Breakthrough Innovation Identification Based on Topic Mutation in Cited Scientific Knowledge (Postprint)

Authors: Zhang Jinzhu, Zhang Xiaolin

Date: 2017-10-11T00:00:00+00:00

Abstract

Objective: In technology fields closely related to basic research, it is necessary to identify breakthrough innovations based on the thematic content of cited scientific knowledge in patent information. Method: Extract keywords and disciplinary classifications from patent science citations to represent cited scientific knowledge; identify themes of cited scientific knowledge within keyword co-occurrence networks and disciplinary classification combinations; propose a calculation method for thematic mutation degree based on keywords and disciplinary classifications, select themes with high mutation degree as the technology themes where breakthrough innovations occur, and thereby identify breakthrough innovations. Results: In the field of nanoelectronics, identified nanocircuit-related themes that have been confirmed as breakthrough innovations, namely nanocircuit materials and preparation themes such as nanowires, carbon nanotubes, and computable circuits, which represent the result of multidisciplinary integration across materials science, chemistry, optics, biology, and applied physics, thereby validating the effectiveness of the method. Limitations: The accuracy of cited scientific knowledge extraction, preprocessing, and matching needs to be improved, and the generalizability of the method requires validation in other fields. Conclusion: This method represents an important improvement and supplement for identifying breakthrough innovations based on patent information and can be extended and applied to identify breakthrough innovations in other technology fields closely related to basic research.

Full Text

Radical Innovation Identification Based on Topic Mutation of Scientific Knowledge Cited in Patents

Zhang Jinzhu[1], **Zhang Xiaolin**[2]
[1](School of Economics and Management, Nanjing University of Science & Technology, Nanjing 210094, China)
[2](National Science Library, Chinese Academy of Sciences, Beijing 100190, China)

Abstract

[**Objective**] In technical fields closely related to basic research, radical innovation must be identified from the thematic content of scientific knowledge cited in patents. [**Methods**] This study extracts keywords and subject classifications from patent scientific references to represent cited scientific knowledge, identifies topics within keyword co-occurrence networks and subject classification combinations, proposes a computational method for topic mutation degree based on keywords and subject classifications, selects topics with high mutation degree as technical themes where breakthrough innovation occurs, and thereby identifies radical innovation. [**Results**] In the field of nanoelectronics, the method successfully identified nano-circuit-related themes that have been confirmed as radical innovations, including nano-wires, carbon nanotubes, and computable circuits—specifically, materials and fabrication themes for nano-circuits. These represent the convergence of multiple disciplines including materials science, chemistry, optics, biology, and applied physics, thereby validating the effectiveness of the proposed method. [**Limitations**] The accuracy of cited scientific knowledge extraction, preprocessing, and matching needs improvement, and the method' s generalizability requires validation in other domains. [**Conclusions**] This method represents an important improvement and supplement to radical innovation identification based on patent information and can be extended to other technology fields closely related to basic research.

Keywords: Radical innovation; Scientific knowledge cited in patents; Topic mutation; Mutation degree; Nanoelectronics

---

1. Introduction

Radical innovation (RI) refers to discontinuous changes in technical themes such as methods, products, equipment, and materials, triggering performance leaps or functional transformations that ultimately lead to discontinuous changes in markets, products, services, and business models [1-2]. Identifying fields and topics where radical innovation may occur is crucial for planning technology development directions and priority themes, avoiding potentially obsolete technologies, and optimizing R&D deployment. Such identification provides important reference value for national science and technology planning, high-tech industry identification, and enterprise development strategies [3].

