# #LearnDeeply : A Multi-Modal Deep Learning Approach to Predicting the Popularity of Posts on Instagram

Alexandra Paul   Austin Tung
Qi Feng   Yiran Zhou

## ABSTRACT

In the social media age, post virality (measured in the number of likes received) has become an increasingly important metric for both power users such as influencers and celebrities - for whom likes translate into monetary gain - and for standard users as well. Existing research on post popularity focuses primarily on Flickr. However, there are few promising results for similar studies done in the scope of Instagram.

The dearth of studies on post virality on Instagram can be attributed to the lack of a comprehensive, publicly available dataset, the difficulty of post collection due to API limits, and the complexity of the underlying dynamics of this platform which makes it difficult to approximate a function that accurately explains the phenomenon of post virality. Further, we build upon an existing dataset of Instagram posts to introduce one that is updated and larger in scale. Building upon existing research, we run a variety of experiments using different deep learning models that take a post-content-centric approach to predicting the number of likes a post will receive on Instagram.

Our findings seem to indicate that a post's metadata, particularly context and time, are far more important than post content (image, captions, etc.) in influencing virality.

## RELATED WORK

Much research has been done in the realm of predicting popularity of posts on Flickr. *Sequential Prediction of Social Media Popularity with Deep Temporal Context Networks*[4] proposes a deep temporal context network that investigates the prediction of popularity of posts on Flickr based on user-photo time sequences. Their research yielded promising results that significantly outperformed existing prediction algorithms. If we had more data, we would have likely attempted to apply this approach to the problem of predicting popularity on Instagram.

There is a lack of papers that yield promising results for doing the same for posts on Instagram. Specifically, the papers *Predicting the Popularity of Instagram Posts for a Lifestyle Magazine Using Deep Learning*[1] and *Instagram Popularity Prediction via Neural Networks and Regression Analysis*[2], attempt to address this problem by focusing on virality prediction for a very limited dataset- 1,280 posts from exclusively the GQ India Instagram account and 3,411 posts labeled as scenery in the respective studies. We hypothesized that the limited data led to the creation of a model that failed to generalize the phenomenon in its entirety. However, certain components of each model proposed in these papers, particularly the analysis of social data[2] , seemed promising. We hypothesized that visual analysis proved ineffective in both of these papers due to the homogeneity of their datasets.

*Multimodal Popularity Prediction of Brand-related Social Media Posts*[3] seemed to be a far more comprehensive study, and found that visual and textual features are complementary in predicting the popularity of a post. However, they take a brand specific approach that again fails to generalize the problem. Their findings encouraged us to take a content-centric approach in our research. This only a brief synopsis of all the papers we referenced in our research.

## DATASET

The lack of a comprehensive, publicly available dataset of Instagram posts necessitated the generation

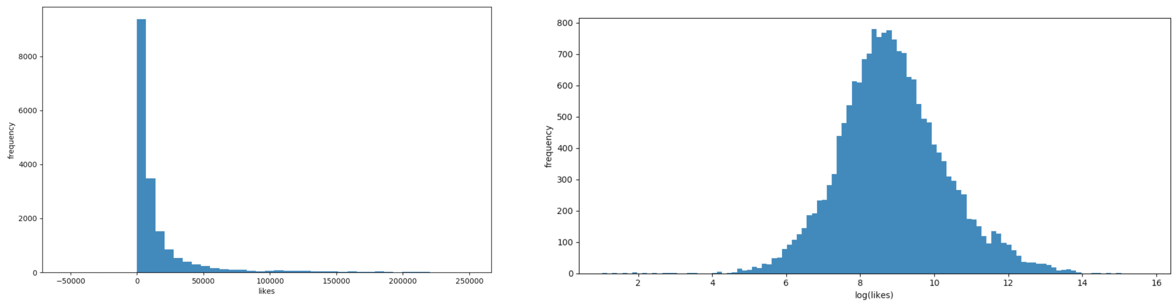

Figure 1: Comparison between the distribution of target labels unnormalized (left) and log scaled target labels (right).

of a new dataset. We built upon the dataset[5] introduced by C. Dugué, cleaning it of faulty data (ex: removing posts where the number of likes less is than 0) , updating the content (all image links were broken), and scraping more posts. The official Instagram API limits users to 200 requests per hour, so instead, we adapted an existing web-scraper[6] that utilizes Selenium to crawl through Instagram posts, extracting all relevant information regarding the post and the poster.

A list of 'influencers' was obtained from a combination of general knowledge of their existence as well as further research for curated lists[7,8] of such 'influencers' on Instagram. We scraped the most recent posts for each user on this list. The dataset collected consists of approximately 18,170 Instagram posts across 1,307 users. For each post, the image, the number of likes received, hashtags, post caption, other user mentions, and time (month, weekday, and hour) of posting were collected, along with the poster's follower and following count, the poster's total post count, and the poster's average likes per post.

