Journal of Informetrics 14 (2020) 101039

Contents lists available at ScienceDirect

Journal of Informetrics

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

Predicting the citation counts of individual papers via a BP neural network

Xuanmin Ruan[a], Yuanyang Zhu[b], Jiang Li[a], Ying Cheng[a,*]

[a] School of Information Management, Nanjing University, Nanjing, 210023, China
[b] School of Management and Engineering, Nanjing University, Nanjing, 210023, China

ARTICLE INFO

Article history:
Received 18 October 2019
Received in revised form 24 March 2020
Accepted 25 March 2020
Available online 27 April 2020

Keywords:
Citation prediction
Neural network
XGBoost
Linear regression

ABSTRACT

Predicting the citation counts of academic papers is of considerable significance to scientific evaluation. This study used a four-layer Back Propagation (BP) neural network model to predict the five-year citations of 49,834 papers in the library, information and documentation field indexed by the CSSCI database and published from 2000 to 2013. We extracted six paper features, two journal features, nine author features, eight reference features, and five early citation features to make the prediction. The empirical experiments showed that the performance of the BP neural network is significantly better than those of the six baseline models. In terms of the prediction effect, the accuracy of the model at predicting infrequently cited papers was higher than that for frequently cited ones. We determined that five essential features have significant effects on the prediction performance of the model, i.e., 'citations in the first two years', 'first-cited age', 'paper length', 'month of publication', and 'self-citations of journals', and the other features contribute only slightly to the prediction.

1. Introduction

It is widely accepted that academic papers are carriers of knowledge, and their impacts are of significant importance to scientific evaluation, especially for such outcomes as promotion, recruitment decisions, and grant allocation (Bai, Zhang, & Lee, 2019; Li, Lin, Yan, & Yeh, 2015). Their impact also plays a crucial role for researchers when selecting high-quality papers from databases. Among a variety of impact metrics, the citation count is acknowledged as the primary one (Didegah & Thelwall, 2013b) and is considered to be one of the most standard, objective and straightforward measurements (Yan, Huang, Tang, Zhang, & Li, 2012). However, the accumulation of citations takes time (Fu & Aliferis, 2010). Thus, it is valuable to predict the citation counts of academic papers shortly after their publication.

Previously, the prediction of citations was conducted using two different approaches. One is classification approaches, where machine learning methods, such as decision trees (Wang, Yu, & Yu, 2011) and SVM (Fu & Aliferis, 2010), were applied to classify papers according to the magnitude of their citation counts. The other is regression approaches, where linear regressions (Bornmann, Leydesdorff, & Wang, 2014; Lokker, McKibbon, McKinlay, Wilczynski, & Haynes, 2008; Yan et al., 2012) and negative binomial regressions (Onodera & Yoshikane, 2015) were widely used for citation prediction. In addition, though previous studies have found that many factors are significantly associated with the citation counts of scientific papers

* Corresponding author.
E-mail address: chengy@nju.edu.cn (Y. Cheng).

https://doi.org/10.1016/j.joi.2020.101039

(Tahamtan, Afshar, & Ahamdzadeh, 2016), which factors are relatively essential to the prediction of citation counts has not been determined.

Recently, the resurgence of artificial neural networks, especially the progress in deep neural network techniques, has provided a new method for citation prediction. However, few studies focus on the use of deep neural network techniques in the prediction of citation counts, and their performance needs to be further examined. The Back Propagation (BP) neural network, which is one of the most common artificial neural networks, has been widely used for regression and prediction in many fields, outperforming other models, such as linear regression models (Geng, Bose, & Chen, 2015; Lee & Choeh, 2014; Mai, Tian, Lee, & Ma, 2019; Wong & Chan, 2015; Wong, Chan, & Lacka, 2017). Hence, we aim to improve the citation prediction accuracy by using a BP neural network and find the features that play essential roles in the prediction. The research questions are as follows.

1) How does the BP neural network perform at predicting the citations of academic papers?
2) Does the BP neural network have better performance than the baseline models (e.g., a linear regression)?
3) What is the comparative importance of the selected features for citation prediction?

We use a multilayer BP neural network to predict the citations of academic papers. First, we select 49,834 papers in the library, information and documentation field published from 2000 to 2013 and indexed in the Chinese Social Science Citation Index database (hereafter CSSCI) (Su, Deng, & Shen, 2014). Second, we extract six article features, two journal features, eight reference features, nine author features, and five early citation features from the database. Finally, a four-layer BP neural network model is applied to predict the five-year citations.

2. Related work

2.1. Classification approach for citation prediction

Some researchers have regarded citation prediction as a classification problem and used machine learning methods to predict the classifications of papers based on citations. For example, Fu and Aliferis (2008; 2010) used the Support Vector Machine (SVM) to predict whether the ten-year citations of a paper would be higher than a certain number (e.g., 20, 50, 100 or 500) in the biomedical field. The prediction AUC was from 0.857 to 0.918. Ibanez, Larranaga, and Bielza (2009) divided papers into three categories, including those with few citations (one or less), some citations (2–4), and many citations (more than four). Machine learning methods, such as the Naive Bayes, logistic regression, decision trees, and K-Nearest Neighbors (KNN), were adopted to predict the citations from the first year to the first four years. The results showed that the logistic regression and Naive Bayes achieved the highest accuracy. Wang et al. (2011) divided 219 papers into high, medium, and low groups in the field of astronomy and astrophysics. These researchers used a multiclassifier system consisting of five decision trees classifiers to make the prediction and obtained high classification accuracy. Wang, Fan, Zeng, and Di (2019) used a neural network model to predict the highly cited papers of the ESI and achieved a satisfactory result. Furthermore, some researchers developed classification criteria based on the contribution that a paper made to their author's *h*-index. For example, in the study of Dong, Johnson, and Chawla (2015), papers whose citation counts were higher than the author's *h*-index were classified into the positive class, and vice versa. These authors used a logistic regression, a Random Forest (RF), and bagged decision trees in their research and found that the content features were the most important predictors followed by the journal features.

These classification approaches are concise in their operations and often have high classification accuracy, which meets the requirements in some tasks such as identifying frequently cited papers. However, these coarse-grained approaches have the following shortcomings. First, there is no scientific standard for the classification boundaries, which leads to the inability to compare different studies and restrains the application of the research results. Second, a classification result is a simplified treatment of the citations without an in-depth analysis of the citation patterns or trajectories.

2.2. Regression approach for citation prediction

The linear regression model is one of the most commonly used methods. For example, Lokker et al. (2008) used 17 reference-related and three journal-related features to predict the two-year citations of clinical papers. The results showed that the R^2 for the training set was 0.60, and that for the test set was 0.56. According to the regression coefficients, Yu, Yu, Li, and Wang (2014) found the six most essential features for citation prediction, including the citations in the first two years, the number of references, the IF-5 (five-year impact factor), the reciprocal of the first-cited age, the number of authors, and the total citations for the first author. Bornmnn et al. (2014) predicted the percentile of citations of a paper in the 31st year after publication using the percentiles in the previous years. According to the changes in the R^2, these researchers found that the Journal Impact Factor (JIF) and the citations in the first two years were critical. Abramo et al. (2019) used the citations in the early stage and the JIF to predict the citations after ten years. These researchers found that the prediction accuracy was relatively high by using the citations in the first three years only. As the citation window increases, the role of the JIF gradually decreased until it was almost negligible. Since the number of citations is non-negative, its distribution is skewed, and its variance is usually more substantial than the mean (Didegah & Thelwall, 2013b; Onodera & Yoshikane, 2015). The negative binomial regression model has become another popular model. For example, Onodera and Yoshikane (2015) predicted the citations received by an article in the sixth and eleventh years after publication in six selected subject fields. The Price index and the number of references are essential predictors. The value of the pseudo-R^2 ranged from 0.23 to 0.54. In addition to the models above, some machine learning methods were also used to solve the problem. For example, Bai et al. (2019) used the Gradient Boosting Decision Trees (GBDT) model in the prediction. Chakraborty, Kumar, Goyal, Ganguly, and Mukherjee (2014) used the Support Vector Regression (SVR) for the five-year citation prediction task. The R^2 of the model was 0.71, and the Mean Squared Error (MSE) was 4.08. Li et al. (2015) also used the SVR to predict the total citations in the 10th, 11th, and 12th years, respectively, in two different ways. One was to predict the citations directly, and the other was to develop a prediction model after leveraging the citation count trend of a paper. The R^2 of the model ranged from 0.67 to 0.68 in the latter. Yan et al. (2012) used the linear regression, KNN, SVR, and Gaussian Process Regression (GRP) models to predict the citations in the first, fifth, and tenth years after publication, respectively. The results showed that the performance of nonlinear algorithms (e.g., SVR and GRP) was better than that of the linear ones. The R^2 of the GRP model in predicting five-year citations was as high as 0.869. Robson and Mousques (2016) used the RF model to predict citations in the field of environmental modeling. The selected features can only predict a small percentage of the variations in citations (less than 30 %).

