



Patent Quality Valuation with Deep Learning Models

Hongjie Lin, Hao Wang, Dongfang Du, Han Wu,
Biao Chang, and Enhong Chen^(✉)

Anhui Province Key Laboratory of Big Data Analysis and Application,
University of Science and Technology of China, Hefei, China
{hjljin,wanghao3,dfdu,wuhanhan,chbiao}@mail.ustc.edu.cn,
cheneh@ustc.edu.cn

Abstract. Patenting is of significant importance to protect intellectual properties for individuals, organizations and companies. One of practical demands is to automatically evaluate the quality of new patents, i.e., patent valuation, which can be used for patent indemnification and patent portfolio. However, to solve this problem, most traditional methods just conducted simple statistical analyses based on patent citation networks, while ignoring much crucial information, such as patent text materials and many other useful attributes. To that end, in this paper, we propose a Deep Learning based Patent Quality Valuation (DLPQV) model which can integrate the above information to evaluate the quality of patents. It consists of two parts: Attribute Network Embedding (ANE) and Attention-based Convolutional Neural Network (ACNN). ANE learns the patent embedding from citation networks and attributes, and ACNN extracts the semantic representation from patent text materials. Then their outputs are concatenated to predict the quality of new patents. The experimental results on a real-world patent dataset show our method outperforms baselines significantly with respect to patent valuation.

Keywords: Patent quality valuation · Attribute network embedding
Convolutional Neural Network · Patent citation network

1 Introduction

With regard to industry research and development, patent application is one of the most significant sources of key technologies for protecting intellectual properties. And patenting is also an important asset for companies in the knowledge economy. Over the past few decades, with the rapid development of multifaceted technology in different application domains, a large amount of patents are applied and authorized. They serve as one of the crucial intellectual property components for individuals, organizations and companies. Many compa-

nies, especially burgeoning firms, apply several thousands of patents each year¹. The granted patent information is open to public and can be available for professional organizations in various countries or regions around the world. For instance, World Intellectual Property Organization (WIPO)² reported over 2 million total patent applications authorized worldwide within a year [15]. The researches which contrapose patent information are more and more important in order to make fair and credible valuation results available to investors.

In fact, questions involving patent mining have intrigued scholars for decades, and there have been many influential academic researches in this area, including patent retrieval [7], patent classification [4], patent visualization and cross-language patent mining [8] and patent valuation [1, 10]. In this work, we devoted to exploring this deeper and hope to make further support for the last topic, patent valuation, a common process of evaluating the quality of patent documents.

Indeed, assessing the value of a patent is crucial both at the licensing stage and during the resolution of a patent infringement lawsuit [20], and it is undeniable that business community have paid much concern about this question because of its considerable significance, so they might hire many professional patent analysts engaging in this. Obviously, patent valuation is a non-trivial task which requires tremendous amount of human efforts. What's more, it is necessary for patent analysts to have a certain degree of expertise in different research domains, including information retrieval, data mining, domain-specific technologies, and business intelligence [32]. As a result, it is of great significance to evaluate the potential value of a given patent automatically, which is the goal of this work.

However, there are many challenges to solve this question. First of all, different from general text analysis, patent document contains dozens of special features, including structured items and unstructured items [32]. The structured items are uniform in semantics and format (such as patent number, inventor, assignee, application date, grant date and classification code) and the unstructured ones consist of text content in different length (including claims, abstracts, and descriptions of the invention.). Second, there contains much useful information in patent citation network, but how to model it and make it effectively contribute to patent valuation is kind of difficult, which is one of the technicality goal of our framework modeling.

As mentioned above, there are indeed previous works focusing on patent valuation, while most of them just focus on one aspect of patent value, such as statistical analysis [28] and text mining [13]. As far as concerned, none of the existing works [13, 20] takes into account both the patent text materials and the citation networks in terms of finding more valuable patents. To solve all these problems with addressing the challenges above, we propose Deep Learn-

¹ <http://www.ificlaims.com/>, in 2015, IBM received 8,088 granted U.S. patents, followed by Samsung (5,518), Canon (3,665), Qualcomm (2,897), Google (2,835), Intel (2,784), LG (2,428), Microsoft (2,398).

² <http://www.wipo.int>.

ing based Patent Quality Valuation (DLPQV) model to evaluate patent quality, which extracts the patent attribute network embedding by Attribute Network Embedding (ANE) and analyzes patent text materials by Attention-based Convolutional Neural Network (ACNN).

Specially, given the text materials, citation relations and meta features of patents, we first design an unified CNN-based and ANE-based architecture to exploit the semantic representations and network embedding for all patents. Then we qualify the quality valuation contribution of each sentence to the title by utilizing an attention strategy. Next, train DLPQV and generate the quality valuation prediction for each patent. Finally, extensive experiments on a large-scale real-world dataset validate both the effectiveness and explanatory power of our proposed framework. The main contributions of this paper could be summarized as:

- (1) We are the first one to apply deep learning method to patent document analysis, which is an ingenious piece of work combining the strength of deep learning and patent characteristics.
- (2) We present novel attribute network embedding for learning the low-dimensional vectors of patent citation networks, which is one of the most important components of patent valuation.
- (3) We propose a unified framework to combine attribute network embedding and deep learning based CNN methods, which allows jointly modeling patent information for patent quality valuation.
- (4) The extensive experiments in a real patent dataset show the proposed method outperforms baselines significantly.

