Personalized Citation Recommendation via Convolutional Neural Networks

Jun Yin and Xiaoming Li

School of Electronic Engineering and Computer Science, Peking University, Beijing 100871, P.R. China
{jun.yin, lxm}@pku.edu.cn

Abstract. Automatic citation recommendation based on citation context, together with consideration of users' preference and writing patterns is an emerging research topic. In this paper, we propose a novel personalized convolutional neural networks (p-CNN) discriminatively trained by maximizing the conditional likelihood of the cited documents given a citation context. The proposed model not only nicely represents the hierarchical structures of sentences with their layer-by-layer composition and pooling, but also includes authorship information. It includes each paper's author into our neural network's input layer and thus can generate semantic content features and representative author features simultaneously. The results show that the proposed model can effectively captures salient representations and hence significantly outperforms several baseline methods in citation recommendation task in terms of recall and Mean Average Precision rates.

Keywords: Convolutional Neural Networks, Citation Recommendation, Personalization

1 Introduction

With the impressive expanding speed of research papers producing annually, finding out which papers to cite is not a trivial task. For newcomers in a research area, this is especially a challenging task. Under such circumstances a system automatically recommending candidate citations is in great need. Figure 1 shows an example of citation recommendation. Given a piece of citation context and a set of papers in the left part of Fig. 1, the right part is the recommended result that we expect our proposed method outputs. Some researchers have been already aware of the necessity of citation recommendation and they have figured out various algorithms[14, 9, 16, 7, 2].

Unfortunately, simply solving out the citation recommendation task via lexical matching encounters various problems or difficulties with observation that the same concept is often expressed using different expressions and word usages in different papers. Latent semantic models such as latent semantic analysis (LSA) are able to represent documents. Based on this, probabilistic topic models such as probabilistic LSA (PLSA), Latent Dirichlet Allocation(LDA), and Bi-Lingual Topic Model(BLTM) are developed and applied in semantic matching.

Recent days deep learning models have shown great potential on learning effective representations and achieved state-of-the-art performance in computer vision[17] and natural language processing applications[1]. In spite of deep model's more appealing capability in automatically learning feature representation, they are inferior to shallow models such as in collaborative filtering from considering implicit authorship and similarity relationship between items. This calls for integrating deep learning with context aware recommendation by performing deep learning collaboratively.

Our research contributions can be summarized as follows:

1) To the best of our knowledge, this work is a successful attempt in applying the CNN-like method to citation recommendation task. This method avoids a lot of time-consuming feature engineering to represent semantic content and author preference.

2) We propose a novel p-CNN which includes paper's author into input layer and thus can generate semantic content features and representative author features simultaneously. Personalization citation recommendation is achieved via this approach.

2 Related Work

2.1 Citation Recommendation

There have been different approaches dealing with citation recommendation task, which can be grouped into three categories. The first category uses a graphical framework. Each research paper is represented as a node with citation relationship regarded as links between them. The recommendation task is cast as link prediction[13, 9, 16, 7]. The second category usually utilizes various kinds of content based semantic analysis techniques. [14] applys Topic Model. [8] uses translation model. [2] formalizes this problem under the retrieval framework.

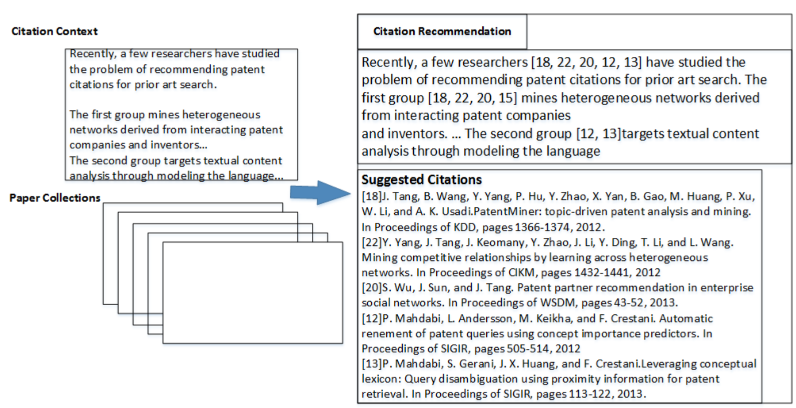

Fig. 1. One example of citation recommendation

From IR view of point, citation context is regarded as query while the target paper is retrieved document corpus. The third category concentrates on personalized aspect of this task, in which collaborative filtering is a widely used method. Diverse kinds of features are designed [18] to enhance recommendation performance yet have the disadvantage of energy-consuming.

