Neural Citation Network for Context-Aware Citation Recommendation

Travis Ebesu, Yi Fang

Department of Computer Engineering
Santa Clara University
{tebesu,yfang}@scu.edu

Santa Clara University
School of Engineering

Problem

Given a short passage of text (citation context), we recommend a list of high-quality candidate papers to cite or fill the citation placeholder. For example:

Citation Context: ...a great deal of recent research builds upon Ponte and Croft's initial proposal [Citation Placeholder] wherein the rank of a document d is based on the probability...

Cited Paper: Jay M. Ponte and W. Bruce Croft. 1998. A Language Modeling Approach to Information Retrieval. (SIGIR '98).

Key Contributions

- We propose a flexible encoder-decoder architecture Neural Citation Network (NCN) embodying a robust representation of the citation context with a max time delay neural network, further augmented with an attention mechanism, author networks and a gated recurrent unit decoder.
- We demonstrate the effectiveness of integrating the author networks on the large scale RefSeer dataset significantly outperforming baselines across four different metrics.

Neural Citation Network

Encoder, a Time Delay Neural Network (TDNN) $f(\cdot)$, capturing phrase level semantics over the citation context $\mathbf{X}^q$ with a sliding window of convolutions followed by max pooling

$$o_k = \mathrm{ReLU}(\mathbf{w}^\intercal \mathbf{x}^q_{k:k+l-1} + b_k); \qquad \hat{o} = \max\{o_1, \ldots, o_{n-l+1}\}$$

Next a fully connected layer captures complex nonlinear interactions between each phrase

$$\mathbf{s}_j = \tanh(\mathbf{U}_{s_j}\hat{\mathbf{o}}_j + \mathbf{b}_{s_j})$$

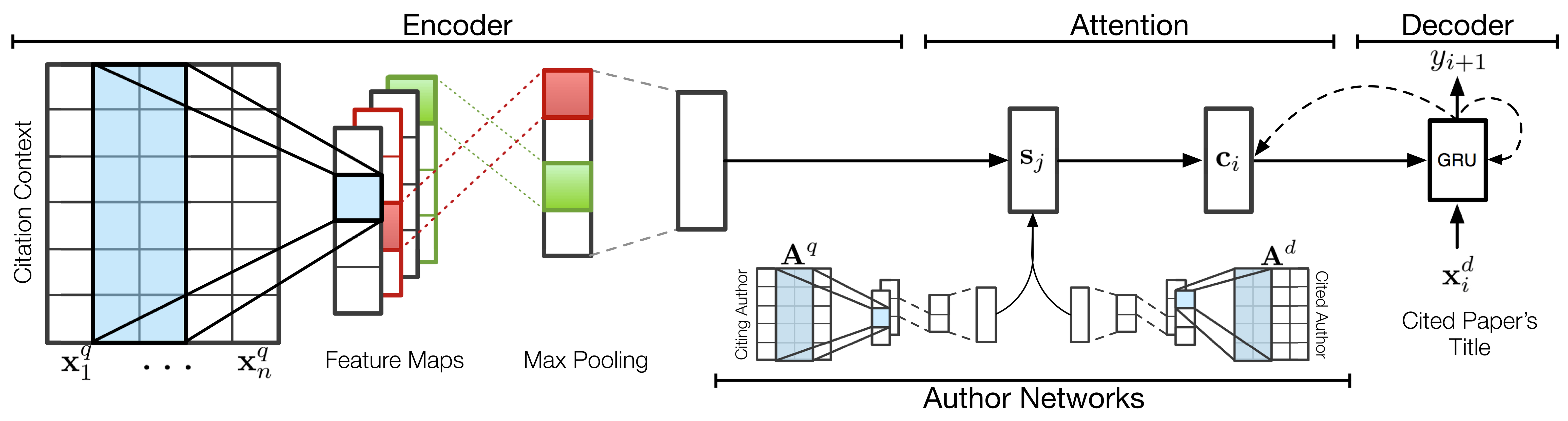
Author Networks encode both the citing author(s) $\mathbf{A}^q$, and cited author(s) $\mathbf{A}^d$, with two separate TDNNs concatenated with the citation context

$$\mathbf{s}_j = [f(\mathbf{X}^q) \oplus f(\mathbf{A}^q) \oplus f(\mathbf{A}^d)]_j$$

Attention Mechanism iteratively considers the importance of the citation context, citing author(s) and cited author(s) during the decoding process

$$\mathbf{c}_i = \sum_j \alpha_{ij}\mathbf{s}_j \qquad \text{where } \alpha_{ij} = \mathrm{softmax}(\mathbf{v}^\intercal \tanh(\mathbf{W}_a \mathbf{h}_{i-1} + \mathbf{U}_a \mathbf{s}_j))$$

Decoder leverages a Gated Recurrent Unit (GRU) to conditionally score the cited paper's title given the citation contexts and authors.



