Scientometrics (2021) 126:1867–1896
https://doi.org/10.1007/s11192-020-03842-6

Anticipating multi-technology convergence: a machine learning approach using patent information

Changyong Lee[1] · Suckwon Hong[2] · Juram Kim[3]

Received: 5 November 2019 / Accepted: 14 December 2020 / Published online: 5 February 2021
© Akadémiai Kiadó, Budapest, Hungary 2021

Abstract

Technology convergence has been the subject of many prior studies, yet most have focussed on the structural patterns of convergence between a pair of technologies rather than the dynamic aspects of multi-technology convergence. This study proposes a machine learning approach to anticipating multi-technology convergence using patent information. For this, a patent database is first constructed using the United States Patent and Trademark Office database, distinguishing the primary class from other patent classes to consider the direction of multi-technology convergence. Second, association rule mining is employed to construct technology ecology networks describing the significant structural patterns of multi-technology convergence for different time periods in the form of a primary patent class → supplementary patent classes. Third, the technology ecology networks between the periods are compared to identify implications on the changing patterns of multi-technology convergence. Finally, link prediction analysis based on logistic regression models is utilised to provide insight into the prospects of multi-technology convergence by identifying the links to be added to or removed from the network. Based on this, we also discuss the characteristics of the proposed approach and the technological impact and uncertainty of the identified patterns of multi-technology convergence. The case of drug, bio-affecting, and body treating compositions technology is presented herein.

Keywords Multi-technology convergence · Machine learning approach · Patent information · Technology ecology network · Association rule mining · Link prediction analysis

✉ Changyong Lee
changyong@sogang.ac.kr

Suckwon Hong
amoeba94@unist.ac.kr

Juram Kim
juram92@unist.ac.kr

1 Graduate School of Management of Technology, Sogang University, 35 Baekbeom-ro, Mapo-gu, Seoul 04107, Republic of Korea

2 Department of Management Engineering, Ulsan National Institute of Science and Technology, 50 UNIST-gil, Ulsan 44919, Republic of Korea

3 Department of Industrial Engineering, Ulsan National Institute of Science and Technology, 50 UNIST-gil, Ulsan 44919, Republic of Korea

Introduction

The concept of innovation ecosystems has gained increasing attention in academia and practice. Innovation ecosystems are defined as the evolving set of actors, activities, artefacts, and the institutions and relations that are important for the innovation performance of actors (Granstrand and Holgersson 2020). Among various perspectives and dimensions (e.g. science, technology, and business communities) (Oh et al. 2016; Xu et al. 2018), the integration of technologies in innovation ecosystems is delineated as a key element that influences the emergence of relationships among different actors and the development of innovation in the whole ecosystem (Adner and Kapoor 2010). This is especially true in recent technology-based industries since technology is a critical resource at the basis of firm and business growth (Adner 2006).

Prior literature presents that understanding the technology dimension in innovation ecosystems helps to assess the innovation capacities of innovation ecosystems and improve value propositions creating a new offering or reconfiguring ecosystem partners (Xu et al. 2018). In particular, technology convergence is considered a major phenomenon in understanding the technology dimension of innovation ecosystems, as innovation boundaries are blurred by merging and overlapping technologies (Athreye and Keeble 2000). Even though the impact of technology convergence on innovation ecosystems may vary to a great extent, it can be highly disruptive, and even minimal responses may require a long lead time (Hacklin et al. 2005). For instance, Ericsson and Siemens disappeared from the mobile phone market when internet browsing became mobile, while other companies, such as Apple, Google, and Samsung, have dominated the market (Hacklin et al. 2013).

A variety of models and methods have been presented for measuring technology convergence by using different data sources and methods. However, while prior studies have proved useful for providing empirical evidence of technology convergence and for deepening the understanding of the technology dimension of innovation ecosystems, most studies have focussed on the analysis of structural patterns of technology convergence (i.e. the relations between technologies at a certain point in time), rather than the assessment of dynamic aspects of technology convergence (i.e. changes in the relations between technologies over time). Although some studies have examined the dynamic aspects of technology convergence (Jeong et al. 2015; Kwon et al. 2020; No and Park 2010), previous methods have been limited to identifying the changes in core technologies driving convergence by comparing simple index values between different time periods. Hence, new methods should be developed to examine the significant changing patterns of technology convergence, and to envision the evolution of innovation ecosystems.

Another drawback is associated with the modelling of the structural patterns of technology convergence. Even though convergence occurs between more than two segments (Curran et al. 2010), relatively little attention has been paid to the method of measuring multi-technology convergence. Some studies have presented algorithms to measure multi-technology convergence (Kim and Lee 2017; Kim et al. 2011), yet they are limited to aggregating the patterns of convergence between a pair of technologies based on common converging segments. Therefore, new methods should be developed to measure multi-technology convergence by considering multiple converging segments jointly and to provide a fair reflection of the technology dimension of innovation ecosystems.

We propose a machine learning approach to anticipating multi-technology convergence using patent information. The premise of this study is twofold: (1) large-scale patent co-classification analysis can provide objective and reliable information on multi-technology convergence (Kwon et al. 2020); and (2) machine learning models can identify significant changing patterns and provide indications of the prospects of multi-technology convergence (Lee Han et al. 2015). Specifically, association rule mining (ARM) is employed to construct technology ecology networks that describe the significant structural patterns of multi-technology convergence in the form of a primary patent class → supplementary patent classes. The technology ecology network is a directed and valued network where a source and a target node represent a primary patent class and supplementary patent classes and a link represents the convergence intensity between them. The technology ecology networks are developed for different time periods to specify implications of the changing patterns of multi-technology convergence, whereas supervised link prediction analysis based on logistic regression models is developed to gain insight into the prospects of multi-technology convergence by identifying links that will be added to or removed from the network. Based on this, we also discuss the characteristics of the proposed approach and the technological impact uncertainty of the identified patterns of multi-technology convergence. A case study of the patents on drug, bio-affecting, and body treating compositions technology confirms that the proposed approach can provide a more accurate and reliable depiction of multi-technology convergence, and can guide organisations toward responding to the evolution of innovation ecosystems.

The remainder of this paper is organised as follows. "Background" section presents the research background. "Methodology" section explains our methodology, illustrated by a case study of drug, bio-affecting, and body treating compositions technology in "Case study" section. "Discussion" section discusses the characteristics of the proposed approach and the technological impact and uncertainty of the patterns of multi-technology convergence identified. Finally, "Conclusion" section concludes with limitations and suggests scope for future research.

Background

Many models and methods have been presented to provide empirical evidence of technology convergence. Among various data sources such as patents (Curran and Leker 2011; Geum et al. 2012; Jeong et al. 2015), news articles (Kim et al. 2015), and Wikipedia hyperlinks (Kim et al. 2019), patents have been considered the most important data source since they provide detailed information on the technology development and innovation activities of a wide spectrum of fields in a highly structured format (Caviggioli 2016; Kwon et al. 2020; Lee et al. 2012; No and Park 2010).

Two main approaches have been used to analyse patents to examine technology convergence. First, patent citation network analysis employs citing-cited relations as a proxy for knowledge flows between technological entities. No and Park (2010) identified the trajectory patterns of technology fusion from patent citation networks using in and out degree centrality measures for a specific patent class. Kim et al. (2014) recognised core technologies that drive convergence using centrality-focussed patent citation network analysis. Second, patent co-classification network analysis examines the co-occurrence of different subject-classification codes assigned to patents. Geum et al. (2012) measured the intensity and coverage of technology convergence via patent co-classification network analysis. Jeong et al. (2015) examined the time-varying status of inter-sector and inter-field technology convergence using patent co-classification network analysis. Lee Han et al. (2015) identified potential converging technologies by applying link prediction analysis to patent co-classification networks. In addition, Kwon et al. (2020) combined large-scale patent co-classification analysis with concordance between patent classes and industrial sectors, centrality and brokerage analysis, and link prediction analysis to anticipate technology-driven industry convergence.

While the results of prior studies have proved quite useful for many different purposes, they are subject to the following limitations. First, although patent co-classification network analysis has certain advantages over patent citation network analysis in that (1) the citation information may be biased due to legal issues related to the inventions (Jeong et al. 2015) and (2) recent patents are naturally less likely to be cited by other patents (Kwon et al. 2020), patent co-classification-based approaches do not consider the direction of technology convergence. It should be noted that convergence between primary technology C1 and supplementary technology C2 may have different meanings from that of the opposite (e.g. drugs for skincare and skincare lotions for nonmedical purposes) (Lee and Lee 2019). Hence, any developed approach should consider the direction of technology convergence to provide more detailed information on the relations between technologies and the role of actors within innovation ecosystems. Second, most studies have focussed on the analysis of structural patterns rather than the assessment of dynamic aspects of technology convergence. Some studies have examined the dynamic aspects of technology convergence (Jeong et al. 2015; Kwon et al. 2020; No and Park 2010), yet they have been limited to comparing simple index values (e.g. degree centrality scores) between different time periods, thereby being ineffective and unreliable in measuring significant changing patterns. For instance, although the number of co-occurrences of patent classes C1 and C2 has increased, the intensity of technology convergence between C1 and C2 may have actually decreased if the number of co-occurrences is smaller than the number of occurrences of C1 and C2. Hence, any approach that is proposed should allow the comparison of analysis results for different time periods. Finally, despite the fact that convergence occurs between more than two technologies (Curran et al. 2010), previous studies have predominantly focussed on convergence between a pair of technologies (Geum et al. 2012; Kim et al. 2019). Only a few studies have presented algorithms to measure multi-technology convergence (Kim and Lee 2017, Kim et al. 2011), yet they are limited to aggregating convergence patterns between a pair of technologies based on common patent classes. It should be noted that convergence between C1 and C2, between C2 and C3, and between C3 and C1 do not necessarily mean convergence among C1, C2, and C3. For instance, the number of co-occurrences of C1, C2, and C3 can be small even when the number of co-occurrences of C1 and C2, C2 and C3, and C3 and C1 are high. Therefore, any approach that is developed should examine multi-technology convergence in a more concrete way and further present guidelines on defining the scope of innovation ecosystems and developing collaboration strategies.

These limitations refer to the underlying motivation and are addressed in this study. Given that the primary class (i.e. the class on the forefront) best represents the field in which the patented invention can be applied (Lee and Lee 2019),[1] we distinguish the primary class from other supplementary classes to consider the direction of

[1] According to the United States Patent and Trademark Office, every US patent has one and only one primary class (i.e. first bold class) that represents the main idea of invention described in the patent. It is double-vetted and reliable since the primary class is used for routing the application along the patent office. If there is a mistake in primary classification, the examiner will reject the patent, and it will be reclassified and routed to a different examiner (http://www.acclaimip.com/the-us-patent-classification-system-class-types/).

