Temporal Evolution, Research Themes, and Emerging Trends in Case-Based Reasoning Literature

Gu Dongxiao¹ Liu Bo¹ Bichindaritz Isabelle² Liang Changyong^{1,3,*}

¹gudongxiao@hfut.edu.cn (Gu), 592619476@qq.com (Liu) School of Management, Hefei University of Technology, Hefei (China)

²*ibichind@oswego.edu (Bichindaritz)* Computer Science Department, SUNY Oswego, New York (USA)

^{3,*} mikehfut0551@163.com (Liang) National Joint Engineering Research Center for Intelligent Decision and Information Systems, Hefei (China)

**Corresponding author.*

Abstract

Case-based reasoning, a methodology of artificial intelligence, is applicable to various fields, such as fault diagnosis, medical and health decision support, engineering aided design, and risk pre-alert, etc. In recently years, this area has attracted contributions from many researchers, but little work has described the overall development of case-based reasoning research through informetrics or literature visualization. To analyze the temporal evolution, research themes, and emerging trends in case-based reasoning, in this paper, we completed an informetrics analysis based on 4460 articles about case-based reasoning published from 2000 to 2015 in SCI-E, SSCI, CPCI-S and CPCI-SSH, the sub-databases of Web of Science, using visual knowledge maps and informetrics methods. This paper summarizes conclusions on the temporal evolution, research themes, and emerging trends for case-based reasoning. The results will help researchers rapidly grip the overall development and future directions of case-based reasoning.

Keywords

Case-Based Reasoning, Informetric Analysis, Knowledge Map, Co-citation Analysis, Emerging Trends

Introduction

Case-based reasoning (CBR), an important methodology from the artificial intelligence (AI) field, is a problem-solving process based on the reuse of solutions from similar previously solved cases. It is formalized from a reasoning standpoint in a four-step process – retrieve, reuse, revise and retain (R, et al., 2005). Modeling human reasoning, CBR involves reusing previous experiences, called historical cases, to solve a current problem: Retrieve similar cases from memory stored cases that are relevant to the target problem; Adapt the retrieved solutions to fit the new situation; Test the adaption in the real world and, if necessary, revise; Store the resulting experience as a new case. CBR, based on the reuse of cases as a criterion for evaluating the newest solutions and as a basis for prediction of future errors, is highly applicable in many situations (Marir, 1994).

A case-based reasoning system is a decision support system designed with case-based reasoning as its core principle, and its major purpose is to provide solutions and support decision making, CBR systems greatly improve efficiency of the problem-solving process. Early in 1983, Janet Kolodner (1983) constructed CYRUS, the first CBR system, based on the dynamic memory model, which was firstly conceived by Roger Schank (1982) in his theory of the dynamic memory. After that pioneering work, in time other CBR systems were built, such as MEDIATOR (Simpson, & Lee, 1985), CHEF (Hammond, 1986), PERSUADER (Sycara, 1987), CASEY (Koton, 2005), JULIA (Hinrichs, 1992), and so on. Following, researchers at many institutions have modeled other CBR systems derived from this work. In

the late 1980s and the early 1990s, CBR concepts impacted numerous fields such as computer science (Ros, Arcos, Mantaras, & Veloso, 2009), medicine (Gu, Liang, & Zhao, 2017), engineering (Gu, Liang, Bichindaritz, Zuo, & Wang, 2012), jurisprudence (Branting, 2003), environmental science (Toro, Meire, Gálvez, & Fdez-Riverola, 2013), public administration and policy (Amailef, & Lu, 2013), business administration and E-commerce (Li, & Sun, 2009). In the new millennium, CBR systems have continually integrated with other AI techniques including artificial neural networks (ANN) (Henriet et al., 2012), rule-based reasoning (RBR) (Kumar, Singh, & Sanyal, 2009), genetic algorithms (Ahn, & Kim, 2009) and others, CBR research has been focused on projects concerning knowledge discovery, case representation, reasoning and meta-reasoning models, retrieval algorithms, similarity assessment, case adaptation and management, and applications.

