Journal of Scientometric Res. 2021; 10(1):18-26
http://www.jscires.org

Research Article

Research Fronts of Computer Science: A Scientometric Analysis

Wanjun Xia[1,*], Yanping Jiang[2], Weifeng Zhu[2], Shuang Zhang[2], Tianrui Li[1]

[1]School of Information Science and Technology, Southwest Jiaotong University, Chengdu, CHINA.
[2]Southwest Jiaotong University, Chengdu, CHINA.

ABSTRACT
Computer science and technology have developed rapidly in the past few decades and shown an increasing tendency of interdisciplinary research in the community. Research fronts of Computer Science (CS) have attracted the attention of scientists from different background and it is a big challenge for them to discover the development trends. The study uses scientometric methods and a combination of macro and micro analysis to detect the research fronts of CS based on the data from Scopus and Scival database. Macro analysis focuses on leading countries and institutions by scholarly output and citation count. Micro analysis pays attention to the performance of institutions and their competitors in research fronts and helps researchers understand frontier topics of specific research field. This paper provides a comprehensive and finer-grained analysis about the research frontier topics of CS domain. The insights obtained from the analysis are for both researchers and policy makers.

Keywords: Computer Science, Research fronts, Direct citation, Scientometric analysis, Scival.

Correspondence
Wanjun Xia
School of Information Science and Technology, Southwest Jiaotong University, Chengdu-611756, CHINA.
Email: 13058723@qq.com

Received: 20-11-2020
Revised: 12-01-2021
Accepted: 01-02-2021
DOI: 10.5530/jscires.10.1.3

INTRODUCTION

Research fronts are a key problem of scientific research and guide the direction of scientific development. It is generally considered that research fronts are the most advanced, newest and potential topic or field in scientific research. Price[1] first proposed the concept of research fronts and mentioned the research front as “growing tip, or epidermal layer” of the literature. Garfield[2] defined research fronts as “co-citation clusters and the documents that cite them.” Chen[3] defined the research fronts as a set of emergent dynamic concepts and potential research problems. Upham and Small[4] considered that research fronts represented the most dynamic areas of science and technology. How to grasp the research fronts scientifically and accurately has attracted the attention of researchers and policy makers. The study of research fronts can help researchers understand how their academic fields emerged and how they are currently developing.[5] Research managers and policy makers need to keep abreast of the progress and dynamics in scientific research, rationally allocate resources, effectively evaluate scientific achievements and promote scientific progress with limited resources. Timely grasping the research fronts can help them make more reasonable plans.

CS is the most active and fastest growing part in the field of science and technology. The application of computer science and information technology has penetrated into all aspects of social life and become an important engine to promote social progress. The progress of computer science has not only improved people’s living standards and production efficiency, but also been considered as a key factor of the development of countries. The frontiers of CS have attracted extensive attention from academia and industry. It is a great challenge for researchers from different background to discover the research fronts of CS. Therefore, for researchers and policy makers, it is of great significance to assess research fronts of CS.

This paper tries to make a finer-grained analysis of research fronts in CS, which help us see the detailed structure and dynamics of recent science at the research problem level. Our research aims at deal with the following research questions:

- What are the research fronts or hotspots in CS?
- Which countries and institutions are leading in research frontier topics of CS?
- Which frontier topics are our institutions and the competitors currently active in?
- What are frontier topics in a research field?

- What is the development profile of a frontier topic?

This paper is organized as follows. In "Related work", we discuss previous related works about research fronts detection in CS. Section "Methodology and data" introduces the methodology followed and data used. Section "Results and discussion" describes analysis of research fronts in CS and several case studies in micro and macro levels respectively. Section "Conclusion" presents a summary of the work.

RELATED WORK

Scientists have proposed various methods to detect research fronts, which can be divided into two approaches: Citation-based and term-based. On the one hand, the use of citation-based methods which help us understand the structure of science has a rich history, including co-citation, bibliographic coupling and direct citation.[6–8] Research fronts can also be expressed by the emergence of new topics or changes in the relationship between keywords. In CS, several studies have analyzed the topic trends or research fronts. Tattershall *et al.*[9] explored a stock market-inspired burst detection methodology to the free text of a large corpus of CS abstracts which is gathered from DBLP. 2.6 million articles from 1988 to 2017 were used to detect popularity of research topics. It turns out that topics such as "deep learning", "word embedding" and "fog computing" were in the top bursty terms list and would rise in popularity in the future. Wu *et al.*[10] studied the research topic trends by analyzing the evolution of topics the authors with uninterrupted and continuous presence worked on. They found that the community showed an increasing tendency of interdisciplinary research. Hoonlor *et al.*[11] utilized bursty words detection to study trends of CS. On the other hand, text-based analysis is applied to identify major research topics of CS in some countries. Uddin *et al.*[12,13] identified topic trends through frequency of keywords in Mexico and the SAARC countries. Based on the research output data of 100 most productive institutions from India and the world over different time periods, Singh *et al.*[14] implemented the burst detection algorithm to analyze research topics.

