



Article

Data-Driven Technology Roadmaps to Identify Potential Technology Opportunities for Hyperuricemia Drugs

Lijie Feng ¹, Weiyu Zhao ^{2,*}, Jinfeng Wang ^{3,*} , Kuo-Yi Lin ⁴, Yanan Guo ^{5,6,*} and Luyao Zhang ^{7,8}

- ¹ Logistics Engineering College, Shanghai Maritime University, 1550 Haigang Avenue, Pudong District, Shanghai 201306, China
- ² Institute of Logistics Science and Engineering, Shanghai Maritime University, 1550 Haigang Avenue, Pudong District, Shanghai 201306, China
- ³ China Institute of FTZ Supply Chain, Shanghai Maritime University, 1550 Haigang Avenue, Pudong District, Shanghai 201306, China
- ⁴ School of Business, Guilin University of Electronic Technology, Guilin 541004, China
- ⁵ School of Life Sciences, Shanghai University, Shanghai 200444, China
- ⁶ School of Life Sciences, Zhengzhou University, No. 100 Science Avenue, Zhengzhou 450001, China
- ⁷ School of Economics and Management, Shanghai Maritime University, 1550 Haigang Avenue, Pudong District, Shanghai 201306, China
- ⁸ School of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou 450016, China
- * Correspondence: 202040510009@stu.shmtu.edu.cn (W.Z.); wangjinfeng@shmtu.edu.cn (J.W.); guoyan@zzu.edu.cn (Y.G.)



Citation: Feng, L.; Zhao, W.; Wang, J.; Lin, K.-Y.; Guo, Y.; Zhang, L. Data-Driven Technology Roadmaps to Identify Potential Technology Opportunities for Hyperuricemia Drugs. *Pharmaceuticals* **2022**, *15*, 1357. <https://doi.org/10.3390/ph15111357>

Academic Editors: Silviya Petrova Zustiak and Serge Mordon

Received: 20 September 2022

Accepted: 31 October 2022

Published: 3 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Hyperuricemia is a metabolic disease with an increasing incidence in recent years. It is critical to identify potential technology opportunities for hyperuricemia drugs to assist drug innovation. A technology roadmap (TRM) can efficiently integrate data analysis tools to track recent technology trends and identify potential technology opportunities. Therefore, this paper proposes a systematic data-driven TRM approach to identify potential technology opportunities for hyperuricemia drugs. This data-driven TRM includes the following three aspects: layer mapping, content mapping and opportunity finding. First we deal with layer mapping. The BERT model is used to map the collected literature, patents and commercial hyperuricemia drugs data into the technology layer and market layer in TRM. The SAO model is then used to analyze the semantics of technology and market layer for hyperuricemia drugs. We then deal with content mapping. The BTM model is used to identify the core SAO component topics of hyperuricemia in technology and market dimensions. Finally, we consider opportunity finding. The link prediction model is used to identify potential technological opportunities for hyperuricemia drugs. This data-driven TRM effectively identifies potential technology opportunities for hyperuricemia drugs and suggests pathways to realize these opportunities. The results indicate that resurrecting the pseudogene of human uric acid oxidase and reducing the toxicity of small molecule drugs will be potential opportunities for hyperuricemia drugs. Based on the identified potential opportunities, comparing the DNA sequences from different sources and discovering the critical amino acid site that affects enzyme activity will be helpful in realizing these opportunities. Therefore, this research provides an attractive option analysis technology opportunity for hyperuricemia drugs.

Keywords: data-driven TRM; SAO analysis; link prediction; hyperuricemia drug; human uric acid oxidase

1. Introduction

The amount of uric acid in the body needs to be kept at a stable level. While the synthesis of uric acid increases or the amount of uric acid excreted from the body decreases, the concentration of uric acid in the blood increases [1–3]. A person is considered to have hyperuricemia when the uric acid level in the blood exceeds the average levels [4–7]. The

increased intake of high-fat, high-protein, and high-sugar food leads to metabolic disorders and a rising risk of hyperuricemia [8–10]. It is estimated that there are currently about 17.7 million hyperuricemia patients worldwide. As can be seen, hyperuricemia positively correlates to many other potential diseases, such as obesity, hypertension, diabetes, cardiovascular disease, and chronic kidney disease [11]. The primary approach in treating hyperuricemia is to rebalance uric acid synthesis and excretion. Drugs for treating hyperuricemia are divided into three categories: xanthine oxidase inhibitors, urate anion transporter 1 (URAT1) and urate oxidase. However, most of these chemical drugs for hyperuricemia have specific side effects, and the oxidase-based drugs produce antibodies with long-term use. Neither of them can dissolve uric stones deposited in the joints [12,13]. Technological innovation for hyperuricemia drugs is imperative.

The interest in developing hyperuricemia drugs is continuously growing. Monitoring hyperuricemia pharmaceutical technologies are fundamental to analyzing and identifying potential technology opportunities. It is helpful to narrow the scope of technology research topics and reduce R&D risks, which not only supports the decision-making of hyperuricemia drug research but also helps to avoid unnecessary R&D costs [14–17]. Various data analysis tools were explored to analyze potential technology opportunities efficiently [18–20], such as bibliometrics [21], citation analysis [22], technology roadmap (TRM) [23], Biterm Topic Model (BTM) [24], Bidirectional Encoder Representation from Transformers (BERT) [25,26], Subject-action-object analysis (SAO), [27] and link prediction [15,28,29]. These tools integrate mathematics, statistics, computer science, and operations research in technology opportunity analysis [16,30–33]. However, these tools do not work independently, but often combine into a new, more efficient analysis path.

TRM is a comprehensive approach to capturing changes in technology and markets over time in an integrated manner. It is not only a flexible approach to analyzing technologies and market requirements [34,35], but it can also create a more effective way to track and analyze the latest technology trends by integrating data analysis tools [36–38]. Pharmaceutical technology innovations are influenced by various factors, such as changing customer expectations, uncertain intellectual property (IP) procedures, unconsidered technology changes, and resource requirements [23,39–41]. It is necessary to use TRM to analyze technology opportunity trends from both technological and market dimensions. In addition, TRM is a valuable tool in shortening the technology development cycle, discovering drug targets, and optimizing resource consumption. There are three major categories of research on TRMs: theory-based, case study-focused, and data/methodology-specific [15]. Among numerous extensions of TRM, data integration is a notable trend. Some researchers have employed the data-driven approach in TRM. For example, Yu developed a patent roadmap for the competitive market and patent layout planning analysis [42]. Zhou traced the innovation path of Solid lipid nanoparticles [37].

Despite the contributions of previous research using TRM to analyze TOA, data-driven TRM have some limitations in three areas: processing data data-driven TRM, identifying potential opportunities, and selecting the data source. From the data processing standpoint, most previous research has adopted keyword-based network analysis approaches for data-driven TRM. However, traditional keyword-based analysis approaches neglected to express the relationships between technologies and the market, inadequately reflecting the contexts. Compared with identifying potential technology opportunities, most researchers relied on current trends from the perspective of technology forecasting. The technology opportunity analysis (TOA) process contains six main stages. It includes data acquisition, description, potential relationship extraction, visualization, analysis of results, and identifying potential technical opportunities. Instead of identifying potential technology opportunities, the current TRM focuses on the first five steps in TOA. Technology opportunities are primarily based on technology hotspots, while technology opportunities often exist in potential connections. From the perspective of data source selection, technology and the market have become complex. Some critical technical information exists not only in patents, but also in literature and market reports. Similarly, market information is hidden not

only in market reports but is also embedded in patents and literature. Therefore, it is necessary to construct a comprehensive data-driven TRM to automatically, quickly, and accurately extract technology and market data from a large amount of literature, patent, and commercial data.

Therefore, this article is proposing a systematic method for developing data-driven TRM to identify potential technology opportunities for hyperuricemia drugs which contains three stages. The first stage is layer mapping. We map the literature, patent, and commercial data into the technology and market layers based on BERT and semantic analysis for the technology layer and market layer based on SAO. The second is content mapping. We identify topics of SAO components for technology and market layers based on BTM. The last stage is opportunity finding. We identify possible links between unconnected nodes based on link prediction. This data-driven TRM effectively identifies potential technology opportunities for hyperuricemia drugs and suggests pathways to realize these opportunities.

The rest of this paper is organized as follows. Section 2 describes the relevant theoretical background on TOA, TRM, and data-driven TRM regarding BTM, SAO, and link prediction. Section 3 outlines our proposed approach for a data-driven TRM. It explains integrating BTM, BERT, SAO, and link prediction to analyze technological opportunities for hypouricemic drugs. A case study of technology prediction related to hyperuricemia is then presented in Section 4. Section 5 discusses our discovery and extension of the study. Section 6 summarizes the paper, looks at possible future research, and provides some limitations of our study.

2. Theoretical Background

2.1. Technology Opportunity Analysis

Technology opportunity analysis (TOA) helps researchers and organizations explore potential technological opportunities. It also enables a better understanding of scientific and technological developments by deeply mining valid information in publications, patents, and the literature [43].

Many researchers have developed effective methods to identify and predict technology opportunities. Early research used qualitative analysis methods that relied on expert experiences, such as Delphi and Workshop. While in specialized fields, specialist opinion can provide creative foresight for analyzing technology opportunities. However, information increased steeply. It is impossible to consistently identify technology trends based on expert knowledge alone, which is time-consuming and costly. In addition, specialist judgment is often limited by personal expertise and bias. Sometimes consensus cannot be reached.

Bibliometrics was first introduced to analyze technological opportunities for emerging technologies [44], which is used to evaluate R&D activities by counting the number of authors and literature citation relationships [45,46]. This method has been widely used for the analysis of technology evolution trajectories in energy [47], conductive polymer nanocomposites [48], and robotics [49]. Bibliometrics provides quantitative data and objective evidence to evaluate technical opportunities and reach a consensus of experts. However, bibliometrics cannot extract the meaning of documents in-depth but can only reflect information such as the flow of knowledge, citations of literature, and patents. However, it has a time lag.

Then text mining is then introduced in TOA, which is suitable for unstructured text data analysis and can extract text features in-depth [50]. There are many data analysis techniques such as machine learning and natural language processes arose. Some research has focused on developing automated and semi-automated data analysis methods in TOA [51–53]. Among them, principal component analysis (PCA) and text clustering are often used to extract topic information [37]. The similarity is used to measure connections between technical topics [54]. For example, Wu predicted evolutionary relationships between stem cell themes based on LAD, HMM, and co-occurrence theory [55]. Du used the topic models to predict potential topics for new drugs [22]. Zheng presented text mining

tools to reveal possible innovation pathways and commercial applications of solid lipid nanoparticle drugs [56]. Zheng reviewed the importance of machine learning in facilitating the translation of bioenergy and biofuel innovations [57].

TOA has evolved over a long period of time and has been enriched by many scholars. To analyze potential technology opportunities efficiently, various data analysis tools were explored. These tools integrate mathematics, statistics, computer science, and operations research in TOA. However, rather than working independently, these methods are often combined into a new, more efficient analysis path.

2.2. Technology Roadmap and Data-Driven Technology Roadmap

TRM is a time-based multi-layer chart that can be integrated with various data analysis tools to form a more efficient analysis path in TOA. TRM-based technology opportunity analysis can better identify the dynamic distribution of technology. It can also predict technology development trends and identify potential technology opportunities in a time series [34,40,58]. It usually consists of three layers: the market, product, and technology [39,59] layers, as shown in Figure 1. The existing research of TRM in TOA is mainly categorized into the following streams: theory-based, case study-focused, and data/methodology-specific.

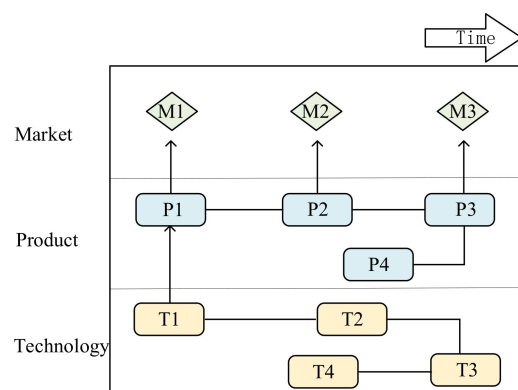


Figure 1. Structure of technology roadmap.

Regarding theory-based TRM, some previous studies have focused on the concept and process of TRM [60,61]. To support market-pull and technology-driven innovation, new frameworks for TRM have been proposed, such as T-Plan TRM [62], learning-based TRM [63], or umbrella-based TRM [61].

In terms of a case study-focused on TRM, some previous studies have focused on applying TRM in different industries or sectors [59,64–67]. To accommodate the domain-specific and case-specific needs in other areas, the customization of TRMs has been developed. There are TRMs in the aeronautical and aerospace sectors [68] and in robotics technologies in the power sector [69]. There are also TRMs in agile hardware development [67] and pharmaceutical technology landscape development [70].