In addition, Wang, Song, and Barabasi (2013) derived a mechanistic model for the citation dynamics based on three fundamental mechanisms, namely, preferential attachment, aging, and fitness. The model predicted the future citations of an individual paper accurately, and its performance was better than those of basic models, such as the logistic model. Abrishami and Aliakbary (2019) considered the issue as a 'sequence-to-sequence' problem and used the RNN to predict future citation 'sequences' based on the citation 'sequences' in the first few years after publication. Xu, Li, Jiang, Ge, and Cai (2019) used the features of heterogeneous bibliographic networks and a Convolutional Neural Network (CNN) to predict the ten-year citation counts of an individual paper and improved the prediction accuracy by 5 % compared with the baseline models.

2.3. BP neural network

The BP neural network was proposed by Rumelhart, Hinton, and Williams (1986) and is one of the most widely used artificial neural network models in prediction tasks. The application fields of this network include finance (Mai et al., 2019), accounting (Geng et al., 2015), management science (Wong & Chan, 2015), e-commerce (Lee & Choeh, 2014), and computer science (Wong et al., 2017). The artificial neural network has also been extensively applied in bibliometrics. For example, Rokach, Kalech, Blank, and Stern (2011) combined the outputs of three learning techniques (logistic regressions, decision trees, and artificial neural networks) to predict the next AAAI fellowship winners in the field of artificial intelligence. Zhang, Yuan, Chang, and Ken (2012) used patent-related features to predict the performance of an enterprise via a three-layer neural network, which achieved an excellent performance. Dang and Ignat (2016) used the Doc2Vec technique and a BP neural network with four hidden layers to evaluate the quality of Wikipedia articles. The performance of the BP neural network was better than those of the KNN and the Classification and Regression Tree. A variety of deep neural networks, including the CNN, LSTM, CNN-LSTM, multilayer LSTM (stacked LSTM), and deep neural network, were used by Wang and Li (2019) to evaluate the quality of Wikipedia articles. The experimental results showed that the latter presented outstanding accuracy and training speed.

Unlike the linear regression model, the BP neural network has no strict requirements for the data distribution. The prediction results of this network are usually robust. In addition, the performance of shallow machine learning models (e.g., the SVM and LR) depends on the feature engineering quality. However, it is not easy for human experts to design useful features. By contrast, the deep neural network has an advantage in its feature learning, i.e., it can automatically transform the initial "bottom" feature representation into a "high-level" feature through a multilevel and nonlinear transformation. A fully connected feed-forward neural network can approximate any continuous function at any desired level of precision (Hornik, Stinchcombe, & White, 1989). Meanwhile, several studies have confirmed that the BP neural network has the advantage in prediction. For example, Wong and Chan (2015) found that the performance of the BP neural network was significantly better than those of the linear regression and SVR. Lee and Choeh (2014) and Wong et al. (2017) also found that the prediction performance of the BP neural network model was better than that of the linear regression model. Furthermore, the features used in our study are neither a sequence form value nor multidimensional data. Thus, network structures, such as the CNN and RNN, seem to not be suitable for our study. Therefore, this paper conducts an empirical study by applying the BP neural network to citation prediction.

3. Data and methodology

3.1. Dataset

The CSSCI is one of the essential Chinese citation index databases for 28 disciplines in social sciences and humanities. This study selected the reviews and research articles published from 2000 to 2013 in the library, information and documentation field from the CSSCI. Since the journals indexed in the database are updated every two years, we selected 15 everlasting journals to ensure the homogeneity of the features. The publication data before 2013 and the citation data before 2018 were

Table 1
Description of features and prediction target.

NO.	Feature	Description	Category
x_1	Document type	Document type of the article: research article (0) or review (1)	
x_2	Paper length	Number of pages in the article	**Paper features**
x_3	Title length	Number of words in the article title	
x_4	Funding	The funding an article received: national funding, provincial funding, other funding, and no funding	
x_5	Month of publication	The publication month of the paper	
x_6	Punctuation in the title	Has punctuation in the article title: Yes (1) or No (0)	
x_7	Journal Impact Factor	The total citations received by the papers published in the first two years of the journal divided by the number of papers published in the first two years of the journal	**Journal features**
x_8	Number of publications in the journal	The number of papers in the journal in a year	
x_9	Number of authors	The number of authors for the paper	
x_{10}	Productivity	The number of papers published by the first author before the current one	
x_{11}	Previous citations	The number of citations received by the first author before the current paper	
x_{12}	Previous citations per article	The average citations per paper published by the first author before the current article	**Author features**
x_{13}	*h*-index	*h*-index of the first author before the current article	
x_{14}	Number of organizations	The number of institutions for the paper	
x_{15}	Total number of papers produced in the institution	Total number of papers published in the institution of the first author	
x_{16}	Impact of faculty members—max *h*-index	The max *h*-index among the authors affiliated with the institution of the first author	
x_{17}	Impact of faculty members—average *h*-index	The average *h*-index among the authors affiliated with the institution of the first author	
x_{18}	Number of references	The number of references in the paper	
x_{19}	Reference age	The average age of the references in the paper	
x_{20}	Self-citations of journals	Percentage of the reference papers published in the journal in which the current article was published	
x_{21}	Percentage of journal articles	Percentage of journal articles in the reference list	**Reference features**
x_{22}	Percentage of conference papers	Percentage of conference papers in the reference list	
x_{23}	Percentage of monographs	Percentage of monographs in the reference list	
x_{24}	Percentage of online resources	Percentage of online resources in the reference list	
x_{25}	Percentage of dissertations	Percentage of dissertations in the reference list	
x_{26}	First-cited age	Reciprocal of the years between the paper's publication year and its first citation year	
x_{27}	Citations in the first year	The number of citations in the first year after publication.	**Early citation features**
x_{28}	Citations in the first two years	The number of citations in the first two years after publication.	
x_{29}	No. of citing journals in the first year	The number of citing journals in the first year after publication	
x_{30}	No. of citing journals in the first two years	The number of citing journals in the first two years after publication	
Y	Five-year citations	The number of citations in the first five years after publication.	**Prediction target**

selected because we used a five-year citation window. In addition, we could not extract the author-related features from papers without the 'first author' and 'author institution' fields; therefore, we excluded them. Finally, 49,834 papers were obtained and randomly divided into training and test sets at a ratio of 4:1.

3.2. Features and predicted values

Previous studies showed that the number of citations of a paper in the five years after its publication was a critical reflection of its quality (Wang et al., 2011; Wang, Yu, An, & Yu, 2012) and a five-year interval was sufficient for a distinct polarization pattern to occur (Aksnes, 2003). Therefore, we used a five-year citation window in the prediction experiment.

There are abundant studies of the factors affecting citation counts. For example, Tahamtan et al. (2016) systematically reviewed these studies and summarized 28 related factors. Xie et al. (2019) discussed the influence of 66 factors on the citations of Chinese papers. We selected the features mentioned in Tahamtan et al. (2016) and Xie et al. (2019) if they met the following criteria: (1) no manual coding, and (2) available from the CSSCI database. Based on the two criteria, we obtained 30 features and listed them in Table 1.

