Social media popularity prediction and recommendation

Interim Report 4

Group 11

Nitigya Pant - 2017355

Shashwat Jain - 2017103

***Abstract*—The aim of the paper is to determine the relevance of difference feature in determining the popularity of the image using the results and source code of paper "Social Media Popularity Prediction: A Multiple Feature Fusion Approach with Deep Neural Networks"**

Keywords*— *social media, feature extraction, Popularity

1. Introduction

Diverse variety of social media posts with varied distribution of likes, views and comments. Using models to analyse social media posts to determine their popularity.

1.1 Motivation

Internet penetration is increasing across the globe adding to the ever increasing base of social media users. Predicting popularity of posts in social media is highly significant considering the increased usage of social media and the new businesses that have emerged from it. The task of predicting popularity is useful for different types of users from the common user looking to increase their likes to large conglomerates and social media influencers working to increase their audience.

1.2 Proposed directions

We will be using the features of posts to determine popularity of posts. The features range from image aesthetics, user profile characteristics to tags associated with the image. These features will be fused with our deep neural network model to predict popularity. We will ultimately study the significance of individual features to analyse the impact of each.

1.3 Challenges

Several factors contribute to a post becoming popular from image aesthetics to the tags used by the user. Firstly, There exists no single factor that predicts the popularity of social media posts, our first task is to analyse what factors may affect the popularity of posts. Secondly, combining all the features like the NIMA model to measure image aesthetics and BERT model for text.

2. Literature Review

In recent years, there has been a shift in focus towards predicting popularity from social media platforms. They employ a similar pipeline to compute popularity scores for different types of social media content. We would focus on literature focused on predicting popularity of social media posts.

First is the kind which focuses on extracting features and then employs a regressor. We also studied some deep learning based models for popularity prediction in social media. Most of the popularity prediction studies only consider single mode like converting text into account without exploring other modalities which might limit their performance of popularity prediction.

Rsheed and Muhammad [34] built a model that can predict the popularity of news articles on Twitter by first classifying features of news articles into internal and external features and then use decision tree for prediction.

Some work has also been proposed to predict the popularity of photos. For instance, Kholsa et al. [33] predicted the popularity of photos by learning regression on both image content and user context.

3. Proposed Solution

3.1 Feature Extraction

1. *Visual Features*

    1.1. Image Features: We will extract image features from the post using ResNet-101 which is trained on ImageNet.

    1.2. Aesthetic Score: We will use the Nima model which is trained on the AVA database which will predict the aesthetic score of the image.

    1.3 Image Popularity - This is a very crucial point for a post to be viral. We will give an image popularity score to each image using the IIPA model, which is trained on Instagram using the deep ranking method and it gives a human-level performance.

2. *Text Feature*

    2.1. Deep Text Features: Posts have a caption, tag, and category. We use the BERT model to extract the text features of the post.

    2.2. Tags Count and Caption Length: We will use the number of tags and the caption length as the auxiliary features

3. *Numerical Features*

    3.1. Follower Count

    3.2. Following Count

    3.3. Likes Count

    3.4. Temporal Spatial Feature: How long the post has been posted for and geographical location of that post.

In addition to the Deep neural network-based regressor for the popularity prediction of the post, the project also aims to build a recommendation system, which will make certain recommendations to users to enhance the chances of their posts to be popular using feature extraction from the past popular posts of the particular user and available dataset of the collection of popular posts with asymmetric weightage to the result obtained from users past post.

The research paper aimed to predict the popularity of the post ,we tend to carry this research forward with the aim to provide the end user the correct recommendation before/while posting so that the user could optimise his/her post for the maximum attention.

4. Methodology

Labels	Description
Image id	Column 0
Category	Column 1
Tags	Column 2
Title	Column 3
Time and spatial	Column 4:6
User information	Column 7:9
Numerical features	Column 10:11
Image score	Column 12:13

First we will check the relevance of the feature by removing or fudging the values of a feature column and then calculating the mean of difference of the old and new values.If this value is significant then we assume that respective feature plays an important role in determining the popularity score of the post.

