A deep-learning based citation count prediction model with paper metadata semantic features

Anqi Ma[1] · Yu Liu[1] · Xiujuan Xu[1] · Tao Dong[1]

Abstract

Predicting the impact of academic papers can help scholars quickly identify the high-quality papers in the field. How to develop efficient predictive model for evaluating potential papers has attracted increasing attention in academia. Many studies have shown that early citations contribute to improving the performance of predicting the long-term impact of a paper. Besides early citations, some bibliometric features and altmetric features have also been explored for predicting the impact of academic papers. Furthermore, paper metadata text such as title, abstract and keyword contains valuable information which has effect on its citation count. However, present studies ignore the semantic information contained in the metadata text. In this paper, we propose a novel citation prediction model based on paper metadata text to predict the long-term citation count, and the core of our model is to obtain the semantic information from the metadata text. We use deep learning techniques to encode the metadata text, and then further extract high-level semantic features for learning the citation prediction task. We also integrate early citations for improving the prediction performance of the model. We show that our proposed model outperforms the state-of-the-art models in predicting the long-term citation count of the papers, and metadata semantic features are effective for improving the accuracy of the citation prediction models.

✉ Yu Liu
  yuliu@dlut.edu.cn

  Anqi Ma
  maanqi@mail.dlut.edu.cn

  Xiujuan Xu
  xjxu@dlut.edu.cn

  Tao Dong
  dongtao2019@mail.dlut.edu.cn

[1] School of Software, Dalian University of Technology, Dalian 116621, China

Introduction

With the rapid increase of academic papers, evaluating the impacts of academic papers has become a hot issue. Citation count is one of the most commonly used indicators for evaluating paper's impact, a paper with more citation count is considered to have higher impact. Researchers have proposed many classical indicators based on the citation count for impact evaluation of scholars, journals and papers (Braun et al. 2006; Egghe, 2006; Garfield, 2006; Hirsch, 2005; Yan & Ding, 2010). Since the citation count is easy to obtain from the literature databases, scholars usually regard this simple, standard and objective indicator as a key factor on locating the papers for reading. With the rapid increase in the number of academic papers, scholars usually need to identify high-impact papers in advance, these papers can inspire scholars to breed research ideas to an certain extent, so that helping with planing their research directions better (R. Abrishami & Aliakbary, 2019; Hu et al. 2020; Yan et al. 2012). Predicting the citation count of papers can help scholars capture the high-quality papers in the field (Ruan et al., 2020). As Abrishami and Aliakbary (2019) said: "By predicting the citation count of a paper, we can evaluate the future impact of the paper authors, with potential applications in hiring researchers and faculties, and granting awards and funds", and some existing works also hold the similar viewpoint (Bai et al., 2019; Clauset et al., 2017; Ruan et al., 2020; Xiao et al., 2016). Therefore, it is also of important reference value for peer review process to solve the above-mentioned fundamental problems.

Previous studies have focused on building effective citation count prediction models for exploring the citation patterns of academic papers (Bai et al., 2019; Cao et al., 2016; Chen, 2015; Yan et al., 2011), they chose machine learning algorithms such as k-Nearest Neighbor (KNN), XGBoost and Gradient Boosting Regression Trees (GBRT). For example, Yan et al. (2011) first presented the citation count prediction task. They predicted the future citation count for publications by employing several machine learning regression models. The predictive models were based on author feature, venue feature, paper feature and time feature. In recent years, deep learning (Lecun et al., 2015) has made significant achievement in various fields such as face recognition (Wang et al., 2019c; Wen et al., 2019), speech recognition (Chen et al., 2019; Jati et al. 2019), machine translation (Platanios et al., 2020; Zeng et al., 2020), sentiment analysis (Tang et al., 2019; Zhu et al., 2019) and text generation (Guo et al., 2018; Yu et al., 2017), etc. Therefore, deep learning techniques are also considered to solve this citation count prediction problem (Abrishami & Aliakbary, 2019; Li et al., 2019; Ruan et al., 2020; Yuan et al., 2018). Li et al. (2019) obtained bibliometric features at three levels from academic heterogeneous network, and then used Convolutional Neural Network (CNN) for capturing implicit relations between different features to predict long-term citation count. This is the first study which introduced CNN to the citation prediction problem. A multi-layer neural network was employed to predict five-year citations of CSSCI papers (Ruan et al., 2020). This method extracted a total of 30 features in five categories to tackle the prediction problem. Moreover, they selected five features with a significant impact on the prediction performance of the model from all the 30 features, i.e., the number of citations in the first two years, the time window from the publish year to the first citation year, the publication month, and the journal self-citation rate. The finding shows that the prediction performance of the model only with these five features is slightly worse than that of the model using all features.

Many features have been explored for predicting citation count. For example, early citations have shown that early citations contribute to improving the performance of predicting the long-term impact of a paper in many previous works (Abramo et al., 2019; Bornmann et al., 2014; Newman, 2014; Stegehuis et al., 2015). Except for early citations, some bibliometric features and altmetric features have also been explored for predicting the impact of academic papers (Hassan et al., 2019; Wang et al. 2019a, b; Wu et al. 2019). With the opening of the peer view process, the text information such as peer review text was also used to enhance the effectiveness of the citation count predictive model. Li et al. (2019b) first considered extracting the semantic representation from the peer view text for enhancing the effectiveness of citation prediction model. They also constructed the wide component according to the topic distribution, author influence and so on. The results show that peer view text is useful for this task. However, this method cannot be applied to most journal and conference papers due to the unpublished open peer review text of these papers.

Actually, paper metadata text such as title, abstract and keywords also contains valuable information which has effect on the future impact of papers (Fronzetti Colladon et al., 2020; Hu et al., 2020; Sohrabi & Iraj, 2017), and are easier to obtain than the peer review text. Fronzetti Colladon et al. (2020) first explored the sentiment metric of the abstract text, which is calculated by the regularized sum of sentiment value of each word in the abstract text given by the VADER lexion. Although it was the pioneer work which considered the impact of the sentiment on the long-term citation count of a paper, the semantic information contained in the abstract context was still ignored. Metadata text is the common and easily accessible information of a paper, it demonstrates the research problem, proposed method and improved result, and it is the most directly part for researchers to gain the information which they concerned such as research task and conclusion. However, present studies ignore the semantic information contained in the metadata text. Hence, how to effectively extract text features with semantic information from the metadata is crucial.

Citation prediction task includes predicting cumulative citations under given citation time window and predicting long-term citation sequence. In this paper, we consider the long-term citation prediction problem, and many researchers have made efforts on this task(Abrishami & Aliakbary, 2019; Cao et al., 2016). And we propose a novel citation prediction model based on paper metadata text to predict the long-term citation count, and the core of our model is to obtain the semantic information from the metadata text. We choose deep learning techniques for semantic features extraction and citation prediction. The long-term citation count prediction task is defined as a multi-output regression problem in supervised machine learning, i.e., the prediction model outputs a sequence which contains the annual citation count received by the paper in each year. We use Doc2Vec algorithm to encode the metadata text, and then apply Bi-directional Long Short Term Memory (Bi-LSTM) with attention mechanism to further extract high-level semantic features for learning the citation prediction task. We also integrate early citations for improving the prediction performance of the model. We compare the proposed model with other five popular citation prediction methods to verify the accuracy of the proposed model. The experiment results show that our proposed model outperforms the existing state-of-the-art models, and metadata semantic features are effective for improving the accuracy of the citation prediction models.

