Beyond tasks, methods, and metrics: extracting metrics-driven mechanism from the abstracts of AI articles

Yongqiang Ma(presenter), Jiawei Liu,

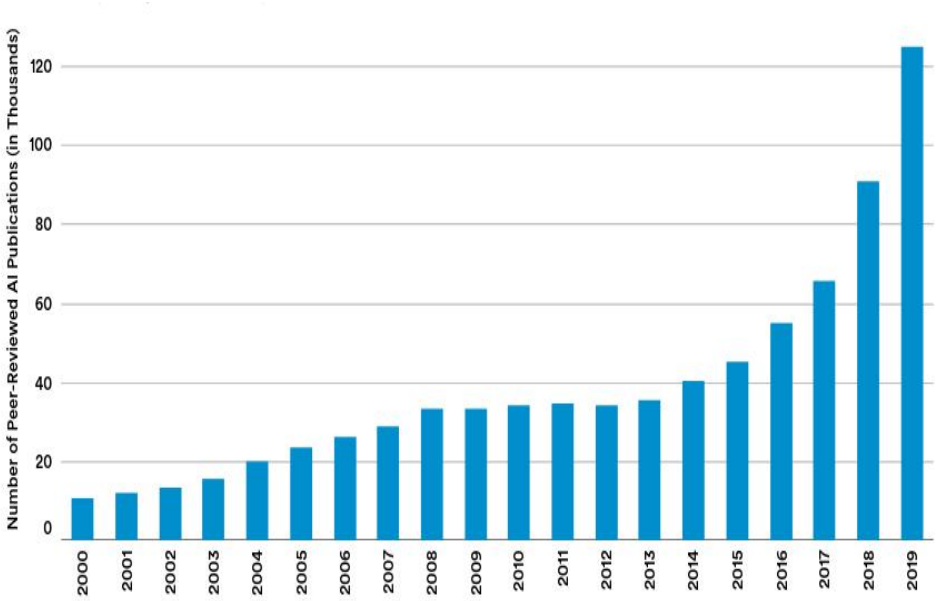
Wei Lu, Qikai Cheng

Wuhan University, China

3rd Workshop on Extraction and Evaluation of Knowledge Entities from Scientific Documents
EEKE2022 @ JCDL2022

Introduction

- In artificial intelligence (AI), peer-reviewed scientific publications has grown by nearly **12 times** in the past 20 years.
- The papers number read by scientists per year has **reached a peak (264 papers per year)**.
- Automated extracting key insight from scientific publications is required.



NUMBER of PEER-REVIEWED AI PUBLICATIONS

Zhang, D. , Mishra, S. , Brynjolfsson, E. , Etchemendy, J. , Ganguli, D. , & Grosz, B. , et al. (2021). The ai index 2021 annual report.
Noorden, R. V. . (2014). Scientists may be reaching a peak in reading habits. Nature News.

Literature Review

- Scientific information extraction: Extracting key insight from scientific publications .

- 3 disciplines in scientific information extraction:
  Computer Science,Biomedicine , and Chemistry.

- Named entity recognition (NER): Extraction of entities, i.e., entity mentions and their types.

  'we apply multi-task learning $_{Method}$ *to text classification* $_{Task}$*'*

- Relation extraction (RE): Extraction of the relation between entity pairs.

  Relation(multi-task learning, text classification) = Used-for

Literature Review

- **Tasks, Methods, Datasets and Metrics** are the primary domain-specific entity types in AI community.
  - SCIERC dataset, SCIREX dataset
- Isolated tasks, methods, datasets, metrics and shallow relations between entities only reflect the **descriptive information**.
  - Gábor et al. proposed the *usage, result, part-whole, and compare* relations.
  - Mondal et al. proposed the *evaluated-On* and *evaluated By* relations.

