Integrating Graph based Algorithm and Transformer Models for Abstractive Summarization

Sayed Ayaan Ahmed Sha*, Sangeetha Sivanesan*,
Anand Kumar Madasamy+, Navya Binu+
*National Institute of Technology, Tiruchirappalli
+National Institute of Technology karnataka, Surathkal
405425002@nitt.edu

Abstract

Summarizing legal documents is a challenging and critical task in the field of Natural Language Processing(NLP). On top of that generating abstractive summaries for legal judgments poses a significant challenge to researchers as there is limitation in the number of input tokens for various language models. In this paper we experimented with two models namely BART base model finetuned on CNN DailyMail dataset along with TextRank and `pegasus_indian_legal`, a finetuned version of legal-pegasus on indian legal judgments for generating abstractive summaries for Indian legal documents as part of the **JUST-NLP 2025 - Shared Task on Legal Summarization**. BART+TextRank outperformed `pegasus_indian_legal` with a score of 18.84.

1 Introduction

Legal texts have special characteristics that set them apart from other document. Compared to other domains, legal documents are typically longer and more thorough. They employ a lot of domain-specific jargon, acronyms, and citations or references, and their language is complicated(Akter et al., 2025). Summarization of legal documents is a fundamental and challenging task in legal practice as cases are complex and lengthy in nature(Shukla et al., 2022). A majority of India's population lacks a strong command of the English language (Datta et al., 2023) and it is difficult for a layman to comprehend the complex structure of a judgment. Thus, it is crucial to summarize legal documents in plain and understandable english. Legal text summarization focuses on extracting and highlighting the essential points of a legal document in a brief and clear manner, enabling quick decision-making. Creating technologies that can handle documents from specific nations or languages is one of the targets of region-specific legal summarization(Akter et al., 2025). Region-specific legal documents, summarization strategies, and summarization methodologies are the three basic components of legal summarization.

General summarization strategies identified in legal summarization involve extractive, abstractive, and hybrid approaches. The extractive method(Cheng and Lapata, 2016) involves directly copying significant sentences from the source document and combining them to create the output summary. Meanwhile the abstractive strategy(Rush et al., 2015) mimics human understanding by interpreting the source document and producing a summary based on its important concepts. By reconstructing a summary using certain key information taken from the original document, the hybrid technique seeks to combine the advantages of both approaches. However, there are a variety of summarization techniques, including rank-based, graph-based, transformer-based, and others.

2 Related Work

Wide range of solutions have been proposed for the summarization of indian as well as foreign legal documents. Among the extractive family, there are unsupervised summarization approaches such as Reduction(Jing, 2000) and graphical approach LexRank(Erkan and Radev, 2004). SummaRuNNer(Nallapati et al., 2017) and BERTSum(Liu and Lapata, 2019) are two supervised neural summarizers that approach document summarization as a binary classification issue (in-summary vs. out-of-summary). Some abstractive models include (Lewis et al., 2020) and (Zhang et al., 2020). Generating abstractive summaries specifically for indian legal documents, (Shukla et al., 2022) created the IN-Abs corpus containing 7,130 case documents, together with their headnotes/summaries. IL-TUR(Joshi et al., 2024) a benchmark for legal tasks, included the SOTA models for indian legal

Proceedings of the 1st Workshop on NLP for Empowering Justice (JUST-NLP 2025), pages 186–190
December 24, 2025

text summarization. One of the early works in legal text summarization include identification of the thematic structure to find the argumentative themes of the judgement(Farzindar and Lapalme, 2004). The relevant sentences were extracted for each theme and presented as a table-style summary. (Verma et al., 2022) suggested a two-step method in which sentences were first clustered by similarity using a partitional technique; key sentences from each cluster were then chosen based on text feature scores, and similarity was measured using a linear combination of normalised Google distance and word mover's distance.

3 Dataset

The **InLSum**(Indian Legal Summarization) dataset was provided by the organizers of the workshop. It had 3 splits - train (1200 datapoints), validation (200 datapoints), and test (400 datapoints). Each entry correspond to one court judgment with a unique ID and with two files for each split (train, validation, test): each split contained the full text judgments and the other contained the gold(reference) summaries. Each summary is an abstractive human-written summary of the respective judgment.

4 Task and Evaluation

The task focused on generating abstractive summaries for the indian court judgments in English. Participants were required to train language models that can read legal judgments and produce concise, coherent, and fluent summaries of approximately 500 words. The models were trained on the training split of the dataset and the predictions for the validation set and train set were submitted for evaluation.