2 Related Work

Generally, the related work can be classified into the following two parts, i.e., patent citation network studies in patent quality valuation and text mining techniques for patent analysis.

2.1 Patent Citation Network in Patent Quality Valuation

Many scholars have suggested that patent citation counts are strongly relevant to patent value or patent quality [1, 10, 12, 22]. Sterzi [29], who proposed that the number of times a patent has been cited by other patents is significantly associated with the value of the patent, trying to solve data truncation problems by using year dummies; these dummies represented the period from the priority year up to 3 years, the period from the priority year up to 6 years, and the period from 7 years to the search year. Fischer and Henkel [6] used the natural logarithms of the number of forward citations +1 to reduce the skewness of the distribution of patent citation counts. The number of citations made by other firms or researchers in a similar field for up to 5 years after the publication date showed a considerable association with economic patent value [19, 29], and late

citations those made after 5 years since a patent was granted showed a strong relationship with the market value of a firm [11, 29]. In addition, Karki [17] considered the number of citations to reflect a patents technological influence on subsequent inventions. The number of backward citations signifies references that are quoted by the relevant patent, and a variety of technological information is expected to contribute to high patent quality [2]. Based on all the previous works, we can tell that the number of patent citations can reflect patent value in terms of novelty.

However, the common limitation of these works is that these methods are usually based on statistical analysis using historical citation information in order to explore some specific relationships between patent citation count and patent value, and there still need extensive and unified approaches to synthetically measuring patent quality, which is what we devote to. Different from them, our study adopts both the citation networks with the patent meta features and abundant patent documents to predict the potential patent value, trying to reveal more deeper insights in this problem using attribute network embedding method.

2.2 Text Mining Techniques for Patent Analysis

One of the crucial steps in our framework is the understanding and representations of patent text materials, which aims at automatically processing patent document inputs and producing textual outputs. Most of the previous researches in this area are based on bag-of-words or LDA. Hasan et al. [13] built a patent ranking software, named COA (Claim Originality Analysis) that rates a patent based on its value by measuring the recency and the impact of the important phrases. Shaparenko et al. [27] discovered important documents in a document collection, which are clustered by their word bags. They find that a document is important if it has fewer similar documents published before it, and has more similar documents published after it. Specifically, Tang et al. [31] designed and implement a general topic-driven framework for analyzing and mining the heterogeneous patent network. Besides, to assess the technology prospecting of a company, Jin et al. [16] proposed an Assignee-Location-Topic (ALT) Model to extract emerging technology terms from patent documents of different companies, which are also based on LDA method.

However, these existing methods fail to display the relationships among words or sentences in patent documents, which is exactly the strengths of deep learning methods in NLP (Natural Language Processing) field.

Combining the above two points, in this work, we adopt both the citation networks with the patent meta features and abundant patent documents and propose a novel framework (DLPQV) of patent quality valuation, consisting of Attribute Network Embedding (ANE) and Attention-based Convolutional Neural Network (ACNN), which mixes patent text materials, meta features and citation network together to carry our point of comprehensive valuation of given patents.

3 Deep Learning Based Patent Quality Valuation (DLPQV) Framework

In this section, we first detailedly introduce the Patent Quality Analysis task, and then we introduce the technical details of DLPQV. The DLPQV model consists of Attribute Network Embedding (ANE) and Attention-based Convolutional Neural Network (ACNN).

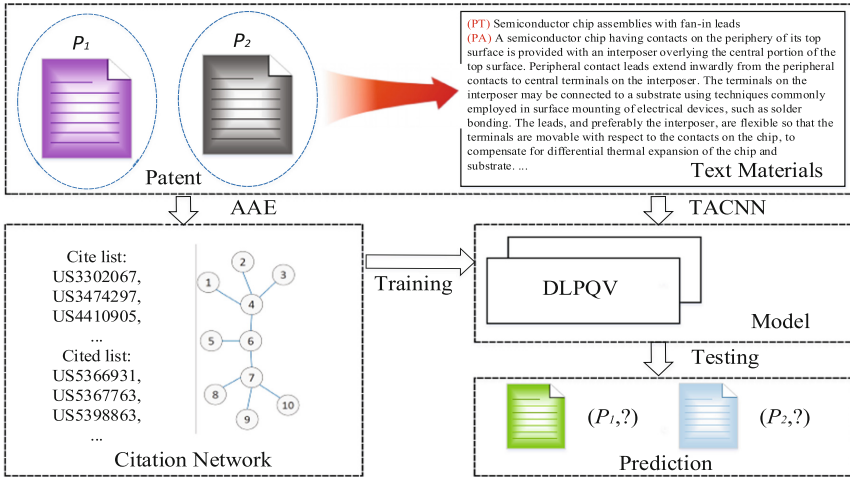


Fig. 1. The flowchart overview of our work

3.1 Problem and Study Overview

Traditional patent citation analysis can work on different applications for patents. For instance, if a patent have a high citation count, the cited patent probably have high chance to be a foundation of the citing patents. That is to say, highly-cited patents are possibly more important compared with those less ones. Therefore, we regard forward citation within two decades after authorization as patent quality with normalization.