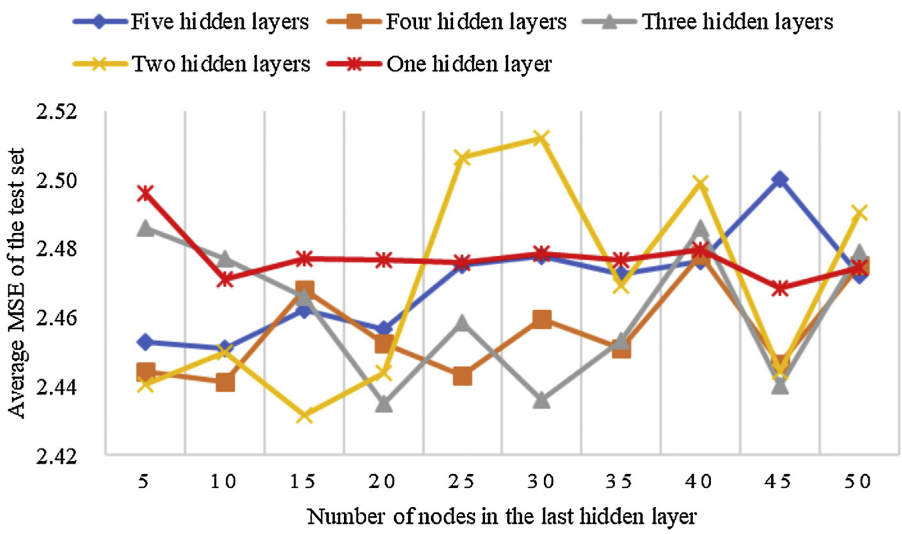

Fig. 1. Average MSE of the fivefold cross-validation of the models with different numbers of hidden layers and hidden units. The curves are colored based on the number of hidden layers. In the two-hidden-layer models, the number of hidden units in the first hidden layer is fixed as 15. In the three-hidden-layer models, the numbers of hidden units in the first two hidden layers are fixed as 15 and 45, respectively. In the four-hidden-layer models, the numbers of hidden units in the first three hidden layers are fixed as 15, 45, and 30, respectively. In the five-hidden-layer models, the numbers of hidden units in the first four hidden layers are fixed as 15, 45, 30, and 15, respectively.

1) Paper features. The features of a paper exert an essential influence on citations. Many studies have shown that frequently cited papers were characterized by short titles, titles with punctuation, long pages, being published at the beginning of the year, or being financially supported by research funding (Amara, Landry, & Halilem, 2015; Didegah & Thelwall, 2013b; Rostami, Mohammadpoorasl, & Hajizadeh, 2014; de Araujo et al., 2012; Vanclay, 2013). In addition, the number of citations of a review paper is usually higher than that of a research paper (Vanclay, 2013). Accordingly, this paper selected six paper features, i.e., x_1-x_6.
2) Journal features. Most studies showed that papers published in journals with high JIFs were cited more (Didegah & Thelwall, 2013b; Vanclay, 2013; Yu et al., 2014). The total number of papers in the journals is also a predictor of citations (Wang et al., 2011). Therefore, we selected the JIF (x_7) and the number of papers published in the journal (x_8) as the journal features.
3) Author features. The author's reputation exerts a significant influence on a paper's citations. Researchers usually follow the work of well-established scholars (Li et al., 2015). Thus, authors with a good reputation usually receive more citations (Xie et al., 2019; Yu et al., 2014). The features of authors' affiliations also affect citations. For example, productivity, as well as the reputation of the organization, might influence citations (Amara et al., 2015). In addition, the cooperation of researchers has a positive effect on citations. Papers with more authors (Didegah & Thelwall, 2013b; Lokker et al., 2008; Yu et al., 2014) or more institutions (Didegah & Thelwall, 2013b) receive more citations in some fields. As a result, we selected nine author features, i.e., x_9-x_{17}, for the citation prediction.
4) Reference features. Previous studies showed that papers with more references (Didegah & Thelwall, 2013b; Lokker et al., 2008; Onodera & Yoshikane, 2015; Yu et al., 2014) or a higher proportion of monographs (Xie et al., 2019) received more citations. In addition, the more recent the references are, the greater the number of citations (Roth, Wu, & Lozano, 2012). Moreover, the journal self-citation rate (Vanclay, 2013) also affects the citations. Therefore, we selected the number of references (x_{18}), the average age (x_{19}) of the references, the journal self-citation rate (x_{20}), and the proportion of different types of references (x_{21}-x_{25}) as reference-related features.
5) Early citation features. There is a close relationship between early-citations and subsequent-citations (Yu et al., 2014). In general, the greater the number of early-citations is, the greater the number of subsequent-citations. Therefore, we selected the first-cited age (x_{26}), the citations in the first year and first two years (x_{27}-x_{28}), and the numbers of citing journals in the first year and the first two years (x_{29}-x_{30}) as the early-citation features.

3.3. Parameters and evaluation

3.3.1. Parameters

We used fivefold cross-validation to determine the appropriate number of hidden layers and the number of hidden units in each hidden layer according to the average MSE of the test set while the other parameters are fixed. As shown in Fig. 1, overall, the best performance of the structures with only one hidden layer is worse than those with multiple hidden layers. It seems that the best performance of the neural network with two or more hidden layers shows little difference. Therefore, we selected a neural network with two hidden layers, and the number of nodes per layer is set to be 30-15-45-1.

Table 2
Parameters of the BP neural network model.

Parameters	Values
Number of units in each layer	30-15-45-1
Activation function in the hidden layers	*leaky-relu*
Regularization coefficient (λ)	0.0015
Learning rate (lr)	Initial value: 0.01; loss<2.5, lr = 0.005; loss<2.3, lr = 0.001; loss<2.1, lr = 0.0005; and loss<2.0, lr = 0.0001.
Optimizer	Adam (β_1 = 0.9, β_2 = 0.9999)
Batch size	64

The learning rate exerts a significant influence on the performance of the model. A high learning rate accelerates the training, but it may lead the model to converge to a locally optimal solution. In contrast, a low learning rate makes the model converge to the globally optimal solution, but the training time is relatively long. To balance the training speed and effect of the model, a dynamic learning rate is adopted. A high learning rate (0.01) is chosen at the beginning of training, and it gradually decreases according to the value of the loss function (Table 2). L2 regularization is an effective way to prevent overfitting (Eq. (1)). The regularization coefficient (λ) is experimentally set as 0.003.

$$\text{loss} = \frac{1}{m}\sum_{i=1}^{i=m}(y_i - \hat{y}_i)^2 + \lambda\sum_j w_j^2, \tag{1}$$

where $y_i$represents the observed number of citations, $\hat{y}_i$indicates the predicted number of citations and w_j mean s the weight of neuron *j*.

The choice of the optimizer also affects the training speed and results. One basic optimization algorithm is the gradient descent method, which updates the parameters along the direction of the gradient descent. However, this method has two drawbacks. One drawback is that it is easy to converge to a locally optimal solution. The other drawback is that the learning rate used for different parameters is the same. To address these problems, researchers have proposed a variety of improved optimization algorithms, among which Adam (Kingma & Ba, 2014) is an adaptive-learning-rate optimization algorithm that updates various parameters with different learning rates. Its excellent performance has been widely confirmed in previous studies. Thus, the Adam optimizer was selected, and the batch size was set to 64.

The activation function converts the input into a nonlinear output, which is vital for enhancing the expressive power of the network. In the inchoate phase, the *sigmoid* and *tanh* functions were mostly used in artificial neural network models. However, both of them experience the vanishing gradient problem, and it limits the network depth. Krizhevsky, Sutskever, and Hinton (2012) used the *relu* activation function — to solve this problem. This function is widely used in the field of deep learning. The disadvantage of the *relu* function is that the value of the negative gradient is always zero, which may cause the 'necrosis' of the neurons. Therefore, variants of the *relu* function have appeared, among which the *leaky relu* has outstanding performance at solving the vanishing gradient problem and neuron 'necrosis' and accelerating the model convergence. Thus, it is adopted in this study.

3.3.2. Comparisons and evaluation

To compare the performance of the BP neural network with other methods, we selected models that were verified to have excellent performance in previous studies as baselines, i.e., the Linear Regression (LR) (Abramo et al., 2019; Lokker et al., 2008; Yan et al., 2012; Yu et al., 2014), XGBoost (eXtreme Gradient Boosting) (Geng, Jing, Jin, & Luo, 2018), RF (Robson & Mousques, 2016), the SVR (Chakraborty et al., 2014; Li et al., 2015; Yan et al., 2012), and the KNN (Yan et al., 2012). The training and testing processes of the five models were implemented through the algorithm library encapsulated in scikit-learn[1] (Pedregosa et al., 2011). To ensure the satisfactory performance of the baseline models, the random search method (Bergstra & Bengio, 2012) was used to determine the parameters. In the parameter selection process, the parameters with a specific distribution were randomly sampled, and those with a set of numerical values were sampled with equal probabilities. After sampling, all combinations of the chosen parameters were traversed, and the optimal parameter combination is determined via a fivefold cross-validation method. The LR has no parameters. The properties of the parameters of the other four models are shown in Table 3. The complete experimental process is shown in Fig. 2.