# METHOD

## Data Processing

The number of likes, which functioned as the target label, followed a power law distribution. The nature of variability of data distributed in this manner is unideal for learning purposes, so we log scale the target labels as suggested in Mazloom[6]. This forces the data to behave more closely to a normal distribution. Such a scaling introduces less variability than would a power law distributed dataset (see Figure 1). In an effort to stabilize neural net learning and improve performance, metadata and image pixel values were also normalized to fit within a [0, 1] and [-1, 1] range respectively.

The metadata parameters were normalized by dividing by the maximum occurrence of each respective parameter. The time of posting (month, weekday, and hour) was translated into one-hot vectors. The hours of the day were grouped into 6 groups to reflect the different relative time periods (early morning, late morning, early evening, etc.) within the day. The one-hot vectors for the hour of posting was a reflection of these groupings. Hashtags were assigned weights based on their respective popularities and accumulated as a single parameter to represent the the hashtags used in a post. Although hashtag popularity cannot be defined in a straightforward manner, we assigned scores to hashtags based on their ranking in a list[1] 5 of the most popular hashtags.

The images were downsampled to a 224 by 224 square image before pixel value normalization.

## Proposed Model

A post-content-centric approach was the basis of our model. We considered the effect of the image itself, in terms of objects present and aesthetic value. We also considered the post's metadata as a framing for the context of the post. We experimented with a variety of single modal and multi-modal models that utilize these factors in an attempt predict virality. We arrived at the following multi-modal model (Figure 2), but also experimented with every possible combination of the individual metadata, image classification, and aesthetic scoring models.

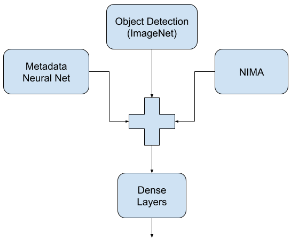

Figure 2: Proposed multi-modal model for predicting Instagram post virality.

## Metadata Model

The aim of including metadata was to supplement the image data and provide a contextual framework for the post. For the metadata module, we took into consideration the weight of the hashtags used, length of the caption, number of mentions, month, day, and hour of posting, as well as the follow and following count of the poster and the average number of likes the user received per post in our dataset. This data was fed through a network of fully-connected layers and returns an auxiliary output of how the context of a post affected its virality.

## Image Classification Model

Continuing with content-centric approach, we figured that the content of the posted image would correlate significantly with virality. We utilized Inception-ResnetV2 pre-trained on ImageNet weights to predict the objects within an image. This pre-trained model returns 5 objects with the highest probability of being in the input image. The returned predictions were highly specific (golden retriever versus yellow lab), so the predictions were mapped upwards in the hierarchy of synsets to obtain a more general category, meaning the detections we integrated into our models were something similar to 'animal' as opposed to 'german shepherd dog'. These detections were fed into an embedding layer and then through a series of fully-connected layers to obtain the preliminary output of how the objects within an image affect post virality.

## NIMA: Neural Image Assessment

Google's Neural Image Assessment is a deep CNN that is trained to predict the aesthetic and technical quality of a photo with high correlation to human perception[13]. We hypothesized that the human-like rating of image quality would accurately capture how users 'feel' about the photo in the post they are viewing, and as such, would correlate to whether they chose to like it or not.

We used transfer learning to attempt to exploit what had been learned in NIMA to improve generalization in our problem[11]. Because our dataset is small and similar to the original dataset, we used MobileNet pretrained on NIMA weights[9] as a fixed feature extractor[10]. We did so by removing the last fully-connected layer and retraining the last layer.

# RESULTS

We experimented with all possible combinations of the 3 submodules including individual testing. —(table or figure here with results) Table X displays our results of our experimented models compared to a published model.

Table 1: Comparing our model(s) with existing researcher model.

| | MAE | MSE |
|---|---|---|
| **Our Models** | | |
| Metadata NN | 0.345 | 0.283 |
| Image class NN | 1.099 | 2.079 |
| NIMA (transfer learning) | 1.424 | 3.400 |
| Combined Metadata & Image Class | 0.338 | 0.252 |
| Combined Metadat & NIMA | 0.373 | 0.282 |
| Combined all 3 | 0.347 | 0.281 |
| **Researcher Model** | | |
| Research Model[14] | 1.00 | 2.06 |

Our best performing model exhibits results on par with published models. It is important to note this comparison is relative due to differences in datasets and error metrics. Keeping in mind that the target labels were scaled by a logarithmic function, the losses are likewise in a log-scale and therefore, a bit difficult to interpret. For reference a mean absolute error loss of 0.35 corresponds to roughly 25% accuracy in a testing dataset of 6,000 samples, where consideration for correctness is a prediction that lands within ± 10% of its corresponding target label. Amongst the models and combination of models tested, the metadata alone performed the best. The combination of all the models performed only marginally better and is ultimately a negligible improvement in accuracy that requires far greater training time. While the results from even the best performing model were not ideal, it's worth noting that the accurate predictions were very close to their respective target labels, and the incorrect predictions were still within the same order of its target label (prediction of 4 for a target of 14, and prediction of 200 for a prediction of 150). From these results, we were able to conclude that the most significant element in predicting the virality of a post is the context of the post, not the content as we had falsely presumed. Furthermore, 2 specific parameters within the metadata held the highest weight: the poster's average likes per post and the time of posting. This leads us to believe that an LSTM model, like the one proposed in Wu[4], might have been better suited for predicting post virality due to the significance of the poster's history and the time of posting.

