De-Biased Court's View Generation with Causality

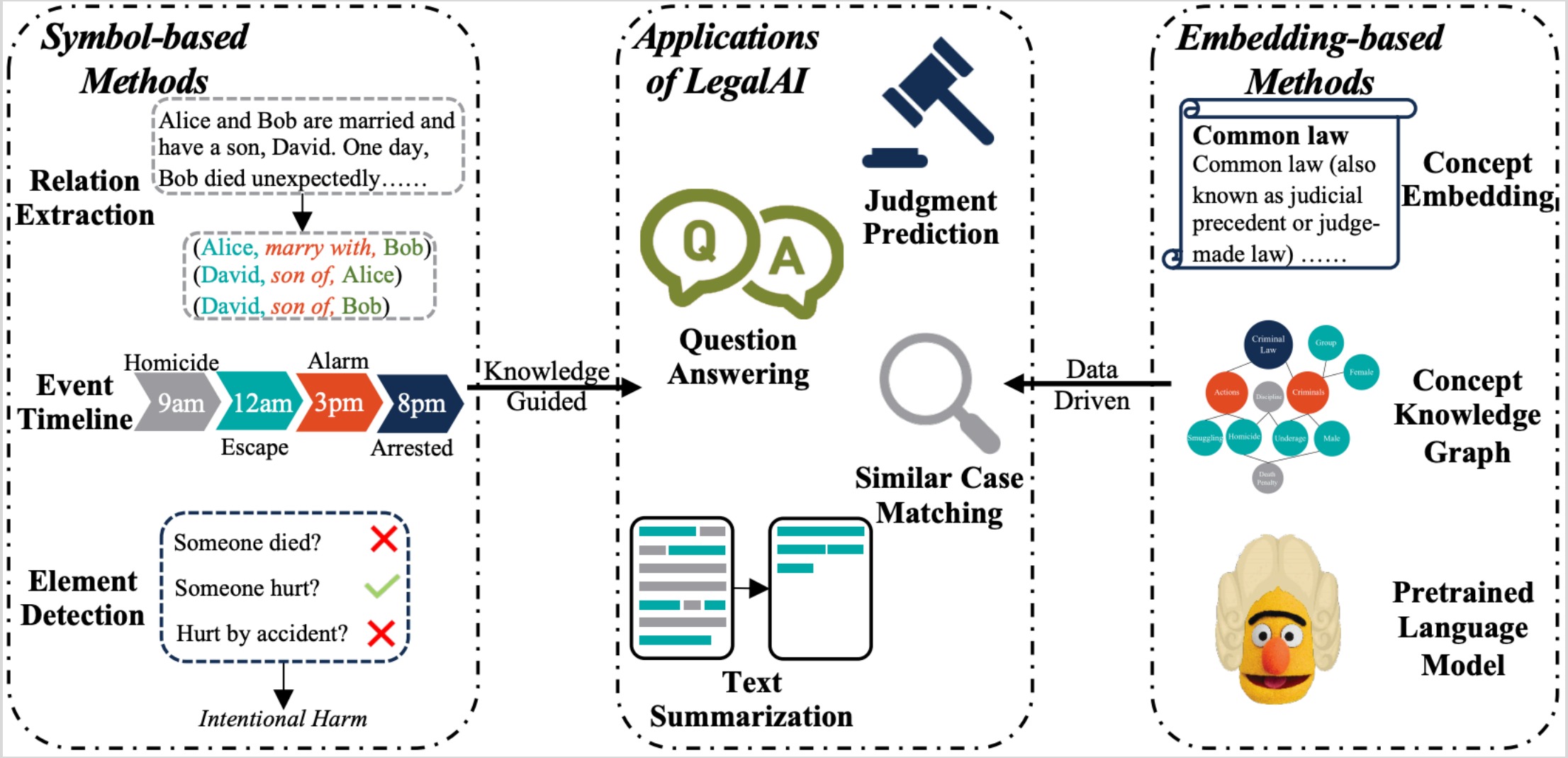
Yiquan Wu[1], Kun Kuang[1], Yating Zhang[2], Xiaozhong Liu[3], Changlong Sun[2], Jun Xiao[1], Yueting Zhuang[1], Luo Si[2], Fei Wu[1]

Zhejiang University[1], Alibaba Group[2], Indiana University Bloomington[3]

Introduction

Court’s View Generation

Legal AI



[1] How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence.

