Article

Exploring the Research Trend of Smart Factory with Topic Modeling

Hyun-Lim Yang [1] [iD], Tai-Woo Chang [2],* [iD] and Yerim Choi [2]

[1] Department of Information and Communication Engineering, DGIST, Daegu 42988, Korea; lims1@dgist.ac.kr
[2] Department of Industrial and Management Engineering/Intelligence and Manufacturing Research Center, Kyonggi University, Suwon, Gyeonggi 16227, Korea; yrchoi@kgu.ac.kr
* Correspondence: keenbee@kgu.ac.kr; Tel.: +82-31-249-9754

check for
updates

Abstract: Growing competition among manufacturing businesses and the advent of the Fourth Industrial Revolution has meant that many countries are conducting various research projects to understand how to introduce and populate smart factories. Smart factories are expected to provide a way of solving the manufacturing industries' complex problems, to take a role in breakthroughs in factories and to carry on a sustainable business. Smart factories are currently in the introduction stage, so we should follow up on the majorities and check their tendencies. However, smart-factory research is an interdisciplinary field that should be studied by researchers with diverse backgrounds in various domains. Thus, studying the past and present overall research trends of smart factory studies is required for their successful introduction and sustainable research. In this study, we explored the research trends of smart factories in both international and specifically Korean research, as an example of a nation case, to determine the major research directions. We determined trends using latent semantic analysis, which is a known topic-modeling technique, and analyzed the trends with regression-based methods. As a result, we could read the clear trends by analyzing existing studies related to smart factories. In addition, it is possible to compare research trends in Korea and international research trends for the commonly appeared topics, such as 'ICT' (Information and Communications Technology) and 'R&D (Research and Development)/Technology Innovation'. We expect that the quantitative analysis results and suggestions presented in this study can be used to formulate strategies for the future diffusion of smart factories.

Keywords: research trend analysis; smart factory; industry 4.0; topic modeling; latent semantic analysis

1. Introduction

Recently, research on smart factories has been actively pursued with the advent of the Fourth Industrial Revolution. Radziwon et al. [1] defines a smart factory as one that provides flexible and adaptable production processes that can solve problems in an increasingly complex world. As problems that require solutions in manufacturing industries become increasingly complicated, many problems that cannot be solved by a conventional single technology are quickly being solved, and various studies are under consideration. Examples of smart-factory research according to their purposes and domains are shown in Table 1. Wang et al. [2] and Hwang et al. [3] exploited data from factories and smart devices to make factories smarter. Jung et al. [4] and Goryachev et al. [5] considered the traditional operational management techniques to resolve problems in smart factories. Shrouf et al. [6] tried to solve the environmental issues in smart factories, and Lehmhus et al. [7] studied additive manufacturing techniques, such as a 3D printing, for customized production in smart factories. Tao and Zhang [8] explored a digital twin shop-floor reference model for smart manufacturing,

while Wang et al. [9] presented backlash error based predictive maintenance system using machine learning technique. In order of to make better use of advanced technologies in the factory, as part of the Industry 4.0 manufacturing system, Lee et al. [10] proposed a five-level guideline for implementation of cyber physical systems (CPS) in smart factories from data acquisition with sensor connections to self-configuration of machines.

Table 1. Researches on diverse topics related to smart factory.

Reference	Main Purpose	Domain
Wang et al. [2]	Build smart factory frameworks using Internet of Things (IoT) techniques.	Flexible production using data from various devices including machines, conveyers, and sensors.
Jung et al. [4]	Develop an analytical method with which to measure the performance of smart factories.	Operational strategies that consider Supply Chain Operations Reference (SCOR) from the Supply Chain Council.
Shrouf et al. [6]	Identify the energy usage in factories more efficiently.	Measuring energy consumption and CO_2 emissions.
Goryachev et al. [5]	Maximize productivity and efficiency in smart factories.	Resource management and production planning.
Lehmhus et al. [7]	Study current additive manufacturing systems.	Additive manufacturing for customized production.
Hwang et al. [3]	Develop a performance measurement system for IoT and smart-factory environments.	Performance measurement of data and planning information from factories.
Tao and Zhang [8]	Make a series of operation in manufacturing process visible.	Design a digital twin method for physical-virtual converged shop-floor.
Wang et al. [9]	Minimize the operational and resource cost of manufacturing.	Build a machine learning model for predictive maintenance with backlash error data.
Lee et al. [10]	Enhance equipment efficiency, reliability, and product quality of manufacturing systems.	Unified five-level architecture guideline for implementation of cyber physical systems (CPS) in smart factories.

As can be seen from the above examples, smart-factory research is an interdisciplinary class that is performed by researchers from diverse backgrounds and is conducted based on research results from various domains. Its interdisciplinary nature means that understanding technology development in a domain and understanding trends in the other domains are both important. However, it is difficult to follow research trends in multiple domains. Therefore, researchers will find the provision of previous and current research trends of overall smart-factory studies by analyzing research papers related to smart factories useful.

These needs are evidenced by a number of studies that have been conducted to identify trends in areas related to smart factories. O'Donovan et al. [11] analyzed distribution of publications in big data technologies from manufacturing using systematic mapping method. Four filters were applied with several inclusion and exclusion criteria, and finally classified the type of analytics with high-level taxonomy for business analytics to find the tendency of publications. Nguyen et al. [12] also investigated distribution of publications in big-data analytics from supply chain management (SCM) using a qualitative context analysis method. The number of papers are explored by year, supply chain function, and big data analytics techniques. Preuveneers et al. [13] tried to find the trends in manufacturing environments with literature review, and suggested several opportunities for future research on intelligent manufacturing. Unfortunately, however, traditional studies rely on qualitative analysis results and it is worrisome that might miss the emerging trends in the new paradigm of Industry 4.0. To address those concerns, we used a topic-modeling method, which can extract topics based on data to consider all possible topics from literatures.

There have been attempts to analyze research papers using the topic-modeling method in which the topics of documents are extracted by reducing the dimension of a high-dimensional vector that represents the document. Dimension reduction is performed by clustering words that have similar meanings [14]. Jeong and Song [15] attempted to derive the differences in the technology-diffusion rate between academia and industry by analyzing the trends of various bibliographies, such as patents, news articles, and academic papers. Kulkarni et al. [16] tried to identify the research trends of

operation management by analyzing papers in related fields. Most recently, Wang et al. [17] utilized a topic-modeling method to extract topics from personal-library data.

Popular methods for topic modeling include Latent Semantic Analysis (LSA) [18], Latent Dirichlet Allocation (LDA) [19], and autoencoder [20]. LSA utilizes singular value decomposition (SVD) technique to reduce the dimensions and find latent meaning in documents. LDA extracts the topics of documents using a Dirichlet distribution to improve some of the limitations of LSA. Autoencoder is a data-compression model that is based on deep learning that performs well when big data is available.

