Intrinsic Image Popularity Assessment

Aarushii Agrawal
IIIT Delhi
India
aarushii17324@iiitd.ac.in

Shahid Nawaz Khan
IIIT Delhi
India
shahid17102@iiitd.ac.in

Sudhanshu Mohan
IIIT Delhi
India
sudhanshu17110@iiitd.ac.in

***Abstract*—In the paper Intrinsic Image Popularity Assessment, the aim was to single out the contribution of visual content to image popularity. We plan to replicate their model and further train it for Twitter analysis. Specifically, we first describe a probabilistic method to generate popularity-discriminable image pairs using pictures from tweets of various popularity from trending hashtags on Twitter. We then develop computational models for Image comparison based on deep neural networks, optimizing for ranking consistency with millions of popularity-discriminable image pairs. Popularity of the image was measured using a combination of retweets and likes on that tweet. A validation accuracy of upto 70 percent was achieved using a pre-trained ResNet-50 model.**

***Index Terms*—Intrinsic image popularity, popularity-discriminable image pairs, Twitter**

I. Introduction

The last decade was seen to be a site of explosion in terms of online media and photo sharing, where platforms like Instagram and Flickr emerged as full-blown businesses for a lot content makers. In such a culture, predicting the possibility of a picture going viral on the internet is of crucial value. Pictures that surface on these platforms are very diverse in nature, this is evident from the varied number of likes, comments and shares they receive. These differences can be explained based on numerous factors ranging from the actual visual content of the image to user's influence to the time and location of the post.

Previous work done with a similar problem statement, has been on social media sites like Flickr or Instagram. In this paper, we will focus solely on Twitter, one of the most famous social media platforms in today's day and age. It is well known that on Twitter, the number of followers and time of posting has a role to play in defining the reach of the tweet. However, in this work, we will constrain ourselves to the visual components of the image in a tweet. Since we've taken images from trending hashtags, the number of retweets will be an appropriate measure of the popularity achieved. However, we will consider a combination of number of retweets and number of likes to determine the final popularity.

One of the most evident challenges faced when working with a problem statement related to virality is the adaptation of ever-evolving pop-cultural context. Predicting meme virality faces a similar problem, modern humor being different from generic humor. Another challenge encountered when working with a platform like Twitter is its varied usage. The most millennial use of Twitter is as a medium for expression of thought. There are also some amateur as well as professional photographers and content-makers. Verified users use Twitter differently, avidly for for promotions purposes and anything they tweet is likely to go 'viral'. Youth activist accounts are the highlight right now. The list goes on. Training a model that works accurately despite all this variation is a challenging task.

Identify applicable funding agency here. If none, delete this.

II. Literature Review

In the paper "Intrinsic Image Popularity Assessment" by Ding, Ma and Wang [1], they focus only on the visual component of the image to predict the possibility of it going viral. Their database consists of posts collected from over 200,000 active users on Instagram and accommodates features like download time, post URL, user ID, content type, upload time, caption, number of likes, and the number of comments. As a result, the final dataset consists of over 200 million posts with varying popularity metrics. In the model used, they have constructed pairs of images (Popularity Discriminable Image Pairs, PDIP) based on a criteria and trained a deep neural network on the same. The pairs were selected based on a probabilistic method to ensure that the popularity-discriminability of all the pairs in the dataset. Apart from this, constraints like the posts should be from the same user, and the post must have been posted within ten days prior to the dataset creation were also kept in mind. The paper adopted ResNet-50 as the default Deep Neural Network architecture, and obtained the predicted intrinsic popularity score by replacing the final ResNet-50 layer with a fully connected layer of one output. They achieved an accuracy of around 76% with this model.

In the paper "Retweet Wars: Tweet Popularity Prediction via Dynamic Multimodal Regression" by Bansal, Wang and Frahm [2], they use a combination of visual features and contextual features containing data modalities like tweet language semantics, embedded images, author' social relationships, and the diffusion process of tweets to predict the popularity of a tweet. In the paper, they have proposed a joint-embedding neural network. This neural network combines 3 types of data: visual, textual, and social cues. They have also optimised a novel Poisson regression loss for training purposes. Their model consists of advanced CNN and LSTM models along with the joint embedding proposed. The evaluation metric used by them are Spearman's Ranking and MAPE (mean absolute percentage error). Their results were: Spearman's Ranking of 0.358 and MAPE value of 0.084 on the Twitter2015 dataset.

