Computers & Industrial Engineering 176 (2023) 108974

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

Combining machine learning with a pharmaceutical technology roadmap to analyze technological innovation opportunities

Lijie Feng [a,b], Weiyu Zhao [b,*], Jinfeng Wang [c,*], Jian Feng [c], Yanan Guo [d,e]

[a] Logistics Engineering College, Shanghai Maritime University, 1550 Haigang Avenue, Pudong District, Shanghai 201306, China
[b] Institute of Logistics Science and Engineering, Shanghai Maritime University, 1550 Haigang Avenue, Pudong District, Shanghai 201306, China
[c] China Institute of FTZ Supply Chain, Shanghai Maritime University, 1550 Haigang Avenue, Pudong District, Shanghai 201306, China
[d] School of life sciences, Shanghai University, No. 99 Shangda Road Baoshan District, Shanghai 200444, China
[e] School of Life Sciences, Zhengzhou University, No. 100 Science Avenue, Zhengzhou, Henan 450001, China

ARTICLE INFO

Keywords:
Pharmaceutical technology innovation
Pharmaceutical technology opportunity analysis
Technology roadmap (TRM)
Bidirectional Encoder Representation from Transformers (BERT)
Latent Dirichlet Allocation (LDA)

ABSTRACT

Technology roadmaps (TRM) with flexible architectural structures and development processes can effectively mitigate innovation dilemmas in pharmaceutical technology innovation opportunity analysis, such as long innovation cycles and inefficient translation. However, pharmaceutical technology innovation has become complicated and increasingly diverse. The current TRMs do not adapt well to this new reality. In response, this study proposes a systematic and specific framework to develop a pharmaceutical technology roadmap. Compared with standardized TRM, this TRM proposes three improvements. The first extension is the designing of the layers. The second extension is the selecting of the data source. The last extension is the processing of data sources. To validate the proposed framework, a case study in the field of hyperuricemia drugs was conducted. Our analysis focuses on the construction of a pharmaceutical TRM framework and a review of technology topics from multiple data sources that can supply quantitative information to explore the current trends. The analyzed results can assist the research and development professionals to predict technological opportunities and support experts to make more reasonable decision-making for a particular pharmaceutical domain.

1. Introduction

Pharmaceutical technological innovation is becoming increasingly important in promoting a transformation of scientific discoveries into new drugs and improved public health. Pharmaceutical technology innovation has the following features: high technology, high risk, high investment, long cycle, and high reward (Ye et al., 2012). As a result of the research and development (R&D) of a new bioengineered drug, the average success rate is only five to ten percent. It takes more than one to three billion U.S. dollars for the investment and a time span in excess of ten years (Liu et al., 2020). Once pharmaceutical technology innovation is successfully developed, it not only can have a high return on profits (more than 10 times), but it can also greatly improve people's health. Government and business enterprises make substantial investments in pharmaceutical technological innovation (Bowen & Casadevall, 2015). Although these expenditures keep increasing, the pace of translation of new pharmaceutical technologies into new drugs has been lagging behind their discovery (Cmwda et al., 2021). The government and pharmaceutical enterprises are facing increasing challenges to reduce unnecessary innovation costs. Accordingly, identifying how pharmaceutical innovation evolves, accurate analysis, and prediction of pharmaceutical technological opportunities have become critical factors in promoting technology innovation, such as narrowing down the research topics, minimizing risk (Duda et al., 2014), and avoiding unnecessary costs (Kim & Geum, 2021).

To analyze pharmaceutical technology opportunities and identify development paths by tracking the latest development trends of technology (Jia et al., 2018; Wang & Zhao, 2021; Zhou et al., 2019), researchers have proposed many methods. These include co-occurrence theory (Wu et al., 2014), bibliometric investigation (Raza et al., 2021), citation analysis (Du et al., 2019), net effect analysis (Zhou et al., 2019), technology roadmap (TRM) (Alejandro et al., 2022), Latent Dirichlet Allocation (LDA) (Wang et al., 2021), Hidden Markov Model (HMM) (Wu et al., 2014), Subject-action-object (SAO) (Zhou et al., 2019) and machine learning (ML) tools, which are helpful in pharmaceutical innovation. TRM can integrate other methods to analyze pharmaceutical

* Corresponding authors.
E-mail addresses: zhaoweiyu2022@qq.com (W. Zhao), wangjinfeng@shmtu.edu.cn (J. Wang).

https://doi.org/10.1016/j.cie.2022.108974
Received 3 February 2022; Accepted 29 December 2022
Available online 7 January 2023

technology opportunities based on time and layer series. Pharmaceutical technology become complicated and increasingly diverse. TRM may be more suitable for the features of pharmaceutical technological innovation such as those consisting of specific R&D stages, clear product generation, and various types of technologies (Jia et al., 2018).

Some efforts have been made to analyze pharmaceutical technological opportunities using TRM (Alejandro et al., 2022; Borschiver et al., 2019). However, there are several limitations of using pharmaceutical TRM from three aspects: layers design, data sources selection, and data processes for TRM. From the perspective of the designing of layers, previous standardized TRM have often classified technology layer into three layers: technology, product, and market layers (Kim & Geum, 2021). However, these previous works have neglected to develop a clinical trial layer. Clinical trials constitute an essential part of pharmaceutical innovation as well as technology, productization, and marketization (Zhao et al., 2019). To accommodate the pharmaceutical's domain-specific and case-specific needs, it is critical to alter the generic TRM's layers design flexibly and develop a clinical trial layer (Kim & Geum, 2021). Clinical trials can bridge the gap between technology and productization, and it is necessary to develop a clinical trial layer to accomplish this. From the perspective of data source selecting for pharmaceutical TRM, only a few studies combined patents and literature with commercial data. Patents, literature, and commercial data are embedded with valuable technical, clinical, product, and market information, which are recognized as separate critical data sources of TRM in the course of pharmaceutical technology analysis. Taking these three as data sources to analyze pharmaceutical technology trends can provide a more comprehensive insight from technology to market, especially for complex and diverse pharmaceutical technologies. From the standpoint of the data process for TRM, it is essential to combine ML with TRM to minimize the experts' judgments. Although experts are essential for analyzing pharmaceutical technology opportunities, excessive reliance on experts results in low replicability and objectivity. Approaches such as ML were introduced to support an objective and efficient TRM. ML tools such as topic modeling have been used efficiently for TRM in previous work (Hao et al., 2021). However, classifying data into different layers still relied on experts' judgments, resulting in a long and costly procedure (Wang, 2022).

In response, the main purpose of this article is to propose a systematic and domain-specific TRM for pharmaceutical technological opportunities analysis (TOA). Compared with standardization TRM, this TRM proposes three improvements. The first extension is the designing of the layers. This article constructs a comprehensive pharmaceutical TRM with six layers: technology, clinical trial, product, market, categories, and future direction. The second extension is the selecting of the data source. This article selects patents, paper, and commercial data as the data sources. The last extension is the processing of data sources. This paper proposes a novel approach to integrating Bidirectional Encoder Representation from Transformers (BERT) and LDA to support an objective and efficient ML-driven TRM. This means that the investigators can not only identify promising and hot technologies but can also narrow down one drug's future development direction comprehensively and effectively with the help of pharmaceutical TRM.

This paper is organized into the following parts: Section 2 provides a brief overview of the theoretical background that describes the previous literature regarding pharmaceutical TOA, pharmaceutical TRM, and ML. Section 3 describes the proposed approach, which contains a framework, and the overall process of constructing the pharmaceutical TRM and analyzing the pharmaceutical technology opportunities using this TRM. Section 4 chooses the hyperuricemia drug as an example to illustrate our approach. Section 5 discusses our findings and extension research. Section 6 concludes this paper with an outlook on possible future research and provides some limitations of our research.

2. Theoretical background

2.1. Pharmaceutical technological innovation opportunities analysis

Pharmaceutical technological innovation has promoted countless scientific discoveries, transforming them into clinical medicine, drugs, vaccines, and diagnostic reagents (Rake, 2017). It exists throughout the entire process of pharmaceutical technology R&D and plays an increasingly important role in providing efficient healthcare and constantly improving patients' health conditions. Identifying pharmaceutical technology opportunities can drive pharmaceutical technological innovation, improve the performance of pharmaceutical products, and improve product manufacturing processes, in addition to other benefits. Enterprises and investigators need finer insights into identifying and predicting hot spots, evolving paths, and analyzing future development direction (Jing et al., 2019). However, with the development of big data, many enterprises continue to have an incomplete understanding of how pharmaceutical technology evolves and the real pharmaceutical technological opportunities that can transform into a product.

Many researchers have been developing effective methods to analyze the path of how pharmaceutical innovation evolves and to identify and predict where pharmaceutical technological opportunities exist. Previous studies have focused on qualitative methods, such as Delphi, to gain domain experts' views and judgments on pharmaceutical technologies, which was time-consuming and labor-intensive and will also reduce the replicability of the prediction process. To overcome these shortcomings, Alan and Michael (1995) developed a bibliometric approach to analyze the emerging technological opportunities, which has also been widely used in the field of pharmaceutical technology (Bruno & Arns, 2019; Yeung et al., 2021). The citation relationship between different kinds of scientific papers, applied patents could provide an objective basis and quantitative data for identifying potential technological opportunities in the pharmaceutical field.

However, bibliometrics takes the relationship between citation and co-occurrence into account and it applies to the analysis of the contents of literature and patents only (Liu & Feng, 2022). With the development of big data, pharmaceutical TOA requires not only an extraction of the features of the documents but also an understanding of the meaning hidden behind the documents. Therefore, text mining is introduced (Leem et al., 2021). Coupled with publications, patents, and commercial data, Aaldering and Song (2019) used TRM combined with text mining tools to reveal the potential innovation pathways and commercial applications of solid lipid nanoparticle drugs. From the perspective of industrial technology, cultural diversity, product differentiation, and drug supervision, there have been studies that have used text mining tools to analyze the ways of promoting medical technology innovation (Allmendinger, 2021).

With the development of ML, some studies shifted their interests to develop automation and semi-automation approaches in pharmaceutical TOA. Zhengxin et al. (2022) explains the importance of ML in promoting the innovative conversion of bioenergy and biofuels in the form of a review. Considering medical technology's domain and topic characteristics, Kun et al. (2022) used ML and text clustering to fill the gaps in the emerging topics of medical technology. Wang et al. (2021) analyzes how drug innovation develops and tracks the evolutionary path of biomedical technology between papers and patents through topic evolution. Combining patents with clinical trials data, Houssein et al. (2021) introduced the EMR2vec platform to identify medical technology opportunities. In summary, approaches such as TRM, ML, and text mining are playing an increasingly important role in pharmaceutical TOA.

2.2. Pharmaceutical technology roadmap

TRM has been proved to be a critical technique for identifying

promising technology opportunities (Kerr & Phaal, 2020), promoting production innovation, and analyzing marketing (Jin et al., 2015). Compared with other tools, analyzing pharmaceutical technology opportunities based on TRM can better address the dynamic distribution of pharmaceutical technology topics in time series. Therefore, TRM has gradually been used in pharmaceutical TOA, such as creating the pharmaceutical technology landscape (Tierney et al., 2013), and analyzing the developmental trend and derivatives market of hyaluronic acid (Borschiver et al., 2019).