LSA is particularly suitable for analyzing research papers for the following reasons: First, the number of topics in a research field are expected to be fewer than dozens, and LSA has outperformed LDA both quantitatively and qualitatively in many previous studies considering the small number of topics [21–23]. Second, autoencoder's performance will be limited since there are relatively few papers on smart factories. Finally, some of LSA's disadvantages are that it cannot grasp semantic ambiguity in words, and can only extract one topic in a document [24]; however, these are addressed in this problem as the words used in the single research field's papers are unlikely to be ambiguous and the paper mainly discusses a single topic. Taking these characteristics into consideration, LSA has been frequently used to explore research trends in various fields, such as operation management [16], building information modeling [25], behavior operation [26], and SCM [27].

Therefore, this study identifies the research trends of smart factories by analyzing research papers using a two-step approach. First, topics are extracted from research papers related to smart factories using LSA. Second, the trends of the topics over time are measured via regression analysis. This article's contributions are twofold: mainly, the trends of smart factory-related research in progress are analyzed, making them easy to understand; in addition, international research trends and the case of an individual country are compared. Specifically, we compare and analyze the results of the research trend analysis of smart factories on sets of international and Korean research papers to illustrate the differences between an international flow and a national case. Thus, two types of real-world datasets are utilized, one set of international research papers and one of Korean research papers.

The reminder of the paper is organized as follows. Section 2 introduces the trend analysis framework and methods and Section 3 presents the results of the research-trend analysis on international and Korean papers. The paper is concluded in Section 4.

2. Analysis on Smart Factory Research Trends

2.1. Research Framework

This study identifies the research trends of smart factories by analyzing papers related to smart factories. Note that we only considered the abstracts of papers to grasp their main concepts. The overall flow of the analysis is as shown in Figure 1. Data were collected from an online scientific research database, ScienceDirect, for international researches and two other representative databases for Korean researches. After collecting the considered data and eliminating redundancies, stopwords were removed and synonyms were processed prior to trend analysis. Next, two-step trend analysis was performed; topics were extracted from preprocessed data with the LSA topic-modeling method and a regression algorithm was applied to the extracted topics to identify their trends. In addition, comparative analysis was conducted on the results that had been obtained by performing trend analysis on the two types of research paper (international and Korean).
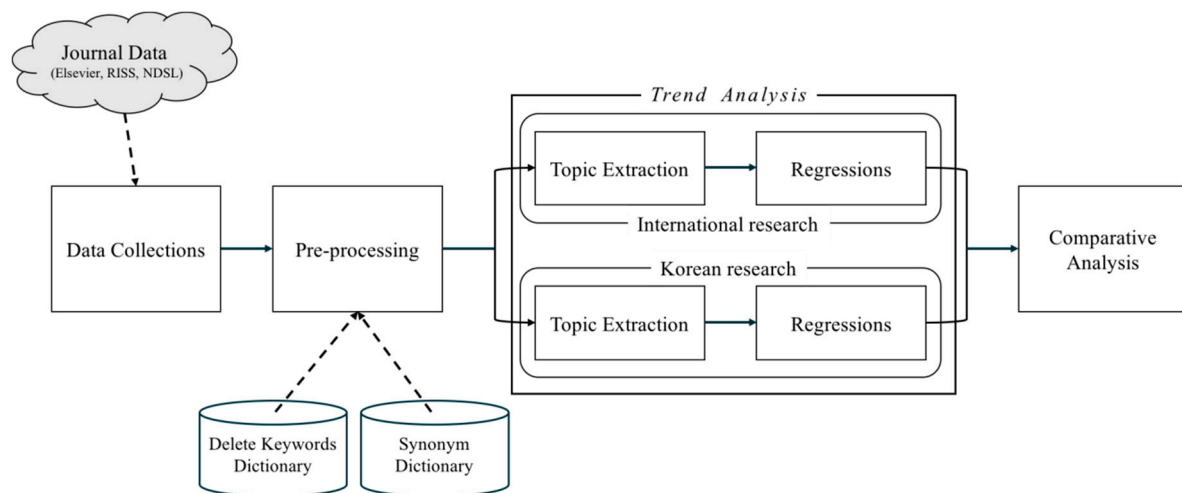
Figure 1. Overview of the research framework.

2.2. Data Collection

We analyzed international research trends on smart factories by collecting the abstracts of research papers and their metadata through ScienceDirect, which is an academic digital library operated by Elsevier. Korean research papers were also analyzed by collecting the abstracts of Korean research papers and their metadata through a Research Information Sharing Service (RISS, www.riss.kr) and National Digital Science Library (NDSL, www.ndsl.kr). The research papers that contained keywords such as "Smart Factory", "Smart Manufacturing", "Industry 4.0", and "Smart Logistics" were collected. An additional keyword, "Manufacturing innovation 3.0", was utilized when collecting the Korean research papers because it refers to a project by the Korean government that encourages smart-factory research and applications.

Table 2 shows the number of collected research papers in terms of year, range, and international or Korean status. In total, 2488 international papers and 404 Korean papers were collected and utilized after removing duplicates and unrelated ones; publications between 1995 and 2016 were considered. It can be observed that the number of international research papers related to smart factories increased, while those among Korean research papers remained relatively consistent.

Table 2. Number of collected papers related to smart factories (from 1995 to 2016).

	~2010	2011	2012	2013	2014	2015	2016
International	290	69	94	143	217	583	1092
Korean	81	30	41	44	50	74	84

2.3. Data Preprocessiong

The collected abstracts of research papers were preprocessed prior to topic modeling. First, morpheme analysis was performed; specifically, we adopted a dictionary-based morpheme analysis in which a predefined dictionary was used to detect or infer the meaning of words. Adding technical words into the dictionary such as "Industry 4.0", "smart factory", "smart manufacturing", "Information and Communications Technology (ICT)", "Supply Chain Management (SCM)", "Enterprise Resource Planning (ERP)", and "Manufacturing Execution System (MES)" meant that we could consider the manufacturing industry's terminology. In addition, stopwords were removed and synonyms were collapsed into a single word. An example list of stopwords and synonyms has been provided in Tables 3 and 4.

Second, a term-document matrix (TDM), whose element indicates the frequency of the words that occurred in our data, was developed to convert text data into a numeric matrix. In the TDM, rows correspond to keywords and columns corresponding to research papers. TDM is a sparse matrix in which most elements are zero, and it is an asymmetric matrix with a larger number of rows than columns. In addition, a large difference exists among the elements since certain keywords appear more frequently than others. Therefore, weight regularization is required for effective analysis. Term frequency-inverse document frequency (TF-IDF) and the logarithm of term frequency (TF) values (log (TF)) were used since LSA extracts the topic of a document based solely on keyword frequency, and long research papers can therefore have a high value in all dimensions in TDM.