A. Details of Baseline Paper

The paper we have chosen titled "Intrinsic Image Popularity Assessment" [1] achieved an accuracy of 76.65% which outperforms the human accuracy of 72.40%. The testing is done in pairs, i.e. the model returns the virality scores on the images in the pair. The testing set consisted of 50,000 PDIPs and the model used was ResNet-50.

III. Proposed Solution

The solution we propose is to use the popularity-discriminable image pairs(PDIP) method as mentioned by Ding's paper [1] and use the method for image popularity prediction on Twitter after fine tuning it in order to apply it on images on twitter. We use the pretrained model provided by the paper [1] and run back-propagation on it to use it for Twitter images. Despite the differences in the two platforms in terms of user demographics and purpose of use, the restrictions placed by Ding, Wang et, al. [1] while creating the PDIP dataset ensure that the dataset is not entirely platform dependant and therefore their model proves to be an excellent starting point to move forward with creating a popularity prediction model for any platform. Since previous work done on Twitter has been oriented to include contextual data as well, we propose to simplify the process and predict image popularity on Twitter strictly based on only visual components. Firstly, we will retrieve images from Twitter from trending (or famous) hashtags using Twitter APIs. We will then form PDIPs by implementing the method followed in our baseline paper. After that, we will train our model which consists of a Deep Learning Network for each picture. The network focuses solely on the visual aspects of the image.

$$Q_A \leq Q_B \tag{1}$$

$$Where\ Q_X\ is\ the\ popularity\ score\ of\ Image\ X$$

This gives us a direct relationship between the popularity metric of the two images in a PDIP.

Popularity of an image (p) will be determined by a combination of number of retweets (r) and number of likes (l) on the tweet, with a greater emphasis on the number of retweets. The greater emphasis on number of retweets was because our literature review suggested that number of retweets are a better measure for virality on Twitter and the paper "Retweet Wars: Tweet Popularity Pre-diction via Dynamic Multimodal Regression" [2] by Bansal, Wang, et al. suggested the same. The formula for the same is:

$$p = (0.7) * r + (0.3) * l \tag{2}$$

We tried a total of 3 combinations: (0.6, 0.4), (0.7, 0.3) and (0.8, 0.2) as the weights for number of retweets and number of likes respectively. However, the second combination gave the closest value to the ground truth.

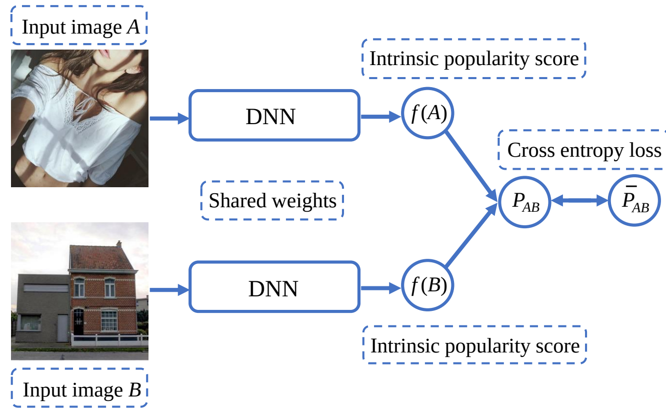


Fig. 1: Architecture of the DNN models for Ding's proposed solution [1]. The two streams have the same model of ResNet-50 as the DNN, The model parameters are shared between them, and is optimized by minimizing the cross entropy loss.

This is done to incorporate those tweets that have a high number of likes but might not be retweeted as much, eg. pictures of the one's own self, etc. and also to break ties that arise when only number of retweets is considered.

IV. Dataset Collection

We formed a set of trending hashtags on Twitter or famous hashtags (often used). We queried the Twitter API to retrieve the images based on the hashtags. The pictures were downloaded in .jpg format. Our final dataset consisted of over 32,000 images.

V. PDIP Generation

After acquiring the images, we paired images according to a probabilistic criteria similar to the described in the baseline paper [1] to generate PDIPs (popularity-discriminable image pairs). The criteria is as follows, the images should be:

- from the same user;
- posted within ten days of each other;
- have captions with length differing by a maximum of five words;
- have at least one hashtag in common;

A majority of PDIPs were formed on these conditions alone, but to make sure we create variety, we added a couple additional condition for about 1/3rd of the PDIPs generated. These were:

- the number of retweets should be different in the two tweets;
- at least one tweet should have more than 20 retweets.

This was done taking into consideration that most tweets scraped had number of retweets in single digits and some of the pairs had equal number of retweets as well.

