Evolution Analysis of Technical Topics Based on Non-negative Matrix Factorization (Postprint)

Authors: Wang Yuanyuan, Zhao Yajuan

Date: 2023-08-26T00:00:00+00:00

Abstract

[Purpose/Significance] Analyzing the evolution process of technical topics can help trace the development trajectory of technology, which holds significant importance for fostering innovation and predicting technology development trends. However, limited research exists on analyzing the evolution trajectory of technical topics from a semantic perspective. Therefore, this study investigates the evolution process of technical topics from a semantic perspective.

[Method/Process] This paper proposes an improved dynamic non-negative matrix factorization model based on non-negative matrix factorization for dynamic topic modeling of patent texts, and utilizes the TextRank algorithm to extract noun phrases for annotation, thereby enhancing the interpretability of the extracted technical topics. On this basis, the technical evolution trajectory is calculated using word vectors and visualized.

[Result/Conclusion] An empirical analysis was conducted on five-party patents from 2002, 2005, 2008, 2011, and 2014, identifying 65 technical topics and their evolution trajectories, which demonstrates the feasibility of the proposed method.

Full Text

Preamble

Volume 62, Issue 10, May 2018

Evolutionary Analysis of Technological Topics Based on Non-negative Matrix Factorization

Wang Yuanyuan, Zhao Yajuan

National Science Library, Chinese Academy of Sciences, Beijing 100190

School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190

Abstract

[**Purpose/Significance**] Analyzing the evolution of technological topics enables us to trace the development trajectory of technologies, which is essential for fostering innovation and forecasting technological trends. However, existing research rarely examines technological topic evolution from a semantic perspective. This study therefore analyzes the evolution of technological topics from a semantic viewpoint. [**Method/Process**] We propose an improved dynamic non-negative matrix factorization model for dynamic topic modeling of patent texts, and employ the TextRank algorithm to extract noun phrases for labeling, thereby enhancing the interpretability of extracted technological topics. Building upon this, we calculate technological evolution trajectories using word vectors and visualize the results. [**Result/Conclusion**] An empirical analysis of five-country patents from 2002, 2005, 2008, 2011, and 2014 identified 65 technological topics and their evolutionary trajectories, demonstrating the feasibility of the proposed method.

Keywords: technological topic evolution; non-negative matrix factorization; topic model; dynamic topic analysis
Classification Number: G254.11
DOI: 10.13266/j.issn.0252-3116.2018.10.013

Modern technological development is advancing at an unprecedented pace, with increasing technology flows, collaborations, and integration across industries. Technological interconnections are becoming more intricate, as technological progress in one industry is closely related to changes in others. As primary drivers of social innovation, enterprises face the constant challenge of developing new products through continuous innovation. Consequently, technological complexity and diversity are growing daily, the pace of technological innovation is accelerating, its intensity is increasing, and uncertainty in technological development continues to rise. Technological topic analysis constitutes a crucial component of patent intelligence analysis, encompassing two main aspects: technological topic distribution and technological topic evolution analysis. While technological topic distribution focuses on static characteristics, technological topic evolution analysis is more comprehensive, including examination of evolutionary processes, prediction of development trends, and identification of emerging technological topics. Understanding the mechanisms of technological topic evolution is vital for innovation development.

Patents serve as carriers of technical, legal, and commercial information, representing important data resources for technological topic evolution research. The growing volume of patents and increasing technological complexity pose significant challenges for analysis. With advancements in text mining and semantic analysis, topic models (such as LDA and NMF) have been widely and successfully applied across various domains, including social media and scientific literature, greatly improving the efficiency of mining and understanding semantic information in unstructured text data. This provides a valuable approach for technological topic evolution analysis—examining the dynamic evolution of technological topics through semantic analysis of patent text content. This study proposes an improved dynamic non-negative matrix factorization approach from a semantic perspective, dividing patent texts into different time windows to extract window topics, then deriving dynamic topics based on these window topics to explore the dynamic evolution of technological topics.