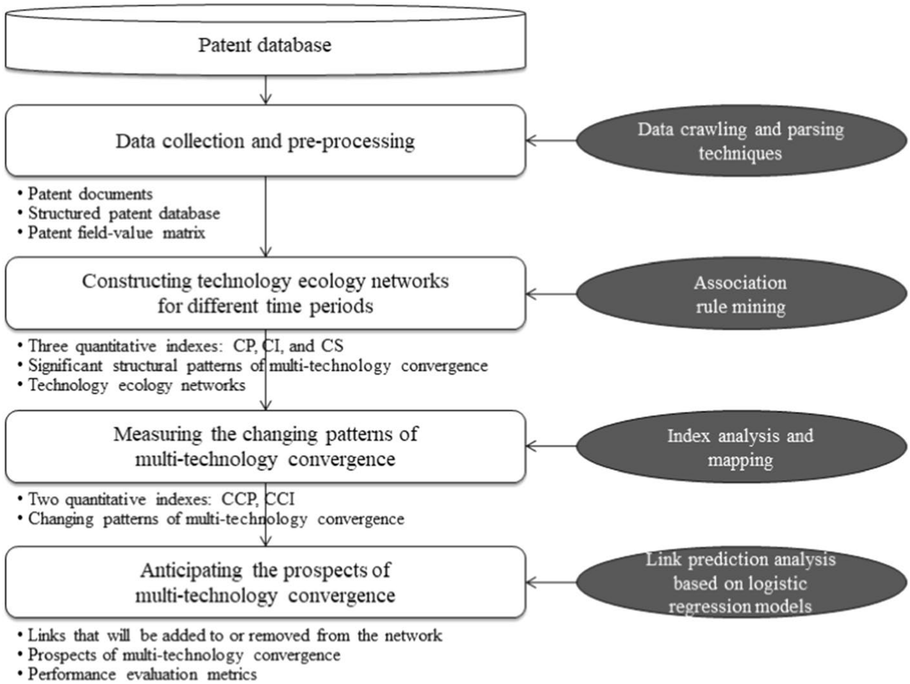

Fig. 1 Overall process of the proposed approach

multi-technology convergence trend in the form of a primary patent class → supplementary patent classes. ARM is employed to identify the significant structural patterns of multi-technology convergence by examining the ratio of the observed co-occurrence frequency of a primary patent class and supplementary patent classes to that expected if they were independent. Consequently, technology ecology networks constructed in this manner represent the significant structural patterns of multi-technology convergence, providing a more accurate and reliable representation of the technology dimension of innovation ecosystems. Moreover, the analysis results are compared between the periods to measure the changing patterns of multi-technology convergence while link prediction is employed to offer insights into the prospects of multi-technology convergence, guiding organisations to respond to the evolution of innovation ecosystems.

Methodology

As per Fig. 1, the proposed approach is executed in four discrete steps: (1) data collection and pre-processing, (2) constructing technology ecology networks for different time periods, (3) measuring the changing patterns of multi-technology convergence, and (4) anticipating the prospects of multi-technology convergence.

Data collection and pre-processing

Given a technology field of interest, the relevant patents (Set 1) are collected based on certain search criteria. Since the patents collected at this stage are in either HTML or XML formats, they are parsed according to the types of information and are stored in a structured patent database. Additionally, the patents that cite Set 1 patents (Set 2) are collected, parsed, and stored in the same manner to analyse the technological impact and uncertainty of the multi-technology convergence identified. The resulting patent database thereby includes information on both citing and cited patents.

The main input of the proposed approach is a patent field-value matrix constructed from the patent database. The patent field-value matrix consists of three parts: (1) basic information, (2) converging fields, and (3) technological value. First, basic information represents patent numbers and issued dates, which define the scope of analysis and the time period of interest. Second, the converging fields part includes patent classification information because it provides a rich and reliable picture of the specific technology fields of a patent (Aharonson and Schilling 2016). Here, we differentiate the primary class (i.e. the class at the forefront in bold font) from the other classes for fine-grained characterisation of patented inventions (Lee and Lee 2019). Consequently, the converging fields part from the patent that belongs to C1 as a primary class and C2 as a supplementary class has a different value from that of the patent that belongs to C2 as a primary class and C1 as a supplementary class. This distinction serves as the basis for identifying the direction of technology convergence. Finally, the technological value part contains the time-series forward citations to assess the technological impact and uncertainty of the patterns of the multi-technology convergence identified (Narin et al. 1987; Shin et al. 2013).

Constructing technology ecology networks for different time periods

ARM, which was developed by Agrawal et al. (1993), is an unsupervised method that discovers significant relations among items from a dataset. This method extracts association rules, X → Y, where X is an antecedent and Y is a consequent, using three general measures: support, confidence, and lift. In technology and innovation management research, this method has been successful in identifying significant relations between technologies for different purposes such as technology opportunity analysis (Kim et al. 2017), technology trend analysis (Shih et al. 2010), and technology impact analysis (Kim et al. 2011).

With regards to the analysis of technology convergence, ARM examines the detailed characteristics of multi-technology convergence in a quantitative manner. Specifically, although patent co-classification network analysis is based on the number of co-occurrences of patent classes, this method is limited to measuring convergence between a pair of technologies, and is unable to compare the analysis results across different time periods. For example, can we confirm that patent classes C1, C2, and C3 have converged if the number of co-occurrences of C1 and C2, C1 and C3, and C2 and C3 are 200, 150, and 180 respectively? Or that convergence between patent classes C1 and C2 becomes more significant than before, if the number of co-occurrences of C1 and C2 has increased from 100 to 200? To answer these questions, the occurrence frequency of individual patent classes and the co-occurrence frequency of the possible combinations of patent classes should be considered jointly.

We adopt ARM to examine the structural patterns of multi-technology convergence in the form of $\left(X_{\text{primary}} \rightarrow Y_{\text{supplementary}}\right)$ Given that the patent field-value matrix distinguishes the primary class from the other patent classes, we modify the three general indicators of ARM, namely, support, confidence, and lift, as follows. First, convergence predominance (CP) is based on the support indicator and measures the usefulness of the discovered structured patterns of multi-technology convergence. This index is defined as the ratio of the number of patents of patent class X as a primary class, X_{primary} and patent class Y as supplementary classes, $Y_{\text{supplementary}}$ to the total number of patents (i.e. the probability of co-occurrence of patent classes X and Y), as shown in Eq. (1). Second, convergence intensity (CI) is based on the confidence indicator and measures the certainty of the discovered structural patterns of multi-technology convergence. This index is referred to as the ratio of the number of patents related to patent class Y as supplementary classes among the patents of patent class X as a primary class (i.e. the conditional probability of Y given X), as shown in Eq. (2). Finally, convergence significance (CS) is based on the lift indicator and represents the statistical dependence between patent class X as a primary class and patent class Y as supplementary classes. This index is calculated by dividing the CI by the probability of occurrence of patent class Y as supplementary classes, as shown in Eq. (3).

$$\mathrm{CP}_{\left(X_{\text{primary}} \rightarrow Y_{\text{supplementary}}\right)} = \frac{\#\text{ of patents}\left(X_{\text{primary}},\ Y_{\text{supplementary}}\right)}{\#\text{ of patents}} \tag{1}$$

$$\mathrm{CI}_{\left(X_{\text{primary}} \rightarrow Y_{\text{supplementary}}\right)} = \frac{\#\text{ of patents}\left(X_{\text{primary}}, Y_{\text{supplementary}}\right)}{\#\text{ of patents}\left(X_{\text{primary}}\right)} \tag{2}$$

$$\mathrm{CS}_{\left(X_{\text{primary}} \rightarrow Y_{\text{supplementary}}\right)} = \frac{\mathrm{CI}_{\left(X_{\text{primary}} \rightarrow Y_{\text{supplementary}}\right)}}{\#\text{ of patents }\left(Y_{\text{supplementary}}\right)} = \frac{\#\text{ of patents }\left(X_{\text{primary}},\ Y_{\text{supplementary}}\right)}{\#\text{ of patents }\left(X_{\text{primary}}\right) \cdot \#\text{ of patents }\left(Y_{\text{supplementary}}\right)} \tag{3}$$

Using these indexes, the procedure for constructing a technology ecology network from the converging fields part of the matrix consists of three steps, as exemplified in "Appendix 1". First, the frequently co-occurred structural patterns that exceed the prescribed threshold values for CP are identified based on the Apriori algorithm. Specifically, this algorithm identifies the frequent individual primary and supplementary classes and extends them to larger co-occurred patterns as long as the patterns appear frequently in the database. Second, the patterns with CS values greater than one, are selected as significant among the frequently co-occurred structured patterns, after their CI values are calculated. Here, the threshold values for CP and CI should be carefully determined as hyperparameters. For instance, if pattern $X \rightarrow Y$ has a high CP value and a low CI value, the convergence between patent classes X and Y is less likely to be significant. If pattern $X \rightarrow Y$ has a low value of CP and a high value of CI, convergence between patent classes X and Y may not be worth considering for further analysis. That is, using large threshold values for CP and CI may create more meaningful results by restricting the scope of analysis to major patterns of multi-technology convergence. In contrast, using small threshold values will provide a practical solution by generating more patterns at micro-level. Finally, the identified significant structural patterns are represented as a technology ecology network for a comprehensive view of multi-technology convergence.

The resulting technology ecology network is a directed and valued network that represents the significant structural patterns of multi-technology convergence. Specifically, a source and a target node represent a primary patent class and supplementary patent classes; a link represents the CI among the associated classes; and the size of a node indicates the number of corresponding patent classes in the identified patterns, which can be regarded as the overall importance (OI) of the patent class in convergence settings. The technology ecology networks are constructed for different time periods, based on which analysis results are compared between the periods to present implications for the changing patterns of multi-technology convergence.

Measuring the changing patterns of multi-technology convergence

Technological changes do not occur at random, but are affected by many factors such as state-of-the-art technologies already in use, R&D activities and efforts, technological advances achieved by firms and organisations, and significant regularities (Dosi 1984). Following this seminal work, researchers have examined the patterns of technology evolution in different contexts—for example, recombinant search based on technology landscapes (Fleming 2001; Fleming and Sorenson 2001), technology development paths (Barberá-Tomás et al. 2011; Choi and Park 2009), and technology cluster evolution (Érdi et al. 2013)—to direct investment and to reduce risks associated with technology development. However, the changing patterns of convergence is seldomly investigated, except for the identification of core technologies that drive convergence over time.

