Expert Systems with Applications 40 (2013) 3160–3173

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Knowledge management vs. data mining: Research trend, forecast and citation approach

Hsu-Hao Tsai *

Department of Commerce Technology and Management, Chihlee Institute of Technology, No. 313, Sec. 1, Wunhua Rd., Banciao District, New Taipei City 22050, Taiwan, ROC

ARTICLE INFO

Keywords:
Knowledge management
Data mining
Research trend analysis
Bibliometric methodology

ABSTRACT

Knowledge management (KM) and data mining (DM) have become more important today, however, there are few comprehensive researches and categorization schemes to discuss the characteristics for both of them. Using a bibliometric approach, this paper analyzes KM and DM research trends, forecasts and citations from 1989 to 2009 by locating headings "knowledge management" and "data mining" in topics in the SSCI database. The bibliometric analytical technique was used to examine these two topics in SSCI journals from 1989 to 2009, we found 1393 articles with KM and 1181 articles with DM. This paper implemented and classified KM and DM articles using the following eight categories—publication year, citation, country/territory, document type, institute name, language, source title and subject area—for different distribution status in order to explore the differences and how KM and DM technologies have developed in this period and to analyze KM and DM technology tendencies under the above result. Also, the paper performs the K–S test to check whether the distribution of author article production follows Lotka's law. The research findings can be extended to investigate author productivity by analyzing variables such as chronological and academic age, number and frequency of previous publications, access to research grants, job status, etc. In such a way characteristics of high, medium and low publishing activity of authors can be identified. Besides, these findings will also help to judge scientific research trends and understand the scale of development of research in KM and DM through comparing the increases of the article author. Based on the above information, governments and enterprises may infer collective tendencies and demands for scientific researcher in KM and DM to formulate appropriate training strategies and policies in the future. This analysis provides a roadmap for future research, abstracts technology trend information and facilitates knowledge accumulations, therefore the future research can concentrated in core categories. This implies that the phenomenon "success breeds success" is more common in higher quality publications.

1. Introduction

1.1. Knowledge management

Knowledge management (KM) does not carry its name accidentally because management normally means that 'something' has to be managed (Wiig, Hoog, & Spex, 1997). Since Polanyi's discussion of the distinction between explicit and tacit knowledge (Polanyi, 1966), researchers were developed a set of management definitions, concepts, activities, stages, circulations, and procedures, all directed towards dealing with objects in order to describe the framework of KM as the KM methodology. Different KM working definitions, paradigms, frameworks, concepts, objects, propositions, perspectives, measurements, impacts, have been described for investigating the question of: What is KM? What are its methods and techniques? What is its value? And what are its functions for supporting individual and organizations in managing their knowledge (Drew, 1999; Heijst, Spek, & Kruizinga, 1997; Hendriks & Vriens, 1999; Johannessen, Olsen, & Olaisen, 1999; Liao, 2002; Liebowitz, 2001; Liebowitz & Wright, 1999; Nonaka, Umemoto, & Senoo, 1996; Rubenstein-Montano et al., 2001; Wiig, 1997; Wiig et al., 1997; Wilkins, Wegen, & Hoog, 1997).

For example, the concept of 'the knowledge-creating company' is a management paradigm for the emerging 'knowledge society', and information technology can help implement this concept (Nonaka et al., 1996). Some articles have investigated issues concerning the definition and measurement of knowledge assets and intellectual capital (Liebowitz & Wright, 1999; Wilkins et al., 1997). A conceptual framework presents KM as consisting of a repertoire of methods, techniques, and tools with four activities performed sequentially (Wiig et al., 1997). These are also combined with another extension of KM working definitions and its historical development (Wiig, 1997). From the organizational perspective, corporate memories can act as a tool for KM on three types of learning in organizations: individual learning, learning through

* Tel.: +886 2 22576167; fax: +886 2 22521777.
E-mail address: simon705@mail.chihlee.edu.tw

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.12.029

direct communication, and learning using a knowledge repository (Heijst et al., 1997). Another example is innovation theory based on organizational vision and KM, which facilitates development-integration and application of knowledge (Johannessen et al., 1999). For strategy, Drew explores how managers might build KM into the strategy process of their firms with a knowledge perspective and established strategy tools (Drew, 1999). Furthermore, a systems thinking framework for KM has been developed, providing suggestions for what a general KM framework should include (Rubenstein-Montano et al., 2001). Also, the emergence and future of KM, and its link to artificial intelligence been discussed (Liebowitz, 2001). Knowledge inertia (KI), means stemming from the use of routine problem solving procedures, stagnant knowledge sources, and following past experience or knowledge. It may enable or inhibit an organization's or an individual's ability on problem solving (Liao, 2002). On the other hand, the organizational impact of KM and its limits on knowledge-based systems are discussed in order to address the issue of how knowledge engineering relates to a perspective of KM (Hendriks & Vriens, 1999). These methodologies offer technological frameworks with qualitative research methods and explore their content by broadening the research horizon with different perspectives on KM research issues.

1.2. Data mining

Data mining (DM) is an interdisciplinary field that combines artificial intelligence, database management, data visualization, machine learning, mathematic algorithms, and statistics. DM, also known as knowledge discovery in databases (KDD) (Chen, Han, & Yu, 1996; Fayyad, Piatetsky-Shapiro, & Smyth, 1996a), is a rapidly emerging field. This technology provides different methodologies for decision-making, problem solving, analysis, planning, diagnosis, detection, integration, prevention, learning, and innovation.

This technology is motivated by the need of new techniques to help analyze, understand or even visualize the huge amounts of stored data gathered from business and scientific applications. It is the process of discovering interesting knowledge, such as patterns, associations, changes, anomalies and significant structures from large amounts of data stored in databases, data warehouses, or other information repositories. It can be used to help companies to make better decisions to stay competitive in the marketplace. The major DM functions that are developed in commercial and research communities include summarization, association, classification, prediction and clustering. These functions can be implemented using a variety of technologies, such as database-oriented techniques, machine learning and statistical techniques (Fayyad, Piatetsky-Shapiro, & Smyth, 1996b).

DM was defined by Turban, Aronson, Liang, and Sharda (2007, p. 305) as a process that uses statistical, mathematical, artificial intelligence and machine-learning techniques to extract and identify useful information and subsequently gain knowledge from large databases. In an effort to develop new insights into practice-performance relationships, DM was used to investigate improvement programs, strategic priorities, environmental factors, manufacturing performance dimensions and their interactions (Hajirezaie, Mohammad, Husseini, Barfourosh, & Karimi, 2010). Berson, Smith, and Thearling (2000), Lejeune (2001), Ahmed (2004) and Berry and Linoff (2004) also defined DM as the process of extracting or detecting hidden patterns or information from large databases. With an enormous amount of customer data, DM technology can provide business intelligence to generate new opportunities (Bortiz & Kennedy, 1995; Fletcher & Goss, 1993; Langley and Simon, 1995; Lau, Wong, Hui, & Pun, 2003; Salchenberger, Cinar, & Lash, 1992; Su, Hsu, & Tsai, 2002; Tam & Kiang, 1992; Zhang, Hu, Patuwo, & Indro, 1999).

Recently, a number of DM applications and prototypes have been developed for a variety of domains (Brachman, Khabaza, Kloesgen, Piatetsky-Shapiro, & Simoudis, 1996) including marketing, banking, finance, manufacturing and health care. In addition, DM has also been applied to other types of data such as time-series, spatial, telecommunications, web, and multimedia data. In general, the DM process, and the DM technique and function to be applied depend very much on the application domain and the nature of the data available.

1.3. Relationship between KM and DM

Most KM and DM techniques involve learning patterns from existing data or information, and are therefore built upon the foundation of machine learning and artificial intelligence. The primary KM and DM techniques that can be used by the organizations include statistical analysis, pattern discovery and outcome prediction. A variety of non-typical data can be similarly monitored. Before the advent of DM and KM techniques, the organizations relied almost exclusively on human expertise. It was believed that these domain experts could effectively convert their collected data into usable knowledge. As the different types of data collected grew in scope, the organizations sought to find more practical methods to make sense of what they had. This led first to the employment of in-house statisticians who created better measures of performance and better decision-making criteria. One way that these measures were used was to augment the decision-making of domain experts with additional knowledge and provide them with a competitive advantage. Armed with this knowledge, it was not a far step for organizations to begin harnessing more practical methods of extracting knowledge using DM techniques. These techniques allowed organizations to begin to predict and/or forecast under specific conditions.

