A General Methodology for Technology Opportunity Discovery Based on Opportunity Evaluation and Optimization

Haiying Ren, Luyao Zhang, and Qian Wang

***Abstract*—Technology opportunities are important drivers of technological advances. Consequently, methods for technology opportunity discovery (TOD) are proposed to discover new types of technology opportunities and to design criteria for defining and evaluating technology opportunities, providing R&D teams and innovators with a plethora of inventive ideas. However, current TOD methods have some common limitations. First, the criteria for defining technology opportunities are typically restrictive, thus may exclude some promising candidates. Second, most criteria for evaluating opportunities lack empirical evidence. In this article, we propose a general methodology for discovering technology opportunities that addresses these limitations. We create a less restrictive technology opportunity space (TOS), built evaluation models for each candidate by learning from historical data, and use optimization techniques to search the TOS for the best technology opportunities. We then implement the proposed methodology in a case study that discovered firm-specific technology opportunities in neural network technology. We present technology opportunities as connected subnetworks of subject–action–object based knowledge networks; designed industry-level, firm-specific and patent-specific evaluation criteria; use random forest to develop the evaluation model from historical patents; and apply ant colony optimization to find the best opportunities. The case shows the feasibility and effectiveness of the general methodology for TOD.**

***Index Terms*—Knowledge networks, opportunity evaluation, opportunity optimization, technology opportunity discovery (TOD), technology opportunity space (TOS).**

I. Introduction

TECHNOLOGY opportunities are sets of possibilities for technological advances and have been recognized as having a significant impact on innovation [1], [2], [3] at the industry and firm levels [4]. Technology opportunity discovery (TOD), also known as technological opportunities analysis [5], aims to find and select the best opportunities for industries or firms from voluminous data. TOD can complement the traditional subjective ideation process of researchers and engineers and increase innovation productivity.

In recent years, quantitative TOD methods have been advancing on several fronts. First, in addition to methods for discovering industry-level technology opportunities [6], [7], [8], more firm-specific TOD methods have been explored [9], [10], [11]. Second, new definitions of technology opportunities have been proposed, such as technological sequences [12] and products [10]. Third, more criteria for evaluating technology opportunities have been developed, as exemplified by the two opportunity-specific and eight firm-specific criteria proposed by Lee et al. [10]. These quantitative TOD methods release the heavy burden on experts and provide policymakers and executives with more objective technology opportunities.

Despite these advances, most TOD studies are faced with two common issues. The first concerns the definition or qualification procedure of technology opportunities, i.e., how knowledge elements (or sets of knowledge elements) are qualified as domain-level technology opportunities or firm-specific technology opportunities. Current methods normally define technology opportunities as those (sets of) elements satisfying certain "desired" criteria or features (such as vacancy or novelty), thus greatly reducing the number of candidate technology opportunities (CTOs). Although intuitive and convenient, the desired criteria or features may filter out some good candidates that do not satisfy them [13]. The second issue concerns how to select criteria for defining and evaluating technology opportunities. In this study, they are referred to as definition criteria and evaluation criteria, respectively. In most TOD methods, both definition criteria and evaluation criteria are proposed and used without empirical testing for efficacy (except in a few studies, such as the article presented in [14]). Since different industries and firms may be situated in different competitive environments and faced with heterogeneous factors influencing opportunity, the importance of the proposed criteria may vary.

In this study, we attempt to resolve the above issues. We propose a general methodology for TOD in which a broader definition of technology opportunities is adopted to prevent good candidates from exclusion, evaluation of technology opportunities is learned from historical data, and search algorithms are used to discover better technology opportunities. We demonstrate the effectiveness of the methodology with a firm-specific TOD case study in neural network technology. In this case, we

Manuscript received 29 August 2022; revised 18 November 2022 and 7 February 2023; accepted 10 March 2023. Date of publication 13 April 2023; date of current version 5 April 2024. This work was supported by Beijing Municipal Natural Science Foundation under Grant 9192003. Review of this manuscript was arranged by Department Editor M. Dabic. *(Corresponding author: Haiying Ren.)*

The authors are with the School of Economics and Management, Beijing University of Technology, Beijing 100124, China (e-mail: 1123868020@qq.com; 571253536@qq.com; 2962666451@qq.com).

Digital Object Identifier 10.1109/TEM.2023.3262257

extract subject–action–object (SAO) based knowledge structures from patents for constructing industrial-level knowledge networks (IKNs) and firm-specific knowledge networks and represent CTOs. Then, we find the important evaluation criteria by machine learning with historical patent data and build an opportunity optimization model. Finally, we use evolutionary optimization techniques to search the knowledge networks for the best technology opportunities for a focal firm.

II. Literature Review

Quantitative TOD methods have been developing steadily in both defining technology opportunities and evaluating technology opportunities. The former determines how knowledge elements are qualified as CTOs, while the latter applies various criteria for judging the value of these candidates. Currently, most TOD methods define and evaluate CTOs in sequence.

A. Defining Technology Opportunities

In defining technology opportunities, industry-level and firm-specific TOD methods often use distinct opportunity criteria or features. On the one hand, since industry-level technology opportunities presumably exist in the pioneering "region" of technological domains, criteria, such as vacancy [7], [15], [16], gaps [4], [17], [18], outliers [6], [14], [19], [20], new combinations [21], or predicted links [22], [23], are considered crucial for defining technology opportunities. Vacancy or "vacuum" was one of the first criteria for defining technology opportunities. Lee et al. [15] constructed patent maps based on a principal component analysis (PCA) of patent keyword vectors and defined technology opportunities as vacancies in the maps, where the vacancies were judged subjectively. Son et al. [7] and Yoon and Magee [16] enhanced the objectivity of the vacancies ("vacuums") in patent maps through generative topographic mapping (GTM) by which each "vacuum" was represented as a blank grid in the GTM-based patent maps. Gaps between science and technology were also explored. Wang et al. [4], Shibata et al. [17], and Li et al. [18] identified scientific and technological clusters from papers and patents, respectively, and defined technology opportunities as the gaps between the two sets of clusters, especially the scientific clusters that are absent from technological clusters. Outliers, new combinations, and predicted links were other forms of technology opportunities. Yoon and Kim [6] constructed patent maps by measuring the semantic similarities between the SAO structures of patents and used the "outlierness" (novelty) index to define technology opportunities. Lee et al. [19] built morphological patent contexts to select novel patents with high local outlier factors as technology opportunities. Rodriguez et al. [20] clustered patents in attributed patent citation networks and selected patents with high outlier ranking as technology opportunities. Lee and Lee [14] applied the outlier criterion to patent classifications and defined technology opportunities as outliers in the patent class landscape. Song et al. [21] first selected reference technologies that were similar to a target technology (both represented by F-terms) and then created technology opportunities by combining the reference and the target technologies. Han et al. [22] and Ma et al. [23] defined technology opportunities as the predicted links in an SAO network and a co-occurrence network of feature terms, respectively. It should be noted that the industry-level technology opportunities are not always novel or outliers. For example, emerging technologies [24], [25] contain recent but promising industry-level technology opportunities. Park et al. [26] defined technology opportunities as the extant patent topics that would reduce the potential risks brought by emerging technologies.

On the other hand, firm-specific technology opportunities are considered more realistic and should fit the technology portfolio of a focal firm [27]. Therefore, most firm-specific TOD methods define CTOs as knowledge elements similar to or connected to a focal firm's technology capabilities. Yoon and Park [28] proposed identifying technology opportunities at both the industry and firm levels with morphology analysis. In their study, the industry-level opportunities were selected from the new morphologies in the whole technology domain, while the firm-level opportunities were discovered by using the patent of a focal firm and one of its competitors. Lee et al. [29] built a technological attribute–application table and identified firm-specific technology opportunities by matching it with keywords and action objects of a target firm. Yoon et al. [9] built a TOD knowledge base from patents from many fields, designed a TOD logic that derived technology opportunities from a focal firm's existing technologies or products (ETPs) using semantic functional similarities, and identified four types of TOD paths from the firm's current ETPs. Park and Yoon [27] represented the technology portfolios of a focal firm as the international patent classification (IPC) codes it covered, and those untapped IPCs with high preference scores by using collaborative filtering were selected as technology opportunities. Choi et al. [12] generated a sequence database containing a firms' dynamic change in focus technology fields, matched it with a precedence enterprise (PE) sequence database, and identified the PE sequences similar to that of the target firm as technology opportunities. Lee et al. [10] used a patent-product database to measure the similarity of products via word2vec, and products similar to a firm's target product were selected as technology opportunities. Lee et al. [30] constructed a universal F-term network with large-scale patents issued by the Japanese Patent Office and an F-term network representing a target firm's technology portfolio and calculated link prediction scores to identify F-terms (technology opportunities) that could connect to the firm in the future.

B. Evaluating Technology Opportunities

In evaluating technology opportunities, a few studies analyzed and evaluated the technology opportunities qualitatively [4], [17], [26], [28], while most research designed quantitative criteria for evaluating the technology opportunities. Industry-level TOD methods use evaluation criteria that are considered important for technological or commercial advantages. Industry-level criteria include indicators, such as technological criticality [15], technological trends [15], patent citations [19], patent claims [19], importance of products in industry [31], technological growth [18], [21], [24], [25], [27], competitive level [10], [27], applicability [21], pace of technological change [10], technology impact [30], network position [13], and various novelty indicators, that are also used in opportunity definition processes [6], [16], [22], [23], [24], [25].

Firm-specific TOD methods applied criteria that evaluated the similarity or compatibility of technology opportunities and firm capabilities, and some methods included both industry-level and firm-specific evaluation criteria. Firm-specific criteria consisted of semantic similarity [9], the importance of opportunities in a target firm [31], heterogeneity [27], [30], and so forth. Choi et al. [12] designed five indices for measuring the compatibility of the technological paths (technology opportunities) with the focal firm: the parallax index, focus technological similarity index, concentration ratio, technology capability index, and recency index. Lee et al. [10] gathered eight indicators from the literature and modified them to assess the value of firm-specific technology opportunities: technology share, technological coverage, technological newness, technological strength, collaboration intensity, collaboration coverage, international scope, and protection scope.

