

Social Media Popularity Prediction based on Multi modal Self Attention Mechanisms

¹PALLANTLA HEMANKITHA, ²NEELAM VIMALA, ³NEMALIPURI SUMASRI, ⁴MEKALA VENKATESHWARLU, ⁵MODI GNANADEEP, ⁶Dr SHALINI.D

^{1,2,3,4,5}Btech STUDENT, VISAKHA INSTITUTE OF ENGINEERING AND TECHNOLOGY,
DEPARTMENT OF COMPUTER SCIENCE And ENGINEERING, VISAKHAPATNAM – 530027

⁶Professor, VISAKHA INSTITUTE OF ENGINEERING AND TECHNOLOGY,
DEPARTMENT OF COMPUTER SCIENCE AND SYSTEMS ENGINEERING, VISAKHAPATNAM – 530027

ABSTRACT

Social media popularity prediction is a significant task due to its many practical applications in the real world like ads, recommendatory systems, and trend prediction. Nevertheless, it is a difficult task since social media is influenced by various factors that are not easy to simulate (e.g. content quality, applicability to audiences, actual events). Other methods mostly take the greedy approach to incorporate as many factors and modalities as possible into their model but equally treat these features. To address this phenomenon, our designed method utilizes the self-attention mechanism to automatically and effectively combine various features to attain higher performance for the prediction of the popularity of a post, where the features utilized in our model can be broadly classified into two modalities, semantic (text) and numerical features. Using large-scale experiments and ablation studies on the training and testing sets of the difficult ACM Multimedia SMPD 2020 Challenge dataset, the testing results prove that the proposed method is significantly better than other methods. Machine learning is a significant part of the emerging discipline of data science. By applying statistical approaches, various type of algorithms is trained to classify or predict, and to reveal fundamental insights in this project. These insights then inform decision making within applications and companies, hopefully affecting core growth metrics. Machine learning algorithms construct a model from this project data, or training data, in an attempt to predict or make decisions without actually being programmed to do so. Machine learning algorithms apply to very large varieties of datasets for which it is not easy or even impossible to create traditional algorithms that will carry out the required tasks.

Keywords:- SMPD 2020, OCR,SVM,Decesion Tree, content quality

1. INTRODUCTION

SOCIAL media offers a public platform to share information with one another easily, and people spend a lot of time daily on social media platforms. As social media takes up most of the daily life of contemporary people, many are interested in studying how to pull data from social media. One example of data that might be obtained from social media is the popularity score. In particular, this score indicates the number of views of a post, and the higher number of views, the more influential. Social media popularity prediction (SMP) is estimating the popularity score based on the data observed for a given social media post. Estimating the popularity score is hard because of the many and complex factors that affect popularity. Quality of content and relevance to viewers are some of the factors, and these are difficult to measure. Other factors such as real-life events are tough to include in a prediction model. New SMP approaches seek to address such complicated factors by incorporating additional modalities [4, 5, 7, 12, 17], including images [14, 39], relation networks [25], time context [13], tags, and categories. While it is a good way to the papers, it adds complexity to the model, namely, architecture, memory usage, number of modules, etc. Alternatively, the paper [7, 26, 27, 28, 29, 30] is also a multi-modal approach but in its pipeline, it represented images as captions (i.e. texts). Different modalities could be converted to another modality using existing technologies. Image captioning converts

images to texts. There exist speech-to-text methods already. We could get various numeric values from the social graph of a post, for instance, the neighbors of each node. In addition, user information may influence the popularity of posts. According to most studies, there is strong correlation between image popularity and users [20, 32, 33]. One reason is that the users themselves also have followers, various users will have varied follower numbers. Typically, posts created by the user with the most followers will have a better possibility to be viewed more and liked more. And the time and space data also might have the impact on popularity, so the previous post has to appeal more people's attentions, if the user submits the post from a featured position, then more attention would also be captured.