To date, there has been many review studies regarding the CBR methodology, the construction and application of CBR systems (Gu, Liang, & Zhao, 2017; Gu, Liang, Bichindaritz, Zuo, & Wang, 2012), and CBR integration with other methods (Kumar, Singh, & Sanyal, 2009; Ahn, & Kim, 2009; Wei, Mahmud, & Raj, 2014). Most existing relevant literature reviews have focused on CBR technology and the applications of CBR, but have hardly described the overall development of CBR as a field. For example, Watson and Marir (1994) in 1994 described the development of CBR technology in the last century. Chen and Burrell (2001) in 2001 analyzed the development of CBR applications with artificial neural networks, Greene et al. (2008) in 2008 summarized the evolution of research themes in the CBR conference literature (ICCBR, ECCBR, and EWCBR) published from 1993 to 2008, however, they did not analyze developments and emerging trends of CBR research using informetrics and visualization approaches (Kim, & Chen, 2015; Fang, 2015), In contrast to published articles, this paper has four advantages: (1)Use informetrics and visualization approaches for analysis; (2)Analyze up-to-date, the literature records in this paper published from 2000 to 2015; (3) More literature types, including proceedings papers, journal articles and reviews; (4)More comprehensive analysis, containing temporal evolution, literature co-citation, journals co-citation, research themes, and emerging trends etc.

To make up the research gap mentioned above, and also explore the overall development and future directions of CBR technology, in this study, we conducted an informetrics analysis based on published articles in CBR and investigated the implicit knowledge associated with CBR methodology. We collected 4460 articles from 4 databases including SCI-E, SSCI, CPCI-S and CPCI-SSH, which are the sub-databases of Thomson Reuter's Web of Science (WOS), and conducted literature data analysis with the HistCite, CiteSpace, Netdraw bibliometric tools, among others. After that, we summarized the temporal evolution, research themes, and emerging trends for CBR research in the 21st century. The results of this paper will be helpful for relevant researchers in CBR, and also to promote research and development in CBR.

The rest of this paper is organized as follows: In section 2, we introduce the methods and tools we used in this research, and also explain the process of data collection in detail. In section 3, we thoroughly present the results of the informetrics analysis of literature data associated with CBR using HistCite, CiteSpace, etc, including changes in published articles over time, knowledge domain visualization (co-citation analysis of literature), core journals and co-citation analysis, and research trends evolution analysis. Finally, this paper concludes with a summary of findings and some future directions in Section 4.

Methodology

Literature data

The literature data for this study were retrieved from SCI-E, SSCI, CPCI-S, and CPCI-SSH, the sub-databases of the Web of Science, on November 25, 2016. The detailed retrieval process is as follows:

Firstly, retrieve relevant literature by "topic = (case-based reason*)" in SCI-E, SSCI, CPCI-S and CPCI-SSH, the sub-databases of WOS, with time span from 2000 to 2016, which returned 4465 articles;

Secondly, eliminate 5 repeated or invalid articles by using HistCite software;

Finally, the remaining 4460 articles are pertinent and were used for informetrics analysis.

Methods and tools

There are mainly two methods, informetrics and visualization analysis, and two tools, HistCite and CiteSpace, used in this paper, Informetrics combines mathematics, statistics and philology, to quantitatively analyze knowledge carriers and quantify the implicit knowledge in literature data, focusing on the object of the analysis, which in this research consists in literature data (Mingers, & Leydesdorff, 2015). In this case, informetrics is also called bibliometrics, a branch of informetrics. In the early 1900s, the method of quantitative literature analysis was created first to allow researchers mainly to count and classify documents according to quantitative statistical methods. After the 1960s, some researchers integrated bibliometrics concepts into statistics, and since then bibliometrics has become a multidisciplinary study of statistics and metrics, is to mathematically mine the implicit knowledge from an abundant literature and to statistically infer the characteristics and prospects of a specific subject. In this paper, we mapped the evolution and development of CBR research based on informetrics and visualization analysis.

Statistical analysis with HistCite. HistCite is a web-based software enabling researchers to analyze the overall view of literature records from the Web of Science. Its major feature is statistical analysis (Garfield, 2009; Bornmann, & Marx, 2012). By using HistCite to analyze literature data, we obtained a large volume of information about the development of CBR. The overview of 4460 articles is as follows: there are 2966 keywords in 7 languages, contributed by 5895 authors from 1728 institutions, published in 1718 journals, and circulated in 89 countries/regions during the years 2000 to 2015, and the references of these papers are 49212.

