Automatic Academic Paper Rating Based on Modularized Hierarchical Convolutional Neural Network

Pengcheng Yang[1], Xu Sun[1,2], Wei Li[2], Shuming Ma[2]

[1]Deep Learning Lab, Beijing Institute of Big Data Research, Peking University

[2]MOE Key Laboratory of Computational Linguistics, Peking University

{yang_pc, xusun, liweitj47, shumingma}@pku.edu.cn

Abstract

●**Task:**

Automatically determine whether to accept an academic paper.

●**Motivation:**

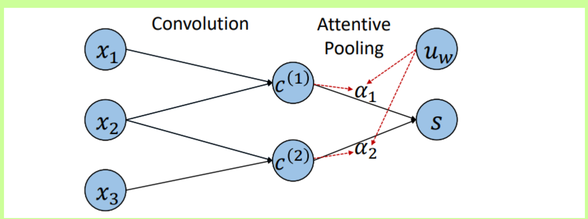
- More and more academic papers are being submitted to conferences and journals.
- Evaluating papers by professionals is time-consuming and can cause inequality due to the personal factors.

●**Proposal:**

- A new dataset for automatic academic paper rating.
- A modularized hierarchical convolutional neural network.

Attention-Based Convolutional Neural Network

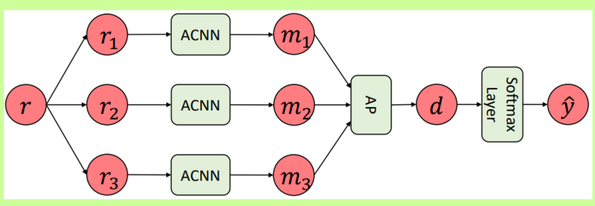
●**Model:**



$(x_1, x_2, \cdots, x_m)$ is the input sequence, u_w is parameter vector, s is the high slevel representation of the whole sequence.

Modularized Hierarchical Convolutional Neural Network

●**Model:**



ACNN denotes attention-based CNN and AP denotes attentive pooling. r_i and m_i represent the token sequence and high ssssslevel representation of the i-th module, respectively. d denotes the final representation of the source paper.

Proposed Model

●**Modularize:**

The source paper r **—>** Several modules $(r_1, r_2, \cdots, r_l)$.

●**Module representation:**

- Input the token sequence r_i.
- **ACNN:** word level —> sentence level —> module level.
- Output the module sslevel representation m_i.

●**Aggregation and classification:**

- Aggregate $(m_1, m_2, \cdots, m_l)$ to d.
- Perform classification based on d.

Experiment

●**Result:**

Models	Accuracy	Models	Accuracy
RP	50.0%	Logistic	60.0%
CART	58.6%	KNN	60.3%
MNB	58.3%	GNB	58.5%
SVM	61.6%	AdaBoost	58.9%
Bagging	59.4%	LSTM	60.5%
CNN	61.3%	C-LSTM	60.8%
MHCNN	**67.7%**		

●**Conclusion:**

- The proposed MHCNN outperforms all baselines.
- The modularized hierarchical structure and attention mechanism are of great help to improve accuracy.

Ablation Study

●**Result:**

Models	Accuracy	Decline
MHCNN	67.7%	−−
w/o Attention	66.8%*	↓0.9%
w/o Module	61.3%*	↓6.4%

●**Conclusion:**

- Either the modularized hierarchical structure or the attention mechanism is of great help to improve accuracy.
- The modularized hierarchical structure of the model is beneficial to obtain better representations by incorporating structure knowledge of the source paper.

Comparison of Various Parts of the Source Paper

●**Result:**

Contexts	Accuracy	Decline
Full data	67.7%	−−
w/o *Title*	66.6%*	↓1.1%
w/o *Abstract*	65.5%*	↓2.2%
w/o *Authors*	64.6%*	↓3.1%
w/o *Introduction*	65.7%*	↓2.0%
w/o *Related work*	66.0%*	↓1.7%
w/o *Methods*	66.2%*	↓1.5%
w/o *Conclusion*	65.0%*	↓2.7%

●**Conclusion:**

- Except for *authors*, the two most significant modules affecting acceptance are *conclusions* and *abstract*.
- The impact of the *title* is the smallest.

Conclusion

We propose the task of automatic academic paper rating (AAPR), which aims to automatically determine whether to accept academic papers. We propose a novel modularized hierarchical CNN for this task to make use of the structure of a source paper. Experimental results show that the proposed model outperforms various baselines by a large margin. In addition, we find that the conclusion and abstract parts have the most influence on whether the source paper can be accepted when setting aside the factor of authors.