Computers in Industry 140 (2022) 103677

Contents lists available at ScienceDirect

Computers in Industry

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

Identification of emerging business areas for business opportunity analysis: An approach based on language model and local outlier factor

Jaewoong Choi[a], Byeongki Jeong[b], Janghyeok Yoon[c,*]

[a] Computational Science Research Center, Korea Institute of Science and Technology, Seoul, Republic of Korea
[b] Optimization&Analytics Office, SK Innovation, Seoul, Republic of Korea
[c] Department of Industrial Engineering, Konkuk University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 27 December 2021
Received in revised form 20 March 2022
Accepted 5 April 2022
Available online 14 April 2022

Keywords:
Business opportunity
Designated goods and services
Novelty indicator
Language model
Local outlier factor

ABSTRACT

Emerging business areas are early indicators of potential business opportunities, which are considered key to formulating new business strategies and envisioning near-future business environments. However, existing methods for analysing business opportunities solely depend on the opinion and knowledge of experts, which are time-consuming and labour-intensive. In academia, recent years have witnessed a significant increase in attempts to identify emerging business areas as near-future business opportunities with data-driven approaches. Although successful innovation requires sources of novelty, how to measure the novelty of business areas has barely been investigated in the literature. As a solution, we propose an approach to identifying emerging business areas with high novelty with a systematic process and quantitative outcomes. At the heart of the proposed approach is the composite use of the language model and local outlier factor (LOF). The meaning of business opportunities become more explicit by identifying emerging business areas composed of novel goods and services, with implications for the business operation stage. Finally, business opportunity maps are developed based on recency and visibility values, thereby investigating the implications as business opportunities. A case study of the trademarks related to scientific apparatus is presented to illustrate the proposed approach. The systematic process and quantitative results are expected to be employed in practice as a complementary tool, serving as a cornerstone for analysing business opportunities using trademarks.

1. Introduction

The significance of discovering business opportunities has been growing due to the uncertainty and volatility of the business environment. Companies are increasingly paying attention to the rapidly evolving business environment by keeping an eye on recent business item developments of competitors. However, as the relationship between products, companies and customers in the business environment become more complex and the life cycle of products or services shortens (Adner, 2013), traditional expert-centric methods become time-consuming and labour-intensive. As a result, managers and industrial practitioners are increasingly embracing a data-driven approach to access reliable results concerning business areas with opportunities. Over the years, it has received much attention in the field of technology management to discover business opportunities with data-driven approaches. Here, business opportunity is a source of innovativeness that arises exogenously from changes in the competitive business environment in which the industry or market of interest exists (Jones and Barnir, 2019).

The implications of data-driven business opportunity analysis differ depending on data sources. Specifically, there are four kinds of sources used to address business opportunities: patents, web news, social media and business initiatives. First, many researchers used patents as a reliable source of technological information to identify business opportunities at the stage of technology development, or technology commercialisation. Patents, the outputs of R&D projects, implicate the technologies or functions necessary to develop a product or service that will be used in future business (Lee and Lee, 2017; Seol et al., 2011). Next, web news and social media may have slightly different implications depending on the platforms, but they mainly deal with futuristic content or post-launch customer reactions (Yoon, 2012; Jeong et al., 2019). Of course, these are not standardised data, so it is necessary to arbitrarily set the scope of analysis by selectively collecting data on technologies, products, and brands, but they can provide hints on business opportunities at the

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

https://doi.org/10.1016/j.compind.2022.103677

stage of market monitoring or forecasting. Lastly, business initiatives of a set of start-ups are used to find hints of promising early-stage business ideas (Lee and Sohn, 2019), under the premise that new technology-based firms or venture companies are the most agile and fast actors in new business areas.

Given that clues of business opportunities also exist in the current or near-future business landscape, it is necessary to utilise data that indicates the business areas the company is currently engaged in or intends to do soon for business opportunity analysis. To this end, we adopt trademark data as a new data source for defining business areas and identifying business opportunities. Here, it is noteworthy that the motive of trademark applications is not only to indicate and differentiate the origin of ongoing business areas but also to announce and pre-empt new areas to be launched in the near future (Seip et al., 2018). In addition, trademarks, as known as an identifier of goods and services (G&S), represent the business items of companies according to a standardised form. Considering these points, trademarks can be more actively utilised as a useful material for analysing current as well as near-future business landscape and can benefit significantly from the application of computational methods. Also, the ways of data-driven business opportunity analysis leave several research avenues in defining business landscape, identifying business areas, and assessing business opportunities.

These deficiencies necessitate the development of a systematic way to define and identify business opportunities. To this end, we propose a novel approach to detecting emerging business areas with a systematic process and quantitative results. The meaning of business opportunities become clearer by identifying the emerging business areas composed of a set of novel goods and services, rather than interpreting a set of ambiguous keywords as a particular business area. At the heart of the proposed approach is the language model and local outlier factor (LOF). The language model is adopted to define a business area by grouping business items with similar meanings and the LOF is used to measure the degree of novelty in a quantitative manner. Hence the novelty in this study is defined as the degree of novelty of business items compared to prior goods and services (Seol et al., 2011; Lee and Lee, 2015). That is, the novelty of business areas is measured according to the degree to which business items are similar or different in keyword usage patterns. This indicator will provide new insights on business growth or new business development, which is similar to the purpose of the novelty indicator representing the inimitable degree of patented innovation in the R&D stage (Petruzzelli et al., 2015).

By exploiting the synergetic merits of the language model and LOF and interpreting novel goods and services as evidence of emerging business areas, the business landscape in the near future becomes explicit. In addition, since the proposed approach consists of statistical and scientific methods for each step, it can be developed as a simpler and more efficient software system, thereby allowing people unfamiliar with computational methods to augment the research results. Also, we expect that the systematic process and quantitative results in the proposed approach can aid managers or industrial practitioners to envision near future business trends and serve as a basis for developing more practical outcomes. This study may provide creative input to the fuzzy-front end stages of designing business models. Of course, the definition of a business model varies with the architecture, method, description, and conceptual tool or model for value creation (Landoni et al., 2020), but emerging goods or services that change over time can be appropriately used for the idea of most business models. Although there is nothing new under the sun, the current or near-future emerging business areas identified in this study can provide new insights into what to offer customers (e.g., value proposition) in reinventing or innovating existing business models (Lantano et al., 2022).

2. Background

2.1. Theoretical background on data-driven business opportunity analysis

In today's highly competitive and dynamic business environment, entrepreneurial companies need to innovate through the identification and exploitation of opportunities to succeed (Jones and Barnir, 2019). Opportunity literature largely presents the source of opportunities from two perspectives: opportunity creation and discovery. Opportunity creation deals with endogenous opportunities arising from the entrepreneur's subjective perception and relies on entrepreneurs' experimentation based on past experiences or imaginations about the future (Alvarez and Barney, 2007). On the other hand, opportunity discovery deals with opportunities exogenously generated from the imperfections resulting from changes in the context in which the industry or market of interest exists (Chiles et al., 2010). In this study, our focus lies in the emergence of new business areas in the business landscape, and quantitatively evaluate the potential as opportunities, which is close to the concept of opportunity discovery literature. There have been many attempts to provide objective evidence regarding business opportunities with data-driven approaches. Most of data-driven approaches to business opportunity discovery include the process of defining business landscape, identifying business areas and evaluating business opportunities as follows.

