Journal of Informetrics 14 (2020) 101004

Contents lists available at ScienceDirect

Journal of Informetrics

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

Regular article

Identification of highly-cited papers using topic-model-based and bibliometric features: the consideration of keyword popularity

Ya-Han Hu[a,b,c], Chun-Tien Tai[d,e], Kang Ernest Liu[f,*], Cheng-Fang Cai[d]

[a] *Department of Information Management, National Central University, Taoyuan City, Taiwan 320, R.O.C.*
[b] *Center for Innovative Research on Aging Society, National Chung Cheng University, Chiayi, Taiwan 621, R.O.C.*
[c] *MOST AI Biomedical Research Center at National Cheng Kung University, Taiwan 701, R.O.C.*
[d] *Department of Information Management, National Chung Cheng University, Chiayi, Taiwan 621, R.O.C.*
[e] *Chiayi Chang Gung Memorial Hospital, Chiayi, Taiwan 613, R.O.C.*
[f] *Department of Agricultural Economics, National Taiwan University Taipei, Taiwan 106, R.O.C.*

ARTICLE INFO

Article history:
Received 1 April 2019
Received in revised form
24 December 2019
Accepted 25 December 2019

Keywords:
highly-cited papers
keyword popularity
supervised learning
binary classification
topic model

ABSTRACT

The number of received citations have been used as an indicator of the impact of academic publications. Developing tools to find papers that have the potential to become highly-cited has recently attracted increasing scientific attention. Topics of concern by scholars may change over time in accordance with research trends, resulting in changes in received citations. Author-defined keywords, title and abstract provide valuable information about a research article. This study performs a latent Dirichlet allocation technique to extract topics and keywords from articles; five keyword popularity (KP) features are defined as indicators of emerging trends of articles. Binary classification models are utilized to predict papers that were highly-cited or less highly-cited by a number of supervised learning techniques. We empirically compare KP features of articles with other commonly used journal-related and author-related features proposed in previous studies. The results show that, with KP features, the prediction models are more effective than those with journal and/or author features, especially in the management information system discipline.

1. Introduction

As the Internet flourishes, journal related information (e.g., metadata) in various domains can be obtained by searching literature databases such as Web of Science (WOS) and Scopus. Google Scholar (https://scholar.google.com/) and Research-Gate (https://www.researchgate.net/) also provide platforms to search for scientific papers. Due to the continuous growth of published academic papers and their easy access in recent decades, users have increasingly spent time searching for influential papers (Rodríguez-Bolívar, Alcaide-Muñoz, & Cobo, 2018). In scientific areas, one of the most commonly used approaches to locating important papers of interest is the citation count, which helps researchers determine which paper is significant to read. The basic idea behind this metric is that the more citations a paper receives, the higher its impact.

* Corresponding author at: National Taiwan University, Department of Agricultural Economics, No. 1, Section 4, Roosevelt Road, Taipei 10617, Taiwan. Tel.: +886-2-33662657; fax: +886-2-23628496.

E-mail addresses: yahan.hu@mis.ccu.edu.tw (Y.-H. Hu), flamquit@hotmail.com (C.-T. Tai), kangernestliu@ntu.edu.tw (K.E. Liu), u9933116@gmail.com (C.-F. Cai).

https://doi.org/10.1016/j.joi.2019.101004
1751-1577/

Consequently, citation counts have become a popular indicator of the impact of academic publications (Stegehuis, Litvak, & Waltman, 2015). Developing effective citation prediction tools has attracted increasing attention in the field of scientometrics (e.g., Fu & Aliferis, 2010; Wang et al., 2012; Newman, 2014; Kosteas, 2018; Wang, Wang, & Chen, 2019; Abrishami & Aliakbary, 2019).

Previous research on citation prediction models can be divided into two categories: prediction of citation counts and identification of highly-cited papers. As for citation count prediction, some citation prediction studies have developed models and tested their effectiveness. For example, Abrishami and Aliakbary (2019) utilized deep neural network learning techniques to predict long-term citation counts of a paper based only on its early citation counts. Bai, Zhang, and Lee (2019) introduced the Paper Potential Index model to explore citation patterns over time in physics. Some studies have focused on identifying factors that affect citation counts of scientific papers. Kosteas (2018) claimed that short-term citations (between 1 and 2 years after publication) predicted long-run citation counts better than features such as journal impact factors (JIF) or other journal rankings for economic journals. In the biomedical field, Fu and Aliferis (2010) showed that using a mixture of content-based and bibliometric features has good performance for predicting future citation counts.

As for the identification of highly-cited papers, some researchers have developed tools for predicting frequently cited papers. For example, Newman (2014) utilized the statistical z-scores within a field or topic to detect which scientific papers would be highly-cited in the future, even if they have not currently received substantial citations. Wang et al. (2012) built up a case-based classifier with a soft fuzzy rough set technique to recognize highly-cited papers. Other researchers have attempted to create more efficient indicators to better predict highly-cited papers. Wang et al. (2019) utilized the neural network algorithms to investigate the effectiveness of the selected factors in predicting the ESI (Essential Science Indicators) highly-cited papers. Their findings showed that the team factor is important in the long term whereas the potential leader factor is important in the short term. Wang et al. (2019) combined twenty bibliometric features obtained through WOS and three of the five alternative metrics, i.e., the saved, viewed, and discussed times in the first two years after publication, collected from PLOS ALMs (Article-level Metrics), to predict the future success of papers using supervised learning techniques. Their experimental results showed that using the combined features of both can better predict the highly-cited papers.

Since keywords contain the core information that authors are most concerned about, keyword analyses have been used to identify the subjective focus and to discover scientific research hotspots and trends (e.g., Tian, Wen, & Hong, 2008; Li, Ding, Feng, Wang, & Ho, 2009; Chang, Huang, & Lin, 2015). Especially, frequency of keywords has been widely used as the primary metric in signaling research trends (e.g., Zhang et al., 2014; Huang & Zhao, 2019). Keywords identified by burst detection methods were utilized as indicators of emerging trends over time (Zhou & Zhao, 2015). Topics of concern by scholars may change over time in accordance with research trends, resulting in changes in received citations and JIF (Finardi, 2014).

The effect of author-defined keywords on citation counts has recently been studied. Uddin and Khan (2016) utilized correlation analyses and regression techniques to investigate the impact of author-defined keywords on citation counts. Applying to the obesity research domain, their findings showed that most metrics defined by using author-defined keywords were statistically significant correlated to citation counts. Sohrabi and Iraj (2017) also utilized author-defined keywords; they created two variables measuring the keyword repetition in abstracts and the keyword frequency per journal. Employing both the logistic regression (LGR) and the least-squares linear regression models, the effects of both variables were statistically significant in predicting citation counts.

In addition to author-defined keywords, title and abstract also provide valuable information in searching influential papers. As text mining techniques have become increasingly available, computational semantics methods such as latent semantic analysis and topic models can be used to identify the main themes and to extract article keywords (Natale, Fiore, & Hofherr, 2012). This study attempts to utilize these article keywords extracted by using latent Dirichlet allocation (LDA) algorithm and topic models to create features of keyword popularity (KP) of articles.

As for keyword popularity, Huang and Zhao (2019) cited the definition of "popular" from the Merriam-Webster dictionary as "of or relating to the general public" and "frequently encountered or widely accepted". They used R package "PAFit", a Bayesian statistical method, to measure keyword popularity and compared the results with the frequency of keywords in the network analysis of ecological topics. Their findings revealed that KP has a stronger power in detecting ecological hotspots than frequency. Unlike Huang and Zhao (2019), we adopt the keyword popularities of articles as predictors in this study. Specifically, we create and define five KP features, namely the published popularity (PP), topic popularity (TP), news popularity (NP), web page popularity (WPP) and video popularity (VP) to construct keyword popularities of articles and examine whether or not these KP features can improve the power of identification of highly-cited papers.