In addition, the RNN (Abrishami & Aliakbary, 2019) was also chosen as a baseline because it is an artificial neural network method and outperforms state-of-the-art citation prediction methods. Notably, the prediction target of our study is a numerical value, rather than a sequence of consecutive values of citation counts. Therefore, we simplify the sequence-to-sequence RNN model proposed by Abrishami and Aliakbary (2019) to a many-to-one RNN model and used only two sequence-related features, i.e., 'citations in the first year'(x_{27}) and 'citations in the second year'(x_{28} -x_{27}), to train and test the model. The parameters of the RNN model are shown in Table 3.

[1] https://scikit-learn.org/stable/

Table 3
Parameters in the baseline models.

Models	Parameters	Description	Default value	Search range	Setting value
XGBoost	n_estimators	The number of estimators in the models.	100	(10,300)	71
	min_child_weight	The minimum sum of the instance weights (Hessian) needed in a child.	1	(1,30)	3
	max_depth	The maximum depth of the tree.	3	(1,30)	2
RF	n_estimators	The number of trees in the forest.	10	(10,300)	98
	max_depth	The maximum depth of the tree. If None, then the nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples.	None	(1,30)	18
	min_samples_split	The minimum number of samples required to split an internal node.	2	(1,30)	17
	min_samples_leaf	The minimum number of samples required to be at a leaf node.	1	(1,30)	9
KNN	n_neighbors	The default number of neighbors to use for neighbors' queries.	5	(1,30)	9
	weights	Weight function used in prediction. Possible values: 'uniform' or 'distance'	Uniform	'uniform', 'distance'	'distance'
	leaf_size	Leaf size passed to BallTree or KDTree. It affects the construction and query speeds, as well as the memory required to store the tree. The optimal value depends on the nature of the problem.	30	(5,50)	29
SVR	C	Regularization parameter.	1	(1,30)	28
	gamma	Kernel coefficient for the 'rbf', 'poly' and 'sigmoid'. If gamma is 'auto', then 1/n_features will be used instead.	'auto'	(0,1)	0.703
	kernel	Specifies the kernel type to be used in the algorithm.	rbf	'rbf', 'linear'	'rbf'
RNN	Learning rate	Parameters that control the learning speed of the model.	/	/	0.0001
	Batch size	The number of samples in one training batch.	/	/	64
	Number of hidden units	The number of units in the hidden layer.	/	/	512
	Dropout rate	The random dropout probability of network units.	/	/	0.2
	Optimizer	Method used to train the network	/	/	Adam

The evaluation of the model involves the fitness of the model on the training set, and the prediction performance on the test set. The evaluation indicators include the MSE, MAE and R^2 (Eqs. (2)–(4) and are defined as

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{i=N}(y_i - \hat{y}_i)^2 \tag{2}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{i=N}|y_i - \hat{y}_i|, \tag{3}$$

and

$$R^2 = 1 - \frac{\sum_{i=1}^{i=N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{i=N}(y_i - \bar{y}_i)^2} \tag{4}$$

where $y_i$represents the actual number of citations, $\hat{y}_i$indicates the predicted number of citations, $\bar{y}_i$ indicates the average number of observed citations per paper and N represents the number of samples.

4. Results

4.1. Descriptive analysis

The descriptive analysis of the continuous variables is shown in Table 4.

1) **Paper features.** The number of pages in a paper varies from one to twenty-three and the mean is 3.843. The values of the title length range from three to seventy-six with a mean of 18.109 and a median of 17.
2) **Journal features.** The impact factors of journals are between 0.017 and 2.037. Each journal average published 344 papers in a year.
3) **Author features.** The number of authors in a paper shows a skewed distribution with a range from one to seven and a mean of 1.706. In addition, the values of author features x_{10}-x_{13} are sparse, i.e., more than half of them are 0, indicating

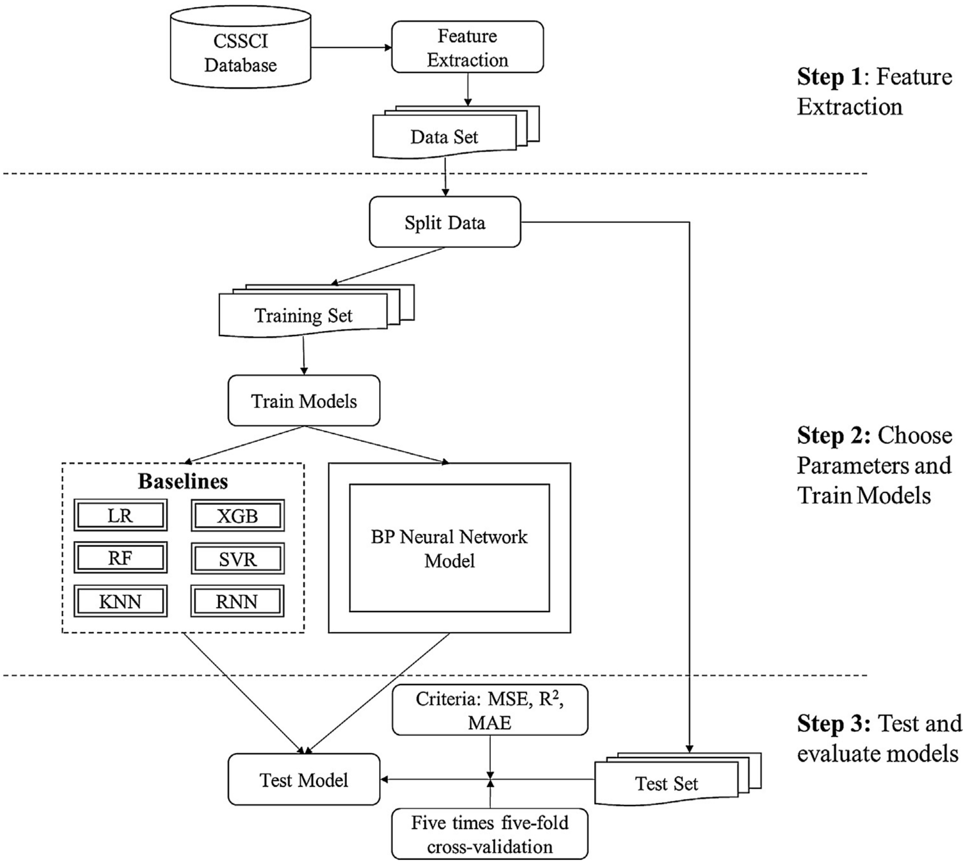

Fig. 2. Experimental process.

Table 4
Descriptive statistics of the continuous variables.

Features	Maximum	Minimum	Median	Mean	Standard deviation
Paper length (x_2)	23	1	4	3.843	1.932
Title length (x_3)	76	3	17	18.109	6.595
Month of publication (x_5)	12	1	7	6.730	3.432
JIF (x_7)	2.037	0.017	0.403	0.471	0.277
Number of publications in the journal (x_8)	830	79	322	344.337	191.266
Number of authors (x_9)	7	1	1	1.706	0.930
Productivity (x_{10})	327	0	1	5.889	17.571
Previous citations (x_{11})	1527	0	0	11.01	63.285
Previous citations per article (x_{12})	30	0	0	0.466	1.110
h- index (x_{13})	17	0	0	0.754	1.641
Number of organizations (x_{14})	6	1	1	1.139	0.392
Total number of papers produced in the institution (x_{15})	29,174	0	750	3409.626	5502.636
Impact of faculty members—max h-index(x_{16})	23	0	3	4.604	4.173
Impact of faculty members—average h-index (x_{17})	3	0	0.308	0.330	0.234
Number of references (x_{18})	145	0	8	9.537	8.315
Reference age (x_{19})	89.333	0	4	5.266	5.258
Self-citations of journals (x_{20})	1	0	0	0.049	0.113
Percentage of journal articles (x_{21})	1	0	0.600	0.548	0.338
Percentage of conference papers (x_{22})	1	0	0	0.026	0.091
Percentage of monographs (x_{23})	1	0	0.091	0.196	0.268
Percentage of online resources (x_{24})	1	0	0	0.13	0.233
Percentage of dissertations (x_{25})	1	0	0	0.013	0.050
First cited age (x_{26})	1	0	0.333	0.413	0.408
Citations in the first year (x_{27})	26	0	0	0.447	1.021
Citations in the first two years (x_{28})	54	0	0	0.978	1.89
No. of citing journals in the first year (x_{29})	11	0	0	0.405	0.847
No. of citing journals in the first two years (x_{30})	16	0	0	0.825	1.354
Five-year citations (Y)	104	0	1	2.011	3.685

Table 5
Descriptive statistics of the categorical variables.

