doi:10.3772/j.issn.1006-6748.2022.01.011

Study on the detection methods of S&T frontier from the multi-dimensional perspective[①]

ZENG Wen(曾　文), ZHENG Jia[②], WANG Dawei, XIONG Shuling, ZHANG Lei, WEI Xiaoqi
(Institute of Scientific and Technical Information of China, Beijing 100038, P. R. China)

Abstract

The development of network and information technology has brought changes to the information environment. The sources of information are becoming more diverse, and intelligence acquisition will be more complicated. The intelligence reflected by different dimensions of scientific and technological (S&T) data will have their own focuses. It has become inevitable to carry out the multi-dimensional research of S&T frontier, which is also a current research hotspot. This paper uses quantitative and qualitative research methods to conduct research and analysis of S&T frontier detection from three dimensions including S&T research projects, S&T papers and patents, and proposes related research methods and development tools. This work analyzes the S&T frontiers in the field of artificial intelligence and draws conclusions based on the analysis results of real and effective S&T data in three dimensions.

Key words: multi-dimensional, scientific and technological (S&T), frontier, artificial intelligence

0 Introduction

The detection of scientific and technological (S&T) frontiers is one of the main tasks of intelligence research to serve the national S&T strategic decision-making. Its purpose is to support the S&T strategic decision-making departments to study and predict S&T frontiers, grasp the general trend, seize opportunities, and make strategic planning and deployment of science and technology as soon as possible. The existing research methods basically use quantitative or qualitative research methods. The problems of quantitative research methods are mainly content analysis of data, which has strong objectivity. However, due to the single data type, the results of detection are prone to deviations. Forecasting research or activities such as technology foresight, research institutions and implementers adopts qualitative analysis method based on expert experience. Such methods are usually based on macro-policy data, and the data are limited. The limitations of the expert field background will inevitably lead to subjective biases of results. Based on this, the multi-dimensional data is used as a research object, and a combination of quantitative and qualitative research method is adopted to form a S&T frontier research method and system.

1 Related work

Through the investigation of S&T literature at home and abroad, it is found that the research on the S&T frontiers is mostly focused on the detection of research frontiers or technological frontier. The detection of frontier mostly uses S&T papers as the main source of intelligence[1], and the detection of technological fronts uses patents as the main source of intelligence. Most of the detection methods are based on quantitative or qualitative analysis methods. The detection of quantitative analysis methods focuses on the quantitative research and judgment of hotspot and high-level S&T frontier research or technology, and the detection of qualitative analysis methods focuses on subjective S&T frontier research or technology predict.

1.1 Review of quantitative analysis methods

(1) Methods based on citations: there are mainly methods based on co-citation[2-3], document coupling[4], direct citation[5-6] and the main path analysis method of citations[7-8]. The problem of above methods is the delay of citation time. For these methods it takes a certain time for papers to reach a certain citation frequency or high citation, so it is impossible to detect

[①] Supported by the National Natural Science Foundation of China (No. 72074201).

[②] To whom correspondence should be addressed. E-mail: zhengjia@istic.ac.cn.

Received on July 18, 2021

and identify emerging research hotspots and research trends.

(2) Methods based on topic words (keywords): there are mainly burst word monitoring methods[9-10], co-word method[11], and topic model based on latent Dirichlet allocation (LDA)[12]. These methods do not have the problem of the delay of citation time. They use natural language processing technology and topic statistical models to detect important research topics (keywords) from text data, but because words can express different meanings in different contexts, the expression meaning of a single word is not specific, and the users need to judge whether it is the frontier research of the field based on their own domain knowledge. Therefore, the interpretation of the detection results varies from person to person, that is, these methods have more subjective factors.

(3) Compound methods based on citation and subject terms: there are mainly methods of combining common words and citations[13], combining subject terms and citations[14], and combining full text and citations[15]. Although these methods make up for the respective shortcomings of the citation-based and topic-based methods, this method becomes complicated, cumbersome, and poor in practice when it is applied with a large amount of data.

(4) Method based on patents measurement: design and use patents evaluation indicators to determine technology frontiers in a certain field through citation analysis or cluster analysis[16]. Citation analysis has a certain delay in patent documents, and cluster analysis is affected by the number of clusters or cluster spacing, which makes the quality of clustering poor.

