

Patent Recommendation Based on Boundary-Spanning Technology Search: An Empirical Study from the Robotics Field

Qinghua DONG

*School of Economics and Management, Beijing University of Technology, Beijing 100124, China;
School of Arts and Sciences, Beijing Institute of Fashion Technology, Beijing 100029, China
E-mail: jcbdqh@bift.edu.cn*

You ZHANG

*School of Economics and Management, Beijing University of Technology, Beijing 100124, China
E-mail: zhangyou09@outlook.com*

Xin ZHANG*

*Institute of Mathematics and Physics, Beijing Union University, Beijing 100101, China
E-mail: zhangx@buu.edu.cn*

Abstract A three-dimensional boundary-spanning technology search model including search depth, scope and height is established, and a quantitative calculation method is proposed to dynamically describe an organisation's technology search behaviour and demand characteristics. Organisations are clustered by types as technical, comprehensive, or professional using k-means based on technology search behaviour. Recommendation strategies for various types of organisations are proposed based on this, and the search and supply libraries of each organisation are built by considering their type and search contents. The semantic similarity between patents in different libraries is calculated using a Word2Vec and TextRank model to achieve patent recommendations. An empirical study of the robotics field shows a recommendation accuracy of 0.751, and the accuracy of the technical, comprehensive, and professional types is 0.8282, 0.5389 and 0.7723, respectively. This study considers an organisation's dynamic search behaviour and makes class-based recommendations, with a low computational complexity and strong interpretability.

Keywords technology search; patent recommendation; semantic similarity; robotics

1 Introduction

Patent transaction is increasingly important in facilitating knowledge transfer among innovative organisations and optimizing patent technology resource allocations between large and small organisations^[1]. According to 2022 China Patent Investigation Report issued by the Strategic Planning Department of the State Intellectual Property Office, only 798,000 invention patents were authorized from 1.619 million applications in 2022, and few were transferred with a very low valid patent transfer rate of 11.5% in China. With industrial transformation and upgrading the technological demands of enterprises are growing rapidly^[2]. How to promote

Received September 20, 2023, accepted November 17, 2023

* Corresponding author

the docking of supply and demand and enhance the application of patent transformation has become an urgent problem^[3]. Much of the recent research on patent recommendation has concerned this problem. Recommendation systems, also known as recommender systems, are a specific type of information filtering system^[4]. They are designed to present users with suggestions tailored to their needs, preferences, interests, or observed behaviors^[5]. These systems help in managing the problem of information overload by selectively filtering significant information from a vast, dynamically generated pool of data^[6].

While boundary-spanning technology search is an external technology-related information collection activity driven by current or potential technical needs, can be used in recommendation systems to acquire and use external technical resources^[7]. It can target a specific technical field. Using boundary-spanning technology search, we can analyse and learn the dynamic behaviour and unique preferences of an organisation's technology demand, and research patent technology recommendations based on demand, which can reduce the search cost of the supplier and consumer and increase the patent transfer rate^[8].

This study uses the patent transfer data of organisations to describe and calculate the multidimensional characteristics of an organisational boundary-spanning technology search, cluster organisations based on dynamic demand characteristics, and propose patent recommendation strategies and models for various types of organisations based on the semantic similarity of supply and demand patents.

2 Literature Review

2.1 Research on Recommendations

Traditional personalized recommendation methods are divided into three categories, overall. 1) Proposed by Goldberg, et al. in 1992, collaborative filtering-based recommendations were applied to Tapestry mail system which can filter useful emails for users^[9]. The idea is that users will have the preferences in the future that they had in the past. We find similar users or items by mining users' historical annotation information, and using the rating information of items from similar users to predict the current user's preference^[10]. 2) Content-based recommendations were proposed by Lang, et al. in 1995 by extracting user attributes or product features, learning user interest models, examining the matching degree between user-profiles and alternative recommended products, and recommending the product with the highest matching degree to the user^[11,12]. This method requires further exploration of user behaviour and its implicit interest tendency, and its interpretability must be improved^[13]. 3) Hybrid recommendation is based on the former two methods, which can incorporate the advantages while inheriting the disadvantages of both methods. For example, Balabanovic, et al. proposed a method in 1997 that combined and optimized various techniques to obtain recommendations^[14]. Commonly used hybrid methods are weighted, transformed, stacked, feature-expanded, feature combination type, and meta-level type. These methods have a high computational complexity.

The similarity calculation of user or product features based on semantic similarity is a key step in recommendation research. The Word2Vec method is a widely used neural network-based process model^[15]. The idea is to learn semantic features from the context corpus so the computer can automatically construct the mapping relationship between context and target

words and form word embedding. In the vector space, the semantic similarity of words is characterised by the distance between two vectors, as calculated by cosine similarity. The TextRank algorithm can be used to extract keywords in the text to construct a word network according to the sequence in which words appear. If there is a context between two words, a link is established, whose weight is related to the frequency with which two words are adjacent in the text. TextRank performs well at identifying keywords for software development tasks and providing a keyword list that can be accurately searched by developers^[16].

2.2 Features of Boundary-Spanning Technology Search

Boundary-spanning technology search focuses on knowledge distance and knowledge type. Knowledge distance measures the relative distance between an organisation's search knowledge and existing knowledge in the time, space, and cognitive dimensions. Knowledge type focuses on the type of knowledge for which the organisation searches, which is the content attributes of search knowledge. Knowledge search is divided into the search of market, technological, and supplier knowledge^[17].