1. We take the values of each column and replace the values with random data and store this as a copy of the original data file for each column multiple times .
2. We calculate the popularity score for each copy of the original data .
3. We calculate the standard deviation of each of the popularity score of each post in original data file and the copy
4. We take the average of the standard deviation obtained by randomising same column
5. The standard deviation is used to calculate the influence score of the feature
6. we select the top features with maximum influence score greater than 0.7
7. We mutate the values of the top influence score on train file with different percentages
8. Analyse the change in the average popularity score on the train file with change in average value of the feature

9. Obtain the local maxima from the above analysis of the popularity score for change in each feature
10. Build a recommendation system such that the recommendation to shift the feature value towards the local maxima is given

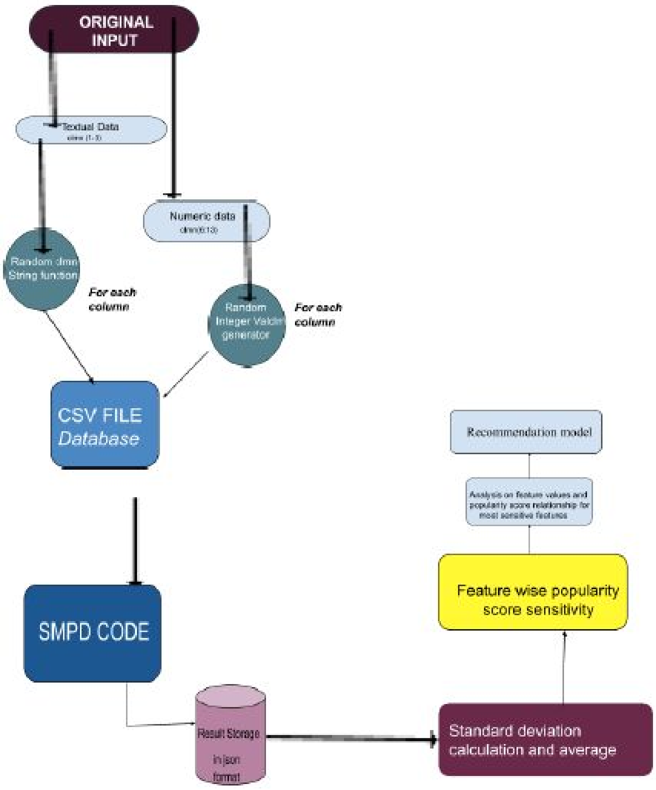

Figure 1.1 (Architecture of the system)

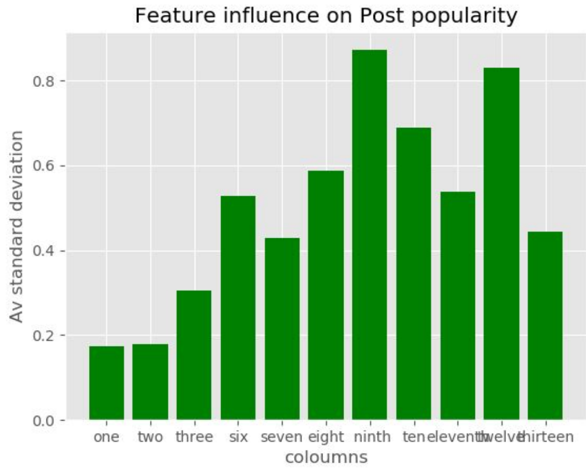

Figure 1.2 (Bar Graph)

Feature Column	Influence score
One	0.17
Two	0.18
Three	0.30
Six	0.52
Seven	0.43
Eight	0.59
Nine	0.87
Ten	0.68
Eleven	0.53
Twelve	0.83
Thirteen	0.44

Table 1.1

5. DATA ANALYSIS

LABELS	Columns	Score
Category	1	0.17
Tags	2	0.18
Title	3	0.30
Time and spatial	6	0.52
User Information	7:9	0.63
Numerical features	10:11	0.60
Image Score	12:13	0.63

Table 1.2

We can observe from table 1.2 that the user information and image score has the highest score , highlighting that these two play the most crucial role in determining the image popularity.

