Text Generation from Knowledge Graphs with Graph Transformers

NAACL19

Rik Koncel-Kedziorski , Dhanush Bekal , Yi Luan , Mirella Lapata , and Hannaneh Hajishirzi

University of Washington

University of Edinburgh

Allen Institute for Artificial Intelligence

https://www.youtube.com/watch?v=BiRyvB2NmCM

Reporter：Xiachong Feng

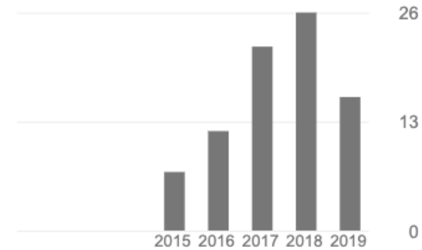
Outline

- Author
- Motivation
- Task
- Dataset
- Model
- Experiments
- Conclusion

Author

Cited by

	All	Since 2014
Citations	84	83
h-index	5	5
i10-index	2	2



- Rik Koncel-Kedziorski
- Lives on a sailboat
- University of Washington Ph.D. Winter 2019

<u>**Selected Publications**</u>

Rik Koncel-Kedziorski, Dhanush Bekal, Yi Luan, Mirella Lapata, and Hannaneh Hajishirzi.
Text Generation from Knowledge Graphs. Under Review

Sachin Metha, Rik Koncel-Kedziorski, Mohammad Rastegari, and Hannaneh Hajishirzi.
Pyramidal Recurrent Units for Language Modeling. EMNLP 2018

Rik Koncel-Kedziorski, Ioannis Konstas, Luke Zettlemoyer, and Hannaneh Hajishirzi.
A Theme-Rewriting Approach for Generating Math Word Problems. EMNLP 2016

Aaron Jaech, Rik Koncel-Kedziorski, and Mari Ostendorf.
Phonological Pun-derstanding. NAACL 2016

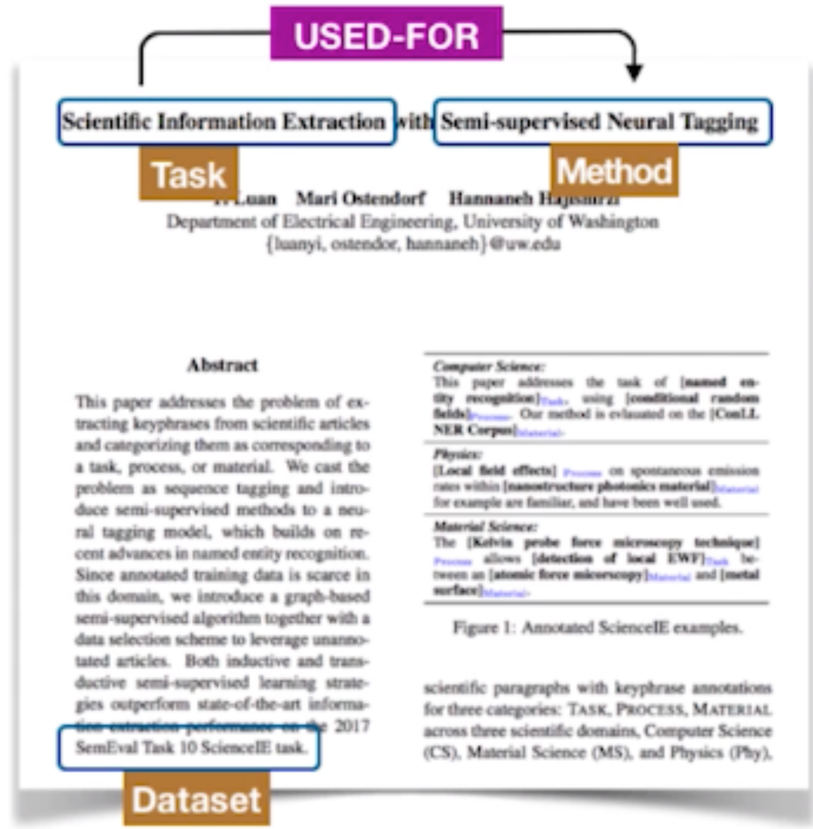
Rik Koncel-Kedziorski, Subhro Roy, Aida Amini, Nate Kushman, and Hannaneh Hajishirzi.
MAWPS: A Math Word Problem Repository. NAACL 2016

Rik Koncel-Kedziorski, Hannaneh Hajishirzi, Ashish Sabharwal, Oren Etzioni, and Siena Dumas Ang.
Parsing Algebraic Word Problems into Equations. TACL 2015.

R. Koncel-Kedziorski, Hannaneh Hajishirzi, and Ali Farhadi. 2014.
Multi-Resolution Language Grounding with weak supervision. EMNLP 2014.