Studies have found that innovation is a key driver for organizations to survive and thrive in competitive markets^[18]. Through boundary-spanning technology search, organisations can meet the inner demands of product innovation^[19]. Wei, et al. found the linkage of technology search, innovation, and position in local and global technology networks based on a time-lag structure^[20]. Liu, et al. learned the impact of enterprises' search strategies on technological innovation^[21]. Recommendations based on technology searches are absent from past studies. Xiao used patent licensing and transfer data, based on fitness landscape theory, extended the 2D search model proposed by Katila and Ahuja to construct a 3D search model based on search depth, scope, and height to investigate the relationship between organisations' technology search characteristics and technology trading network positions^[22,23]. In current technology search models, search depth is related to how deeply a firm accesses external knowledge. Search scope/breadth measures the extent of access to a wide range of external knowledge. Search height measures the value of an organization's technology search. Based on existing 3D technology search models, we use patent transfer data of the robotic fields to describe the dynamic technology search behaviour of an organisation and make multi-dimensional recommendations through organisational behaviour clustering.

2.3 Research Review

Research provides a theoretical basis to determine, classify, and quantify multidimensional boundary-spanning technology search indices of organisations, as well as methodological support for patent recommendation research based on semantic similarity measures of patent texts^[24]. However, there are still blank areas in the research of the effective combination of boundary-spanning technology search and technology recommendations.

This paper proposes a patent recommendation model based on boundary-spanning technology search, and data in the robotics field were collected for empirical study. The framework of the research is shown in Figure 1.

The remainder of this study is organised as follows. Part 1 discusses the patent transfer data used to build a three-dimensional technology search model organised by time series. Part

2 explains the development of such a model of search depth, scope, and height, and proposes a calculation method to analyse the dynamic search characteristics of demand organisations. Part 3 clusters organisations based on multi-dimensional boundary-spanning technology search characteristics, and proposes recommendation strategies for various types of organisations to enhance the interpretability of recommended results. Part 4 discusses how we built search and supply libraries for organisations, and calculates the semantic similarity between patents in different libraries using a Word2Vec and TextRank model to achieve patent recommendations.

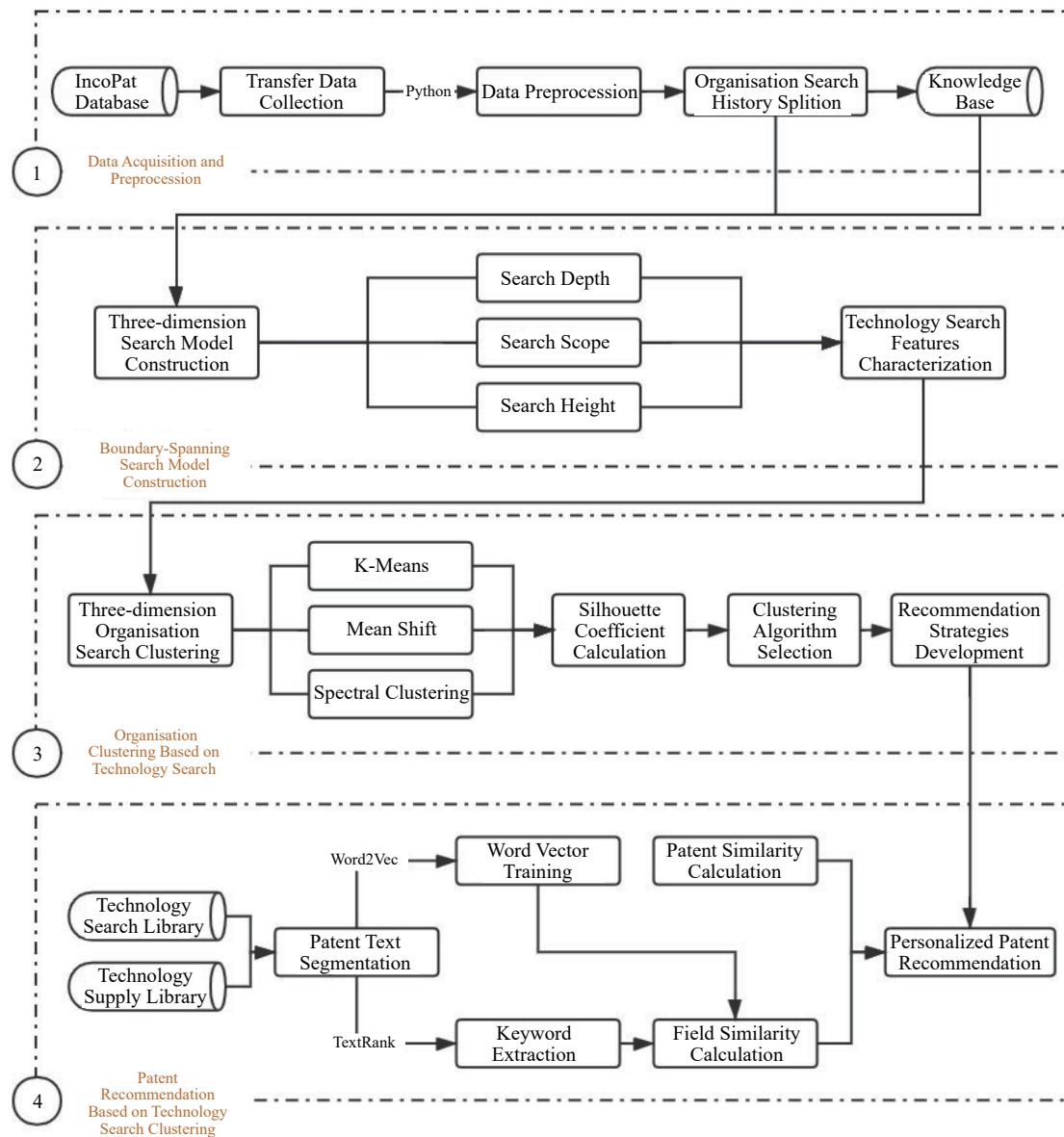


Figure 1 Research framework

3 Empirical Study of Patent Recommendation in Robotics

3.1 Data Acquisition and Proccession

Patent transfer data were acquired based on the IncoPat patent database, which contains information from more than 100 million patents of organisations and regions in 120 countries. We searched 5,514 valid patents in the robotics field from December 2009 to December 2019 in China. Transferors and transferees are a simple collection of all organisations, and are not available directly for each transfer record. Therefore, we used Python to crawl multiple transfer records and store them separately.