The abovementioned studies used keyword frequency or burst words to detect topic trends or frontier topics. But sometimes it is not easy to fully express the meaning of research fronts through a single keyword. Burst words such as "artificial intelligence", "semantic web" and "data mining", cover a wide range and are not precise enough to represent the specific meaning of research fronts. And it is also not easy to know the institution's involvement in research fronts and effectively evaluate the relationship between research direction and research fronts.

Our work combines with the direct citation method, which has been proven to detect large and young emerging clusters earlier and show better performance in detecting research fronts.[15] Different from the previous methods of using a single keyword, our work uses three words simultaneously to express a topic, which can represent the topic more accurately. In addition, leading countries and institutions in each frontier topic and the performance of institutions and their competitors in research fronts are also discussed. As far as we know, our work is the first to make a comprehensive and finer-grained analysis of research fronts in CS.

METHODOLOGY AND DATA

Subject classification is based on All Science Journal Classification (ASJC) subject system of Scopus database and CS is one of 27 major subject areas in ASJC classification. This study uses Scival, a scientific analysis tool, to generate research topics of CS. It is based upon a direct citation analysis of 75 million literature data in Scopus database from 1996 forward. Different topics were generated by clustering the direct citation references. The prominence of each topic, which is an indicator of the momentum of a particular field, is calculated by citation count, views count and journal impact based on the literature data during 2017-2019. Research fronts could be obtained through threshold setting of the prominence value. Data collection time was October 30, 2020. Analyzing schema is depicted in Figure 1.

Creation of topics

The general process of the direct citation is as follows: First, a list of citing-cited pairs is created. Each pair is assigned a weight based on the link relationship and the weight a_{ij} between each pair of papers i and j is set to $1/k$ where k is the number of edges for the paper j.[16] Then, papers are assigned to clusters via VOS algorithm, which uses a variant of modularity-based clustering and attempts to maximize the ratio of links within clusters to links between clusters.[17] Each topic consists of a

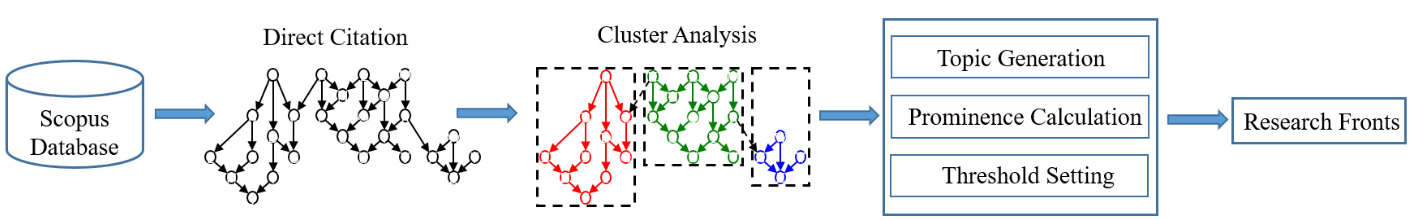

Figure 1: Analyzing schema of this study.

set of publications with a common focused intellectual interest and one publication can only belong to one topic.

Generation of topic names

Each topic is named by extracting from Elsevier Fingerprint Engine, which uses Natural Language Processing techniques to mine the text of titles, abstracts and keywords in the literature. Each topic consists of three distinctive keyphrases. The first two are generally high-frequency keyphrases which are selected to provide a macro-level description of the topic in the research field. And the third keyphrase is a more specific description of the topic.

Calculation of topic prominence

Prominence is calculated by the combination of recent citation count, recent Scopus views count and CiteScore value. Scopus views count is the sum of abstract views and clicks on the link to view the full-text. CiteScore is an indicator to evaluate the academic influence of a journal. To calculate prominence, the following variables are considered by the topic and year *n*:[16]

- Citation Count in year *n* to papers published in *n* and *n*–1;
- Scopus Views Count in year *n* to papers published in *n* and *n*–1;
- Average CiteScore for year *n*.

According to the analysis of Klavan's and Boyack,[16] the three variables (Citation count, Views count, CiteScore) are highly strongly correlated, through three-variables factor analysis, the normalized factor scores are calculated as 0.495, 0.391, 0.114, respectively, which represent the weight of each variable. Then prominence of topic *j* in year *n* is calculated as the following equation:

$$p_j=0.495(C_j-mean(C_j))/stdev(C_j)+0.391(V_j-mean(V_j))/stdev(V_j)+0.114(CS_j-mean(CS_j))/stdev(CS_j)$$

where *cj* is citation count to articles in cluster *j* published in years *n* and *n*–1, *vj* is Scopus views count to articles in cluster *j* published in years *n* and *n*–1 and *csj* is average CiteScore for articles in cluster *j* published in year *n*. These raw values are log-transformed into the values used in the formula as $Cj = ln(cj + 1)$, $Vj = \ln(vj + 1)$ and $CSj = ln(csj + 1)$.