In this context, a more detailed analysis should be conducted to obtain richer information and to assist decision-making in the design and development of innovation ecosystems, although changes in the structural patterns of multi-technology convergence can be identified from the technology ecology networks for different time periods. To examine the changing patterns of multi-technology convergence, we categorise the structural patterns into three broad types according to their CS values, as follows. First, the patterns with CS values less than one in period $t-1$ and greater than one in period t are regarded as *emergence*. Second, the patterns with CS values greater than one in period $t-1$ and less than one in period t are classified as *disappearance*. Finally, the patterns with CS values greater than one in both periods are considered as *persistence*. The patterns of *persistence* can be further classified into four sub-categories in terms of changes in CP and CI values. The change in CP (CCP) measures the degree to which the corresponding pattern becomes a mainstream trend, as defined in Eq. (4). The change in CI (CCI) represents the degree to which the corresponding pattern intensifies, as shown in Eq. (5).

$$\mathrm{CCP}_{\left(X_{\text{primary}} \to Y_{\text{supplementary}}\right)_{\Delta t}} = \mathrm{CP}_{\left(X_{\text{primary}} \to Y_{\text{supplementary}}\right)_{t+\Delta t}} - \mathrm{CP}_{\left(X_{\text{primary}} \to Y_{\text{supplementary}}\right)_{t}} \tag{4}$$

$$\mathrm{CCI}_{\left(X_{\text{primary}} \to Y_{\text{supplementary}}\right)_{\Delta t}} = \mathrm{CI}_{\left(X_{\text{primary}} \to Y_{\text{supplementary}}\right)_{t+\Delta t}} - \mathrm{CI}_{\left(X_{\text{primary}} \to Y_{\text{supplementary}}\right)_{t}} \tag{5}$$

Anticipating the prospects of multi-technology convergence

One of the most integral questions in technology convergence research is: *which technology fields will converge in the future?* Based on this, we have defined a computational problem for multi-technology convergence: given the technology ecology networks for the first and second periods, is it possible to identify the links that will exist in the network for the third period? To answer the question, first, all possible combinations of patent classes are generated to consider different changing patterns of technology convergence. For instance, some combinations may become significant during the third period, although they were not significant during the first and second periods. Here, we consider the combinations comprising of a primary and a supplementary patent class, although the method is not limited to this level, and can allow more complex patterns. Second, we compute the CP, CI, and CS values during the first and second periods and the CCP, CCI, and change in CS (CCS) values between the periods for the generated combinations. Finally, by using the index scores computed in the second step as inputs, we develop a logistic regression model as a classifier, to identify the significant patterns of multi-technology convergence during the third period (i.e. the patterns with CP, CI, and CS values over the threshold values).

The distinct characteristics of the proposed link prediction analysis regarding previous link prediction methods (Kim et al. 2019; Lee Han et al. 2015) are as follows. First, the proposed approach considers various factors describing the changing patterns of multi-technology convergence that are extracted from multiple technology ecology networks (e.g. CI among the associated classes and the difference between periods), while previous methods are based on similarity scores and rely solely on the network information at a certain point in time (e.g. existence and absence of links). Second, the proposed approach can present information about the likely timings of multi-technology convergence by considering time periods.

The proposed method is assessed by examining several quantitative performance metrics after a confusion matrix is constructed. Considering that the number of patterns for each class (i.e. significant and insignificant) is imbalanced, we use a fivefold stratified sampling technique for performance evaluation. Moreover, in addition to basic metrics such as accuracy, precision, and recall, we employ the F_1 score (Powers 2011) and Youden's J statistic (Youden 1950) to compensate for classes with different weights in our case study. Specifically, the F_1 score measures the overall effectiveness of a classifier and is defined as the harmonic average of precision and recall, as shown in Eq. (6). This measure presents the best value at 1 and the worst at 0.

$$F_1 \text{ score} = 2 \times \frac{1}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{6}$$

The Youden's J statistic also captures the effectiveness of a classifier, as shown in Eq. (7). It ranges from -1 to 1 (perfect classification), with $J=0$, meaning that the classifier provides the same proportion of positive results for classes regardless of true conditions.

$$\text{Youden's } J = \text{Specificity} + \text{Sensitivity} - 1 \tag{7}$$

Case study

We conducted a case study of drug, bio-affecting, and body treating compositions technology for three reasons. First, multi-technology convergence has advanced more rapidly in this field over time (Jiang and Luan 2018). In particular, this technology is closely related to multiple technologies such as molecular biology technology, microbiology technology, natural resins or derivatives technology, and peptide or protein technology (Kim and Lee

2017). Second, given the substantial development costs, risks, and the rate of high development in this area, industrial practitioners call for high-quality and well-organised information to assist with decision-making in the design and development of innovation ecosystems (Chen and Chang 2010). In this context, measuring multi-technology convergence is a crucial activity for defining innovation boundaries and developing collaboration strategies (e.g. sharing the cost and risk of technology development and reducing the time to market). Finally, patents are especially important in this area, since a patent normally equals a product and the manufacturing process is relatively easy to replicate (Chaudhuri 2005), which means that the technology dimension of innovation ecosystems is more important than other dimensions. Therefore, drug, bio-affecting, and body treating compositions technology serve as a suitable test case for assessing the feasibility and utility of the proposed approach.

Data collection and pre-processing

The United States Patent and Trademark Office database served as a data collection source in this study, since the United States is the world's largest patent market and is considered appropriate for analysing international technologies (Lee Kang et al. 2015). A total of 39612 patents of 424 class (entitled 'drug, bio-affecting, and body treating compositions') (Set 1) were first collected over the reference period of 2006–2014. These patents were then parsed based on their structures to differentiate each patent document by content, while details about patent number, issued date, primary class, supplementary classes, and citations were stored in a structured patent database. The patents that cite Set 1 patents (Set 2) were also collected, pre-processed, and stored in the same manner.

The patent field-value matrix for drug, bio-affecting, and body treating compositions technology was constructed as per Table 1. The patents were grouped according to time periods (2006–2008, 2009–2011, and 2012–2014), and each group was analysed separately in the second step. With respect to the converging fields part, various levels of patent classification, such as patent classes and subclasses, can serve as a unit of analysis. Considering the trade-off between class-level analysis and subclass-level analysis (Lee et al. 2012), mainline subclasses were selected as a unit of analysis in this study. Mainline subclass-level analysis has proved successful in enabling fine-grained characterisation of patented inventions (Aharonson and Schilling 2016; Lee and Lee 2019; Lee et al. 2018). As for the technological value part, the forward citation counts of patents were extracted since the patents were published.

Constructing technology ecology networks for different time periods

This step identifies the significant structural patterns of multi-technology convergence for the three periods via the modified ARM. A total of 351, 269, and 286 patterns, with CP, CI, and CS values greater than 0.003, 0.01, and 1, respectively, were identified as significant (Table 2) based on the results of the sensitivity analysis (Table 3).

The technology ecology networks for the three periods present different structural patterns of multi-technology convergence, as shown in Fig. 2. In the network, white and grey nodes represent single (e.g. 424/184.1) and multiple patent classes (e.g. {436/501, 424/9.1, 530/350}); blue and red links indicate within-technology convergence where the primary patent class also appears in the supplementary patent class part (e.g. 424/43 → 424/400) and between-technology convergence where the primary patent class is

Table 1 Part of the patent field-value matrix

Basic information		Converging fields		Technological value				
Patent number	Issued year	Primary	Supplementary	2007	2008	…	2016	2017
7101573	2006	424/400	424/400, 424/400, 424/400, 424/400	0	0	…	8	2
7122188	2006	424/130.1	424/130.1, 424/130.1	0	0	…	0	1
7294704	2007	530/350	424/130.1, 530/350, 530/350, 530/350	0	0	…	0	2
7297330	2007	424/93.1	424/520, 514/1	0	0	…	0	0
7297537	2007	435/320.1	424/184.1, 435/235.1, 435/325	0	4	…	3	5
7470431	2008	424/184.1	424/184.1, 424/9.1	–	0	…	0	0
7470441	2008	424/725		–	0	…	3	0
7470675	2008	514/1	424/93.1	–	0	…	0	1
…	…	…	…	…	…	…	…	…
7628995	2009	424/184.1		–	–	…	4	2
7625865	2009	514/1	424/1.11, 530/300	–	–	…	11	5
7608683	2009	530/300	424/184.1, 530/350	–	–	…	0	0
7840247	2010	600/300	378/1, 378/1, 424/9.1	–	–	…	10	0
7850954	2010	424/78.02	424/400, 424/59, 424/62	–	–	…	4	1
8071135	2011	424/520		–	–	…	11	2
8080055	2011	623/1.1	424/400, 623/1.1, 623/1.1	–	–	…	2	0
8084021	2011	424/85.1	530/350, 530/350	–	–	…	0	1
…	…	…	…	…	…	…	…	…
8119767	2012	530/350	424/184.1	–	–	…	1	0
8147861	2012	424/400	424/600, 427/2.1	–	–	…	24	9
8584670	2013	128/200.14	128/200.24, 424/43, 424/400	–	–	…	0	0
8586056	2013	424/184.1	424/184.1, 424/184.1, 424/184.1, 530/350	–	–	…	0	2
8916209	2014	424/725	424/725	–	–	…	0	0
8921526	2014	530/350	424/130.1, 424/130.1, 424/130.1, 530/350, 530/350	–	–	…	16	2

Table 2 Part of the significant structural patterns of multi-technology convergence

Significant structural pattern	CP	CI	CS	Period
{424/184.1} → {424/184.1, 435/235.1}	0.005	0.049	4.437	1
{424/184.1} → {435/320.1, 435/41}	0.005	0.049	1.789	1
{424/70.1} → {424/400}	0.005	0.314	1.157	1
{424/130.1} → {435/325}	0.006	0.080	1.222	1
{435/4} → {424/184.1}	0.013	0.440	2.422	1
{435/4} → {424/184.1, 435/4}	0.006	0.215	8.410	1
{435/4} → {424/130.1, 530/350}	0.004	0.135	1.931	1
…	…	…	…	…
{424/1.11} → {424/1.11}	0.010	0.670	17.554	2
{435/243} → {424/93.1}	0.005	0.747	10.830	2
{435/4} →{424/9.1}	0.010	0.308	5.077	2
{435/4} → {424/184.1, 435/4}	0.006	0.215	8.410	2
{530/350} → {424/130.1, 530/350}	0.020	0.440	6.056	2
{530/350} → {424/178.1}	0.006	0.138	5.325	2
{600/300} → {424/9.1}	0.003	0.854	14.064	2
…	…	…	…	…
{424/130.1} → {435/325, 530/350}	0.006	0.064	5.013	3
{424/43} → {424/400}	0.004	0.477	1.913	3
{424/85.1} → {530/350}	0.005	0.462	3.653	3
{435/4} → {424/9.1, 422/50}	0.003	0.154	45.407	3
{435/4} → {424/9.1}	0.008	0.406	8.618	3
{510/108} → {510/108}	0.003	0.934	98.907	3
{514/1} → {424/93.1}	0.005	0.090	1.490	3

Table 3 Numbers of the nodes identified with different threshold values

		Threshold value for CI		
		0.005	0.01	0.015
(a) First period (2006–2008)				
Threshold value for CP	0.002	593	593	593
	0.003	351	351	351
	0.004	249	249	249
(b) Second period (2009–2011)				
Threshold value for CP	0.002	516	511	511
	0.003	269	269	269
	0.004	174	174	174
(c) Third period (2012–2014)				
Threshold value for CP	0.002	495	488	470
	0.003	286	286	278
	0.004	165	165	165

