Applied Energy 313 (2022) 118898

Contents lists available at ScienceDirect

Applied Energy

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

Investigating emerging hydrogen technology topics and comparing national level technological focus: Patent analysis using a structural topic model

Hyunhong Choi [a], JongRoul Woo [b,*]

[a] Department of Industrial and Management Systems Engineering, Kyung Hee University, 1732 Deogyeong-daero, Giheung-gu, Yongin, Gyeonggi 17104, South Korea
[b] Energy Environment Policy and Technology, Graduate School of Energy and Environment (KU-KIST Green School), Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841, South Korea

HIGHLIGHTS

- Identifies trending and impactful topics in various fields of hydrogen technology.
- Analyzes technological portfolio of key countries in the hydrogen fields.
- Performs a technology correlation analysis of the identified technology topics.
- Proposes a decision framework for policymakers to categorize identified topics.

ARTICLE INFO

Keywords:
Hydrogen technology
Research and development strategy
Structural topic model
Patent analysis

ABSTRACT

Hydrogen technology has recently attracted great attention as a new energy technology with a potential to transform existing energy systems. However, since hydrogen technology is based on a concept encompassing a broad range of topics, such as hydrogen production, storage, distribution, and utilization, it is difficult for individual field experts to analyze it from an integrated viewpoint. In this study, we used a semiautomated, unsupervised learning approach for patent data analysis to identify specific technology topics in various fields of hydrogen technology and to analyze the technological focus of key countries. Thus, we collected 17 281 hydrogen technology patents from the last decade (2010–2019) from the United States Patent and Trademark Office (USPTO) and input the text in the title and abstract of the collected patents along with their metadata to a structural topic model. Consequently, we identified various hydrogen technology topics estimated based on the co-occurrence pattern of words within patents and represented as probabilistic word distribution. Furthermore, the metadata of the collected patents were used to identify topics that were newer or more impactful and appeared more frequently in patents from a certain country. After identifying technology topics, we also estimated the technology maturity rate (TMR) of each topic to measure its remaining potential. Among the 40 latent technology topics identified from the collected patents, 11 topics showed increasing proportion over time (new and trending) and five topics were highly cited by other patents (impactful). Furthermore, based on the analysis results, implications were presented for hydrogen research and development (R&D) strategy by (1) comparing the technology portfolios of key countries, (2) performing a technology correlation analysis of the identified hydrogen technology topics, and (3) proposing a decision framework for policymakers to categorize identified topics. We found that the primary technological focus was on fuel cell technologies for South Korea and Japan whereas hydrogen production technologies for France and the United States. Furthermore, technological rivalry patterns between key countries may differ notably depending on their specific technology topics or fields, highlighting the need for considering such aspects in developing R&D strategies. Finally, based on the proposed decision framework, we identified which technology topic to continue to focus on and consider easier expansion.

* Corresponding author.
E-mail address: jrwoo@korea.ac.kr (J. Woo).

https://doi.org/10.1016/j.apenergy.2022.118898
Received 3 August 2021; Received in revised form 20 February 2022; Accepted 4 March 2022
Available online 12 March 2022

1. Introduction

As various environmental issues have emerged because of fossil fuel-based energy systems, and as the international community has been continuously making efforts to solve such issues, there is growing interest in hydrogen technology and hydrogen energy systems, which are expected to reform the fossil fuel-based energy systems into environment-friendly systems and create a new high-value-added market [1–4]. However, hydrogen technology has not yet reached its full-scale commercialization stage, and resolving uncertainties through technology development is a prerequisite for the full-scale application of future hydrogen energy systems. Thus, governments around the world are establishing and implementing national level policies to support hydrogen technology development and promote related industries for securing a competitive edge in this new technology field.

However, "hydrogen technology" is not a term that designates a specific technology but is a generic concept covering a broad range of technology fields, including the production, storage, distribution, and application of hydrogen. Particularly, hydrogen production and application approaches are diversified and are closely related to existing industries. For example, hydrogen can be produced by reforming or processing conventional fossil fuels, and existing petrochemical companies are showing strength in this sector. In addition, hydrogen can be produced through the water electrolysis approach using renewable energy (wind power, photovoltaics, etc.), and existing power generation and other new companies are competing in this sector. Furthermore, in terms of hydrogen application, companies in various fields, such as the automotive, chemical, and energy industry, are using their core capabilities to enter the hydrogen technology field.

The hydrogen technology field is deeply connected with other existing technology fields and covers a broad range of technology topics, thus making it important for different countries to identify specific technology topics and prepare their response strategies. Many countries aim to create relevant industrial markets, gain energy independence, and build eco-friendly energy systems through developing hydrogen technologies. However, if international technological competitiveness is not secured, the benefits that a country can gain by implementing hydrogen technology and hydrogen energy systems may be reduced significantly. Thus, policymakers need to identify the relationship patterns of technologies in different sectors and untapped technological opportunities to provide strategic investments and supports. However, it is difficult for a few experts to evaluate a broad range of technological content. Therefore, this study aims to present a framework that can support the decision making of key policymakers using a semi-automated, unsupervised learning approach.

Specifically, this study conducted a "content-based" technology analysis through an in-depth text analysis using hydrogen technology patent data, with a focus on key countries. As the hydrogen technology field has not fully entered the commercialization stage yet, a preemptive research and development (R&D) investment decision based on scientific technology analysis may have a greater strategic significance. One of the key objectives of this study is to propose a framework that enables such preemptive technology analysis and R&D strategy proposals for the emerging technology sector. Some previous studies used patent data for technology analysis to conduct technology trend analysis using patent classification code and several patents for each code [57,58]. However, such analysis, not considering the content of patents, may have limitations for the emerging technology sector, which is not likely to have detailed dedicated classification codes for all specific technology topics. On the other hand, some other previous studies used text information (title and abstract) in the patent for their analysis. For example, some extracted keywords to analyze development patterns [15] or draw technology map [33]. Some others used a probabilistic topic model to identify specific topics in the target technology sector [27–29]. However, the keyword-based approach may have limitations on understanding contextual meaning of words in a broad technology sector, and analysis using a basic probabilistic topic model cannot provide implications about the background of topic prevalence and a correlation between identified topics.

This study bridges such gap by using a structural topic model (STM) to identify key technology topics that appear in the title and abstract of the hydrogen technology patents and compared the hydrogen technology topics of focus for each country. By using the topic model, the study can identify specific topics within the classification code or identify convergent topics from different classification codes. Moreover, by specifically using STM and incorporating national and patent-level variables in the model, we can identify emerging technology topics (both in terms of trend and impact) among the identified topics and provide national level technology focus through empirical validation. Furthermore, a technology correlation analysis based on the co-occurrence pattern of topics within patents can be conducted to provide R&D strategy implications based on the identified relationship between different topics. Finally, we propose a decision framework that can categorize identified topics considering whether a topic is trending or impactful, maturity level of the topic, and technology portfolio of a country.

The rest of this paper is organized as follows: Section 2 discusses the background of this study, and Section 3 describes the data and methodology used in this study. Section 4 presents the key findings of this study. Finally, Section 5 summarizes the main points of this study and presents policy implications.

2. Background

2.1. Hydrogen technology

Hydrogen technology is a generic term for technologies related to the production and usage of hydrogen. The demand for hydrogen and the investment in hydrogen technology have been continuously rising. Specifically, the global pure hydrogen demand in 2015 was approximately 70 million tons, more than tripled from that in 1975 [6]. Furthermore, related R&D investments have been active in recent years, nearly doubling in five years from USD 494 billion to USD 903 billion in 2020 [7].

Hydrogen is used in a wide variety of applications. Particularly, its various uses in the transportation and energy sectors are based on the fuel cell technology, which is a key application technology. Specifically, hydrogen fuel cell electric vehicle (FCEV) technology in both light- and heavy-duty vehicles [8–10] and high-capacity energy storage systems (ESS) using hydrogen fuel cells [11,12] have received great attention. However, for establishing hydrogen energy systems as new alternative energy systems, it is necessary to develop not only these application technologies using hydrogen energy systems but also technologies for supply chain stabilization encompassing the stable production, storage, and distribution of hydrogen [13,14].

First, hydrogen is not a naturally occurring primary energy source and thus requires production. The most popular methods include production of hydrogen as a byproduct of industrial processes and separation of hydrogen from the main components of fossil fuels (natural gas and coal). However, because these methods also emit CO_2 as an end-product (gray hydrogen production), there has been a growing interest in eco-friendly methods that emit less CO_2 (green hydrogen production) [11,16]. Specifically, there are methods of producing hydrogen from water electrolysis using renewable power sources (photovoltaics, wind power, etc.) [16,17] and from waste [18,19]. Last, there are issues regarding hydrogen storage and distribution. Because hydrogen is a gas, technology development is underway to develop an efficient storage method (compression, liquefaction, adsorption, etc.) and distribution (pipelines, pressure vessels, etc.) [11,20,21].

Hydrogen technology is expected to revolutionize future energy systems, and because of the huge long-term economic effects expected from constructing relevant infrastructures, companies, or countries that preemptively secure reliable technical capabilities have been forecasted to gain enormous economic benefits. Thus, the government and companies in each country should properly identify areas where the R&D investments of major competitors are focused on and areas where they should concentrate on making efficient R&D investment decisions. Based on these motives, this study uses the text part of the patent data to analyze the R&D trends of hydrogen technology in the last decade and specific hydrogen technologies of focus for each country to suggest relevant policies.

2.2. Topic model

A topic model is an unsupervised learning approach that extracts multiple latent topics from a large number of documents. The most widely used topic model is the Latent Dirichlet Allocation (LDA) model proposed by Blei et al. [22] It has been used for text analysis in various fields, for example, in online reviews for products or services [23,24], social media [25], and research reports [26]. Godin et al. [25] applied a topic model to Twitter data to identify various latent topics in tweets and proposed a method of recommending multiple hashtags for a tweet based on the results. Further, Kim et al. [26] used the abstract part of policy research reports in South Korea to conduct an LDA analysis separately for three time periods and investigated similarity between the topics of adjacent time periods to connect topics from different time periods. Based on the results, they analyzed how details of policy research topics in South Korea change by time.

