

# A Survey of Knowledge Tracking Based on Deep Learning

Juwei Dao\*

College of Information, Yunnan Normal University, Kunming, 650000, China.

\* Corresponding author: Email: 804079054@qq.com.

**Abstract:** In order to fully understand the research status in the field of knowledge tracking based on deep learning in China, this paper uses CiteSpace visualization analysis software to conduct visualization analysis and text analysis on 118 literatures with the theme of "knowledge tracking based on deep learning" included in CNKI from 2012 to 2022. The results show that "artificial intelligence", "intelligent education", "deep learning" and "big data" are the research hotspots of knowledge tracking based on deep learning in China. The correlation between its frontier research and the development of the times rose rapidly from 2017 to 2021. In the future research, researchers should focus on the optimization of the model, the application of basic research and the practical application of the research results.

**Keywords:** Knowledge tracking; Deep learning; CiteSpace.

## 1. Introduction

The popularization of education informatization has led to massive educational data generated by online education platforms, which makes it difficult for educators to track the knowledge status of each student and provide personalized learning guidance. Moreover, the types of knowledge in the online learning system are complex, and students need to search and retrieve from various redundant information by themselves, which leads to the diversification of learning resources and learning paths, but not necessarily effective. Therefore, it is necessary to use scientific methods to analyze and excavate pertinently (4).

Knowledge tracing (KT) is a powerful tool to realize online wisdom education, which can personalize the effective learning path for students, and track the current knowledge state of students according to their answers, so as to realize individualized teaching (6). In view of its great significance to education, knowledge tracking has been widely used in industry-related online education platforms.

Based on CiteSpace software, this study studies the field of knowledge tracking based on deep learning in the past two decades, and understands the development of knowledge tracking based on deep learning by analyzing keyword co-occurrence and research preface. The existing problems and the future development direction of the research are discussed and prospected.

## 2. Data Sources and Research Methods

### 2.1. Data sources

The data in this paper comes from China National Knowledge Infrastructure (CNKI), the search method is keyword search, the time range is from 2002 to 2022, the carrier form is "journal", the source categories are: SCI source journals, EI source journals, Peking University core, Chinese Social Sciences Citation Index (CSSCI). Firstly, the two keywords of "deep learning" and "knowledge tracking" were jointly searched, and 118 articles were found (the deadline of data retrieval was June 26, 2022). After removing the duplicate articles, 118 articles were obtained for research.

### 2.2. Research methods

This paper uses literature research method and CiteSpace (version 5.8.R3) software to analyze. CiteSpace is an information visualization software developed by Chen Chaomei [3] of Drexel University in the United States using Java language, which is mainly based on co-citation analysis theory and routing network algorithm, and can help researchers quickly mine key information from a large number of literatures. The software is suitable for multivariate, time-sharing and dynamic complex network analysis, and is widely used to analyze the co-citation relationship among academic papers, journals and authors, and to map scientific knowledge.

In this study, the publication annual trend chart in CNKI and CiteSpace software were used to conduct visual analysis and text analysis of the literature. Firstly, check the annual trend chart of the selected literature published in CNKI, and explain the literature published in the past 20 years based on deep learning knowledge tracking. Secondly, the research samples are exported in Refworks format, and the CiteSpace visualization software is used to draw and quantitatively analyze the visual atlas of the literature on knowledge tracking based on deep learning, and the keyword co-occurrence map, time zone map and emergent word map of knowledge tracking based on deep learning in the past 20 years are obtained. According to the results of this text analysis, the research hotspots and evolution trends of knowledge tracking based on deep learning in China are summarized.

## 3. Research results and analysis

### 3.1. Analysis of the time of issuing documents

Through the statistics of the number of relevant literatures published in each time period, it is helpful to understand the degree of concern and trend in the research field. The annual publication of knowledge tracking based on deep learning in China is as follows: Figure As shown.

From the annual change of the number of publications, we can see that although the references of knowledge tracking

based on in-depth learning can be traced back a long way, from the global research point of view, the focus of knowledge tracking in 2015 began to shift from the traditional probability model to in-depth learning. Piech (C. Piech) et al. [4] of Stanford University first introduced the Long short-term Memory (LSTM), a variant of recurrent neural network, into the knowledge tracking task, and proposed the first deep knowledge tracking model. In 2016, he published the first Chinese research results in related fields in China.



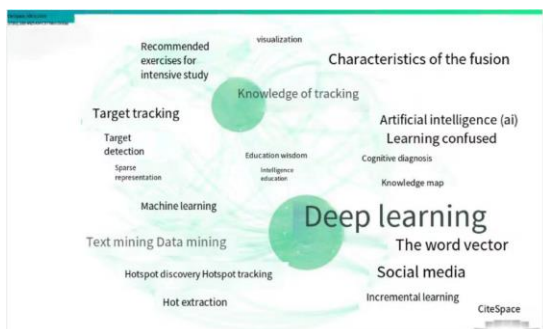
**Figure 1.** Annual changes in the number of papers on knowledge tracking based on deep learning from 2002 to 2022

From 2016 to 2018, the domestic research in this field is really concerned, the research heat is not high enough, and the number of articles is very small, with only two articles per year in 2016 and 2017, and four in 2018.