2.2 Deep Learning

Recently, various kinds of deep learning methods have achieved great success in speech, image and natural language processing[1]. By exploiting deep architectures, deep learning techniques are capable to discover the hidden structures from training data and features at different levels of abstractions useful for the target tasks. Convolutional neural networks (CNN) are designed utilize layers with convolving filters that are applied to local features[6]. CNN models have subsequently been utilized and demonstrated to achieve excellent results in NLP tasks, such as question answering [19], search query retrieval[12], sentence modeling and matching [4, 5], and several other traditional NLP tasks[1].

3 Problem Formulation

We formalize this task as context-aware recommendation problem. The input query is short sentences (**citation context**). The task is to return a ranked list of **cited papers** as candidates to users. Deliberately avoiding from heuristic feature designing, we turn to deep convolutional network architecture to learn sentence and author representation under supervised learning framework.

4 Our proposed approaches

The model architecture is illustrated in Fig. 2. By using convolution and max-pooling architecture, local contextual information at the word n-gram level is modeled first. Then, salient local features in a word sequence and author information are combined to form a global low-dimensional feature vector. Finally, a multi-layer perceptron (MLP) is introduced here as a nonlinear similarity function computing the matching degree by comparing the global feature vector[5].

Input Layer. The input layer takes words in the form of embedding vectors. In our work, we set the maximum length of sentences to 50 words. For sentences shorter than that, we put zero padding.

For citation context, we get input layer denoted as l_t by concatenating each word word embedding of citation context and its author. For cited documents the input layer is generated by concatenating its content and author. More specifically, for each citation context cc, the input layer with its window indexed by t to upcoming convolution layer is given by the concatenated vector

$$l_t = \Big[\sum_i \{a_{cc}\}, w_t^T, w_{t+1}^T, w_{t+2}^T\Big] \tag{1}$$

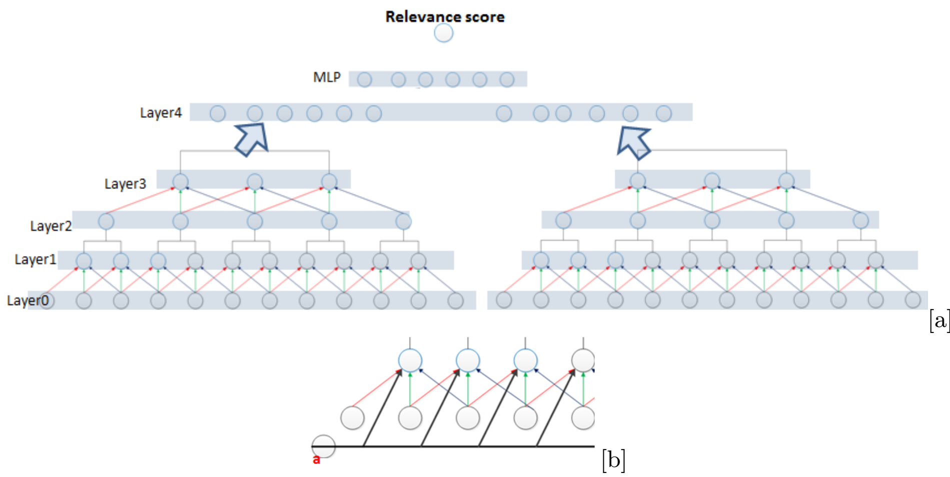

Fig. 2. (a)We use CNN mapping word sequence to a low-dimensional vector in semantic space; (b)The convolution units of multi- modal convolution layer of p-CNN

Here we simply average the author vectors as an expansion in the input layer. The author vector is initialized as the average of title word vectors from his or her ever published papers. The author vectors are model parameters and updated during the training phrase with respect to the model cost. Our experiments show that learning task-specific author vectors through fine-tuning offers performance gains compared with static author vectors.

Convolution. The convolution operation can be viewed as extracting local features from a sliding window of width k_1. Contextual feature vector h_t in this layer is computed by:

$$h_i^{(l)} = \sigma(W^{(l)} h_i^{(l-1)} + b^{(l)}) \tag{2}$$

where: $h_i^{(l)}$ is the output vector of feature maps for location i in Layer l, $\sigma(\cdot)$ is the activation function, W^l is the parameters on Layer l, $h_i^{(l-1)}$ is the segment of Layer-$(l-1)$ for the convolution layer at location i. For the first convolution layer: $h_i^{(0)} = x_{i:i+k_1-1} = \left[x_i^T, x_{i+1}^T, ..., x_{i+k_1-1}^T\right]$ concatenates the word vectors for k_1 width of sliding window from input sentence x.