Published papers, approved patents, and commercial data have been three major data sources in innovation opportunities analysis of pharmaceutical technology based on TRM. Each drug innovation is accompanied, on average, by 19 journal publications and 23 patent applications (Sternitzke, 2010). Paper publication and patent application filing peak when pharmaceutical innovation is transformed successfully (Du et al., 2019). Throughout the entire process of pharmaceutical innovation, commercial reports are extremely critical (Choi & Hwang, 2014). No matter whether the pharmaceutical technological innovation was transformed successfully, the related patents, papers, and commercial data can provide inspiration. The existing TOA chooses patents (Hao et al., 2021), or chooses publications, patents, and commercial data as data sources (Zhou et al., 2019), rarely from the perspective of patents, literature, and commercial data. However, patents, literature, and commercial data are embedded with valuable technical, clinical, and product and market information.

The previous TRM has generally been constructed using the "market-product-technology" framework in the innovation opportunities analysis of pharmaceutical technology based on TRM (Zhou et al., 2019). However, pharmaceutical technology is different from other technologies in critical innovation process dimensions. Clinical trials are observations or experiments made in clinical research that help to determine the safety and efficacy of new pharmaceutical technology (Guo, 2022). Typically, about 37 percent of the total costs and over 70 percent of the total R&D time in drug innovation are spent on clinical trials (Zhao et al., 2019). Clinical trials are critical dimensions as well as technology, product, and market dimensions (Jin et al., 2017). To accommodate these domain-specific needs (Lee & Park, 2005), it is necessary to develop a clinical trial layer. Jia et al. (2018) have constructed a pharmaceutical TRM by using the "technology-clinical trial-product-market" structure. However, it has shown limited success at analyzing development trends from the perspective of the entire pharmaceutical industry using qualitative methods, such as lacking quantitative data. Therefore, it is necessary to construct an effective quantitative pharmaceutical TRM.

With the development of big data analysis, approaches such as bibliometrics (Kim & Geum, 2021), text mining (Kim & Geum, 2021), similarity measurement (Jeong & Yoon, 2015), ML (Hao et al., 2021) and link prediction (Kim & Geum, 2021) have been widely employed to support quantitative and intelligent TRM. Compared with the qualitative TRM which relied heavily on domain experts, such as Delphi (Kim & Geum, 2021), Analytic Hierarchy Process (Lee & Geum, 2017), and scenario-based (Mcdowall, 2012) approaches, quantitative and intelligent TRM can minimize the participation of experts. It can provide quantitative data to identify technological opportunities efficiently for both investigators and enterprises. It can also support experts to make more reasonable decision-making for a particular pharmaceutical domain.

2.3. Machine learning (ML)

Machine learning (ML) techniques have been increasingly used in supporting quantitative and intelligent TRM (Feng, 2022). Classification and topic clustering are two fundamental tasks in ML (Chen, Wu, et al., 2017). Some studies have used ML and text classification models in TRM layers classification (Sze et al., 2021). Others have applied ML and topic clustering models to identify topics, among other approaches (Kun et al., 2022; Zhengxin et al., 2022).

When classifying the pharmaceutical technological text, Liu et al. (2020) proposed a combinatorial *Self-Organizing Map–Kernel Principal Component Analysis–Support Vector Machine* (SOM-KPCA-SVM) ML model to classify the patent of the biomedicine industry. Liu, Zhou, et al. (2019) relied on the generative adversarial network (GAN) and deep neural network (DNN) to improve the multi-classification accuracy with small sample sizes in classifying cancer stages. As the task of pharmaceutical technological text classification becomes more clearly defined, more ML methods are adopted in TRM layers classification. Tingting et al. (2021) introduced the naïve Bayesian algorithm to group the SAO structures into different layers of TRM. However, previous studies were limited by factors such as small sample sizes, inefficient calculation of models, and high reliance on domain experts. Google proposed the Bidirectional Encoder Representation from Transformers (BERT) model to improve the classification accuracy of ML based on a small sample size in 2018. The BERT model has been proved to have excellent performance in NLP downstream tasks, such as text classification and text summarization (Guihua et al., 2020). Nevertheless, the pre-trained Bert model still needs to be improved in domain-specific classification tasks, such as pharmaceutical technology, which calls for pre-trained domain datasets (Kumar et al., 2021). The more accurate approach to using BERT is a fine-tuning approach in pre-trained domain-specific small datasets (Shah et al., 2021). To classify pharmaceutical data accurately into different layers, we created a small domain-specific training set and used BERT based on fine-tuning (Tan et al., 2021).

When identifying promising pharmaceutical topics, more unsupervised ML techniques were proposed, containing probabilistic-based topic modeling and text clustering (Zhang et al., 2016). Due to its advantages in extracting hot and potential technical topics from large volumes of textual data automatically (Chen, Zhang, et al., 2017), even when confronted with the domain-specific dataset (Lijun & Yafeng, 2017; Park & Kremer, 2017), Latent Dirichlet Allocation (LDA) has become one of the most widely used topic modeling techniques. While it has proven to be an effective technique in topic modeling, the LDA does not take time sequences into account, and it cannot capture the trend of technical topics over time (Wang et al., 2021). TRM is a time-based multilayered chart that, together with the LDA topic model, can track the trend of core topics over time (Yu & Zhang, 2019).

In light of this history, the main purpose of this article is to propose a systematic and domain-specific TRM for pharmaceutical TOA constructing a comprehensive pharmaceutical TRM with six layers, selecting patents, paper, and commercial data as the data source, and proposing a novel approach to integrating BERT and LDA to support an ML-driven TRM. From the analysis, we can examine how the topic evolves and gain insights into the future direction of the drugs.

3. Methodology

This study proposed a method for developing pharmaceutical TRM with six layers based on BERT and LDA analysis. The framework is an excellent guide to analyze pharmaceutical technology opportunities and their evolving paths. The proposed framework consists of three stages, as is illustrated in Fig. 1: (1) pharmaceutical technical data grouping, (2) pharmaceutical technical topics clustering and extracting, and (3) pharmaceutical technical topics evolution analysis.

3.1. Pharmaceutical technical data grouping based on BERT

3.1.1. Data retrieving and collecting

Data retrieving and collecting should be the starting phase of this study since our work involves technology, clinical trials, products, and marketing. We chose three different kinds of databases as the data sources: Medline (Kun et al., 2022), Derwent Innovations Index (DII) (Herrera-Vallejera & Gorbea-Portal, 2021), and Abstracts of Business Information (ABI) (Ferrero & Sison, 2014). In this study, the Medline

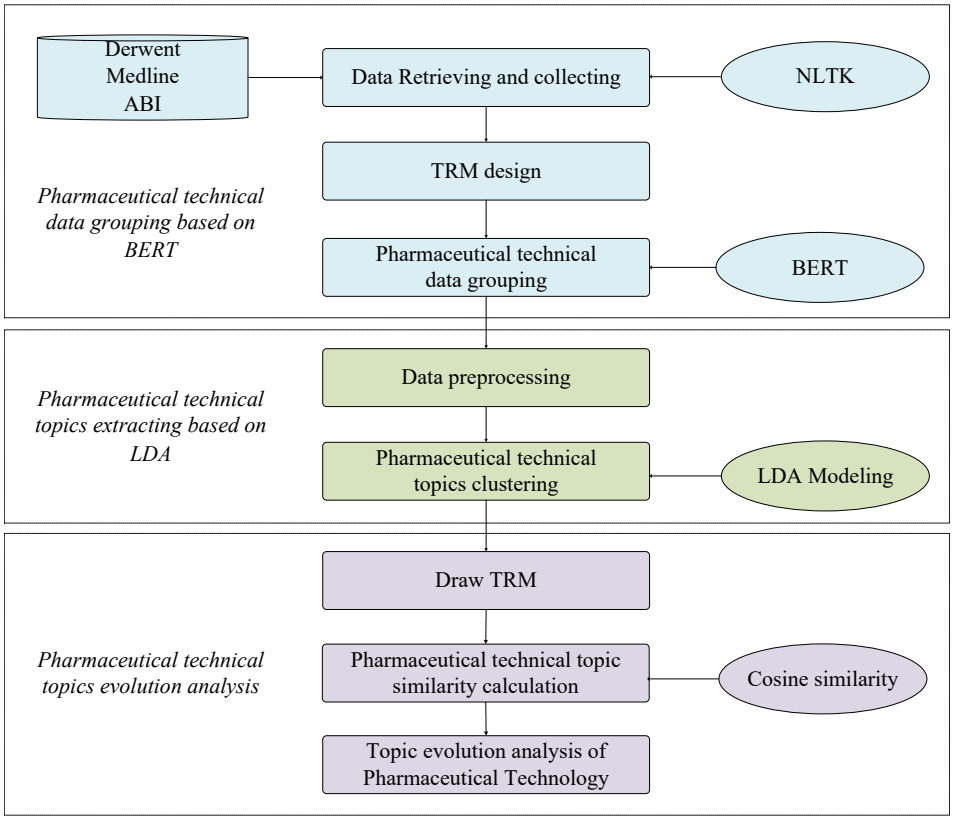

Fig. 1. Research framework.

paper was selected because it is the premier literature collection in the field of biomedicine, 70 %–80 % of which have English abstracts (Andrej & Dimitar, 2019). DII patents were selected because it is not only authoritative but also a comprehensive patent information database. It contains more than 96 percent of patents related to biomedicine, pharmaceutical, among others. ABI was selected because it includes more than 1,000 business magazines and financial newspapers worldwide, which are related Next, the literature data will be retrieved from MED using Mesh terms (Medical Subject Headings terms), patents data from DII using the International Patent Classification Number (IPC) search algorithm (Zhinan et al., 2019), and the commercial data were retrieved from ABI searching by the key terms.

3.1.2. TRM design

The first step is to generate the hierarchical structure of TRM. This study aims to analyze the future opportunities of pharmaceutical technology smoothly by constructing an intensive and in-detail domain-specific pharmaceutical TRM (Ali & Gittelman, 2016). Considering the critical role played by clinical trials, this study developed a clinical trial layer externally, and the vertical axis consists of six layers: Technology (L_1), Clinical Trial (L_2), Product (L_3), Market (L_4), Categories (L_5), and Future direction (L_6) layers (Tierney et al., 2013; Wang et al., 2015), each layer having its characteristics as shown in Table 1.

Next, the main objective was to divide the stage of TRM. This study chose the technology life cycle for the division of time stages. It can normally be classified into four periods: emerging, growth, maturity, and saturation period (Hao et al., 2021). After obtaining the data required, this study calculated the cumulative number of scientific papers, patents, and commercial data each year, respectively. The life cycle S-curve was drawn and the whole set of data will be split into S_M subsets based on the results above.

$$S_M \in \{S_1 \ldots S_M\}, M = 1, \ldots, 4 \quad (1)$$

Table 1
Layers of pharmaceutical TRM.

Layers	Categories	Description
L_1	Technology	Drug R&D capabilities, R&D achievements, R&D process, preclinical research, pathological research
L_2	Clinical Trial	Clinical trial phase I, Clinical trial phase II, Clinical trial phase III, Clinical trial phase IV
L_3	Product	Product layout, Product generation
L_4	Market	Sales, market share, and market competitiveness
L_5	Categories	Key technology categories
L_6	Future direction	The future direction of the technology

Where subset S_M represents this subset is in the Mth stage of TRM.