- Business landscape: Previous studies have focused on the process of selectively collecting data on a specific technology, product or topic in defining the business landscape, which is the scope of exploring business opportunities (Lee and Sohn, 2019). Most patent-based methods for business landscaping start from a set of patents representing a particular technology domain (Lee and Lee, 2017). In the case of social media or web news, the business landscape is defined by collecting a dataset based on tag, label or keywords related to a product, brand or topic (Yoon, 2012; Jeong et al., 2019). Whilst previous scoping methods for exploring business opportunities have made a significant contribution, their utility is somewhat limited since the boundaries of the business landscape are not clearly defined. Even if data was systematically collected through a patent search formula composed of related keywords representing a specific business, it is difficult to reproduce in the field. This causes difficulties and ambiguity in defining business areas and interpreting business opportunities as stated below.
- Business area definition: The identification and interpretation of business opportunities can still be time-consuming and labour intensive, depending on what level this step provides the detailed business areas present in the overall business landscape (Seol et al., 2011). It is noteworthy that there is no definitive answer to guide users in segmenting business areas at this step. How many and diverse business areas exist in the business landscape? A business area is defined differently depending on the analytical method or the intention, knowledge or opinion of the researchers, even with a similar business landscape. As a result, the implications of business opportunities can be interpreted differently, because business opportunities are defined by interpreting keywords or classification codes representing the areas corresponding to business opportunities. Of course, some pioneering researchers have attempted to define business areas in a systematic manner by applying topic modelling techniques such as latent Dirichlet allocation (LDA) to text data (Jeong et al., 2019). At its most basic, this approach provides a keyword set called topic by deriving a topic-keyword distribution and a document-topic distribution for pre-selected keywords. In most cases, the business areas identified through this process are difficult to

interpret as feasible areas and much effort is required to link them to practical business opportunities. Accordingly, the more specific and interpretable results for a business item or area should be provided rather than a keyword set or classification code, to facilitate the process of defining a business area.

- Business opportunity assessment: Because the interpretation of business areas is difficult, the assessment of business opportunities is also expert-centric and relies on tracebacks to keyword usage patterns or documents that are relatively relevant to the opportunity area (Lee and Lee, 2017). Previous studies related to business opportunity analysis highlight the utility of document-level analysis as a contribution because it is easy to analyse growth, popularity, and visibility of business areas (Yoon, 2012). However, although successful innovation can be characterised by high levels of novelty (Amit and Zott, 2001), little has been done about how to quantify the novelty of business areas in the relevant field.

2.2. Methodological background

In this study, emerging business areas are mainly addressed as potential business opportunities. As successful innovation is strongly related to the source of novelty, business areas consisting of novel goods and services are first identified. Here, our focus is on how to measure the novelty of various goods and services in a quantitative way. Accordingly, the language model is adopted to effectively analyse numerous and variously written business items, and the local outlier factor method is utilised to measure novelty in a quantitative manner.

2.2.1. Language model

A pre-trained language model was adopted to effectively analyse the text information of numerous business items because there was a large number of unique business items in the designated G&S statements. The unprocessed use of these business items would make it difficult to interpret the results of emerging business areas. Therefore, we employed a well-established pre-trained semantic language model to represent the items in vector space and group similar items as the individual business areas.

The vectorisation of text data can be conducted through the rudimentary method—Bag-of-Words (BoW), which assumes each word of the text as a dimension (Harris, 1954). Unfortunately, the BoW is susceptible to a curse of dimensionality when the number of unique words increases by thousands or more (Zhao and Mao, 2017). This fatal demerit leads to computational problems in terms of space and time complexity. In addition, this method discards the order of words and does not reflect the context, making it difficult to distinguish synonymous or same words that are differently arranged. As a remedy, language modelling has been recently developed with the advances of deep learning and the use of big data. The language model constructs an embedded vector in a fixed N-dimension to reflect the semantic context of words. A representative language model known as Word2Vec was developed by (Mikolov et al., 2013). This pioneering method introduced Negative sampling, which learns more accurate vectors for frequent words—instead of the hierarchical SoftMax—to facilitate both faster training and better representation of uncommon words. The success of Word2Vec ignited the continuous development of semantic language models (Chen et al., 2020) that has led to the evolution of such models based on the attention mechanism (Bahdanau et al., 2014) or transformer (Vaswani et al., 2017). Recently, universal language models such as Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2018), a robustly optimised BERT training approach (RoBERTa) (Liu et al., 2019), and text-to-text transfer transformer (T5) (Raffel et al., 2019) have been developed.

2.2.2. Local outlier factor

The LOF is a density-based outlier detection method (Breunig et al., 2000) and was adopted in this study to identify novel business items. This method detects local outliers by calculating the outlines as the degree of how isolated an object is from its surrounding neighbours. By considering the local information that has often been overlooked in other methods, LOF has been used in several applications (Pokrajac et al., 2007; Lazarevic et al., 2003). In this study, when business areas composed of goods' names are embedded through a language model, it may be difficult to judge the novelty of a business area based on distance alone. Because even if goods have similar expressions (closer by distance in embedding), a small word change or difference in expression can intensively represent a large technological development, which can mean a new business area. Therefore, the density-based LOF method instead of the distance-based clustering methods is utilised to identify clusters with arbitrary shapes and find local outliers.

The LOF operates in the following procedure. First, the Euclidean distance of the k-th nearest neighbours from an object p is computed and defined as *k-distance (p)*, where the parameter k representing the number of nearest neighbours can be adjusted. The set of k-th nearest neighbours (kNN(p)) is constructed by objects within the *k-distance* from p. Thereafter, the reachability distance of p to an object o in kNN(p) is computed as follows:

$$reachDist_k(p, o) = \max\{k - distance(o), d(p, o)\}, \tag{1}$$

where $d(p, o)$ is the Euclidean distance between p and o. The local reachability density ($lrd_k(p)$) of p is computed as follows:

$$lrd_k(p) = \frac{k}{\sum_{o \in kNN(p)} reachDist_k(p, o)} \quad . \tag{2}$$

Lastly, the LOF of p for k surrounding neighbours is computed as follows:

$$LOF(p) = \frac{1/k \sum_{o \in kNN(p)} lrd_k(o)}{lrd_k(p)} \quad . \tag{3}$$

According to Eq. (3), LOF(p) is the ratio of the average density of kNN(p) to the density of p. If p is inlier, its LOF value would be close to 1 because the densities in the numerator and denominator are similar. Otherwise, its LOF value would be larger than 1 because of the very small relative density of the outlier p to kNN(p). Thus, an object p being far from other objects has a large LOF value and is likely to be an outlier. Fig. 1 shows three cases of LOF computation. In case 1, the location of the object p in the middle of a dense area results in large values for both $lrd_k(o)$ and $lrd_k(p)$, and contrastingly a small value for the LOF. Moreover, in case 2, where the object p is in the middle of a sparse area, both $lrd_k(o)$ and $lrd_k(p)$ have small values, and thus, the LOF value is small. Lastly, in case 3, the object p is in a sparse area, but its neighbouring objects form dense clusters. Accordingly, $lrd_k(o)$ is large, but $lrd_k(p)$ is relatively small, which makes the object p an outlier with a large LOF value. Therefore, such a relative density-based method is effective in detecting both local and global outliers.