Max-Pooling. To retain only the most useful local features produced by the convolutional layers, we take the max-pooling in every non-overlapping two-unit windows following each convolution. It computes as follows:

$$h_i^{(l)} = max\left\{h_t^{(l-1)}(i), h_t^{(l-1)}(i+1)\right\}, \tag{3}$$

MLP. These are fully connected layers. The input of which is the concatenation of fixed length vector of convolution-max pooling output from both citation context and cited paper. MLP can be seen as a nonlinear score function computing the probability of citation context references cited papers[11, 12].

4.1 Training

We employ a discriminative training strategy with a large margin objective. The model parameters are learned to minimize the pair-wise loss function written as follows:

$$Loss(\theta) = max\{0, 1 + s(cc, D_j^-) - s(cc, D^+)\} \tag{4}$$

where $s(cc, D)$ is the predicted relevance score for pair (cc, D) computed as the output of MLP. The negative sampling process chooses the documents written by totally different authors as negative samples, which gets better performance than random negative sampling.

The dimension of word vector and author vector is 100. The network architecture of all CNN-related methods is configured as two for convolution, two for pooling, and two for MLP. Dropout proved to be such a good regularizer shown in our experiment, which is also consistent with [5].

5 Experiments

5.1 Dataset

We use the same dataset as [18]. Following [15], one citation context consists of three sentences around the a citation placeholder. We combine the title and abstract as the content of cited papers. The dataset includes 73236 citation relationships which send out by the papers from the 10 seed venues[1]. We filtered 35362 distinct words and 2,191 distinct authors who appear in our data set more than 5 times. Rare words and authors appearing less than 5 times are replaced by "UNK_W" and "UNK_A".

5.2 Baseline Methods

We compare our proposed methods with several popular methods as follows. All the baseline models are trained on the same training set as our proposed model.

Language model (**LM**): is one-gram language model.

Translation Model (**TM**): [8] proposed a translation model to overcome the language gap between citation contexts and cited papers.

word2vec: word2vec[2][10] is an unsupervised model. We calculate the matching degree of citation context and cited documents with cosine similarity followed by sum of words as input .

DSSM: [3] compute the (query, document) pair score by cosine similarity between their semantic vectors via conventional CNN.

CNN: [1] uses principled CNN to get sentence representation without consideration of author information.

Table 1. Performance of Citation Recommendation

	MAP	**Recall@10**
rand	0.007	0.000
LM	0.299	0.376
TM	0.501	0.594
word2vec	0.657	0.779
DSSM	0.710	0.801
CNN	0.723	0.832
p-CNN	**0.762**	**0.857**

5.3 Experimental Results

The evaluation metrics aims to assess the positions of the true citations in the ranking list for each given context. We report standard metrics of recall and MAP on the test set as the experimental results. The main results are summarized in Table 1. We can see that our *p*-CNN is the best performer, beating other methods in terms of MAP and Recall@10. Besides, supervised learning on citation relationship is essential for obtaining superior personalized citation recommendation performance, i.e., *p*-CNN and CNN all outperforms word2vec. By comparing *p*-CNN with CNN, we also see the the effectiveness of learned author feature vectors.

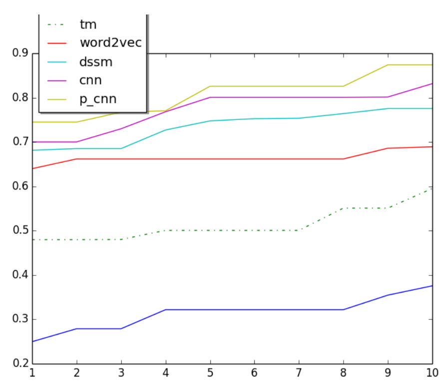

Fig. 3. Recall value of every position at one to ten.

We then further investigated the performance of *p*-CNN using recall at different positions with experimental results shown in Fig. 3. We can see *p*-CNN is stable on recall and achieve the best performance at all positions. Specifically, we observe that even at position one, the *p*-CNN still can have effective recommendation.