3.1.3. Pharmaceutical technical data grouping based on BERT

3.1.3.1. Pre-processing the collected data. The purpose of this stage is to clean and extract the related data feature of the scientific papers, patents, and commercial data obtained from 3.1.2. First, we will collect abstracts and published data of papers, patents, and commercial data in S_M groups. According to the features of pharmaceutical technology, we will remove the non-English, non-text, and redundant and noise data in the dataset. The cleaned papers, patents, and commercials data will be converted into text format compatible with text mining. Finally, the Natural Language Toolkit (NLTK) package will be imported to cut the extracted abstract into short sentences.

3.1.3.2. Pharmaceutical technical data grouping based on BERT. An appropriate BERT classifier will be selected for a short sentence. Where L_1 represents technology, L_2 represents the clinical trial, L_3 represents the product, and L_4 represents the market subset. L_0 represents the data that is irrelevant to L_1, L_2, L_3, and L_4 subsets. Based on the corpus trained

by Google, domain experts who have been devoted to pharmaceutical technological innovation and hyperuricemia drug innovation for more than 10 years will be invited. With the help of the domain experts, we will build a small training set. Representative sentences were selected to construct the domain-specific training set. Finally, the downstream tasks of the BERT model are pre-trained based on fine-tuning using the training set. Each subset is classified into five groups based on the pre-trained BERT model.

3.2. Pharmaceutical technical topics clustering and extracting

3.2.1. Pre-processing the data

The classified data in 3.1.3 are divided into $m \times n$ subsets, represented by G_{MN}.

$$G_{MN} \in \{G_{1\times1}, ...G_n, ..., G_{m\times1}, ..., G_{m\times n}\}, (m, n = 1, 2, ...) \quad (2)$$

Where G_{MN} represents this subset is in the Nth stage and the M th sublayers of the pharmaceutical TRM. For example, if the pharmaceutical TRM owns four different layers (technology, clinical trial, product, and marketing) and three different time stages (emerging, growth, maturity), the data collected will divide into 12 sub-datasets. According to the features of the pharmaceutical technology, preprocessed the $m \times n$ sub-datasets group by group, which involves word segmentation, verb reduction, manually combining similarity words, and deleting meaningless and extreme words based on term frequency–inverse document frequency (TF-IDF) (Kim et al., 2020). The most important step will be to build a domain stop words list with the help of the domain expert, which can be employed to reduce noise information and provide more refined data later (Liu, Liao, et al., 2019).

3.2.2. Pharmaceutical technical topics clustering and extracting based on LDA

Perplexity is one of the most common and reliable metrics to evaluate the quality of the LDA model (Lim et al., 2021). The performance of the LDA model is often excellent with the least perplexity. The optimal number of topics depends on the minimized perplexity. It is also necessary to define the hyperparameters α and β in the LDA model. LDA models have been proven to perform well and have good topic clustering results when $\alpha = 50/K$, and $\beta = 0.01$ by Griffiths, where K represents the number of topics (Schoggl et al., 2020; Wang et al., 2020). We pre-trained the LDA model based on the parameter values selected above and extracted potential pharmaceutical technology topics for each group. Each pharmaceutical technical topic is represented by a set of related words (Feng et al., 2021). Finally, the LDAvis package is employed to visualize the results of LDA topic models. Each topic is represented by a circle, where figures on the circles represent the topic numbers. The distance between circles reflects the discrimination between topics. The size of the circle reflects the strength of related pharmaceutical technology activities of topics.

3.3. Pharmaceutical technical topics evolution analysis

3.3.1. Pharmaceutical technical topic similarity calculation

After topic clustering, the cosine similarity is selected in this paper to calculate the similarity between topics, which evaluates the similarity of topics by calculating the cosine value between two topics (Frank et al., 2021). This method is based on a proposition that the smaller the cosine similarity between two topics, the greater their similarity (Frank et al., 2021). Based on the similarity calculation results above, it is critical to set an appropriate threshold value. With the help of this threshold, we can determine whether there will be a potential link between the two topics. As a result, we can identify how the pharmaceutical technology topics evolve in different stages and different layers in TRM, with the help of cosine similarity between various pharmaceutical technical topics.

3.3.2. Draw pharmaceutical TRM and analysis the pharmaceutical technology development trend

The first step is to visualize the LDAvis cycles in a two-dimensional TRM in time series and layer series. An example of the final visualization output is displayed in Fig. 4. The second step is to identify the potential link between two topics in the same sublayers or different sublayers. The cosine value of two topics that are greater than the threshold was regarded as the ones having potential connections, and they are illustrated by lines with different colors in TRM, where a different color means different categories.

Based on the outcomes above, this study describes the importance of pharmaceutical technical topics through their continuity and intensity (Hao et al., 2021). Technical continuity reflects the survival state of a topic, which means how many stages the pharmaceutical technology topics have gone through. Going through more stages means good performance in topic continuity. Technical strength refers to the innovation activity intensity and attention in different pharmaceutical technical topics, which is represented by the size of the circle. Then, according to the analysis of pharmaceutical technical continuity and strength, the technology topics are divided into four types: key topics, hot topics, breakthrough topics, and blank topics. Topics with high technical continuity and high technical strength are recognized as key topics; the ones with high technical continuity but poor technical strength are recognized as hot topics; the ones with high technical strength but poor technical continuity are breakthrough topics; the ones with both poor technical continuity and poor technical strength are blank topics. In this study, topics with continuity equal to or more than three stages are considered to have high continuity, and the ones with equal to or more than 10 % in LDAvis are considered to have high strength. The classification of pharmaceutical topics is shown in Table 2.

Next, by combining the identified pharmaceutical technological topics from layer L_1 to layer L_4 in pharmaceutical TRM with the classification results obtained for pharmaceutical topics, the investigators can identify the main categories in specific technology which is summarized in Layer L_5. Finally, through the dynamic analysis of the first five layers, the future development direction of the medical technology innovation is summarized and displayed in L_6. The final TRM will be presented above in Fig. 4.

4. Application in the hyperuricemia drug

Nowadays, with the change of lifestyles, the incidence rate of hyperuricemia has been increasing year by year. It is estimated that by 2030, the number of patients with hyperuricemia in the world will reach 1.42 billion (Chen-Xu et al., 2019). Meanwhile, drugs used to treat hyperuricemia have a long research and development cycle, and the side effects of existing drugs need to be improved urgently. Therefore, this study tracks the emergence and evolution of technology topics of hyperuricemia drugs, hoping to provide a reference for the development of hyperuricemia drugs in the future.

4.1. Pharmaceutical technical data grouping based on BERT

4.1.1. Data retrieving and collecting

This paper collects the academic papers from Medline, the patents from Derwent, and commercial data from ABI. When data collecting, this paper used the Mesh term 'MH = (Gout OR Hyperuricemia)' as the

Table 2
The classification of pharmaceutical topics.

Classification	Technical Continuity	Technical Strength
Key topics	High	High
Hot topics	High	Poor
Breakthrough topics	Poor	High
Blank topics	Poor	Poor

query to search the medical scientific papers from Medline, and 5445 papers were retrieved from the database from 2010 to 2020. The International Patent Classification Number 'IP = (A61P-019/06)' was used as the query to search the patent data from DII, and 4706 published patents were retrieved from the database from 2010 to 2020. Commercial data published from 2010 to 2020 were retrieved from ABI using 'hyperuricemia' as targeting query keywords. Data types are limited collected from company news, blogs, podcasts, and websites, newspapers, country reports, business cases, business plans, market reports, market research, annual reports, reports, government and official documents, news, commentaries, editorials, industry reports, reviews and interviews in English, and 3758 papers are collected in total. The data were downloaded on January 22, 2021.

A total of 85 percent of the academic data and 36 percent of the commercial data were found with abstracts. After consulting the domain experts, the quality of the data with abstracts was significantly better than the ones without abstracts. This paper kept the data with abstracts only: 4488 with academic data, 4706 with patents data, and 1335 with commercial data.

4.1.2. TRM design

The cumulative number of scientific papers, patents, and business news related to hyperuricemia drugs each year are shown in Fig. 2. In terms of the technology life cycle, the technology innovation activities of hyperuricemia drugs from 2010 to 2020 were divided into three periods: emerging (2010–2013), growth (2014–2018), and maturity (2019–2020). Next, we put the scientific papers, patents, and commercial data in one corpus. Based on the results of technology stage division, the whole corpus was split into three corpora and was represented by S_1 (emerging period), S_2 (growth period), and S_3 (maturity period).

4.1.3. Pharmaceutical technical data grouping based on BERT

4.1.3.1. Pre-processing the collected data. The tech-related data features were extracted, including published data and abstracts in S_1, S_2, and S_3. Next, the extracted abstracts of S_1, S_2, and S_3, were cut into short sentences with Python's NLTK package separately, removing duplicate sentences and leaving the relevant sentences without data noise. Consequently, the 2541 text data in S_1, 5915 text data in S_2, and 2073 text data in S_3 were converted into 16,745 short sentences in S_1, 42,905 short sentences in S_2, and 17,146 short sentences in S_3.

4.1.3.2. Pharmaceutical technical data grouping based on BERT. Based on the general corpus trained by Google, this study invited three domain experts who have focused on pharmaceutical technological innovation and hyperuricemia drug development research for more than 10 years. The experts jointly constructed a small training set. The classification description of hyperuricemia drugs was refined in detail, and 5068 representative short sentences were selected from 76,793 short sentences data. Next, the training set was manually and efficiently annotated by the domain experts. In hopes of improving downstream performance, the BERT model was used based on fine-tuning. With the help of the domain training set, we pre-trained the BERT model, and each subset (S_1, S_2, S_3) was classified into five groups, represented by L_1, L_2, L_3, L_4, L_0. We keep the categories L_1, L_2, L_3, and L_4 only per stage.

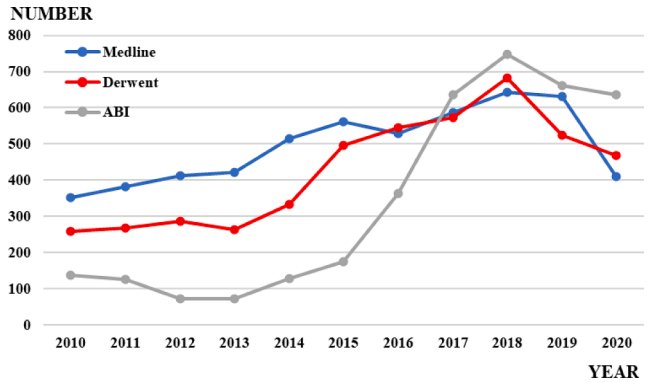

Fig.2. Statistical results of papers, patents, and commercials data related to hyperuricemia drugs.

After classification, 3000 sentences were selected as validation sets from the dataset randomly. The validation set is manually annotated by the same group of experts who constructed the training set above. It can be proven that the BERT model has high accuracy and fine granularity, and the classification accuracy is more than 90 %. Finally, the 76,793 short sentences were divided into 12 subsets, which are presented in Table 3. For instance, G_{43} represents that the group of data is located in the fourth sublayer of the pharmaceutical TRM (market layer), and in the third time stage (2019–2020).

4.2. Pharmaceutical technical topics clustering and extracting

4.2.1. Pre-processing data

This study preprocessed the 12 sub-datasets group by group (Kim et al., 2020). When data were preprocessed, we excluded numbers and basic verbs based on 972 commonly used stop-words, stemming, and lemmatization. Domain-specific words such as "hyperuricemia" and "therapeutic", which are meaningless in analyzing the hyperuricemia drug trend, were designated as domain-stop-words and removed from the analysis. An extra 2188 domain-specific stop-words are added based on 927 commonly used English stop-words. According to the domain-specific stop-words, we employed TF-IDF to exclude words that were either too common or too rare.