Features	Value	Frequency	Percentage
Document type (x_1)	Research article	48,703	97.73%
	Review	1131	2.27%
Funding (x_4)	National funding	7010	14.07%
	Provincial funding	4599	9.23%
	Other funding	3305	6.63%
	No funding	34,920	70.07%
Punctuation in the title (x_6)	No	38,776	77.81%
	Yes	11,058	22.19%

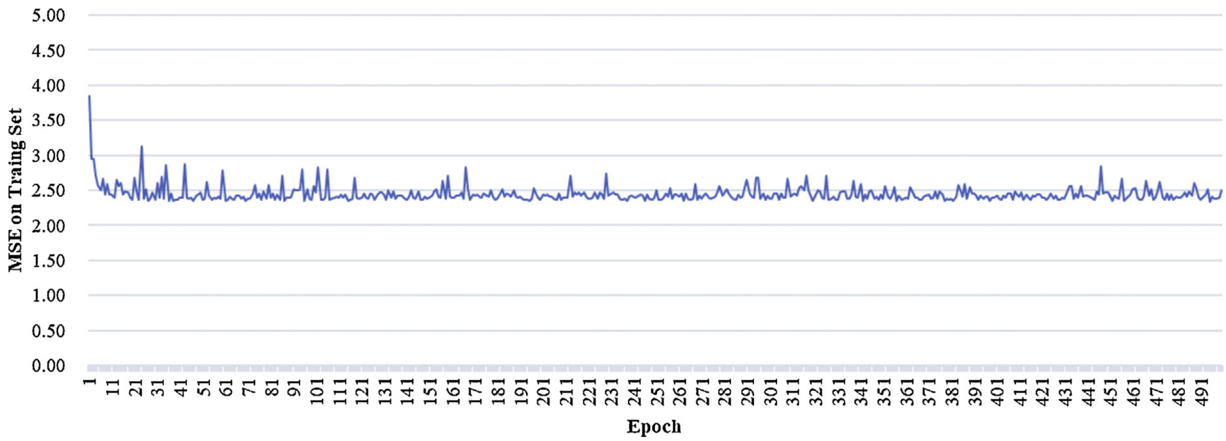

Fig. 3. MSE during the model training process.

that only a few researchers have strong academic influence. For the features of organizations, the distribution of the number of organization (x_{14}) in a paper is also skewed with a range from one to six. The values of features x_{15}-x_{17} have large standard deviations, indicating that their values vary significantly from their means.

4) **Reference features.** The number of references in a paper ranges from zero to 145 with a mean of eight. The average value of the reference age is four. In addition, the average self-citation rates of journals are small, i.e., only 4.9 % of the references are self-citations of journals. Among the varieties of reference types, the percentage of journal articles in the reference list is the largest, followed by books, and the percentages of other reference types are small.
5) **Early citation features.** The five early-citation features present skewed distributions that mean that most papers are not cited frequently in the early stage, and only a few papers are highly cited.

There are three categorical variables in the selected features, as listed in Table 5. The number of research papers is significantly higher than that of review papers. The percentages of papers with different funding sources from high to low are no funding > national funding > provincial funding > other funding. In addition, the number of papers without punctuation in the titles is significantly larger than that with punctuations.

4.2. Prediction results

The change of the MSE for the training set during the training process is shown in Fig. 3. In the batch training mode, the MSE values show a wavelike decrease and they no longer decreased after 500 training epochs. The model with the smallest MSE on the test set was selected as the optimal model for further research.

In the optimal model, the MSEs for the training and test set are 2.357 and 2.584, respectively. The corresponding R^2s are 0.819 and 0.837, respectively. To analyze the prediction performance in depth, 100 samples with the smallest (Fig. 4) and with the largest (Fig. 5) errors were selected for further research. The error is the absolute value of the difference between the predicted and the observed citations. For the 100 best-predicted samples, the predicted results completely match the observed citations, among which the number of observed citations ranges from 0 to 6 and 89 samples have no citations. For the 100 samples with the worst prediction performance, as the number of observed citations increases, the predicted values show a wavelike increase, among which the number of observed citations ranges from 8 to 104.

As shown in Fig. 5, the BP neural network usually estimates lower values for the citation counts. However, as shown in Fig. 6, the proportions of high and low estimates are roughly equal among the whole test set, i.e., the predicted citation counts of 40.4 % of the samples are higher than the actual ones.

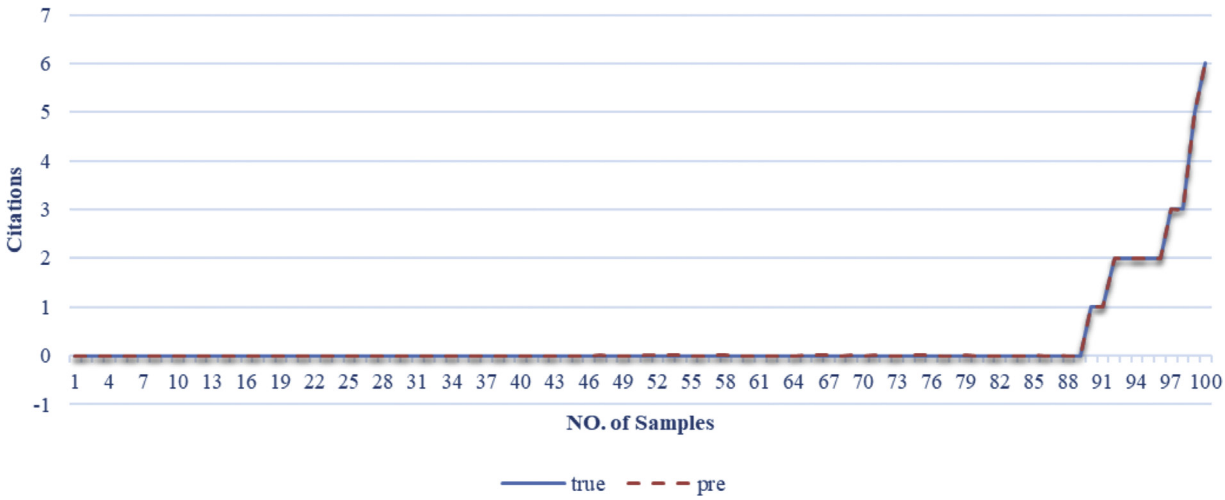

Fig. 4. 100 samples with the best prediction performance.

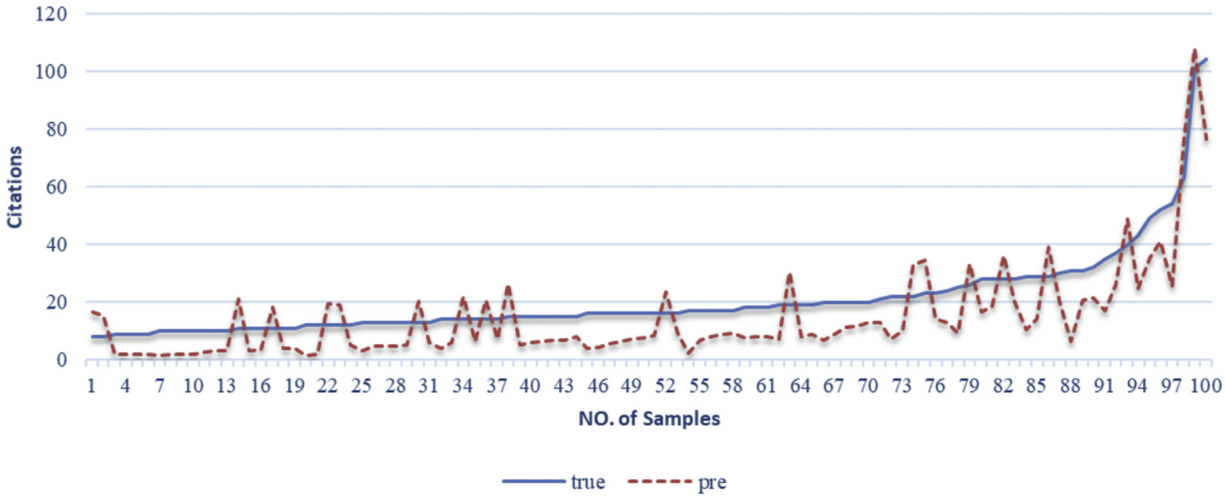

Fig. 5. 100 samples with the worst prediction performance.