The submitted predictions were evaluated by the organizers using three standard metrics: (i) ROUGE-2, (ii) ROUGE-L (Lin, 2004) and (iii) BLEU(Papineni et al., 2002).

5 Experimentation

For producing summaries for the given dataset, we experimented with two approaches: (i) We use BART (Lewis et al., 2020) base model finetuned on DailyMail dataset[1] along with TextRank(Mihalcea and Tarau, 2004) for generating the desired output summaries. Bart leverages a standard seq2seq architecture with a bidirectional encoder(like BERT(Devlin et al., 2019)) and a left-to-right decoder (like GPT(Radford et al., 2018)). BART has proven to produce effective results when finetuned for text generation tasks. TextRank(Mihalcea and Tarau, 2004) is a graphical unsupervised extractive summarization strategy. It leverages embeddings for generating the similarity scores between the sentences and stores them in a matrix. The similarity matrix is then converted into a graph, with sentences acting as vertices and the scores as edges for computing the sentence rank. (ii) `pegasus_indian_legal`[2], a finetuned version of legal-pegasus on indian legal judgments. The legal-pegasus is also a further finetuned version of Pegasus(Zhang et al.). We have provided the github link for the codes and experiments performed for this paper[3]

5.1 Training and Validation phase

For the first approach, we first finetuned the BART-base-cnn model on the training data and then used the TextRank algorithm to extract top ranked sentences from the validation set. We had input token constraint of 1024 and assigned generated summary length to maximum of 768 tokens. To generate the embeddings for the judgments, we utilized all-MiniLM-L6-v2, a sentence transformer model. For each judgment, we compute a similarity matrix from the sentence embeddings for it's constituent sentences. The similarity scores were computed using cosine similarity. We select top n sentences for each judgment based on the scores using TextRank and the resultant sentences were fed to the model for generating the abstractive summaries. To choose the optimal n value, we first calculated the average number of sentences in the reference summaries of train set and test set and found it to be approximately 20. We experimented with n values 20 and 30. When n = 30, the results for the validation set were slightly lower when n is 20. The n value we finally opted was 20.

In the second approach, we finetuned the pegasus_indian_legal model on the train dataset provided by the organizers and ran the finetuned model on the validation set to generate the summary. Here also the maximum input token sequence is 1024 tokens and the summary length was intilalized to

[1] https://huggingface.co/ainize/bart-base-cnn

[2] https://huggingface.co/akhilm97/pegasus_indian_legal

[3] https://github.com/Ayaan2123/Shared_Task-2025

maximum of 768 tokens.

During the training phase, the HuggingFace implementation of BART/pegasus_indian_legal applies automatic truncation when the input exceeds the maximum length of the models. Thus, for input sequences longer than 1024 tokens, the models only used the first 1024 tokens.

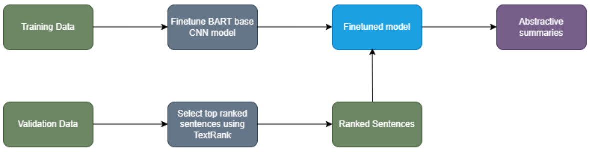

Figure 1: Summary Generation Workflow

5.2 Testing Phase

The task was to generate abstractive summaries of approximately 500 words for the test data provided by the organizers. The respective finetuned models were finally tested using the test set.

6 Evaluation and Analysis

The organizers used three set of standard metrics for evaluating the performance of the summarization models: (i) ROUGE-2 (ii) ROUGE-L (Lin, 2004) and (iii) BLEU(Papineni et al., 2002). Our work finished 9th in the final rankings with an average score of 18.84

BART+TextRank proved to be the best performer among the two approaches we incorporated for the desired results. Table 1 gives the performance of the models on the validation as well as test set. BART along with TextRank performs better on validation data while it drops slightly in the testing phase. The first approach still yields better results than PEGASUS_INDIAN_LEGAL. Appendix A provides sample summaries generated for a particular judgment using both the approaches.

Algorithm	R-2	R-L	BLEU	Average
Performance during validation phase				
BART + TEXTRANK	21.37	22.42	15.64	19.81
PEGASUS_INDIAN_LEGAL	14.39	23.08	9.19	15.55
Performance during testing phase				
BART + TEXTRANK	20.37	22.49	13.67	18.84
PEGASUS_INDIAN_LEGAL	15.64	23.27	11.75	16.89

Table 1: Performance comparison of summarization methods.

