Context-Aware Citation And Cooperator Recommendation

Junjie Wang
Department of Computer Science
Shanghai Jiao Tong University
dreamboy.gns@sjtu.edu.cn

1 Introduction

As the number of scientific publications increases rapidly, it becomes more and more difficult to search for relevant works or have a comprehensive knowledge about the most recent progress in specific domain. On the one hand, search engines or information retrieval methods are highly keyword dependent. On the other hand, following specific top conferences or journals limit the scope of relevant research. In this work, different from research which makes recommendations based on user profile and paper feature, we focus on context-aware citation recommendation. We provide personalized, high-quality recommendations for each incoming query(context) based on only titles and authors of candidate papers.
Classic information retrieval methods are highly keyword dependent and are unable to capture dynamic changes of word meanings. Such dynamic changes mainly lie in two aspects: 1) the meaning of the same phrase may be quite different across different ages. For example, in 1950s, "neural network" may more refer to biomedical mechanism in human bodies, while currently we may come up with "deep learning" and "artificial intelligence". Generally techniques that based on keywords overlap can not handle such variance. 2) there may exist a large gap between the word usage of different authors. Therefore the system may fail to make proper recommendations if the authors of the context and cited document have quite different writing habits.
To capture the latent semantics behind incoming context, we build our work on Neural Citation Network(NCN)[4]. Specifically, we make some improvements on the network architecture and propose an efficient pipeline for citation and cooperator recommendation. Experiments on Refseer dataset show that our method is more effective than traditional IR methods and NCN.

2 Related Work

Paper Recommendation. Many researchers focus on building a paper recommendation based on user profile(built by user behavior or tags) and paper features. There are mainly 4 kinds of methods in the literature. 1)[8][7][3] use content-based filtering methods(CBF). They first build up the item representation(i.e. paper) based on word frequency and co-occurrence, then describe users by their behaviors(likes or dislikes) and finally generate recommendations by computing similarities between users and papers. 2) collaborative filtering(CF). The basic idea of CF is that if users A and B make ratings on some common items, theirs interests will be considered similar. If there are some items existing in user B's record but not in user A's, these items can be recommended to user A. There are two strategies, user-based and item-based. In the former one, we first find neighbouring users and re-rank items based on their ratings. In the latter strategy[2][9], we exploit rating history and recommend based on item similarity. Finally, recommendations are made based on similarities. 3) Graph-based methods. Graphs are constructed based on user behaviors(e.g. interests in specific works, having common authors) instead of paper features or user profiles in the heterogeneous graph. Then algorithms like random walk, PaperRank for citation network are used to generate proposals. 4) Hybrid method(HM). Two or more above techniques are combined to provide more high-quality recommendations.

Citation Recommendation. A wide range of technologies have been applied to the task of citation recommendation, including traditional IR, topic modeling, statistical translation machine(SMT)[5], deep neural networks. SMT aims to learn an alignment from the citation context to cited document, giving the probability of requiring a citation. NMT offers a general framework to process parallel pairs of arbitrary length sequences. The source sequence is encoded into fixed length features and then be translated by a decoder into target sequence conditioned on previous states. In machine translation task usually RNNs are used as the encoder/decoder. [1] also proposes to use attention mechanism between the encoder-decoder architecture to strengthen the representation ability.

3 Method

3.1 Review Of NCN

We first review the framework of neural citation network[4]. The overall pipeline is shown in 1. For each incoming pair, NCN first embeds context(by an embedding layer) into $C \in \mathbb{R}^{n \times q}$ and embeds cited title into $T \in \mathbb{R}^{n \times q}$. Then NCN uses several one-dimensional filters with strides $\{l_1, l_2, ..., l_k\}$ to convolve C, covering information from various windows. After convolution, NCN obtains $\{o_1, o_2, ..., o_k\}$ where $o_i \in \mathbb{R}^{(n-l_i+1) \times h}$ and h is hidden size. Then NCN performs max-pooling over each output, yielding $\{c_1, c_2, ..., c_k\}$ where $c_i \in \mathbb{R}^h$. Then encoded results and title embeddings are passed on to an attention decoder. To combine the information from citing author and cited authors, NCN also use similar encoder network to encode author embeddings. To capture the sequential relationship, GRU is employed as the decoder, and at each timestep, hideen state is conditioned on encoded results(by the attention mechanism) and previous decoder states. The decoder also outputs a probability distribution over the whole vocabulary. Then distributions from different timestep are averaged and used to score each context-paper pair.