4.3. Feature importance

We follow three steps to compare the relative importance of the selected features. First, we randomly divided the data into training and test data sets ten times at a ratio of 4:1 and obtained ten paired training and test sets. Second, in each group of data, we sequentially dropped one feature and used the remaining 29 features ('leave-one-out model') to train the model using the training set, after which the MSE of the test set was calculated. The models using all 30 features were used as the basis of the comparison ('30-feature model'). As a result, ten groups of results were obtained and each of them contains 31 models. Third, the paired sample *t*-test was used to determine whether there are differences between each 'leave-one-out model' and the '30-feature model'. As shown in Table 6, the MSEs of the five leave-one-out models are significantly higher than that of the '30-feature model', i.e., 'citations in the first two years' (x_{28}), 'first-cited age' (x_{26}), 'paper length' (x_2), 'month of publication' (x_5), and 'self-citations of journals' (x_{20}). In other words, the five features have significant effects on the prediction performance of the model, while the other features barely contribute to the prediction.

4.4. Comparisons

The fitting and prediction performance of the baselines and the BP neural network model are shown in Table 7. Compared with the baseline models, the BP neural network model has a slightly worse fitting performance than the RF, SVR, and XGB on the training set and a better fitting performance than the LR and RNN; and it has better prediction performance than the six baseline models. The MSE on the test set of the BP neural network model is reduced by 8.76 % for XGB, 17.26 % for the RF, 26.67 % for the SVR, 11.29 % for the LR, 60.60 % for the KNN, and 16.21 % for the RNN. In addition, the changes of the R^2

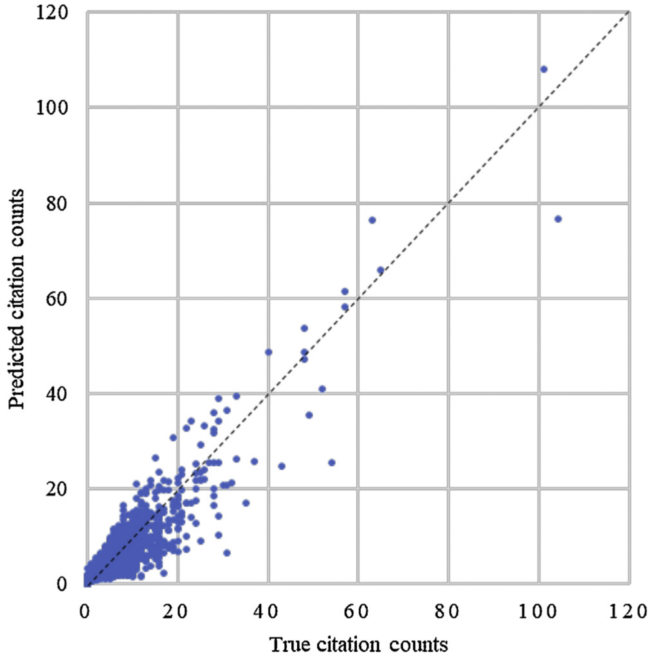

Fig. 6. Scatter of the predicted citation counts and actual citation counts.

Table 6
Difference between each 'leave-one-out model' and the '30-feature model'.

No. of the dropped feature	Difference in the mean MSE	No. of the dropped feature	Difference in the mean MSE
x_1	0.004	x_{16}	−0.004
x_2	0.036**	x_{17}	0.003
x_3	0.006	x_{18}	0.004
x_4	0.000	x_{19}	0.006
x_5	0.031**	x_{20}	0.028*
x_6	−0.003	x_{21}	−0.001
x_7	0.005	x_{22}	0.013
x_8	−0.002	x_{23}	−0.005
x_9	0.007	x_{24}	0.013
x_{10}	−0.002	x_{25}	0.002
x_{11}	−0.004	x_{26}	0.261***
x_{12}	−0.003	x_{27}	−0.002
x_{13}	0.003	x_{28}	0.599***
x_{14}	0.012	x_{29}	−0.002
x_{15}	−0.001	x_{30}	−0.005

Notes: * indicates $p < 0.05$, ** indicates $p < 0.01$, and *** indicates $p < 0.001$.

Table 7
Fitting and prediction performances for different models.

Models	Training		Test	
	MSE	R^2	MSE	R^2
KNN	/	/	6.559	0.587
SVR	2.094	0.839	3.524	0.778
RF	1.595	0.877	3.123	0.803
LR	2.710	0.792	2.913	0.817
XGB	2.305	0.823	2.832	0.822
RNN	2.895	0.777	3.145	0.802
BP_five_features	2.455	0.811	2.600	0.836
BP	2.357	0.819	2.584	0.837

for the LR, BP, SVR, XGB, RF, and RNN models between the test and training sets are 2.5 %, 1.8 %, 6.1 %,0.1 %, 7.4 %, and 2.5 %, respectively, indicating that the models have excellent generalization capabilities (applicability to other datasets).

We trained another model (BP_five_features) using only the five features mentioned in Section 4.3 that have a significant effect on the prediction performance and compared its performance with those of other models. As shown in Table 7, the

Table 8
Performance of each model at predicting the top 1% of the high-impact papers.

Models	MAE	MSE
XGB	8.227	107.715
RF	8.486	138.254
SVR	8.467	155.223
KNN	15.437	348.900
LR	7.417	88.554
RNN	7.550	97.550
BP_five_features	7.411	83.870
BP	7.370	84.635

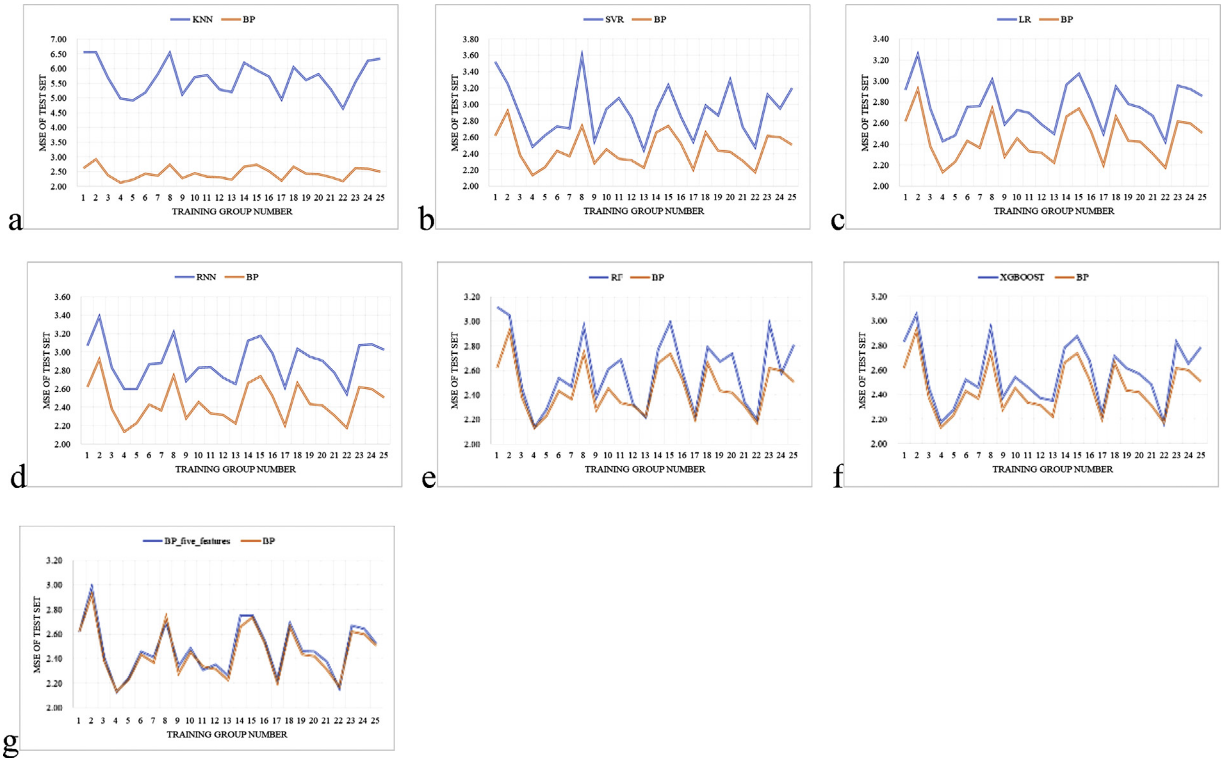

Fig. 7. Results of the five times five-fold cross-validation. a-g, the comparisons of the performance between the BP neural network and the KNN, SVR, LR, RNN, RF, XGBoost, and BP_five_features, respectively, according to the MSE on the test set.

performance of 'BP_five_features' on the test set is slightly worse than that of the BP model, but it outperforms the other six models.