1 Literature Review

Patents are the primary data source for technological topic evolution research. Based on how existing studies utilize patent information, we categorize technological topic evolution research into three types: classification-based approaches, citation-based approaches, and text content-based approaches.

Patent classification provides a simple and universal technical classification system based on the technical content disclosed in patents. Classification-based technological topic analysis primarily involves statistical analysis and co-classification analysis. Patent co-classification refers to the co-occurrence of different patent classification codes (such as IPC) in a single patent, indicating connections between different technical directions that can be used to analyze technological topics. K. Suzuki et al. employed IPC co-occurrence to study technology fusion in development. S. Jeong et al. used the Jaccard coefficient to examine the strength of IPC co-occurrence relationships and analyzed network density characteristics to understand temporal changes in different technological topics and major types of technology fusion. W. S. Lee et al. conducted link prediction analysis on IPC co-occurrence networks to forecast potential emerging technological topics and used topic analysis to extract keywords for identifying future emerging fields. B. Huang et al. applied association rule analysis to examine IPC co-occurrence in information technology and biotechnology, analyzing technology topic characteristics from support, confidence, and lift perspectives.

Citation-based approaches analyze citation relationships between patent documents and between patents and scientific literature. Citation relationships reflect technology flows, and constructing citation networks enables analysis of technological topic evolution trajectories. P. L. Chang et al. combined hierarchical and non-hierarchical clustering based on patent citation relationships to cluster target domain patents into three clusters, generating technology topics for each cluster and constructing network diagrams of relationships within each cluster. C. Choi et al. proposed a patent citation network-based method for identifying technological topic evolution paths. Y. Geum et al. and Zhai Dongsheng et al. analyzed technology fusion based on knowledge flows in citation networks between technology categories. Beyond direct citation relationships, networks based on co-citation and bibliographic coupling relationships have also been used to construct patent networks representing technological topic similarity. Overall, current citation-based technological topic analysis methods fall into three categories: clustering based on citation relationships followed by evolution analysis; identifying knowledge flow main paths in citation networks to map evolution trajectories; and employing social network analysis to evaluate evolution stages.

While classification- and citation-based approaches can identify macro-level development trends, they fail to reveal specific evolutionary details. Text content-based analysis addresses this limitation, making mining of hidden information in patent texts a primary means of technological topic evolution analysis. Current text-based approaches mainly include term frequency analysis and keyword co-occurrence analysis. Luan Chunjuan mapped the evolution of co-occurrence networks in solar technology using keyword co-occurrence and social network analysis. Han Hongqi et al. proposed a strategic diagram analysis method based on co-occurrence of patent technology feature terms to study topic evolution. S. H. Chen et al. clustered patent texts from different time windows and combined citation network analysis to examine evolution processes.

This study proposes an improved NMF-based model for technological topic evolution analysis from a semantic perspective. NMF is an unsupervised method that decomposes non-negative matrices into lower-dimensional non-negative factors, widely applied in image processing and latent topic extraction from text corpora. Direct application of NMF for topic modeling is static and cannot reflect temporal evolution. This paper proposes an improved dynamic non-negative matrix factorization (Dynamic NMF) approach that first divides patent texts into different time windows to extract window topics, then derives dynamic topics from these window topics to explore dynamic evolution from a semantic perspective.

2 Research Method

The overall research framework is shown in Figure 1 [Figure 1: see original paper], comprising four main steps: (1) training word vectors on extracted patent text data for subsequent topic coherence evaluation and similarity calculation; (2) applying dynamic NMF to patent texts for dynamic topic modeling to obtain dynamic and window topics, where the number of topics is determined using word vector-based topic coherence evaluation metrics; (3) labeling extracted topics with noun phrases extracted via TextRank to enhance interpretability; and (4) calculating and visualizing technological evolution trajectories. Key steps are detailed below.