4.2.2. Cluster and extract the technical topics of hyperuricemia drugs based on LDA

This study imported the Sklearn package in Python to pre-train the LDA model, with the hyperparameters setting $\alpha = 50/K$, and $\beta = 0.01$(Wang et al., 2020). The number of topics in each subset is defined by perplexity. Despite that a smaller perplexity value is more accurate as a result of topic clustering, the number of topics will be too large to reduce dimensionality. Therefore, the optimal number of topics for each subset is limited from 2 to 50. Next, we calculated the topic perplexity subset by subset. The topic perplexity value and the curve are shown in and Fig. 3. Where the vertical axis in Fig. 3 is made up of four sublayers (technology(L_1), clinical trial(L_2), product (L_3), and market sublayers (L_4)), and its horizontal axis can be used to show time stages such as 2010–2013 (S_1), 2014–2018 (S_2), and 2019–2020 (S_3).

After calculation, it is easy to see that when the number of topics is 5, 3, 4, 2, 2, 2, 10, 5, 15, 8, 3, and 13, the clustering result of each sub-dataset G_{11}, G_{12}, G_{13}, G_{21}, G_{22}, G_{23}, G_{31}, G_{32}, G_{33}, G_{41}, G_{42}, G_{43} based on LDA is the best. The outcomes of the LDA model for each subset are shown in Tables 4–7. Each topic is described using thirty related words, only 5 main contributing keywords are listed below. At last, this study uses LDAvis to visualize the result of topic modeling in 12 groups. The visualization result is shown in Fig. 4.

Table 3
The group result of different subsets.

layers	Time Stages		
	Emerging stage 2010–2013(S_1)	Rising stage 2014–2018(S_2)	Maturing stage 2019–2020(S_3)
Technology (L_1)	$Group_{11}(G_{11})$	$Group_{12}(G_{12})$	$Group_{13}(G_{13})$
Clinical Trial (L_2)	$Group_{21}(G_{21})$	$Group_{22}(G_{22})$	$Group_{23}(G_{23})$
Product (L_3)	$Group_{31}(G_{31})$	$Group_{32}(G_{32})$	$Group_{33}(G_{33})$
Market (L_4)	$Group_{41}(G_{41})$	$Group_{42}(G_{42})$	$Group_{43}(G_{43})$

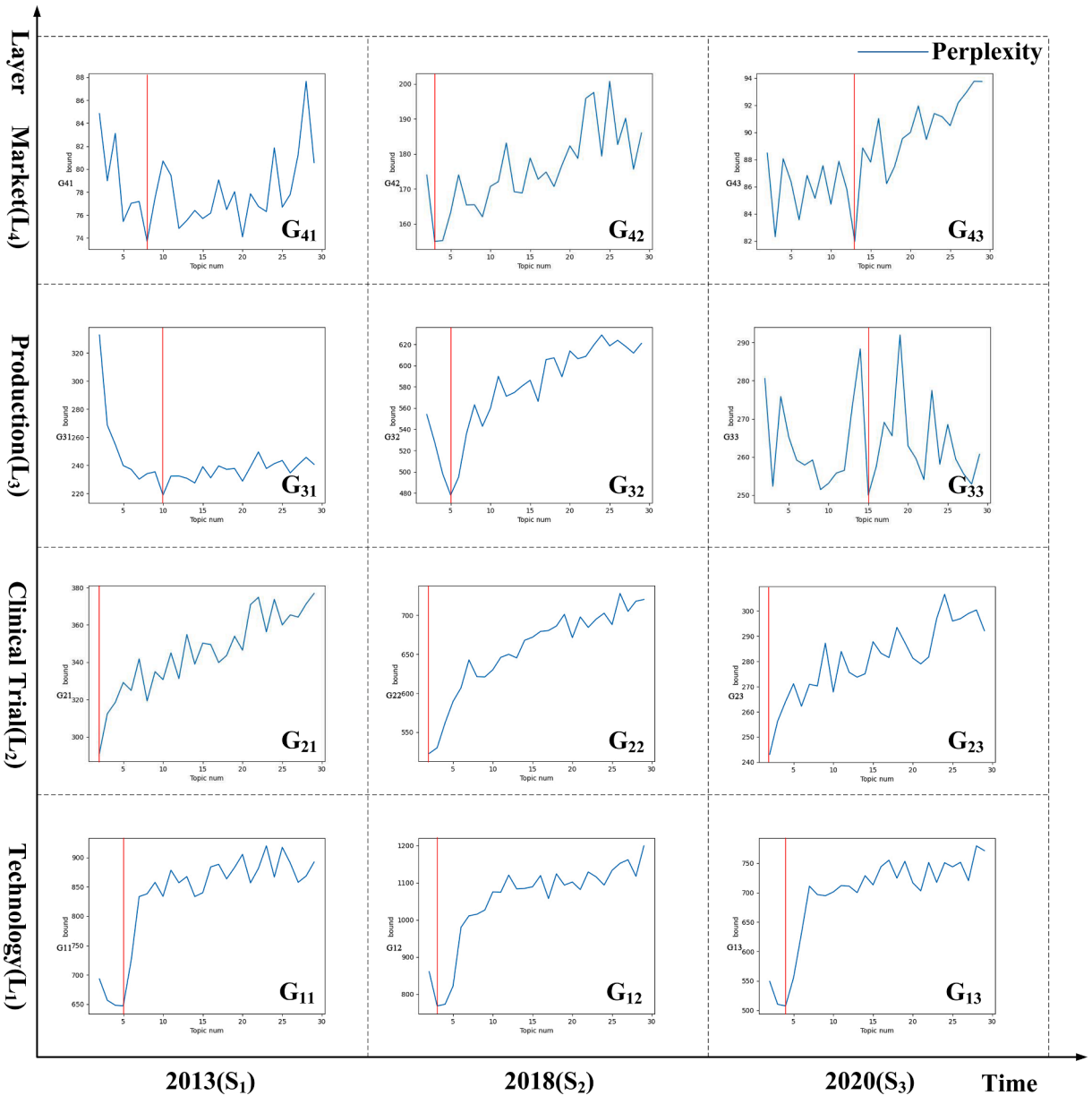

Fig. 3. Topic perplexity curve.

4.3. Analyze the topic evolution of hyperuricemia drugs

4.3.1. Calculate the topic similarity of hyperuricemia drugs

After obtaining different visualization results for each subset, this study chooses cosine similarity to calculate the similarity between topics in different sublayers and different stages. The calculation result among topics of hyperuricemia drugs in the same sublayers but different stages is illustrated in Tables 8–11. The calculation result among topics of hyperuricemia drugs in different sublayers is illustrated in Tables 12–14.

4.3.2. Draw TRM and analysis of the pharmaceutical technology development trend

The first step is to organize the visualization result of LDAvis in TRM. It is critical to organize the visualization result by LDAvis from L_1, L_2, L_3, to L_4 layers of the pharmaceutical TRM and organize them in time series from S_1, S_2, to S_3 in Fig. 4.

The second step is to identify the potential link between two topics in the same sublayers. Based on the results in Tables 8–11, the cosine similarity threshold is set from 0.5 to 1. If the similarity value between two topics is equal to or greater than 0.5, there will be a connection between two topics in the same layers but in different stages. The topics that have connections will be connected by arrows of different colors. For example, in the technology sublayer, since the cosine similarity between the topics G_{13}–T_1 and G_{12}–T_1 is 0.5741, G_{12}–T_1 and G_{11}–T_5 is 0.5774, both of which are greater than 0.5 and they are all in the same sublayers in TRM, so G_{13}–T_1 evolved from topics G_{12}–T_1, and G_{11}–T_5. These topics are all about protein drugs, and they are linked by purple arrows in Fig. 4. Similarly, G_{13}–T_2 and G_{13}–T_4 are all evolved from topics G_{12}–T_3 and G_{11}–T_2. These topics are all about small molecule drugs, so they are linked by green arrows in Fig. 4. G_{13}–T_3 is evolved from topics G_{12}–T_2 and G_{11}–T_4. These topics are all about Chinese medicine, and they are linked by orange arrows. If they do not belong to the above three categories, the links between topics are drawn with brown-yellow arrows.

The potential links between topics in different sublayers are identified similarly. From Tables 12–14, if the similarity value between two topics in different layers is equal to or greater than 0.35, there will be a connection between the two topics in different layers. The topics that

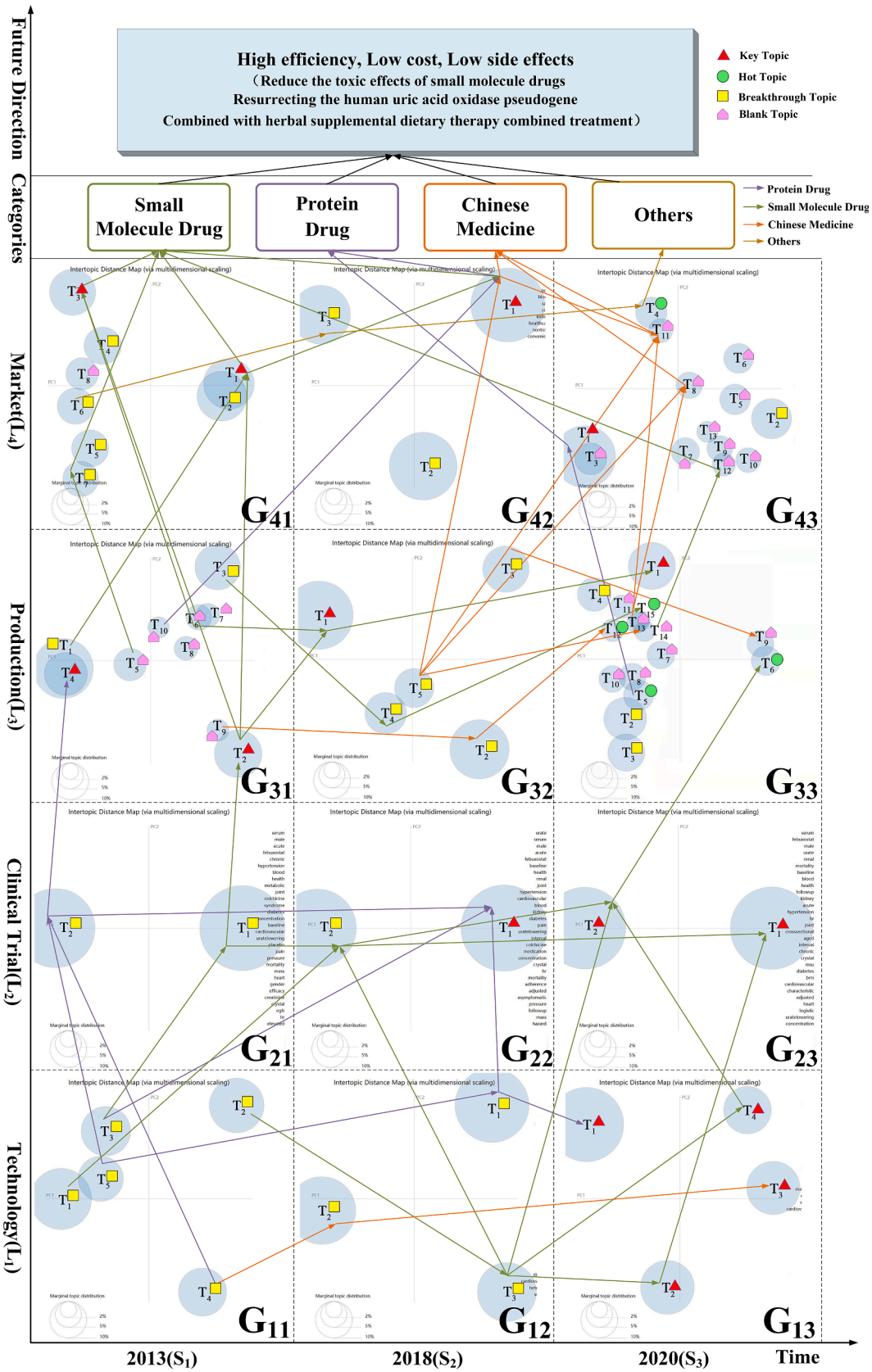

Fig. 4. The TRM with six layers.