Selection of research fronts

The percentile is calculated after sorting by the topic prominence. The higher the prominence percentile, the more attention the topic receives and the better its growth momentum. The prominence percentile is calculated based on citation count and views count of publications in the past two years, which reflects characteristics of high attention and novelty, so it can represent research fronts. According to experiences, topics with prominence percentile greater than 90 are considered as research fronts, while those greater than 99 are hot research fronts.[18,19]

RESULTS AND DISCUSSION

Q1: What are research fronts or hotspots in CS?

CS covers 15,460 topics, accounting for 16% of the total topics, including more than 500 research frontier topics (prominence percentile>90) and 136 hot research frontier topics (prominence percentile>99). Table 1 lists the 20 research frontier topics with highest prominent percentiles, we can see that it covers popular fields such as deep learning, block chain, natural language processing, recommendation systems and Internet of Thing, etc. The topics listed are a more granular portfolio analysis of a research field, e.g. hot topics related to natural language processing, include named entity recognition and product viewpoint mining, etc.

According to the statistics of the scholarly output of the top 500 frontier topics, the least is 680, the most is 19,030 and the median is 1048. Table 1 shows that the scholarly output of the top 20 topics is almost all over 2000 during 2017-2019, except for topic "Interatomic Potential; Potential Energy Surface; Material Science", which is a cross-topic and mainly about the application of machine learning in the field of materials science. Journals in the field of materials science often have a high CiteScore, so this topic also has a high prominence value. Previous study showed that there was a moderate positive correlation between the number of publications and the prominence ranking of topics,[18] that is, the more the number of publications in a topic, the higher the prominence value might be. At the same time, Field-Weighted Citation Impact (FWCI, a normalized impact indicator, more than 1.0 of which indicates publications have been cited more than the global average for similar publications) of these topics are all above 1 and some of them more than 2, indicating that these topics also have a higher influence.

Figure 2 shows the top 100 frontier topics by the scholarly output in CS. Each bubble represents a topic and the size of the bubble indicates the scholarly output of a topic. The position of the bubble is determined by the ASJC subject. The closer the bubble is to the center, the more multidisciplinary the topic is. For example, the biggest bubble is "Object Detection; CNN; IOU", which has the most scholarly output. The topic "Exome; Copy Number Variation; Whole Genome Sequencing" belongs to bioinformatics, showing strong multidisciplinary characteristic.

Q2: Which countries and institutions are leading in research frontier topics of CS?

Figure 3 shows the top 10 countries by scholarly output during 2017-2019. China, the United States and India are the most

Table 1: Top 20 frontier topics in CS.

Rank	Topic	Scholarly Output	FWCI	Citation Count	Views Count	Average CiteScore	Prominence Percentile
1	Object Detection; CNN; IOU	19030	2.5	36,939	65,352	5.68	99.999
2	Bitcoin; Ethereum; Blockchain	5755	2.99	11,243	77,910	3.25	99.978
3	Demand Response; Demand Side Management; Energy Trading	4633	1.47	7,210	27,097	6.34	99.955
4	Edge Computing; Task Scheduling; Location Awareness	4307	2.64	9,737	19,675	5.43	99.953
5	Electronic Word-Of-Mouth; Online Review; Brand Community	3523	1.41	4,447	51,612	4.23	99.941
6	Fog Computing; Block chain; Internet Of Thing	4462	2.01	6,983	33,558	3.09	99.938
7	Aggregation Operator; Pythagorean; Group Decision Making	2768	2.8	9,108	16,741	4.45	99.934
8	Technology Acceptance Model; Mobile Payment; UTAUT	4186	1.35	4,709	53,649	2.98	99.933
9	Rolling Bearing; Rotating Machinery; Fault Diagnose	3805	1.67	6,671	20,517	3.54	99.912
10	Collaborative Filtering; Recommende System; Implicit Feedback	5483	1.54	5,994	20,446	3.11	99.891
11	Exome; Copy Number Variation; Whole Genome Sequencing	2314	1.37	4,791	10,253	8.06	99.855
12	Consensus Problem; Formation Control; Output Regulation	3980	1.23	5,074	12,400	5.4	99.853
13	Landsat; Land Cover; Cropland	2866	1.01	3,442	15,544	7.95	99.852
14	Ad Hoc Network; Unmanned Aerial Vehicle; Base Station	2312	2.68	5,919	9,137	5.8	99.851
15	Sentiment Classification; Named Entity Recognition; Entailment	5249	1.78	6,258	14,953	2.58	99.847
16	Smart City; Municipal Administration; Internet Of Thing	2864	1.66	3,287	32,724	3.02	99.846
17	Sentiment Classification; Opinion Mining; Product Review	4830	1.22	3,904	27,378	2.54	99.839
18	Software Defined Networking; Traffic Engineering; Denial-Of-Service Attack	4653	1.67	4,532	13,944	4.6	99.832
19	Boltzmann Machine; Belief Network; Generative	2857	1.87	5,047	11,889	4.41	99.827
20	Interatomic Potential; Potential Energy Surface; Material Science	953	2.57	3,833	10,866	8	99.824

