PaperRobot: Incremental Draft Generation of Scientific Ideas

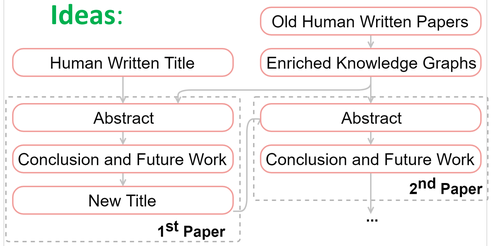
Qingyun Wang[1], Lifu Huang[1], Zhiying Jiang[1], Kevin Knight[2], Heng Ji[1,3], Mohit Bansal[4], Yi Luan[5]

[1] Rensselaer Polytechnic Institute, [2] DiDi Labs, [3] University of Illinois at Urbana-Champaign [4] University of North Carolina at Chapel Hill [5] University of Washington

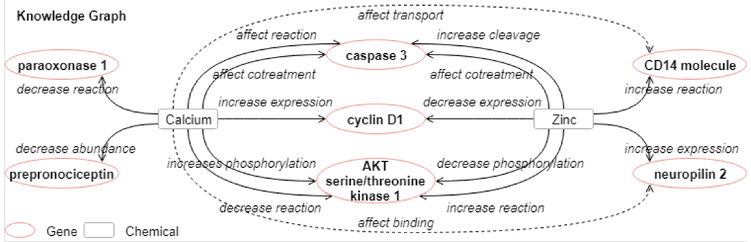
kevinknight@didiglobal.com, hengji@illinois.edu

➢ Motivation

- **Read Existing Papers**:
  - More than **500K** paper are published at PubMed every year.
  - **Build background knowledge bases**
- **Create New Ideas**:
  - More than **60%** of 6.4 million papers are incremental work.
  - **Predicts new links**
- **Write a New Paper about New Ideas**:



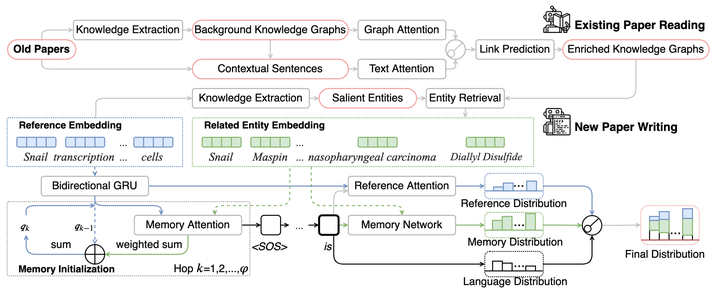
➢ Link Prediction



Contextual Sentence: So, Ca^{2+}possibly ***promoted*** **caspases** activation upstream of ***cytochrome c*** release, but inactivated **caspase** activity by calpain and/or fast depletion of ATP; whereas Zn^{2+} blocked the ***activation of procaspase-3*** with no visible change in the level of **cytochrome c**, and the block possibly resulted from its direct inhibition on **caspase-3** enzyme.

- **Graph structure Encoder**:
  - Capture importance of each neighbor's feature similar to graph attention network (Velickovic et al., 2018)
- **Contextual text encoder**:
  - Apply Bi-LSTM to contextual text
  - Use bilinear attention to get context representation
- **Gated combination:**
  - Use an entity-dependent gate

➢ Architecture Overview



- **Memory initialization:**
  - Filter irrelevant entities
  - Multihop attention draw better correlation
- **Memory network**:
  - Capture multihop attention of related entities
- **Reference attention:**
  - Capture soft attention of reference text

➢ Repetition removal

- **Coverage loss** (See et al., 2017)
  - Avoid any entity receiving attention multiple times
- New and simple **masking method**
  - Remove repetition during the test time

➢ Human Post-Editing

- It took the domain expert **40 minutes** to edit **50 abstracts**
  - Scores by comparing the abstracts before and after human editing