3. Proposed approach

3.1. Concept

Considering that the landscape of the near-future business environment can be envisioned with the current emerging business areas (Li, 2009), the starting point for business opportunity analysis is aligned with the early identification of emerging business areas (Clark, 2003). In this regard, emerging business areas have been presented and analysed as clues to potential business opportunities. However, this method had low feasibility and reproducibility in

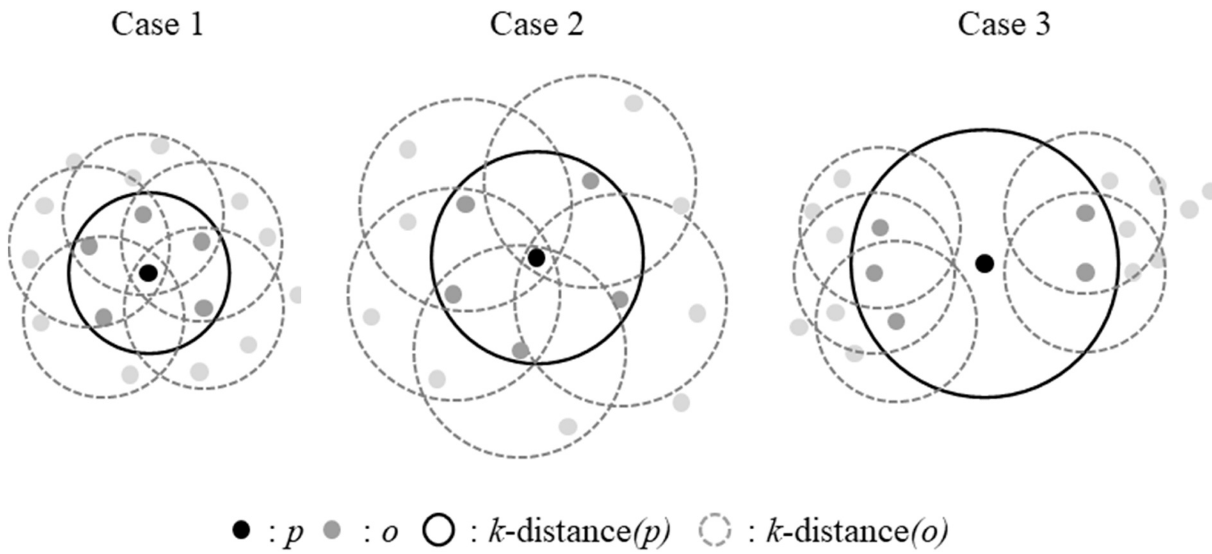

Fig. 1. LOF computation examples.

practice because the process of defining and analysing business landscape and business areas tends to be expert-dependent and vague. In addition, given the growing complexity of the business landscape in which business opportunities are explored, the time, cost and effort associated with traditional approaches result in inefficiencies. Therefore, such processes should be implemented with better quality and well-organised data sources. Hence, there may be a clear need for a rethinking of the traditional data source used in business opportunity analysis, and a new approach specifically focusing on business opportunities operating in the rapidly changing business environment is required.

In this context, we started this research with the basic premise that data-driven results can promise to provide helpful insights regarding business opportunities. We suggest an approach for identifying business opportunities in a systematic manner, by analysing emerging business areas with trademarks rather than traditional data sources. In particular, when the keywords in the G&S statements of trademarks represent individual business areas (Lee and Lee, 2017; Jeong et al., 2021; Ko et al., 2020; Jeong et al., 2019), novel word usage patterns are investigated to identify potential emerging business areas. The meanings of business opportunities become more explicit by identifying the emerging business areas composed of novel G&S, rather than interpreting a set of ambiguous keywords as a particular business area. By exploiting the synergetic merits of the language model and LOF, the proposed approach finds emerging business areas in a systematic and quantitative manner. The differences between previous studies and present research are summarised in Table 1.

Here, it is noteworthy that the purpose of the proposed approach is not to discern whether business areas are emerging, but rather to identify opportunity areas that are relatively likely to be emerging. In addition, it needs to be highlighted that the focus of this study is to find business opportunities that will lead the business environment in the near future. This is the reason why we analysed trademark data close to the business operation stage rather than analysing patents, web news, or social media used in previous studies. Computational methods and results in this study can guide experts to emerging business areas with high potential for business opportunity, but their role should be limited to supporting the decision-making process of experts and providing evidence from a large amount of data that humans cannot all cover. The knowledge and experience of experts or managers are still essential to determine whether the emerging business areas presented by the proposed approach are indeed business opportunities and crystallise them to successful business innovation or development.

3.2. Data

This study employed the United States Patent and Trademark Office (USPTO) database as the main data source of trademarks. The USPTO has been managing trademarks since 1870 and has an electronic database format that can retrieve detailed trademark information such as legal events, designated G&S statements, and classification codes (Graham et al., 2013). Among the various trademark information, the international classification code of the Nice Classification (NCL) data is employed to narrow business landscape, as it categorises goods and services into 45 classes (1–35: goods, 36–45: services). In addition, the filing time information is used to cover the timing of business operations, while the designated G&S statements are used to capture specific business items and define business areas.

Table 1
Comparisons between previous studies and present research.

Factor	Previous studies	Present research
Data	Patents, web news, social media and business initiatives	The designated G&S statements of trademarks
Approach	Detecting business opportunities by measuring growth, popularity, and visibility for each topic, business area or classification code	Identifying emerging business areas composed of novel goods and services
Method	Text mining methods from keyword extraction to topic modelling such as LDA	The composite use of language model and LOF
Results	Opportunity business areas that are represented as a set of keywords or classification code	Emerging business areas with quantitative novelty indicators
Implications	Business opportunities at the stage of technology development, or post-launch market monitoring	Business opportunities at the early stage of business operation – near the time of trademark application

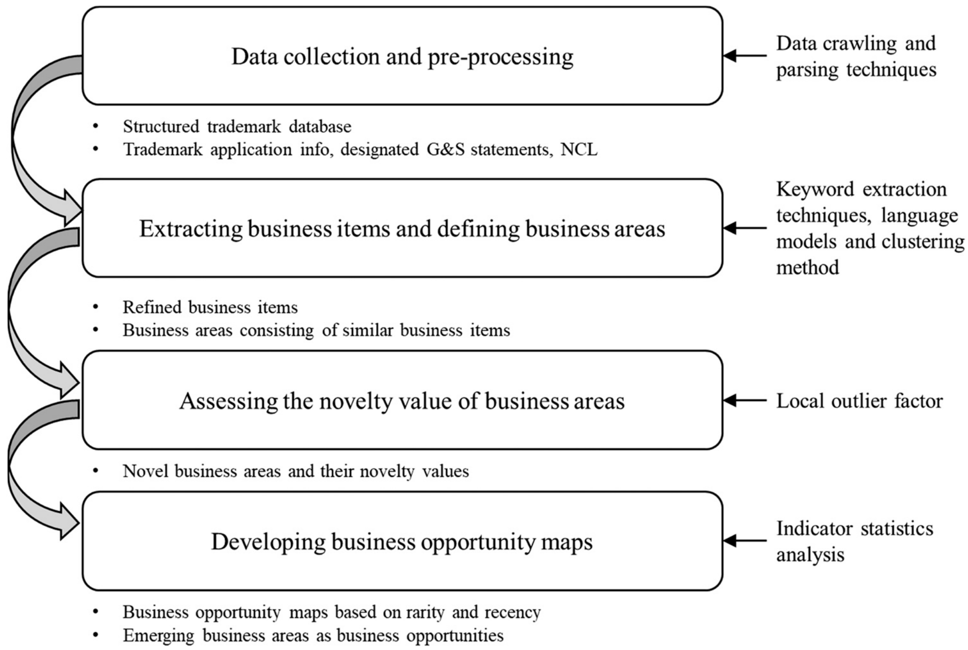

Fig. 2. The overall process of the proposed approach.