2. Material and methodology

2.1. Research material

Weingart (2003), Weingart (2004) pointed at the very influential role of the monopolist citation data producer ISI (Institute for Scientific Information, now Thomson Scientific) as its commercialization of these data (Adam 2002) rapidly increased the non-expert use of bibliometric analysis such as rankings. The materials used in this study were accessed from the database of the Social Science Citation Index (SSCI), obtained by subscription from the ISI, Web of Science, Philadelphia, PA, USA. In this study, we discuss the papers published in the period from 1989 to 2009 because there was no data prior to that year. The Social Sciences Citation Index is a multidisciplinary index to the journal article of the social sciences. It fully indexes over 1950 journals across 50 social sciences disciplines. It also indexes individually selected, relevant items from over 3300 of the world's leading scientific and technical journals.

2.2. Research methodology

Pritchard (1969, p. 349) defined bibliometrics as "the application of mathematics and statistical methods to books and other media of communication." Broadus (1987, p. 376) defined bibliometrics as "the quantitative study of physical published units, or of bibliographic units, or of the surrogates for either." Bibliometric techniques have been used primarily by information scientists to study the growth and distribution of the scientific article. Researchers may use bibliometric methods of evaluation to determine the influence of a single writer, for example, or to describe the relationship between two or more writers or works. Besides, properly designed and constructed (Moed & Van Leeuwen, 1995; Van Raan 1996; Van Raan 2000, chap. 15), bibliometrics can be applied as a powerful support tool to peer review. Also for interdisciplinary research fields this is certainly possible (Van Raan & Van Leeuwen 2002). One common way of conducting bibliometric research is to use the Social Science Citation Index (SSCI), the Science Citation Index (SCI) or the Arts and Humanities Citation Index (A&HCI) to trace citations.

There are some researches to analyze the trends and forecasts by using bibliometric methodology, such as e-commerce, supply chain management, customer relationship management and data mining. (Tsai & Chang, 2011; Tsai & Chi, 2011; Tsai & Chi, 2012; Tsai, 2011; Tsai, 2012).

2.2.1. Lotka's law

Lotka's law describes the frequency of publication by authors in a given field. It states that "the number (of authors) making n contributions is about $1/n^2$ of those making one; and the proportion of all contributors, that make a single contribution, is about 60%" (Lotka 1926). Lotka's law is stated by the following formula is where y is the number of authors making x contributions, the exponent n and the constant c are parameters to be estimated from a given set of author productivity data. This means that out of all the authors in a given field, about 60% will have just one publication, about 15% will have two publications ($1/2^2$ times 0.60), about 7% of authors will have three publications ($1/3^2$ times 0.60), and so on. Lotka's law, when applied to large bodies of article over a fairly long period of time, can be accurate in general, but not statistically exact. It is often used to estimate the frequency with which authors will appear in an online catalog (Potter 1988).

Lotka's law is generally used for understanding the productivity patterns of authors in a bibliography (Coille 1977; Gupta 1987; Nicholls 1989; Pao 1985; Rao 1980; Vlachy 1978). In this article, Lotka's law is chosen to perform bibliometric analysis to check the number of publications versus accumulated authors between 1989 and 2009 to perform an author productivity inspection to collect the results for research tendency in the near future. To verify the analysis, the paper implements the K–S test to evaluate whether the result matches Lotka's law.

2.2.2. Research architecture

Using a bibliometric approach, the paper analyzes KM and DM technology trends, forecasts and citations from 1989 to 2009 by locating heading "knowledge management" and "data mining" in topics in the SSCI database. The bibliometric analytical technique was used to examine these two topics in SSCI journals from 1989 to 2009, we found of 1393 articles with KM in the keywords and 1181 articles with DM. This paper surveys and classifies KM and DM articles using the following eight categories––publication year, citation, document type, country/territory, institute name, language, source title and subject area––for different distribution status in order to explore the difference and how KM and DM technologies and applications have developed in this period and to analyze KM and DM technology trend under the above result.

As a verification of its analysis, the paper implements the following steps to check whether the analysis follows Lotka's law:

(1) Collect data
(2) List author & article distribution table
(3) Calculation the value of n (slope)

According to Lotka's law, the generalized formula is $x^n y = c$, the value of n is -2. The parameter n of applied field is calculated by the least square-method using the following formula (Pao, 1985):

$$n = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - (\sum X)^2} \tag{1}$$

N is the number of pairs of data, X is the logarithm of publications (x) and Y is the logarithm of authors (y).

The least-square method is used to estimate the best value for the slope of a regression line which is the exponent n for Lotka's law (Pao, 1985). The slope is usually calculated without data points representing authors of high productivity. Since values of the slope change with different number of points for the same set of data, we have made several computations of n. The median or the mean values of n can also be identified as the best slope for the observed distribution (Pao, 1985). Different values of n produce different values of the constant c.

(4) Calculation the value of c

According to Lotka's law, the generalized formula is $x^n y = c$, the value of c is 0.6079. The parameter c of applied field is calculated using the following formula (Pao, 1985):

$$c = \frac{1}{\sum_1^{p-1} \frac{1}{x^n} + \frac{1}{(n-1)p^{n-1}} + \frac{1}{2p^n} + \frac{n}{24(p-1)^{n+1}}} \tag{2}$$

p is the 20, n is the value obtained in (3) Calculation the value of n, and x is the number of publications.

(5) Utilizing the K–S (Kolmogorov–Smirnov, K–S) test to evaluate whether the analysis matches Lotka's law

Pao (1985) suggests the K–S test, a goodness-of-fit statistical test to assert that the observed author productivity distribution is not significantly different from a theoretical distribution. The hypothesis concerns a comparison between observed and expected frequencies. The test allows the determination of the associated probability that the observed maximum deviation occurs within the limits of chance. The maximum deviation between the cumulative proportions of the observed and theoretical frequency is determined by the following formula (Pao, 1985):

$$D = \mathrm{Max}|\mathrm{Fo}(x) - \mathrm{Sn}(x)| \tag{3}$$

Fo(x) = theoretical cumulative frequency

Sn(x) = observed cumulative frequency

The test is performed at the 0.01 level of significance. When sample size is greater than 35, the critical value of significance is calculated by the following formula (Pao, 1985):

$$\text{The critical value at the 0.01 level of significance} = \frac{1.63}{\sqrt{\sum y}} \tag{4}$$

$\sum y$ = the total population under study

If the maximum deviation falls within the critical value the null hypothesis that the data set conforms to Lotka's law can be accepted at a certain level of significance. But if it exceeds the critical value the null hypothesis must be rejected at a certain level of significance and concluded that the observed distribution is significantly different from the theoretical distribution.

The analysis provides a roadmap for future researches, abstracts technology trend information and facilitates knowledge accumulation, therefore the future research can concentrated in core categories. This implies that the phenomenon "success breeds success" is more common in higher quality publications.

3. Results

3.1. Distribution by publication year

As Fig. 1 shows, the article production on both KM and DM has been rising since 1995. The article distribution can be divided into three segments to show the trend of development: (1) from 1989 to 1994, (2) from 1995 to 2001 and from 2002 to 2009 for KM

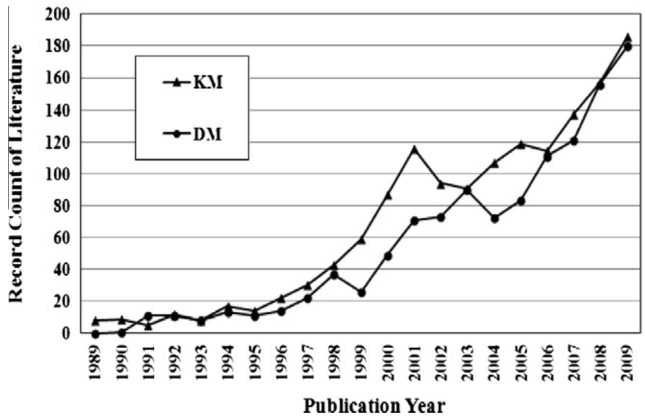

Fig. 1. The tendency chart of publication growth of KM and DM (Source: SSCI database).

research domain and (1) from 1989 to 1998, (2) from 1999 to 2003 and from 2004 to 2009 for DM research domain. From 1989 to 1995, KM and DM did not draw many researchers' attention. After 1995, the publication productivity per annum steadily increased, was followed by fast growth between 1996 and 2003, and very sharp growth in 2006, and rapidly peaked in 2009. In the periods, the amount of article on KM is always larger than on DM. The status implicates that KM has great potential to grow in the future.

3.2. Distribution by citation

From Fig. 2, we can see that the citation distribution of KM and DM is not easy to recognize between 1989 and 1998, followed by a dramatic growth and rapidly peaked in 2009 of KM. The result indicates that KM will become more popular than DM in the future.