Main contributions of this paper include: (1) We propose a novel citation count prediction model, which employs Doc2Vec and Bi-LSTM with attention mechanism for metadata semantic features extraction and citation prediction; (2) We combine early citations and

metadata semantic features of academic papers applying for predicting the long-term citation count of papers; (3) We verify the correctness and superiority of the proposed model over the existing baseline models in the citation count prediction task by running the comparison experiments.

The paper is organized as follows: In Related work Section we discuss previous work related to our research. Dataset Section describes the dataset used in our experiment. Methodology Section introduces the proposed model. Result and discussion Section presents the comparison results with the state-of-the-art models. In Conclusion Section we conclude the paper and describe the future studies.

Related work

Predicting the impact of academic papers

Previous studies on predicting the impact of academic papers based on the citation count metric are mainly categorized into two aspects: identifying the highly-cited paper and predicting citation count.

Identifying the highly-cited papers

Identifying the highly-cited papers can help researchers track research trends. Generally, recognizing highly cited papers is defined as a binary classification problem. In this problem, many efforts have also been devoted to the design of the methods for identifying the highly-cited papers in the future. For example, Newman (2014) detected highly-cited papers in a field using z-score which was calculated by the short-term citation count of the papers. Wang et al. (2012) proposed a case-based classifier (CBC) based on case-based reasoning (CBR) and soft fuzzy rough set (SFRS), then used the classifier to predict whether the papers from four different journals in different fields were highly-cited papers (HCPs), medium-cited papers (MCPs) or low-cited papers (LCPs) within 15 years of publication.

Moreover, the efficiency of many features has been investigated extensively in this task. For example, Wang et al. (2019b) collected twenty-three bibliometric indices from Web of Science (WOS) and alternative indices from Article-level Metrics to identify the highly-cited papers in the Public Library of Science (PLOS) by using three supervised machine learning methods. Their results showed that both bibliometric indices and alternative metrics were well predictive, and the combination of both was considered to be better. Hassan et al. (2019) designed eleven features extracted from the altmetric data to distinguish the highly-cited articles, and the user influence feature was proved to be the most important feature in classification. Hu et al. (2020) defined five keyword popularity (KP) features for the first time, and combined KP features with author-based and journal-based bibliometric features for identification of highly-cited papers. Their experimental results showed that KP features can make the model more predictive, especially in the management information system (MIS) discipline. It might be that many new topics and concepts are often introduced in the interdisciplinary fields, thus KP features can provide more positive impact on the MIS papers. Wang et al. (2019a) explored the ability of four factors (impact of the first author, scientific impact of the potential leader, scientific impact of the team and the relevance of authors' existing papers) on predicting ESI highly-cited papers based on

neural network. They found the potential leader factor played a more important role in the short term, while the team factor was more important in the long term.

Predicting citation count

Existing works for predicting citation count can be divided into two categories according to their input information. The first category used multiple features as the input information. Bornmann et al. (2014) improved the citation impact measurement by considering journal impact, the number of authors, the number of cited references and the number of pages. Bornmann et al. (2012) found that citation counts were correlated with the citation performance of the cited references, the language of the publishing journal, the specific chemical subfield, and the reputation for the authors by using multiple regression analysis. Stegehuis et al. (2015) took two predictors, i.e., impact factor and citation count in the first year for predicting the long-term citation of the papers in Physics. Compared with only using one indicator, combining both two factors can make the regression model fit better. Additionally, the issue whether other indicators are predictive or not will be investigated. Bai et al. (2019) introduced the Paper Potential Index (PPI) model to explore the citation pattern evolving over time based on inherent quality of scholarly paper, scholarly paper impact decaying over time, early citations and early citers' impact factors.

In the second category, only the citation count of the paper was used as input feature. For example, Cao et al. (2016) found the most matched papers according to the early short-term citations, then used the citation patterns of the most matched papers to predict the future citation count of a paper. The first way is using the annual average number of citations of the most matched papers as the prediction results; the second way is dividing the most matched papers into three groups using Gaussian Mixture Model algorithm, and then finding out the centroid which is the most similar to the predicted target according to the early citation pattern, and using the future citations of this centroid as the prediction results. Abrishami and Aliakbary (2019) used sequence to sequence model for predicting future citation count of a paper based on its early citations, and this method outperforms state-of-the-arts methods. Moreover, it is also proved that the more sufficient input information the model has, the more accurate prediction the model outcomes.

The text of academic paper affects its future impact

The title and abstract metadata is an important section used for obtaining the valuable information about a paper (Hu et al. 2020), and it plays an important role in attracting researchers to read the paper, thus it affects the citation count (Haggan, 2004; Habibzadeh & Yadollahie, 2010; Jamali & Nikzad, 2011; Weinberger et al., 2015; Letchford et al., 2016; Colladon et al., 2020). Habibzadeh and Yadollahie (2010) analyzed the relevance between the length of the paper title in medical journals and citation count by using linear regression analysis. In their dataset, longer title can conduct with higher citations, and this phenomenon occurred more in journals with high journal impact factor (JIF). Jamali & Nikzad, (2011) explored three main title types of PLoS journal papers, i.e., descriptive, declarative and interrogative title. The conclusion shows that papers with the type of descriptive or declarative titles were more citable compared to that with the type of question titles. It was the first study on the relationship between the title type of the paper and the citations. As Fronzetti Colladon et al. (2020) mentioned, "If the title plays an important role as a 'touch point' for attracting the reader towards the manuscript, the abstract

should do so even more by 'advertising' its content and encouraging the full reading of the paper". Several studies also considered that the length of paper's abstract influenced the citations. Weinberger et al. (2015) constructed large abstract corpus from eight disciplines and found that papers with fewer words and fewer sentences in the abstract induce less citations, while short sentences have a positive impact on the citations of the papers only in Mathematics and Physics fields. On the contrary, Letchford et al. (2016) declared that shorter abstracts with more common words are cited more slightly.

All the above literature neglects the impact of the context semantic information contained in the title and abstract metadata text on the long-term citation count, thus our main target is to construct the comprehensive semantic features of the metadata text. In this task, an important problem is how to effectively extract semantic features from the metadata. Different text representation methods extract different text features, several existing popular methods for extracting text features include Latent Dirichlet Allocation (LDA) (Hu et al. 2020), bag of words (BOW) model, Term Frequency-Inverse Document Frequency (TF-IDF) (Yahav et al. 2019) and Word2Vec (Li et al., 2018; Zhang et al., 2018), etc. These existing text feature extraction techniques are word-level, and they often take the sum (TF-IDF weighted or not) or average of the feature vectors of words in a sentence when constructing the sentence-level feature vector. Both two methods have a disadvantage that they do not consider the order of the words in a sentence. As an extension of Word-2Vec (Mikolov et al., 2013), Doc2Vec (Le & Mikolov, 2014) can take the semantic relationship between words in a sentence into account to convert the sentence into a numeric vector with semantic information. It is widely used in many natural language processing (NLP) tasks such as sentiment analysis and text documents clustering (Aikawa et al., 2019; Karvelis et al., 2018; Lau & Baldwin, 2016; Markov et al., 2017; Stiebellehner et al., 2018) and achieves superior performance. Therefore Doc2Vec model is adopted in our work.