However, since 2018, the heat in this field has increased greatly, and the number of articles has begun to rise rapidly. Since 2020, affected by the COVID-19 epidemic, online education has reached an unprecedented scale. How to effectively provide personalized learning services while carrying out large-scale online education has become an urgent and important demand. Knowledge tracking has become an important example to meet this need [5], during which research results have increased almost exponentially, and still maintain a rapid growth trend today.

Neural networks based on deep learning have been proved to be able to achieve certain results in performance, and are far superior to traditional algorithms. The deep network randomly samples from the massive student data on the ITS platform as training data samples. With the increase of student practice records, the original samples lack real-time and cannot track the knowledge growth of students in time. Faced with the increasing number of students' learning records, how to extract representative practice records, how to use enhanced sampling to strengthen the training process of deep learning network, how to collect data sets, and how to enhance data samples are also the main problems faced by knowledge tracking algorithms based on deep learning.

### 3.2. Research hotspots of knowledge tracking based on deep learning



**Figure 2.** Keyword co-occurrence map of knowledge tracking research based on deep learning in China from 2012 to 2022

Keyword co-occurrence analysis is a high-level summary of common words or noun phrases appearing in the literature to determine the relationship between topics in the discipline represented by the research literature data. The co-occurrence analysis of keywords is helpful to understand the research hotspots in this field. The scientific knowledge map of keyword co-occurrence in knowledge tracking research based on deep learning in China is shown in Figure 2, with 148 nodes and 299 links.

From the keyword co-occurrence graph, we can see that the two largest keyword nodes are "deep learning" and "knowledge tracking", which are the keywords used in searching literature, and it is also very reasonable to be the largest node.

As a new field with rapid development in the past decade, deep learning has attracted more and more researchers' attention. It has obvious advantages over shallow models in feature extraction and modeling. Deep learning is good at mining more and more abstract feature representations from the original input data, and these representations have good generalization ability. It overcomes some problems that were considered difficult to solve in artificial intelligence in the past.

Knowledge tracking refers to the use of computer models to automatically measure students' knowledge level and judge students' mastery of knowledge according to the interaction records between students and exercises, which is characterized by automation and personalization. Many researchers pay attention to the powerful feature extraction ability of deep learning and apply it to the field of knowledge tracking, which is called deep learning-based knowledge tracing or deep learning-based knowledge tracing (DLKT). Compared with the traditional machine learning model, knowledge tracking based on deep learning does not require manually labeled knowledge component (KC) information, and can capture more complex knowledge representations of students, and can also discover and utilize the association information between KCs. At present [6], the research of knowledge tracking based on deep learning has become a hot topic in the field of knowledge tracking.

In addition, the larger and more relevant nodes are: "wisdom education", "artificial intelligence", "knowledge diagnosis" and so on.

As an important part of intelligent education, knowledge tracking has been widely studied and applied in the fields of artificial intelligence tutoring system, distance education and online teaching platform. It can personalize effective learning paths for students and track the current state of knowledge mastery of students according to their answers, so as to further realize individualized teaching and personalized teaching, which is of great significance to wisdom education.

According to the different modeling methods, the existing knowledge tracking models can be divided into two categories: traditional machine learning methods and deep learning methods. Bayesian knowledge pursuit [7] (Bayesian Knowledge Tracing, BKT) and additive factor model (AFM) [8] are two representative models in traditional machine learning methods. They have been widely used to assess the knowledge level of learners in intelligent tutoring systems. However, since 2015, knowledge tracking based on deep learning has become the "main force" of knowledge tracking research. The reason is that the rapid development of artificial intelligence technology and in-depth learning in recent years has led to the enabling education of artificial intelligence and

promoted the rapid development of knowledge tracking applications. Since the introduction of deep learning into knowledge tracking, knowledge tracking has developed by leaps and bounds, and more and more emerging intelligent technologies have been integrated into knowledge tracking, which provides new ideas and opportunities for further promoting the development of intelligent education.



**Figure 3.** Keyword co-occurrence map of knowledge tracking research based on deep learning in China from 2012 to 2022

After clustering the scientific knowledge map of keyword co-occurrence, the keyword cluster map is obtained as shown in Figure 3. There are six categories in total, which are sorted according to the number of nodes contained in each category: "Sao method", "knowledge tracking", "target tracking", "learning emotion", and "machine learning" and "incremental learning".