Note that the data analysis in the study was conducted using R (v.3.2.4), which is a statistical programming language, and that morphological analysis was carried out with Stanford CoreNLP POS tagger [28] and Hannanum morphological analyzer.

Table 3. Examples of deleted keywords.

Deleted Keywords	Reason
Meanwhile, first, next, otherwise, furthermore, etc.	Stopwords
Abstract, paper, recently, rather than, future, etc.	Meaningless keywords

Table 4. Examples of synonym processing.

Synonyms	Integrated Keyword
R&D, Research and development, development, technology improvement	R&D
Internet of Things, IoT, IIoT, industrial IoT, industrial internet, network infra	IoT
Science technology, tech, core-technology, manufacturing technology	Technology
Operation field, manufacturing facilities, operation site	Factory
Raw data, big data	Data

2.4. Topic Extraction

2.4.1. LSA and SVD

LSA utilizes SVD to reduce the dimensions of a word space and find latent meanings of a document by abstracting keywords. SVD is a method for matrix decomposition that is widely used for data compression. LSA exploits the nature of SVD to infer documents' topics.

Let D be a $t \times d$ TDM with the collected data, where t is the number of keywords in data, and d is the number of research papers. Then, $D^T D$ is both a square matrix and a symmetric matrix. According to matrix diagonalization, it can then diagonalize the matrix $D^T D$ as Equation (1). Note that P is the orthogonal matrix, which is a square matrix with elements that are orthogonal unit vectors.

$$P^T \left(D^T D \right) P = E \tag{1}$$

where, P is orthogonal, and $E = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}$.

The eigenvalue σ is $\sigma_i = \sqrt{\lambda_i}$. Let r be the number of nonzero σ, and n be the total number of σ. If a specific eigenvalue σ_k has nonzero value in the range $1 \leq k \leq r$, and zero or almost zero value in the range $(r+1) \leq k \leq n$, it can approximate $\sigma_{r+1}, \cdots, \sigma_n$ as zero. Using this approximation procedure lets us obtain D' as Equation (2).

$$D = U\Sigma V^T = U[\sigma_1, \cdots, \sigma_k, 0, \cdots, 0]V^T \cong \sum_{i=1}^{k} u_i \sigma_i v_i^T = D' \tag{2}$$

where U and V are orthogonal, and the columns of each contain left-singular vectors and right-singular vectors.

Through the operations, the dimension of the research paper is reduced and the topic is thus extracted. As shown in Figure 2, after decomposing TDM into SVD and compressing it into k dimensions, it can reconstruct the approximate k dimensional vector space and analyze the distribution of documents in this space. In other words, decomposing the projected TDM D and reducing the dimension using extracted top-k eigenvalues as in the shaded area in Figure 2 allows us to then find inherent topics in the reduced dimension.
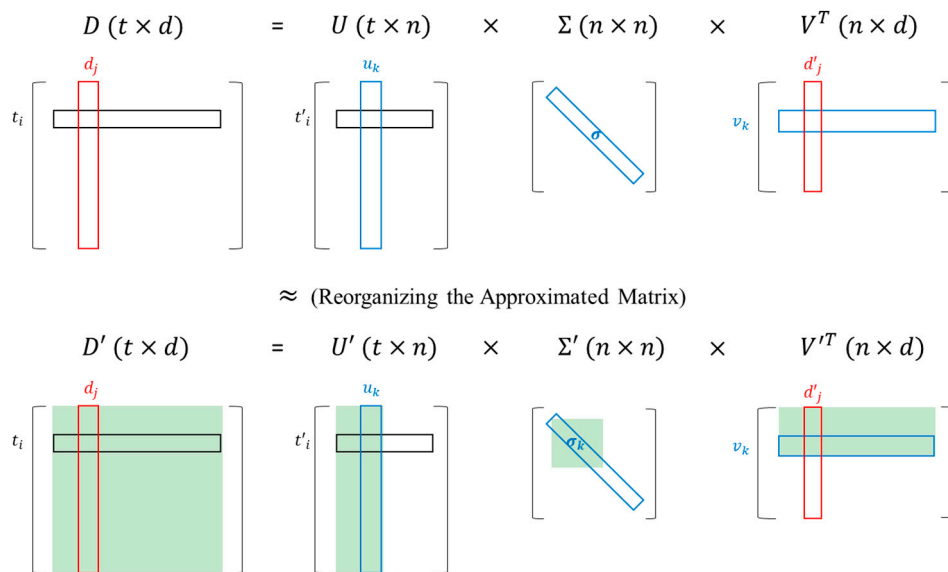
Figure 2. Concept of Latent Semantic Analysis (LSA)-based topic modeling using singular value decomposition (SVD).

2.4.2. Mapping the Topics

Each selected eigenvalue extracted from LSA can be regarded as a single topic, which is a compressed unknown feature from keywords and research papers. Therefore, a topic can be inferred based on keywords that have relatively high value in the corresponding dimension. We have manually checked the keywords that have high values in each dimension, and assigned a topic to each dimension. In addition, we conducted validation by examining the subjects of research papers as described in Section 2. In summary, a dimension that has been reduced via LSA can be regarded as an unknown feature, and the topics that are contained in each dimension can be inferred based on highly related keywords (those that have high value) to the unknown feature. In the remainder of this paper, topic and dimension are synonymous.

2.5. Topic-Based Trend Analysis

In this study, the number of papers has been aggregated by year, grouped topics from LSA, and analyzed the trends of topics using the Locally Weighted Scatterplot Smoothing (LOWESS) method. LOWESS was first suggested by Cleveland and Devlin [29] and is a nonparametric smoothing method. It is very flexible and easy to use because it does not presuppose a parametric form of data. The regression line is estimated using local regression with only part of the data, and this is fit with the weighted least square. Furthermore, it can find a robust model against outliers via interactive reweighting. However, obtaining a simple formation of the fitted regression is not possible and it is difficult to understand a model's parameters. There is also the disadvantage that the analyst's

subjectivity can be involved in the hyperparameter decision. Several attempts have been made in the past to grasp trends though the LOWESS method despite these limitations. Most notably, Bongioanni et al. [30] analyzed the trend of scientific production by country with the LOWESS method using the SCOPUS database.

The number of research papers per topic to be proposed in this study changes sensitively according to their age, so the noise in the data is large. In addition, least-square linear regression does not fit the good line because the correlation of data over time is weak and polynomial regression is difficult to determine for the polynomial. Therefore, we have tried to estimate the trend using the LOWESS method.