3.3. Procedure

The entire process of the proposed approach is described, while brief descriptions of each step are provided. As shown in Fig. 2, the proposed approach utilises various data analytics methods. Considering the complexities of the proposed approach, we designed it to proceed in a series of four discrete steps: data collection and pre-processing; extracting business items and defining business areas; assessing the novelty value of business areas; developing business opportunity maps.

3.3.1. Data collection and pre-processing

Since the novelty or *emergingness* of business areas covered in this study is a comparative concept, it is necessary to first determine the scope in which to explore business opportunities, namely the business landscape. Given the focal business landscape, trademarks of interest are collected from the USPTO database based on some search criteria. A business landscape can be narrowly specified as a set of trademarks including specific keywords, or broadly as a set of trademarks corresponding to an NCL code. Afterwards, trademark data needs to be pre-processed as structured data and unstructured text data are mixed in the data. Therefore, we parse the raw data and store it in a database, which contains both structured items (e.g., NCL codes, application numbers) and unstructured items (e.g., title, designated G&S).

3.3.2. Extracting business items and defining business areas

This study employs the designated G&S statements of trademarks to identify the business items that companies will operate on now or in the near future, namely goods and services. Although the USPTO provides a Trademark Identification Manual that provides general representations of designated goods and services, many applicants use new or creative keywords or noun phrases to differentiate their business items. Consequently, the number of business items in different expressions is so huge that we cannot analyse each of them in manual. For this reason, how to extract items from each trademark in a structured form and how to group similar items to define a business area are dealt with in this step, which is composed of a series of four sub-steps as follows.

(1) Refinement of the designated G&S statements: The designated G&S statement of a trademark holds the set of business items (i.e., goods or services) covered by the trademark. However, sometimes there is an irregular expression of legal actions (e.g., intent to use, use in commerce, priority application) in the statements, which is not related to the business item and needs to be removed by pre-processing. For example, in the statement of a certain trademark such as "(Based on 44(e) Priority Application) Smart phones; Television receivers; Monitors for computers; Commercial monitors", the priority information in parentheses may be removed to facilitate the analysis of business items.
(2) Extraction of business items: After pre-processing, text data consists of the only business items in the designated G&S statements. The delimiter present in the statement allows the identification and extraction of individual business items. As a result, the statement document can be transformed into a structured form such as a set of business items. For example, a statement ("Smart phones; Television receivers; Monitors for computers; Commercial monitors") could be transformed into a list of smart phones, television receiver, monitors for computers and commercial monitors.
(3) Vector representation of business items using pre-trained language model: The extracted business items may be items in the Trademark Identification Manual but may be new expressions created for differentiation. Therefore, the number of unique business items is too large, and it is difficult to manually analyse them for business area definition. As a solution, a pre-trained language model and clustering method are employed to embed business items into vector space models and to group similar items in a group, respectively. Regarding the pre-trained language models, there are various state-of-the-art models such as BERT (Devlin et al., 2018), RoBERTa (Liu et al., 2019), and ELECTRA (Clark et al., 2020), which return business items expressed as keywords or noun phrases as n-dimensional vectors. As a result, business items with similar meanings or expressions are closely located on the vector space model.
(4) Definition of business areas by clustering similar business items: Although the business items are converted into a comparable vector form, there are still too many items to analyse. Considering this, high-level clusters representing business areas

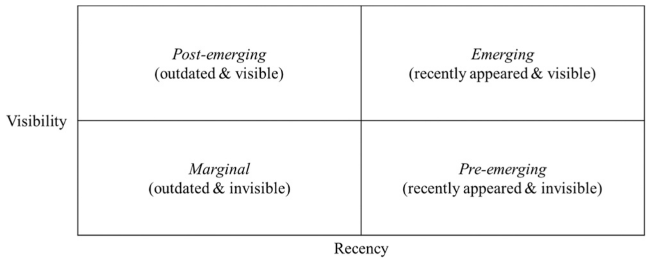

Fig. 3. Structure of business opportunity map.

are generated by grouping similar business items. Unsupervised learning-based clustering methods such as k-means, k-medoids can be used to group business items. Here, the number of business item clusters k (i.e., the number of business areas) can be statistically determined. As a result, k business areas composed of similar business items are generated, and interpretation of each business area can be easily performed based on business items having a close vector distance from the centre point of each cluster.

3.3.3. Assessing the novelty value of business areas

This step assesses whether a business area is relatively novel compared to other business areas according to the process of LOF explained in the section 'Local outlier factor'. The unit of analysis in this step is business areas, and the input of LOF is the centre value of each business area, which is calculated based on the vector values of the constituent business items in the previous step. In calculating the LOF value of a business area, the number of nearest neighbours k is determined by experts. In the proposed approach, k is considered as the number of business areas, which corresponds to the majority in relation to novel business areas. This parameter can be determined through statistical methods, but qualitative judgements provide more flexibility in practical use. Considering that the criteria may be subjective depending on the context of business opportunity analysis and business areas, this process is often performed in manual. For example, if a company is conducting exploratory research to detect novel business areas, a large value of k may allow more adjacent areas to be considered, resulting in fruitful results from a macroscopic point of view. On the other hand, for a company interested in incremental innovation, narrowing the scope of analysis to adjacent business areas may be a practical solution. In summary, qualitative judgements of experts are effective in determining the value of k, and this process is driven by iterative test results using a manageable number of business areas to facilitate the consensus-building on the outcomes. Finally, it is possible to quantitatively assess the novelty value of all business areas in this way.

3.3.4. Development of business opportunity maps

Novel business areas have a high potential for emerging business opportunities, but their possibilities can be enriched when integrated with additional information such as radical growth. In this regard, recency and visibility information has been frequently employed to capture the radical growth of opportunity areas (Yoon, 2012; Hiltunen, 2008). Recency indicates how recently a business area appeared in the business environment, as measured by the time of appearance of the analysis subject. Visibility represents the degree to which a business area stands out in the overall business environment, as measured by the frequency of a subject. Put together, if a novel business area has recently emerged and is clearly visible, it has a high potential for business opportunities. In that vein, we adapt these concepts to quantitative indicators in the context of business opportunity analysis. Recency is defined as the degree to which the business area is recently recognised by other actors, i.e., recently seen in trademarks. Considering that many scholars noted that *emergingness* can be conceived as an up to date (original) one that does not readily align with the understanding contained in the coherent configuration of the accumulated environment, assessment of recency can inform fresh and new areas from the prevailing trends in the current business landscape (Rotolo et al., 2015; Halaweh, 2013). Visibility, on the other hand, is defined as the degree to which a business is observed in the business landscape – the results of business innovation and operation from past to present – captured by trademark data. As many scholars referred to *emergingness* as vague and uncertain information about the future landscape that is usually hidden among the weak signals of the complex environment and that gradually emerges to form a sort of intelligence for competitive edges (Schiebel et al., 2010; Roche et al., 2010). The appraisal of visibility can guide organisations toward pervasive (including notions of trend-leading, influential, popular, etc.) or signalling (information about the future landscape of the business areas that are likely to be emerging enough to require a paradigm shift) areas.