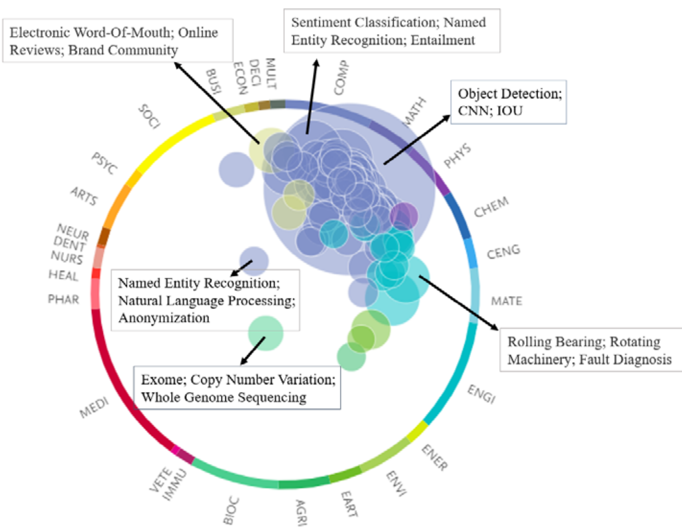

Figure 2: Top 100 frontier topics by scholarly output in CS.

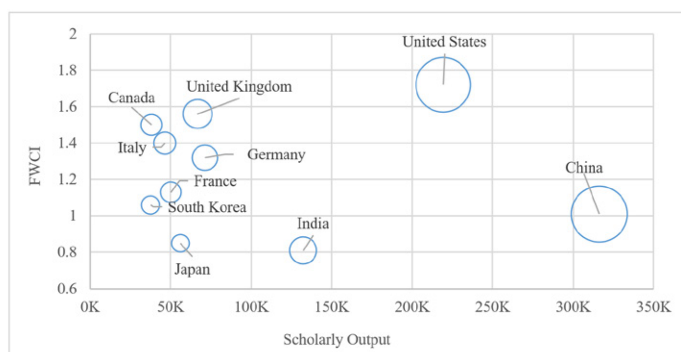

Figure 3: Top 10 most productive countries in CS during 2017-2019.

productive countries, especially China, which is far ahead. For FWCI, the United States, the United Kingdom and Canada have greater influence, while China only reaches the average level (FWCI=1).

According to the scholar output of 136 hot research fronts, China ranks first in 91 frontier topics, far ahead of other countries. The United States has 32 leading topics, India with 6, Germany with 3, Italy, Malaysia, Russia and the United Kingdom have 1 respectively (Figure 4). And according to citation count of publications in hot research fronts, China leads in 68 topics and the United States leads in 59 (Figure 5), indicating that China has an advantage in the number of publications, while the United States has a higher impact in CS domain.

Among global research institutions, Chinese Academy of Sciences leads in 27 hot research fronts by scholarly output.

Beijing University of Posts and Telecommunications ranks the first in the academic sector with 12 hot research frontiers, followed by CNRS with 7 leading topics. There is a total of 11 institutions with more than three leading topics (Figure 6). According to citation count, Chinese Academy of Sciences leads in 7 hot topics, Alphabet Inc. leads in 5, followed by Beihang University, Harvard University and University of California at Berkeley with 4 leading topics (Figure 7). Chinese Academy of Sciences' leading position by citation count is less obvious than that by the scholarly output.

Q3: Which research frontier topics are our institutions and the competitors currently active in? - A case study

This study takes Southwest Jiaotong University as an example and Beijing Jiaotong University is selected as a benchmarking institution. This question focuses on the topics of two universities as key contributors. If the institution has at least 1/3 as many papers (or 1/3 as many citations) as the top publishing institution (or the top cited institution) in a topic, it is considered as a key contributor in a topic. This study selects 6 topics for which the two universities are the key contributors and with large amount of the scholarly output in this domain. Table 2 and Table 3 show that prominence percentiles of the most productive topics of two universities are all greater than 90, indicating that their main research directions are the current frontier topics. Beijing Jiaotong University mostly concentrates in the field of ground transportation. According to the scholarly output, it ranks first in the topics of scheduling, traffic flow prediction and pedestrian evacuation, etc. Southwest Jiaotong University mainly contributes in fault diagnosis, magnetic levitation, traffic scheduling and rough sets, ranking first in incomplete information system.

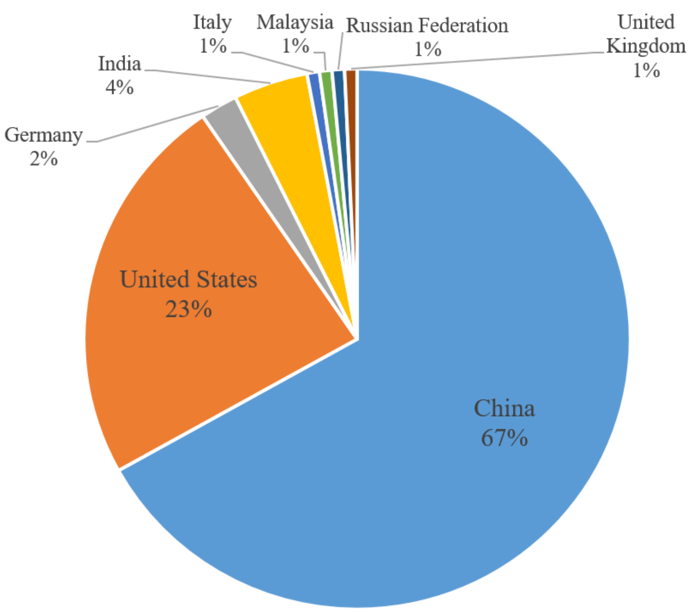

Figure 4: Leading countries by scholarly output.

