NCE Journal of Science and Engineering (NJSE), Vol. 2, June 2025

A Stacked Transformers-based Nepali News Title Generation

Pawan Acharya[1,*] and Subarna Shakya[2,*]

[1]Nepal College of Information Technology, Lalitpur, Nepal
[1]Institute of Engineering, Tribhuvan University, Lalitpur, Nepal
[*]Correspondence: pawanacharya0123@gmail.com and drss@ioe.edu.np
Manuscript received January 8, 2025; accepted April 25, 2025

***Abstract*—News organizations face the challenge of generating engaging and informative titles for their articles on a daily basis. This task is difficult since it requires a deep understanding of the content, the audience, and the cultural context. In recent times, many approaches have been developed that have proven to be state-of-the-art for news title generation. However, a comparative analysis of these approaches for low-resource languages like Nepali is currently missing. This paper proposes a Nepali news title generation system based on transformers architecture. The proposed system uses a stacked approach, which includes generation and refinement modules. The generation module generates a candidate title according to the given Nepali news article. The generated candidate title is dispatched to the refinement module in order to output the syntactically refined title. At last, BLEU and ROUGE scores of the titles that would have been generated by the refinement model against the reference title are calculated in order to know its performance. This dataset can be collected from some popular Nepali news portals through web scraping for training transformer-based models in the Nepali language.**

***Keywords*—Low-resource language, transformers, ROUGE, BLEU, and web scraping.**

I. INTRODUCTION

GENERATING effective news headlines is a specific text summarization challenge crucial in NLP, especially with the growing volume of textual data from news stories, blog posts, and social media. The rise of digital media has led to an overwhelming number of news stories from diverse sources, necessitating advanced, scalable algorithms. Headlines must be concise, informative, and engaging, capturing complex sentence structures and contextual references. While extractive methods often lack clarity and cohesion, abstractive approaches can generate coherent, readable summaries by creating new language and paraphrasing information. This study utilizes a stacked design with transformer-based modules for generation and refinement, employing self-attention layers and contextual word embeddings to produce high-quality headlines.

II. DATA COLLECTION AND PREPROCESSING

A novel corpus of Nepali news was created for this research by scraping articles and titles from various Nepali news web portals, as no suitable dataset with Nepali text and summaries was available.

Data preprocessing involves converting raw data into a useful, efficient, and understandable format by removing noisy information such as stop words, extra white space, special characters, emoticons, symbols, and non-Nepali characters. This cleaning process enhances model performance. A comprehensive dataset of 224,500 news article-title pairs was compiled from prominent Nepali news portals. Initially, 120,000 pairs were used to train the generative model. An additional 100,000 pairs were generated by passing new articles through the trained model, creating a refined dataset for further improvement. The remaining 4,500 pairs were reserved for unbiased evaluation of both the generative and refinement models.

II. METHODOLOGY

The proposed model for title generation consists of two modules: Generation and Refinement, followed by a Scoring function. The article body is fed to the Generation module to produce a candidate title. This candidate title is then passed to the Refinement module, which de-noises and improves it syntactically and semantically. Finally, the Scoring function evaluates titles from both modules using BLEU and Rouge scores.

A. Generative Models

Transformers are used in the text-to-text framework for title generation by concatenating inputs and targets with a separator. Since titling is a sub-task of summarization, transformers are effective for this task. Each dataset pair includes an article xi and its corresponding title yi. Articles are tokenized and embedded into vectors using FastText embeddings in the encoder. The lengths of the input article and target title sequences, denoted as a and b, respectively, define the dimensions of the embeddings and positional encoding for each token.