Hence business opportunity maps are constructed to explore the implications of emerging business areas, combining the novelty indicator with the recency and visibility indicators. In the maps where novel business areas identified in the previous step are nodes, the values of novelty, recency and visibility are used to represent the node size, the horizontal coordinate value and the vertical coordinate value. As shown in Fig. 3, each novel business area is mapped to one of four areas on the map: *emerging area, post-emerging area, marginal area* and *pre-emerging area*. First, the business areas that have high recency and visibility are defined as novel emerging business areas. These areas are considered the most valuable business opportunities, as they are emerging business areas that, despite their recent appearance, have a leading position in current business trends. Second, the business areas that have low recency and high visibility are defined as post-emerging business areas. These are currently in an influential position, but they are likely to be saturated. With regard to these areas, business opportunities for a niche market through a paradigm shift can be obtained through thorough investigation. Marginal areas consist of business items that have been observed for a long time in trademark data (low recency) and have a marginal position in business trends (low visibility). These areas may be excluded from potential business opportunities unless significant innovation occurs. Pre-emerging areas consist of novel goods and services that have not yet been observed in trademarks (high recency), but they are small in number in the business landscape (low visibility). Crystallising these areas into business opportunities requires constant observation.

As the business opportunity maps show a cross-section at the point of analysis, continuous data and results update is required for use in practice. If a particular business area already exists in the in-house business operation stage, it may be removed from the map. Also, if a business area is excluded from analysis by the judgement of experts or industrial practitioners, the LOF and indicator calculation procedure must be performed again and placed on the map. Despite the necessity of additional works, it is noteworthy that once the scope for exploring opportunities (i.e., business landscape) is defined, maps and other outcomes are fully reproducible, and a systematic process enables new data to be added and analysed. These business opportunity maps are expected to reduce the uncertainty in decision making in the business opportunity analysis process.

4. Case study

A case study of the trademarks about apparatus and instruments for scientific and research purposes (NCL 09) is conducted to illustrate the proposed approach for the following reasons. First, in the classification code, a number of technology-intensive products and services such as virtual reality devices (Lee et al., 2019), lithium-ion batteries (Wrålsen et al., 2021), Bluetooth-enabled smart devices (Aheleroff et al., 2020), artificial intelligence applications (Munan Li and Zhou, 2021), and cryptocurrency (Kim et al., 2020), which are recently in the spotlight, exist and relevant trademarks are gradually increasing. According to the Trademark Statistical Report (Office, 2019), NCL 09 is a growing business landscape as trademarks are being filed steadily and incrementally. Second, finding emerging business areas and exploring business opportunities among state-of-the-art equipment for scientific/research purposes equipped with multiple technologies or functions is an important task in market analysis and competitive intelligence of companies. Therefore, analysing trademarks for apparatus and instruments for scientific and research purposes is important for timely analysis of business opportunities. Consequently, 134,898 filed trademarks covered by NCL 09, and their designated G&S statements were selectively used for analysis in this case study. The latest dataset available at the time of analysis (January 2017 to December 2018) was collected and set as the analysis period to identify emerging business areas.

4.1. Results of novel business area identification

A total of 254,063 business items were extracted from the designated G&S statements and refined based on the document frequency (DF) and the Trademark Identification manual. In selecting business items, we removed items with a DF of 1, that is, items that appear only in one trademark. Since a DF value of 1 is a trivial number, more frequent and similar items are expected to replace it. Next, the Trademark Identification manual provides the names of general business items, we selected items that do not exist in the manual, as they are likely to be differentiated goods and services. As a result, a refined dataset consisting of 73,048 business items was utilised in the later steps. Thereafter, the pre-trained language model of RoBERTa (Liu et al., 2019) was applied to represent the refined business items in a dense embedding. This robust model can represent the semantic context of a business item composed of a noun or noun phrase well. The business items were represented as a fixed-dimensional vector of 1024.

Next, k-means clustering method was applied on the vectorized business items for clustering similar business items into several business areas. Here, it is noteworthy that it is necessary to carefully determine the number of business areas according to the experts' empirical judgements based on inertia values, as the hyperparameter directly affects the level of concrete implications as business opportunities. We used the centroid of each business area as an input of LOF to identify which business area is relatively novel. In the LOF, the number of local neighbours was determined by 20 according to the experts' opinion. Consequently, 103 business areas were identified as novel, and their main items are shown in Table 2. Novel business areas such as 'monopod', 'Bluetooth communication device', 'robot exoskeleton suits', and 'lithium-ion battery' were identified, meaning that they were unique and creative in a semantic context compared to other non-novel business areas.

4.2. Results of business opportunity map development

A business opportunity map was developed for the 103 novel business areas by identifying their trademarks and measuring the recency and visibility, as shown in Fig. 4. The specific indicator values of each business area are summarised in Table 3. The average values of the normalised recency and visibility values were used for categorising the business areas in the map.

In the business opportunity map, the novel business areas were mapped into one of four quadrants. First, there were 16 business areas in *emerging area*, with relatively high recency and visibility. These are expected to be the most emerging area for business opportunities and should be thoroughly examined by analysing the existence of first-movers or competitors. Second, 33 business areas in *pre-emerging area* need to be continuously monitored for the purpose of early identification of emerging business areas in the future. Third, 40 business areas determined to belong to *marginal area* are mostly excluded from the business opportunity analysis. Lastly, 14 business areas observed in post-emerging area need to be scrutinised to see if they are indeed a saturated market. Specifically, in *emerging area*, 'software for encryption (687)', 'smart watch (745)', and 'software for artificial intelligence (662)' are noteworthy as novel areas with high potential for business opportunities. In the case of *pre-emerging area*, there are 'lithium-ion batteries (719)', 'robotic exoskeleton suits (948)', 'voice-controlled smart home devices (991)', and 'self-service checkout terminals (988)', which require continuous monitoring as their visible signals are weak. Specific business items in emerging business areas are presented in Appendix A.

4.3. Results of quantitative and qualitative validation

In this section, the results of validating whether business areas identified via the proposed approach are emerging in practice are provided. The validation in this study was carried out in two ways, quantitative comparison and qualitative further investigation. In short, the former is about comparing the proposed indicator of *emergingness* with the prior indicator such as community index between emerging business areas and pre-emerging ones. The latter is to investigate whether companies in emerging business areas are still active. Here, it should be noted that the purpose of the proposed approach is to identify opportunity areas that are relatively likely to be emerging, not to discern whether business areas are emerging. Determining how emerging a business area is difficult and prone to subjective because in general, *emergingness* is judged through subsequent analysis. In addition, the feasibility and utility of the proposed approach cannot be easily confirmed since it provides potential business opportunities with the potential for new value

Table 2
Parts of novel business areas and their main business items.

Business area	Representative business items	Novelty
Monopod	Selfie sticks and hand-held monopods for smart phones and cameras	1.366
Survival horror game software	Survival horror video game	1.344
Golf simulator	Golf scopes, range finders for golf, and golf training aids	1.338
Bluetooth communication devices	Bluetooth enabled devices and Bluetooth speakers.	1.333
Robot exoskeleton suits	Powered robotic exoskeleton suits worn by humans for enhancing the strength and endurance	1.332
Lithium-ion battery	Lithium-ion polymer primary batteries	1.282
Business management intelligence	Computer software that provides real-time, integrated business management intelligence	1.270
Stand-alone voice-controlled information devices	Computer hardware consisting of stand-alone voice-controlled information devices, namely, cloud-connected and voice-controlled smart audio speakers with virtual personal assistant capabilities	1.253
Humanoid robots with artificial intelligence	Humanoid robots with artificial intelligence, namely, robots for personal, educational and hobby use and structural parts thereof	1.251
Electronic baby monitoring devices	Audio and video devices for monitoring babies	1.237