Regarding data/methodology specific to TRM, some previous studies have focused on integrating data analysis tools to develop an efficient TRM [58,71–73]. The TRM has excellent flexibility in the structure and development process. Various tools can be flexibly selected to build a TRM according to different purposes of TOA. For example, researchers have integrated various tools into TRM to accommodate the complex business environment and the rise of big data. Means such as technology mining (TM), analytic hierarchy process (AHP) [74], business model canvas (BMC) [75], cross impact analysis (CIA) [76], and fuzzy set theory [58] have been employed. Some studies used technology mining-based patents analysis for technology roadmaps to explore AI-healthcare innovation [77]. Furthermore, some studies employ tools such as Bayesian network and topic modeling to develop a risk-adaptive technology roadmap under deep uncertainty [78].

With the rise of big data analytics and the rapid change in the business environment, TRM integrating data and analysis methodology for TOA is becoming increasingly popular. More and more researchers concentrate on the importance of data to TRM [79]. The data-driven technology roadmap is gradually proposed [80]. The data source for data-driven TRM is increasingly diversified, mainly including patents, literature, and commercial data. Data analysis tasks can be used for data-driven TRM, such as text classification, summarizing, key information extraction, topic clustering, semantic analysis, navigation, topic visualization, and node linking [15,71,81]. To put the data source selected into proper layers of TRM, some studies choose data analysis tasks, such as text classification models [82]. Some studies employ data analysis tasks to identify potential technology topics for TRM, such as text clustering tools [57,83]. Some studies used semantic analysis tools to extract critical technology information, such as SAO [84,85]. And to identify potential technology opportunities, some studies adopt link prediction [15].

2.3. Data Analysis Techniques and Data-Driven Technology Roadmap

Data analysis techniques have been increasingly employed to support quantitative and intelligent data-driven TRM.

2.3.1. Bidirectional Encoder Representations for Transformers with the Data-Driven Technology Roadmap

Data analysis tools, such as text classification models, can be used to put the data source selected into proper layers of data-driven TRM. Classification models such as support vector machine (SVM) [86], k-nearest neighbor (KNN) [87,88], Hidden Markov [89], and Bayesian [44,90] can be employed.

To improve classification accuracy, these text classification models should be trained based on a massive manually labeled training dataset [86], which is time-consuming, labor-intensive, and costly. The low accuracy of the classification model is often caused by the small sample size, inefficient model computation, and high reliance on domain experts. To improve the accuracy of classification based on small sample training set models, the BERT model is proposing in 2018. The model has been widely used for its excellent performance in text classification [91]. When dealing with domain-specific classification tasks such as pharmaceutical technology, it is required to construct small domain sample datasets and pre-train the model [17,92,93].

This paper intends to create a small domain-specific training set. This will be followed by training the BERT model based on fine-tuning [94] to accurately classify pharmaceutical data into proper layers of data-driven TRM.

2.3.2. Subject-Action-Object Analysis with the Data-Driven Technology Roadmap

Data analysis tools, such as SAO, can extract critical technology information for layer mapping of data-driven TRM [95]. Initially, the SAO structure was widely used to analyze technical documents such as patents, present valid technical information, critical technocratic findings, and represent the relationships between technical elements [96–102]. Guo constructed SAO chains to identify future directions of technology [103]. Wang identified technology opportunities based on SAO and the morphological matrix [104,105]. Natural language processing techniques have enabled SAO structures to express rich semantic information compared with topics. Therefore, it is considered an effective tool for identifying critical technical inter-elements in a corpus [106].

Subsequently, the SAO structure has been extended to many other fields, such as patent similarity analysis [85,107] and patent network analysis [108]. It can also apply to technology tree analysis [96,109], technology trend analysis [110], online review demand extraction [27], and M&A target selection [101].

Although SAO is widely used in TRM with TOA, there are still some limitations. When constructing the data-driven TRM, if the SAO structures are adopted directly without refining, the TRM is likely has a large amount of redundancy. It is inappropriate for efficient

analysis and needs further refinement [111]. Therefore, it is necessary to identify topics of SAO components for different layers of data-driven TRM based on topic modeling tools [27,105].

2.3.3. Biterm Topic Model with the Data-Driven Technology Roadmap

Data analysis tools, such as topic models, can identify potential technology topics for data-driven TRM [112,113]. Identifying topics of SAO components for different layers via a topic modeling tool can help researchers understand the target domain effectively. It can help to extract which areas the technical solution focuses on, how the solutions work, and which parts of the solution are the targets [44,52]. The most popular topic modeling techniques are LDA [81]. The LDA model has been widely used in various fields, such as text mining, bioinformatics, and image processing. The model has proven effective in extracting topics from large amounts of text data and analyzing technical topic changes [114,115].

However, the data is increasingly various, and short text data has emerged and been exploded. Sparse and unbalanced texts characterize short texts. The accuracy of extracting topics using the same topic modeling algorithms as long texts, such as LDA, is low. There is an urgent need to propose a topic model suitable for short texts [24]. Yan proposed a topic model algorithm, BTM, that is more suitable for short text clustering [116]. The model enhances the learning efficiency of the topic model by calculating the unordered co-occurrence word pairs. It effectively solves the semantic sparsity problem of short texts [117]. BTM can automatically extract hot and potentially technical topics from large amounts of short text data, even in the face of domain-specific datasets. BTM has become one of the most widely used short text modeling technologies [118].

BTM analyzes the technologies and market topics based on keywords that can't reflect the contexts. To reflect the contextual semantics of the topics, BTM is more effective when combined with SAO [108]. However, the SAO structure consists of phrases. BTM is suitable for identifying topics of SAO components, which helps reduce the redundancy of SAO effectively. Some limitations exist in using BTM and SAO for technical opportunity analysis. They only consider the existing relationships and links, and pay less attention to identifying potential technology opportunities in automation [102]. However, technology opportunities possibly exist in potential connections [111]. SAO and BTM need to be combined with predictive tools for better performance in TOA, such as link prediction.

2.3.4. Link Prediction with Data-Driven Technology Roadmap

Data analysis tools, such as link prediction, can identify potential connections for data-driven TRM. Link prediction is a technique for discovering nodes or links in a network that are currently unknown but may be connected in the future. It has been well developed and applied in social network analysis and TOA [119,120]. There are three significant categories of link prediction: link prediction based on similarity, maximum likelihood estimation, and probabilistic models. Link prediction based on maximum likelihood estimation is unsuitable for massive amounts of data with low prediction accuracy. Link prediction based on probabilistic models often relies on external attributes of nodes, which are often difficult to obtain. In contrast, similarity-based link prediction is more accurate and is widely used [121].

With the development of classification models, Hansen found that link prediction can be constructed based on them [122]. Supervised classification models based on Bayesian, neural networks, and support vector machines (SVM) [123] can be employed to improve model accuracy. Subsequently, scholars began to compare the performance of link prediction based on classification models in processing domain datasets. In TOA, Yoon compared similarity-based and SVM-based link prediction performance. We can try to identify potential technology opportunities based on more classification models, such as Lightgbm with link prediction [123].

Link prediction is increasingly becoming a research hotspot in technology prediction. It has been widely used in the technical analysis of biological and medical patent data [124]. Shibata performed link prediction analysis in five large citation networks [125]. Xiao combined SAO with link prediction for identifying technical opportunities in skin melanoma [111]. Ma proposed a link-prediction-based technical knowledge network framework to predict potential technical opportunities in Alzheimer's disease [126].

3. Methodology

This article proposed a systematic method for developing data-driven TRM to identify potential technology opportunities for hyperuricemia drugs. It contains three stages. The first is layer mapping. We classify the literature, patent, and commercial data into layers based on BERT and semantic analysis for the technology layer and market layer based on SAO. The second stage is content mapping. We identify topics of SAO components for technology and market layers based on BTM. The last stage is opportunity finding. We identify possible links between unconnected nodes based on link prediction. The data-driven TRM benefits technology needs assessment and technology response development in the technology roadmap process. The proposed model consists of three modules, as in Figure 2.

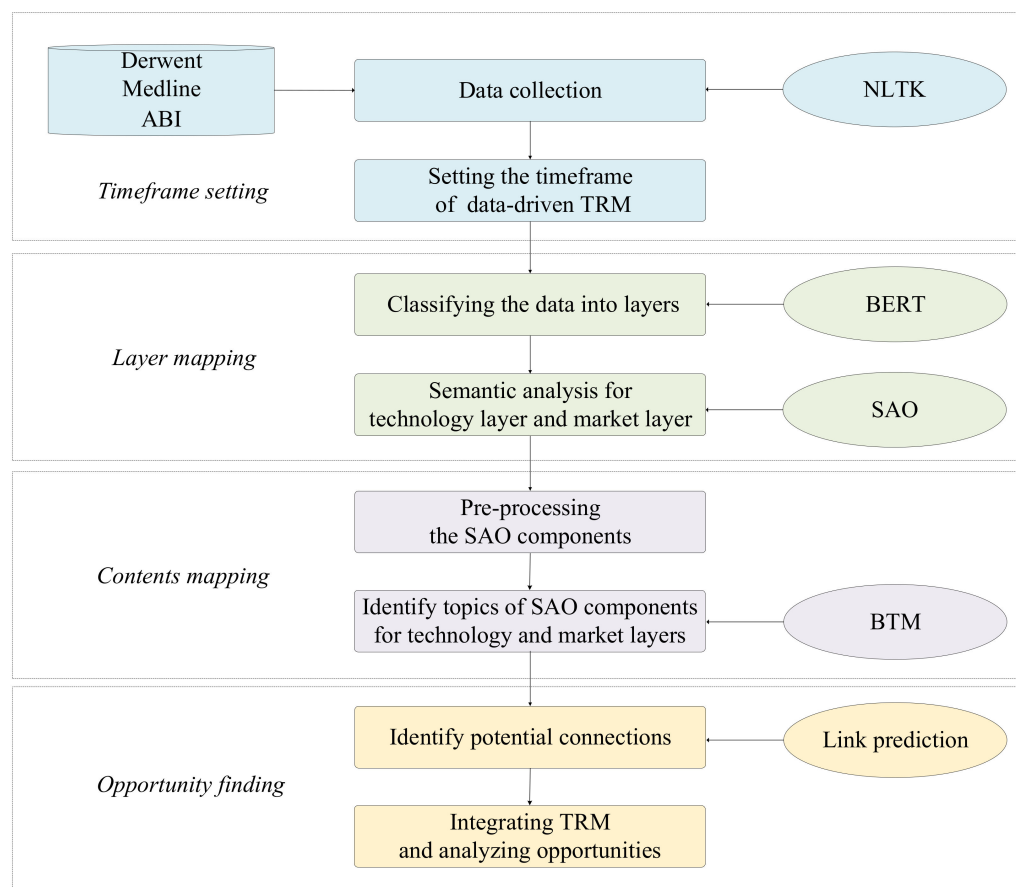


Figure 2. Research framework.

3.1. Collecting and Pre-Processing Data for Technology and Market

3.1.1. Data Collection

The numerous databases of patents, academic papers, journals, and business reports contain voluminous technology and market information. Their abstracts are also stored in a structured database format, making them a beneficial source for data analysis. Since this study calls for technical and market-related data in developing a data-driven technology roadmap, we employ Medline, Derwent Innovations Index (DII), and Abstracts of Business

Information (ABI) databases as data source collection. We then use the different search queries related to the research topic to download relevant scientific papers, patents and business journals and reports.

3.1.2. Setting the Timeframe of Data-Driven TRM

Considering the complex and dynamic technology replacement and market changes, one core of the TRM is setting the time frame. The technology opportunity analysis can be done within each time frame. Different rules, such as S-curve, can set the time frame. In the S-curve, the generation and development of technology have their pattern and trajectory. The stages of the technology life cycle are predictable and iterative. Building a time frame based on the S-curve helps to identify technology opportunities in a forward-looking manner. Therefore, this study uses the s-curve-based model to determine the development stages of R&D activities of technologies, and to identify technology and market trends within the altered time frame [127].

3.2. Layer Mapping

3.2.1. Classifying the Data into Layers Based on BERT

In the second module, BERT, the text classification algorithm based on fine-tuning is used to classify the tech-related and market-related data into layers for data-driven TRM. The previous research on TRM using data-driven approaches tends to use abstracts. The abstract is an overview of the full text and facilitates a rapid discovery of high-value information with low-value density data.

This section consists of the following steps: First, the abstracts are separately extracted from the database in a timeframe. After that, we pre-process the tech-related and market-related data in text format in the timeframe. This article chooses the sent tokenize module of the Natural Language Toolkit (NLTK) in the Python package to divide the abstracts into sentences. We then conducted a BERT model to classify technical and market data. Even if BERT performs well in classification tasks based on the Google corpus, organizing technical and market data requires domain training sets, such as pharmaceutical data. To construct the domain training set, domain experts will invite, and 30 percent of sentences will be extracted randomly from the entire data set. The extracted training set will then be manually labeled with technology-related, market-related, and irrelevant data. After that, we will pre-train the BERT model. Only the technical-related and market-related data will be left. Finally, the whole dataset will be divided into several subsets and classified into technology and market layers of data-driven TRM in the timeframe.