(5) Other methods: mainly include neural network model method[17], knowledge map method[18], essential science indicators (ESI) method[19], alternative metrology method[20], computer processing technology, and quantitative analysis method combined with the development of visualization tools. These research methods have improved the first four methods to a certain extent, but the effect is still limited.

1.2 Review of qualitative analysis methods

The realization of S&T frontier detection is supported by expert knowledge and cognition, and experts in related fields are organized by international organizations, governments, authoritative institutions or departments, S&T research management institutions, and S&T funding institutions to conduct on-site discussions, online discussions, communication surveys, interviews and other methods. It gathers the knowledge and wisdom of experts in different fields, and uses qualitative methods such as Delphi and forward-looking methods to form conclusions of prediction.

2 Main research methods and related tool development

2.1 Main research methods

This paper identifies three types of S&T data, i.e., S&T project, papers, and patents. Among them, S&T projects are the concrete implementation and concrete manifestation of a country's strategic layout in science and technology frontier, and have a certain industry or social application demand orientation, and represent the future direction of frontier. The S&T data in the field of artificial intelligence are used in this paper. Judging from the hotspot of the global research field, artificial intelligence is the core driving force of the current global industrial intelligent development, and it is also the cause of a new round of technological change and industrial revolution. In terms of research methods, this paper uses a combination of quantitative analysis and qualitative judgment. The quantitative analysis method uses word frequency statistics and topic model method (LDA) to quantitatively analyze multi-dimensional S&T data. The quantitative analysis method is mainly through the fusion of term frequency-inverse document frequency (TF-IDF), LDA, and other algorithms, as shown in Fig. 1. Functions, such as data splitting and data clustering, statistics and analysis of data in different dimensions, are used to analyze the development trend of S&T frontier research in the field of artificial intelligence with the help of expert opinions.

LDA is a topic generation model proposed in Ref. [12], which represents text based on the implicit semantics of text data, and is a semantic dimensionality reduction technology. The LDA method assumes that text data is composed of multiple topics, and each topic is composed of multiple words. The generation of a text is to select a topic with a certain probability, then selects a word under this topic with a certain probability and keep looping, finally gets a text. The use of LDA is a reverse process of generation. The process is to find the topic distribution of the text and the word distribution corresponding to the topic based on a text. The LDA topic model is a three-layer Bayesian network model, including a three-layer structure of text, topic, and words. The core idea is that each document corresponds to a topic vector that obeys the Dirichlet distribution, and each topic corresponds to a word that obeys the Dirichlet distribution. The model is shown in Fig. 2.

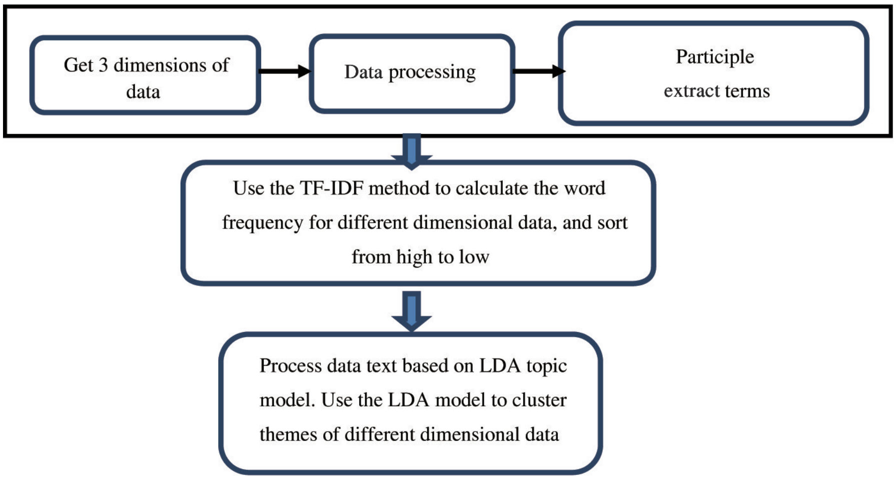

Get 3 dimensions of data
Data processing
Participle extract terms
Use the TF-IDF method to calculate the word frequency for different dimensional data, and sort from high to low
Process data text based on LDA topic model. Use the LDA model to cluster themes of different dimensional data

Fig. 1 The basic process of quantitative analysis method

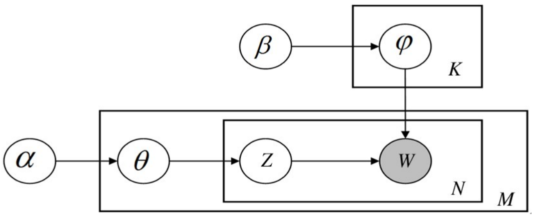


Fig. 2 Framework of LDA topic model

The explanation of Fig. 2 is shown below.

