nepBART: Nepali Bi-Directional Autoregressive Transformer

1. Dataset

For this project, I created scripts to extract news data from 30 different news portals. I was able to extract around 2.2M training dataset along with a 250K evaluation set.

2. Model

The model I chose is Facebook's Encoder-Decoder Transformer, BART. The model parameters have been scaled down to ~61.5 M.

2.1 Hyperparameters

Hyperparameter	Value
Vocabulary Size	32,768
Sequence Length	512
Embedding Dimension	512
Encoder Layers	6
Decoder Layers	6
Encoder FFN Dimension	2048
Decoder FFN Dimension	2048
Dropout	0.1
Epochs	5
Batch Size	32
Warmup Steps	5,000
Learning Rate	1e-4
L2 Regularization	0.01

2.2 Training Objective

The training objective is token infilling as discussed in the original BART paper.

2.3 Loss Function

The model is trained with the Cross-Entropy loss function.

Please visit the GitHub repo to access the scraping, cleaning, and training scripts.
The model is still being trained, and the report will be continuously updated.