Patents are important carriers of technological innovation, and patent information analysis constitutes a key method for radical innovation identification. Existing approaches primarily utilize technical knowledge within patent information to identify breakthrough innovations, such as mutations in patent classification clustering, patent topic clustering, changes in patentee collaboration and cross-boundary cooperation, and structural changes in cited patents and cited patent categories [5]. However, in technology fields closely related to basic research—such as nanoscience, genetic engineering, medicine, and biotechnology—mutations in scientific knowledge or changes in scientific principles play a more significant role in guiding technological innovation and overcoming technical bottlenecks [6]. Therefore, radical innovation identification must begin with scientific knowledge cited in patents. For instance, most technological inventions during the Second Industrial Revolution were built upon scientific theories, such as electromagnetic induction leading to a series of electrical application technologies and X-rays leading to X-ray photography. Technical novelty is an important characteristic of radical innovation, and the novelty of cited scientific knowledge is a crucial influencing factor of technical novelty [7]. Consequently, researchers have used indicators such as science linkage, the number of newly published scientific papers cited by patents, and the temporal distribution of patent scientific citations [8-9] to calculate the overall dependence of patents on emerging scientific knowledge from patent scientific references, thereby representing technical novelty. Higher dependence indicates a greater likelihood of radical innovation. On the other hand, from the content perspective of cited scientific knowledge, researchers have identified the timing of radical innovation and representative keywords by comparing keyword differences in patent scientific citations across different time periods [10], though further identification of specific technical themes and fields from the perspective of topic mutation in cited scientific knowledge is needed. Another characteristic of radical innovation is multidisciplinary crossover, where recombination occurs between knowledge domains that normally do not interact, making breakthrough innovations more likely [11]. While patent classification combinations are typically used to identify technological convergence [12], using combinations of subject classifications from patent scientific citations to identify technical fields where radical innovation occurs requires further exploration.

Thus, this study extracts keywords and subject classifications from patent scientific references and their relationships to represent cited scientific knowledge, identifies topics within keyword co-occurrence networks and subject classification combinations, proposes computational methods for topic mutation degree based on new versus repeated keywords and new versus repeated subject classification combinations, selects topics with high mutation degree as technical themes where radical innovation occurs, and thereby improves and supplements radical innovation identification methods based on patent information.

2. Identifying Radical Innovation Through Topic Mutation of Cited Scientific Knowledge

If the topics of scientific knowledge cited in current patents mutate relative to those in previous patents, this suggests that current patents may apply new scientific principles, technologies, and methods, making them potential radical innovations. Therefore, this study uses keywords and subject classifications to represent cited scientific knowledge, identifies topics within this knowledge, calculates the mutation degree of different topics, and identifies technical themes and fields most likely to generate radical innovation. The research framework is illustrated in [Figure 1: see original paper].

[Figure 1: see original paper] Research framework for identifying radical innovation through topic mutation of cited scientific knowledge

2.1 Topic Mutation of Cited Scientific Knowledge in Keyword Co-occurrence Networks By clustering the keyword co-occurrence network, multiple keyword topics in a technical field can be obtained. Calculating the mutation degree of these topics and selecting those with high mutation degree enables identification of potential radical innovations. Since the mutation concerns scientific knowledge cited by patents, keyword co-occurrence refers not to keywords appearing in the same patent scientific reference, but to keywords appearing across all scientific references cited by a single patent. Topic identification employs the Louvain community detection algorithm [13] to cluster the keyword co-occurrence network. The following sections explain the computational methods for topic mutation degree based on keywords, including calculations based on new keywords and repeated keywords.

(1) New Keyword-Based Topic Mutation Degree Calculation

The more new keywords a cited scientific knowledge topic contains in the current time period relative to the previous period, the higher its mutation degree, indicating higher novelty of the technological innovation citing that topic and a greater likelihood of radical innovation. As shown in [Figure 2: see original paper], for time periods t and $t+1$, C and C_1 represent all keywords in each period, and C_1 represents the keyword set for topic i obtained in period $t+1$. The mutation degree calculation for C_1 is explained below. The frequency of new keywords also affects mutation degree calculation, where frequency indicates the number of patent scientific references containing that keyword, denoted as $w(k)$ and $w_1(k)$ for keyword k in periods t and $t+1$, respectively. Keywords in period t are represented by circles, while those in $t+1$ are represented by triangles.