2.1 Non-negative Matrix Factorization (NMF)

Given a corpus containing n documents, we first construct a document-term matrix $A \in \mathbb{R}^{n\times m}$, where m represents the vocabulary size. NMF of matrix A produces a rank-k approximation in the form of a product of two non-negative factors, i.e., $A \approx WH$. NMF aims to minimize the reconstruction error between

A and WH. The rows of factor $H \in \mathbb{R}^{k\times m}$ can be interpreted as k topics, with each topic defined as non-negative weights for the m terms in the vocabulary. Sorting each row by term weights yields the top-n terms representing each topic. The columns of matrix $W \in \mathbb{R}^{n\times k}$ represent the weights of n documents for each topic, linking documents to corresponding topics. NMF algorithms typically initialize with random factors, causing convergence to different local optima and resulting instability. This study uses Non-negative Double Singular Value Decomposition (NNDSVD) [28] to generate initialization factors, improving topic quality.

A critical parameter in topic models is the number of topics k, which directly determines extraction results. Too small a k yields overly broad topics, while too large a k produces excessive, highly similar topics. J. Chang et al. [29] used topic coherence for comparing topics across different k values, assuming terms within the same topic are related and using this relatedness to evaluate topic quality. D. O' Callaghan et al. [30] proposed a Word2Vec-based topic coherence evaluation method (Topic Coherence via Word2Vec, TC-W2V) that assesses topic quality by evaluating correlations among top-n terms. Generally, higher similarity among terms indicates greater semantic coherence. This study uses TC-W2V to determine k, as shown in equations (1) and (2). Each topic is represented by the top t terms, and for a single topic t_h, coherence is the average cosine similarity between all pairs of top t terms, where term vectors are computed by Word2Vec [31]:

$$coh(t_h) = \frac{2}{t(t-1)} \sum_{i=1}^{t-1} \sum_{j=i+1}^{t} cos(wv_i, wv_j)$$

For a topic model T consisting of k topics, the overall coherence score is the average of individual topic coherences:

$$coh(T) = \frac{1}{k} \sum_{h=1}^{k} coh(t_h)$$

Given a range $[k_{min}, k_{max}]$, the optimal k can be determined by the maximum TC-W2V value.

2.2 Dynamic Non-negative Matrix Factorization

Dynamic NMF is a two-layer NMF process. For temporally sequenced patent texts (e.g., by year), we first apply NMF to topic model patents in each fixed time window, then treat the extracted topics as documents and apply NMF again to model these window-level outputs, extracting dynamic technological topics across all time windows. The dynamic NMF process is illustrated in Figure 2 [Figure 2: see original paper].

First Layer: For temporal data, we first partition data into fixed-length time windows. Regarding window division, there are two main approaches: overlapping [32] and non-overlapping [33]. Overlapping windows may overlook short-cycle topics and ignore topic states at each time point. This study adopts non-overlapping partitioning into τ time windows $\{T_1, \dots, T_\tau\}$. For each time window T_i, we apply NMF to generate a window topic model M_i containing k_i window topics, where parameter k_i is determined by equation (2). This first layer produces consecutive window topic models $\{M_1, \dots, M_\tau\}$.

Second Layer: For each window topic model' s factor H_i, we treat the rows of H_i (the k_i window topics) as topic-documents (topics represented by terms, thus viewable as documents) to construct a compressed representation of the original corpus. The topic-term matrix B is constructed as follows: (1) Build an empty matrix B; (2) For each window topic model M_i: for each window topic in M_i, select the top t terms from the corresponding NMF factor H_i' s row vector, set other term weights to 0, and add this vector as a new row to B; (3) After adding all topics from all window models, delete columns in B that are all zeros (terms that never appear in any window topic' s top t terms).