Furthermore, some studies applied a topic model to the patent data, similar to our study. While various approaches have been made using patent data to identify promising technology fields, many of them focused on existing patent classification codes or citation information to investigate the relationship between different technology areas. For example, Kim and Bae [63] utilized k-means clustering based on classification codes and then conducted a citation analysis between clusters. Noh and Lee [65] investigated various indices that can be calculated from various patent metadata (number of new classification pairs, number of subclasses, citation frequency, etc.) to identify promising technology. Studies utilizing topic models focus on a different aspect, which is the content of the patent represented by text data. Such an approach can provide different insights compared to previous approaches. Suominen et al. [27] used approximately 160,000 United States Patent and Trademark Office (USPTO) registered patents of leading telecommunication firms, such as Samsung, Apple, Google, and Microsoft, collected over 14 years. They visualized and compared the knowledge profile of each firm based on the 75 topics derived from an LDA analysis and identified the technology fields each company was focusing on and the technology topics that had emerged and declined in the telecommunication sector. Furthermore, Kim et al. [28] applied LDA to the patents related to the greenhouse gas reduction technology to identify 50 topics and analyzed the word composition of the most important topics among them. Song and Suh [59] also applied LDA to patent data. To be specific, they conducted an LDA-based network analysis to patents for safety technologies. Moreover, Kang et al. [60] combined LDA and citation analysis to investigate the pathway in carbon capture, storage and utilization (CCUS) technologies. Erzurumlu and Pachamanova [61] applied LDA to patent data and linked it to licensing that was databased to investigate the commercial viability of identified topics in the healthcare sector. Chen et al. [62] focused on technological forecasting of a specific country by applying LDA to Australian patents from 2000 to 2014, and observed the pattern of topic prevalence by time. Zhang et al. [29] divided the blockchain technology-related patents into several time periods and then applied LDA to each time period to analyze the correlation of topics between different time periods, similar to Kim et al. [26].

Although various interesting studies have been conducted using LDA, this study used a STM. Among the topic models, the STM has been proposed relatively recently and can be useful, especially in the social science field. It can extract latent topics from documents more reliably than LDA, analyze how topic prevalence changes depending on the characteristics (metadata) of each document, and derive results with significant implications depending on the research hypothesis and key variable settings. For example, Farrell used an STM to analyze how the corporate funding of climate change-related research studies had an influence on the topics being investigated more frequently [30]. They observed that the corporate-sponsored papers included several specific topics skeptical about climate change, and the opinions (topics) on climate change were polarized over time because of these papers. Furthermore, STMs can be useful in the marketing field. Korfiatis et al. [31] used an STM to analyze not only the review data of passengers who used airline services but also the metadata such as the ratings for specific parameters related to the experience (friendliness, cleanliness, entertainment, and food) and flight (such as seat class and flight distance). In their study, they analyzed the prevalence of specific topics closely related to higher or lower ratings. Moreover, they visualized the service quality map of different airline companies to observe similarities and differences in their service quality and provided marketing implications based on the results. Furthermore, Kuhn [32] used the aviation incident record data to identify topics that prevailed depending on individual flight circumstances, such as flight purpose, flight stage, and flight time.

National R&D policy decision makers may have some idea about a specific technology field, but it is difficult for them to understand all the details about the whole field. Patent review by domain experts is one of the most widely used approach to overcome such difficulties, but it can be very expensive. Moreover, when the scope of technology is wide, as with hydrogen technology, it may be difficult for an expert in a certain field to make a comprehensive field assessment. Therefore, in this study, we used the text part of the hydrogen technology patents to summarize and describe the key content by identifying latent technology topics in patents. Subsequently, we compared the results at the national level to provide practical implications that can aid R&D decision making.

3. Research gap, research question, and contribution of this research

According to our review on relevant literature about R&D technology opportunity discovery with focus on energy sector, previous studies have three major research gaps.

First, existing studies using patent data usually relied on existing classification codes, where it may have limited performance when applied to wide technology area such as the hydrogen technology sector. To be specific, some studies investigated patents based on existing patent classification codes (citation analysis, trend analysis), to investigate R&D trend and relationship between patents from different classification codes [53,55,57,63]. Whereas such classification codes can be useful in many applications, its performance may be limited when such codes cannot represent the interest of the study properly. For example, the scope of the study may be hard to be matched with existing codes or the study may want to investigate detailed topics among patents in the same classification code. Therefore, this study used a STM to identify detailed technology topics using the text data in the titles and abstracts of patents, to overcome such limitations.

Second, limited systematic analysis for the whole hydrogen technology sector has been made using patent data, while it is gaining high attention from many firms and countries. Previous studies investigating the energy sector focused on other fields usually more commercialized, such as solar or wind power [47,50–52,57], electric vehicles [33,48,54], or other specific application technologies [49,53,60,64]. Considering the recent global attention to the hydrogen technology sector, this study aims to focus on the hydrogen technology sector, considering wide spectrum of technologies including hydrogen production, storage, distribution, and application.

Third, limited national level content-based analysis has been made to identify emerging technology topics and investigate a different technology portfolio of different countries. Many previous studies using patents conducted investigation with focus on specific countries or did not consider nationality [26,29,46,50,62,64], or focused on firm-level analysis [27,48,51]. Moreover, many studies have compared the number of patents or simple indicators between key countries [49,52,54,56–58] content-based investigation which can provide various implications between key countries was limited. Even when they conducted text analysis including topic identification [27–29], investigation and comparison of a national technology portfolio was limited. Therefore, this study uses the STM and compared a national level technology portfolio between key countries including the US, Japan, Korea, Germany, and France. Moreover, we propose a decision framework for policymakers to help their R&D strategy planning considering the key characteristics of the technology topic and technology portfolio of the country of concern.

To summarize, to address the above-mentioned gaps, we use patent data and STM to investigate the hydrogen technology field and propose a semiautomated approach to help the decision making of policymakers—the research question addressed in this study are:.

1. What are the emerging and/or high potential technology topics in the hydrogen technology sector?
2. How do the national-level technological focus of key countries in the hydrogen technology sector and their respective strategies differ from each other?
3. How can we construct a semiautomated approach using patent data to help policymakers make R&D decisions?

4. Methods

Research framework used in this study comprised four steps: 1) data collection, 2) data refinement, 3) technology topic identification, 4) technology maturity rate estimation, and 5) visualization.

4.1. Data collection

This study collected patents related to hydrogen technology for the last decade (2010–2019) from the USPTO. However, hydrogen technology is a concept that includes a wide range of technologies related to production, storage, distribution, and application of hydrogen. One of the key concerns in technology market analysis using patent data is selecting patents of interest, especially when the target technology field is wide, as in this study. Among various approaches suggested in previous literature, the most widely adopted method of selecting patents is via keyword search or classification code [33]. In this study, we used a patent classification code to select patents of interest as the keyword search approach was not applicable[1].

Various classification codes are used to group and categorize patents with similar content. For example, the International Patent Classification (IPC) code is the most widely accepted patent classification code worldwide, and the Cooperative Patent Classification (CPC) code is a more detailed classification code mainly developed by USPTO and European Union Intellectual Properties Office (EUIPO). One of the key differences between the IPC and CPC codes is the existence of section Y, which comprises patents related to new technologies or cross-sectional technologies, in the CPC code[2]. The existence of section Y in the CPC code was important for this study as hydrogen technologies are also included in section Y as one of the new technology fields. Specifically, CPC code Y02E60 is described as “hydrogen technology.” Therefore, this study collected hydrogen technology patents using the CPC code.

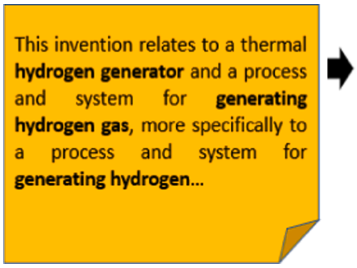
However, as many hydrogen technologies emerge in relation with existing technologies or industries, it cannot be said that hydrogen technologies specified in section Y represent the whole hydrogen technology field. Therefore, this study categorized the hydrogen technology into three fields: 1) hydrogen production, 2) hydrogen storage and distribution, and 3) hydrogen application. Subsequently, we searched CPC codes that could represent each technology field. To be specific, we conducted a keyword search in the USPTO database to identify key CPC codes for each field using representative keywords (e.g., hydrogen production, hydrogen storage, hydrogen fuel cell, etc.). By reviewing the top CPC codes appearing in patents from search queries, we finalized the CPC codes to be considered in this study (Table 1). Whereas hydrogen can be applied in various fields other than fuel cells such as hydrogen fueled gas turbines [42,43], fuel cells still comprise the vast majority of applications. Therefore, this study focused on fuel cells for hydrogen applications, which also have distinctive CPC codes.

Table 1
Cooperative Patent Classification (CPC) codes of hydrogen technology patents.

Type	CPC code	Description
Hydrogen production	C01B3/	Hydrogen; Gaseous mixtures of hydrogen; Separation of hydrogen from its mixtures; Purification of hydrogen
	Y02E60/36	Hydrogen production from non-carbon containing sources though approaches such as water electrolysis
	Y02E70/10	Other energy conversion or management systems reducing GHG emissions - Hydrogen from electrolysis with non-fossil energy, e.g., photovoltaic, wind, and nuclear power
Hydrogen storage and distribution	Y02E60/32	Hydrogen storage
	Y02E60/34	Hydrogen distribution
Hydrogen application	Y02E60/50	Fuel cells
	H01M8	Fuel cells; Manufacture thereof
	H01M2250	Fuel cells for particular applications; Specific features of fuel cell system
	Y02T90/3	Application of fuel cell technology to transportation

[1] The term “hydrogen” is a very general term as it is a chemical element. Therefore, the term appears frequently in patents with very little relevance to the concept of hydrogen technology in this study. Moreover, it is hard to define specific keywords for all fields of hydrogen technology.

[2] According to USPTO, section Y refers to: GENERAL TAGGING OF NEW TECHNOLOGICAL DEVELOPMENTS; GENERAL TAGGING OF CROSS-SECTIONAL TECHNOLOGIES SPANNING OVER SEVERAL SECTIONS OF THE IPC; TECHNICAL SUBJECTS COVERED BY FORMER USPC CROSS-REFERENCE ART COLLECTIONS [XRACs] AND DIGESTS.



	hydrogen	gas	generat	...	stack
Patent 1	5	7	0	...	0
Patent 2	4	2	9	...	7
Patent 3	10	0	2	...	0
...	...	...	...	...	...
Patent N	9	0	7	...	14

Fig. 1. Example of document-term matrix derived from patents.

4.2. Data refining

After collecting hydrogen technology patents, we converted the text in the titles and abstracts of these patents into the input form required by topic models. Specifically, topic models require input data in a DTM (document-term matrix) format, which shows the term frequency of each term (word) for each document. This is because topic models identify latent topics that exist within the document set based on the co-occurrence pattern of words. An example of DTM is shown in Fig. 1. In DTM, each row represents one document (patent), each column represents each word, and each cell represents the term frequency of each word in each document (patent). In this study, we conducted pre-processing to select the words that best explain the hydrogen technology sector. First, we stemmed the words with various forms to group words with similar meaning (e.g., generates, generation, generating → generate). Next, we excluded the words that appeared in less than 1% of the total patents collected to reduce bias. Finally, we excluded general terms that had no meaningful implications. After preprocessing, we formed a 507-word DTM of 17 281 hydrogen technology patents from 2010 to 2019.