C. Evaluating Technology Opportunities

Most of the aforementioned industry-level and firm-specific TOD methods first used one or a few criteria for defining CTOs and then applied additional criteria for evaluating the candidates. The assumption of this dominant "define–evaluate" strategy is that technology opportunities should have some key characteristics, therefore satisfying one or a few conspicuous criteria appealing to the decision makers of technology. This strategy has the advantage of fitting the human thinking style, which makes it easier to adopt. Another strength is its efficiency in the sense that only a few candidates need further evaluation. However, the limitation is that the definition criteria may filter out some good technology opportunities that do not satisfy them.

A second limitation of the current TOD methods, which is closely related to the above limitation, is that most methods do not assess the importance of the definition criteria and evaluation criteria of technology opportunities. The key issue is whether the definition criteria are indeed more important than the evaluation criteria to the value of technology opportunities. Several studies, e.g., the article presented in [10], used previous research to support the choice of their evaluation criteria. However, there has been insufficient support for the definition criteria in most TOD studies, and the importance of most definition criteria and evaluation criteria has not been empirically assessed before use. We posit that because technologies and industries have different development trajectories and competition environments from each other, the criteria for influencing technology opportunities may vary. If some evaluation criteria were in fact more important than the definition criteria, the more important criteria should have been used as definition criteria instead. Even so, the dominant "define–evaluate" strategy still hinders the discovery of more technology opportunities.

D. Proposed Strategy

Recently, a few TOD studies have departed from the above dominant strategy. Yun et al. [11] and Seo et al. [31] evaluated all product associations and expired patents, respectively. Lee and Lee [14] developed naïve Bayes models that evaluate technology opportunities based on the forward citations of the existing patented inventions, but they defined technology opportunities using an outlier criterion. Ren and Zhao [13] first proposed reorganizing the "define–evaluate" processes and using historical technology opportunities to assess the importance of all criteria. The current study proposes a general TOD strategy building on these studies. The basic idea is to find the most important criteria for evaluating technology opportunities with historical data and then search for the best technology opportunities among all the sets of knowledge elements. We termed this an "evaluate–search" strategy. Fig. 1 compares our strategy with the dominant strategy.

Fig. 1(a) and (b) shows the dominant "define–evaluate" strategy for discovering industry-level and firm-specific technology opportunities, respectively. Both strategies implicitly assume that there are important cutoff criteria for CTOs. However, this cutoff process (step ①) may filter out better opportunities lying in the "blank" areas in Fig. 1(a) or (b).

In the "evaluate–search" strategy, as illustrated in Fig. 1(c), neither clear-cut distinctions between opportunities and nonopportunities are made nor are there clear boundaries between industry-level and firm-specific opportunities. Instead, the strategy first uses historical technology opportunities to find the most important criteria for evaluating industry-level or firm-specific technology opportunities and then applies these criteria to search for the best ones from among all possible technology opportunities. This strategy will be implemented into the general methodology for TOD presented in Section III.

III. Methodology

We now formally propose a general methodology for TOD. It consists of four major steps:

1) collecting knowledge elements;
2) representing CTOs;
3) finding evaluation models for CTOs;
4) searching for the best technology opportunities.

This methodology is general because the knowledge elements, technology opportunities, evaluation criteria, and search procedure are presented in general forms, i.e., free from actual knowledge sources, data types, and technology opportunity representations, to encompass a wide range of implementations. This methodology is illustrated in Fig. 2, where its right panel demonstrates a hypothetical microcase of TOD. It is assumed that the goal of the microcase is to discover the best novel combinations of certain patents in a technical field.

A. Collecting Knowledge Elements

In this step, knowledge elements are collected. A "knowledge element" is a unit of knowledge in one or multiple technological domains depending on the scope of a TOD study. Examples of

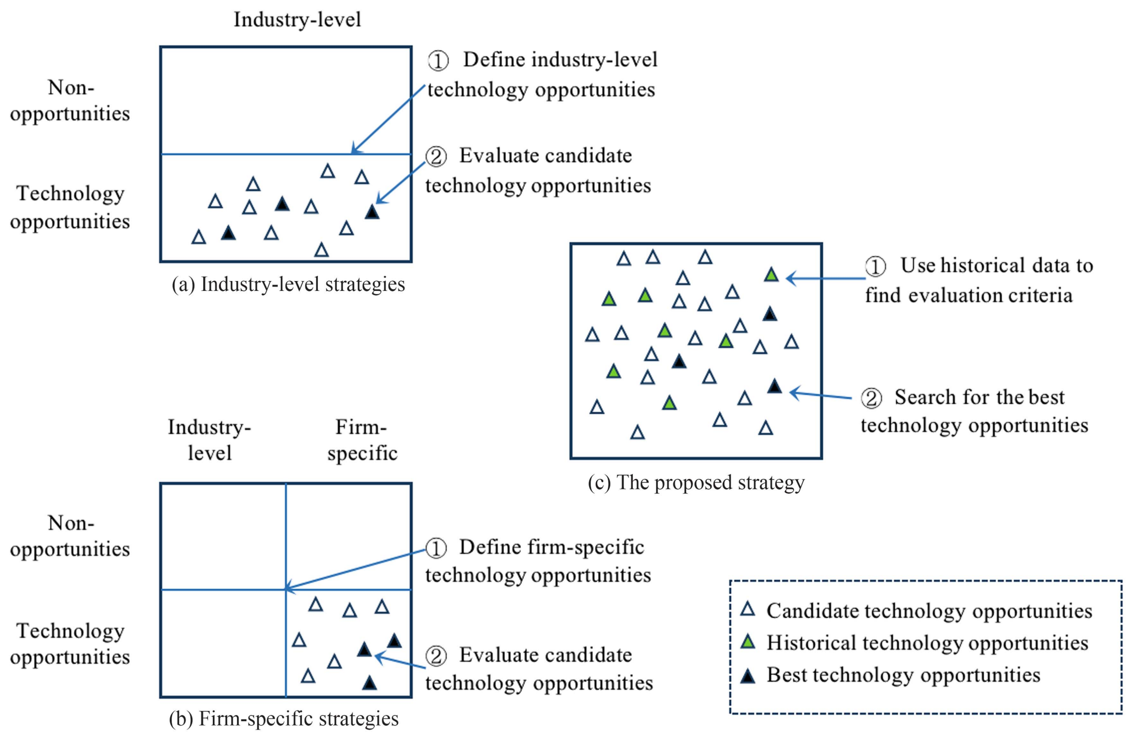

Fig. 1. Illustration of the "define–evaluate" strategies compared with the proposed "evaluate–search" strategy.

knowledge elements include technical concepts, words, patents, application dates, academic papers, inventors or authors, materials, instruments, products, firms, institutions, countries, and so on. For the microcase, the knowledge elements are shown in the right panel of Fig. 2(a) as the short lines, each representing a patent.

Collecting knowledge elements is by no means a trivial task. Before knowledge elements are ready for TOD, a series of actions should be taken as follows:

1) determining the scope of technical fields in the TOD study;
2) choosing data sources, such as patents, academic papers, industrial reports, and/or other Internet resources;
3) acquiring and preprocessing the data for TOD;
4) extracting knowledge elements and constructing knowledge structures [13].

B. Representing CTOs

In this step, all CTOs are represented as subsets of a set of knowledge elements, and constraints may be used to filter out unsuitable CTOs and form a technology opportunity space (TOS).

1) Representing All CTOs: A CTO is represented as a member of the power set $P(\mathrm{K})$, where K is the set of knowledge elements in the focal technological domain(s) of a TOD study. In other words, a CTO is a subset of K. For instance, let us assume that only three knowledge elements, a, b, and c, are considered in a TOD study. $\mathrm{K} = \{a, b, c\}$. $P(\mathrm{K})$, the power set of K, has eight members, namely, $\{a\}$, $\{b\}$, $\{c\}$, $\{a, b\}$, $\{b, c\}$, $\{a, c\}$, $\{a, b, c\}$, and the empty set. Therefore, there are seven CTOs (corresponds to "nonopportunity") in $P(\mathrm{K})$. In this study, the sequence or possible internal relationships of knowledge elements are not considered, and each member of $P(\mathrm{K})$ maps to a unique technology opportunity.

2) Filtering Out Unsuitable CTOs: In the above CTO concept, any member of the $P(\mathrm{K})$ can be a candidate for a TOD study. However, since there are 2^{K} members in $P(\mathrm{K})$, it is infeasible to search and evaluate all the members when $|\mathrm{K}|$ is large. This is precisely the reason why most TOD studies use certain criteria (such as vacancy or outlierness) to reduce a TOS to a manageable level. Therefore, there exists a tradeoff between strictly defined and unrestricted CTOs. Our strategy is to find the least restrictive definition criteria-called "constraints" in optimization theory that only filter out the members that are unsuitable for the purpose of the TOD practice, and the suitable CTOs form a TOS. Here, "a CTO is unsuitable" means that its value cannot be measured or is not applicable in the TODs industrial setting.

