InterIIT: Adobe Behaviour Simulation Challenge

Team Number 52

Introduction

Timely and targeted content generation is essential to yield desired marketing results for any business. This is carried through higher user engagement which in turn drives sales and broader customer outreach. Scoring high on user engagement KPIs boosts brand credibility, reach and product recall. This also helps marketers strengthen on algorithmic favorability which for a social media company decides on the post that gets propelled based on the engagement it drives. Therefore, it is necessary to have a method/tool to assess and predict the user behaviour of the published content as well as generate content for the expected behavioural pattern. Therefore, this document presents a solution which tried to solve the problem of behavior simulation (Task-1) and content simulation (Task-2), thereby helping marketers to estimate user engagement on their social media content as well as create content that elicits the desired key performance indicators (KPI) from the audience.

Dataset

Brands use Twitter to post marketing content about their products to serve several purposes, including ongoing product campaigns, sales, offers, discounts, brand building, community engagement, etc. User engagement on Twitter is quantified by metrics like user likes, retweets, comments, mentions, follows, clicks on embedded media and links. For this, sampled tweets posted in the last five years from Twitter enterprise accounts were used. Each sample contains tweet ID, company name, username, timestamp, tweet text, media links and user likes.

Proposed Methodology

The high level design diagram of the proposed approach is as shown in Figure 1

As shown, the input consists of various modalities of inputs such as images, videos, GIFs and Text. Each of these modalities must be handled uniquely.

The proposed methodology tries to efficiently tackle this by projecting the various modalities of data onto a common modality, which is here Text. Therefore, here the data provided in the form of Image and Video are converted into Text format, using Image and Video Captioning.

Using advanced methods for Image and Video Captioning, the methodology mainly works over the idea of capturing and representing images as text, and therefore, bringing the entire data into a common parameter space which can be effectively used for the intended purpose.

Further the proposed approach has an efficient working due to the type of data used, which is tweets. In general, these tweets have context in them via hashtags and account information, which would therefore, help in interlinking the visual data and its text representation.

Now, once all the data is processed as text, the entire data is passed to separate models to tackle each of the individual tasks, which are Behaviour Prediction and Content Generation. For the Behaviour Prediction, Regression models are used while for Content Generation Large Language Models (LLMs) are made use of to which data is communicated in the form of prompts.

Implementation

The implementation mainly included of implementing Machine Learning and Deep Learning Models for Image Captioning, Behaviour Prediction and Content Generation. The details are as follows:

Data Pre-processing.

- **Date Calculation:** The model calculates the temporal distance by referencing the post date against a specific date, quantifying the number of days before the reference date.
- **Hashtag Encoding:** Extraction of hashtags from post captions, followed by one-hot encoding to represent their presence or absence in the captions.
- **Caption Length:** Determination of the character length of the post captions.
- **Textual Data Generation:** Caption extraction from images and videos, merged with original text, creating a unified corpus for analysis.
- **Text Analysis:** Utilization of the **wordninja** library after preprocessing the corpus to derive two key scores: polarizer and summarizer

Image Captioning.

The image captioning makes use of a pre-trained transformer model, specifically the BLIP-2 Model for generating Image Captions. This model takes input the image after loading it from the URL. BLIP-2 achieves state-of-the-art performance on a variety of vision-language tasks, surpassing existing methods with significantly fewer trainable parameters.