3.2.2. Semantic Analysis for the Technology Layer and Market Layer Based on SAO

The SAO structures are a machine learning technique. It is always employed to obtain objective, structure, and effect data from text and then converts that information into structured text data. The SAO structures consider contextual meaning, which is superior to the keyword-based analysis. Thus, we chose the SAO technique to extract technical and market semantic structures, reflecting the contexts.

This section consists of the following steps: first, we will extract the SAO structures. SAOs cannot be extracted without the help of a parser, which can analyze textual data through regular syntax rules. In this study, we employ the Stanford Parser, (Stanford Parser, 3.9.2-models; package for extracting SAO structures) available as an open-source package for sentence separation [128]. Next, we extract the SAOs with a series of linguistic algorithms from each sentence in the timeframe. After that, we will filter, clean, and combine the SAO structures. The data used for technical analysis is likely to be very large with a low-value density. It contains all of the SAO structures. It is not appropriate for high-efficient analysis and needs to be filtered, cleaned, and combined. Therefore, we delete the duplicated technology and market SAO structures; only the unique SAOs are left. However, in the base of the dependency parser, the SAO (subject + action + object), SO

(subject + action), and AO (adjective + object) structures all collected [129]. We remove the technology and market SAOs without subjects, actions, or objects, such as SO and AO.

3.3. Contents Mapping

3.3.1. Pre-Processing the SAO Components

There will be vast redundancy if the contents mapping is solely based on SAO semantic analysis to identify potential technology opportunities. Hence, in this module, in addition to filtering the SAO structure above, we will employ a text clustering approach toward the dimensionality reduction of SAOs. Short text data is sparse text and an imbalance of data. Text clustering algorithms such as LDA cannot extract topics of short texts. Most of the SAO structures are phrases. We selected BTM, more appropriate for short text clustering, to extract technology and market SAOs components topics.

This section consists of the following steps: first, we will divide the remaining technology and market SAOs into several subsets. We will then remove the data noise and pre-process the sub-datasets group by group, such as word token, stemming, lemmatization, and excluding stop words.

3.3.2. Identify Topics of SAO Components for Technology and Market Layers Based on BTM

We conducted a BTM-based topic model to identify meaningful core and potential technology and market topics automatically. Perplexity is a crucial index to evaluate the clustering effect of topic modeling models. Nonetheless, perplexity cannot explain the semantic coherence of words for each topic on a non-probabilistic model. Topic coherence can describe it. Therefore, we chose coherence as a metric to evaluate the BTM model's effectiveness.

This section consists of the following steps: first, we evaluate the value of coherence while varying the number of topics to determine the optimal number. Second, we pre-train the BTM model. Lastly, we identify topics of SAO components for technology and market layers based on BTM.

3.4. Opportunity Finding

3.4.1. Identify Potential Connections Based on Link Prediction

The core of building a data-driven roadmap is to predict possible technology opportunities. However, content mapping only analyzes past data and ignores potential future technological opportunities. Therefore, in this module, we chose link prediction to predict the potential links between unlinked nodes.

This section consists of the following steps: first, we select the results of SAO pre-processing in Section 3.3.1 as a train set. We then train the model based on link prediction to construct the overall network. Finally, we identify the probability of potential links for all unlinked topic nodes based on the trained link prediction model.

3.4.2. Integrating TRM and Analyzing Technology Opportunities

It is challenging to analyze such a large-scale network, so we only keep and interpret technology and market subnetworks. Subnetwork nodes are selected from the topics extracted by BTM.

This section consists of the following steps: first, we map the technology and market subnetworks in a time series and layer series in two-dimensional data-driven TRM. Figure 3 shows an example of the final visualization. As shown in the figure, the technology roadmap is divided into two layers, technology and market. For example, the technical layer is divided into three sub-layers, from sub-layer S, sub-layer A to sub-layer O, along the vertical axis. The horizontal axis is the time series arranged in the timeframe—the same for the market layer.

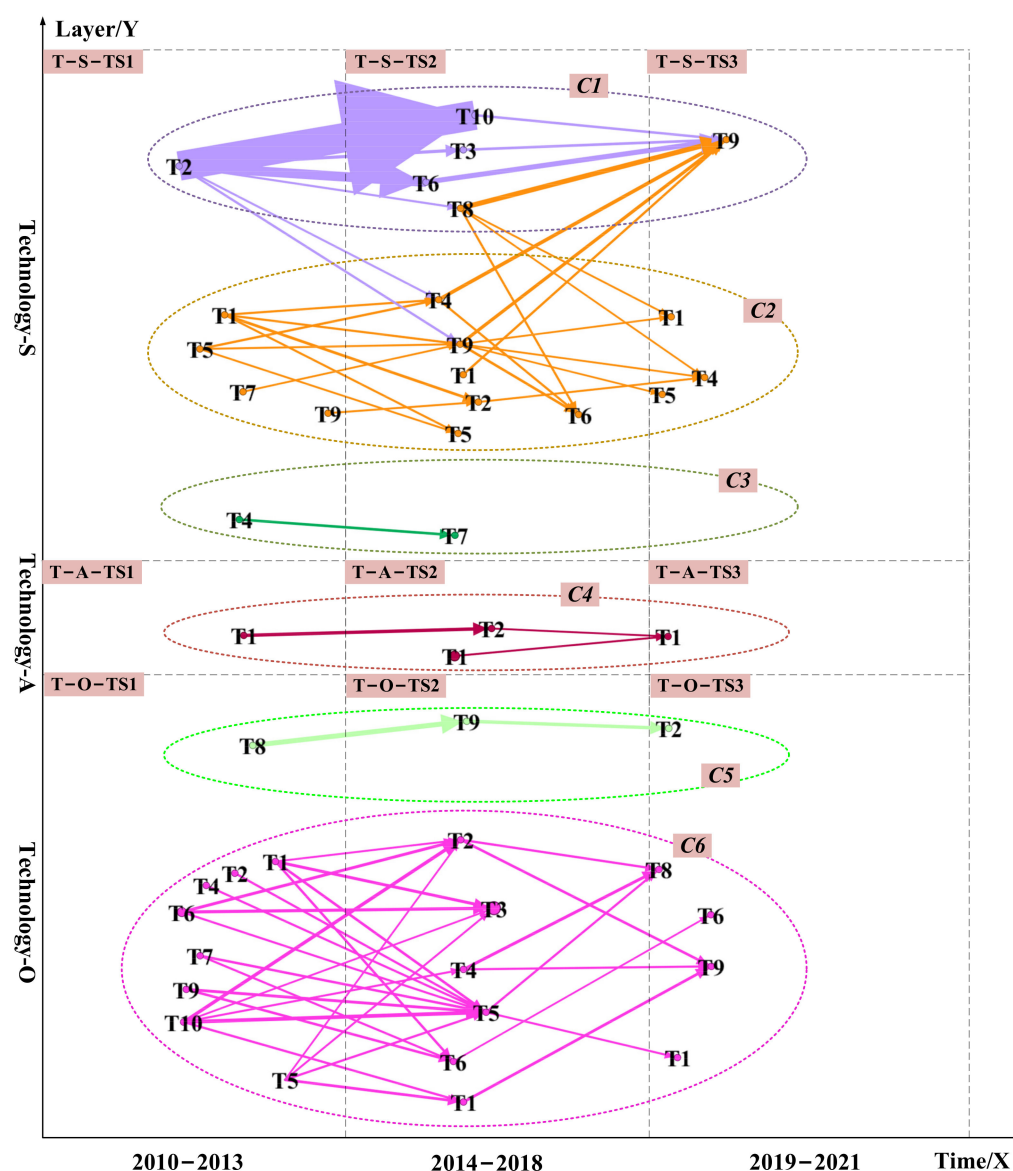


Figure 3. The data-driven TRM for the technology layer. The technical layer was divided into six communities represented by C_1 , C_2 , C_3 , C_4 , C_5 , and C_6 . Dashed cycles with different colors highlight diverse communities. The edges' width and arrows' width represent the probability of a potential link between unconnected nodes. The wider the edges and arrows, the higher the likelihood of a potential link. Arrows and edges with different colors represent different community themes.

Second, we select the nodes and edges in the technology and market subnetworks where potential links exist and visualize them in technical and market layers. If there is a potential link between two unlinked topics, the two nodes are linked with a line with arrows. The edges' width and arrows represent the probability of a potential link between unconnected nodes. The wider the edges and arrows, the higher the likelihood of a potential link. We use these connections as possible directions for future technology and market development.

We then divide the link prediction visualization results into different communities in technical and market layers [130,131]. Arrow edges with different colors represent different community themes. Dashed cycles with different colors highlight diverse communities.

Lastly, we analyze possible opportunities for future technology and market development based on the final technology roadmap.

4. Illustrative Example

4.1. Collecting and Pre-Processing Data for Technology and Market

4.1.1. Data Collection

The dataset of this study was derived from three distinct databases extracted from Medline, Derwent, and ABI, respectively. We then used the Mesh term ‘MH = (gout OR hyperuricemia)’ as a search query from Medline. We chose the International Patent Classification Number ‘IP = (A61P-019/06)’ as a search query from Derwent. And we selected the keyword “hyperuricemia” as a search query for ABI. The cutoff date was 31 December 2021. Any data beyond that date are not part of this study. Therefore, a set of 6124 hyperuricemia-related essays, a collection of 5158 hyperuricemia-related patent data, and 4582 hyperuricemia-related commercial data were extracted, as shown in Table 1. The study keeps the dataset with abstracts. It includes 5066 literature abstracts, 5158 patent abstracts, and 1447 commercial data. The databases used in this study are presented in Table 1. The literature abstracts, patent abstracts, and commercial data were analyzed for technology opportunities.

Table 1. Database for data-driven TRM.

Type	Data Source	Retrieval Strategy	Count
Paper	Medline	MH = (gout OR hyperuricemia)	6124
Patent	Derwent	IP = (A61P-019/06)	5158
Market	ABI	hyperuricemia	4582

4.1.2. Setting the Timeframe of Data-Driven TRM

This study divides the part of the dataset with abstracts related to hyperuricemia drugs into three sub-periods according to the S-curve, as shown in Figure 4. They are the stable period (2010–2013), the rising period (2014–2018), and the fluctuating period (2019–2021), which are represented by TS₁, TS₂, and TS₃, respectively. This study then analyzes the technology and market trends from period to period.

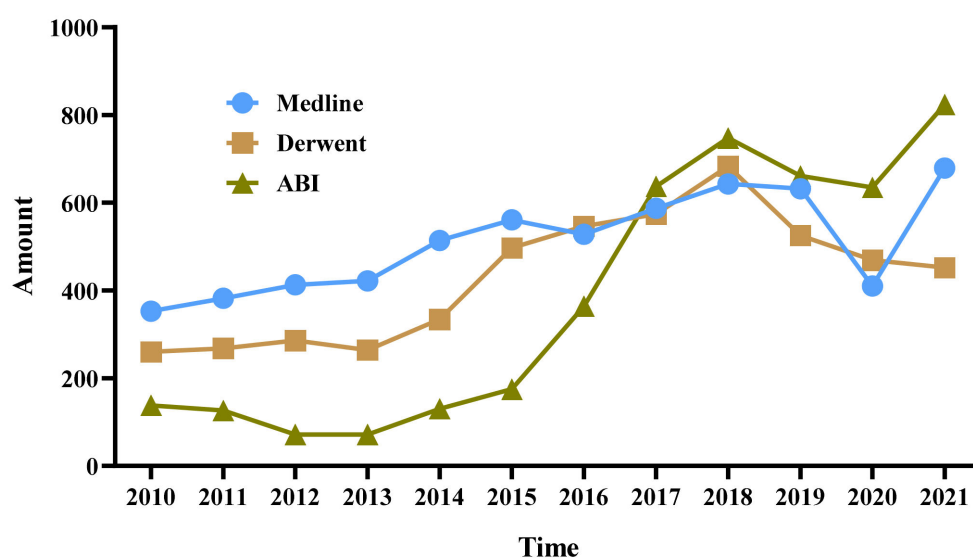


Figure 4. Statistical results of papers, patents, and commercial data related to hyperuricemia drugs.

4.2. Layer Mapping

4.2.1. Classifying the Data into Layers Based on BERT

In the second module, the abstracts are separately extracted from the tech-related and market-related data of TS₁, TS₂, and TS₃ for text analysis. After that, we pre-processed the data extracted from TS₁, TS₂, and TS₃ by dividing the 11,671 abstracts into

85,656 individual sentences with Python's NLTK package. We then conducted a BERT model to classify technical and market data. Even if BERT is beneficial for classifying technical and market data, how to build a training set cannot do without the help of experts. Three experts engaged in introducing hyperuricemia drugs for more than ten years were invited to construct the training set. With the help of the domain experts, we reviewed the data sentence-by-sentence and extracted 30 percent sentence randomly from the entire data set as the training set. The extracted training set is then manually labeled and classified into technical-related, market-related, and irrelevant data, denoted by C_1 and C_2 , and C_0 , respectively.

