Review Summarization using GPT2

Overview:

In this project, the goal is to utilize the Amazon Fine Food Reviews dataset for training a GPT-2 model to generate summaries of reviews. The process involves cleaning and preprocessing the text data, dividing the dataset into training and testing sets, implementing a custom dataset class for training, fine-tuning the GPT-2 model, experimenting with different hyperparameters, and evaluating the model's performance using ROUGE scores.

Approach:

Data Preprocessing:
Clean and preprocess the 'Text' and 'Summary' columns from the dataset by converting text to lowercase, removing punctuation, tokenization, removing stopwords, and lemmatization.
Model Training:
Initialize a GPT-2 tokenizer and model from Hugging Face.
Divide the dataset into a training set (75%) and a testing set (25%).
Implement a custom dataset class to prepare the data for training.
Fine-tune the GPT-2 model on the review dataset to generate summaries. Experiment with different hyperparameters such as learning rate, batch size, and number of epochs to optimize performance.
Evaluation:
After training, compute ROUGE scores on the test set to assess the model's overall performance.
Compare the generated summaries with the actual summaries using ROUGE metrics (ROUGE-1, ROUGE-2, ROUGE-L).
Methodologies:

Data Cleaning and Preprocessing:
Lowercasing the text, removing punctuation, tokenization, removing stopwords, and lemmatization are performed using NLTK library functions.
These steps help in standardizing the text data and removing noise, making it suitable for model training.
Model Training:
PyTorch and the Transformers library are used for model training.
The GPT-2 model is fine-tuned on the review dataset using the Adam optimizer and cross-entropy loss.
Hyperparameters such as learning rate, batch size, and number of epochs are tuned to optimize the model's performance.
The training process involves feeding input sequences to the model and predicting the next token in the sequence.
Evaluation:
ROUGE (Recall-Oriented Understudy for Gisting Evaluation) scores are computed to evaluate the quality of the generated summaries.
ROUGE metrics measure the overlap between the generated summary and the reference summary in terms of n-grams (unigrams, bigrams, and longest common subsequences).
Assumptions:

The Amazon Fine Food Reviews dataset is assumed to be representative of the review data, and the quality of summaries generated by the model will depend on the quality and diversity of the dataset.
The GPT-2 model architecture is assumed to be suitable for the task of review summarization, and fine-tuning it on the dataset will improve its performance.
The ROUGE scores are assumed to provide a reasonable evaluation of the model's performance, although they may not capture all aspects of summary quality.
Results:

The results of the model training and evaluation process will include:

Generated summaries for a subset of reviews from the test set.
ROUGE scores (ROUGE-1, ROUGE-2, ROUGE-L) indicating the precision, recall, and F1-score of the generated summaries compared to the actual summaries.
Experimentation with different hyperparameters to find the optimal configuration for the model.
Insights into the strengths and limitations of the model based on the evaluation results.
By following this approach, we aim to build an effective review summarization model using GPT-2 that can provide concise and informative summaries of Amazon Fine Food Reviews.