Definition 1. Formally, given a set of patents with corresponding text materials including title (PT), abstract (PA), citation networks and patent meta features. And each patent has a quality valuation record obtained from cited amount with normalization (see Table 1). Our goal is to leverage the information of patent P_i available to train a prediction model \mathcal{M} (i.e., DLPQV), which can be effectively used to valuate the quality of patents in the new granted patents.

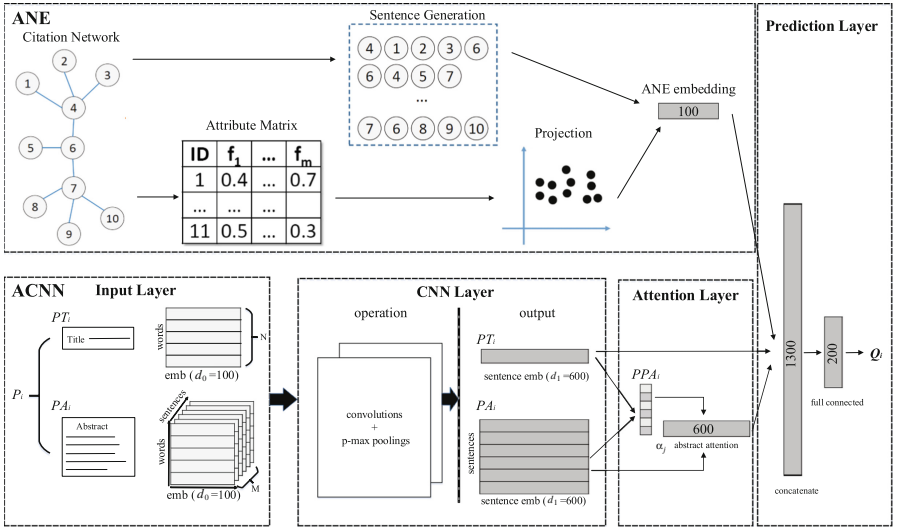
As is shown in Fig. 1, our solution is a two-stage framework, which includes a training stage and a testing stage: (1) In the training stage, given patent features including text materials, citation network and patent meta features

Table 1. Examples of patent instances with text, citation and attributive information.

Quality (Q)	Patent (P)	Text materials		Citation network		Meta features		
		Title (PT)	Abstract (PA)	Cite	Cited	WIPO	Claims	Nber
0.8761	US5148265	Semiconductor chip...	A semiconductor chip having contacts on...	US3302067...	US5367763...	A	39	Meth
0.1205	US5366931	Fabrication method...	This invention relates to a technology...	US4177480...	US5895966...	B2	5	Meth
?	US5477611	Method of forming...	A method for creating an interface...	US4079511...	US5776796...	B1	39	Elec

(see Table 1), we propose DLPQV to represent the text materials of each patent P_i and embedding the attribute network so as to evaluate patent quality Q_i . (2) In the testing stage, after the training of DLPQV is completed, for each new granted patent, DLPQV could estimate its quality with the available patent features.

Our DLPQV detailed framework is showed in Fig. 2, and we will introduce the model specifically in the following description, which covers Attribute Network Embedding (ANE) and Attention-based Convolutional Neural Network (ACNN).


Fig. 2. Deep Learning based Patent Quality Valuation (DLPQV) framework

3.2 Attribute Network Embedding for Citation Network

Definition 2 (ANE for Citation Network). Treating the granted patents as nodes and citation relations among them as edges respectively, we construct a citation network and use the proposed attribute network embedding method to learn the patent representation. Our citation network representation problem is formalized as follows. Given a citation network $G = (V, E, F)$, where V is the sets of nodes, E is the sets of edges and $F = \{f_1, f_2, \dots, f_{|V|}\}$ represents the sets of features of size m for each node. We aim to learn a low-dimensional vector representation $u_v \in R^d$ for each node $v \in V$ in G , where d is much smaller than $|V|$.

Attribute Network Embedding Framework. For citation network, we propose a Attribute Network Embedding model (ANE) that incorporates the node attributes, whose framework is shown in Fig. 3. Firstly, different from the *sentences generation* (like *word2vec*) method used in previous work, we propose the *sentences generation* method based on nodes’ neighbors. We can preserve the citation network structure based on these sentences, so that nodes with the similar neighborhoods will have the similar citation network embedding. Then, in order to incorporate the attributes of nodes into citation network embedding, we take nodes’ attributes as the initial input and utilize the mapping function to project it into the node embedding space. Finally, through the optimization of the model, we obtain the citation network embedding which can simultaneously preserve the citation network structure and reflect the similarity of node attributes. In the following section, we will introduce our model in detail.