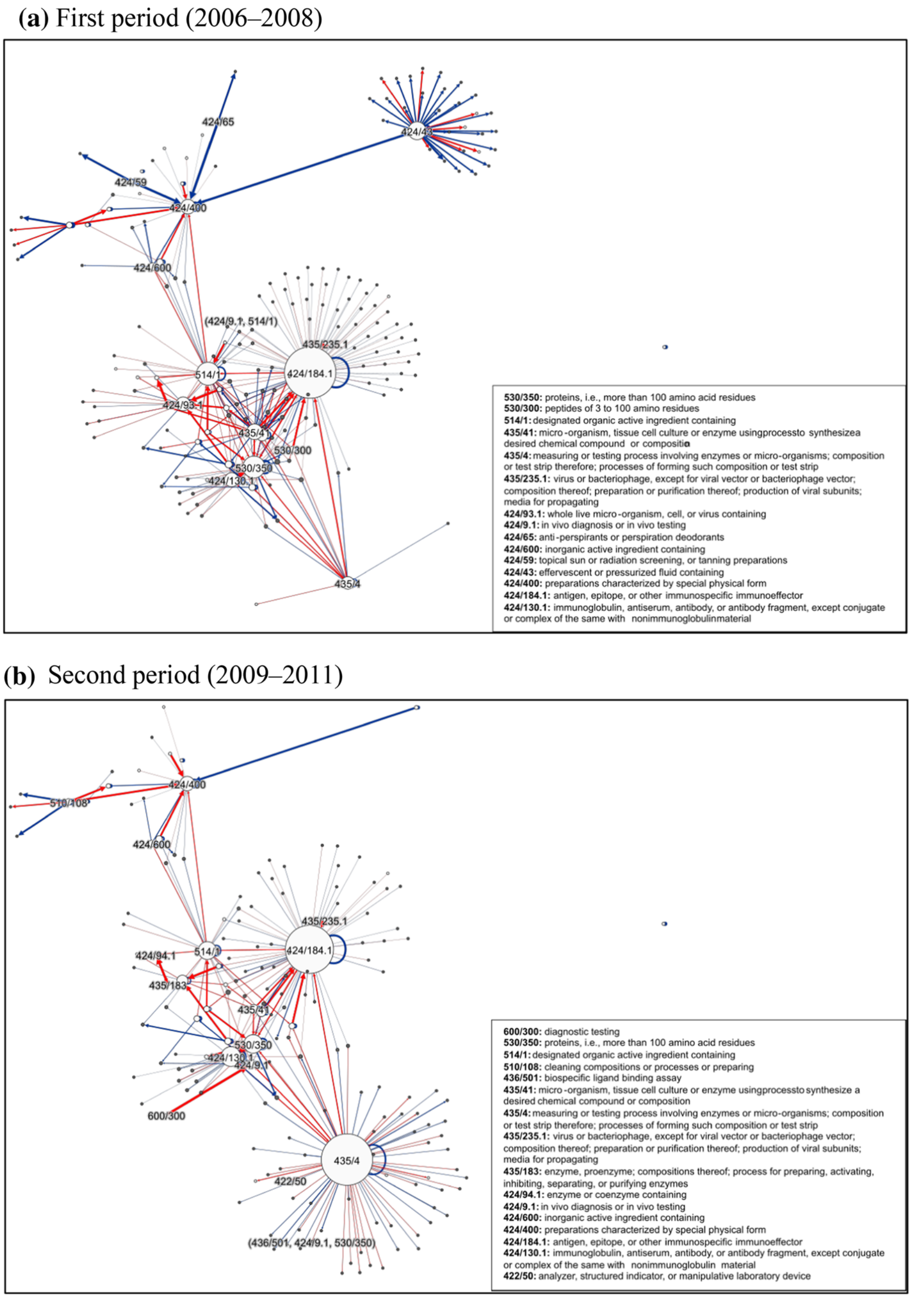

Fig. 2 Technology ecology networks

(c) Third period (2012–2014)

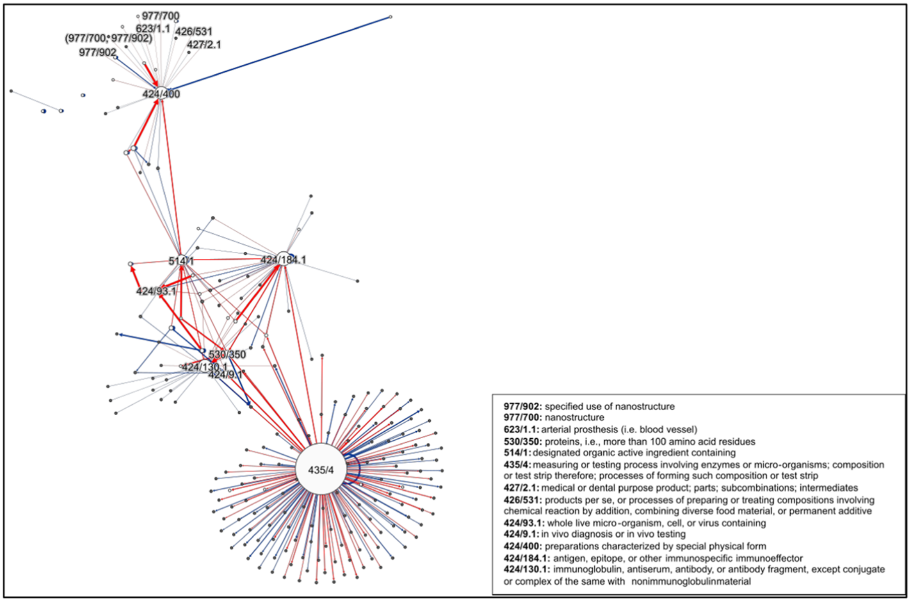

Fig. 2 (continued)

not included in the patent class segment (e.g. 435/41 → 530/350), respectively. Also, the positions of nodes were fixed in the networks to allow for effective comparisons of the structural patterns between the periods, and the description of USPC mainline subclasses is summarised in "Appendix 1". The three technology ecology networks show that convergence occurs in different patterns and that the technological scope in this field increases over time. The patent classes driving convergence during the first period include 424/184.1 (antigen, epitope, or other immune-specific immune transfer factor) (OI = 0.141), 514/1 (designated organic active ingredient containing) (OI = 0.060), and 530/350 (proteins, i.e. more than 100 amino acid residues) (OI = 0.057). Among others, 424/184.1 is worth noting. As a primary technology, it links to 530/300 (peptides of 3 to 100 amino acid residues) and {424/9.1 (in vivo diagnosis or in vivo testing), 514/1}, whereas 424/184.1, as a supplementary technology, links to 435/4 (measuring or testing progress involving enzymes or micro-organisms; composition or test strip therefore; processes of forming such composition or test strip) and 435/235.1 (virus or bacteriophage, except for viral vector or bacteriophage vector; composition thereof; preparation or purification thereof; production of viral subunits; media for propagating). Moreover, very strong links exist between the following patent classes: 424/65 (anti-perspirants or perspiration deodorants) → 424/400 (preparations characterised by special physical form) (CP = 0.004, CI = 0.911, CS = 3.353), 424/59 (topical sun or radiation screening, or tanning preparations) → 424/400 (CP = 0.011, CI = 0.851, CS = 3.131), and 435/235.1 → 424/184.1 (CP = 0.003, CI = 0.842, CS = 4.636). Here, it should be noted that the proposed approach can discern the direction of convergence and identify not only the convergence between two technologies but also between more than two technologies (e.g. 424/184.1 → {424/9.1, 514/1}). The analysis results are considered crucial in effectively establishing innovation ecosystems, informing decision-making in defining innovation boundaries, clarifying the product and application sectors, and specifying the role of actors within innovation ecosystems.

The patent classes that drive convergence during the second period are quite different from those in the first period. During the second period, enzymes and microorganisms have become an integral part of drug, bio-affecting, and body treating compositions technology because of the ability to promote reactions more quickly and effectively. Several industrial sectors such as cosmetics, animal feed, and agricultural industries have shown interest in enzymes and microorganisms for specific applications. The technology ecology network for this period confirms that the most significant patterns involved in patent classes 424/43 (effervescent or pressurised fluid containing) during the first period become insignificant, while 435/4 (measuring or testing process involving enzymes or micro-organisms; composition or test strip therefore; processes of forming such composition or test strip) emerges as one of the most important patent classes in driving convergence (OI = 0.139). In the network, 435.4, as both primary and supplementary technologies, has many relations with other classes such as 435/41 (micro-organism, tissue cell culture or enzyme using process to synthesize a desired chemical compound or composition), 436/501 (biospecific ligand binding assay), 424/9.1, 530/350}, and {436/518 (involving an insoluble carrier for immobilising immunochemicals), 422/50 (analyser, structured indicator, or manipulative laboratory device)}. In addition, 422/50, which is not associated with any significant pattern during the first period, becomes involved with 16 significant patterns during the second period. Here, it should be noted that many of these patterns have small CP values but relatively large CI values that cannot be easily identified by conventional patent co-classification network analysis. There are also changes in the list of strong links, although some of the strong links during the first period remain strong during the second period (e.g. 435/235.1 → 424/184.1). The strongest links during the second period exist between the following patent classes: 600/300 (diagnostic testing) → 435/4, 435/235.1 → 424/184.1, and 435/183 (enzyme [e.g. ligases (6.), etc.], proenzyme; compositions thereof; process for preparing, activating, inhibiting, separating, or purifying enzymes) → 424/94.1 (enzyme or coenzyme containing).

Finally, convergence with patent class 435/4 becomes the dominant trend during the third period, while the most significant patterns involved in 424/184.1 are no longer significant. The value of OI of 435/4 almost doubles from 0.139 to 0.255. Patent class 424/400 is also noteworthy since it has relations with many new classes such as 426/531 (products per se, or processes of preparing or treating compositions involving chemical reaction by addition, combining diverse food material, or permanent additive), 623/1.1 (arterial prosthesis [i.e. blood vessel]), 427/2.1 (medical or dental purpose product; parts; subcombinations; intermediates [e.g. balloon catheter, splint]), 977/700 (nanostructure), and 977/902 (specified use of nanostructure). Moreover, convergence with nanotechnology (e.g. 977/700 and 977/902) become significant during the third period. Exemplary patents are 8768806 (lubricating gels consisting of nanoparticles to deliver the cure to a target tissue) and 8834902 (biodegradable in vivo supporting device using nanoparticles).

