Research Article

Attention-Based Personalized Encoder-Decoder Model for Local Citation Recommendation

Libin Yang,1 Zeqing Zhang,2 Xiaoyan Cai,1 and Tao Dai3

1School of Automation, Northwestern Polytechnical University, Xi’an, Shaanxi, China
2School of Telecommunications Engineering, Xidian University, Xi’an, Shaanxi, China
3School of Software Engineering, Xi’an Jiaotong University, Xi’an, Shaanxi, China

Correspondence should be addressed to Xiaoyan Cai; xiaoyanc@nwpu.edu.cn

Received 13 February 2019; Revised 16 April 2019; Accepted 16 May 2019; Published 3 June 2019

Academic Editor: Maciej Lawrynczuk

Copyright © 2019 Libin Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With a tremendous growth in the number of scientific papers, researchers have to spend too much time and struggle to find the appropriate papers they are looking for. Local citation recommendation that provides a list of references based on a text segment could alleviate the problem. Most existing local citation recommendation approaches concentrate on how to narrow the semantic difference between the scientific papers’ and citation context’s text content, completely neglecting other information. Inspired by the successful use of the encoder-decoder framework in machine translation, we develop an attention-based encoder-decoder (AED) model for local citation recommendation. The proposed AED model integrates venue information and author information in attention mechanism and learns relations between variable-length texts of the two text objects, i.e., citation contexts and scientific papers. Specifically, we first construct an encoder to represent a citation context as a vector in a low-dimensional space; after that, we construct an attention mechanism integrating venue information and author information and use RNN to construct a decoder, then we map the decoder’s output into a softmax layer, and score the scientific papers. Finally, we select papers which have high scores and generate a recommended reference paper list. We conduct experiments on the DBLP and ACL Anthology Network (AAN) datasets, and the results illustrate that the performance of the proposed approach is better than the other three state-of-the-art approaches.

1. Introduction

The rapid development of science and technology leads to a substantial increase in the number of scientific papers. Researchers have to spend a lot of time and effort to find relevant papers. Many researchers use keyword search via scholar search engines, such as Google Scholar and Microsoft Academic Search, but there exist many instances where such a keyword query is either over broad, returning many papers that are loosely relevant to what the researcher really need, or too narrow, filtering many potentially relevant papers out or returning nothing at all [1–5]. To alleviate the above problems, many studies presented citation recommendation approaches which use a manuscript or a text segment instead of a set of keywords as query [6–8]. Citation recommendation approaches are categorized into two major types: local citation recommendation, which recommends relevant papers based on a citation context [9–11], and global citation recommendation, which recommends relevant papers based on a given manuscript [12–15]. This study focuses on local citation recommendation.

Most local citation recommendation approaches utilize scientific papers’ and citation context’s text content only [9, 16–18] and concentrate on how to narrow the semantic difference between the two text objects, completely neglecting other information. For instance, each author has his own word usage, grammar structure, writing style, and personal citation preference, while each venue has its own topic, and it only publishes papers related to that topic. However, such information which has been neglected by researchers may have direct influence on the local citation recommendation task’s performance, helping researchers find more appropriate references for the given citation context and yielding better performance.
Recent studies show the encoder-decoder framework performs well in the field of machine translation [19–21] as it learns relations between pairs of variable-length text. In this paper, we deem the scientific papers and citation context as parallel pairs, integrate several information related to scientific papers in attention mechanism, and propose an attention-based encoder-decoder (AED) model to score scientific papers via the venue information, author information, and the given citation context information, in turn, to enhance the performance of local citation recommendation. In AED model, we first construct an encoder which utilizes TDNN [7] to represent a citation context as a vector in a low-dimensional space, then we construct an attention mechanism integrating venue information and author information, and finally, we map the decoder’s output into a softmax layer and get the score value of scientific papers. A recommended reference paper list is generated based on high-scoring scientific papers. To summarize, our major contributions are as follows:

1. An attention mechanism is introduced in the encoder-decoder framework; the attention mechanism combines author and venue information.
2. An attention-based encoder-decoder (AED) model is proposed for local citation recommendation.
3. Experimental results illustrate the efficiency of the proposed AED model.