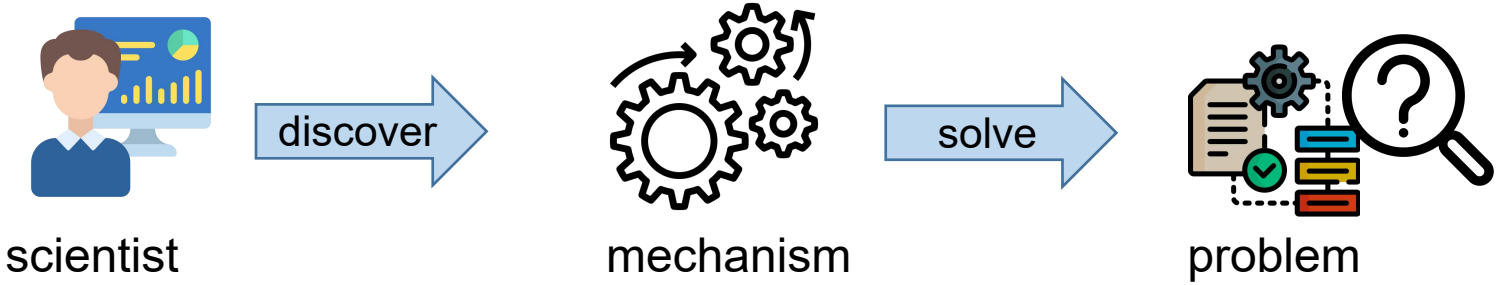
What is the research task in paper X?

What is the research method in paper X?

What datasets are used in paper X?

Application scenarios of descriptive information are primary to 'What' questions.

Mechanism in Science



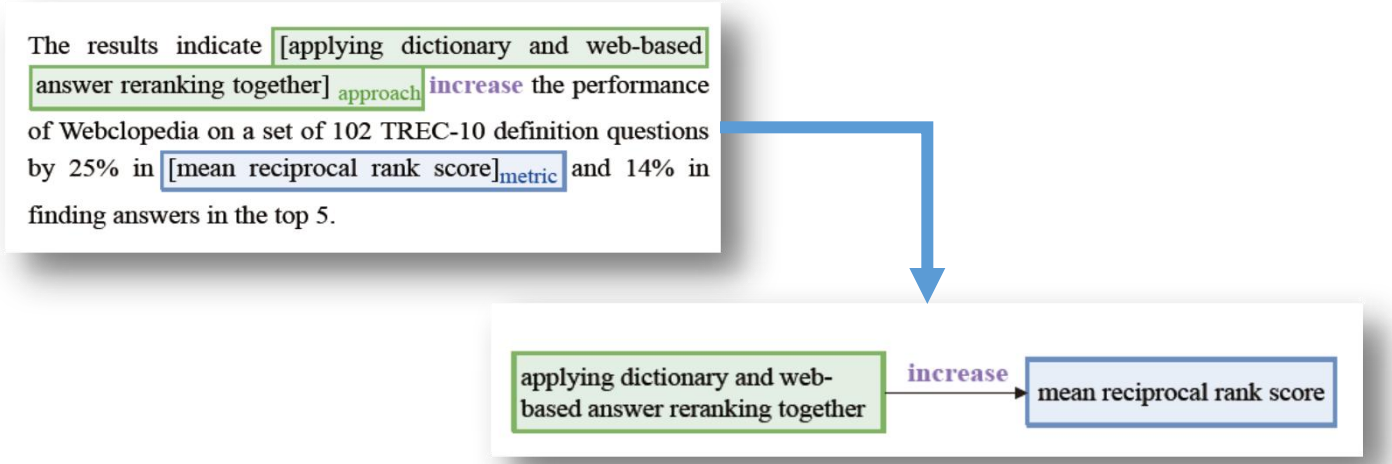
- The **object of AI research** is to discover the mechanism between problem and solution.
- **Metrics-driven mechanism** : procedural information for optimizing specific metrics.
- **Gap: Current entity and relation types are disable to cover the procedural information**

Research Question :

1. how to represent the metrics-driven mechanism in AI.
2. how to extract the metrics-driven mechanism in AI.