The BERT model is pre-trained, and each subset (TS_1 , TS_2 , TS_3) was divided into three categories. This study calls for technical and market-related data in developing a data-driven technology roadmap. The irrelevant data in C_0 is useless to our research [15]. Thus, after classification, we only kept 54,026 tech-related sentences in C_1 and 31,630 market-related sentences in C_2 . Finally, the whole dataset is divided into six subsets and classified into different layers of data-driven TRM. C_1 in TS_1 means tech-related data in 2010–2013, C_1 in TS_2 means tech-related data in 2014–2018, C_1 in TS_3 means tech-related data in 2019–2021, C_2 in TS_1 means market-related data in 2010–2013, C_2 in TS_2 means market-related data in 2014–2018, and C_2 in TS_3 means market-related data in 2019–2021.

4.2.2. Semantic Analysis for the Technology Layer and Market Layer Based on SAO

After classifying the data into different layers of data-driven TRM, we chose SAO to extract technical and market semantic structures, reflecting the contexts. The SAO structures are retrieved from each sentence employing word annotation. The word annotation is performed using the Stanford parser, employed by many investigators. In reliance on the Stanford parser, 54,026 technology SAO structures and 31,630 market SAO structures were extracted in the timeframe. The original SAOs are heavily redundant. To analyze them efficiently, we filtered, cleaned, and combined the SAOs. We then deleted 1022 duplicates of technology SAO structures and 1681 market structures. A total of 41,826 unique technology SAO structures and 10,621 unique market SAO structures remained. We are removing the 35,483 invalid technology SAO records and 9545 market SAO records with no subjects, actions, or objects. There were 6343 technology structures and 1076 market structures left, as shown in Tables 2 and 3.

Table 2. Number of SAOs for the technology layer.

Time Series	Sentences	SAO Components			
		Original	Duplicate	Incomplete	Reserved
2010–2013	12,161	9912	202	8263	1447
2014–2018	24,441	19,902	596	16,408	2898
2019–2021	17,424	13,034	224	10,812	1998
In total	54,026	42,848	1022	35,483	6343

Table 3. Number of SAOs for market layer.

Time Series	Sentences	SAO Components			
		Original	Duplicate	Incomplete	Reserved
2010–2013	4371	2439	219	1960	260
2014–2018	17,799	6315	1035	4767	513
2019–2021	9460	3548	427	2818	303
In total	31,630	12,302	1681	9545	1076

4.3. Contents Mapping

4.3.1. Pre-Processing the SAO Components

In the third module, we extract technology and market SAO components topics in the timeframe. Thus, we divided the 6343 technology SAOs and 1076 market SAOs into

19 subsets, as shown in Table 4. For example, T-S-TS₁ meant the technology S components related data in 2010–2013, and M-S-TS₁ represents the market S components associated data in 2010–2013. To remove the data noise, we pre-processed the 18 sub-datasets group by group, such as word token, stemming, lemmatization, and excluding stop words. In addition to the 972 basic stop-words, we designated 2188 domain-specific stop-words and excluded them from analysis, such as ‘uric acid’, ‘hyperuricemia’, ‘treatment’, etc.

Table 4. The group result of SAO components for the technology and market layer.

Layer	SAO Components	Time Series	Group
Technology	S	2010–2013	T-S-TS ₁
		2014–2018	T-S-TS ₂
		2019–2021	T-S-TS ₃
	A	2010–2013	T-A-TS ₁
		2014–2018	T-A-TS ₂
		2019–2021	T-A-TS ₃
	O	2010–2013	T-O-TS ₁
		2014–2018	T-O-TS ₂
		2019–2021	T-O-TS ₃
Market	S	2010–2013	M-S-TS ₁
		2014–2018	M-S-TS ₂
		2019–2021	M-S-TS ₃
	A	2010–2013	M-A-TS ₁
		2014–2018	M-A-TS ₂
		2019–2021	M-A-TS ₃
	O	2010–2013	M-O-TS ₁
		2014–2018	M-O-TS ₂
		2019–2021	M-O-TS ₃

4.3.2. Identify Topics of SAO Components for Technology and Market Layers Based on BTM

We conducted a BTM-based topic model to identify meaningful core and potential technology and market topics automatically. To determine the optimal number of topics, we evaluate the value of coherence while varying the number of topics. The maximization of the coherence value defines the number of optimal topics. We then calculated the coherence value subset by subset, as shown in Table 5, Figures 5 and 6.

Table 5. Topic coherence value for each sub-set.

Technology			Market		
Group	Number of Topics	Coherence	Group	Number of Topics	Coherence
T-S-TS ₁	9	−45.59	M-S-TS ₁	9	−16.61
T-A-TS ₁	1	12.03	M-A-TS ₁	1	−8.79
T-O-TS ₁	10	−61.45	M-O-TS ₁	7	−2.38
T-S-TS ₂	10	−78.55	M-S-TS ₂	9	−35.95
T-A-TS ₂	2	−30.04	M-A-TS ₂	1	−8.27
T-O-TS ₂	10	−87.32	M-O-TS ₂	2	2.24
T-S-TS ₃	9	−64.03	M-S-TS ₃	10	−24.46
T-A-TS ₃	1	−16.53	M-A-TS ₃	1	−26.92
T-O-TS ₃	9	−70.89	M-O-TS ₃	2	15.48

As a result, 9, 1, 10, 10, 2, 10, 9, 1 and 9 topics were derived for sub-set T-S-TS₁, T-A-TS₁, T-O-TS₁, T-S-TS₂, T-A-TS₂, T-O-TS₂, T-S-TS₃, T-A-TS₃, and T-O-TS₃. 9, 1, 7, 9, 1, 2, 10, 1, and 2 topics were derived for sub-set M-S-TS₁, M-A-TS₁, M-O-TS₁, M-S-TS₂, M-A-TS₂, M-O-TS₂, M-S-TS₃, M-A-TS₃, and M-O-TS₃, respectively. The critical and potential

topics of SAO components for technology and market layers are shown in Supplementary Tables S1–S6. Take Supplementary Table S6 as an example. The topics identified can be defined as O components for the market layer in the time-based framework. There are topics in stage TS₁ such as key amino acid mutations (M-O-TS₁-T₁), reduced side effects of small molecule drugs (M-O-TS₁-T₂), Chinese medicine treatment (M-O-TS₁-T₃), the need to develop drugs with low nephrotoxicity (M-O-TS₁-T₄), the need to develop immunogenic drugs with low (M-O-TS₁-T₅), drug market risk (M-O-TS₁-T₆), improve patients' quality of life (M-O-TS₁-T₇). There are small molecule drug products such as Zurampic and Duzallo (M-O-TS₂-T₁); enterprises need to find key amino acid sites (M-O-TS₂-T₂) topics in stage TS₂. There are reduced drug side effects (M-O-TS₃-T₁) and market demand for drugs with low immunogenicity (M-O-TS₃-T₂) topics in stage TS₃. Identifying SAO component topics at the technical and market layers helps each investigator understand the complex topics in his field and throughout hyperuricemia drug development.

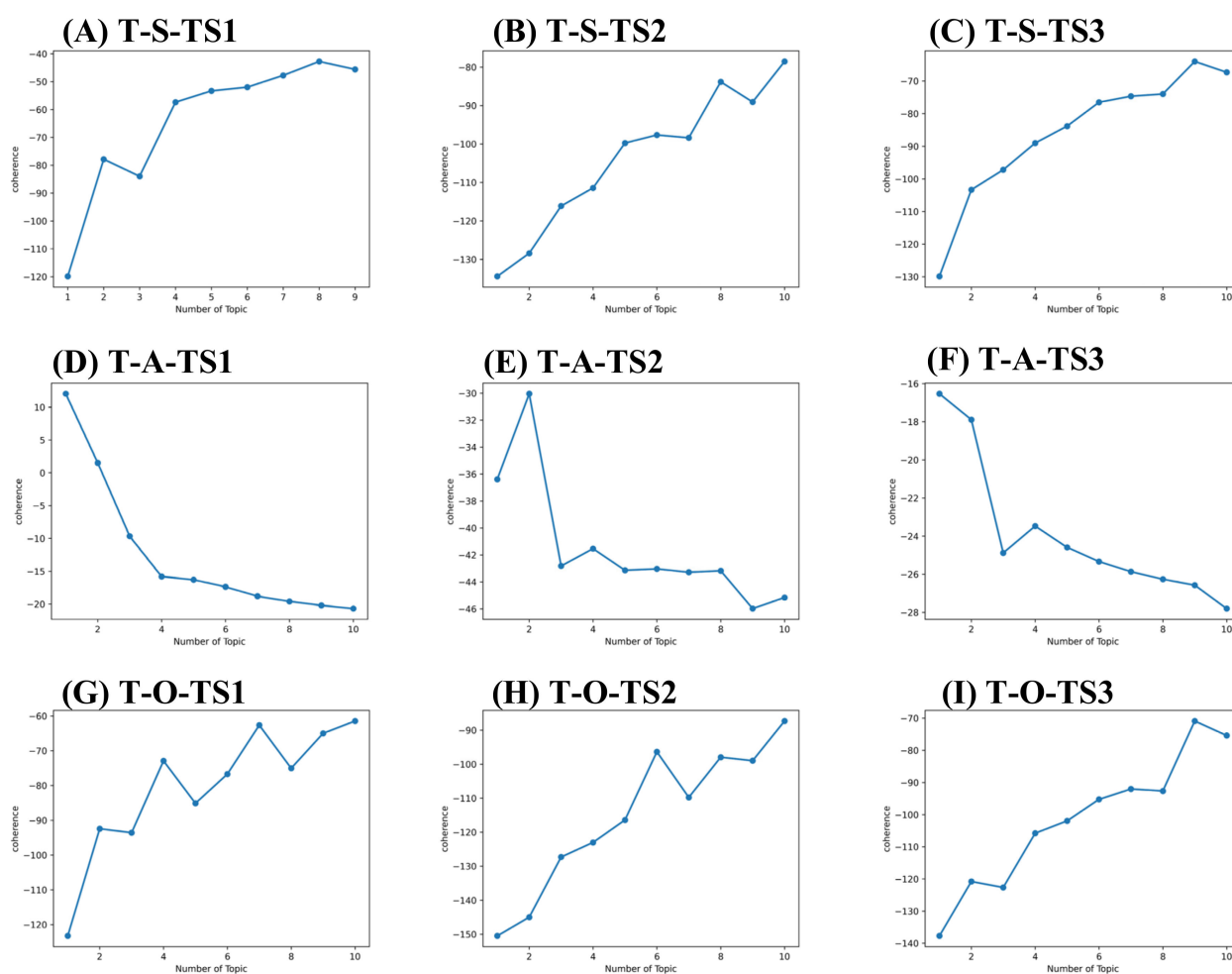


Figure 5. Topic coherence curve for technology layer. (A) Topic coherence value of S components for technology layer in 2010–2013 (T-S-TS1). (B) Topic coherence value of S components for technology layer in 2014–2018 (T-S-TS2). (C) Topic coherence value of S components for technology layer in 2019–2021 (T-S-TS3). (D) Topic coherence value of A components for technology layer in 2010–2013 (T-A-TS1). (E) Topic coherence value of A components for technology layer in 2014–2018 (T-A-TS2). (F) Topic coherence value of A components for technology layer in 2019–2021 (T-A-TS3). (G) Topic coherence value of O components for technology layer in 2010–2013 (T-O-TS1). (H) Topic coherence value of O components for technology layer in 2014–2018 (T-O-TS2). (I) Topic coherence value of O components for technology layer in 2019–2021 (T-O-TS3).