The aim of this study is to develop prediction models that recognize future highly-cited papers using journal, author, and KP features. The uniqueness of the current paper is threefold. First, keywords contain valuable information but have rarely been used as a predictor in citation prediction models. This study is one of a few studies making best use of keywords in predicting future highly-cited papers. Second, different from utilizing author-defined keywords in previous studies, this paper takes advantage of article information (title, abstract, and author-defined keywords) and online information platforms (Google Scholar, Google Trends, and ResearchGate) to generate five KP features via techniques of text mining and probabilistic topic models. To the best of our knowledge, these metrics are newly constructed and thus have never been used in evaluating citation prediction models. Third, this study features a wide range of data sources, including WOS, Scopus, Google Scholar, Google Trends, and ResearchGate, which makes the current study more integrated than older ones.

Table 1
Comparison of related works.

Work	Journal feature						Author feature	Keyword feature	Data source
	JIF	5-JIF	EF	AI	SJR	SNIP	h-index		
Bai et al. (2019)	√						√		American Physical Society dataset
Wang et al. (2019)	√						√		Public Library of Science (PLOS)
Yin, Gong, and Wang (2018))	√						√	√	SCI-E and SSCI databases
Sohrabi and Iraj (2017)					√			√	Scopus
Bornmann, Leydesdorff, and Wang (2014))	√								WOS
Dorta-González, Dorta-González, Santos-Peñate, and Suárez-Vega (2014))		√	√						Journal citation reports
Tsai (2014)	√	√					√		WOS
Wang, Song, and Barabási (2013))	√						√		Science citation index
Acuna, Allesina, and Kording (2012))							√		Academic-tree.org
Wang et al. (2012)	√						√		Science citation index
Beliakov and James (2011)	√	√	√	√					Australian research council
Leydesdorff and Bornmann (2011)	√					√			WOS
Fu and Aliferis (2010)	√								WOS MEDLINE
Our study	√	√			√		√	√	WOS Scopus Google Scholar Google Trends ResearchGate

Note: JIF: journal impact factor; 5-JIF: 5-year journal impact factor; EF: Eigenfactor; AI: Article Influence; SJR: SCImago Journal Rank; SNIP: Source Normalized Impact per Paper.

The rest of this paper is organized as follows. Section 2 reviews recent studies on citation count prediction and feature extraction from the text of research articles. Section 3 entails the preparation of data, variables, and experimental setup. Section 4 provides evaluation results and discussion. Section 5 concludes our study.

2. Literature overview

2.1. Features affecting citation counts

Several categories of features were extensively investigated in predicting citation counts. Among them, journal-related and author-related features were frequently used in previous studies. As shown in Table 1, most studies have utilized either journal or author features. Among journal features, the primary ones were journal impact factor (JIF) and 5-year JIF (5-JIF) in the collected studies whereas other features such as Eigenfactor (EF), Article Influence (AI), SCImago Journal Rank (SJR) and Source Normalized Impact per Paper (SNIP) have been rarely adopted. Even though author-defined keywords were used in citation count prediction models, to the best of our knowledge, keyword popularities of articles have never been applied to any of the related works and thus KP will be further investigated in this study in addition to journal and author predictors. Five KP features were created and included in our prediction models. The goal of the current study is to examine whether or not these new KP features can provide additional predicting power.

2.2. Feature extraction from the content of academic papers

Feature extraction techniques were commonly adopted in previous studies in order to retrieve information from the content of academic papers. Table 2 lists several related works employing different feature extraction techniques such as the latent Dirichlet allocation (LDA) (e.g., Jiang, Qiang, & Lin, 2016; Kar, Nunes, & Ribeiro, 2015; Song & Ding, 2014), word embedding (Zhang et al., 2018), and term frequency-inverse document frequency (TF-IDF) (Kim, Ha, Lee, Jo, & El-Saddik, 2011) techniques. Among all well-known techniques, the LDA has been extensively applied in research on machine learning and information retrieval (Blei, Ng, & Jordan, 2003). LDA is a probabilistic topic model that can be used to process large files and identify relevant topic structure to generate shorter file descriptions. Because LDA alleviates the drawbacks of both the complexity in excessive computational burden and a lack of foundation of statistical modeling observed by latent semantic

Table 2
Comparison of feature extraction techniques in related works.

Work	Technique	Data source
Liu, Tian, Kong, Lee, and Xia (2019))	LDA	8 top-tier MIS journals and conferences
Iqbal et al. (2019)	LDA	ACM, IEEE Xplore, Scopus and CrossRef
Zhang et al. (2018)	Word2Vec	WOS and National Science Foundation
Jiang et al. (2016)	LDA	WOS
Kar et al. (2015)	LDA	Wikipedia
Song and Ding (2014)	LDA	WOS
Zhang, Zhang, Gao, Guo, and Sun (2012))	LDA	CiteULike
Kim et al. (2011)	TF-IDF	National Science Foundation
Pan and Li (2010)	LDA	Biometrics research papers
Liang, Yang, Chen, and Ku (2008))	Semantic Analysis	National digital library of theses and dissertations

Note: LDA: latent Dirichlet allocation; Word2Vec: Word to Vector; TF-IDF: term frequency-inverse document frequency.

Table 3
Number of published articles from selected journals by field of research.

Field of Research	Journal	Number of Published Articles				
		2009	2010	2011	2012	Total
marketing	JM	25	25	25	25	100
	JMR	25	25	25	25	100
	MS	25	25	25	25	100
MIS	ISR	26	47	47	66	186
	MISQ	31	32	45	52	160
	JMIS	25	25	25	25	100
	Total	157	179	192	218	746

Note: JM: Journal of Marketing; JMR: Journal of Marketing Research; MS: Marketing Science; ISR: Information Systems Research; MISQ: MIS Quarterly; JMIS: Journal of Management Information Systems.

analysis, LDA has been increasingly popular in recent empirical studies. The LDA technique is therefore adopted to construct the KP features in the current study.

3. Data and Research Methods

3.1. Dataset

There were 46 journals listed in *Financial Times* in 2012. Among them, we selected two fields of research, i.e. marketing and management information system (MIS) to explore discipline differences in citation prediction models. The three top influential journals from each discipline were chosen for this study, including *Journal of Marketing* (JM), *Journal of Marketing Research* (JMR), *Marketing Science* (MS), *Information Systems Research* (ISR), *MIS Quarterly* (MISQ), and *Journal of Management Information Systems* (JMIS). All papers published between 2009 and 2012 were included in our experiments. Table 3 shows the number of published articles in each year. During this 4-year span, the number of papers published in marketing was unchanged with 25 articles each year whereas the number of papers published in MIS increased except for the JMIS.

3.2. Text preprocessing for article metadata

Fig. 1 shows the text preprocessing steps. The four major phases of text pre-processing comprised of (1) extracting article metadata (title, abstract, and keywords) from WOS, (2) conducting text preprocessing techniques to generate article-term matrix, (3) applying LDA to reduce number of features and to obtain the final set of keywords, and (4) retrieving keyword popularities by accessing external search engines.

The first step was to collect metadata of target articles from WOS. The abstract, title and author-defined keywords of each article were combined in the corpus. Next, a number of text preprocessing tasks were performed, including word segmentation, stemming, and part-of-speech tagging.

In word segmentation, an input word string was tokenized into single terms. After that, the Porter stemming algorithm reverted each term to its original form (Porter, 1980). The stemming process reduced inflectional or derivational forms of a word to its original form, which significantly reduced the feature space of a document. The Stanford Part-of-Speech Tagger was applied to assign parts of speech to each word (e.g., nouns, verbs, and adjectives) (Zheng, Kang, & Kim, 2009). Because nouns (NN), proper nouns (NP), and adjectives (JJ) are more capable of differentiating academic paper contents than other parts of speech, the present study further extracted text containing NN, NP, and JJ in the articles (Chen & Lin, 2010). At the end of step 2, each article was represented by a feature vector listing all NN, NP, and JJ terms. Therefore, the article corpus was transformed into an article-term matrix.