$$P(y_i|\mathbf{X}^q, \mathbf{X}^d, \mathbf{A}^q, \mathbf{A}^d) = \frac{1}{m} \sum_i P(y_{\leq i}|s)$$

where $\mathbf{X}^q, \mathbf{X}_d$ denote context and cited title embeddings, $\mathbf{A}^q, \mathbf{A}^d$ denote the author embeddings of context and cited title.

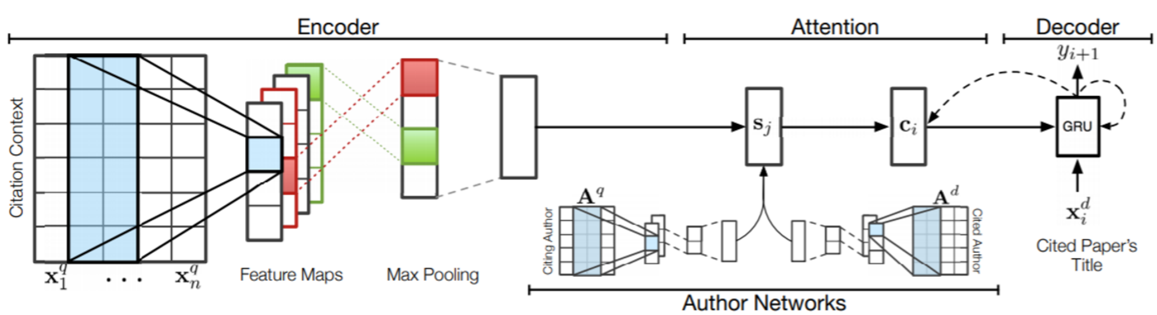

Figure 1: Network In NCN

3.2 Proposed Network

Intuitively, we think there is much room for improvement in terms of the network architecture of NCN and we propose our network as follows:

1. **Average Pooling.** NCN use several filters of different region sizes to convolve the input sequence and perform max-pooling over different word windows for each region size. The problem with max-pooling is that it assumes only one word window should be activated and remaining ones should be suppressed, while usually in incoming contexts, only the entire sequence can have a continuous and complete semantic interpretation. We use average pooling over the entire sequence to obtain a more reasonable feature vector.

2. **One-dimensional convolution.** NCN performs one-dimensional convolution over input context to capture spatial relationship and semantics, but the cited title is ignored. Intuitively, the cited title should also take care of its context. Based on this motivation, we perform one-dimensional convolution on target titles before passing them to the decoder.

3. **Projection before attention decoder.** NCN concatenates feature representations of contexts and authors directly after corresponding encoders. Intuitively, there is a gap between features from two different sources(i.e. authors and contexts) and direct concatenation may not be a good choice. Therefore we propose to project them to common subsapce(i.e. use affine projection layers) to get a more reasonable and compatible concatenation.

The modified network is illustrated in 2.