In a transfer record that includes multiple transferors or transferees, to split $U_1 \rightarrow U_2 U_3$ into $U_1 \rightarrow U_2$ and $U_1 \rightarrow U_3$, and to split $U_1 U_2 \rightarrow U_3$ into $U_1 \rightarrow U_3$ and $U_2 \rightarrow U_3$ respectively, where U_1 , U_2 , U_3 represent the transferor or the transferee and \rightarrow represents the transfer direction. Then records that involved individual transferor or transferee were deleted to obtain inter-organisational technology transfer records and organisation list.

3.2 Boundary-Spanning Search Model Construction

We constructed a three-dimensional boundary-spanning technology search model consisting of search depth, scope, and height. We discuss each index below.

3.2.1 Search Depth

Search depth is defined as the degree to which search revisits a firm's prior knowledge. A higher depth connotes a higher frequency that organisation searches for field-specific knowledge. We measure it by the ratio of the number of patents inside an organisation's knowledge base to the number of all patents purchased by the organisation. As patent transfer activity in robotics became active after 2015 and most organisations purchase patents infrequently, we set 2019 as the base year for the calculation of the index, and set year 2019 to t . Years 2009 and 2018, as the starting and ending years, are set to $t-10$ and $t-1$, respectively. $KB_i = \{F_{i1}, F_{i2}, \dots, F_{ij}\}$ represents the knowledge base of organisation i , which is the set of IPC purchased by i from year $t-10$ to $t-1$, and j is the total number of IPC inside KB_i .

Figure 2 shows an example that the first three digits of the IPC code of a patent are selected as the field to which it belongs to. Patent 1 belongs to field A01, patent 2 to A01 and A02, and patent 3 to A02. The solid-line links between organisations and patents represent patent transfers from year $t-10$ to $t-1$. The knowledge base of organisations 1 is A01; organisations 2 is A01 and A02.

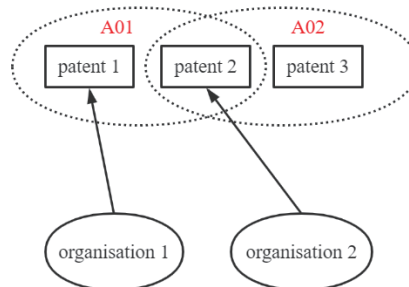


Figure 2 Meaning of knowledge base

Therefore, the formula for the search depth of i is proposed as

$$\begin{cases} D_i = (History_{it} \cap KB_i), \\ Depth_i = \frac{n(D_i)}{\sum_{k=t-10}^t n(History_{ik})}, \end{cases} \quad (1)$$

where $History_{it}$ refers to the IPC of patents transferred by organisation i in year t , D_i refers to the intersection of $History_{it}$ and KB_i , $n(D_i)$ to the number of patents in D_i , and $n(History_{ik})$ to the number of patents transferred by i during year k .

3.2.2 Search Scope

Search scope is defined as the degree of new knowledge that is explored. We measure it by the ratio of the number of patents outside the knowledge base to the number of all patents purchased.

Figure 3 shows an example, where dotted links between organisations and patents represent patent transfers in year t . Both organisation 1 and 2 bought patent 3 in year t . Patent 3 belongs to the field of A02, so A02 is a part of organisation 2's knowledge base, while it is outside organisation 1's knowledge base. Therefore, the link from organisation 1 to patent 3 is a "scope search", and the link from organisation 2 to patent 3 is a "depth search".

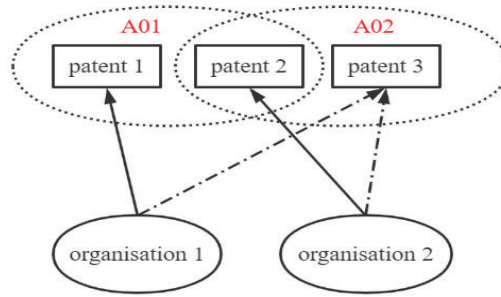


Figure 3 Meaning of search depth and scope

Therefore, the search scope of i is

$$\begin{cases} S_i = (History_{it} - D_i), \\ Scope_i = \frac{n(S_i)}{\sum_{k=t-10}^t n(History_{ik})}, \end{cases} \quad (2)$$

where S_i is the difference set of $History_{it}$ and D_i , which represents IPC transferred in year t but not included in KB_i .

Figure 4 is an example of the search depth and scope. The dotted line represents the organisation's patent transfer from year $t-10$ to $t-1$, and the solid line represents patent transfers in year t . According to formulas (1) and (2), the search depth and scope of organisations 1, 2, and 3 are 0, 1/2; 1/3, 1/3; and 1/4, 1/2, respectively.