Table 4
Related meaning and representative words of the technology in three stages (G_{11} to G_{13}).

Group	Number	Topic meanings	Main contributing keywords
G_{11}	G_{11}-T_1	Animal models in research of Topiroxostat	Serum, alkyl, febuxostat, blood, concentration…
	G_{11}-T_2	Research on complications caused by taking small molecule drugs	Syndrome, disorder, chronic, cancer, renal…
	G_{11}-T_3	Preclinical studies	Urate, agent, crystal, joint, carrier…
	G_{11}-T_4	Research on active ingredients of Chinese medicine and protein drugs	Compound, extract, powder, water, salt…
	G_{11}-T_5	Research on active ingredients of protein drugs	Protein, amino, ring, renal, halo…
G_{12}	G_{12}-T_1	Research on engineering human urate Oxidase	Alkyl, compound, crystal, urate, halo…
	G_{12}-T_2	Research on the natural active of Chinese medicine and small molecules	Water, extract, powder, raw, ethanol…
	G_{12}-T_3	Pre-clinical studies of Lesinurad, and Arhalofenate	Serum, renal, kidney, cancer, chronic…
G_{13}	G_{13}-T_1	Research on uric acid Oxidase from Aspergillus flavus	Compound, alkyl, mouse, protein, gene…
	G_{13}-T_2	Hyperuricemia small molecule drugs market	Kidney, renal, urate, chronic, cardiovascular…
	G_{13}-T_3	Research on extraction method of Chinese medicine	Extract, water, powder, raw, liquid…
	G_{13}-T_4	Small molecule drug complications study	Serum, blood, metabolic, disorder, potential…

Table 5
Related meaning and representative words of the clinical trial in three stages (G_{21} to G_{23}).

Group	Number	Topic meanings	Main contributing keywords
G_{21}	G_{21}-T_1	Clinical studies of Topiroxostat	Urate, acute, health, joint, chronic…
	G_{21}-T_2	Post-marketing safety surveillance of Pegliticase	Serum, male, febuxostat, baseline, renal…
G_{22}	G_{22}-T_1	Real-world utilization and outcomes of Pegliticase	Urate, acute, health, joint, chronic…
	G_{22}-T_2	Clinical trials of gloprba, real-world utilization and outcomes of Colchicine, and Febrestat	Serum, male, febuxostat, baseline, renal…
G_{23}	G_{23}-T_1	Clinical trials of Ahalofenate	Febuxostat, male, renal, cardiovascular, mortality…
	G_{23}-T_2	Clinical trials of Dotinarad	Serum, urate, blood, acute, joint…

have connections will be linked by purple, green, orange, and brown-yellow arrows, which represent the topics' links between protein drugs, small molecule drugs, Chinese medicine, and other categories. For example, since the cosine similarity between the topics G_{11}–T_3 and G_{21}–T_1 is 0.4089, G_{21}–T_1 and G_{31}–T_2 is 0.4045, G_{31}–T_2 and G_{41}–T_1 is 0.4225, all of which are greater than 0.35 and they are all in different layers in TRM. Therefore, G_{41}–T_1 evolved from topics G_{31}–T_2, G_{21}–T_1, and G_{11}–T_3. These topics are all about small molecule drugs, so they are linked by green arrows. Next, based on the analysis above, the potential link between the remaining topics was marked from L_1 to L_4 sublayers.

We then analyzed the importance of 72 topics based on technology continuity and technology strength proposed. As illustrated in Fig. 4, the key topics in the hyperuricemia drugs include G_{13}–T_1, G_{13}–T_2, G_{13}–T_3, G_{13}–T_4, G_{22}–T_1, G_{23}–T_1, G_{23}–T_2, G_{31}–T_1, G_{31}–T_2, G_{31}–T_4, G_{32}–T_1, G_{33}–T_1, G_{41}–T_1, G_{41}–T_3, G_{42}–T_1, and G_{43}–T_1. The 16 topics that were equal to or greater than 10 % are considered to have high continuity and high strength, as they have experienced three or more stages in evolution. Therefore, the above 16 topics with high technology continuity and

Table 6
Related meaning and representative words of the product in three stages (G_{31} to G_{33}).

Group	Number	Topic meanings	Main contributing keywords
G_{31}	G_{31}-T_1	Small molecule pharmaceutical excipients 1	Compound, salt, amino, febuxostat, crystal…
	G_{31}-T_2	Metabolic syndrome caused by taking small molecule drugs	Blood, disorder, diabetes, syndrome, metabolic…
	G_{31}-T_3	Topiroxostat, Sulfinpyrazone, and other small molecule drugs	Alkyl, alkoxy, halo, cycloalkyl, ring…
	G_{31}-T_4	Combination of multiple small molecules and protein drugs	Acute, serum, oxidase, colchicine, xanthine…
	G_{31}-T_5	Small molecule pharmaceutical excipients 2	Agent, active, granule, liquid, ingredient…
	G_{31}-T_6	Hyperuricemia and possible complications	Extract, protein, antibody, binding, polypeptide…
	G_{31}-T_7	Polyethylene glycol recombinant uricase	Cancer, syndrome, chronic, disorder, acute…
	G_{31}-T_8	Chinese medicine-food therapy combination, relieve drug toxicity	Powder, drink, potential, leaf, tea…
	G_{31}-T_9	Chinese medicine treatment-removing dampness and promoting diuresis	Radix, angelica, rhizome, seed, herba…
	G_{31}-T_{10}	Adverse protein drug reactions	Pain, skin, inflammation, pill, heat…
G_{32}	G_{32}-T_1	Metabolic syndrome caused by taking small molecule drugs	Cancer, disorder, syndrome, chronic, kidney…
	G_{32}-T_2	Chinese medicine treatment-antipyretic medicine	Radix, angelica, rhizome, seed, stem…
	G_{32}-T_3	Chinese medicine-food therapy combination	Extract, powder, pain, leaf, mass…
	G_{32}-T_4	Zurampic, Duzallo, and other small molecule drugs	Alkyl, derivative, halo, cycloalkyl, agent…
	G_{32}-T_5	Chinese medicine treatment-removing dampness and promoting diuresis	Blood, herb, acute, pill, granule…
G_{33}	G_{33}-T_1	Complications expected after taking small molecule drugs	Syndrome, disorder, chronic, diabetes, tumo…
	G_{33}-T_2	Technical protein drugs	Compound, agent, salt, active, ingredient…
	G_{33}-T_3	Chinese medicine – mixed medication	Rhizome, bark, astragalus, licorice, yam…
	G_{33}-T_4	Chinese antipyretic medicine treatment-food therapy combination	Leaf, mulberry, tuckahoe, kudzu, antibody…
	G_{33}-T_5	Hepatotoxicity of protein drugs	Kidney, liver, inflammation, bone, medicament…
	G_{33}-T_6	Small molecules drugs: Benzbromarone	Oil, acute, urate, heteroatoms, serum, renal…
	G_{33}-T_7	Small molecules drugs: Topiroxostat	Beverage, sodium, white, chicory, solid…
	G_{33}-T_8	Small molecules drugs: Colchicine	Seed, pain, glabrate, smilacis, joint…
	G_{33}-T_9	Chinese medicine combined with food extract is used in diet therapy	Cancer, extract, powder, lowering, kinase…
	G_{33}-T_{10}	Chinese medicine active ingredient extraction	Derivative, wine, mass, antigout, cold…
	G_{33}-T_{11}	Small molecules drugs: Dotinurad	Drink, regulating, flower, balsam, garden…
	G_{33}-T_{12}	Chinese medicine treatment-promoting blood circulation and removing blood stasis	Radix, herba, angelica, stem, safflower…
	G_{33}-T_{13}	Chinese medicine treatment-removing dampness and promoting diuresis	Blood, tea, healthcare, pain, ginseng…
	G_{33}-T_{14}		

(*continued on next page*)

Table 6 (*continued*)

Group	Number	Topic meanings	Main contributing keywords
		Chinese medicine-Food therapy combination	Health, granule, febuxostat, corn, sunflower…
	G_{33}-T_{15}	Small molecules drugs: Gloperba	Alkyl, alkoxy, halo, cycloalkyl, heteroaryl…

Table 7
Related meaning and representative words of the market in three stages (G_{41} to G_{43}).

Group	Number	Topic meanings	Main contributing keywords
G_{41}	G_{41}-T_1	Acute hyperuricemia symptom relief	Blood, pain, stasis, heat, kidney…
	G_{41}-T_2	Improve the hyperuricemia patients' quality of life	Compound, currency, dollar, cost, device…
	G_{41}-T_3	Drugs for the treatment of chronic uric acid	Syndrome, disorder, diabetes, chronic, acute…
	G_{41}-T_4	Adverse renal effects of small molecule drugs	Convenient, financial, vietnam, fast, sale…
	G_{41}-T_5	Adverse hepatic effects of small molecule drugs	Improved, bioavailability, Japan, solubility, efficacy…
	G_{41}-T_6	Drug development cooperation	Healthcare, tenjin, acquisition, rating, subsidiary…
	G_{41}-T_7	Cost control of small molecule drug treatment	Safe, agent, stable, cost-effective, quality…
	G_{41}-T_8	Japan's future development of Xanthine and Urate Oxidase	Nontoxic, serum, growth, quick, health…
G_{42}	G_{42}-T_1	Drugs used to relieve symptoms of acute hyperuricemia	Pain, blood, kidney, joint, wine, bone…
	G_{42}-T_2	Demand development direction: high bioavailability, cost-effectiveness, safety.	Safe, cost, nontoxic, convenient, quick…
	G_{42}-T_3	Co-development, financing, and strong alliance	Healthcare, partnership, amp, acquisition, financing…
G_{43}	G_{43}-T_1	Market development direction: reduction of liver damage	Kidney, recurrence, prevents, liquor, natural…
	G_{43}-T_2	Product R & D demand affected by economic impact and cost	Swot, financial, compound, wine, bioavailability…
	G_{43}-T_3	Chinese herbal medicine for hyperuricemia	Powder, efficacy, foot, pill, warming…
	G_{43}-T_4	Looking for partners across companies and transnational	Healthcare, amp, alliance, comprehensive, acquisition…
	G_{43}-T_5	Japanese Kampo medicines	Tenjin, pharma, nippon, chemiphar, trademark…
	G_{43}-T_6	Market demand direction: low toxicity:	Convenient, process, toxicity, taste, disorder…
	G_{43}-T_7	Market demand: Reduce spleen toxicity	Safe, avoid, liver, kidney, nontoxic…
	G_{43}-T_8	Chinese medicine alleviates side effects	Blood, promotes, circulation, absorption, lowering…
	G_{43}-T_9	Market demand: Reduce liver toxicity	Cancer, skin, price, gel, property…
	G_{43}-T_{10}	Market demand for Chinese herbal compound drugs	Cost, raw, agent, wide, quick…
	G_{43}-T_{11}	Acute hyperuricemia needs to increase the immunity by Chinese medicine	Pain, blood, heat, swelling, stasis…
	G_{43}-T_{12}	Market demand for protein drugs with low immunogenicity	Health, anti-inflammatory, analgesic, metabolic, inflammation…
	G_{43}-T_{13}	Global hyperuricemia drug market risk	Japan, subsidiary, ointment, allena, extract…

