

Mid Prep : Adobe

Team 99

Inter IIT TechMeet 12.0

December, 2023

Table of Contents

- 1 Overview
- 2 Solution Pipeline
- 3 Behaviour Prediction Head
- 4 Content Generation Head
- 5 Future Scope

An Overview of Our Solution

- Our solution consists of two parts:
 - Behaviour Prediction Head
 - Content Generation Head
- **Behaviour Prediction Head** is used to perform the regression required to predict the likes on a given tweet.
- **Content Generation Head** is used to predict the tweet from a given set of inputs.

Solution Pipeline

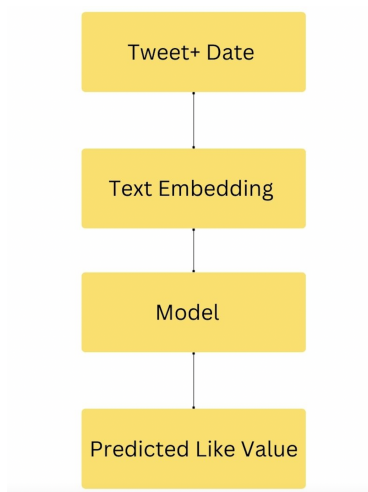


Figure: Behaviour Prediction Head

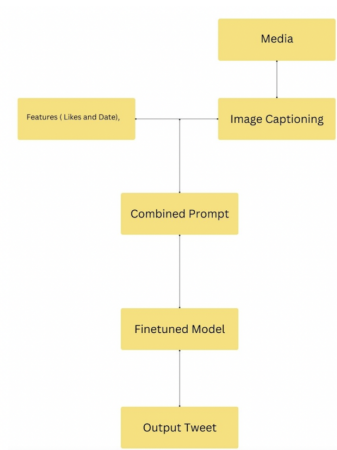


Figure: Content Simulation Head

Behaviour Prediction Head

- **Baseline Solution:** A simple feed forward neural network taking vectorized text as input and performing regression to predict likes.
- Extensive training data coverage across the domain and the limited number of layers in our model result in a significantly subpar performance.
- **New Approach:** Downstream by clustering followed by training with image data on independent models for each cluster.

Behaviour Prediction Head

- **Clustering:** Ideal Number of Clusters found using the Elbow Method.

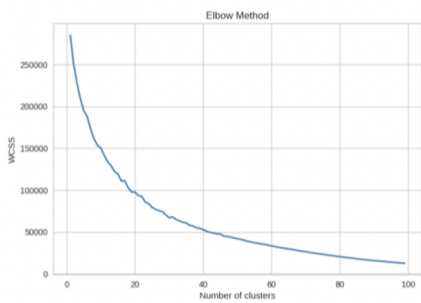


Figure: Finding the ideal number of clusters

- For capturing image data, we used two methods:
 - Preprocessing image into embeddings using MobileNetV2
 - Generating text from images using BLIP and using this text as an input to the regression model in vectorized form

- For Task 2 Our solution consists of two parts:
 - Image Captioner
 - Content Generator
- **Image Captioner** is used to generate captions for the images
- **Tweet Generator** is used to generate the content.

- **Baseline Solution:** Use of VIT+GPT2 for image captioning.
- The combined VIT+GPT2 model exhibited challenges in accurately interpreting images with complex scenes, often leading to hallucinations. Further research is needed to address these limitations and enhance the model's robustness in diverse visual contexts.
- **New Approach:** In our comparison of GIT, BLIP, and VIT+GPT2 models, GIT demonstrated superior performance, albeit with longer inference times. BLIP presented a balanced trade-off, offering a more efficient inference process while providing sufficient image details.

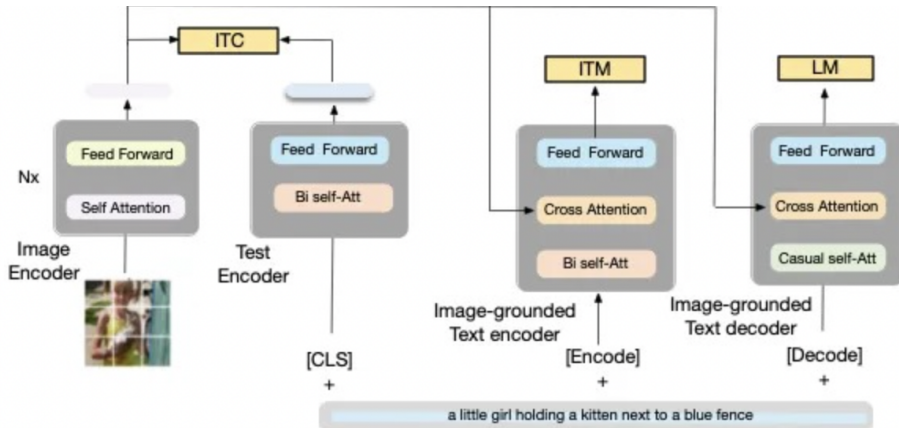


Figure: Architecture Diagram of BLIP

Image Captioning



Figure: Caption for above figure: blackberry logo on a dark background

- **Baseline Solution:** Using GPT-2 without fine-tuning for content generation results in suboptimal output, including inaccuracies in hashtag generation. This limitation stems from GPT-2's lack of domain-specific knowledge and context. Fine-tuning on relevant data or exploring other text generation techniques improved performance.
- **New Approach:** Fine-tuning GPT-2 LLM on specific data has proven effective, resulting in content generation at an acceptable level. This process enhances the model's ability to produce more contextually relevant and accurate outputs tailored to the domain of the fine-tuned data.

GPT 2

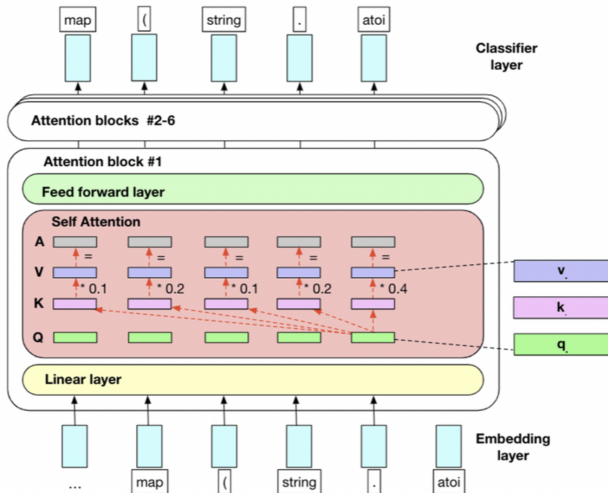


Figure: GPT 2 Architecture

Future Scope: Psychographic Clustering

- Psychographic clustering is the process of categorizing persons according to psychological and social criteria, such as their interests, values, lifestyles, and personalities.
- Segmentation goes beyond psychographic clustering by splitting the audience into discrete subgroups that have similar traits. Segmentation is a crucial component in constructing resilient prediction models. By comprehending the varied tastes among groups, we may customize our models to achieve more accuracy and personalization.
- Forecasting the number of likes by analyzing tweet metadata. Traditional models typically take into account variables such as posting time, frequency, and content kind. However, psychographic segmentation adds an additional level of complexity by integrating user preferences, emotions, and contextual comprehension.

Thank You!