The University of Queensland Australia | Create Change

# Effective Representation Learning for Legal Case Retrieval

Yanran Tang

The University of Queensland

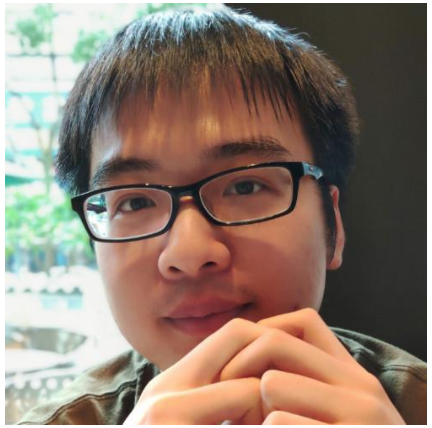
THUIR Seminar, May 28, 2024

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

# Team Members

**Yanran Tang**

EECS, UQ

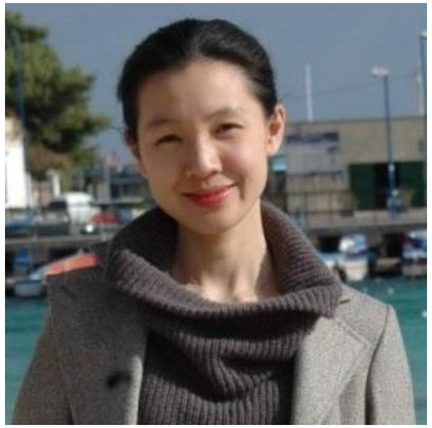


**Dr. Ruihong Qiu**

EECS, UQ

**Prof. Xue Li**

EECS, UQ



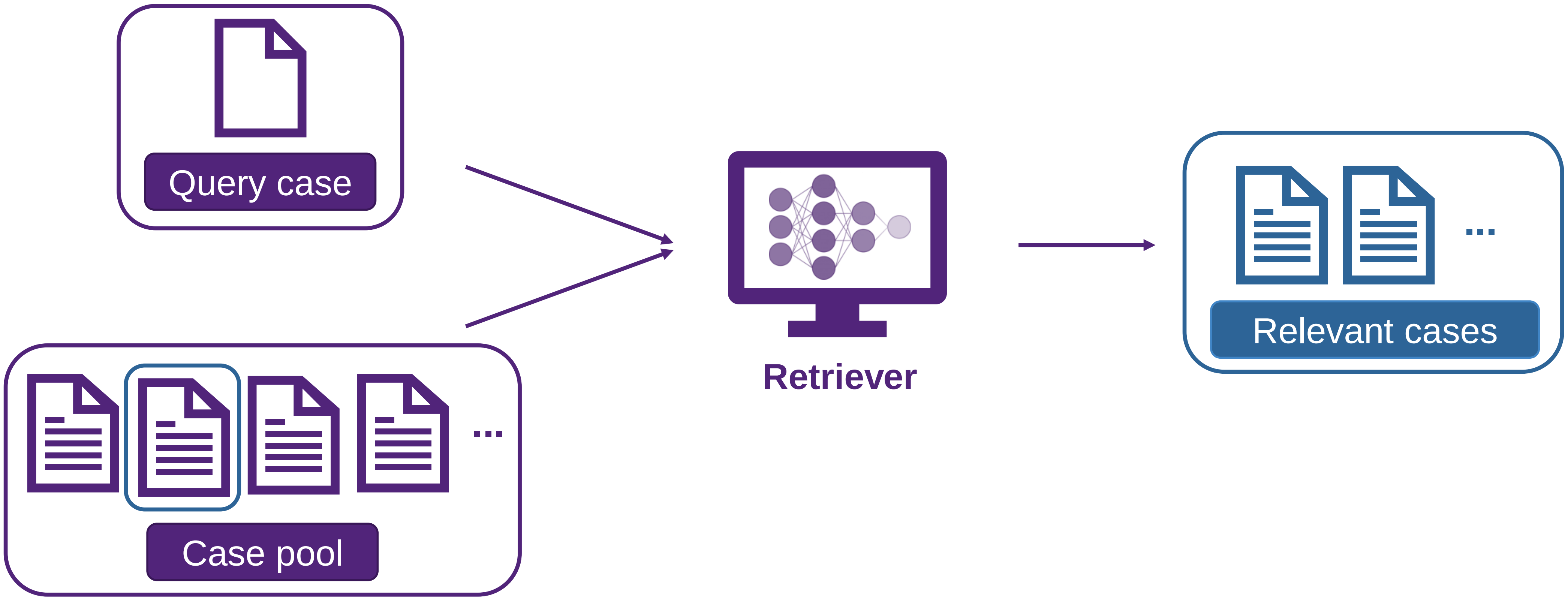
**Prof. Helen Huang**

EECS, UQ

# Contents


| | |
|---|---|
| 0 | Introduction |
| 1 | Related Work |
| 2 | Research 1: PromptCase |
| 3 | Research 2: CaseGNN |
| 4 | Research 3: CaseGNN++ |
| 5 | Research 4: CaseLink |
| 6 | Key Takeaways |

# Legal Case Retrieval Workflow



# Related work

# Related Work in Information Retrieval

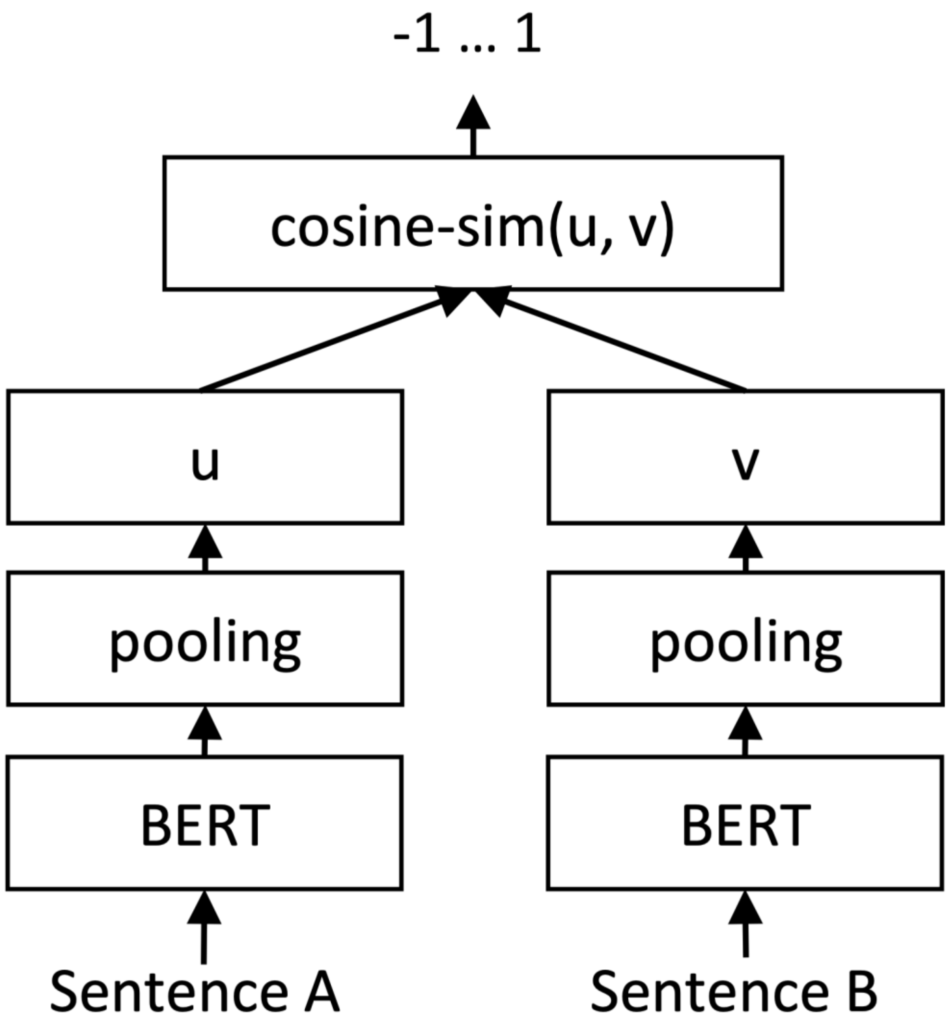
- **Sparse Retrieval**
  - TF-IDF [1]
  - BM25 [2]
  - LMIR [3]

[1] Gerard Salton and Christopher Buckley. “Term-weighting approaches in automatic text retrieval”, Information Processing & Management, 1988
[2] Stephen E Robertson and Steve Walker, “Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval”, In SIGIR, 1994
[3] Fei Song and W Bruce Croft, “A general language model for information retrieval”, In CIKM,1999

# Related Work in Information Retrieval

- **Dense Retrieval**
  - Sentence-BERT [4] :