7 Conclusion

Summarizing legal documents is a critical task, facing significant challenges due to the volume and complexity of legal documents. In this paper we tested two models for generating the abstractive summaries for Indian legal judgments and analyzed their performance. There is a lot of room for future development and progress in the area of legal document summarization.

Limitations

Our work has several limitations that should be considered during the results analysis. Both BART and PEGASUS models are constrained by a maximum input length of 1024 tokens, while the legal judgments in the dataset are significantly longer. In the first approach (BART+TextRank), we partially mitigated this problem during inference by using an extractive then abstractive pipeline. However, in Approach 2 (PEGASUS_INDIAN_LEGAL), the model relied on the truncation behavior of the tokenizer, which limits the model's ability to reason over the entire judgment. This disparity in long-context handling likely contributed to the performance differences between the two approaches.

Acknowledgments

We extend our gratitude to the organizers of the **JUST-NLP 2025 - Shared Task on Legal Summarization** for providing the data and evaluation resources. We also thank our team members for their assistance and valuable contributions.

References

Mousumi Akter, Erion Çano, Erik Weber, Dennis Dobler, and Ivan Habernal. 2025. A comprehensive survey on legal summarization: Challenges and future directions. *arXiv preprint arXiv:2501.17830*.

Jianpeng Cheng and Mirella Lapata. 2016. Neural summarization by extracting sentences and words. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 484–494, Berlin, Germany. Association for Computational Linguistics.

Debtanu Datta, Shubham Soni, Rajdeep Mukherjee, and Saptarshi Ghosh. 2023. MILDSum: A novel benchmark dataset for multilingual summarization of Indian legal case judgments. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 5291–5302, Singapore. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the*

North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers), pages 4171–4186.

Günes Erkan and Dragomir R Radev. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of artificial intelligence research*, 22:457–479.

Atefeh Farzindar and Guy Lapalme. 2004. Letsum, an automatic legal text summarizing. *Legal knowledge and information systems: JURIX*, page 11.

Hongyan Jing. 2000. Sentence reduction for automatic text summarization. In *Sixth Applied Natural Language Processing Conference*, pages 310–315, Seattle, Washington, USA. Association for Computational Linguistics.

Abhinav Joshi, Shounak Paul, Akshat Sharma, Pawan Goyal, Saptarshi Ghosh, and Ashutosh Modi. 2024. IL-TUR: Benchmark for Indian legal text understanding and reasoning. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11460–11499, Bangkok, Thailand. Association for Computational Linguistics.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Yang Liu and Mirella Lapata. 2019. Text summarization with pretrained encoders. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3730–3740, Hong Kong, China. Association for Computational Linguistics.

Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 404–411, Barcelona, Spain. Association for Computational Linguistics.

Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. 2017. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, and 1 others. 2018. Improving language understanding by generative pre-training.

Alexander M. Rush, Sumit Chopra, and Jason Weston. 2015. A neural attention model for abstractive sentence summarization. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 379–389, Lisbon, Portugal. Association for Computational Linguistics.

Abhay Shukla, Paheli Bhattacharya, Soham Poddar, Rajdeep Mukherjee, Kripabandhu Ghosh, Pawan Goyal, and Saptarshi Ghosh. 2022. Legal case document summarization: Extractive and abstractive methods and their evaluation. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1048–1064, Online only. Association for Computational Linguistics.

Pradeepika Verma, Anshul Verma, and Sukomal Pal. 2022. An approach for extractive text summarization using fuzzy evolutionary and clustering algorithms. *Applied Soft Computing*, 120:108670.

J Zhang, Y Zhao, M Saleh, and PJ Liu. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. arxiv 2019. *arXiv preprint arXiv:1912.08777*.

Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter Liu. 2020. PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 11328–11339. PMLR.

