Hash-tag Generation for Social Media Content

Mentor : Shashank Gupta

Members

- Harshil Jain

- Syed Ahmad

- Kaleemullah Mohammed

- Krishna Chaitanya Pappu

Project Description

The project takes Social media content having text and images and generates hash-tags describing the content. It is not a classification problem where we assign one of the predefined set of labels, rather we generate new hash-tags each time using a trained language model.

System Design

Crawler

The data is crawled from twitter using tweepy API based on user-name. The text and URL to the image are saved.

Later, we separate the tweets having hash-tags. So, each data entry has

< plain text, image , <set of hash-tags> >.

Training

We separately extract features from the image as well as text.

Once the features are got, we train a language model based on the collected data to generate a set of hash-tags.

Functioning

Once the system is fully trained, given a social media content, we get the textual and image features and feed them into the language model, which predicts the hash-tags for this content.

System Diagram