We organize the rest of this paper as follows. Section 2 reviews related work. Section 3 constructs an attention-based encoder-decoder model. Section 4 illustrates the proposed attention-based encoder-decoder model to solve the local citation recommendation task. Section 5 illustrates the experimental results, and Section 6 concludes the paper.

2. Related Work

Local citation recommendation is to find an ordered list of scientific papers from a dataset, and these scientific papers can be acted as candidate citations for a given citation context [11, 22]. The citation context can be the citation sentence itself or the composition of several preceding sentences and succeeding sentences of the citation sentence [23]. Most of the existing works focus on considering content semantic relationship between the two text objects, i.e., citation context and scientific papers, and recommend relevant references.

He et al. [24] evaluated the relevance between a given citation context and a paper and proposed a local citation recommendation model. Furthermore, they developed to automatically analyze a manuscript that lacks a bibliography and to recognize candidate locations in the manuscript where citations are needed [25]. Duma et al. [26] used the function of Core Scientific Concepts to local citation recommendation. Tang et al. [27] developed a bilingual embedding model, learning latent semantics of citations and contexts. Peng et al. [28] explored utilizing both knowledge-based methods and word-embedding similarity measures to model the relatedness between the papers and the citation contexts. Zhou [29] constructed an ActiveCite system, which integrates collaborative filtering, content-based filtering, and citation analysis approaches. ActiveCite uses current citation sentence as the query sentence for local citation recommendation and extracts topic of the paper as query sentence for global citation recommendation. The advantages of ActiveCite are that it is more intelligent and can minimize the interruption time of the paper writing process. Meanwhile, the author conducted usability research from the user’s point of view and put forward some suggestions on the constructions and optimization of citation recommendation system. Huang et al. [30] developed RefSeer which suggests candidate citations based on the input queries. RefSeer presents both topic-based global recommendation and citation-context based local recommendation.

As there exists vocabulary gap between the paper’s content and the citation context’s content, Lu et al. [31] proposed a translation retrieval model to bridge this gap. Huang et al. [32] proposed to use unique IDs to denote the cited papers; they regarded these papers as “word” in a novel language, and they used a translation model to estimate the translation probability of an ID based on the citing words. Furthermore, they built a neural network-based citation recommendation model by using the distributed representations of words and documents [10], but all the above approaches concentrate on citation contexts’ content information and scientific papers’ content information only, ignoring venue information and author information. Ebesu and Fang [11] solved the local citation recommendation problem by proposing a neural citation network, which utilized content information and author information, but ignoring venue information. In this paper, we deem scientific papers as candidate papers and introduce an attention mechanism incorporating venue information and author information, and then we propose a novel model which contains encoder, attention mechanism, and decoder to solve the local citation recommendation task.

3. Attention-Based Encoder-Decoder Model

Inspired by the encoder-decoder framework that has been successfully applied in neural machine translation [33, 34], we utilize the encoder-decoder framework in our citation recommendation task. We also combine the attention mechanism in the framework to integrate venue information and author information. Figure 1 shows the architecture of the proposed AED model.

3.1. Encoder. For each given citation context, we use TDNN [35] to obtain convolution on all possible word windows. Then, we use max-pooling nonlinear mapping to extract feature representation for each convolutional word window. Particularly, for a citation context whose length is \( \ell \), we denote the word embedding of the \( t \)-th word in the citation context as \( \hat{x}_t \) and denote the embedding of the citation context as \( \hat{x}_C = \hat{x}_{c_1} \odot \cdots \odot \hat{x}_{c_{\ell}} \). A convolutional filter \( W^c \in \mathbb{R}^{k \times g} \) (\( g \) is the dimension of word vector) slides over
$k$ words at a time over all possible window lengths, i.e., $\{x_{i}, x_{i+k}, \ldots, x_{i+k+1}\}$, and the convolutional layer is defined as

$$
o_i = \text{ReLU}(W_t^o x_{i+k-1} + b_i),$$

$$\tilde{o} = \max\{o_i, \ldots, o_{i-k} + b_i\},$$

where ReLU is the activation function, $o_i$ is the $i$-th feature vector, and $o \in \mathbb{R}^{1+k}$. We apply max-pooling operation to obtain the largest feature value extracted by each filter and repeat $p$ times yielding $\tilde{o}_j \in \mathbb{R}^p$. The aim of TDNN is to represent the citation context $X_c$ as a low-dimensional vector representation $s_j$, i.e.,

$$s_j = \text{tanh}(U_j \tilde{o}_j + b_j).$$

Finally, $f(X_c)$ gets different granularity of phrases (i.e., bigrams and trigrams) by using a variety of filters $L = \{l_1, \ldots, l_d\}$. The phrase level representation captured by the TDNN can improve the performance of semantic representation and meanwhile reduce computational time.

