Scientific Discourse Tagging for Evidence Extraction

Xiangci Li, Gully Burns & Nanyun Peng
University of Texas at Dallas, Chan Zuckerburg Initiative, University of California Los Angeles
USC University of Southern California
Information Sciences Institute

Introduction

- Papers that describe original experiment work provide the crucial raw material for subsequent scientific research.
- Automatic information extraction (IE) from biomedical literature is a crucial step to help researchers use the growing amount of literature efficiently.
- Traditional IE methods focus on entities, relations and events, but do not take into account the evidence derived from experiments that the authors use to sell their claims.
- Thus, it's important to recognize rhetorical components of scientific discourse, which distinguish observations vs. implications, or claims vs. hypotheses.
- Scientific discourse tagging is a task that tags "sentences" in a scientific article with different rhetorical components of scientific discourses.

lc3 , the mammalian atg8 homolog , undergoes a set of modifications resulting in conversion from lc3i to lc3ii during autophagy 42 . **[fact]** to further test the function of rag in autophagy **[goal]** we examined the lc3 modification in hek293 cells . **[method]** expression of raga ql and ragc sn inhibited lc3 conversion in response to amino acid starvation (fig. 7e) . **[result]** furthermore , expression of raga tn and ragc ql enhanced lc3 conversion even in the presence of amino acids . **[result]** these results are consistent with the data observed in drosophila and further demonstrate a role of the rag gtpases in autophagy regulation in response to nutrient signals **[implication]**

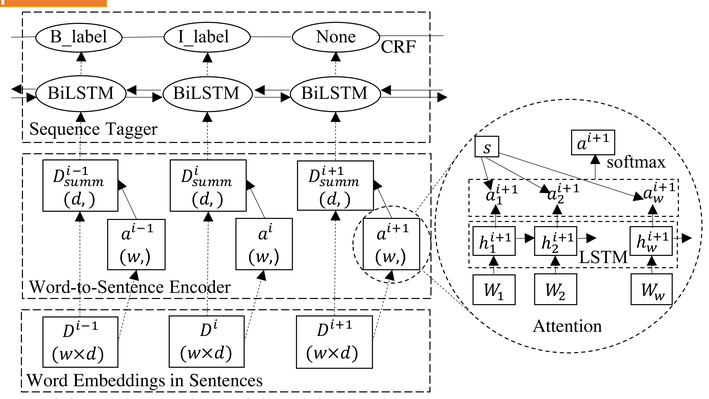
An example of a paragraph tagged with discourse types.

Type	Definition
Goal	Research goal
Fact	A known fact, a statement taken to be true by the author
Result	The outcome of an experiment
Hypothesis	A claim proposed by the author
Method	Experimental method
Problem	An unresolved or contradictory issue
Implication	An interpretation of the results
None	Anything else

Eight label taxonomy defined by De Waard and Maat (2012) and Dasigi et al. (2017).

Scientific Discourse Tagging

Approach



Sequence tagging problem
- Each example is a paragraph
- Tag each "sentence"

Model
- SciBERT embedding
- LSTM-Attention as word-to-sentence encoder
- Sentence-level BiLSTM-CRF sequence tagger

Datasets

SciDT
- Extended by us
- Derived from Pathway Logic and INTACT databases
- Small (634 paragraphs)

PubMed-RCT
- Large (20k abstracts)
- Labels natively come with abstracts

CODA-19
- 10k abstracts
- Human annotated on the abstracts of CORD-19

SciDT	
result	2053
method	1542
implication	797
fact	732
hypothesis	514
goal	225
problem	193
none	68

CODA-19	
finding / contribution	50018
background	46082
method	16216
purpose	8252
other	2486

PubMed-RCT	
methods	79214
results	77507
conclusions	36321
background	28797
objective	18548

Scientific Discourse Tagging

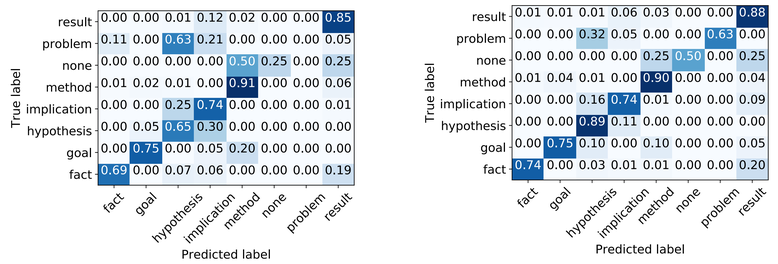
Experimental Results

Model		RCT	SciDT
CRF			0.679
SVM			0.737
Dasigi et al. (2017)			0.791
HSLN-CNN		0.922	
HSLN-RNN		0.926	
Srivastava et al. (2019)		0.928	
Embedding	**Attention**		
BioGloVe	No Context	0.901	0.745
BioGloVe	RNN	0.909	0.763
BioGloVe	LSTM	0.913	0.794
BioBERT	No Context	0.909	0.794
BioBERT	RNN	0.915	0.775
BioBERT	LSTM	0.927	0.794
SciBERT	No Context	0.918	0.806
SciBERT	RNN	0.922	0.817
SciBERT	**LSTM**	**0.951**	**0.841**

Scientific discourse tagging performance measured by test F1 score on PubMed 20k RCT and SciDT dataset.