    **Sentence embedding** of a query interacts with sentence embedding of a document.

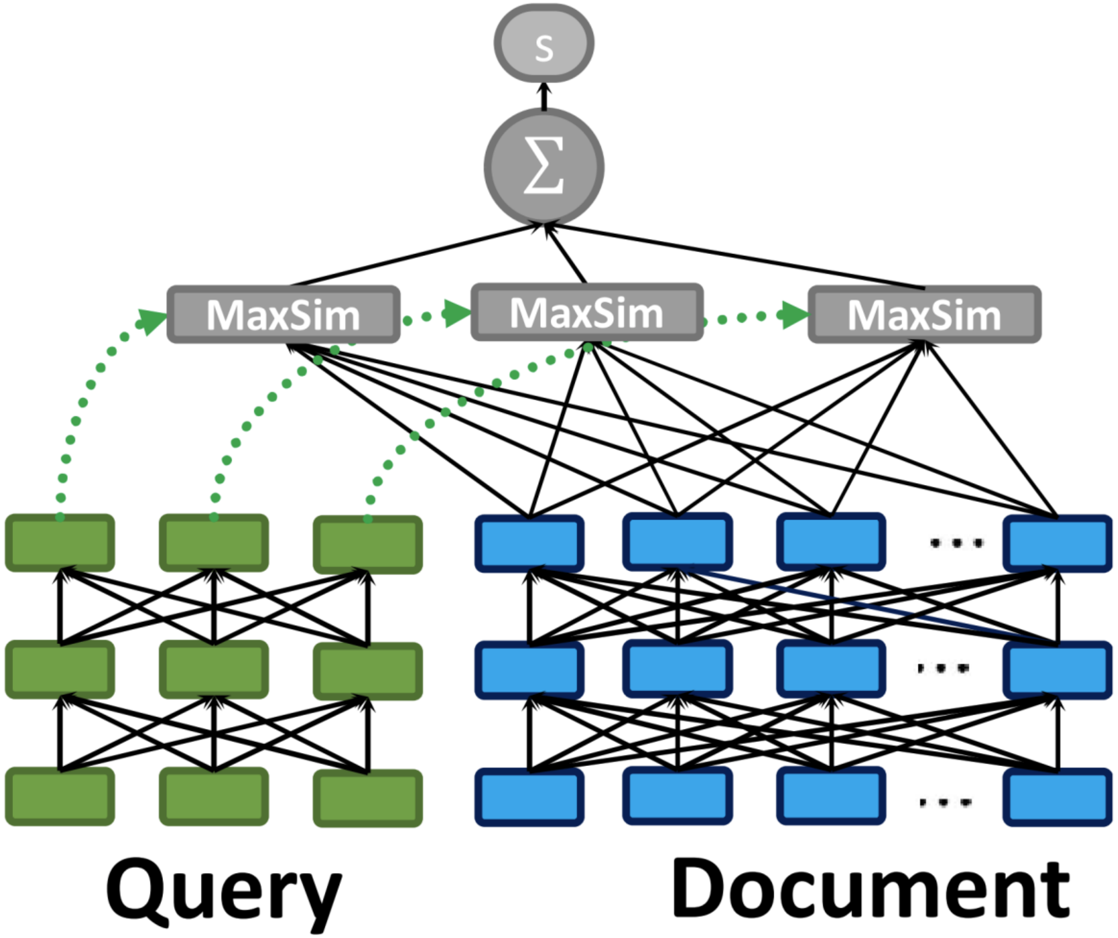



[4] Nils Reimers, Iryna Gurevych, “Sentence BERT: Sentence Embeddings using Siamese BERT-Networks”, In EMNLP, 2019

# Related Work in Information Retrieval

- **Dense Retrieval**
  - ColBERT [5] :

    Every **word embedding** of a query interacts with all word embeddings of a document.




[5] Omar Khattab, Matei Zaharia, "ColBERT: Efficient and Effective Passage Search via Contextualized Late Interaction over BERT", In SIGIR, 2020

# Summary

- Pros
  - High accuracy on normal IR tasks
  - Easy to apply on LCR

- Cons
  - No legal **expert knowledge**
  - For sparse retrieval: No **semantic**, which is very important for revealing legal relationship
  - For dense retrieval: Cases are **too long** to directly utilized dense information retrieval models.

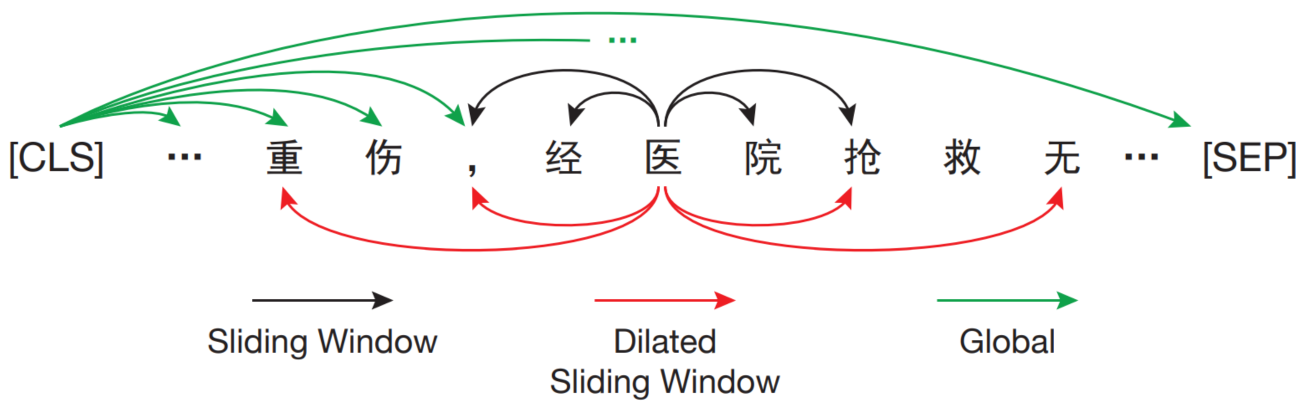
# Related Work in Legal Case Retrieval

- **Legal pre-trained model**
  - LEGAL-BERT [6] :
    - Pretrained with a large number of English legal corpus
    - 12 GB of diverse English legal text
    - Totally 355k pieces of UK legislation, European legislation and us court cases, etc.

[6] Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, Ion Androutsopoulos, "LEGAL-BERT: The Muppets straight out of Law School", In EMNLP (Findings), 2020

# Related Work in Legal Case Retrieval

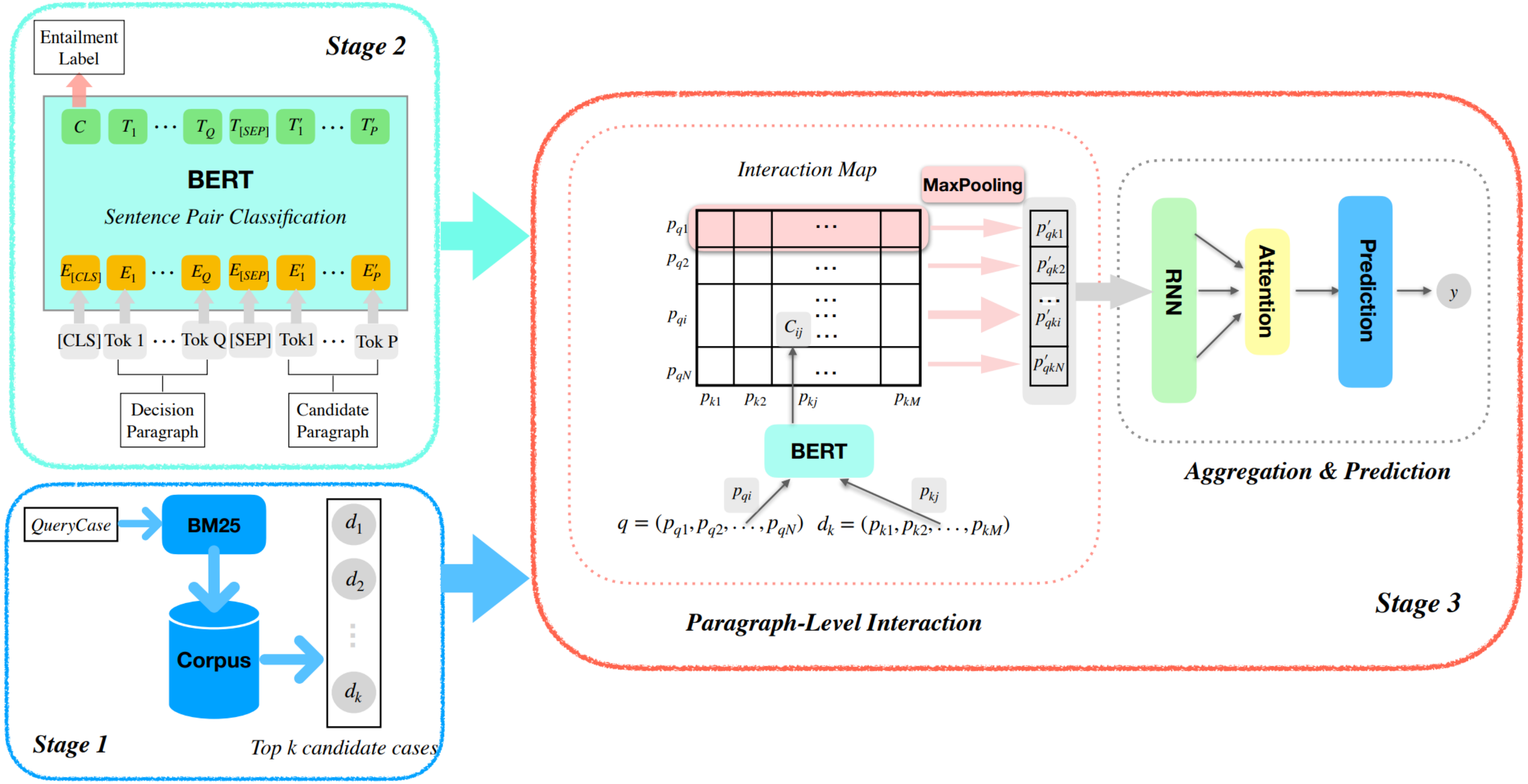
- **Legal pre-trained model**
  – Lawformer [7] :
    - Pretrained with Chinese legal corpus
    - Based model: Longformer
    - Combination of the three types of attention mechanism