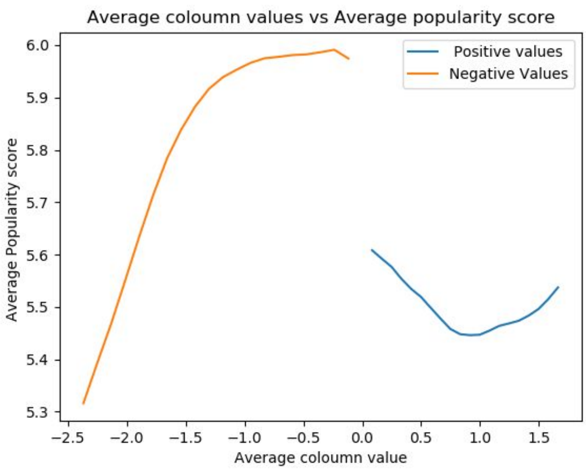

Figure 1.3

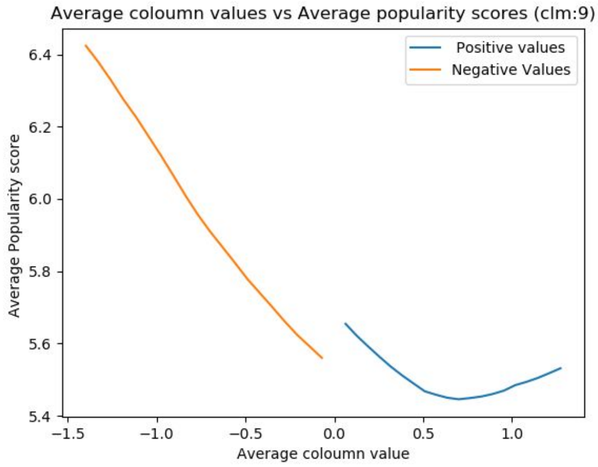

Figure 1.4

The Figure 1.4 and Figure 1.3 displays the relationship between the average column value after mutation and the average popularity for column 12 and 9 respectively.

6. Result

The recommendation system was on the inferences from the above analysis . The recommendation system ran for 5 test files with 100 test posts in each.

Testfile 1	76%
Testfile 2	77%
Testfile 3	80%
Testfile 4	77%
Testfile 5	58%

Table 1.3

Table 1.3 displays the recorded percentage of popularity score improvement for each test file.

7. Contribution

Shashwat Jain (2017103) - formulating the Methodology and implementation of recommendation system and it's testing (70%)

Nitigya Pant-Feature extraction and literature Survey (30%)

8. Data set and code reference

We are gratuitous to Prof. Ding Kenyan of The City University of Hong Kong to provide us the dataset with numerical values and images along with the source code which has helped us use the data points in providing inference for recommendation for social media popularity of posts.

8. References

[1] MP Challenge Organization. 2019. Social Media Prediction Challenge. http: //smp-challenge.com

Saeideh Bakhshi, David A Shamma, and Eric Gilbert. 2014. Faces engage us:

Photos with faces attract more likes and comments on Instagram. In SIGCHI

Conference on Human Factors in Computing Systems. 965–974.

[2] Ethem F Can, Hüseyin Oktay, and R Manmatha. 2013. Predicting retweet count

using visual cues. In ACM International Conference on Information & Knowledge

Management. 1481–1484.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert:

Pre-training of deep bidirectional transformers for language understanding.

arXiv:1810.04805 (2018).

[4] Keyan Ding, Kede Ma, and Shiqi Wang. 2019. Intrinsic Image Popularity Assess-

ment. ACM International Conference on Multimedia.

[5] Francesco Gelli, Tiberio Uricchio, Marco Bertini, Alberto Del Bimbo, and Shih-Fu

Chang. 2015. Image popularity prediction in social media using sentiment and

context features. In ACM International Conference on Multimedia. 907–910.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual

learning for image recognition. In IEEE Conference on Computer Vision and Pattern

Recognition. 770–778.

[7] Jack Hessel, Lillian Lee, and David Mimno. 2017. Cats and captions vs. creators

and the clock: Comparing multimodal content to context in predicting relative

popularity. In International Conference on World Wide Web. 927–936.

[8] Shintami Chusnul Hidayati, Yi-Ling Chen, Chao-Lung Yang, and Kai-Lung Hua.

2017. Popularity meter: An influence-and aesthetics-aware social media popular-

ity predictor. In ACM International Conference on Multimedia. 1918–1923.

[9] Chih-Chung Hsu, Chia-Yen Lee, Ting-Xuan Liao, Jun-Yi Lee, Tsai-Yne Hou, Ying-

Chu Kuo, Jing-Wen Lin, Ching-Yi Hsueh, Zhong-Xuan Zhang, and Hsiang-Chin

Chien. 2018. An iterative refinement approach for social media headline predic-

tion. In ACM International Conference on Multimedia. 2008–2012.

[10] Chih-Chung Hsu, Ying-Chin Lee, Ping-En Lu, Shian-Shin Lu, Hsiao-Ting Lai,

Chihg-Chu Huang, Chun Wang, Yang-Jiun Lin, and Weng-Tai Su. 2017. Social me-

dia prediction based on residual learning and random forest. In ACM International

Conference on Multimedia. 1865–1870.