Matrix B has dimensions $n' \times m'$, where $n' = \sum_{i=1}^{\tau} k_i$ is the number of topic-documents and $m' \ll m$ is the subset of remaining terms. Retaining only top t terms leverages representative terms from each time window while excluding low-significance terms, reducing computational cost for the second factorization. The second-layer NMF on matrix B extracts k' dynamic topics. The TW-W2V coherence measure determines parameter k'. The decomposition result $B \approx UV$ is interpreted as: the top-n terms in each row of factor V represent dynamic topics; column values in factor U indicate the relevance of each window topic to each dynamic topic.

Linking dynamic topics to window topics enables tracking topic evolution over time. First, based on values in each row of factor U, each window topic is associated with the dynamic topic for which it has the maximum weight. Similarly, patent documents can be linked to window topics, thereby linking dynamic topics to patent documents. The two-layer NMF topic modeling process outputs: (1) τ window topic models, each containing k_i window topics, with each window topic having associated documents and represented by top t terms; (2) k' dynamic topics, each associated with a set of window topics and linked documents.

4 Results and Analysis

4.1 Determining the Number of Topics k

As described above, this study uses the TC-W2V topic coherence metric to automatically determine k. We first train Word2Vec on all patent text data, using the top 20 terms per topic to calculate TC-W2V values. In each time window, considering computational efficiency, we set the topic number range to $k \in [80, 180]$ with step size 5 to generate window topic models with different topic counts and identify the optimal k value (the k with highest TC-W2V).

Figure 4 [Figure 4: see original paper] shows TC-W2V scores for different k values in the 2014 window topic model, with the maximum occurring at $k = 110$. Similarly, the optimal k values for 2002, 2005, 2008, and 2011 are 105, 120, 100, and 110, respectively. The dynamic topic model' s topic count range is set to $k \in [50, 150]$ with step size 5, yielding an optimal k of 65. This study extracted 65 dynamic topics, denoted as D01, D02, ···, D65, and 545 window topics, denoted as "year + topic number" (e.g., 2002_{01}, 2002_{02}, ···, 2014_{110}).

4.2 Technological Topic Evolution Analysis

This study extracted 65 dynamic topics using the proposed dynamic NMF model. To further analyze their evolution, we employed two screening criteria: topic intensity and technology fusion degree. Topic intensity refers to the number of patents associated with a topic. Technology fusion degree can be measured in various ways, each revealing different characteristics of technology fusion [36]. Entropy measures technology fusion by assessing the distribution of a technological direction across different technology categories. Following E. J. Han and S. Y. Sohn [37] and Y. Cho and M. Kim [3], we use entropy of patent classification information (IPC) to measure technology fusion degree for each topic.

Specifically, we link identified technological topics to patents (associating each patent with its highest-weight topic), then link these topics to patent classification information (IPC), obtaining frequency distribution information of IPCs associated with each topic. Based on this IPC distribution, we compute entropy. This study measures technology fusion degree using 4-digit IPC codes, as shown in equation (3), where $P(x_i)$ represents the frequency of a technology category (4-digit IPC):

$$H(X) = -\sum_{i} P(x_i) \log(P(x_i))$$

Table 1 shows the topic intensity and technology fusion degree for all 65 dynamic topics. To analyze the evolution of identified topics, we examine dynamic topics D64 and D59, which exhibit both high topic intensity and technology fusion degree.

Topic models typically represent topics using top-n terms [27, 38-39]. However, individual terms are often too broad; phrases provide more complete and precise semantic expression, especially for technical terminology in patents. Topic labeling helps humans understand topic meanings [40-43]. This study uses the TextRank algorithm [44] to label NMF-generated topics. TextRank calculates term importance, and adjacent important terms form phrases. After part-of-speech tagging and syntactic filtering, we obtain noun phrases.

Dynamic topics D59 and D64 are shown in Table 2 , where NMF representations use top 20 terms, and TextRank representations use extracted noun phrases. Comparison shows that while NMF-generated terms can express topic content, word-based representations remain overly broad, making semantic meanings incomplete. Noun phrases provide clearer, more accurate, and complete semantic expression. To further explore topic content, we extracted the most relevant patents to aid understanding. Dynamic topic D64 represents automotive production and manufacturing technologies, while D59 represents electrical equipment technologies.