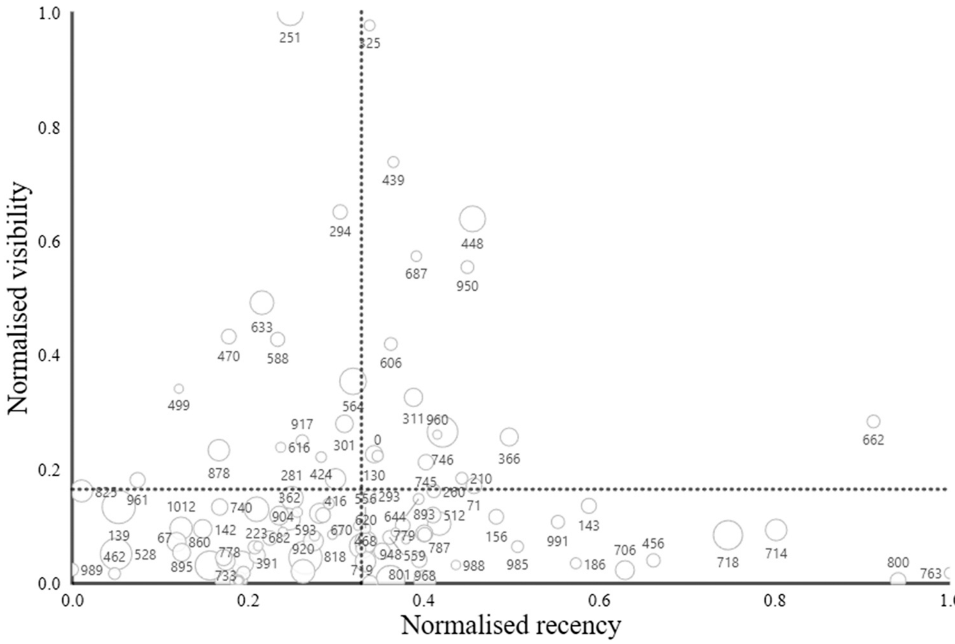

Fig. 4. Results of developing business opportunity map.

Table 3
Results of measuring novelty, recency, and visibility.

Identification No.	Novelty	Recency	Visibility
593	1.3887	0.2658	0.0453
139	1.3873	0.0531	0.1338
528	1.3660	0.0501	0.0513
746	1.3508	0.4219	0.2656
778	1.3443	0.1793	0.0101
…	…	…	…
512	1.1994	0.4111	0.1197
402	1.1988	0.3394	0.0000
801	1.1972	0.3954	0.0412
745	1.1971	0.4030	0.2123
143	1.1961	0.5887	0.1358
…	…	…	…
904	1.1624	0.2401	0.0905
499	1.1623	0.1216	0.3410
911	1.1617	0.3803	0.0765
372	1.1611	0.1896	0.0050
813	1.1608	0.3684	0.0895

Table 4
Summary of t-test for community scores.

Group	N	Mean	Standard deviation	Standard error
Emerging business areas	16	247	160	40
Pre-emerging business areas	33	39.1	27.4	4.8

$t = 5.11$, $p = 0.000$, DF = 38.

creation although they are not yet monopolised or saturated. As a remedy, taking into account that the number of relevant players (i.e., communities) that increase activity in a business area is positively correlated with *emergingness* (Porter et al., 2019; Moehrle and Caferoglu, 2019), we performed a t-test to statistically compare the average values of the number of applicants in the two different sets of business areas (*emerging area* vs *pre-emerging area*). Here, *pre-emerging* area was used as a comparison group, because it is most likely to develop into emerging area in the future, but not yet to have *emergingness* characteristics now. We performed a two-tailed Student's t-test with unequal sample sizes and unequal variances to test the null hypothesis indicating the mean values of the two sets are equal. Table 4 shows that there were significant differences in the community scores between the two sets, roughly supporting the argument that the proposed approach detects emerging business areas.

The results for the emerging business areas showed interesting examples, but they do not prove the quality of the proposed approach. Therefore, in this regard, we investigated the quality of the results of the business opportunity analysis by tracking the subsequent activities of major companies in emerging business areas. Although the progress of business operations may vary from company to company for the identified emerging business areas, the areas are likely to have become a major business area for companies at the time of analysis (December 2018) or in the near future. For example, an area regarding solar panels for electricity production is identified as emerging business area by all three indicators (i.e., novelty, recency and visibility). This area was strongly associated with Hanwha Q cells who has been actively engaged in business activities regarding solar power systems with frequent mergers and acquisitions since entering the solar energy industry. Hanwha Q cells has launched a solar panel technology with maximised efficiency over conventional panels, thereby guaranteeing high performance under any weather conditions. Software for encryption consisting of communication software, computer hardware, security tokens, and encryption key management software is also identified as emerging business area. In the current business landscape, the scope of application of cryptocurrency is rapidly expanding and is expected to revolutionise all industries. At the time of analysis, the pioneer HashiCorp, Inc., has become a major company in providing data encryption solutions to protect sensitive data with centralised key management. In addition, Tencent and Tessian, which were initially involved in the area, are now providing data security and encryption businesses to their customers. Also, lithium-ion batteries, smart home devices and self-checkout terminals have been identified as pre-emerging business areas. Although these areas are recognised as candidates for emerging business areas then, they have become integral parts of many products or services in recent years. Particularly, the lithium-ion battery market is gradually growing due to the rapid spread of electric vehicles and increased demand for smartphones and wearable devices. Also, the smart home device market has grown rapidly with the surge in Internet of Things devices and smartisation movements and has high growth expectations due to its high utilities in public facilities such as hotels and hospitals as well as individual users. Most of the self-service checkout terminals are in the form of a touch screen equipped with a network function, providing various types of information, and are called kiosks. As the demand for non-face-to-face consumption is rapidly increasing due to the timely technological development, the kiosk market, which is used for payment such as orders and reservations in stores, restaurants, and coffee shops, is rapidly growing.

5. Discussions

5.1. Implications for theory and practice

The proposed approach provides new understandings of how to identify emerging business opportunities in the business landscape, thereby leaving theoretical and practical implications. First, the proposed approach improves upon previous one-off methods into an interactive and adaptable structure that enables experts to create, adjust, and interpret the results of data analysis based on their knowledge and judgement, providing theoretical implications on business opportunity discovery literature. Specifically, unlike the previous one-off methods of identifying new business opportunities for a single product, service, or brand of interest, the proposed approach can broadly explore various business items as potential opportunities. To this end, we used pre-trained language models to construct a business landscape consisting of numerous business items as a scope for navigating business opportunities. To the best of our knowledge, this study is a pioneering attempt to apply the latest pre-trained language model, to identify business opportunities. Specifically, the analysis of the designated G&S statement advances the level of results from the keyword level to the business item level. The proposed approach can guide managers to strategic decision-making by providing business opportunities with easily interpretable and practical results of goods/service level. In addition, the systematic process and scientific methods of this study enable the development of quantitative evaluation metrics of the business areas. Although the main research question was which business area is the opportunity, the proposed approach and results may be useful for other purposes such as technology foresight, future signal detection and competitive analysis.

Second, the proposed approach has practical implications, since its systematic structure enables its development into automated software, which augments the business opportunity analysis process for domain experts who lack complex database management or data analytics skills. The proposed approach can be developed as an interactive software system that returns intermediate results through functional unit execution for each process of constructing a business landscape, defining business areas, and evaluating *emergingness* based on experts' feedbacks. Of course, the developed software can provide the final results of business opportunity analysis (e.g., business opportunity maps, *emergingness* indicators) directly from the pre-established business landscape, but it is necessary to invest sufficient time in precisely building the dataset and tuning the parameters for adequate reproduction. In practice, organisations need to periodically update their business landscape and thus trademark database to utilise more up-to-date business items for analysis and to continuously monitor the latest trends. Fortunately, the proposed approach allows organisations to adjust the scope of dataset and easily add new data when constructing their unique business landscapes. Once the new trademark data of interest is determined, the only additional work is to identify their designated G&S statements and redraw the business landscape.