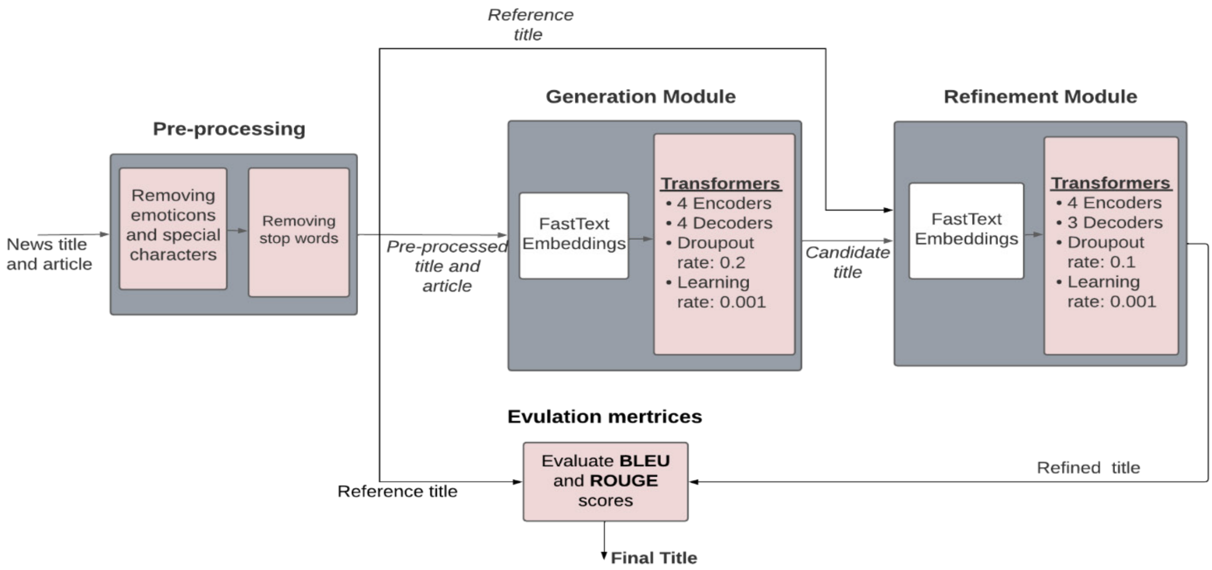

Fig. 1. Proposed block diagram with pre-processing block followed by generative and refinement modules and an evaluation module

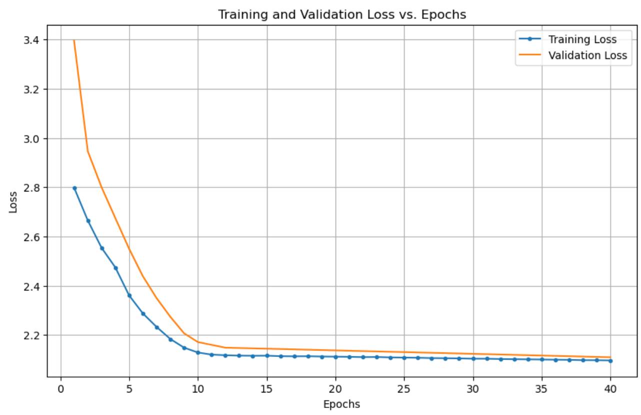

Fig. 2. Generative Model cross-entropy loss over the number of epochs.

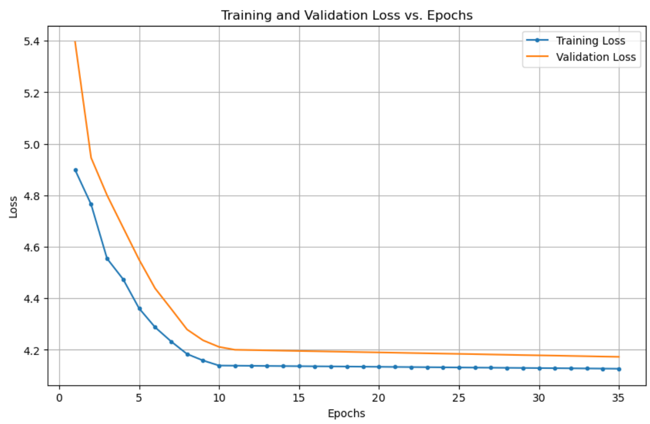

Fig. 3. Refined Model cross-entropy loss over the number of epochs.

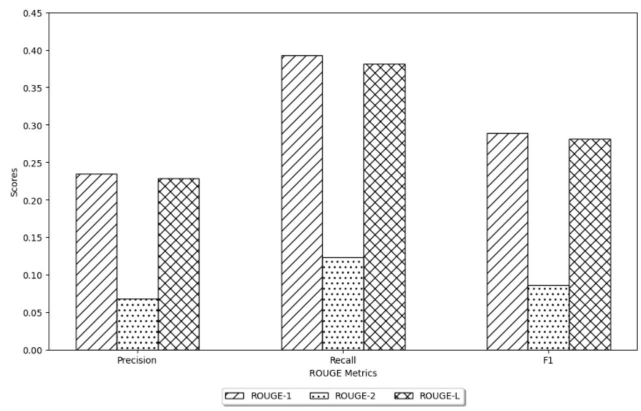

Fig. 4. Generative Model ROUGE scores bar graph

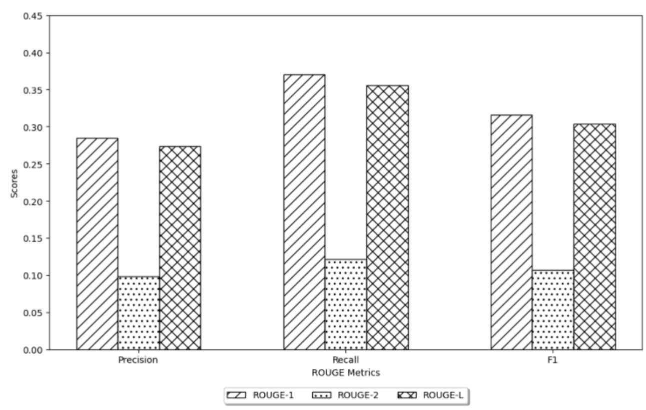

Fig. 5. Refinement Model ROUGE scores bar graph

B. Refinement Module

For training the Refinement model, a specialized dataset was created using outputs from the Generation module. Sample titles generated by the Generation module served as inputs, while the corresponding original titles were used as ground truth. This approach trains the Refinement model to improve and correct the generated titles, aligning them with high-quality originals. This method enhances syntactical coherence and relevance, creating a robust feedback loop that iteratively improves title quality. The Refinement model focuses on minimizing discrepancies between generated and original titles, using a similar architecture to the Generation model to fine-tune and polish titles for better accuracy and readability.