M is the total amount of text data.

K is number of topics set.

N is the number of words in a selected text.

θ is the topic vector distribution of a text.

φ is the topic distribution of each text, usually is set independently.

α is hyperparmeters of the prior Dirichlet distribution of the topic distribution of each paper, usually is set independently.

β is hyperparmeters of the prior Dirichlet distribution for each topic word distribution, usually is set independently.

Z refers to the theme given by the word in the text, which is a hidden variable.

W refers to the word of the text.

For each document, the topic generation process is as follows.

(1) Randomly assign values to all data texts and topics.

(2) For the n-th word in the m-th text, randomly assign its topic Z_k, and calculate model probability.

(3) Try to enumerate all topics Z, select a topic for the n-th word in the m-th document based on the results of these probabilities, use Gibbs Sampling formula to sample, and select the topic Z_k.

(4) If the topic Z_k chosen at this time is different from the randomly assigned at the beginning, it will have an impact on θ and φ, θ and φ will in turn affect the calculation of model probability.

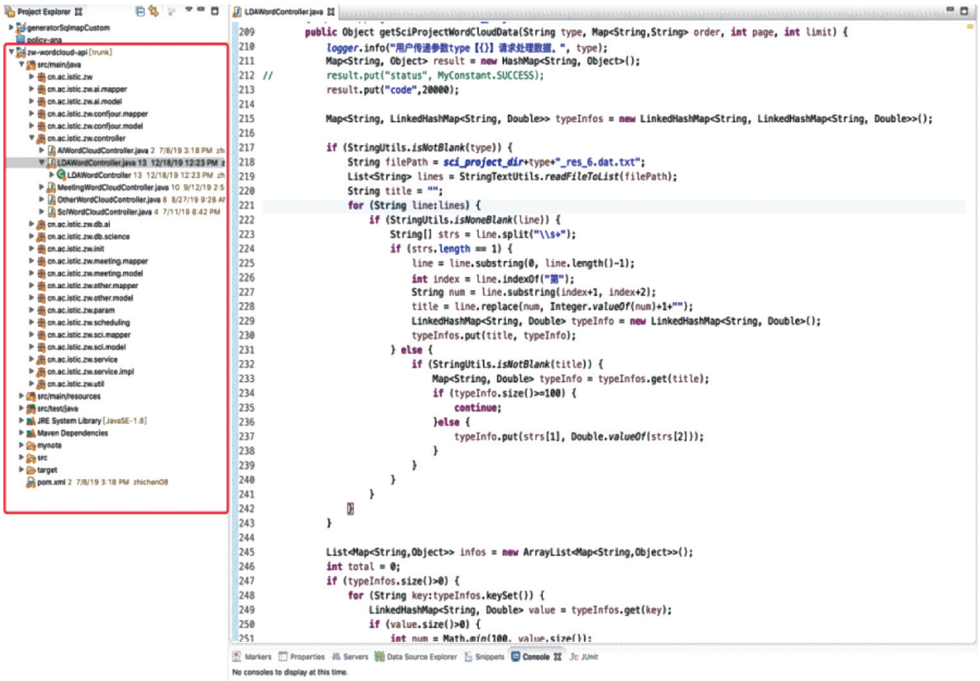
(5) Repeat Steps (3) and (4) until Gibbs Sampling converges and ends.

(6) Output the final θ and φ.