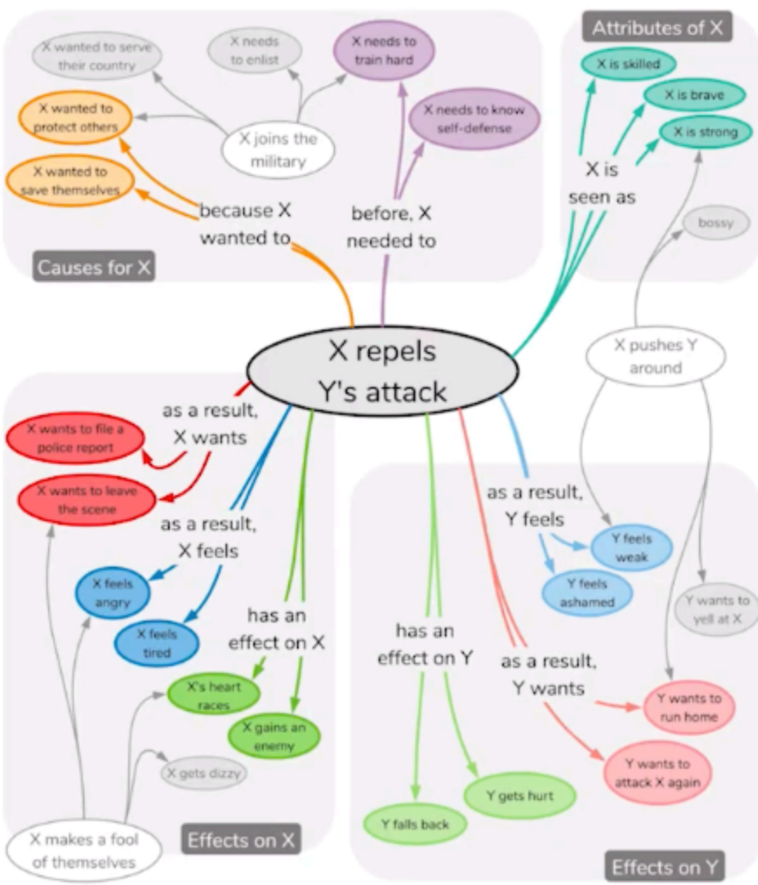
Knowledge

World Events

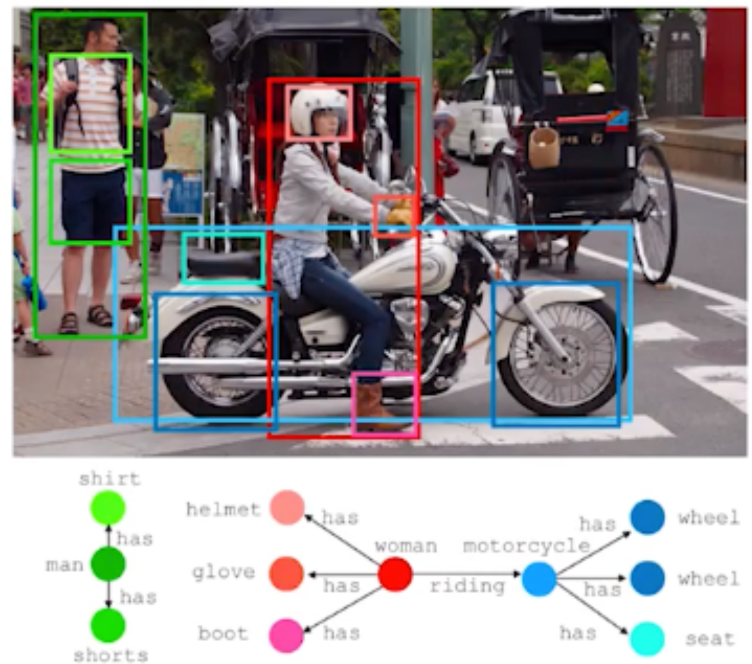


Science

Knowledge



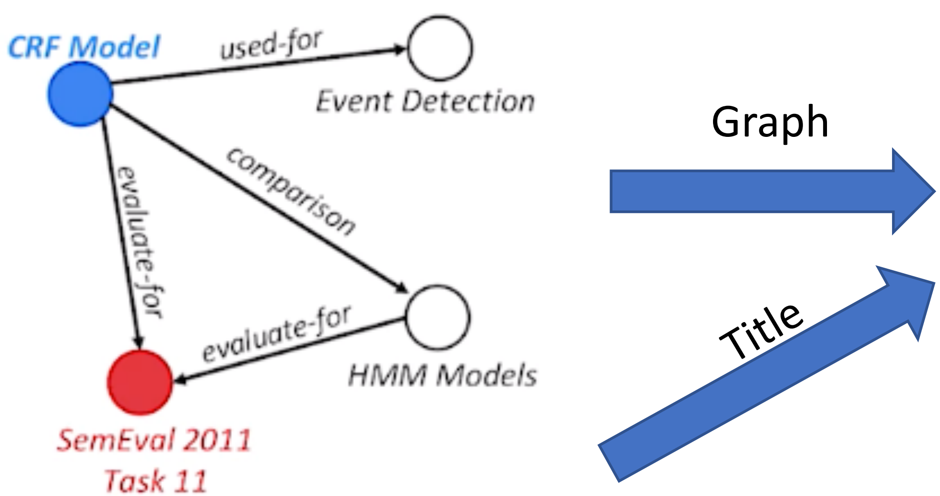
Common Sense



Multi-media

Task

- **Input**
  - **Title** of a scientific article;
  - **Knowledge graph** constructed by an automatic information extraction system;
- **Output**
  - **Abstract** (text);



Abstract

We present a CRF model for Event Detection tasks. Our model utilizes such and such features and can outperform standard HMM models by 110% on SemEval Task 11 Dataset. ...