[7] Chaojun Xiao, Xueyu Hu, Zhiyuan Liu, Cunchao Tu and Maosong Sun, "Lawformer: A pre-trained language model for chinese legal long documents", AI Open, 2021

# Related Work in Legal Case Retrieval

- **Bert-based model**
  - BERT-PLI [8]
    - Encode paragraphs with BERT
    - Paragraph-level interaction

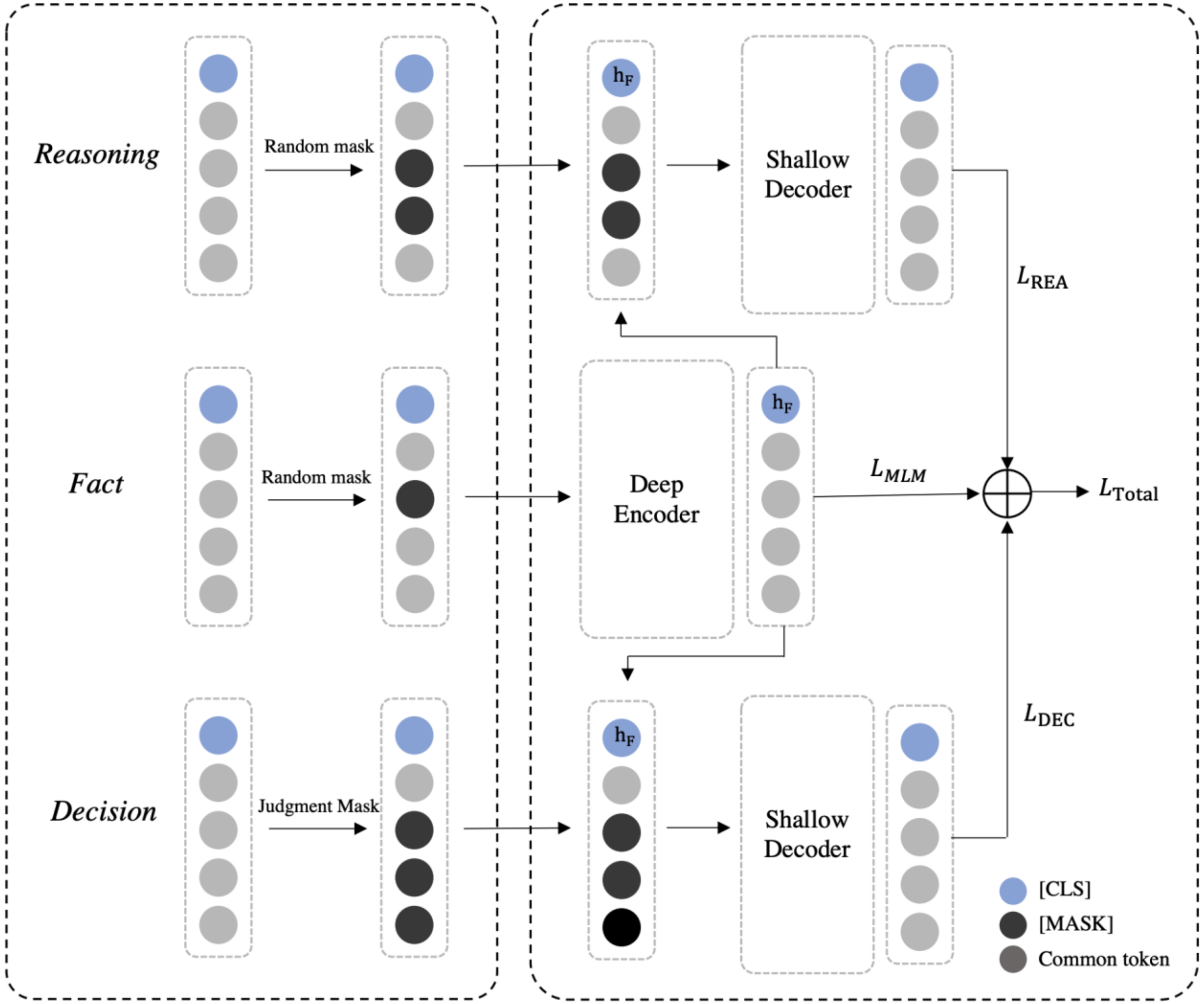





[8] Yunqiu Shao, Jiaxin Mao, Yiqun Liu, Weizhi Ma, Ken Satoh, Min Zhang and Shaoping Ma, “BERT-PLI: Modeling Paragraph-Level Interactions for Legal Case Retrieval”, In IJCAI, 2020

# Related Work in Legal Case Retrieval

- **Bert-based model**
  - SAILER [9]
    - Generation pretraining




[9] Haitao Li, Qingyao Ai, Jia Chen, Qian Dong, Yueyue Wu, Yiqun Liu, Chong Chen, Qi Tian, "SAILER: Structure-aware Pre-trained Language Model for Legal Case Retrieval", In SIGIR, 2023

# Summary

- Pros
  - Better accuracy with semantics by **legal corpus pre-training**
  - Dividing case text for **lengthy** problem

- Cons
  - Case text dividing → loss of legal **context** information & case **global** view

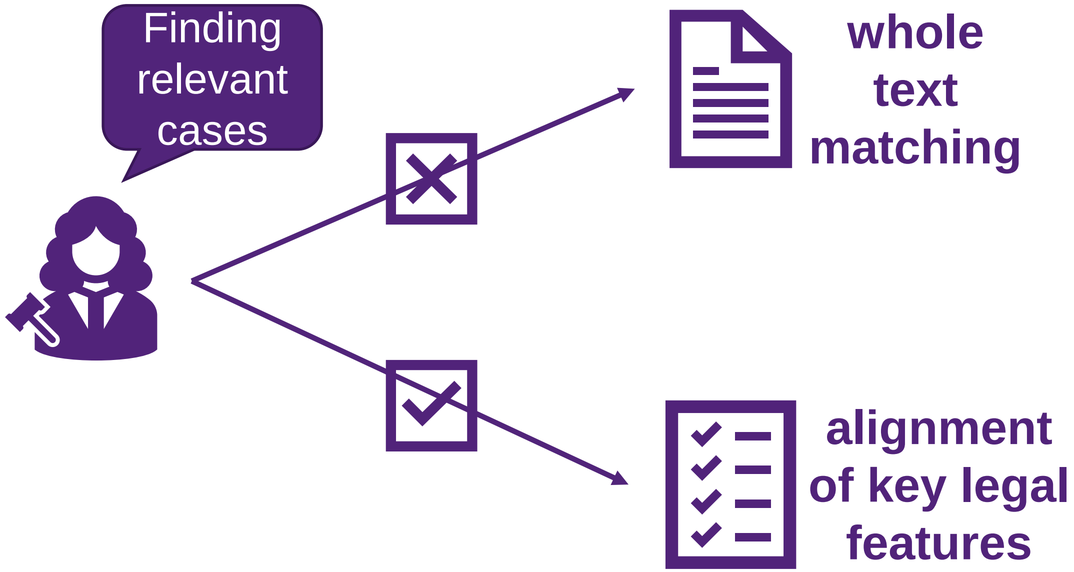
# Research 1

## PromptCase:
## Prompt-based effective input reformulation for legal case retrieval

Yanran Tang, Ruihong Qiu, and Xue Li, “Prompt-Based Effective Input Reformulation for Legal Case Retrieval”, in ADC 2023

# Challenges

- Determining factors of relevant cases:




- Input limitation of language models:

  Case needs to be truncated or divided into paragraphs

  → Loss of legal information

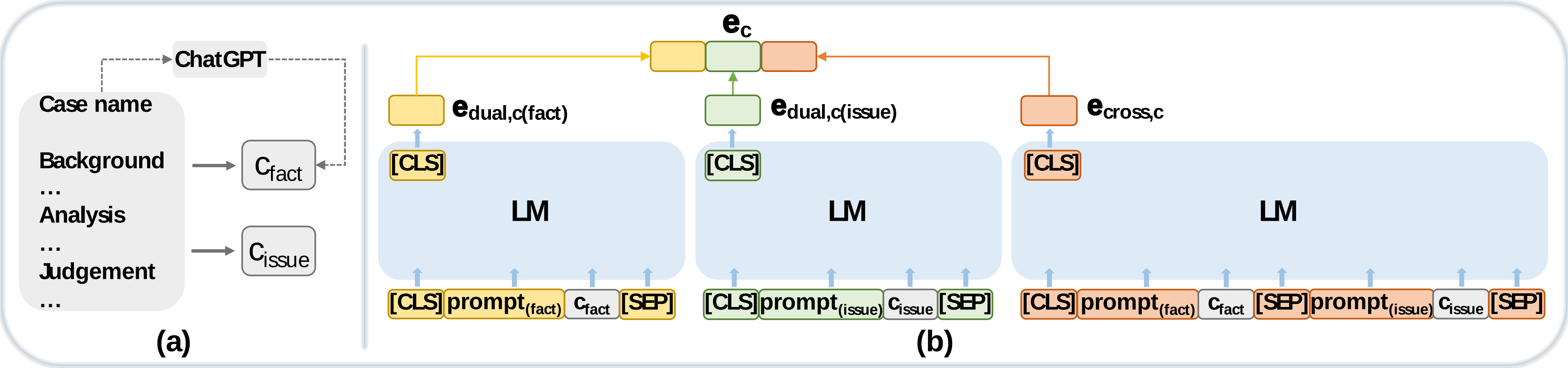
# Solution

- **Legal facts** and **legal issues** are considered as the determining factors:
  - Legal facts: Detailed process of a case → Case summary
  - Legal issues: Dispute points between the parties → Precedents / Charges