4.3. Technology topic identification using an STM

The STM is a probabilistic topic model proposed by Roberts et al. [5]. It is an extended version of the most widely used topic model, LDA [22]. The general structure of probabilistic topic models, including STM and LDA, is presented in Fig. 2. In a probabilistic topic model, the document set is defined as a probabilistic relationship between the three-level hierarchy of word-topic-document. To be specific, each document can be represented as a probabilistic distribution of topics, and each topic can be represented as a probabilistic distribution of words. Considering that a document set typically comprises a few hundreds to thousands of words, a probabilistic topic model can also be considered as a dimension-reduction method that defines topics (fewer than words) as a probabilistic distribution of words and uses defined topics to explain documents. Moreover, a researcher should determine the theme of each topic by observing the top words in each topic, which makes this method an exploratory, unsupervised learning approach.

In probabilistic topic models, key parameters estimated are: word distribution in each topic (beta–blue lines in Fig. 2) and topic distribution in each document (theta–orange lines in Fig. 2). By observing the estimated betas for each topic, the researcher may infer the theme of each topic[3]. Moreover, by observing the estimated thetas for the document set, the researcher may analyze the topic prevalence pattern of the document set. In LDA, a homogeneous multivariate distribution (Dirichlet distribution) is assumed as a prior distribution for estimating betas and thetas. In contrast, an STM assumes a heterogeneous distribution by considering a continuous distribution, allowing diversion based on a document's metadata. Thus, an STM can incorporate the influence of a document's metadata in parameter estimation.

Thus, this study incorporated variables in Table 2 as covariates that influence parameter estimation. We selected covariates to analyze three key factors in hydrogen technology: newness (trending), impact, and national technological focus. First, the year of patent registration was considered to capture the newness (trend) of patents. Next, the adjusted citations by other patents[4] were considered to capture the impact of patents. Finally, dummy variables representing the big five countries with the largest number of hydrogen technology patents for the last decade (the United States, Japan, Korea, Germany, and France) were considered to capture the national technological focus.

These covariates are incorporated in the model to influence topic estimation, and the statistical significance of covariates in topic prevalence can be analyzed through a simple post-estimation analysis. Specifically, linear regression was conducted to check the significance of covariates on topic prevalence by setting the topic prevalence of each topic as a dependent variable and covariates in Table 2 as independent variables. For example, if regression analysis for the prevalence of topic 1 suggests that the parameters for registration year and Korea dummy variable are significant with a positive value, it can be inferred that topic 1 is a newer (trending) topic whose prevalence increases by time and is more easily found (prevails) in patents from Korea, compared to other countries[5]. Moreover, if a certain topic prevails in patents from a specific country, it can be assumed that the country is focusing on that specific technological topic. Based on such results of the STM, this study aimed to not only identify specific technological topics in the hydrogen technology sector but also analyze the newness, impact, and national technological focus for each topic identified.

4.4. Technology maturity rate estimation

In this review, we assess technology maturity rate (TMR) of the identified topics from the patents. Technology maturity can be measured using S-curve models including Gompertz model, which is suitable for

[3] For example, if estimated betas suggest that words such as storage, vessel, pressure, and compress showed a high proportion in a specific topic, it can be inferred that this topic is related to pressure vessel for hydrogen storage.

[4] As patents registered in the past had more time to be cited by other patents than recent patents, it is unfair to use the absolute number of citations by other patents as a covariate to represent the impact of patents. Therefore, this study adjusted the number of citations by other patents based on annual average citations by other patents. Specifically, we first calculated the annual average citations by other patents for patents registered between 2010 and 2019. Next, we divided each patent's citations by other patents with the annual average value of the patent's registration year. By conducting such adjustments, we were able to identify more impactful patents within same registration year. For example, average citations by other patents was 4.16 for patents registered in 2010, and 0.90 for patents registered in 2015. Then, even if patent A, which was registered in 2010, and patent B, which was registered in 2015, may have same number of citations by other patents ($n = 10$), their adjusted citations by other patents differ significantly as 2.40 ($10 \div 4.16$) for patent A and 11.11 ($10 \div 0.90$) for patent B.

[5] Results of post-estimation analysis for this study are presented in ESI, Supplementary note 2.

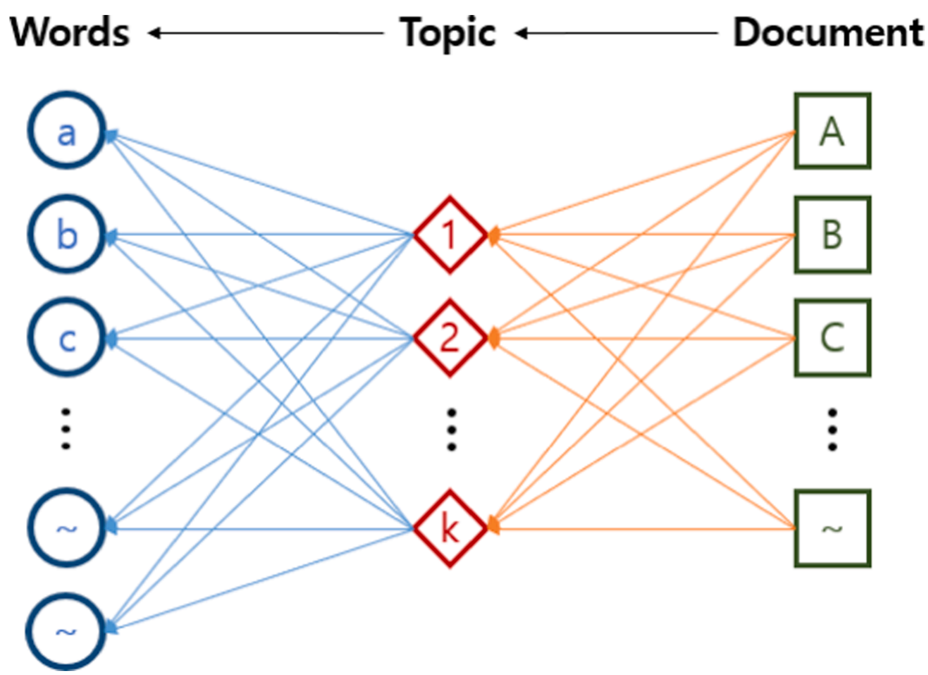


Fig. 2. General structure of a topic model.

Table 2
Covariates considered in the structural topic model.

Covariate	Variable type	Description	Linked factor
Year of registration	Continuous	Timing of patent registration in USPTO (year)	Newness (trending)
Adjusted citations by other patents	Continuous	Normalized value of the number of citations by other USPTO patents (normalized via annual average citations by other USPTO patents)	Impact
United States	Dummy	Dummy variable which has a value of 1 when the patent is registered by an assignee from the United States and 0 otherwise	National technological focus
Japan	Dummy	Dummy variable which has a value of 1 when the patent is registered by an assignee from Japan and 0 otherwise	
Korea	Dummy	Dummy variable which has a value of 1 when the patent is registered by an assignee from Korea and 0 otherwise	
Germany	Dummy	Dummy variable which has a value of 1 when the patent is registered by an assignee from Germany and 0 otherwise	
France	Dummy	Dummy variable which has a value of 1 when the patent is registered by an assignee from France, and 0 otherwise	

fitting the S-curve of patent growth [38–40]. The Gompertz model is defined by equation (1).

$$P(t) = me^{-ae^{-bt}} + \varepsilon_t \tag{1}$$

where $P(t)$ represents the cumulative number of patents that belong to each identified topic at time$t$,a,b, and m are parameters of the model. Here m represents the upper limit (i.e., saturation level) of S-curve. These parameters can be estimated by the non-linear least squares method. To calculate the cumulative number of patents for each topic, we multiplied the annual total number of topics with topic proportion each year to get annual number of patents for each identified topic.

Based on the estimated parameters, the TMR of identified topics can be calculated using equation (2) using the estimated upper limit$m$..

$$TMR(t) = \frac{P(2019)}{m} \tag{2}$$

where $P(2019)$ represents the cumulative number of patents of each identified topic at year 2019 (the last year of the patent data we collected). The TMR has a value between 0 and 1, and indicates how much a technology has approached its maximum level of development [38,39,41]. Therefore, lower TMR implies there is more potential for the topic for future development.

4.5. Visualization

While the interpretation of identified topics itself can provide useful implications, this study aimed to provide additional intuitive results and implications using visualization approaches, with a focus on big five countries leading in hydrogen technology (the United States, Japan, Korea, Germany, and France). To be specific, this study attempted to 1) visualize the topic prevalence pattern of big five countries and compare the national technology portfolio and 2) visualize the correlation between topics to propose implications on the national R&D strategy. For the first approach, we compared the national technological focus using the radar chart and normalized topic prevalence for patents from each country to provide intuitive visualization results for decision makers. For the latter approach, we used the co-occurrence pattern of topics in patents to visualize the correlation between technology topics in a graph. This aspect cannot be analyzed using conventional LDA, which assumes independence between topics, but can be analyzed using an STM, which allows for a correlation analysis between topics. By utilizing such visualization approaches, this study aimed to identify key technology topics for each country, compare national technology focus, and conduct technology correlation analysis to implicate national R&D strategy.

5. Results and discussion

Results of this study are presented in the following order. First, we present the descriptive results of the hydrogen technology patents collected to provide general information about the current status of the hydrogen technology sector (Section 4.1). Subsequently, specific technological topics in the hydrogen technology sector were identified using an STM, and the details about key technology topics that were trending or impactful are presented (Section 4.2). Next, a comparison of technology portfolios and key technologies for big five countries was conducted (Section 4.3). Finally, a technology correlation analysis considering the correlation between identified topics was conducted to provide implications on national R&D strategy (Section 4.4).

5.1. Descriptive results

This study collected 17 281 hydrogen technology patents registered from 2010 to 2019 based on the CPC codes presented in Table 1. Before presenting the results of the STM analysis, this section provides key descriptive results to provide a general status of the hydrogen technology. First, the rankings of the patents registered in various technology fields are provided in Fig. 3[6]. The nationality of patents was determined by the nationality of the first named assignee. The results showed high technological capability for the United States and Japan, followed by Korea, Germany, and France. To be specific, the United States ranked first overall and in the production and storage & distribution fields. In the application field, the United States ranked second, following Japan. Korea, Germany, and France ranked third to fifth depending on the fields; whereas Korea showed competency in the application field, France showed competency in the production and storage & distribution fields. Overall, patents from these big five countries in the hydrogen technology sector accounted for 84% of the total patents registered.