One way of determining the suitability (the constraints) of CTOs is to investigate the representations of historical technology opportunities used for assessing the importance of the evaluation criteria for CTOs (this will be explained in Section III-C). Our suggestion is to set the constraints so that CTOs share the same evaluation model with the historical technology opportunities. For example, there are 110 knowledge elements, i.e., patents in the microcase [in the right panel of Fig. 2(a)]. Obviously, it is infeasible to investigate all 2^{110} possible patent combinations (CTOs). However, if the historical patent combinations can be empirically assessed, and they consist of small, "circle-like" patent clusters, we can filter out isolated patents [in Fig. 2(a)] and set the constraints of CTOs to be the patent clusters [in

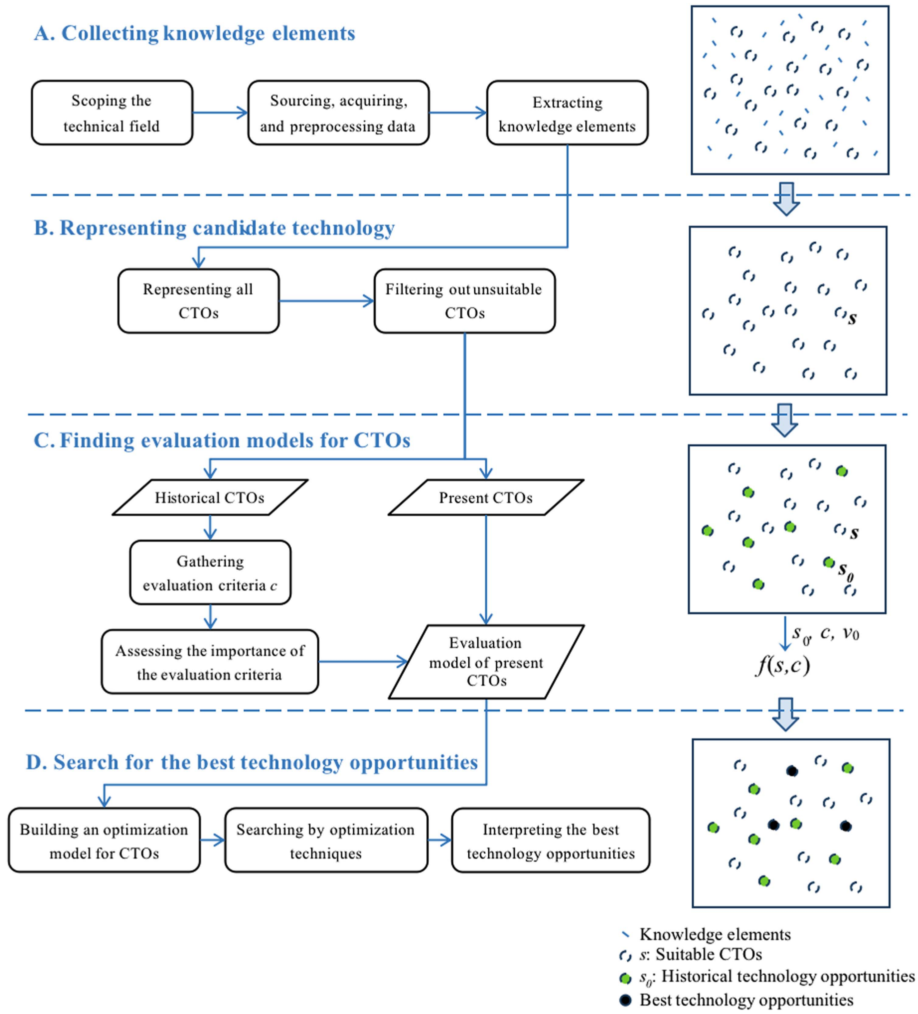

Fig. 2. Overall procedure of the proposed TOD methodology.

Fig. 2(b)], resulting in a TOS of 20 potential CTOs, which is much easier to search than the original $P(\mathrm{K})$.

The above procedure of using constraints to form TOS appears similar to the dominant "define–evaluate" strategy. However, the proposed procedure has two standout features. First, the constraints on CTOs in the general TOD strategy need not be applied in cases of small knowledge sets and can be adjusted on a spectrum from the least restrictive to the most restrictive constraints. Second, the constraints do not have to be set *prior to* the evaluation process. Instead, they can be determined while building evaluation models for CTOs, as will be further demonstrated in the case study in Section IV.

C. Finding the Evaluation Model for CTOs

In this step, the evaluation criteria for CTOs are gathered, and their importance is assessed or modeled with historical technology opportunities. This is a crucial step in our methodology.

1) Gathering the Evaluation Criteria for CTOs: Currently, there is no unique way of measuring the value of various types of CTOs. Most TOD studies use multiple criteria to evaluate CTOs. In the proposed methodology, all definition criteria and evaluation criteria in the TOD literature may be used, depending on the knowledge sources and representations of opportunities (see Section II-A and B for example criteria used in the literature). Note that both industry-level and firm-specific criteria could be used in a firm-specific TOD study. If new representations of CTOs, such as the patent combinations in the microcase, are explored, some new criteria would have to be designed.

Regardless of whether using the existing criteria or designing new ones, the management literature provides rich empirical results about what criteria may affect the innovation outcome. Although technology opportunities have various representations, such as products, patents, keywords, patent classifications, and knowledge combinations, some general results are similar on various levels of innovation. For example, novelty positively affects innovation performance at the patent [32], [33], product [34], and firm levels [35], [36]. This suggests that various criteria measuring novelty could be applied in a particular TOD study.

With respect to the microcase, since CTOs are represented by patent combinations, the mean, median, or maximum of novelty indicators of all patents in a combination can be tried as evaluation criteria for measuring novelty. Other patent indicators, such as forward citations and number of IPCs, can also be integrated for designing criteria for patent combinations. All the evaluation criteria are denoted as set c in the right panel of Fig. 2(c).

2) Assessing the Importance of the Evaluation Criteria With Historical Technology Opportunities: In this substep, the importance of the evaluation criteria for CTOs is assessed with historical technology opportunities, resulting in an evaluation model for CTOs. If the (qualitative) relationship between the criteria and the value of CTOs are already known, well-established multiple-criteria decision-making (MCDM) techniques can be used to assess the importance of evaluation criteria and compare candidate CTOs. For example, PCA groups related criteria into components and can assign importance to each component based on the percentage of variation it explains. Data envelopment analysis (DEA), a collection of weight-free MCDM models, is also a good choice if the evaluation criteria can be categorized as inputs or outputs. PCA and DEA are data-driven methods that are "empirical" in nature. The analytical hierarchy process is a more subjective method in which experts assess the relative importance of criteria by pairwise comparison. However, this approach can reveal the technological preferences of firms or R&D teams and, hence, is an alternative to data-driven evaluation methods.

However, if a TOD study explores new evaluation criteria for CTOs (as described in substep 1), the MCDM techniques may not apply because the relationship between the new evaluation criteria and the value of CTOs is unknown. The importance of the new criteria must be assessed to evaluate historical technology opportunities and predict the potential value of CTOs. Our general methodology for TOD suggests assessing the new criteria with the ex post value of CTOs and the historical technology opportunities in the same technological domain(s). First, the value of historical CTOs is measured. Commercial success (such as product sales) is a desired measure, but substantial effort has to be made to obtain the commercial value and attribute it to individual CTOs. Next, regression models or machine learning models are used to assess the impact of each evaluation criterion (predictor variable) on the value of historical technology opportunities (predicted variable) with historical data. Regression models have the strength of statistical rigor, although their predictive accuracy is often less than that of the machine learning models. The drawback of machine learning models is their black-box nature. The ending date of the historical data should be as close to the present as possible (to minimize extrapolation error) but should leave enough time for the value of technology opportunities to manifest. Finally, various models are tested for statistical power or predictive accuracy, and the best evaluation model is selected for building an optimization model for CTOs. In the microcase, for instance, the historical CTOs [the green patent clusters, such as s_0 in Fig. 2(c)] are identified, their criteria c are calculated, and their value v_0 are measured (v_0 can be measured by the mean value of all patents in a cluster). Regression or machine learning methods can be used to obtain the mapping f from s_0 and c to v_0, which becomes the evaluation model for all the 20 CTOs (7 historical and 13 novel).

D. Searching for the Best Technology Opportunities

In this step, the evaluation model obtained from Step 3 is used to build an optimization model for CTOs. The best technology opportunities are searched and discovered by optimization techniques and are finally interpreted.

1) Building an Optimization Model for CTOs: In the general methodology for TOD, the number of CTOs in the TOS is normally large. Therefore, an optimization model (*M*) is constructed to search for the best technology opportunities, stated as follows:

$$\text{Maximize } Z = f(s, c) \qquad (M)$$
$$\text{s.t.} \quad s \in \text{TOS}$$

where the objective function *f* is the evaluation model of CTO *s* (the decision variable) and criteria set c; s is subject to (s.t.) constraints that define suitable CTOs in the TOS; *Z* denotes the value of s. The objective function of (*M*) aims to maximize Z. It is advisable to search for the top *M* technology opportunities with the largest Z values to provide R&D teams with multiple valuable alternatives.

2) Searching for the Best Technology Opportunities by Optimization Techniques: In the microcase example, there are only 13 novel CTOs [the patent clusters not in green in Fig. 2(d)], and the optimization model (*M*) can be solved just by enumeration with the top three technology opportunities marked in black. In general, however, since TOS is a subset of the power set *P*(*K*), the optimization problem (*M*) belongs to a class of subset selection problems, which is NP-complete [37] in general. Classical optimization methods that search for the exact optimum solution cannot solve (*M*) efficiently. Fortunately, many evolutionary algorithms, such as genetic algorithms and ant colony optimization (ACO), can be used to obtain multiple high-quality solutions.

3) Interpreting the Best Technology Opportunities: In the final procedure of the general methodology for TOD, the best technology opportunities obtained from the optimization model are analyzed and interpreted. The most reliable interpreters are the end users of the TOD study, such as the R&D team members and managers. These experts will analyze the knowledge elements embodied in or related to the technology opportunities, be it materials, methods, functions, or people. It is possible that the technology opportunities might provoke interesting, imaginative, and creative ideas beyond the opportunities themselves, triggering new opportunities unforeseen by the proposed procedure.

Another way to interpret a technology opportunity is to break it down into subsets of knowledge elements, i.e., partial ideas, each of which generally contains more specific technological information and is easier to analyze and understand. Ultimately, the overall meaning of the technology opportunities can be interpreted as a combination of all the partial ideas.

In addition, the general methodology for TOD also provides some key information that may facilitate the interpretation procedure. The important evaluation criteria can be calculated for the best opportunities and their components, and the extreme values for the criteria can help understand and interpret the conspicuous features of the technology opportunities.

Finally, the interpretation substep also serves as a validation of the TOD study. If the end users of TOD find the best opportunities ordinary, infeasible, or incomprehensible, the data should be checked and/or the method modified to obtain better and more meaningful and useful results.

IV. Case Study

This section describes the implementation of the proposed general methodology for TOD in a case study to demonstrate the effectiveness of the methodology. The implemented TOD method aimed to discover the best firm-specific technology opportunities in neural network technologies that would contain key ideas for developing new patents. All the data processing and analysis were coded with Python.