- **Identify** legal facts and legal issues → Feed into language model

- Use **prompt** to preserve legal context:
  - “The legal facts are: ” + legal facts
  - “The legal issues are: ” + legal issues

# PromptCase Framework



# Experiment Setting: Datasets

- English: COLIEE2023 [10]

**Lafond v. Muskeg Lake Cree Na3on (2008), 330 F.T.R. 60 (FC)**

**Background**
On February 13, 2006, the applicant was elected as a councillor to the MLCN Band Council
for a term of three years. The respondent Band is located in the province of Saskatchewan…

**Analysis**
Does this Court have jurisdiction over the present application? In order to determine the jurisdiction of the Federal Court in this matter, it is imperative to…
Indeed this was recognized by the Federal Court of Appeal in FRAGMENT_SUPPRESSED,
where it held that FRAGMENT_SUPPRESSED. I agree that the Chief does have inherent…

**Order**
For these reasons, the application for judicial review of Chief Ledoux's decision will be
allowed.

- Chinese: LeCaRD [11]

**李月航容留他人吸毒一案 (Case name)**

**案件基本情况 (Background)**
长乐市人民检察院指控： 1、2017 年 9 月 25 日 22 时许，被告人李月航在其租住的长乐市某街道某村某公寓房间内，容留王某吸食甲基苯丙胺（俗称"冰毒"）。 2、2017 年 10 月 19 日晚，被告人李月航在其租住的长乐市某街道某村某公寓房间内，容留王某… **经审理查明**： 1、2017 年 9 月 25 日 22 时许，被告人李月航在其租住的长 …

**裁判分析过程 (Analysis)**
本院认为，被告人李月航多次为他人吸食毒品提供场所，其行为已构成容留他人吸毒罪。长乐市人民检察院指控的罪名成立，应依法追究被告人李月航的刑事责任。被告人李月航因涉嫌吸毒被公安机关抓获，主动向公安机关供述了尚未被掌握的其容留他人吸毒的犯罪事实，视为自动投案，系自首，依法可从轻处罚；被告人李月航被公安 …

**判决结果 (Judgement)**
被告人李月航犯容留他人吸毒罪，判处拘役五个月，并处罚金人民币三千元。

[10] https://sites.ualberta.ca/~rabelo/COLIEE2023/
[11] Yixiao Ma, Yunqiu Shao, Yueyue Wu, Yiqun Liu, Ruizhe Zhang, Min Zhang, Shaoping Ma, "LeCaRD: A Legal Case Retrieval Dataset for Chinese Law System". In SIGIR, 2021

# Experiment Setting: Metrics

- Precision: $\frac{TP}{TP+FP}$
- Recall: $\frac{TP}{TP+FN}$
- F1: $2 \times \frac{precision \times recall}{precision + recall}$
- Macro F1: $\frac{1}{N}\sum_{i=1}^{N} \mathrm{F1}_i$
- Mean Average Precision (MAP) @K: $\frac{1}{N}\sum_{i=1}^{N} \mathrm{AP}_i$
- Mean Reciprocal Rank (MRR) @K: $\frac{1}{N}\sum_{v_{label} \in S_{test}} \frac{1}{Rank(v_{label})}$
- Normalized Discounted Cumulative Gain (NDCG) @K: $\frac{DCG@K}{IDCG@K}$

# Experiment Setting

### Baselines

- BM25
- BERT [12]
- Lawformer
- LEGAL-BERT
- Mono-T5 [13]
- SAILER

### Two-stage experiments

- Top 10 retrieved cases by BM25 as the first stage result

[12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, In NAACL-HLT, 2019
[13] Rodrigo Frassetto Nogueira, Zhiying Jiang, Ronak Pradeep, Jimmy Lin, “Document Ranking with a Pretrained Sequence-to-Sequence Model”, In EMNLP (Findings), 2020

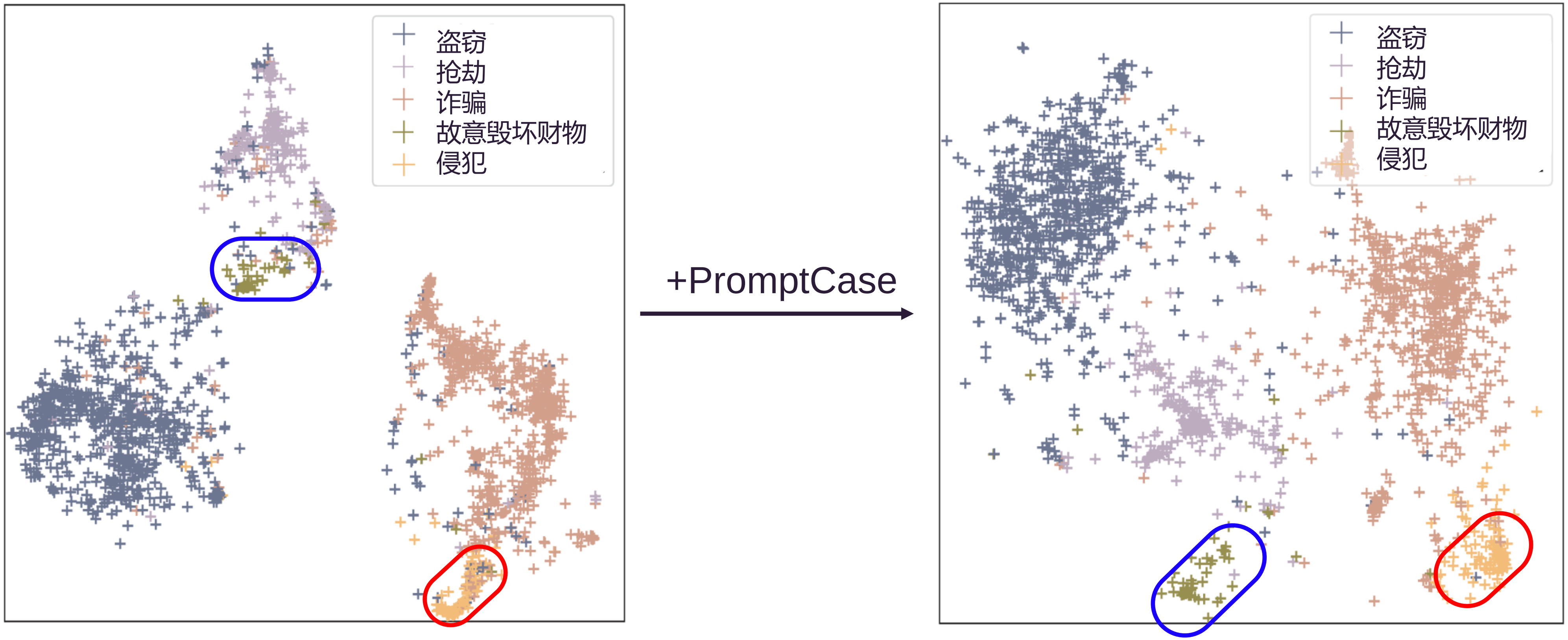
# Overall Performance

| Methods | LeCaRD@5 | | | | | | |
|---|---|---|---|---|---|---|---|
| | P@5 | R@5 | Mi-F1 | Ma-F1 | MRR@5 | MAP | NDCG@5 |
| BM25 | 40.0 | 19.2 | 26.0 | 30.5 | 58.3 | 48.5 | 45.9 |
| +PromptCase | 41.3 | 19.9 | 26.8 | 31.7 | 60.6 | 58.8 | 65.2 |
| BERT | 38.7 | 18.6 | 25.1 | 26.7 | 57.4 | 54.3 | 61.0 |
| +PromptCase | 46.2 | 22.2 | 30.0 | 35.4 | 64.4 | 61.2 | 67.9 |
| Lawformer | 29.0 | 13.9 | 18.8 | 19.5 | 43.6 | 41.9 | 48.2 |
| +PromptCase | 38.9 | 18.7 | 25.3 | 30.7 | 62.0 | 59.7 | 64.0 |
| SAILER | 46.7 | 22.5 | 30.4 | 37.1 | 67.9 | 65.4 | 70.1 |
| +PromptCase | 51.6 | 24.8 | 33.5 | 43.0 | 71.1 | 67.6 | 74.2 |
| Two-stage SAILER | 47.8 | 23.0 | 31.1 | 36.1 | 67.3 | 64.4 | 70.6 |
| +PromptCase | 51.0 | 24.6 | 33.2 | 38.7 | 70.7 | 67.9 | 73.5 |

| Methods | COLIEE2023 | | | | | | |
|---|---|---|---|---|---|---|---|
| | P@5 | R@5 | Mi-F1 | Ma-F1 | MRR@5 | MAP | NDCG@5 |
| BM25 | 16.5 | 30.6 | 21.4 | 22.2 | 23.1 | 20.4 | 23.7 |
| +PromptCase | 17.0 | 31.5 | 22.1 | 23.0 | 24.2 | 21.6 | 24.4 |
| BERT | 2.07 | 3.84 | 2.69 | 2.57 | 5.51 | 5.48 | 6.25 |
| +PromptCase | 2.38 | 4.42 | 3.10 | 3.02 | 6.33 | 6.25 | 7.21 |
| LEGAL-BERT | 4.64 | 8.61 | 6.03 | 6.03 | 11.4 | 11.3 | 13.6 |
| +PromptCase | 4.83 | 8.96 | 6.28 | 6.44 | 13.4 | 13.4 | 15.5 |
| MonoT5 | 0.38 | 0.70 | 0.49 | 0.47 | 1.17 | 1.33 | 0.61 |
| +PromptCase | 0.56 | 1.05 | 0.73 | 0.72 | 1.63 | 1.43 | 0.89 |
| SAILER | 12.8 | 23.7 | 16.6 | 17.0 | 25.9 | 25.3 | 29.3 |
| +PromptCase | 16.0 | 29.7 | 20.8 | 21.5 | 32.7 | 32.0 | 36.2 |
| Two-stage SAILER | 19.6 | 32.6 | 24.5 | 23.5 | 37.3 | 36.1 | 40.8 |
| +PromptCase | 21.8 | 36.3 | 27.2 | 26.5 | 39.9 | 38.7 | 44.0 |

Plug-and-play and improve consistently

# PromptCase Case Study




After utilising PromptCase, case embeddings evenly distributed corresponding to 5 charges as 5 clusters.