5.2. Implementation and customisation of the proposed approach

This study suggests a systematic approach to identifying business opportunities with quantitative outcomes and scientific methods. The proposed approach has the following advantages over previous studies. First, the proposed approach allows experts to effectively explore business opportunities in the broader business landscape, which can significantly reduce the time and efforts involved in business opportunity analysis. Second, the proposed approach can serve as a useful complementary instrument to support experts' decision-making in the business opportunity analysis process, since the emerging areas identified have a high potential of opportunity. Particularly, because the interactive steps, where experts are able to explore potential opportunities in a way that defines the business landscape and crystallise opportunities based on their knowledge and judgements, are embedded throughout the proposed approach, its practical implications will be greater than previous methods. Third, the outcomes can offer more practical assistance with decision-making in business opportunity analysis over the previous methods, because potential business opportunities are provided as a set of goods or services rather than a set of keywords.

Nevertheless, since the newly developed method should be deployed with caution in practice, industrial practitioners should consider several factors when applying the proposed approach. Above all, the following statements are most important: (1) organisations should examine as many opportunities as possible within their accessible business landscape for successful business opportunity analysis, and (2) the implications of business opportunities identified in this study are quite different from those addressed by existing methods. In this respect, the outcomes of the proposed approach can expand the horizons of experts and enrich their ideas for generating new business items. In addition, the proposed and existing approaches are not exclusive and can be combined and utilised. For example, in the results of identifying emerging business areas, prior approaches (e.g., social media mining for product/market/competitor monitoring) can be applied to crystallise the business opportunities. Second, the business landscape analysis presented in this study is designed as an interactive and adaptable structure in identifying feasible business opportunities. For implementation and customisation in the field, it should be provided in a systematic way to generate a dataset that accurately reflects the scope of feasible business opportunities, as the dataset represents the scope of exploring business opportunities and thus the outcomes such as emerging business areas may differ depending on the business landscape. Third, although the advantage of the proposed method lies in that it effectively supports labour-intensive and time-consuming business opportunity analysis, it still requires some expert intervention and manual work when examining business areas to interpret and crystallise the identified business opportunities. The value of the presented results can be enriched when other business area analysis methods and evaluation metrics are integrated. For example, the patent-trademark linking database can be used to identify underlying technologies and competitors' technologies for identified business areas (Ko et al., 2020). Last but not least, the proposed approach has tuneable parameters that can determine the level of business opportunities (e.g., embedding dimensions of business items, number of business areas). Although the use of LOF and Euclidean distance is appropriate considering the distribution of embedding values in business domains and the purpose of the study, customisation on distance measurement and outlier detection will be required according to data.

6. Conclusions

The role of business opportunity analysis is becoming increasingly significant as a means for providing early insight for emerging business areas and competitive edges. This study proposed a business landscaping approach to business opportunity analysis based on quantitative outcomes and scientific methods. A central tenet of this study is that emerging business areas have many sources of novelty, and analysis of these characteristics provides valuable insights into business opportunity analysis. Accordingly, a business landscape, the scope of exploring business opportunities, was defined as a vector space model where goods and services with similar semantic context are located close and grouped into a business area. The business opportunities were identified according to (1) novelty assessment and (2) business opportunity map development. Finally, the business areas in *emerging area* were mainly addressed as business opportunities. The case study on apparatus for scientific and research purposes showed that the proposed approach enables an extensive search for business opportunities in the complex business environment.

The contributions of this study are two-fold. From an academic perspective, this study expands the applicability of previous methods suitable for business opportunity analysis in the stage of technology development or market forecasting to an approach that enables an extensive search of business opportunities in the business operation stage. The proposed approach is designed to be executed in a reproducible, and interactive way where step-by-step results are generated, adjusted and adapted based on expert judgement and intervention. In addition, this study is the first attempt to apply text mining to the designated G&S statements in the field of business opportunity discovery. The approach can be utilised for technology foresight or future signal detection, by adopting data sources such as patents, news or scientific papers and modifying several techniques. From a practical perspective, the proposed approach can be developed into an automated software system, facilitating use by those unfamiliar with complex techniques. The developed software and quantitative results can be useful in broadening the search for business opportunities and supporting expert decision-making.

Despite the contributions, this study is marked by several limitations, unfortunately. First, with regard to emerging business areas, this study focused on how to identify opportunity areas with the potential of emerging. The definitive answers for emerging business areas cannot be easily obtained through the proposed approach. The value of the proposed approach will be enriched if combined with lagging indicators of whether a business area has subsequently emerged. Secondly, regarding dimensions of opportunities, the proposed business opportunities need to be crystallised and refined, by analysing data on the market, technologies, and customers. Third, in terms of data analytics, the proposed approach provided acceptable results for business opportunities, but there is room for improvement by adopting the latest techniques. Although the proposed approach benefited from the use of LOF and language models, obtaining appropriate results in a simple and intuitive way, the computational complexity can increase as the number of data or embedding dimensions increases, which requires appropriate follow-up studies. Fourth, regarding the automation, most of the proposed approach can be automated, but some tasks (e.g., selecting business items, interpreting business domains) are still left to the judgement of experts. To this end, it would be helpful to integrate other methods such as statistical pre-processing techniques and text abstraction models and to develop reproducible guidelines, although the participation of experts is essential for proper use in practice. Lastly, this study conducted a single case study and cannot fully investigate the performance and utility of the proposed approach because the business opportunities identified in this study have not yet been demonstrated but may have potential as opportunities.

CRediT authorship contribution statement

Jaewoong Choi: Conceptualization, Methodology, Programming, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Byeongki Jeong:** Conceptualization, Methodology, Data curation, Programming, Investigation. **Janghyeok Yoon:** Conceptualization, Resources, Writing – review & editing, Supervision, 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 Ministry of Education (NRF-2021R1A2C101002711). This work was supported by the Human Resources Program in Energy Technology of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20204010600220).

Appendix A

See Appendix Table 1.

Appendix Table 1
Parts of business items for emerging business areas.

Business area	Business items	Distance from the centroid of business area
'software for encryption (687)'	software for encryption	3.9917
	encryption software, namely, software for encryption	3.9938
	communications software for encryption	4.1205
	…	
	encryption software enabling the secure transmission of digital information	8.3076
	encryption cellular phone hardware	8.5922
	computer network-attached storage (nsa) hardware	9.9834
'smart watch (745)'	smart watch	5.9923
	smart bands	6.0283
	smart bracelets	6.5748
	…	…
	smart wearable activity trackers	7.0412
	smart personal assistants	8.7088
	smart watch bands	9.9906
'software for artificial intelligence (682)'	computer software platforms for artificial intelligence	3.2386
	computer software and hardware for artificial intelligence	3.3961
	computer software for artificial intelligence applications	3.4629
	…	…
	artificial intelligence software for use in cognitive therapy	7.95444
	artificial intelligence apparatus	8.0432
	a self-modifying computerised artificial intelligence system, such as a decentralised computerised artificial intelligence system	8.0813

References

Adner, R., 2013. The wide lens: What successful innovators see that others miss, Penguin.