In this work, we had suggested a network that utilized the semantic (text) and the numerical (number) modality to infer a social media post's popularity leveraging the self-attention mechanism. Because of the difference in data types, we separated the data into semantic and numerical paths. In the semantic branch, the content of images is mapped to caption words and tags, all the text features are transformed into tokens, each token has a corresponding word embedding [23], because the attention mechanism [9] is proven to effectively extract contextual information, to better aggregate sequence of embedding, we also design a feature attention mechanism for the purpose, which can handle dispensing recurrence, and convolutions completely. Only the semantic features modality is not enough for some social media post types, so we utilized the numerical features too which can be easily mapped into scalars, like timestamps, geo location. After preprocessing, we extracted and combined the features in both modalities respectively, and build two models to compute the popularity score. The contributions of this work are 3 fold:

2. LITERATURE SURVEY AND RELATED WORK

Social media sites create enormous amounts of data in a variety of formats, such as text, images, videos, and user interactions. Social media popularity prediction has attracted much interest in recent years due to its use in marketing, content suggestion, and information spreading. Traditional methods mostly focused on textual and metadata-based features, such as user interaction data, post time, and textual sentiment analysis. Early research utilized machine learning algorithms including regression models, decision trees, and support vector machines to forecast content popularity. These algorithms usually performed poorly in capturing the intricate dependencies between varied modalities and hence were not very predictive in nature.

With deep learning emerging as a research trend, researchers now utilize neural network-based methods, mainly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), to process text and image information. CNNs have been extensively utilized to feature extraction from images, and RNNs and LSTM networks have been used to process temporal relationships in text data. But these architectures lack the capability to combine multiple modalities efficiently, so the attention mechanism was invented. The Transformer model and self-attention mechanisms are introduced to change the game of natural language processing (NLP) and have also been applied to multi-modal tasks.

Multi-modal self-attention mechanisms have proven to be a robust method of social media popularity prediction by identifying cross-modal dependencies among textual, visual, and metadata features. Experiments that utilize Vision Transformers (ViTs) for image feature extraction and Transformer-based encoders for text processing have shown enhanced accuracy in user engagement prediction. Multi-modal fusion methods like cross-attention and late fusion have been used to combine information from various modalities in an effective manner. In addition, breakthroughs such as the Multimodal Transformer and BERT-based vision-language models have demonstrated better performance in processing intricate social media content.

3. Implementation Methodology

Khosla et al. [1] employed the image content and user context to forecast the popularity of images using millions of images. They systematically examined the effect of low-level, middle-level, and high-level features on prediction performance. Wu et al. [2] integrated a variety of time-scale dynamics into a sequential popularity prediction. Van Zwol examined the behaviors of users' social behavior on Flickr in a study [3]. He disclosed that the photos gained the most of their views in the first two days of uploading them. Furthermore, the images became popular depending on the owners' contacts and the social groups where he or she belonged. Various works are read on other websites as well. Hessel et al. [4] compared that the integration of visual and textual modalities usually results in optimal accuracies in predicting relative popularity on Reddit. Mazloom et al. [5] suggested that there exist a number of significant features, referred to as engagement parameters, including sentiment, vividness, and

entertainment. They employed the parameters for popularity prediction of posts related to brands on Instagram.

Most researchers forecasted social media popularity using ACM Multimedia Challenge 2019 or prior [29, 30, 31, 35]. For instance, Hsu et al. [7] used word-to-vector models to represent the text information and image semantic features obtained by image caption. Ding et al. [15] combined textural and numerical information with deep neural network methods to forecast the popularity score. Li et al. [19] introduced a Doc2Vec model and a good text-based feature fusion engineering, but these papers simply concatenated the various types of features and then input them into the regression model, without taking into account the correlation between various features. Hsu et al. [21] introduced an iterative refinement approach to offset prediction error and [22] calculated the view count of a post by residual learning. But this work only took a few limited forms of social media data, and there are plenty of useful data that can enhance the prediction performance.

