Neural Multi-task Learning for Citation Function and Provenance

Xuan Su, Animesh Prasad, Min-Yen Kan, Kazunari Sugiyama

School of Computing, National University of Singapore

Two Citation Analysis Tasks

Citation Function: In academia, why do authors make citations? The task of citation function assigns one out of a set of predefined rhetorical roles to a given citation.

Citation Provenance: The task of citation provenance identifies the cited texts in the cited paper, corresponding to a given citation.

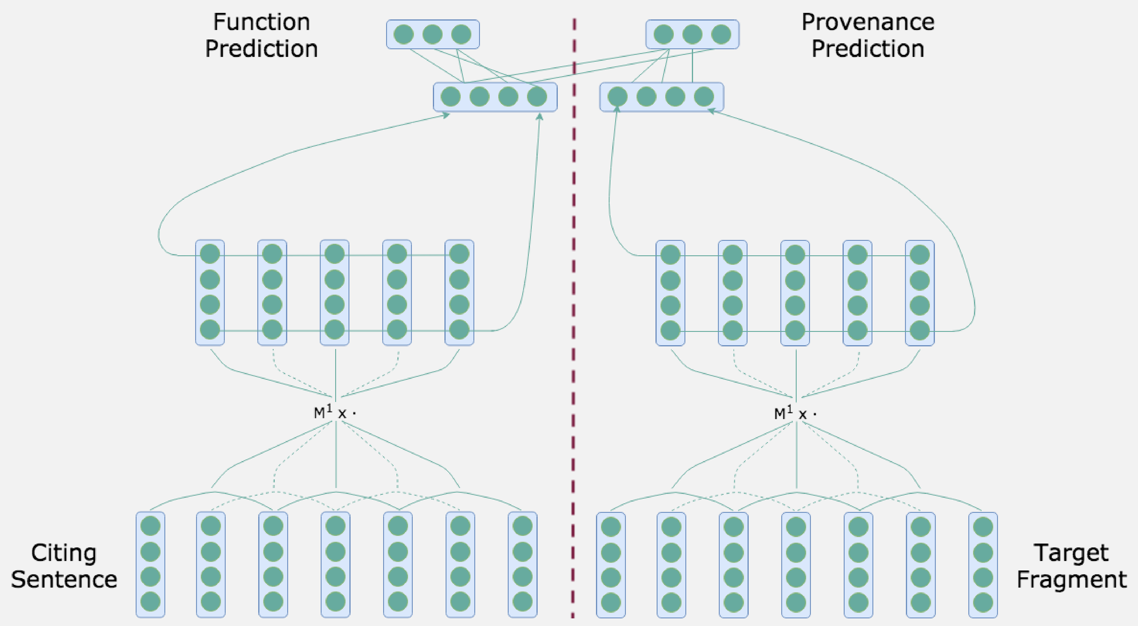
Our Contribution: Most existing approaches to these two tasks use conventional machine learning models. We are the first to apply multi-task learning [1] on top of a convolutional neural network to deal with both tasks.

Dataset

Citation Function: We adopt the classification scheme in [2]: (*Weak*)ness, Compare and Contrast (*CoCo*), (*Pos*)itive, and (*Neut*)ral. For example, a *Weak* citation reveals a drawback of the cited paper. We manually annotated data to create our dataset of 1,432 instances: 1,011 *Neut*, 295 *Pos*, 95 *CoCo* and 31 *Weak*.

Citation Provenance: We use a binary scheme, *Prov* and *-Prov*. Given a citation, a text fragment is labeled *Prov* if it contains evidence for the cited information. We directly use the public dataset in the CL-SciSumm Shared Task 2016 [3], where Task 1A requires participants to identify texts in the cited paper containing evidence. Since the dataset only contains positive (*+Prov*) instances, we source for our own *-Prov* instances by randomly sampling three negative instances from each cited paper. The final dataset has 608 *+Prov* and 885 *-Prov* instances (1,493 in total).