4.2.1 Quantitative Analysis of Technological Topic Evolution

Figure 5 [Figure 5: see original paper] shows the evolution of patent counts over time for dynamic topics D59 (“Electrical Equipment Technology”) and D64 (“Automotive Production and Manufacturing Technology”). Higher patent counts indicate hotter, more important technologies. Figure 5 reveals that “Automotive Production and Manufacturing Technology” patents increased continuously from 2002 to 2011, indicating rapid development, but declined sharply in 2014, suggesting transition from rapid to stable development. According to Figure 3’ s statistics on all five-country patents, the overall number also declined in 2014, so the decline in automotive technology patents may also reflect this overall trend. “Electrical Equipment Technology” patents grew rapidly between 2002 and 2005, then stabilized, indicating slowed development and entry into a stable phase.

4.2.2 Content Analysis of Technological Topic Evolution

Dynamic topic analysis based on NMF can reveal both document set content and dynamic topic evolution across time windows. We calculate similarity between topics in adjacent time windows using Word2Vec and visualize results with Graphviz. For any window topic th, its vector representation is $\sum_{i\in t} wv_i$, where t represents the topic’ s top t terms and wv_i is the i-th term’ s vector. As shown in equation (4), similarity between two adjacent window topics th and th' is computed using cosine similarity. In visualization, if topics in adjacent windows meet a certain similarity threshold, they are considered to have an evolutionary relationship; higher similarity indicates stronger relationships, represented by thicker connecting lines.

$$sim(th, th') = cos(th, th')$$

To clearly show evolution paths, we tested similarity thresholds of 0.3, 0.5, and 0.7. As shown in Figure 6 [Figure 6: see original paper], overly high thresholds obscure relationships between topics in adjacent windows, making evolution trajectories unclear; overly low thresholds introduce unnecessary connections, complicating evolution patterns. Therefore, we set the similarity threshold to 0.5: if similarity between two topics in adjacent windows exceeds 0.5, they are considered related, with higher similarity indicating closer connections (thicker lines).

Figure 7 [Figure 7: see original paper] shows evolution trajectories for "Electrical Equipment Technology" and "Automotive Production and Manufacturing Technology" across time windows. Each window topic is represented by a circle; columns represent time windows (2002, 2005, 2008, 2011, 2014 from left to right). Circle size is proportional to the number of associated documents, with larger circles indicating hotter, more important topics. "Automotive Production and Manufacturing Technology" maps to 21 window topics (5, 5, 4, 3, and 4 topics across the five windows), while "Electrical Equipment Technology" maps to 30 window topics (5, 6, 3, 7, and 9 topics).

We identify five evolution patterns for window topics: emergence, expansion, fusion, continuation, and decline [31]. **Emergence** indicates no evolutionary relationship exists with previous window topics, such as "Semiconductor/Resistor Devices" and "Digital Signal Processing" in the electrical equipment trajectory, and "Fabric/Fiber/Coating Materials" in the automotive trajectory. **Expansion** indicates a topic evolves into multiple topics in the next window, such as "Engine/Power Transmission Technology" in 2011 expanding into "Electric Vehicle Power and Control Technology" and "Power Generation/Turbine Technology" in 2014. **Fusion** indicates a topic evolves from multiple topics in the previous window, such as "Electric Vehicle Power and Control Technology" in 2014 fusing from "Engine/Power Drive Technology" and "Power Control/Automotive Assist Technology" in 2011, and "Solar Cells" in 2011 fusing from "Electrical Components" and "Semiconductor Devices" in 2008. **Continuation** indicates exactly one topic in the next window has an evolutionary relationship with the current topic, such as "Optical Storage/Control Device Technology" (2002) to "Optical Drive Technology" (2005), and "Solar Cell Technology" (2011) to "Solar Cell Technology" (2014). **Decline** indicates no evolutionary relationship with any topic in the next window, such as "Optical Drive Technology," "Fabric/Fiber/Coating Materials," and "Rubber/Tire Technology."