Visualization analysis with CiteSpace. CiteSpace is a Java application for analyzing and visualizing developments and trends in scientific literature, it was developed by Professor Chen Chaomei (Chen, 2006) as a tool for knowledge visualization. By using CiteSpace, we can mine valuable information from literature records and visualize developments and trends of CBR research, which are also two main tasks in this paper. The main analysis contents include changes over time of published articles, co-occurrence of knowledge carriers (two levels: references and journals), and evolution analysis of research hot topics.

Results

Distribution over time

To explore publication trends of CBR research in the 21st century, we examined the temporal changes in 4460 articles, which describe the yearly number of CBR publications, reflecting some changes in research interests of global scholars, and also revealing the future development trends in CBR. In this section, the major bibliometrics indicator is the annual number of articles (referred to as Recs for records) – see Figure 1.



Fig 1. Temporal distribution of the CBR Recs (records) between 2000 and 2015

From Figure 1, we mainly divide this time series into two stages. Shown with the left arrow, the first stage from 2000 to 2006, is a period of sustained development, all of the Recs rose steadily except an abnormal high point in 2001. CBR technology, as an advanced topic, attracted larger numbers of relevant researchers at this stage, who produced large quantities of excellent academic papers, such as (Hui, & Jha, 2000; Humphreys, Mcivor, & Chan, 2003; Chow, Choy, Lee, & Lau, 2006). The second stage from 2006 to 2014, is a period of ups and downs, showing mainly three changes: a dive in 2007, a continuous rise again from 2008 to 2009, and a slow decrease from 2010 to 2013. It shows that most researchers have not contributed satisfactorily to innovation in CBR since 2009, and we also found a similar situation from the annual total global citation scores of articles. In addition, owing to the fact that part of the papers published in 2015 and 2016 may not have been included, we do not have a complete analysis of 2015 and 2016. However, it is notable that the quantity of literature rose again in 2014, and in consideration of the emerging research themes in recent years, such as big data, cloud computing, Internet of Things, and smart health, by integrating CBR technology with these emerging field, 2014 might be a turning point and may hold that it is possible for the research of CBR to rise again.

Co-occurrence of knowledge carriers

There are two carriers analyzed in this section, references co-citation and journal co-citation. The references co-citation is essential for informetrics in order to investigate the knowledge base of the CBR field and distinguish the leading edge (Zhu, & Hua, 2017), while the journal co-citation can distinguish the core journals, the marginal journals and the relative preference between them, allowing researchers to rapidly identify important documents and the key journals available for their contributions.

Knowledge base of subject development. If one document cites two other documents together, these two documents are co-cited. The more co-citations two documents receive, the more likely they are semantically related (Small, 2003). Document co-citation indicates the knowledge base of a subject or a research field, which is the stepping-stone of insights into research (Chen, 2006). Document co-citation measures a spatial data assemblage of documents using citation relationships.

Early in 1965, Price (1965) proposed the concept of the research frontier to depict dynamic nature of academic research. He claimed that the research frontier of one field was built on 40-50 documents published in recent years. While the knowledge base is a concept which

benefits to further distinguish the nature of the research frontier (Persson, 1994). If the research frontier is defined as the development of a research field, accordingly the knowledge base consist of the references of the research frontier, professor Chen Chaomei (Chen, 2006), the software developer of CiteSpace, redefined the knowledge base of the research frontier as the quote path of the references, i.e, literature co-citation. Which means the research frontier and the knowledge base can be found through co-citation analysis.



Fig 2. The network of CBR references co-citation

Figure 2 shows the CiteSpace results of references co-citation of 4460 articles, analyzing the top 50 cited references for each one-year time slice. The threshold value of the cited documents is Freq>=17 and retained more than 30 nodes. The colored lines represent the years of the first co-citations, and the rings of the nodes are co-citation frequencies. Obviously, there are two major color groups in the network of literature co-citation, a cold area and a warm area, which represent the knowledge base and research frontier respectively. In the blue area, all the cited references were published in the early CBR stage, and also belong to highly cited documents; as a result, these cited references are the knowledge base of CBR research. In the red area, overall cited references were contributed in recent years; furthermore, these documents are highly cited documents, consequently all the cited references are the research frontier of CBR research, shown in the red area. According to above results, tables 1a and 1b illustrate the top 33 highly cited documents.