3.3. Distribution by country/territory

It is notable that the same three countries ranked in the top three for both KM and DM from 1989 to 2009. Table 1 shows the US at the top with 461 (33.09%) in KM and 551 (46.66%) in DM, following by England, with 226 (16.220%) and 108 (9.14%), respectively. Canada ranks third with 81 (7.07%) in KM and Taiwan ranks third with 104 (8.81%) in DM. Behind them, Australia, the PRC and Germany are also major academic providers in the field of KM and DM.

In Figs. 3 and 4, we can find the article distribution of the top five countries/territories in each year for KM and DM. The US leads both fields and is followed by England. The result indicates that the US is still the main country/territory in both KM and DM research domain.

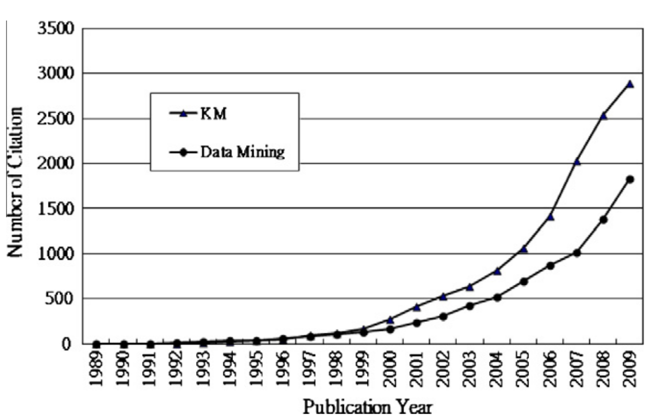

Fig. 2. Citation status in each year of KM and DM (Source: SSCI database).

3.4. Distribution by institution name

Table 2 is easy to summarize: Harvard University, the University of Illinois, the University of Texas and the University of North Carolina are the top scholarly affiliations both in KM and DM research domain. The distribution of institutions shows that the US is still the most productive country in the world in KM and DM.

Regarding the relationship between article production and citations in KM, there are only fifteen articles from Harvard University, thirteen articles from University of Illinois, ten articles from University of Texas and nine articles from University of N. Carolina in KM, but their citations are 704 times, 549 times, 618 times and 409 times in the domain. The others almost follow the article production ranking accordingly (Fig. 5). On the other side, there are only nine articles from Yale University in DM, but it has the largest amount of citations (717 times) in the domain. The others in DM follow their rank accordingly (Fig. 6).

3.5. Distribution by document type

In Table 3, the distribution of document types from 1989 to 2009 indicates that the most popular publication document type is "Article" (894 articles, 64.18% in KM and 936 articles, 79.25% in DM). The result demonstrates that the article is the major tendency of document type in KM and DM research domain.

3.6. Distribution by language

In Table 4, the majority language for KM and DM researchers is English, with 1327 articles (95.26%) in KM and 1149 articles (97.29%) in DM. Clearly, English is still the main trend in both KM and DM research domain.

3.7. Distribution by subject area

Table 5 offers critical information for future research tendencies in KM and DM, allowing researchers a better understanding of the distribution of the top 25 subject areas in future research. The top three subject areas for KM research domains are information science & library science (260 articles, 22.02%), followed by computer science & information system (251 articles, 21.25%) and operations research & management science (168 articles, 14.23%). Besides, this paper's analysis suggests that there are other important research disciplines for KM article production such as management, computer science & artificial intelligence, economics, computer science & interdisciplinary applications, public environmental & occupational health and engineering and electrical & electronic.

On the other hand, the top three DM research domains are information science & library science (260 articles, 22.02%), followed by computer science & information system (251 articles, 21.25%) and operations research & management science (168 articles, 14.23%). Analysis reveals that there are many additional research domains for DM article production, such as management, computer science & artificial intelligence, economics, computer science & interdisciplinary applications, public environmental & occupational health and engineering and electrical & electronic.

As Fig. 7 illustrates, KM citations follow article production ranking in the top 25 subjects, except for business (24.70 average citations per article). From Fig. 8, we can find that DM citations follow article production ranking in the top 25 subjects, except for statistics & probability (57.48 average citations per article), social sciences & mathematical methods (32.09 average citations per article), economics (12.26 average citations per article), computer

Table 1
Distribution of top 25 countries/territories from 1989 to 2009. *Source:* SSCI database.

Ranking	Knowledge management			Data mining		
	Country/territory	NP	% of 1393 (%)	Country/territory	NP	% of 1181 (%)
1	The US	461	33.09	The US	551	46.66
2	England	226	16.22	England	108	9.14
3	Canada	82	5.89	Taiwan	104	8.81
4	Taiwan	76	5.46	Canada	67	5.67
5	Australia	55	3.95	The PRC	54	4.57
6	The PRC	47	3.37	Australia	47	3.98
7	Germany	45	3.23	Germany	32	2.71
8	Netherlands	37	2.66	South Korea	32	2.71
9	Spain	35	2.51	Spain	27	2.29
10	Sweden	31	2.23	Netherlands	21	1.78
11	France	28	2.01	Belgium	20	1.69
12	New Zealand	22	1.58	France	20	1.69
13	Italy	21	1.51	Japan	18	1.52
14	South Africa	21	1.51	Italy	17	1.44
15	South Korea	21	1.51	Brazil	13	1.10
16	Scotland	20	1.44	Scotland	13	1.10
17	Singapore	18	1.29	South Africa	13	1.10
18	Norway	17	1.22	Sweden	12	1.02
19	Greece	16	1.15	Turkey	12	1.02
20	Brazil	15	1.08	India	11	0.93
21	Denmark	14	1.01	Slovenia	11	0.93
22	Finland	13	0.93	Austria	10	0.85
23	India	12	0.86	Finland	10	0.85
24	Japan	12	0.86	Singapore	10	0.85
25	Switzerland	12	0.86	Wales	10	0.85

NP = number of publication.

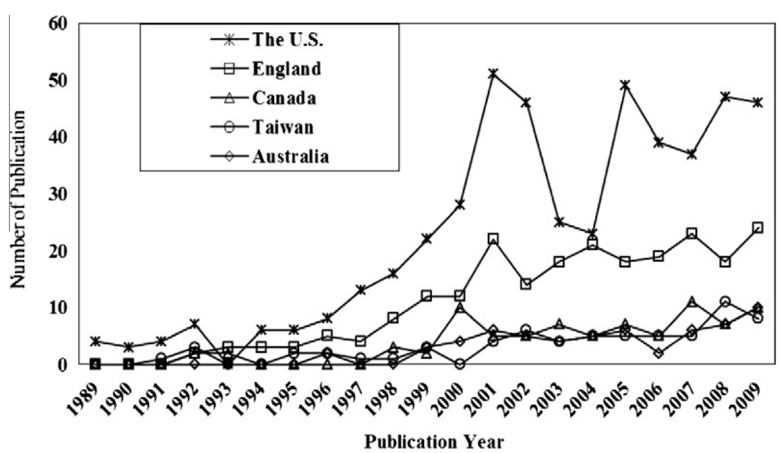

Fig. 3. Publication distribution of top five countries/territories in KM (Source: SSCI database).

science & artificial intelligence (10.79 average citations per article), engineering, electrical & electronic (9.05average citations per article) and computer science & information systems (7.73 average citations per article).

Analysis of the top five subjects––management, information science & library science, operations research & management science and computer science & information systems––shows that these subjects are repeated in KM and DM research, indicating that these subjects will become the most important category for KM and DM researchers.

3.8. Distribution by source title

Table 6 highlights information on trends for KM and DM, allowing researchers to closely approach the distribution of the top 25 source titles in future research. The top three KM research journals are *Expert Systems with Applications* (69 articles, 5.84%), followed by *Journal of the American Medical Informatics Association* (35 articles, 2.96%) and *Journal of Operation Research Society* (26 articles, 2.20%). In addition, there are a significant number of research sources for KM article production such as *Journal of the American Society for Information and Technology*, *Information Processing & Management*, *International Journal of Geographical Information Science*, *Journal of Information Science*, *Online Information Review*, *Information & Management, Decision Support Systems* and *Resources Policy*.

In the meantime, the top three DM research journals are *Expert Systems with Applications* (69 articles, 5.84%), followed by *Journal of the American Medical Informatics Association* (35 articles, 2.96%) and *Journal of Operation Research Society* (26 articles, 2.20%). Moreover, it also find out that there are a lot of research sources for DM article production such as *Journal of the American Society for Information and Technology*, *Information Processing & Management*, *International Journal of Geographical Information Science*, *Journal of Information Science*, *Online Information Review*, *Information & Management*, *Decision Support Systems* and *Resources Policy*.