Measuring the changing patterns of multi-technology convergence

The patterns were classified into three types according to their CS values, as shown in Fig. 3. Only major patterns are presented in the figures owing to a lack of space. Although managerial guidelines may differ across the organisational context, organisations should investigate the changing patterns in the direction shifting from *emergence*

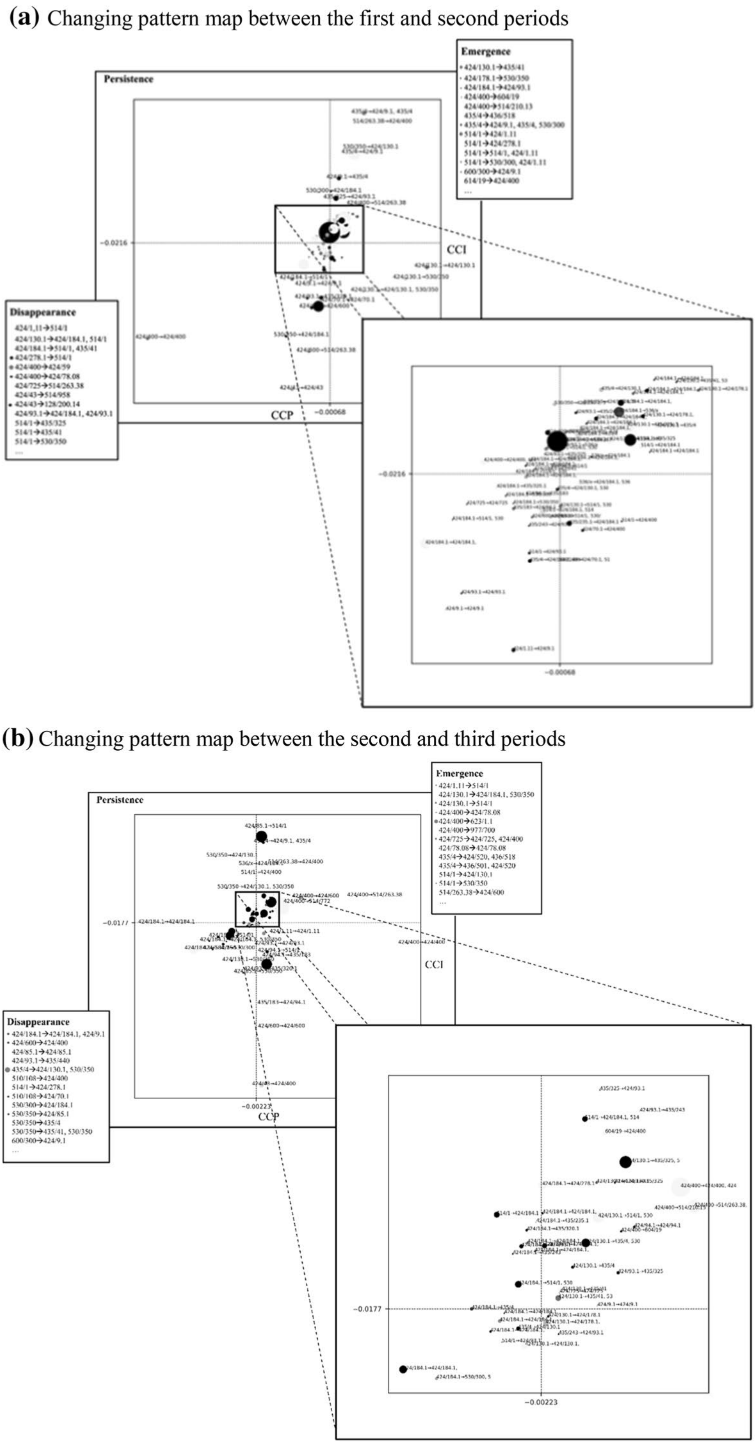

Fig. 3 Changing patterns of technology convergence

and *disappearance* to *persistence*. First, a total of 107 and 110 structural patterns of multi-technology convergence associated with *emergence* between the first and second periods and between the second and third periods. For instance, patterns, 424/178.1 (conjugate or complex of monoclonal or polyclonal antibody, immunoglobulin, or fragment thereof with nonimmunoglobulin material) → 530/350, 600/300 → 424/9.1, and 435/4 → 436/518 (involving an insoluble carrier for immobilising immunochemicals) emerges during the second period. Secondly, 159 and 93 structural patterns are categorised as *disappearance* between the first and second periods and between the second and third periods. For instance, patterns, 424/1.11 (radionuclide or intended radionuclide containing; adjuvant or carrier compositions; intermediate or preparatory compositions) → 514/1 and 424/93.1 (whole live microorganism, cell, or virus containing) → {424/184.1, 424/93.1} are identified as significant only during the first period. The patterns of *emergence* and *disappearance* can be highly disruptive and may threaten the sustainability of innovation ecosystems. Organisations should identify the probable future conditions that may be driven by emerging and disappearing patterns and examine their direct and indirect impacts on innovation ecosystems and organisational innovation capabilities. Finally, the remaining 162 and 176 patterns are classified as *persistence* between the first and second periods and between the second and third periods. The patterns of this category require continuous monitoring. Organisations should develop and adjust innovation ecosystems to increase the fitness of ecosystems to changing environments and reveal favourable future conditions. They were further divided into four sub-categories according to their CCP and CCI values for in-depth investigation of the changing patterns of multi-technology convergence. The patterns with high CCP and CCI are the dominant trends of convergence, which should be traced continuously. Specifically, for patterns with high CCP values, the direction and quality of technology development should be monitored to sustain a competitive edge and to establish benchmarking and differentiation strategies. For patterns with high CCI values, organisations should assess the potential of converging segments as the dominant trend or find niche markets to which the relevant technologies can be applied.

Anticipating the prospects of multi-technology convergence

Given the CP, CI, and CS values during the first and second periods, and the CCP, CCI, and CCS values between the periods as inputs, a logistic regression-based link prediction analysis was conducted to anticipate the prospects of multi-technology convergence. Here, the maximum number of classes of a pattern was restricted to three, although the proposed approach can allow for more complex analyses. The possible combinations of patent classes were classified into two categories, significant and insignificant, as reported in Table 4.

Table 5 reports on the performance of the proposed link prediction analysis. Although there are differences in the degree of performance across different classes, the results confirm that the proposed link prediction analysis is effective in identifying the links that will exist in the network for the third period using the limited quantitative indexes that are extracted from the networks for the first and second periods. Hence, we are confident that the proposed approaches are useful not only for measuring the changing patterns but also for anticipating the prospects of multi-technological convergence.

Table 4 Part of the results of link prediction analysis

Combinations	CP (period 1)	CI (period 1)	CS (period 1)	CP (period 2)	CI (period 2)	CS (period 2)	CCP	CCI	CCS	Predicted condition in period 3	Actual condition in period 3
424/1.11 → 424/9.1	0.006	0.350	5.501	0.004	0.256	4.212	−0.002	−0.095	−1.289	Significant	Insignificant
424/1.11 → 428/357	0	0	0	0.000	0.011	2.487	0.000	0.011	2.487	Insignificant	Insignificant
424/130.1 → 514/1	0.012	0.172	0.993	0.012	0.118	0.806	−0.000	−0.054	−0.186	Significant	Insignificant
424/400 → 424/400	0.112	0.636	2.340	0.083	0.487	2.056	−0.030	−0.149	−0.284	Significant	Significant
424/65 → 424/600	0.000	0.067	1.480	0.000	0.150	4.162	−0.000	0.083	2.682	Insignificant	Insignificant
424/85.1 → 530/350	0.009	0.509	2.678	0.009	0.547	3.034	0.000	0.038	0.357	Significant	Significant
424/94.1 → 435/320.1	0.002	0.139	2.222	0.001	0.050	1.437	−0.002	−0.089	−0.785	Insignificant	Insignificant
…	…	…	…	…	…	…	…	…	…	…	…
435/4 → 424/9.1	0.006	0.211	3.312	0.009	0.308	5.076	0.003	0.097	1.764	Significant	Significant
435/4 → 530/350	0.009	0.295	1.549	0.009	0.283	1.568	−0.000	−0.012	0.019	Significant	Significant
514/263.38 → 424/400	0.009	0.483	1.779	0.014	0.621	2.621	0.005	0.137	0.842	Significant	Significant
514/1 → 424/1.11	0.002	0.039	1.175	0.011	0.150	3.929	0.008	0.111	2.754	Insignificant	Significant
800/8 → 424/9.1	0.001	0.556	8.724	0.001	0.600	9.885	−0.000	0.044	1.160	Insignificant	Insignificant
800/8 → 530/350	0.000	0.056	0.292	0	0	0	−0.000	−0.056	−0.292	Insignificant	Insignificant

Table 5 Results of performance evaluation

		True condition in the third period Significant	Insignificant
(a) Confusion matrix			
Predicted class	Significant	65	13
	Insignificant	30	84,552

Type	Accuracy	Precision	Recall	Specificity	*F* measure	Youden's *J* statistic
(b) Quantitative performance metrics						
All	0.999	0.833	0.684	0.999	0.751	0.684
424 → 424	0.994	0.737	0.583	0.998	0.651	0.581
424 → non-424	0.999	0.813	0.684	0.999	0.743	0.684
Non-424 → 424	0.999	0.684	0.722	0.999	0.703	0.722
Non-424 → non-424	0.999	0.765	0.867	0.999	0.813	0.867

Discussion

Characteristics of the proposed approach

Several considerations should be made before applying and deploying a new method in practice. The characteristics of the proposed approach are summarised in innovation ecosystem contexts by comparing the results of patent co-classification network analysis (Fig. 4) and the proposed approach, as follows. First, the proposed approach identifies the direction of technology convergence trend based on the distinction between the primary and other classes of a patent. Specifically, the convergence between patent class X as a primary class and patent class Y as supplementary classes, is considered different compared to the opposite. The results of our analysis show that only 22%, 25%, and 18% of the significant patterns identified are significant in both directions (i.e. $X \rightarrow Y$ and $Y \rightarrow X$) during the first, second, and third periods, respectively. Hence, the technology ecology network can provide more detailed information on the relations between technologies, and can assist decision-making on the role of actors within innovation ecosystems and collaboration strategies. Second, conventional patent co-classification network analysis does not provide information on multi-technology convergence, since this method is based solely on the number of co-occurrences of two different patent classes. In this respect, the proposed ARM approach can provide a fine-grained characterisation of the technology dimension of innovation ecosystems by measuring the significant patterns of multi-technology convergence. It is noteworthy that 63%, 59%, and 67% of the patterns identified during the first, second, and third periods represent the convergence between more than two technologies, which cannot be easily identified by conventional patent co-classification network analysis, but should be considered in developing innovation ecosystems. Therefore, the proposed approach allows the scope and boundaries of innovation ecosystems to be defined more accurately. Third, the number of co-occurrences of two different patent classes should be converted into binary values on the basis of a certain cut-off value in patent co-classification network analysis. Here, using large cut-off values increases the risk of missing significant patterns of technology convergence, whereas small cut-off values increase the

