Towards Argument-aware Abstractive Summarization of Long Legal Opinions with Summary Reranking

Mohamed Elaraby , Yang Zhong, and Diane Litman

mse30@pitt.edu
yaz118@pitt.edu
dlitman@pitt.edu

Motivation

- Long legal opinions have **implicit argument roles.**
- **Those roles** are vital components of the summaries.
- We propose a **second stage reranking framework** that ranks multiple candidate summaries based on their alignment with the input's argument roles.

Dataset

- CanLII dataset with **1049** legal opinion/summary pairs.
- Both opinions and summaries are annotated for argument roles, using the IRC scheme [1].

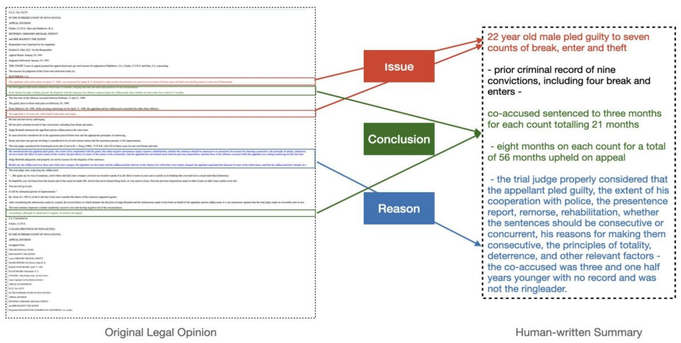
Argument Roles Annotation [1]

The summary is dominated by arguments.

The applicant operated commercial private billiards club with liquor licence and served bar food. The applicant applied to amend the liquor licence to another category to allow public access and slightly longer hours of operation. The Municipality refused to issue a letter to the effect that … would result in a change in use requiring a development agreement. Application granted; There was no evidence that the proposal would result in a nuisance or increased traffic, neither of which could be presumed.

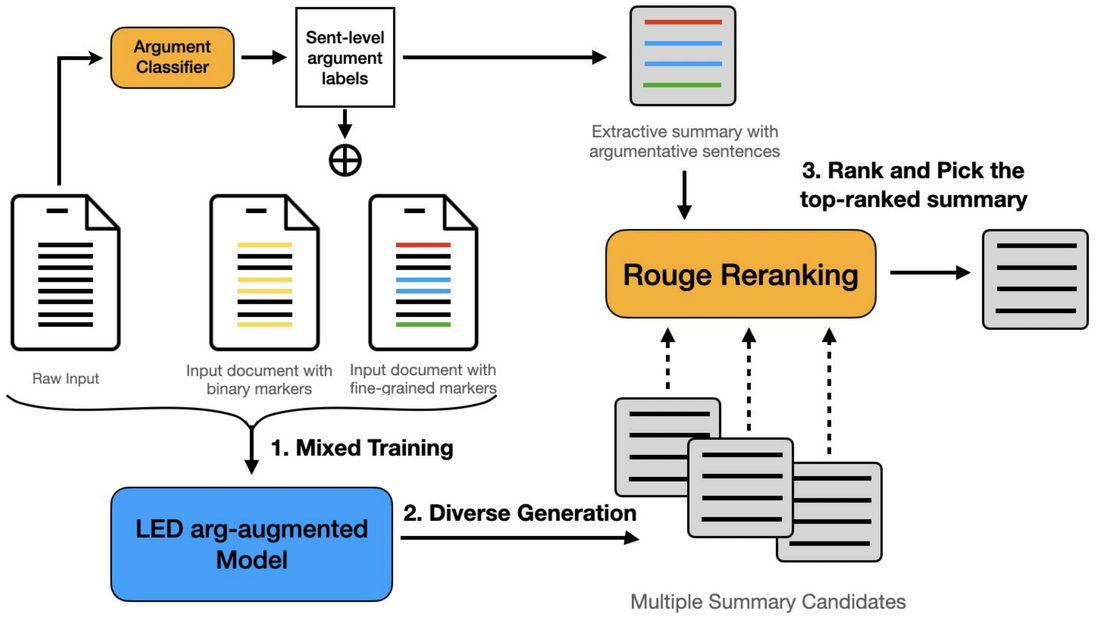
Issue Conclusion Reason

Argument Roles in Opinions



Original Legal Opinion Human-written Summary

Our Proposed Framework and Components



Augmenting Training Data with Arguments (Mixed Training)

Legal opinion	Sent 1 | Sent 2(Issue sent) | Sent 3 | …
Raw	Sent 1 | Sent 2 | Sent 3 | …
+ binary markers	Sent 1 | **<IRC>** Issue sent **</IRC>** | Sent 3 | …
+ fine-grained markers	Sent 1 | **<Issue>** Issue sent **</Issue>** | Sent 3 | …

Three times larger blended training data.

Diverse Generation and Reranking

- We generate summaries using a **diverse beam size of 1 to 5**.
- For **Reranking**, we first employ a sentence-level argument role classifier to extract sentences from opinions with argument roles, treated it as an **extractive summary**.
- We compute the **R-1** between the generated summary and the extractive one.

References

[1] Huihui Xu, Joramir Savelka, and Kevin Ashley: Toward Summarizing Case Decisions via Extracting Argument Issues, Reasons, and Conclusions, ICAIL 2021
[2] Mohamed Elaraby and Diane Litman, ArgLegalSumm: Improving Abstractive Summarization of Legal Documents with Argument Mining, COLING 2022.

Results

- We report our results based on a **5-fold cross validation.**
- We employ **ROUGE** and **BERTScore** to compare our models to previous models.
- We present the system outputs utilizing predicted markers at inference time, more results on gold markers in the paper.

Model	R-1	R-2	R-L	BS
Finetune LED [2]	*47.33*	*22.80*	*46.48*	*86.43*
LED + binary marker [2]	*48.85*	*24.74*	*45.82*	*86.79*
LED + fine-grained markers [2]	*49.02*	*24.92*	*45.92*	*86.86*
Baseline ranking	49.79	25.13	46.63	86.87
LED-arg-augmented model	50.52	24.82	47.19	86.85
LED-arg-augmented model + beam search	**54.13**	**27.02**	**50.14**	**87.38**

- Our framework results **(bolded)**, are significantly better than previously introduced baselines [2] *(italicized)***.**
- Simply ranking among the three baseline models' outputs improves the performance.
- Mixed Training with different input forms (LED-arg-augmented) improves performance.
- Beam search improves results by further diversifying the candidates for ranking over the (LED-arg-augmented) model.

Acknowledgments

CanLII NSF amazon