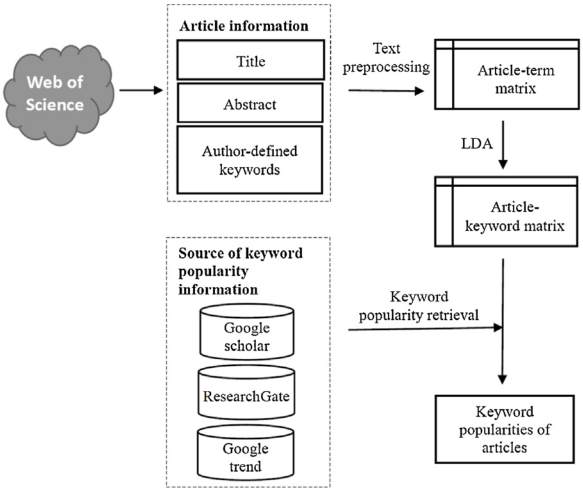


Fig. 1. Text preprocessing steps.

Table 4
The variable definition and types.

Category	Variable	Name	Definition
Journal (J)	JIF	Journal impact factor	Journal impact factor in the published year of the article
	5-JIF	5-year journal impact factor	5-year journal impact factor in the published year of the article
	JTC	Journal total citation counts	Journal total citation counts in the published year of the article
	SJR	SCImago Journal Rank	SCImago Journal Rank in the published year of the article
Author (A)	HCA	Author H-index	H-index of corresponding author in the published year of the article
	NCA	Author publications	Number of papers published by corresponding author in the published year of the article
	CCA	Author citations	Citation counts of corresponding author in the published year of the article
Keyword popularity (KP)	TP	Topic popularity	Weighted mean of the numbers of questions from ResearchGate matching article keywords with corresponding probabilities
	PP	Published popularity	Weighted mean of the numbers of search results of the article keywords from Google Scholar with corresponding probabilities
	NP	News popularity	Weighted mean of the degrees of news popularity of article keywords from Google Trends with corresponding probabilities
	WPP	Web page popularity	Weighted mean of the degrees of web page popularity of article keywords from Google Trends with corresponding probabilities
	VP	Video popularity	Weighted mean of the degrees of video popularity of article keywords from Google Trends with corresponding probabilities

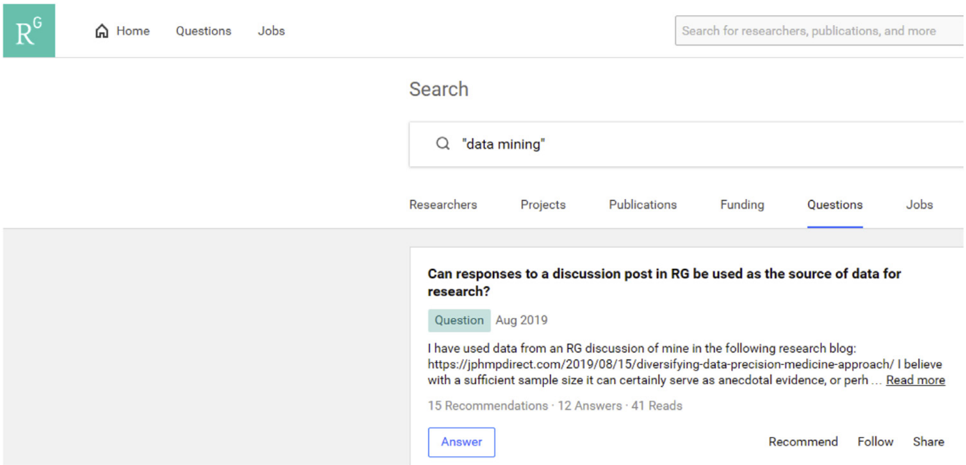
The third step used LDA to identify meaningful keywords from the article-term matrix, which is a useful approach for dimension reduction. This study employed JGibbLDA developed by Phan and Nguyen (2008) to construct the topic models through Gibbs sampling. JGibbLDA is an open-source code written in Java. The whole LDA process ended when the topic model was fully converged; keywords and topics as well as the probability distributions of each keyword within the topic (φ) and that of the topic within the paper (θ) were generated. Based on the set of keywords discovered by LDA, the article corpus can be represented by the article-keyword matrix.

The final step retrieved keyword popularities by searching through ResearchGate, Google Scholar and Google Trends. Fig. 2 illustrates the results of querying the term "data mining" on ResearchGate, Google Scholar and Google Trends. Through the use of web crawler and their application programming interfaces, we searched each keyword on the websites during February 2017 and obtained its popularity values. As a result, keyword popularities of articles were calculated using the formulas discussed in the next section.

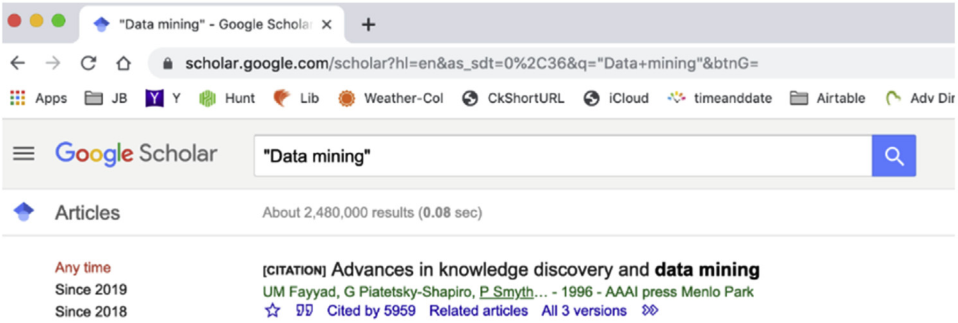
3.3. Variables

The dependent variable was defined as whether the total citation count of a paper is ranked in the top q% of the selected research field for each of the six years since publication. The citation count of each article was accessed from Scopus. The value of threshold q was set as 25, 33, and 50, respectively. Specifically, a paper was classified as *highly-cited* if its citation count was ranked in the top q%; whereas it was defined as *less highly-cited* if being ranked in the bottom q%.

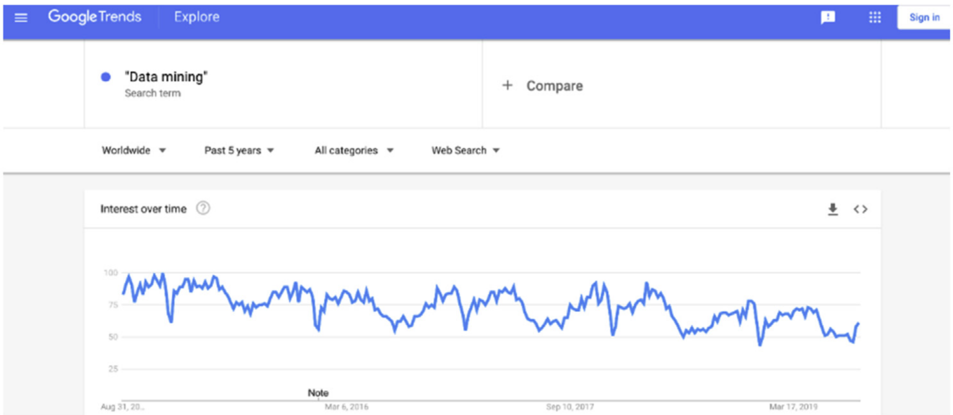
This study examined the effect of the following three independent variable sets on identifying highly-cited papers: journal features (J), author features (A), and keyword popularity features (KP) (see Table 4). Among the journal features,



(a) ResearchGate



(b) Google Scholar



(c) Google Trends

Fig. 2. Screenshots of searching ResearchGate, Google Scholar and Google Trends.

journal impact factor (JIF), 5-year journal impact factor (5-JIF), and journal total citation counts (JTC) were retrieved from WOS; SCImago Journal Rank (SJR) were accessed on SCImagoJR. Note that the values of journal features vary each year. For each article in the dataset, we only consider the values of journal features in the published year of the article. As for author features, the h-index of the corresponding author (HCA), the number of papers published by the corresponding author (NCA), and the citation counts of the corresponding author (CCA) were collected from Scopus.

Table 5
Confusion matrix.