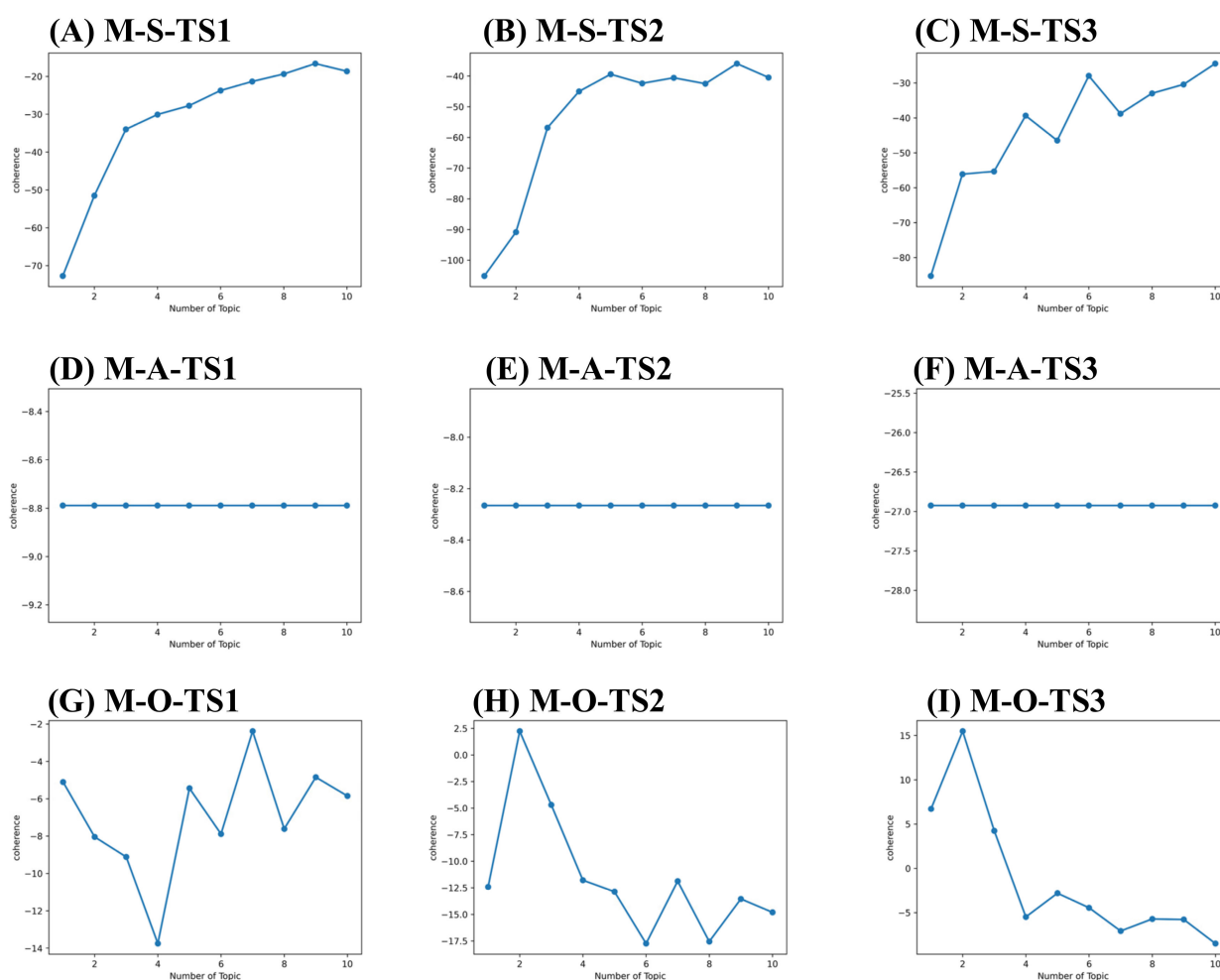


Figure 6. Topic coherence curve for market layer. (A) Topic coherence value of S components for market layer in 2010–2013 (M-S-TS1). (B) Topic coherence value of S components for market layer in 2014–2018 (M-S-TS2). (C) Topic coherence value of S components for market layer in 2019–2021 (M-S-TS3). (D) Topic coherence value of A components for market layer in 2010–2013 (M-A-TS1). (E) Topic coherence value of A components for market layer in 2014–2018 (M-A-TS2). (F) Topic coherence value of A components for market layer in 2019–2021 (M-A-TS3). (G) Topic coherence value of O components for market layer in 2010–2013 (M-O-TS1). (H) Topic coherence value of O components for market layer in 2014–2018 (M-O-TS2). (I) Topic coherence value of O components for market layer in 2019–2021 (M-O-TS3).

4.4. Opportunity Finding

4.4.1. Identify Potential Connections Based on Link Prediction

The last section forecasts possible technical opportunities for hyperuricemia. We constructed the entire network based on link prediction. To build the whole network, we selected the data set pre-processed in Section 4.3.1 to train the model in technology and market layers based on link prediction. The entire technology network consists of 7693 nodes and 17,190 edges. The whole market network consists of 1868 nodes and 2907 edges.

4.4.2. Integrating TRM and Analyzing Opportunities

We select some nodes and edges in the technology and market sub-networks for visualization. Only the nodes and edges that are potentially connected are selected for visualization from the 61 technology topics and 42 market topics extracted in Section 4.3.2.

Next, in the timeframe, we arrange the visualizations in technology and market layers of data-driven TRM from sub-layer S to sub-layer O, as shown in Figures 3 and 7. Finally, we

divided the communities. The technical layer was divided into six communities represented by C_1 , C_2 , C_3 , C_4 , C_5 , and C_6 . The marketing layer was divided into five communities represented by C_7 , C_8 , C_9 , C_{10} , and C_{11} .

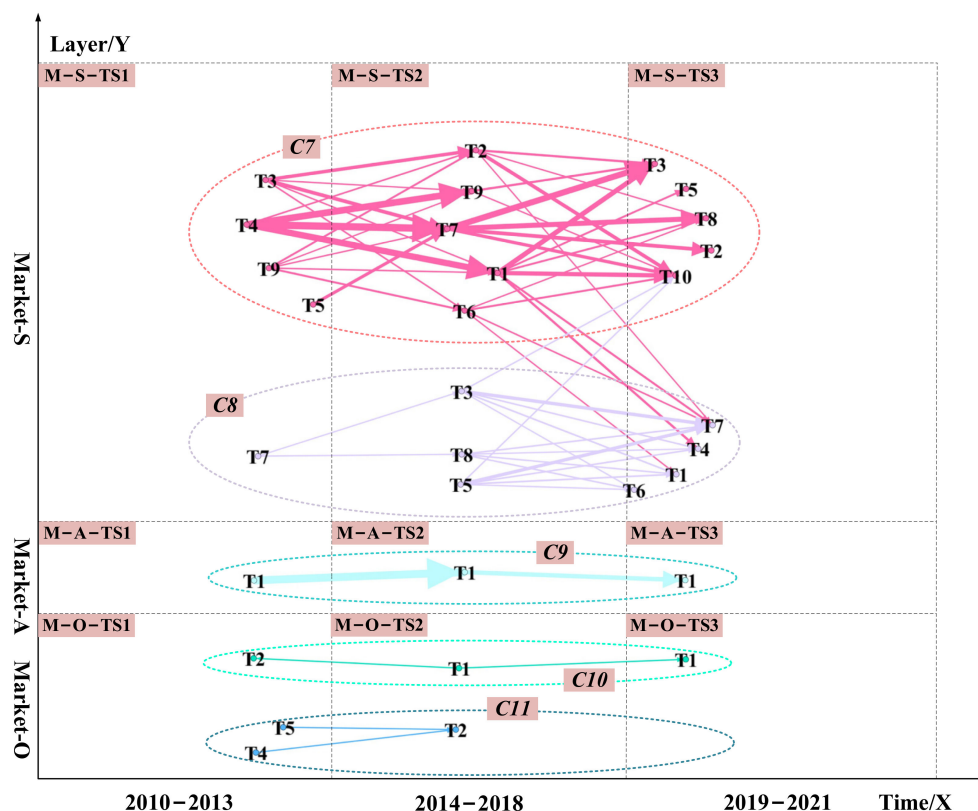


Figure 7. The data-driven TRM for the market layer. The marketing layer was divided into five communities represented by C_7 , C_8 , C_9 , C_{10} , and C_{11} . Dashed cycles with different colors highlight diverse communities. The edges' width and arrows' width represent the probability of a potential link between unconnected nodes. The wider the edges and arrows, the higher the likelihood of a potential link. Arrows and edges with different colors represent different community themes.

Figures 3 and 7 integrate semantic analysis, topic modeling, and link prediction results and show the final technology roadmap. We analyzed the 11 communities in the technology and market layers. Several technological opportunities can be identified. The technical opportunities in the C_1 , C_4 , C_7 , and C_{10} communities are mostly related to chemical-based drugs such as small molecule drugs. The technological opportunities in the C_2 , C_3 , C_6 , C_8 , and C_{11} communities are primarily associated with biological-based medications such as protein drugs.

Take protein drugs related communities as an example. The technology opportunities for hyperuricemia drugs are protein drugs related to the C_2 , C_3 , C_6 , C_8 , and C_{11} communities. From 2010 to 2013, the opportunities focus on how to extend the half-life of protein drugs in patients and how to develop new protein drugs. Protein drug research is mainly based on recombinant *Aspergillus flavus* uricase and PEG-modified urate oxidase derived from pig baboons (T-S-TS₁-T₅ in C_2 , M-S-TS₁-T₇ in C_8). All uric acid oxidases on the market are foreign proteins (M-O-TS₁-T₄ in C_{11} , M-O-TS₁-T₅ in C_{11}). It is easy to produce antibodies after long-term administration of foreign protein. The immunogenicity of uric acid oxidase limits its use, and there are many adverse reactions after entering the market (T-S-TS₁-T₁ in C_2). From 2014 to 2018, it was discovered that active urate oxidase drugs with relatively low immunogenicity could be obtained by “resurrecting” human urate oxidase (M-O-TS₂-T₂ in C_{11}). And it ultimately achieve the purpose of treating hyperuricemia and gout (T-S-TS₂-T₁, T-S-TS₂-T₂, T-S-TS₂-T₄, and T-S-TS₂-T₅ in C_2 , T-A-TS₂-T₁ in

C₄). From 2019 to 2021, while trying to restore human urate oxidase activity, researchers continued to study how to reduce the half-life of existing protein drugs. We can derive urate oxidase from *Aspergillus flavus* (T-S-TS₃-T₁ in C₂, T-S-TS₃-T₄, and T-S-TS₃-T₅ in C₂). How to eliminate or reduce the immunogenicity of existing urate oxidase and obtain active and non-immunogenic human urate oxidase drugs will be the future direction of protein drug development. At these three stages, developers' attention to restoring human uric acid oxidase activity is greater than the market demand. In the future, a comparison of DNA sequences of uric acid oxidase from different sources could be considered to discover the critical amino acid sites that affect the enzyme activity (T-S-TS₂-T₁ in C₂). The essential amino acid sites were then mutated (M-O-TS₂-T₂ in C₂). After each completed mutation, human uric acid oxidase (T-A-TS₂-T₂ in C₄) was induced, and the activity of uric acid oxidase (T-S-TS₃-T₁ in C₂) was assayed after affinity purification. We can resurrect the human uric acid oxidase pseudogene through the above pathway. We can also overcome the disadvantage that existing oxidase drugs are immunogenic and produce antibodies when used for a long time. We can obtain human uric acid oxidase with high drug activity but low immunogenicity to improve the treatment of hyperuricemia and gout.

The technology opportunities for hyperuricemia drugs are small molecule drugs related to the C₁, C₄, C₇, and C₁₀ communities. Similarly, how to reduce the side effects of small molecule drugs will be the future direction of small molecule drug R&D, such as lowering hepatotoxicity and nephrotoxicity (M-O-TS₁-T₂). At these three stages, developers' attention to reducing the side effects of small molecule drugs is lower than the market demand. Reducing the side effects of small molecule drugs is critical in the future. We can try to discover new structures from Chinese medicine or in combination with Chinese medicine, such as combining heat-clearing, dampness-relieving herbs with small molecule drugs (M-S-TS₃-T₁₀ in C₇, M-S-TS₃-T₁₀ in C₃). In addition, it would be good to try to change diet patterns combined with small molecule drug therapy. We can reduce the intake of high purine foods and prevent a eutrophication diet (T-S-TS₂-T₆ in C₁, T-O-TS₃-T₂ in C₅, M-S-TS₂-T₉ in C₇, M-S-TS₃-T₇ in C₈).

5. Discussion

This paper presents a systematic approach to developing a data-driven TRM to identify potential technology opportunities for hyperuricemia drugs. Despite the contributions of previous research, we can extend the existing data-driven TRM from three aspects. These are identifying potential opportunities with data-driven TRM, the process, and the data source of data-driven TRM.

Compared with current trends, we chose link prediction from the perspective of technology forecasting to identify potential technology opportunities for opportunity finding of data-driven TRM. SAO considers existing technology connections, while technology opportunities are often hidden in potential relationships. SAO needs to be combined with link prediction to predict technology opportunities better. Therefore, we identify possible links between unconnected nodes based on link prediction. Based on the results of the link predictions, we should focus on resurrecting the pseudogene of human uric acid oxidase and reducing the toxicity of small molecule drugs in the future.

