Comparative Analysis of Text Generation Models to Generate News Headline

Group 25

Paarthvi Sharma
Prajwal Shenoy
Sanath Naik
Saiteja Kura

Introduction

- With the rapid advancement of automatic text generators and various transformer models, numerous real-world applications have emerged, streamlining processes and reducing manual labor.

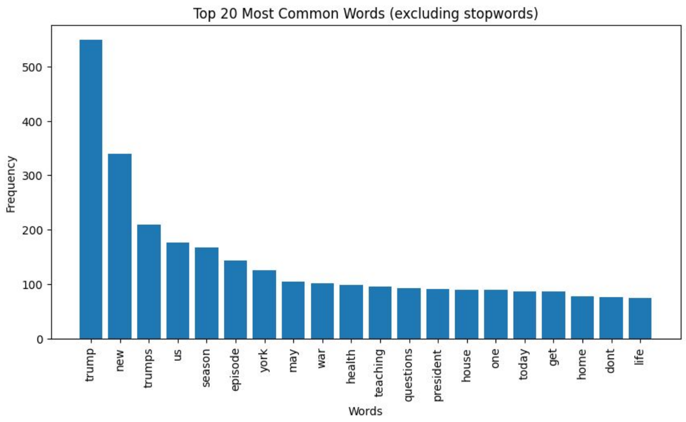
- Our project involves using sequence-to-sequence models to generate headlines, employing models such as LSTM, Bi-directional LSTM, GRUs, and the GPT-2 large model.

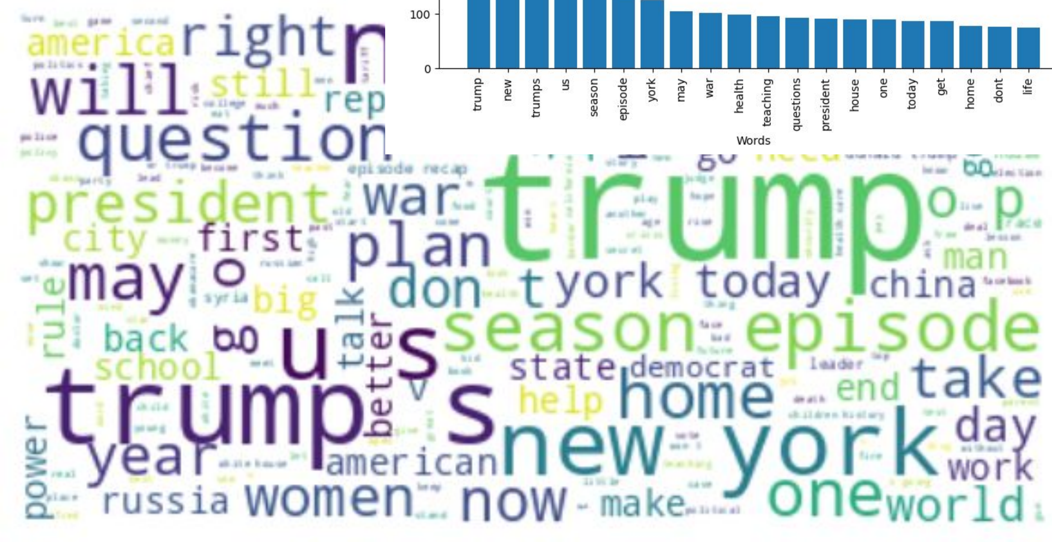
- We do a comparative analysis of the performances of all these models to understand their strengths and limitations in different contexts. Additionally, we also assess their efficiency and accuracy in generating contextually relevant and engaging headlines

Dataset and Preprocessing

- **Dataset Utilization:** Used the New York Times Article and Comment dataset to generate news headlines. This dataset comprises around 16 attributes and includes more than 9,000 articles.