[1] ACL, CIKM, EMNLP, ICDE, ICDM, KDD, SIGIR, VLDB, WSDM, WWW
[2] https://code.google.com/p/word2vec/

5.4 Parameter Analysis

We investigate the parameter sensitivity to test the stability of our model with results shown in Fig. 4. We can see that CNN and p-CNN are not sensitive to dimension n when d is greater than 100. Actually, $d = 100$ is enough to make our models achieve their approximate optimal performance, while increasing d results in no higher performance but training time consuming.

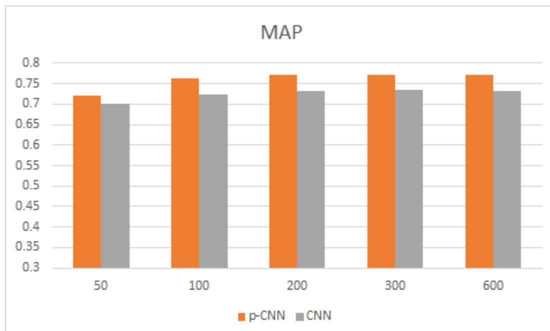

Fig. 4. Parameter sensitivity w.r.t. dimension.

6 Conclusions

Our work is a successful attempt in applying the CNN-like method to citation recommendation task. Our proposed p-CNNs approach is able to automatically learn the representative features for each author and integrate them into computing relevance score between citation context and candidate cited documents. Experimental results show that the proposed convolutional neural network can significantly improve the recommendation performance.

Acknowledgments This work has been partially supported by the 973 Program under Grant No. 2014CB340405 and National Natural Science Foundation of China under Grant No. U1536201.

7 References

References

1. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (almost) from scratch. The Journal of Machine Learning Research 12, 2493–2537 (2011)
2. He, Q., Pei, J., Kifer, D., Mitra, P., Giles, L.: Context-aware citation recommendation. In: Proceedings of the 19th international conference on World wide web. pp. 421–430. ACM (2010)

3. Huang, P.S., He, X., Gao, J., Deng, L., Acero, A., Heck, L.: Learning deep structured semantic models for web search using clickthrough data. In: Proceedings of the 22nd ACM international conference on Conference on information & knowledge management. pp. 2333–2338. ACM (2013)
4. Kalchbrenner, N., Grefenstette, E., Blunsom, P.: A convolutional neural network for modelling sentences. arXiv preprint arXiv:1404.2188 (2014)
5. Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882 (2014)
6. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE 86(11), 2278–2324 (1998)
7. Liu, X., Yu, Y., Guo, C., Sun, Y.: Meta-path-based ranking with pseudo relevance feedback on heterogeneous graph for citation recommendation. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. pp. 121–130. ACM (2014)
8. Lu, Y., He, J., Shan, D., Yan, H.: Recommending citations with translation model. In: Proceedings of the 20th ACM international conference on Information and knowledge management. pp. 2017–2020. ACM (2011)
9. Mahdabi, P., Crestani, F.: Query-driven mining of citation networks for patent citation retrieval and recommendation. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. pp. 1659–1668. ACM (2014)
10. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
11. Shen, Y., He, X., Gao, J., Deng, L., Mesnil, G.: A latent semantic model with convolutional-pooling structure for information retrieval. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. pp. 101–110. ACM (2014)
12. Shen, Y., He, X., Gao, J., Deng, L., Mesnil, G.: Learning semantic representations using convolutional neural networks for web search. In: Proceedings of the 23rd International Conference on World Wide Web. pp. 373–374. ACM (2014)
13. Strohman, T., Croft, W.B., Jensen, D.: Recommending citations for academic papers. In: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. pp. 705–706. ACM (2007)
14. Tang, J., Zhang, J.: A discriminative approach to topic-based citation recommendation. In: Advances in knowledge discovery and data mining, pp. 572–579. Springer (2009)
15. Tang, X., Wan, X., Zhang, X.: Cross-language context-aware citation recommendation in scientific articles. In: Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval. pp. 817–826. ACM (2014)
16. Wang, S., Lei, Z., Lee, W.C.: Exploring legal patent citations for patent valuation. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. pp. 1379–1388. ACM (2014)
17. Wang, X., Wang, Y.: Improving content-based and hybrid music recommendation using deep learning. In: Proceedings of the 22nd ACM international conference on Multimedia. pp. 627–636. ACM (2014)
18. Yan, R., Yan, H., et al.: Guess what you will cite: Personalized citation recommendation based on users preference. In: Information Retrieval Technology, pp. 428–439. Springer (2013)
19. Yin, J., Jiang, X., Lu, Z., Shang, L., Li, H., Li, X.: Neural generative question answering. arXiv preprint arXiv:1512.01337 (2015)