A. Collecting Knowledge Elements

The following list outlines the different types of graphics published in IEEE journals. They are categorized based on their construction and use of color/shades of gray.

1) Determining the Scope of Technical Fields in the TOD Study and Choosing the Data Source: Neural networks are computer programs that are inspired by the natural neural network in the brain and are at the forefront of cognitive computing [38]. To scope this technical field, we consulted the subcategories in the neural network domain supplied by AMiner [39], a knowledge base of artificial intelligence developed by Tsinghua University. From these subcategories, we obtained the keywords related to neural networks, including “deep neural network,” “artificial neural network,” “Boltzmann machine,” “deep learning,” “convolutional neural network (CNN),” etc., and their abbreviations. We used these keywords to delineate the targeted knowledge field.

To choose a data source for the case study, we used the Derwent Innovation Index (DII) because it is a technological resource widely used in academics and industry.

In the case study, we wanted to discover firm-specific technology opportunities. The focal firm is Beijing Kuangshi Technology Co., Ltd., (Kuangshi Tech for short). Kuangshi Tech was a startup company specializing in software design, and it was a new player in neural network technology with only three patents in the field. Therefore, Kuangshi Tech wanted to explore the best technology opportunities it could capture and develop patents accordingly.

2) Acquiring and Preprocessing the Data for TOD: We used the above keywords as search terms in the topic field of DII. A total of 16 565 patents were obtained from 1973 to December 31, 2020. Most of the patents were applied after 2015 and at an accelerating speed, indicating the rapid development of the field and many technology opportunities therein. We also downloaded all three patents of Kuangshi Tech since they all represent the technology capability of the firm related to neural networks.

Then, the titles and abstracts of the patents were extracted for preprocessing, which included cleansing of unrelated words, abbreviation replacement, coreference resolution, and other operations. Important information on the patents, such as patent numbers, application years, assignees, and priority countries, was saved for use in later procedures.

3) Extracting Knowledge Elements and Constructing Knowledge Structures: In this study, we extracted all the SAOs from the preprocessed texts of each patent as knowledge elements, used these SAOs to construct a knowledge network for each single patent, and then constructed knowledge networks for all the patents in the technical domain as well as for the patents of a focal firm. These steps are described as follows. The extraction of SAOs and the construction of knowledge networks were coded with Python packages, especially the NLTK, StanfordParser, and NetworkX.

a) Extracting SAOs: SAO structures can show the components and functions in a patent [40]. SAOs have been widely used in patent analysis [41], technological planning [42], [43], and TOD [22]. Most SAO extraction methods focus on the subjects and objects that are connected by verbal phrases while overlooking the modifiers (especially prepositional phrases) in the sentence. In this case study, we extracted not only regular SAOs but also prepositional phrases that contained additional technical and functional information in the patents. The following example compares the common SAO extraction methods with our approach.

Original sentence:

“The system has a sensing mechanism for outputting sensing signal to a control unit.”

Regular SAO extracted:

system (S), has (A), sensing mechanism (O).

SAOs extracted in this study:

system (S), has (A), sensing mechanism (O);

sensing mechanism (S), for outputting (A), sensing signal (O);

sensing signal (S), to (A), control unit (O).

The example showed that our approach extracted more useful knowledge elements from the content of patents, therefore facilitating the discovery of detailed, patent-level technology opportunities. To implement this SAO extraction method, we parsed a sentence with Stanford NLP Parser to obtain its constituent tree and extracted noun phrases (NP tags) separated by neighboring verbs (VB-type tags) or prepositions (P tags) in the tree. The verbs and prepositions were assigned “A” labels, and the separated noun phrases were assigned “S” or “O” labels. For the purpose of TOD, the distinction between “S” and “O” was not crucial, and “A” labels did not necessarily mean “actions.”

We extracted all SAOs for all the patents in the neural network domain. To facilitate the construction process of knowledge networks, all words were stemmed, and the stop words were removed from the noun phrases.

b) Constructing knowledge networks for single patents: After the extraction of SAOs, we constructed knowledge networks for single patents. First, the knowledge elements labeled with “S” and “O” were taken as the nodes of the knowledge

TABLE I
CRITERIA FOR EVALUATING FIRM-SPECIFIC TECHNOLOGY OPPORTUNITIES

Category	Subcategory	Meaning and symbol
Industry-level criteria	Novelty	Recent nodes (*recent_n*), novel nodes (*novel_n*), novel edges (*novel_e*), median application year (*year_n_median*)
	Conventionality	Year range of edges (*e_range*), mean strength of nodes (*strength_median*), maximum weight of edges (*weight_max*), edge importance (*e_importance*), maximum citations of nodes (*cited_n_max*)
	Network position	Mean eigenvalue of nodes (*eigen_mean*), mean PageRank of nodes (*pr_mean*)
Firm-specific criteria	Firm's R&D capability	Inventors (*inventors*), experience (*A_year_min*), codeveloped patents (*A_co_pats*)
	Compatibility with firm's capability	Binary match with nodes (*A_is_match_n*), mean match with nodes (*A_match_n_mean*), maximum match with nodes (*A_match_n_max*), unmatched nodes (*A_no_match_n*), total knowledge exploitation (*explo_sum*), codeveloped nodes (*A_co_match*)
Patent-specific criteria		Application year (*Year*), priority country (*country*), claims in the patent (*claims*), number of IPCs (*IPCs*), backward citations (*back_cite*)

networks, and those labeled with "A" were taken as the edges. Then, all the same nodes with "S" or "O" labels for a patent were combined into one node and connected into a knowledge network for the patent. It should be noted that the nodes and edges of the two SAOs were all combined when both "S" and "O" were in common, even if the two "As" were different. Finally, the corresponding phrases, SAO labels, patent numbers, IPCs, assignees, application years, priority countries, and other patent information were added to all the nodes and edges as the node attributes and the edge attributes. These attributes were used to evaluate technology opportunities in later steps.

c) Constructing knowledge networks for the industry and firms: Using the same procedure of constructing knowledge networks for single patents, we combined SAOs of two sets of patents and constructed IKNs and firm-level knowledge networks (FKNs).

First, IKNs were constructed with all patents in the neural network industry. Two IKNs were constructed, one for assessing the importance of the evaluation criteria with historical technology opportunities (using all patents up to 2015 in the study) and the other for discovering future technology opportunities with all available patents in the industry. We called these IKNs the historical industry-level knowledge network (HIKN) and the present industry-level knowledge network (PIKN), respectively.

Next, two types of FKNs were constructed with the patents owned by particular firms: historical firm-specific knowledge networks (HFKNs) and the present firm-specific knowledge network (PFKN). One HFKN is constructed for a firm that owns the historical patents that are used for assessing the importance of the evaluation criteria. One PFKN is constructed for the focal firm (Kuangshi Tech in our case) to discover future technology opportunities. The time ranges of HFKN and PFKN are the same as those of HIKN and PFKN, respectively.

It should be noted that before combining the SAOs in various patent sets for constructing the HIKN, PIKN, HFKN, and PFKN, we "clumped" similar phrases with "S" or "O" labels of all the SAOs in the industry following a term-clumping procedure proposed by Trumbach and Payne [44]. For example, the phrases "first neural network" and "neural network" were clumped together. The purpose of combining similar concepts was to improve the connectivity (hence, new potential opportunities) of the knowledge network. In the case study, the similarities between words were calculated based on their WordNet hierarchies. When clumped nodes (and possibly edges) were combined, the original phrases and associated node/edge attributes were put together as the new attributes of the clumped nodes and edges. Saving all the relevant attributes would enable future interpretation of technology opportunities.

B. Representing CTOs

1) Representing CTOs: After the knowledge elements were collected in the form of IKNs, CTOs were represented. In the general sense, any subnetwork of PIKN (even including PIKN itself) could represent a technology opportunity, be it industry level or firm specific. However, this representation would result in an extremely large P(K) that was essentially intractable. Therefore, some constraints were added to the CTOs to form the TOS in the next substep.

2) Filtering Out Unsuitable CTOs to Form a TOS: As observed in [13], the titles of patents were good proxies for technology opportunities. Therefore, we used the knowledge networks constructed from the titles of historical patents to represent these historical technology opportunities, and each historical technology opportunity was mapped to its corresponding patent.

Since the goal of the case study was to develop new patents from technology opportunities, the potential technology opportunities should be similar in form to the historical opportunities. However, there were no titles for the "future patents." By noting that the syntactic relations in the titles of patents form connected subnetworks (as was the case for historical technology opportunities), we represented the firm-specific CTOs using connected subnetworks of the PIKN. The connectedness constraint of the subnetworks was rational since unconnected knowledge elements could not express logical, congruent ideas. As the mean number of SAOs in a title was close to 6 in the neural network dataset, we constrained the number of nodes in the CTOs to 6. CTOs not conforming to the two constraints were filtered out, resulting in a much more manageable TOS.

C. Finding the Evaluation Model for CTOs

1) Gathering the Evaluation Criteria for CTOs: To evaluate the firm-specific technology opportunities represented in the last step, we chose to design and explore many criteria because there had been few TOD studies that used subnetworks to represent technology opportunities. The 25 criteria, as listed in Table I, were gathered based on the previous theoretical and empirical studies, and were classified into three categories: industry-level criteria, firm-specific criteria, and patent-level criteria. All the criteria were calculated with NetworkX in Python. Due to space limitation, the detailed rationale and calculations of these criteria were described in Appendix A.

2) Assessing the Importance of the Evaluation Criteria With Historical Technology Opportunities: In this substep, the HIKN and HFKN (with patents up to 2015) were used to construct the evaluation model of CTOs, and the historical technology opportunities were obtained from all 3735 patents in the neural network domain in 2016–2017 (755 patents for 2016 and 2969 for 2017).

We used the above criteria as predictor variables for historical firm-specific technology opportunity s_i because they were all available once s_i was developed into patent p_i. However, the technical value of s_i had not been shown at the point of application of p_i. In the case study, forward citations, a widely accepted ex post measure of the value of patents [33], [45], [46], were used to compute the predicted variable, Cited, in the evaluation model. Specifically, Cited equaled 1 for patents with forward citation counts of at least 1 and equaled 0 for those with no forward citation. The citation counts were measured at the end of 2020. In the case study, all patents in 2016–2017 were used as samples in building the evaluation model, resulting in 2455 patents with Cited $= 0$ and 1280 patents with Cited $= 1$.