		Predicted class	
		highly-cited	less highly-cited
Actual class	**highly-cited**	True Positive (TP)	False Negative (FN)
	less highly-cited	False Positive (FP)	True Negative (TN)

As for the keyword popularity features, five variables were created in this study, namely topic popularity (TP), published popularity (PP), news popularity (NP), web page popularity (WPP) and video popularity (VP). After text preprocessing mentioned in Section 3.2, all of the articles were represented by a collection of keywords determined by LDA. Let M be the number of topics and N the number of keywords in each topic in LDA. The $k_{m,n}$ $(m \in M,\ n \in N)$ is defined as the n-th keyword in the m-th topic. Given an article p, the probability of topic m occurring in p is defined as θ_m^p; the probability of keyword $k_{m,n}$ occurring in topic m is φ_n^m.

Assume that an article p was published in year Y. For each keyword $k_{m,n}$, we searched ResearchGate to obtain the number of questions (denoted as $top_{m,n}$) related to $k_{m,n}$. The number of search results for querying $k_{m,n}$ filtered by year Y, denoted as $pub_{m,n}^Y$, can be retrieved from Google Scholar, where Y indicates the year of the paper being published. In Google Trends, similarly, we query $k_{m,n}$ and set the filter of web search (i.e., Google News, Google Web Pages and YouTube) to get the degrees of news popularity ($new_{m,n}^Y$), web pages popularity ($web_{m,n}^Y$) and video popularity ($vid_{m,n}^Y$), for the selected year Y. As a result, the topic popularity (TP_Y^p), published popularity (PP_Y^p), news popularity (NP_Y^p), web page popularity (WPP_Y^p) and video popularity (VP_Y^p) can be mathematically formulated as:

$$\mathrm{TP}_Y^p = \Sigma_m \Sigma_n top_{m,n} \cdot \theta_m^p \cdot \varphi_n^m \tag{1}$$

$$\mathrm{PP}_Y^p = \Sigma_m \Sigma_n pub_{m,n}^Y \cdot \theta_m^p \cdot \varphi_n^m \tag{2}$$

$$\mathrm{NP}_Y^p = \Sigma_m \Sigma_n new_{m,n}^Y \cdot \theta_m^p \cdot \varphi_n^m \tag{3}$$

$$\mathrm{WPP}_Y^p = \Sigma_m \Sigma_n web_{m,n}^Y \cdot \theta_m^p \cdot \varphi_n^m \tag{4}$$

$$\mathrm{VP}_Y^p = \Sigma_m \Sigma_n vid_{m,n}^Y \cdot \theta_m^p \cdot \varphi_n^m \tag{5}$$

3.4. Experiment setup and performance measure

On the basis of disciplines, this study assigned the experimental data into two groups: marketing and MIS. In order to test whether the keyword popularity can improve the prediction performance of classifiers, seven feature sets were formulated in this study: journal features only (J), author features only (A), keyword popularity features only (KP), J + A, J + KP, A + KP, and J + A + KP.

In the use of JGibbLDA, this study set the numbers of topics and the number of keywords for each topic as 50 and 30, respectively. The Orange, a Python-based data mining and machine learning software, was used to construct classification models (Demšar et al., 2013). Four well-known classification techniques were selected in our experiments, namely C4.5 (Quinlan, 2014), logistic regression (LGR) (Lemeshow & Hosmer, 1998), support vector machine (SVM) (Vapnik, 2013), and artificial neural network (ANN) (Rumelhart, Hinton, & Williams, 1986).

A 10-fold cross-validation method was used to assess the prediction performance of the classifiers (Kretschmann, Fleischmann, & Apweiler, 2001). It randomly divides the whole dataset into 10 disjoint sets of approximately equal size, where one set is selected as a test set and the remaining nine sets are combined as the training set. This process is repeated 10 times by selecting different fold as a test set, and the average of performance measures can be obtained.

There are a number of widely accepted indicators for evaluating the performance measure of classification models. Using the confusion matrix in Table 5, the classification performance can be evaluated using the following three metrics:

$$precision{=}TP/(TP{+}FP) \tag{6}$$

$$recall{=}TP/(TP{+}FN) \tag{7}$$

$$F\text{-}measure{=}2.precision.recall/(precision{+}recall) \tag{8}$$

Precision is the ratio of correctly predicted highly-cited papers to the total predicted highly-cited papers. Recall is the ratio of correctly predicted highly-cited papers to all of the highly-cited papers in the dataset. F-measure is the harmonic mean of precision and recall. In addition, the receiver operating characteristic (ROC) curve analysis is a fundamental statistical method for measuring the reliability of binary classification models. In this study, the area under the ROC curve (AUC) was also included as an indicator of classification performance. According to Hosmer and Lemeshow (2004), the evaluation performance is defined as "excellent" if AUC ≥ 0.9, "good" if 0.9 > AUC ≥ 0.8, "fair" if 0.8 > AUC ≥ 0.7, "poor" if 0.7 > AUC ≥ 0.6, and "very poor" if AUC < 0.6.

Table 6
Classification performance results using AUC: top 25%.

Field of Research	Classifier	Period	Feature set						
			J	A	KP	J+A	J+KP	A+KP	J+A+KP
Marketing	C4.5	Y+1	0.550	0.534	0.487	0.571	0.549	0.554	0.564
		Y+2	0.723	0.562	0.477	0.671	0.596	0.529	0.655
		Y+3	0.701	0.625	0.481	0.707	0.552	0.542	0.657
		Y+4	0.743	0.605	**0.540**	0.671	0.573	0.534	0.617
		Y+5	**0.745**	**0.652**	0.535	0.687	**0.670**	**0.564**	**0.695**
		Y+6	0.727	0.620	0.428	**0.753**	0.573	0.529	0.683
	LGR	Y+1	0.589	0.610	0.488	0.668	0.600	0.598	0.665
		Y+2	0.711	0.683	0.458	0.744	0.691	0.657	**0.743**
		Y+3	0.697	0.671	0.458	0.727	0.685	0.632	0.720
		Y+4	**0.724**	0.643	0.508	0.746	**0.717**	0.619	0.739
		Y+5	0.699	**0.686**	0.548	0.755	0.695	**0.659**	0.741
		Y+6	0.711	**0.686**	**0.556**	**0.769**	0.680	0.658	0.740
	SMO	Y+1	0.508	0.578	0.472	0.552	0.424	0.531	0.485
		Y+2	0.668	0.682	**0.512**	0.701	0.679	0.613	0.718
		Y+3	0.677	0.662	0.465	0.716	0.692	0.585	0.710
		Y+4	**0.752**	0.654	0.463	**0.776**	**0.760**	0.582	**0.761**
		Y+5	0.716	0.688	0.408	0.738	0.735	0.602	0.754
		Y+6	0.720	**0.690**	0.401	0.768	0.712	**0.624**	0.749
	ANN	Y+1	0.574	0.585	0.541	0.625	0.550	0.598	0.609
		Y+2	0.727	0.678	0.473	0.769	0.649	**0.617**	0.721
		Y+3	0.719	0.685	0.473	0.765	0.663	0.574	0.730
		Y+4	**0.759**	0.657	0.507	0.783	0.709	0.579	0.745
		Y+5	0.743	**0.696**	0.527	0.784	**0.724**	0.616	0.746
		Y+6	**0.759**	0.686	**0.547**	**0.795**	0.720	0.611	**0.756**
MIS	C4.5	Y+1	0.787	0.540	0.599	0.729	0.662	0.595	0.670
		Y+2	0.773	**0.627**	0.545	0.713	0.716	**0.611**	0.658
		Y+3	**0.833**	0.592	0.573	0.758	**0.754**	0.598	0.768
		Y+4	0.832	0.572	0.563	0.780	0.729	0.573	0.700
		Y+5	0.822	0.594	**0.642**	0.751	0.752	0.608	0.717
		Y+6	0.820	0.573	0.630	**0.800**	0.722	0.628	**0.782**
	LGR	Y+1	0.797	0.636	0.649	0.829	0.841	0.686	0.842
		Y+2	0.795	0.667	0.670	0.830	0.843	0.722	0.845
		Y+3	**0.817**	0.664	0.668	**0.860**	0.852	0.712	0.868
		Y+4	0.814	0.656	0.675	0.848	0.855	0.722	0.859
		Y+5	0.803	**0.669**	0.676	0.857	0.848	0.727	0.860
		Y+6	0.811	0.664	**0.697**	0.856	**0.860**	**0.741**	**0.870**
	SMO	Y+1	0.748	0.597	0.633	0.779	0.785	0.652	0.794
		Y+2	0.736	**0.640**	0.658	0.810	0.792	0.700	0.807
		Y+3	**0.771**	0.624	0.656	0.818	0.820	0.675	0.833
		Y+4	0.754	0.614	0.674	0.813	**0.825**	0.699	0.837
		Y+5	0.734	0.626	0.675	0.807	0.792	**0.704**	0.823
		Y+6	0.740	0.628	**0.679**	**0.826**	0.814	**0.704**	**0.845**
	ANN	Y+1	0.804	0.643	0.621	0.833	0.817	0.665	0.824
		Y+2	0.800	**0.661**	0.661	0.821	0.821	0.719	0.819
		Y+3	0.811	0.659	0.665	**0.861**	0.836	0.701	0.856
		Y+4	**0.834**	0.645	0.622	0.855	**0.848**	0.711	**0.864**
		Y+5	0.818	0.659	0.646	0.848	0.838	0.695	0.856
		Y+6	0.820	**0.661**	**0.690**	0.859	0.847	**0.742**	0.857