Title: Event Detection with Conditional Random Fields

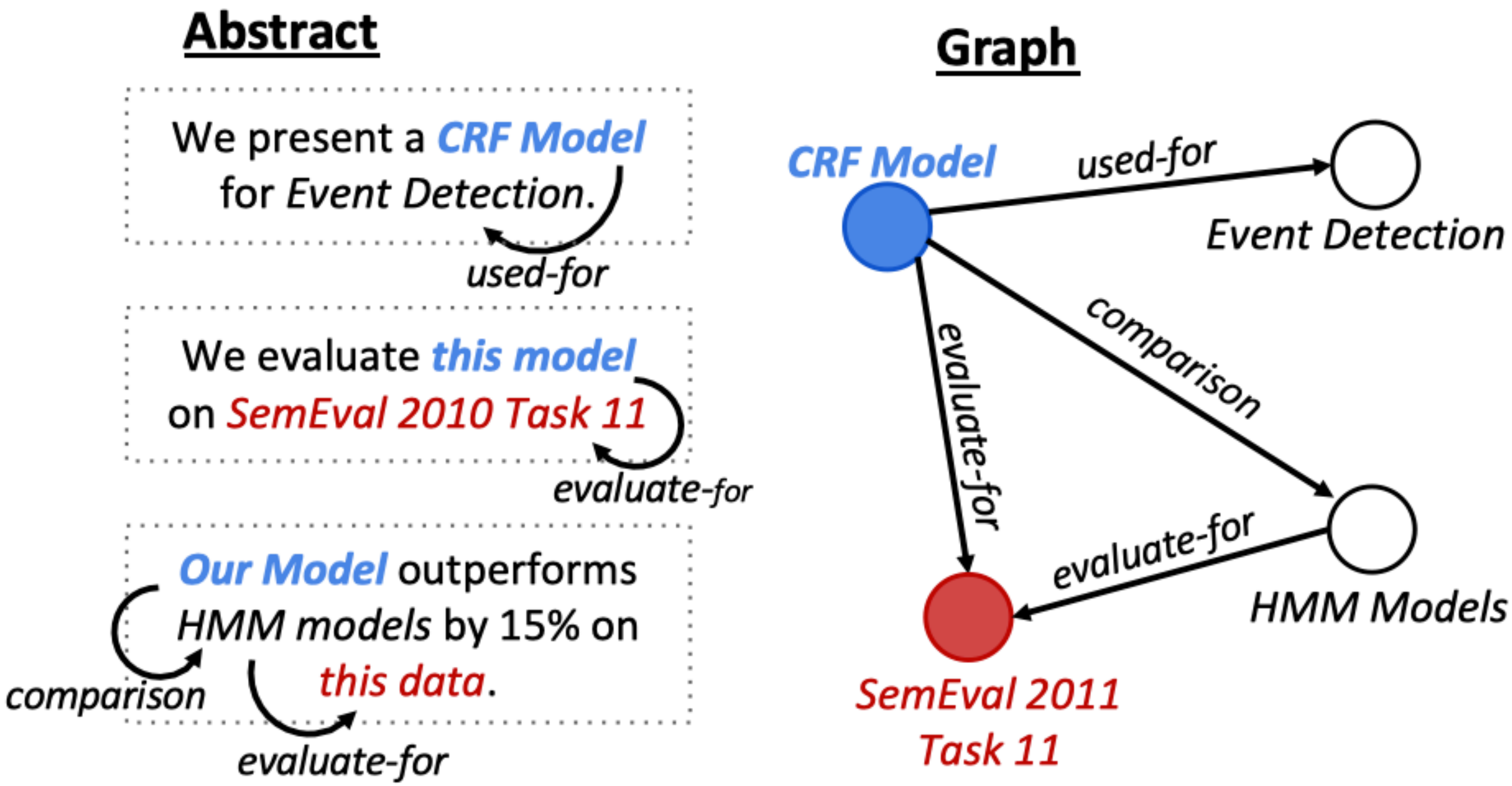
Dataset

- **A**bstract **GEN**eration **DA**taset (AGENDA) Dataset
- 12 top AI conferences
- **SciIE** system : a state-of-the-art science domain information extraction system.
  - NER、Co-Reference、Relations