BLUE1	BLUE2	BLUE3	BLUE4	ROUGE	TER
59.6%	58.1%	56.7%	55.4%	73.3%	44.6%

➢ Example

	Human (Neal et al., 2012)	System
Title	***Snail*** *transcription factor negatively regulates* ***maspin tumor suppressor*** *in* ***human prostate cancer cells***	
Entities	Related: ***nasopharyngeal carcinoma****; diallyl disulfide*	
Abstract	*Background:* ***Maspin****, a putative* ***tumor suppressor*** *that is down-regulated in breast and* ***prostate cancer****, has been associated with decreased cell motility.* ***Snail*** *transcription factor is a zinc finger protein that is increased in breast cancer and is associated with increased tumor motility and invasion by induction of epithelial-mesenchymal transition (EMT).We investigated the molecular mechanisms by which* ***Snail*** *increases* ***tumor*** *motility and invasion utilizing* ***prostate cancer*** *cells. Methods: Expression levels were analyzed by* ***RT-PCR*** *and* ***western blot*** *analyses. Cell motility and invasion assays were performed , while* ***Snail*** *regulation and binding to* ***maspin*** *promoter was analyzed by luciferase reporter and chromatin immunoprecipitation (ChIP) assays. Results:* ***Snail*** *protein expression was higher in different* ***prostate cancer*** *cells lines as compared to normal prostate epithelial cells.*	*Background:* ***Snail*** *is a multifunctional protein that plays an important role in the pathogenesis of* ***prostate cancer****. However, it has been shown to be associated with poor prognosis. The purpose of this study was to investigate the effect of negatively on the expression of* ***maspin*** *in* ***human nasopharyngeal carcinoma*** *cell lines. Methods: Quantitative real-time PCR and western blot analysis were used to determine whether the demethylating agent was investigated by quantitative* ***RT-PCR*** *(qRT-PCR) and* ***Western blotting****. Results showed that the binding protein plays a significant role in the regulation of* ***tumor*** *growth and progression.*
Conclusion and Future work	*Collectively, our results indicate for the first time that* ***Snail*** *can negatively regulate* ***maspin*** *through direct promoter repression resulting in increased migration and invasion in* ***prostate cancer*** *cells. This study reveals a novel mechanism of how Snail may function and show the importance of therapeutic targeting of* ***Snail*** *signaling in future.*	*In summary, our study demonstrates that* ***Snail*** *negatively inhibited the expression of* ***Maspin*** *in* ***human nasopharyngeal carcinoma*** *cell lines and in vitro. Our results indicate that the combination of the demethylating agent might be a potential therapeutic target for the treatment of* ***prostate cancer****.*
New Title	*Role of* ***maspin*** *in cancer (Berardi et al., 2013)*	*The role of* ***nasopharyngeal carcinoma*** *in the rat model of* ***prostate cancer*** *cells*

➢ Automatic Evaluation

Model	Title-to-Abstract		Abstract-to-Conclusion and Future Work		Conclusion and Future Work-to-Title	
	Perplexity	METEOR	Perplexity	METEOR	Perplexity	METEOR
Seq2seq	19.6	9.1%	44.4	8.6%	49.7	6.0%
Editing Network	18.8	9.2%	30.5	8.7%	55.7	5.5%
Pointer Network	146.7	8.5%	74.0	8.1%	47.1	6.6%
Our Method (-Repetition Removal)	13.4	12.4%	24.9	**12.3**%	31.8	7.4%
Our Method	**11.5**	**13.0**%	**18.3**	11.2%	**14.8**	**8.9**%

➢ Turing Test

Task	Input		Output	Domain Expert	Non-expert
End-to-End	Human Title	Different	Abstract (1st)	10%	30%
		Same		30%	16%
	System Abstract	Different	Conclusion and Future work	12%	0%
		Same		8%	8%
	System Conclusion and Future work	Different	Title	12%	2%
		Same		12%	25%
	System Title	Different	Abstract (2nd)	14%	4%
Diagnostic	Human Abstract	Different	Conclusion and Future work	12%	14%
		Same		24%	20%
	Human Conclusion and Future work	Different	Title	8%	12%
		Same		2%	10%

Code and Dataset: https://github.com/EagleW/PaperRobot

➢ Remaining Challenges

- **Human output is more vivid**
  - **Human:** *Does HPV play any role in the initiation or prognosis of endometrial adenocarcinomas?*
  - **System:** *The role of HPV in the treatment of endometrial adenocarcinomas*
- **Human output is more concrete**
  - **Human:** *etumorType , An Algorithm of Discriminating Cancer Types for Circulating Tumor Cells or Cell-free DNAs in Blood*
  - **System:** *Gastrointestinal Stromal tumor initiation: A Review of the Literature*
- **System generates a few incorrect abbreviations**
  - **E.g.** *"chronic kidney disease(UC)"*
- **System generates some incorrect numbers and pronouns**
  - **E.g.** " **Group 1 , Group B**", "A 63-year-old **man** was referred to our hospital ... **she** was treated with the use of ..."