Note: MIS: management information system; LGR: logistic regression; SMO: support vector machine; ANN: artificial neural network; Y + *t* (*t* = 1,2,. . .,6): the period *t* to be evaluated after published in year Y; J: journal feature set; A: author feature set; KP: keyword popularity feature set.

4. Results and discussion

4.1. Evaluation results of classification performance

According to our designed experiments, binary classification models were developed to predict future citations. Three thresholds were used to classify the top 25%, 33%, and 50% of the highly-cited papers. The experimental results of the top 25% are presented in Table 6 whereas those of the top 33% and 50% are shown in Appendix Tables. Table 6 displays AUC measures according to disciplines (marketing and MIS), supervised classifiers (ANN, C4.5, LGR, and SMO), years after publication (denoted by Y + 1, Y + 2, Y + 3, Y + 4, Y + 5, Y + 6), and seven different feature sets (i.e., J, A, KP, J + A, J + KP, A + KP, J + A + KP). The highest AUC values among the six periods for each of the classifiers and feature sets are depicted in bold numbers.

This study used the AUC to assess the effectiveness of the selected four classification techniques. The top panel of Table 6 showed that, for marketing, the highest AUC scores distributed unevenly among the four classifiers. ANN classifier earned the highest scores in three out of seven feature sets, i.e., J, A, and J + A; whereas both LGR and SMO classifiers received the

Table 7
Classification performance results for highly-cited papers.

Field of Research			**Feature set**						
	Metric	**Period**	**J**	**A**	**KP**	**J+A**	**J+KP**	**A+KP**	**J+A+KP**
	precision	Y+1	0.333	0.680	0.300	0.410	0.370	0.472	0.449
Marketing		Y+2	0.673	0.683	0.373	0.667	0.603	0.583	0.619
		Y+3	0.689	0.679	0.487	0.727	0.647	0.565	0.706
		Y+4	0.754	0.649	0.500	0.761	0.687	0.603	0.716
		Y+5	0.810	0.667	0.521	0.761	0.648	0.600	0.701
		Y+6	0.810	0.696	0.547	0.776	0.652	0.582	0.681
	recall	Y+1	0.145	0.246	0.130	0.232	0.246	0.246	0.319
		Y+2	0.544	0.412	0.279	0.618	0.559	0.412	0.574
		Y+3	0.568	0.514	0.500	0.649	0.595	0.527	0.649
		Y+4	0.722	0.514	0.458	0.708	0.639	0.569	0.667
		Y+5	0.635	0.486	0.500	0.689	0.622	0.568	0.635
		Y+6	0.627	0.520	0.547	0.693	0.600	0.520	0.653
	F-measure	Y+1	0.202	0.362	0.182	0.296	0.296	0.324	**0.373**
		Y+2	0.602	0.514	0.319	**0.641**	0.580	0.483	0.595
		Y+3	0.622	0.585	0.493	**0.686**	0.620	0.545	0.676
		Y+4	**0.738**	0.574	0.478	0.734	0.662	0.586	0.691
		Y+5	0.712	0.562	0.510	**0.723**	0.634	0.583	0.667
		Y+6	0.707	0.595	0.547	**0.732**	0.625	0.549	0.667
	precision	Y+1	0.747	0.667	0.500	0.719	0.742	0.620	0.731
MIS		Y+2	0.773	0.739	0.683	0.805	0.759	0.684	0.783
		Y+3	0.793	0.698	0.595	0.783	0.791	0.635	0.763
		Y+4	0.795	0.721	0.689	0.775	0.791	0.647	0.786
		Y+5	0.776	0.773	0.622	0.857	0.771	0643	0.824
		Y+6	0.800	0.791	0.660	0.821	0.795	0.661	0.829
	recall	Y+1	0.740	0.280	0.180	0.690	0.720	0.310	0.680
		Y+2	0.708	0.354	0.292	0.646	0.688	0.406	0.677
		Y+3	0.753	0.309	0.227	0.742	0.742	0.340	0.732
		Y+4	0.745	0.330	0.330	0.734	0.723	0.351	0.702
		Y+5	0.725	0.374	0.253	0.659	0.703	0.396	0.670
		Y+6	0.731	0.366	0.333	0.742	0.710	0.398	0.731
	F-measure	Y+1	**0.744**	0.394	0.265	0.704	0.731	0.413	0.705
		Y+2	**0.739**	0.479	0.409	0.717	0.721	0.510	0.726
		Y+3	**0.772**	0.429	0.328	0.762	0.766	0.443	0.747
		Y+4	**0.769**	0.453	0.446	0.754	0.756	0.455	0.742
		Y+5	**0.750**	0.504	0.359	0.745	0.736	0.490	0.739
		Y+6	0.764	0.500	0.443	**0.780**	0.750	0.497	0.777

Note: MIS: management information system; Y+*t* (*t* = 1,2,. . .,6): the period *t* to be evaluated after published in year Y; J: journal feature set; A: author feature set; KP: keyword popularity feature set.

highest scores in two feature sets each, i.e., KP and A + KP for LGR and J + KP and J + A + KP for SMO. In addition, for the top 33% and 50% of the highly-cited papers (Tables A1 and A2), the ANN algorithm yielded more of the highest AUC scores, five and four out of seven feature sets, respectively. As for MIS, the bottom panel of Table 6 presents the comparison results. The LGR algorithm yielded the highest AUC scores among four out of seven different feature sets, namely, A, KP, J + KP, and J + A + KP. If the highly-cited papers were classified using 33% and 50% as thresholds, the LGR almost won all the feature sets (Tables A1 and A2). Our results suggest that, for marketing, the ANN algorithm is generally more effective compared with C4.5, SMO and LGR classifiers and that the LGR dominates other classifiers for MIS. As a result, we further narrowed down our discussion of the comparative results using the ANN for marketing and the LGR for MIS, respectively.

Our performance comparison of prediction models turned to the six different years after publication, denoted as Y + 1, Y + 2, Y + 3, Y + 4, Y + 5, and Y + 6, respectively. Among the seven feature sets, the highest scores happened mostly in Y + 6 for both fields of research. The overall highest value was 0.795 in the J + A feature set for marketing and 0.870 in J + A + KP for MIS in Table 6. Looking closer at the AUC values of feature sets, some commonalities and differences were found in disciplines. Regardless of years in predicting future citations, AUC metrics in J + A were found to be higher than those in J for both marketing and MIS, suggesting that inclusion of both journal and author features may make a better prediction than using either journal or author feature sets alone. In addition, for marketing, feature sets of J + KP, A + KP, and J + A + KP had mostly lower AUC scores than those of J, A, and J + A, respectively; whereas for MIS, AUC scores with additional KP features were universally higher than those without. This finding reveals that inclusion of the newly constructed KP features may improve the overall classification prediction of future papers but depends on disciplines.