Post the PDIP generation, we acquired the number of retweets and likes of the images (tweets) and calculated the popularity of the images according to the formula described.

VI. DNN-BASED COMPUTATIONAL MODEL

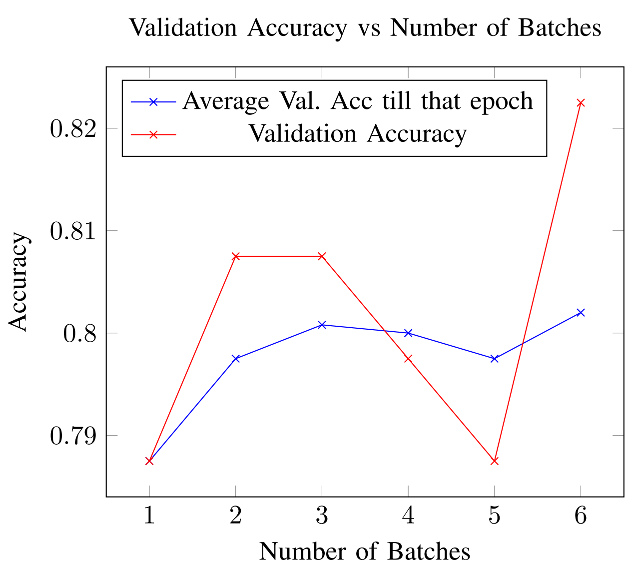
We used the same model used in the paper "Intrinsic Image Popularity Assessment" [1] by Ding, Ma and Wang. The model employs the *learning-to-rank* technique to learn rankings. In particular, the model uses a **pairwise learning-to-rank approach**, and it takes the relative rank between two images as an input, and in turn minimizes the pairs that incorrectly ranked. The two images are taken as inputs in a RGB image format, a *predicted intrinsic popularity score* is calculate for both the images and finally their difference is calculated (O_{AB}). During the training and testing, the weights of both the streams are shared. The network architecture of the both the streams is identical. The difference (O_{AB}) is converted to a probability using a logistic function:

$$P_{AB} = \frac{\exp\left(O_{AB}\right)}{1 + \exp\left(O_{AB}\right)}. \tag{3}$$

The loss function used in the model is binary cross entropy loss. The default DNN architecture is ResNet-50. Additional details of the model can be found in the paper "Intrinsic Image Popularity Assesment" paper by Ding et, al. [1]

VII. RESULTS

We ran the pre-trained Resnet50 model (trained on 5 million images from Instagram) on the training set consisting of 15,000 PDIPs. The pre-trained model weights were available on the GitHub account of author [1] and was of size 90 MB. Upon running this model on the train data, we achieved a validation accuracy of around 80.17%.



The SOTA model (Intrinsic Image Popularity Assessment [1]), reported an accuracy of 76.65%. Our model beats the SOTA model. We think this can be attributed to selectively picking images of varying popularity from Twitter instead of randomly scraping images. Our model performance as compared to past models that solved a similar problem can be visualised in Fig. 2.

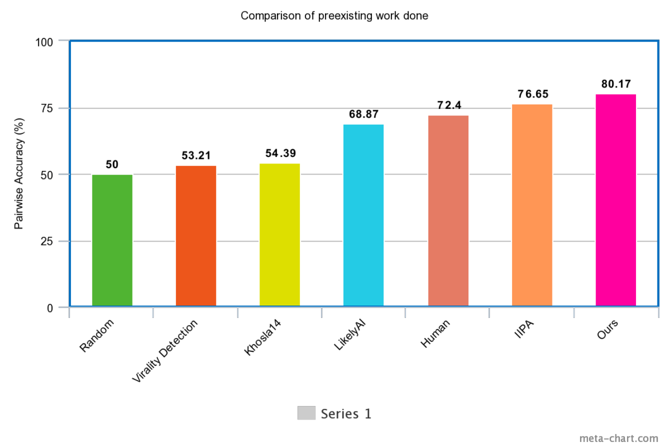


Fig. 2

REFERENCES

[1] Ding, Keyan, et al. "Intrinsic Image Popularity Assessment." Proceedings of the 27th ACM International Conference on Multimedia, 2019, doi:10.1145/3343031.3351007.

[2] Bansal, Wang et al. "Retweet Wars: Tweet Popularity Prediction via Dynamic Multimodal Regression" IEEE Winter Conference on Applications of Computer Vision (WACV), 2018, DOI: 10.1109/WACV.2018.00204 ...