high technology strength were recognized as key topics, which are marked with red triangles. Similarly, there were five hot topics in total: G_{33}–T_5, G_{33}–T_6, G_{33}–T_{12}, G_{33}–T_{15}, and G_{43}–T_4 with high technology continuity but low technology strength, which are marked with green circles. There were twenty-five breakthrough topics in total: G_{11}–T_1, G_{11}–T_2, G_{11}–T_3, G_{11}–T_4, G_{11}–T_5, G_{12}–T_1, G_{12}–T_2, G_{12}–T_3, G_{21}–T_1, G_{21}–T_2, G_{22}–T_2, G_{31}–T_3, G_{32}–T_2, G_{32}–T_3, G_{32}–T_4, G_{32}–T_5, G_{33}–T_2, G_{33}–T_3, G_{33}–T_4, G_{41}–T_2, G_{41}–T_4, G_{41}–T_5, G_{41}–T_6, G_{41}–T_7, and G_{43}–T_2, with low technology continuity but high technical strength, which are marked with yellow quadrilaterals. There are still 26 topics with low technology continuity but high technical strength, which are blank topics that are represented by rose-red pentagons.

With the help of the domain experts, we summarized the key categories of hyperuricemia drugs: Chinese medicines, small molecule drugs, and protein drugs. Next, we connected the key, hot, breakthrough, and blank topics in market sublayers with proper categories, where orange arrows represent Chinese medicines, green arrows represent small molecule drugs, purple arrows represent protein drugs, and brown-yellow arrows represent other categories in L_5.

According to the analysis, small molecule drugs were the core category concerned by investigators in the years from 2010 to 2020. There are many small-molecule topics in key topics, hot topics, and breakthrough topics. From 2010 to 2013, the topic has been focused on the research and development of drugs targeting xanthine oxidase or URAT1. However, these small molecule drugs have great toxicity, side effects, and adverse reactions. They should be used with caution in clinical practice, and liver damage has been reported on the market. From 2014 to 2018, it was found that the combined use of small molecules was better than the single-use in reducing blood uric acid levels. However, small-molecule chemical drugs have a certain degree of side effects, and none of them can dissolve the uric acid stones that have been deposited in the body. From 2019 to 2020, with the targeting of xanthine oxidase or URAT1, some companies and R&D institutions are still attempting to develop related small molecule drugs and striving to reduce side effects such as liver and kidney injuries caused by existing small molecule drugs. Learning how to reduce the side effects of small molecule drugs will be the future direction for small molecule drug research and development. Similarly, determining how to eliminate or reduce the immunogenicity of existing urate oxidase and how to obtain active and non-immunogenic human urate oxidase drugs will be the future direction for protein drug development. Finally, learning how to combine dietary formulas, Chinese medicine prescriptions with small molecule and protein drugs to treat hyperuricemia is the future direction for Chinese medicine research and development.

The pharmaceutical TMR with six layers is an important tool for analyzing pharmaceutical technology development trends. The analysis of the pharmaceutical TRM demonstrates that reducing the toxicity of small molecule drugs, resurrecting the pseudogene of human uric acid oxidase, and adopting combination therapy with Chinese medicines supplementary dietary therapy will be the future direction for hyperuricemia drugs, which is illustrated in L6. It highlights that the R&D direction of drugs for the treatment of hyperuricemia will gradually shift from therapeutic and rehabilitative drugs to preventive and detection drugs, reflecting the transformation process of pharmaceutical product innovation to the "4P" trend of predictability, prevention, personalization, and participation in medical treatment (Krishna N, Durga et al., 2018; Denicolai, Previtali et al., 2020).

5. Discussion

Prior works have proposed many methods to analyze pharmaceutical opportunities, for example, qualitative methods, bibliometric approach, text mining, TRM, among others. The TRM can analyze the pharmaceutical opportunities better than other tools in both phase and time series. Some efforts have been made to analyze pharmaceutical opportunities using TRM. However, these studies have some limitations from

Table 8
Topic Similarity Calculation Result of technology among three stages (G_{11} to G_{12}, G_{12} to G_{13}).

Topic	G_{11}-T_1	G_{11}-T_2	G_{11}-T_3	G_{11}-T_4	G_{11}-T_5	G_{13}-T_1	G_{13}-T_2	G_{13}-T_3	G_{13}-T_4
G_{12}-T_1	0.0543	0.0664	0.4089	0.2571	0.5774	0.5741	0.3017	0.0154	0.1285
G_{12}-T_2	0.0852	0.0200	0.0777	0.6925	0.0131	0.0822	0.0419	0.9376	0.0349
G_{12}-T_3	0.4655	0.6907	0.1899	0.0945	0.1870	0.2054	0.6258	0.0241	0.5622

Table 9
Topic Similarity Calculation Result of clinical trial among three stages (G_{21} to G_{22}, G_{22} to G_{23}).

Topic	G_{21}-T_1	G_{21}-T_2	G_{23}-T_1	G_{23}-T_2
G_{22}-T_1	0.1579	0.7587	0.2369	0.4968
G_{22}-T_2	0.8998	0.2384	0.6661	0.5579

Table 10
Topic Similarity Calculation Result of product among three stages (G_{31} to G_{32}, G_{32} to G_{33}).

Topic	G_{32}-T_1	G_{32}-T_2	G_{32}-T_3	G_{32}-T_4	G_{32}-T_5
G_{31}-T_1	0.0293	0.0036	0.0482	0.3334	0.0025
G_{31}-T_2	0.5201	0.0021	0.0279	0.0032	0.3562
G_{31}-T_3	0.0019	0.0011	0.0017	0.8331	0.0016
G_{31}-T_4	0.1233	0.0048	0.0832	0.1576	0.1450
G_{31}-T_5	0.0748	0.0059	0.0588	0.2050	0.0790
G_{31}-T_6	0.8706	0.0024	0.0078	0.0028	0.0534
G_{31}-T_7	0.0394	0.0104	0.3805	0.0210	0.0645
G_{31}-T_8	0.0146	0.1088	0.3626	0.0238	0.0666
G_{31}-T_9	0.0035	0.9402	0.0190	0.0019	0.0744
G_{31}-T_{10}	0.2460	0.0029	0.3104	0.0077	0.1495
G_{33}-T_1	0.6625	0.0106	0.0191	0.0023	0.0331
G_{33}-T_2	0.0889	0.0045	0.0490	0.2954	0.0464
G_{33}-T_3	0.0035	0.3454	0.0059	0.0022	0.0452
G_{33}-T_4	0.0047	0.0683	0.2946	0.0219	0.0026
G_{33}-T_5	0.2661	0.0042	0.0412	0.0071	0.1303
G_{33}-T_6	0.1608	0.0068	0.1343	0.0922	0.1011
G_{33}-T_7	0.0147	0.0655	0.0829	0.0218	0.0478
G_{33}-T_8	0.0998	0.1851	0.2131	0.0051	0.0365
G_{33}-T_9	0.2410	0.0037	0.5293	0.0022	0.0054
G_{33}-T_{10}	0.0053	0.0559	0.1016	0.2463	0.0028
G_{33}-T_{11}	0.0072	0.0612	0.0039	0.0086	0.1388
G_{33}-T_{12}	0.0028	0.8352	0.0031	0.0040	0.0377
G_{33}-T_{13}	0.0326	0.0742	0.0816	0.0024	0.5884
G_{33}-T_{14}	0.0214	0.0471	0.0397	0.0409	0.0767
G_{33}-T_{15}	0.0015	0.0015	0.0016	0.7621	0.0015

Table 11
Topic Similarity Calculation Result of the market among three stages (G_{41} to G_{42}, G_{42} to G_{43}).

Topic	G_{42}-T_1	G_{42}-T_2	G_{42}-T_3
G_{41}-T_1	0.7998	0.0188	0.0073
G_{41}-T_2	0.0107	0.2583	0.0348
G_{41}-T_3	0.3433	0.0320	0.1403
G_{41}-T_4	0.0129	0.2434	0.0784
G_{41}-T_5	0.0183	0.2149	0.1171
G_{41}-T_6	0.0229	0.0521	0.6409
G_{41}-T_7	0.0637	0.4848	0.0095
G_{41}-T_8	0.1270	0.2613	0.0865
G_{43}-T_1	0.1774	0.1812	0.0123
G_{43}-T_2	0.0340	0.0439	0.0468
G_{43}-T_3	0.0399	0.0943	0.0125
G_{43}-T_4	0.0244	0.0074	0.7362
G_{43}-T_5	0.0093	0.0242	0.1553
G_{43}-T_6	0.0256	0.3558	0.0560
G_{43}-T_7	0.1207	0.4623	0.0064
G_{43}-T_8	0.4956	0.0366	0.0057
G_{43}-T_9	0.0585	0.0951	0.0610
G_{43}-T_{10}	0.0368	0.3871	0.0259
G_{43}-T_{11}	0.8716	0.0095	0.0067
G_{43}-T_{12}	0.1687	0.0141	0.1397
G_{43}-T_{13}	0.0611	0.0217	0.2339

Table 12
Topic Similarity Calculation Result of topics among L_1 to L_2.

Topic	G_{21}-T_1	G_{21}-T_2	G_{22}-T_1	G_{22}-T_2	G_{23}-T_1	G_{23}-T_2
G_{11}-T_1	0.0543	0.0852	0.0842	0.6040	0.1995	0.5807
G_{11}-T_2	0.0664	0.0200	0.2669	0.3054	0.3314	0.1993
G_{11}-T_3	0.4089	0.0777	0.4674	0.0340	0.0601	0.2852
G_{11}-T_4	0.2571	0.0693	0.0385	0.0431	0.0411	0.0136
G_{11}-T_5	0.0131	0.5774	0.1112	0.0992	0.0909	0.0805
G_{12}-T_1	0	0	0.4038	0.0971	0.1414	0.2478
G_{12}-T_2	0	0	0.0545	0.0588	0.0699	0.0362
G_{12}-T_3	0	0	0.2919	0.5935	0.3188	0.5384
G_{13}-T_1	0	0	0	0	0.0558	0.0724
G_{13}-T_2	0	0	0	0	0.4264	0.3645
G_{13}-T_3	0	0	0	0	0.0110	0.0123
G_{13}-T_4	0	0	0	0	0.2095	0.6493

three aspects: layers design, data sources selection, and data processes for TRM. Pharmaceutical technology innovation usually consists of specific stages, with many types of technologies and clear product generation. We propose a systematic and domain-specific TRM in pharmaceutical TOA in three extensions rather than the standardization TRM: the designing of the layers, the selecting of the data source, and the processing of data sources.