Dataset

In recent years, many outstanding academic achievements have emerged in artificial intelligence field. Hence, we selected a total of 20 journals at A, B levels in the field of Artificial Intelligence from China Computer Federation catalogue 2019, they are Artificial Intelligence (AI), IEEE Trans on Pattern Analysis and Machine Intelligence (TPAMI), International Journal of Computer Vision (IJCV), Journal of Machine Learning Research (JMLR), Autonomous Agents and Multi-Agent Systems (AAAMS), Computer Vision and Image Understanding (CVIU), Data and Knowledge Engineering (DKE), IEEE Transactions on Audio, Speech, and Language Processing (TASLP), IEEE Transactions on Evolutionary Computation (TEC), IEEE Transactions on Fuzzy Systems (TFS), International Journal of Approximate Reasoning (IJAR), Journal of Artificial Intelligence Research (JAIR), Journal of Speech, Language, and Hearing Research (JSLHR) respectively. An academic paper may have different citations in various literature databases due to the different data sources of the literature databases. As for this problem, we consider using the widely-used Scopus database which is a scientific literature database to obtain the required data in our experiment. Scopus covers many well-known journal and conference papers, and it is the largest literature and citations database in the world. Another important reason for choosing Scopus as our data source is that Scopus adopts the powerful author name disambiguation algorithm, which provides each author with an unique Elsevier EID. We choose the pybliometrics python

Table 1 Descriptive statistics of the dataset

		2000	2001	2002	2003	2004	2005	2006	Total
Publications		1,114	1,145	1,225	1,321	1,293	1,372	1,647	9,117
Sentences		9,477	9,250	9,712	10,685	10,364	11,026	13,571	74,085
Citations	Maximum	4,864	9,597	11,555	10,105	28,760	4,856	7,499	28,760
	Average	93.363	87.615	111.576	100.960	130.847	90.881	92.794	101.029
	Std. dev	287.547	347.636	473.971	373.951	855.009	243.541	304.751	453.109
		Training set					Test set		

library developed by Rose and Kitchin (2019) to extract the title and abstract along with the citation sequence which included the citations for 14 years from the year when the paper was published. Assuming that paper A was cited by paper B, in this case, if paper A receives more than 1 citation by the paper B, we took the citation count of paper A caused by paper B as 1 time. We summarized the detail statistics of the dataset in Table 1. As shown in Table 1, there were 9,117 papers in the dataset, and the number of the sentences of the papers published in each year was counted. These papers were published from 2000 to 2006, and we counted the citation count of each paper in the 14-year citation window. And we excluded the paper which has no citations. Lastly, We considered 6098 published papers between 2000 and 2004 as the training set, and 3019 papers published from 2005 to 2006 as the test set. Since the dataset has a limited number of papers, we only randomly select 10% training data as the validation set.

Figure 1 shows the citation pattern of the sampled papers. We randomly select one paper from each journal. According to Fig. 1, we can intuitively find each sample has its own pattern of citations which is slightly different from others. How to design a citation prediction method which leads to better results? It deserves study.
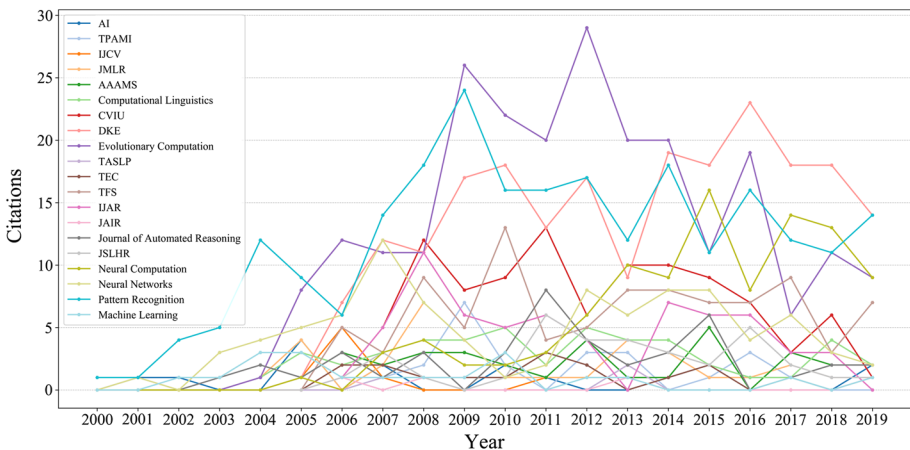

Fig.1 Annual citation count of 20 sampled papers

Methodology

Problem definition

Our task is to predict the long-term citation count of an academic paper based on its metadata semantic features and early citations. In other words, we define the set of metadata semantic features as F which is extracted from the title and abstract, and then the citation count of the paper in year t after its publication is denoted as c_t. Therefore, we want to predict the future citation count $c_{k+1}, c_{k+2}, ..., c_n$ of the paper according to its known early citations $c_0, c_1, ..., c_k$ and the feature set F. Figure 2 shows the basic architecture of the proposed model.
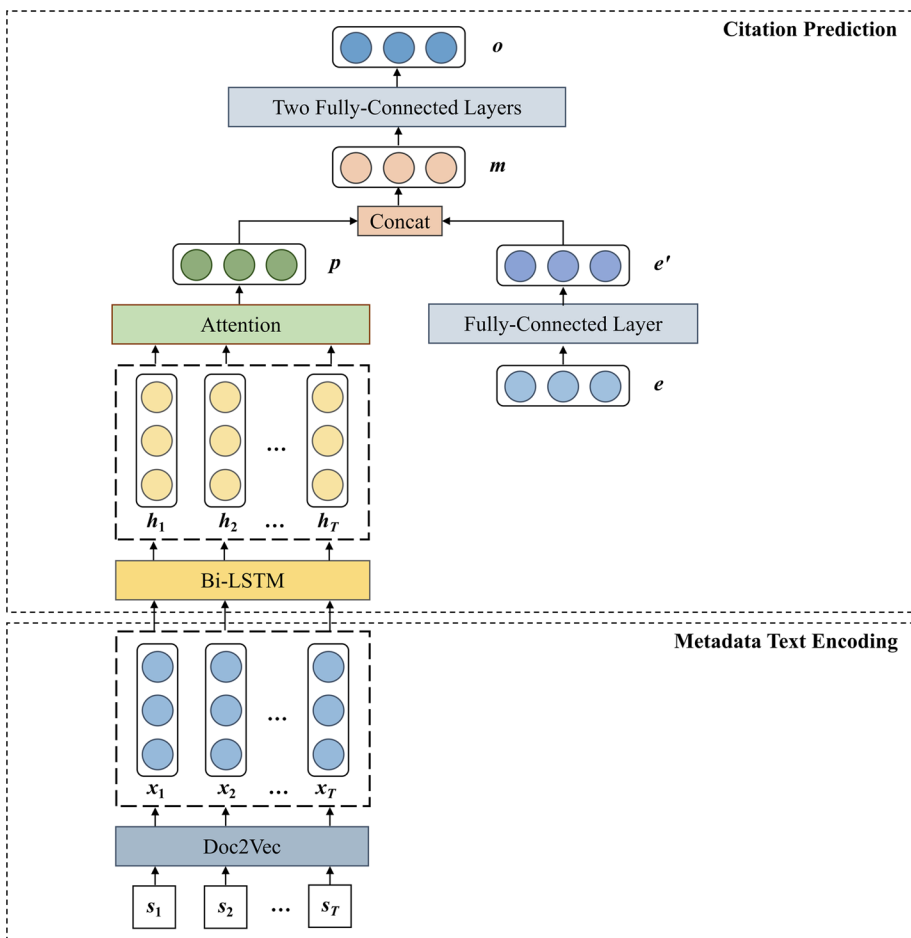

Fig. 2 The basic architecture of the proposed prediction model

Deep learning-based predictive model

This section presents our deep learning-based predictive model. Figure 2 shows the basic architecture of the proposed model.