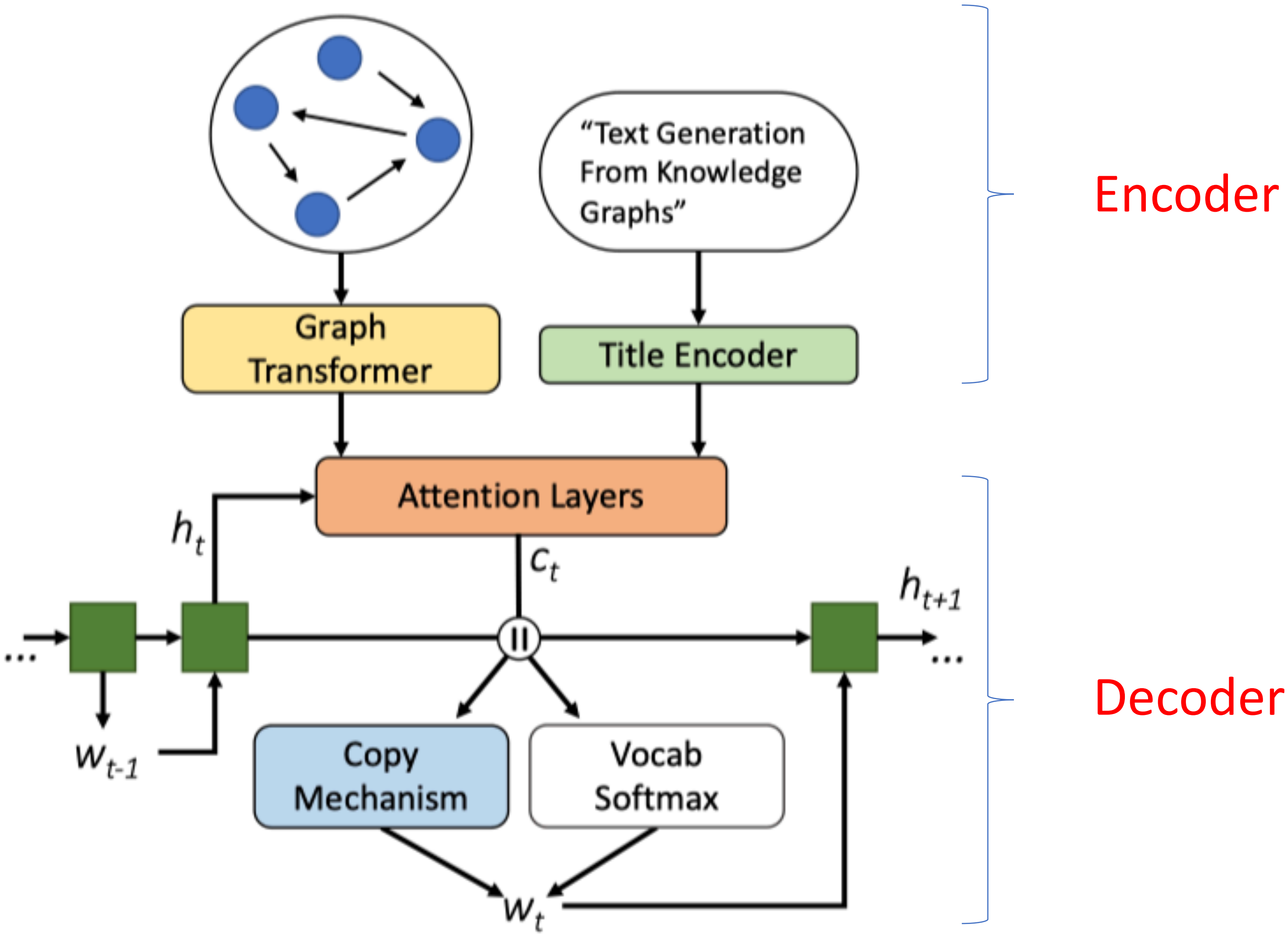
	Title	Abstract	KG
Vocab	29K	77K	54K
Tokens	413K	5.8M	1.2M
Entities	-	-	518K
Avg Length	9.9	141.2	-
Avg #Vertices	-	-	12.42
Avg #Edges	-	-	4.43

Dataset

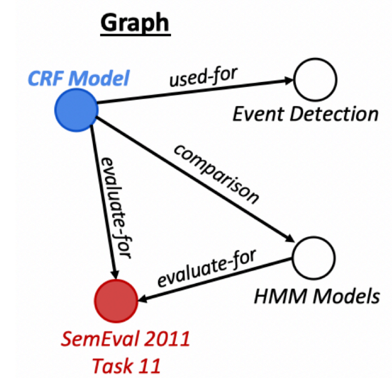
Title: Event Detection with Conditional Random Fields