At first, we expanded on the design of the layers. Clinical research is an essential part of pharmaceutical innovation. To accommodate the pharmaceutical's domain-specificity (Lee & Park, 2005; Lee et al., 2013), it is necessary to develop clinical trial layers (Aghil Hamidi et al., 2021; Al-Humadi, 2017; Palucki et al., 2010). The previous TRM is generally constructed in the "market-product-technology" framework (Alejandro et al., 2022; Zhou et al., 2019), although there have been researchers who have constructed a research framework of "market-product-clinical trial-technology" (Jia et al., 2018). However, only a qualitative method has been adopted to analyze the development of the entire pharmaceutical industry from a macro perspective. In this study, combined with the quantitative method, we propose a domain-specific TRM to identify technological opportunities, which is suited for certain types of pharmaceutical technological innovation from a micro perspective. Finally, the feasibility of the method was validated with a hyperuricemia drug innovation.

Furthermore, we enriched the data source. Whether pharmaceutical innovations can transform into products successfully, the related papers, patents, and commercial data can provide positive or negative enlightenment indications for the investigators (Sternitzke, 2010). Paper, patents, and commercial data are major data sources of technology, clinical trial, product, and market sublayers in pharmaceutical TRM. They are complementary and well organized. In previous studies, when selecting data sources for the TRM, some investigators used patents and papers, and some investigators used research data, patents, and commercial publications as data sources (Wang et al., 2021; Zhou et al., 2019). Few researchers chose data sources from the aspect of papers, patents, and commercial data. This article combined multiple data sources: Medline, DII, and ABI commercial databases to describe the evolution of how the basic technology pharmaceutical topics transition to clinical trials, to commercial products, and then advance toward the market over time. This provides a new perspective for analyzing the pharmaceutical TOA using the TRM (Milshina & Vishnevskiy, 2017; Son & Lee, 2019).

Finally, we adjusted the data process. Although there have been studies that combined TRM with SAO to analyze TOA or identify

Table 13
Topic Similarity Calculation Result of topics among L_2 to L_3.

Topic	G_{31}-T_1	G_{31}-T_2	…	G_{32}-T_1	G_{32}-T_2	…	G_{33}-T_1	G_{33}-T_2	…
G_{21}-T_1	0.0231	0.4045	…	0.2033	0.0033	…	0.2041	0.0168	…
G_{21}-T_2	0.2054	0.1164	…	0.2672	0.0151	…	0.1028	0.0917	…
G_{22}-T_1	0	0	…	0.2234	0.0169	…	0.0735	0.0412	…
G_{22}-T_2	0	0	…	0.2105	0.0018	…	0.1754	0.0178	…
G_{23}-T_1	0	0	…	0	0	…	0.1589	0.0149	…
G_{23}-T_2	0	0	…	0	0	…	0.0635	0.0257	…

Table 14
Topic Similarity Calculation Result of topics among L_3 to L_4.

Topic	G_{41}-T_1	G_{41}-T_2	…	G_{42}-T_1	G_{42}-T_2	…	G_{43}-T_1	G_{43}-T_2	…
G_{31}-T_1	0.0020	0.4098	…	0.0105	0.1711	…	0.0020	0.1018	…
G_{31}-T_2	0.4225	0.0166	…	0.3673	0.0278	…	0.1167	0.0044	…
……	……	……	…	……	……	…	……	……	…
G_{32}-T_1	0	0	…	0.3271	0.0508	…	0.1533	0.0120	…
G_{32}-T_2	0	0	…	0.0269	0.0150	…	0.0014	0.0189	…
……	……	……	…	……	…	…	…	……	…
G_{33}-T_1	0	0	…	0	0	…	0.0180	0.0019	…
G_{33}-T_2	0	0	…	0	0	…	0.0165	0.1167	…
……	……	…..	…	…	…	…	……	……	…

potential commercial opportunities by using keywords (Wang et al., 2015; Zhou et al., 2019). However, expert judgment is required when classifying keywords into different layers of the TRM. Although expert judgment is essential, relying on expert judgment exclusively may produce many obstacles, such as being heavily influenced by experts' personal experiences, inefficient data analysis, and having difficulty reaching consensus among different experts. Inspired by this aspect, we attempted to develop an ML-driven TRM, a hybrid approach based on BERT, LDA, and cosine similarity. The BERT model, in particular, was chosen to classify the pre-processed literature, patents, and commercial data into different sublayers of the TRM. Based on Google's large prediction database, domain experts in pharmaceutical technological innovation and hyperuricemia drugs R&D were invited to build a small training set and classify more than 70,000 short sentences into proper sublayers. After testing, the accuracy in the classification of hyperuricemia drug data was more than 90 percent, which is a high efficiency that is suitable for massive data.

6. Conclusion

Pharmaceutical technological innovation is becoming increasingly important in promoting transforming scientific discoveries into marketable drugs and better public health. Although the expenditures spent on pharmaceutical innovation keep increasing, the pace of translation of new pharmaceutical technologies into drugs has been lagging behind the discovery. Understanding how pharmaceutical technology topics evolve helps us to analyze technology opportunities, which has implications for investigators and business enterprises. In response to this challenge, we proposed a novel approach to develop a pharmaceutical TRM. The proposed framework consists of three stages. First, the pharmaceutical TRM is designed with six layers (technology, clinical trial, product, market, key categories, and future direction). Using paper, patent, and commercial data as the data sources, the collected data were classified into proper layers and stages of the pharmaceutical TRM based on the BERT model. Second, clustering and extracting the potential technology topics in the TRM based on LDA were conducted. The final stage was drawing the TRM, evaluating the importance of potential technology topics, and analyzing the evolution path. An illustrative case study relating to hyperuricemia drugs was employed to validate the feasibility of analyzing our pharmaceutical drugs. The case study demonstrated that reducing the toxicity of small molecule drugs, resurrecting the pseudogene of human uric acid oxidase, and adopting combination therapy with Chinese medicines supplementary dietary therapy may be the future direction for hyperuricemia drugs. This approach can be effectively used by investigators, enterprises, and governments working for pharmaceutical technology innovation, all of whom want to explore trajectories of pharmaceutical innovation progress. Except that the research framework developed in this study is capable of providing reliable, quantitative information, as well as scientific evidence for pharmaceutical innovation at a micro level.

Despite its contributions, there are still some limitations of this study, which may be addressed in future research. First, literature, patents, and commercial databases were selected as the data sources of the pharmaceutical TRM. In the future, because product descriptions have various databases, such as the Drug Bank, we can extend more databases into the analysis. Second, although our study aimed to minimize subjective opinions, time, and efforts by an expert with a systematic approach through combing ML tools with pharmaceutical TRM, nevertheless the opinions of domain experts were still required in building small sample training sets. In the future, we can employ more advanced techniques with the aim of more accurate forecasting. Finally, the pharmaceutical TRM should be extended and validated with more cases in the future, which can provide more general guidelines for the investigators.

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 Innovation Method Fund of China (Project No. 2018IM020300, 2019IM020200); the Joint Funds of National Natural Science Foundation of China (Project No. U1904210-4); Shanghai Science and Technology Program (Project No. 20040501300); the China-National Natural Science Foundations (Project No. U1604187 , 62173253); and National Key Research and Development Program (grant number 2022YFF0608700).

References

Aaldering, L. J., & Song, C. H. (2019). Tracing the technological development trajectory in post-lithium-ion battery technologies: A patent-based approach. *Journal of Cleaner Production, 241*, Article 118343.

Aghil Hamidi, H., Jalal, A., & Ali Bonyadi, N. (2021). Pharmaceutical R&D project portfolio selection and scheduling under uncertainty: A robust possibilistic optimization approach. *Computers & Industrial Engineering, 155*, Article 107114.

Alan, L. P., & Michael, J. D. (1995). Technology opportunities analysis. *Technological Forecasting and Social Change, 49*, 237–255.

Alejandro, T. D., Lukas, G., Dietmar, W.H., & David, H. (2022). Technology roadmap for the development of a 3D cell culture workstation for a biomedical industry startup. *Technological Forecasting and Social Change*, *174*, 121213.

Al-Humadi, N. (2017). Pre-clinical toxicology considerations for vaccine development. *Vaccine, 35*.

Ali, A., & Gittelman, M. (2016). Research paradigms and useful inventions in medicine: Patents and licensing by teams of clinical and basic scientists in Academic Medical Centers. *Research Policy, 45*, 1499–1511.

Allmendinger, A. (2021). Opportunities in an evolving pharmaceutical development landscape: Product differentiation of biopharmaceutical drug products. *Pharmaceutical Research, 38*, 739–757.

Andrej, K., & Dimitar, H. (2019). Disentangling the evolution of MEDLINE bibliographic database: A complex network perspective. *Journal of Biomedical Informatics, 89*, 101–113.

Borschiver, S., Vasconcelos, R. C., Silva, F. C., Freitas, G. C., Santos, P. E., & do Bomfim, R. O. (2019). Technology roadmap for hyaluronic acid and its derivatives market. *Biofuels, Bioproducts and Biorefining*, *13*, 435–444.

Bowen, A., & Casadevall, A. (2015). Increasing disparities between resource inputs and outcomes, as measured by certain health deliverables, in biomedical research. *Proceedings of the National Academy of Sciences of the United States of America, 112*, 11335–11340.

Bruno, S., Maria Teresinha Arns, S., Amanda Trojan, F., & Rafael Henrique Palma, L. (2019). Data mining and machine learning techniques applied to public health problems: A bibliometric analysis from 2009 to 2018. *Computers & Industrial Engineering*, *138*, 106120.

Chen, F., Wu, H., Dou, R., & Li, M. (2017). A high-dimensional classification approach based on class-dependent feature subspace. *Industrial Management & Data Systems, 117*, 2325–2339.

Chen, H., Zhang, G., Zhu, D., & Lu, J. (2017). Topic-based technological forecasting based on patent data: A case study of Australian patents from 2000 to 2014. *Technological Forecasting & Social Change*. S0040162517303244.

Chen-Xu, M., Yokose, C., Rai, S. K., Pillinger, M. H., & Choi, H. K. (2019). Contemporary prevalence of gout and hyperuricemia in the United States and decadal trends: The National Health and Nutrition Examination Survey, 2007–2016. *Arthritis & Rheumatology, 71*, 991–999.

Choi, J., & Hwang, Y. S. (2014). Patent keyword network analysis for improving technology development efficiency. *Technological Forecasting & Social Change, 83*, 170–182.

Cmwda, B., Dp, C., Nd, B., & Llb, D. (2021). Methodology for constructing scenarios for health policy research: The case of coverage decision-making for drugs for rare diseases in Canada. *Technological Forecasting and Social Change, 171*, 171.

Du, J., Li, P., Guo, Q., & Tang, X. (2019). Measuring the knowledge translation and convergence in pharmaceutical innovation by funding-science-technology-innovation linkages analysis. *Journal of Informetrics, 13*, 132–148.

Duda, G. N., Grainger, D. W., Frisk, M. L., Bruckner-Tuderman, L., Carr, A., Dirnagl, U., Einhaupl, K. M., Gottschalk, S., Gruskin, E., & Huber, C. (2014). Changing the mindset in life sciences toward translation: A consensus. *Science Translational Medicine, 6*, 264cm212.

Feng, J., Liu, Z., & Feng, L. (2021). Identifying opportunities for sustainable business models in manufacturing: Application of patent analysis and generative topographic mapping. *Sustainable Production and Consumption*.