We used random forest [47], a machine learning method, to build an evaluation model for historical technology opportunities. The random forest model was coded with Scikit-learn in Python.

We used all 25 evaluation criteria defined in the last substep as the predictor variables and Cited as the predicted variable in the random forest classification model. Of all the patent samples, 80% of patents were randomly drawn into a training set and 20% into a testing set. The synthetic minority oversampling technique was used to train the unbalanced dataset, and grid search was applied to find the best model parameters, such as the number of decision trees, the maximum tree depth, and the maximum number of tree leaves. The parameters were "optimized" such that the recall for the category "Cited $= 1$" was high (so that we did not miss good opportunities), while the precision was reasonable. The random forest model was coded with Scikit-learn in Python. The resulting evaluation model has a recall of 0.74 and a precision of 0.48 for good technology opportunities (Cited $= 1$). We compensated for the relatively low precision by picking the highest ranked opportunities in the last step.

Fig. 3 shows, in descending order, the importance of the top 20 important evaluation criteria in the model, which accounted for 90% of the importance weights. Fig. 3 shows that industry-level, firm-specific, and patent-specific criteria all impacted the value of the patents, with combined importance weights of 36.2%, 21.4%, and 32.1%, respectively. Among these important criteria, back_cite, claims, county_CN, and Year were patent specific and did not measure the technology opportunities that preceded patent applications. However, industry-level and firm-specific criteria that directly measured the technology opportunities still explained more than half of the value of the patents, which meant the value of technology opportunities correlated strongly with that of the corresponding patents. Among the industry-level and firm-specific criteria, firm experience (A_year_min), the mean eigenvalue of nodes (eigen_mean), and the mean PageRank of nodes (pr_mean) are among the most important. The most important criteria could be applied to analyze the best opportunities.

D. Searching for the Best Technology Opportunities

1) Building an Optimization Model for CTOs: The evaluation model constructed with the random forest in Section IV-C was used to predict the potential value of future technology opportunities, assuming that it did not change over the past few years. Specifically, the evaluation model was set as the objective function Z = f(s, c) in Model (*M*), where s and c represent the future technological opportunities and the evaluation criteria, respectively. Note that we used the percentage of votes obtained from the random forest as f instead of the 0–1 categorical values to further identify "good opportunities." The constraints in (*M*) referred to the connectivity of the subnetworks and the number of nodes in the subnetworks (which was 6).

There was one remaining issue in calculating Z with the evaluation model, which was the unavailability of a few evaluation criteria, including Year, country, claims, IPCs, and back_cite. These criteria were unavailable in the TOD process because they could be calculated only when the patent developed from the future technology opportunity *s* was filed. In this case, we set Year to 2017 and country_CN to 1 (because Kuangshi Tech was likely to file patents in China). The values of other criteria were set to their means of all sample patents in 2016–2017, which is a standard treatment for missing values. This process completed the construction of the optimization model.

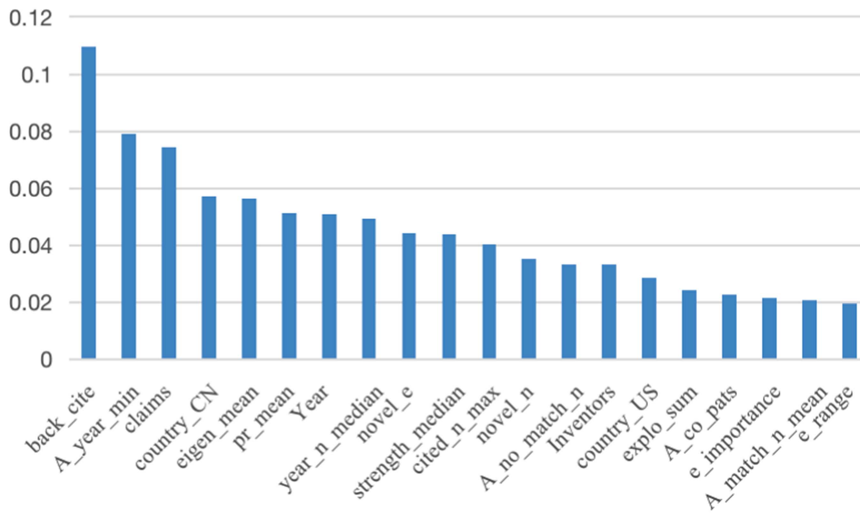

Fig. 3. Importance of evaluation criteria with historical technology opportunities.

2) Searching for the Best Technology Opportunities by Optimization Techniques: As mentioned in Section III-D, it was extremely difficult to obtain the exact optimum solution for Model (*M*). However, evolutionary algorithms, such as ACO, could be used to obtain multiple good solutions [48].

In the case study, we adopted the ACO algorithm proposed in [13] for searching knowledge networks. We set the initial pheromone concentration to 0.1, the pheromone factor to 2, the heuristic factor to 3, the pheromone volatilization coefficient to 0.95, the upper limit of pheromone to 4, and the lower limit to 0.01. Moreover, the number of ants (solutions) in each generation was set to 1200, and the pheromones of the top 120 solutions of each generation were increased. At the end of each generation, the best 500 solutions over all generations were updated and saved as the candidates in the final CTO ranking. We ran the ACO for 500 generations. The ACO algorithms were implemented with Python.

We showed the top ten technology opportunities in the neural network for Kuangshi Tech, the focal firm (see Table II). Due to space limitations, only the nodes are listed. It can be seen that the best technology opportunities had a high percentage of positive votes (Z values), demonstrating high potential value for the focal firm.

3) Interpreting the Best Technology Opportunities: We demonstrated the interpretation of the best technology opportunities with Opportunity #1 (O_1 in short), which had the highest *Z* value. The network representation for O_1 is shown in Fig. 4.

O_1 could be interpreted in multiple ways. Professional developers in the neural network and computer technology would be stimulated to form their own creative ideas, sometimes adding or deleting some knowledge elements in O_1. However, for those who were not inspired (such as the authors, who were familiar with neural network algorithms but not with computer architecture and systems), matching the partial ideas of the subnetwork of O_1 with their related patents would aid the interpretation and ideation process.

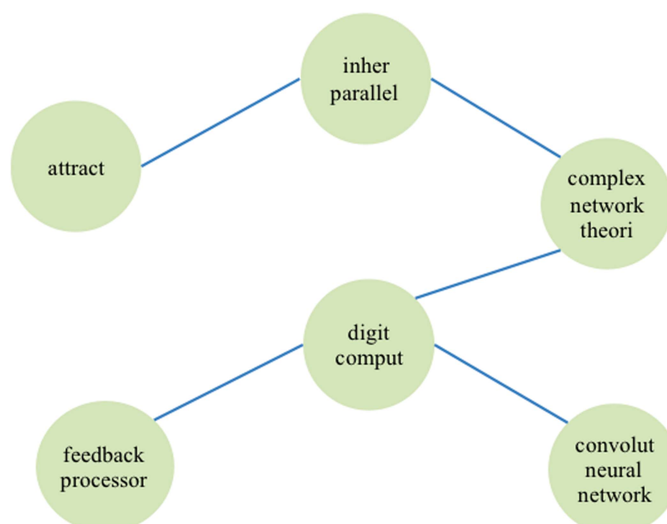

Fig. 4. Subnetwork of opportunity #1 (O_1).

The detailed matching process for O_1 was described in Appendix B. In short, O_1 could be broken down into three partial ideas: “feedback processor - digit comput,” “convolut neural network - digit comput,” and “attract - inher parallel - complex network theori - digit comput.” The first partial idea could be interpreted as using feedback processors in digital computing. The second partial idea, i.e., “convolut neural network - digit comput,” could be that CNNs and digital computers should be linked together. The third partial idea, i.e., “attract - inher parallel - complex network theori - digit comput,” was that the complex network theory could improve the performance of digital computers with (attractive) inherent parallelism. These partial ideas could be integrated to create a new technology opportunity, namely, O_1, which could be a digital computing device with feedback processors implementing one or more CNNs, which could take advantage of complex network theory

TABLE II
TOP TEN TECHNOLOGY OPPORTUNITIES IN NEURAL NETWORK TECHNOLOGY

No.	List of nodes in the technology opportunities	Z
1	[attract, inher parallel, complex network theori, digit comput, feedback processor, convolut neural network]	0.954
2	[radar wave, deep learn, android malici applic, method, continu posit airway pressur, breath disord]	0.934
3	[bear posit, eddi current sensor, horizont, imag, actual class, defect valu]	0.914
4	[long short term memori predict model, cloud comput data center energi save method, data block, analysi, convolut neural network, devic]	0.914
5	[gray threshold, abdomin bodi imag liver segment method, splice data, method, signal process, deep neural network]	0.907
6	[imag, posit, refer node, signal strength, channel, linear constraint]	0.907
7	[reagent, cell, respons, convolut neural network, behavior data, activ unit]	0.907
8	[human hand, model, output unit, convolut neural network, output neuron, network]	0.907
9	[bottomup salienc element, bayesian framework, processor, comput readabl storag media, convolut neural network, execut]	0.907
10	[convolut process oper, train imag, learn devic, convolut neural network, target function, complex coeffici]	0.907

* This list only contains the nodes and objective values of each of the top 10 technology opportunities.

and attractive inherent parallelism, therefore improving the computing efficiency. O_1 contained key ideas that could be developed into one or more patents.

It should be noted that the breakdown, analysis, and integration of partial ideas in the technology opportunities were a trial-and-error process. Other breakdowns or understandings of the knowledge elements were possible and might result in different interpretations.