As machine learning or deep learning develops rapidly, various vision-based works show up, for instance, Lin et al. [37] utilized several residual dense blocks to remove patterns. Yeh et al. [38] used a visual attention module to promote image classification ability. Ortis et al. [40] took visual and text data into account to carry out sentiment analysis using the SVM classifier, and Katsurai et al. [41] utilized the SentiWordNet to obtain sentiment information and combined the visual and text views to classify the post belongs positive or negative via SVM as well, but the SVM model cannot support the large-scale dataset, and it is difficult to extend to high dimensional data.

In 2016, He et al. [10] introduced a new deep network architecture, Residual Network (ResNet), in general, the deeper network will have better performance, but there is a degradation issue: when the layer number increases, the accuracy will drop. ResNet introduces an identity mapping mechanism to address issues of gradient vanishing and explosion.

4. Proposed Methodology

In this work, we presented a network that takes advantage of semantic (text) and numerical (number) modalities to predict the popularity of a social media post according to the self-attention mechanism. Because of the data type mismatch, we separated the data into semantic and numerical branches. In semantic branch, contents of the image are passed to caption texts and tags, all text features are transformed into tokens, and each token owns an associated with word embedding [23], as the attention mechanism [9] is demonstrated effective to obtain contextual information, to aggregate the sequence of embedding better, we also propose a feature attention mechanism for the aim, which could handle dispensing recurrence and convolutions completely. With only semantic features modality, it is not enough for certain kinds of social media posts, therefore we utilized the numerical features as well which can be easily transformed into scalars like timestamps, geolocation. We then preprocessed and extracted and combined the features in both modalities respectively, and construct two models to compute the popularity score. The contributions of this work are 3 fold. We designed a network that adopts an attention mechanism and exploits multiple features in two modalities to perform model ensemble, the network can be easily extended to include more different modalities furthermore, which is able to solve problems with heavy categories.

We examined the impact of semantic features on performance of the model. Additionally, we also created more numerical features, the outcome confirms the extracted features are helpful to enhance our network performance. We showed that our approach performs better than the other current best approaches in Social Media Popularity Dataset.