Aheleroff, S., Xu, X., Lu, Y., Aristizabal, M., Velásquez, J.P., Joa, B., Valencia, Y., 2020. IoT-enabled smart appliances under industry 4.0: a case study. Adv. Eng. Inform. 43, 101043.

Alvarez, S.A., Barney, J.B., 2007. Discovery and creation: alternative theories of entrepreneurial action. Strateg. Entrep. J. 1, 11–26.

Amit, R., Zott, C., 2001. Value creation in e-business. Strateg. Manag. J. 22, 493–520.

Bahdanau, D., Cho, K., Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate, arXiv preprint arXiv:.

Breunig, M.M., Kriegel, H.-P., Ng, R.T., Sander, J., 2000. LOF: identifying density-based local outliers. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, pp. 93–104.

Chen, Y., Jin, Y., Sun, X., 2020. Language model based interactive estimation of distribution algorithm. Knowl. Based Syst., 105980.

Chiles, T.H., Vultee, D.M., Gupta, V.K., Greening, D.W., Tuggle, C.S., 2010. The philosophical foundations of a radical Austrian approach to entrepreneurship. J. Manag. Inq. 19, 138–164.

Clark, G., 2003. The disruption opportunity. MIT Sloan Manag. Rev. 44, 27.

Clark, K., Luong, M.-T., Le, Q.V., Manning, C.D., 2020. Electra: Pre-training text encoders as discriminators rather than generators, arXiv preprint arXiv:.10555.

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:.04805.

Graham, S.J., Hancock, G., Marco, A.C., Myers, A.F., 2013. The USPTO trademark case files dataset: descriptions, lessons, and insights. J. Econ. Strategy 22, 669–705.

Halaweh, M., 2013. Emerging technology: what is it. J. Technol. Manag. Innov. 8, 108–115.

Harris, Z.S., 1954. Distributional structure. Word 10, 146–162.

Hiltunen, E., 2008. The future sign and its three dimensions. Futures 40, 247–260.

Jeong, B., Yoon, J., Lee, J.-M., 2019. Social media mining for product planning: a product opportunity mining approach based on topic modeling and sentiment analysis. Int. J. Inf. Manag. 48, 280–290.

Jeong, B., Ko, N., Son, C., Yoon, J., 2021. Trademark-based framework to uncover business diversification opportunities: application of deep link prediction and competitive intelligence analysis. Comput. Ind. 124, 103356.

Jeong, Y., Park, I., Yoon, B., 2019. Identifying emerging Research and Business Development (R&BD) areas based on topic modeling and visualization with intellectual property right data. Technol. Forecast. Soc. Change 146, 655–672.

Jones, R.J., Barnir, A., 2019. Properties of opportunity creation and discovery: Comparing variation in contexts of innovativeness. Technovation 79, 1–10.

Kim, S., Park, H., Lee, J., 2020. Word2vec-based latent semantic analysis (W2V-LSA) for topic modeling: a study on blockchain technology trend analysis. Expert Syst. Appl. 152, 113401.

Ko, N., Jeong, B., Yoon, J., Son, C., 2020. Patent-trademark linking framework for business competition analysis. Comput. Ind. 122, 103242.

Landoni, P., Dell'era, C., Frattini, F., Petruzzelli, A.M., Verganti, R., Manelli, L., 2020. Business model innovation in cultural and creative industries: insights from three leading mobile gaming firms. Technovation 92, 102084.

Lantano, F., Petruzzelli, A.M., Panniello, U., 2022. Business model innovation in video-game consoles to face the threats of mobile gaming: evidence from the case of Sony PlayStation. Technol. Forecast. Soc. Change 174, 121210.

Lazarevic, A., Ertoz, L., Kumar, V., Ozgur, A., Srivastava, J., 2003. A comparative study of anomaly detection schemes in network intrusion detection. In: Proceedings of the 2003 SIAM International Conference on Data Mining, SIAM, pp. 25–36.

Lee, C., Lee, H., 2015. Novelty-focussed document mapping to identify new service opportunities. Serv. Ind. J. 35, 345–361.

Lee, J., Kim, J., Choi, J.Y., 2019. The adoption of virtual reality devices: the technology acceptance model integrating enjoyment, social interaction, and strength of the social ties. Telemat. Inform. 39, 37–48.

Lee, M., Lee, S., 2017. Identifying new business opportunities from competitor intelligence: an integrated use of patent and trademark databases. Technol. Forecast. Change 119, 170–183.

Lee, W.S., Sohn, S.Y., 2019. Discovering emerging business ideas based on crowd-funded software projects. Decis. Support Syst. 116, 102–113.

Li, Y.-R., 2009. The technological roadmap of Cisco's business ecosystem. Technovation 29, 379–386.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., 2019. Roberta: A robustly optimized bert pretraining approach, arXiv preprint arXiv:.11692.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J., 2013. Distributed representations of words and phrases and their compositionality. Adv. Neural Inf. Process. Syst. 3111–3119.

Moehrle, M.G., Caferoglu, H., 2019. Technological speciation as a source for emerging technologies. Using semantic patent analysis for the case of camera technology. Technol. Forecast. Change 146, 776–784.

Munan Li, W.W., Zhou, Keyu, 2021. Exploring the technology emergence related to artificial intelligence: a perspective of coupling analyses. Technol. Forecast. Change 172, 121064.

Office W.I.P., 2019. World Intellectual Property Indicators 2019 - TRADEMARKS.

Petruzzelli, A.M., Rotolo, D., Albino, V., 2015. Determinants of patent citations in biotechnology: An analysis of patent influence across the industrial and organizational boundaries. Technol. Forecast. Soc. Change 91, 208–221.

Pokrajac, D., Lazarevic, A., Latecki, L.J., 2007. Incremental local outlier detection for data streams. In: Proceeding of the 2007 IEEE symposium on computational intelligence and data mining, IEEE, pp. 504–515.

Porter, A.L., Garner, J., Carley, S.F., Newman, N.C., 2019. Emergence scoring to identify frontier R&D topics and key players. Technol. Forecast. Change 146, 628–643.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P.J., 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. arXiv preprint arXiv: 10683.

Roche, I., Besagni, D., François, C., Hörlesberger, M., Schiebel, E., 2010. Identification and characterisation of technological topics in the field of Molecular Biology. Scientometrics 82, 663–676.

Rotolo, D., Hicks, D., Martin, B.R., 2015. What is an emerging technology? Res. Policy 44, 1827–1843.
Schiebel, E., Hörlesberger, M., Roche, I., François, C., Besagni, D., 2010. An advanced diffusion model to identify emergent research issues: the case of optoelectronic devices. Scientometrics 83, 765–781.
Seip, M., Castaldi, C., Flikkema, M., De Man, A.-P., 2018. The timing of trademark application in innovation processes. Technovation 72, 34–45.
Seol, H., Lee, S., Kim, C., 2011. Identifying new business areas using patent information: A DEA and text mining approach. Expert Syst. Appl. 38, 2933–2941.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need, Advances in neural information processing systems, pp. 5998–6008.
Wrålsen, B., Prieto-Sandoval, V., Mejia-Villa, A., O'Born, R., Hellström, M., Faessler, B., 2021. Circular business models for lithium-ion batteries-Stakeholders, barriers, and drivers. J. Clean. Prod. 317, 128393.
Yoon, J., 2012. Detecting weak signals for long-term business opportunities using text mining of Web news. Expert Syst. Appl. 39, 12543–12550.
Zhao, R., Mao, K., 2017. Fuzzy bag-of-words model for document representation. IEEE Trans. Fuzzy Syst. 26, 794–804.