2.2 Development of quantitative analysis software

The quantitative analysis software for data is developed by using the separation of front and back ends, and is mainly divided into three main functional modules, i. e. data processing and analysis module, data preprocessing and interface generation module, and data display module. In the data processing and analysis module, Python language is mainly used to develop and integrate statistical analysis algorithms such as TF-IDF and LDA. This module mainly implements functions such as text splitting and data clustering, and provides the underlying data support for the entire software. In the data preprocessing and interface generation module, Java language is mainly used for research and development, which mainly realizes the functions of data preprocessing, data storage, and interface generation. The data display module adopts the currently popular VUE framework to realize the visual display of data. An example of the software structure of the tool is shown in Fig. 3.

Data quantitative analysis software can currently realize the management and analysis of data such as S&T strategies, papers, patents, projects in the field of artificial intelligence and life sciences. It realizes the basic visual display function. Taking the field of artificial intelligence as an example, there are 4 submenus under the menu, i. e. S&T strategies, S&T papers, S&T patents, and S&T projects. The module interface in the field of artificial intelligence is shown in Fig. 4 and Fig. 5.


Project Explorer
generatorSqlmapCustom
policy-ana
zw-wordcloud-api [trunk]
src/main/java
cn.ac.istic.zw
cn.ac.istic.zw.ai.mapper
cn.ac.istic.zw.ai.model
cn.ac.istic.zw.confjour.mapper
cn.ac.istic.zw.confjour.model
cn.ac.istic.zw.controller
AIWordCloudController.java 2 7/8/19 3:18 PM zh
LDAWordController.java 13 12/18/19 12:23 PM z
LDAWordController 13 12/18/19 12:23 PM zh
MeetingWordCloudController.java 10 9/12/19 2:5
OtherWordCloudController.java 8 8/27/19 9:28 AI
SciWordCloudController.java 4 7/11/19 8:42 PM
cn.ac.istic.zw.db.ai
cn.ac.istic.zw.db.science
cn.ac.istic.zw.init
cn.ac.istic.zw.meeting.mapper
cn.ac.istic.zw.meeting.model
cn.ac.istic.zw.other.mapper
cn.ac.istic.zw.other.model
cn.ac.istic.zw.param
cn.ac.istic.zw.scheduling
cn.ac.istic.zw.sci.mapper
cn.ac.istic.zw.sci.model
cn.ac.istic.zw.service
cn.ac.istic.zw.service.impl
cn.ac.istic.zw.util
src/main/resources
src/test/java
JRE System Library [JavaSE-1.8]
Maven Dependencies
mynote
src
target
pom.xml 2 7/8/19 3:18 PM zhichen08
LDAWordController.java
Markers
Properties
Servers
Data Source Explorer
Snippets
Console
JUnit

```
    public Object getSciProjectWordCloudData(String type, Map<String,String> order, int page, int limit) {
        logger.info("用户传递参数type [{}] 请求处理数据。", type);
        Map<String, Object> result = new HashMap<String, Object>();
//      result.put("status", MyConstant.SUCCESS);
        result.put("code",20000);

        Map<String, LinkedHashMap<String, Double>> typeInfos = new LinkedHashMap<String, LinkedHashMap<String, Double>>();

        if (StringUtils.isNotBlank(type)) {
            String filePath = sci_project_dir+type+"_res_6.dat.txt";
            List<String> lines = StringTextUtils.readFileToList(filePath);
            String title = "";
            for (String line:lines) {
                if (StringUtils.isNoneBlank(line)) {
                    String[] strs = line.split("\\s+");
                    if (strs.length == 1) {
                        line = line.substring(0, line.length()-1);
                        int index = line.indexOf("第");
                        String num = line.substring(index+1, index+2);
                        title = line.replace(num, Integer.valueOf(num)+1+"");
                        LinkedHashMap<String, Double> typeInfo = new LinkedHashMap<String, Double>();
                        typeInfos.put(title, typeInfo);
                    } else {
                        if (StringUtils.isNotBlank(title)) {
                            Map<String, Double> typeInfo = typeInfos.get(title);
                            if (typeInfo.size()>=100) {
                                continue;
                            }else {
                                typeInfo.put(strs[1], Double.valueOf(strs[2]));
                            }
                        }
                    }
                }
            }
        }

        List<Map<String,Object>> infos = new ArrayList<Map<String,Object>>();
        int total = 0;
        if (typeInfos.size()>0) {
            for (String key:typeInfos.keySet()) {
                LinkedHashMap<String, Double> value = typeInfos.get(key);
                if (value.size()>0) {
                    int num = Math.min(100, value.size());
```

No consoles to display at this time.