From Figure 7, we extract core evolution paths for both technological fields. In "Electrical Equipment Technology," core paths include "Semiconductor Devices," "Solar Cells," and "Medical Assist Devices." In "Automotive Production and Manufacturing Technology," core paths include "Electric Vehicle Power and Control Technology," "Magnetic Materials and Devices," and "Power Generation/Turbine Technology." Taking three core evolution paths in automotive technology as examples:

1. **Electric Vehicle Power and Control Technology** (Figure 8 Figure 8: see original paper): "Electromagnetic Technology" and "Motor Technology" fuse into "Electromagnetic Actuator Technology," which then continues to "Magnetic Material/Power Technology." Meanwhile, "Motor Technology" and "Drive/Power Transmission Technology" fuse into "Engine Drive and Power Transmission/Control Technology," which expands into "Power Drive/Transmission Technology" and "Power Device/Automotive Control

Technology." These then fuse into "Electric Vehicle Power and Control Technology."

2. **Magnetic Materials and Devices** (Figure 8 Figure 8: see original paper): "Electromagnetic Technology" and "Motor Technology" fuse into "Electromagnetic Actuator Technology," then continue evolving into "Magnetic Materials and Devices Technology."
3. **Power Generation/Turbine Technology** (Figure 8 Figure 8: see original paper): "Electromagnetic Technology" and "Motor Technology" fuse into "Electromagnetic Actuator Technology," which continues to "Magnetic Material/Power Technology." "Drive/Power Transmission Technology" continues to "Power Drive/Transmission Technology," then fuses with "Magnetic Material/Power Technology" into "Engine/Power Drive Technology," finally continuing to "Power Generation/Turbine Technology."

Although analyzed separately, these three core evolution paths show many intersections, indicating close interconnections. Other technologies appear more isolated, such as rubber/tire materials (2014_{107}, 2008_{99}) and functional material technologies like fabrics and coatings (2005_{112}), which occupy noncore positions and rarely connect with other technologies.

5 Conclusion and Outlook

This study proposes an improved dynamic NMF model from a semantic perspective for dynamic topic modeling of patent texts to analyze technological topic evolution. The approach comprises five steps: (1) training word vectors via Word2Vec to obtain distributed representations for determining topic number k and calculating topic similarity; (2) dynamic topic modeling using improved dynamic NMF to extract dynamic and window topics; (3) labeling topics with TextRank-extracted noun phrases to enhance interpretability; (4) calculating evolution trajectories using word vectors and visualizing with Graphviz; (5) empirical analysis using five-country patent data from 2002, 2005, 2008, 2011, and 2014.

The method effectively utilizes patent text content to automatically identify technological topics and their evolution paths. Several limitations require further research: (1) The two-step process of first extracting topics then labeling with phrases, while effective, is cumbersome; future work should focus on directly generating topic phrases. (2) This study uses only text information, ignoring valuable information like patent classifications; incorporating such information to improve model precision and effectiveness needs further investigation.

References

[1] ROSENBERG N. Technological change in the machine tool industry, 1840-1910[J]. The journal of economic history, 1963, 23(4): 414-443.

[2] ROSENBERG N. Exploring the black box: technology, economics, and history[M]. Cambridge: Cambridge University Press, 1994.

[3] CHO Y, KIM M. Entropy and gravity concepts as new methodological indexes to investigate technological convergence: patent network-based approach[J/OL]. PLOS ONE, 2014, 9(6): e98009 [2016-10-26]. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0098009.

[4] HU Apei, ZHANG Jing, LEI Xiaoping, et al. A Review of Patent Technology Theme Analysis Based on Text Mining[J]. Journal of Intelligence, 2013(12): 88-92.