Metadata sentence encoding

As Fig. 2 shows, the first step is metadata sentence encoding. We employ Doc2Vec (Le & Mikolov, 2014), which is an extension of Word2Vec, to extract semantic features from the metadata sentences. Doc2Vec is a neural probabilistic language model based on the distributional hypothesis, which states that words with similar contexts have similar semantic meanings. Doc2Vec considers the semantic relationship between words in a sentence to convert the sentence into a unique vector with semantic information by a shallow neural network, and the word vectors are shared among all the sentences.

We first introduce Word2Vec algorithm, the foundation of Doc2Vec algorithm. The training objective of Word2Vec algorithm is to maximize the average log probability given the sequence of words $w_1, w_2, w_3, \ldots, w_T$:

$$\frac{1}{T} \sum_{t=k}^{T-k} \log p\left(w_t | w_{t-k}, \ldots, w_{t+k}\right) \tag{1}$$

Thus we can predict the w_T by the softmax multi-class classifier:

$$p\left(w_t | w_{t-k}, \ldots, w_{t+k}\right) = \frac{e^{y_{wt}}}{\sum_i e^{y_{wi}}} \tag{2}$$

where y_{wi} is the un-normalized log-probability of the output word w_i, and the calculating formulation of y_{wi} is denoted as:

$$y_{wi} = b + Vf\left(w_{i-k}, \ldots, w_{i+k}; W\right) \tag{3}$$
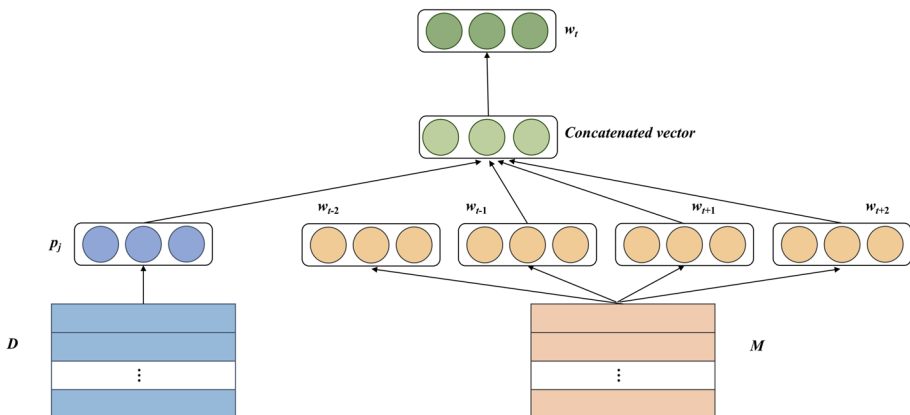

Fig. 3 The framework for Doc2Vec algorithm

where b is the bias vector, V is the weight matrix and f is the operation of concatenating or averaging the word vectors extracted from the matrix W.

As shown in Fig. 3, Doc2Vec algorithm concentrates the paragraph vector with the word vectors for predicting the target word vector. Hence, the difference from Word2Vec is that y_{wi} is computed as:

$$y_{wi} = b + Vf\left(w_{i-k}, \dots, w_{i+k}, p_i; W, D\right) \tag{4}$$

where p_i is the current sentence vector extracted from the matrix D.

First, we perform the preprocessing steps on the sentences in the metadata text including removing stopwords, lemmatization, word segmentation and filter punctuation. Two ways are presented for generating the sentence-level numeric vector: Distributed Bag of Words (DBOW) and Distributed Memory (DM), we use DM algorithm to vectorize each sentence in the metadata text rather than the whole metadata paragraph in order to make the extracted text features contain richer semantic information. As a result, each sentence is represented as a sentence vector with 200 dimensions. In other words, the sentence sequence $\{s_1, s_2, \dots, s_T\}$ is transformed into the sequence of sentence vectors $\{x_1, x_2, \dots, x_T\}$.

Citation prediction

As Fig. 2 shows, the second step is citation prediction. Inspired by the previous work (Zhou et al., 2016), we use Bi-LSTM with attention mechanism to further extract high-level features from the sequence of sentence vectors generated by Doc2Vec in this step.

LSTM is a kind of neural network used to process sequence data which was introduced by (Hochreiter & Schmidhuber, 1997). As one of the variants of RNN, LSTM is capable of capturing the long-range dependence in sequence data. It contains three kinds of gates, i.e., forget gate f_t, input gate i_t and output gate o_t at time step t. These gates are used to control the cell state c_t. f_t determines which information in the previous cell state c_{t-1} will be discarded. The calculation formula of f_t is given as

$$f_t = \sigma\left(W_{fx}x_t + W_{fh}h_{t-1} + b_f\right) \tag{5}$$

where W_{fx} and W_{fh} are the weight matrices, b_f is bias vector, h_{t-1} is the hidden state at the previous time step t-1 and σ is the sigmoid activation function. i_t decides which information can be added to the current cell state c_t. The calculation formula of i_t can be expressed as

$$i_t = \sigma\left(W_{ix}x_t + W_{ih}h_{t-1} + b_i\right) \tag{6}$$

where W_{ix} and W_{ih} are the weight matrices, b_i is the bias vector. Then the current cell state ct can be updated using the previous cell state c_{t-1} and the new candidate information $\tilde{c}_t$. The update concept is denoted as

$$\tilde{c}_t = \tanh\left(W_{cx}x_t + W_{ch}h_{t-1} + b_c\right) \tag{7}$$

$$c_t = f_t c_{t-1} + i_t \tilde{c}_t \tag{8}$$

where W_{cx} and W_{ch} are the weight matrices, b_c is the bias vector. Lastly, o_t maps the current cell state c_t to the hidden state h_t.

$$o_t = \sigma\left(W_{ox}x_t + W_{oh}h_{t-1} + b_o\right) \tag{9}$$

$$h_t = o_t \tanh\left(c_t\right) \tag{10}$$

where W_{ox} and W_{oh} are the weight matrices, b_o is the bias vector. Bi-LSTM consists of two LSTM layers, one is in the forward direction from left to right and the other is in the backward direction from right to left. Hence, the sequence of final hidden state vectors $H = \{h_1, h_2, \ldots, h_T\}$ is computed as

$$H = \left[\overrightarrow{H} \oplus \overleftarrow{H}\right] \tag{11}$$

where $\overrightarrow{H} = \{\vec{h}_1, \vec{h}_2, \ldots, \vec{h}_T\}, \overleftarrow{H} = \{\overleftarrow{h}_1, \overleftarrow{h}_2, \ldots, \overleftarrow{h}_T\}$, and $\oplus$ means the element-wise sum operation. As a consequence, a sequence of 128-dimensional hidden state vectors H containing high-level features of sentences is obtained after using Bi-LSTM.

Next, we use the attention neural network to merge the sequence of sentence vectors H produced by Bi-LSTM into a semantic vector p representing the whole metadata paragraph. Attention mechanism (Bahdanau et al. 2015) simulates an important characteristic of human perception, i.e., it usually focuses on the certain parts of the text instead of the whole text. Attention mechanism assigns different weight for each sentence vector in the sequence according to the importance of sentence to capture the significant semantic information in the metadata text. The vector p is computed by

$$\alpha = soft\max\left(H^T W_\alpha h_T\right) \tag{12}$$

$$p = \tanh\left(W_p H\alpha + W_p h_T + b_p\right) \tag{13}$$

where W_α is the weight matrix of α, W_p and b_p are the weight matrix and bias vector of p respectively, H^T is the transpose of H, and h_T is the hidden state vector of Bi-LSTM at the last time step. By using the attention neural network, we get the final paragraph-level vector p used for citation prediction. We set the dimensions of the vector p as 128.