Fig. 3 Example of software structure

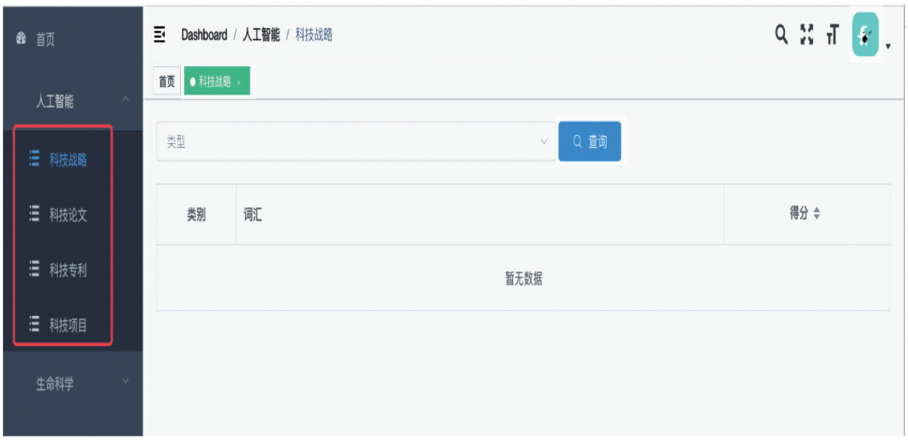


Fig. 4 Example of software interface

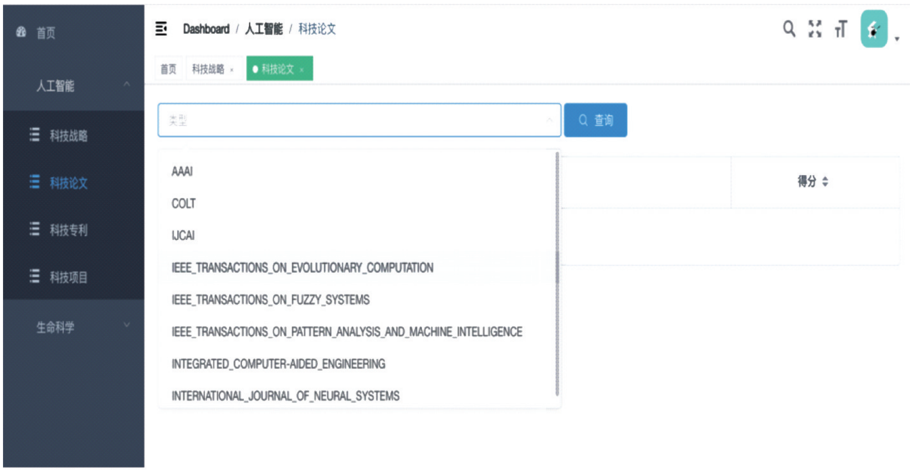
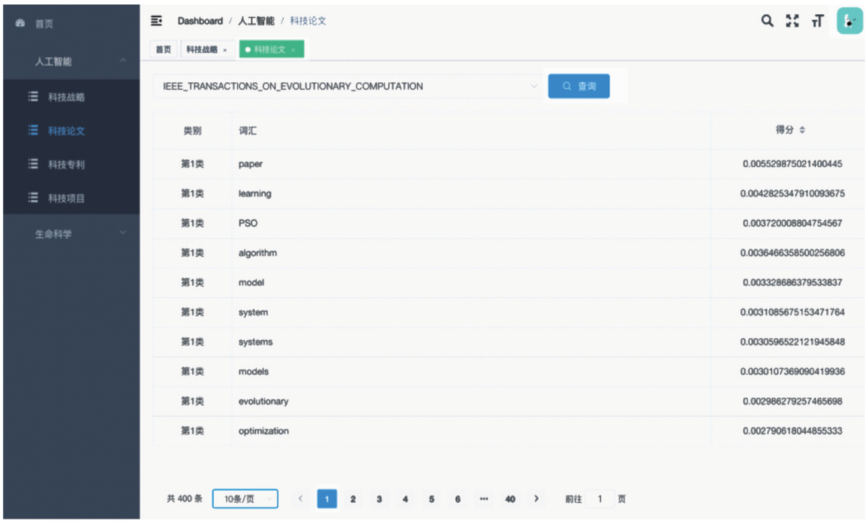