3. Analysis Results

In this study, we tried to identify research trends using LOWESS based on topics extracted using LSA. As the number of published research papers increased, it was considered a major trend topic, and trends were derived from the percentage of research papers by year. We applied the same method to international research papers and Korean research papers and compared the results.

3.1. International Research Trends

Analyzing the international research trends related to smart factories allowed us to derive the relationship between keywords that were inferred probabilistically using LSA based on 2488 abstracts of collected research papers. Among various k dimensions, we selected eight dimensions that were considered as having a good relationship between words, and we inferred each dimension's topics with keywords that have a high value for each dimension. Each 0 to 1 value next to a keyword was the relative influence in the corresponding dimension; the larger the value, the greater the likelihood that certain keywords would appear in that topic. We defined the inherent topics of eight dimensions as "IoT or ICT", "Alternative Energy", "R&D/Technology Innovation", "Biological or Chemical", "Logistics", "Operations management (related to S/W)", "Environmental conservation", and "Manufacturing (related to H/W)". The top 16 keywords considered for topic inferring are listed below. The side 0–1 value is the relative importance of a keyword in that dimension.

[1] Expected Topic: IoT or ICT

Keywords: sensor (0.2263), smart (0.1987), network (0.1735), IoT (0.1709), device (0.1698), data (0.1564), internet (0.1519), communicate (0.1513), system (0.1503), applicability (0.1360), wireless (0.1305), service (0.1278), inform (0.1240), technology (0.1159), computable (0.1153), mobile (0.1113)

[2] Expected Topic: Alternative Energy

Keywords: energy (0.4435), power (0.3309), grid (0.2734), electric (0.2697), renewable (0.2369), generate (0.1514), consumption (0.1349), wind (0.1257), solar (0.1163), heat (0.1147), demand (0.1141), smart (0.1044), storage (0.0973), source (0.0958), fuel (0.0874), price (0.0861)

[3] Expected Topic: R&D/Technology Innovation

Keywords: industry (0.2551), study (0.1520), firm (0.1468), innovative (0.1337), market (0.1235), factor (0.1231), research (0.1203), govern (0.1154), policy (0.1119), company (0.1063), develop (0.1044), china (0.1039), influence (0.0997), social (0.0989), import (0.0988), strategy (0.0948)

[4] Expected Topic: Biological or Chemical

Keywords: water (0.1413), temperature (0.1398), treatment (0.1295), chemical (0.1069), property (0.1056), result (0.1011), test (0.1004), concentrate (0.0963), cell (0.096), material (0.0928), drug (0.0921), surface (0.0916), condition (0.086), heat (0.0856), acid (0.0834), protein (0.0831)

[5] Expected Topic: Logistics

Keywords: urban (0.3458), transport (0.3385), city (0.3106), freight (0.2886), vehicle (0.1548), area (0.1374), plan (0.1309), sustain (0.1174), European (0.0912), traffic (0.0901), road (0.0881), project (0.0875), policy (0.0868), solution (0.0836), region (0.0830), public (0.0809)

[6]   Expected Topic: Operations management (related to S/W)

Keywords: model (0.2010), proposal (0.1698), supply (0.1659), chain (0.1597), optimal (0.1493), cost (0.1416), problem (0.1373), schedule (0.1304), algorithm (0.1211), decision (0.1194), data (0.1163), perform (0.1135), time (0.1092), custom (0.1009), result (0.1003), simulate (0.0984)

[7]   Expected Topic: Environmental conservation

Keywords: emission (0.4752), CO_2 (0.3437), carbon (0.2430), china (0.1841), reductase (0.1653), steel (0.1419), energy (0.1343), cement (0.1304), consumption (0.1158), iron (0.1089), industry (0.1018), mitigate (0.0933), pollutant (0.0888), dioxide (0.0861), save (0.0838), factor (0.0838)

[8]   Expected Topic: Manufacturing (related to H/W)

Keywords: product (0.3139), manufacturing (0.2470), process (0.2446), factory (0.1489), system (0.1469), material (0.1265), design (0.1213), engine (0.1178), sustain (0.1106), chain (0.1089), supply (0.1084), integral (0.1029), technology (0.1018), develop (0.0931), require (0.0925), tool (0.0901)

Table A1 in Appendix A shows the titles of representative research papers with high values in each dimension to validate whether the topic was set up appropriately. We can confirm that each dimension's expected topic was in line with the actual title of the research papers, and that the expected topic in this study is valid. Therefore, the trends of research papers were analyzed based on these topics.

We analyzed the research trends by year based on the eight defined topics. In Figure 3, the horizontal axis indicates the flow of time and the vertical axis indicates the percentage of the number of published research papers. The trend lines for each topic are as depicted using the LOWESS method.

In the international research trends related to the smart factory by topic, it is analyzed that the research related to "IoT or ICT" have recently decreased. Instead, research about "R&D/Technology Innovation" and "Environmental conservation" is rapidly increasing. In particular, research into "Environmental conservation" was not studied in the past, but this has increased dramatically since around 2005. "Operations management (related to S/W)" research and "Manufacturing (related to H/W)" research are both steadily being studied at a certain minimal level or higher, and the percentage of the number of research studies has gradually increased since about 2004 for both topics.
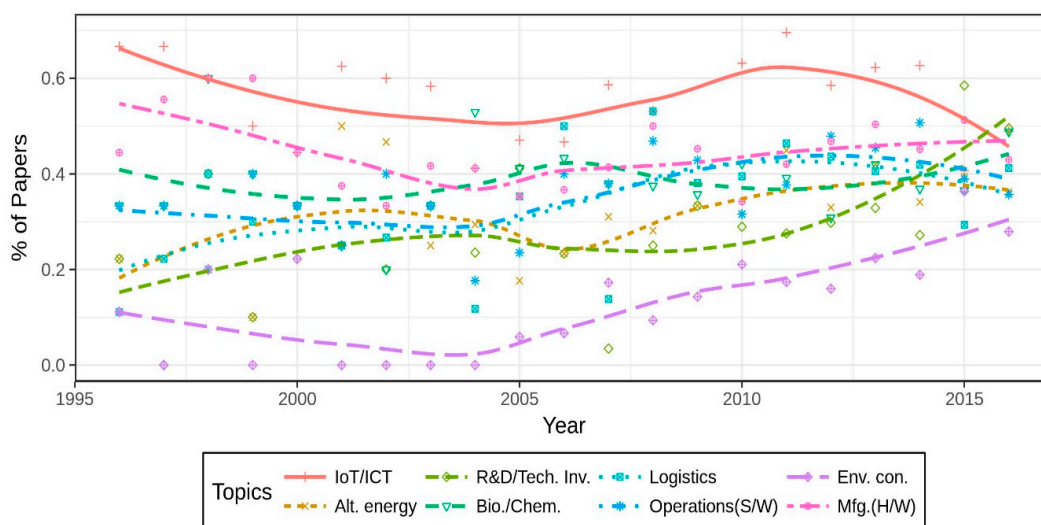

Figure 3. Research trends of international research papers by year.