To assess the quality of the technology opportunities, we presented the top technology opportunities in Table II to three research scholars in the Department of Computer Science at Beijing University of Technology who were experts in computer science and neural network technology. Each of the ten opportunities was presented in a subnetwork format with no further explanations (which could affect the thought process of the experts). The questions for the experts concerned the meaningfulness, novelty, and potential for implementation of these opportunities. All experts agreed that most of the ideas contained in these opportunities were very novel, meaningful, and feasible. One expert thought that Opportunity #2 was difficult to foresee or conceive but could be realistic in the future; another judged Opportunity #8 as ordinary. In general, most technology opportunities were of high quality and could be developed into inventions. Later, we presented these technology opportunities to two professionals in Kuangshi Tech but was asked to provide patent information relevant to the concept nodes for better interpretation of their meanings. They found that the associate patents had good quality and made the opportunities more specific and interesting. They gave rankings of the ten technology opportunities, which were highly correlated with the Z scores in Table II, except that Opportunities #4 and #8 were considered ordinary.

Finally, as mentioned in Section III-D, some important evaluation criteria could be applied for the best technology opportunities and their components to help understand their features. As seen in Table III, the node "convolut neural network" stood out as a distinctive feature of O_1 since it had the highest value of eigenvalue and PageRank centrality, strength, and citations of all the nodes in PIKN. This meant that CNN was a highly conventional knowledge element, which would enhance the usefulness and value of O_1. Indeed, the CNN was an extremely popular source of opportunities, as six out of the top ten opportunities contained it as a knowledge element (c.f. Table II of the main article). Meanwhile, the nodes "attract," "inher parallel," and "feedback processor" were on the other end of the spectrum, well below the mean values of conventionality criteria. Although the earliest application years of these nodes were not recent, they provide O_1 with combinational novelty. We also analyzed the other top ten technology opportunities and found a similar pattern of combining highly conventional nodes with unconventional nodes. This "pattern of success" was proposed by Uzzi et al. [49] in their study of academic papers. Whether this phenomenon is common to good opportunities could be an interesting research question.

TABLE III
SOME IMPORTANT EVALUATION CRITERIA FOR NODES IN O_1

Node (knowledge element)	Eigenvalue	PageRank	Earliest application year	Strength	Citations
Attract	0.00014	7.81E-06	2009	2.16657	2
inher parallel	1.19E-05	9.96E-06	2009	2.71000	2
complex network theori	0.001886	2.93E-05	2009	3.41023	5
digit comput	0.002476	1.97E-05	1991	5.52122	158
feedback processor	0.000435	6.62E-06	1991	1.00000	0
convolut neural network	0.339033	0.017813	2008	9.69097	2544
Maximum of all nodes	0.339033	0.017813	2020	9.69097	2544
Mean of all nodes	0.001158	1.277E-05	2015.2	1.531394	4.928053

V. DISCUSSION

The case study demonstrated the strengths of the proposed general methodology for TOD, especially its "evaluate–search" strategy of using historical data to learn the evaluation models for CTOs and searching the TOS for the best opportunities.

First, it showed the advantage of creating a less restrictive TOS and learning the importance of evaluation criteria. Since the purpose of the case study was to discover firm-specific technology opportunities for a focal firm, we investigated the top ten technology opportunities in terms of compatibility with firm capability. If we had used the "define–evaluate" strategy, only those technology opportunities that were compatible with the firm's technology capabilities would be defined as CTOs. In the case study, however, we found that the best firm-specific technology opportunities might not necessarily conform to the compatibility criteria. In fact, O_1 had only one knowledge element that matched Kuangshi Tech, which was a "convolut neural network." We calculated the unmatched nodes (*A_no_match_n*) for the ten opportunities and found that the average number of matched nodes (knowledge elements) was 1.0, and O_2 did not share any common nodes with Kuangshi Tech's PFKN. One key reason was that *A_no_match_n* (the most important firm compatibility criterion) had a weight of less than 0.04 and played a minor role in determining the value of technology opportunities, as shown in Fig. 3. Therefore, contrary to the assumptions of many firm-specific TOD methods, this case study showed that good firm-specific technology opportunities might not be highly compatible with the firm's capabilities (at least in the field of neural networks). In terms of the general methodology, the proposed model demonstrated that a less restrictive TOS combined with data-driven evaluation models of CTOs could facilitate finding better opportunities.

Second, it showed the advantage of using historical data to build evaluation models for technology opportunities. Because of the surprising unimportance of compatibility criteria in the case study, it was speculated that the value of the firm-specific technology opportunities in other industries might rely more on a firm's knowledge compatibility. In general, the importance of evaluation criteria could be industry specific and might vary greatly among industries. Therefore, we advise R&D teams using historical data to learn and build empirical evaluation models for the technology opportunities in their focal industries before searching.

Finally, the case study showed the importance of using optimization techniques to search for the best technology opportunities. For such a large-scale TOS in the case study, numerous CTOs existed. Using the "define–evaluate" strategy to reduce the search space would sacrifice the quality of the technology opportunities. Optimization research offered a plethora of techniques, many of which could be used to find technology opportunities of high quality in a manageable time.

VI. CONCLUSION

Technology opportunities are important sources of creativity and future technological advances. However, they are "hiding" in the existing knowledge structure. To discover the most valuable technology opportunities, in this article, we proposed a general methodology and implemented the proposed methodology in a case study that discovered firm-specific technology opportunities in neural network technology. The proposed methodology demonstrated the following advantages over the dominant "define–evaluate" strategy.

1) Less restrictive TOS enabled the discovery of better opportunities.
2) Evaluation models of technology opportunities based on the historical data would be more suitable for particular industries and situations.
3) Optimization techniques could aid the R&D team in finding the best technology opportunities from a large number of candidates.

In addition, the TOD in the case study showed some specific and interesting results. For one, the technology opportunities represented by subnetworks of technical concepts were specific to an inventive problem and were easy to interpret. For the other, the top firm-specific technology opportunities did not have to match well with the firm's capabilities.

Despite the contributions of the present study, limitations of the general methodology exist. In particular, it remains difficult to build evaluation models for technology opportunities with historical data because there is no consensus on how to measure the "accurate value" of technology opportunities. Even if the commercial value of products can be collected, attributing or distributing the value of products to individual technologies and associated technology opportunities is by no means an easy task. The opportunity evaluation problem also makes it difficult to validate current TOD methods. We believe that the analysis of the "opportunity–technology–product" value chain based on the heterogenous technological and market data may improve the evaluation of technology opportunities, and this will be a future research direction.

APPENDIX

A. Rationale and Formulae of the Evaluation Criteria

This section described the detailed rationale and formulae of the evaluation criteria in Table I.

Before describing these criteria, we introduce the necessary notations and naming conventions used in Table I and later sections. Let s_i denote a historical technology opportunity *i*; p_i and a_i denote the patent and its assignees (copyright holders) corresponding to s_i. In the symbols of criteria, "*n*" and "*e*" refer to the nodes and edges, respectively, and "*A*" refers to the assignees. Other abbreviations in the symbols are self-explanatory.

Industry-level criteria were calculated from the network information of the HIKN only. There were three subcategories of industry-level criteria: novelty, conventionality, and network position. Most of these criteria were designed based on the related results in the literature showing that novelty [32], [49], conventionality [49], and network position [50] in general impacted the value of technology opportunities. The industry-level criteria were calculated as follows.

1) Recent nodes (*recent_n*) was calculated as the square root of the number of nodes with the earliest application year more recent than 90% of all nodes in HIKN. This criterion measured how much emerging technology was contained in a technology opportunity.
2) Novel nodes (*novel_n*) and novel edges (*novel_e*) were calculated as the square root of the number of nodes or edges that existed only in s_i but not in HIKN. These criteria were considered more novel than *recent_n*.
3) Median application year (*year_n_median*) was calculated as the median of all earliest application years in all the nodes in s_i. The median number of application years reflected the general novelty of knowledge elements.
4) Year range of edges (*e_range*) was calculated as the median of the (application) year ranges of all the edges in s_i. It measured the overall time span of s_i. We posited that long-lasting SAOs would increase the conventionality of the technology opportunity.
5) Mean strength of nodes (*strength_median*) and maximum weight of edges (*weight_max*) were calculated as the mean decayed strength of all the nodes in s_i and the maximum weight of all the edges in s_i, respectively. The higher these criteria, the more conventional knowledge was contained in the technology opportunity, increasing its feasibility. Decayed node strengths gave older knowledge (SAOs) less importance in HIKN. The formula of strength is shown in (1), where *n* is the number of SAOs involving the focal node, $\text{year}_{\max}$ is the last year in HIKN (which was 2015 in the case), and the decay factor α is 0.9

$$\text{strength } = \sum_{j=1}^{n} \alpha^{\text{year}_{\max} - \text{year}_j}. \tag{1}$$

Edge importance (e_importance) was calculated as follows:

$$\text{e\_importance } = \max(\text{range}_j + \text{max\_freq}_\text{j}) \tag{2}$$

where range_j is the year range for edge *j*, max_freq_j denotes the maximum yearly patent count for edge_j, and the maximum is calculated for all edges in s_i. The idea was to combine both the longevity and the climax of an edge to form a single criterion.
6) Maximum citations of nodes (cited_n_max) was calculated as follows: First, the total number of forward citations of all the patents involving each node in s_i was computed. Then, the maximum of these totals was taken as *cited_n_max*. A higher value would indicate a better historical impact of the knowledge elements in a technology opportunity.
7) Mean eigenvalue of nodes (*eigen_mean*) and mean PageRank of nodes (*pr_mean*) were calculated as the mean eigenvalue and PageRank of all nodes in s_i, respectively. These popular statistics measure the structural positions of nodes in networks and can be used to measure the conventionality of a technology opportunity. The PageRank of a patent in the citation network had a significant impact on the forward citations it received [51].

Firm-specific criteria were calculated from the network information of both the HFKN and HIKN. These criteria can be grouped into two subcategories: firm's R&D capability and compatibility with firm's capability. Again, some criteria had support in the literature, while others were of explorative nature. The firm-specific criteria were calculated as follows.