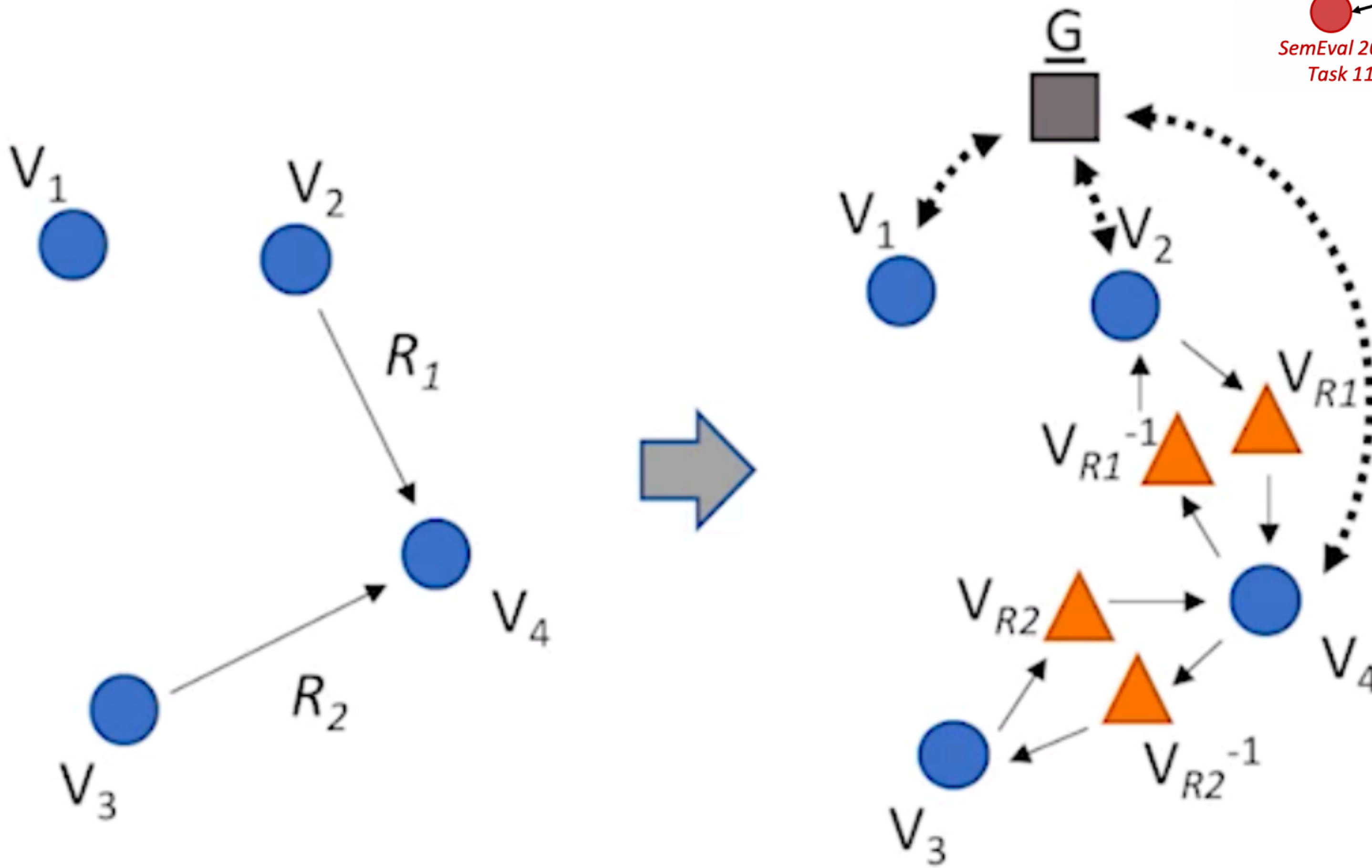


Model-GraphWriter



Graph Preparation





disconnected labeled graph

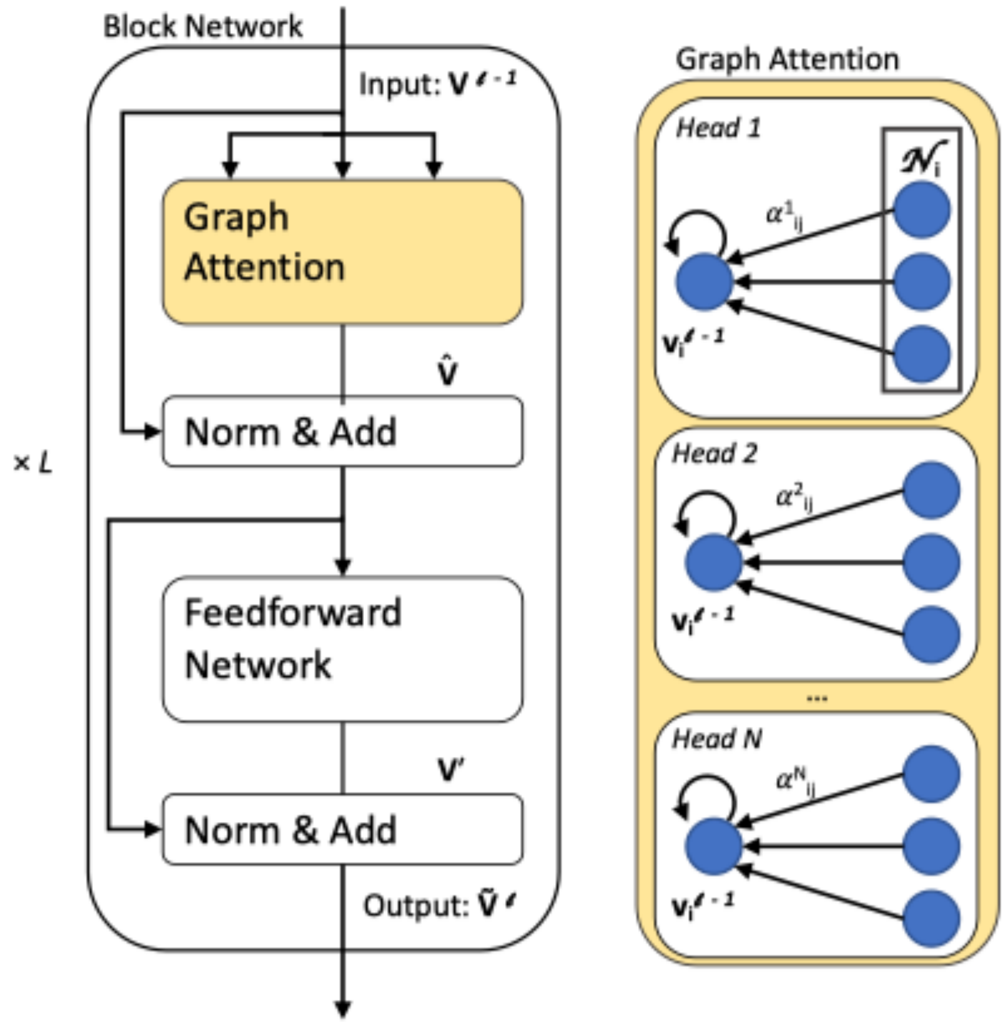
connected unlabeled graph

Embedding Vertices, Encoding Title

- **Relation**： forward- and backward-looking, two embeddings per relation
- **Entities** correspond to scientific terms which are often multi-word expressions.
- **Bidirectional RNN** run over embeddings of each word

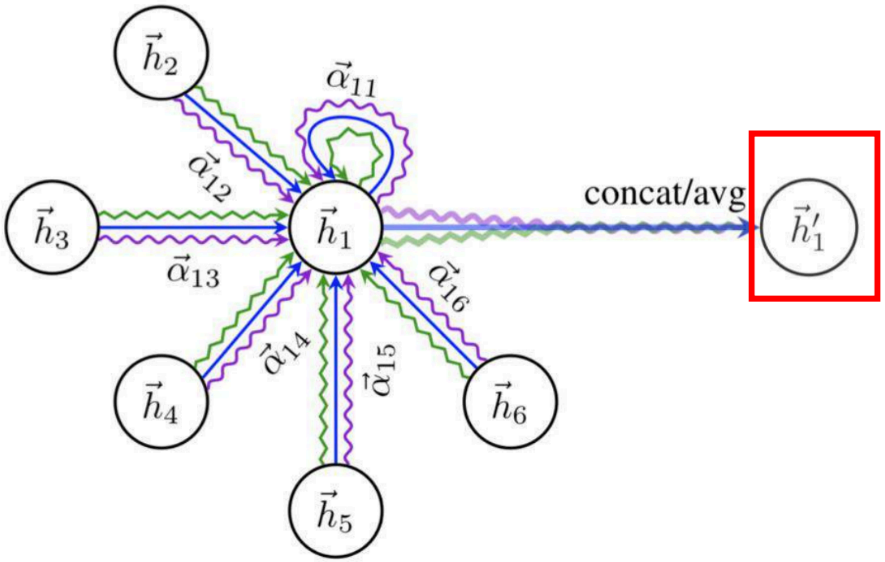
- The **title** input is also a short string, and so we encode it with another **BiRNN**