3.2. Attention Mechanism. The authors and the venue of a scientific paper have essential impact on the scientific paper. Usually, researchers will track other researchers or research groups which have similar research interests with them. Similarly, the venue of the scientific paper also plays a vital role in whether the scientific paper should be cited, and the authoritative venues are far better than general venues. Therefore, we construct an attention mechanism based on the citation context’s authors, the papers’ authors, and the papers’ venue. We concatenate the vector representations of citation context, citation context’s author, the paper’s authors, and the paper’s venue, yielding

$$s_j = [f(X_c) \oplus f(A^c) \oplus f(A^d) \oplus f(V^d)]_j,$$

where $A^c, A^d,$ and $V^d$ denote the vector representation of the authors of citation context, the vector representation of the scientific paper’s author, and the vector representation of scientific paper’s venue. The interaction of citation context, authors, and venues is occurred in decoding process.

3.3. Decoder. Since RNN has ability to learn current word from the previous words in the sequence and it can consider the encoder’s representation and its internal state at the same time, we use RNN as the decoder. We denote $x^d$ as the word embedding of the $i$-th word in the scientific papers and apply gated recurrent unit (GRU) [36] to solve gradient exploding or vanishing problem. The attention mechanism leans a weighted interpolation $c_i$ based on the encoder’s representation, i.e.,

$$c_i = \sum_j \alpha_{ij}s_j,$$

where $\alpha_{ij}$ is the output value derived from the softmax function and the $i$-th word must be aligned with the $j$-th output.

4. Attention-Based Encoder-Decoder Model for Local Citation Recommendation Approach

We map the output of the RNN decoder to the softmax layer, obtaining

$$P(y_i | y_{<i}, s) = \text{softmax}(Vh_i),$$

where $P(y_i | y_{<i}, s)$ is the conditional probability of all previous words in the scientific papers prior to the $i$-th word.
We simultaneously train the encoder-decoder using stochastic gradient descent (SGD) [37] and maximize the following equation:

$$\log P(y | X^c, X^d, A^c, A^d, V^d) = \sum_i^m \log P(y_i | y_{<i}, s), \quad (6)$$

where $X^d$ is the scientific papers’ vector representation and $m$ is the word number of the scientific paper. Once the network is fully trained, the scientific papers $y$ can be scored based on the citation context’s vector representations $X^c$, author information $A^c$ and $A^d$, and venue information $V^d$ via equation (6). Those papers which have higher raking score values are selected to generate a recommendation list.

5. Experiments

5.1. Datasets. We use AAN (http://tangra.cs.yale.edu/newaan/) and DBLP (https://dblp.uni-trier.de/) datasets to estimate the AED model-based approach’s performance. As DBLP dataset contains about 1 million bibliography records from the DBLP web set, we select a subset of it, i.e., computer security (ISI, NDSS, AREX, ACSAC FC, and SP), information retrieval (AIRS, CIKM, SIGIR, JCDL, ICTIR, ECIR, TREC, and WSDM), computer vision (ECCV, MM, CVPR, ICPR, ICIP, ACCV, and ICCV), networks and communications (MOBICOM, INFOCOM, ICNP, SECON, ICC, ICDCS, GLOBECOM, and SIGCOMM), and machine learning (ICDE, PAKDD, ICDM, SIGKDD, ICML, NIPS, and WSDM).

Those papers which have no abstracts or titles are removed from the datasets, and the remaining papers are used for experiments. We divide DBLP dataset as a test set and a training set; the training set contains 57,162 papers which are published in 2014 and 2015. AAN dataset is divided in the same way, and the training set of AAN contains 11,197 papers published from 1965 to 2013; the test set contains 1,358 papers. We choose the preceding three sentences and succeeding three sentences of each citation placeholder in each paper, and they constitute the local citation context. Besides, we extract each paper’s abstract and title as its text content.

5.2. Evaluation Index. We use the following three evaluation indices to evaluate the local citation recommendation’s performance.