首页
人工智能
科技战略
科技论文
科技专利
科技项目
生命科学
Dashboard / 人工智能 / 科技论文
首页
科技战略
科技论文
类型
查询
得分
AAAI
COLT
IJCAI
IEEE_TRANSACTIONS_ON_EVOLUTIONARY_COMPUTATION
IEEE_TRANSACTIONS_ON_FUZZY_SYSTEMS
IEEE_TRANSACTIONS_ON_PATTERN_ANALYSIS_AND_MACHINE_INTELLIGENCE
INTEGRATED_COMPUTER-AIDED_ENGINEERING
INTERNATIONAL_JOURNAL_OF_NEURAL_SYSTEMS

Fig. 5 Data display interface of sub-menu

The display interface of data analysis result is shown in Fig. 6. In the display interface, relevant information after data analysis will be displayed. The analysis results of vocabulary in each type of data are arranged and displayed numerically according to the score automatically calculated by the software from high to low. In the display interface, there is a paging display function, and the users can choose to display a page or choose the number of data records displayed on each page.



类别	词汇	得分
第1类	paper	0.005529875021400445
第1类	learning	0.0042825347910093675
第1类	PSO	0.003720008804754567
第1类	algorithm	0.0036466358500256806
第1类	model	0.003328686379533837
第1类	system	0.0031085675153471764
第1类	systems	0.0030596522121945848
第1类	models	0.0030107369090419936
第1类	evolutionary	0.002986279257465698
第1类	optimization	0.002790618044855333

Fig. 6 Data display interface of analysis result

3 Research results

3.1 Frontier analysis of artificial intelligence based on the dimensions of papers and patents

In terms of S&T papers, this paper selects top international conference papers for analysis. In the field of artificial intelligence, in addition to academic journal papers, researchers also attach importance to top international conferences. These top conference papers can reflect hotspot research directions, the latest methods and technologies, and represent the latest advances and trends of research or technology in the field of artificial intelligence. Based on the above, through network investigation and experts consultation, this work selects the international journals of Engineering Village Complex Index Library, the Association for the Advance of Artificial Intelligence (AAAI), IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE International Conference on Computer Vision (ICCV), International Conference on Machine Learning (ICLM), and International Joint Conference on Artificial Intelligence (IJCAI) as data sources, and the actual span of the papers published is from 2011 to 2018. After data correlation calculation[21] and data cleaning, a total of 16 871 valid data are finally obtained.

In 2019, the World Intellectual Property Organization (WIPO), in order to understand the development of the field of artificial intelligence, defines artificial intelligence in three aspects including artificial intelligence technology, artificial intelligence application, artificial intelligence application field based on patents data. This work compares and analyzes results of the quantitative calculation about patents and papers, and obtains the following conclusions.

(1) The main frontier of artificial intelligence technology is machine learning, among which deep learning, latent representation, unsupervised learning, support vector machine, neural network, reinforcement learning, multi-task learning, etc. are the frontier hotspot technologies. In terms of artificial intelligence applications, the application direction of frontier is computer vision, among which biometric recognition, scene understanding, general computer vision, predictive reality, image and video segmentation, feature recognition, target tracking, etc. are frontier application directions and hotspots.

(2) The most common combinations in patents are deep learning and computer vision, computer vision and transportation, telecommunications and security, ontology engineering and natural language processing, life science research and machine learning.

(3) Machine learning is the main artificial intelligence technology, and more than one-third of all identified inventions are machine learning. The industries involved in the application fields are transportation, life sciences, personal equipment, computing and hu-

man-computer interaction (HCI), banking, security, industrial manufacturing, agriculture, and networking.

3.2 Frontiers of artificial intelligence in the dimension of S&T projects

This paper selects the open project database of 4 research institutions including the National Science Foundation (NSF), the European Commission, The Engineering and Physical Sciences Research Council (EPSRC) and the UK Research and Innovation Agency (UKRI) as the data source, and the search involves project data of artificial intelligence. The span of project selected from 2016 to 2019. After data correlation calculation[21] and data cleaning, a total of 2102 project data related to artificial intelligence is finally obtained. Through quantitative analysis, it is found that the frontiers of artificial intelligence with the most project investment are semantics, robotics, multi-modality, electronics, space technology, wireless technology, life and medical sciences, transportation, the Internet and other application fields.