The prefix "smart-", as used in the term "smart factory", has long been used in many cases as a term to refer to connections between independent devices or combinations of additional functions and features [1] that are often associated with "IoT or ICT." However, the announcement of the Industry 4.0 vision in Germany in 2011 has meant that the focus of smart factories has begun to spread to various fields such as "Operations", "Manufacturing", "Technology Innovation", "R&D", and "Environmental conservation" in addition to "IoT or ICT". In this manner, it can be analyzed that research that takes a strategic approach to the introduction of smart factories can be actively carried out and environmental factors have been emphasized for sustainability in recent years.

3.2. Korean Research Trends

A previous study [31] analyzed Korean research papers related to smart factories with publications between 1995 to September 2016. We expand the analysis period to the end of 2016 which be compatible to the international analysis, and analyzed the trend with same analytical methods. The research trends in Korea were analyzed by examining 404 abstracts with same procedure as in international research trend analysis. In Korean trend analysis, they selected five dimensions that were considered as having a good relationship between words. The defined topics were "Policies by Country", "Economic Change", "R&D", "ICT", and "Enterprise Innovation". The top 10 keywords considered for topic inference were as listed below. The 0–1 value on the side, which was the same in international trends analysis, was the relative importance of the keyword in that dimension. The same method, which examined the actual title of the research papers, was used to validate the topics' appropriateness.

[1]　Expected Topic: Policies by Country

Keywords: manufacturing industry (0.2724), Korea (0.2220), Germany (0.2159), USA (0.1967), policy (0.1981), China (0.1766), investment (0.1713), industry (0.1679), competitiveness (0.1621), Japan (0.1518)

[2]　Expected Topic: Economic Change

Keywords: ratio (0.2423), finance (0.2321), inventory (0.2217), level (0.1772), labor (0.1719), economy (0.1707), growth (0.1642), import (0.1421), deepening (0.142), increase (0.1415)

[3]　Expected Topic: R&D

Keywords: bioengineering (0.3404), technology transfer (0.2647), government funded institute (0.2467), scientific (0.2438), table (0.2287), knowledge base (0.2001), university (0.1913), commercialization (0.1872), best practice (0.1872), venture business (0.1667)

[4]　Expected Topic: ICT

Keywords: mobile (0.2346), smart (0.2177), system (0.2175), service (0.217), design (0.1988), RFID (0.1863), device (0.1728), sensor (0.1690), IoT (0.1574), internet (0.1455)

[5]　Expected Topic: Enterprise Innovation

Keywords: first (0.3719), innovation (0.2752), conglomerate (0.2638), service industry (0.2562), SME (0.2473), enterprise (0.1936), difference (0.1861), importance (0.1827), survey (0.1813), new product (0.1803)

We analyzed the research trends by year with the extracted topics as described above. As shown in Figure 4 research related to "ICT" has been actively researched recently, and it can be seen that ICT-related research has gradually increased. Prior to 2000, few studies dealt with "ICT", but these were consistently high in the early 2000 s and have continued to the present. In addition, topics of "Economic Change" and "R&D" has been analyzed as increasing trend, while research related to "Policies by Country" as decreasing recently. This can be interpreted as follows. In the early 2000s, there were a lot of researchers conducting case studies from the United States and Germany, which are recognized as leaders in smart-factory technology, and China, which is called the factory of the world. However, since the later 2000s, research seems to be focusing more on other topics based on lessons

learned from overseas cases. Other topics have shown a declining trend, and the topic of "Enterprise Innovation" in particular was the most actively studied in the early 2000s; however, the proportion of related research decreased sharply around 2006.
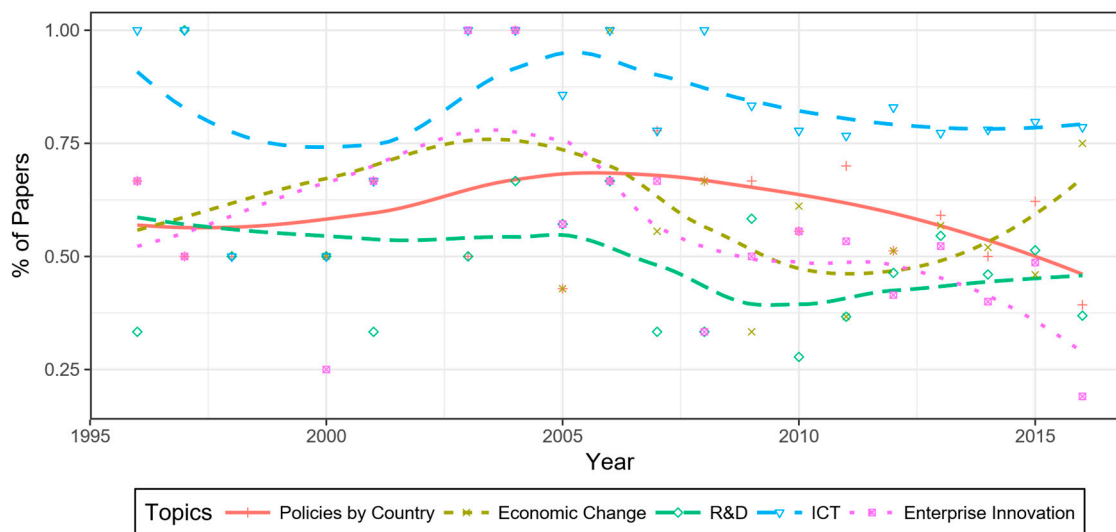
Figure 4. Research trends of Korean research papers by year.

The reason for these phenomena can be interpreted that Korea is a secondary mover compared to the United States and Germany, and Korea had steadily carried out research about policy analysis or case studies by benchmarking existing foreign best practices. It seems that various studies are being carried out to reliably deploy the smart factory and secure-leading technologies. In addition, it has been analyzed that there are many research studies into exploiting excellent ICT infrastructure, such as an ICT powerhouse.

On the other hand, topics related to the underlying technology of the smart factory, such as operations or manufacturing, have not been elucidated, but topic such as economic changes have addressed. This seems to be a very good reflection of Korea's economic situation which has achieved rapid economic growth. Korea quickly absorbs and applies outstanding overseas technology to domestic industries, but unlike other countries, such as Germany or the United States, it lacks a long-term investment in manufacturing. Korea is also focused on exports because of its relatively small market size. Those reasons seem to have lead economic change to emerge as a major topic than the other technological issues.