After the attention neural network, we apply a fully-collected (FC) layer with 32 neurons to the early citations vector e. Then, a vector m generated by concatenating the vector p and the vector e' is adopted for predicting the future citation count of the paper.

$$m = [p, e'] \tag{14}$$

Finally, two FC layers are added to the model. The former layer containing 16 neurons is for enhancing the learning ability of the model, and the latter layer with several neurons is the output layer which gives the citation prediction result o.

$$m' = relu\left(W'm + b'\right) \tag{15}$$

$$o = W_o m' + b_o \tag{16}$$

where W' and b' are the weight matrix and bias vector of m', W_o and b_o are the weight matrix and bias vector of the output o. As a result, we obtain the predicted value of the future citation count of the paper.

We choose Adam optimization algorithm to train the model, and the value of initial learning rate is 0.005 with a decrease value of 3×10^{-6}. In terms of loss function, we

use the Mean Squared Error (MSE) for training the proposed model. And the model is trained on a single GTX-1080Ti GPU for 300 epochs with batches of 64 samples, and the training time of one epoch is 2 seconds. In addition, all FC layers except the output layer use the Rectified Linear Unit (ReLU) activation function for adding some nonlinearity to the outputs. After a series of experiments, we take the set of hyper-parameters with the highest prediction accuracy as the final hyper-parameters of the proposed model.

Results and discussion

Model measurement metrics

We choose four common evaluation metrics in regression problem to evaluate the performance of the proposed model. They are Root Mean squared error (RMSE), Mean absolute error (MAE), coefficient of determination (R^2) and Normalized Discounted cumulative gain (NDCG)@m respectively. RMSE is more sensitive to the exception value. MAE measures the deviation between the predicted values and the actual values. R^2 measures the fitness of the prediction model. NDCG considers the order of the predicted top m highly-cited papers, it gives the paper with higher rank a greater weight. The definitions of four metrics are as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{17}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{18}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{19}$$

$$NDCG@m = \frac{\sum_{i=1}^{m} \frac{p_i}{\log_2 (i+1)}}{\sum_{i=1}^{m} \frac{a_i}{\log_2 (i+1)}} \tag{20}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, $\bar{y}$ the average of all actual values, n is the number of samples, p_i is the actual value of the i-th paper in the predicted highly-cited papers and a_i is the actual value of the i-th paper in the actual highly-cited papers.

Comparison results with baselines

For evaluating the proposed model, we compare its performance against five baselines. The evaluation results are measured by annual RMSE, MAE and R^2 values of the predicted citation count in subsequent 8 years respectively. We select Gradient Boosting Regression Trees (GBRT) (Friedman, 2001), XGBoost (Chen & Guestrin, 2016), Bi-LSTM (Graves, 2012), NNCP (Abrishami & Aliakbary, 2019) and the BP neural prediction model proposed by Ruan et al. (2020). We use the citation count for 6 years from the year when the paper was published, along with metadata semantic features to predict the citation count

for the subsequent 8 years, i.e., $n = 13$ and $k = 5$ (refer to Problem definition Section). It should be emphasized that GBRT and XGBoost models only use the embedding vector of the paragraph-level metadata text generated by Doc2Vec algorithm. Bi-LSTM model takes the hidden state h_T generated by the last time step T as the paragraph-level vector. Moreover, NNCP model only takes early citations as the input information, and it has already outperformed the existing methods, such as Cao et al. (2016). Notably, the prediction target of Ruan's model is the five-year citation count, which is a single value, rather than a sequence of of citation counts in consecutive years. Therefore, we convert Ruan's model into a mutioutput regression model, and the input is consistent with that of the GBRT and XGBoost model.

Figure 4 illustrates the comparison results of the six prediction models. We can observe that the proposed model (named BIL_A) outperforms all competing models with three metrics, and as time increases, the prediction performances of all the five models show a downward trend. GBRT and XGBoost baselines perform worst, and Ruan's model perform better than them slightly, which indicates that extracting the sentence-level metadata semantic features can get more precise semantic information compared with extracting the paragraph-level metadata semantic features. Additionally, our model using attention mechanism can further capture key information from the metadata text of the paper compared with Bi-LSTM model. By comparing the proposed model with NNCP model, we know that metadata semantic features are effective for improving the accuracy of the citation prediction model.

In order to evaluate the accuracy in predicting citations of highly-cited papers, we selected top 20, 50 and 100 highly-cited papers respectively from each journal, i.e., $m = 20$, $m = 50$ and $m = 100$. The evaluation results are measured by the average of NDCG@20, NDCG@50 and NDCG@100 values of the predicted citation count in subsequent 8 years respectively. Figure 5 shows the citation prediction results on highly-cited papers according to all three NDCG@m values. The proposed model still has a better predictive performance against the five baselines, NDCG@20 is 84.4%, NDCG@50 is 87.0% and NDCG@100 is 87.5%. And the average accuracy of all the baselines for the prediction of highly-cited papers has also reached over 68.2%.