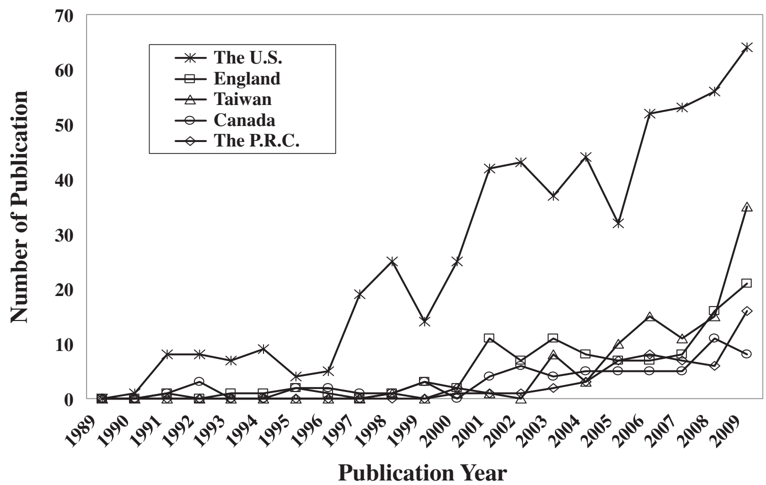

Fig. 4. Publication distribution of top five countries/territories in DM (Source: SSCI database).

Table 2
Distribution of the top 25 institutions for KM and DM from 1989 to 2009. *Source:* SSCI database.

Ranking	Knowledge management				Data mining			
	Institution name	NP	% of 1393 (%)	Country	Institution name	NP	% of 1181 (%)	Country
1	University of Warwick	16	1.15	The UK	NIOSH	17	1.44	The US
2	Harvard University	15	1.08	The US	Pennsylvania State University	17	1.44	The US
3	Rutgers State University	15	1.08	The US	University of Wisconsin	17	1.44	The US
4	University of Toronto	14	1.01	Canada	University of Illinois	13	1.10	The US
5	University of Illinois	13	0.93	The US	Columbia University	12	1.02	The US
6	University of Loughborough	12	0.86	The UK	National Central University	12	1.02	Taiwan
7	University of Sheffield	12	0.86	The UK	University of Pennsylvania	12	1.02	The US
8	Hong Kong Polytech University	11	0.79	The PRC	National Chiao Tung University	11	0.93	Taiwan
9	University of Manchester	11	0.79	The UK	Purdue University	11	0.93	The US
10	Napier University	10	0.72	The UK	Monash University	10	0.85	Australia
11	National Cheng Kung University	10	0.72	Taiwan	University of Texas	10	0.85	The US
12	University of Leeds	10	0.72	The UK	Duke University	9	0.76	The US
13	University of Pretoria	10	0.72	South Africa	Tamkang University	9	0.76	Taiwan
14	University of Texas	10	0.72	The US	University of N. Carolina	9	0.76	The US
15	University of Washington	10	0.72	The US	University of Western Ontario	9	0.76	Canada
16	City University of London	9	0.65	The UK	Yale University	9	0.76	The US
17	McGill University	9	0.65	Canada	Virginia Commonwealth University	9	0.76	The US
18	Michigan State University	9	0.65	The US	City University of Hong Kong	8	0.68	The PRC
19	NanYang Technical University	9	0.65	Singapore	Harvard University	8	0.68	The US
20	University of Minnesota	9	0.65	The US	NanYang Technology University	8	0.68	Singapore
21	University of N. Carolina	9	0.65	The US	National Sun Yat-Sen University	8	0.68	Taiwan
22	Indiana University	8	0.57	The US.	ONR	8	0.68	The US
23	Korea Advanced Institute of Science and Technology	8	0.57	South Korea	Syracuse University	8	0.68	The US
24	University of Nebraska	8	0.57	The US	University of Arizona	8	0.68	The US
25	University of Nottingham	8	0.57	The UK	University of Hong Kong	8	0.68	The PRC

NP = number of publication.

In Fig. 9, KM citations follow article production ranking in the top 25 sources, except for *International Journal of Technology Management* (3.00 average citations per article), *Decision Support Systems* (15.68 average citations per article), *Long Range Planning* (26.07 average citations per article), *Information & Management* (16.57 average citations per article) *Journal of Management Studies* (28.64 average citations per article) and *Journal of Management Information Systems* (74.18 average citations per article).

As Fig. 10 shows, we can find that DM citations article production ranking in the top 25 sources, except for *Decision Support Systems* (20.00 average citations per article), *Information & Management* (14.75 average citations per article), *Journal of the American Society for Information Science* (12.45 average citations per article), *International Journal of Geographical Information Science* (9.70 average citations per article) and *Scientometrics* (9.00 average citations per article).

An analysis of the top five journal sources—Expert Systems with Applications, Journal of Operation Research Society and Journal of the American Society for Information and Technology are repeated within KM and DM research domain—demonstrates that these sources will become the most critical journals for KM and DM researchers.

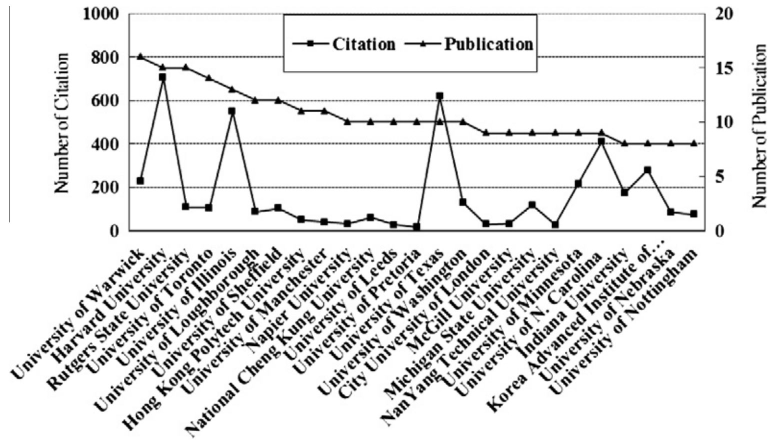

Fig. 5. Publication and citation distribution of top 25 institutions in KM (Source: SSCI database).

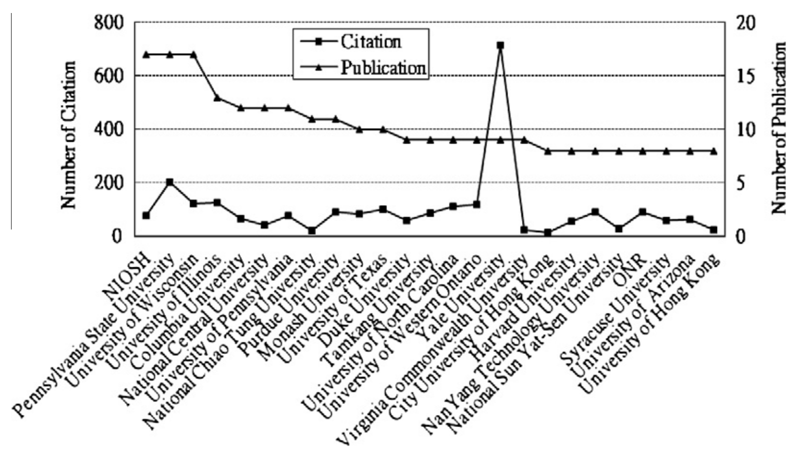

Fig. 6. Publication and citation distribution of top 25 institutions in DM (Source: SSCI database).

Table 3
Distribution of document types from 1989 to 2009. source: SSCI database.

Knowledge management			Data mining		
Document type	NP	% of 1393 (%)	Document type	NP	% of 1181 (%)
Article	894	64.18	Article	936	79.25
Book review	211	15.15	Proceedings paper	106	8.98
Proceedings paper	81	5.81	Book review	50	4.23
Editorial material	73	5.24	Review	41	3.472
Review	62	4.45	Meeting abstract	23	1.95
Meeting abstract	51	3.66	Editorial material	19	1.61
News Item	7	0.50	News item	2	0.17
Correction	5	0.36	Correction	1	0.08
Letter	3	0.22	Note	1	0.08
Note	3	0.22	Reprint	1	0.08
Bibliography	1	0.07	Software review	1	0.08
Reprint	1	0.07	Total	1181	100
Software review	1	0.07			
Total	1393	100			

NP = number of publication.

4. Discussion

The section implements the steps which are demonstrated in Section 2.2.2 to verify whether the distribution of author article production follows Lotka's law in KM and DM research domain.

4.1. Knowledge management

(1) Collect data and
(2) List author & article distribution table

Table 4
Distribution of languages from 1989 to 2009. *Source:* SSCI database.