3.3. Comparative Study

Through the abovementioned trend analysis, different topics have been extracted amongst international research trends and Korean research trends. Regarding the number of topics, eight were extracted for international research trends and five for Korean research trends. In international research trends, topics of production and manufacturing, or the internal factors of smart factories were extracted as major topics; however, topics on strategies and policies or external factors that influence smart factories appeared as major topics in Korea. The topics of research papers that were related to production and manufacturing, which are the foundation of smart factories, are steadily increasing in international research trends. However, few studies have proceeded in Korea and the extracted topics have been minor. Furthermore, while international research on environmental conservation and alternative energy has increased since the mid-2000s, research in Korea has not.

In addition, the following two topics can be compared more directly. The graph depicted in Figure 5 is a direct comparison of international research trends and Korean research trends with common topics. Note that "R&D" and "Enterprise Innovation" topics from Korean research trends

are merged to single topic to compare with related topic in international trend "R&D/Technology Innovation". Although "IoT or ICT"-related research has been declining internationally, the Korean research related to "ICT" is gradually increasing. In addition, the percentage of research on topics such as "R&D/Technology Innovation" has increased sharply in international research trends, while few such studies have recently been conducted in Korea.

(**a**) ICT

(**b**) R&D/Technology Innovation
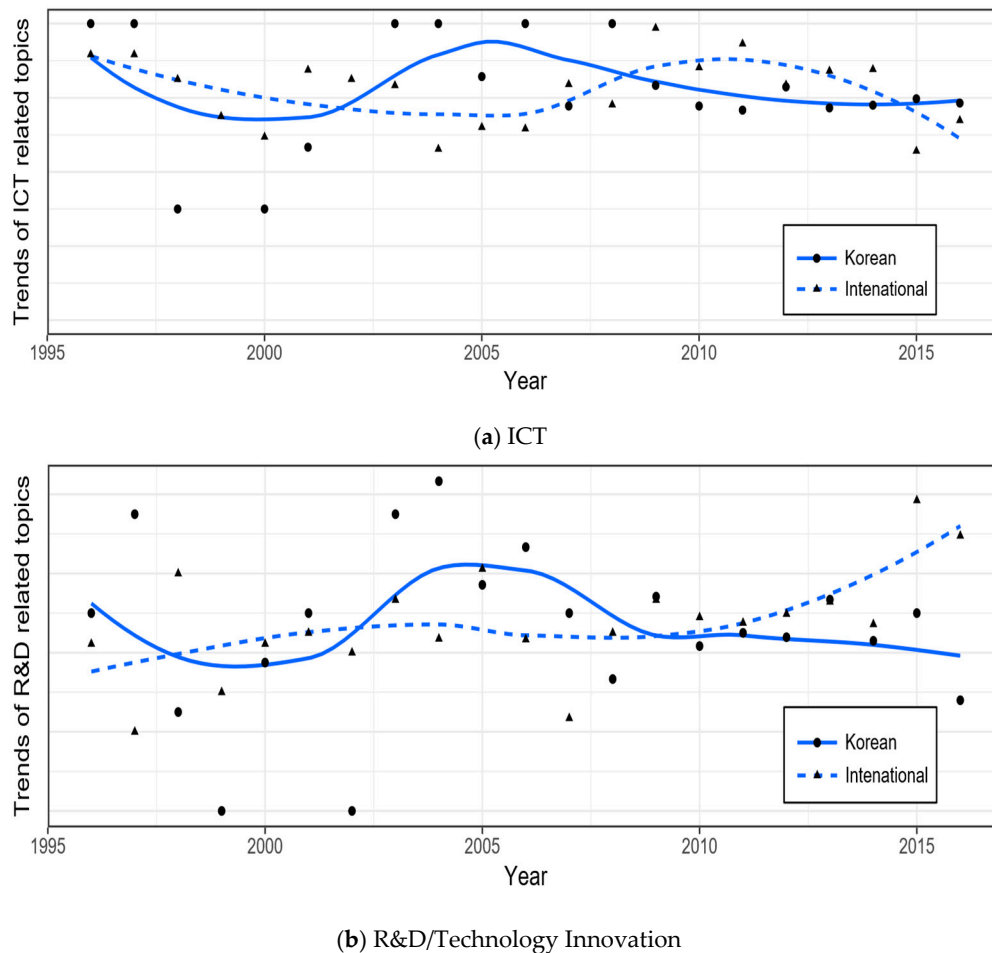
Figure 5. Comparison of research trends among international and Korean research papers in two commonly appeared topics.

In international research trends, there are as many studies about R&D as there are about ICT technology, which is a term for representing the prefix "smart-" before. It can be seen that there has been rapid increase in recent years in articles on sustainable R&D-related research to effectively support developed ICT technology for the era of Industry 4.0. In Korea, however, the focus seems to be on securing the additive technologies of smart factory and R&D-related research seems to still be lacking.

4. Conclusions

The smart factory is one of the most rapidly developed and variously studied fields in response to the recent paradigm shift of the so-called Fourth Industrial Revolution. The effective incorporation of the properties of interdisciplinary study and technological development will require grasping overall research trends. We collected research papers related to smart factories in this study and extracted the underlying topics using LSA. In addition, we analyzed the research trends over time through a regression algorithm.

Smart factory research topics were derived as a result of analyzing 2488 international research papers as follows: "IoT or ICT", "Alternative Energy", "R&D/Technology Innovation", "Biological or Chemical", "Logistics", "Operations management (related to S/W)", "Environmental conservation", "Manufacturing (related to H/W)". On the international side, research into "IoT or ICT" has recently been decreasing and the main avenue of research has been into "R&D/Technology Innovation" and "Environmental conservation". Studies related to "Operations management (related to S/W)" and "Manufacturing (related to H/W)" have steadily been progressing, and their recent trends have gradually increased.

The results of 404 Korean research papers in five topics are as follows: "Policies by Country", "Economic Changes", "R&D", "ICT", and "Enterprise Innovation". Recent studies related to "ICT", "Policies by Country", and "Economic Changes" have been conducted at a high rate, and research is underway to introduce and spread smart factories efficiently by benchmarking advanced foreign cases and exploiting excellent ICT infrastructure.