To compare the prediction performance for highly cited papers, we observe the prediction results of the papers receiving more than 17 citations (104 papers ranking in the top 1.0 % among test papers). As shown in Table 8, the MAE and MSE of the BP neural network are slightly lower than those of the LR and significantly lower than those of the other four models, which suggest that BP outperforms other models at predicting highly cited papers.

To further determine whether the advantage of the BP neural network is significant, five times fivefold cross-validation was used, and 25 group results were obtained. As shown in Fig. 7a–g, the MSE of the BP neural network model is the smallest in each result group. The differences in the MSEs are small for the XGBoost, RF, LR, SVR, RNN and BP. However, the performance of the KNN is clearly inferior to the other models. As shown in Table 9, the results of the paired sample *t*-test show that the MSE on the test set of the BP neural network model is significantly smaller than those of the six baseline models. In addition, the 'BP_five_features' model is significantly and slightly worse than the BP model but it outperforms the six baseline models.

5. Discussion and conclusions

This study used a four-layer BP neural network model to predict the five-year citations of 49,834 papers published from 2000 to 2013 in the library, information and documentation field indexed by the CSSCI database. We extracted six paper features, two journal features, nine author features, eight reference features, and five early-citation features for the prediction

Table 9
Paired sample *t*-test results of the models.

Models	Mean (Standard deviation)	Compared to BP (t)	Compared to BP_five_features (t)
BP	2.451(0.204)	/	−4.342***
XGBoost	2.570(0.244)	8.589***	6.106***
RF	2.597(0.291)	5.289***	4.031***
SVR	2.913(0.320)	11.448***	9.976***
LR	2.763(0.214)	40.198***	34.633***
KNN	5.670(0.556)	40.782***	40.219***
RNN	2.898(0.220)	48.371***	38.558***
BP_five_features	2.477(0.042)	−4.342***	/

Notes: *** indicates $p < 0.001$.

task. One of our contributions is that we improved the citation prediction accuracy by applying the BP neural network. The empirical results showed that the BP neural network significantly outperformed the other six baselines (XGBoost, RF, LR, SVR, KNN, and RNN) in terms of both the whole dataset of papers and the high-impact papers. Additionally, its generalizability is comparatively better. The outstanding performance of the BP neural network is attributed to its advantage in automatically learning features through a multilevel and nonlinear transformation.

The other is that this study determined five essential features for citation prediction, i.e., ‘citations in the first two years’, ‘first-cited age’, ‘paper length’, ‘month of publication’, and ‘self-citations of journals’. Based on the five factors, the model showed that a two-year citation window is sufficient for predicting the medium-term impact of an individual paper. The crucial role that early citations, i.e., ‘citations in the first two years’ and ‘first-cited age’, play in citation prediction has been confirmed in previous studies (Abramo et al., 2019; Yu et al., 2014); however, the other three features were rarely mentioned in previous studies.

It is crucial to emphasize that the dataset used in our study is comprised of only Chinese papers; therefore, we observed some findings different from those in previous studies that were mainly based on English publications. The journal features were identified to be insignificant to citation prediction, which is inconsistent with studies of Bornmann et al. (2014), Didegah and Thelwall (2013a, 2013b) and Yu et al. (2014) whose datasets mainly covered journals indexed in the Web of Science database. Though the JIF has been widely used to measure the impact of scientific papers and scholars and allocate research findings (Holden, Rosenberg, Barker, & Onghena, 2006; Kirchhof, Bornfeld, & Grehn, 2007), our results indicate that such journal features as the JIF encounter difficulty in predicting the citation counts of the Chinese articles published in this database. It is also worth noticing that we simplified the RNN models proposed by Abrishami and Aliakbary (2019) and used only citations in the first two years to train the model. Our prediction target is different from theirs, i.e., we aim to predict the five-year citation counts of papers, while they attempted to predict a sequence number of citations. Thus, the performance of the RNN model is probably underestimated, and the comparison needs further exploration.

The experiment results show that it is feasible to use the BP neural network to predict the five-year citations of Chinese academic papers since their publication. Compared with other machine learning methods used in the previous studies to predict English papers, it achieved better performance and its generalizability was excellent. Moreover, the high prediction accuracy shows that this approach has the potential to predict the impact of Chinese research works or academics and therefore is useful in such situations as literature retrieval and scientific evaluation. The results also showed that such nonlinear models as the BP, XGBoost, and RF performed better than the linear model (LR) did, which is in line with the work on English academic papers (Yan et al., 2012).

There are at least two limitations in this study. One limitation is that the dataset used in this study only contains Chinese papers from the CSSCI in the library, information and documentation discipline. Whether the advantages of the BP neural network still exist in predicting the citation counts of papers from other subjects or other languages requires further studies. The other is that due to the ‘black box’ nature of the BP neural network, the exact relationship between the selected features and the number of citations has not been determined.

Author contributions

Xuanmin Ruan: Collected the data, Contributed data or analysis tools, Performed the analysis, Wrote the paper.
Yuanyang Zhu: Contributed data or analysis tools, Performed the analysis.
Jiang Li: Conceived and designed the analysis, Other contribution.
Ying Cheng: Conceived and designed the analysis, Other contribution.

Acknowledgments

This work was partially supported by the National Social Science Foundation of China (17BTQ014).

References

Abramo, G., D'Angelo, C. A., & Felici, G. (2019). Predicting publication long-term impact through a combination of early citations and journal impact factor. *Journal of Informetrics*, *13*(1), 32–49. http://dx.doi.org/10.1016/j.joi.2018.11.003

Abrishami, A., & Aliakbary, S. (2019). Predicting citation counts based on deep neural network learning techniques. *Journal of Informetrics*, *13*(2), 485–499. http://dx.doi.org/10.1016/j.joi.2019.02.011

Aksnes, D. W. (2003). Characteristics of highly cited papers. *Research Evaluation*, *12*(3), 159–170. http://dx.doi.org/10.3152/147154403781776645

Amara, N., Landry, R., & Halilem, N. (2015). What can university administrators do to increase the publication and citation scores of their faculty members? *Scientometrics*, *103*(2), 489–530. http://dx.doi.org/10.1007/s11192-015-1537-2

Bai, X. M., Zhang, F. L., & Lee, I. (2019). Predicting the citations of scholarly paper. *Journal of Informetrics*, *13*(1), 407–418. http://dx.doi.org/10.1016/j.joi.2019.01.010

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*, 281–305.

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. http://dx.doi.org/10.1016/j.joi.2013.11.005

Chakraborty, T., Kumar, S., Goyal, P., Ganguly, N., & Mukherjee, A. (2014). Towards a stratified learning approach to predict future citation counts. *2014 Ieee/Acm Joint Conference on Digital Libraries (Jcdl)*, 351–360.

Dang, Q. V., & Ignat, C. L. (2016). Quality assessment of wikipedia articles without feature engineering. *2016 Ieee/Acm Joint Conference on Digital Libraries (Jcdl)*, 27–30. http://dx.doi.org/10.1145/2910896.2910917

de Araujo, C. G. S., de Oliveira, B. R. R., Brito, L. V. D., da Matta, T. T., Viana, B. F., de Souza, C. P., . . . & Portugal, E. D. M. (2012). Two-year citations of JAPPL original articles: Evidence of a relative age effect. *Journal of Applied Physiology*, *112*(9), 1434–1436. http://dx.doi.org/10.1152/japplphysiol.01491.2011

Didegah, F., & Thelwall, M. (2013a). Determinants of research citation impact in nanoscience and nanotechnology. *Journal of the American Society for Information Science and Technology*, *64*(5), 1055–1064. http://dx.doi.org/10.1002/asi.22806

Didegah, F., & Thelwall, M. (2013b). Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *Journal of Informetrics*, *7*(4), 861–873. http://dx.doi.org/10.1016/j.joi.2013.08.006

Dong, Y. X., Johnson, R. A., & Chawla, N. V. (2015). Will this paper increase your h-index? Machine learning and knowledge discovery in databases. *Pt Iii*, *9286*, 259–263. http://dx.doi.org/10.1007/978-3-319-23461-8_26

Fu, L. D., & Aliferis, C. (2008). Models for predicting and explaining citation count of biomedical articles. *AMIA . . . Annual Symposium Proceedings. AMIA Symposium*, 222–226.