Graph Transformer



GAT

- Graph attention networks ICLR 2018 GAT

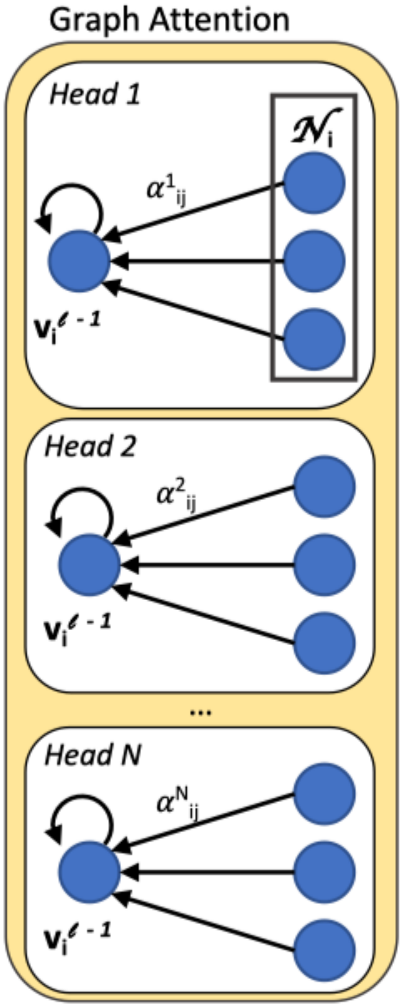


[Figure from Veličković et al. (ICLR 2018)]

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\vec{\mathbf{a}}^T[\mathbf{W}\vec{h}_i \| \mathbf{W}\vec{h}_j]\right)\right)}{\sum_{k\in\mathcal{N}_i}\exp\left(\text{LeakyReLU}\left(\vec{\mathbf{a}}^T[\mathbf{W}\vec{h}_i \| \mathbf{W}\vec{h}_k]\right)\right)}$$

$$\vec{h}'_i = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j\in\mathcal{N}_i}\alpha_{ij}^k \mathbf{W}^k \vec{h}_j\right)$$