Knowledge management			Data mining		
Language	NP	% of 1393 (%)	Language	NP	% of 1181 (%)
English	1327	95.26	English	1149	97.29
German	32	2.30	Spanish	12	1.02
Spanish	12	0.86	German	5	0.42
Portuguese	8	0.57	Slovak	4	0.34
French	5	0.36	Japanese	3	0.25
Czech	3	0.22	Czech	2	0.17
Danish	1	0.07	French	2	0.17
Norwegian	1	0.07	Portuguese	2	0.17
Russian	1	0.07	Russian	1	0.08
Slovak	1	0.07	Slovene	1	0.08
Swedish	1	0.07	Total	1181	100
Turkish	1	0.07			
Total	1393	100			

NP = number of publication.

Author quantity is calculated by the equality method from 1393 articles retrieved by the SSCI index. Altogether, 2549 authors on KM are included. See Table 7 for reference.

(3) Calculation the value of n (slope)

In Table 8, we list the number of authors and the number of publications by one author for calculation of the exponent n with topic as "knowledge management" in SSCI database. The results of the calculations in Table 8 can be brought into Eq. (1) to calculate the value of n:

$$\mathrm{n} = \frac{15(3.08) - (5.83)(9.69)}{15(4.78) - (5.83)(5.83)} \quad (5)$$

Then we can find $n = -3.194592051$.

(4) Calculation the value of c

The value of c is calculated by using Eq. (2), where $P = 20$, $x = 1,2,3,4,5,6,7,8$ and $n = 3.194592051$, then we can find $c = 0.857063311$.

With $n = -3.194592051$ and $c = 0.\ 857063311$, the Lotka's law equation of KM is:

$$f(x) = 0.857063311/x^{3.194592051} \quad (6)$$

When the result is compared to Table 7, we can see that authors with only one article account for 89.25% (100% − 10.75% = 89.25%), which almost matches the primitive c value 85.70% generated by Lotka's law. The values for n and c can be calculated by the least squares law and then brought into further analysis for Lotka's law compliance.

According to Pao (1989), the absolute value of n should be between 1.2 and 3.8, as given by the generalized Lotka's law. The result indicates that n (=3.194592051) is between 1.2 and 3.8 and is matched the reference data by observation. The detail distribution chart is shown in Fig. 11.

(5) Utilize the K–S test to evaluate whether the analysis matches Lotka's law

We use Eq. (3) to evaluate whether the analysis matches Lotka's law. From Table 9, we can find D ($D = \mathrm{Max}|\mathrm{Fo}(x) - \mathrm{Sn}(x)|) = 0.0354$. According to the K–S test, the critical value at 0.01 level of significance is calculated by Eq. (4):

$$1.63/\sqrt{2549} = 0.03228514 \quad (7)$$

The maximum deviation found is 0.0354 which exceeds the critical value of 0.03228514 at the 0.01 level of significance. Therefore, the null hypothesis must be rejected and concluded that the KM data do not fit Lotka's law (Potter 1981).

4.2. Data mining

(1) Collect data and
(2) List author & article distribution table

Table 5
Distribution of top 25 subject areas from 1989 to 2009. source: SSCI database.

Ranking	Knowledge management			Data mining		
	Subject area	NP	% of 1393 (%)	Subject area	NP	% of 1181 (%)
1	Management	459	32.95	Information science & library science	260	22.02
2	Information science & library science	366	26.27	Computer science, information systems	251	21.25
3	Computer science, information systems	270	19.38	Operations research & management science	168	14.23
4	Operations research & management science	178	12.78	Management	149	12.62
5	Business	165	11.84	Computer science, artificial intelligence	132	11.18
6	Engineering, industrial	71	5.10	Economics	112	9.48
7	Engineering, multidisciplinary	71	5.10	Computer science, interdisciplinary applications	103	8.72
8	Computer science, artificial intelligence	68	4.88	Public, environmental & occupational health	85	7.20
9	Computer science, interdisciplinary applications	55	3.95	Engineering, electrical & electronic	82	6.94
10	Economics	51	3.66	Environmental studies	68	5.76
11	Nursing	51	3.66	Business	56	4.74
12	Planning & development	40	2.87	Geography	52	4.40
13	Environmental studies	39	2.80	Medical informatics	49	4.15
14	Education & educational research	36	2.58	Environmental sciences	38	3.22
15	Social sciences, interdisciplinary	36	2.58	Social sciences, mathematical methods	35	2.96
16	Engineering, electrical & electronic	35	2.51	Ergonomics	34	2.88
17	Sociology	28	2.01	Engineering, industrial	33	2.79
18	Health care sciences & services	24	1.72	Planning & development	31	2.62
19	Psychology, applied	24	1.72	Education & educational research	30	2.54
20	Anthropology	23	1.65	Social sciences, interdisciplinary	30	2.54
21	Psychology, multidisciplinary	22	1.58	Sociology	30	2.54
22	Public, environmental & occupational health	22	1.58	Mathematics, interdisciplinary applications	26	2.20
23	Computer science, cybernetics	20	1.44	Geography, physical	24	2.03
24	Medical informatics	20	1.44	Computer science, cybernetics	23	1.95
25	Computer science, theory & methods	19	1.36	Statistics & probability	21	1.78

NP = number of publication.

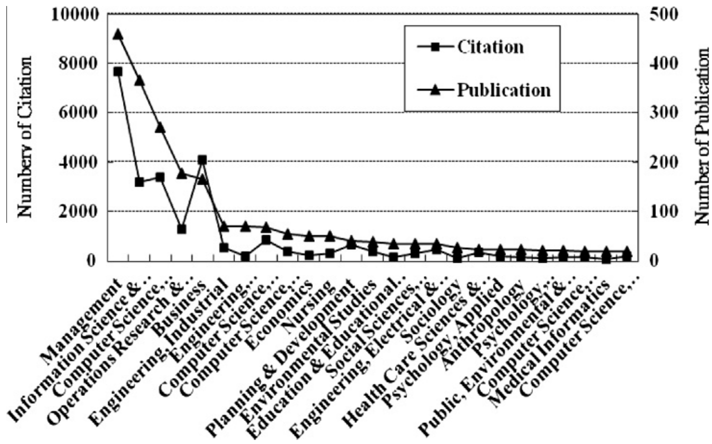

Fig. 7. Publication and citation distribution of top 25 subjects in KM (Source: SSCI database).

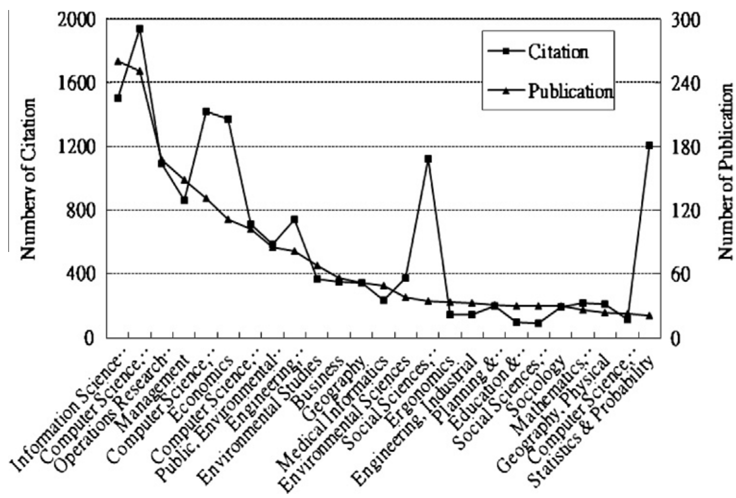

Fig. 8. Publication and citation distribution of top 25 subjects in DM (Source: SSCI database).

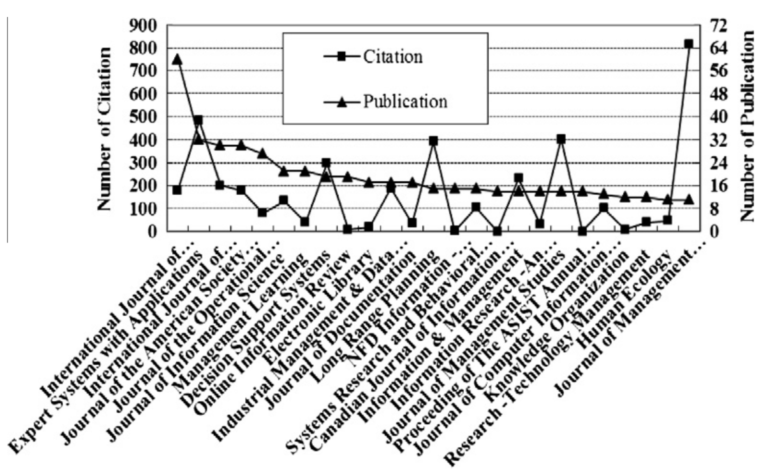

Fig. 9. Publication and citation distribution of top 25 sources in KM (Source: SSCI database).

Author quantity is calculated by the equality method from 1181 articles retrieved by the SSCI index. Altogether, 2519 authors on DM are included. See Table 10 for reference.