New Keyword Quantity Mutation Rate: This is measured by the proportion of new keywords in a topic relative to all keywords in that topic in period $t+1$. A larger proportion indicates higher topic mutation degree, while a smaller proportion indicates lower mutation degree. Relative to all keywords *C* in period *t*, the number of repeated keywords in topic *i* in period *t+1* is |*C* 1 *C* |, so the number of new keywords in topic *i* is |*C* 1 | - |*C* 1 *C* |. Normalizing by the total number of keywords in topic *i* in period *t+1* (|*C* 1 |) yields the new keyword quantity mutation rate:

$$\text{New Keyword Quantity Mutation Rate} = 1 - \frac{|C_{(t+1)i} \cap C_t|}{|C_{(t+1)i}|}$$

New Keyword Frequency Mutation Rate: Building upon quantity mutation, frequency mutation considers the impact of each new keyword' s frequency on mutation degree. A larger proportion of new keyword frequency sum indicates higher mutation degree, while a smaller proportion indicates lower mutation degree. Relative to all keywords *C* in period *t*, the sum of frequencies for repeated keywords in topic *i* in period *t+1* is * *{k C 1 C } w 1(k)*, so the sum of frequencies for new keywords is {k C 1 } w 1(k) - *{k C 1 C } w 1(k).* Normalizing by the total keyword frequency sum in topic *i* in period* t+1 *(* {k C 1 } w 1(k)) yields the new keyword frequency mutation rate:

$$\text{New Keyword Frequency Mutation Rate} = \frac{\sum_{k \in C_{(t+1)i}} w_{t+1}(k) - \sum_{k \in C_{(t+1)i} \cap C_t} w_{t+1}(k)}{\sum_{k \in C_{(t+1)i}} w_{t+1}(k)}$$

(2) Repeated Keyword-Based Topic Mutation Degree Calculation

The greater the frequency change of repeated keywords in a cited scientific knowledge topic in the current period relative to the previous period, the higher the topic' s mutation degree, indicating potential breakthrough progress in the technological innovation citing that topic and a greater likelihood of radical innovation. The mutation degree of repeated keywords is calculated through repeated keyword frequency mutation rate: larger frequency changes indicate higher mutation degree, while smaller changes indicate lower mutation degree. As shown in [Figure 2: see original paper], relative to all keywords *C* in period *t*, the frequency change for repeated keywords in topic *i* in period *t+1* is * *{k C 1 C } (w 1(k) - w (k)).* Normalizing by the total keyword frequency sum in topic *i* in period* t+1 *(* {k C 1 } w 1(k)) yields the repeated keyword frequency mutation rate:

$$\text{Repeated Keyword Frequency Mutation Rate} = \frac{\sum_{k \in C_{(t+1)i} \cap C_t} (w_{t+1}(k) - w_t(k))}{\sum_{k \in C_{(t+1)i}} w_{t+1}(k)}$$

Note: Normalizing by the repeated keyword frequency sum in period *t+1* (_{k C 1 C } w 1(k)) could artificially inflate mutation degree for topics containing few repeated keywords with large frequency changes. For example, if topic i contains 100 keywords with only one repeated keyword whose frequency increases from 1 in period t to 100 in $t+1$, the frequency mutation rate would be (100-1)/100 = 0.99, which clearly does not reflect reality.

2.2 Topic Mutation of Cited Scientific Knowledge in Subject Classification Combinations Subject classification combinations of cited scientific knowledge represent knowledge domain recombination underlying patents. Calculating the mutation degree of these combinations and selecting those with high mutation degree enables identification of potential radical innovation fields. Each patent cites one or more scientific references, each with corresponding subject classifications that form subject classification combinations. Similar to keyword-based topic mutation calculation, subject classification combination mutation includes two approaches: new combination-based and repeated combination-based mutation degree calculation.