5.2.1. Recall@N. It is defined as the ratio of ground-truth papers appearing in top $N$ papers of the recommended list. For this, we evaluate the recall index through $N = 30, 70, 100$.

5.2.2. Mean Average Precision (MAP). It measures average precision reflecting the rank position regarding the retrieval recommendation list. This indicator is based on the position of the corresponding label values for the top $K$ recommended reference papers. We measure the indicator with $K = 40$.

5.2.3. Normalized Discounted Cumulative Gain (NDCG). It measures the performance of a recommender system based on the graded relevance of the recommended papers. We calculate NDCG value using the following formula:

$$\text{NDCG}_n = \frac{1}{n} \sum_{j=1}^{n} \frac{2^{R(j)} - 1}{\log(1 + j)}, \quad (7)$$

where $R(j)$ represents the $j$-th paper’s rating value in the ranking list. If the retrieved paper is relevant, then $R(j) = 1$; otherwise $R(j) = 0$. $n$ is the position and $Z_n$ is a normalization factor.

5.3. Experiments and Discussion. For the proposed AED model, we set the dimension of embeddings, memory cell sizes, batch sizes, and feature maps to 128.

5.3.1. Performance of the AED-Based Citation Recommendation Approach. First, we aim to test whether other information beyond content information can enhance the citation recommendation’s performance, so we use scientific papers’ content information only. We run four sets of experiments: (1) AED-C, which incorporates content information of scientific papers, (2) AED-CA, which incorporates author and content information of scientific papers, (3) AED-CV, which incorporates content and venue information of scientific papers, and (4) AED-CAV, which incorporates scientific papers’ author, venue, and content information. Table 1 shows the results of the above methods.

As scientific papers’ content information can only provide coarse information, it leads to low similarity between the two text objects, i.e., scientific papers and citation context; thus, the AED-C approach performs poorest. Experimental results show the performance of AED-CV is inferior to that of AED-CA. We deem the reason is that the author information can provide fine-grained information than the venue information. We found that when we fully utilize scientific papers’ author, venue, and content information, we can obtain more accurate reference papers based on the given citation context.

Besides, we investigate whether personal information based on citation context can result in more relevant and individualized references. We denote the personalized citation context as $c_1 = [c_i, c_a]$ and denote the non-personalized citation context as $c_2 = [c_i]$. When an academic newcomer who does not have any publication, submits a citation context, the personalized AED based approach automatically reduces to non-personalized AED-based approach, since the citation context contains only content information $c_i$. Table 2 illustrates the experimental results.

5.3.2. Comparison with Other Baseline Approaches. To validate the efficiency of the proposed AED-based approach, we compare it with the following approaches: (1) Context-Aware Relevance (CAR) model [24], which is a non-parametric probabilistic model to recommend references for a given text segment; (2) Neural Citation Network (NCN) model [11], which integrates author information to enhance
the performance local citation recommendation performance, and (3) Translation Retrieval (TR) model [31], which narrows the semantic gap between the paper’s content and the citation context. Experimental results are listed in Table 3.

From the table, we found that CAR-based approach performs poorest; this can be due to the CAR approach using vector representation based on bag-of-words that loses the valuable semantic information between scientific papers and the given citation contexts. TR approach is better than CAR approach as the TR approach evaluates the semantic relevance between scientific papers and the given citation contexts. However, both CAR approach and TR approach perform worse than NCN approach. This is because CAR approach and TR approach concentrate on scientific papers’ content information only, but NCN approach utilizes author information besides content information. We found that the AED approach performs best in these four approaches. We attribute it to that the AED approach simultaneously utilizes multiple information related of scientific papers.

6. Conclusion

We develop an attention-based encoder-decoder (AED) model for local citation recommendation in this paper; it first constructs an encoder to represent a citation context as a vector in a low-dimensional space, then it constructs an attention mechanism integrating venue information and author information, and uses RNN to construct a decoder; finally, it maps the decoder’s output into a softmax layer and scores the scientific papers via the given citation context, author information, and venue information. A recommended reference paper list is generated based on scientific papers which have high score values. We conduct experiments on the DBLP and ACL Anthology Network (AAN) datasets, and the results illustrate that the performance of AED based approach is better than the other three baseline approaches.