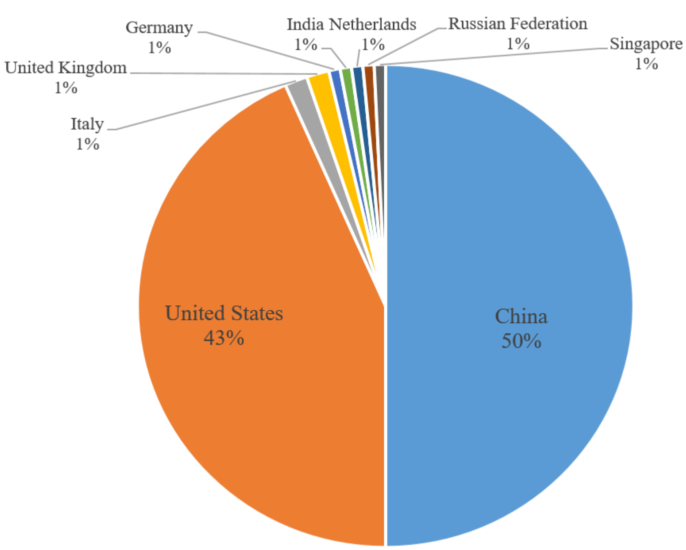

Figure 5: Leading countries by citation count.

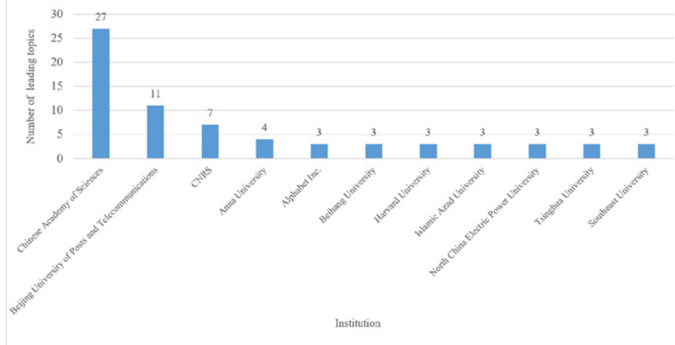

Figure 6: Leading institutions by scholarly output.

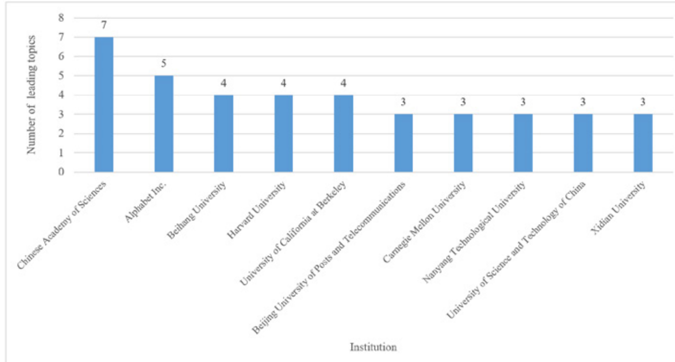

Figure 7: Leading institutions by citation count.

Q4: What are frontier topics in a research field？-A case study

Analyzing the distribution of publications in a certain research field over time can show the development and changes of the research problem and also help us discover the key nodes and emerging trends. In this study, we take Natural Language Processing (NLP) as an example to explore the key nodes and emerging trends of this research problem. NLP is an important direction in the field of CS and is regarded as one of the core problems of AI-complete. There are five hot topics related to NLP, as shown in Table 4.

Figure 8 shows the trends of these five topics in the past 10 years, which have been increased gradually. Two topics with the most scholarly output are T.108 and T.1614, with more than 2000 in 2019, respectively. T.108 is about sentiment analysis and opinion mining, which is the computational study of people's opinions, appraisals, attitudes, emotions toward entities, individuals, issues, events, topics and their attributes.[20] At present, a large number of comments have been accumulated on e-commerce sites. Comment information is closely related to people's daily life and is widely used by consumers and business organizations. When ordinary consumers buy a certain product or service, they generally refer to the comment information of previous users to obtain a feedback. Review information on e-commerce sites generally has better structure and is widely used by academia and industry. Before 2017, this topic is the most productive among the five hot topics.

During 2017-2019, T.1614 was the most productive topic. It is mainly about the words and sentences embedding representation which can be applied to various downstream tasks, such as sentiment classification and textual entailment. It has become an important part of NLP system based on deep learning. In 2013, Google released a tool, word2vec, for word vector calculation, providing an efficient method for learning high-quality word vector representation from a large amount of unstructured text data, which has attracted great attention of industry and academia.[21,22] This topic entered a period of rapid development after 2013. Then from GloVe,[23] ELMo[24] to Bert,[25] which were the most cited in 2014, 2018 and 2019 respectively in Scopus, language representation achieves milestone development and has become one of the fastest growing topics at present. T.4431 focuses on the application of NLP in medical data, especially electronic medical records. It can be seen that although it is one of hot topics in NLP field,

Table 2: Main research topics of Beijing Jiaotong University.