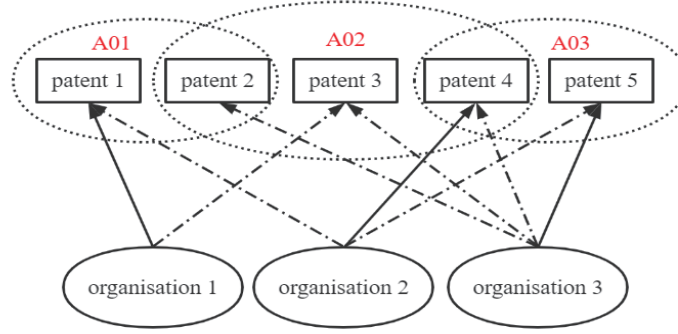


Figure 4 Example of search depth and scope

3.2.3 Search Height

Search height measures the value of patents an organisation searches for. Organisations with high search technology value often own the core or leading technology of their industry. On the one hand, the accumulated high-value technology identification capabilities and experience will help the organisation to purchase new technologies with development potential at a lower cost; on the other hand, the consistency of organisations with a high search value at the business-strategy level makes it easier to share patent pools, which will further reduce the technology search cost^[25].

This study measures the index by the average annual citation frequencies of patents in an organisation's knowledge base. As shown in Figure 5, the solid link between organisations and patents represents patent purchases. The average annual citation is 1.3 in the A01 field, and 0.9 in the A02 field. The search height of organisations 1, 2, and 3 are 1.3, 1.1, and 0.9, respectively. Since the search depth and scope are both between 0 and 1, we normalise the search height to obtain new values of organisations 1, 2, and 3 of 1, 0.5, and 0, respectively.

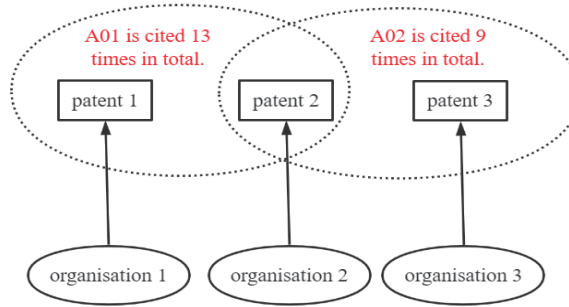


Figure 5 Example of search height

Therefore, the search height of i is proposed as

$$\begin{cases} H_i = \{History_{it-10}, History_{it-9}, \dots, History_{it-1}\}, \\ Height_i = \frac{\sum_{r=1}^{n(H_i)} Citation_{pr}/10}{\sum_{k=t-10}^{t-1} n(History_{ik})}, \end{cases} \quad (3)$$

where H_i represents the IPC set of patents transferred by organisation i from year $t - 10$ to

$t - 1$, p_r are patents belonging to H_i and transferred by i , $n(H_i)$ represents the number of patents whose IPC code belong to H_i , and $Citation_{p_r}$ represents the citation frequency of p_r .

We used all valid patents in the robotics industry from year 2009 to 2019. After filtering organisations with no technology transfer during 2009–2018 and 2019, 59 robotic organisations are obtained that can make technology recommendations. The empirical analysis is based on their technology transfer data from 2009 to 2019. Some results from calculating the three-dimensional boundary-spanning technology search index are shown in Table 1.

Table 1 Three-dimensional technology search indices for some organisations in the robotics field

Organisation	Depth	Scope	Height
Tsinghua University	0.000	0.333	0.375
Nanjing Yusheng Robot Technology Co., Ltd	0.111	0.667	0.573
Nanjing University of Technology	0.000	0.750	0.516
Shenzhen Zhongzhi Kechuang Robot Co., Ltd	0.000	0.750	0.422
Zhongtian Ocean System Co., Ltd	0.500	0.000	0.375
Ocean Exploration Robot Dongtai Co., Ltd	0.500	0.000	1.000
Guangdong Institute of New Materials	0.500	0.000	0.000
Harbin Institute of technology Robot Group (Harbin)	0.333	0.333	1.000
Asset Management Co., Ltd			
Suzhou Huiteng Intellectual Property Consulting Co., Ltd	0.000	0.667	0.500
Guangdong Gaochang Intellectual Property Operation Co., Ltd	0.000	0.100	0.257
Aerospace Science and Engineering Intelligent Robot Co., Ltd	0.000	0.500	0.625
Harbin Institute of Technology	0.000	0.667	0.719
Ningbo Intelligent Manufacturing Industry Research Institute	0.200	0.000	0.458
Shandong Haitian Intelligent Engineering Co., Ltd	0.500	0.000	0.094
Harbin Enhanced Lianzhi Investment Entity (limited partnership)	0.250	0.250	0.813

3.3 Organisation Clustering Based on Technology Search

3.3.1 Cluster Results Analysis

Organisations are clustered based on search indices with three methods including k -means, mean shift, and spectral clustering. The silhouette coefficient is used as the evaluation index of the clustering method, and ranges from -1 to $+1$, where a high value indicates that an object is well matched to its cluster and poorly matched to neighbouring clusters^[26]. The results are shown in Figure 6. It is found that when organisations are clustered into three categories by k -means, the silhouette coefficient is the largest and the clustering effect is the best.

We plot the clustering results of k -means, $K = 3$, in three-dimensional space in Figure 7. The X-, Y-, and Z-axis are the search depth, scope, and height, respectively. Points represent organisations, and colours correspond to different clusters in the clustering results. Blue, red, and black organisations have a high search depth, scope, and height, respectively. Organisations are divided into three types.