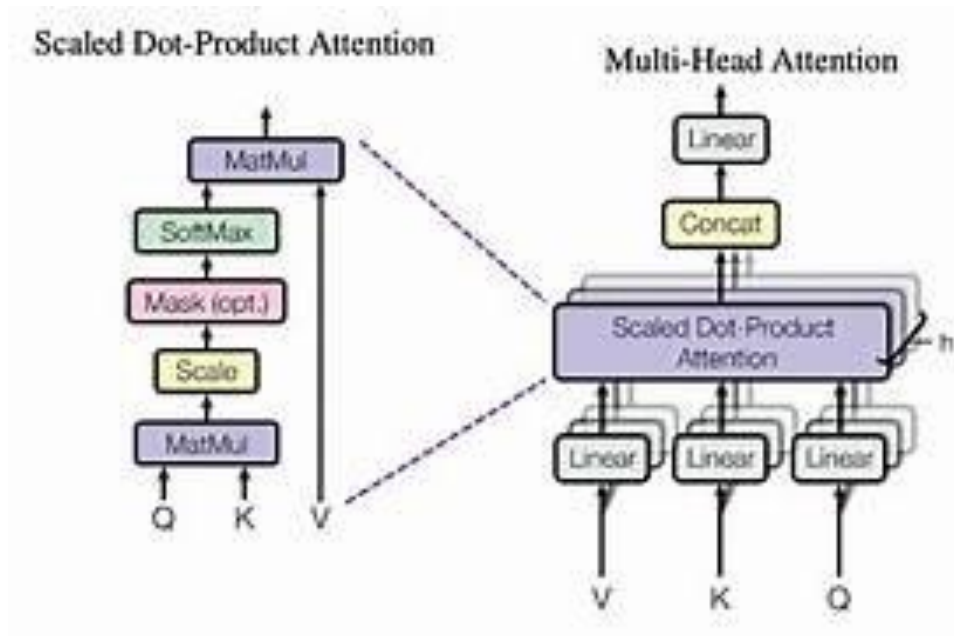


FIG1- SYSTEM ARCHITECTURE

5. METHODOLOGIES

5.1 MODULE

5.1.1 Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as

Login, Browse Datasets and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Predicted Social Media Popularity Posts, View Predicted Social Media Popularity Ratio, Download Predicted Data Sets, View Social Media Popularity Type Ratio Results, View All Remote Users.

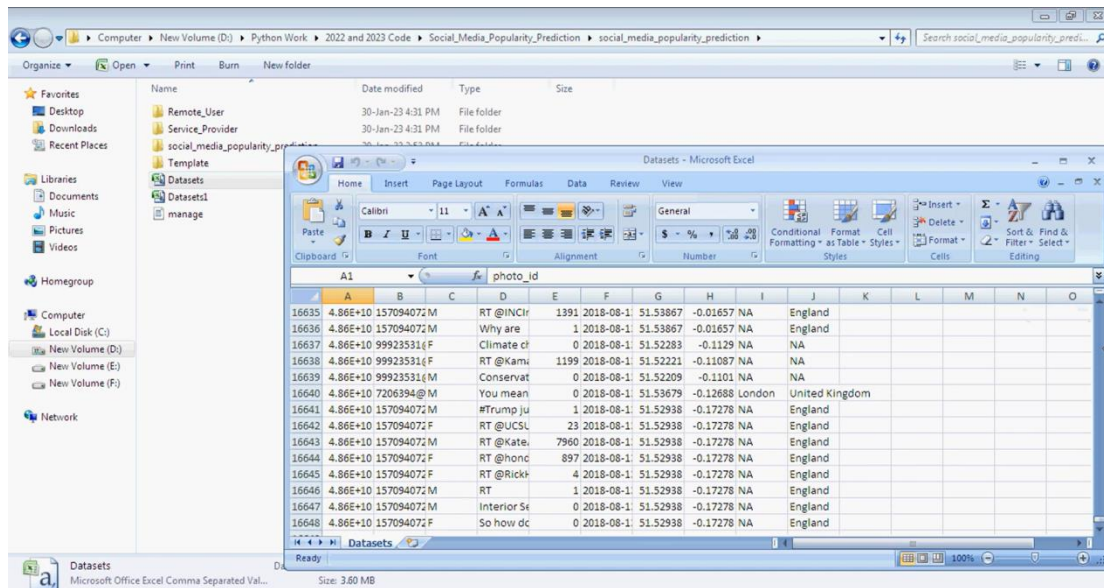
5.1.2 View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