[5] XIAO Huwei. Patent Map Methods and Applications[M]. Shanghai: Shanghai Jiao Tong University Press, 2011.

[6] SUZUKI K, SAKATA J, HOSOYA J. An empirical analysis on progress of technology fusion[C]//IEEE. Third International Conference on Digital Information Management (ICDIM). Piscataway: IEEE Xplore, 2009: 2433-2440.

[7] JEONG S, KIM JC, CHOI JY. Technology convergence: what developmental stage are we in?[J]. Scientometrics, 2015, 104(3): 841-871.

[8] LEE WS, HAN EJ, SOHN SY. Predicting the pattern of technology convergence: patent-based analysis[J]. Technological forecasting and social change, 2015, 100: 317-329.

[9] HUANG B, HUANG L. Patent-based association analysis for technological convergence of IT and BT[C]//IEEE. 2015 International Conference on Logistics, Informatics and Service Sciences (LISS). Piscataway: IEEE Xplore, 2015: 1-6.

[10] NARIN F. Patent bibliometrics[J]. Scientometrics, 1994, 30(1): 147-155.

[11] CHANG PL, WU CC, LEU HJ. Using patent analyses to monitor technological trends in an emerging field of technology: a case of carbon nanotube field emission display[J]. Scientometrics, 2010, 82(1): 5-19.

[12] CHOI C, PARK Y. Monitoring the organic structure of technology based on the patent development paths[J]. Technological forecasting and social change, 2009, 76(6): 754-768.

[13] GEUM Y, KIM C, LEE S, et al. Technological convergence of IT and BT: evidence from patent analysis[J]. ETRI journal, 2012, 34(3): 439-449.

[14] ZHAI Dongsheng, CAI Liwei, ZHANG Jie, et al. Research on Identification Model of Technology Convergence Innovation Orbit Based on Patents: A Case Study of Cloud Computing Technology[J]. Journal of the China Society for Scientific and Technical Information, 2015, 34(4): 352-360.

[15] MOGEE ME, KOLAR RG. Patent co-citation analysis of Eli Lilly & Co. patents[J]. Expert opinion on therapeutic patents, 1999, 9(3): 291-305.

[16] HUANG MH, CHIANG LY, CHEN DZ. Constructing a patent citation map using bibliographic coupling: a study of Taiwan' s high-tech companies[J]. Scientometrics, 2003, 58(3): 489-509.

[17] 陈亮, 张志强. 技术演化研究方法进展分析 [J]. 图书情报工作, 2012, 56(17): 59-66.

[18] MARTINELLI A. An emerging paradigm or just another trajectory? Understanding the nature of technological changes using engineering heuristics in the communications switching industry[J]. Research policy, 2012, 41(2): 414-429.

[19] MINA A, RAMLOGAN R, TAMPUBOLON G, et al. Mapping evolutionary trajectories: applications to the growth and transformation of medical knowledge[J]. Research policy, 2007, 36(5): 789-806.

[20] SUH N, LEE PC. Dynamic and quantitative exploration on technology evolution mechanism: the case of electrical conducting polymer nanocomposite[C]//IEEE. 2009 Portland International Conference on Management of Engineering & Technology (PICMET). Piscataway: IEEE Xplore, 2009: 937-939.

[21] BATAGELJ V. Analyzing the structure of US patents network[M]//Data science and classification. Springer, Berlin, Heidelberg, 2006: 141-148.

[22] ABBAS A, ZHANG L, KHAN SU. A literature review on the state-of-the-art in patent analysis[J]. World patent information, 2014, 37(4): 3-13.

[23] LUAN Chunjuan. Analysis of Global Solar Technology Network and Key Technology Evolution Based on Patent Co-occurrence[J]. Journal of the China Society for Scientific and Technical Information, 2013, 32(1): 68-79.

[24] HAN Hongqi, AN Xiaomi, ZHU Donghua, et al. Strategic Diagram Analysis Method for Co-occurrence of Patent Technology Terms[J]. Application Research of Computers, 2011, 28(2): 576-579.