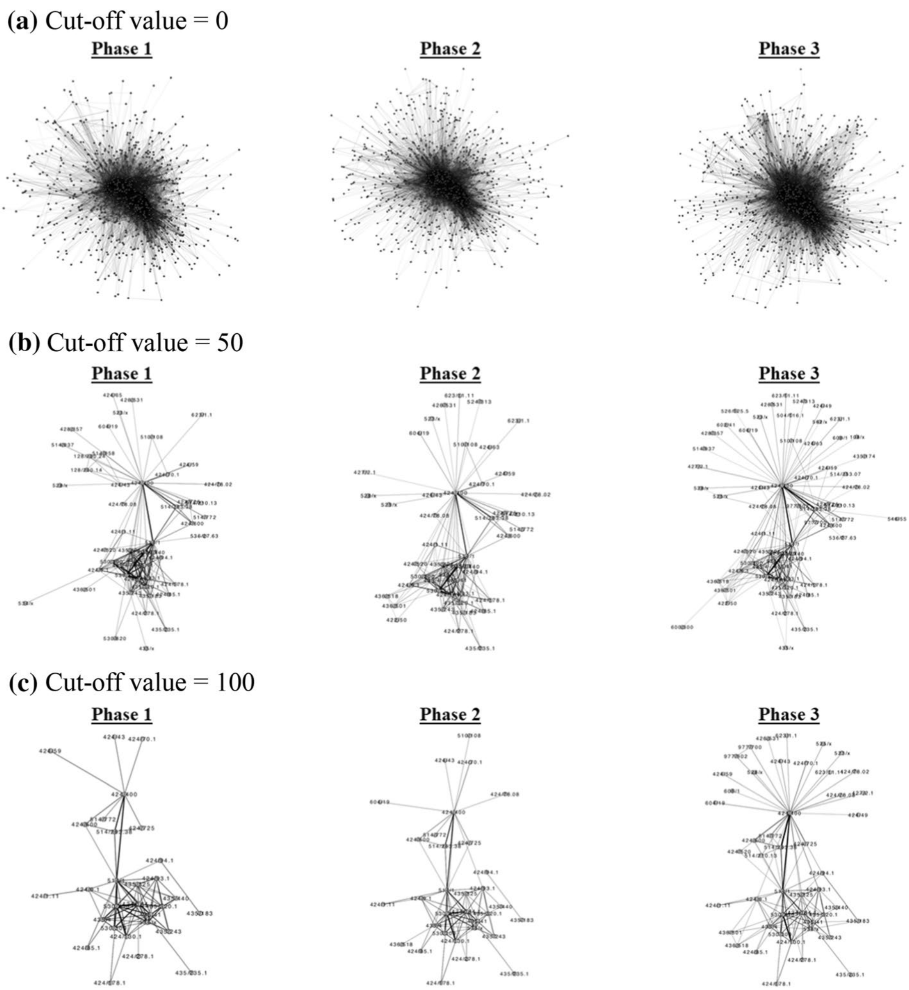

Fig. 4 Results of patent co-classification network analysis

possibility of including insignificant patterns (Jeong et al. 2015; Kwon et al. 2020; No and Park 2010). The proposed approach overcomes these limitations by identifying the significant structural patterns of technology convergence based on the statistical dependence between technology fields (i.e. CS values). Although our approach also requires threshold values for CP and CI, they are not used to identify significant patterns but to frame the scope of analysis. Moreover, the values of the proposed indexes such as CP, CI, and CS can be compared between the periods, since they are based on the relative value (i.e. not absolute value). Consequently, the proposed approach presents a more accurate and reliable representation of the technological dimension of innovation ecosystems, and guides organisations towards responding to the evolution of innovation ecosystems. Finally, the proposed approach examines the scope of technology convergence. Among the patterns identified during the first, second, and third periods, 57%, 53%, and 54% of the patterns represent within-technology convergence where the primary patent class also appears in

Table 6 Technological impact and uncertainty of the patterns

Significant structural pattern	TI	TU	Period
{424/184.1} → {424/184.1, 435/235.1}	0.533	1.807	1
{424/184.1} → {435/320.1, 435/41}	0	0	1
{424/70.1} → {424/400}	0.222	0.548	1
{424/130.1} → {435/325}	1.333	1.528	1
{435/4} → {424/184.1}	0.231	0.583	1
{435/4} → {424/184.1, 435/4}	0.400	1.056	1
{435/4} → {424/130.1, 530/350}	0.250	0.452	1
…	…	…	…
{424/1.11} → {424/1.11}	0.319	0.501	2
{435/243} → {424/93.1}	0.146	0.545	2
{435/4} →{424/9.1}	1.278	3.478	2
{435/4} → {424/184.1, 435/4}	0.083	0.289	2
{530/350} → {424/130.1, 530/350}	0.375	1.003	2
{530/350} → {424/178.1}	0	0	2
{600/300} → {424/9.1}	0.222	0.732	2
…	…	…	…
{424/130.1} → {435/325, 530/350}	0.333	0.724	3
{424/400} → {424/400, 424/600}	2.727	6.341	3
{424/85.1} → {530/350}	0.194	0.468	3
{435/4} → {424/9.1, 422/50}	0	0	3
{435/4} → {424/9.1}	0.056	0.236	3
{510/108} → {510/108}	0	0	3
{514/1} → {424/93.1}	0.083	0.279	3

the supplementary patent class part, while the remaining 43%, 47%, and 46% indicate between-technology convergence where the primary patent class is not included in the patent class segment.

Technological impact and uncertainty of multi-technology convergence

Although the proposed approach has certain advantages over previous methods in understanding the technology dimensions of innovation ecosystems, given organisations' resource limitations, the identification of impactful and significant patterns in the technology ecology network is required in innovation ecosystem contexts (Hacklin et al. 2009). In this respect, many empirical studies have shown a significant positive relationship between the number of forward citations of a patent and the technological impact or economic value of patented technology (Narin et al. 1987). Additionally, recent studies have adopted the standard deviation or variance of the forward citations of a patent as a proxy for the technological uncertainty or risk of patented technology (Jang et al. 2017; Shin et al. 2013). Following this convention, we measure the technological impact (TI) and the technological uncertainty (TU) of the patterns of multi-technology convergence as the mean and the standard deviation of forward citations of the relevant patents, as defined in Eqs. (4) and (5):

Table 7 Descriptive statistics for the technological impact and uncertainty of the patterns

Period	Index	Min	25th	Median	75th	Max	Mean	S.D.	Skewness	Kurtosis
1	TI	0	0	0	0	2.222	0.084	0.257	4.966	30.236
	TU	0	0	0	0	3.833	0.175	0.483	4.494	25.161
2	TI	0	0	0	0.456	3.478	0.287	0.534	2.629	8.595
	TU	0	0	0	0.167	1.933	0.131	0.274	3.213	12.470
3	TI	0	0	0	0.148	2.727	0.135	0.319	4.345	24.164
	TU	0	0	0	0.407	6.341	0.350	0.809	4.171	21.897

$$\mathrm{TI}_{\left(X_{\text{primary}} \rightarrow Y_{\text{supplementary}}\right)} = \frac{\sum_{i=1}^{\#\text{ of patents}\left(X_{\text{primary}},\ Y_{\text{supplementary}}\right)} \text{forward citations of patent}_i}{\#\text{ of patents }\left(X_{\text{primary}},\ Y_{\text{supplementary}}\right)} \tag{8}$$

$$\mathrm{TU}_{\left(X_{\text{primary}} \rightarrow Y_{\text{supplementary}}\right)} = \sqrt{\frac{\sum_{i=1}^{\#\text{ of patents}\left(X_{\text{primary}},\ Y_{\text{supplementary}}\right)} \left(\text{forward citations of patent}_i - \mathrm{TI}_{\left(X_{\text{primary}} \rightarrow Y_{\text{supplementary}}\right)}\right)^2}{\#\text{ of patents}\left(X_{\text{primary}},\ Y_{\text{supplementary}}\right)}} \tag{9}$$

Table 6 shows the TI and TU of the identified patterns. For this, we employed the number of forward citations of patents over the next 3 years, after the patents were issued, to aid decision-making in short-term technology planning. As per the results of prior studies (Harhoff et al. 2003; Karki 1997; Lee et al. 2012), the skewed distributions to the right are observed for the TI and TU of the patterns for the three periods, as reported in Table 7. The average values of the TI and TU of the patterns are 0.084 and 0.175 for the first period, 0.287 and 0.131 for the second period, and 0.135 and 0.350 for the third period. In terms of the TI, only 67, 79, and 75 patterns have an impact above the average values for the first, second, and third periods, respectively.

Moreover, the changes in TI and TU values should be considered to explore the implications of the evolution of innovation ecosystems. In this respect, the change in TI (CTI) denotes the degree to which the corresponding pattern becomes technologically more impactful, as defined in Eq. (10). Patterns with high CTI values are likely to become future key technologies. The change in CU value (CCU) indicates the degree to which the corresponding pattern becomes technologically more uncertain, as shown in Eq. (11). The risk and uncertainty factors that affect technology development and commercialisation should be identified for patterns with high CTU values.

$$\mathrm{CTI}_{\left(X_{\text{primary}} \rightarrow Y_{\text{supplementary}}\right)_{\Delta t}} = \mathrm{TI}_{\left(X_{\text{primary}} \rightarrow Y_{\text{supplementary}}\right)_{t+\Delta t}} - \mathrm{TI}_{\left(X_{\text{primary}} \rightarrow Y_{\text{supplementary}}\right)_{t}} \tag{10}$$

$$\mathrm{CTU}_{\left(X_{\text{primary}} \rightarrow Y_{\text{supplementary}}\right)_{\Delta t}} = \mathrm{TU}_{\left(X_{\text{primary}} \rightarrow Y_{\text{supplementary}}\right)_{t+\Delta t}} - \mathrm{TU}_{\left(X_{\text{primary}} \rightarrow Y_{\text{supplementary}}\right)_{t}} \tag{11}$$

In Fig. 3, the size of a circle is proportional to its CTI value, and the shade of a circle's colour is proportional to its CTU value. The TI and TU of the patterns vary between periods. Among the top ten impactful patterns during the third period, no patterns are found in the top ten lists of the first and second periods, except for 424/400 → {424/400, 424/600 (inorganic active ingredient containing), which lasts from the second period. Moreover, different patterns present different changing behaviours of TU. For instance, converging pattern, 424/130.1 (immunoglobulin, antiserum, antibody, or antibody fragment, except conjugate or complex of the same with nonimmunoglobulin material) → 435/325 (animal cell, per se [e.g. cell lines, etc.]; composition thereof; process of propagating, maintaining or preserving an animal cell or composition thereof; process of isolating or separating an animal cell or composition thereof; process of preparing a composition containing an animal cell; culture media thereof), has the highest level of TU during the first period (TI = 1.333, TU = 1.528), its TU decreases during the second period (TI = 0.111, TU = 0.333), and the converging pattern becomes insignificant during the third period. In contrast, pattern 424/400 → {424/400, 424/600} has a low level of TU during the first period (TI = 0.167, TU = 0.389), its TU increases during the second period (TI = 0.833, TU = 1.749), and the patterns are most impactful with the highest level of TU during the third period (TI = 2.727, TU = 6.341).