# Conclusion of Research 1

- **Legal facts** and **legal issues** are determining factors for legal case retrieval.

- **PromptCase** effectively encodes the legal features.

# Research 2

## CaseGNN:
## Graph neural networks for legal case retrieval with text-attributed graphs

Yanran Tang, Ruihong Qiu, Yilun Liu, Xue Li, and Zi Huang, “CaseGNN: Graph Neural Networks for Legal Case Retrieval with Text-Attributed Graphs”, in ECIR 2024.

# Challenges

- Legal **structural** information:
  - High-order interactions of **elements** in a case: parties, crime activities and evidences
- **Lengthy** legal text limitation:

| Datasets | LeCaRD | COLIEE2023 |
|---|---|---|
| Language | Chinese | English |
| Avg. length/case | 8,275 | 5,566 |
| Largest length of cases | 99,163 | 61,965 |
| Avg. relevant cases/query | 10.33 | 2.69 |

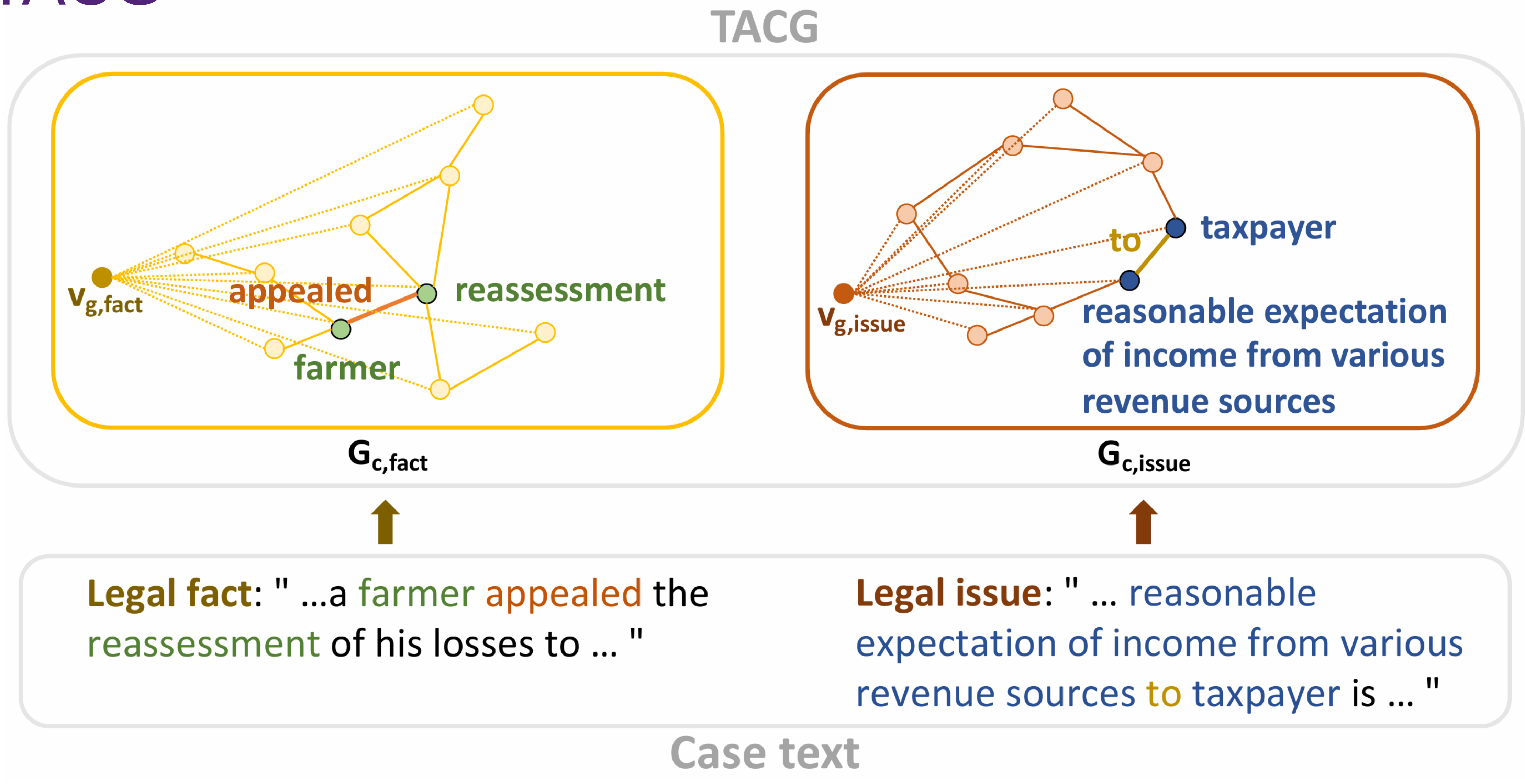
# Solution

- Graph is an effective data structure to incorporate **structural** information for legal cases.

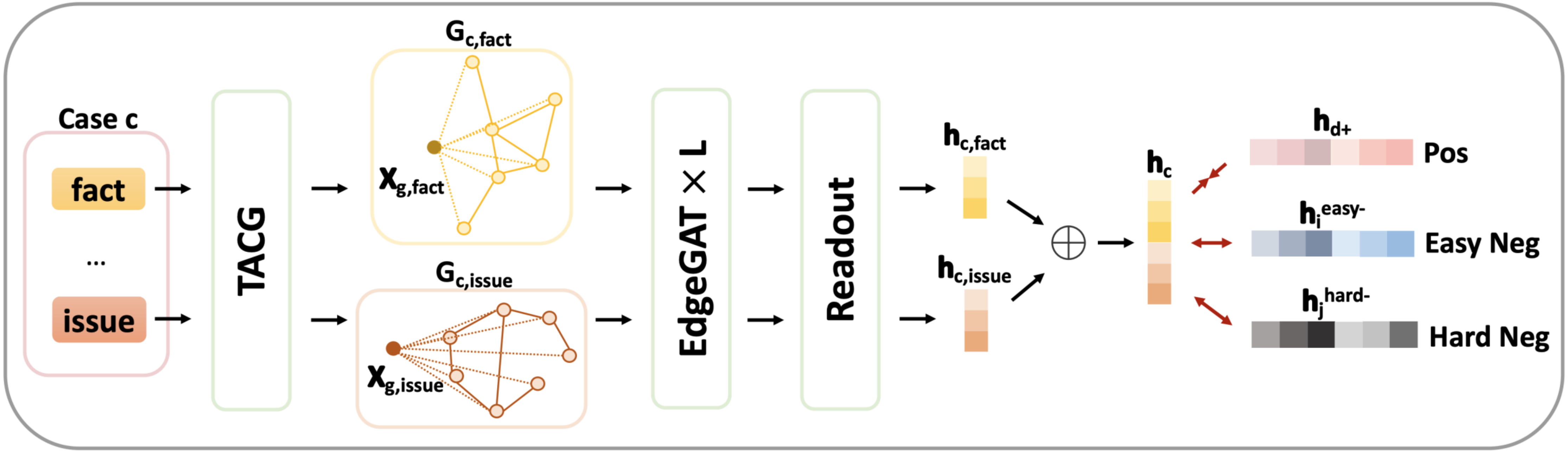
- Transform a legal case into a **Text-Attributed Case Graph (TACG)**.

- An **Edge Graph Attention Layer (EdgeGAT)** and a readout function are proposed to obtain a graph level case representation.

# TACG


TACG
$v_{g,fact}$
appealed
reassessment
farmer
$G_{c,fact}$
$v_{g,issue}$
to
taxpayer
reasonable expectation of income from various revenue sources
$G_{c,issue}$
Legal fact: " ...a farmer appealed the reassessment of his losses to ... "
Legal issue: " ... reasonable expectation of income from various revenue sources to taxpayer is ... "
Case text

# CaseGNN Framework



# Experiment Setting

- Metrics and baselines: follow PromptCase

- Datasets:
  - COLIEE2022 [14] and COLIEE2023
  - LeCaRD is not used due to no sufficient foundational and open-sourced relation extraction tool for Chinese

| Datasets | COLIEE2022 | | COLIEE2023 | |
|---|---|---|---|---|
| | train | test | train | test |
| # Query | 898 | 300 | 959 | 319 |
| # Candidates | 4415 | 1563 | 4400 | 1335 |
| # Avg. relevant cases | 4.68 | 4.21 | 4.68 | 2.69 |
| Avg. length (# token) | 6724 | 6785 | 6532 | 5566 |
| Largest length (# token) | 127934 | 85136 | 127934 | 61965 |