Models



1. For a single text sequence, we use a simple convolutional neural network (**CNN**) as base model (*cf.* individual halves in above figure).
2. For citation provenance, we use a double CNN ("**dCNN**") architecture, since both the citation context and the cited paper fragment need to be considered. Two CNNs accept and process the two inputs separately, but combine at the FC layer to generate class predictions.
3. We claim that there is a relationship between the function and provenance labels for a given citation. For example, a (*Pos*)itive citation is unlikely to refer to sentences with negative

Models

connotations. This motivates us to apply multi-task learning (MTL) to share weights across tasks which improves learning efficiency and prediction accuracy [1].

Experiments

We use 5-fold cross validation, and evaluate our models based on precision, recall, and F_1 scores weighted over all classes. Baselines are from [2] and [4]. CNN models have at-pat or superior performance compared to the baselines; MTL brings further improvement of about 1%.

Model	BL	CNN	MTL
Precision	68.28%	68.78% ± 0.51%	69.55% ± 0.61%
Recall	69.40%	68.65% ± 0.68%	72.33% ± 0.36%
F_1	68.70%	68.31% ± 0.52%	69.63% ± 0.47%

Table: Citation Function

Model	BL	dCNN	MTL
Precision	71.82%	79.36% ± 1.71%	79.47% ± 1.37%
Recall	72.13%	79.07% ± 1.84%	79.53% ± 1.36%
F_1	71.68%	78.55% ± 1.67%	79.38% ± 1.36%

Table: Citation Provenance

Examples

Citing Sentence	Actual	BL	MTL
(a1) We show that the performance of our approach (using simple lexical features) is comparable to that of the state-of-art statistical MT system (Koehn et al., 2007).	*CoCo*	*Pos*	*CoCo*
(a2) Errors have been shown to have a significant impact on predicting learner level (Yannakoudakis et al., 2011).	*Neut*	*Weak*	*Neut*

Citing Sentence	Target Fragment	Actual	BL	dCNN
(b1) Bigrams have recently been shown to be very successful features in supervised word sense disambiguation (Pedersen, 2001).	This paper shows that the combination of a simple bigrams and a standard decision tree learning algorithm results in accurate word sense disambiguation.	+*Prov*	+*Prov*	+*Prov*
(b2) A number of automatically acquired inference/rule paraphrase collections are available, such as (Szpektor et al., 2004).	In this paper, we will propose an unsupervised method to discover paraphrases from a large untagged corpus.	+*Prov*	−*Prov*	+*Prov*

Examples

Example (b2) shows advantages of a deep learning model using word embeddings. There are almost no word overlap between the two text inputs, hence the baseline relying on word overlap fails to classify it correctly. However, phrases such as "automatically acquired", "inference" share a common meaning as "unsupervised". GloVe embeddings successfully capture semantic relationships.

Conclusions

We leverage our key insight of the relationship between citation function and provenance, and employ MTL on top of neural models, resulting in performance improvement.

References

[1] Rich Caruana.
Multitask Learning.
Learning to Learn, pages 95–133, 1998.

[2] Eric Yulianto.
Citation Typing.
Bachelor's thesis. School of Computing, National University of Singapore, 2012.

[3] Kokil Jaidka, Muthu Kumar Chandrasekaran, Sajal Rustagi, and Min-Yen Kan.
Insights from CL-SciSumm 2016: the Faceted Scientific Document Summarization Shared Task.
International Journal on Digital Libraries, 19(2-3):163–171, 2018.

[4] Heng Wee Low.
Citation Provenance.
Bachelor's thesis. School of Computing, National University of Singapore, 2011.

[5] Animesh Prasad.
WING-NUS at CL-SciSumm 2017: Learning from Syntactic and Semantic Similarity for Citation Contextualization.
In *Proc. of BIRNDL*, pages 26–32, 2017.

Contact Information

- Email: suxuan@comp.nus.edu.sg

NUS National University of Singapore