[11] Feitao Huang, Junhong Chen, Zehang Lin, Peipei Kang, and Zhenguo Yang.

2018. Random forest exploiting post-related and user-related features for social

media popularity prediction. In ACM International Conference on Multimedia.

2013–2017.

[12] Aditya Khosla, Atish Das Sarma, and Raffay Hamid. 2014. What makes an image

popular?. In International Conference on World Wide Web. 867–876.

[13] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic opti-

mization. arXiv:1412.6980 (2014).

[14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. ImageNet classifica-

tion with deep convolutional neural networks. In Advances in Neural Information

Processing Systems. 1097–1105.

[15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. nature

521, 7553 (2015), 436.

[16] Liuwu Li, Sihong Huang, Ziliang He, and Wenyin Liu. 2018. An effective text-

based characterization combined with numerical features for social media head-line prediction. In ACM International Conference on Multimedia. 2003–2007.

[17] Liuwu Li, Runwei Situ, Junyan Gao, Zhenguo Yang, and Wenyin Liu. 2017. A

hybrid model combining convolutional neural network with xgboost for pre-

dicting social media popularity. In ACM International Conference on Multimedia.

1912–1917.

[18] Masoud Mazloom, Bouke Hendriks, and Marcel Worring. 2017. Multimodal

context-aware recommender for post popularity prediction in social media. In

Thematic Workshops of ACM Multimedia. 236–244.

[19] Masoud Mazloom, Robert Rietveld, Stevan Rudinac, Marcel Worring, and

Willemijn Van Dolen. 2016. Multimodal popularity prediction of brand-related

social media posts. In ACM International Conference on Multimedia. 197–201.

[20] Philip J McParlane, Yashar Moshfeghi, and Joemon M Jose. 2014. Nobody comes

here anymore, it's too crowded; predicting image popularity on Flickr. In ACM

International Conference on Multimedia Retrieval. 385–391.

[21] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013.

Distributed representations of words and phrases and their compositionality. In

Advances in neural information processing systems. 3111–3119.

[22] Naila Murray, Luca Marchesotti, and Florent Perronnin. 2012. AVA: A large-scale

database for aesthetic visual analysis. In IEEE Conference on Computer Vision and

Pattern Recognition. 2408–2415.

[23] SMHP Challenge Organization. 2018. Social Media Headline Prediction. https:

//social-media-prediction.github.io/PredictionChallenge

[24] SMP Challenge Organization. 2019. Social Media Prediction Challenge. http:

//smp-challenge.com

[25] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove:

Global vectors for word representation. In Conference on empirical methods in

natural language processing (EMNLP). 1532–1543.

[26] Hossein Talebi and Peyman Milanfar. 2018. NIMA: Neural image assessment.

IEEE Transactions on Image Processing 27, 8 (2018), 3998–4011.

[27] Wen Wang and Wei Zhang. 2017. Combining multiple features for image popu-

larity prediction in social media. In ACM International Conference on Multimedia.

1901–1905.

[28] Bo Wu. 2017. TPIC: A social media dataset for temporal popularity prediction.

https://social-media-prediction.github.io/TPIC2017

[29] Bo Wu, Wen-Huang Cheng, Yongdong Zhang, and Tao Mei. 2016. Time matters:

Multi-scale temporalization of social media popularity. In ACM International

Conference on Multimedia. 1336–1344.

[30] Bo Wu, Wen-Huang Cheng, Yongdong Zhang, Huang Qiushi, Li Jintao, and Tao

Mei. 2017. Sequential prediction of social media popularity with deep tempo-

ral context networks. In International Joint Conference on Artificial Intelligence

(IJCAI).

[31] Bo Wu, Tao Mei, Wen-Huang Cheng, and Yongdong Zhang. 2016. Unfolding

temporal dynamics: Predicting social media popularity using multi-scale temporal

decomposition.. In AAAI Conference on Artificial Intelligence. 272–278.

[32] Wei Zhang, Wen Wang, Jun Wang, and Hongyuan Zha. 2018. User-guided

hierarchical attention network for multi-modal social image popularity prediction.

In International Conference on World Wide Web. 1277–1

[33] A. Khosla, A. Das Sarma, and R. Hamid. What makes an image popular? In ACM WWW, pages 867–876, 2014.

[34] N. A. Rsheed and M. B. Khan. Predicting the popularity of trending arabic news on twitter. In ACM MEDES, pages 15–19, 2014