Next, the patents collected from the USPTO also include specific information about the assignee for each patent. Table 3 provides rankings for the top five firms with the highest number of patents registered. The results showed that the firm with the highest number of hydrogen technology patents registered was Toyota (Japan), followed by GM (United States), Honda (Japan), Hyundai (Korea), and Samsung SDI (Korea). As one can observe, global automotive companies led the technology development, followed by a battery manufacturer (Samsung SDI). This is likely because patents in hydrogen application (fuel cell) for vehicles (fuel cell electric vehicles, FCEVs) constitute an overall large proportion in the hydrogen technology sector. One point to note is that whereas the number of patents in total was the highest for the United States and Japan, with a significant gap compared with other countries, two Korean firms ranked among the top five firms. This implies that Korea's hydrogen technology development is led by a few key companies, compared to other countries.

While the hydrogen application field showed the same top five companies with the overall ranking, other fields showed different companies in the top five ranks. In the hydrogen production field, conventional oil & gas companies, such as Shell Oil Company (United States), Air Products and Chemicals (United States), and Praxair Technology (United States), were observed along with a relatively new company that specializes in the hydrogen sector, Intelligent Energy (United Kingdom). In contrast, the hydrogen storage & distribution field was led by both global automotive companies that led the application field and oil & gas companies that led the production field. This is because the storage & distribution field is closely related with both production and application fields as a supplementary technology (e.g., technologies that produce hydrogen and store them in other forms, technologies that can stably distribute hydrogen to FCEVs, etc.). Finally, many companies were observed at the top ranks in multiple fields. For example, Honda ranked fifth in the production field and third in both storage & distribution and application fields, showing its interdisciplinary competency throughout multiple hydrogen technology fields.

As presented in this section, descriptive results can also provide insights into the hydrogen technology sector. However, such analysis cannot provide more practical results that can be used to design an R&D strategy, such as emerging technologies in the field, national technological focus, or correlation between different technologies. Moreover, considerable time and effort are required to manually review the content in all patents. Therefore, this study aimed to conduct a text analysis of the title and abstract of the patents using an STM to identify latent technology topics within the hydrogen technology sector. The identified technology topics were then used to identify the national technological focus and to analyze the technology correlation pattern in an unsupervised, semiautomated manner.

5.2. Technology topic identification

This section provides the key results of the technology topics identified using the STM. In most topic models, including STM and LDA, a researcher must determine the number of topics in a document set before the estimation, and this is one of most difficult parts of using topic models. There is no distinctive approach for determining the optimal number of topics (K). Some previous literature determined K by consulting experts or referencing previous literature [26]. In contrast, other studies compared multiple model performance indices of models with different K values to determine the preferred K value [31,32]. This study followed the latter approach by comparing the held-out likelihood, residual, and semantic coherence of models with different K values, increasing from 5 to 100, with an increment of 5. Semantic coherence is an index used in various text analytics, with a high value implying a high co-occurrence tendency of top (probable) words in a given topic [34]. Moreover, this index correlate with s well with human judgment for topic quality. Analysis of the preferred K value showed that whereas the held-out likelihood and residual improved, the semantic coherence got worse, as the K value increased. However, as K gets greater than 40, the improvement in held-out likelihood and residual diminishes, whereas the semantic coherence decreases significantly. Therefore, this study assumed there were 40 topics in the hydrogen technology patents collected ($K = 40$), considering the balance between the considered indices.

In most topic models, researchers observe the word distribution within a topic to identify the theme of each topic. While words with higher absolute probability (PROB) can provide intuitive implications, previous studies also used the FREX (Frequency and Exclusivity) index to identify the theme of topics. The FREX index is higher for the words that have a higher probability within a specific topic (frequency) compared to other topics (exclusivity) [35,36]. The FREX index can also easily identify the theme of each topic as key technological words tend to have high FREX values. This study observed top 10 words in both PROB and FREX for each topic to determine the theme of each topic. The themes of topics were first identified by the authors and then confirmed by multiple experts in the energy field. Then, we fitted the cumulative number of patents of each identified topic from 2010 to 2019 using the Gompertz model and calculated the TMR of each topic.

Owing to space constraints, this section provides the summarized results of key topics. We have total of 40 topics identified, but we only summarize either trending or impactful topics in the main text. Details of all 40 topics, including the top 10 words in PROB and FREX, are presented in ESI, Supplementary note 1 and Table S1. We selected the key emerging technologies based on the post-estimation results of STM. To be specific, we checked the significance of covariates presented in Table 2. Detailed results of the post-estimation analysis are presented in ESI, Supplementary note 2. Using the results of the post-estimation

[6] Number of patents by the fields do not exactly add up to be overall number of patents as many patents are included in multiple fields (i.e., they have multiple CPC codes from different fields).

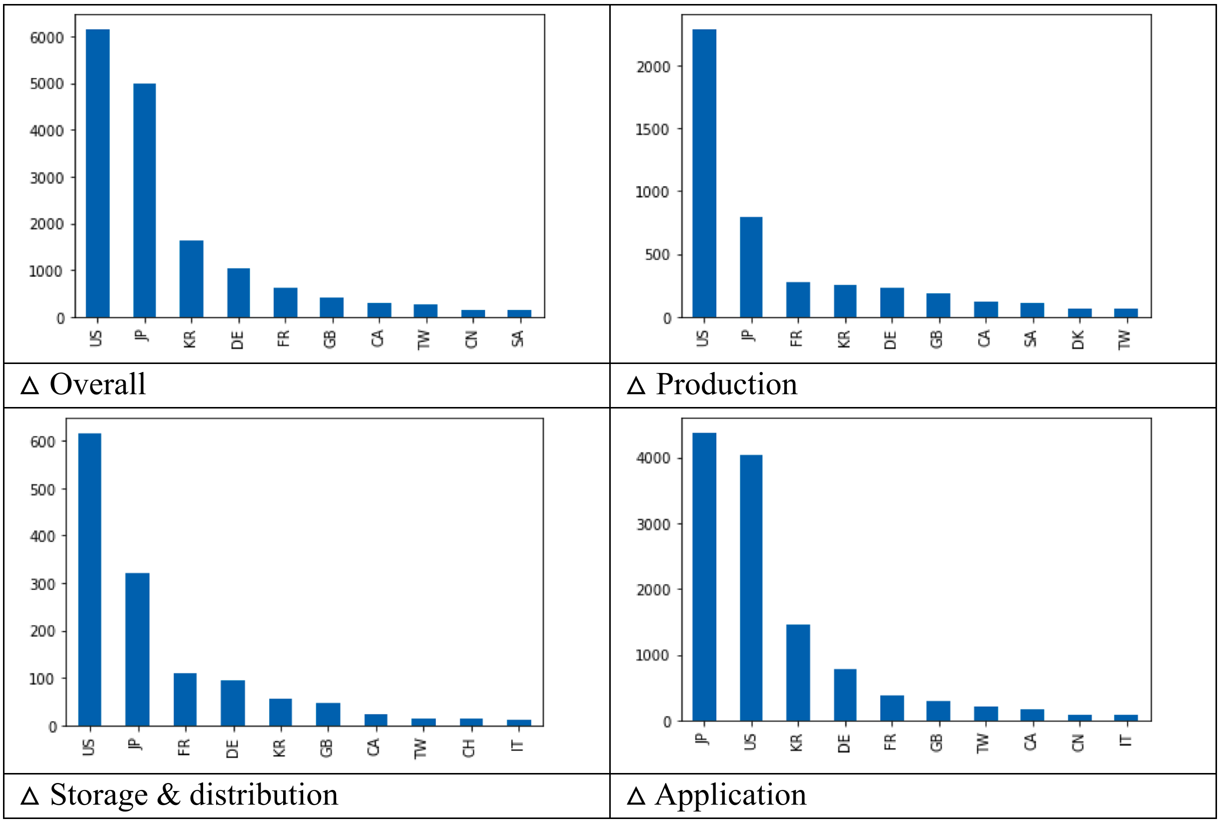


Fig. 3. Top 10 countries with the highest number of hydrogen technology patents registered from 2010 to 2019. Note: US – the United States, JP – Japan, KR – Korea (Republic of), DE – Germany, FR – France, GB – the United Kingdom, CA – Canada, TW – Taiwan, CN – China, SA – Saudi Arabia, DK – Denmark, CH – Switzerland, AT - Austria, IT - Italy.

Table 3
Top five firms with the highest number of hydrogen technology patents registered from 2010 to 2019.

Rank	Total	Production	Storage & distribution	Application
1	Toyota Jidosha Kabushiki Kaisha (Japan)	Praxair Technology, Inc. (United States)	Toyota Jidosha Kabushiki Kaisha (Japan)	Toyota Jidosha Kabushiki Kaisha (Japan)
2	GM Global Technology Operations LLC (United States)	Intelligent Energy Ltd. (United Kingdom)	GM Global Technology Operations LLC (United States)	GM Global Technology Operations LLC (United States)
3	Honda Motor Co., Ltd. (Japan)	Shell Oil Company (United States)	Honda Motor Co., Ltd. (Japan)	Honda Motor Co., Ltd. (Japan)
4	Hyundai Motor Company (Korea)	Air Products and Chemicals, Inc. (United States)	Intelligent Energy Ltd. (United Kingdom)	Hyundai Motor Company (Korea)
5	Samsung SDI Co., Ltd. (Korea)	Honda Motor Co., Ltd. (Japan)	Air Products and Chemicals, Inc. (United States)	Samsung SDI Co., Ltd. (Korea)

analysis, we identified trending (newer) topics whose proportion in the document set increase by time[7] and impactful topics that are more frequently cited by other patents[8]. To be specific, 11 trending topics (2, 7, 9, 11, 14, 19, 25, 27, 30, 31, and 36) and five impactful topics (2, 7, 11, 12, and 29) were identified, whereas three (2, 7, 11) were both trending and impactful. The 13 key hydrogen technology topics either trending or impactful were grouped with larger topics, and their summarized theme was provided along with the information about their trending or impactful nature and the TMR in Table 4. Moreover, the list of key countries for each topic are also provided based on the parameter estimates for country dummies in the post-estimation analysis.

The results showed that three topics were related to hydrogen production and 10 topics were related to fuel cell (hydrogen application). Those related to fuel cell could be further divided into those related to manufacturing (5), management system (3), and new types (2). Topics related to hydrogen storage and distribution (e.g., topic 10, 16, 22, 38, …) were not included in Table 4. This implies such technologies did not significantly increase by time, or got significantly cited more by others. However, detailed results per topic in Supplementary note 1 suggest these topics also did not decrease by time or cited less by others. This implies that such topics have a stable position in the hydrogen technology sector.