(3) Calculation the value of n (slope)

In Table 11, we list the number of authors and the number of publications by one author for calculation of the exponent n with topic as "data mining" in SSCI database. The results of the calculations in Table 5 can be brought into Eq. (1) to calculate the value of n:

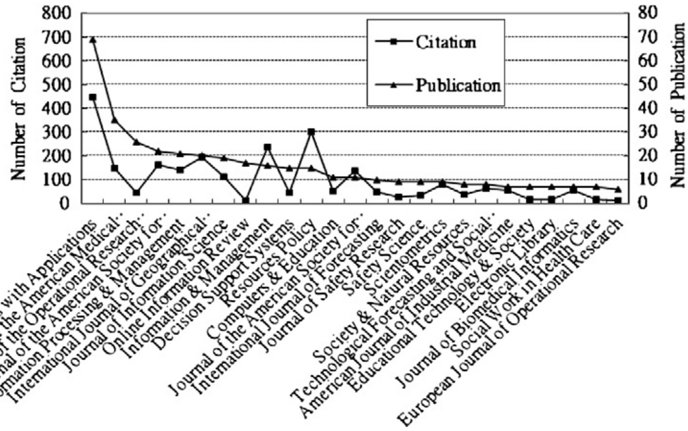

Fig. 10. Publication and citation distribution of top 25 sources in DM (Source: SSCI database).

Table 6
Distribution of top 25 source titles from 1989 to 2009. *Source:* SSCI database.

Ranking	Knowledge management			Data mining		
	Source title	NP	% of 1393 (%)	Source title	NP	% of 1181 (%)
1	*International Journal of Technology Management*	60	4.31	*Expert Systems with Applications*	69	5.84
2	*Expert Systems with Applications*	32	2.30	*Journal of the American Medical Informatics Association*	35	2.96
3	*International Journal of Information Management*	30	2.15	*Journal of the Operational Research Society*	26	2.20
4	*Journal of the American Society for Information Science and Technology*	30	2.15	*Journal of the American Society for Information Science and Technology*	22	1.86
5	*Journal of the Operational Research Society*	27	1.94	*Information Processing & Management*	21	1.78
6	*Journal of Information Science*	21	1.51	*International Journal of Geographical Information Science*	20	1.69
7	*Management Learning*	21	1.51	*Journal of Information Science*	19	1.61
8	*Decision Support Systems*	19	1.36	*Online Information Review*	17	1.44
9	*Online Information Review*	19	1.36	*Information & Management*	16	1.35
10	*Electronic library*	17	1.22	*Decision Support Systems*	15	1.27
11	*Industrial Management & Data Systems*	17	1.22	*Resources Policy*	15	1.27
12	*Journal of Documentation*	17	1.22	*Computers & Education*	11	0.93
13	*Long Range Planning*	15	1.08	*Journal of the American Society for Information Science*	11	0.93
14	*NFD Information – WISSENSCHAFT UND Praxis*	15	1.08	*International Journal of Forecasting*	10	0.85
15	*Systems Research and Behavioral Science*	15	1.08	*Journal of Safety Research*	9	0.76
16	*Canadian Journal of Information and Library Science*	14	1.01	*Safety Science*	9	0.76
17	*Information & Management*	14	1.01	*Scientometrics*	9	0.76
18	*Information Research – An International Electronic Journal*	14	1.01	*Society & Natural Resources*	8	0.68
19	*Journal of Management Studies*	14	1.01	*Technological Forecasting and Social Change*	8	0.68
20	*Proceeding of The ASIST Annual Meeting*	14	1.01	*American Journal of Industrial Medicine*	7	0.59
21	*Journal of Computer Information Systems*	13	0.93	*Educational Technology & Society*	7	0.59
22	*Knowledge Organization*	12	0.86	*Electronic Library*	7	0.59
23	*Research Technology Management*	12	0.86	*Journal of Biomedical Informatics*	7	0.59
24	*Human Ecology*	11	0.79	*Social Work in Health Care*	7	0.59
25	*Journal of Management Information Systems*	11	0.79	*European Journal of Operational Research*	6	0.51

NP = number of publication.

Table 7
Calculation of author productivity of KM.

NP	Author (s)	(NP) ∗ (author)	Accumulated record	Accumulated record (%)	Accumulated author(s)	Accumulated author(s) (%)
15	1	15	15	0.51	1	0.04
9	1	9	24	0.81	2	0.08
7	2	14	38	1.29	4	0.16
6	1	6	44	1.49	5	0.20
5	8	40	84	2.84	13	0.51
4	16	64	148	5.01	29	1.14
3	41	123	271	9.17	70	2.75
2	204	408	679	22.99	274	10.75
1	2275	2275	2954	100.00	2549	100.00

NP = number of publication.

$$n = \frac{9(3.26) - (5.56)(9.87)}{9(4.22) - (5.56)^2} \tag{8}$$

Then we can find $n = -3.629488955$.

(4) Calculation the value of c

Table 8
Calculation of the exponent n for KM.

x (NP)	y (author)	$X = \log(x)$	$Y = \log(y)$	XY	XX
15	1	1.18	0.00	0.00	1.38
9	1	0.95	0.00	0.00	0.91
7	2	0.85	0.30	0.25	0.71
6	1	0.78	0.00	0.00	0.61
5	8	0.70	0.90	0.63	0.49
4	16	0.60	1.20	0.72	0.36
3	41	0.48	1.61	0.77	0.23
2	204	0.30	2.31	0.70	0.09
1	2275	0.00	3.36	0.00	0.00
Total	2549	5.83	9.69	3.08	4.78

x = number of publication; y = author; X = logarithm of x; Y = logarithm of y.

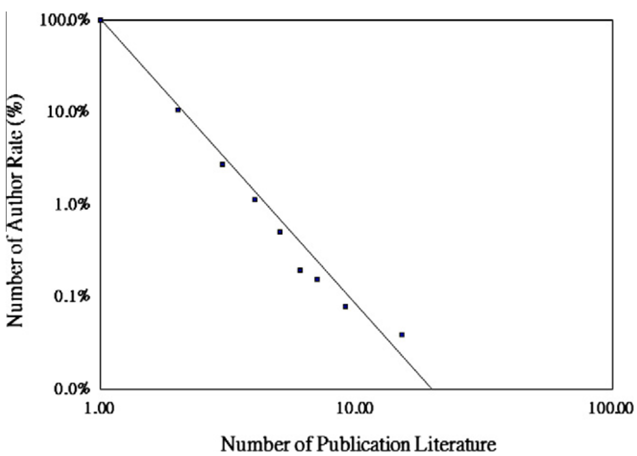

Fig. 11. Distribution of literature productivity of author on KM research aspect.

Table 9
The K–S test for KM.

NP	Author (s)	KM (observed)	Sn(x)	KM (expected)	Fo(x)	D
1	2275	0.8925	0.8925	0.8571	0.8571	0.0354
2	204	0.0800	0.9725	0.0936	0.9507	0.0219
3	41	0.0161	0.9886	0.0256	0.9763	0.0123
4	16	0.0063	0.9949	0.0102	0.9865	0.0084
5	8	0.0031	0.9980	0.0050	0.9915	0.0065
6	1	0.0004	0.9984	0.0028	0.9943	0.0041
7	2	0.0008	0.9992	0.0017	0.9961	0.0032
9	1	0.0004	0.9996	0.0011	0.9972	0.0024
15	1	0.0004	1.0000	0.0008	0.9979	0.0021

NP: number of publication; KM: author productivity of KM; Sn(x) = observed cumulative frequency; Fo(x) = theoretical cumulative frequency; D = maximum deviation.

Table 11
Calculation of the exponent n for DM.

x (NP)	y (Author)	$X = \log(x)$	$Y = \log(y)$	XY	XX
9	1	0.95	0.00	0.00	0.91
8	0	0.90	0.00	0.00	0.82
7	2	0.85	0.30	0.25	0.71
6	3	0.78	0.48	0.37	0.61
5	6	0.70	0.78	0.54	0.49
4	12	0.60	1.08	0.65	0.36
3	37	0.48	1.57	0.75	0.23
2	206	0.30	2.31	0.70	0.09
1	2252	0.00	3.35	0.00	0.00
Total	2519	5.56	9.87	3.26	4.22

x = number of publication; y = author; X = logarithm of x; Y = logarithm of y.

