Team 6 Project Proposal

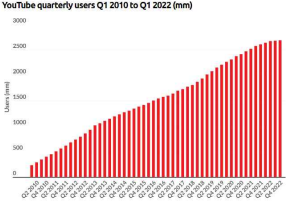
YouTube View Count Prediction

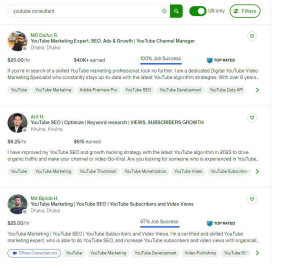
Suho Park
Chaejin Lim
Jaejoon Yoo
Jeongyeop Seo
Jihyun Lee

1. Background

- Reached 2.5 billion active users in Q2 2021
  - Many consultants in UpWork
  - **How could I get more views for my video?**

- Idea: View Count Prediction Model





1. Background

- “Machine Learning enabled models for YouTube Ranking Mechanism and Views Prediction.” - Gupta et al., 2022
  - Analytic study using IBM Watson Model(AutoML)
  - Text and numerical features
- “Engagement and popularity dynamics of YouTube videos and sensitivity to meta-data” - Hoiles et al., 2017
  - Study on feature importance of thumbnail and meta-level features
- “Predicting and Characterizing Early Growth of YouTube Videos” - Kharkar et al., 2020
  - Predict view growth patterns using thumbnail analysis (e.g., presence of text, major colors)

2. Proposed Method

- Predict the number of views of a YouTube video, prior to uploading
  - Features: **thumbnail**, number of subscribers, title, uploaded time period
  - Predictor model: Random forest, MLP, or GNN

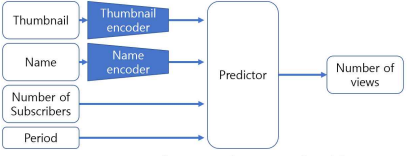


Figure 1. Overall structure of our proposal model

2. Proposed Method

- Utilization of Thumbnail - Image Encoder
  - Pretrained image feature encoders
    - Targeted on image classification tasks
  - Contrastive Learning
    - Specialized encoder for thumbnail feature extraction

- Targeted Catagory: Korean Mukbang (FoodCasting)
  - Nation and catagory of video highly affects view counts

3. Contribution

1. Utilize image features from thumbnail that highly affect view counts
2. Learn features targeted for Mukbang thumbnails using contrastive learning
3. Build own dataset using browser engines and YouTube V3 API

4. Dataset

- Web crawler using BeautifulSoup, Selenium, Youtube V3 API
- Crawl search list on YouTube with keyword “Mukbang”
  - Data preprocessing: filter out unrelated videos
- Query videos’ features and target value
  - Title, number of subscribers, thumbnail image, creation time, view count

5. Evaluation Method

- Loss function and metric
  1. MSE of logarithm of view counts
  2. Cross-Entrophy of softmax to classes, splitting database by view counts to 10 classes
- Baseline models
  - "Youtube Views Prediction Machine Learning" - Putra, Kaggle
    - Opensource model using Random Forest
    - Text and numerical features

6. QnA

- Thank you.