Graph Attention



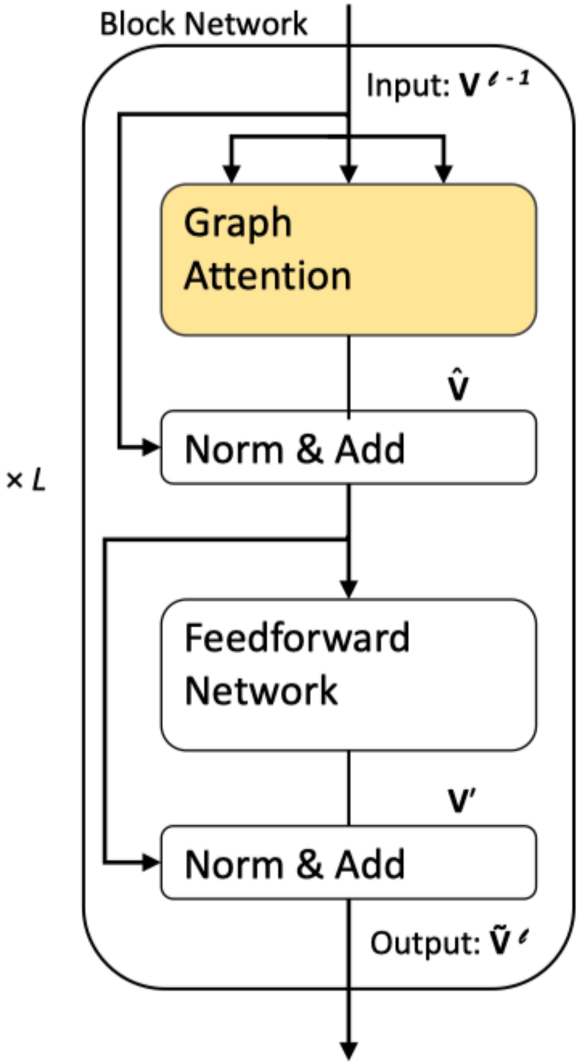
concat

$$\hat{\mathbf{v}}_i = \mathbf{v}_i + \Big\Vert_{n=1}^{N} \sum_{j \in \mathcal{N}_i} \alpha_{ij}^n \mathbf{W}_V^n \mathbf{v}_j$$

$$\alpha_{ij}^n = a^n(\mathbf{v}_i, \mathbf{v}_j)$$

$$a(\mathbf{q}_i, \mathbf{k}_j) = \frac{\exp((\mathbf{W}_K \mathbf{k}_j)^\top \mathbf{W}_Q \mathbf{q}_i)}{\sum_{z \in \mathcal{N}_i} \exp((\mathbf{W}_K \mathbf{k}_z)^\top \mathbf{W}_Q \mathbf{q}_i)}$$