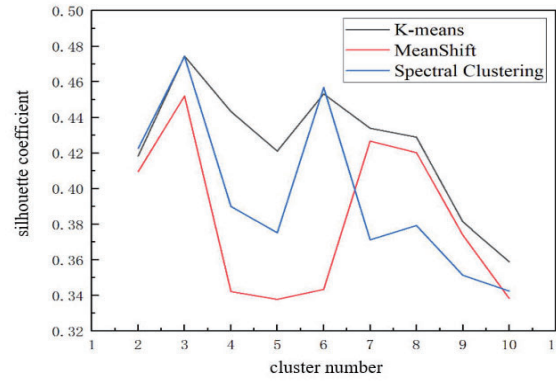


Figure 6 Results of three clustering methods

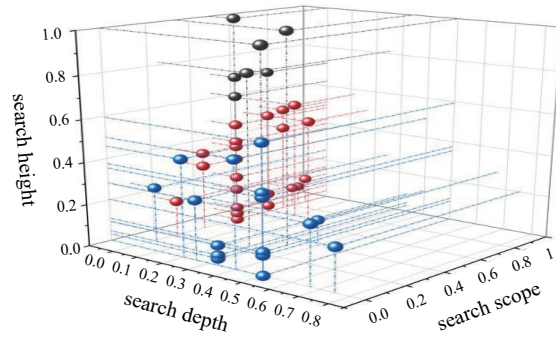


Figure 7 *k*-means clustering result

Type I (technical type) includes 16 organisations. This type of search has high depth and scope, indicating that the organisation focuses on the accumulation of technical searches in certain fields.

Type II (comprehensive type) includes 29 organisations, mostly of medium or large scale, with searches of high scope and low depth. Some organisations have high search height, indicating a search of wider range that is more likely to be cross-field exploratory.

Type III (professional type) includes 14 organisations that usually are universities or research institutes. This type of search has high height, and relatively low depth and scope, indicating that the search technology value of this type of organisation is relatively high.

3.3.2 Recommendation Strategy

The essence of the recommendation strategy is to predict the search behaviour in 2019 based on the historical search contents of an organisation, and to compare it to the actual search to calculate the similarity between patents. We design recommendation strategies for different types of organisations, fully considering the types of knowledge they search for.

We constructed technology search libraries for each organisation to store the patents of its historical search. The search library of organisation *i* is the collection of patent texts (including title and abstract) transferred by *i* from year 2009 to 2018,

$$SL_i = History_{it-10} \cup History_{it-9} \cup \dots \cup History_{it-1} \quad (4)$$

We built technology supply libraries for each of the three types of organisations, as follows. From the 5514 patents, the primary supply library of organisation i was filtered including patents that were not transferred by i in years 2009–2018 (because patents transferred during this period will not need to be recommended) and all patents of 2019. According to the organisations' search characteristics, we further filtered to obtain the supply library, as follows.

For technical organisations, we filtered the field inside an organisation's knowledge base: $DSL_i = \{p_1, p_2, \dots, p_i, \dots\}$, $D(p_i) \in KB_i$, where $D(p_i)$ indicates that the IPC of p_i belongs to the supply library of technical-type organisation i .

For comprehensive organisations, all patents in the two neighbouring fields of the knowledge base of organisation are selected as the supply library: $SSL_i = \{p_1, p_2, \dots, p_i, \dots\}$, $S(p_i) \in \varphi(KB_i)$, where $S(p_i)$ indicates that the IPC of p_i belongs to the supply library of comprehensive-type organisation i , and $\varphi(KB_i)$ refers to the two neighbouring fields with the largest similarity to fields in KB_i .

For professional organisations, all the high-value patents are selected from the primary supply library, $HSL = \{p_1, p_2, \dots, p_i, \dots\}$, $citation_{p_i} \geq 1$.

3.4 Patent Recommendation Based on Technology Search Clustering

Considering an organisation's type and the content of the search knowledge, we calculated the semantic similarity of patent texts between each organisation's search and supply libraries, and selected patents with high similarity for further recommendation.

The steps are as follows.

3.4.1 Word Embedding Training

The corpus for word embedding training is the Chinese Wikipedia corpus and 5514 patent texts, including titles and abstracts. Through the Harbin Institute of Technology (HIT) stop words, we used Python for word segmentation, and used the Word2Vec model based on Gensim to train the word embedding of the corpus. Before training, we set the word embedding dimension to 50 and the training window size to 5.

3.4.2 Patent Similarity Calculation

The Word2vec model can only obtain the embedding of each word and calculate the semantic similarity between embeddings. Hence, it is necessary to transform word embedding to patent embedding by calculating the average embedding of a word in patent text,

$$w_{p_1} = \frac{\sum_{i=1}^{n(p_1)} wv_i}{n(p_1)}, \quad (5)$$

where wv_i represents the word embedding of the i^{th} word segmentation in a patent text, and $n(p_1)$ is the total number of words in patent text 1.

We use the cosine similarity of the patent embedding to calculate the semantic similarity,

$$sim(w_{p_1}, w_{p_2}) = \frac{\sum_{i=1}^{n_p} (x_i \times y_i)}{\sqrt{\sum_{i=1}^{n_p} (x_i)^2} \times \sqrt{\sum_{i=1}^{n_p} (y_i)^2}}, \quad (6)$$

where x_i and y_i represent the i^{th} element in patent embedding w_{p_1} and w_{p_2} , respectively; n_p is the dimension of the word embedding; and $x_i \times y_i$ represents the dot multiplication of two patent embeddings. The value range of the word embedding similarity is $[0, 1]$. The closer it is to 1, the more similar the embedding is. Part of a patent similarity matrix is shown in Table 2.