1) Inventors (*inventors*) were calculated as the number of inventors for p_i, the corresponding patent of s_i. It has been shown that the value of a patent increases with the number of inventors and their experience [45], [52].
2) Experience (*A_year_min*) was calculated as the earliest application year of all the patents filed by the assignee (*firm*) of p_i. This criterion is correlated with a firm's patent portfolio, which was shown to positively influence the value of a patent [45].
3) Codeveloped patents (*A_co_pats*) were calculated as the number of coapplied patents of the assignee of p_i. The assumption was that the more patents a firm codeveloped with other firms, the better the firm's ability to leverage heterogeneous technical knowledge, therefore capitalizing on new technology opportunities. This criterion was used by Lee et al. [30] and Ernst [53].

4) Binary match with nodes (*A_is_match_n*) was set to 1 if at least one node in s_i was contained in the HFKN of a_i and 0 otherwise. The criterion measured whether the technology opportunity matched the existing capability (patents) of the firm.
5) Mean match with nodes (*A_match_n_mean*) was calculated as follows. First, the number of patents in the HFKN of a_i containing each node in s_i was calculated; then, the mean value of all nodes in s_i was taken as *A_match_n_mean*. Compared with *A_is_match_n*, *A_match_n_mean* was assumed to be a more accurate measure.
6) Maximum match with nodes (*A_match_n_max*) was calculated in the same way as *A_match_n_mean* except that the maximum was used instead of the mean.
7) Unmatched nodes (*A_no_match_n*) was calculated as the number of nodes in s_i that were not contained in the HFKN of a_i. *A_is_match_n*, *A_match_n_mean*, *A_match_n_max*, and *A_no_match_n* measured the fit between a technology opportunity and the technological capability of the firm. It was assumed that the more knowledge in s_i was in the firm's technological capability, the more successful the patent would be.
8) Total knowledge exploitation (explo_sum) was calculated as follows:

$$\text{explo\_sum} = \sum_{j=1}^{m(p_i)} n(p_i)_j \tag{3}$$

where $m(p_i)$ is the number of IPCs of p_i and $n(p_i)j$ is the number of patents owned by a_i that belong to IPC j. explo_sum measured the knowledge exploitation by the firm on the technological class level and was firm-specific criterion akin to firm capability.
9) Maximum citations of nodes (*cited_n_max*) were calculated as follows. First, the total number of forward citations of all the patents involving each node in s_i was computed. Then, the maximum of these totals was taken as *cited_n_max*. A higher value would indicate a better historical impact of the knowledge elements in a technology opportunity.
10) Codeveloped nodes (*A_co_match*) were calculated as the number of nodes (in s_i) whose assignees in HIKN had codeveloped patents with a_i prior to p_i. It was assumed that if more nodes of the technology opportunity were "in the hands of" the partners of the firm, the in-house capability of the firm would be compensated in developing the new technology.

Patent-specific criteria included application year (*Year*), priority country (*country*), claims in the patent (*claims*), number of IPCs (*IPCs*), and backward citations (*back_cite*). These criteria were introduced because our technology opportunities were mapped to corresponding patents, and the above criteria were shown to significantly affect the quality or impact of patents. Therefore, these criteria were expected to improve the predictive power of the evaluation mode.

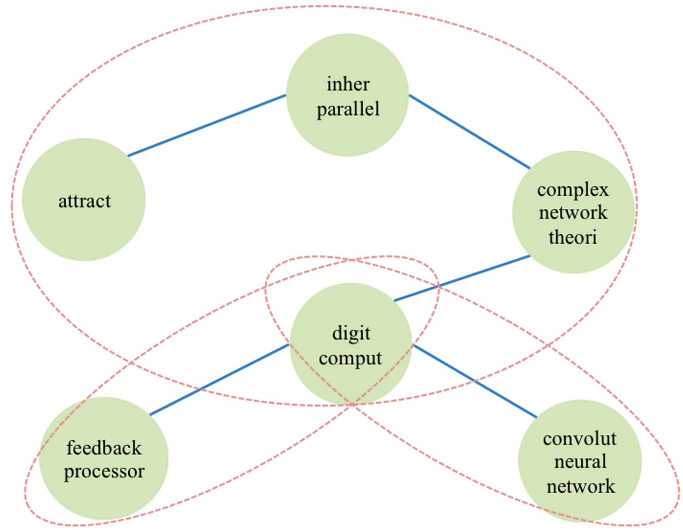

Fig. 5. Subnetwork of Opportunity #1.

1) Application year (*Year*) was the application year of p_i and was often used as a control variable in regression models for patent quality.
2) Priority country (*country*) was the priority country of p_i, which was categorical. In a later stage, it would be represented by binary variables, such as country_CN and country_US, to indicate patents filed in major priority countries and the World Intellectual Property Organization (*country_WO*). We used this criterion to explore the country effect on the value of technology opportunities.
3) Claims in the patent (*claims*) were the number of claims in p_i. Studies have shown that the number of claims has a positive effect on patent quality [32].
4) Number of IPCs (*IPCs*) was the number of the four-digit IPC subclasses assigned to p_i. Patents with more IPCs tended to be more generic and might have wider technological coverage and, therefore, were more likely to be cited [54].
5) Backward citations (*back_cite*) was the number of patents cited by p_i. Many studies found that this criterion positively influenced the forward citation a patent received [32] because patents with large back_cite tended to improve upon many inventions, thus increasing their usefulness.

B. Detailed Interpretation of Technology Opportunity O_1

This section described the detailed process of matching the partial ideas in O_1 (Fig. 5; built on Fig. 4 in the main article) with their related patents in the knowledge networks and the subsequent interpretation.

In a possible interpretation, O_1 could be broken down into three partial ideas: "feedback processor - digit comput," "convolut neural network - digit comput," and "attract - inher parallel - complex network theori - digit comput," as shown in the ellipses in Fig. 5. The first partial idea could be interpreted as *using feedback processors in digital computing*. This idea was not novel, as the edge "feedback processor - digit comput" in PIKN came from a European Patent Application 91116405 (published in April 1992 and was in the patent set of the PIKN), titled "Digital architecture for an artificial neural network," proposed using feedback and feedforward processors in a digital neural network computing architecture.

The second partial idea, i.e., "convolut neural network - digit comput," could be that *CNNs and digital computers should be linked together*. CNNs implemented as software programs running on general-purpose computers were conventional ideas. However, CNN-based special-purpose digital computers seemed unusual and could be desirable to increase computational efficiency and reduce silicon area. To evaluate this idea, we searched DII for recent patents having similar ideas. One such patent was U.S. Patent 11144819 titled "CNN hardware configuration." It described a hardware implementation of CNN used in certain applications for high efficiency and limited resource usage, therefore supporting the idea. Interestingly, the patent was published in November 2017 and was in the PIKN, but the patent text did not contain terms, such as "digit∗ comput∗" and, therefore, did not form the edge "convolut neural network - digit comput" in the PIKN. Instead, the edge came from U.S. Patent Application 20190114748 in which CNNs were used for "computer-generated digital content." By checking the term-clumping procedure in Section IV-A, we found that the latter phrase was merged into the node "digit comput" by virtue of similarity. This finding showed that term clumping could improve the connectivity of SAO-based knowledge networks such that more opportunities could be discovered.

The third partial idea, i.e., "attract - inher parallel - complex network theori - digit comput," was that *the complex network theory could improve the performance of digital computers with (attractive) inherent parallelism*. Although this idea seemed complex and unconventional, some inventions actually started incorporated this idea. For example, in a recent U.S. Patent Application 20210049465, titled "Self-optimizing and self-programming computing systems: A combined compiler, complex networks, and machine learning approach," complex network theory was applied to detect communities of heterogeneous computing units, and, as the application stated, "tasks suitable to run on CPUs are formed by community detection to minimize data movement overhead." Moreover, US20210049465 claimed to exploit the parallelism with complex network theory and other techniques. Since this patent application was published in February 2021 (not in our PIKN), our prediction of the new technology opportunity contained in O_1 indeed came true.

Although most of the above partial ideas had been proposed in previous patents, they could be integrated to create a novel technology opportunity, namely, O_1. Hence, the interpretation of O_1 could be *a digital computing device with feedback processors implementing one or more CNNs, which could take advantage of complex network theory and attractive inherent parallelism, therefore improving the computing efficiency*. O_1 contained key ideas that could be developed into one or more patents.

References

[1] A. K. Klevorick, R. C. Levin, R. R. Nelson, and S. G. Winter, "On the sources and significance of interindustry differences in technological opportunities," *Res. Policy*, vol. 24, pp. 185–205, 1995.

[2] S. A. Zahra, "Goverance, ownership, and corporate entrepreneurship: The moderating impact of industry technological opportunities," *Acad. Manage. J.*, vol. 39, no. 6, pp. 1713–1735, 1996.

[3] M. Nieto and P. Quevedo, "Absorptive capacity, technological opportunity, knowledge spillovers, and innovative effort," *Technovation*, vol. 25, pp. 1141–1157, 2005.

[4] M.-Y. Wang, S.-C. Fang, and Y.-H. Chang, "Exploring technological opportunities by mining the gaps between science and technology: Microalgal biofuels," *Technol. Forecasting Social Change*, vol. 92, pp. 182–195, 2015.

[5] A. L. Porter and M. J. Detampel, "Technology opportunities analysis," *Technol. Forecasting Social Change*, vol. 49, no. 3, pp. 237–255, 1995.

[6] J. Yoon and K. Kim, "Detecting signals of new technological opportunities using semantic patent analysis and outlier detection," *Scientometrics*, vol. 90, no. 2, pp. 445–461, 2012.

[7] C. Son, Y. Suh, J. Jeon, and Y. Park, "Development of a GTM-based patent map for identifying patent vacuums," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 2489–2500, 2012.

[8] X. Wang et al., "Combining SAO semantic analysis and morphology analysis to identify technology opportunities," *Scientometrics*, vol. 111, pp. 3–24, 2017.

[9] J. Yoon, H. Park, W. Seo, J.-M. Lee, B.-Y. Coh, and J. Kim, "Technology opportunity discovery (TOD) from existing technologies and products: A function-based TOD framework," *Technol. Forecasting Social Change*, vol. 100, pp. 153–167, 2015.

[10] C. Lee, D. Jeon, J. M. Ahn, and O. Kwon, "Navigating a product landscape for technology opportunity analysis: A word2vec approach using an integrated patent-product database," *Technovation*, vol. 96–97, 2020, Art. no. 102140.