Conclusion

This study proposed a machine learning approach to anticipating multi-technology convergence by using patent information. To this end, technology ecology networks were first generated, via ARM, using patent co-classification information. The technology ecology networks were compared between the periods to present implications on the changing patterns of multi-technology convergence, whereas supervised link prediction analysis based on logistic regression models was utilised to provide insight into the prospects of multi-technology convergence by identifying the links to be added to or removed from the network. The specific case of drug, bio-affecting, and body treating compositions technology confirmed that the proposed approach enables fine-grained identification of the changing patterns of multi-technology convergence.

The contributions of this study are threefold. From a theoretical perspective, the proposed approach contributes to technology mapping and forecasting research by extending the identification of structural patterns of convergence between a pair of technologies to the examination of dynamic aspects of multi-technology convergence. By integrating patent co-classification and citation information, the proposed approach measures the changing patterns of multi-technology convergence in terms of CS, CP, CI, TI, and TU. The quantitative indexes are found to be effective in measuring the changing patterns of technology convergence, and in anticipating future converging segments. The systematic processes and quantitative indexes developed in this study are expected to serve as a basis for technology convergence research, and to provide a better understanding of the technology dimensions of innovation ecosystems. From a methodological perspective, this study developed a machine learning approach to technology mapping and forecasting in convergence contexts using patent information. Specifically, the developed approach distinguishes the primary class from other patent classes to consider the direction of technology convergence trend, and modifies ARM to discover the significant structural patterns of multi-technology convergence based on the statistical dependence between patent classes. Moreover, we developed a link prediction method, using various factors that can be extracted from technology ecology networks, to anticipate technology convergence. Although this study focussed on examining the changing patterns of multi-technology convergence, the approach could be useful for many different purposes to be achieved by patent analysis, such as technology trend analysis and technology opportunity analysis. Finally, from a practical standpoint, a software system was developed to automate most parts of the proposed approach, enabling quick assessment of dynamic aspects of multi-technology convergence. The software system could prove useful as a complementary tool for supporting expert decision-making, and to allow even those unfamiliar with patent databases and computational models to measure the changing patterns of multi-technology convergence.

Despite the contributions, this study has certain limitations that should be complemented by future research. First, although this study is one of the earliest attempts to model multi-technology convergence, the proposed approach is limited to the analysis of convergence between a primary patent class and multiple supplementary patent classes. However, multiple classes can be involved in convergence processes as primary technologies. This should be elaborated further by developing an algorithm to integrate multiple individual patterns of convergence with identical supplementary, but different primary patent classes into a convergence pattern. Second, this study focusses on the three main types of changing patterns of multi-technology convergence. Other types of patterns and the factors that affect the evolution of innovation ecosystems should be investigated to guide organisations to respond in turn. Moreover, other items of patent data, such as assignees, can be employed to diversify the scope of analysis (e.g. *who* drives technology convergence). Third, many issues regarding how to improve the performance and efficiency of our method. For example, sequential pattern mining can be employed to measure the significant changing patterns of multi-technology convergence instead of comparing technology ecology networks between prescribed time periods. Support vector machines and neural networks can be used to consider the nonlinear classification boundary in the link prediction process. Moreover, the Apriori algorithm employed in this study is simple and historically significant, but not effective for massive data. More efficient models based on the frequent pattern growth (FP-growth) algorithm (Han et al. 2007) such as dynamic FP-growth (Gyorodi et al. 2003), FP-bonsai (Bonchi and Geothals 2004), and AFOPT (Liu et al. 2003) should be incorporated into our method. In addition, different resampling techniques (e.g. under- and over-sampling) and performance evaluation metrics should be considered to handle the imbalanced dataset, although this study used a fivefold stratified sampling technique as well as the F_1 score and Youden's J statistic. Fourth, the results and implications derived from the proposed approach may differ according to the prescribed threshold values for CP and CI. Moreover, when determining the threshold values, the field size should be considered in order to present field-specific implications, since different classes have different numbers of patents. Finally, this study considered a single case study of drug, bio-affecting, and body treating compositions technology. Further testing of different technologies with different types predictors and data generation techniques are essential to confirm the feasibility and utility of the proposed approach and to identify the best performing model.

Appendix 1: An example of constructing a technology ecology network

We present a step-by-step example of constructing a technology ecology network describing the significant structural patterns of multi-technology convergence. Five patents are used in this example, as shown in Table 8, and the threshold values for CP and CI are set to 0.2 and 0.5.

Table 8 Five patents used in this example

Basic information		Converging fields	
Patent number	Issued year	Primary	Supplementary
7037495	2006	424/130.1	424/184.1, 514/1
7264812	2007	424/184.1	530/855
7455836	2008	424/130.1	424/184.1
7387777	2008	424/130.1	424/184.1, 514/1, 530/350
7375075	2008	424/184.1	514/1, 530/350

First, we identify frequently co-occurred structural patterns based on the Apriori algorithm. Specifically, the frequent individual primary and supplementary classes (i.e. one-item set) that exceed the prescribed threshold values for CP (i.e. 0.2) are identified; the frequent patterns with one primary class and one supplementary class (i.e. two-item set) are identified from the one-item set; the two-item set is extended to the three-item set with one primary class and two supplementary classes by adding one supplementary class at a time. This extension process terminates when no further extensions are found. Table 9 shows the results of Apriori algorithm employed in this study, where the frequently co-occurred structural patterns derived from the five patents are highlighted in bold.

Next, we compute the values of CI and CS for the frequently co-occurred patterns, as shown in Table 10. Three patterns are identified as significant as their CI and CS values are greater than the threshold value for CI (i.e. 0.5) and one, respectively.

Table 9 Results of the Apriori algorithm employed in this study

One-item set (CP)	Two-item set (CP)	Three-item set (CP)
424/130.1$^{\text{Primary}}$ (0.6)	**424/130.1$^{\text{Primary}}$ → 424/184.1$^{\text{Supplementary}}$ (0.6)**	**424/130.1$^{\text{Primary}}$ → 424/184.1$^{\text{Supplementary}}$, 514/1$^{\text{Supplementary}}$ (0.4)**
424/184.1$^{\text{Primary}}$ (0.4)	**424/130.1$^{\text{Primary}}$ → 514/1$^{\text{Supplementary}}$ (0.4)**	
424/184.1$^{\text{Supplementary}}$ (0.6)	424/130.1$^{\text{Primary}}$ → 530/350$^{\text{Supplementary}}$ (0.2)	
514/1$^{\text{Supplementary}}$ (0.6)	424/184.1$^{\text{Primary}}$ → 424/184.1$^{\text{Supplementary}}$ (0)	
530/350$^{\text{Supplementary}}$ (0.4)	424/184.1$^{\text{Primary}}$ → 514/1$^{\text{Supplementary}}$ (0.2)	
530/855$^{\text{Supplementary}}$ (0.2)	424/184.1$^{\text{Primary}}$ → 530/350$^{\text{Supplementary}}$ (0.2)	

Table 10 List of the significant structural patterns

Frequently co-occurred patterns	CI	CS
424/130.1$^{\text{Primary}}$ → 424/184.1$^{\text{Supplementary}}$	$\frac{3/5}{3/5} = 1$	$\frac{1}{3/5} = \frac{5}{3}$
424/130.1$^{\text{Primary}}$ → 514/1$^{\text{Supplementary}}$	$\frac{2/5}{3/5} = \frac{2}{3}$	$\frac{2/3}{3/5} = \frac{10}{9}$
424/130.1$^{\text{Primary}}$ → 424/184.1$^{\text{Supplementary}}$, 514/1$^{\text{Supplementary}}$	$\frac{2/5}{3/5} = \frac{2}{3}$	$\frac{2/3}{2/5} = \frac{5}{3}$

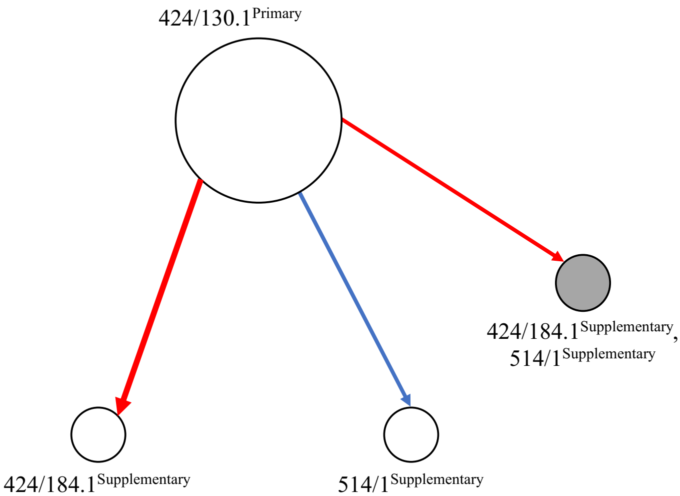

Fig. 5 The resulting technology ecology network

Finally, the significant structural patterns are represented as the technology ecology network. In Fig. 5, a source and a target node are a primary class and supplementary classes; the size of a node represents the number of the corresponding pattern's occurrence; a link represents the CI among the associated classes. White and grey nodes denote single and multiple patent classes; blue and red links indicate within-technology convergence where the primary patent class also appears in the supplementary patent class part and between-technology convergence where the primary patent class is not included in the patent class segment, respectively.

Appendix 2

See Table 11.