[14] https://sites.ualberta.ca/~rabelo/COLIEE2022/

# Overall Performance

| Methods | COLIEE2022 | | | | | | | COLIEE2023 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P@5 | R@5 | Mi-F1 | Ma-F1 | MRR@5 | MAP | NDCG@5 | P@5 | R@5 | Mi-F1 | Ma-F1 | MRR@5 | MAP | NDCG@5 |
| **One-stage** | | | | | | | | | | | | | | |
| BM25 | 17.9 | 21.2 | 19.4 | 21.4 | 23.6 | 25.4 | 33.6 | 16.5 | 30.6 | 21.4 | 22.2 | 23.1 | 20.4 | 23.7 |
| LEGAL-BERT | 4.47 | 5.30 | 4.85 | 5.38 | 7.42 | 7.47 | 10.9 | 4.64 | 8.61 | 6.03 | 6.03 | 11.4 | 11.3 | 13.6 |
| MonoT5 | 0.71 | 0.65 | 0.60 | 0.79 | 1.39 | 1.41 | 1.73 | 0.38 | 0.70 | 0.49 | 0.47 | 1.17 | 1.33 | 0.61 |
| SAILER | 16.6 | 15.2 | 14.0 | 16.8 | 17.2 | 18.5 | 25.1 | 12.8 | 23.7 | 16.6 | 17.0 | 25.9 | 25.3 | 29.3 |
| PromptCase | 17.1 | 20.3 | 18.5 | 20.5 | 35.1 | 33.9 | 38.7 | 16.0 | 29.7 | 20.8 | 21.5 | 32.7 | 32.0 | 36.2 |
| CaseGNN (Ours) | **35.5**±0.2 | **42.1**±0.2 | **38.4**±0.3 | **42.4**±0.1 | **66.8**±0.8 | **64.4**±0.9 | **69.3**±0.8 | **17.7**±0.7 | **32.8**±0.7 | **23.0**±0.5 | **23.6**±0.5 | **38.9**±1.1 | **37.7**±0.8 | **42.8**±0.7 |
| **Two-stage** | | | | | | | | | | | | | | |
| SAILER | **23.8** | 25.7 | 24.7 | 25.2 | 43.9 | 42.7 | 48.4 | 19.6 | 32.6 | 24.5 | 23.5 | 37.3 | 36.1 | 40.8 |
| PromptCase | 23.5 | 25.3 | 24.4 | **30.3** | 41.2 | 39.6 | 45.1 | **21.8** | 36.3 | **27.2** | 26.5 | 39.9 | 38.7 | 44.0 |
| CaseGNN (Ours) | 22.9±0.1 | **27.2**±0.1 | **24.9**±0.1 | 27.0±0.1 | **54.9**±0.4 | **54.0**±0.5 | **57.3**±0.6 | 20.2±0.2 | **37.6**±0.5 | 26.3±0.3 | **27.3**±0.2 | **45.8**±0.9 | **44.4**±0.8 | **49.6**±0.8 |

- CaseGNN outperforms other baselines.
- CaseGNN does not benefit from **two-stage** retrieval in COLIEE2022, since BM25 cannot provide a useful first stage result.

# Conclusion of Research 2

- Legal **structural** information is important and can be utilised by **graph neural network**.

- Case graphs help avoid **lengthy** case text and **preserve** legal context.

# Research 3

## CaseGNN++:
## Graph Contrastive Learning for Legal Case Retrieval with Graph Augmentation



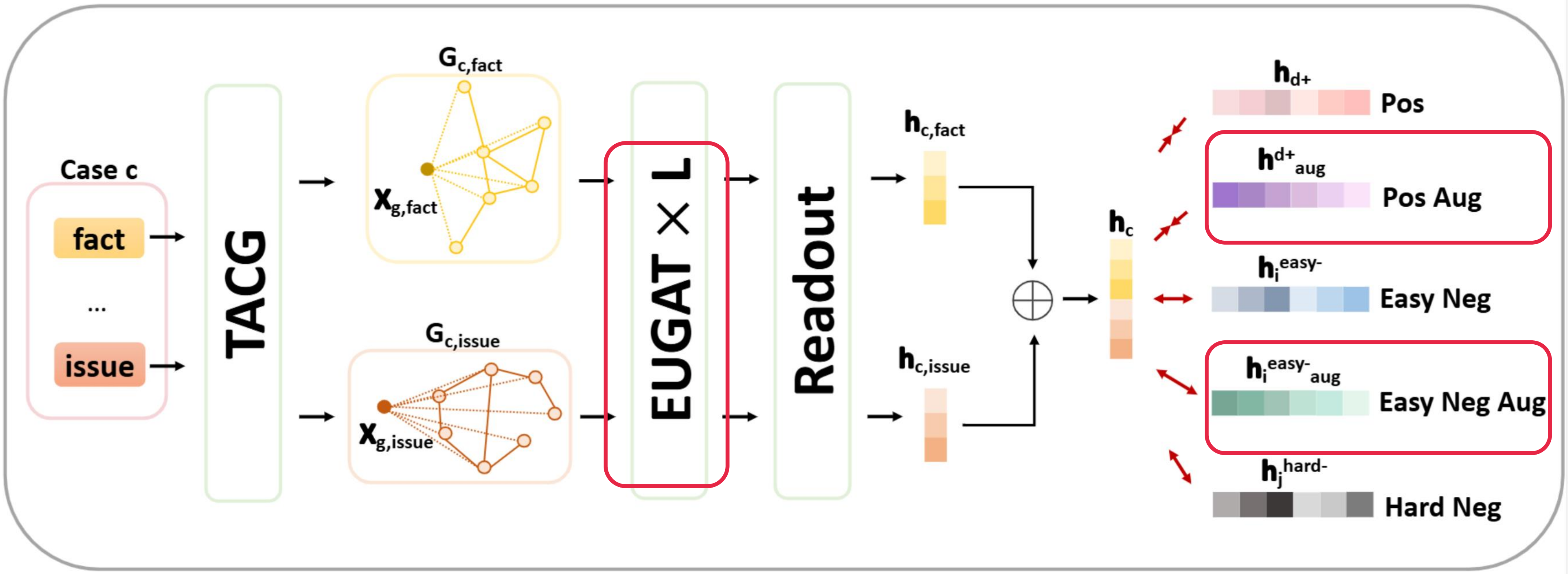
Yanran Tang, Ruihong Qiu, Yilun Liu, Xue Li and Zi Huang, "CaseGNN++: Graph Contrastive Learning for Legal Case Retrieval with Graph Augmentation", CoRR abs/2307.11572, 2024,under review of TOIS

# Challenges

- The underutilization of rich **edge information** within text-attributed case graphs limits CaseGNN to generate informative case representation

- The **inadequacy of labelled data** in legal datasets hinders the training of CaseGNN model.

# CaseGNN++ Framework




- EUGAT
  - Comprehensively update node and edge features during graph modelling
- Graph Contrastive Learning & Graph Augmentation:
  - Edge Dropping
  - Feature Masking: node or edge feature

# Overall Performance

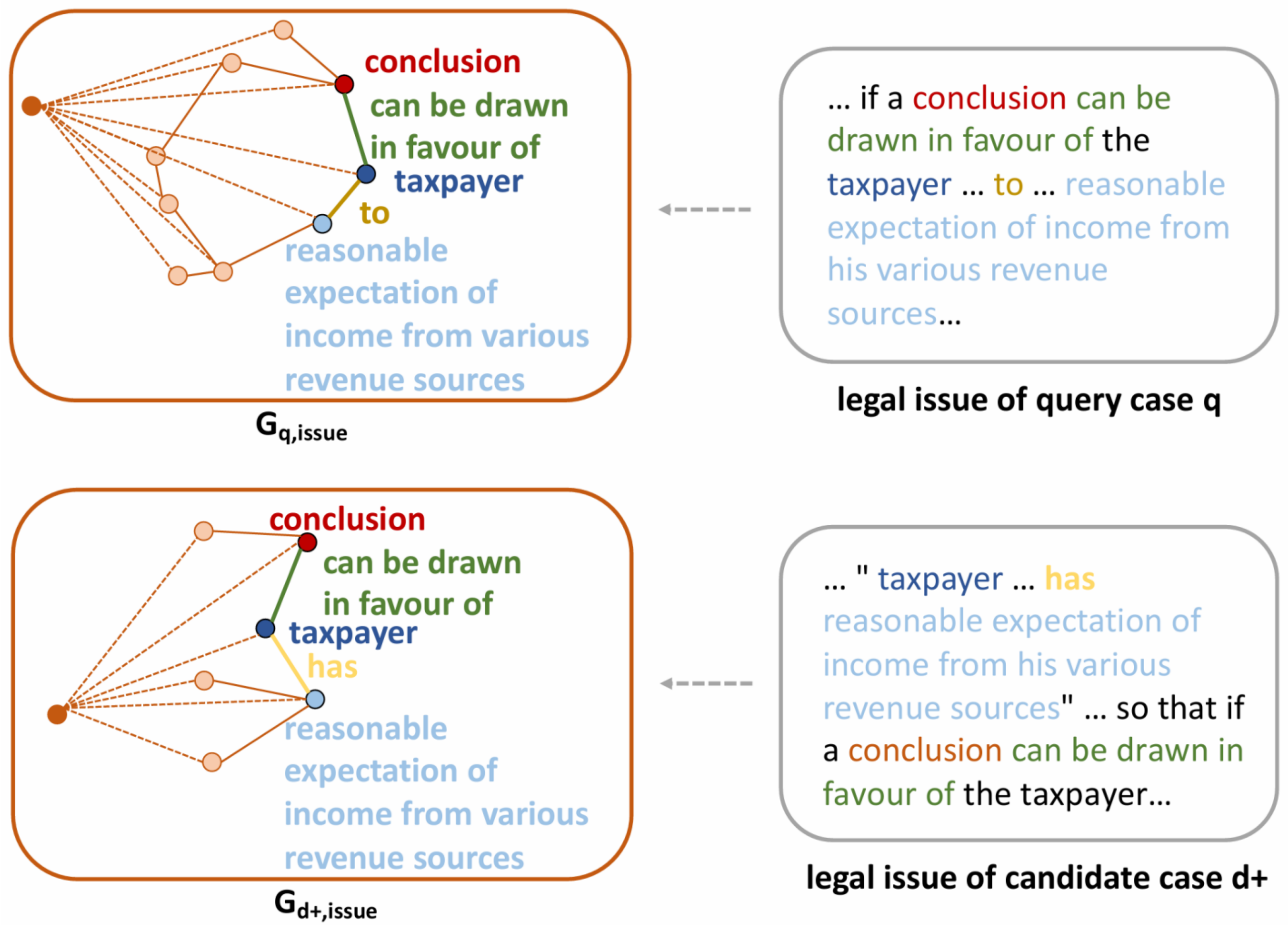
| Methods | COLIEE2022 | | | | | | |
|---|---|---|---|---|---|---|---|
| | P@5 | R@5 | Mi-F1 | Ma-F1 | MRR@5 | MAP | NDCG@5 |
| **One-stage** | | | | | | | |
| BM25 | <u>17.9</u> | <u>21.2</u> | <u>19.4</u> | <u>21.4</u> | 23.6 | 25.4 | 33.6 |
| LEGAL-BERT | 4.47 | 5.30 | 4.85 | 5.38 | 7.42 | 7.47 | 10.9 |
| MonoT5 | 0.71 | 0.65 | 0.60 | 0.79 | 1.39 | 1.41 | 1.73 |
| SAILER | 16.6 | 15.2 | 14.0 | 16.8 | 17.2 | 18.5 | 25.1 |
| PromptCase | 17.1 | 20.3 | 18.5 | 20.5 | <u>35.1</u> | <u>33.9</u> | <u>38.7</u> |
| CaseGNN (Ours) | 35.5±0.2 | 42.1±0.2 | 38.4±0.3 | 42.4±0.1 | 66.8±0.8 | 64.4±0.9 | 69.3±0.8 |
| CaseGNN++ (Ours) | **36.5**±0.6 | **43.3**±0.7 | **39.6**±0.6 | **43.8**±0.7 | **68.1**±1.1 | **65.3**±1.1 | **70.8**±1.1 |
| **Two-stage** | | | | | | | |
| SAILER | <u>23.8</u> | <u>25.7</u> | <u>24.7</u> | 25.2 | <u>43.9</u> | <u>42.7</u> | <u>48.4</u> |
| PromptCase | 23.5 | 25.3 | 24.4 | <u>**30.3**</u> | 41.2 | 39.6 | 45.1 |
| CaseGNN (Ours) | 22.9±0.1 | 27.2±0.1 | 24.9±0.1 | 27.0±0.1 | 54.9±0.4 | 54.0±0.5 | 57.3±0.6 |
| CaseGNN++ (Ours) | **24.8**±0.1 | **29.4**±0.1 | **26.9**±0.1 | 29.3±0.1 | **55.6**±0.6 | **54.3**±0.3 | **58.1**±0.3 |