5.1.2 Remote User

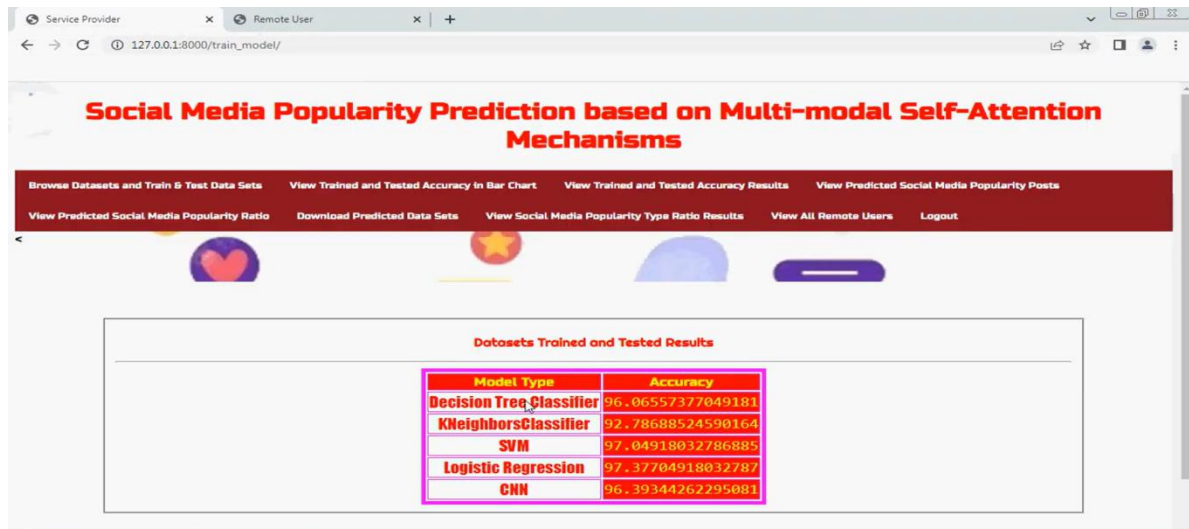
In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT SOCIAL MEDIA POPULARITY, VIEW YOUR PROFILE.

6. RESULTS AND DISCUSSION SCREEN SHOTS



| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|-------|----------|-----------|---|-------------|------|-----------|----------|----------|--------|----------------|---|---|---|---|---|
| 16635 | 4.86E+10 | 157094072 | M | RT @INC | 1391 | 2018-08-1 | 51.53867 | -0.01657 | NA | England | | | | | |
| 16636 | 4.86E+10 | 157094072 | M | Why are | 1 | 2018-08-1 | 51.53867 | -0.01657 | NA | England | | | | | |
| 16637 | 4.86E+10 | 999235314 | F | Climate c | 0 | 2018-08-1 | 51.52283 | -0.1129 | NA | NA | | | | | |
| 16638 | 4.86E+10 | 999235314 | F | RT @Kam | 1199 | 2018-08-1 | 51.52221 | -0.11087 | NA | NA | | | | | |
| 16639 | 4.86E+10 | 999235314 | M | Conservat | 0 | 2018-08-1 | 51.52209 | -0.1101 | NA | NA | | | | | |
| 16640 | 4.86E+10 | 7206394 | M | You mean | 0 | 2018-08-1 | 51.53679 | -0.12688 | London | United Kingdom | | | | | |
| 16641 | 4.86E+10 | 157094072 | M | #Trump ju | 1 | 2018-08-1 | 51.52938 | -0.17278 | NA | England | | | | | |
| 16642 | 4.86E+10 | 157094072 | F | RT @UCSL | 23 | 2018-08-1 | 51.52938 | -0.17278 | NA | England | | | | | |
| 16643 | 4.86E+10 | 157094072 | M | RT @Kate | 7960 | 2018-08-1 | 51.52938 | -0.17278 | NA | England | | | | | |
| 16644 | 4.86E+10 | 157094072 | F | RT @hond | 897 | 2018-08-1 | 51.52938 | -0.17278 | NA | England | | | | | |
| 16645 | 4.86E+10 | 157094072 | F | RT @Rick | 4 | 2018-08-1 | 51.52938 | -0.17278 | NA | England | | | | | |
| 16646 | 4.86E+10 | 157094072 | M | RT | 1 | 2018-08-1 | 51.52938 | -0.17278 | NA | England | | | | | |
| 16647 | 4.86E+10 | 157094072 | M | Interior Se | 0 | 2018-08-1 | 51.52938 | -0.17278 | NA | England | | | | | |
| 16648 | 4.86E+10 | 157094072 | F | So how dc | 0 | 2018-08-1 | 51.52938 | -0.17278 | NA | England | | | | | |

Fig 2:- sample dataset



| Model Type | Accuracy |
|--------------------------|-------------------|
| Decision Tree Classifier | 96.06557377049181 |
| KNeighborsClassifier | 92.78688524590164 |
| SVM | 97.04918032786885 |
| Logistic Regression | 97.37704918032787 |
| CNN | 96.39344262295081 |

