Glocal: Incorporating Global Information in Local Convolution for Keyphrase Extraction

Animesh Prasad and Min-Yen Kan
School of Computing, National University of Singapore
{animesh, kanmy}@comp.nus.edu.sg

❖ Introduction

➢ **Motivation**:
Sequential labeling models fail to scale well for large documents.

➢ **Current State of the Art**:
Feature-rich candidate selection and ranking techniques or sequence labeling over a subset of the document (e.g., Abstract)

➢ **Key Observations:**
- structure of the graph formed by the text concurrence is not exploited.
- random walk scores are used as-is, without incorporation of semantic features or fine-tuning via supervision.

➢ **Proposed Glocal (global–local *portmanteau*)**
- ✓ Incorporate random walks as importance (Global)
- ✓ Learn from node specific features (Local: Self)
- ✓ Use neighborhood to enhance node-specific features (Local: Neighborhood)

❖ Method

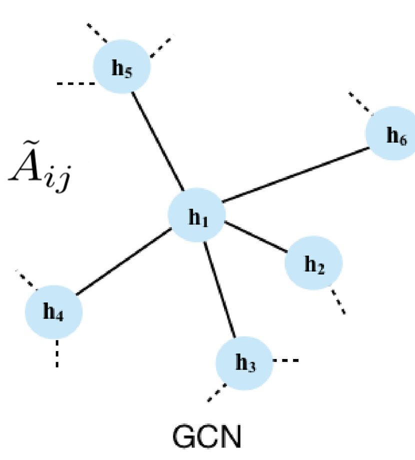
➢ **Graph Convolutional Network**

$$y \approx \theta\left(I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}\right) x$$

$$\tilde{A} = A + I_N \qquad \tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$$

$$H^{(0)} = X$$

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}\right)$$



GCN

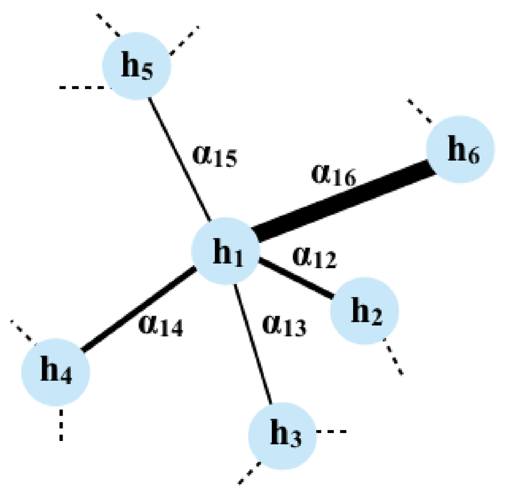
Aggregates features in a well-defined neighborhood over normalized adjacency

➢ **Graph Attention Network**

$$a_{ij} = attn(W\vec{h}_i, W\vec{h}_j)$$

$$\alpha_{ij} = \frac{\exp(a_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(a_{ik})}$$

$$\vec{h}_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} W \vec{h}_j^{(l)}\right)$$



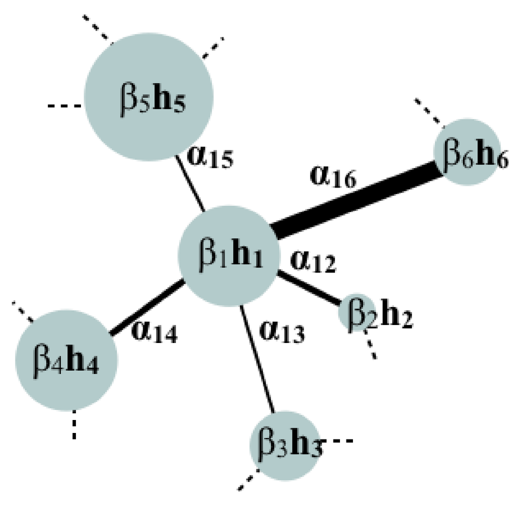
GAT

Aggregates features in a soft neighborhood over normalized pair-wise parametrized adjacency

➢ **Glocal Convolution**

$$\beta_j = TextRank(i)$$

$$\vec{h}_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}_i} \beta_j \alpha_{ij} \mathbf{W} \vec{h}_j^{(l)}\right)$$



Glocal Convolution

Aggregates features in a soft neighborhood over normalized pair-wise parametrized adjacency and scales features to their global importance