(1) New Subject Classification Combination-Based Topic Mutation Degree Calculation

The more frequently a new subject classification combination appears in the current period relative to the previous period, the higher the mutation degree of the corresponding cited scientific knowledge topic, indicating higher cross-fertilization of the technological innovation citing that topic and a greater likelihood of radical innovation in related technical fields. As shown in [Figure 3: see original paper], for time periods t and $t+1$, C and $C\ _1$ represent all subject classification combinations in each period, and $C\ \ _1$ represents the ith subject classification combination in period $t+1$. The mutation degree calculation for $C\ \ _1$ is explained below. The frequency of subject classification combinations indicates the number of patents containing that combination, denoted as $w\ (i)$ and $w\ _1(i)$ for combination i in periods t and $t+1$, respectively. Subject classifications in period t are represented by circles, while those in $t+1$ are represented by triangles.

The more new subject classification combinations appear in period $t+1$, the higher the topic mutation degree; fewer new combinations indicate lower mutation degree. The new combination-based topic mutation degree is calculated by the new combination frequency mutation rate. As shown in [Figure 3: see original paper], using the appearance count of new subject classification combinations in period $t+1$ ($w\ _1(i)$) as the basis yields:

$$\text{New Subject Classification Combination Frequency Mutation Rate} = w_{t+1}(i)$$

(2) Repeated Subject Classification Combination-Based Topic Mutation Degree Calculation

The greater the frequency change of a repeated subject classification combination in the current period relative to the previous period, the higher the mutation degree of the corresponding cited scientific knowledge topic, indicating potential breakthrough progress in the technological innovation citing that topic and a greater likelihood of radical innovation in related technical fields.

The repeated combination-based topic mutation degree is calculated through repeated combination frequency mutation rate: larger frequency changes indicate higher mutation degree, while smaller changes indicate lower mutation degree. As shown in [Figure 3: see original paper], relative to all subject classification combinations C in period t, the frequency change for repeated combination i in period *t+1* is $w_{\ 1}(i)$ - $w\ (i)$. Normalizing by the appearance frequency of combination i in period *t+1* ($w_{\ 1}(i)$) yields the repeated combination frequency mutation rate:

$$\text{Repeated Subject Classification Combination Frequency Mutation Rate} = \frac{w_{t+1}(i) - w_t(i)}{w_{t+1}(i)}$$

---

3. Empirical Analysis in the Field of Nanoelectronics

Based on the characteristics of radical innovation, the selection of a validation domain must satisfy three criteria: representative radical innovations confirmed by authoritative sources; high dependence on scientific knowledge (i.e., high science linkage); and being a hot, cutting-edge research area where radical innovations are more likely to occur.

In nanoelectronics, nano-circuits composed of nanowires, carbon nanotube and nanowire-based logic circuits, and computable circuits using single-molecule transistors have been recognized as radical innovations, ranking first in *Science* magazine' s "Breakthrough of the Year" list in 2001 [14]. Additionally, the proportion of scientific citations in nanotechnology patents exceeds the average, demonstrating the field's science-based nature and its substantial dependence on basic research [15]. Nanotechnology is also a current hot research area, with nanoelectronics being its cutting-edge domain [16]. Based on these considerations, this study validates the proposed method using nanoelectronics data, treating nano-circuits as the radical innovation in this field. If topic mutation of cited scientific knowledge occurs precisely in areas related to nanowires, carbon nanotubes, and computable circuits, the feasibility of identifying radical innovation through cited scientific knowledge topic mutation will be demonstrated.

3.1 Extraction of Cited Scientific Knowledge Patents related to nanoelectronics were retrieved and downloaded from Thomson Innovation , and patents containing non-patent references were selected. Titles of non-patent references were extracted through rule-based matching, and these titles were used to retrieve corresponding keywords and subject classifications from the Science Citation Index (SCI) database. Shirabe employed a similar method to obtain metadata for patent scientific references [17].