If highly-cited papers attract more attention than less highly-cited ones, correct prediction of highly-cited papers is relatively more important than that of less highly-cited papers. Table 7 shows the prediction performance specific to the target class (i.e., highly-cited papers) using the three metrics, i.e., precision, recall, and F-measure. Because F-measure incorporates information of both precision and recall, F-measure is commonly used to compare the overall performance. For marketing, the best F-measure among the seven feature sets in Y + 1 was 0.373 for J + A + KP, showing an ineffective performance in prediction. The F-measure increased to its peak in Y + 4 for the feature set J and the highest scores in the last two periods, i.e., Y + 5 and Y + 6, both appeared for J + A. The maximum F-measures took place mostly for the J + A feature set in terms of time windows. However, for MIS, the highest F-measure metrics were concentrated for the journal feature set during the first five years of prediction whereas the largest one in Y + 6 was 0.780 for J + A. In addition, the overall highest F-measure was 0.738 in the feature set of journals only in Y + 4 for marketing and 0.780 in J + A feature set in Y + 6 for MIS.

4.2. Discussion

What types of features can better identify future highly-cited papers: journal-related features, author-related features or keyword-related features? This study utilized supervised classification techniques to perform a binary prediction of future highly-cited papers versus less highly-cited ones.

The experiment results reveal that, overall, the combination of two or more feature categories would make a better prediction than just one feature category alone. According to AUC metric, prediction models using ANN classifier with combined journal and author features in year 6 after publication produced the highest AUC scores in marketing, whereas in MIS, the overall highest AUC scores were generated by the LGR classifier with all three features included during all six years after publication. Therefore, keyword popularity provided valuable information in making prediction models even more precise, at least to a certain extent.

Our experimental results may provide additional evidence to echo the findings by Choi, Yi, and Lee (2011) who analyzed the characteristics of MIS journal papers and observed that because MIS is regarded as interdisciplinary, the rapid changes in this domain often introduce many new topics and concepts. As a result, MIS researchers follow these new topics in their research, and thus keyword popularity is likely to induce a high citation count from other researchers. Comparing the results between marketing and MIS, the overall evaluation using AUC showed that in addition to journal and author features, KP features can provide more information to make predictions more precise in MIS but not in marketing. Therefore, the present study reveals that MIS papers are probably influenced more by keyword popularity than are marketing papers.

If detecting frequently cited papers is the most important task, F-measure is thus more appropriate than AUC in evaluating the overall performance. Accordingly, the experiment results reveal that using both journal and author features is likely the most effective in marketing and using journal-related features only is probably the most effective in MIS. However, the largest F-measure was 0.738 using journal-related features only in marketing and 0.780 using both journal and author features in MIS. Our findings imply that journal-related features would still play the crucial role in prediction; however, author and KP features can provide additional information in prediction performance.

Finally, citation years after publication may also affect evaluation results. The discipline matters, for example, the social sciences take much longer to be recognized and cited than in biomedical fields (Glänzel & Schoepflin, 1995). According to the overall evaluation metrics, AUC and F-measure, predicting performance of six years after publication (Y + 6) produced the highest scores except for the performance of predicting highly-cited papers in marketing (0.738 in Y + 4). Since a short time window of one or two years may provide unfair evaluation in some fields of research (Wang, 2013), our results also reveal that longer years after publication (say 5-6 years) can better predict future citations than shorter years (e.g., 1-3 years). It is also worth noting that marketing and MIS are not covered in the related literature, the current paper provides additional evidence in predicting future citations.

5. Conclusion

With the growth of academic research development and increasingly rapid and convenient Internet access, Internet (online) search results are prone to yield nonessential information or academic papers with less reference value. This phenomenon highlights the importance of locating influential papers. The present study pioneers the investigation of keyword popularity to gain an understanding of its influence on predicting papers that are highly or less highly cited. We selected articles from both marketing and MIS disciplines to examine whether or not these KP features can improve the power of identification of highly-cited papers. Specifically, this study compared keyword popularities of articles with other commonly used journal-related and author-related features proposed in previous studies and analyzed the influence of each category of features on citation prediction of academic papers. Binary classification models were estimated by using supervised classification techniques to compare the predictive power of identifying highly-cited papers for each model.

Our evaluation results show that, with KP features, the prediction models are more effective than those with journal and/or author features, especially in the MIS discipline. Therefore, keyword popularity features can improve the effectiveness of the prediction models, which implies that consideration of article keywords can be of help to correctly detect citation counts of papers if a research is more interdisciplinary.

To further enhance the effectiveness of the prediction models, future studies are recommended to expand the experimental dataset for assessment and prediction or to examine other journal disciplines. Thresholds for identifying highly-cited papers may also be expanded to the top 10%, 5%, or 1% for comparison purposes. Other computational methods that reinforce the keyword popularities of articles and alternative data mining techniques may be considered to obtain more favorable research outcomes. Finally, it is worth noting that, based on the information recorded in WOS, all of the first authors in our selected journals are also the corresponding authors. This situation does not suit all other journals in most disciplines. Researchers must be cautious on this issue when author features are considered in future studies.

Author contributions

Ya-Han Hu: Conceived and designed the analysis; Contributed data or analysis tools; Performed the analysis; Wrote the paper.
Chun-Tien Tai: Conceived and designed the analysis; Contributed data or analysis tools; Performed the analysis; Wrote the paper.
Kang Ernest Liu: Conceived and designed the analysis; Wrote the paper.
Cheng-Fang Cai: Collected the data; Contributed data or analysis tools; Performed the analysis.

Acknowledgements

This research was supported in part by the Ministry of Science and Technology of the Republic of China (grant number MOST 107-2410-H-194 -054 -MY2) and the Center for Innovative Research on Aging Society from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

Appendix A.

Table A1
Classification performance results using AUC: top 33%.

Field of Research			Feature set						
	Classifier	Period	J	A	KP	J+A	J+KP	A+KP	J+A+KP
marketing	C4.5	Y+1	0.559	0.503	0.506	0.540	0.530	0.519	0.547
		Y+2	0.675	0.523	0.494	0.604	0.598	0.548	0.603
		Y+3	0.723	0.564	**0.527**	0.604	0.610	0.537	0.611
		Y+4	0.734	0.543	0.441	**0.675**	0.579	**0.596**	**0.688**
		Y+5	**0.736**	0.585	0.497	0.654	**0.623**	0.576	0.665
		Y+6	0.735	**0.617**	0.500	0.629	0.568	0.509	0.678
	LGR	Y+1	0.578	0.608	0.474	0.640	0.582	0.587	0.642
		Y+2	0.655	**0.667**	0.534	0.711	0.647	0.652	0.707
		Y+3	0.689	0.661	0.497	0.733	0.685	0.624	0.726
		Y+4	0.706	0.662	0.487	**0.759**	0.692	0.636	**0.746**
		Y+5	0.705	0.663	0.527	0.746	0.695	**0.658**	**0.746**
		Y+6	**0.711**	0.636	**0.567**	0.736	**0.702**	0.627	0.741
	SMO	Y+1	0.497	0.620	0.434	0.592	0.533	0.562	0.579
		Y+2	0.611	0.656	0.438	0.651	0.622	**0.613**	0.668
		Y+3	0.703	0.662	0.428	0.707	0.674	0.604	0.712
		Y+4	0.695	**0.666**	0.372	**0.719**	0.705	0.568	0.745
		Y+5	0.708	0.653	**0.473**	0.714	**0.722**	0.600	**0.750**
		Y+6	**0.717**	0.610	0.471	0.717	0.704	0.563	0.741
	ANN	Y+1	0.579	0.590	0.541	0.642	0.571	0.578	0.643
		Y+2	0.690	0.659	**0.561**	0.727	0.672	**0.657**	0.733
		Y+3	0.719	**0.680**	0.478	0.755	0.674	0.596	0.732
		Y+4	0.734	0.658	0.471	**0.782**	0.690	0.586	0.757
		Y+5	0.738	0.670	0.532	0.774	**0.739**	0.631	**0.784**
		Y+6	**0.750**	0.630	0.520	0.778	0.729	0.573	0.772
MIS	C4.5	Y+1	**0.786**	0.500	0.594	0.752	0.684	0.570	0.674
		Y+2	0.784	0.515	0.566	0.720	**0.734**	0.547	0.686
		Y+3	**0.786**	**0.612**	0.586	**0.764**	0.697	**0.643**	**0.703**
		Y+4	0.751	0.571	0.562	0.674	0.690	0610	0.683
		Y+5	0.754	0.603	0.525	0.691	0.636	0.582	0.688
		Y+6	0.758	0.536	**0.610**	0.696	0.602	0.571	0.650
	LGR	Y+1	0.778	0.647	0.634	0.829	0.814	0.682	0.840
		Y+2	**0.791**	0.651	0.637	**0.846**	**0.823**	0.690	**0.854**
		Y+3	0.783	0.646	0.632	0.836	0.814	0.679	0.838
		Y+4	0.758	**0.677**	0.642	0.832	0.790	**0.723**	0.840
		Y+5	0.754	0.667	**0.645**	0.819	0.796	0.711	0.828
		Y+6	0.750	0.656	0.638	0.814	0.791	0.699	0.826
	SMO	Y+1	0.746	0.607	0.592	0.767	0.776	0.619	0.796
		Y+2	0.740	**0.642**	0.588	0.771	**0.793**	0.668	**0.808**
		Y+3	**0.747**	0.620	0.618	0.778	0.781	0.654	0.782
		Y+4	0.746	0.638	0.614	**0.784**	0.756	0.674	0.777
		Y+5	0.717	0.626	**0.627**	0.763	0.740	**0.684**	0.771
		Y+6	0.723	0.613	0.611	0.758	0.745	0.676	0.764