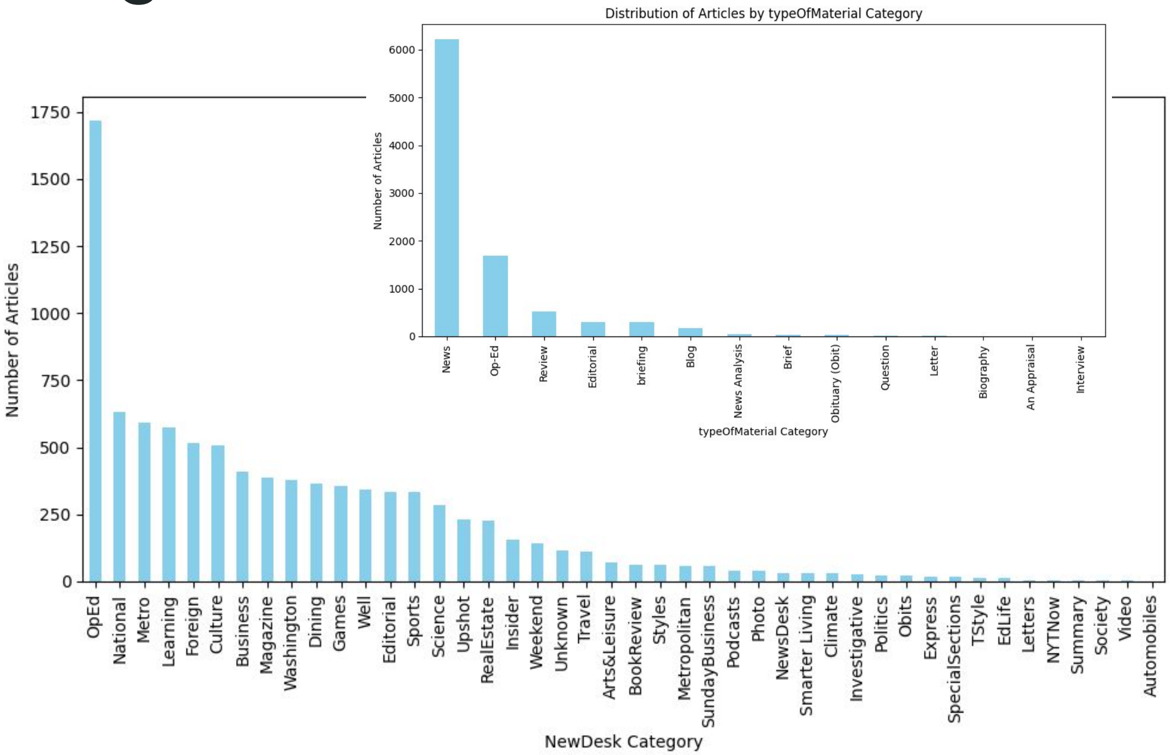
- **Data Cleaning:** Excluded articles that had an 'Unknown' heading to maintain the quality and relevance of the dataset used for model training.





Dataset and Preprocessing

- **Preprocessing :** Implemented general preprocessing steps across all models, which included removing stopwords and tokenizing the words. Specific preprocessing adjustments were made based on the requirements of each individual model.



Distribution of Articles by Categories

Model Selection

- **LSTM (Long Short-Term Memory) :** LSTM is a type of recurrent neural network (RNN) capable of learning order dependence in sequence prediction problems. They are good at capturing long-range dependencies in text data which makes them suitable for this project where we are generating coherent and contextually relevant headlines where past context is key.

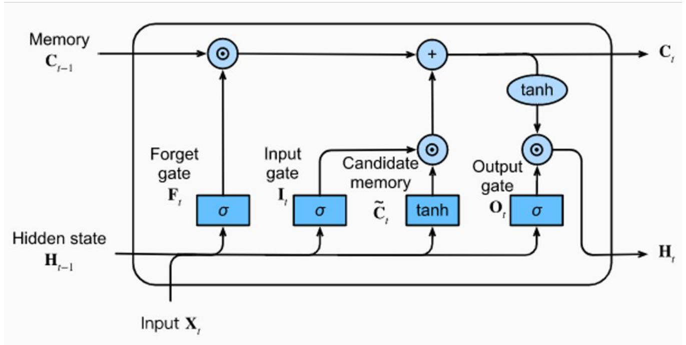
- **Bi-directional LSTM :** This approach basically extends the previous model i.e. the traditional LSTM by processing data but in both forward and backward directions which effectively increases the amount of information available to the network.

Model Selection

- **GRUs (Gated Recurrent Units) :** GRUs are a streamlined variant of LSTMs designed to use fewer parameters and simplify the learning process without sacrificing the capability to capture temporal dependencies. They provide a balance between computational efficiency and performance in learning dependencies for shorter sequences which is why they have been also used in this project.