Task Definition

PLAINTIFF'S CLAIM	The plaintiff A claimed that the defendant B should return the loan of $29,500 *Principle Claim* and the corresponding interest *Interest Claim*.
FACT DESCRIPTION	After the hearing, the court held the facts as follows: The defendant B borrowed $29,500 from the plaintiff A, and agreed to return after one month. After the loan expired, the defendant failed to return *Fact*.
COURT'S VIEW	The court concluded that the loan relationship between the plaintiff A and the defendant B is valid. The defendant failed to return the money on time *Rationale*. Therefore, the plaintiff's claim on principle was supported *Acceptance* according to law. The court did not support the plaintiff's claim on interest *Rejection* because the evidence was insufficient *Rationale*.

Input:

- Plaintiff's claim
- Fact description

Output:

- Court's View, which consists of
  - Rationale
  - Judgment

Court's view generation is a specific text generation task

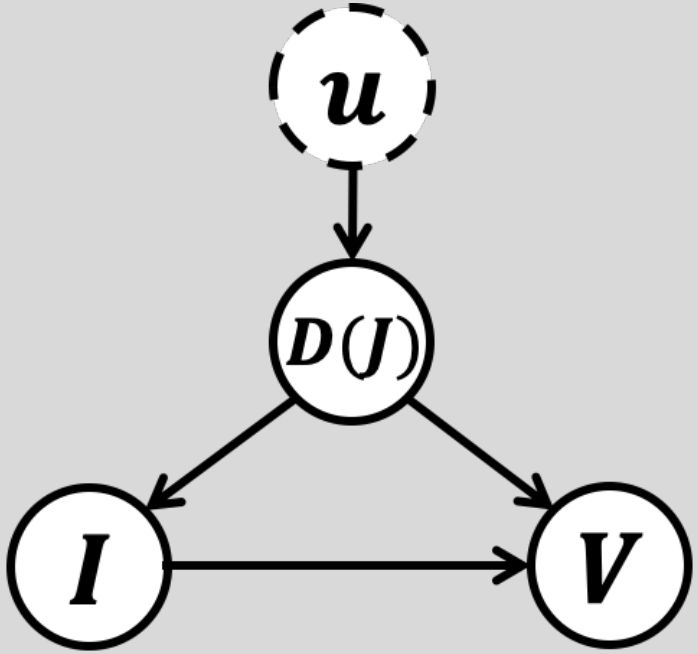
Challenges

PLAINTIFF'S CLAIM	The plaintiff A claimed that the defendant B should return the loan of $29,500 *Principle Claim* and the corresponding interest *Interest Claim*.
FACT DESCRIPTION	After the hearing, the court held the facts as follows: The defendant B borrowed $29,500 from the plaintiff A, and agreed to return after one month. After the loan expired, the defendant failed to return *Fact*.
COURT'S VIEW	The court concluded that the loan relationship between the plaintiff A and the defendant B is valid. The defendant failed to return the money on time *Rationale*. Therefore, the plaintiff's claim on principle was supported *Acceptance* according to law. The court did not support the plaintiff's claim on interest *Rejection* because the evidence was insufficient *Rationale*.