First, the topics related to hydrogen production were hydrogen production technologies using different sources, namely biomass (topic 2), natural gas (topic 9), and ammonia (topic 19). Specifically, topic 2

[7] A topic is considered trending if the parameter estimate for the registration year in the post-estimation analysis has a significant positive value.

[8] A topic is considered impactful if the parameter estimate for the adjusted citations by other patents in the post-estimation analysis has a significant positive value.

Table 4
Key emerging hydrogen technology topics identified.

Group		Topic	Trending	Impactful	TMR	Key countries
Hydrogen production		2: Hydrogen production from biomass	O	O	69.7%	
		9: Hydrogen production from natural gas	O		71.5%	
		19: Hydrogen production from ammonia	O		68.5%	
Fuel cell (hydrogen application)	Manufacturing	14: Fuel cell coating solution for metal parts	O		64.5%	
		27: Sandwich structure of electrodes and electrolytes for fuel cell	O		62.1%	KR > JP
		30: Fuel cell blocks with plurality	O		65.1%	DE
		31: Fuel cell body frame assembly	O		56.2%	JP > KR
		36: Fuel cell sealing	O		70.3%	JP > DE > US
	System	12: Electric charging and storage for FCEVs		O	67.5%	DE > JP > US
		25: Fuel cell control system	O		69.4%	JP > KR > US
		29: Engine systems for better efficiency		O	69.9%	US
	New types	7: Redox flow battery	O	O	53.4%	JP > US
		11: Lithium-air battery	O	O	37.6%	KR > DE > US > JP

Note: US – the United States, JP – Japan, KR – Korea (Republic of), DE – Germany, FR – France.

was about generating hydrogen from biomass (or waste) through thermal decomposition and gasification; topic 9 was about gaining hydrogen from methane (CH_4), the core constituent of natural gas; and topic 19 was about separating hydrogen from ammonia (NH_3). Among these topics, hydrogen separation from ammonia was also related to the storage & distribution of hydrogen as ammonia is easier to store and distribute than hydrogen because of its higher storage density and better infrastructure. For example, one may produce hydrogen, store it as ammonia, transport it to demand location, and then separate hydrogen from ammonia near the demand location.

Next, half of the key fuel cell-related topics were related to fuel cell manufacturing. Various topics related to coating, sealing, and assembly of fuel cells were identified, and these topics were trending but not impactful. However, the topics related to the management system of fuel cells, for example charging and storing electricity for FCEVs (topic 12) or engine systems for better efficiency (topic 29), were relatively more impactful. Finally, the topics related to next-generation new types of fuel cell (topic 7 and 11) were both trending and impactful.

Moreover, three topics identified as both trending and impactful were those related to 1) hydrogen production from biomass (topic 2), 2) redox flow battery (topic 7), and 3) lithium-air battery (topic 11). In other words, the trending and impactful topics in the hydrogen technology sector were related to the environment-friendly production of hydrogen and new types of fuel cells. While other green production technologies, including electrolysis using renewable energy, are still in an early stage, hydrogen production from biomass is currently gaining huge attention as it is a more viable option compared to others. Moreover, the new fuel cell types had different strengths and application areas compared to a polymer electrolyte membrane (PEM) battery, which is dominant in the current market (mainly used for batteries for FCEVs). For example, a redox flow battery (topic 7) has competency when the desired capacity for energy storage gets bigger, and thus it is gaining attention from the large-scale ESS market. In contrast, a lithium-air battery has competency in its high energy density and lightweight, and thus it is gaining attention from the aviation industry, including the drone market.

The results of calculating TMR for all 40 topics, showed a minimum value of 37.6%, maximum value of 88.2%, median value of 70.1%, and lower Q1 value of 65.7%. The results show that emerging topics in Table 4 generally showed lower TMR, and those related to fuel cell manufacturing or new type of fuel cells an even lower TMR. Such results imply that emerging topics in Table 4 usually also had high future potential. Among top 10 topics with the lowest TMR, 7 were already included in Table 4. The remaining three topics not included in Table 4 were: topic 38 (inlet & outlet filter for fuel cell), 8 (platinum catalyst for fuel cell), and 15 (fuel cell separator molding).

Finally, based on the results of STM post-estimation analysis, we can also identify which of the big five countries focus on the key topics identified[9]. Results show strong competency of Japan on fuel cell-related topics, followed by Korea and Germany. No countries were found to exhibit a clear focus on the topics related to hydrogen production other than the big five countries; the United States showed high focus for topics 2 and 14, whereas France showed high focus for topic 19 compared to other big five countries when we observe the topic probability of patents from each country. Based on the results of this section, the analysts and decision makers of each country may identify key technology topics and potential competitors for each topic in a semi-automated way.

5.3. National technology portfolio

This section aims to analyze how national technology portfolio differs among big five countries. To observe how each country focuses on each technology topic more comprehensively, we used radar charts for visualization. Before drawing radar charts, we first constructed a dendrogram of 40 topics based on the similarity of word distribution within topic, to group topics with similar content together. Subsequently, we placed topics in the radar chart using the results of the dendrogram to flock similar topics together. Moreover, when representing topics in the radar chart, we normalized the value by defining the topic focus index. For example, topic 3 has a proportion of 2.48% for all patents collected, whereas it has a proportion of 5.03% for the patents from Korea. Here, a topic focus index of topic 3 for Korea is 2.02 ($5.03 \div 2.48$). We calculated the topic focus index for all topics for the big five countries and then drew a radar chart for each country. The results are presented in Fig. 4.

The results intuitively show the national focus by technology sector. As we placed topics using the dendrogram based on word distribution similarity between topics, similar topics flock together in the radar chart. To be specific, the upper left side comprises the topics related to fuel cell management systems, the straight left side comprises the topics related to storage & distribution of hydrogen, and the bottom left side comprises the topics related to hydrogen production. The right side of the chart comprises the topics related to fuel cell manufacturing; whereas the upper side is more related to the chemical part, the bottom side is more related to the mechanical part of fuel cell manufacturing. Considering such allocation of topics, one can easily get the general view of each country's technology portfolio in the hydrogen technology

[9] A topic is considered to be focused by the big five countries if parameter estimate for country dummies in the post-estimation analysis has a significant positive value.

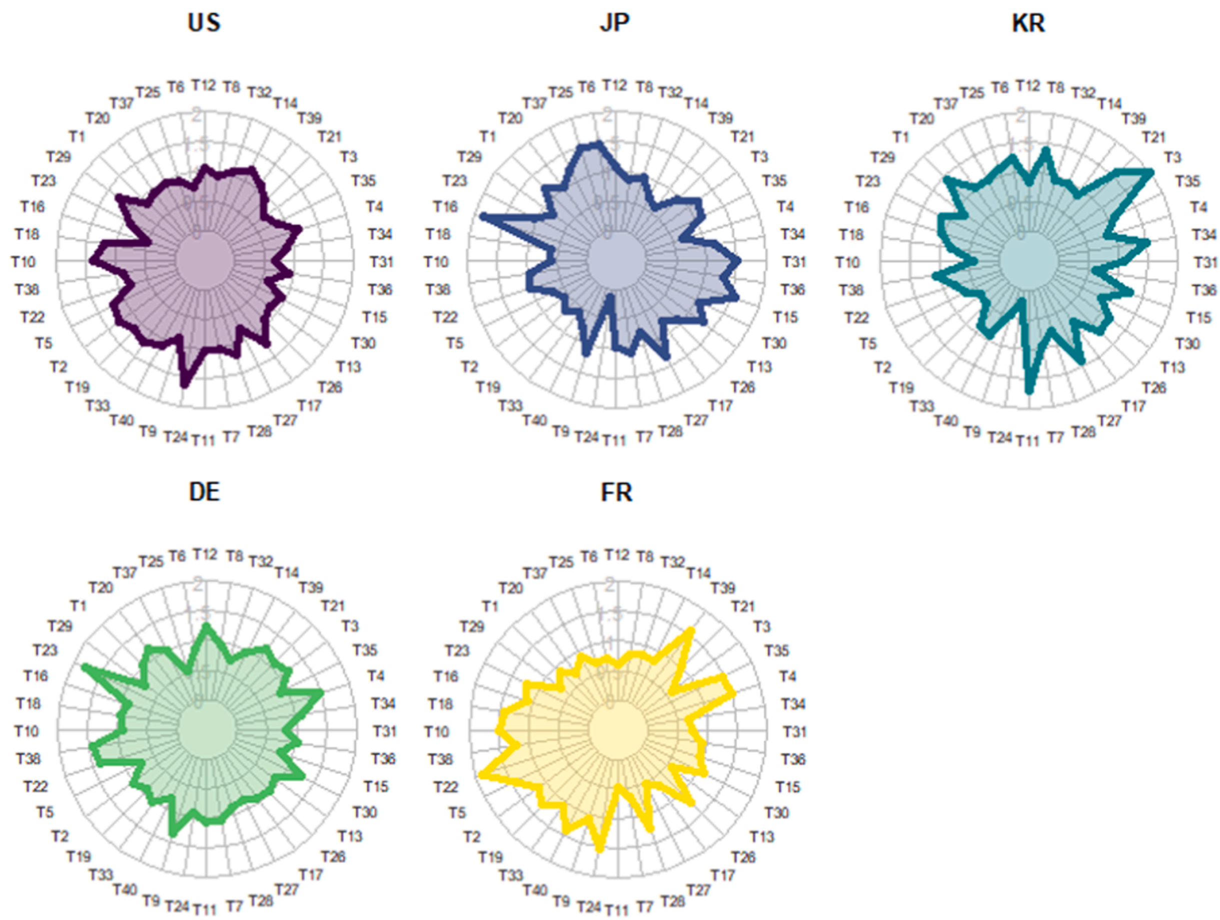


Fig. 4. Technology profile comparison between key countries.

sector.

First, the United States showed balanced technology portfolio, while mildly focusing on hydrogen production (bottom left) and storage (straight left). In contrast, Japan strongly focused on some specific fields in hydrogen technology, such as fuel cell management system (upper left) and mechanical part of fuel cell (bottom right). Owing to this, Japan focused little on the chemical part of fuel cell and hydrogen production, storage, and distribution fields (straight left and bottom left). Korea focused on fuel cell manufacturing and management systems, while lying low stress on hydrogen production, storage, and distribution, as done by Japan (straight left and bottom left). Germany also showed balanced technology profile like the United States, while focusing on some of the selected topics in each field. In contrast, France strongly focused on hydrogen production (bottom left), storage (straight left), and chemical part of fuel cell manufacturing (upper right), without focusing much on fuel cell management system (upper left) and mechanical part of fuel cell manufacturing (bottom right).

Besides the results obtained from Fig. 4, it is important to elucidate specific details about the topics that countries specifically focus on. Table 5 shows details about the top three topics with the highest topic focus index for each of the big five countries. In the notes, we have marked the topics trending/diminishing or impactful/impactless (based on the post-estimation results of STM) and those with high topic focus index rank (within 5) from the other big five countries.