From the perspective of the data process, compared with keyword-based analysis, we chose SAO to extract critical technology information for layer mapping of data-driven TRM reflecting the contexts. The SAO structure was widely used to analyze documents such as patents, online review demand extraction, and paper. Although SAO is commonly used in TOA, the SAO structures must be refined for efferent analysis. Therefore, this article identifies topics of SAO components for different layers of data-driven TRM based on BTM. It can reduce the redundancy of SAO effectively. Besides that, it can also help to extract which areas the technical solution focuses on, how the solution works, and which parts of the solution are the targets of SAOs. Based on the potential opportunities identified by the link prediction, the realization path of the opportunity is inferred by the SAO structure. For example, it is critical to resurrect the pseudogene of human uric acid oxidase. We could

consider comparing DNA sequences of different sources of uric acid oxidase to discover the critical amino acid site that affects enzyme activity (T-S-TS₂-T₁ in C₂). We would then mutate the essential amino acid site (M-O-TS₂-T₂ in C₂) and induce the expression of uric acid oxidase (T-A-TS₂-T₂ in C₄) affinity purification to analyze the enzyme activity (T-S-TS₃-T₁ in C₂).

From the perspective of data source selection, compared with selecting patents as technical data and commercial reports as market data, we choose patent, literature, and commercial reports as technical and market data of data-driven TRM. We distinguish technology and market data from the vast amount of literature, patent, and commercial data automatically based on BERT and combined with the small domain training set to train BERT models to classify hyperuricemia drugs with high accuracy. To identify potential technology opportunities in concert with market demand, it is necessary to analyze technology opportunities from both market and technology perspectives.

6. Conclusions

It is essential to assist hyperuricemia developers via a data-driven TRM for TOA. However, less attention has been paid to integrating multiple analytical tools within a data-driven TRM to identify potential technology opportunities automatically, such as SAO, BTM, and link prediction. This study extends the existing data-driven TRM from several aspects to fill this gap. There are data process, technology forecasting, and data source selection. And we try to respond to the following questions in this study. First, we illustrate how to build a semantic-based data-driven TRM and identify topics of SAO components based on BTM. Second, we point out how to identify potential technology opportunity points through link prediction. Last, we demonstrate how to extract technical and market information automatically based on text classification tools from various patents, literature, and business data. Therefore, this research provides an attractive option for hyperuricemia drugs TOA. It is critical to narrow down the technology research topics, reducing R&D risks, and supporting the decision-making of hyperuricemia drug research.

Despite the promise of data-driven TRM in hyperuricemia drugs' TOA, challenges have remained in one aspect that can be addressed by future research. First, the SAO structure is complex. It is challenging for machine learning models to identify semantic relations from complex sentences. In future research, we will focus on exploring how to extract SAO structures from complex sentences in the future. Second, patent, literature, and commercial data are not real-time data because of the time lag. It is difficult to obtain the latest technology and market information for TOA. In the future, we will consider developing a dynamic data-driven TRM, such as by adopting dynamic topic models for the topic analysis of the SAO structure. Finally, this paper selects patent, technical and commercial data for data source selection. In the future, more diversified data sources, such as online reviews, drug instructions, etc., will be selected to complement the existing analysis effectively.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ph15111357/s1>, Table S1. Topic result of S components for technology layer in the time-based framework; Table S2. Topic result of A components for technology layer in the time-based framework; Table S3. Topic result of O components for technology layer in the time-based framework; Table S4. Topic result of S components for market layer in the time-based framework; Table S5. Topic result of A components for market layer in the time-based framework; Table S6. Topic result of O components for market layer in the time-based framework.

Author Contributions: Data curation, W.Z.; Formal analysis, W.Z.; Investigation, W.Z.; Methodology, W.Z.; Resources, L.F., J.W. and Y.G.; Software, W.Z.; Supervision, K.-Y.L., Y.G. and L.Z.; Validation, W.Z.; Visualization, W.Z.; Writing—original draft, W.Z.; Writing—review & editing, W.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The authors gratefully acknowledge support from the Innovation Method Fund of China (Project No. 2018IM020300, 2019IM020200), Shanghai Science and Technology Program (Project No. 20040501300), and the China-National Natural Science Foundations (Project No. U1604187, 62173253). National Key Research and Development Program, grant number 2022YFF0608700.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article and Supplementary Material.

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

References

1. Feng, X.; Yang, Y.; Xie, H.; Zhuang, S.; Fang, Y.; Dai, Y.; Jiang, P.; Chen, H.; Tang, H.; Tang, L. The Association Between Hyperuricemia and Obesity Metabolic Phenotypes in Chinese General Population: A Retrospective Analysis. *Front. Nutr.* **2022**, *9*, 773220. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Al-Amodi, Y.A.; Hosny, K.M.; Alharbi, W.S.; Safo, M.K.; El-Say, K. Investigating the Potential of Transmucosal Delivery of Febuxostat from Oral Lyophilized Tablets Loaded with a Self-Nanoemulsifying Delivery System. *Pharmaceutics* **2020**, *12*, 534. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Galindo, T.; Reyna, J.; Weyer, A. Evidence for Transient Receptor Potential (TRP) Channel Contribution to Arthritis Pain and Pathogenesis. *Pharmaceutics* **2018**, *11*, 105. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Tátrai, P.; Erdő, F.; Dörnyei, G.; Krajcsi, P. Modulation of Urate Transport by Drugs. *Pharmaceutics* **2021**, *13*, 899. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Yang, B.; Kwon, I. Thermostable and Long-Circulating Albumin-Conjugated *Arthrobacter globiformis* Urate Oxidase. *Pharmaceutics* **2021**, *13*, 1298. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Da Cruz, R.M.D.; Mendonca, F.J.B.; De Melo, N.B.; Scotti, L.; De Araujo, R.S.A.; De Almeida, R.N.; De Moura, R.O. Thiophene-Based Compounds with Potential Anti-Inflammatory Activity. *Pharmaceutics* **2021**, *14*, 692. [\[CrossRef\]](#)
7. Kanbay, M.; Jensen, T.; Solak, Y.; Le, M.; Roncal-Jimenez, C.; Rivard, C.; Lanaspá, M.A.; Nakagawa, T.; Johnson, R.J. Uric acid in metabolic syndrome: From an innocent bystander to a central player. *Eur. J. Intern. Med.* **2016**, *29*, 3–8. [\[CrossRef\]](#)
8. Cong, R.; Zhang, X.; Song, Z.; Chen, S.; Liu, G.; Liu, Y.; Pang, X.; Dong, F.; Xing, W.; Wang, Y.; et al. Assessing the Causal Effects of Adipokines on Uric Acid and Gout: A Two-Sample Mendelian Randomization Study. *Nutrients* **2022**, *14*, 1091. [\[CrossRef\]](#)
9. Ye, C.; Huang, X.; Wang, R.; Halimulati, M.; Aihemaitijiang, S.; Zhang, Z. Dietary Inflammatory Index and the Risk of Hyperuricemia: A Cross-Sectional Study in Chinese Adult Residents. *Nutrients* **2021**, *13*, 4504. [\[CrossRef\]](#)
10. Shi, J.; He, L.; Yu, D.; Ju, L.; Guo, Q.; Piao, W.; Xu, X.; Zhao, L.; Yuan, X.; Cao, Q.; et al. Prevalence and Correlates of Metabolic Syndrome and Its Components in Chinese Children and Adolescents Aged 7–17: The China National Nutrition and Health Survey of Children and Lactating Mothers from 2016–2017. *Nutrients* **2022**, *14*, 3348. [\[CrossRef\]](#)
11. Chen, H.-W.; Chen, Y.-C.; Lee, J.-T.; Yang, F.M.; Kao, C.-Y.; Chou, Y.-H.; Chu, T.-Y.; Juan, Y.-S.; Wu, W.-J. Prediction of the Uric Acid Component in Nephrolithiasis Using Simple Clinical Information about Metabolic Disorder and Obesity: A Machine Learning-Based Model. *Nutrients* **2022**, *14*, 1829. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Toyoda, Y.; Takada, T.; Saito, H.; Hirata, H.; Ota-Kontani, A.; Tsuchiya, Y.; Suzuki, H. Identification of Inhibitory Activities of Dietary Flavonoids against URAT1, a Renal Urate Re-Absorber: In Vitro Screening and Fractional Approach Focused on Rooibos Leaves. *Nutrients* **2022**, *14*, 575. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Fahmy, U.; Aldawsari, H.; Badr-Eldin, S.; Ahmed, O.; Alhakamy, N.; Alsulimani, H.; Caraci, F.; Caruso, G. The Encapsulation of Febuxostat into Emulsomes Strongly Enhances the Cytotoxic Potential of the Drug on HCT 116 Colon Cancer Cells. *Pharmaceutics* **2020**, *12*, 956. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Duda, G.N.; Grainger, D.W.; Frisk, M.L.; Bruckner-Tuderman, L.; Carr, A.; Dirnagl, U.; Einhüpl, K.M.; Gottschalk, S.; Gruskin, E.; Huber, C.; et al. Changing the Mindset in Life Sciences Toward Translation: A Consensus. *Sci. Transl. Med.* **2014**, *6*, 264cm12. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Kim, J.; Geum, Y. How to develop data-driven technology roadmaps: The integration of topic modeling and link prediction. *Technol. Forecast. Soc. Chang.* **2021**, *171*, 120972. [\[CrossRef\]](#)
16. Lara, R.A.N.; Beltrán, J.A.; Brizuela, C.A.; Del Rio, G. Relevant Features of Polypharmacologic Human-Target Antimicrobials Discovered by Machine-Learning Techniques. *Pharmaceutics* **2020**, *13*, 204. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Luo, L.X.; Zheng, T.Y.; Wang, Q.; Liao, Y.L.; Zheng, X.Q.; Zhong, A.; Huang, Z.N.; Luo, H. Virtual Screening Based on Machine Learning Explores Mangrove Natural Products as KRAS(G12C) Inhibitors. *Pharmaceutics* **2022**, *15*, 584. [\[CrossRef\]](#)
18. Walsh, E.I.; Cherbuin, N. Mapping the Literature on Nutritional Interventions in Cognitive Health: A Data-Driven Approach. *Nutrients* **2018**, *11*, 38. [\[CrossRef\]](#)
19. Janssen, A.; Bennis, F.C.; Mathôt, R.A.A. Adoption of Machine Learning in Pharmacometrics: An Overview of Recent Implementations and Their Considerations. *Pharmaceutics* **2022**, *14*, 1814. [\[CrossRef\]](#)
20. Ko, Y.K.; Gim, J.-A. New Drug Development and Clinical Trial Design by Applying Genomic Information Management. *Pharmaceutics* **2022**, *14*, 1539. [\[CrossRef\]](#)