- **GPT-2 (Generative Pre-trained Transformer 2):** GPT-2 is an advanced transformer-based model renowned for its ability to generate text that mimics human-like context and structure. With the increasing popularity of transformer models, it becomes necessary for us to test them alongside traditional models to understand their computational strengths

LSTM



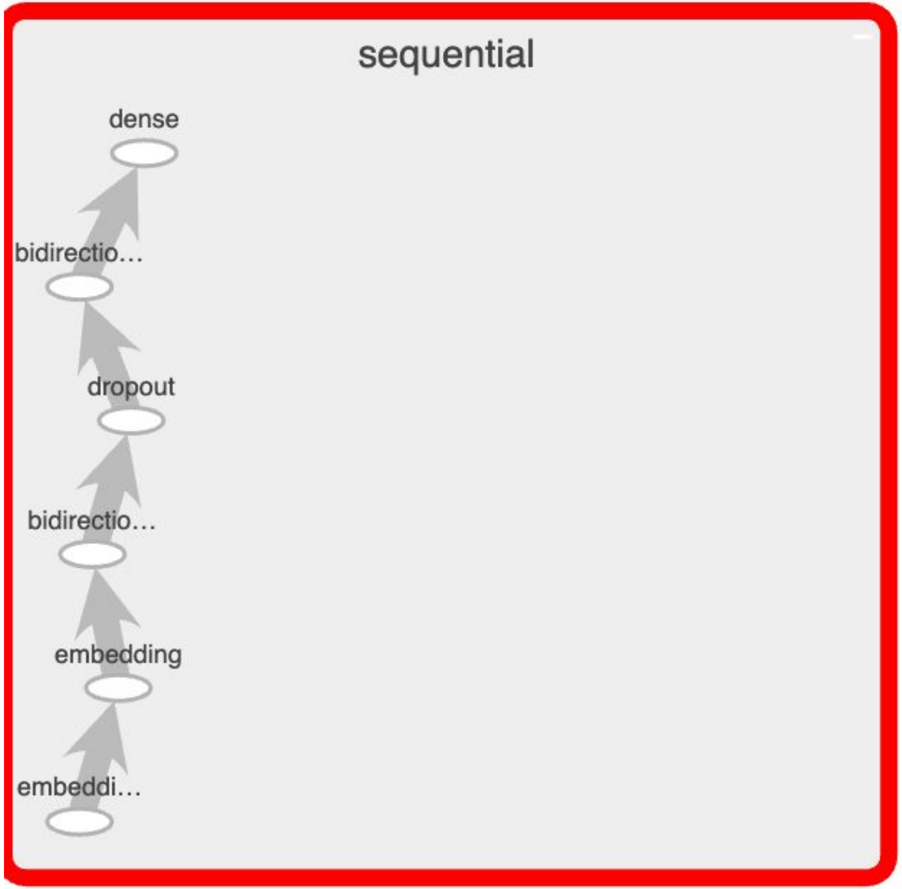
- LSTM is type of RNN designed to overcome the vanishing gradient problem
- It consists of 3 gates - Forget, Input and Output
    - Forget Gate: Decides which information is discarded from the cell state.
    - Input Gate: Decides which new information is added to the cell state.
    - Output Gate: Determines the output from the current cell state and the hidden state

- The model comprises an Embedding layer, one LSTM layer, a Dropout layer, and a Dense output layer.
- Embedding layer converts tokens into vectors of dimension 10
- LSTM layer consists of 100 LSMT units where input dimension is 10 and the output is 100
- Dropout layer drops an output 10% of the time, this prevents overfitting
- Dense layers maps the output of the LSTM to a Softmax function which predicts the next word in the sequence from the vocabulary

Layer (type)	Output Shape	Param #
embedding_2 (Embedding)	(None, 23, 10)	100160
lstm_2 (LSTM)	(None, 100)	44400
dropout_2 (Dropout)	(None, 100)	0
dense_2 (Dense)	(None, 10016)	1011616