- There exists '**no claim, no trial**' principle in civil legal systems
  - court's view should only focus on the facts related to the claims
- The **imbalance** of judgment in civil cases
  - over 76% of cases were supported in private lending
  - would blind the training of the model by focusing on the supported cases while ignoring the non-supported cases

Imbalance: Mechanism Confounding Bias

- Imbalance between supported and non-supported cases
  - Lead to confounding bias during model training
- Understanding confounding bias with a causal graph:
  - u: unobserved data generation mechanism
  - D(J): judgment in dataset
  - I: input (i.e., plaintiff's claim and fact description)
  - V: court's view
- Understanding confounding bias mathematically
  - j: judgment (support and non-support):



$$P(V|I) = \sum_j P(V|I,j)P(j|I) \xrightarrow{P(j=1|I)\approx 1} P(V|I) \approx P(V|I,j=1)$$

Method

Attentional and Counterfactual based Natural Language Generation

Attentional and Counterfactual based NLG

- Attentional encoder:
  - Claim-aware attention
- Counterfactual decoder:
  - Back-door adjustment: from observation to intervention
  - Cut the dependence between D(J) and I via counterfactual modeling

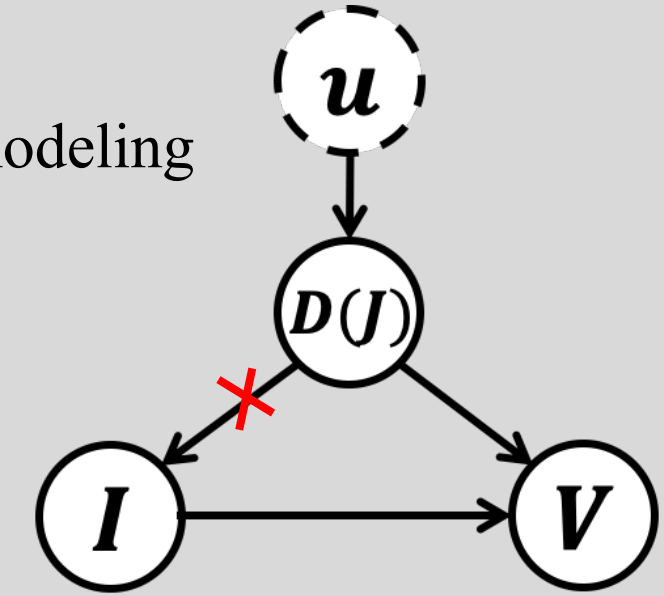
$$P(V|I) = \sum_j P(V|I,j)P(j|I)$$

Back-door

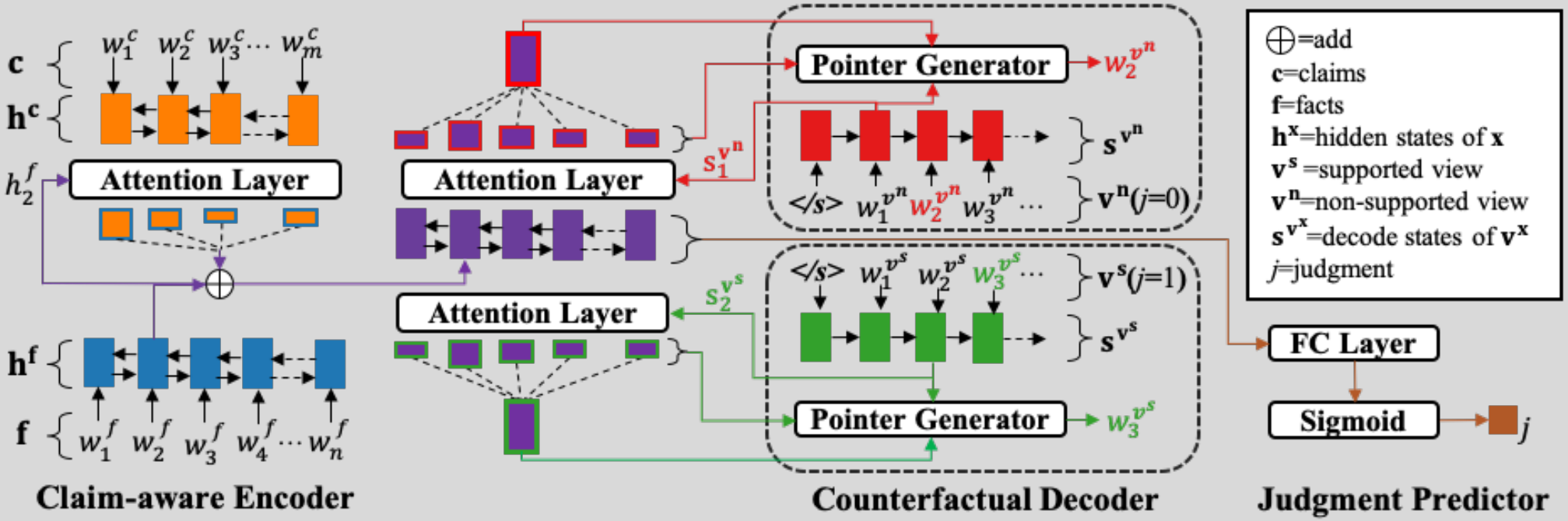
$$P(V|do(I)) = \sum_j P(V|I,j)P(j)$$

Binary j

$$P(V|do(I)) = P(V|I,j=0)P(j=0) + P(V|I,j=1)P(j=1)$$



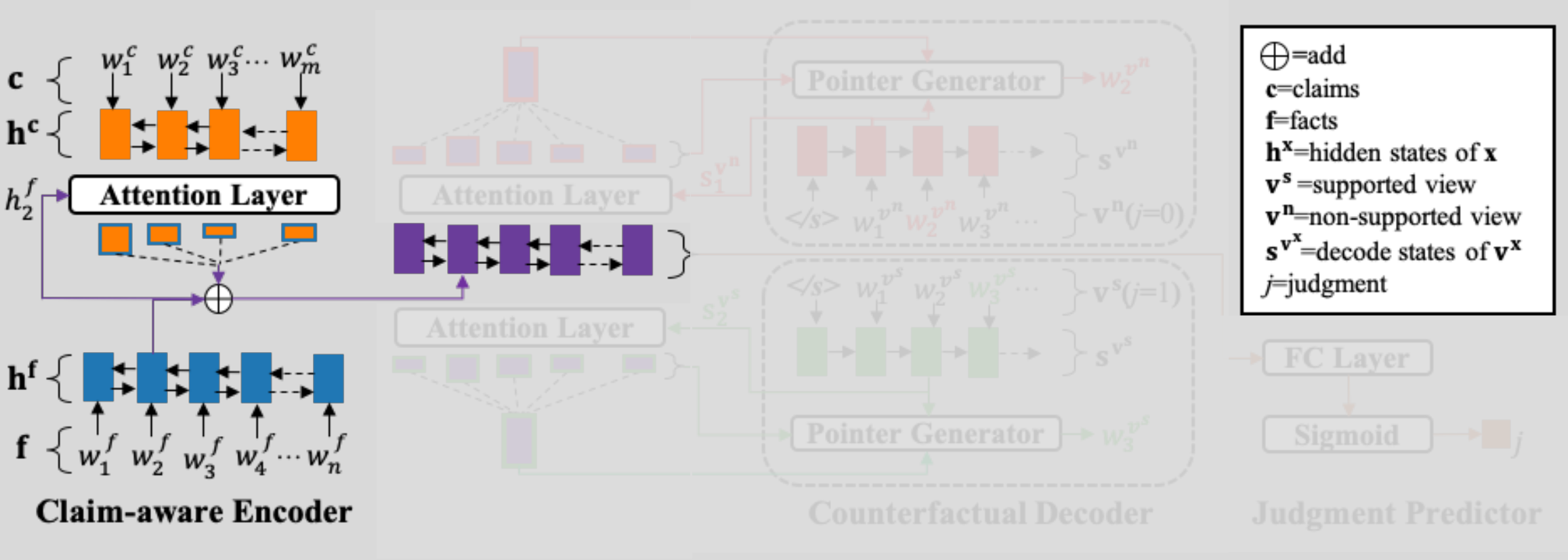
Our Framework



AC-NLG is a multi-task model with:

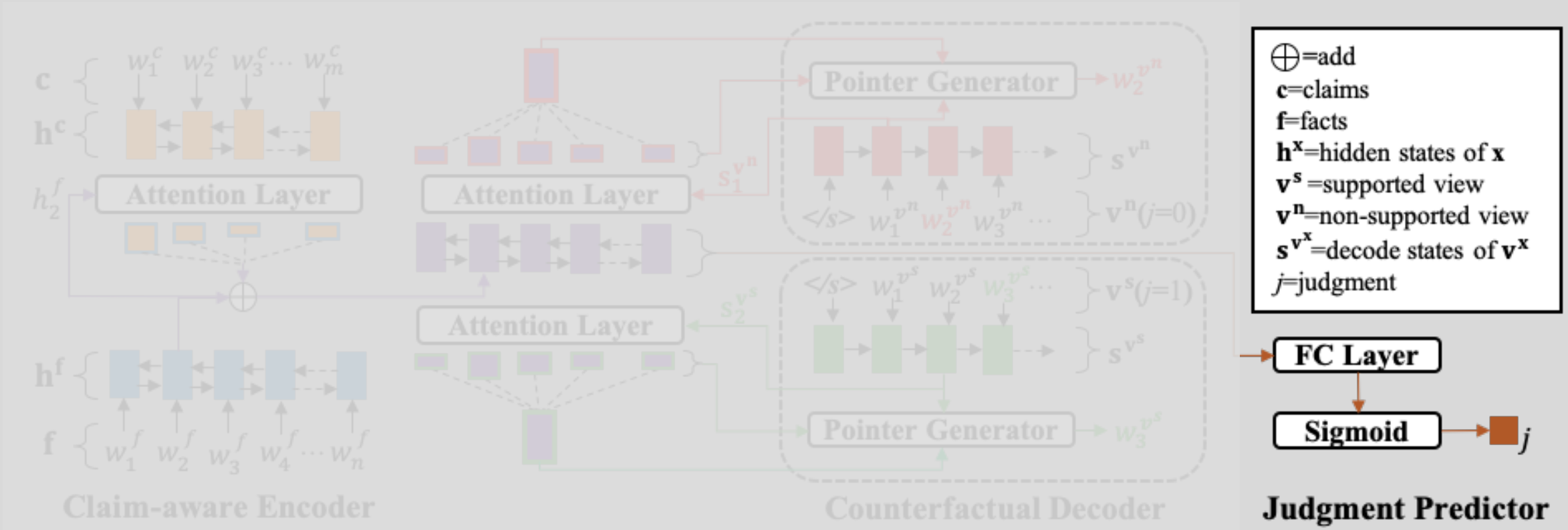
- Claim-aware encoder
  - Claim embedding
  - Fact embedding
  - Claim-Fact attention
- Counterfactual decoders
  - Supportive court's view generation
  - Non-supportive court's view generation
- Judgment predictor