Year	Author	Freq	Title	Journal/Book
1993	Kolodner J	109	Case Based Reasoning	Book
1994	Aamodt A	84	Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches	AI Commun
1995	Smyth B	32	Remembering to forget: A competence-preserving case deletion policy for case-based reasoning systems	14 th Int Joint C AI

Table 1a. Cited references of knowledge base (partial)

1996	Leake DB	73	CBR in context: The present and future	Case Based Reasoning
1997	Watson I	88	Applying Case-Based Reasoning: Techniques for Enterprise Systems	Book
1998	Lenz M	24	Case-Based Reasoning Technology: From Foundations to Applications	Case Based Reasoning
1999	Watson I	34	CBR is a methodology not a technology	Knowl-Based Syst
2001	Aha DW	30	Conversational case-based Reasoning	Appl Intell
2002	Chiu CC	31	A case-based customer classification approach for direct marketing	Expert Syst Appl
2003	Corchado JM	17	Constructing deliberative agents with case-based reasoning technology	Int J Intell Syst
2004	Pal SK	38	Foundations of Soft Case-Based	Book

Year	Author	Freq	Title	Journal
2005	De Mantaras RL	25	Retrieval, reuse, revision, and retention in case-based reasoning	Knowl Eng Rev
2006	Bichindaritz I	40	Case-based reasoning in the health sciences: What's next?	Artif Intell Med
2007	Diaz-Agudo B	18	Building CBR systems with JCOLIBRI	Sci Comput Program
2009	Ahn H	24	Global optimization of case-based reasoning for breast cytology diagnosis	Expert Syst Appl
2011	Begum S	19	Case-based reasoning systems in the health sciences: A survey of recent trends and developments	IEEE T Syst Man C Part C: A&R

There are 78 clusters among the co-cited documents in figure 2. Table 2 illustrates the top 10 clusters (the size of clusters is: more than 20). The size reflects the number of cluster nodes, and the silhouette is the contour value of the clusters, which is a measure of network homogeneity. The network homogeneity is proportional to the silhouette: the closer the silhouette is to 1, the higher is the network homogeneity. If the silhouette is not less than 0.5, the cluster is reliable; if not less than 0.7, highly reliable. The label is the result of topic extraction by the LLR (Log-Likelihood Ratio) algorithm. Mean Year is the average of the years of citation. Most of the cited references of literature records about CBR research were included in the top 21 clusters (covering 85.4% of all the documents), which means that these cluster terms can represent the core content of the knowledge base and the research frontier.

Table 2. Top 10 clusters of cited reference	es
---------------------------------------------	----

ID	Size	Silhouette	Label (LLR)	Mean Year
----	------	------------	-------------	-----------

1	28	0.497	recommender system (2219.56) case-based reasoning-perspective (1651.08)	2000
2	24	0.663	possible failure (1414.44) multiple proportion case-basing (1414.44)	2008
3	24	0.527	supporting ontological integration (921.52) proficient knowledge (921.52)	2005
4	23	0.652	pattern classification (2235.31) retrieval strategies (2235.31)	1997
5	23	0.522	dos attack (1433.1) agent-based intrusion detection mechanism soap message (1179.33)	2006
6	23	0.441	resource allocation (1117.67) learning automata (1117.67)	2004
7	22	0.545	multi-modal reasoning system (2689.72) utility problem (2460.56)	1997
8	22	0.519	adaptation methodology (2255.01) environmental emergency preparedness (2241.37)	2008
9	21	0.647	supply network (6232.05) supplier relationship management (3443.7)	1998
10	21	0.648	diabetes management (3121.39) knowledge management (3021.8)	1998

Core journals and trends. If a document in one journal cites documents in two other journals, these two journals are co-cited (Small, 2003). Journal co-citation indicates many factors, such as the main knowledge source for disciplinary development, specific journals for a research field, and academic spheres consisting of journal clusters, distinguishing core journals and marginal journals. Figure 3 shows the CiteSpace results for the literature data. A time slice is a span of 1 year, the threshold value of time slices is the top 100, the type of network pruning is pathfinder, and the threshold value of journals display is co-citation frequency > 150 (to provide a brief and clear layout to understand).