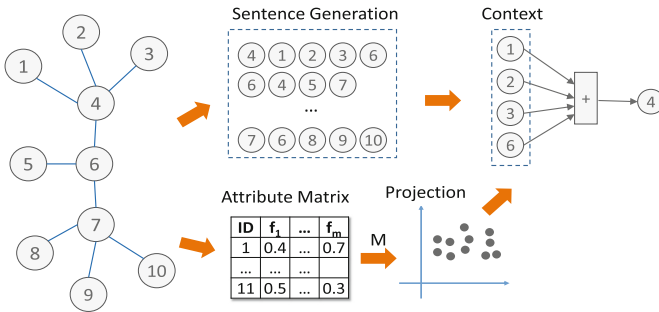


Fig. 3. ANE model framework

Sentences Generation. In previous network representation learning research works, there are two main ways to learn the network structure information. Like Deepwalk [25], node2vec [9] etc., they sample uniformly a random node $v \in G$ as the root of random walk and generate a truncated random walk sequences as the training *Sentences* to learn the node embedding. They are based on the

assumption that the node is similar to the surrounding nodes in window size k , which we think is too strong for some network structures, like the network in Fig. 3. The other is to learn the network embedding by preserving the First-order Proximity or the Second-order Proximity, like [3, 30], etc. However, these methods only consider the similarity between the node and their neighborhoods and don't consider the similarity between nodes' neighborhoods. In order to alleviate these problems, inspired by [26], we proposed the *sentences generation* method based on nodes' neighborhoods as follows:

We use each node as the root once, and take the random permutations of the root node's neighborhoods into the sentence. Each generated sentences has the form: $[v_{root}, v_1, \dots, v_n]$, where $\forall 1 \leq i \leq n, v_i$ is the neighborhood of v_{root} . Take the node 2 in Fig. 3 as an example, [3, 4, 1] is a permutation of the node 2's neighborhoods and [2, 3, 4, 1] is an instance of sentence generation of node 2. Also, it is important to note that the nodes in root node's neighborhoods should be no explicit order. So we set the number of permutations of root node's neighborhood to be N^P . The larger the value N^P is selected, the more evenly distributed root node's neighborhoods are in generated sentences.

ANE Model Formulation. Here, we describe how the ANE model incorporates the node attributes into citation network embedding. For each node in generated sentences, the ANE model predicts the center node v_i given a representation of the surrounding context nodes $v \in \{v_{i-k}, \dots, v_{i+k}\} \setminus \{v_i\}$, where k is the window size of context nodes. The objective function of ANE model is to maximize the average log probability of the center node v_i given the context nodes $context(v_i)$ for all the sentences $s \in S$, which is defined as following:

$$L = -\frac{1}{|s|} \sum_{i=1}^{|s|} \log p(v_i | context(v_i)) = -\frac{1}{|s|} \sum_{i=1}^{|s|} \log \frac{\exp u_i'^T u_{context(i)}}{\sum_{j=1}^{|V|} \exp u_j'^T u_{context(i)}} \quad (1)$$

where u_i' is 'output' vector representation of node v_i , $u_{context(i)}$ is vector representation of context words of node v_i and $|V|$ is the number of citation network nodes as well as the number of patents.

In order to make full use of the nodes' own attributes, as shown in Fig. 3, we take the nodes' attributes as the initial input of the model. Then we transform it to node embedding space with the use of transformation matrix M , where we have:

$$u_i = M^T f_i \quad (2)$$

where u_i is the 'input' vector representation of the node v_i , f_i is attribute value of node v_i . And $M \in R^{m \times d}$, where m is the node attributes dimension, and d is the dimension of u_i .

Furthermore, we defined $u_{context(i)}$ as weighted average of the 'input' vector representation of context nodes:

$$u_{context(i)} = \frac{1}{2k} \sum_{j \in [i-k, i+k] \setminus \{i\}} u_j \quad (3)$$

Finally, by minimizing Eq. (1), we obtain ‘input’ representation u_i and ‘output’ vector representation u'_i for node $v_i \in V$, and both of them can be regraded as low-dimensional representation of node. Therefore, we utilize the concatenation of them as the citation network embedding, and each patent is represented by a citation network embedding.

Model Optimization. Next, we introduce the details of how to use the Stochastic Gradient Descent method(SGD) to train the ANE model. Then we present the algorithm framework and time complexity of the model.

Approximation by Negative Sampling: Optimizing the Eq.(1) is computationally expensive, because the denominator of $p(v_i|context(v_i))$ requires summation over all the nodes in citation network, which the number of node is usually very large. To address this problem, we adopt the approach of negative sampling proposed in [23], which select negative samples according to the noisy distribution $P_n(v)$ for each node contexts. As a result, the $\log p(v_i|context(v_i))$ in Eq. (1) is replaced by the following objective function:

$$L_i = \log \sigma(u_i'^T u_{context(i)}) + \sum_{t=1}^{neg} E_{v_t \sim P_n(v)} [\log \sigma(-u_t'^T u_{context(i)})] \quad (4)$$

where $\sigma(x) = 1/(1 + \exp(-x))$, neg is the number of negative samples. And we set the node noisy distribution $P_n(v) \propto d_v^{3/4}$ as proposed in [23], where d_v is the out-degree of node v .

We employ the widely used Adaptive Moment Estimation (Adam) algorithm [18] to optimize the Eq. (4). In each step, the Adam algorithm samples a mini-batch of training instance(center-context) and then update the model parameter by walking along the descending gradient direction,

$$u_i'^{t+1} = u_i'^t - \eta \cdot \frac{\partial L_i}{\partial u_i'} \quad (5)$$

where u' is the ‘output’ vector representation of the node v_i , and t is the iteration times. η is the learning rate, which is automatically adjusted in Adam algorithm.