This study makes two main contributions. First, we could read the clear trends by analyzing existing studies related to smart factories more quantitatively. This has been the first attempt to emphasize the importance of international trend analysis in smart factories to the best of our knowledge. Second, we could find that it is possible to compare research trends in Korea, which is a special example of a nation, and international research trends using a topic-modeling method. From our analysis, international research has tended to deal with the internal factors of smart factories while Korean research has mainly coped with the external factors that have influenced smart factories because Korean researchers have made efforts to utilize Korea's excellent ICT infrastructure and the tendency to benchmark-existing policies and cases from foreign countries as a second mover.

Based on the results from this study, the strategic recommendation about research on smart factories in Korea can be suggested as follows: First, while the percentage of research papers related to ICT has decreased in international research trends, this has mainly been studied in Korea. If they continue to invest in studying ICT-related fields as current research trends suggest, they will be able to compete with the global market in the era of the Fourth Industrial Revolution. Second, research on manufacturing and operations management, which are fundamental in smart factories, are being carried out in international research studies, but the proportion of such research in Korea is small, so related topics cannot be derived. As smart-factory technology aims to improve factories, it has been necessary to strengthen research on manufacturing and operational efficiencies in Korea. Finally, research about environmental conservation and alternative energy should be actively carried out to secure competitiveness for exporting smart factory technologies and to sustainably develop the industry.

On the other hand, international research on smart factories should be encouraged by more ICT-related research. Recently, it is expected that artificial intelligence (AI) technologies will have a similar or superior performance to people in a certain area, such as image-recognition tasks [32], and AI can contribute to increasing factory productivity [33]. Through research on ICT that can effectively apply AI technology on factories, it will enable sustainable development in manufacturing.

This study's limitations are as follows: First, there were too few research papers available for analysis, thus various topics could not be derived, and there has been some incompatibility in conducting regression analysis through the LOWESS algorithm. It was not able to derive various trends as much as the topics studied in the traditional manufacturing fields. In addition, there is no definite metric with which to define topics from the extracted dimension, and it was necessary to manually check labeled topics in correspondence with the actual titles of research papers according to the subject. Finally, the significance of the paper was not considered, thus failing to reflect the impact of important topics with a small number of papers.

Smart factories have been actively researched recently; thus, if more research papers accumulate in the future, it will be possible to understand trends by applying a deep learning-based approach, such as an autoencoder, that can employ more inferential reasoning. Follow-up research is being

conducted to consider the importance of paper using the citation record. In addition, it is expected that further research will be able to conduct a comparative analysis on research trends both in Korea and in various countries through suggested methods.

Author Contributions: H.-L.Y. and T.-W.C. did the initial design, data collection, and analysis on the research; H.-L.Y. performed experiments; H.-L.Y. and T.-W.C. analyzed the data; Y.C. came up with structuring the framework, improved the design, and introduced analytical strategy; H.-L.Y. prepared the bulk of the manuscript and wrote the draft of the paper, while T.-W.C. and Y.C. contributed write-up for the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. The extracted topics and titles of the research papers included in the topics.

Extracted Topics	Title of the Research Papers
IoT or ICT	- Cognitive radio for "M2M" and Internet of Things: A survey - On the interplay of Internet of Things and cloud computing: A systematic mapping study - An agile information processing framework for high pressure die casting applications in modern manufacturing systems - Intelligent multi-sensor control system based on innovative technology integration via ZigBee and Wi-Fi networks - Internet of Things (IoT): A vision, architectural elements, and future directions
Alternative energy	- Application of solar energy in the oil industry—Current status and future prospects - Review of renewable energy industry in Beijing: Development status, obstacles and proposals - Role of smart grid in renewable energy: An overview - Optimal planning and scheduling of energy hub in presence of wind, storage and demand response under uncertainty - Multilevel awareness of energy used in production processes
R&D/Technology Innovation	- Dynamic patterns of industry convergence: Evidence from a large amount of unstructured data - Innovation performance and influencing factors of low-carbon technological innovation under the global value chain: A case of Chinese manufacturing industry - The impact of sectoral changes on individual competences: A reflective scenario-based approach in the creative industries - Unwanted effects of European Union environmental policy to promote a postcarbon industry. The case of energy in the European ceramic tile sector - Buyer supplier relationship and supply chain sustainability: empirical study of Indian automobile industry
Biological or Chemical	- African leafy vegetables as bio-factories for silver nanoparticles: A case study on Amaranthus dubius C Mart. Ex Thell - A feasibility study of ultrafiltration/reverse osmosis (UF/RO)-based wastewater treatment and reuse in the metal finishing industry - Morphology and performances of electrospun polyethylene glycol/poly (dl-lactide) phase change ultrafine fibers for thermal energy storage - Performance assessment of thermophotovoltaic application in steel industry - Radio-frequency tag with optoelectronic interface for distributed wireless chemical and biological sensor applications
Logistics	- Clean last mile transport and logistics management for smart and efficient local governments in Europe - METRO.FREIGHT.2020—Strategies for strengthening rail infrastructure for freight transport in urban regions - Designing new models for energy efficiency in urban freight transport for smart cities and its application to the Spanish case - Smart city logistics on cloud computing model - Urban big data and the development of city intelligence

Table A1. *Cont.*

Extracted Topics	Title of the Research Papers
Operations management (related to S/W)	- Reducing schedule instability by identifying and omitting complexity-adding information flows at the supplier–customer interface - Robust dynamic schedule coordination control in the supply chain - A job assignment model for conveyor-aided picking system - 'RFID' together with multi-agent systems to control global value chains - Optimizing the production scheduling of a single machine to minimize total energy consumption costs
Environmental conservation	- 'CO$_2$' emissions from China's iron and steel industry - Quantifying 'CO$_2$' emissions from China's cement industry - Comparative analysis of carbon dioxide emission factors for energy industries in European Union countries - 'CO$_2$' emissions and mitigation potential of the Chinese manufacturing industry - An Integrated Carbon Accounting and Mitigation Framework for Greening the Industry
Manufacturing (related to H/W)	- Customized Smartness: A Survey on Links between Additive Manufacturing and Sensor Integration - Factory Templates for Digital Factories Framework - A Seamless Convergence of the Digital and Physical Factory Aiming in Personalized Product Emergence Process (PPEP) for Smart Products within 'ESB' Logistics Learning Factory at Reutlingen University - Engineering Apps for Advanced Industrial Engineering - A knowledge-based tool for designing cyber physical production systems