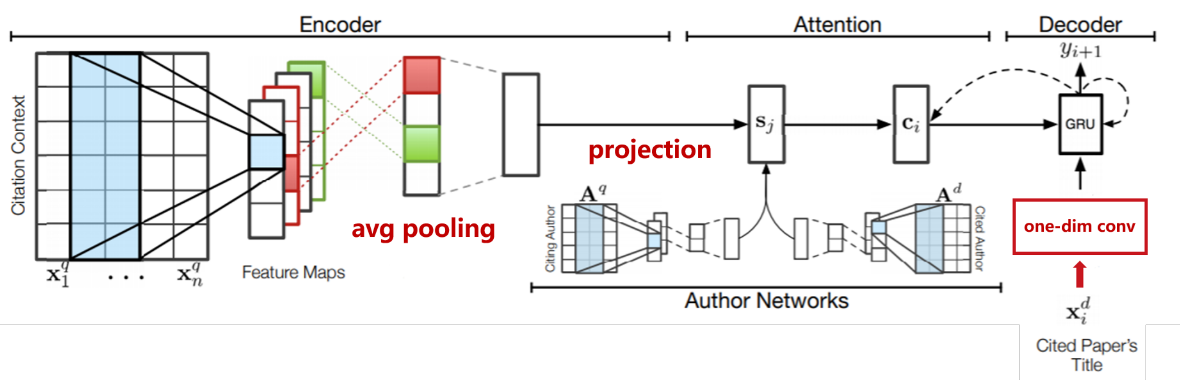

Figure 2: Proposed Network

3.3 Proposed Pipeline

3.3.1 Cooperator Recommendation

Obviously it makes no sense to design a similar encoder-decoder network to recommend authors(since probability over author vocabulary is meaningless). Therefore we utilize previous paper recommendation results to recommend potential cooperators. Formally, we rank the candidate authors as follows and take the top ones.

$$S_a = \sum_{a \in P_a} S_p$$

where S_p is the paper score given by our model, and P_a is the paper set written by author a.

3.3.2 Efficient Inference

Generally we have a large database containing candidate papers and their authors. It should take a long time if we compose each document with incoming context and pass on to our model. An intuitive idea is to pre-process composed pairs using traditional IR methods(e.g. BM-25) and only take the top ones.

3.3.3 Overall Pipeline

The overall pipeline is shown in 3. We first use traditional IR method to filter out most possible cited documents and compose them with incoming context one by one, pass on the our model and obtain paper scores. Paper scores are used to make paper recommendations, compute author scores and make cooperator recommendations.

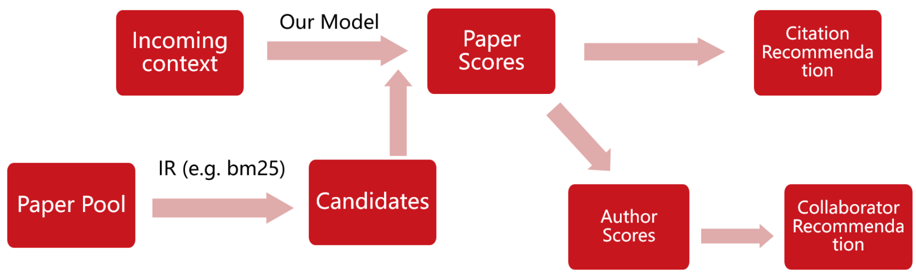

Figure 3: Proposed Pipeline

4 Experiments

4.1 Dataset And Setup

We evaluate proposed network on RefSeer dataset[6]. After preprocessing invalid entries, we obtain 4,549,267 context pairs with 855,735 papers in a citation-cited relation. We divide these pairs by year, and use 4,258,383 pairs before 2013 for training, 141,957 pairs in 2013 for testing.

4.2 Results

The training loss curve is shown in 4. We can observe that proposed network converges much faster than NCN and converges to a much lower loss value, proving the effectiveness of our proposed network. Also quantitative results are summarized in Tab. 1. The results by our implementation is much worse compared with original results released by [4]. A possible reason is that, since we don't have titles of cited documents in pre-processed dataset, we use random sampling instead for each incoming context and true positive may not be included in sampled candidates.