Fig 3. Journal co-citation network

Figure 3 illustrates highly cited journals in CBR, the colorful lines indicate the relationship of co-citation, the different colors represent co-citation year, the frequency of journals co-citation is represented by the size of nodes, and the thickness of the colorful layer indicates annual co-citation frequency, As a result, we can determine the core journals in CBR, including CASE BASED REASONING (Freq 1041), AI COMMUN (Freq 989), LECT NOTES ARTIF INT (Freq 812), EXPERT SYST APPL (Freq 750), ARTIF INTELL (Freq 527), KNOWL BASED SYST (Freq 520), ENG APPL ARTIF INTELL (Freq 283), ARTIF INTELL REV (Freq 269), DECIS SUPPORT SYST (Freq 259), ARTIF INTELL MED (Freq 255), COMMUN ACM (Freq 253), MACH LEARN (Freq 252), APPL INTELL (Freq 236), AI MAG (Freq 234), EUR J OPER RES (Freq 232), IEEE T KNOWL DATA EN (Freq 215), INFORM SCIENCES (Freq 176), IEEE EXPERT (Freq 159), and FUZZY SET SYST (Freq 153).



Fig 4a. Rising journals (partial)



Fig 4b. Declining & remaining journals (partial)

According to the colorful rings in figure 3, we analyzed the change of co-citations in the top 24 journals, and randomly selected 12 journals to draw the change curve of co-citation history, which includes 6 rising journals (shown in figure 4a) and 6 declining or remaining journals (shown in figure 4b). In recent years, the co-citation frequency of several journals has been decreasing gradually, mainly including APPL INTELL, ARTIF INTELL MED, ARTIF INTELL, CASE BASED REASONING, COMMUN ACM, ENG APPL ARTIF INTELL, IEEE EXPERT, IEEE T SYST MAN CYB, LECT NOTES ARTIF INT, and MACH LEARN, while the co-citation frequency of some journals has been increasing gradually, mainly containing AI COMMUN, ARTIF INTELL REV, DECIS SUPPORT SYST, EUR J OPER RES, EXPERT SYST APPL, FUZZY SET SYST, IEEE T KNOWL DATA EN, INFORM SCIENCES, KNOWL ENG REV, and KNOWL BASED SYST. We also found that the preference of the journals with decreasing trend is more inclined to the research of CBR method, while the journals with increasing trend prefer to receive the study about CBR application.

Research focus

The keywords associated with a document provide a summary, which can intuitively present its major research content. According to co-word analysis of the keywords of several documents in a research field, we can trace the major contents of this research field during a certain period, and also explore the potential trends for the future by tracking the changes of the keywords co-occurrence frequency over time. Co-word means two or more keywords appearing in one document together. Co-word analysis is a text-based analysis method, which counts the co-occurrence frequency of a pair of words in several documents to measure the relationship between the keywords (Wu, & Leu, 2014).

To analyze the research focus in CBR, we used co-words analysis with the CiteSpace software. The analysis process of CiteSpace includes three steps, extracting keywords, building the matrix of co-words, and drawing the network of co-occurrences (Wu, & Leu, 2014). The relevant parameters of CiteSpace are: time span from 2000 to 2015, one year per time slice, select top100 keywords, the co-occurrence frequency per time slice, the network pruning type is pathfinder. The analysis results are shown in Figure 5, and the colored bar at the top of the figure corresponds to 16 years, from 2000-2015; the colored circles represent the keywords (a.k.a. the nodes); the bigger nodes indicate the higher co-occurrence frequency; the thickness of colored layers refers to the frequency of the nodes in various years; the lines between two nodes indicate the co-occurrence relationship of two keywords in one document;

the color of lines indicates the first year of the keywords co-occurrence. The results are shown in Table 3 (frequency of the top 20 keywords) and Figure 5 (the co-occurrence network of keywords).