3.3 The results of quantitative and qualitative analysis on frontiers about artificial intelligence

(1) Image recognition based on deep neural network

Image recognition is a recent hotspot topic in the field of artificial intelligence. In particular, deep neural network algorithms based on convolutional neural networks and recurrent neural networks are very active in the field of image recognition. At present, the hotspot topics of this research mainly focus on the two aspects of human facial recognition and action recognition, by means of deep learning related algorithms to realize the calculation of human movement, posture and spatial features. From the perspective of technical application, related researches mainly focus on pedestrian re-recognition technology, image segmentation, image retrieval, visual tracking, and 3D perspective reconstruction.

(2) Natural language processing technology based on semantic understanding

Natural language processing is developing from the understanding of word meaning and sentence meaning to the direction of multiple semantics and knowledge discovery. Hot topics involve learning methods for natural language processing tasks, interpretable natural language processing based on semantic analysis, knowledge and common sense, context modeling, and multiple rounds of semantic understanding. From the perspective of specific application scenarios, the application of basic researches such as machine translation based on neural networks, natural communication between humans and machines using natural language, machine reading comprehension, information retrieval, knowledge question and answer, automatic abstracting, etc., are relatively research hotspots.

(3) Unsupervised learning algorithm

Machine learning methods have presented two major development directions, supervised learning and unsupervised learning. In terms of unsupervised learning, automatic classification of multi-dimensional data, automatic data indexing, and hidden structure of learning data are research hotspots. From the perspective of learning characteristics of data, there are two evolutionary characteristics of machine learning, the transformation of existing labeled data to unlabeled data, the transformation of synthetic data to real data. At present, because the accumulation of real basic data is far from meeting the needs of model training, synthetic data is the data basis for unsupervised machine learning. The field of semantic segmentation algorithms, fully convolutional adaptive networks and unsupervised adaptive semantic segmentation are frontier hotspots.

(4) Brain-like intelligence

Research on brain-like intelligence based on bionics has long been incorporated into the artificial intelligence development strategy of various countries for the key layout. From an overall point of view, global brain-like intelligence has two frontier areas, including brain-like computing and brain-like bionic control. Among them, the research of brain-like computing mostly starts from the two directions of neural network and neuron, and mostly starts from the perspective of brain mechanism and functional structure. Brain chips are the main frontier hotspot of brain-like computing in hardware. The brain-computer interface is also one of the frontier hotspots. The human-computer hybrid system is controlled through the brain-computer interface. Most of the frontier research in brain-like bionic control comes from the field of intelligent robots. Research on multi-modal perception and brain-like autonomous decision-making based on perceptual information is one of the frontier hotspots. In addition, the development of new materials, new transmission devices, new materials and new transmission systems have enabled a series of breakthroughs in muscle-like drivers.

4 Conclusion

From the perspective of information research methods, this paper combines quantitative analysis and qualitative analysis methods, extends the single-dimensional detection to multi-dimensional detection of S&T frontier, breaks through the limitations of the existing detection method framework, and applies it to frontier research in the field of artificial intelligence. Research results show that artificial intelligence technology has shifted from theoretical research to commercial product and service applications focuses on the field of machine learning and computer vision. What needs to be pointed out is that there is currently no unified standard or model for the reliability or validity of the detection results of S&T frontiers, and most of researches use expert evaluation methods. For the credibility of the detection results of the frontiers of S&T, the paper adopts the method combining quantitative analysis of data and expert opinions. In terms of obtaining multi-source data, the paper adopts the research method proposed in Ref. [21], and combined with the terms of artificial intelligence provided by domain experts. These information is used to filter or retrieve data to ensure the validity of the multi-dimensional data source and at the same time ensure the validity of the data analysis results. In the future, we will conduct further research on the evaluation methods of the S&T frontiers.