21. Raza, M.S.; Khahro, S.H.; Memon, S.A.; Ali, T.H.; Memon, N.A. Global trends in research on carbon footprint of buildings during 1971–2021: A bibliometric investigation. *Environ. Sci. Pollut. Res.* **2021**, *28*, 63227–63236. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Du, J.; Li, P.; Guo, Q.; Tang, X. Measuring the knowledge translation and convergence in pharmaceutical innovation by funding-science-technology-innovation linkages analysis. *J. Inf.* **2018**, *13*, 132–148. [\[CrossRef\]](#)
23. Ramos, A.G.; Daim, T.; Gaats, L.; Hutmacher, D.W.; Hackenberger, D. Technology roadmap for the development of a 3D cell culture workstation for a biomedical industry startup. *Technol. Forecast. Soc. Chang.* **2021**, *174*, 121213. [\[CrossRef\]](#)
24. Cheng, X.; Yan, X.; Lan, Y.; Guo, J. BTM: Topic Modeling over Short Texts. *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 2928–2941. [\[CrossRef\]](#)
25. Kang, H.; Goo, S.; Lee, H.; Chae, J.-W.; Yun, H.-Y.; Jung, S. Fine-tuning of BERT Model to Accurately Predict Drug–Target Interactions. *Pharmaceutics* **2022**, *14*, 1710. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Guven, Z.A.; Unalir, M.O. Natural language based analysis of SQuAD: An analytical approach for BERT. *Expert Syst. Appl.* **2022**, *195*, 116592. [\[CrossRef\]](#)
27. Jang, H.; Park, S.; Yoon, B. Exploring Technology Opportunities Based on User Needs: Application of Opinion Mining and SAO Analysis. *Eng. Manag. J.* **2022**, 1–14. [\[CrossRef\]](#)
28. McCoy, K.; Gudapati, S.; He, L.; Horlander, E.; Kartchner, D.; Kulkarni, S.; Mehra, N.; Prakash, J.; Thenot, H.; Vanga, S.; et al. Biomedical Text Link Prediction for Drug Discovery: A Case Study with COVID-19. *Pharmaceutics* **2021**, *13*, 794. [\[CrossRef\]](#)
29. Nuñez de Villavicencio-Díaz, T.; Rabalski, A.J.; Litchfield, D.W. Protein Kinase CK2: Intricate Relationships within Regulatory Cellular Networks. *Pharmaceutics* **2017**, *10*, 27. [\[CrossRef\]](#)
30. Russo, S.; Bonassi, S. Prospects and Pitfalls of Machine Learning in Nutritional Epidemiology. *Nutrients* **2022**, *14*, 1705. [\[CrossRef\]](#)
31. Otani, K.; Kanno, K.; Akutsu, T.; Ohdaira, H.; Suzuki, Y.; Urashima, M. Applying Machine Learning to Determine 25(OH)D Threshold Levels Using Data from the AMATERASU Vitamin D Supplementation Trial in Patients with Digestive Tract Cancer. *Nutrients* **2022**, *14*, 1689. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Hamed, A.A.; Fandy, T.E.; Tkaczuk, K.L.; Verspoor, K.; Lee, B.S. COVID-19 Drug Repurposing: A Network-Based Framework for Exploring Biomedical Literature and Clinical Trials for Possible Treatments. *Pharmaceutics* **2022**, *14*, 567. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Keutzer, L.; You, H.; Farnoud, A.; Nyberg, J.; Wicha, S.G.; Maher-Edwards, G.; Vlasakakis, G.; Moghaddam, G.K.; Svensson, E.M.; Menden, M.P.; et al. Machine Learning and Pharmacometrics for Prediction of Pharmacokinetic Data: Differences, Similarities and Challenges Illustrated with Rifampicin. *Pharmaceutics* **2022**, *14*, 1530. [\[CrossRef\]](#)
34. Liu, Z.; Wang, Y.; Feng, J. Vehicle-type strategies for manufacturer’s car sharing. *Kybernetes* **2022**, ahead-of-print. [\[CrossRef\]](#)
35. Meqbil, Y.J.; Rijn, R.M.V. Opportunities and Challenges for In Silico Drug Discovery at Delta Opioid Receptors. *Pharmaceutics* **2022**, *15*, 873. [\[CrossRef\]](#)
36. Jia, X.; Chen, J.; Tang, X. Construction of Technology Road-Map(TRM) Specific to Pharmaceutical Industry. *Sci. Technol. Manag. Res.* **2018**, *11*, 128–133.
37. Zhou, X.; Huang, L.; Porter, A.; Vicente-Gomila, J.M. Tracing the system transformations and innovation pathways of an emerging technology: Solid lipid nanoparticles. *Technol. Forecast. Soc. Chang.* **2019**, *146*, 785–794. [\[CrossRef\]](#)
38. Wang, L.-Y.; Zhao, D. Cross-domain function analysis and trend study in Chinese construction industry based on patent semantic analysis. *Technol. Forecast. Soc. Chang.* **2020**, *162*, 120331. [\[CrossRef\]](#)
39. Han, M.; Geum, Y. Roadmapping for Data: Concept and Typology of Data-Integrated Smart-Service Roadmaps. *IEEE Trans. Eng. Manag.* **2020**, *69*, 142–154. [\[CrossRef\]](#)
40. Borschiver, S.; Vasconcelos, R.C.; Silva, F.C.; Freitas, G.C.; Santos, P.E.; Bomfim, R.O.D. Technology roadmap for hyaluronic acid and its derivatives market. *Biofuels Bioprod. Biorefining* **2018**, *13*, 435–444. [\[CrossRef\]](#)
41. Rincón-López, J.; Almanza-Arjona, Y.C.; Riascos, A.P.; Rojas-Aguirre, Y. When Cyclodextrins Met Data Science: Unveiling Their Pharmaceutical Applications through Network Science and Text-Mining. *Pharmaceutics* **2021**, *13*, 1297. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Yu, X.; Zhang, B. Obtaining advantages from technology revolution: A patent roadmap for competition analysis and strategy planning. *Technol. Forecast. Soc. Chang.* **2019**, *145*, 273–283. [\[CrossRef\]](#)
43. Han, X.; Zhu, D.; Wang, X.; Li, J.; Qiao, Y. Technology Opportunity Analysis: Combining SAO Networks and Link Prediction. *IEEE Trans. Eng. Manag.* **2019**, *68*, 1288–1298. [\[CrossRef\]](#)
44. Ma, T.T.; Zhou, X.; Liu, J.; Lou, Z.K.; Hua, Z.T.; Wang, R.T. Combining topic modeling and SAO semantic analysis to identify technological opportunities of emerging technologies. *Technol. Forecast. Soc. Chang.* **2021**, *173*, 121159. [\[CrossRef\]](#)
45. Ma, T.T.; Porter, A.L.; Guo, Y.; Ready, J.; Xu, C.; Gao, L.D. A technology opportunities analysis model: Applied to dye-sensitised solar cells for China. *Technol. Anal. Strateg. Manag.* **2014**, *26*, 87–104. [\[CrossRef\]](#)
46. Shibata, N.; Kajikawa, Y.; Takeda, Y.; Matsushima, K. Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation* **2008**, *28*, 758–775. [\[CrossRef\]](#)
47. Kajikawa, Y.; Yoshikawa, J.; Takeda, Y.; Matsushima, K. Tracking emerging technologies in energy research: Toward a roadmap for sustainable energy. *Technol. Forecast. Soc. Chang.* **2008**, *75*, 771–782. [\[CrossRef\]](#)
48. Lee, P.-C.; Su, H.-N.; Wu, F.-S. Quantitative mapping of patented technology—The case of electrical conducting polymer nanocomposite. *Technol. Forecast. Soc. Chang.* **2010**, *77*, 466–478. [\[CrossRef\]](#)
49. Jeong, Y.; Park, I.; Yoon, B. Identifying emerging Research and Business Development (R&BD) areas based on topic modeling and visualization with intellectual property right data. *Technol. Forecast. Soc. Chang.* **2019**, *146*, 655–672. [\[CrossRef\]](#)

50. Yoon, B.; Magee, C.L. Exploring technology opportunities by visualizing patent information based on generative topographic mapping and link prediction. *Technol. Forecast. Soc. Chang.* **2018**, *132*, 105–117. [\[CrossRef\]](#)
51. Cheng, Y.-Y.; Qu, H.-B.; Zhang, B.-L. Chinese medicine industry 4.0: Advancing digital pharmaceutical manufacture toward intelligent pharmaceutical manufacture. *China J. Chin. Mater. Med.* **2016**, *41*, 1–5.
52. Zhang, Y.; Zhang, G.; Chen, H.; Porter, A.L.; Zhu, D.; Lu, J. Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research. *Technol. Forecast. Soc. Chang.* **2016**, *105*, 179–191. [\[CrossRef\]](#)
53. Pavlinek, M.; Podgorelec, V. Text classification method based on self-training and LDA topic models. *Expert Syst. Appl.* **2017**, *80*, 83–93. [\[CrossRef\]](#)
54. Zhang, Y.; Robinson, D.K.; Porter, A.L.; Zhu, D.; Zhang, G.; Lu, J. Technology roadmapping for competitive technical intelligence. *Technol. Forecast. Soc. Chang.* **2016**, *110*, 175–186. [\[CrossRef\]](#)
55. Wu, Q.; Zhang, C.; Hong, Q.; Chen, L. Topic evolution based on LDA and HMM and its application in stem cell research. *J. Inf. Sci.* **2014**, *40*, 611–620. [\[CrossRef\]](#)
56. Aaldering, L.J.; Song, C.H. Tracing the technological development trajectory in post-lithium-ion battery technologies: A patent-based approach. *J. Clean. Prod.* **2019**, *241*, 118343. [\[CrossRef\]](#)
57. Wang, Z.X.; Peng, X.G.; Xia, A.; Shah, A.A.; Huang, Y.; Zhu, X.Q.; Zhu, X.; Liao, Q. The role of machine learning to boost the bioenergy and biofuels conversion. *Bioresour. Technol.* **2022**, *343*, 126099. [\[CrossRef\]](#)
58. Son, W.; Lee, S. Integrating fuzzy-set theory into technology roadmap development to support decision-making. *Technol. Anal. Strat. Manag.* **2016**, *31*, 447–461. [\[CrossRef\]](#)
59. Cho, Y.; Yoon, S.-P.; Kim, K.-S. An industrial technology roadmap for supporting public R&D planning. *Technol. Forecast. Soc. Chang.* **2016**, *107*, 1–12. [\[CrossRef\]](#)
60. Valerio, K.G.D.; Da Silva, C.E.S.; Neves, S.M. Overview on the technology roadmapping (TRM) literature: Gaps and perspectives. *Technol. Anal. Strateg. Manag.* **2020**, *1*, 1–12.
61. Milshina, Y.; Vishnevskiy, K. Roadmapping in fast changing environments—The case of the Russian media industry. *J. Eng. Technol. Manag.* **2019**, *52*, 32–47. [\[CrossRef\]](#)
62. Phaal, R.; Farrukh, C.; Probert, D. Characterisation of technology roadmaps: Purpose and format. In Proceedings of the PICMET '01. Portland International Conference on Management of Engineering and Technology. Proceedings Vol.1: Book of Summaries (IEEE Cat. No.01CH37199), Portland, OR, USA, 29 July–2 August 2001. [\[CrossRef\]](#)
63. Ghazinoory, S.; Dastranj, N.; Saghafi, F.; Kulshreshtha, A.; Hasanzadeh, A. Technology roadmapping architecture based on technological learning: Case study of social banking in Iran. *Technol. Forecast. Soc. Chang.* **2017**, *122*, 231–242. [\[CrossRef\]](#)
64. Lee, J.H.; Phaal, R.; Lee, S.-H. An integrated service-device-technology roadmap for smart city development. *Technol. Forecast. Soc. Chang.* **2013**, *80*, 286–306. [\[CrossRef\]](#)
65. Wang, J.; Li, K.; Feng, L. Tracing the technological trajectory of coal slurry pipeline transportation technology: An HMM-based topic modeling approach. *Front. Energy Res.* **2022**, *10*, 1303. [\[CrossRef\]](#)
66. Kerr, C.; Phaal, R.; Thams, K. Customising and deploying roadmapping in an organisational setting: The LEGO Group experience. *J. Eng. Technol. Manag.* **2019**, *52*, 48–60. [\[CrossRef\]](#)
67. Pearson, R.; Costley, A.; Phaal, R.; Nuttall, W. Technology Roadmapping for mission-led agile hardware development: A case study of a commercial fusion energy start-up. *Technol. Forecast. Soc. Chang.* **2020**, *158*, 120064. [\[CrossRef\]](#)
68. Aleina, S.C.; Viola, N.; Fusaro, R.; Longo, J.; Saccoccia, G. Basis for a methodology for roadmaps generation for hypersonic and re-entry space transportation systems. *Technol. Forecast. Soc. Chang.* **2018**, *128*, 208–225. [\[CrossRef\]](#)
69. Daim, T.U.; Yoon, B.-S.; Lindenberg, J.; Grizzi, R.; Estep, J.; Oliver, T. Strategic roadmapping of robotics technologies for the power industry: A multicriteria technology assessment. *Technol. Forecast. Soc. Chang.* **2018**, *131*, 49–66. [\[CrossRef\]](#)
70. Tierney, R.; Hermina, W.; Walsh, S. The pharmaceutical technology landscape: A new form of technology roadmapping. *Technol. Forecast. Soc. Chang.* **2013**, *80*, 194–211. [\[CrossRef\]](#)
71. Jeong, Y.; Yoon, B. Development of patent roadmap based on technology roadmap by analyzing patterns of patent development. *Technovation* **2015**, *39–40*, 37–52. [\[CrossRef\]](#)
72. De Alcantara, D.P.; Martens, M.L. Technology Roadmapping (TRM): A systematic review of the literature. *Technol. Forecast. Soc. Chang.* **2019**, *138*, 127–238. [\[CrossRef\]](#)
73. Pereira, C.G.; Lavoie, J.R.; Garces, E.; Basso, F.; Dabić, M.; Porto, G.S.; Daim, T. Forecasting of emerging therapeutic monoclonal antibodies patents based on a decision model. *Technol. Forecast. Soc. Chang.* **2019**, *139*, 185–199. [\[CrossRef\]](#)
74. Martin, H.; Daim, T.U. Technology roadmap development process (TRDP) for the service sector: A conceptual framework. *Technol. Soc.* **2012**, *34*, 94–105. [\[CrossRef\]](#)
75. Willyard, C.H.; McClees, C.W. Motorola's Technology Roadmap Process. *Res. Manag.* **1987**, *30*, 13–19. [\[CrossRef\]](#)
76. Lee, H.; Geum, Y. Development of the scenario-based technology roadmap considering layer heterogeneity: An approach using CIA and AHP. *Technol. Forecast. Soc. Chang.* **2017**, *117*, 12–24. [\[CrossRef\]](#)
77. Wang, Y.-H.; Lin, G.-Y. Exploring AI-healthcare innovation: Natural language processing-based patents analysis for technology-driven roadmapping. *Kybernetes* **2022**. [\[CrossRef\]](#)
78. Jeong, Y.; Jang, H.; Yoon, B. Developing a risk-adaptive technology roadmap using a Bayesian network and topic modeling under deep uncertainty. *Scientometrics* **2021**, *126*, 3697–3722. [\[CrossRef\]](#)