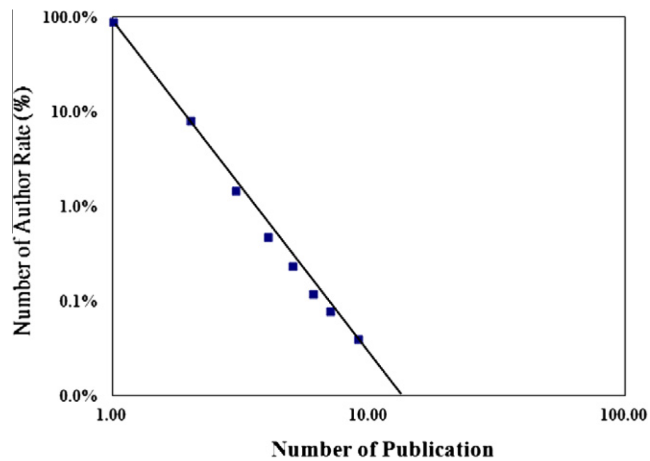

Fig. 12. Distribution of literature productivity of author on DM research aspect.

The value of c is calculated by using Eq. (2), where $P = 20$, $x = 1,2,3,4,5,6,7,8$ and $n = 3.629488955$, then we can find $c = 0.892795157$.

With $n = -3.629488955$ and $c = 0.892795157$, the Lotka's law equation of DM is:

$$f(x) = 0.892795157/x^{3.629488955} \tag{9}$$

When the result is compared to Table 10, we can see that authors with only one article account for 89.40% (100% – 10.60% = 89.40%), which almost matches the primitive c value 89.28% generated by Lotka's law. The values for n and c can be calculated by the least squares law and then brought into further analysis for Lotka's law compliance.

According to Pao (1989), the absolute value of n should be between 1.2 and 3.8, as given by the generalized Lotka's law. The result indicates that n (=3.629488955) is between 1.2 and 3.8 and is

Table 10
Calculation of author productivity of DM.

NP	Author (s)	(NP) * (author)	Accumulated record	Accumulated record (%)	Accumulated author(s)	Accumulated author(s) (%)
9	1	9	9	0.31	1	0.04
8	0	0	9	0.31	1	0.04
7	2	14	23	0.79	3	0.12
6	3	18	41	1.42	6	0.24
5	6	30	71	2.45	12	0.48
4	12	48	119	4.11	24	0.95
3	37	111	230	7.95	61	2.42
2	206	412	642	22.18	267	10.60
1	2252	2252	2894	100.00	2519	100.00

NP = number of publication.

Table 12
The K–S test for DM.

NP	Author (s)	Data mining (observed)	Sn(x)	Data mining (expected)	Fo(x)	D
1	2252	0.8940	0.8940	0.8928	0.8928	0.0012
2	206	0.0818	0.9758	0.0721	0.9649	0.0109
3	37	0.0147	0.9905	0.0166	0.9815	0.0090
4	12	0.0048	0.9952	0.0058	0.9873	0.0079
5	6	0.0024	0.9976	0.0026	0.9899	0.0077
6	3	0.0012	0.9988	0.0013	0.9913	0.0076
7	2	0.0008	0.9996	0.0008	0.9920	0.0076
8	0	0.0000	0.9996	0.0005	0.9925	0.0071
9	1	0.0004	1.0000	0.0003	0.9928	0.0072

NP = number of publication; Data mining = author productivity of data mining; Sn(x) = observed cumulative frequency; Fo(x) = theoretical cumulative frequency; D = maximum deviation.

matched the reference data by observation. The detail distribution chart is shown in Fig. 12.

(5) Utilize the K–S test to evaluate whether the analysis matches Lotka's law

We use Eq. (3) to evaluate whether the analysis matches Lotka's law. From Table 12, we can find D ($D = \mathrm{Max}|\mathrm{Fo}(x) - \mathrm{Sn}(x)|$) = 0.0109. According to the K–S test, the critical value at 0.01 level of significance is calculated by using Eq. (4):

$$1.63/\sqrt{2519} = 0.032477 \tag{10}$$

4.3. Discussion

(1) According to Lotka's methodology, the value of the exponent n for KM is estimated 3.194592051 and the constant c computed 0.857063311. Using the K–S test it is found that at the 0.01 level of significance the maximum deviation is 0.0354 which exceeds the critical value of 0.03228514. Therefore, it can be concluded that the author productivity distribution of KM does not conform to Lotka's law.
(2) Based on Lotka's methodology, the value of the exponent n for DM is estimated 3.629488955 and the constant c computed 0. 892795157. Using the K–S test it is found that at the 0.01 level of significance the maximum deviation is 0.0109 which falls within the critical value of 0.032477. Therefore, it can be concluded that the author productivity distribution of DM fits Lotka's law.
(3) The reason why KM does not match is that the number of authors who publish only one article is too large; as a result, the difference between the observed value and the expected value becomes greater than the K–S test critical value. This outcome causes the DM distribution to diverge from the slope of Lotka's law.

5. Conclusion

Using a bibliometric approach, this paper analyzes KM and DM research trends, forecasts and citations from 1989 to 2009 by locating headings "knowledge management" and "data mining" in topics in the SSCI database. The bibliometric analytical technique was used to examine these two topics in SSCI journals from 1989 to 2009, we found 1393 articles with KM and 1181 articles with DM. This paper implemented and classified KM and DM articles using the following eight categories—publication year, citation, country/territory, document type, institute name, language, source title and subject area—for different distribution status in order to explore the differences and how KM and DM technologies and applications have developed in this period and to analyze KM and DM technology tendencies under the above result. Also, the paper performs the K–S test to check whether the analysis follows Lotka's law.

The results in this paper have several important implications. Compared to DM, KM has more potential to grow up and becomes more popular. The article is the main tendency of document type in both KM and DM research. Clearly, English is still the major trend of language in both KM and DM research.

On the basis of the countries/territories, the US, England, Canada and Taiwan are the top four countries/territories in both KM and DM research. Besides, Australia, the PRC and Germany are also the major academic article providers in KM and DM.

Regarding the institutions, Harvard University, the University of Pennsylvania, the University of Texas, and the University of North Carolina are the specific scholarly affiliations in both KM and DM research. Analysis of the locations of these affiliations shows that the US is still the most productive country within the research aspect of KM and DM in the world. Regarding to the relationship between article production and citation in KM, there are only fifteen articles from Harvard University, thirteen articles from University of Illinois, ten articles from University of Texas and nine articles from University of N. Carolina in KM, but their citations are 704 times, 549 times, 618 times and 409 times in the domain. The others almost follow the article production ranking accordingly. On the other side, there are only nine articles from Yale University in DM, however their citations are the largest amount in the domain. The others in DM almost follow the article production ranking accordingly.

Judging from the subjects, the most relevant disciplines for KM and DM subject category provided by management, information science & library science, operations research & management science and computer science & information systems and will become the most important categories for KM and DM researchers. The KM citation follows the article production ranking except for business. In the meantime, we can find that DM citation follows the article production ranking except for statistics & probability, social sciences & mathematical methods, economics, computer science & artificial intelligence, engineering, electrical & electronic and computer science & information systems.

Based on the sources, the most enthusiastic supports for KM and DM scholarly publishing enterprise come from Expert Systems with Applications, Journal of Operation Research Society and Journal of the American Society for Information and Technology which are repeated during KM and DM research domain and will turn into the most critical journals for KM and DM researchers. The KM citation follows the article production ranking except for International Journal of Technology Management, Decision Support Systems, Long Range Planning, Information & Management Journal of Management Studies and Journal of Management Information Systems. On the other hand, we can find that DM citation follows the article production ranking except for Decision Support Systems, Information & Management, Journal of the American Society for Information Science, International Journal of Geographical Information Science and Scientometrics.

According to the K–S test, the result shows that the author productivity distribution predicted by Lotka holds for KM, but not for DM. The reason why DM does not fit Lotka's law is the amount of authors in DM who published one article is too large. The result causes that the difference between observed value and expected value becomes greater than the K–S test critical value. The outcome diverges DM distribution from the slope of Lotka's law.

The research findings can be extended to investigate author productivity by analyzing variables such as chronological and academic age, number and frequency of previous publications, access to research grants, job status, etc. In such a way characteristics of high, medium and low publishing activity of authors can be identified. Besides, these findings will also help to judge scientific research trends and understand the scale of development of research in KM and DM through comparing the increases of the article author. Based on the above information, governments and enterprises may infer collective tendencies and demands for scientific researcher in KM and DM to formulate the appropriate training strategies and policies in the future.

The analysis provides a roadmap for future research, abstracts technology trend information and facilitates knowledge accumulation, therefore the future research can concentrated in core categories. This implies that the phenomenon "success breeds success" is more common in higher quality publications.

5.1. Limitation of the study

The results and conclusion are limited and not intended to be exclusive. SSCI journals adopt stringent journal reviewing criteria, the articles might take two years from submission to publication. Besides, the SSCI database does not collect conference proceedings in education. Therefore, findings in this study may not reflect the most recent research trends.