The full name of Sao is subject-action-object, which is an important semantic recognition structure in the field of natural language processing. It can deeply explore the evolution and composition relationship between technology topics, so as to identify potentially related technology innovation combinations. [9]

There are many evidences in psychological research that show that learners' learning emotions are influenced by learners' knowledge and goals, and are closely related to learners' learning outcomes. Early studies used learners' emotions to predict learning outcomes, with the aim of trying to explain human behavior in the learning process by understanding human emotional and cognitive processes. [10] Learner's affective state is accompanied by learning behavior, and learning emotions may systematically affect the way learners process learning materials. In the process of teaching, educators intervene in the negative emotions of learners in time and help students maintain the positive emotions of learners, so that learners will be more likely to achieve their learning goals successfully.

The main difference between incremental learning and batch learning is the way the trained model treats new input data. The incremental learning model learns directly on the basis of the existing model. The batch learning model needs to abandon the existing model and re-learn the new data together with the old data. [11]

### 3.3. Research development stage and future research space

In order to analyze the academic frontier information of knowledge tracking research based on deep learning in China, the emergent word map of artificial intelligence applied to teaching management and evaluation is derived through CiteSpace software. Figures shown, there are 11 emergent words under the theme of artificial intelligence applied to

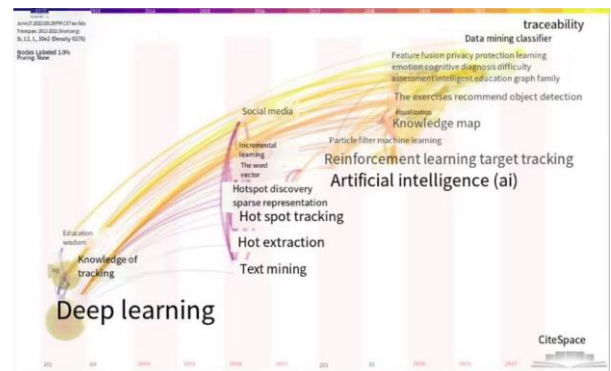
teaching management and evaluation. According to the chronological order of the beginning of the emergent words, they are "semantic mining", "concept", "graph convolution", "SAO method", 'patent analysis', 'intelligent education', 'artificial intelligence', 'exercise recommendation', 'target tracking', 'target detection' and 'cognitive diagnosis'.

**Top 11 Keywords with the Strongest Citation Bursts**

Keywords	Year	Strength	Begin	End	2002 - 2022
Semantic Mining	2002	4.7	2002	2015	Red bar from 2002 to 2015
concept	2002	4.42	2002	2015	Red bar from 2002 to 2015
Convolution Graph	2002	4.14	2002	2015	Red bar from 2002 to 2015
SAO act	2002	4.14	2002	2015	Red bar from 2002 to 2015
Patent Analysis	2002	4.14	2002	2015	Red bar from 2002 to 2015
Smart Education	2002	2.99	2002	2015	Red bar from 2002 to 2015
Artificial Intelligence	2002	1.25	2018	2019	Red bar from 2018 to 2019
Exercise Recommendation	2002	1.36	2019	2020	Red bar from 2019 to 2020
Target Tracking	2002	2.23	2020	2020	Red bar at 2020
Object Detection	2002	1.83	2020	2022	Red bar from 2020 to 2022
Cognitive Diagnosis	2002	1.66	2021	2022	Red bar from 2021 to 2022

**Figure 4.** Emergent Word Map of Knowledge Tracking Research Based on Deep Learning in China from 2012 to 2022

As can be seen from Figure 3, the emergent words in the knowledge tracking study based on deep learning can be roughly divided into three categories. The first category is the emergent words that appeared before 2015 and ended in 2015, including "semantic mining", "concept", "graph convolution", "SAO method" and "patent analysis". The second category is the emergent words whose outbreak duration is between 2018 and 2020, including "artificial intelligence", "exercise recommendation" and "target tracking". The third category is the keywords that broke out after 2020 and continue to be popular to this day, including "target detection" and "cognitive diagnosis".



**Figure 5.** Keyword co-occurrence time zone of knowledge tracking research based on deep learning in China

It can be seen from Figure 4 that the key words of "deep learning", "knowledge tracking" and "wisdom education" can be traced back to a long time ago, and the research results accumulated by predecessors have provided many advanced ideas and methods for the research in this field. Since the domestic research on knowledge tracking based on in-depth learning was officially launched in 2016, the early domestic attention in this field has focused on semantic analysis, hot spot tracking, incremental learning, reinforcement learning, etc. After 2019, the technical methods and ideas concerned in this field have changed. It turns more to target tracking, target detection, knowledge mapping, cognitive diagnosis and so on. In the last year's research, new concerns have emerged-privacy protection and traceability.