## DISCUSSION

Initially, we started with trying to approach the problem from a classification perspective. We grouped the posts into bins, represented by ranges of likes, and had the model attempt to predict the bin in which the post fell in. After this prediction was made, further analysis over the statistics and distribution of the individual predicted bin could be analyzed to further improved the estimate of the virality of the post. We experimented with various methods in creating bins, ranging from even bin sizes to uneven bin sizes with uniform distribution within the bins. However, these approaches all had about the same accuracy as randomly choosing a bin and as such, we pivoted to framing this as a regression problem. An surprising conclusion was the minimal effect of a post's content on virality predictions. Results from neither the image classification or the aesthetic scoring submodule provided desirable results, which contradicted findings from existing research[3]. However, it was interesting to see that the top and bottom 3 objects that corresponded to post virality seemed to fall within our initial assumptions. While it does seem that seem that objects within the images does have some effect, the effect seems to be negligible in popularity predictions.

Table 2: Top and bottom 3 objects detected within images and their respective weights (summed) that corresponded to post virality.

| Object Family | weight | WordNetID |
|---:|:---:|:---:|
| Domestic animal | 1.613 | n01317541 |
| Institution | 1.517 | n03574555 |
| Alcohol | 1.006 | n07884567 |
| Weapon | -1.047 | n04565375 |
| Substance | -1.131 | n00020090 |
| Tableware | -1.434 | n04381994 |

In truth, none of the tested combination of models performed as we had hoped. Comment count was also mentioned in related research[2] to possess strong correlation to post popularity, however we had trouble scraping comments with our adapted web-crawler. Another interesting avenue, would be to use comment count as an alternative target label. We believe that since it takes more effort to leave a comment on Instagram compared to a simple double-tap for likes, it is likely a more predictable measure of post popularity. As earlier stated, an LSTM might have been a better approach for this problem. However, with the dataset we generated, we only had 1,307 users, which is insufficient data to confidently support conclusions from such a network.

# REFERENCES

[1]: "A Gentle Introduction to Transfer Learning for Deep Learning." Machine Learning Mastery. November 25, 2018. https://machinelearningmastery.com/transfer-learning-for-deep-learning/.

[2]: Qian, C. J., M. A. Penza, J. D. Tang, and C. M. Ferri. "Instagram Popularity Prediction via Neural Networks and Regression Analysis."

[3]: Mazloom, M., R. Rietveld, S. Rudinac, M. Worring, and W. Dolen. "Multimodal Popularity Prediction of Brand-related Social Media Posts."

[4]: Wu, B., W. Cheng, Y. Zhang, Q. Huang, J. Li, and T. Mei. "Sequential Prediction for Social Media Popularity with Deep Temporal Context Networks."

[5]: gvsi. GitHub. https://github.com/gvsi/instagram-like-predictor/blob/master/dataset.csv.

[6]: Grossmann, T. GitHub. https://github.com/timgrossmann/instagram-profilecrawl.

[7]: "100 Best Instagram Accounts." Rolling Stone. https://www.rollingstone.com/interactive/features-the-100-best-instagram-accounts/.

[8]: Urgo. "Top 500 Most Followed Instagram Channels (Sorted by Followers Count)." Unboxtherapy YouTube Stats, Channel Statistics - Socialblade.com. https://socialblade.com/instagram/top/500/followers.

[9]: Titu1994. GitHub. https://github.com/titu1994/neural-image-assessment.

[10]: CS231n Convolutional Neural Networks for Visual Recognition. http://cs231n.github.io/transfer-learning/.

[11]: Donges, Niklas. "Transfer Learning – Towards Data Science." Towards Data Science. April 23, 2018. https://towardsdatascience.com/transfer-learning-946518f95666.

[12]: "A Gentle Introduction to Transfer Learning for Deep Learning." Machine Learning Mastery. November 25, 2018. https://machinelearningmastery.com/transfer-learning-for-deep-learning/.

[13]: "Introducing NIMA: Neural Image Assessment." Google AI Blog. December 18, 2017. https://ai.googleblog.com/2017/12/introducing-nima-neural-image-assessment.html

[14]: Meghawat, M., Y. Yin, S. Yadav, R. Ratn Shah, D. Mahata, and R. Zimmermann. "A Multimodal Approach to Predict Social Media Popularity." June 16, 2018.

[15]: ralbertazzi. GitHub. https://github.com/ralbertazzi/instagramlikeprediction/blob/master/IG/datasets/hashtags.txt.