[11] S. Yun, K. Song, C. Kim, and S. Lee, "From stones to jewellery: Investigating technology opportunities from expired patents," *Technovation*, vol. 103, 2021, Art. no. 102235.

[12] J. Choi, B. Jeong, and J. Yoon, "Technology opportunity discovery under the dynamic change of focus technology fields: Application of sequential pattern mining to patent classifications," *Technol. Forecasting Social Change*, vol. 148, 2019, Art. no. 119737.

[13] H. Ren and Y. Zhao, "Technology opportunity discovery based on constructing, evaluating, and searching knowledge networks," *Technovation*, vol. 101, 2021, Art. no. 102196.

[14] C. Lee and G. Lee, "Technology opportunity analysis based on recombinant search: Patent landscape analysis for idea generation," *Scientometrics*, vol. 121, no. 2, pp. 603–632, 2019.

[15] S. Lee, B. Yoon, and Y. Park, "An approach to discovering new technology opportunities: Keyword-based patent map approach," *Technovation*, vol. 29, no. 6/7, pp. 481–497, 2009.

[16] B. Yoon and C. L. Magee, "Exploring technology opportunities by visualizing patent information based on generative topographic mapping and link prediction," *Technol. Forecasting Social Change*, vol. 132, pp. 105–117, 2018.

[17] N. Shibata, Y. Kajikawa, and I. Sakata, "Extracting the commercialization gap between science and technology—Case study of a solar cell," *Technol. Forecasting Social Change*, vol. 77, no. 7, pp. 1147–1155, 2010.

[18] X. Li, Q. Xie, T. Daim, and L. Huang, "Forecasting technology trends using text mining of the gaps between science and technology: The case of perovskite solar cell technology," *Technol. Forecasting Social Change*, vol. 146, pp. 432–449, 2019.

[19] C. Lee, B. Kang, and J. Shin, "Novelty-focused patent mapping for technology opportunity analysis," *Technol. Forecasting Social Change*, vol. 90, pp. 355–365, 2015.

[20] A. Rodriguez et al., "Patent clustering and outlier ranking methodologies for attributed patent citation networks for technology opportunity discovery," *IEEE Trans. Eng. Manage.*, vol. 63, no. 4, pp. 426–437, Nov. 2016.

[21] K. Song, K. S. Kim, and S. Lee, "Discovering new technology opportunities based on patents: Text-mining and F-term analysis," *Technovation*, vol. 60–61, pp. 1–14, 2017.

[22] X. Han, D. Zhu, X. Wang, J. Li, and Y. Qiao, "Technology opportunity analysis: Combining SAO networks and link prediction," *IEEE Trans. Eng. Manage.*, vol. 68, no. 5, pp. 1288–1298, Oct. 2021.

[23] J. Ma, N. F. Abrams, A. L. Porter, D. Zhu, and D. Farrell, “Identifying translational indicators and technology opportunities for nanomedical research using tech mining: The case of gold nanostructures,” *Technol. Forecast. Soc. Change*, vol. 146, pp. 767–775, 2019.

[24] D. Rotolo, D. Hicks, and B. R. Martin, “What is an emerging technology,” *Res. Policy*, vol. 44, pp. 1827–1843, 2015.

[25] S. F. Carley, N. C. Newman, A. L. Porter, and J. G. Garner, “An indicator of technical emergence,” *Scientometrics*, vol. 115, no. 1, pp. 35–49, 2018.

[26] I. Park, B. Yoon, S. Kim, and H. Seol, “Technological opportunities discovery for safety through topic modeling and opinion mining in the fourth industrial revolution: The case of artificial intelligence,” *IEEE Trans. Eng. Manage.*, vol. 68, no. 5, pp. 1504–1519, Oct. 2021.

[27] Y. Park and J. Yoon, “Application technology opportunity discovery from technology portfolios: Use of patent classification and collaborative filtering,” *Technol. Forecasting Social Change*, vol. 118, pp. 170–183, 2017.

[28] B. Yoon and Y. Park, “A systematic approach for identifying technology opportunities: Keyword-based morphology analysis,” *Technol. Forecasting Social Change*, vol. 72, no. 2, pp. 145–160, 2005.

[29] Y. Lee, S. Y. Kim, I. Song, Y. Park, and J. Shin, “Technology opportunity identification customized to the technological capability of SMEs through two-stage patent analysis,” *Scientometrics*, vol. 100, no. 1, pp. 227–244, 2014.

[30] J. Lee, N. Ko, J. Yoon, and C. Son, “An approach for discovering firm-specific technology opportunities: Application of link prediction to F-term networks,” *Technol. Forecasting Social Change*, vol. 168, 2021, Art. no. 120746.

[31] W. Seo, J. Yoon, H. Park, B.-Y. Coh, J.-M. Lee, and O.-J. Kwon, “Product opportunity identification based on internal capabilities using text mining and association rule mining,” *Technol. Forecasting Social Change*, vol. 105, pp. 94–104, 2016.

[32] S. Kaplan and K. Vakili, “The double-edged sword of recombination in breakthrough innovation,” *Strategic Manage. J.*, vol. 36, no. 10, pp. 1435–1457, 2015.

[33] H. J. Jung and J. Lee, “The quest for originality: A new typology of knowledge search and breakthrough inventions,” *Acad. Manage. J.*, vol. 59, no. 5, pp. 1725–1753, 2016.

[34] H. Evanschitzky, M. Eisend, R. J. Calantone, and Y. Jiang, “Success factors of product innovation: An updated meta-analysis,” *J. Product Innov. Manage.*, vol. 29, pp. 21–37, 2012.

[35] P. Therrien, D. Doloreux, and T. Chamberlin, “Innovation novelty and (commercial) performance in the service sector: A Canadian firm-level analysis,” *Technovation*, vol. 31, no. 12, pp. 655–665, 2011.

[36] D. Kafetzopoulos, K. Gotzamani, and V. Gkana, “Relationship between quality management, innovation and competitiveness: Evidence from Greek companies,” *J. Manuf. Technol. Manage.*, vol. 26, no. 8, pp. 1177–1200, 2015.

[37] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York, NY, USA: Freeman, 1979.

[38] V. Zwass, “Neural network, Encyclopedia Britannica,” 2020. [Online]. Available: https://www.britannica.com/technology/neural-network

[39] AMiner, 2021. [Online]. Available: https://www.aminer.cn/

[40] G. Cascini, A. Fantechi, and E. Spinicci, “Natural language processing of patents and technical documentation,” in *Proc. Int. Workshop Document Anal. Syst.*, 2004, vol. 3163, pp. 508–520.

[41] S. Choi, J. Yoon, K. Kim, J. Y. Lee, and C.-H. Kim, “SAO network analysis of patents for technology trends identification: A case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells,” *Scientometrics*, vol. 88, pp. 863–883, 2011.

[42] S. Choi, H. Park, D. Kang, J. Y. Lee, and K. Kim, “An SAO-based text mining approach to building a technology tree for technology planning,” *Expert. Syst. Appl.*, vol. 39, no. 13, pp. 11443–11455, 2012.

[43] H. Park, K. Kim, S. Choi, and J. Yoon, “A patent intelligence system for strategic technology planning,” *Expert. Syst. Appl.*, vol. 40, no. 7, pp. 2373–2390, 2013.

[44] C. C. Trumbach and D. Payne, “Identifying synonymous concepts in preparation for technology mining,” *J. Inf. Sci.*, vol. 33, no. 6, pp. 660–677, 2007.

[45] E. Ferrucci and F. Lissoni, “Foreign inventors in Europe and the United States: Diversity and patent quality,” *Res. Policy*, vol. 48, no. 9, 2019, Art. no. 103774.

[46] M. Trajtenberg, “A penny for your quotes: Patent citations and the value of innovations,” *RAND J. Econ.*, vol. 21, no. 1, pp. 172–187, 1990.

[47] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[48] C. Solnon and D. Bridge, “An ant colony optimization meta-heuristic for subset selection problems,” in *System Engineering Using Particle Swarm Optimization*. Hauppauge, NY, USA: Nova Sci., 2006.

[49] B. Uzzi, S. Mukherjee, M. Stringer, and B. Jones, “A typical combinations and scientific impact,” *Science*, vol. 342, no. 6157, pp. 468–472, 2013.

[50] R. S. Burt, “Structural holes and good ideas,” *Amer. J. Sociol.*, vol. 110, no. 2, pp. 349–399, 2004.

[51] M. S. Mariani, M. Medo, and F. Lafond, “Early identification of important patents: Design and validation of citation network metrics,” *Technol. Forecasting Social Change*, vol. 146, pp. 644–654, 2019.

[52] J. Singh and L. Fleming, “Lone inventors as sources of breakthroughs: Myth or reality?,” *Manage. Sci.*, vol. 56, no. 1, pp. 41–56, 2010.

[53] H. Ernst, “Patent information for strategic technology management,” *World Pat. Inf.*, vol. 25, no. 3, pp. 233–242, 2003.

[54] W. Schoenmakers and G. Duysters, “The technological origins of radical inventions,” *Res. Policy*, vol. 39, no. 8, pp. 1051–1059, 2010.

Haiying Ren received the Ph.D. degree in industrial engineering from the University of South Florida, Tampa, FL, USA, in 2000.

He is currently an Associate Professor of management science and engineering with the School of Economics and Management, Beijing University of Technology, Beijing, China. His articles were published on Technovation, IIE Transactions, Science Research Management, among others. His research interests include management of technology, invention and innovation, technological forecasting, patent analysis, and text mining.

Luyao Zhang received the B.S. degree in management from the School of Computer Science, Liaocheng University, Liaocheng, China, in 2020. She is currently working toward the M.S. degree in management science and engineering with the School of Management and Economics, Beijing University of Technology, Beijing, China.

Her research interests include information management and patent text mining.

Qian Wang received the B.S. degree in management from Yangzhou University, Yangzhou, China, in 2018, and the M.S. degree in management science and engineering from the School of Management and Economics, Beijing University of Technology, Beijing, China, in 2021.

Her research interests include technology foresight and natural language processing.