From the above change process, it can be seen that the technical focus in the field of knowledge tracking based on

deep learning has changed rapidly. In just a few years, many research teams have adopted more than ten different technical orientations to optimize the existing deep learning knowledge tracking model and put forward a variety of new answers.

At the guiding ideology level, there has also been a great change, from the initial focus on the excavation of knowledge content itself to the later gradual shift to focus on the learning process of students, the pursuit of in-depth learning knowledge tracking model and the organic combination of online learning platform, the integration of more humanized design, to the recent start of technology. More attention should be paid to the protection of students' privacy in the practical application of this technology. This transformation also echoes the historical process of our cognitive development from objectivism to humanism. The realization of technology is important, but we cannot ignore that technology serves people and should be people-oriented.

#### 4. Conclusion

As an important part of intelligent education, knowledge tracking has been widely studied and applied in the fields of artificial intelligence tutoring system, distance education and online teaching platform. However, knowledge tracking based on deep learning is a new idea that has just been born and has flourished in recent years. Many different researchers around the world have built new models or optimized existing models from different entry points, providing rich theoretical research for knowledge tracking based on deep learning.

At present, there are still various problems in knowledge tracking based on deep learning: (1) from the data level, there are problems of interdisciplinary and imbalance in knowledge tracking; (2) from the model level, there are problems of systematization and standardization in knowledge tracking; (3) from the decision-making level, knowledge tracking has problems of weak interpretation and poor transfer ability; (4) From the perspective of application, knowledge has the problems of unclear practice path and field bias.

Based on this, this paper puts forward the following suggestions: (1) introduce fuzzy logic to model the uncertainty of knowledge quantification in different disciplines, so as to carry out knowledge tracking based on fuzzy logic architecture, integrate small sample intelligent learning techniques such as transfer learning, and enhance the ability of knowledge tracking to deal with small samples and imbalance, so as to enhance the applicability of knowledge tracking; (2) Comprehensive use of pedagogy, psychology, computer and other multi-disciplinary knowledge, top-down design of a unified theoretical system from the top, referring to learner modeling standards, learning system architecture and service interfaces, and using software engineering methods to develop knowledge tracking architecture and interface standards; (3) introducing an interpretable machine learning method to explain the decision of knowledge tracking, fusing domain knowledge such as pedagogy,

psychology and the like as the explanation support of the decision, introducing an adaptive learning mechanism, and constructing a knowledge tracking method with strong transfer capability through domain adaptation; (4) Through empirical research, knowledge tracking is applied to the real intelligent environment to explore the effective application path of knowledge tracking, using intelligent data enhancement technology to eliminate the impact of bias, and combining with multi-modal analysis method to build a multi-knowledge tracking model, so as to build a universal knowledge tracking model.

#### References

- [1] BAKHSHINATEGH B, ZAIANE O R, ELATIA S, et al. Educational data mining applications and tasks: A survey of the last 10 years[J]. *Education and Information Technologies*, 2018, 23: 537-553.
- [2] Zeng Fanzhi, Xu Luqian, Zhou Yan, et al. Review of Knowledge Tracing Model for Wisdom Education [J]. *Computer Science and Exploration*: 1-27.
- [3] Chaomei. Chen. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature[J]. *Journal of the American Society for Information Science and Technology*, 2006, 3(57):359-377.
- [4] Piech C, et al. Deep knowledge tracing [C]// Cortes C, et al. *Proceedings of the 28th International Conference on Neural Information Processing Systems*. Cambridge: MIT Press, 2015:503-513.
- [5] Wang Zhifeng, Xiong Shasha, Zuo Mingzhang, et al. Knowledge Tracking from the Perspective of Wisdom Education: Current Situation, Framework and Trend [J]. *Journal of Distance Education*, 2021, 39 (5): 45-54.
- [6] PIECH C, SPENCER J, HUANG J, et al. Deep knowledge Tracing[J]. *Computer Science*, 2015, 3(3):19-23.
- [7] ROWE J P, LESTER J C. Modeling user knowledge with dynamic Bayesian networks in interactive narrative environments[C]// *AAAI Conference on Artificial Intelligence & Interactive Digital Entertainment*, 2010.
- [8] SCHODDE T, BERGMANN K, KOPP S. Adaptive robot language tutoring based on Bayesian knowledge tracing and predictive decision-making[C]// *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 2017:128-136.
- [9] MA T, ZHOU X, LIU J, et al. Combining topic modeling and SAO semantic analysis to identify technological opportunities of emerging technologies[J]. *Technological forecasting and social change*, 2021, 173:121159.
- [10] ALEPIS E, VIRVOU M, KABASSI K. Mobile education: Towards affective bi-modal interaction for adaptivity [J]. *International Journal of Interactive Mobile Technologies*, 2009, 3(2):51-56.
- [11] Liu Peilei. Research on Social Media Hotspot Detection and Tracking Technology [D]. National University of Defense Technology, 2016.