(1) **Patent Data Retrieval and Acquisition**: In the Derwent World Patents Index (DWPI), nanoelectronics-related patents were retrieved using classification codes, with application dates from January 1, 1995, to December 31, 2005, and patent document type codes A1, A2, A9, B1, B2, E, H. The search yielded 9,359 patents. The retrieval expression was:

```
EC=((B82Y001000)) AND ADB>=(19950101) AND ADB<=(20051231) AND KI=(A1 or A2 or A9 or B1 or B2
```

(2) **Title Extraction from Non-Patent References**: Patents without non-patent references were excluded. Among the 9,359 patents, 4,723 (approximately 50%) contained non-patent references, corresponding to 34,577 non-patent references. After removing webpages, patent applications, and unidentifiable references, 32,601 references remained. Titles were then identified from these references using multiple matching rules, successfully extracting titles from 29,355 references. The specific matching rules and corresponding counts are shown in , with matching performed in the order of rule type numbers.

Title extraction rules for non-patent references and corresponding counts

(3) **Representation of Cited Scientific Knowledge**: After removing references with title length of 1, 29,054 references remained. Matching these titles in the SCI database yielded 15,525 patent scientific references, with a matching success rate of 53.4%. The keywords, subject classifications, and their relationships from these references were used to represent cited scientific knowledge.

3.2 Identification of Keyword Topics for Radical Innovation Analysis revealed that 1998 and 2001 exhibited the largest numbers of new keywords and new subject classification combinations relative to the previous year, with the greatest growth rates, suggesting higher likelihoods of radical innovation. Moreover, nano-circuits were recognized as the top breakthrough innovation in 2001. Therefore, keywords from scientific papers cited in 2001 and their co-occurrence relationships were selected for clustering to obtain topics, and the potential for radical innovation in each topic was calculated to analyze those with high mutation degree. Clustering of 2001 keywords yielded 105 topics; the topics with highest mutation degree are described below, with the top 20 keywords by mutation degree translated into English.

(1) New Keyword-Based Radical Innovation Identification

Using formulas (1)-(3), Topic 21 exhibited a new keyword quantity mutation rate of 0.96, new keyword frequency mutation rate of 0.92, and repeated keyword frequency mutation rate of 0.63. This topic contained 113 keywords, including 109 new keywords. The top 20 keywords by mutation degree and their frequencies are shown in . This topic demonstrates multidisciplinary crossover among chemistry, optics, and nanoelectronics, enabling industrial application of nano-circuits. Specifically, "supramolecular systems" can achieve "photoinduced electron transfer," making intramolecular electron transfer and energy transfer possible and providing foundational conditions for industrial nano-circuit applications. "Fullerenes" and "silica" are important materials and sources for carbon nanotubes—tiny hollow tubes with enormous potential applications in electronics and a new structural form of "fullerene." Using "redox properties" from organic chemistry, metals and oxides can first be filled inside carbon nanotubes (e.g., "highly reduced organometallic complexes," "organic transition metal complexes," and "organic electrophiles"), and then the carbon layers can be etched away to prepare the finest nanoscale wires or entirely new one-dimensional materials for future molecular or nanoelectronics devices. Some carbon nanotubes themselves can serve as nanoscale wires, enabling micro-wires prepared using carbon nanotubes or related technologies to be placed on silicon chips for producing more complex circuits.

Top 20 keywords by mutation degree in Topic 21 (2001) (frequency changes)

Topic 5 exhibited a new keyword quantity mutation rate of 0.93, new keyword frequency mutation rate of 0.88, and repeated keyword frequency mutation rate of 0.18. This topic contained 190 keywords, including 177 new keywords. The top 20 keywords by mutation degree are shown in . This topic demonstrates multidisciplinary crossover among chemistry, biology, and nanoelectronics, enabling nanoscale material manipulation for biological applications and providing foundations for molecular transistor research—a component of nano-circuits. "Chemical force microscopy" and "atomic force microscopy" represent the core content of this topic. Atomic force microscopy is a nanoscale high-resolution scanning probe microscope, 1,000 times better than the optical diffraction limit, with the key feature that the probe tip' s radius of curvature is at the nanoscale, providing true "three-dimensional" surface mapping and serving as the most important tool for nanoscale material manipulation and imaging. It can be used to study biological macromolecules and even living biological tissues, with applications in "biomolecules" and "structural biology." Meanwhile, "chemical force microscopy" enables molecular function design. "Instrument standardization" and "equipment calibration" are critical challenges for both microscopes, and results require analysis using multiple techniques such as "chemometrics," "multivariate calibration," and "partial least squares regression." Nano-circuits similarly require atomic force microscopy and chemical force microscopy for nanoscale operations, laying foundations for material manipulation and molecular circuit research related to nano-circuits (e.g., molecular transistors).