Fig 3:- Accuracy of different machine learning Algorithm

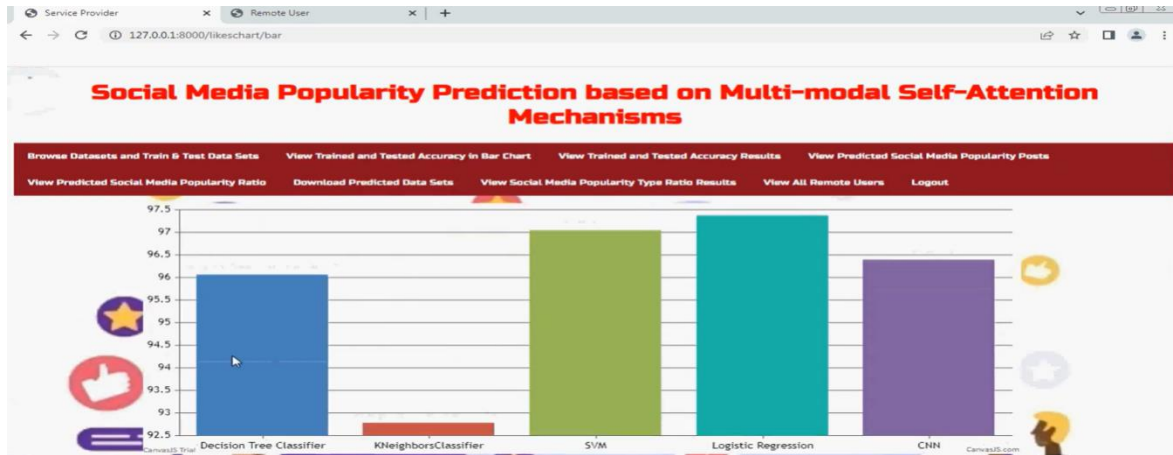


Fig 4:- Accuracy Graph of different machine learning Algorithm

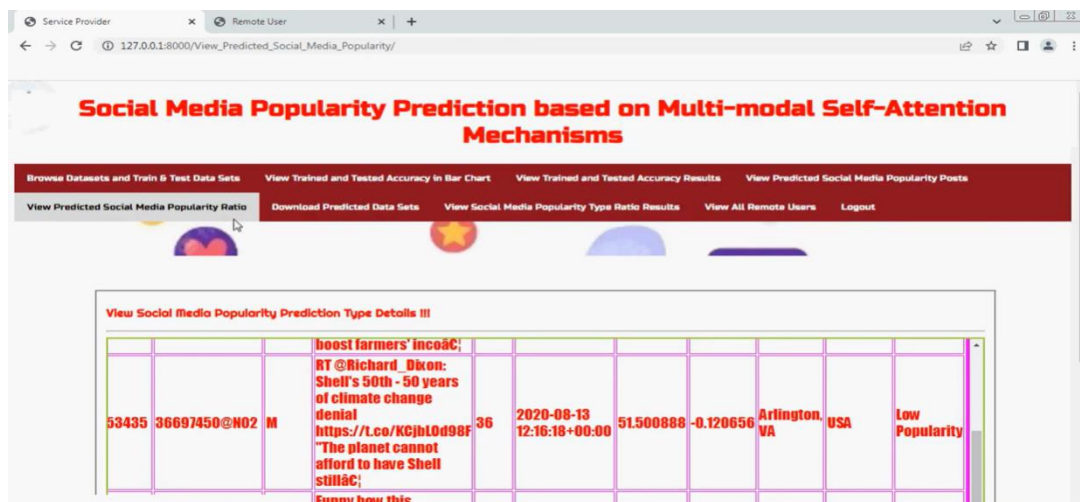


Fig 5:- predicted popularity of the test dataset

7. CONCLUSION AND FUTURE SCOPE

7.1 CONCLUSION

In this work, we introduced a social media popularity prediction approach with multi-modal input and attention-based mechanisms. Our approach specifically employs semantic and numerical features to calculate the popularity score. Semantic features are text-based and sequential so attention-based networks (i.e. Transformer) have good synergy with this task. We also transformed images to semantic features with existing image captioning algorithms. In addition, we extended the already present numerical attributes to enhance our model's performance. We proved that our model works fairly good in comparison with other state-of-the-art solutions.