Topic	Scholarly Output	FWCI	Prominence percentile	Rank
Timetabling; Rescheduling; Urban Rail Transit	176	1.14	96.927	1
High-Speed Train; Urban Rail Transit; Regenerative Braking	176	0.93	94.224	1
High-Speed Railway; Handover; Local Thermodynamic Equilibrium	129	1.45	95.734	1
Rolling Bearing; Rotating Machinery; Fault Diagnosis	83	1.22	99.912	4
Pedestrian Flow; Evacuation; Crowds	81	0.92	99.162	1
Traffic Flow; Travel Time; Advanced Traveler Information Systems	72	2.4	99.293	1

Table 3: Main research topics of Southwest Jiaotong University.

Topic	Scholarly Output	FWCI	Prominence percentile	Rank
Rolling Bearing; Rotating Machinery; Fault Diagnosis	86	1.91	99.912	3
Multiple Access; Power Allocation; Successive Interference Cancellation	64	5.28	99.732	8
Timetabling; Rescheduling; Urban Rail Transit	63	0.52	96.927	2
Pedestrian Flow; Evacuation; Crowds	49	1.27	99.162	3
Attribute Reduction; Fuzzy Rough Sets; Incomplete Information System	47	3.15	98.9	1
Magnetic Levitation; Guideways; Suspension Systems	44	0.46	90.868	2

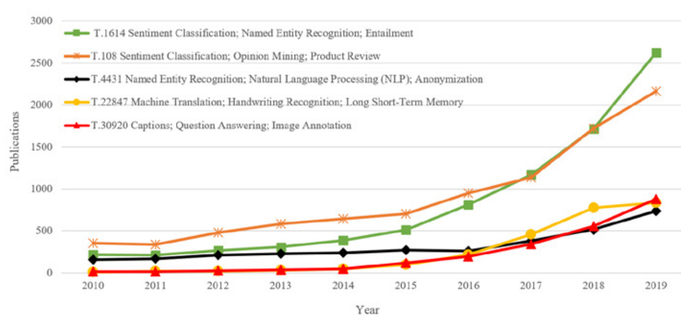

Figure 8: Scholarly output trends of 5 hot frontier topics in NLP during 2010-2019.

Table 4: Hot frontier topics in NLP.

Topic	Topic Number	Citation Count	Scopus Views Count	Average CiteScore	Prominence percentile
Sentiment Classification; Named Entity Recognition; Entailment	T.1614	6,258	14,953	2.58	99.847
Sentiment Classification; Opinion Mining; Product Review	T.108	3,904	27,378	2.54	99.839
Captions; Question Answering; Image Annotation	T.30920	3,340	3,827	11.23	99.660
Machine Translation; Handwriting Recognition; Long Short-Term Memory	T.22847	2,834	6,483	2.87	99.414
Named Entity Recognition; Natural Language Processing; Anonymization	T.4431	1,924	8,201	4.08	99.387

but it is not developing as fast as T.108 and T.1614 (Figure 8). The main reason is that this field faces problems such as non-open data, data islands, data privacy and ethical issues. Data quality, structuring and standardization of medical records are the primary issues that need to be resolved. In September 2016, the Laboratory for Computational Physiology of MIT released the third edition of MIMIC-III (Medical Information Mark for Intensive Care) data set, comprising information relating to patients admitted to critical care units at a large tertiary care hospital.[26] Since then, the number of publications on this topic has increased.

In contrast, T.30920 and T.22847 are not the most productive, but they still get a high prominence. T.22847 is about machine translation and the most cited paper comes from Google's implementation of an English to French translation task with a multilayered Long Short-Term Memory (LSTM) in 2014.[27] This article proposed the use of RNN Encoder-Decoder in neural machine translation (NMT), that was, the well-known Seq2Seq model and also laid the foundation for NMT. Since 2015, the number of publications of this topic has gradually increased. Especially in 2016, Google released Google Neural Machine Translation (GNMT), which meant that NMT has become the absolute mainstream of modern machine translation and one of the most popular topics in the current NLP field. T.30920 mainly focuses on image captioning, which can automatically describe the content of images. It is a fundamental problem in artificial intelligence that connects computer vision and NLP. The most frequently cited publication in this topic comes from Microsoft' open data set, namely COCO (Common objects in context), providing a data basis for subsequent image caption research. Starting from the work of *Show and Tell*[28] published in 2015, the field of image description has developed rapidly in recent years. Models have been improved gradually by adding attention mechanism, visual sentinel, improved CNN, reinforcement learning and object detection. This topic combines two major directions of artificial intelligence: Computer vision and NLP. It is also one of hot frontier topics in NLP field.