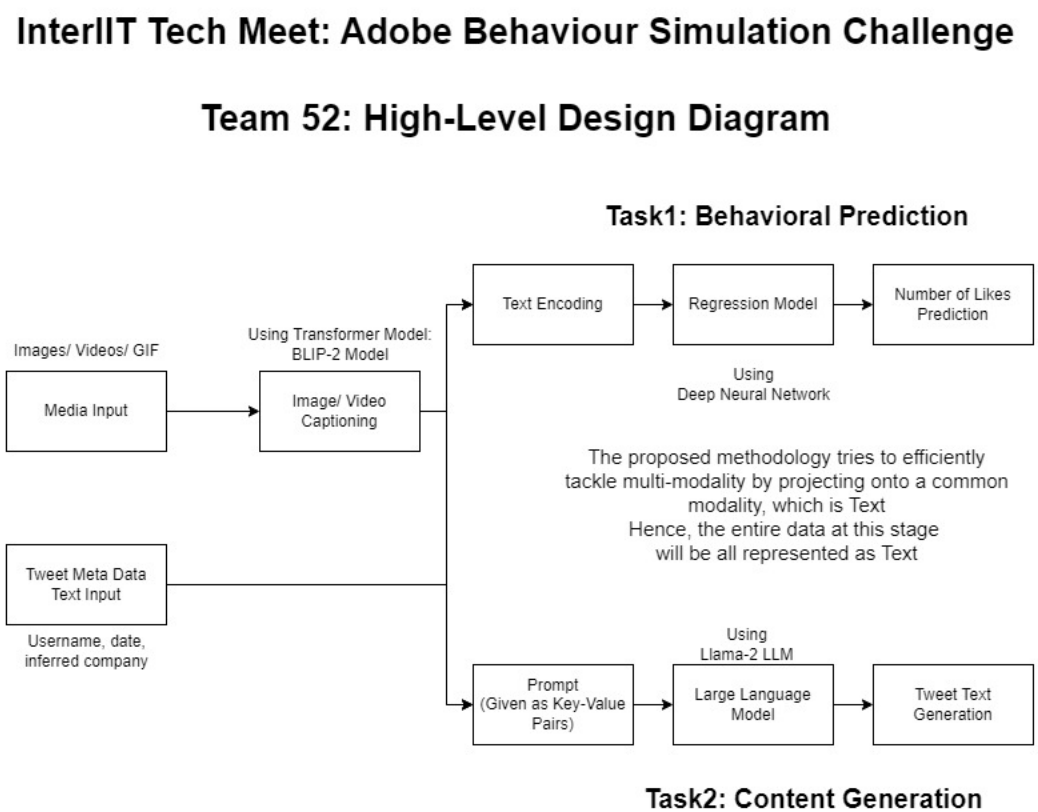

Fig. 1. High Level Design Diagram

Behaviour Prediction.

The predictive model employs a deep neural network architecture, integrating several key elements to enhance performance and prevent overfitting:
Six input parameters are utilized:

- Days before the reference date when the post occurred.
- One-hot encodings representing the inferred company name.
- One-hot encodings indicating the presence of hashtags.
- Length of the post caption.
- Two scores obtained from the **wordninja** library: polarizer and summarizer.

The deep learning architecture employed involves a neural network comprising layers equipped with Dropout and Batch Normalization
The model's primary output is the prediction of the number of likes a social media post is expected to receive based on the amalgamation of these input parameters and the trained neural network

Prompt Engineering.

Before passing the data to the Large Language Model (described in the next section), there is a need to create a prompt for input data to effectively communicate it to the LLM. The prompt used for the current usecase is

"You are the tweet predictor bot for company all over the world. You have vast knowledge of how the tweet are designed. You should predict the tweet depend on the likes it got, image caption given by the user and utilize your knowledge to know about the company at that time if you can. You should predict the tweet such that it was written by a human. Tweet should have maximum of 3 sentences.
Generate a tweet for a social media post based on the following information:
Likes:
Image Captions:
Date and Time:
Inferred Company:
Username :
Craft a compelling and engaging tweet that incorporates the given image caption ideas, and try to include ideas if get from the inferred company name as well as username. Ensure the tweet is suitable for a social media platform and captures the audience's attention. You don't need to provide any explanation."
The results are also on the basis of this prompt

Content Generation.

For the task of content generation, the implementation makes use of a Large Language Model (LLM), specifically the Llama-2 LLM. This LLM generates tweets on the basis of the prompt mentioned above. Llama 2 was pretrained on publicly available online data sources.

Experimentation

Behaviour Prediction.

For Behaviour Prediction, both Machine Learning models as well as Deep Learning Models were tested The model was trained on a subset of the dataset and the results are as shown in Figure 2 were obtained

```
Random Forest Regression Metrics:
Mean Absolute Error: 145.6495
Mean Squared Error: 41810.912985
Root Mean Squared Error: 204.477169838102
R-squared: 0.2790101878757373
```

Fig. 2. Random Forest RMSE

The Deep Learning Model performance was as seen in Figure 3

```
Epoch 1/100
9375/9375 [==============================] - 26s 3ms/step - loss:     1038.0000 - mae: 1082.1580 - ac
curacy: 69.0672
Epoch 2/100
8314/9375 [=========================>....] - ETA: 2s - loss:     3282.0000 - mae: 1088.2264 - accurac
y: 64.0673
```

Fig. 3. Deep Learning Mean Absolute Error

Content Generation.

The final content generation results for a few tweets are as shown in Figure 4

```
Tweet:
Ready for a relaxing getaway? Treat yourself to a luxurious stay at @ClarenceHouse, where every room is a sanctuary. #ClarenceHouse #GetawayGoals #

Tweet: Breaking News! A shocking discovery has been made in the depths of the ocean! Scientists have found a new species of fish that glows in the dark!
 Just generate the tweet.

Here is the generated tweet:

" We're on a roll! Our Reds are unstoppable, and we're loving every minute of it! #LFC #LiverpoolFC #Anfield #Kop #YNWA"

Tweet:
"Every child deserves a fair chance at life. Let's make sure they get it! #UNICEFROSA #EducationForAll #InclusiveLearning"
```

Fig. 4. Sample Tweet Generated

Conclusion

The document proposes a simple yet efficient architecture to tackle the problem. Although the model was trained and tested on a subset, the generated output showed decent results. The code is available here: https://github.com/MoyankGiri/MP4_AdobeChallenge