Analysis of the usefulness of metadata semantic features

To further examine whether or not metadata semantic features are useful for the citation prediction performance improvement, we design two different feature sets, i.e., early citations (E) and early citations + metadata semantic features (E + M) as the input data of GBRT, XGBoost, Ruan's model and our model respectively (Bi-LSTM structure is the same as our model structure after removing the metadata semantic features). The evaluation results are measured by the average of RMSE and R^2 values of the predicted citation count in subsequent 8 years respectively. As shown in Table 2, metadata semantic features contribute to improving the prediction performance of all four models. It is worth noting that the performance of the proposed model is indeed improved with RMSE from 16.460 to 15.677 and R from 0.712 to 0.739. This experimental result reveals that metadata semantic features can improve the accuracy of future citation count prediction of academic papers to an extent. Furthermore, we also find that the prediction performance of the proposed model with early citations input is also acceptable, thus both two categories of features are necessary. Additionally, the proposed model can get the best performance in terms of both two different feature sets, which can also demonstrate the capability of our model.
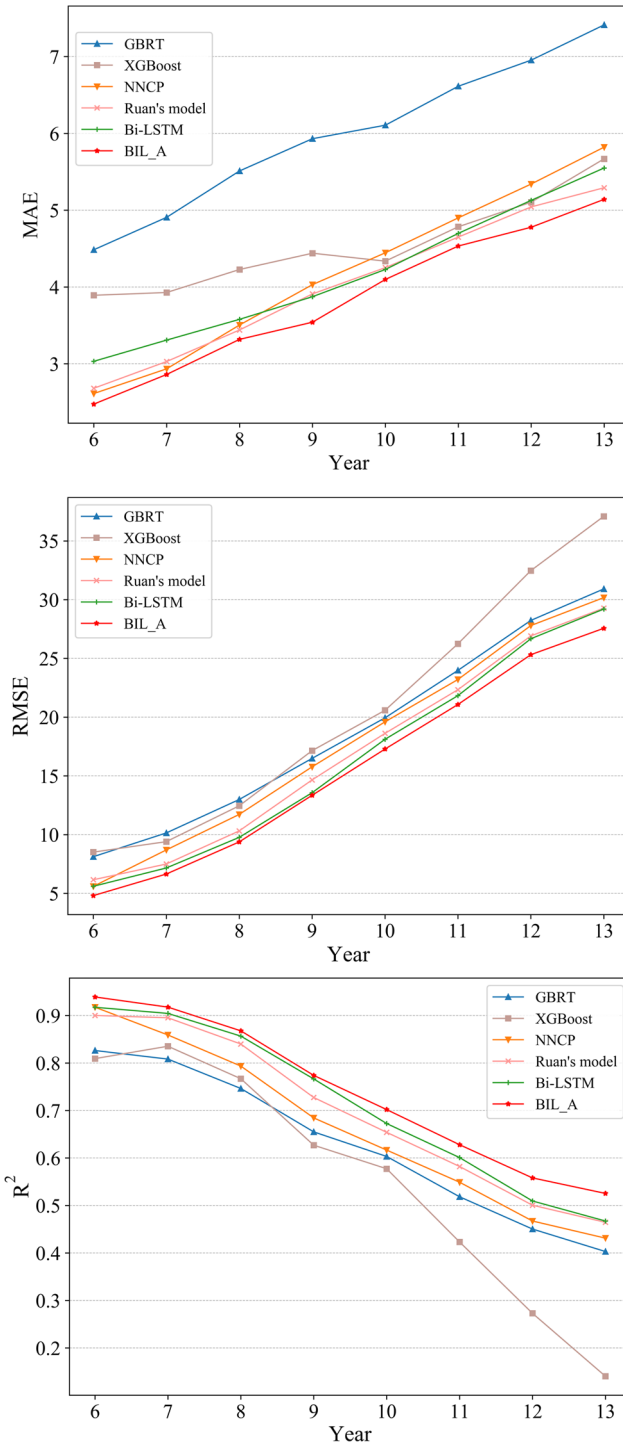
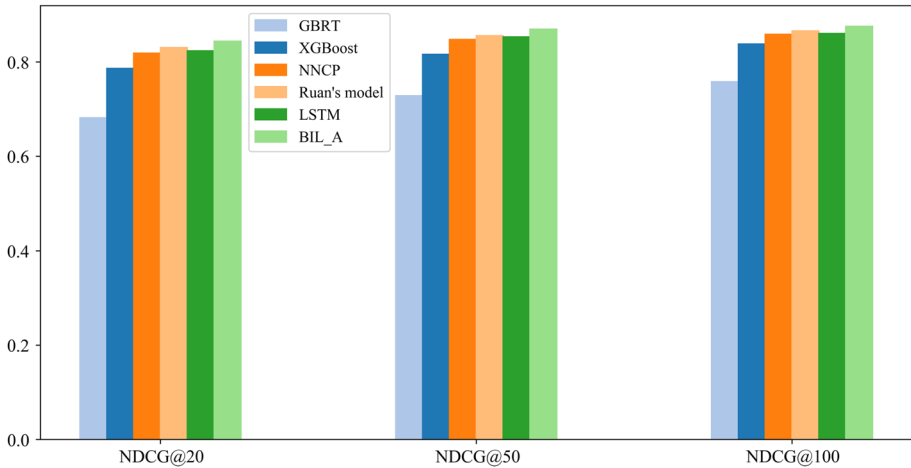**Fig. 4** Comparison results for the six prediction models

Fig. 5 Comparison results of citation prediction for highly-cited papers

Table 2 The prediction results of the proposed model under different feature sets

	GBRT		XGBoost		Ruan's model		BIL_A	
RMSE	E	19.029	E	20.698	E	18.659	E	16.460
	$E+M$	18.851	$E+M$	20.487	$E+M$	16.972	$E+M$	15.677
R^2	E	0.618	E	0.548	E	0.632	E	0.712
	$E+M$	0.626	$E+M$	0.557	$E+M$	0.695	$E+M$	0.739

Note: E, early citations; M, metadata semantic features

Table 3 The prediction performance of the proposed model under different length of early citation sequence

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$ (Our model)
RMSE	30.037	29.079	25.175	23.392	21.408	18.163	15.677
R^2	0.046	0.074	0.323	0.422	0.517	0.652	0.739

Note: k is the length of early citation sequence

The effect of different early citation sequence length on prediction performance

Which length of early citations can cause the best prediction performance? In the final experiment, we analyze the effect of different length of early citation sequence on prediction performance of the proposed model. In order to facilitate analysis, the evaluation results are measured by the average of RMSE and R^2 values of the predicted citation count in subsequent 8 years respectively. As Table 3 shows, with the length of early citation sequence increasing, the prediction performance of the propose model improves, and this fact is consistent with the experiment result of Abrishami and Aliakbary (2019). And we also find that only using semantic information performs worst, which indicates that semantic information cannot be used independently but give an auxiliary support for predicting long-term citation count.

Discussion

Citation prediction is still a challenging task, effective model for recognizing citation pattern is necessary for this problem. For solving this problem, we design a deep-learning based future citation prediction model using paper metadata text and early citations. The first experimental result indicates that our proposed model is more predictive than other five baselines, and has the ability to learn the comprehensive semantic representation of the text. Among these five baselines, GBRT and XGBoost models are traditional machine learning models, while the proposed model based on deep learning techniques is more powerful and has better performance than these traditional machine learning models. Furthermore, the proposed model takes sentence-level vectors as input rather than vectorizing the whole metadata paragraph. By using several repeated memory units, Bi-LSTM can model the text by using the transferred information from the first time step to the last time step and then uses attention mechanism to further extract important features from the sequence of hidden state vectors. Therefore, this method can improve the accuracy of the proposed citation prediction model. We note that NNCP model only takes early citations as the input information, which may be not sufficient for learning the citation patterns of academic papers. Therefore, more categories of features are necessary for the citation prediction task. However, we also find that NNCP model still has higher prediction accuracy than GBRT and XGBoost models according to MAE, RMSE and R^2 metrics. We speculate that both inputs and outputs of this model are the sequence of the citations, and NNCP adopts the sequence-to-sequence model built with "Encoder-Decoder" architecture, thus it can learn the well-fitting prediction function from the citation data. Although Ruan's model takes a four-layer neural network (including one input layer, two hidden layer and one output layer), it does not seem to be more suitable for extracting the semantic information from the metadata text than our model. It would be more applicable for learning high-level features from the independent feature set rather than the data with sequence pattern. To a certain degree, highly-cited papers stand for the authority in the field, and draw a lot of attention among academic papers. Therefore, it is particularly important to accurately predict the highly-cited papers. Our proposed model demonstrates the outstanding effectiveness with higher NDCG@20, NDCG@50 and NDCG@100 values.

We can know from the second and third experiments that metadata semantic features are essential for improving the citation prediction models, but only using semantic features cannot achieve accurate prediction, which indicates that metadata semantic information is not the most important indicator in terms of effectiveness. Therefore, metadata semantic information plays a support role in citation prediction task. Generally, metadata summarizes the research, and the semantic information contained in the metadata text of a paper attracts scholars to read and cite it to an extent. In addition, since the traditional machine learning methods treat the text embedding as an independent feature set, these methods cannot extract the semantic dependencies between the sentence embeddings. By contrast, the proposed model based on deep learning techniques can automatically learn better semantic representation of the metadata text, and thus it is more effective. And when text semantic features are discarded, our model degenerates into a model with only two-layer neural network, thus the prediction performance is notably reduced. However, our model still outperforms Ruan's model in this situation, which illustrates a point that when 6-year early citations input matches with a 4-layer neural network structure, over-fitting may occur. In summary, the proposed method is more suitable for extracting high-level text features with semantic information. This is also the main reason why many NLP tasks employ attention mechanism in text sequence modeling techniques. It is important to emphasize

that these models only with citations input also get acceptable prediction accuracy, and in terms of the proposed model, the prediction performance is much better than that only with semantic information. Therefore, early citation count of a paper is an important feature for predicting the long-term citations (Abrishami & Aliakbary, 2019; Cao et al., 2016; Newman, 2014). When both metadata semantic features and early citations are taken as the input information, all the four models perform better. Therefore, combining metadata semantic features and early citations can provide more sufficient information to recognize more comprehensive citation pattern of academic papers.