Block networks

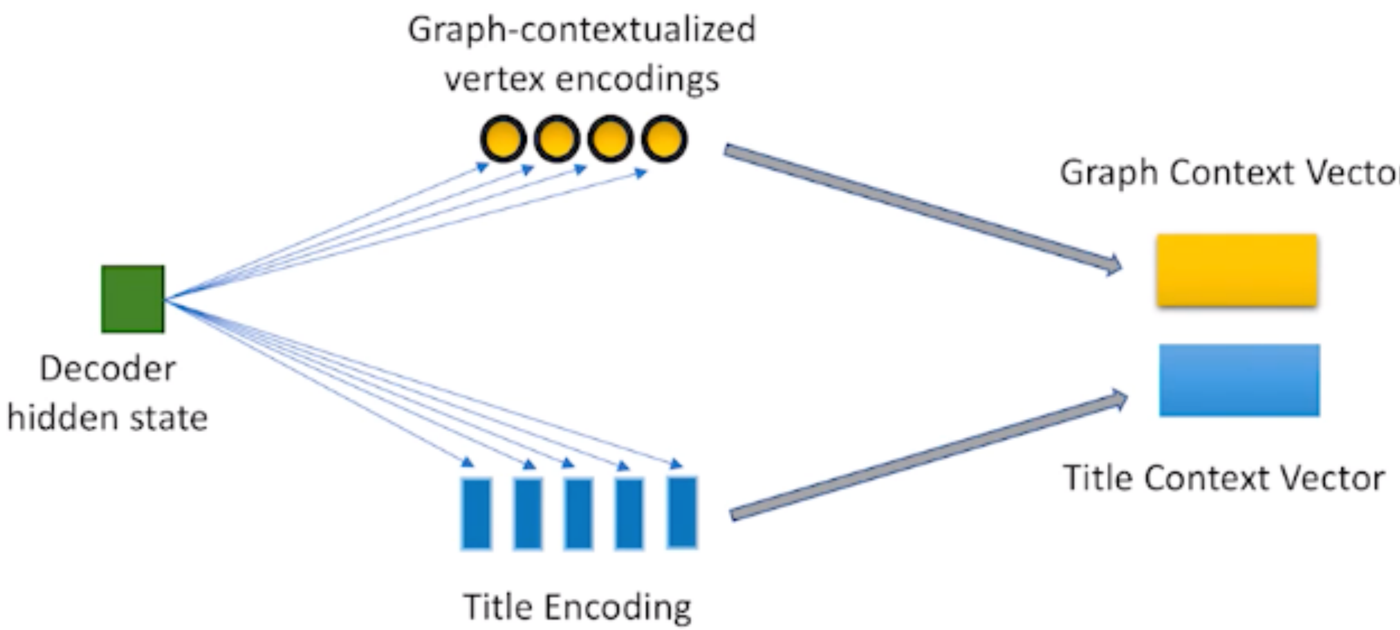


global contextualization

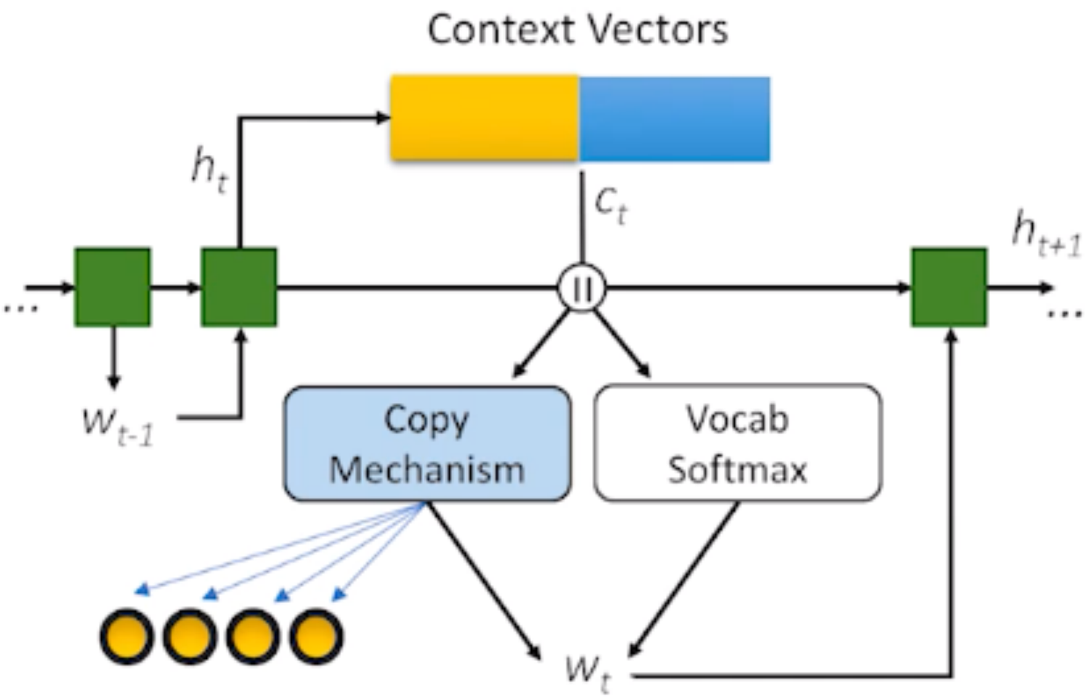
$$
\begin{aligned}
\tilde{\mathbf{v}}_i &= \mathrm{LayerNorm}(\mathbf{v}'_i + \mathrm{LayerNorm}(\hat{\mathbf{v}}_i)) \\
\mathbf{v}'_i &= \mathrm{FFN}(\mathrm{LayerNorm}(\hat{\mathbf{v}}_i))
\end{aligned}
$$

Decoder

- At each decoding timestep t we use decoder hidden state ht to compute context vectors cg and cs for the graph and title sequence



Copy



$$\mathbf{c}_t = [\mathbf{c}_g \| \mathbf{c}_s]$$

$$p = \sigma(\mathbf{W}_{copy}[\mathbf{h}_t \| \mathbf{c}_t] + b_{copy})$$

$$p * \alpha^{copy} + (1-p) * \alpha^{vocab}$$

entities

Experiments

- Evaluation Metrics
- Human evaluation
  - Grammar
  - Fluency
  - Coherence
  - Informativeness
- Automatic metrics
  - BLEU
  - METEOR

Baselines

- **GAT** : PReLU activations stacked between 6 self-attention layers.
- **EntityWriter** : uses only entities and title (no graph)
- **Rewriter** : uses only the document title

	BLEU	METEOR
GraphWriter	**14.3** $\pm$ 1.01	**18.8** $\pm$ 0.28
GAT	12.2 $\pm$ 0.44	17.2 $\pm$ 0.63
EntityWriter	10.38	16.53
Rewriter	1.05	8.38

Does Knowledge Help?

	Best	Worst
Rewriter (No knowledge)	12%	64%
GraphWriter (Knowledge)	24%	36%
Human Authored	64%	0%

Table 3: Does knowledge improve generation? Human evaluations of best and worst abstract.

	Win	Lose	Tie
Structure	63%	17%	20%
Informativeness	43%	23%	33%
Grammar	63%	23%	13%
Overall	63%	17%	20%

Table 4: Human Judgments of GraphWriter and EntityWriter models.

Conclusion

- Propose a new graph transformer encoder that applies the successful sequence transformer to graph structured inputs.
- Provide a large dataset of knowledge graphs paired with scientific texts for further study.

Thanks!