Table 2 Part of patent similarity matrix

Patent Index	Patent 0	Patent 1	Patent 2	Patent 3	Patent 4	Patent 5
Patent 0	1	0.7495	0.8039	0.7139	0.5721	0.7711
Patent 1		1	0.9854	0.8859	0.8373	0.9002
Patent 2			1	0.8991	0.8549	0.9253
Patent 3				1	0.8783	0.9358
Patent 4					1	0.9082
Patent 5						1

3.4.3 Field Similarity Calculation

When constructing the supply libraries of technical-type organisations, it is necessary to select from the fields with the highest similarity to the IPC fields in an organisation's knowledge base. Therefore, we must calculate field similarity before patent similarity.

In natural language processing based on word embedding, the dimension of the word embedding will affect the accuracy of the language feature representation. Because the calculation of field similarity needs to involve the whole field's title and abstract text to embed words, and using relatively low dimensional embedding to describe these texts will cause information loss. The TextRank algorithm is used to extract a field's keywords, and keyword embedding similarity represents the similarity between fields.

The TextRank algorithm is used to extract the keyword tables from all patent texts in the field, and to calculate the vector similarity between keyword tables. A network through the neighbouring relationships between words is constructed, and PageRank is used to iteratively calculate the rank value of each node, and to rank that to get the keyword. The rank value is calculated as

$$PR(V_i) = (1 - d) + d * \sum_{j \in \ln(V_j)} \frac{w_{ji}}{|\text{Out}(V_j)|} PR(V_j), \quad (7)$$

where $PR(V_i)$ is the rank value of node V_i ; $\ln(V_i)$ is the set of predecessor nodes of V_i ; $\text{Out}(V_j)$ is the set of successor nodes of V_j ; d is the damping factor, which ranges from 0 to 1; and w_{ji} indicates the importance of the connection between nodes. In the network constructed by TextRank, the node type is the word segmentation in the patent text, and the weight w_{ji} is determined by the similarity of the word embedding of the two-word embedding. Taking the F01 field as an example, we calculated the keyword table shown in Table 3.

Table 3 Keyword table of the F01 field

heating	air pump	gas	air cylinder
solar energy	water pump	connect	track
join	curved surface	plug post	plunger
motor	rotor	piston	solar energy
steam turbine	reciprocating motion	gas turbine lubrication	linkage mechanism

After the keywords are selected, the keyword table is used as text data, and the field keyword embedding is calculated by formula (5). We calculate the similarity between fields by formula (6), with results as shown in Table 4. The patents in the two most similar fields are selected as the supply library for comprehensive-type organisations.

Table 4 Part of field similarities

IPC	A01	A21	A22	A23	A24
A01	1	0.9084	0.9224	0.8912	0.8881
A21		1	0.9446	0.9412	0.9049
A22			1	0.9228	0.9010
A23				1	0.9085
A24					1

3.4.4 Recommendation Accuracy

The recommendation accuracy is

$$Accuracy = \frac{\sum L * T}{\sum T}, \quad (8)$$

where T is the actual transfer matrix between organisations and patents in 2019; $L = (O'_1, O'_2, \dots, O'_i, \dots, O'_m)$ is a matrix composed of the patent embedding recommended for all organisations; n and m are the total numbers of patents (5514) and organisations (59), respectively^[27]. The calculation process of L is as follows.

We calculated similarities between patents in the search and supply libraries. For organisation i , we calculated

$$C_i = \begin{pmatrix} \text{sim}(l_1, d_1) & \cdots & \text{sim}(l_1, d_k) & \cdots & \text{sim}(l_1, d_{n(SD_i)}) \\ \vdots & & \vdots & & \vdots \\ \text{sim}(l_j, d_1) & \cdots & \text{sim}(l_j, d_k) & \cdots & \text{sim}(l_j, d_{n(SD_i)}) \\ \vdots & & \vdots & & \vdots \\ \text{sim}(l_{n(SL_i)}, d_1) & \cdots & \text{sim}(l_{n(SL_i)}, d_k) & \cdots & \text{sim}(l_{n(SL_i)}, d_{n(SD_i)}) \end{pmatrix}, \quad (9)$$

where $\text{sim}(l_j, d_k)$, $l_j \in SL_i$, $d_k \in SD_i$ in matrix C_i is the similarity between patents in organisation's search library SL_i and supply library SD_i ; $n(SL_i)$ is the number of patents in its search library of organisation i and $n(SD_i)$ is that in supply library.

From each column of C_i , that is from C_{i1}^T to $C_{in(SD_i)}^T$, for example $C_{i1}^T = (\text{sim}(l_1, d_1), \text{sim}(l_2, d_1), \dots, \text{sim}(l_{n(SL_i)}, d_1))$, $l \in SL_i$, $d \in SD_i$. We took the maximum value for each patent in supply libraries, namely the similarity between each patent of organization i , to obtain the similarity vector $O_i = (\text{Max}\{C_{i1}^T\}, \text{Max}\{C_{i2}^T\}, \dots, \text{Max}\{C_{in(SD_i)}^T\})$. We took out the patents with the K highest similarity in O_i as recommended patents to get the top K matching patent in search library, and mapped the similarity vector O_i of length $n(SD_i)$ to n dimensions (complete with 0) to obtain the recommended vector O'_i of organisation i .

Taking K values of 5, 10, 15, \dots , 60, we show the trend of accuracy with the K value in Figure 8.