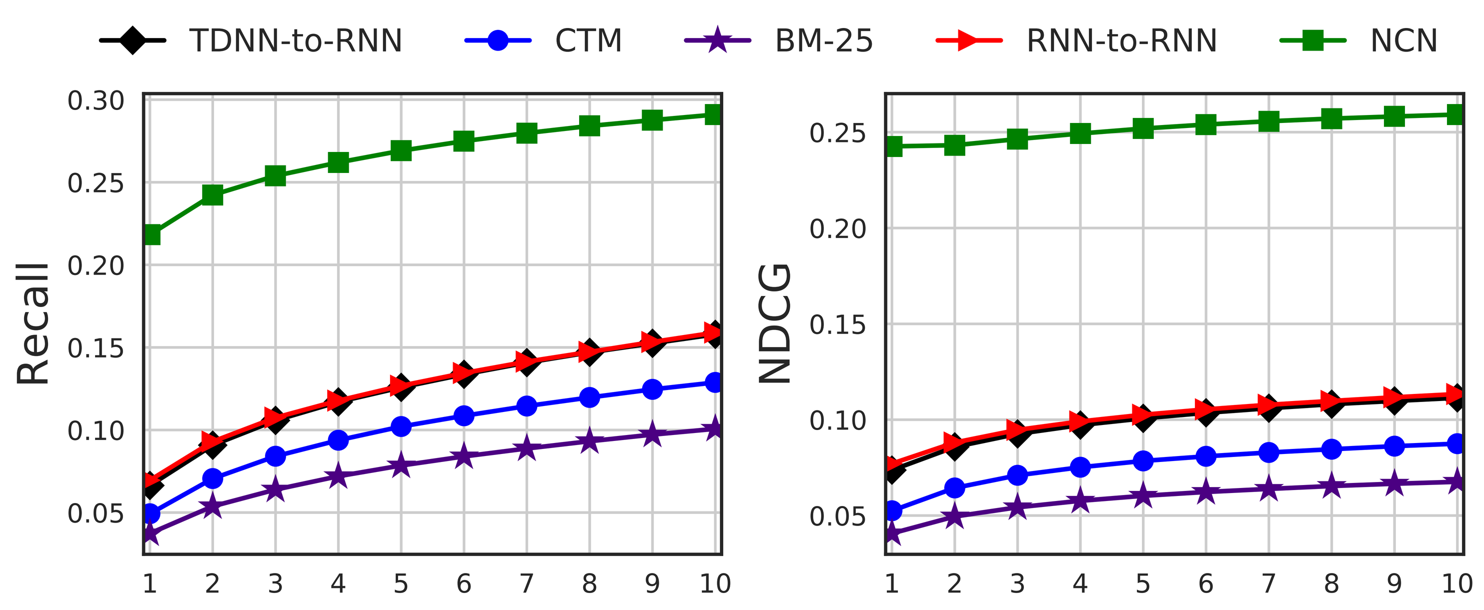
Neural Citation Network's (NCN) architecture with the attention mechanism and author networks. The dashed arrows represent recurrent dependencies.

Experiments

Results on the Refseer dataset consisting of 4,549,267 context pairs with 855,735 papers in a citation-cited relation.

	Recall	MAP	MRR	NDCG
BM-25	0.1007	0.0556	0.0606	0.0676
CTM	0.1288	0.0726	0.0777	0.0875
RNN-to-RNN	0.1590	0.0958	0.1054	0.1134
TDNN-to-RNN	0.1579	0.0935	0.1032	0.1114
Neural Citation Network	**0.2910**	**0.2418**	**0.2667**	**0.2592**

Performance comparison of the top 10 recommendations against baselines.



Recall and NDCG as the number of recommendations vary from 1 to 10.

Context:	"find a distribution over the latent variables that is close to the posterior of interest [Citation Placeholder]. Variational methods provide effective approximations in topic models and nonparametric Bayesian models"
NCN:	1. **Graphical models, exponential families, and variational inference** 2. Graphical models and variational methods 3. **An introduction to variational methods for graphical models**
CTM:	1. Indexing by latent semantic analysis 2. **An introduction to variational methods for graphical models** 3. Bayesian data analysis
RNN-to-RNN:	1. **An introduction to variational methods for graphical models** 2. The variational formulation of the Fokker–Planck equation 3. A Bayesian analysis of the multinomial probit model with fully identified parameters

Top 3 recommendations for Neural Citation Network (NCN), CTM, and RNN-to-RNN for the citation context (query), correct recommendations are in bold.