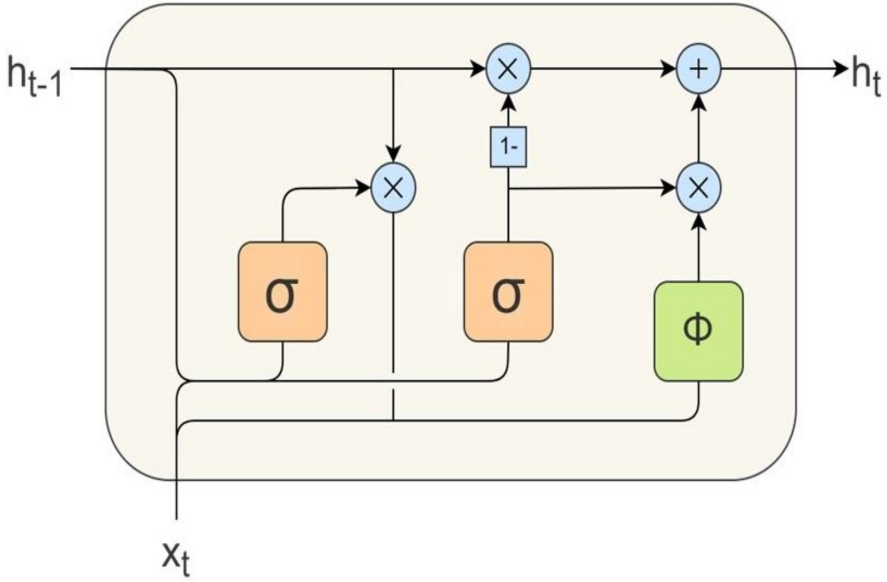
Total params: 1156176 (4.41 MB)
Trainable params: 1156176 (4.41 MB)
Non-trainable params: 0 (0.00 Byte)

Bidirectional LSTM(Conceptual Graph)



- **Embedding Layer**: Maps words to a 100-dimensional vector space to capture linguistic meanings.
- **Bi-directional LSTM (First Layer)**: Analyzes text from both directions with 256 units, enhancing context understanding.
- **Dropout Layer**: Reduces overfitting by randomly omitting 40% of LSTM units from each update cycle.
- **Bi-directional LSTM (Second Layer)**: A subsequent 128-unit layer to refine the sequence representation.
- **Dense Output Layer:** Transforms LSTM output to a vocabulary-sized probability distribution for word prediction.

GRU

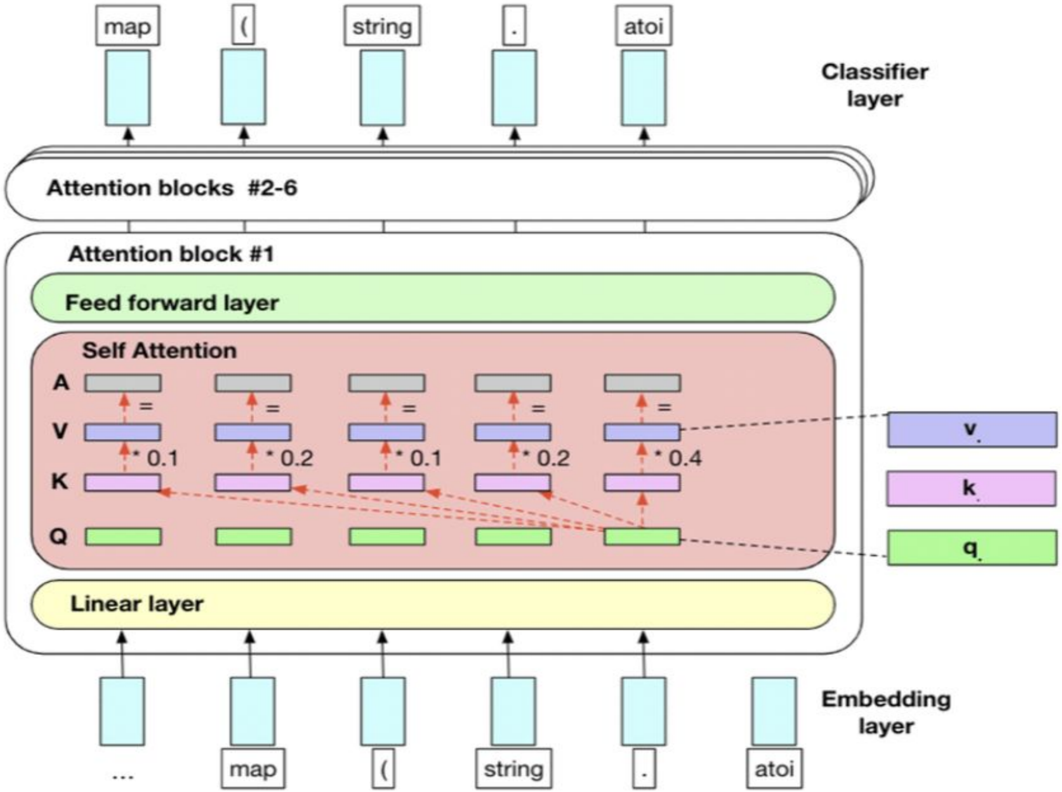


- **Architecture Simplification:**
  Merges key gates into a single mechanism, simplifying LSTM design without losing performance.
  Optimally processes sequential data, excelling in tasks with important time dependencies like speech and text.