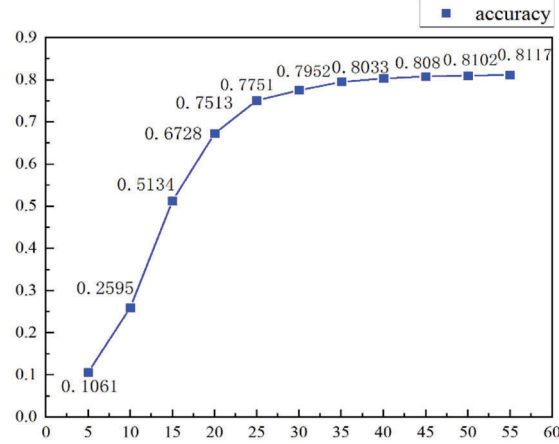


Figure 8 Trend of accuracy with K value

The recommended accuracy in Figure 8 gradually converges starting at $K = 25$. Therefore, we choose $K = 25$ as the upper limit of the number of patents recommended for each organisation, and the accuracy of the recommended model is 0.751. Calculating the average of the recommended accuracy of different types of organisations, we obtain that the recommended accuracy of technical, comprehensive, and professional organisations is 0.828, 0.539, and 0.772, respectively. The accuracy of comprehensive organisations is low. Possible reasons are as follows. 1) When using TextRank to calculate the field similarity, the number of keyword tables is too small, which leads to deviations in the calculation of field similarity. 2) Only two neighbouring fields are selected in the supply library of the organisation. The range is small, resulting in lower recommended accuracy.

3.4.5 Recommendation Results Analysis

We chose one of the three types of organisations and tracked the transfer results in 2019 to compare the recommendation results with the actual transfers, as shown in Figure 9. Nodes represent patents, and the node colour and size respectively represent fields of patents and similarity between patents and an organisation's search library. The node in red box indicates that the patent has been recommended and actually transferred in 2019. The node in blue box means that the patent actually transferred in 2019, but is not recommended.

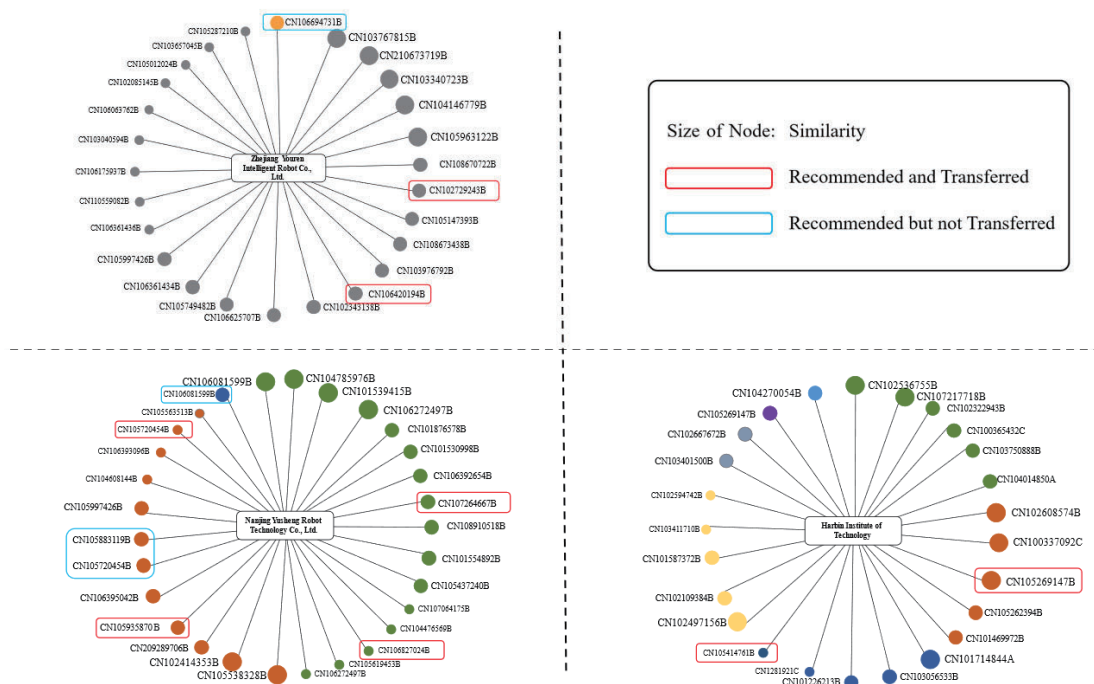


Figure 9 Recommendation result and actual transfer in 2019

From Figure 9, we found the following. For Zhejiang Youren Intelligent Robot Co., Ltd., which is of the technical type, three patents were transferred in 2019, of which two were in the list of recommended top 25 patents. This organisation's technology search demand mainly focuses on field A61. Nanjing Yusheng Robot Technology Co., Ltd., which is of the comprehensive type, transferred seven patents in 2019, of which four were in the recommended list. This organisation mainly focuses on fields B65 and B25. Harbin Institute of Technology, which is of the professional type, transferred two patents in 2019, and both were among the top 25 recommended patents. This organisation is involved in fields such as F04, G01, H02, and B65, which shows a variety of demand. The results show that the recommendation accuracy of the technical and professional organisations is relatively high, and that of the comprehensive type is relatively low. From the recommendation results of Nanjing Yusheng Robot Technology Co., Ltd., we also found that some patents have been transferred but not recommended because their fields are not in the organisation's supply library. So, the recommendation accuracy of the comprehensive type can be improved by adding more neighbouring fields to the organisation's supply library.