Keyword	Freq	Keyword	Freq
case based reasoning	500	decision support	15
system	122	classification	14
expert system	40	machine learning	14
model	35	knowledge based system	14
design	33	selection	11
neural network	27	similarity	11
knowledge management	25	framework	10
artificial intelligence	24	knowledge representation	10
retrieval	24	genetic algorithm	10
knowledge	21	information retrieval	9

Table	3.	The	top	20	keywords
			F		



Fig 5. The co-occurrence network of keywords

From Table 3 and Figure 5, we can find that the top 3 keywords are case based reasoning, system, and expert system, and the ratio of case-based reasoning is maximum, showing that the core of CBR research lies in CBR methodology and its application.

In addition, according to the sequential evolution of keywords, we summarized two trends in CBR research development from 2000 to 2015. The first one is that the research in CBR has placed more emphasis in actual applications and fulfilling real demands from society. For example, from 2000 to 2009, the research focus of CBR concentrated on the methodological layer, the typical keywords containing case-based reasoning, knowledge management (KM), information retrieval, similarity, classification and model. However, since 2010, the research focus has been to emphasize problem solving, for instance, decision support, expert system, fault diagnosis, health service, and so forth. The second trend shows increased integration with other techniques or methods. Before 2010, the research in CBR was involved in its core methodology, such as case based reasoning, retrieval, classification, selection, similarity, and

knowledge representation, which are all fundamental CBR topics. However, since 2010, CBR research has been applied more broadly, and more relevant research topics have been related to CBR integration with other techniques, including neural networks, machine learning, artificial intelligence, ontology, data mining, genetic algorithm, and others.

Conclusion

In this study, we have conducted an informetrics analysis to explore the temporal distribution and emerging trends of CBR research. This research has analyzed literature data from 4460 papers published from 2000 to 2015 and indexed by the databases SCI-E, SSCI, CPCI-S and CPCI-SSH. From CiteSpace, HistCite, and visual graph, we can summarize several results as follows:

With technical research evolution over time, until 2014 the number of published papers associated with CBR had decreased since 2006, such that in the last 5 years, the development of CBR research generally presented this tendency steadily. However, we found that CBR research still has great value in consideration of two aspects; the first consideration is the pull factor, referring to the research actual demand, for instance, health-care management, artificial intelligence systems, text-based sentiment analysis, etc; the push factor is the second consideration, which is the emergence of new technologies, which includes big data, cloud computing, the Internet of Things, etc.

In the co-occurrence analysis of knowledge carriers, on the one hand, we summarized several typical references about the knowledge base and the research frontier in CBR field. We also listed relevant details of its typical references. We identified the major research contents of these typical references through clustering; on the other hand, we analyzed the changes in the co-citation frequency of the core journals.

We summarized two trends of development in CBR research from 2000 to 2015 through the evolution of the core keywords over time. First, the research in CBR has placed more attention in creating actual applications and thus fulfilling actual demands of society. Second, more integration with other technologies or methods has been taking place.

In conclusion, we have conducted a comprehensive and systematic analysis and discussion about the development of CBR in 21st century. The informetrics analysis and visualization based on historical literature data will help scholars understand the general development, research hot topics, and potential future directions in the area of CBR. To foster CBR research, according to the analysis results in this paper and the development of emerging information techniques, we suggest two directions for future work.

First, the research of depth information integration and knowledge services for high-dimensional dynamic space-time cases. Extensive application of big data technology and general development of Internet of things made sequential cases with temporal-spatial trait present four obvious features, the explosive growth of data volume, the high-dimensional data structure, the complex data types, and the dynamic evolution of case data, and the existing system of case-based reasoning cannot meet such demands for processing large-scale data. On the one hand, the age of big data has brought great challenges for data processing in case-based reasoning, the development of case-based reasoning has to make new innovation and change to realize operation and maintenance of large-scale case base, for instance, the innovation of representation method for high-dimensional heterogeneity and time series cases, the change of organization and storage method for large-scale case base, the innovation of efficient case-retrieve model for case-knowledge quickly obtain and visualization for retrieve results based on larger-scale case base; on the other hand, the rise of Internet of Things and wearable device provides a new research direction for CBR system, the Internet of Things is all things are connected in brief, its foundation and core is still Internet, and wearable device is representative product of information technology in the era of the Internet of Things. CBR system based on big data and cloud computing not only need to processing static and historical cases, and also need to analyzing dynamic and real-time data, the Internet of Things and wearable device enable case base to real-time obtain case data, but it is a challenge for CBR system to connect with wearable device and Internet of Things at the moment.