Model	Test F1
Huang et al. (2020)	0.749
Ours	**0.885**
Zero-shot Prediction from RCT	0.760
Zero-shot Prediction from SciDT	**0.761**
PubMed-RCT pre-train	**0.909**

Strong transfer learning performance on CODA-19 dataset.



Confusion matrix on SciDT test data. Left: Dasigi et al. (2017). Right: Our scientific discourse tagger.

Claim Extraction

Dataset

- Binary sequence tagging problem
- Dataset
  - 1500 abstracts from MEDLINE database
  - Human-annotated

Experiments

Baseline Model
- Average pooling of the GloVe embedding
- BiLSTM-CRF model
- Pretrained on PubMed 200k RCT

We directly used our scientific discourse tagger.

Conclusion

Leveraging scientific discourse tagging can improve the performance

Panax notoginseng saponins (PNS) are components derived from Chinese herb panax notoginseng and play important roles in the cure of wounds. However, how PNS plays this function is still unclear. In this study, we used MTT assay, wound healing assay, western blot, quantitative real time PCR and enzyme-linked immunosorbent assay to detect the effects of PNS on the proliferation, migration and expression of collagen and fibronectin of anterior cruciate ligament (ACL) fibroblasts as well as the underlying mechanism. We found that PNS promoted the proliferation and migration of ACL fibroblasts and increased the expression levels of collagen and fibronectin. **Further mechanism study indicates that PNS might play its function through the phosphorylation of PI3K, AKT and ERK. This study provides a possible mechanism for the function of PNS and lays foundation for further study on the function of panax notoginseng.**

An example abstract with claim sentences highlighted in claim-extraction dataset

Model	Test F1
Achakulvisut et al. (2019)	0.790
Ours (No pre-train)	0.791
Ours (PubMed-RCT pre-train)	**0.828**

Claim extraction performance measured by binary F1 score.

Evidence Fragment Detection

Task

Associate experimental figures and the descriptions in the text (evidence fragment).

•Challenges
- The labels are not determined, and the number of labels vary
- Some clause semantically refer to sub-figures, but only implicitly.
- Very small dataset
  - 210 paragraphs

To ensure equal loading of protein in all lanes **[goal, O]** …… cell supernatants from uninfected and … **[method, O]** As seen in Figure 1C , SHIP-/- **[result, B]** BMMs produced significantly elevated levels of … **[result, I]** Similar results were obtained with peritoneal **[result, I]** that SHIP-deficient macrophages make more … **[implication, I]** — Fragment for Fig. 1C

Protein-matched cell lysates from SHIP+/+ … **[method, B]** BMMs were analyzed by … (Figure 1D) . **[goal, I]** These data indicate **[implication, I]** that SHIP is a negative regulator of IL-12 , IL-6 , … **[implication, I]** — Fragment for Fig. 1D

The regulatory influence of … (MOIs ; Figure 2A-2C) , **[result, B]** suggesting that the negative regulatory effect of … **[implication, I]** — Fragment for Fig. 2A, 2B, 2C

In addition , the viability of the bacteria is not … **[implication, B]** as macrophages infected with … (Figure 2D-2F) . **[result, I]** Likewise , SHIP also down-regulated **[result, I]** F. tularensis LVS-induced … (Figure 2D-2F) . **[result, I]** — Fragment for Fig. 2D, 2F

Approach

Block-based assumption
- Clauses in a block share the same sub-figure codes.
- Each sub-figure code is explicitly mentioned at least once within each block.
- Only need to determine the beginning and end of each block, by tagging BIO tags.

Experimental Results

Model	BIO F1	Test F1
Burns et al. (2017)	N/A	0.75
Ours (W/O Discourse Tags)	0.750	0.742
Ours (W/ Discourse Tags)	**0.821**	**0.807**

Evidence fragment detection performance measured by micro F1 score. Our block-based decoding method achieves 0.94 F1 using ground truth BIO sequences.

Baseline model
- Rule-based tagger

Ours model
- Feature-based CRF
- Take discourse tags as features

Conclusion

Discourse tags drastically improves the tagging performance.

Discussion

•Scientific discourse tagging facilitates extracting "evidence fragments"
•"Evidence fragment" as independent documents can be cataloged, indexed and reused.
•"Evidence fragment" as a part of the evidence along with the figure and figure captions can be easily reused by *other papers* as a "Micropublication".
•Ultimately, this dramatically increases the amount of primary evidence used to generate individual claims and therefore improves the quality of those claims.

Summary

•Scientific discourse tagging
- Novel tagger
- SOTA result
- Strong transferability to other datasets

•Benefit of scientific discourse tagging on downstream-tasks
- Claim-extraction
- Evidence fragment detection

•Evidence fragment for higher quality claims