Metrics-driven mechanism Schema

- Two patterns : express the metrics-driven mechanism in the abstract of AI papers.
  - ➢ Direct description:
  - ➢ Indirect description:



- **Representation Schema**

metrics-driven mechanism triple: ***<operation ,effect, direction>***

- ➢ ***operation***: the entity that refers to the innovative model, method, or approach proposed in a paper.
- ➢ ***effect***: the entity that refers to the metrics evaluating operation's effectiveness or performance.
- ➢ ***direction***: the relation between the operation entity and the effect entity,
  direction ∈ { positive effect, negative effect, other}

Problem Definition

Input: Sentences S in papers abstract

Output: <*operation ,effect, direction*>, Salient tasks of research papers

Problem 1: Metrics-driven mechanism triple extraction:

subtask 1. Identify sentences S^+ containing metric-driven mechanism from the all sentences in abstract.

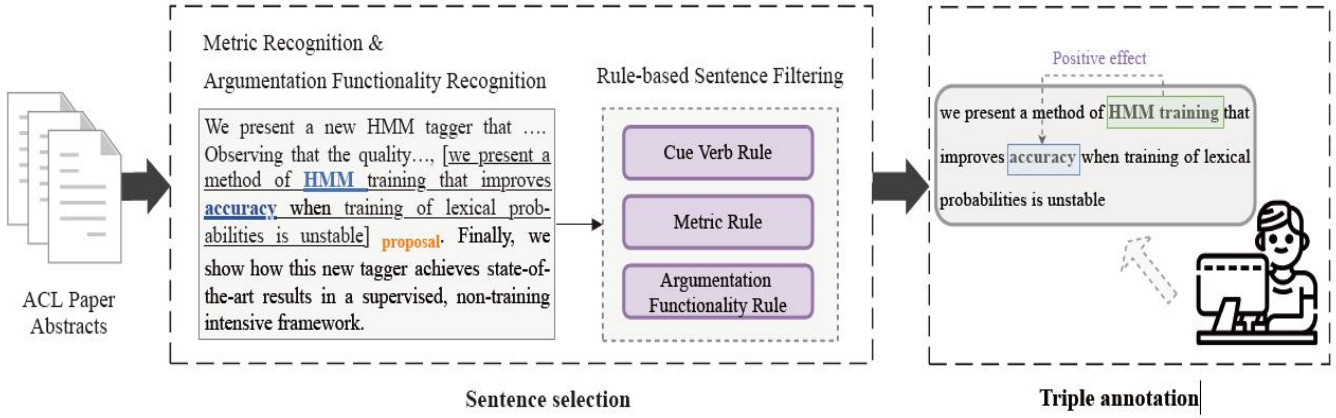
subtask 2. Extract metric-driven mechanism triple from S^+.

Problem 2: Salient tasks recognition. For an abstract D , map the D to a predefined salient tasks list.

Dataset

Metric-driven mechanism triple extraction:

- Sentence selection: three rules (Cue Verb Rule, Metric Rule and Argumentation Functionality Rule). Our rules prioritize the recall.
- Triple annotation: use the BRAT (https://brat.nlplab.org)



Salient tasks recognition:

We use the paper metadata released in Papers With Code, in which papers are assigned several tasks from the predefined taxonomy.

Dataset

Metrics-driven mechanism triple extraction:

- We focus on metrics-driven mechanism triples explicitly existing in the single sentence.

Statistics of dataset for subtask 1

Statistics items	Number
Total # of Sentences	4,163
# of Mechanism Sentences	1,032
# of Non-mechanism Sentences	3,131

Statistics of dataset for subtask 2

Statistics items	Number
Total # of Sentences	1,032
# of Operation Entities	1,214
# of Effect Entities	1,311
# of Relations	1,486

Salient tasks recognition:

- 2327 tasks in artificial intelligence
- 131k papers abstracts

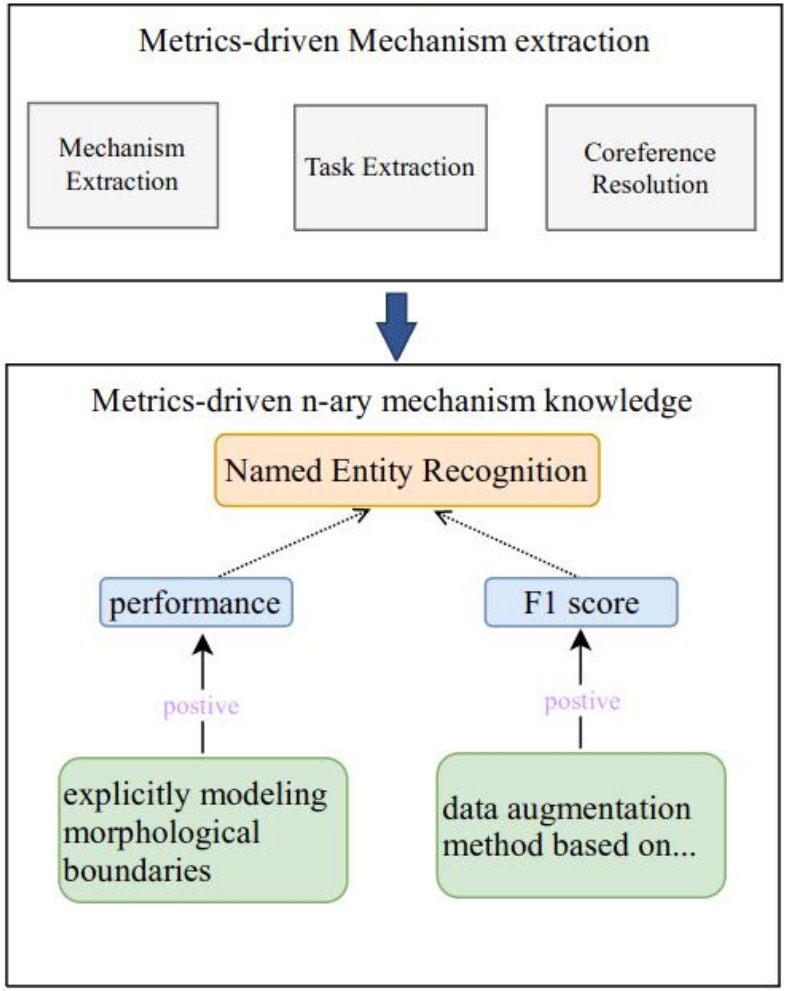
Pipeline Model Details

1. Identify sentences containing mechanism.
   Pretrained model : SciBERT

2. Extract the mechanism triples from the recognized sentences.
   Joint entity and relation extraction model: SpERT

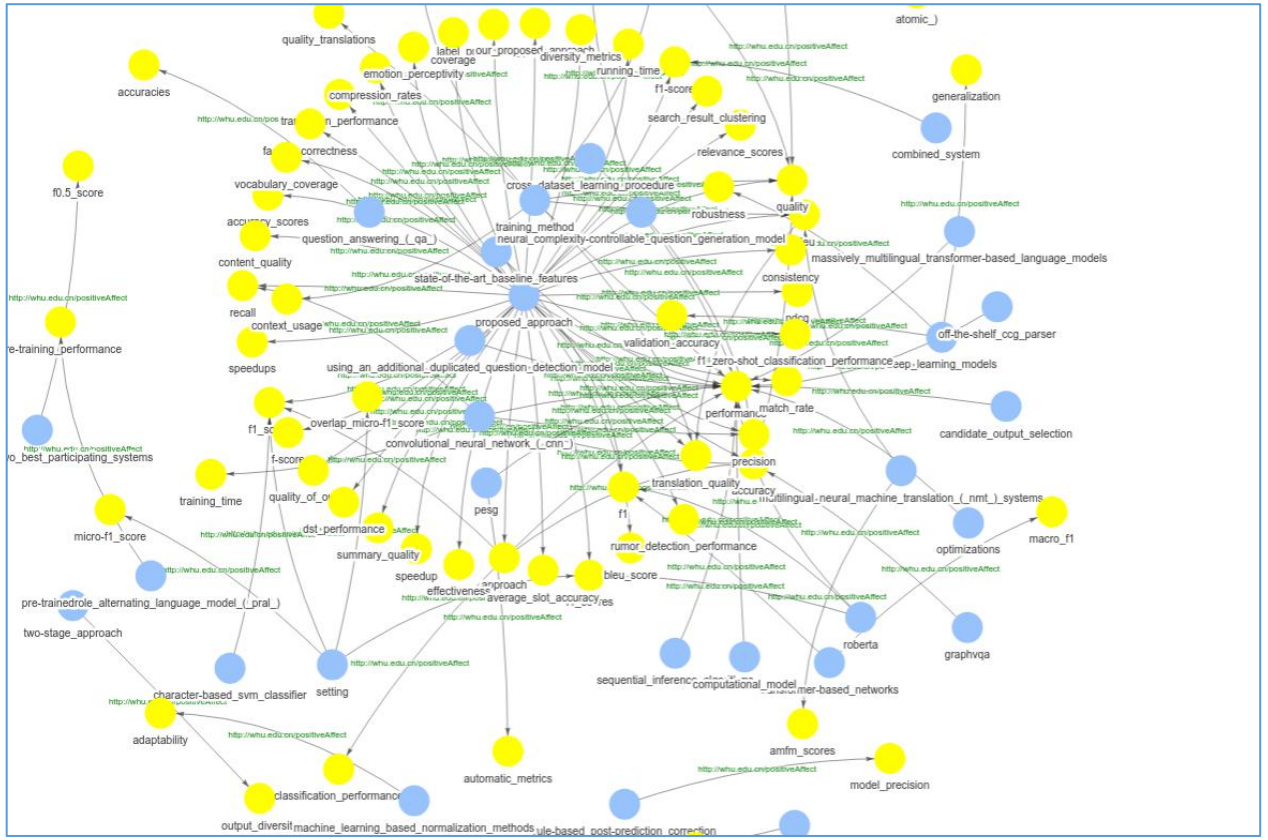
3. Task recognition problem is formalized as a multi-label classification task based on SciBERT.

4. Coreference resolution: Some *operation* entities are pronouns (e.g., “our model”, “proposed method”).
   AllenNLP coreference resolution model.



Application

- Source: 26k ACL paper abstracts *http://aclanthology.org*
- 24k <operation ,effect, direction,task> n-ary mechanism relations.
- For the abstract without n-ary mechanism relations, pseudo n-ary mechanism relations in the form <method, metric, direction, task> are extracted using SpERT .
- Based on the PwC hierarchical task taxonomy, our knowledge base supports automatic tasks semantic extension



NLPMKB

Result & analysis

- Mechanism sentence extraction model achieves an 89 F1 score on the testset.