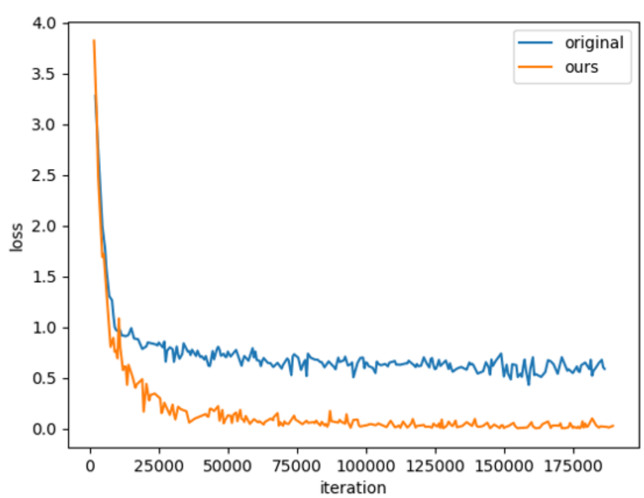

Figure 4: Training Loss Curve

Method	MAP	MRR
NCN	0.2418	0.2667
NCN(ours)	0.0144	0.0211
Ours	0.0172	0.0224

Table 1: Qualitative Results Of Our Model And NCN

5 Conclusion

In this paper we focus on the task of context-aware citation recommendation. We build up our network and method on the basis of neural citation network. Experiments validate the effectiveness of our method.

References

[1] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. In Y. Bengio and Y. LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL `http://arxiv.org/abs/1409.0473`.

[2] J. Beel, B. Gipp, S. Langer, and C. Breitinger. Research-paper recommender systems: a literature survey. *Int. J. on Digital Libraries*, 17(4):305–338, 2016. doi: 10.1007/s00799-015-0156-0. URL `https://doi.org/10.1007/s00799-015-0156-0`.

[3] C. Caragea, F. A. Bulgarov, A. Godea, and S. D. Gollapalli. Citation-enhanced keyphrase extraction from research papers: A supervised approach. In A. Moschitti, B. Pang, and W. Daelemans, editors, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1435–1446. ACL, 2014. doi: 10.3115/v1/d14-1150. URL `https://doi.org/10.3115/v1/d14-1150`.

[4] T. Ebesu and Y. Fang. Neural citation network for context-aware citation recommendation. In N. Kando, T. Sakai, H. Joho, H. Li, A. P. de Vries, and R. W. White, editors, *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, August 7-11, 2017*, pages 1093–1096. ACM, 2017. doi: 10.1145/3077136.3080730. URL `https://doi.org/10.1145/3077136.3080730`.

[5] J. He, J. Nie, Y. Lu, and W. X. Zhao. Position-aligned translation model for citation recommendation. In L. Calderón-Benavides, C. N. González-Caro, E. Chávez, and N. Ziviani, editors, *String Processing and Information Retrieval - 19th International Symposium, SPIRE 2012, Cartagena de Indias, Colombia, October 21-25, 2012. Proceedings*, volume 7608 of *Lecture Notes in Computer Science*, pages 251–263. Springer, 2012. doi: 10.1007/978-3-642-34109-0_27. URL `https://doi.org/10.1007/978-3-642-34109-0_27`.

[6] W. Huang, Z. Wu, C. Liang, P. Mitra, and C. L. Giles. A neural probabilistic model for context based citation recommendation. In B. Bonet and S. Koenig, editors, *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA*, pages 2404–2410. AAAI Press, 2015. URL `http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9737`.

[7] P. Jomsri, S. Sanguansintukul, and W. Choochaiwattana. A framework for tag-based research paper recommender system: An IR approach. In *24th IEEE International Conference on Advanced Information Networking and Applications Workshops, WAINA 2010, Perth, Australia, 20-13 April 2010*, pages 103–108. IEEE Computer Society, 2010. doi: 10.1109/WAINA.2010.35. URL `https://doi.org/10.1109/WAINA.2010.35`.

[8] M. J. Pazzani and D. Billsus. Content-based recommendation systems. In P. Brusilovsky, A. Kobsa, and W. Nejdl, editors, *The Adaptive Web, Methods and Strategies of Web Personalization*, volume 4321 of *Lecture Notes in Computer Science*, pages 325–341. Springer, 2007. doi: 10.1007/978-3-540-72079-9_10. URL `https://doi.org/10.1007/978-3-540-72079-9_10`.

[9] D. Valcarce, J. Parapar, and Á. Barreiro. Item-based relevance modelling of recommendations for getting rid of long tail products. *Knowl. Based Syst.*, 103:41–51, 2016. doi: 10.1016/j.knosys.2016.03.021. URL `https://doi.org/10.1016/j.knosys.2016.03.021`.