Reference

[1] SUN M W. Exploration on frontier research field of informatics[C] // International Conference on Economic Development and Education Management, Dalian, China , 2019: 525-528

[2] NAVONIL M, KORINA K, PAUL F. Exploring the modeling and simulation knowledge base through journal co-citation analysis [J]. *Scientometrics*, 2014, 98 (3): 2145-2159

[3] TRUJILLO C M, LONG T M. Document Co-citation analysis to enhance trans-disciplinary research [J]. *Science Advances*, 2018, 4(1) : e1701130

[4] MALIK K H, ALI D. Anomaly detection in heterogeneous bibliographic information networks using co-evolution pattern mining[J]. *Scientometrics*, 2017,113(1):149-175

[5] SHIBATA N, KAJIKAWA Y, TAKEDA Y, et al. Detecting emerging research fronts based on topological measures in citation networks of scientific publications [J]. *Technology Innovation*, 2008, 28(11): 758-775

[6] AYSE K F, WEI L W, STUART M. Technological forecasting-a review[EB/OL]. http://web. mit. edu/ smadnick/OldFiles/www/wp/2008-15. pdf: MIT, (2008-09-01),[2021-05-18]

[7] VIKTOR B, ANASTASIYA S, OLGA E, et al. Russia on world research front of industrial scientific direction[C] //International Conference on Actual Issues of Mechanical Engineering, Suzhou, China, 2017: 113-119

[8] LU L Y, LIU J S. A novel approach to identify research fronts of tourism literature[C] //2015 International Conference Management of Engineering and Technology, Portland, USA, 2015: 2211-2217

[9] NAOKI S, YUYA K, YOSHIYUKI T, et al. Detecting emerging research fronts in regenerative medicine by the citation network analysis of scientific publications[C] // International Conference on Mangement of Engineering and Technology, Portland, USA, 2009: 2964-2976

[10] KLEINBERG J. Bursty and hierarchical structure in streams [J]. *Data Mining and Knowledge Discovery*, 2003, 7(4): 373-397

[11] ZHAO L M, ZHANG H. Research frontier of Chinese digital library in big data context [J]. *Information Science*, 2019, 37(3): 97-104 (In Chinese)

[12] BLEI D M. Probabilistic topic models[J]. *Communications of the ACM*, 2012, 55(4): 77-84

[13] BAI J E, YAN D W, CHEN Q. Trend prediction of emerging topics based on topic model and curve fitting [J]. *Information studies: Theory & Application*, 2020,43 (7) :130-136,193

[14] Ma T, Cao J M. Knowledge transfer research evolution path combing and frontier hot spot analysis[J]. *Soft Science*. 2016(2): 121-125 (In Chinese)

[15] VAN D, BESSELAAR P, HEIMERIKS G. Mapping research topics using word reference co-occurrences: a method and an exploratory case study[J]. *Scientometrics*, 2006, 68 (3): 377-393

[16] KONG D J, LI M, ZHENG W J. To identify technology frontier for mass-customized production service converged with artificial intelligence based on patent data mining [C]//The 15th International Conference on Service Systems and Service Management, Hangzhou, China, 2018: 1-6

[17] FUJIMAGARI H, FUJITA K. Detecting research fronts using neural network model for weighted citation network analysis[J]. *Journal of Information Processing*, 2015, 23 (6): 753-758

[18] PARK I, LEE K, YOON B. Exploring promising research frontiers based on knowledge maps in the solar cell technology field [J]. *Sustainability*, 2015, 7 (10): 13660-13689

[19] SUN W, WU L, HAO X. Agricultural research front detection and national cooperative performance analysis based on ESI[C]//2016 3rd International Conference on Systems and Informatics, Shanghai, China, 2016: 1095-1100

[20] ZHANG D, LENG F H. Research front tracking method and empirical research based on citing papers[J]. *Information Science*, 2020,38(4): 62-69 (In Chinese)

[21] WANG X L, DONG C, ZENG W, et al. Survey of data a value evaluation methods based on open source scientific and technological information[C]//The 5th International Conference of Pioneering Computer Scientists, Engineers and Educators, Guilin, China, 2019: 172-185

ZENG Wen, born in 1973. She received her Ph. D degree in Shenyang Institute of Automation, Chinese Academy of Sciences in 2009. She also received M. S. degree from the College of Information Science and Engineering at Northeastern University in 2003. Her current research interests include S&T information theory and method, S&T frontier, regional innovation and development research.