Research on KM and DM has been carried out since the 1960s and 1970s, and even before that date. However, this study used only one search term each ("knowledge management" and "data mining") to analyze KM and DM publications from 1989 to 2009 collected in the SSCI databases at that time. Future studies with greater resources, using more search terms, are needed to expand these findings. (See Table 12.)

References

Adam, D. (2002). The counting house. *Nature, 415*, 726–729.

Ahmed, S. R. (2004). Effectiveness of neural network types for prediction of business failure. *Information Technology: Coding and Computing, 2*, 455–459.

Berry, M. J. A., & Linoff, G. S. (2004). *Data mining techniques second edition - for marketing, sales, and customer relationship management*. New York: Wiley.

Berson, A., Smith, S., & Thearling, K. (2000). *Building data mining applications for CRM*. New York: McGraw-Hill.

Bortiz, J. E., & Kennedy, D. B. (1995). Effectiveness of neural network types for prediction of business failure. *Expert Systems with Applications, 9*, 503–512.

Brachman, R. J., Khabaza, T., Kloesgen, W., Piatetsky-Shapiro, G., & Simoudis, E. (1996). Mining business databases. *Communications of the ACM, 39*(11), 42–48.

Broadus, R. N. (1987). Toward a definition of bibliometrics. *Scientometrics, 12*(5–6), 373–379.

Chen, M. S., Han, J., & Yu, P. S. (1996). Data mining: an overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering, 8*(6), 866–883.

Coille, R. C. (1977). Lotka's frequency distribution of scientific productivity. *Journal of American Society for Information Science, 28*, 366–370.

Drew, S. (1999). Building knowledge management into strategy: Making sense of a new perspective. *Long Range Planning, 32*(1), 130–136.

Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996a). From data mining to knowledge discovery: an overview. In *Advances in knowledge discovery and data mining* (pp. 1-34). California: American Association for Artificial Intelligence.

Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996b). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM, 39*(11), 27–34.

Fletcher, D., & Goss, E. (1993). Forecasting with neural networks: An application using bankruptcy data. *Information and Management, 24*(3), 159–167.

Gupta, D. K. (1987). Lotka's law and productivity of entomological research in Nigeria for the period 1900–1973. *Scientometrics, 12*, 33–46.

Hajirezaie, M., Mohammad, S., Husseini, M., Barfourosh, A. A., & Karimi, B. (2010). Modeling and evaluating the strategic effects of improvement programs on the manufacturing performance using neural networks. *African Journal of Business Management, 4*(4), 414–424.

Heijst, G., Spek, R., & Kruizinga, E. (1997). Corporate memories as a tool for knowledge management. *Expert Systems with Applications, 13*(1), 41–54.

Hendriks, P. H. J., & Vriens, D. J. (1999). Knowledge-based systems and knowledge management: Friends or foes? *Information and Management, 35*, 113–125.

Johannessen, J. A., Olsen, B., & Olaisen (1999). Aspects of innovation theory based on knowledge-management. *International Journal of Information Management, 19*, 121–139.

Langley, P., & Simon, H. A. (1995). Applications of machine learning and rule induction. *Communication of the ACM, 38*(11), 54–64.

Lau, H. C. W., Wong, C. W. Y., Hui, I. K., & Pun, K. F. (2003). Design and implementation of an integrated knowledge system. *Knowledge-Based Systems, 16*, 69–76.

Lejeune, M. A. P. M. (2001). Measuring the impact of data mining on churn management. *Internet Research: Electronic Networking Applications and Policy, 11*, 375–387.

Liao, S. H. (2002). Problem solving and knowledge inertia. *Expert Systems with Applications, 22*, 21–31.

Liebowitz, J. (2001). Knowledge management and its link to artificial intelligence. *Expert Systems with Applications, 20*, 1–6.

Liebowitz, J., & Wright, K. (1999). Does measuring knowledge make cents? *Expert Systems with Applications, 17*, 99–103.

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences, 16*(12), 317–324.

Moed, H. F., & Van Leeuwen, Th. N. (1995). Improving the accuracy of the institute for scientific information's journal impact factors. *Journal of the American Society for Information Science, 46*, 461–467.

Nicholls, P. T. (1989). Bibliometric modeling processes and empirical validity of Lotka's law. *Journal of American Society for Information Science, 40*(6), 379–385.

Nonaka, I., Umemoto, K., & Senoo, D. (1996). From information processing to knowledge creation: A paradigm shift in business management. *Technology in Society, 18*(2), 203–218.

Pao, M. L. (1985). Lotka's law, a testing procedure. *Information Processing and Management, 21*, 305–320.

Pao, M. L. (1989). *Concept of information retrieve*. Colorado: Libraries Unlimited.

Polanyi, M. (1966). *The tacit dimension*. London: Routledge and Kegan Paul.

Potter, W. G. (1981). Lotka's law revisited. *Library Trends, 30*(1), 21–39.

Potter, W. G. (1988). Of making many books there is no end: Bibliometrics and libraries. *The Journal of Academic Librarianship, 14*. 238a–238c.

Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of Documentation, 25*(4), 348–349.

Rao, I. K. R. (1980). The distribution of scientific productivity and social change. *Journal of American Society for Information Science, 31*, 111–122.

Rubenstein-Montano, B., Liebowitz, J., Buchwalter, J., McCaw, D., Newman, B., & Rebeck, K. (2001). A systems thinking framework for knowledge management. *Decision Support Systems, 31*, 5–16.

Salchenberger, L. M., Cinar, E. M., & Lash, N. A. (1992). Neural networks: A new tool for predicting thrift failures. *Decision Sciences, 23*, 899–916.

Su, C. T., Hsu, H. H., & Tsai, C. H. (2002). Knowledge mining from trained neural networks. *Journal of Computer Information Systems, 42*, 61–70.

Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: The case of bank failure predictions. *Management Science, 38*, 926–947.

Tsai, H. H. (2011). Research trends analysis by comparing data mining and customer relationship management through bibliometric methodology. *Scientometrics, 87*(3), 425–450.

Tsai, H. H. (2012). Global data mining: An empirical study of current trends, future forecasts and technology diffusions. *Expert Systems with Applications, 39*(9), 8172–8181.

Tsai, H. H., & Chang, J. K. (2011). E-Commerce research trend forecasting: A study of bibliometric methodology. *International Journal of Digital Content Technology and its Application, 5*(1), 101–111.

Tsai, H. H., & Chi, Y. P. (2011). Trend analysis of supply chain management by bibliometric methodology. *International Journal of Digital Content Technology and its Application, 5*(1), 285–295.

Tsai, H. H., & Chi, Y. P. (2012). An empirical study of research trends and forecasts: Customer relationship management. *African Journal of Business Management, 6*(4), 1418–1427.

Turban, E., Aronson, J. E., Liang, T. P., & Sharda, R. (2007). *Decision support and business intelligence systems* (8th ed.). Taiwan: Pearson Education.

Van Raan, A. F. J. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics, 36*, 397–420.

Van Raan, A. F. J. (2000). The Pandora's box of citation analysis: Measuring scientific excellence, the last evil? In B. Cronin & H. B. Atkins (Eds.), *The web of knowledge. A festschrift in honor of eugene garfield* (pp. 301–319). New Jersey: ASIS Monograph Series.

Van Raan, A. F. J., & Van Leeuwen, Th. N. (2002). Assessment of the scientific basis of interdisciplinary, applied research. Application of bibliometric methods in nutrition and food research. *Research Policy, 31*, 611–632.

Vlachy, J. (1978). Frequency distribution of scientific performance. A bibliography of Lotka's law and related phenomena. *Scientometrics, 1*, 109–130.

Weingart, P. (2003). *Evaluation of research performance: The danger of numbers*. In *Bibliometric analysis in science and research. Applications, benefits and limitations*. *Second conference of the central library* (pp. 7–19). Jülich: Forschungszentrum.

Weingart, P. (2004). Impact of bibliometrics upon the science system: Inadvertent consequences? In H. F. Moed, W. Glanzel, & U. Schmoch (Eds.), *Handbook on quantitative science and technology research*. The Netherlands: Kluwer Academic Publishers.

Wiig, K. M. (1997). Knowledge management: where did it come and where will it do? *Expert Systems with Applications, 13*(1), 1–14.

Wiig, K. M., Hoog, R., & Spex, R. (1997). Supporting knowledge management: A selection of methods and techniques. *Expert Systems with Applications, 13*(1), 15–27.

Wilkins, J., Wegen, B., & Hoog, R. (1997). Understanding and valuing knowledge assets: Overview and method. *Expert Systems with Applications, 13*(1), 55–72.

Zhang, G., Hu, M. Y., Patuwo, B. E., & Indro, D. C. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross validation analysis. *European Journal of Operational Research, 116*, 16–32.