3.3 Attention-Based Convolutional Neural Network

Through ANE for citation network, a patent P_i as a node in citation network is represented as a representation vector u'_i , which is expressed as PU_i in the following description. In this subsection, we will introduce the specific components of ACNN of DLPQV, which deals with text materials to obtain the representation of patents. As shown in Fig. 2, ACNN can be divided into four components, i.e., *Input Layer*, *CNN Layer*, *Attention Layer* and *Prediction Layer*. The following will cover concrete content about the four layer, especially CNN Layer and Attention Layer.

Definition 3 (ACNN of DLPQV). Given a dataset of patents with text materials including patent titles (PT), patent abstracts (PA) and patent attribute citation network embedding (PU), and each patent P_i has a quality valuation Q_i (e.g., 0.8761) obtained from the normalized cited amount (see Table 1), we aim at leveraging the information of patents to train a prediction model based on ACNN, which can estimate the qualities of patents.

Input Layer. The input to ACNN is the title text and all abstract text of a patent P_i , i.e., title(PT_i) and abstract(PA_i). Specifically, the abstract text PA_i is expressed to a sequence of sentences $PA_i = \{s_1, s_2, \dots, s_M\}$ where M is the sequence length. And the title PT_i is an individual sentence. Considered to sentence constituents, each sentence consists of a sequence of words $s = \{w_1, w_2, \dots, w_N\}$ where $w_i \in \mathbb{R}^{d_0}$ is obtained from d_0 -dimensional pre-trained word embedding and N is the length of sentence. Finally, the title of a patent is translated into a matrix $PT_i \in \mathbb{R}^{N \times d_0}$, and the abstract of a patent is represented by a tensor $PA_i \in \mathbb{R}^{M \times N \times d_0}$.

CNN Layer. We aim at learning each sentence representation from word embedding in CNN Layer. Reasonably, we choose CNN-based model to learn sentence embedding with following reasons: (1) Because of convolution-pooling operations, CNN works better on considering dominated information of each sentence from local to global views. Usually, sentence is well represented by local key words. (2) CNN leverages shared convolution filter for training model, so it can reduce the complexity compared with other deep learning architecture, such as DNN or RNN [21]. (3) CNN is suitable for learning the interactions between words and deeply mining the semantic representations for sentences.

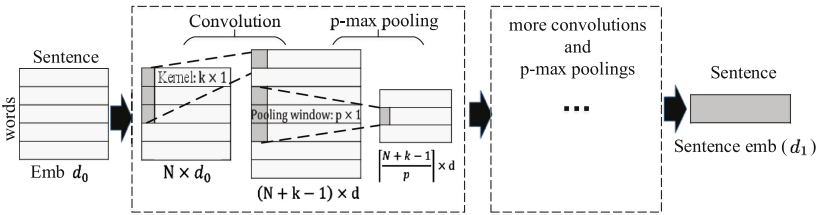


Fig. 4. CNN Layer, which contains several layers of convolution and p-max pooling.

As shown in Fig. 2, we design CNN Layer as a traditional model [5] that selects several layers of convolution and p -max pooling. Then each sentence is represented as a fixed length vector. Next, we will introduce the detail of the convolution-pooling operation in CNN Layer.

Specifically, we analyze the first convolution-pooling operation, and the other more operations are similar to that. In Input Layer, we transform a sentence

into a sentence matrix input $s \in \mathbb{R}^{N \times d_0}$ as the input of CNN Layer (showed in Fig. 4), then the wide convolution operates on a sliding window of every k words with a kernel $k \times 1$. Through the first convolution operation, the input sentence $s = w_1, w_2, \dots, w_N$ is transformed to a new hidden sequence, i.e., $e^c = \{\vec{e}_1^c, \vec{e}_2^c, \dots, \vec{e}_{N+k-1}^c\}$, where:

$$\vec{e}_i^c = ReLU(\mathbf{G} \cdot [w_{i-k+1} \oplus \dots \oplus w_i] + \mathbf{b}), \tag{6}$$

here, $\mathbf{G} \in \mathbb{R}^{d \times k d_0}$, $\mathbf{b} \in \mathbb{R}^d$ are the convolution parameters, and d is the output dimension of the convolution operation. $ReLU(x)$ is a nonlinear activation function which is equal to $ReLU(x) = \max(0, x)$. “ \oplus ” is in order to concatenate k word vectors into a long vector.

After the convolution operation, we obtain a local semantic representation by convoluting sequential k words. Next, we leverage p -max pooling operation to transform the convolution sequence e^c into a new global hidden sequence, i.e., $e^{cp} = \{\vec{e}_1^{cp}, \vec{e}_2^{cp}, \dots, \vec{e}_{[(N+k-1)/p]}^{cp}\}$, where:

$$\vec{e}_i^{cp} = \left[\max \begin{bmatrix} e_{i-p+1,1}^c \\ \dots \\ e_{i,1}^c \end{bmatrix}, \dots, \max \begin{bmatrix} e_{i-p+1,d}^c \\ \dots \\ e_{i,d}^c \end{bmatrix} \right]. \tag{7}$$

Similar to the first convolution-pooling operation, more layers of convolution-pooling processes are merged into the ACNN model to gradually express the global semantic information of sequential words in a sentence. Finally, a sentence consisted of N word embedding is transformed to a vectorial representation $s \in \mathbb{R}^{d_1}$, where d_1 is the output dimension of CNN Layer.