- **Advantages Over Traditional RNNs:**
  Overcomes traditional RNN limitations such as the vanishing gradient problem.
  Requires fewer parameters than complex models, enhancing training efficiency.

- **Applications and Efficiency:**
  Ideal for environments where computational resources are limited (e.g., mobile and embedded systems).
  Balances computational efficiency with robust sequence processing capabilities, suitable for both small and large-scale applications.

GPT2 Architecture



- **GPT-2** is based on the Transformer architecture, which has traditional recurrent layers in favor of multi-headed self-attention mechanisms.

- In our use case, we incorporate the **GPT2LMHead Model** which is created for language modeling tasks.

- The current model being used for the project is : GPT-2 Large which has **774 million** parameters.

Model and Parameter Fine Tuning

- **LSTM :** Embedding dimension: 10, 100 LSTM Units, Dropout layer (10%), Dense Layer - 10016 (vocab size). Optimizer used - Adam. Number of epochs was 200 and batch_size of 64 was used.
- **Bi-directional LSTM :** Embedding Dimension: 100, First Bi-directional LSTM Layer Units: 256, Dropout Layer: 0.4. Second Bi-directional LSTM Layer Units: 128, Dense Output Layer Units: Equal to the vocabulary size, Optimizer: Adam with a learning rate of 0.001. Number of epochs was 200 and batch_size of 64 was used.
- **GRU:** Embedding Dimension: 10, GRU Layer Units: 200, Dropout Layer: 0.2, Dense Output Layer Units: Equal to the vocabulary size, Optimizer: Adam with a learning rate of 0.001. Number of epochs was 100 and batch_size of 32was used.
- **GPT2 :** For GPT, techniques such as Beam Search were tested to generate shorter headlines, including parameters like early stopping, max_length, and eos_token_id. Setting the temperature to 0.8 and top_k to 20 yielded the best results. It’s important to note, however, that while the model produces high-quality outputs, the BLEU score remains low because the model is trained to generate more diverse responses.

Comparative Results

Model	Loss	BLEU Score	Human Verification	Output Quality (low - 1, high - 5)
LSTM	1.348	0.265	Good	2
Bi-directional LSTM	0.57	0.27	Good	3
GRU	1.72	0.34	Below Average	1
GPT2	8.53	0.25	Good	4

Results

LSTM

```
Usa Gymnastics Still Values Medals More Than Girls
Black History Month Pressing Forward On The Heels Of Progress
Donald Trump Master Of Low Expectations
Racing Cars Decline In Value Why Not Homes Too
Minimum Wage Workers Protest The Wimping Of Joggers
Protest Errupt At Sacramento City Hall As Speakers Condemn Killing By Police
North Korea Missile Test Was Short But Informative
```

Bi-directional LSTM

```
United States And China Battle Over Fees Could Cripple Labor
Trump And Staff Rethink Tactics
Attack In Paris Casts A Shadow On French
```

GRU

```
Input: The third season
Model: The Third Season Of The Great Creator Of Social Media To Help You Become A Better Baker
Input: Raising capital to
Model: Raising Capital To 3 Health Care Bill Is A Genius Of 2018 Plans For The United States
Input: New York Today Sees
Model: New York Today Sees A Deal To All In The United States'
```

GPT2

```
Input reference: Donald Trump Announces
Model: Donald Trump Announces His Presidential Nomination
Input reference: India or
Model: India or India-china Relations
```

Conclusion

- Our comprehensive study on text generation models has unveiled distinct capabilities and limitations inherent in LSTM, Bi-directional LSTM, GRU, and GPT-2 models. Our findings underscore the sophistication of these models in generating coherent and contextual news headlines, striking a balance between complexity and computational efficiency

Future Work

- Enhanced Model Optimization
- Expanding Dataset Diversity