The results showed that the top topics of focus for the United States were related to hydrogen production, storage, and energy efficiency. While the topic related to energy efficiency (topic 29) was impactful, the United States should be warned about topic 24 and 10 since they are also being focused upon by France. All the top topics of focus for Japan and Korea were related to fuel cell, and they did not overlap with other countries. While some of these topics were trending (topics 31 and 11) or impactful (topic 11), a topic focused upon by Korea was diminishing by time (topic 3). Germany also showed the biggest focus on fuel cell-related topics. However, one topic of focus (topic 4) was also focused upon by France and had low impact; therefore, this point should be considered in the country's R&D strategy. Finally, France showed the biggest focus on hydrogen production, storage, and chemical part of fuel cell. However, it should be aware of the result that one of the topics of focus is diminishing by time (topic 39) and other topics are being focused upon by other countries (the United States and Germany).

5.4. Technology correlation pattern analysis

In topic models, each document can be explained with a probability distribution of multiple topics. In other words, multiple topics may coexist in a single patent. Such coexistence of topics may follow a certain pattern depending on technological and other factors of these technology topics. This section aims to visualize such correlation patterns along with the national technological focus, to provide R&D strategy implications in the hydrogen technology sector. Such analysis cannot be performed with LDA, since LDA assumes independence between the identified topics. However, as this study used an STM, we could detect such correlation between topics to conduct a correlation pattern analysis. Moreover, this study also used the results of the post-estimation of STM to select and visualize a single focusing country among the big five

Table 5
Hydrogen technology topics focused by big five countries.

	Topic number (topic focus index)	Topic	Notes
United States	24 (1.61)	Reactor for syngas stream (for hydrogen production)	*Ranked #3 in France*
	10 (1.36)	Pressure vessel for hydrogen storage	*Ranked #5 in France*
	29 (1.28)	Engine systems for better efficiency	Impactful
Japan	16 (1.84)	Valve and discharging technology for hydrogen supply	
	15 (1.58)	Fuel cell separator molding	
	31 (1.52)	Fuel cell body frame assembly	Trending
Korea	3 (2.02)	Sulfonic acid polymer for fuel cell	Diminishing
	11 (1.70)	Lithium-air battery	Trending & impactful
	21 (1.59)	Ion-exchange (semi-permeable) membrane for fuel cell	
Germany	23 (1.80)	Fuel cell cooling & coolant	
	4 (1.49)	Bipolar plates (for fuel cell)	Impactless *Ranked #4 in France*
	38 (1.44)	Inlet & outlet filter for fuel cell	
France	22 (1.90)	Hydrogen liquefaction and storage	*Ranked #4 in Germany*
	39 (1.58)	Stable composition of chemical compounds	Diminishing
	24 (1.54)	Reactor for syngas stream (for hydrogen production)	*Ranked #1 in the United States*

countries for each topic[10].

Visualization results are presented in Fig. 5. Fig. 5-(a) on the left side shows only topic correlation, and Fig. 5-(b) on the right side shows the national technological focus. Each node represents each topic, and the edge between the two nodes indicates that two topics appear in the same patent above a certain level. We used the method proposed by Zhao et al. to estimate high-dimensional undirected graphs [37]. Briefly, we estimated the graph based on semiparametric transformation of topic proportions. The results show some clusters are formed based on the co-occurrence pattern of topics. Specifically, we identified five clusters, which were related to 1) hydrogen production, 2) hydrogen storage & distribution, 3) fuel cell–chemical, 4) fuel cell–mechanical, and 5) fuel cell–system. Moreover, we also highlighted key technologies mentioned in Table 4, by marking trending topics with red arrow and impactful topics with yellow star. For TMR, we used the color of nodes to represent TMR level and marked lower Q1 topics with blue diamond.

Fig. 5 shows the overall view of technology competition between key countries for each technology sector. First, it could be observed that the topic nodes linked together tend to be focused by the same country. Such patterns emerge since multiple technology topics frequently appear together in a single patent because these technology topics have high technological (prior or resulting technology) or infrastructural (sharing the same infrastructure) correlation, which makes them likely to be developed by the same organization. With such a viewpoint, a country seeking to expand its technological focus may consider expanding to topics adjacent to topics that the country is already focusing on. For example, if Korea wants to put more focus on fuel cell–mechanical field technologies, one may consider topic 4 or 31, which is adjacent to topic 34 already focused on by Korea. Moreover, some kind of surround attack strategy can be used. For example, if Japan wants to increase focus on fuel cell–system field, it can preferentially consider expansion to topic 12, which is connected to two topics that Japan focuses on (topic 6 and 25). Finally, the results in Fig. 5 can also identify key rivals for each technology fields. For example, France may consider different key rivals for each field, namely the United States for hydrogen production, Germany for hydrogen storage, and Korea for fuel cell–chemical, when planning its R&D strategy.

Finally, a few stand-alone topics not connected with any other topics were also observed. Among these stand-alone topics, the trending ones were likely to be new technologies, different from the existing ones. These included topics 7 and 11 in this study that were topics related to redox flow and lithium-air fuel cell battery, whereas the PEM fuel cell is a dominant fuel cell technology. Moreover, these topics were highly impactful topics (cited more frequently by others). As presented in this section, a correlation analysis between the identified topics and corresponding visualization can provide various implications based on both clustered and stand-alone topics.

5.5. *Categorization of identified topics and R&D decision making of policymakers*

In this section, we aim to provide a framework that can help policymakers decide R&D directions. Specifically, considering whether a topic is trending or impactful, its MTR level, and technology portfolio of the country of concern, we propose a decision tree for policymakers to help their R&D decision making (Fig. 6). Specifically, we first divide a given topic by whether it's an emerging (trending or impactful) technology topic. Then we divide it again depending on the given topic's TMR. Here we used 65%, which is a lower Q1 level for the identified topics. Then, for emerging topics with low maturity, we divide again by considering whether the topic is already being focused on by the country of concern. If the country is not focusing on the given topic, we divide again by whether the given topic is correlated (those connected with edges in Fig. 5) with focused topics of the country of concern.

Using such a decision tree, we can categorize identified topics into six types, and we focused on three types (A, B, and C) which may provide important implications to policymakers. First, type A topics are emerging (trending or impactful) and high potential (low TMR) topics already being focused on by the country of concern. For these topics, the country of concern should put continuous focus and not to lose their focus as it has large remaining potential. For Korea, examples of such type A topics are topics 11 (Lithium-air battery) and 27 (sandwich structure of electrodes and electrolytes for fuel cell). Next, type B topics are emerging and high potential topics, but not currently being focused on by the country of concern. However, it is correlated with other focused topics which may make it easier to access. Therefore, type B topics can be considered as key topics the country of concern should first consider expanding based on its current technological portfolio. For Korea, examples of such type B topics are topics 14 (fuel cell coating solution for metal parts) and 31 (fuel cell body frame assembly). Finally, type C topics are not emerging but have high potential. While it is not trending or impactful, countries should monitor these topics as it has low TMR which implies high potential. Examples of such topics are topic 8 (platinum catalyst for fuel cell), 15 (fuel cell separator molding), and 38 (inlet & outlet filter for fuel cell).

[10] In order to define a single focusing country for each topic, this study used the following conditions to determine a specific country as a focusing country: 1) the post-estimation results suggest that a parameter for the specific country dummy is significant with positive value and its value is bigger than significant parameters (if any) for other country dummies; 2) post-estimation results suggest that parameter for the specific country dummy is not significant but parameters for all other country dummies are significant with negative value; 3) post-estimation results suggest that all parameters for country dummies are significant with negative value and the parameter value of a specific country is the largest (smallest absolute value). If none of these conditions were met for a topic, we did not specify a focusing country for that topic.

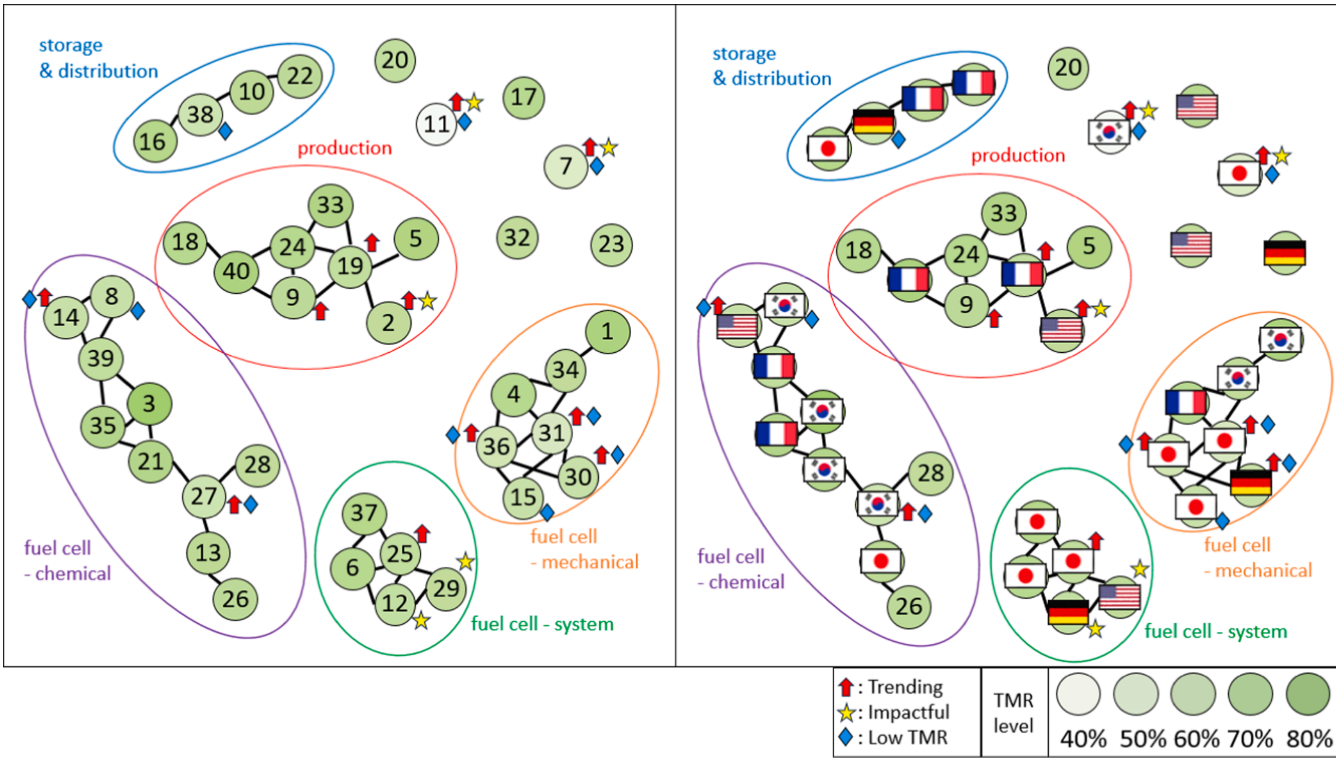


Fig. 5. Topic correlation and national technological focus. Note: Topics with red arrow are those whose proportion increase by time, topics with yellow stars are those with higher number of citations per year, topics with blue diamonds are those with low technology maturity rate (TMR). Brightness of nodes are determined by the TMR level. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