[25] CHEN SH, HUANG MH, CHEN DZ. Identifying and visualizing technology evolution: a case study of smart grid technology[J]. Technological forecasting & social change, 2012, 79(6): 1099-1110.

[26] LEE DD, SEUNG HS. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401: 788-791.

[27] WANG Q, CAO Z, XU J, et al. Group matrix factorization for scalable topic modeling[C]//Proceedings of the 35th international ACM SIGIR conference on research and development in information retrieval. New York: ACM, 2012: 375-384.

[28] BOUTSIDIS C, GALLOPOULOS S. SVD based initialization: a head start for non-negative matrix factorization[J]. Pattern recognition, 2008, 41(4): 1350-1362.

[29] CHANG J, GERRISH S, WANG C, et al. Reading tea leaves: how humans interpret topic models[C]//Proceedings of the 22nd international conference on

neural information processing systems. USA: Curran Associates Inc., 2009: 288-296.

[30] O’ CALLAGHAN D, GREENE D, CARTHY J, et al. An analysis of the coherence of descriptors in topic modeling[J]. Expert systems with applications, 2015, 42(13): 5645-5657.

[31] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[C]//Proceedings of the 26th international conference on neural information processing systems. USA: Curran Associates Inc., 2013: 3111-3119.

[32] PAN Jiaofeng, ZHANG Xiaolin, WANG Xiaomei, et al. Scientific Structure Map 2012[M]. Beijing: Science Press, 2013.

[33] SULO R, BERGER-WOLF T, GROSSMAN R. Meaningful selection of temporal resolution for dynamic networks[C]//Proceedings of the Eighth workshop on mining and learning with graphs. New York: ACM, 2010: 127-136.

[34] LEE HJ, LEE S, YOON B. Technology clustering based on evolutionary patterns: the case of information and communication technologies[J]. Technological forecasting and social change, 2011, 78(6): 953-967.

[35] KIM E, CHO Y, KIM W. Dynamic patterns of technological convergence in printed electronics technologies: patent citation network analysis[J]. Scientometrics, 2014, 98(2): 975-998.

[36] LI Shuying, FANG Shu. Research Progress on Data Analysis Methods for Measuring Technology Convergence and Trends[J]. Data Analysis and Knowledge Discovery, 2017, 1(7): 2-12.

[37] HAN EJ, SOHN SY. Technological convergence in standards for information and communication technologies[J]. Technological forecasting and social change, 2016, 106: 1-10.

[38] BLEI DM, NG AY, JORDAN MI. Latent Dirichlet Allocation[J]. Journal of machine learning research, 2003, 3: 993-1022.

[39] HOFMANN T. Unsupervised learning by probabilistic latent semantic analysis[J]. Machine learning, 2001, 42(1-2): 177-196.

[40] MEI Q, SHEN X, ZHAI C. Automatic labeling of multinomial topic models[C]//Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining. New York: ACM, 2007: 490-499.

[41] EL-KISHKY A, SONG Y, WANG C, et al. Scalable topical phrase mining from text corpora[J]. Proceedings of the VLDB endowment, 2014, 8(3): 305-316.

[42] WANG X, MCCALLUM A, WEI X. Topical n-grams: phrase and topic discovery, with an application to information retrieval[C]//Proceedings of the 2007 Seventh IEEE international conference on data mining. Washington, DC: IEEE Computer Society, 2007: 697-702.

[43] BLEI DM, LAFFERTY JD. Visualizing topics with multi-word expressions[EB/OL]. [2016-10-28] https://arxiv.org/abs/0907.1013.

[44] MIHALCEA R, TARAU P. TextRank: bringing order into texts[C]//Proceedings of the ACL 2004 on interactive poster and demonstration sessions. Stroudsburg: Association for Computational Linguistics, 2004: 20.

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

Source: ChinaXiv —Machine translation. Verify with original.