Multimodal Deep Regression Project

Wong, Louis- lwong64@gatech.edu
Song, Mingyao - msong41@gatech.edu
Xu, Jason - jxu623@gatech.edu
Salih, Ahmed - asalih6@gatech.edu

1. Team Name: Deep Gamma
2. Project Title: Multimodal Deep Regression on TikTok Content Success
3. Project summary:

   Content creators grapple with the challenge of predicting if their time and investments will translate into increased viewership and audience growth, a task made more complex by the hidden algorithms and unpredictable audience interaction of social media platforms. This research's objective is to architect a model that predicts video success, effectively indicating a video's potential virality. By employing advanced convolution techniques for video encoding, leveraging strides in natural language processing models, we're pushing boundaries in deep video content analysis. Our ultimate goal is to construct a multimodal ensemble model, a general video content to regression that can comprehend human generated content and accurately map its nonlinear relation elements to predict success.

4. Implementation:

   The video will first be broken down into visual and audio tracks, with our primary focus on the video visual. The video visual will undergo an autoencoder process, utilizing a convolution-based network architecture, ConvLSTM Autoencoder, to unsupervised pre-training from scratch. This process will encode the context of the video into embedding vectors. Subsequently for the audio, we will leverage a pretrained model, such as Mozilla's Deep Speech or the open-source Whisper, which will be used to create a transcript for the audio, supplementing the project. Afterward, embeddings will be extracted from the visual branch and the audio, which will then be concatenated and input into a transformer-based regression model. The aim is for this model to learn the semantic and non-linear relationships needed to predict video creator success metrics, such as video views. Finally, we will establish a baseline using a less complex model and compare it with our main implementation to evaluate its success.

5. Resources / Related Work & Papers:

   Several studies like "**Instagram Popularity Prediction via Neural Networks and Regression Analysis**" and "**Instagram Post Popularity Trend Analysis and Prediction**" have leveraged CNN+RNN for content success prediction. Paper "**Video Summarization Using Deep Neural Networks"** uses techniques like GANs and auto-encoders have been employed for video summarization. "**Predicting User Participation of TikTok Challenges**" also explores multi-modal such as ResNet and BERT for classification problems. Our research builds on this foundation, incorporating video encoding into regression analysis to explore the deep ensemble architecture and extract insightful features for content popularity prediction.

   - Images and Videos Popularity Prediction: a Deep Learning-Based Approach https://ceur-ws.org/Vol-3102/paper2.pdf

- Instagram Post Popularity Trend Analysis and Prediction using Hashtag, Image Assessment, and User History Features https://iajit.org/PDF/Vol%2018,%20No.%201/19395.pdf
- Instagram Popularity Prediction via Neural Networks and Regression Analysis https://cjqian.github.io/docs/instagram_paper.pdf
- Slapping Cats, Bopping Heads, and Oreo Shakes: Understanding Indicators of Virality in TikTok Short Videos https://arxiv.org/pdf/2111.02452.pdf
- Video Summarization Using Deep Neural Networks: A Survey https://arxiv.org/pdf/2101.06072.pdf
- Deep Learning for Video Classification and Captioning https://arxiv.org/pdf/1609.06782.pdf
- Unbox the Black Box: Predict and Interpret YouTube Viewership Using Deep Learning https://arxiv.org/pdf/2101.01076.pdf
- Will You Dance To The Challenge? Predicting User Participation of TikTok Challenges https://arxiv.org/pdf/2112.13384.pdf

6. Datasets:

   The dataset consists of video data scraped from TikTok using our custom-built Selenium scraper on a Chromium browser. The videos were randomly scraped from the platform using a variety of random hashtag topics, including Sports, Dance, Entertainment, Comedy and Drama, Autos, Fashion, Lifestyle, Pets and Nature, Relationships, Society, Informative, and Music. In total, we scraped approximately 5,000 videos in .mp4 format, amounting to a total size of 32.7 GB. Each video has a unique video ID tag. Alongside the video, we have JSON data that contains the video ID tag and other metadata for the video, such as the URL and video view count according to TikTok.