Claim-aware encoder



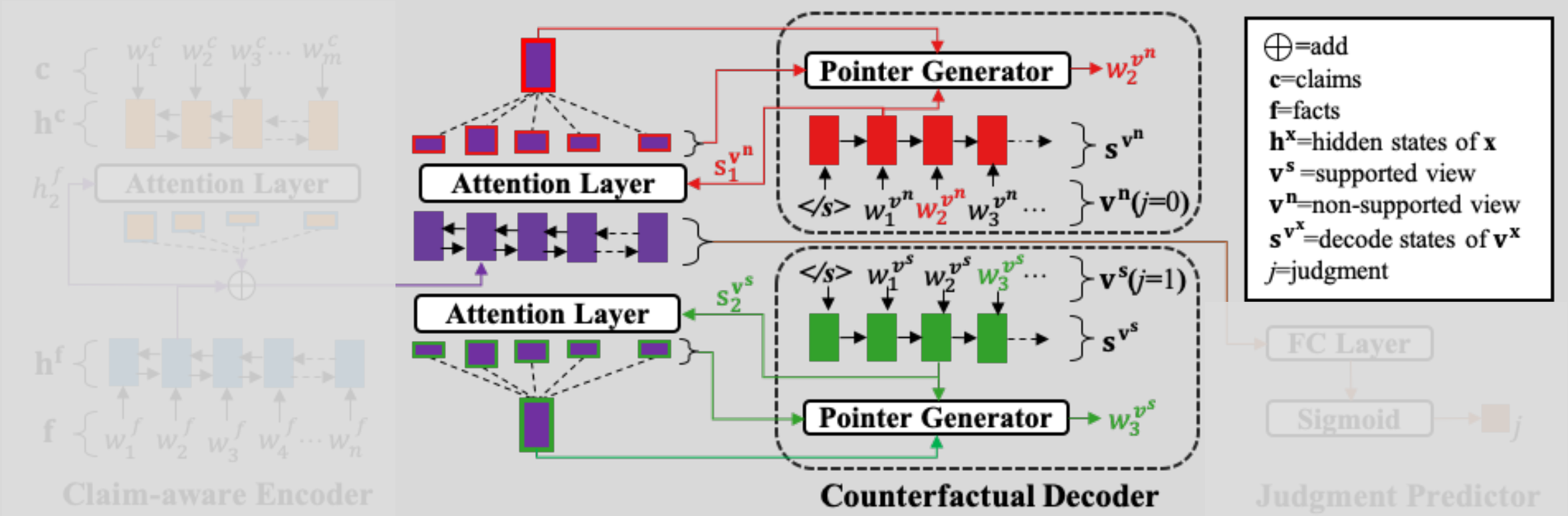
Challenge 1: court's view should only focus on the facts related to the claims

Judgment predictor



$$P(V|do(I)) = P(V|I, j = 0)P(j = 0) + P(V|I, j = 1)P(j = 1)$$

Counterfactual decoders



$$P(V|do(I)) = P(V|I, j = 0)P(j = 0) + P(V|I, j = 1)P(j = 1)$$

Experiment

Dataset Description

- We build a dataset based on raw civil legal documents by following steps:
  - Split legal documents into three parts
  - Human annotation
  - Annotation verification

Type	Result
# Supported case	51087(76%)
# Non-supported case	15817(24%)
Avg. # tokens in claim	77.9
Avg. # tokens in fact	158.0
Avg. # tokens in court's view	194.4

[1] https://wenshu.court.gov.cn/

Metrics

Type	Metric	
Automatic Evaluation	ROUGE	R-1, R-2, R-L
	BLUE	B-1, B-2, B-N
	BERT SCORE	p, r, f1-score
	Acc. of judgment prediction	p, r, f1-score
Human Evaluation	Judgment level, Rational level, Fluency level	

Baselines

Type	Model Name	Description
Comparison Model	S2S	Sequence-to-sequence model
	PGN	Pointer-Generator Network
	S2SwS	Apply oversampling
	PGNwS	Apply oversampling
	AC-NLGwS	Apply oversampling
Ablation Model	AC-NLGw/oD	Remove decoder
	AC-NLGw/oBA	Remove back-door adjustment
	AC-NLGw/oCA	Remove claim-aware attention

Result

Results on court's view generation

Method	ROUGE			BLEU			BERT SCORE		
	R-1	R-2	R-L	B-1	B-2	B-N	p	r	f1
S2S	54.0	35.7	48.3	**65.0**	**57.6**	50.5	89.6	89.5	89.6
S2SwS	51.5	32.0	45.0	63.3	55.6	47.9	83.8	88.8	86.2
PGN	53.3	37.1	48.8	62.0	56.1	50.0	94.0	91.2	92.6
PGNwS	53.2	36.0	48.0	63.1	56.7	50.2	95.7	94.0	94.8
AC-NLGw/oBA	54.1	38.1	49.9	61.8	55.9	49.9	93.6	91.9	92.8
AC-NLGw/oCA	53.7	36.7	49.1	62.1	56.0	49.7	94.5	92.6	93.5
AC-NLGwS	53.7	36.4	48.5	62.8	56.5	50.0	94.0	92.1	93.0
AC-NLG	**55.1**	**38.6**	**50.8**	63.2	57.1	**51.0**	**96.5**	**94.6**	**95.5**