Top 20 keywords by mutation degree in Topic 5 (2001) (frequency changes)

(2) Repeated Keyword-Based Radical Innovation Identification

Topic 22 exhibited a new keyword quantity mutation rate of 0.77, new keyword frequency mutation rate of 0.47, and repeated keyword frequency mutation rate of 0.78. This topic contained 111 keywords, including 86 new keywords. The top 20 keywords by mutation degree are shown in . This topic demonstrates multidisciplinary crossover between chemistry and materials science. In "materials science," tools such as "atomic force microscopy" and "scanning probe microscopy" are used to synthesize various "nanotubes" and nanowires through "nanolithography" and other "photolithography processes," including "silica nanotubes," "single-walled carbon nanotubes," and "gold nanowires." Gold nanotechnology is based on surface plasmon polaritons—nanoplasmonics—typically encapsulated in "silica" "thin films," with enormous potential for manufacturing nanophotonic integrated circuits, attracting widespread attention from the global microelectronics industry. Various synthesis methods for nanotubes and nanowires are closely related to "supramolecular chemistry," such as "laser ablation," sol-gel, and "chemical vapor deposition." "Self-assembled monolayers" and "thiol monolayers" are important methods for preparing nanoparticles, primarily utilizing single-molecule films as substrates to grow organic and inorganic nanoparticles on "organic surfaces." These nanomaterials and fabrication technologies are closely related to nanotubes and nanowires in nano-circuits, providing foundations for their application and development.

Top 20 keywords by mutation degree in Topic 22 (2001) (frequency changes)

Topic 20 exhibited a new keyword quantity mutation rate of 0.69, new keyword frequency mutation rate of 0.31, and repeated keyword frequency mutation rate of 0.74. This topic contained 413 keywords, including 287 new keywords. The top 20 keywords by mutation degree are shown in . This topic primarily addresses "carbon nanotubes," "nanotubes," "tubules," "microtubules," "quantum wires," "crystals," "nanowire arrays," and other nanowire fabrication materials and computations, as well as circuit-related keywords such as "electronic structure," "electronic properties," "logic cell arrays," "circuits," "field emission," and "scattering." These nanotube-related materials and circuit technologies are all closely related to nano-circuits.

Top 20 keywords by mutation degree in Topic 20 (2001) (frequency changes)

3.3 Identification of Subject Classification Topics for Radical Innovation Consistent with keyword topics for radical innovation, this section also uses 2001 data to identify subject classification topics where radical innovation occurs. New and repeated subject classification combinations are used to identify these topics, calculating the potential for radical innovation in each combination and analyzing those with high mutation degree.

(1) New Subject Classification Combination-Based Radical Innovation Identification

Using formula (4) to calculate mutation degree of new subject classification combinations, representative new combinations with frequency greater than 3 are shown in . The combinations most likely to generate radical innovation are marked in bold italics: combinations of chemistry, energy & fuels, chemical engineering, petroleum engineering, and multidisciplinary fields; as well as combinations of crystallography, materials science, characterization & testing materials science, coatings & films materials science, optics, and applied physics. Frequency changes reveal that cross-fertilization among these disciplines makes radical innovation more likely in nanoelectronics. Consistently, among new keyword topics most likely to generate radical innovation, Topic 21 demonstrates multidisciplinary crossover among chemistry, optics, and nanoelectronics, enabling industrial application of nano-circuits, corresponding to the second subject classification combination. Topic 5 shows multidisciplinary crossover among chemistry, biology, and nanoelectronics, enabling nanoscale material manipulation for biological applications and providing foundations for molecular transistor research, corresponding to the eighth subject classification combination. This demonstrates that new subject classification combinations are useful for identifying technical fields of radical innovation and can provide references for advantageous disciplinary technology field layout.