Through CNN Layer, the title of a patent is transformed into a vector $PT_i \in \mathbb{R}^{d_1}$. Meanwhile, the abstract of a patent which contains M sentences is represented by a matrix $PA_i \in \mathbb{R}^{M \times d_1}$. The output form of CNN Layer is showed in Fig. 2.

Attention Layer. After the previous layers’ operation, we obtain sentence representation. However, it is not equally important for the M sentences of the abstract contributing to the patent quality. Therefore, Attention Layer is designed to assign different weights according to the title. Detailedly, the attention representations are modeled as vectors by a weighted sum aggregated result of the sentence representations from abstract perspectives. For example, the abstract attention score PAA_i of a specific patent P_i is represented as follows:

$$PAA_i = \sum_{j=1}^M \alpha_j s_j^{PA_i}, \alpha_j = \cos(s_j^{PA_i}, s^{PT_i}), \tag{8}$$

here, $s_j^{PA_i}$ is the j -th sentence in PA_i , s^{PT_i} is the sentence representation of patent title PT_i ; *Cosine similarity* α_j is denoted as the attention score for measuring the weight of the sentence s_j in abstract PA_i for patent P_i , which means the importance of the contribution to the patent quality.

Prediction Layer. The last layer of ACNN is Prediction Layer, which aims at predicting the quality \widetilde{Q}_i of patent P_i considered the abstract-attention representation PAA_i , the title representation s^{PT_i} and the attribute network embedding PU_i . To be specific, we first merge those three representation vectors into a long vector by concatenation operation, then use a classical full-connected network [14] to learn the overall valuation representation o_i , then predict the quality \widetilde{Q}_i by LeakyReLU function, which we will discuss detailedly in Sect. 4:

$$o_i = ReLU(W_{ReLU} \cdot [PAA_i \oplus s^{PT_i} \oplus PU_i] + b_{ReLU}), \quad (9)$$

$$\widetilde{Q}_i = LeakyReLU(W_{LeakyReLU} \cdot o_i + b_{LeakyReLU}), \quad (10)$$

where W_{ReLU} , b_{ReLU} , $W_{LeakyReLU}$, $b_{LeakyReLU}$ are parameters to tune the network.

And we formulate the function by minimizing the least square loss with a l_2 -regularization term:

$$\mathcal{J}(\Phi) = \sum_{P_i} (Q_i - \widetilde{Q}_i)^2 + \lambda_\Phi \|\Phi_{\mathcal{M}}\|^2, \quad (11)$$

where \mathcal{M} represents the DLPQV that transforms text materials, citation relation and attribute information of patent P_i into predicted patent quality \widetilde{Q}_i (Eq. (10)). $\Phi_{\mathcal{M}}$ denotes all parameters in DLPQV and λ_Φ is the regularization hyperparameter.

4 Experiments

In this section, we first introduce our DLPQV framework settings, then compare the performance of DLPQV against the baseline approaches on patent quality valuation task. At last, we provide a *case study* to visualize the explanatory power of DLPQV.

4.1 Dataset Description

The experimental dataset is supplied by United States Patent and Trademark Office (USPTO)³, which grants US patents to inventors and assignees all over the world since 1976. Patents are classified according to the technical features of patented invention. These classification are mapped to broader, more easily understood technology fields.

For data pre-processing, we extract 51224 patents from USPTO dataset as our experimental dataset including the titles, abstracts, citation relation and meta features. Text materials are cleaned by deleting stop words, and meta features contain WIPO document kind codes, number of claims, categories by National Bureau of Economic Research, authorization year, assign information

³ <http://www.patentsview.org>.

and so on, which are also transformed into one-hot form (8035 dimensions). Lastly, the cited amount of a patent within two decades after granted is normalized as the patent quality.

4.2 Experimental Setup

Word Embedding. The word embedding in Input Layer of ACNN are trained on a large-scale *gigaword* corpus using public *word2vec* tool [23] with the dimension 100.

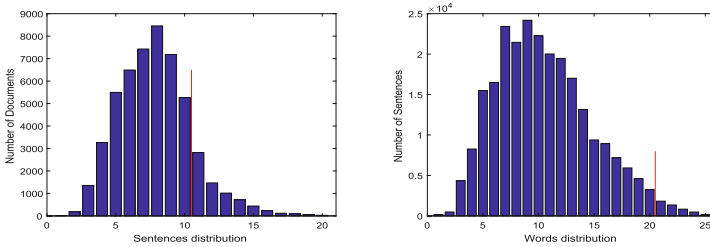


Fig. 5. Statistics of sentence and word distribution.