7.2 Future Work

In this work, we investigated predicting social media popularity via multi-modal self-attention mechanisms. Although our method has achieved encouraging results, there are a number of research directions available that can improve performance and relevance further. One direction is to incorporate more diverse modalities in addition to text, image, and metadata. For example, the inclusion of audio and video capabilities might give a more comprehensive view of social media posts, particularly on platforms such as Instagram and TikTok, where video posts are a major contributor to engagement. Further, data streams in real-time might be utilized to dynamically tune predictive models so that more accurate predictions can be made in fluctuating social media contexts.

A second promising line of work is the investigation of more sophisticated attention mechanisms. Although self-attention proved successful in capturing intra-modal and inter-modal relationships, the integration of transformer-based architectures with graph neural networks (GNNs) could strengthen modeling of relationships among users, content, and engagement patterns. In addition, constructing hierarchical attention mechanisms that take into account global trends as well as localized user interactions can better enable the model to generalize across various social media platforms.

8. REFERENCES

- [1] Aditya Khosla, Atish Das Sarma, and Raffay Hamid, "What makes an image popular?," International Conference on World Wide Web., p.p.867–876. 2014.
- [2] Bo Wu, Wen-Huang Cheng, Yongdong Zhang, and Tao Mei, "Timematters: Multi-scale temporalization of social media popularity," ACM International Conference on Multimedia., p.p. 1336–1344. 2016.
- [3] R. van Zwol, "Flickr: Who is Looking?," IEEE/WIC/ACM International Conference on Web Intelligence., p.p. 184-190. 2017.
- [4] Jack Hessel, Lillian Lee, and David Mimno, "Cats and captions vs. creators and the clock: Comparing multi-modal content to context in predicting relative popularity," International Conference on World Wide Web., p.p. 927–936. 2017.
- [5] Masoud Mazloom, Robert Rietveld, Stevan Rudinac, Marcel Worring, and Willemijn Van Dolen, "Multimodal Popularity Prediction of Brand-related Social Media Posts," ACM International Conference on Multimedia., p.p. 179-201. 2016.
- [6] SMP Challenge Organization. 2020. Social Media Prediction Challenge. Available: <http://smp-challenge.com>
- [7] Chih-Chung Hsu, Li-Wei Kang, Chia-Yen Lee, Jun-Yi Lee, Zhong-Xuan Zhang, and Shao-Min Wu, "Popularity Prediction of Social Media based on Multi-Modal Feature Mining," ACM International Conference on Multimedia., p.p. 2687–2691. 2019.
- [8] Francesco Gelli, Tiberio Uricchio, Marco Bertini, Alberto Del Bimbo, and Shih-Fu Chang, "Image popularity prediction in social media using sentiment and context features," ACM International Conference on Multimedia., p.p. 907–910. 2015.
- [9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin, "Attention is all you need," International Conference on Neural Information Processing Systems., p.p. 6000–6010. 2017.
- [10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," IEEE Conference on Computer Vision and Pattern Recognition., p.p. 770–778. 2016.
- [11] Y. Liu and Myle Ott and Naman Goyal and Jingfei Du and Mandar Joshi and Danqi Chen and Omer Levy and M. Lewis and Luke Zettlemoyer and Veselin Stoyanov, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv preprint arxiv.org/abs/1907.11692, 2019.