Data Availability

The data used in our manuscript can be downloaded from the following two URLs: http://tangra.cs.yale.edu/newaan/ and https://dblp.uni-trier.de/. The two URLs have been also presented in the manuscript (refer Section 5.1).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant nos. 61772429 and 61872296) and in part by the MOE (Ministry of Education in China)
Project of Humanities and Social Sciences (grant no. 18YJCB70001), the China Postdoctoral Science Foundation (grant nos. 2017M63205 and 2017M623241), the Natural Science Basic Research Plan in Shaanxi Province of China (grant no. 2018QJ6031), the Postdoctoral Science Foundation in Shaanxi Province of China (grant nos. 2017BSHEDZZ85 and 2017BSHEDZZ86), and the Fundamental Research Funds for the Central Universities (grant no. 3102018zy026).

References

[1] Z. J. Fu, X. L. Wu, Q. Wang, and K. Ren, “Enabling central keyword-based semantic extension search over encrypted outsourced data,” IEEE Transactions on Information Forensics and Security, vol. 12, no. 12, pp. 2986–2997, 2017.

[2] M. Rusiñol, D. Aldavert, R. Toledo, and J. Lladós, “Efficient segmentation-free keyword spotting in historical document collections,” Pattern Recognition, vol. 48, no. 2, pp. 545–555, 2015.

[3] C. Chen, X. J. Zhu, P. S. Shen et al., “An efficient privacy-preserving ranked keyword search method,” IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 4, pp. 951–963, 2015.

[4] E. A. Khalid and G. Carlos, “Beyond keyword search: discovering relevant scientific literature,” in Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 439–447, San Diego, CA, USA, August 2011.

[5] M. Krallinger, O. Babal, A. Lourenco, J. Oyarzabal, and A. Valencia, “Information retrieval and text mining technologies for chemistry,” Chemical Reviews, vol. 117, no. 12, pp. 7673–7761, 2017.

[6] H. Liu, X. Kong, X. Bai, W. Wang, T. M. Bekele, and F. Xia, “Context-based collaborative filtering for citation recommendation,” IEEE Access, vol. 3, pp. 1695–1703, 2015.

[7] F. Xia, H. F. Liu, I. Lee, and L. B. Cao, “Scientific article recommendation: exploiting common author relations and historical preferences,” IEEE Transactions on Big Data, vol. 2, no. 2, pp. 101–112, 2016.

[8] L. Ma, L. Liao, D. Song, and J. Wang, “A latent entity-document class mixture of experts model for cumulative citation recommendation,” Tsinghua Science and Technology, vol. 23, no. 6, pp. 660–670, 2018.

[9] X. Tang, X. Wan, and X. Zhang, “Cross-language context-aware citation recommendation in scientific articles,” in Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR ’14), pp. 817–826, Gold Coast, Australia, July 2014.

[10] W. Huang, Z. Wu, C. Liang, P. Mitra, and C. Giles, “A neural probabilistic model for context based citation recommendation,” in Proceedings of the 29th AAAI Conference, pp. 2404–2410, Austin, TX, USA, January 2015.

[11] T. Ebesu and Y. Fang, “Neural citation networks for context-aware citation recommendation,” in Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR ’17), pp. 1093–1096, Tokyo, Japan, August 2017.

[12] X. Cai, J. Han, W. Li, R. Zhang, S. Pan, and L. Yang, “A three-layered mutually reinforced model for personalized citation recommendation,” IEEE Transactions on Neural Network and Learning Systems, vol. 29, no. 2, pp. 6026–6037, 2018.

[13] X. Cai, J. Han, S. Pan, and L. Yang, “Heterogeneous information network embedding based personalized query-focused astronomy reference paper recommendation,” International Journal of Computational Intelligence Systems, vol. 11, no. 1, pp. 591–599, 2018.

[14] X. Cai, J. Han, and L. Yang, “Generative adversarial network based heterogeneous bibliographic network representation for personalized citation recommendation,” in Proceedings of the 32nd AAAI Conference, pp. 5747–5754, New Orleans, LA, USA, February 2018.

[15] L. Yang, Y. Zheng, X. Cai, S. Pan, and T. Dai, “Query-oriented citation recommendation based on network correlation,” Journal of Intelligent & Fuzzy Systems, vol. 35, no. 4, pp. 4621–4628, 2018.