# Overall Performance

| Methods | COLIEE2023 | | | | | | |
|---|---|---|---|---|---|---|---|
| | P@5 | R@5 | Mi-F1 | Ma-F1 | MRR@5 | MAP | NDCG@5 |
| One-stage | | | | | | | |
| BM25 | 16.5 | 30.6 | 21.4 | 22.2 | 23.1 | 20.4 | 23.7 |
| LEGAL-BERT | 4.64 | 8.61 | 6.03 | 6.03 | 11.4 | 11.3 | 13.6 |
| MonoT5 | 0.38 | 0.70 | 0.49 | 0.47 | 1.17 | 1.33 | 0.61 |
| SAILER | 12.8 | 23.7 | 16.6 | 17.0 | 25.9 | 25.3 | 29.3 |
| PromptCase | 16.0 | 29.7 | 20.8 | 21.5 | 32.7 | 32.0 | 36.2 |
| CaseGNN (Ours) | 17.7±0.7 | 32.8±0.7 | 23.0±0.5 | 23.6±0.5 | 38.9±1.1 | 37.7±0.8 | 42.8±0.7 |
| CaseGNN++ (Ours) | **18.2**±0.3 | **33.8**±0.4 | **23.7**±0.4 | **24.3**±0.3 | **40.0**±0.2 | **38.9**±0.3 | **43.8**±0.3 |
| **Two-stage** | | | | | | | |
| SAILER | 19.6 | 32.6 | 24.5 | 23.5 | 37.3 | 36.1 | 40.8 |
| PromptCase | **21.8** | 36.3 | **27.2** | 26.5 | 39.9 | 38.7 | 44.0 |
| CaseGNN (Ours) | 20.2±0.2 | 37.6±0.5 | 26.3±0.3 | 27.3±0.2 | 45.8±0.9 | 44.4±0.8 | 49.6±0.8 |
| CaseGNN++ (Ours) | 20.4±0.1 | **37.9**±0.2 | 26.6±0.2 | **27.5**±0.2 | **45.9**±0.4 | **44.5**±0.3 | **49.9**±0.3 |

# CaseGNN & CaseGNN++ Case Study

- Successful retrieval by CaseGNN & CaseGNN++ but not by PromptCase.




- Original **text**: entities and relationships are **far** from each other. Language models are **not** good at **long dependency**.
- TACG: brings multiple entities **together**.

# Research 4

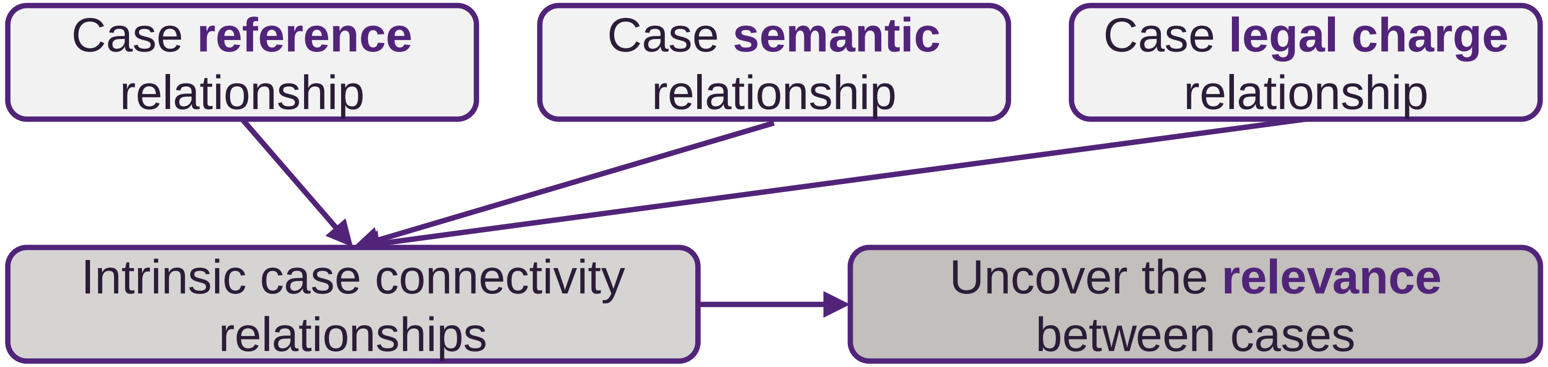
# CaseLink:
# Inductive Graph Learning for Legal Case Retrieval



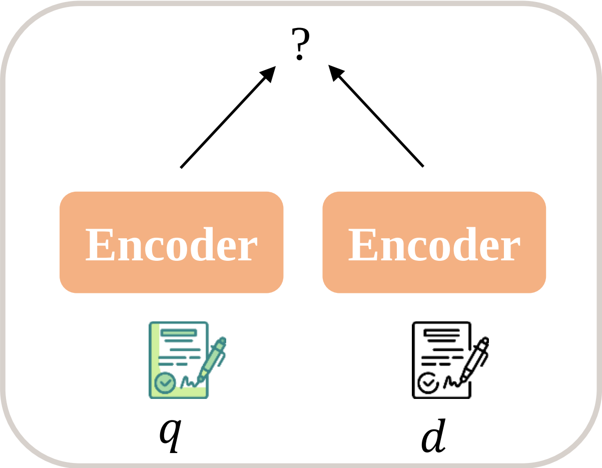
Yanran Tang, Ruihong Qiu, Xue Li, and Zi Huang, “CaseLink: Inductive Graph Learning for Legal Case Retrieval”, in SIGIR 2024

# Challenges

- The intrinsic case connectivity relationships are important for legal case retrieval.

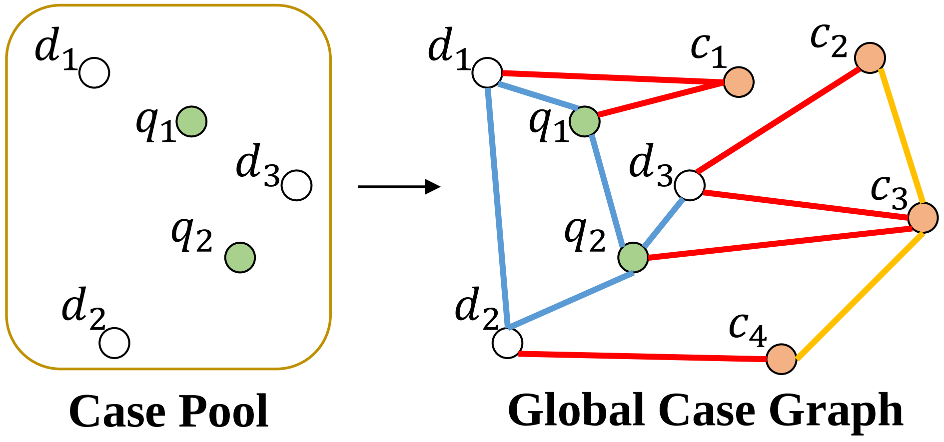



- Not well exploited in general methods.