Second, the research of theories and methods for collaborative CBR system in cloud computing environment. Now, the development in CBR system is facing many challenges, such as the storage and organization for larger-scale case base, distributed retrieval and similarity measures, and multi-agent collaborative operation and maintenance. In consideration of these issue, at first, we suggested that construct novel CBR system based on big data and cloud platform for the organization and storage of larger-scale case base and distributed retrieval and similarity measures; further, to address collaborative smart CBR systems (CS-CBRS), CS-CBRS is integrated with multi-agent systems (MAS) and is actually a novel case-based reasoning technique based on cloud computing and big data analysis. As a powerful analytical technique of big case data for knowledge discovery from multiple heterogeneous case bases located in different agencies and cities, it allows problem solving experiences to be shared among multiple institutes and has the potential to improve the overall performance of knowledge-based reasoning systems compared with traditional CBR systems. As a mechanism that enhances their individual reasoning capabilities, CS-CBRS offers a new paradigm for organizing artificial intelligence applications and may be used to solve important challenges in the area of complex heterogeneous big data from various organizations. The possible key research questions in the CS-CBRS research include problem-oriented mathematical modeling, intelligent case revision and solution generation, the visualization of analysis results, as well as data standardization, data quality assurance, data sharing mechanisms and privacy protection for big historical case data.

In conclusion, if CBR technology could successfully realize integration with emerging information techniques including big data, Internet of Things, and cloud computing, the present situation and mentioned problems about CBR research will be greatly solved, and the performance and efficiency of a CBR system will be greatly improved.

This research is the first review investigating the temporal distribution, emerging trends and new developments of CBR in the 21st century to help scholars better understand the whole development process, current status, and possible directions for future research. Owing to space limitations, we only thoroughly described partial analysis results of CBR, mainly including sequential distribution, co-citation of literature, co-occurrence of journals, and research focus. In addition, we also analyzed the spatial distribution and cooperation network of literature data about CBR, and containing three levels: countries/regions, institutions, and individual. The analysis results of spatial distribution and cooperation network will be presented in a future publication.

Acknowledgments

The data collection, analysis and interpretation of this research are partially supported by the National Natural Science Foundation of China (NSFC) under grant Nos. 71771077, 71331002, 71601061, 71301040 & 71573071 as well as China Social Science Foundation (major project program) under grant No. 12&ZD221.