DLPQV Setting. In ANE of DLPQV, we set patent citation network embedding dimension as 100, and negative sampling number is set as 4 when the maximal length of sentence generation path is 40. In ACNN of DLPQV, we set maximum length $N(M)$ of words (sentences) in sentences (abstracts) as 10 (20) (zero padded when necessary) according to our statistics in Fig. 5, i.e., around 90% sentences (abstracts) contains less than 10 (20) words (sentences). There are four layers of convolution consisted of three wide convolutions and one narrow convolutions and max-pooling. And they are employed for CNN Layer in ACNN to accommodate the sentence length N , where the numbers of the feature maps for four convolutions are (200, 400, 600, 600) respectively. Meanwhile, the kernel size k is set as 3 for all four convolution layers and the pooling window p is set as (2, 2, 2, 1) for each max pooling respectively. We notice that LeakyReLU performances better in the patent quality valuation task, due to the property that it can not only preserve the advantage of ReLU like fast convergence speed, but also retain the informance in the negative axis. $LeakyReLU(x)$ denotes x when $x > 0$, and αx when $x \leq 0$. Further, we choose the value of α as 0.1 by conducting several experiments.

Training Setting. On the basis of the operation [24], we randomly initialize all vector and matrix parameters in ACNN with uniform distribution in the range between $-\sqrt{6/(nin + nout)}$ and $\sqrt{6/(nin + nout)}$, where nin and $nout$ are the numbers of the input and output matrix feature sizes. To measure the performance of DLPQV, we use the widely used *Root Mean Squared Error* (RMSE)

for the comparison of patent quality valuation precision. Overall speaking, the smaller the RMSE is, the better performance the result has.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_i - \tilde{Q}_i)^2}{n}} \tag{12}$$

4.3 Baseline Approaches

To our best knowledge, this is the first work based on deep learning for predicting patent quality valuation based on cited amount, which integrated text materials, citation network and patent meta features, so we verify the effectiveness of each component of DLPQV. The details of comparison are as follows:

- ANE: ANE is a framework without ACNN part, and only use citation network embedding PU_i as the patent embedding to predict the patent quality Q_i .
- ACNN: ACNN only consider text materials without citation relation and patent meta features.
- CNN: CNN is a framework with attention-ignored strategy compared with ACNN. Here, the attention-ignored strategy means the attention parameters α in Eq. (8) are the same for all sentences.
- ANE-CNN: ANE-CNN is a framework with attention-ignored strategy compared with DLPQV.

Both DLPQV and baselines are all implemented by Tensorflow and all experiments are run on a Tesla K20m GPU.

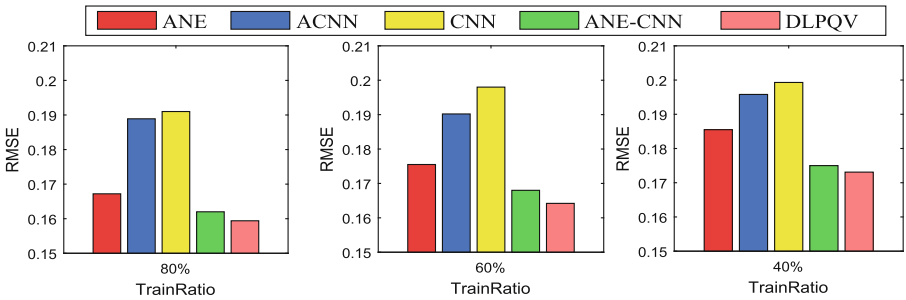


Fig. 6. Overall performance on the patent quality valuation task.

4.4 Experimental Results

Overall Results. To observe the several models’ performance for different data sparsity, we randomly select 80%, 60%, 40% of the extracted patent dataset as the training sets, and the rests as testing sets, respectively. In Fig. 6, we summarize the patent quality valuation results of all models. Obviously, we can see that DLPQV model performs best. Concretely, DLPQV performs better than

ANE, which indicated that the semantic representation from ACNN can provide patents' content features to improve the patent quality valuation accuracy rate. Both attribute information and patent effects are well integrated to enhance the network embedding so that Fig. 6 shows DLPQV beats ACNN, which indicates that ANE is also significant to DLPQV. Meanwhile, ACNN beats CNN as well as DLPQV beats ANE-CNN, which qualifies the contributions of texts with attention strategy. In summary, DLPQV has a best performance in different scale of training sets, and each part of DLPQV provides an important role for enhancing patent quality's forecast accuracy.

5 Conclusions

In this paper, we proposed a novel *Deep Learning based Patent Quality Valuation* (DLPQV) framework to predict patent quality. It is the first one to apply deep learning method to patent quality valuation problem with attribute network embedding and CNN method combined. We firstly design ANE to learn the patent embedding from attribute citation networks. Then, in order to represent text materials, we use a CNN-based architecture for exploiting sentence representations with attention strategy. And we qualified the contributions of abstract sentences to the patent valuation by an attention strategy. Finally, we mix citation network embedding and text representation to generate the patent quality prediction value. Experiments on real-world dataset supplied by USPTO proved that our framework could effectively predict the patent quality. In the future, we will focus on the patent quality variation tendency over time based deep learning method.

Acknowledgements. This research was partially supported by grants from the National Key Research and Development Program of China (Grant No. 2016YFB1000904), and the National Natural Science Foundation of China (Grants No. U1605251 and 61727809).