References

1. Radziwon, A.; Bilberg, A.; Bogers, M.; Madsen, E.S. The smart factory: Exploring adaptive and flexible manufacturing solutions. *Procedia Eng.* **2014**, *69*, 1184–1190. [CrossRef]
2. Wang, S.; Wan, J.; Zhang, D.; Li, D.; Zhang, C. Towards smart factory for industry 4.0: A self-organized multi-agent system with big data based feedback and coordination. *Comput. Netw.* **2016**, *101*, 158–168. [CrossRef]
3. Hwang, G.; Lee, J.; Park, J.; Chang, T. Developing performance measurement system for Internet of Things and smart factory environment. *Int. J. Prod. Res.* **2017**, *55*, 2590–2602. [CrossRef]
4. Jung, K.; Morris, K.C.; Lyons, K.W.; Leong, S.; Cho, H. Mapping strategic goals and operational performance metrics for smart manufacturing systems. *Procedia Comput. Sci.* **2015**, *44*, 184–193. [CrossRef]
5. Goryachev, A.; Kozhevnikov, S.; Kolbova, E.; Kuznetsov, O.; Simonova, E.; Skobelev, P.; Tsarev, A.; Shepilov, Y. "Smart Factory": Intelligent System for Workshop Resource Allocation, Scheduling, Optimization and Controlling in Real Time. *Adv. Mater. Res.* **2013**, *630*, 508–513. [CrossRef]
6. Shrouf, F.; Gong, B.; Ordieres-Meré, J. Multi-level awareness of energy used in production processes. *J. Clean. Prod.* **2017**, *142*, 2570–2585. [CrossRef]
7. Lehmhus, D.; Aumund-Kopp, C.; Petzoldt, F.; Godlinski, D.; Haberkorn, A.; Zöllmer, V.; Busse, M. Customized Smartness: A Survey on links between Additive Manufacturing and Sensor Integration. *Procedia Technol.* **2016**, *26*, 284–301. [CrossRef]
8. Tao, F.; Zhang, M. Digital twin shop-floor: A new shop-floor paradigm towards smart manufacturing. *IEEE Access* **2017**, *5*, 20418–20427. [CrossRef]
9. Wang, K.; Li, Z.; Braaten, J.; Yu, Q. Interpretation and compensation of backlash error data in machine centers for intelligent predictive maintenance using ANNs. *Adv. Manuf.* **2015**, *3*, 97–104. [CrossRef]
10. Lee, J.; Behrad, B.; Hung-An, K. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manuf. Lett.* **2015**, *3*, 18–23. [CrossRef]
11. O'Donovan, P.; Leahy, K.; Bruton, K.; O'Sullivan, D. Big data in manufacturing: A systematic mapping study. *J. Big Data* **2015**, *2*, 20. [CrossRef]
12. Nguyen, T.; Zhou, L.; Spiegler, V.; Ieromonachou, P.; Lin, Y. Big data analytics in supply chain management: A state-of-the-art literature review. *Comput. Oper. Res.* **2018**, *98*, 254–264. [CrossRef]
13. Preuveneers, D.; Ilie-Zudor, E. The intelligent industry of the future: A survey on emerging trends, research challenges and opportunities in Industry 4.0. *J. Ambient Intell. Smart Environ.* **2017**, *9*, 287–298. [CrossRef]

14. Blei, D.M. Probabilistic topic models. *Commun. ACM* **2012**, *55*, 77–84. [CrossRef]

15. Jeong, D.H.; Song, M. Time gap analysis by the topic model-based temporal technique. *J. Inform.* **2014**, *8*, 776–790. [CrossRef]

16. Kulkarni, S.S.; Apte, U.M.; Evangelopoulos, N.E. The use of latent semantic analysis in operations management research. *Decis. Sci.* **2014**, *45*, 971–994. [CrossRef]

17. Wang, H.; Wang, N.; Yeung, D.Y. Collaborative deep learning for recommender systems. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, 10–13 August 2015.

18. Deerwester, S.; Dumais, S.T.; Furnas, G.W.; Landauer, T.K.; Harshman, R. Indexing by latent semantic analysis. *J. Am. Soc. Inf. Sci.* **1990**, *41*, 391–407. [CrossRef]

19. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.

20. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [CrossRef] [PubMed]

21. Stevens, K.; Kegelmeyer, P.; Andrzejewski, D.; Buttler, D. Exploring topic coherence over many models and many topics. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Jeju Island, Korea, 12–14 July 2012; pp. 952–961.

22. Kumar, G.; D'Haro, L.F. Deep autoencoder topic model for short texts. In Proceedings of the International Workshop on Embeddings and Semantics, Alicante, Spain, 15 September 2015.

23. Bergamaschi, S.; Po, L. Comparing LDA and LSA Topic Models for Content-Based Movie Recommendation Systems. In Proceedings of the International Conference on Web Information Systems and Technologies, Barcelona, Spain, 3–5 April 2014; Springer: Cham, Switzerland, 2014; pp. 247–263.

24. Lee, S.; Song, J.; Kim, Y. An empirical comparison of four text mining methods. *J. Comput. Inf. Syst.* **2010**, *1*, 1–10.

25. Yalcinkaya, M.; Singh, V. Patterns and trends in Building Information Modeling (BIM) research: A Latent Semantic Analysis. *Autom. Constr.* **2015**, *59*, 68–80. [CrossRef]

26. Kundu, A.; Jain, V.; Kumar, S.; Chandra, C. A journey from normative to behavioral operations in supply chain management: A review using Latent Semantic Analysis. *Expert Syst. Appl.* **2015**, *42*, 796–809. [CrossRef]

27. Joshi, S.; Sharma, M.; Rathi, S. Forecasting in Service Supply Chain Systems: A State-of-the-Art Review Using Latent Semantic Analysis. *Adv. Bus. Manag.* **2017**, *12*, 181–212.

28. Toutanova, K.; Klein, D.; Manning, C.D.; Singer, Y. Feature-rich part-of-speech tagging with a cyclic dependency network. In Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, Edmonton, AB, Canada, 27 May–1 June 2003; Volume 1, pp. 173–180.

29. Cleveland, W.S.; Devlin, S.J. Locally weighted regression: An approach to regression analysis by local fitting. *J. Am. Stat. Assoc.* **1988**, *83*, 596–610. [CrossRef]

30. Bongioanni, I.; Daraio, C.; Ruocco, G. A quantitative measure to compare the disciplinary profiles of research systems and their evolution over time. *J. Inform.* **2014**, *8*, 710–727. [CrossRef]

31. Chang, T.; Yang, H. Latent Semantic Analysis of Research Papers on Smart Factory. *ICIC Express Lett.* **2017**, *11*, 899–904.

32. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]

33. Posada, J.; Toro, C.; Barandiaran, I.; Oyarzun, D.; Stricker, D.; de Amicis, R.; Pinto, E.B.; Eisert, P.; Döllner, J.; Vallarino, I. Visual computing as a key enabling technology for industrie 4.0 and industrial internet. *IEEE Comput. Graph. Appl.* **2015**, *35*, 26–40. [CrossRef] [PubMed]