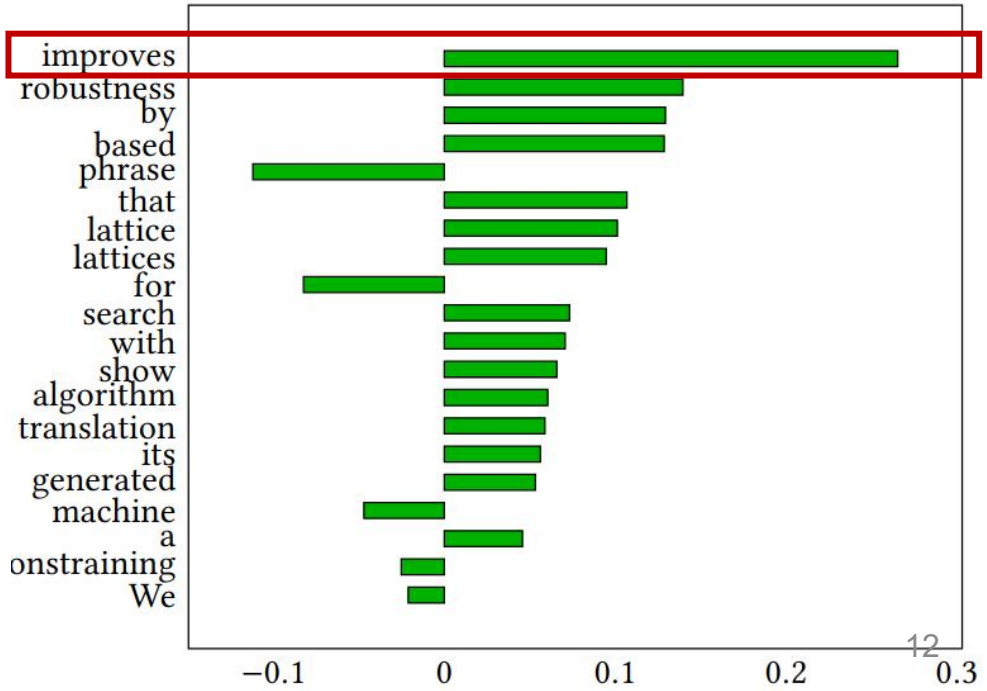
Type	Precision	Recall	F1 score
Non mechanism sent	92.5	93.0	92.7
Mechanism sent	77.8	76.7	77.2
Total	89.0	89.0	89.0

- Adopt the Local Interpretable Model-agnostic Explanations (LIME) : interpret the mechanism sentence extraction model.

- Task recognition model achieves an 87 F1-score.

LIME INPUT:
We present a stack-based lattice search algorithm for NMT and show that constraining its search space with lattices generated by phrase-based machine translation (PBMT) **improves** robustness.

LIME OUTPUT:



Result & analysis

- 78.67 F1 score on both Operation and Effect entity recognition
- 59.80 F1 score on direction recognition.

	Type	Precision	Recall	F1 score
Entities Extraction	Operation	72.0	66.4	69.1
	Effect	86.3	87.6	86.9
	Total	79.8	77.6	78.7
Relations Extraction	Pos_eff	59.7	71.7	65.1
	Neg_eff	50.0	54.2	52.0
	Other	60.0	26.1	36.4
	Total	58.1	61.6	59.8

- Randomly select 100 papers in 2020 COLING and check the extracted metrics-driven mechanism triples.
- An entity is regarded as positive if its type is correct and there is an overlap with the ground truth entity boundary.
- 34 papers that contain metrics-driven mechanism in 100 abstracts.

	Ground truth	
Predicted	Positive	Negative
Positive	27	4
Negative	7	62

Result & analysis

- Metrics-driven mechanism enhenced search engine.
- Test queries:
  - ✓ how to improve F1 on text classification?
  - ✓ how to improve model generalization?
  - ✓ how to decrease training time?
  - ✓ how to improve performance of Named Entity Recognition (NER)?
  - ✓ how to improve BLEU on machine translation?

Query	how to imporve BLEU score
Task	Machine translation

Search

Operation	Source sentence
context-aware ST	Compared to sentence-level ST , context-aware ST obtains better translation quality (+0.18-2.61 BLEU) ...
Transformer optimized model	A Transformer optimized model demonstrated a BLEU score improvement of 7.8 points when compared with a baseline RNN model .

No	Query	Our search engine		Baseline	
		P@3	P@5	P@3	P@5
1	how to improve F1 on text classification?	100	60	33	60
2	how to improve model generalization?	100	80	100	60
3	how to decrease training time?	100	100	100	100
4	how to improve performance of Named Entity Recognition (NER)?	67	80	33	40
5	how to improve BLEU on machine translation?	100	100	100	100
	Avg.	**93**	**84**	73	72

Conclusion

- Introduce a representation schema to express metrics-driven mechanisms.
- Build a dataset in the field of NLP with 1,486 metrics-driven mechanism triples.
- Propose a pipeline to extract metrics-driven mechanisms.
- Construct a metrics-driven mechanism knowledge base (KB) in the NLP domain.

Thank You!

Contact
Yongqiang Ma
Email: mayongqiang@whu.edu.cn