4 Conclusion

Based on the IncoPat patent database, using the patent transfer data of the robotics field, we constructed a three-dimensional boundary-spanning technology search model and provided a quantitative calculation method for search depth, scope, and height. By clustering organisations based on multi-dimensional search features, we identified an organisation's search demand

characteristics; categorised organisations as technical, comprehensive, or professional type; and proposed recommendation strategies for each type. Considering an organisation's technology search characteristics and search content to construct search and supply libraries for different types of organisations, we used a word-embedding model to realize patent recommendations by calculating the semantic similarity between patents in the search and supply libraries, realizing a recommendation accuracy of 0.751. The recommended accuracy of the technical, comprehensive, and professional types was 0.8282, 0.5389 and 0.7723, respectively, which demonstrates the validity of the recommended results. The computational time complexity was reduced because of the classification recommendation.

Only the two fields with the highest similarity to the IPC fields in an organisation's knowledge base were selected as comprehensive organisations' supply libraries, leading directly to lower recommendation accuracy. Future research will increase the number of neighbouring fields to improve the accuracy of recommendation results.

References

- [1] Xiong P P, Wu Y R, Hu H, et al. The novel triangle MGM($1, m, N$) model and its applications. *Journal of Systems Science and Information*, 2022, 10(3): 257–279.
- [2] Dong S, Gong H. Environmental technology diversification and green start-up creation: The perspective of technological structure upgrading. *Science & Technology Progress and Policy*, 2023, 40(7): 1–11.
- [3] Cao X G, Hua L Q, Wen H. Managing pricing of closed-loop supply chain under patent protection. *Journal of Systems Science and Information*, 2015, 3(1): 37–47.
- [4] Gupta S, Dave M. An overview of recommendation system: Methods and techniques. *Advances in Computing and Intelligent Systems: Proceedings of ICACM 2019*, 2020: 231–237.
- [5] Roy D, Dutta M. A systematic review and research perspective on recommender systems. *Journal of Big Data*, 2022: 59.
- [6] Konstan J A, Riedl J. Recommender systems: From algorithms to user experience. *User Modeling and User-Adapted Interaction*, 2012(22): 101–123.
- [7] Ke X, Wang Q, Jin X H. Research on technology scouting and its process in the environment of open innovation. *Library and Information Service*, 2009, 53(24): 21–29.
- [8] Ran C J, Li W. The construction of high-value patent recommendation model for universities from the perspective of university-enterprise cooperation: Taking “new energy vehicles” as the case. *Library and Information Service*, 2023, 67(11): 115–126.
- [9] Goldberg D, Nichols D, Oki B M, et al. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 1992, 35(12): 61–70.
- [10] Liu H L, Guo Y, Ma J. Research progress of similarity algorithm in collaborative filtering. *Computer Engineering and Applications*, 2022, 58(13): 27–35.
- [11] Lang K. Newsweeder: Learning to filter netnews. *Machine Learning Proceedings*, 1995: 331–339.
- [12] Jozani M, Liu C Z, Choo K K R. An empirical study of content-based recommendation systems in mobile app markets. *Decision Support System*, 2023, 169: 113954.
- [13] Cui C S, Wei M, Che L B, et al. Hotel recommendation algorithms based on online reviews and probabilistic linguistic term sets. *Expert Systems with Applications*, 2022, 210: 118503.
- [14] Balabanović M, Shoham Y. Fab: Content-based, collaborative recommendation. *Communications of the ACM*, 1997, 40(3): 66–72.
- [15] He X J, Jiang G R. Technology demand recognition model based on semantic similarity under the supply-demand matching perspective. *Systems Engineering — Theory & Practice*, 2019, 39(2): 476–485.
- [16] Wu C W, Liao L C. TextRank keyword extraction method based on multi-feature fusion. *Lecture Notes in Networks and Systems*, 2022, 235: 493–501.
- [17] Zhu S S, Liu F C, Feng X. A study of enterprises' technology search strategies based on technology position. *Science Research Management*, 2020, 41(4): 182–191.

- [18] Ze R, Kun Z, Boadu F, et al. The effects of boundary-spanning search, network ties, and absorptive capacity for innovation: A moderated mediation examination. *Sustainability*, 2018, 10(11): 3980.
- [19] Zan A, Yao Y H, Chen H H. Knowledge search and firm innovation: The roles of knowledge inertia and knowledge integration capability. *Technology Analysis & Strategic Management*, 2022: 1–16.
- [20] Wei X, Chen W. How does a firm's previous social network position affect innovation? Evidence from Chinese listed companies. *Sustainability*, 2019, 11(4): 1191.
- [21] Liu F C, Feng X, Zhao L S. The evolution of technology search strategy of Huawei and Cisco: Based on technological positions. *Science of Science and Management of S.&T.*, 2019, 40(1): 113–123.
- [22] Xiao Z Q, Zhang S H, Liu F C, et al. The impact of regional technology search characteristics on the technology transaction network: Based on technology position. *Science of Science and Management of S.&T.*, 2019, 40(4): 65.
- [23] Katila R, Ahuja G. Something old, something new: A longitudinal study of search behavior and new product introduction. *Academy of Management Journal*, 2002, 45(6): 1183–1194.
- [24] Chen J, Chen J L, Zhao S, et al. Exploiting word embedding for heterogeneous topic model towards patent recommendation. *Scientometrics*, 2020: 1–18.
- [25] Gambardella A, Harhoff D, Verspagen B. The economic value of patent portfolios. *Journal of Economics & Management Strategy*, 2017, 26(4): 735–756.
- [26] Sun L, Liu M H, Xu J C. K-means clustering algorithm using optimal initial clustering center and contour coefficient. *Fuzzy Systems and Mathematics*, 2022, 36(1): 47–65.
- [27] Xiong X Y, Fu Y C, Liu Z Q. Multi-dimensional recommendation algorithm based on bipartite network projection. *Computer Applications and Software*, 2014, 31(8): 253–256.