Results on judgment prediction

Method	Prediction Acc.					
	Support			Non-support		
	p	r	f1	p	r	f1
w/oD	72.1	81.0	76.3	56.9	44.3	49.8
w/oCA	92.0	**97.2**	94.5	**85.6**	66.0	74.5
wS	86.0	94.3	90.0	62.8	38.6	47.8
AC-NLG	**93.4**	95.9	**94.6**	81.5	**72.9**	**76.9**

Results of human evaluation

Method	Judgment		Rational	Flu.
	Support	Non-support		
PGN	3.34	1.78	3.11	3.41
AC-NLG	**3.52**	**3.24**	**3.25**	**3.50**

Case Study

PLAINTIFF'S CLAIM	The defendant B return the loan of $495,000. The defendant C return the loan together.
FACT DESCRIPTION	After the hearing, the court held the facts as follows: From November 20, 2010 to March 23, 2011, the defendant B successively borrowed a total of $495,000 from the plaintiff A and issued four separate borrowings. The defendant B has not repaid the above loan.
PGN	The court concluded that: The private lending relationship between Plaintiff A and Defendant B, where the subject was appropriate, the content was legitimate, and the meaning was true, should be confirmed to be legal and valid. The two sides did not agreed in writing on a loan period, so the defendant should return the plaintiff's loan in a timely manner within a reasonable period after the plaintiff urged. The defendant's failure to return the loan timely constituted a breach of contract and should bear corresponding civil liabilities. Therefore, the plaintiff 's claim was reasonable and legal, and the court supported it *Acceptance* .
AC-NLG	The court concluded that the subject of the private lending relationship between Plaintiff A and Defendant B was qualified, the content was legal, and the meaning was true. It should be deemed valid. The two sides did not agreed in writing on a loan period, the defendant shall return the loan within a reasonable period after the plaintiff urged. The plaintiff 's claim requesting the defendant to return the loan of $495,000 was in compliance with the law and the court supported it *Acceptance* . However, the court did not support the claim requesting the defendant C to bear the guarantee liability because the evidence was insufficient *Rejection* .
REAL	The court concluded that: The subject of the private lending relationship between Plaintiff A and Defendant B was qualified, the content was legal, and the meaning was true. It should be deemed valid. Defendant should repay the plaintiff's loan within a reasonable period after the plaintiff urged. Therefore, Defendant B should bear the civil liability of returning the plaintiff's loan of $495,000 and paying overdue interest *Acceptance* . The court did not support the plaintiff's claim requesting the defendant C to return the loan together because the evidence was insufficient *Rejection* . Defendant B failed to appear in court after being legally summoned by the court.

Discussion

Ethical Discussion

- Target User
  - The algorithm is designed **assisting** the trial judges for decision making
  - Should **never** 'replace' human judges
- Potential Error
  - The goal of our algorithm is to generate a draft of court's view for trail judge as a reference
  - Judges need to proofread the content generated from algorithm
- Demographic Bias
  - Gender, race etc.
  - Algorithm adoption should be empowered with de-biased pretraining

Conclusion

- We investigate the problem of de-biased court's view generation in civil cases from a causal perspective, considering the issue of confounding bias from judgment imbalance
- We propose a novel AC-NLG model to jointly optimize a claim-aware encoder and a pair of counterfactual decoders for generating a judgment-discriminative court's view by incorporating with a judgment predictive model
- We construct a dataset based on raw civil legal documents with human annotation on the judgment. To motivate other scholars to investigate this novel but important problem, we make the experiment dataset publicly available (https://github.com/wuyiquan/AC-NLG)

Thanks