Table 11 The description of USPC mainline subclasses

USPC mainline subclass	Description
128/200.14	Liquid medicament atomizer or sprayer
128/200.24	Respiratory method or device
378/1	Specific application
422/50	Analyser, structured indicator, or manipulative laboratory device
424/1.11	Radionuclide or intended radionuclide containing; adjuvant or carrier compositions; intermediate or preparatory compositions’
424/130.1	Immunoglobulin, antiserum, antibody, or antibody fragment, except conjugate or complex of the same with nonimmunoglobulin material
424/178.1	Conjugate or complex of monoclonal or polyclonal antibody, immunoglobulin, or fragment thereof with nonimmunoglobulin material
424/184.1	Antigen, epitope or other immune-specific immune transfer factor
424/278.1	Nonspecific immunoeffector, per se (e.g., adjuvant, nonspecific immunostimulator, nonspecific immunopotentiator, nonspecific immunosuppressor, non- specific immunomodulator, etc.); or nonspecific immunoeffector, stabilizer, emulsifier, preservative, carrier, or other additive for a composition containing an immunoglobulin, an antiserum, an antibody, or fragment thereof, an antigen, an epitope, or other immunospecific immunoeffector
424/400	Preparations characterised by special physical form
424/43	Effervescent or pressurised fluid containing
424/49	Dentifrices (includes mouth wash)
424/520	Extract, body fluid, or cellular material of undetermined constitution derived from animal is active ingredient
424/59	Topical sun or radiation screening, or tanning preparations
424/600	Inorganic active ingredient containing
424/62	Bleach for live hair or skin (e.g., peroxides, etc.)
424/65	Anti-perspirants or perspiration deodorants
424/70.1	Live hair or scalp treating compositions (nontherapeutic)
424/725	Plant material or plant extract of undetermined constitution as active ingredient (e.g., herbal remedy, herbal extract, powder, oil, etc.)
424/78.02	Topical body preparation containing solid synthetic organic polymer as designated organic active ingredient (doai)
424/85.1	Lymphokine
424/9.1	In vivo diagnosis or in vivo testing
424/93.1	Whole live micro-organism, cell, or virus containing
424/94.1	Enzyme or coenzyme containing
427/2.1	Medical or dental purpose product; parts; subcombinations; intermediates (e.g., balloon catheter, splint)
435/183	Enzyme (e.g., ligases (6.), etc.), proenzyme; compositions thereof; process for preparing, activating, inhibiting, separating, or purifying enzymes
435/235.1	Virus or bacteriophage, except for viral vector or bacteriophage vector; composition thereof; preparation or purification thereof; production of viral subunits; media for propagating
435/243	Micro-organism, per se (e.g., protozoa, etc.); compositions thereof; process of propagating, maintaining or preserving micro-organisms or compositions thereof; process of preparing or isolating a composition containing a micro-organism; culture media therefor
435/320.1	Vector, per se (e.g., plasmid, hybrid plasmid, cosmid, viral vector, bacteriophage vector, etc.) Bacteriophage vector, etc.)
435/325	Animal cell, per se (e.g., cell lines, etc.)

Table 11 (continued)

USPC mainline subclass	Description
435/4	Measuring or testing process involving enzymes or micro-organisms; composition or test strip therefore; processes of forming such composition or test strip
435/41	Micro-organism, tissue cell culture or enzyme using process to synthesize a desired chemical compound or composition
436/501	Biospecific ligand binding assay
436/518	Involving an insoluble carrier for immobilising immunochemicals
510/108	Cleaning compositions or processes of preparing (e.g., sodium bisulfate component, etc.)
514/1	Designated organic active ingredient containing
530/300	Peptides of 3–100 amino acid residues
530/350	Proteins, i.e., more than 100 amino acid residues
600/300	Diagnostic testing
604/19	Means for introducing or removing material from body for therapeutic purposes (e.g., medicating, irrigating, aspirating, etc.)
623/1.1	Arterial prosthesis (i.e. blood vessel)

Acknowledgements This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korean government (MSIP) (No. 2019R1A6A3A13096839) and and the Sogang University Research Grant of 2020 (No. 202010009.01).

References

Adner, R. (2006). Match your innovation strategy to your innovation ecosystem. *Harvard Business Review, 84*(4), 98.

Adner, R., & Kapoor, R. (2010). Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations. *Strategic Management Journal, 31*(3), 306–333.

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *ACM SIGMOD Record, 22*(2), 207–216.

Aharonson, B. S., & Schilling, M. A. (2016). Mapping the technological landscape: Measuring technology distance, technological footprints, and technology evolution. *Research Policy, 45*(1), 81–96.

Athreye, S., & Keeble, D. (2000). Technological convergence, globalisation and ownership in the UK computer industry. *Technovation, 20*(5), 227–245.

Barberá-Tomás, D., Jiménez-Sáez, F., & Castelló-Molina, I. (2011). Mapping the importance of the real world: The validity of connectivity analysis of patent citations networks. *Research Policy, 40*(3), 473–486.

Bonchi, F., & Geothals, B. (2004). FP-Bonsai: The art of growing and pruning small fp-trees. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 155–160).

Caviggioli, F. (2016). Technology fusion: Identification and analysis of the drivers of technology convergence using patent data. *Technovation, 55,* 22–32.

Chaudhuri, S. (2005). *The WTO and India's pharmaceuticals industry: Patent protection, TRIPS, and developing countries*. Oxford: Oxford University Press.

Chen, Y. S., & Chang, K. C. (2010). The relationship between a firm's patent quality and its market value: The case of US pharmaceutical industry. *Technological Forecasting and Social Change, 77*(1), 20–33.

Choi, C., & Park, Y. (2009). Monitoring the organic structure of technology based on the patent development paths. *Technological Forecasting and Social Change, 76*(6), 754–768.

Curran, C. S., Bröring, S., & Leker, J. (2010). Anticipating converging industries using publicly available data. *Technological Forecasting and Social Change, 77*(3), 385–395.

Curran, C. S., & Leker, J. (2011). Patent indicators for monitoring convergence: Examples from NFF and ICT. *Technological Forecasting and Social Change, 78*(2), 256–273.

Dosi, G. (1984). *Technical change and industrial transformation: The theory and an application to the semiconductor industry*. Berlin: Springer.

Érdi, P., Makovi, K., Somogyvári, Z., Strandburg, K., Tobochnik, J., Volf, P., et al. (2013). Prediction of emerging technologies based on analysis of the US patent citation network. *Scientometrics, 95*(1), 225–242.

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management Science, 47*(1), 117–132.

Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: Evidence from patent data. *Research Policy, 30*(7), 1019–1039.

Geum, Y., Kim, C., Lee, S., & Kim, M. S. (2012). Technological convergence of IT and BT: Evidence from patent analysis. *ETRI Journal, 34*(3), 439–449.

Granstrand, O., & Holgersson, M. (2020). Innovation ecosystems: A conceptual review and a new definition. *Technovation, 90,* 102098.

Gyorodi, C., Gyorodi, R., Cofeey, T., & Holban, S. (2003). Mining association rules using dynamic FP-trees. In *Proceedings of irish signals and systems conference* (pp. 76–81).

Hacklin, F., Battistini, B., & Von Krogh, G. (2013). Strategic choices in converging industries. *MIT Sloan Management Review, 55*(1), 65.

Hacklin, F., Marxt, C., & Fahrni, F. (2009). Coevolutionary cycles of convergence: An extrapolation from the ICT industry. *Technological Forecasting and Social Change, 76*(6), 723–736.

Hacklin, F., Raurich, V., & Marxt, C. (2005). Implications of technological convergence on innovation trajectories: The case of ICT industry. *International Journal of Innovation and Technology Management, 2*(3), 313–330.

Han, J., Cheng, H., Xin, D., & Yan, X. (2007). Frequent pattern mining: Current status and future directions. *Data Mining Knowledge Discovery, 15*(1), 55–86.

Harhoff, D., Scherer, F. M., & Vopel, K. (2003). Citations, family size, opposition and the value of patent rights. *Research Policy, 32*(8), 1343–1363.

Jang, H. J., Woo, H. G., & Lee, C. (2017). Hawkes process-based technology impact analysis. *Journal of Informetrics, 11*(2), 511–529.

Jeong, S., Kim, J. C., & Choi, J. Y. (2015). Technology convergence: What developmental stage are we in? *Scientometrics, 104*(3), 841–871.

Jiang, Q., & Luan, C. (2018). Diffusion, convergence and influence of pharmaceutical innovations: A comparative study of Chinese and US patents. *Globalization and Health, 14*(1), 92.

Karki, M. M. S. (1997). Patent citation analysis: A policy analysis tool. *World Patent Information, 19*(4), 269–272.

Kim, E., Cho, Y., & Kim, W. (2014). Dynamic patterns of technological convergence in printed electronics technologies: Patent citation network. *Scientometrics, 98*(2), 975–998.

Kim, H., Hong, S., Kwon, O., & Lee, C. (2017). Concentric diversification based on technological capabilities: Link analysis of products and technologies. *Technological Forecasting and Social Change, 118,* 246–257.

Kim, J., Kim, S., & Lee, C. (2019). Anticipating technological convergence: Link prediction using Wikipedia hyperlinks. *Technovation, 79,* 25–34.

Kim, J., & Lee, S. (2017). Forecasting and identifying multi-technology convergence based on patent data: The case of IT and BT industries in 2020. *Scientometrics, 111*(1), 47–65.

Kim, N., Lee, H., Kim, W., Lee, H., & Suh, J. H. (2015). Dynamic patterns of industry convergence: Evidence from a large amount of unstructured data. *Research Policy, 44*(9), 1734–1748.

Kim, C., Lee, H., Seol, H., & Lee, C. (2011). Identifying core technologies based on technological cross-impacts: An association rule mining (ARM) and analytic network process (ANP) approach. *Expert Systems with Applications, 38*(10), 12559–12564.

Kwon, O., An, Y., Kim, M., & Lee, C. (2020). Anticipating technology-driven industry convergence: Evidence from large-scale patent analysis. *Technology Analysis and Strategic Management, 32*(4), 363–378.

Lee, C., Cho, Y., Seol, H., & Park, Y. (2012). A stochastic patent citation analysis approach to assessing future technological impacts. *Technological Forecasting and Social Change, 79*(1), 16–29.

Lee, W. S., Han, E. J., & Sohn, S. Y. (2015a). Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents. *Technological Forecasting and Social Change, 100,* 317–329.

Lee, C., Kang, B., & Shin, J. (2015b). Novelty-focused patent mapping for technology opportunity analysis. *Technological Forecasting and Social Change, 90,* 355–365.

Lee, C., Kwon, O., Kim, M., & Kwon, D. (2018). Early identification of emerging technologies: A machine learning approach using multiple patent indicators. *Technological Forecasting and Social Change, 127,* 291–303.

Lee, C., & Lee, G. (2019). Technology opportunity analysis based on recombinant search: Patent landscape analysis for idea generation. *Scientometrics, 121*(2), 603–632.

Liu, G., Lu, H., Yu, J. X., Wang, W., & Xiao, X. (2003). AFOPT: An efficient implementation of pattern growth approach. In *Proceedings of ICDM workshop on frequent itemset mining implementations*.

Narin, F., Noma, E., & Perry, R. (1987). Patents as indicators of corporate technological strength. *Research Policy, 16*(2–4), 143–155.

No, H. J., & Park, Y. (2010). Trajectory patterns of technology fusion: Trend analysis and taxonomical grouping in nanobiotechnology. *Technological Forecasting and Social Change, 77*(1), 63–75.

Oh, D. S., Phillips, F., Park, S., & Lee, E. (2016). Innovation ecosystems: A critical examination. *Technovation, 54,* 1–6.

Powers, D. M. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies, 2*(1), 37–63.

Shih, M. J., Liu, D. R., & Hsu, M. L. (2010). Discovering competitive intelligence by mining changes in patent trends. *Expert Systems with Applications, 37*(4), 2882–2890.

Shin, J., Coh, B. Y., & Lee, C. (2013). Robust future-oriented technology portfolios: Black–Litterman approach. *R&D Management, 43*(5), 409–419.

Xu, G., Wu, Y., Minshall, T., & Zhou, Y. (2018). Exploring innovation ecosystems across science, technology, and business: A case of 3D printing in China. *Technological Forecasting and Social Change, 136,* 208–221.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3*(1), 32–35.