Representative new subject classification combinations and their frequency changes

(2) Repeated Subject Classification Combination-Based Radical Innovation Identification

Using formula (5) to calculate mutation degree of repeated subject classification combinations, those with higher mutation degree are shown in . The repeated combinations most likely to generate radical innovation are marked in bold italics: combinations of electrical & electronic engineering and applied physics; combinations of applied physics and condensed matter physics; and combinations of optics and applied physics. These combinations are more likely to generate radical innovation when combined with nanoelectronics. Consistently, among repeated keyword topics most likely to generate radical innovation, Topic 22 primarily addresses nanomaterials and related fabrication technologies, demonstrating multidisciplinary crossover between chemistry and materials science, corresponding to the fourth repeated subject classification combination. Topic 20 primarily involves nanotube fabrication materials and circuits, related to materials science, electrical & electronic engineering, applied physics, and condensed matter physics—mentioned across almost all repeated subject classification combinations—validating that repeated subject classification combinations can identify technical fields where radical innovation occurs.

Representative repeated subject classification combinations and their mutation degree

---

4. Conclusion and Outlook

Identifying technological mutations through topic mutation of scientific knowledge cited in patents represents an important direction for radical innovation identification. The topics and subject classifications of cited scientific knowledge can represent the technical themes and fields where radical innovation occurs. Therefore, this study uses keywords and subject classifications from patent scientific references to represent cited scientific knowledge, proposes computational methods for topic mutation degree in keyword co-occurrence networks and subject classification combinations, identifies mutated topics within cited scientific knowledge, and uses them to represent technical themes and fields most likely to generate radical innovation. Experiments in nanoelectronics revealed that cited scientific knowledge topics with high mutation degree occurred in technical themes closely related to nano-circuit materials and fabrication, including nanowires, carbon nanotubes, and computable circuits. Subject classification combination mutations primarily occurred in combinations of "chemistry, energy & fuels, crystallography, characterization & testing materials science, coatings & films materials science, optics, applied physics" and "applied physics, condensed matter physics, optics, electrical & electronic engineering." These combinations correspond to the identified technical themes. These results demonstrate that identifying radical innovation through topic mutation of cited scientific knowledge is feasible and valid, and the proposed method can be extended and applied to other technology fields closely related to basic research.

While topic mutation of cited scientific knowledge provides new insights for radical innovation identification, several aspects require further research. First, the accuracy and recall of cited scientific knowledge extraction, preprocessing, and matching need improvement, such as exhaustively enumerating all extraction rules for non-patent reference titles and using fuzzy retrieval in specific databases to return more accurate metadata for patent scientific references. Additionally, patent scientific reference keywords contain numerous broad terms (e.g., growth, development). Since this study uses author keywords supplemented by machine-generated keywords when author keywords are unavailable, these data were not processed; future work should enhance keyword preprocessing. Second, when interpreting the research content of a topic through its keywords, the relationships among keywords are not clearly displayed, increasing interpretation difficulty. Therefore, multiple semantic relationships among keywords need to be extracted to uncover what technical problems these topics solve, supplemented by visualization techniques and expert consultation. Third, while the proposed method can effectively identify radical innovation and provide early warning, further exploration is needed to discover the formation mechanisms of radical innovation for predictive purposes. Finally, validation was conducted only in nanoelectronics; the method' s effectiveness needs validation in other domains or improvement based on specific data characteristics.

References

[1] Christensen C M, Raynor M E. *The Innovator' s Solution: Creating and Sustaining Successful Growth*[M]. Boston: Harvard Business Press, 2003: 2-10.