References

1. Albert, M.B., Avery, D., Narin, F., McAllister, P.: Direct validation of citation counts as indicators of industrially important patents. *Res. Policy* **20**(3), 251–259 (1991)
2. Burke, P.F., Reitzig, M.: Measuring patent assessment quality-analyzing the degree and kind of (in) consistency in patent offices' decision making. *Res. Policy* **36**(9), 1404–1430 (2007)
3. Chang, S., Han, W., Tang, J., Qi, G.-J., Aggarwal, C.C., Huang, T.S.: Heterogeneous network embedding via deep architectures. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 119–128. ACM (2015)
4. Chen, Y.-L., Chang, Y.-C.: A three-phase method for patent classification. *Inf. Process. Manag.* **48**(6), 1017–1030 (2012)

5. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (almost) from scratch. *J. Mach. Learn. Res.* **12**(Aug), 2493–2537 (2011)
6. Fischer, T., Henkel, J.: Patent trolls on markets for technology—an empirical analysis of NPEs’ patent acquisitions. *Res. Policy* **41**(9), 1519–1533 (2012)
7. Fujii, A., Iwayama, M., Kando, N.: Overview of the patent retrieval task at the NTCIR-6 workshop. In: *NTCIR* (2007)
8. Fujii, A., Utiyama, M., Yamamoto, M., Utsuro, T.: Evaluating effects of machine translation accuracy on cross-lingual patent retrieval. In: *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 674–675. ACM (2009)
9. Grover, A., Leskovec, J.: node2vec: scalable feature learning for networks. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 855–864. ACM (2016)
10. Guellec, D., de la Potterie, B.V.P.: Applications, grants and the value of patent. *Econ. Lett.* **69**(1), 109–114 (2000)
11. Hall, B.H., Jaffe, A., Trajtenberg, M.: Market value and patent citations. *RAND J. Econ.* **36**, 16–38 (2005)
12. Harhoff, D., Narin, F., Scherer, F.M., Vopel, K.: Citation frequency and the value of patented inventions. *Rev. Econ. Stat.* **81**(3), 511–515 (1999)
13. Hasan, M.A., Spangler, W.S., Griffin, T., Alba, A.: Coa: finding novel patents through text analysis. In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1175–1184. ACM (2009)
14. Hecht-Nielsen, R., et al.: Theory of the backpropagation neural network. *Neural Netw.* **1**(Supplement-1), 445–448 (1988)
15. Hsia, W.L.: The value of patents toward business. *Intellect. Prop. Manag.* **16**, 20–21 (1998)
16. Jin, B., Ge, Y., Zhu, H., Guo, L., Xiong, H., Zhang, C.: Technology prospecting for high tech companies through patent mining. In: *2014 IEEE International Conference on Data Mining (ICDM)*, pp. 220–229. IEEE (2014)
17. Karki, M.M.S.: Patent citation analysis: a policy analysis tool. *World Patent Inf.* **19**(4), 269–272 (1997)
18. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
19. Lanjouw, J.O., Schankerman, M.: The quality of ideas: measuring innovation with multiple indicators. Technical report, National bureau of economic research (1999)
20. Liu, X., Yan, J., Xiao, S., Wang, X., Zha, H., Chu, S.M.: On predictive patent valuation: forecasting patent citations and their types. In: *AAAI*, pp. 1438–1444 (2017)
21. Ma, L., Lu, Z., Li, H.: Learning to answer questions from image using convolutional neural network. In: *AAAI*, vol. 3, p. 16 (2016)
22. Martinez-Ruiz, A., Aluja-Banet, T.: Toward the definition of a structural equation model of patent value: PLS path modelling with formative constructs. *REVSTAT-Stat. J.* **7**(3), 265–290 (2009)
23. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: *Advances in Neural Information Processing Systems*, pp. 3111–3119 (2013)
24. Montavon, G., Orr, G.B., Müller, K.-R. (eds.): *Neural Networks: Tricks of the Trade*. LNCS, vol. 7700. Springer, Heidelberg (2012). <https://doi.org/10.1007/978-3-642-35289-8>

25. Perozzi, B., Al-Rfou, R., Skiena, S.: Deepwalk: online learning of social representations. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 701–710. ACM (2014)
26. Pimentel, T., Veloso, A., Ziviani, N.: Unsupervised and scalable algorithm for learning node representations (2017)
27. Shaparenko, B., Caruana, R., Gehrke, J., Joachims, T.: Identifying temporal patterns and key players in document collections. In: Proceedings of the IEEE ICDM Workshop on Temporal Data Mining: Algorithms, Theory and Applications (TDM 2005), pp. 165–174 (2005)
28. Stephens, J.C., Jiusto, S.: Assessing innovation in emerging energy technologies: socio-technical dynamics of carbon capture and storage (CCS) and enhanced geothermal systems (EGS) in the USA. *Energy Policy* **38**(4), 2020–2031 (2010)
29. Sterzi, V.: Patent quality and ownership: an analysis of UK faculty patenting. *Res. Policy* **42**(2), 564–576 (2013)
30. Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., Mei, Q.: Line: large-scale information network embedding. In: Proceedings of the 24th International Conference on World Wide Web, pp. 1067–1077. International World Wide Web Conferences Steering Committee (2015)
31. Tang, J., Wang, B., Yang, Y., Hu, P., Zhao, Y., Yan, X., Gao, B., Huang, M., Xu, P., Li, W., et al.: Patentminer: topic-driven patent analysis and mining. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1366–1374. ACM (2012)
32. Zhang, L., Li, L., Li, T.: Patent mining: a survey. *ACM SIGKDD Explor. Newsl.* **16**(2), 1–19 (2015)