Q5: What is the development profile of a topic? - A case study

The topic "Captions; Question Answering; Image Annotation (T.30920)" is a relatively new field and has developed rapidly after 2015 (Figure 8). Figure 9 shows the top 50 key phrases, including question answering, caption, video, semantic, NLP system, computer vision, etc. This topic contains two main tasks: Image captioning and visual question answering. They use a combination of computer vision and NLP technology to deal with images and text in order to get the answer of the image question, which can be applied to image retrieval and life assistance for the visually impaired, etc.

In this topic, the most productive countries are China (741 scholarly output) and the United States (555 scholarly output), both of which account for 73% of all publications during 2017-2019. Most active institutions are shown in Table 5. Chinese Academy of Sciences has the most scholarly output, while Microsoft USA, Facebook Inc, Georgia Institute of Technology and Alphabet Inc. have a higher impact. Georgia Institute of Technology has the highest percentage of highly cited papers and its top 10% citation percentiles is 68.3%. The most active authors are shown in Table 6. These authors not only have a high scholarly output, but also have a high proportion of highly cited papers. The top 5 most active sources by scholarly output are shown in Table 7. The top conferences such as CVPR, ICCV, IJCAI and NIPS are the

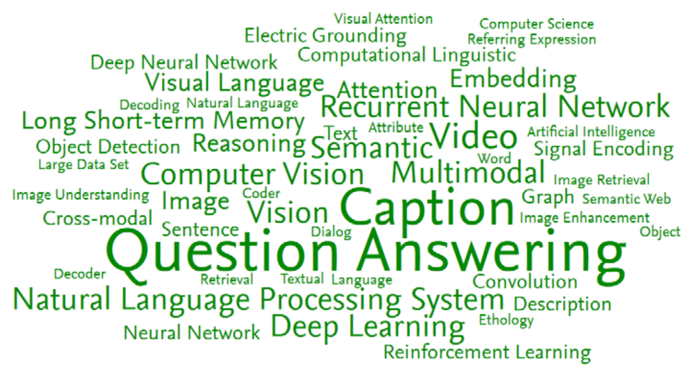

Figure 9: Top 50 keyphrases of Topic T.30920.

Table 5: Top 10 productive institutions in Topic T.30920.

Institution	Scholarly Output	FWCI	Citation Count	Top 10% Citation Percentiles (%)	International Collaboration (%)
Chinese Academy of Sciences	103	2.48	895	24.3	30.1
University of Chinese Academy of Sciences	66	2.2	537	21.2	22.7
Microsoft USA	57	5.61	1599	36.8	49.1
Tsinghua University	56	3.2	592	33.9	58.9
Carnegie Mellon University	55	3.06	825	27.3	38.2
Zhejiang University	52	3.63	992	30.8	34.6
Facebook Inc	49	5.11	1253	49	22.4
Georgia Institute of Technology	41	7.53	1509	68.3	17.1
CAS - Institute of Automation	40	1.57	286	22.5	10
Alphabet Inc.	37	7.64	1227	51.4	27

Table 6: Top 5 productive authors in Topic T.30920.

Author	Affiliation	Scholarly Output	FWCI	Citation Count	Top 10% Citation Percentiles (%)	International Collaboration (%)
Batra, Dhruv	Georgia Institute of Technology	28	7.31	977	67.9	14.3
Parikh, Devi	Georgia Institute of Technology	26	8.98	1252	76.9	15.4
Zhao, Zhou	Zhejiang University	22	2.92	234	36.4	27.3
Zhang, Hanwang	Nanyang Technological University	21	8.27	968	57.1	85.7
Mei, Tao	JD.com Inc	19	4.91	549	52.6	21.1

Table 7: Top 5 productive Scopus sources in Topic T.30920.

Scopus Source	Scholarly Output	FWCI	Citation Count	Top 10% Citation Percentiles (%)
Lecture Notes in Computer Science	192	1.04	381	3.6
Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition	186	6.15	5237	60.2
Proceedings of the IEEE International Conference on Computer Vision	104	3.23	2317	36.5
IJCAI International Joint Conference on Artificial Intelligence	45	2.61	365	22.2
Advances in Neural Information Processing Systems	44	6.52	863	38.6

main contributors, especially CVPR, whose top 10% citation percentile is more than 60%.

CONCLUSION

This study presented research fronts of CS, especially hot frontier topics from both theoretical and empirical perspectives, which was based on direct citation with a global mapping and created accurate topics in a fine-grained way. Leading countries and institutions were selected in terms of scholarly output and citation count. China is the most productive country and USA is the most influential country in CS research. Chinese Academy of Sciences is the leading institution in both scholarly output and citation frequency. Other government research institutions such as CNRS, RIKEN and CSIRO are leading in related hot topics. Corporate sectors are also active in hot frontier topics, such as Alphabet Inc., Ericsson AB, Facebook Inc, AstraZeneca, Lucent and Toshiba, etc. Academic, government and corporate sectors jointly lead the direction of computer science technology. Multi-party participation can quickly transform research output into innovative products and benefit for mankind.