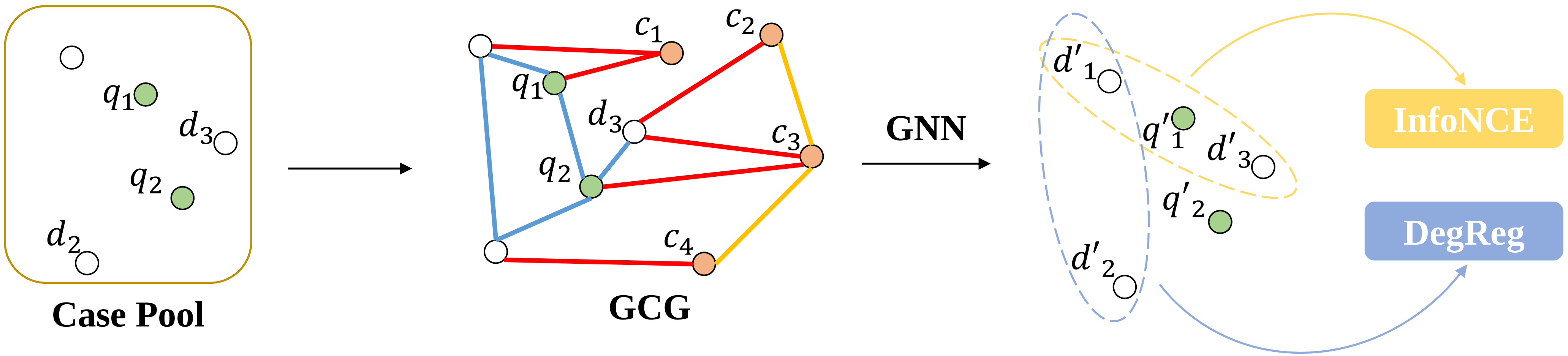


# Solution

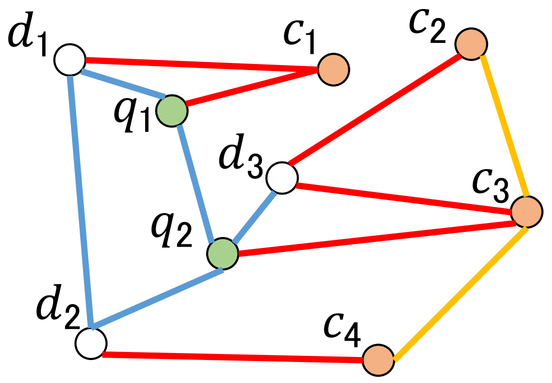
- A pool of cases is converted into a structured graph
  - case-case bm25 (blue)
  - Case-charge (red)
  - Charge-charge (yellow)



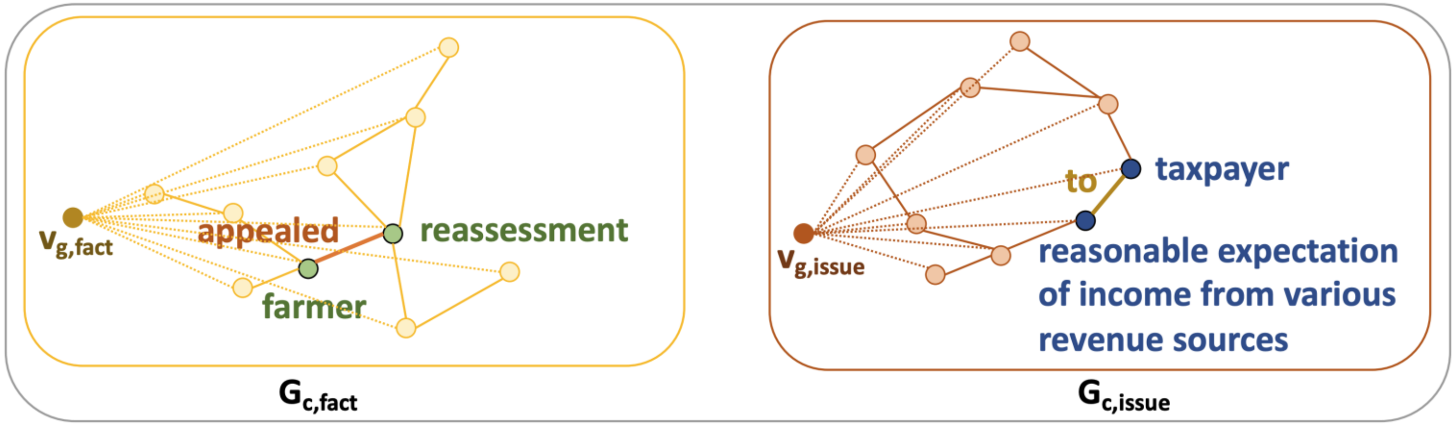
# CaseLink Framework



# GCG Compared with TACG




GCG




TACG

A GCG includes a pool of cases.

Every node is a case.

A TACG stands for a case.

Every node is an entity of the case.

# Degree Regularisation (DegReg)

- Motivation:
  - Real-world **sparse** situation: candidate case will be only related to a small number of query cases of pool → **low degree**
  - Providing the training signal for **candidate** cases

- $\ell_{\text{DegReg}} = \sum(\hat{A}_{candidate}) \downarrow$
  - Minimising the degree of candidate nodes

# Experiment

Settings: the same as CaseGNN

Overall performance:

| Methods | COLIEE2022 | | | | | | | COLIEE2023 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P@5 | R@5 | Mi-F1 | Ma-F1 | MRR@5 | MAP | NDCG@5 | P@5 | R@5 | Mi-F1 | Ma-F1 | MRR@5 | MAP | NDCG@5 |
| **One-stage** | | | | | | | | | | | | | | |
| BM25 | 17.9 | 21.2 | 19.4 | 21.4 | 23.6 | 25.4 | 33.6 | 16.5 | 30.6 | 21.4 | 22.2 | 23.1 | 20.4 | 23.7 |
| LEGAL-BERT | 4.47 | 5.30 | 4.85 | 5.38 | 7.42 | 7.47 | 10.9 | 4.64 | 8.61 | 6.03 | 6.03 | 11.4 | 11.3 | 13.6 |
| MonoT5 | 0.71 | 0.65 | 0.60 | 0.79 | 1.39 | 1.41 | 1.73 | 0.38 | 0.70 | 0.49 | 0.47 | 1.17 | 1.33 | 0.61 |
| SAILER | 16.6 | 15.2 | 14.0 | 16.8 | 17.2 | 18.5 | 25.1 | 12.8 | 23.7 | 16.6 | 17.0 | 25.9 | 25.3 | 29.3 |
| PromptCase | 17.1 | 20.3 | 18.5 | 20.5 | 35.1 | 33.9 | 38.7 | 16.0 | 29.7 | 20.8 | 21.5 | 32.7 | 32.0 | 36.2 |
| CaseGNN | 35.5±0.2 | 42.1±0.2 | 38.4±0.3 | 42.4±0.1 | 66.8±0.8 | 64.4±0.9 | 69.3±0.8 | 17.7±0.7 | 32.8±0.7 | 23.0±0.5 | 23.6±0.5 | 38.9±1.1 | 37.7±0.8 | 42.8±0.7 |
| CaseLink (Ours) | **37.0**±0.1 | **43.9**±0.1 | **40.1**±0.1 | **44.2**±0.1 | **67.3**±0.5 | **65.0**±0.2 | **70.3**±0.1 | **20.9**±0.3 | **38.4**±0.6 | **27.1**±0.3 | **28.2**±0.3 | **45.8**±0.5 | **44.3**±0.7 | **49.8**±0.4 |
| **Two-stage** | | | | | | | | | | | | | | |
| SAILER | 23.8 | 25.7 | 24.7 | 25.2 | 43.9 | 42.7 | 48.4 | 19.6 | 32.6 | 24.5 | 23.5 | 37.3 | 36.1 | 40.8 |
| PromptCase | 23.5 | 25.3 | 24.4 | 30.3 | 41.2 | 39.6 | 45.1 | **21.8** | 36.3 | 27.2 | 26.5 | 39.9 | 38.7 | 44.0 |
| CaseGNN | 22.9±0.1 | 27.2±0.1 | 24.9±0.1 | 27.0±0.1 | 54.9±0.4 | 54.0±0.5 | 57.3±0.6 | 20.2±0.2 | 37.6±0.5 | 26.3±0.3 | 27.3±0.2 | 45.8±0.9 | 44.4±0.8 | 49.6±0.8 |
| CaseLink (Ours) | **24.7**±0.1 | **29.1**±0.1 | **26.8**±0.1 | 29.2±0.1 | **56.0**±0.2 | **55.0**±0.2 | **58.6**±0.1 | 21.0±0.3 | **38.9**±0.5 | 27.1±0.3 | **28.2**±0.3 | **48.8**±0.2 | **47.2**±0.1 | **52.6**±0.1 |

CaseLink performs the best, better than CaseGNN.

Two-stage still suffers from a poor BM25 first-stage ranker.

# Conclusion of Research 4

- Global Case Graph provides effective **connections** among cases.

- **Degree regularisation** can provide effective training signals for candidate cases.

# Key Takeaways

- **Structural** legal information is essential for legal case retrieval.

- Both **intra-case** structural information and **inter-case** structural information can highly be beneficial to legal case retrieval.

# Thank you!

# Q & A

github.com/yanran-tang