Feng, L., et al. (2022). Data-Driven Technology Roadmaps to Identify Potential Technology Opportunities for Hyperuricemia. *Drugs, 15*(11), 1357.

Ferrero, I., & Sison, A. J. G. (2014). A quantitative analysis of authors, schools and themes in virtue ethics articles in business ethics and management journals (1980–2011). *Business Ethics-a European Review, 23*, 375–400.

Frank, B., Philipp, M., & Jörg, F. (2021). Intelligent cost estimation by machine learning in supply management: A structured literature review. *Computers & Industrial Engineering, 160*, Article 107601.

Guihua, W., Hehong, C., Huihui, L., Yang, H., Yanghui, L., & Changjun, W. (2020). Cross domains adversarial learning for Chinese named entity recognition for online medical consultation. *Journal of Biomedical Informatics, 112*, Article 103608.

Guo, R., et al. (2022). A variety of simple and ultra-low-cost methods preparing SLiCE extracts and their application to DNA cloning. *Journal of Microbiological Methods, 106565*.

Hao, Z. A., Td, B., & Ypz, C. (2021). Integrating patent analysis into technology roadmapping: A latent dirichlet allocation based technology assessment and roadmapping in the field of Blockchain - ScienceDirect. *Technological Forecasting and Social Change, 167*.

Herrera-Vallejera, D., & Gorbea-Portal, S. (2021). Pharmacology and Pharmacy: Research and innovation analysis. *Scientometrics, 126*, 2513–2522.

Houssein, D., Rima, K., Rafiqul, H., & Yehia, T. (2021). EMR2vec: Bridging the gap between patient data and clinical trial. *Computers & Industrial Engineering, 156*, Article 107236.

Jeong, Y., & Yoon, B. (2015). Development of patent roadmap based on technology roadmap by analyzing patterns of patent development. *Technovation, 39–40*, 37–52.

Jia, X., Chen, J., & Tang, X. (2018). Construction of technology road-map (TRM) specific to pharmaceutical industry. *Science and Technology Management Research*.

Jin, S., Pazdur, R., & Sridhara, R.J.J. (2017). Re-evaluating eligibility criteria for oncology clinical trials: Analysis of investigational new drug applications in 2015. *Journal of Clinical Oncology*, *35*, JCO2017734186.

Jin, G., Jeong, Y., & Yoon, B. (2015). Technology-driven roadmaps for identifying new product/market opportunities: Use of text mining and quality function deployment. *Advanced Engineering Informatics, 29*, 126–138.

Jing, M., Natalie, F. A., Alan, L. P., Donghua, Z., & Dorothy, F. (2019). Identifying translational indicators and technology opportunities for nanomedical research using tech mining: The case of gold nanostructures. *Technological Forecasting and Social Change, 146*, 767–775.

Kerr, C., & Phaal, R. (2020). Technology roadmapping: Industrial roots, forgotten history and unknown origins. *Technological Forecasting and Social Change, 155*.

Kim, J., & Geum, Y. (2021). How to develop data-driven technology roadmaps: The integration of topic modeling and link prediction. *Technological Forecasting and Social Change, 171*, Article 120972.

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 Systems with Applications, 152*, Article 113401.

Kumar, A., Gupta, P., Balan, R., Neti, L.B.M., & Malapati, A.J.N.P.L. (2021). BERT based semi-supervised hybrid approach for aspect and sentiment classification. *53*, 4207–4224.

Kun, L., Guancan, Y., & Xue, W. (2022). Topics emerged in the biomedical field and their characteristics. *Technological Forecasting and Social Change, 174*, Article 121218.

Lee, H., & Geum, Y. (2017). Development of the scenario-based technology roadmap considering layer heterogeneity: An approach using CIA and AHP. *Technological Forecasting & Social Change, 117*, 12–24.

Lee, S., & Park, Y. (2005). Customization of technology roadmaps according to roadmapping purposes: Overall process and detailed modules. *Technological Forecasting & Social Change, 72*, 567–583.

Lee, J. H., Phaal, R., & Lee, S. H. (2013). An integrated service-device-technology roadmap for smart city development. *Technological Forecasting & Social Change, 80*, 286–306.

Leem, B.-H., Eum, S.-W.-J.-I.-M., & Systems, D. (2021). *Using text mining to measure mobile banking service quality*. Industrial Management & Data Systems.

Lijun, Sun, Yafeng, & Yin (2017). Discovering themes and trends in transportation research using topic modeling. *Transportation Research Part C Emerging Technologies*.

Lim, M.K., Li, Y., Song, X.J.I.M., & Systems, D. (2021). Exploring customer satisfaction in cold chain logistics using a text mining approach. *Industrial Management & Data Systems*.

Liu, B., Lai, M., Wu, J.-L., Fu, C., & Binaykia, A. (2020). Patent analysis and classification prediction of biomedicine industry: SOM-KPCA-SVM model. *Multimedia Tools and Applications, 79*, 10177–10197.

Liu, D.-R., Liao, Y.-S., Lu, J.-Y.-J.-I.-M., & Systems, D. (2019). *Online news recommendations based on topic modeling and online interest adjustment*. Industrial Management & Data Systems.

Liu, Y., Zhou, Y., Liu, X., Dong, F., Wang, C., & Wang, Z. (2019). Wasserstein GAN-based small-sample augmentation for new-generation artificial intelligence: A case study of cancer-staging data in biology. *Engineering*.

Liu, Z., ... Feng, J. (2022). Vehicle-type strategies for manufacturer's car sharing. Kybernetes, ahead-of-print.

Mcdowall, W. (2012). Technology roadmaps for transition management: The case of hydrogen energy. *Technological Forecasting & Social Change, 79*, 530–542.

Milshina, Y., & Vishnevskiy, K. (2017). Roadmapping in fast changing environments – the case of the Russian media industry. *Journal of Engineering & Technology Management, 52*, 32–47.

Palucki, M., Higgins, J. D., Kwong, E., & Templeton, A. C. (2010). Strategies at the interface of drug discovery and development: Early optimization of the solid state phase and preclinical toxicology formulation for potential drug candidates. *Journal of Medicinal Chemistry, 41*.

Park, K., & Kremer, G. E. O. (2017). Text mining-based categorization and user perspective analysis of environmental sustainability indicators for manufacturing and service systems. *Ecological Indicators, 72*, 803–820.

Rake, B. (2017). Determinants of pharmaceutical innovation: The role of technological opportunities revisited. *Journal of Evolutionary Economics, 27*, 691–727.

Raza, M. S., Khahro, S. H., Memon, S. A., Ali, T. H., & Memon, N. A. (2021). Global trends in research on carbon footprint of buildings during 1971–2021: A bibliometric investigation. *Environmental Science and Pollution Research, 28*, 63227–63236.

Schoggl, J. P., Stumpf, L., & Baumgartner, R. J. (2020). The narrative of sustainability and circular economy - A longitudinal review of two decades of research. *Resources Conservation and Recycling, 163*.

Shah, S. M. A., Taju, S. W., Ho, Q. T., Nguyen, T. T. D., & Ou, Y. Y. (2021). GT-Finder: Classify the family of glucose transporters with pre-trained BERT language models. *Computers in Biology and Medicine, 131*.

Son, W., & Lee, S. (2019). Integrating fuzzy-set theory into technology roadmap development to support decision-making. *Technology Analysis & Strategic Management, 31*, 447–461.

Sternitzke, C. (2010). Knowledge sources, patent protection, and commercialization of pharmaceutical innovations. *Research Policy, 39*, 810–821.

Sze, T. M., Phaik-Leng, C., Ai-Vyrn, C., Lai-Meng, L., & Siow-Wee, C. (2021). A review on omics-based biomarkers discovery for Alzheimer's disease from the bioinformatics perspectives: Statistical approach vs machine learning approach. *Computers in Biology and Medicine, 139*.

Tan, X., Zhuang, M., Lu, X., & Mao, T. (2021). An analysis of the emotional evolution of large-scale internet public opinion events based on the BERT-LDA hybrid model. *IEEE Access*, 1-1.

Tierney, R., Hermina, W., & Walsh, S. (2013). The pharmaceutical technology landscape: A new form of technology roadmapping. *Technological Forecasting and Social Change, 80*, 194–211.

Tingting, M., Xiao, Z., Jia, L., Zhenkai, L., Zhaoting, H., & Ruitao, W. (2021). Combining topic modeling and SAO semantic analysis to identify technological opportunities of emerging technologies. *Technological Forecasting and Social Change, 173*, Article 121159.

Wang, X., Qiu, P., Zhu, D., Mitkova, L., Lei, M., & Porter, A. L. (2015). Identification of technology development trends based on subject–action–object analysis: The case of dye-sensitized solar cells. *Technological Forecasting & Social Change, 98*.

Wang, X., Yang, X., Wang, X., Xia, M., & Wang, J. (2020). Evaluating the competitiveness of enterprise's technology based on LDA topic model. *Technology Analysis & Strategic Management, 32*.

Wang, X., Zhang, S., Liu, Y., Du, J., & Huang, H. (2021). How pharmaceutical innovation evolves: The path from science to technological development to marketable drugs. *Technological Forecasting and Social Change, 167*, Article 120698.

Wang, J., et al. (2022). A digital twin modeling approach for smart manufacturing combined with the UNISON framework. *Computers & Industrial Engineering, 169*, 108262.

Wang, L. Y., & Zhao, D. (2021). Cross-domain function analysis and trend study in Chinese construction industry based on patent semantic analysis. *Technological Forecasting and Social Change, 162*, Article 120331.

Wu, Q. Q., Zhang, C., Hong, Q. Q., & Chen, L. Y. (2014). Topic evolution based on LDA and HMM and its application in stem cell research. *Journal of Information Science, 40*, 611–620.

Ye, C., Linas, M., Seza, O., & Gintaras, V. R. (2012). Simulation-optimization approach to clinical trial supply chain management with demand scenario forecast. *Computers & Chemical Engineering, 40*, 82–96.

Yeung, A., Atanasov, A. G., Sheridan, H., Klager, E., & Schaden, E. (2021). Open innovation in medical and pharmaceutical research: A literature landscape analysis. *Frontiers in Pharmacology, 11*, 1–9.

Yu, X., & Zhang, B. (2019). Obtaining advantages from technology revolution: A patent roadmap for competition analysis and strategy planning. *Technological Forecasting and Social Change, 145*, 273–283.

Zhang, Y., Zhang, G., Chen, H., Porter, A. L., Zhu, D., & Lu, J. (2016). Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research. *Technological forecasting and social change, 105*, 179–191.

Zhao, H., Huang, E., Dou, R., & Wu, K. (2019). A multi-objective production planning problem with the consideration of time and cost in clinical trials. *Expert Systems with Applications, 124*, 25–38.

Zhengxin, W., Xinggan, P., Ao, X., Akeel, A. S., Yun, H., Xianqing, Z., … Qiang, L. (2022). The role of machine learning to boost the bioenergy and biofuels conversion. *Bioresource Technology, 343*, Article 126099.

Zhinan, W., Alan, L. P., Xuefeng, W., & Stephen, C. (2019). An approach to identify emergent topics of technological convergence: A case study for 3D printing. *Technological Forecasting and Social Change, 146*, 723–732.

Zhou, X., Huang, L., Porter, A., Vicentegomila, J. M., & Phillips, F. (2019). Tracing the system transformations and innovation pathways of an emerging technology: Solid lipid nanoparticles. *Technological Forecasting and Social Change, 146*.