References

- Ahn, H., & Kim, K. J. (2009). Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach. *Applied Soft Computing*, 9(2), 599-607.
- Amailef, K., & Lu, J. (2013). Ontology-supported case-based reasoning approach for intelligent m-government emergency response services. *Decision Support Systems*, 55(1), 79-97.
- Branting, L. K. (2003). A reduction-graph model of precedent in legal analysis. *Artificial Intelligence*, 150(1), 59-95.
- Bornmann, L., & Marx, W. (2012). HistCite analysis of papers constituting the h-index research front. *Journal of Informetrics*, 6(2), 285-288.
- Chen, C. (2006). CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the Association for Information Science and Technology*, 57(3), 359-377.
- Chen, D., & Burrell, P. (2001). Case-based reasoning system and artificial neural networks: a review. *Neural Computing & Applications*, 10(3), 264-276.
- Chow, H. K. H., Choy, K. L., Lee, W. B., & Lau, K. C. (2006). Design of a RFID case-based resource management system for warehouse operations. *Expert Systems with Applications*, 30(4), 561-576.
- Fang, Y. (2015). Visualizing the structure and the evolving of digital medicine: a scientometrics review. *Scientometrics*, 105(1), 5-21.
- Garfield, E. (2009). From the science of science to scientometrics visualizing the history of science with HistCite software. *Journal of Informetrics*, 3(3), 173-179.
- Greene, D., Freyne, J., Smyth, B., & Cunningham, P. (2008). An Analysis of Research Themes in the CBR Conference Literature. *European Conference on Advances in Case-Based Reasoning*, 5239, 18-43.
- Gu, D. X., Liang, C. Y., Bichindaritz, I., Zuo, C. R., & Wang, J. (2012). A case-based knowledge system for safety evaluation decision making of thermal power plants. *Knowledge-Based Systems*, 26, 185-195.
- Gu, D., Liang, C., & Zhao, H. (2017). A case-based reasoning system based on weighted heterogeneous value distance metric for breast cancer diagnosis. *Artificial Intelligence in Medicine*, 77, 31.
- Hammond, K. J. (1986). CHEF: A model of case-based planning. *National Conference on Artificial Intelligence*. 1, 267-271.
- Hinrichs, T. R. (1992). Problem Solving in Open Worlds. Georgia Institute of Technology.
- Hui, S. C., & Jha, G. (2000). Data mining for customer service support. *Information & Management*, 38(1), 1-13.
- Henriet, J., Leni, P. E., Laurent, R., Roxin, A., Chebel-Morello, B., & Salomon, M., et al. (2012). Adapting numerical representations of lung contours using case-based reasoning and artificial neural networks. *International Conference on Case-Based Reasoning*, 7466, 137-151.
- Humphreys, P., Mcivor, R., & Chan, F. (2003). Using case-based reasoning to evaluate supplier environmental management performance. *Expert Systems with Applications*, 25(2), 141-153.
- Kolodner, J. L. (1983). Reconstructive memory: a computer model. Cognitive Science, 7(4), 281-328.
- Koton, P. A. (2005). Using experience in learning and problem solving. *Massachusetts Institute of Technology*.
- Kim, M. C., & Chen, C. (2015). A scientometric review of emerging trends and new developments in recommendation systems. *Scientometrics*, 104(1), 239-263.
- Kumar, K. A., Singh, Y., & Sanyal, S. (2009). Hybrid approach using case-based reasoning and rule-based reasoning for domain independent clinical decision support in ICU. *Expert Systems with Applications*, 36(1), 65-71.
- Li, H., & Sun, J. (2009). Gaussian case-based reasoning for business failure prediction with empirical data in china. *Information Sciences*, 179(1), 89-108.
- Marir, I. W. F. (1994). Case-based reasoning: a review. *Knowledge Engineering Review*, 9(4), 327-354.
- Mingers, J., & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European Journal of Operational Research*, 246(1), 1-19.
- Price, D. J. (1965). Networks of scientific papers. Science, 149(3683), 510-515.

- Persson, O. (1994). The intellectual base and research fronts of JASIS 1986-1990. *Journal of the Association for Information Science & Technology*, 45(1), 31-38.
- Ros, R., Arcos, J. L., Mantaras, R. L. D., & Veloso, M. (2009). A case-based approach for coordinated action selection in robot soccer. *Artificial Intelligence*, 173(9), 1014-1039.
- R, L. D. M., D, M., D, B., D, L., B, S., & S, C., et al. (2005). Retrieval, reuse, revision and retention in case-based reasoning. *Knowledge Engineering Review*, 20(3), 215-240.
- Schank, R. C. (1982). Dynamic Memory: A Theory of Reminding and Learning in Computers and People. Cambridge University Press.
- Sycara, E. P. (1987). Resolving adversarial conflicts: an approach integrating case-based and analytical methods. *Nuovi Annali Digiene E Microbiologia*, 5(1).
- Small, H. (2003). Paradigms, citations, and maps of science: a personal history. *Journal of the Association for Information Science & Technology*, 54(5), 394–399.
- Simpson, & Lee, R. (1985). A computer model of case-based reasoning in problem solving : an investigation in the domain of dispute mediation. *Georgia Institute of Technology*, 56, 813–824.
- Toro, C. H. F., Meire, S. G., Gálvez, J. F., & Fdez-Riverola, F. (2013). A hybrid artificial intelligence model for river flow forecasting. *Applied Soft Computing*, 13(8), 3449-3458.
- Wu, C. C., & Leu, H. J. (2014). Examining the trends of technological development in hydrogen energy using patent co-word map analysis. *International Journal of Hydrogen Energy*, 39(33), 19262-19269.
- Wei, L. Y., Mahmud, R., & Raj, R. G. (2014). An application of case-based reasoning with machine learning for forensic autopsy. *Expert Systems with Applications*, 41(7), 3497-3505.
- Zhu, J., & Hua, W. (2017). Visualizing the knowledge domain of sustainable development research between 1987 and 2015: a bibliometric analysis. *Scientometrics*, 110(2), 1-22.