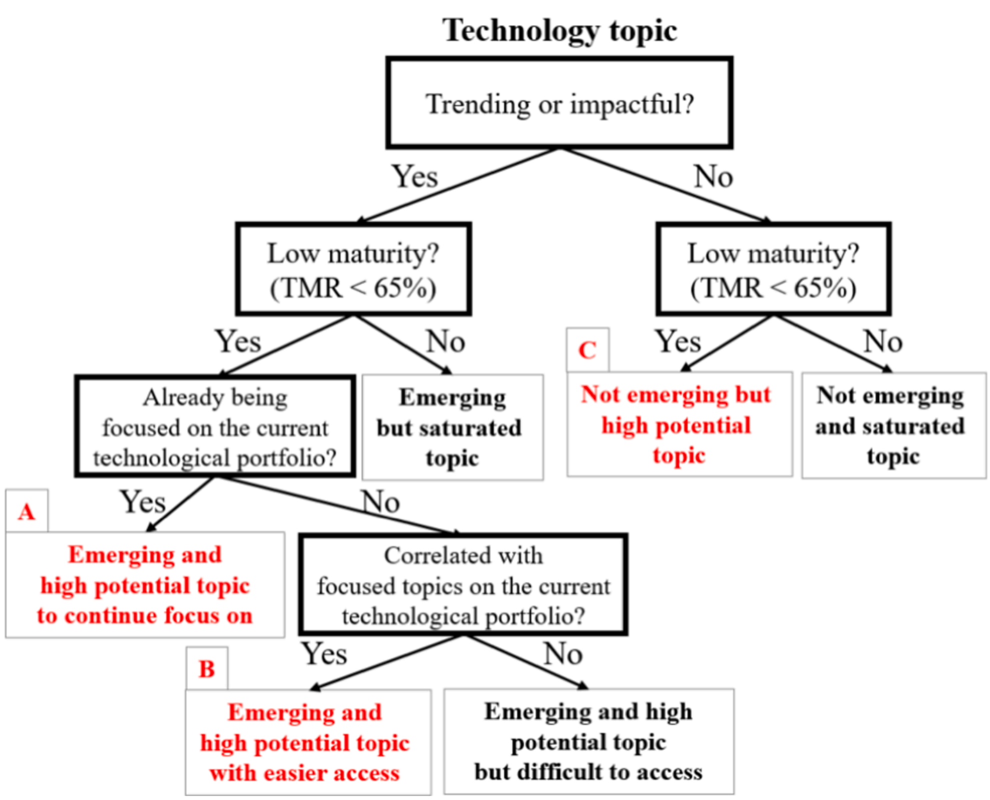


Fig. 6. Decision tree for identified technology topics. Note: TMR is abbreviation for technology maturity rate. 65% is determined by the lower Q1 among whole topics.

6. Conclusion

In this study, we used STM to identify latent technology topics from patents for hydrogen technologies and proposed a method to analyze whether a certain topic is more trending or impactful compared to others and whether a certain country focuses on a specific topic. Moreover, we also investigated TMR of identified topics and proposed decision frameworks for policymakers to help their R&D direction decisions.

The framework proposed in this study may help hydrogen technology policymakers in their decision making. When the scope of technology is wide, as in case of hydrogen technology, and when it is difficult to obtain opinions on technology comprehensively from a small number of experts, the proposed semiautomated approach based on scientific methods may save considerable time and costs. Furthermore, the proposed framework enables the visualization of key analysis results in a more familiar and intuitive form to assist effective decision making. Finally, many new technology fields that have recently attracted attention are convergence technologies that encompass diverse fields and concepts, in a manner similar to the hydrogen technology. Therefore, this framework can be applied to other promising technology fields in the future.

The key contribution of this study is that it proposed an analysis framework that can derive various implications by focusing on the content aspect of patents. In the future, various follow-up studies may be conducted by extending the framework proposed in this study. Specifically, more detailed analysis may be performed focusing on a certain country or region, or studies may be conducted by dividing an entire time period into several phases and observing time-series changes, as described by Kim et al. [26]. Furthermore, whereas this study used the number of citations by other patents as a covariate to represent the impact of patents, future studies may extend their investigation by incorporating more detailed citation information about patents to provide more useful results. Moreover, this study focused on fuel cell technologies for hydrogen applications. Future studies may include or focus on other application fields such as hydrogen gas turbines [42,43] by selecting and including relevant patents to conduct more comprehensive study for the hydrogen technology sector. Finally, this study shows conversion between hydrogen and other chemical substances like ammonia as one of the key topics. Such technologies may open a new pathway for hydrogen usage. Therefore, future studies may include or focus on technology fields related to application of such converted substances from hydrogen. For example, environmentally friendly combustion of ammonia [44,45] made from hydrogen to reduce carbon emissions can also be considered for the investigation of diverse impact pathways of hydrogen technologies.

CRediT authorship contribution statement

Hyunhong Choi: Conceptualization, Methodology, Software, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **JongRoul Woo:** Validation, Formal analysis, Investigation, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF), funded by the Korea government (MSIT) (grant number 2020R1G1A1101899). The authors gratefully acknowledge the support of the KU-KIST School Project (Korea University).

References

[1] Chapman A, Itaoka K, Farabi-Asl H, Fujii Y, Nakahara M. Societal penetration of hydrogen into the future energy system: impacts of policy, technology and carbon targets. Int J Hydrogen Energy 2020;45(7):3883–98. https://doi.org/10.1016/j.ijhydene.2019.12.112.
[2] Lamb JJ, Rytter E, Hillestad M, Lien KM, Burheim OS, Pollet BG. Emerging technology for hydrogen and bioenergy production, in Hydrogen. Biomass Bioenergy Academic Press 2020:55–71.
[3] Staffell I, Scamman D, Velazquez Abad A, Balcombe P, Dodds PE, Ekins P, et al. The role of hydrogen and fuel cells in the global energy system. Energy Environ Sci 2019;12(2):463–91. https://doi.org/10.1039/C8EE01157E.
[4] McPherson M, Johnson N, Strubegger M. The role of electricity storage and hydrogen technologies in enabling global low-carbon energy transitions. Appl Energy 2018;216:649–61. https://doi.org/10.1016/j.apenergy.2018.02.110.
[5] Roberts ME, Stewart BM, Tingley D, Airoldi EM. The structural topic model and applied social science. Advances in neural information processing systems workshop on topic models. Comput Appl Evaluat 2013.
[6] IEA. Global demand for pure hydrogen, https://www.iea.org/data-and-statistics/charts/global-demand-for-pure-hydrogen-1975-2018, [accessed 2021-05-31]; 2021a. Paris: International Energy Agency (IEA), p. 1975–2018.
[7] IEA. Energy technology RD&D budgets, https://www.iea.org/reports/energy-technology-rdd-budgets-overview, [accessed 2021-05-31]; 2021b. Paris: Overview, International Energy Agency (IEA).
[8] Çabukoglu E, Georges G, Küng L, Pareschi G, Boulouchos K. Fuel cell electric vehicles: an option to decarbonize heavy-duty transport? Results from a Swiss case-study. Transp Res D 2019;70:35–48. https://doi.org/10.1016/j.trd.2019.03.004.
[9] Shin J, Hwang WS, Choi H. Can hydrogen fuel vehicles be a sustainable alternative on vehicle market? Comparison of electric and hydrogen fuel cell vehicles. Technol Forecast Soc Change 2019;143:239–48.
[10] Ajanovic A, Haas R. Economic prospects and policy framework for hydrogen as fuel in the transport sector. Energy Pol 2018;123:280–8. https://doi.org/10.1016/j.enpol.2018.08.063.
[11] Zhang F, Zhao P, Niu M, Maddy J. The survey of key technologies in hydrogen energy storage. Int J Hydrogen Energy 2016;41(33):14535–52. https://doi.org/10.1016/j.ijhydene.2016.05.293.
[12] Mayyas A, Wei M, Levis G. Hydrogen as a long-term, large-scale energy storage solution when coupled with renewable energy sources or grids with dynamic electricity pricing schemes. Int J Hydrogen Energy 2020;45(33):16311–25. https://doi.org/10.1016/j.ijhydene.2020.04.163.
[13] Reuß M, Grube T, Robinius M, Stolten D. A hydrogen supply chain with spatial resolution: comparative analysis of infrastructure technologies in Germany. Appl Energy 2019;247:438–53. https://doi.org/10.1016/j.apenergy.2019.04.064.
[14] Robledo CB, Oldenbroek V, Abbruzzese F, van Wijk AJM. Integrating a hydrogen fuel cell electric vehicle with vehicle-to-grid technology, photovoltaic power and a residential building. Appl Energy 2018;215:615–29. https://doi.org/10.1016/j.apenergy.2018.02.038.
[15] Moehrle MG, Caferoglu H. Technological speciation as a source for emerging technologies. Using semantic patent analysis for the case of camera technology. Technol Forecasting. Soc Change 2019;146:776–84. https://doi.org/10.1016/j.techfore.2018.07.049.
[16] Ardo S, Fernandez Rivas D, Modestino MA, Schulze Greiving V, Abdi FF, Alarcon Llado E, et al. Pathways to electrochemical solar-hydrogen technologies. Energy Environ Sci 2018;11(10):2768–83. https://doi.org/10.1039/C7EE03639F.
[17] Shaner MR, Atwater HA, Lewis NS, McFarland EW. A comparative technoeconomic analysis of renewable hydrogen production using solar energy. Energy Environ Sci 2016;9(7):2354–71. https://doi.org/10.1039/C5EE02573G.
[18] Wang C-H, Ok YS, You S, Wang X. The research and development of waste-to-hydrogen technologies and systems. Appl Energy 2020;268:115015. https://doi.org/10.1016/j.apenergy.2020.115015.
[19] Hammond GP, Owen RE, Rathbone RR. Indicative energy technology assessment of hydrogen processing from biogenic municipal waste. Appl Energy 2020;274:115329. https://doi.org/10.1016/j.apenergy.2020.115329.
[20] Chanchetti LF, Leiva DR, Lopes de Faria LI, Ishikawa TT. A scientometric review of research in hydrogen storage materials. Int J Hydrogen Energy 2020;45(8):5356–66.
[21] Abe JO, Popoola API, Ajenifuja E, Popoola OM. Hydrogen energy, economy and storage: review and recommendation. Int J Hydrogen Energy 2019;44(29):15072–86. https://doi.org/10.1016/j.ijhydene.2019.04.068.
[22] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003;3:993–1022.

