently a Ph.D. candidate in the Department of Industrial Engineering at UNIST (Ulsan National Institute of Science and Technology). His research interests include natural language processing, reinforcement learning, and recommender systems.

Chiehyeon Lim is an Associate Professor in the Department of Industrial Engineering and the Graduate School of Artificial Intelligence at UNIST. He obtained his B.S. and Ph.D. from the Department of Industrial and Management Engineering at POSTECH. As part of his postdoctoral experience, he served as an Assistant Project Scientist and Lecturer in the School of Engineering at UC Merced. His research interests include smart service systems, service engineering, and knowledge discovery with data mining.