The evaluation function calculates the relevancy of titles from the Refinement module against reference titles using ROUGE and BLEU metrics. ROUGE measures Recall, Precision, and F1, focusing on n-gram overlap, while BLEU measures precision. ROUGE's multiple indices and recall focus reflect important information, and its forms like ROUGE-N and ROUGE-L offer comprehensive evaluation. BLEU provides a quantitative measure of precision and is language-independent, making it suitable for multilingual tasks. For final judgment, the BLEU score is used due to its simplicity and efficiency.

During the evaluation phase, the generative transformers model achieved a validation BLEU score of 0.2384 across 4,500 validation samples, indicating a substantial alignment with human-generated summaries. This score reflects the model's ability to capture relevant information and linguistic nuances. The ROUGE evaluation scores offer a detailed assessment of the model's performance. For unigram (ROUGE-1), the recall was 0.3928, precision 0.23475, and F1-score 0.28890. For bigram (ROUGE-2), the recall was 0.12359, precision 0.068405, and F1-score 0.085924. The ROUGE-L metric, which measures the longest common subsequence (LCS), was also included in the evaluation.

During evaluation, the refined transformers model achieved a validation BLEU score of 0.2612 across 4,500 samples, indicating strong alignment with human-generated summaries and capturing relevant information and linguistic nuances. The ROUGE scores provide a comprehensive performance overview. For unigram (ROUGE-1), the model had a recall of 0.43729, precision of 0.26203, and F1-score of 0.32208. For bigrams (ROUGE-2), the recall was 0.15448, precision 0.08551, and F1-score 0.1074. The ROUGE-L metric, which evaluates the longest common subsequence (LCS), showed a recall of 0.42331, precision of 0.25413, and F1-score of 0.31211.

V. Conclusion

In conclusion, the stacked model architecture significantly outperforms the previous single model in title generation tasks. Transformer-based techniques have enhanced the quality and efficiency of summaries by better capturing linguistic patterns and dependencies. The stacked model generates more coherent and informative summaries. Future work should explore refining this architecture further, possibly by using the first model to identify key sentences and the second to generate titles, which could offer promising improvements, especially in the Nepali domain.

V. Acknowledgement

I extend my deepest gratitude to my supervisor, Prof. Dr. Subarna Shakya, for his invaluable guidance, support, and encouragement throughout this research. His expertise and dedication have been instrumental in shaping the direction and quality of this work.

I also thank the Nepal College of Information Technology for providing the essential resources and facilities needed for this research. Additionally, I appreciate my friends and colleagues for their valuable insights and moral support, which made this challenging journey more rewarding.

References

[1] Alexey Bukhtiyarov and Ilya Gusev. Advances of transformer-based models for news headline generation. In Artificial Intelligence and Natural Language: 9th Conference, AINL 2020, Helsinki, Finland, October 7–9, 2020, Proceedings 9, pages 54–61. Springer, 2020.

[2] Diksha Khurana, Aditya Koli, Kiran Khatter, and Sukhdev Singh. Natural language processing: State of the art, current trends and challenges. Multimedia tools and applications, 82(3):3713–3744, 2023.

[3] Yashar Mehdad, Giuseppe Carenini, and Raymond Ng. Abstractive summarization of spoken and written conversations based on phrasal queries. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1220–1230, 2014.

[4] Chandra Khatri, Gyanit Singh, and Nish Parikh. Abstractive and extractive text summarization using document context vector and recurrent neural networks. arXiv preprint arXiv:1807.08000, 2018.

[5] K Vimal Kumar and Divakar Yadav. An improvised extractive approach to hindi text summarization. In Information Systems Design and Intelligent Applications: Proceedings of Second International Conference INDIA 2019, Volume 1, pages 291–300. Springer, 2019.

[6] K Raj Mishra, Jayshree Rathi, and Janardan Banjara. Encoder decoder based nepali news headline generation. International Journal of Computer Applications, 175:975–8887, 2020.

[7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. Advances in neural information processing systems, 30, 2017.

[8] Prakhar Mishra, Chaitali Diwan, Srinath Srinivasa, and G Srinivasaraghavan. Automatic title generation for text with pre-trained transformer language model. In 2021 IEEE 15th International Conference on Semantic Computing (ICSC), pages 17–24. IEEE, 2021.

[9] Bipin Timalsina, Nawaraj Paudel, and Tej Bahadur Shahi. Attention based recurrent neural network for nepali text summarization. Journal of Institute of Science and Technology, 27(1):141–148, 2022.

[10] Ming Zhong, Pengfei Liu, Yiran Chen, Danqing Wang, Xipeng Qiu, and Xuanjing Huang. Extractive summarization as text matching. arXiv preprint arXiv:2004.08795, 2020.

[11] Chitran Pokhrel and Rameshwar Adhikari. Automatic extractive text summarization for text in nepali language with bidirectional encoder representation transformers and k-mean clustering.