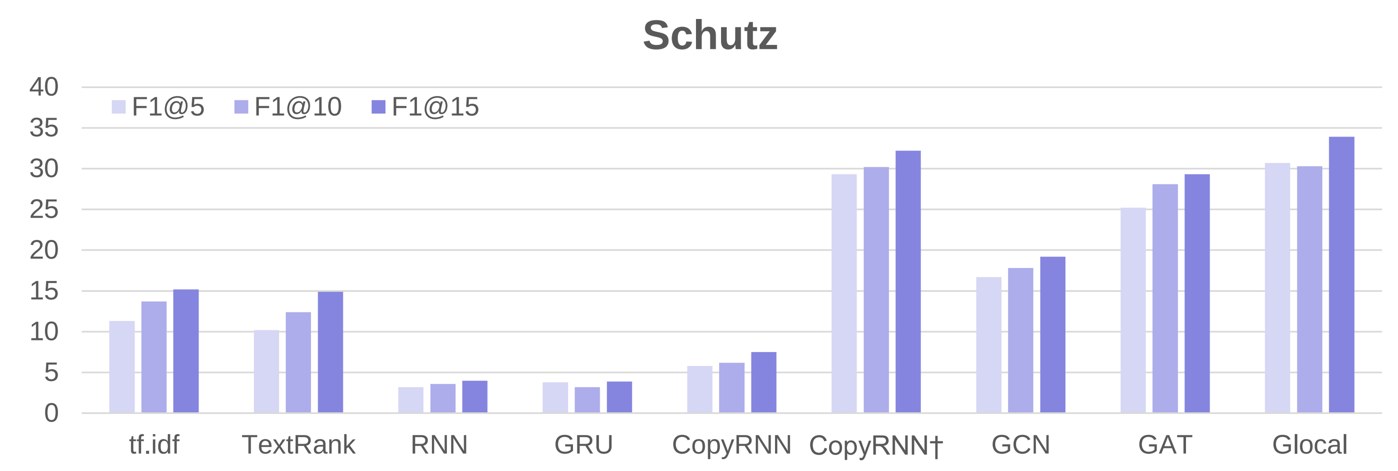
TextRank based importance cannot be learned locally.

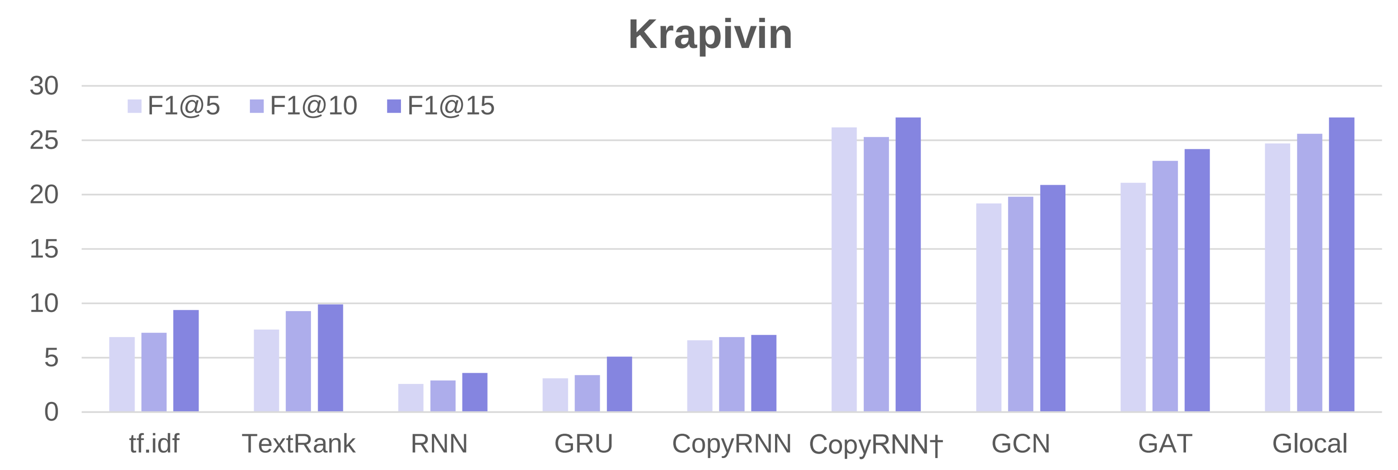
➢ **Steps:**
1) Construct text graph per document where each node is a token and edges are co-occurrences in neighborhood.
2) Train the Glocal model on graph with binary classification target (*keyphrase*, ¬*keyphrase*).
3) During testing, use the output probability to re-rank and generate multi-token keyphrases as:

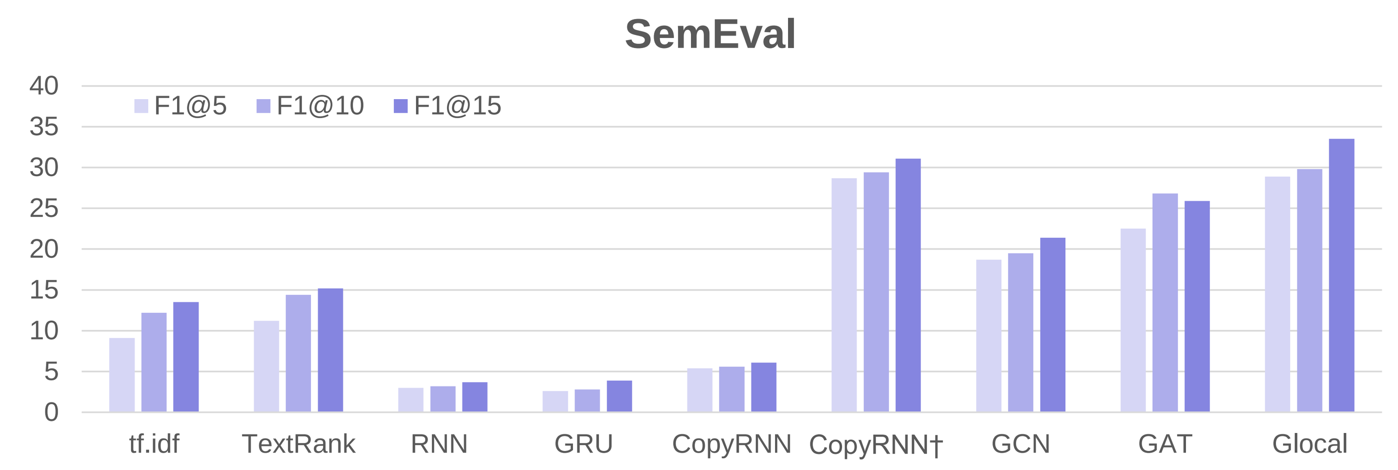
$$R(p) = len(p) * \sum_{w_i \epsilon p} r(w_i)$$

❖ Experiments & Results

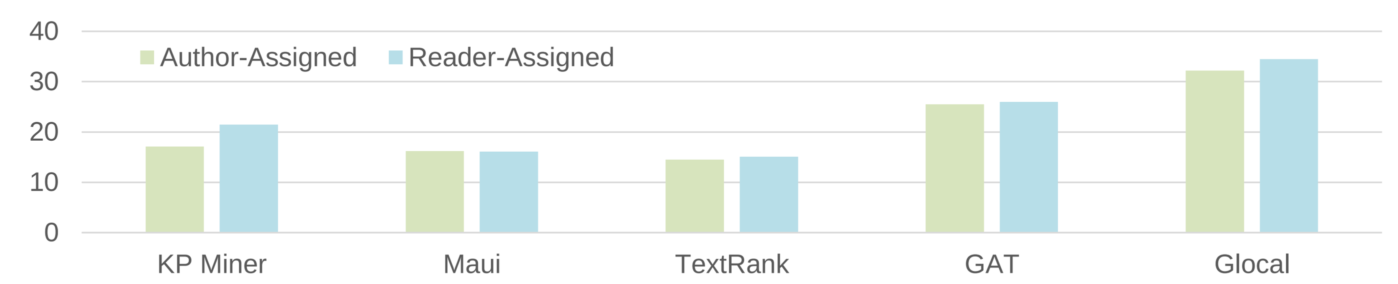
➢ **Results**: Performance on 3 full-text datasets (†abstract only)







Performance Breakdown on the SemEval Dataset



❖ Discussion

➢ **Feature versus Scaling:**
5% gain on average by incorporating as scaling as compared to up to 2% gain as features.

➢ **Graph Node Embeddings:**
TextRank-scaled averaged node neighborhood embedding:

$$\vec{h}_i^{(l+1)} \leftarrow \sigma\left(\sum_{j \in \mathcal{N}_i} \beta j \alpha_{ij} \vec{h}_j^{(l)} \mathbf{W}\right)$$

➢ **Longer, more Meaningful Keyphrases:**
Up to 4% higher ratio of keyphrases of length 3+ and 3% higher keyphrases of length 2.

➢ **Future Directions:**
1) Generalisation of Glocal for text classification, and its use in other extraction/ranking tasks such as summarization.
2) End-to-end graph-based ranking models

References:
[1]: Rada Mihalcea and Paul Tarau. TextRank: Bringing order into text. In *Proceedings of Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2004.
[2]:Rui Meng, Sanqiang Zhao, Shuguang Han, Daqing He, Peter Brusilovsky, and Yu Chi. Deep keyphrase generation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 582–592. Association for Computational Linguistics, 2017.
[3]: Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *Proceedings of International Conference for Learning Representations*, 2016.
[4]: Petar Velicˇkovic ́, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *Proceedings of International Conference on Learning Representations*, 2018.