79. Byungun, Y.; Robert, P. Structuring technological information for technology roadmapping: Data mining approach. *Technol. Anal. Strateg. Manag.* **2013**, *25*, 1119–1137.
80. Geum, Y.; Lee, H.; Lee, Y.; Park, Y. Development of data-driven technology roadmap considering dependency: An ARM-based technology roadmapping. *Technol. Forecast. Soc. Chang.* **2015**, *91*, 264–279. [\[CrossRef\]](#)
81. Zhang, H.; Daim, T.; Zhang, Y.P. Integrating patent analysis into technology roadmapping: A latent dirichlet allocation based technology assessment and roadmapping in the field of Blockchain-ScienceDirect. *Technol. Forecast. Soc. Chang.* **2021**, *167*, 120729. [\[CrossRef\]](#)
82. Tan, M.S.; Cheah, P.L.; Chin, A.; Looi, L.M.; Chang, S.W. A review on omics-based biomarkers discovery for Alzheimer's disease from the bioinformatics perspectives: Statistical approach vs. machine learning approach. *Comput. Biol. Med.* **2021**, *139*, 104947. [\[CrossRef\]](#) [\[PubMed\]](#)
83. Lu, K.; Yang, G.; Wang, X. Topics emerged in the biomedical field and their characteristics. *Technol. Forecast. Soc. Chang.* **2022**, *174*, 121218. [\[CrossRef\]](#)
84. Wang, X.; Ma, P.; Huang, Y.; Guo, J.; Zhu, D.; Porter, A.L.; Wang, Z. Combining SAO semantic analysis and morphology analysis to identify technology opportunities. *Scientometrics* **2017**, *111*, 3–24. [\[CrossRef\]](#)
85. Wang, X.; Qiu, P.; Zhu, D.; Mitkova, L.; Lei, M.; Porter, A.L. Identification of technology development trends based on subject-action-object analysis: The case of dye-sensitized solar cells. *Technol. Forecast. Soc. Chang.* **2015**, *98*, 24–46. [\[CrossRef\]](#)
86. Liu, B.; Lai, M.; Wu, J.-L.; Fu, C.; Binaykia, A. Patent analysis and classification prediction of biomedicine industry: SOM-KPCA-SVM model. *Multimedia Tools Appl.* **2020**, *79*, 10177–10197. [\[CrossRef\]](#)
87. Gomez, J.C. Analysis of the effect of data properties in automated patent classification. *Scientometrics* **2019**, *121*, 1239–1268. [\[CrossRef\]](#)
88. Yun, J.; Geum, Y. Automated classification of patents: A topic modeling approach. *Comput. Ind. Eng.* **2020**, *147*, 106636. [\[CrossRef\]](#)
89. Yang, W.; Chen, L. Machine condition recognition via hidden semi-Markov model. *Comput. Ind. Eng.* **2021**, *158*, 107430. [\[CrossRef\]](#)
90. Ercan, S.; Kayakutlu, G. Patent value analysis using support vector machines. *Soft Comput.* **2014**, *18*, 313–328. [\[CrossRef\]](#)
91. Wen, G.; Chen, H.; Li, H.; Hu, Y.; Li, Y.; Wang, C. Cross domains adversarial learning for Chinese named entity recognition for online medical consultation. *J. Biomed. Inform.* **2020**, *112*, 103608. [\[CrossRef\]](#)
92. Kumar, A.; Gupta, P.; Balan, R.; Neti, L.B.M.; Malapati, A. BERT Based Semi-Supervised Hybrid Approach for Aspect and Sentiment Classification. *Neural Process. Lett.* **2021**, *53*, 4207–4224. [\[CrossRef\]](#)
93. Taju, S.W.; Shah, S.M.A.; Ou, Y.-Y. ActTRANS: Functional classification in active transport proteins based on transfer learning and contextual representations. *Comput. Biol. Chem.* **2021**, *93*, 107537. [\[CrossRef\]](#) [\[PubMed\]](#)
94. Tan, X.; Zhuang, M.; Lu, X.; Mao, T. An Analysis of the Emotional Evolution of Large-Scale Internet Public Opinion Events Based on the BERT-LDA Hybrid Model. *IEEE Access* **2021**, *9*, 15860–15871. [\[CrossRef\]](#)
95. Kim, S.; Yoon, B. Patent infringement analysis using a text mining technique based on SAO structure. *Comput. Ind.* **2021**, *125*, 103379. [\[CrossRef\]](#)
96. Choi, S.; Park, H.; Kang, D.; Lee, J.Y.; Kim, K. An SAO-based text mining approach to building a technology tree for technology planning. *Expert Syst. Appl.* **2012**, *39*, 11443–11455. [\[CrossRef\]](#)
97. Park, H.; Yoon, J.; Kim, K. Identifying patent infringement using SAO based semantic technological similarities. *Scientometrics* **2012**, *90*, 515–529. [\[CrossRef\]](#)
98. Guo, R.; Zhao, W.; Wei, L.; Zhang, S.; Feng, L.; Guo, Y. A variety of simple and ultra-low-cost methods preparing SLiCE extracts and their application to DNA cloning. *J. Microbiol. Methods* **2022**, *202*, 106565. [\[CrossRef\]](#)
99. Kim, H.; Hyeok, Y.; Kim, K. Semantic SAO network of patents for reusability of inventive knowledge. In Proceedings of the 2012 IEEE 6th International Conference on Management of Innovation and Technology (ICMIT), Bali, Indonesia, 11–13 June 2012.
100. Yang, C.; Zhu, D.; Zhang, G. Semantic-Based Technology Trend Analysis. In Proceedings of the 2015 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Taipei, Taiwan, 24–27 November 2015.
101. Huang, L.; Shang, L.; Wang, K.; Porter, A.L.; Zhang, Y. Identifying target for technology mergers and acquisitions using patent information and semantic analysis. In Proceedings of the 2015 Portland International Conference on Management of Engineering and Technology (PICMET), Portland, OR, USA, 2–6 August 2015.
102. Yang, C.; Zhu, D.; Wang, X.; Zhang, Y.; Zhang, G.; Lu, J. Requirement-oriented core technological components' identification based on SAO analysis. *Scientometrics* **2017**, *112*, 1229–1248. [\[CrossRef\]](#)
103. Guo, J.; Wang, X.; Li, Q.; Zhu, D. Subject-action-object-based morphology analysis for determining the direction of technological change. *Technol. Forecast. Soc. Chang.* **2016**, *105*, 27–40. [\[CrossRef\]](#)
104. Lee, Y.; Kim, S.Y.; Song, I.; Park, Y.; Shin, J. Technology opportunity identification customized to the technological capability of SMEs through two-stage patent analysis. *Scientometrics* **2014**, *100*, 227–244. [\[CrossRef\]](#)
105. Vicente-Gomila, J.M.; Artacho-Ramirez, M.A.; Ting, M.; Porter, A.L. Combining tech mining and semantic TRIZ for technology assessment: Dye-sensitized solar cell as a case. *Technol. Forecast. Soc. Chang.* **2021**, *169*, 120826. [\[CrossRef\]](#)
106. Chao, Y.; Cui, H.; Jun, S. An improved SAO network-based method for technology trend analysis: A case study of graphene. *J. Informetr.* **2018**, *12*, 271–286.
107. Cascini, G.; Zini, M. Measuring patent similarity by comparing inventions functional trees. In *Computer-Aided Innovation*; Springer: Boston, MA, USA, 2008; pp. 31–42. [\[CrossRef\]](#)

108. Wang, J.; Wang, X.; Li, W. Research on Potential Adverse Drug Reaction Forecasting Based on SAO Semantic Structure. *IEEE Trans. Eng. Manag.* **2022**, *PP*, 1–14. [\[CrossRef\]](#)
109. Guo, Y.; Yu, M.; Jing, N.; Zhang, S. Production of soluble bioactive mouse leukemia inhibitory factor from *Escherichia coli* using MBP tag. *Protein Expr. Purif.* **2018**, *150*, 86–91. [\[CrossRef\]](#)
110. Feng, J.; Liu, Z.; Feng, L. Identifying opportunities for sustainable business models in manufacturing: Application of patent analysis and generative topographic mapping. *Sustain. Prod. Consum.* **2021**, *27*, 509–522. [\[CrossRef\]](#)
111. Yoon, B.; Kim, S.; Kim, S.; Seol, H. Doc2vec-based link prediction approach using SAO structures: Application to patent network. *Scientometrics* **2021**, *127*, 5385–5414. [\[CrossRef\]](#)
112. Liu, L.; Tang, L.; Dong, W.; Yao, S.; Zhou, W. An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus* **2016**, *5*, 1608. [\[CrossRef\]](#)
113. Anandarajan, M.; Hill, C.; Nolan, T. Probabilistic Topic Models. In *Practical Text Analytics: Maximizing the Value of Text Data*; Springer: Cham, Switzerland, 2019; pp. 117–130.
114. Chen, H.; Zhang, G.; Zhu, D.; Lu, J. Topic-based technological forecasting based on patent data: A case study of Australian patents from 2000 to 2014. *Technol. Forecast. Soc. Chang.* **2017**, *119*, 39–52. [\[CrossRef\]](#)
115. Erzurumlu, S.S.; Pachamano, D. Topic modeling and technology forecasting for assessing the commercial viability of healthcare innovations. *Technol. Forecast. Soc. Chang.* **2020**, *156*, 120041. [\[CrossRef\]](#)
116. Yan, X.; Guo, J.; Lan, Y.; Cheng, X. A bitern topic model for short texts. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013; pp. 1445–1456.
117. Younas, M.; Jawawi, D.N.A.; Ghani, I.; Shah, M.A. Extraction of non-functional requirement using semantic similarity distance. *Neural Comput. Appl.* **2020**, *32*, 7383–7397. [\[CrossRef\]](#)
118. Rashid, J.; Shah, S.M.A.; Irtaza, A. Fuzzy topic modeling approach for text mining over short text. *Inf. Process. Manag.* **2019**, *56*, 102060. [\[CrossRef\]](#)
119. Lü, L.; Zhou, T. Link prediction in complex networks: A survey. *Phys. A Stat. Mech. its Appl.* **2011**, *390*, 1150–1170. [\[CrossRef\]](#)
120. Lü, L.; Jin, C.-H.; Zhou, T. Similarity index based on local paths for link prediction of complex networks. *Phys. Rev. E* **2009**, *80*, 46122. [\[CrossRef\]](#) [\[PubMed\]](#)
121. Liben-Nowelly, D.; Kleinberg, J. The Link Prediction Problem for Social Networks. In Proceedings of the Twelfth International Conference on Information and Knowledge Management, New Orleans, LA, USA, 3–8 November 2003.
122. Thomas, H. The role of patents for bridging the science to market gap. *J. Econ. Behav. Organ.* **2007**, *63*, 624–647.
123. He, T.; Fu, W.; Xu, J.; Zhang, Z.; Zhou, J.; Yin, Y.; Xie, Z. Discovering Interdisciplinary Research Based on Neural Networks. *Front. Bioeng. Biotechnol.* **2022**, *10*, 908733. [\[CrossRef\]](#)
124. Lei, C.; Ruan, J. A novel link prediction algorithm for reconstructing protein–protein interaction networks by topological similarity. *Bioinformatics* **2012**, *29*, 355–364. [\[CrossRef\]](#)
125. Shibata, N.; Kajikawa, Y.; Sakata, I. Link prediction in citation networks. *J. Am. Soc. Inf. Sci. Technol.* **2012**, *63*, 78–85. [\[CrossRef\]](#)
126. Ma, J.; Pan, Y.; Su, C.-Y. Organization-oriented technology opportunities analysis based on predicting patent networks: A case of Alzheimer’s disease. *Scientometrics* **2022**, *127*, 5497–5517. [\[CrossRef\]](#)
127. Chakraborty, S.; Nijssen, E.J.; Valkenburg, R. A systematic review of industry-level applications of technology roadmapping: Evaluation and design propositions for roadmapping practitioners. *Technol. Forecast. Soc. Chang.* **2022**, *179*, 121141. [\[CrossRef\]](#)
128. Kim, S.; Park, I.; Yoon, B. SAO2Vec: Development of an algorithm for embedding the subject–action–object (SAO) structure using Doc2Vec. *PLoS ONE* **2020**, *15*, e0227930. [\[CrossRef\]](#)
129. Kozłowski, D.; Dusdal, J.; Pang, J.; Zilian, A. Semantic and relational spaces in science of science: Deep learning models for article vectorisation. *Scientometrics* **2021**, *126*, 5881–5910. [\[CrossRef\]](#)
130. Blondel, V.D.; Guillaume, J.-L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P10008. [\[CrossRef\]](#)
131. Lambiotte, R.; Delvenne, J.C.; Barahona, M.J.P. Laplacian Dynamics and Multiscale Modular Structure in Networks. *arXiv* **2012**. [\[CrossRef\]](#)