Table A1 (*Continued*)

Field of Research	**Classifier**	**Period**	**Feature set**						
			J	**A**	**KP**	**J+A**	**J+KP**	**A+KP**	**J+A+KP**
	ANN	Y+1	**0.780**	0.638	0.600	0.826	0.798	0.630	0.821
		Y+2	0.775	0.643	0.627	**0.838**	**0.801**	0.681	**0.830**
		Y+3	0.775	0.636	**0.638**	**0.838**	0.793	0.677	0.820
		Y+4	0.754	**0.671**	0.634	0.827	0.762	**0.709**	0.815
		Y+5	0.748	0.660	0.622	0.821	0.761	0.686	0.815
		Y+6	0.749	0.646	0.616	0.814	0.760	0.676	0.800

Note: MIS: management information system; LGR: logistic regression; SMO: support vector machine; ANN: artificial neural network; Y+*t* (*t* = 1,2,. . .,6): the period *t* to be evaluated after published in year Y; J: journal feature set; A: author feature set; KP: keyword popularity feature set.

Table A2
Classification performance results using AUC: top 50%.

Field of Research	**Classifier**	**Period**	**Feature set**						
			J	**A**	**KP**	**J+A**	**J+KP**	**A+KP**	**J+A+KP**
Marketing	**C4.5**	Y+1	0.598	0.510	0.488	0.547	0.518	0.472	0.531
		Y+2	0.666	0.518	0.495	0.566	0.547	**0.610**	0.578
		Y+3	0.683	0.552	**0.509**	0.590	0.564	0.551	**0.605**
		Y+4	0.682	**0.565**	0.436	0.607	0.567	0.497	0.578
		Y+5	0.678	0.544	0.495	0.585	0.581	0.569	0.585
		Y+6	**0.702**	0.545	0.470	**0.626**	**0.620**	0.572	0.547
	LGR	Y+1	0.593	0.622	0.493	0.663	0.594	**0.613**	0.667
		Y+2	0.624	0.583	0.544	0.658	0.652	0.579	0.666
		Y+3	0.662	0.611	**0.549**	0.693	**0.668**	0.603	0.691
		Y+4	0.663	**0.621**	0.531	**0.704**	0.659	0.610	**0.696**
		Y+5	0.669	0.595	0.543	0.684	0.667	0.603	0.686
		Y+6	**0.673**	0.596	0.535	0.699	0.667	0.590	**0.696**
	SMO	Y+1	0.576	0.538	**0.550**	0.604	0.547	0.499	0.591
		Y+2	0.636	0.540	0.476	0.660	0.656	0.487	0.661
		Y+3	0.647	0.544	0.468	0.684	**0.681**	0.564	0.704
		Y+4	0.647	**0.548**	0.468	**0.694**	0.673	**0.575**	**0.709**
		Y+5	**0.672**	0.511	0.438	0.687	0.655	0.527	0.666
		Y+6	**0.672**	0.550	0.460	0.688	0.670	0.554	0.686
	ANN	Y+1	0.617	**0.633**	0.434	0.655	0.574	0.560	0.618
		Y+2	0.671	0.595	0.504	0.669	0.646	0.555	0.667
		Y+3	0.681	0.622	0.521	0.697	0.668	0.600	0.693
		Y+4	0.681	0.650	0.512	**0.719**	0.668	**0.606**	**0.719**
		Y+5	0.683	0.599	**0.536**	0.692	0.659	0.586	0.666
		Y+6	**0.702**	0.591	0.515	0.709	**0.672**	0.557	0.691
MIS	**C4.5**	Y+1	**0.727**	0.481	0.521	**0.660**	0.579	**0.597**	**0.663**
		Y+2	0.672	**0.592**	0.545	0.586	0.586	0.518	0.588
		Y+3	0.665	0.535	0.505	0.619	**0.588**	0.493	0.606
		Y+4	0.678	0.520	0.507	0.639	0.557	0.553	0.585
		Y+5	0.675	0.561	**0.559**	0.591	0.566	0.537	0.571
		Y+6	0.679	0.540	0.530	0.630	0.568	0.552	0.570
	LGR	Y+1	**0.720**	0.615	0.602	**0.767**	**0.749**	0.642	**0.778**
		Y+2	0.697	0.594	0.592	0.730	0.713	0.634	0.742
		Y+3	0.691	0.607	**0.617**	0.743	0.718	0.648	0.754
		Y+4	0.695	0.623	0.614	0.750	0.728	**0.658**	0.764
		Y+5	0.689	0.612	0.600	0.749	0.720	0.647	0.762
		Y+6	0.685	**0.624**	0.594	0.750	0.714	0.644	0.753
	SMO	Y+1	**0.718**	**0.477**	0.487	**0.691**	**0.725**	0.519	**0.706**
		Y+2	0.708	0.428	0.476	0.614	0.647	0.467	0.674
		Y+3	0.695	0.440	0.501	0.650	0.645	0.454	0.667
		Y+4	0.694	0.461	0.499	0.646	0.663	0.469	0.680
		Y+5	0.698	0.458	0.476	0.678	0.680	**0.566**	0.692
		Y+6	0.703	0.467	**0.505**	0.688	0.649	0.521	0.678
	ANN	Y+1	0.681	0.604	0.579	0.739	**0.705**	0.627	0.740
		Y+2	0.712	0.577	0.549	0.726	0.679	0.594	0.707
		Y+3	0.689	0.605	0.588	0.717	0.700	0.636	0.723
		Y+4	**0.713**	0.599	**0.606**	**0.757**	0.697	**0.645**	**0.741**
		Y+5	0.698	0.605	0.561	0.739	0.699	0.636	**0.741**
		Y+6	0.693	**0.609**	0.560	0.748	0.699	0.636	0.728

Note: MIS: management information system; LGR: logistic regression; SMO: support vector machine; ANN: artificial neural network; Y+*t* (*t* = 1,2,. . .,6): the period *t* to be evaluated after published in year Y; J: journal feature set; A: author feature set; KP: keyword popularity feature set.

References

Abrishami, A., & Aliakbary, S. (2019). Predicting citation counts based on deep neural network learning techniques. *Journal of Informetrics*, *13*(2), 485–499.

Acuna, D. E., Allesina, S., & Kording, K. P. (2012). Future impact: Predicting scientific success. *Nature*, *489*(7415), 201.

Bai, X., Zhang, F., & Lee, I. (2019). Predicting the citations of scholarly paper. *Journal of Informetrics*, *13*(1), 407–418.

Beliakov, G., & James, S. (2011). Citation-based journal ranks: the use of fuzzy measures. *Fuzzy Sets and Systems*, *167*(1), 101–119.