[23] Guo Y, Barnes SJ, Jia Q. Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent dirichlet allocation. Tour Manage 2017;59: 467–83. https://doi.org/10.1016/j.tourman.2016.09.009.

[24] Rossetti M, Stella F, Zanker M. Analyzing user reviews in tourism with topic models. Inf Technol Tourism 2016;16(1):5–21. https://doi.org/10.1007/s40558-015-0035-y.

[25] Godin F, Slavkovikj V, De Neve W, Schrauwen B, Van De Walle R. Using topic models for twitter hashtag recommendation. Proceedings of the 22nd international conference on World Wide Web; 2013, p. 593–6.

[26] Kim H, Ahn SJ, Jung WS. Horizon scanning in policy research database with a probabilistic topic model. Technol Forecast Soc Change 2019;146:588–94. https://doi.org/10.1016/j.techfore.2018.02.007.

[27] Suominen A, Toivanen H, Seppänen M. Firms' knowledge profiles: Mapping patent data with unsupervised learning. Technol Forecast Soc Change 2017;115:131–42. https://doi.org/10.1016/j.techfore.2016.09.028.

[28] Kim G, Park S, Jang D. Technology analysis from patent data using latent dirichlet allocation. In: Adv Intell Syst Comput, Lee K, Park SJ, Lee JH, editors Soft Computing in Big Data Processing. Cham: Springer. 2014;271.

[29] Zhang H, Daim T, Zhang Y(. Integrating patent analysis into technology roadmapping: A latent dirichlet allocation-based technology assessment and roadmapping in the field of Blockchain. Soc Change 2021;167:120729. https://doi.org/10.1016/j.techfore.2021.120729.

[30] Farrell J. Corporate funding and ideological polarization about climate change. Proc Natl Acad Sci USA 2016;113(1):92–7. https://doi.org/10.1073/pnas.1509433112.

[31] Korfiatis N, Stamolampros P, Kourouthanassis P, Sagiadinos V. Measuring service quality from unstructured data: A topic modeling application on airline passengers' online reviews. Expert Syst Appl 2019;116:472–86. https://doi.org/10.1016/j.eswa.2018.09.037.

[32] Kuhn KD. Using structural topic modeling to identify latent topics and trends in aviation incident reports. Transp Res C 2018;87:105–22. https://doi.org/10.1016/j.trc.2017.12.018.

[33] Yoon B, Magee CL. Exploring technology opportunities by visualizing patent information based on generative topographic mapping and link prediction. *Technol Forecast* Soc Change 2018;132:105–17. https://doi.org/10.1016/j.techfore.2018.01.019.

[34] Newman D, Lau JH, Grieser K, Baldwin T. Automatic evaluation of topic coherence. In: Human language technologies: The 2010 annual conference of the North American chapter of the Association for Computational Linguistics. Association for Computational Linguistics, p. 100–8.

[35] Bischof J, Airoldi E. Summarizing topical content with word frequency and exclusivity. In: Langford J, Pineau J, editors. Proceedings of the 29th international conference on Machine Learning, Omnipress, New York; 2012, p. 201–8.

[36] Airoldi EM, Bischof JM. Improving and evaluating topic models and other models of text. J Am Stat Assoc 2016;111(516):1381–403. https://doi.org/10.1080/01621459.2015.1051182.

[37] Zhao T, Liu H, Roeder K, Lafferty J, Wasserman L. The huge package for high-dimensional undirected graph estimation in R. J Mach Learn Res 2012;13: 1059–62.

[38] Yuan X, Cai Y. Forecasting the development trend of low emission vehicle technologies: Based on patent data. Technol Forecasting. Soc Change 2021;166: 120651. https://doi.org/10.1016/j.techfore.2021.120651.

[39] Sinigaglia T, Eduardo Santos Martins M, Cezar Mairesse Siluk J. Technological evolution of internal combustion engine vehicle: A patent data analysis. Appl Energy. 2022;306. , .

[40] Yoon J, Jeong B, Lee WH, Kim J. Tracing the evolving trends in electronic skin (e-Skin) technology using growth curve and technology position-based patent bibliometrics. IEEE Access 2018;6:26530–42. https://doi.org/10.1109/ACCESS.2018.2834160.

[41] Nagula M. Forecasting of Fuel cell technology in hybrid and electric vehicles using Gompertz growth curve. J Stat Manag Syst 2016;19(1):73–88. https://doi.org/10.1080/09720510.2014.1001601.

[42] Langston LS. Hydrogen fueled gas turbines. Mech Eng 2019;141:52–4. https://doi.org/10.1115/1.2019-MAR-6.

[43] Koç Y, Yağlı H, Görgülü A, Koç A. Analysing the performance, fuel cost and emission parameters of the 50 MW simple and recuperative gas turbine cycles using natural gas and hydrogen as fuel. Int J Hydr Energy 2020;45:22138–47. https://doi.org/10.1016/j.ijhydene.2020.05.267.

[44] Kenanoğlu R, Baltacioğlu E. An experimental investigation on hydroxy (HHO) enriched ammonia as alternative fuel in gas turbine. Int J Hydr Energy 2021;46 (57):29638–48. https://doi.org/10.1016/j.ijhydene.2020.11.189.

[45] Keller M, Koshi M, Otomo J, Iwasaki H, Mitsumori T, Yamada K. Thermodynamic evaluation of an ammonia-fueled combined-cycle gas turbine process operated under fuel-rich conditions. Energy 2020;194:116894. https://doi.org/10.1016/j.energy.2020.116894.

[46] Li S, Zhang X, Xu H, Fang S, Garces E, Daim T. Measuring strategic technological strength: patent portfolio model. Technol Forecasting Soc Change 2020;157: 120119. https://doi.org/10.1016/j.techfore.2020.120119.

[47] Li X, Xie Q, Daim T, Huang L. Forecasting technology trends using text mining of the gaps between science and technology: the case of perovskite solar cell technology. Technol Forecasting Soc Change 2019;146:432–49. https://doi.org/10.1016/j.techfore.2019.01.012.

[48] Li X, Yuan X. Tracing the technology transfer of battery electric vehicles in China: a patent citation organization network analysis. Energy 2022;239:122265. https://doi.org/10.1016/j.energy.2021.122265.

[49] Mahlia TMI, Syazmi ZAHS, Mofijur M, Abas AEP, Bilad MR, Ong HC, et al. Patent landscape review on biodiesel production: technology updates. Renew Sustain Energy Rev 2020;118:109526. https://doi.org/10.1016/j.rser.2019.109526.

[50] Miyamoto M, Takeuchi K. Climate agreement and technology diffusion: Impact of the Kyoto Protocol on international patent applications for renewable energy technologies. Energy Policy 2019;129:1331–8. https://doi.org/10.1016/j.enpol.2019.02.053.

[51] Nordensvard J, Zhou Y, Zhang X. Innovation core, innovation semi-periphery and technology transfer: the case of wind energy patents. Energy Policy 2018;120: 213–27. https://doi.org/10.1016/j.enpol.2018.04.048.

[52] Shubbak MH. Advances in solar photovoltaics: technology review and patent trends. 109383109383 Renew Sustain Energy Rev 2019;115. https://doi.org/10.1016/j.rser.2019.109383.

[53] Suh JW, Sohn SY, Lee BK. Patent clustering and network analyses to explore nuclear waste management technologies. Energy Policy 2020;146:111794. https://doi.org/10.1016/j.enpol.2020.111794.

[54] Yuan X, Li X. Mapping the technology diffusion of battery electric vehicle based on patent analysis: a perspective of global innovation systems. Energy 2021;222: 119897. https://doi.org/10.1016/j.energy.2021.119897.

[55] Yalcin H, Daim T. Mining research and invention activity for innovation trends: case of blockchain technology. Scientometrics 2021;126(5):3775–806. https://doi.org/10.1007/s11192-021-03876-4.

[56] Zhang G, Duan H, Wang S, Zhang Q. Comparative technological advantages between China and developed areas in respect of energy production: quantitative and qualitative measurements based on patents. Energy 2018;162:1223–33. https://doi.org/10.1016/j.energy.2018.08.081.

[57] Sampaio PGV, González MOA, de Vasconcelos RM, dos Santos MAT, de Toledo JC, Pereira JPP. Photovoltaic technologies: mapping from patent analysis. Renew Sustain Energy Rev 2018;93:215–24. https://doi.org/10.1016/j.rser.2018.05.033.

[58] Fujii H, Managi S. Trends and priority shifts in artificial intelligence technology invention: a global patent analysis. Econ Anal Pol 2018;58:60–9. https://doi.org/10.1016/j.eap.2017.12.006.

[59] Song B, Suh Y. Identifying convergence fields and technologies for industrial safety: LDA-based network analysis. Technol Forecast Soc Chang 2019;138: 115–26. https://doi.org/10.1016/j.techfore.2018.08.013.

[60] Kang J-N, Wei Y-M, Liu L-C, Wang J-W. Observing technology reserves of carbon capture and storage via patent data: Paving the way for carbon neutral. Technol Forecast Soc Chang 2021;171:120933. https://doi.org/10.1016/j.techfore.2021.120933.

[61] Erzurumlu SS, Pachamanova D. Topic modeling and technology forecasting for assessing the commercial viability of healthcare innovations. Technol Forecast Soc Chang 2020;156:120041. https://doi.org/10.1016/j.techfore.2020.120041.

[62] Chen H, Zhang G, Zhu D, Lu J. Topic-based technological forecasting based on patent data: a case study of Australian patents from 2000 to 2014. Technol Forecast Soc Chang 2017;119:39–52. https://doi.org/10.1016/j.techfore.2017.03.009.

[63] Kim G, Bae J. A novel approach to forecast promising technology through patent analysis. Technol Forecast Soc Chang 2017;117:228–37. https://doi.org/10.1016/j.techfore.2016.11.023.

[64] Daim TU, Rueda G, Martin H, Gerdsri P. Forecasting emerging technologies: use of bibliometrics and patent analysis. Technol Forecast Soc Chang 2006;73(8): 981–1012. https://doi.org/10.1016/j.techfore.2006.04.004.

[65] Noh H, Lee S. What constitutes a promising technology in the era of open innovation? An investigation of patent potential from multiple perspectives. Technol Forecast Soc Chang 2020;157:120046. https://doi.org/10.1016/j.techfore.2020.120046.