Conclusion

In this paper, we propose a novel model based on paper metadata text to predict the future citation count. To the best of our knowledge, this is one of the first studies constructing the semantic representation from the metadata text of the academic paper for the citation prediction problem. Specifically, the sentences in the metadata text are encoded with Doc2Vec algorithm, and then the paragraph-level semantic features are extracted from the sentence embeddings by Bi-LSTM with attention mechanism. Lastly, early citations are also integrated to tackle the citation prediction task. We have shown that our proposed model outperforms the existing state-of-the-art models, and metadata semantic features are useful for the citation prediction performance improvement. Our study provides a promising approach for the citation prediction task.

The limitation of this work is that it only explores the journal papers in artificial intelligence field, the experiment results would not be general to academic papers in other fields. As the future study, we will collect more papers to expand the experimental dataset for learning a better citation prediction method. Although our results are promising, we intend to apply Transformer-based models, such as BERT (Devlin et al., 2019), RoBERTa (Liu et al., 2019) and ELECTRA (Clark et al., 2020) in future work to further enhance the prediction effectiveness of the model. Finally, we will also integrate more indicators such as author's h-index, author's citation count, journal impact factor and altmetrics information about papers to improve the performance of the proposed model, and try to employ our model for other valuable applications, such as identifying highly-cited papers, predicting academic rising star and forecasting high-impact research institutions.

Authors' contributions AM Conceptualization, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing—Original Draft, Writing—Review & Editing, Visualization. YL Writing—Review & Editing, Supervision, Project administration, Funding acquisition. XX Writing—Review & Editing, Supervision, Project administration. TD Methodology, Software, Formal analysis, Writing—Review & Editing.

Data availability The code used for this study is available at https://github.com/MatrixLabDUT/CitationPrediction.

Code availability The code used for this study is available at https://github.com/MatrixLabDUT/CitationPrediction.

Declaration

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

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. https://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. https://doi.org/10.1016/j.joi.2019.02.01

Aikawa, K., Kawai, S., & Nobuhara, H. (2019). Multilingual Inappropriate Text Content Detection System Based on Doc2vec. In: *2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)*, pp. 441–442. https://doi.org/10.1109/GCCE46687.2019.9015579

Bahdanau, D., Cho, K. H., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate.In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–15.

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

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. https://doi.org/10.1016/j.joi.2013.11.005

Bornmann, L., Schier, H., Marx, W., & Daniel, H. D. (2012). What factors determine citation counts of publications in chemistry besides their quality? *Journal of Informetrics, 6*(1), 11–18. https://doi.org/10.1016/j.joi.2011.08.004

Braun, T., Glänzel, W., & Schubert, A. (2006). A Hirsch-Type Index for Journals. *Scientometrics, 69*(1), 169–173. https://doi.org/10.1007/s11192-006-0147-4

Cao, X., Chen, Y., & Ray Liu, K. J. (2016). A data analytic approach to quantifying scientific impact. *Journal of Informetrics, 10*(2), 471–484. https://doi.org/10.1016/j.joi.2016.02.006

Chen, J. (2015). Predicting Citation Counts of Papers.In: *2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC)*, pp. 434–440. https://doi.org/10.1109/ICCI-CC.2015.7259421

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Chen, Y., Huang, S., Lee, H., Wang, Y., & Shen, C. (2019). Audio Word2vec : Sequence-to-sequence autoencoding for unsupervised learning of audio segmentation and Representation. *IEEE/ACM Transactions on Audio Speech and Language Processing, 27*(9), 1481–1493. https://doi.org/10.1109/TASLP.2019.2922832

Clark, K., Luong, M.-T., Le, Q. V, & Manning, C. D. (2020). ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. In: BT - 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26–30, 2020. https://openreview.net/forum?id=r1xMH1Btv

Clauset, A., Larremore, D. B., & Sinatra, R. (2017). Data-driven predictions in the science of science. *Science, 355*(6324), 477–480. https://doi.org/10.1126/science.aal4217

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2–7, 2019, Vol 1 (pp. 4171–4186). https://doi.org/10.18653/v1/n19-1423

Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics, 69*(1), 131–152. https://doi.org/10.1007/s11192-006-0144-7

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232.

Fronzetti Colladon, A., D'Angelo, C. A., & Gloor, P. A. (2020). Predicting the future success of scientific publications through social network and semantic analysis. *Scientometrics, 124*(1), 357–377. https://doi.org/10.1007/s11192-020-03479-5

Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA, 295*(1), 90–93. https://doi.org/10.1001/jama.295.1.90

Graves, A. (2012). *Supervised sequence labelling with recurrent neural networks*. Heidelberg: Springer.

Guo, J., Lu, S., Cai, H., Zhang, W., Yu, Y., & Wang, J. (2018). Long text generation via adversarial training with leaked information.In: *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, pp. 5141–5148.

Habibzadeh, F., & Yadollahie, M. (2010). Are shorter article titles more attractive for citations? Cross-sectional study of 22 scientifc journals. *Croatian Medical Journal, 51*(2), 165–170. https://doi.org/10.3325/cmj.2010.51.165

Haggan, M. (2004). Research paper titles in literature, linguistics and science: Dimensions of attraction. *Journal of Pragmatics, 36*(2), 293–317. https://doi.org/10.1016/S0378-2166(03)00090-0

Hassan, S. U., Bowman, T. D., Shabbir, M., Akhtar, A., Imran, M., & Aljohani, N. R. (2019). Influential tweeters in relation to highly cited articles in altmetric big data. *Scientometrics, 119*(1), 481–493. https://doi.org/10.1007/s11192-019-03044-9

Hirsch, J. E. (2005). An index to quantify an individual' s scientific research output. *Proceedings of the National Academy of Sciences of the United States of America, 102*(46), 16569–16572. https://doi.org/10.1073/pnas.0507655102

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hu, Y.-H., Tai, C.-T., Liu, K. E., & Cai, C.-F. (2020). Identification of highly-cited papers using topic-model-based and bibliometric features: the consideration of keyword popularity. *Journal of Informetrics, 14*(1), 101004. https://doi.org/10.1016/j.joi.2019.101004

Jamali, H. R., & Nikzad, M. (2011). Article title type and its relation with the number of downloads and citations. *Scientometrics*, *88*(2), 653–661. https://doi.org/10.1007/s11192-011-0412-z.