[16] C. Caragea, A. Silvescu, P. Mitra, and C. L. Giles, “Can’t see the forest for the trees?: a citation recommendation system,” in Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL ’13), pp. 111–114, Indianapolis, IN, USA, July 2013.

[17] X. Z. Liu, Y. Y. Yu, C. Guo, Y. Z. Sun, and L. C. Gao, “Full-text based context-rich heterogeneous network mining approach for citation recommendation,” in Proceedings of the IEEE/ACM Joint Conference on Digital Libraries, London, UK, September 2014.

[18] A. Livne, V. Gokuladas, J. Teevan, S. T. Dumais, and E. Adar, “CiteSight: supporting contextual citation recommendation using differential search,” in Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR ’14), pp. 807–816, Gold Coast, Australia, July 2014.

[19] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” in Proceedings of the ICLR 2015, pp. 1–15, San Diego, CA, USA, May 2015.

[20] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, “On the properties of neural machine translation: encoder-decoder approaches,” in Proceedings of the SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, pp. 1–9, Doha, Qatar, October 2014.

[21] K. Cho, B. van Merrienboer, C. Gulcehre et al., “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1724–1734, Doha, Qatar, October 2014.

[22] Z. D. Champiri, S. R. Shahamiri, and S. S. B. Salim, “A systematic review of scholar context-aware recommender systems,” Expert Systems with Applications, vol. 42, no. 3, pp. 1743–1758, 2015.

[23] L. Yang, Y. Zheng, X. Cai et al., “A LSTM based model for personalized context-aware citation recommendation,” IEEE Access, vol. 6, pp. 59618–59627, 2018.

[24] Q. He, J. Pei, D. Kifer, P. Mitra, and L. Giles, “Context-aware citation recommendation,” in Proceedings of the of 19th WWW, pp. 421–430, Raleigh, NC, USA, April 2010.

[25] Q. He, D. Kifer, J. Pei, P. Mitra, and C. Giles, “Citation recommendation without author supervision,” in Proceedings of the Fourth ACM International Conference on Web Search and Data Mining (WSDM ’11), pp. 755–764, New York, NY, USA, October 2011.

[26] D. Duma, M. Liakata, A. Clare, J. Ravenscroft, and E. Klein, “Rhetorical classification of anchor text for citation recommendation,” D-Lib Magazine, vol. 22, no. 9/10, 2016.

[27] X. Tang, X. Wan, and X. Zhang, “Cross-language context-aware citation recommendation in scientific articles,” in Computational Intelligence and Neuroscience.
[28] H. Peng, J. Liu, and C. Lin, "News citation recommendation with implicit and explicit semantics," in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pp. 388–398, Berlin, Germany, August 2016.

[29] S. P. Zhou, ActiveCite: An Interactive System for Automatic Citation Suggestion, National University of Singapore, Singapore, 2010.

[30] W. Huang, Z. Wu, P. Mitra, and C. Giles, “ReSeer: a citation recommendation system,” in Proceedings of the of IEEE/ACM Joint Conference on Digital Libraries, pp. 371–374, London, UK, September 2014.

[31] Y. Lu, J. He, D. Shan, and H. Yan, “Recommending citations with translation model,” in Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM ’11), pp. 2017–2020, Glasgow, Scotland, October 2011.

[32] W. Huang, S. Kataria, C. Caragea, P. Mitra, C. Giles, and L. Rokach, “Recommending citations translating papers into references,” in Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM ’12), pp. 1910–1914, Maui, HI, USA, November 2012.

[33] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in Proceedings of the International Conference on Learning Representations (ICLR), San Diego, CA, USA, May 2015.

[34] M. S. H. Ameur, F. Meziane, and A. Guessoum, “Arabic machine transliteration using an attention-based encoder-decoder model,” Procedia Computer Science, vol. 117, pp. 287–297, 2017.

[35] R. Collobert and J. Weston, "A unified architecture for natural language processing: deep neural networks with multitask learning," in Proceedings of the International Conference on Machine Learning, Helsinki, Finland, July 2008.

[36] K. Cho, B. Van Merriënboer, C. Gulcehre et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in Proceedings of the Empirical Methods in Natural Language Processing (EMNLP), Brussels, Belgium, October 2014.

[37] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, Cambridge, MA, USA, 2016.