Appendix

A Comparison of generated summaries

BART+TextRank	PEGASUS_INDIAN_LEGAL
A proposed construction project at NOIDA which did not take off from the drawing board has given rise to proceedings under Article 32 of the Constitution by twenty-five purchasers of commercial premises. The Court made the following observations:The writ petition requires the Court to step into the construction project and ensure that it is duly completed. This would be beyond the remit and competence of the Court under the Constitution."It would be inappropriate for this Court to assume the jurisdiction to supervise the due completion of a construction project especially in facts such as those presented in the present case. This will inevitably draw the Court into the day-to-day supervision of the project, including financing, permissions, and execution—something which lies beyond the ken of judicial review and the competence of a court," the Court said. "The Court must confine itself to its core competencies which consist in the adjudication of disputes amenable to the application of legal standards," the Court added.It added that the writ petition is disposed of, subject to the aforesaid liberty.It was also suggested that the decision of a public authority entrusted with a public duty is amenable to judicial review. But it is quite another hypothesis to postulate that the decision-making authority should be taken over by the Court. The latter is impermissible." The petitioners have sought, inter alia, a writ, order or direction to the respondents to protect the interests and investments of customers/buyers in the larger public interest.A petition application, if any, stands disposed of, the Court noted.In the case, the petitioners had sought to pursue remedies available at law, including the Consumer Protection Act 1986, the Real Estate (Regulation and Development) Act 2016, and the Insolvency and Bankruptcy Code 2016.On the other hand, the Court noted that several provisions of law confer statutory rights on buyers of real estate and invest them with remedies enforceable at law.There was no mention of the clause in the Constitution, which provides for a statutory regime to protect purchasers and created fora which are entrusted with decision-making power.However, it was contended that a decision of a public authority entrusted with public duty was amenable to judicial review, but it is impermissible to take over the decision-making authority. The petitioners were represented by Shiv Kumar Suri and Nikita Thapar.While the Court was hearing the matter, it said that it was not appropriate to assume jurisdiction of the Court under Article 32 of the Constitution."We leave it open to the petitioners to pursue the remedies available in law," the Court said.The writ petition was disposed of and the petition application stood disposed of. The Court added that it would not be appropriate for the Court to take jurisdiction into day-to-day supervision of such a project.The Court added that a writ petition filed by the petitioner was disposed of subject to liberty.Petitioners also submitted that the petition filed against the petitioner stood dismissed and the case was adjourned until the matter was heard by the Court on Thursday."A decision of the petitioner would not be disposed of if the petition was filed before the Court. It was submitted by the counsel of Vinishma Kaul."	The Delhi High Court recently dismissed a petition filed by a group of real estate investors seeking a court-monitored probe into a proposed construction project at NOIDA.A bench of Chief Justice of India (CJI) Dipak Misra and Justice Sanjay Kishan Kaul said that managing a construction project is not within the jurisdiction of the Court under Article 32 of the Constitution. "It would be inappropriate for this Court to assume the jurisdiction to supervise the due completion of a construction project especially in facts such as those presented in the present case.This will inevitably draw the court into the day-to-day supervision of the project, including financing, permissions and execution – something which lies beyond the ken of judicial review and the competence of the court.The court must confine itself to its core competencies which consist in the adjudication of disputes amenable to the application of legal standards.We, therefore, leave it open to the petitioners to pursue the remedies available in law," the Court said.The Court was hearing a petition filed by a group of real estate investors seeking a court-monitored probe into a proposed construction project at NOIDA. The petitioners claimed that the project did not take off from the drawing board and had given rise to proceedings under Article 32 of the Constitution by twenty five purchasers of commercial premises.They sought a writ, order or direction to the respondents to protect the interests and investments of customers/buyers in the larger public interest; (ii) a writ, order or direction for the revival of the project failing which the amounts invested by the petitioners be returned with interest at the rate of 18 % per annum; and (iii) a court-monitored probe.The Court was of the view that the exercise of jurisdiction under Article 32 of the Constitution would not be warranted in the facts of the present case."The writ petition requires the Court to step into the construction project and to ensure that it is duly completed. This would be beyond the remit and competence of the Court under Article 32. Managing a construction project is not within the jurisdiction of the court. Several provisions of law confer statutory rights on purchasers of real estate and invest them with enforceable remedies at law.Several provisions of law confer a statutory regime to protect the rights of purchasers of real estate and created fora which are entrusted with decision-making authority," the Court said."A decision of a public authority which is entrusted with a public duty is amenable to judicial review.But it is quite another hypothesis to postulate that the decision-making authority should be taken over by the court.The latter is impermissible.The latter is impermissible," the Court said.The Court was of the view that managing a construction project is not within the remit and competence of the Court under Article 32 of the Constitution."The court must confine itself to its core competencies which consist in the adjudication of disputes amenable to the application of legal standards.We, therefore, leave it open to the petitioners to pursue the remedies available in law," the Court said.The Court was hearing a petition filed by Shikhil Suri, Shiv Kumar Suri, Madhu Suri, Shilpa Saini, Nikita Thapar, Vinishma Kaul and Vinishma Kaul.

Table 2: Summary comparison of the two methods.