[2] Zhang Xiaolin. The Trends that Will Disrupt Digital Libraries[J]. *Journal of Library Science in China*, 2011, 37(5): 4-12.

[3] Chen Ao, Liu Xielin. How are Radical Technologies Developed? A Literature Survey[J]. *Studies in Science of Science*, 2011, 29(9): 1281-1290.

[4] Dahlin K B, Behrens D M. When is an Invention Really Radical? Defining and Measuring Technological Radicalness[J]. *Research Policy*, 2005, 34(5): 717-737.

[5] Verhoeven D, Bakker J, Veugelers R. Identifying Ex Ante Characteristics of Radical Inventions through Patent-Based Indicators[R]. FEB Research Report-MSI, 2013.

[6] Zhao Zhiyun, Lei Xiaoping. Analysis of Scientific Linkage Between China' s Technology Innovation and Basic Research in Biotechnology Industry Based on Patent Citation[J]. *Journal of the China Society for Scientific and Technical Information*, 2012, 31(12): 1283-1289.

[7] Arts S, Veugelers R. Technology Familiarity, Recombinant Novelty, and Breakthrough Invention[J]. *Industrial and Corporate Change*, 2015, 24(6): 1215-1246.

[8] Sung H Y, Wang C C, Huang M H, et al. Measuring Science-Based Science Linkage and Non-Science-Based Linkage of Patents through Non-Patent References[J]. *Journal of Informetrics*, 2015, 9(3): 488-498.

[9] Huang M H, Yang H W, Chen D Z. Increasing Science and Technology Linkage in Fuel Cells: A Cross Citation Analysis of Papers and Patents[J]. *Journal of Informetrics*, 2015, 9(2): 244-258.

[10] Zhang Jinzhu, Zhang Xiaolin. Radical Innovation Identification Based on Cited Scientific Knowledge Abruption[J]. *Journal of the China Society for Scientific and Technical Information*, 2014, 33(3): 259-266.

[11] Kelley D J, Ali A, Zahra S A. Where do Breakthroughs Come From? Characteristics of High-Potential Inventions[J]. *Journal of Product Innovation Management*, 2013, 30(6): 1212-1226.

[12] Fleming L. Recombinant Uncertainty in Technological Search[J]. *Management Science*, 2001, 47(1): 117-132.

[13] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast Unfolding of Communities in Large Networks[J]. *Journal of Statistical Mechanics: Theory and Experiment*, 2008. DOI: 10.1088/1742-5468/2008/10/P10008.

[14] Service R F. Breakthrough of the Year: Molecules Get Wired[J]. *Science*, 2001, 294(5551): 2442-2443.

[15] Meyer M, Debackere K, Glänzel W. Can Applied Science be 'Good Science' ? Exploring the Relationship Between Patent Citations and Citation Impact in Nanoscience[J]. *Scientometrics*, 2010, 85(2): 527-539.

[16] Palmberg C, Dernis H, Miguet C. Nanotechnology: An Overview Based on Indicators and Statistics[R]. OECD, 2009.

[17] Shirabe M. Identifying SCI Covered Publications Within Non-Patent References in US Utility Patents[J]. *Scientometrics*, 2014, 101(2): 999-1014.

Author Contributions

Zhang Xiaolin and Zhang Jinzhu: Conceived the research idea, designed the study, and finalized the manuscript.
Zhang Jinzhu: Conducted experiments, collected, cleaned, and analyzed data, and drafted the manuscript.

Supporting Data

The supporting data is self-archived by the authors, E-mail: zhangjinzhu@njust.edu.cn.
[1] Zhang Jinzhu. ti_{origin}. Patent raw data.
[2] Zhang Jinzhu. snpr_{{{title}}*{{patent}}}. Metadata of patent scientific references.*
[3] Zhang Jinzhu. skcp{patent}. Cited scientific knowledge data.

Conflict of Interest Statement

All authors declare no conflict of interest.

Manuscript Received: 2016-05-04

Revised Manuscript Received: 2016-05-17

Note: Figure translations are in progress. See original paper for figures.