Jati, A., & Georgiou, P. (2019). Neural predictive coding using convolutional neural networks toward unsupervised learning of speaker characteristics. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 27*(10), 1577–1589. https://doi.org/10.1109/TASLP.2019.2921890

Karvelis, P., Gavrilis, D., Georgoulas, G., & Stylios, C. (2018). Topic recommendation using Doc2Vec. *International Joint Conference on Neural Networks (IJCNN), 2018*, 1–6. https://doi.org/10.1109/IJCNN.2018.8489513

Lau, J. H., & Baldwin, T. (2016). An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation. In: *Proceedings of the 1st Workshop on Representation Learning for {NLP}*, pp. 78–86. https://doi.org/10.18653/v1/W16-1609

Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In: *31st International Conference on Machine Learning, ICML 2014*, vol. *4*, pp. 2931–2939.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Letchford, A., Preis, T., & Moat, H. S. (2016). The advantage of simple paper abstracts. *Journal of Informetrics, 10*(1), 1–8. https://doi.org/10.1016/j.joi.2015.11.001

Li, S., Hu, J., Cui, Y., & Hu, J. (2018). DeepPatent: Patent classification with convolutional neural networks and word embedding. *Scientometrics, 117*(2), 721–744. https://doi.org/10.1007/s11192-018-2905-5

Li, M., Xu, J., Ge, B., Liu, J., Jiang, J., & Zhao, Q. (2019a). A Deep Learning Methodology for Citation Count Prediction with Large-scale Biblio-Features. In: *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 1172–1176. https://doi.org/10.1109/SMC.2019.8913961

Li, S., Zhao, W. X., Yin, E. J., & Wen, J.-R. (2019b). A neural citation count prediction model based on peer review text. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 4914–4924). Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1497

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., et al. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *CoRR*. http://arxiv.org/abs/1907.11692

Markov, I., Gómez-Adorno, H., Posadas-Durán, J.-P., Sidorov, G., & Gelbukh, A. (2017). Author profiling with doc2vec neural network-based document embeddings. In O. Pichardo-Lagunas & S. Miranda-Jiménez (Eds.), *Advances in Soft Computing* (pp. 117–131). Springer International Publishing.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In: *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*, pp. 1–12.

Newman, M. E. J. (2014). Prediction of highly cited papers. *EPL (europhysics Letters), 105*(2), 28002. https://doi.org/10.1209/0295-5075/105/28002

Platanios, E. A., Sachan, M., Neubig, G., & Mitchell, T. M. (2020). Contextual parameter generation for universal neural machine translation.In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*, (2016), pp. 425–435. Doi: https://doi.org/10.18653/v1/d18-1039

Rose, M. E., & Kitchin, J. R. (2019). pybliometrics: scriptable bibliometrics using a Python interface to Scopus. *SoftwareX, 10*, 100263. https://doi.org/10.1016/j.softx.2019.100263

Ruan, X., Zhu, Y., Li, J., & Cheng, Y. (2020). Predicting the citation counts of individual papers via a BP neural network. *Journal of Informetrics, 14*(3), 101039. https://doi.org/10.1016/j.joi.2020.101039

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. https://doi.org/10.1007/s11192-016-2161-5

Stegehuis, C., Litvak, N., & Waltman, L. (2015). Predicting the long-term citation impact of recent publications. *Journal of Informetrics, 9*(3), 642–657. https://doi.org/10.1016/j.joi.2015.06.005

Stiebellehner, S., Wang, J., & Yuan, S. (2018). Learning Continuous User Representations through Hybrid Filtering with doc2vec. *CoRR*. Retrieved from http://arxiv.org/abs/1801.00215

Tang, J., Lu, Z., Su, J., Ge, Y., Song, L., Sun, L., & Luo, J. (2019). Progressive Self-Supervised Attention Learning for Aspect-Level Sentiment Analysis. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 557–566. Doi: https://doi.org/10.18653/v1/P19-1053

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. https://doi.org/10.1016/j.joi.2012.06.002

Wang, F., Fan, Y., Zeng, A., Di, Z., Wang, M., Yu, G., et al. (2019a). Can we predict ESI highly cited publications? *Journal of Informetrics, 118*(1), 109–125. https://doi.org/10.1007/s11192-018-2965-6

Wang, M., Wang, Z., & Chen, G. (2019b). Which can better predict the future success of articles? Bibliometric indices or alternative metrics. *Scientometrics, 119*(3), 1575–1595. https://doi.org/10.1007/s11192-019-03052-9

Wang, Z., Zheng, L., Li, Y., & Wang, S. (2019c). Linkage Based Face Clustering via Graph Convolution Network. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, *1*(c), pp. 1117–1125. https://doi.org/10.1109/CVPR.2019.00121

Weinberger, C. J., Evans, J. A., & Allesina, S. (2015). Ten simple (empirical) rules for writing science. *PLOS Computational Biology, 11*(4), 1–6. https://doi.org/10.1371/journal.pcbi.1004205

Wen, Y., Zhang, K., Li, Z., & Qiao, Y. (2019). A Comprehensive study on center loss for deep face recognition. *International Journal of Computer Vision, 127*(6–7), 668–683. https://doi.org/10.1007/s11263-018-01142-4

Wu, Z., Lin, W., Liu, P., Chen, J., & Mao, L. (2019). Predicting long-term scientific impact based on multifield feature extraction. *IEEE Access, 7*, 51759–51770. https://doi.org/10.1109/ACCESS.2019.2910239

Xiao, S., Yan, J., Li, C., Jin, B., Wang, X., Yang, X., et al. (2016). On Modeling and Predicting Individual Paper Citation Count over Time. In S. Kambhampati (Ed.), *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, {IJCAI} 2016, New York, NY, USA, 9–15 July 2016* (pp. 2676–2682). {IJCAI/AAAI} Press. http://www.ijcai.org/Abstract/16/380

Yahav, I., Shehory, O., & Schwartz, D. (2019). Comments mining with TF-IDF: The inherent bias and its removal. *IEEE Transactions on Knowledge and Data Engineering, 31*(3), 437–450. https://doi.org/10.1109/TKDE.2018.2840127

Yan, E., & Ding, Y. (2010). Measuring scholarly impact in heterogeneous networks. *Proceedings of the American Society for Information Science and Technology, 47*(1), 1–7. https://doi.org/10.1002/meet.14504701033

Yan, R., Huang, C., Tang, J., Zhang, Y., & Li, X. (2012). To Better Stand on the Shoulder of Giants. In: *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 51–60). New York, NY, USA: Association for Computing Machinery. Doi:https://doi.org/10.1145/2232817.2232831

Yan, R., Tang, J., Liu, X., Shan, D., & Li, X. (2011). Citation Count Prediction: Learning to Estimate Future Citations for Literature. In: *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, pp. 1247–1252. Doi: https://doi.org/10.1145/2063576.2063757

Yu, L., Zhang, W., Wang, J., & Yu, Y. (2017). SeqGAN: Sequence generative adversarial nets with policy gradient.In: *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, pp. 2852–2858.

Yuan, S., Tang, J., Zhang, Y., Wang, Y., & Xiao, T. (2018). Modeling and Predicting Citation Count via Recurrent Neural Network with Long Short-Term Memory. *CoRR*, *abs/1811.0*. http://arxiv.org/abs/1811.02129

Zeng, J., Su, J., Wen, H., Liu, Y., Xie, J., Yin, Y., & Zhao, J. (2020). Multi-domain neural machine translation with word-level domain context discrimination. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*, 447–457. Doi: https://doi.org/10.18653/v1/d18-1041

Zhang, Y., Lu, J., Liu, F., Liu, Q., Porter, A., Chen, H., & Zhang, G. (2018). Does deep learning help topic extraction? A kernel k-means clustering method with word embedding. *Journal of Informetrics, 12*(4), 1099–1117. https://doi.org/10.1016/j.joi.2018.09.004

Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., & Xu, B. (2016). Attention-based bidirectional long short-term memory networks for relation classification. In: *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Short Papers*, pp. 207–212.Doi: https://doi.org/10.18653/v1/p16-2034

Zhu, S., Li, S., & Zhou, G. (2019). Adversarial Attention Modeling for Multi-dimensional Emotion Regression. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 471–480. Doi: https://doi.org/10.18653/v1/P19-1045