Blei, D., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Bornmann, L., Leydesdorff, L., & Wang, J. (2014). How to improve the prediction based on citation impact percentiles for years shortly after the publication date? *Journal of Informetrics*, *8*(1), 175–180.

Chang, Y., Huang, M., & Lin, C. (2015). Evolution of research subjects in library and information science based on keyword, bibliographical coupling, and co-citation analyses. *Scientometrics*, *105*(3), 2071–2087.

Chen, P., & Lin, S. (2010). Automatic keyword prediction using google similarity distance. *Expert Systems with Applications*, *37*(3), 1928–1938.

Choi, J., Yi, S., & Lee, K. (2011). Analysis of keyword networks in MIS research and implications for predicting knowledge evolution. *Information & Management*, *48*(8), 371–381.

Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., & Štajdohar, M. (2013). Orange: data mining toolbox in Python. *The Journal of Machine Learning Research*, *14*(1), 2349–2353.

Dorta-González, P., Dorta-González, M. I., Santos-Peñate, D. R., & Suárez-Vega, R. (2014). Journal topic citation potential and between-field comparisons: the topic normalized impact factor. *Journal of Informetrics*, *8*(2), 406–418.

Finardi, U. (2014). On the time evolution of received citations, in different scientific fields: An empirical study. *Journal of Informetrics*, *8*(1), 13–24.

Fu, L., & Aliferis, C. (2010). Using content-based and bibliometric features for machine learning models to predict citation counts in the biomedical literature. *Scientometrics*, *85*(1), 257–270.

Glänzel, W., & Schoepflin, U. (1995). A bibliometric study on ageing and reception processes of scientific literature. *Journal of Information Science*, *21*(1), 37–53.

Hosmer, D., & Lemeshow, S. (2004). *Applied logistic Regression*. New York, NY: John Wiley & Sons.

Huang, T. Y., & Zhao, B. (2019). Measuring popularity of ecological topics in a temporal dynamical knowledge network. *PloS ONE*, *14*(1), e0208370 http://dx.doi.org/10.1371/journal.pone.0208370

Iqbal, W., Qadir, J., Tyson, G., Mian, A. N., Hassan, S. U., & Crowcroft, J. (2019). A bibliometric analysis of publications in computer networking research. *Scientometrics*, *119*(2), 1121–1155.

Jiang, H., Qiang, M., & Lin, P. (2016). Finding academic concerns of the Three Gorges Project based on a topic modeling approach. *Ecological Indicators*, *60*, 693–701.

Kar, M., Nunes, S., & Ribeiro, C. (2015). Summarization of changes in dynamic text collections using Latent Dirichlet Allocation model. *Information Processing and Management*, *50*(6), 809–833.

Kim, H., Ha, I., Lee, K., Jo, G., & El-Saddik, A. (2011). Collaborative user modeling for enhanced content filtering in recommender systems. *Decision Support Systems*, *51*(4), 772–781.

Kosteas, V. (2018). Predicting long-run citation counts for articles in top economics journals. *Scientometrics*, *115*(3), 1395–1412.

Kretschmann, E., Fleischmann, W., & Apweiler, R. (2001). Automatic rule generation for protein annotation with the C4.5 data mining algorithm applied on SWISS-PROT. *Bioinformatics*, *17*(10), 920–926.

Lemeshow, S., & Hosmer, D. (1998). Logistic regression. In *Encyclopedia of biostatistics*.

Leydesdorff, L., & Bornmann, L. (2011). How fractional counting of citations affects the Impact Factor: Normalization in terms of differences in citation potentials among fields of science. *Journal of the American Society for Information Science and Technology*, *62*(2), 217–229.

Li, L. L., Ding, G., Feng, N., Wang, M. H., & Ho, Y. S. (2009). Global stem cell research trend: Bibliometric analysis as a tool for mapping of trends from 1991 to 2006. *Scientometrics*, *80*(1), 39–58.

Liang, T., Yang, Y., Chen, D., & Ku, Y. (2008). A semantic-expansion approach to personalized knowledge recommendation. *Decision Support Systems*, *45*(3), 401–412.

Liu, J., Tian, J., Kong, X., Lee, I., & Xia, F. (2019). Two decades of information systems: a bibliometric review. *Scientometrics*, *118*(2), 617–643.

Natale, F., Fiore, G., & Hofherr, J. (2012). Mapping the research on aquaculture. A bibliometric analysis of aquaculture literature. *Scientometrics*, *90*(3), 983–999.

Newman, M. (2014). Prediction of highly cited papers. *EPL (Europhysics Letters)*, *105*(2), 28002.

Pan, C., & Li, W. (2010). Research paper recommendation with topic analysis. *Computer Design and Applications*, *4*, 264–268.

Phan, X., & Nguyen, C. (2008). *A Java implementation of Latent Dirichlet Allocation (LDA) using Gibbs sampling for parameter estimation and inference. JGibbLDA*. http://jgibblda.sourceforge.net

Porter, M. (1980). An algorithm for suffix stripping. *Program*, *14*(3), 130–137.

Quinlan, J. R. (2014). *C4. 5: programs for machine learning*. Elsevier.

Rodríguez-Bolívar, M., Alcaide-Muñoz, L., & Cobo, M. (2018). Analyzing the scientific evolution and impact of e-Participation research in JCR journals using science mapping. *International Journal of Information Management*, *40*, 111–119.

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533.

Sohrabi, B., & Iraj, H. (2017). The effect of keyword repetition in abstract and keyword frequency per journal in predicting citation counts. *Scientometrics*, *110*(1), 243–251.

Song, M., & Ding, Y. (2014). Topic Modeling: Measuring Scholarly Impact Using a Topical Lens. In *Measuring Scholarly Impact*. pp. 235–257. Cham: Springer.

Stegehuis, C., Litvak, N., & Waltman, L. (2015). Predicting the long-term citation impact of recent publications. *Journal of informetrics*, *9*(3), 642–657.

Tian, Y., Wen, C., & Hong, S. (2008). Global scientific production on GIS research by bibliometric analysis from 1997 to 2006. *Journal of Informetrics*, *2*(1), 65–74.

Tsai, C. (2014). Citation impact analysis of top ranked computer science journals and their rankings. *Journal of Informetrics*, *8*(2), 318–328.

Uddin, S., & Khan, A. (2016). The impact of author-selected keywords on citation counts. *Journal of Informetrics*, *10*(4), 1166–1177.

Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.

Wang, D., Song, C., & Barabási, A. (2013). Quantifying long-term scientific impact. *Science*, *342*(6154), 127–132.

Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, *94*(3), 851–872.

Wang, M., Wang, Z., & Chen, G. (2019). Which can better predict the future success of articles? Bibliometric indices or alternative metrics. *Scientometrics*, *119*(3), 1575–1595.

Wang, M., Yu, G., Xu, J., He, H., Yu, D., & An, S. (2012). Development a case-based classifier for predicting highly cited papers. *Journal of Informetrics*, *6*(4), 586–599.

Yin, J., Gong, L., & Wang, S. (2018). Large-scale assessment of global green innovation research trends from 1981 to 2016: A bibliometric study. *Journal of Cleaner Production*, *197*, 827–841.

Zhang, X., Wang, X., Chen, J., Xie, X., Wang, K., & Wei, Y. (2014). A novel modeling based real option approach for CCS investment evaluation under multiple uncertainties. *Applied Energy*, *113*, 1059–1067.

Zhang, Y., Lu, J., Liu, F., Liu, Q., Porter, A., Chen, H., et al. (2018). Does deep learning help topic extraction? A kernel k-means clustering method with word embedding. *Journal of Informetrics*, *12*(4), 1099–1117.

Zhang, Y., Zhang, B., Gao, K., Guo, P., & Sun, D. (2012). Combining content and relation analysis for recommendation in social tagging systems. *Physica A: Statistical Mechanics and its Applications*, *391*(22), 5759–5768.
Zheng, H., Kang, B., & Kim, H. (2009). Exploiting noun phrases and semantic relationships for text document clustering. *Information Sciences*, *179*(13), 2249–2262.
Zhou, X., & Zhao, G. (2015). Global liposome research in the period of 1995–2014: a bibliometric analysis. *Scientometrics*, *105*(1), 231–248.