From the case studies, we can see that research fronts are useful for researchers and policy makers to make analysis and plan. Research frontier topics can provide researchers with a clear picture of their overall research performance and insight into the momentum of particular topics. Research managers can evaluate the relationship between research direction and research fronts. They can also make comparative assessments of competing institutions. Previous studies have found that topic prominence value has a strong correlation with funding's, which is useful for stakeholders and their needs related to the portfolio planning.[16] Based on analysis of the most productive countries, institutions and authors, researchers and policy makers can look for collaborations with these authors or institutions.

Nevertheless, this paper only utilizes the data in Scopus. In order to identify research fronts more comprehensively and objectively, multi-source data should be applied, such as patent data, science and technology planning texts and fund project data. Meanwhile, prominence doesn't equal to importance, we might have overlooked some low prominences but still important topics. Future research will add more data sources and incorporate more indicators in order to obtain research frontiers more scientifically.

ACKNOWLEDGEMENT

This research was supported by the National Key R&D Program of China (2019YFB2101802), the National Natural Science Foundation of China (No. 61773324) and Teaching Reform Project of Southwest Jiaotong University (No. 20201029). We thank to Elsevier Research Intelligence Customer consultant Chen Ju, for her support and advice.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Price DJDS. Networks of scientific papers. Science. 1965;510-5.
2. Garfield E. Research fronts. Current Contents. 1994;10(41):3-7.
3. Chen C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. Journal of the American Society for Information Science and Technology. 2006;57(3):359-77.

4. Upham SP, Small H. Emerging research fronts in science and technology: Patterns of new knowledge development. Scientometrics. 2010;83(1):15-38.
5. Marrone M. Application of entity linking to identify research fronts and trends. Scientometrics. 2020;122(1):357-79.
6. Small H, Griffith BC. The structure of scientific literatures I: Identifying and graphing specialties. Science Studies. 1974;4(1):17-40.
7. Morris SA, Yen G, Wu Z, Asnake B. Time line visualization of research fronts. Journal of the American Society for Information Science and Technology. 2003;54(5):413-22.
8. Shibata N, Kajikawa Y, Takeda Y, Matsushima K. Comparative study on methods of detecting research fronts using different types of citation. Journal of the American Society for information Science and Technology. 2009;60(3):571-80.
9. Tattershall E, Nenadic G, Stevens RD. Detecting bursty terms in computer science research. Scientometrics. 2020;122(1):681-99.
10. Wu Y, Venkatramanan S, Chiu DM. Research Collaboration and Topic Trends in Computer Science: An Analysis Based on UCP Authors. Proceedings of the 24th International Conference on World Wide Web. Florence, USA: ACM; 2015;1045-50.
11. Hoonlor A, Szymanski BK, Zaki MJ. Trends in computer science research. Communications of the ACM. 2013;56(10):74-83.
12. Uddin A, Singh VK, Pinto D, Olmos I. Scientometric mapping of computer science research in Mexico. Scientometrics. 2015;105(1):97-114.
13. Uddin A, Singh VK. Mapping the Computer Science Research in SAARC Countries. IETE Technical Review. 2014;31(4):287-96.
14. Singh VK, Uddin A, Pinto D. Computer science research: The top 100 institutions in India and in the world. Scientometrics. 2015;104(2):529-53.
15. Shibata N, Kajikawa Y, Takeda Y, Matsushima K. Comparative study on methods of detecting research fronts using different types of citation. Journal of the American Society for Information Science and Technology. 2009;60(3):571-80.
16. Klavans R, Boyack KW. Research portfolio analysis and topic prominence. Journal of Informetrics. 2017;11(4):1158-74.
17. Small H, Boyack KW, Klavans R. Identifying emerging topics in science and technology. Research Policy. 2014;43(8):1450-67.
18. Cui Y, Wang S, Gao X, Yang H, Cao X. Detecting and Characterizing Research Fronts Topics Based on Global-Micro Model (in Chinese). Library and Information Service. 2018;62(15):75-82.
19. Zhang C, Xiao L, He F, Tang Y. The Study of Methods and Practices of Identifying Discipline Research Fronts-Takng Peking University Library as an Example (in Chinese). Library and Information Service. 2020;64(16):36-44.
20. Liu B. Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies. 2012;5(1):1-167.
21. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems. Lake Tahoe, USA. 2013;3111-9.
22. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781. 2013.
23. Pennington J, Socher R, Manning CD. Glove: Global vectors for word representation. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). Doha: Association for Computational Linguistics; 2014;1532-43.
24. Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, *et al*. Deep contextualized word representations. arXiv preprint arXiv:1802.05365. 2018.
25. Devlin J, Chang M, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. 2018.
26. Johnson AE, Pollard TJ, Shen L, Li-Wei HL, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG. MIMIC-III, A freely accessible critical care database. Scientific Data. 2016;3(1):1-9.
27. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. Proceedings of Advances in neural information processing systems. Montreal, Canada; 2014;3104-12.
28. Vinyals O, Toshev A, Bengio S, Erhan D. Show and tell: A neural image caption generator. Proceedings of the IEEE conference on computer vision and pattern recognition. Boston, USA: IEEE; 2015;3156-64.