Fu, L. D., & Aliferis, C. F. (2010). Using content-based and bibliometric features for machine learning models to predict citation counts in the biomedical literature. *Scientometrics*, *85*(1), 257–270. http://dx.doi.org/10.1007/s11192-010-0160-5

Geng, R. B., Bose, I., & Chen, X. (2015). Prediction of financial distress: An empirical study of listed Chinese companies using data mining. *European Journal of Operational Research*, *241*(1), 236–247. http://dx.doi.org/10.1016/j.ejor.2014.08.016

Geng, Q., Jing, R., Jin, J., & Luo, Q. (2018). Citation prediction and influencing factors analysis on academic papers. *Library and Information Service*, *62*(14), 29–40.

Holden, G., Rosenberg, G., Barker, K., & Onghena, P. (2006). Should decisions about your hiring, reappointment, tenure, or promotion use the impact factor score as a proxy indicator of the impact of your scholarship? *MedGenMed: Medscape General Medicine*, *8*(3), 21.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366. http://dx.doi.org/10.1016/0893-6080(89)90020-8

Ibanez, A., Larranaga, P., & Bielza, C. (2009). Predicting citation count of Bioinformatics papers within four years of publication. *Bioinformatics*, *25*(24), 3303–3309. http://dx.doi.org/10.1093/bioinformatics/btp585

Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization*. arXiv e-prints. https://ui.adsabs.harvard.edu/abs/2014arXiv1412.6980K

Kirchhof, B., Bornfeld, N., & Grehn, F. (2007). The delicate topic of the impact factor. *Graefes Archive for Clinical and Experimental Ophthalmology*, *245*(7), 925–927. http://dx.doi.org/10.1007/s00417-007-0618-1

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In NIPS (pp. 1105–2097).

Lee, S., & Choeh, J. Y. (2014). Predicting the helpfulness of online reviews using multilayer perceptron neural networks. *Expert Systems with Applications*, *41*(6), 3041–3046. http://dx.doi.org/10.1016/j.eswa.2013.10.034

Li, C. T., Lin, Y. J., Yan, R., & Yeh, M. Y. (2015). Trend-based citation count prediction for research articles. *Advances in Knowledge Discovery and Data Mining, Part I*, *9077*, 659–671. http://dx.doi.org/10.1007/978-3-319-18038-0_51

Lokker, C., McKibbon, K. A., McKinlay, R. J., Wilczynski, N. L., & Haynes, R. B. (2008). Prediction of citation counts for clinical articles at two years using data available within three weeks of publication: retrospective cohort study. *British Medical Journal*, *336*(7645), 655–657. http://dx.doi.org/10.1136/bmj.39482.526713.BE

Mai, F., Tian, S. N., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*, *274*(2), 743–758. http://dx.doi.org/10.1016/j.ejor.2018.10.024

Onodera, N., & Yoshikane, F. (2015). Factors affecting citation rates of research articles. *Journal of the Association for Information Science and Technology*, *66*(4), 739–764. http://dx.doi.org/10.1002/asi.23209

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Robson, B. J., & Mousques, A. (2016). Can we predict citation counts of environmental modelling papers? Fourteen bibliographic and categorical variables predict less than 30% of the variability in citation counts. *Environmental Modelling & Software*, *75*, 94–104. http://dx.doi.org/10.1016/j.envsoft.2015.10.007

Rokach, L., Kalech, M., Blank, I., & Stern, R. (2011). Who is going to win the next association for the advancement of artificial intelligence fellowship award? Evaluating researchers by mining bibliographic data. *Journal of the American Society for Information Science and Technology*, *62*(12), 2456–2470. http://dx.doi.org/10.1002/asi.21638

Rostami, F., Mohammadpoorasl, A., & Hajizadeh, M. (2014). The effect of characteristics of title on citation rates of articles. *Scientometrics*, *98*(3), 2007–2010. http://dx.doi.org/10.1007/s11192-013-1118-1

Roth, C., Wu, J., & Lozano, S. (2012). Assessing impact and quality from local dynamics of citation networks. *Journal of Informetrics*, *6*(1), 111–120. http://dx.doi.org/10.1016/j.joi.2011.08.005

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. http://dx.doi.org/10.1038/323533a0

Su, X. N., Deng, S. H., & Shen, S. (2014). The design and application value of the Chinese social science citation index. *Scientometrics*, *98*(3), 1567–1582. http://dx.doi.org/10.1007/s11192-012-0921-4

Tahamtan, I., Afshar, A. S., & Ahamdzadeh, K. (2016). Factors affecting number of citations: A comprehensive review of the literature. *Scientometrics*, *107*(3), 1195–1225. http://dx.doi.org/10.1007/s11192-016-1889-2

Vanclay, J. K. (2013). Factors affecting citation rates in environmental science. *Journal of Informetrics*, *7*(2), 265–271. http://dx.doi.org/10.1016/j.joi.2012.11.009

Wang, P., & Li, X. (2019). Assessing the quality of information on Wikipedia. *Journal of the Association for Information Science and Technology*.

Wang, D. S., Song, C. M., & Barabasi, A. L. (2013). Quantifying long-term scientific impact. *Science*, *342*(6154), 127–132. http://dx.doi.org/10.1126/science.1237825

Wang, F. H., Fan, Y., Zeng, A., & Di, Z. R. (2019). Can we predict ESI highly cited publications? *Scientometrics*, *118*(1), 109–125. http://dx.doi.org/10.1007/s11192-018-2965-6

Wang, M. Y., Yu, G., An, S., & Yu, D. R. (2012). Discovery of factors influencing citation impact based on a soft fuzzy rough set model. *Scientometrics*, *93*(3), 635–644. http://dx.doi.org/10.1007/s11192-012-0766-x

Wang, M. Y., Yu, G., & Yu, D. R. (2011). Mining typical features for highly cited papers. *Scientometrics*, *87*(3), 695–706. http://dx.doi.org/10.1007/s11192-011-0366-1

Wong, T. C., & Chan, A. H. S. (2015). A neural network-based methodology of quantifying the association between the design variables and the users' performances. *International Journal of Production Research*, *53*(13), 4050–4067. http://dx.doi.org/10.1080/00207543.2014.988886

Wong, T. C., Chan, H. K., & Lacka, E. (2017). An ANN-based approach of interpreting user-generated comments from social media. *Applied Soft Computing*, *52*, 1169–1180. http://dx.doi.org/10.1016/j.asoc.2016.09.011

Xie, J., Gong, K. L., Li, J., Ke, Q., Kang, H. C., & Cheng, Y. (2019). A probe into 66 factors which are possibly associated with the number of citations an article received. *Scientometrics*, *119*(3), 1429–1454. http://dx.doi.org/10.1007/s11192-019-03094-z

Xu, J., Li, M., Jiang, J., Ge, B., & Cai, M. (2019). Early prediction of scientific impact based on multi-bibliographic features and convolutional neural network. *IEEE Access*, *7*, 92248–92258. http://dx.doi.org/10.1109/access.2019.2927011

Yan, R., Huang, C., Tang, J., Zhang, Y., & Li, X. (2012). To better stand on the shoulder of giants. In Acm/ieee-cs joint conference on digital libraries (pp. 51–60).

Yu, T., Yu, G., Li, P. Y., & Wang, L. (2014). Citation impact prediction for scientific papers using stepwise regression analysis. *Scientometrics*, *101*(2), 1233–1252. http://dx.doi.org/10.1007/s11192-014-1279-6

Zhang, S. F., Yuan, C. C., Chang, K. C., & Ken, Y. (2012). Exploring the nonlinear effects of patent H index, patent citations, and essential technological strength on corporate performance by using artificial neural network. *Journal of Informetrics*, *6*(4), 485–495. http://dx.doi.org/10.1016/j.joi.2012.03.006