ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

Social media popularity prediction based on multi modal self attention mechanisms

[1] Mrs.A.N.RAMA MANI [2] JAVVADI LAKSHMI KAVYA

[1]Associate Professor , S.V.K.P & Dr K.S. Raju Arts & Science College(A),Penugonda, W.G.District,Andhra Pradesh,ramasrinivasu2@gmail.com

[2]PG, scholar , S.V.K.P & Dr K.S. Raju Arts & Science College(A), Penugonda,W.G.District,Andhra Pradesh,kavyajavvadi79@gmail.com

ABSTRACT

Popularity prediction using social media is an important task because of its wide range of real-world applications such as advertisements, recommendation systems, and trend analysis. However, this task is challenging because social media is affected by multiple factors that cannot be easily modeled (e.g. quality of content, relevance to viewers, real-life events). Usually, other methods adopt the greedy approach

to include as many modalities and factors as possible into their model but treat these features equally. To solve this phenomenon, our proposed method leverages the self-attention mechanism to effectively and automatically fuse different features to achieve better performance for the popularity prediction of a post, where the features used in our model can be mainly categorized into two modalities, semantic (text)

and numeric features. With extensive experiments and ablation studies on the training and testing data of the challenging ACM Multimedia SMPD 2020 Challenge dataset, the evaluation results demonstrate the effectiveness of the proposed approach as compared with other methods.

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, different type of algorithms is trained to make classifications or predictions, and to uncover key insights in this project. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics.

Machine learning algorithms build a model based on this project data, known as training

data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of datasets, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

1.INTRODUCTION

SOCIAL media provides a public platform to easily exchange information with each other, and nowadays people spend a lot of time every day on various social media platforms. Since social media occupies a large part of the daily lives of modern people, many people are interested in researching how to extract data from social media. An example of information that could be gained from social media is the popularity score. Specifically, this score tells how many people viewed a post, and a larger number of views means more influence. Social media popularity prediction (SMP) is the task of estimating the popularity score using the available data of 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. Recent SMP methods attempt to tackle these complex factors by adding more modalities [4, 5, 7, 12, 17], such as images [14, 39], relationship networks [25], temporal context [13], tags, and categories.

Although increasing the number of modalities is a good approach to the works, it also increases the complexity of the model, in terms of architecture, memory consumption, 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. From

the social graph of a post, we could extract different numeric values, such as the number of the neighbors for each node.

Moreover, the popularity of posts may be affected by user information. Many studies have shown that there is a high correlation between image popularity and users [20, 32, 33]. One of the reasons is that the users have their own followers, different users may have different numbers of followers. Generally, posts written by the user with more followers have a higher chance to receive more views and likes. And the temporal and spatial information may affect the popularity as well, the earlier post should get more

people's attention, and if the user uploads the post in a special location, it will attract more attention too.

In this paper, we proposed a network that exploits semantic (text) and numerical (number) modalities to estimate the popularity of a social media post based on the self-attention mechanism. Due to the data type discrepancy, we divided the data into semantic and numerical branches. In the semantic branch, the image contents are transferred to caption texts and tags, all of the textual features are converted into tokens, each token has an associated with word embedding [23], since the attention mechanism [9] is shown effective to extract contextual information, to better aggregate the sequence of embedding, we also develop a feature attention mechanism for the purpose, which can deal with dispensing recurrence, and convolutions entirely. Using only the semantic features modality is not sufficient for some types of social media posts, so we used the numerical features as well which can be easily converted into scalars, such as timestamps, geo location. After preprocessing, we extracted and fused the features in both modalities respectively, and assemble two models to calculate 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 analyzed the influence of semantic features on the model performance. Moreover, we generated additional numerical features, the result indicates the derived features are beneficial to improve our network performance. We demonstrated that our method outperforms the other state-of-the-art methods in Social Media Popularity Dataset.

2.LITERATURE SURVEY

The prediction of social media popularity has garnered significant attention due to its implications in marketing, content creation, and information dissemination. Traditional methods primarily employed statistical models and regression techniques based on basic features such as post time, number of followers, and historical engagement data. However, with the advent of multi-modal data—including text, images, videos, and metadata—researchers have explored more sophisticated approaches. Multi-modal self-attention mechanisms have emerged as a promising method to enhance prediction accuracy by dynamically integrating and weighting diverse data types.

Self-attention mechanisms, which form the core of transformer models, allow for the computation of attention scores that highlight the importance of different parts of the input data. When applied to multi-modal data, these mechanisms can adjust the focus on various modalities (such as text, images, and metadata) to improve predictive performance. Various integration techniques, including concatenation, co-attention, and hierarchical attention, have been developed to leverage the strengths of each modality effectively.

Notable advancements include the adaptation of transformer models, such as BERT, for multi-modal contexts. Models like ViLT (Vision-and-Language Transformer) and MMBT (Multi-ModalBitransformers) demonstrate how visual and textual data can be processed together to

predict social media engagement metrics. These models have shown promising results in applications ranging from marketing and advertising to content creation and information dissemination.

Despite these advancements, challenges remain in integrating heterogeneous data sources and handling large-scale data efficiently. Future research directions include developing more sophisticated multi-modal fusion techniques, improving the scalability and efficiency of models, and enhancing the personalization of predictions by incorporating user-specific preferences and behaviors.

Overall, the integration of multi-modal self-attention mechanisms in social media popularity prediction represents a significant leap forward, offering more accurate and nuanced insights into user engagement. Continued research in this field holds the potential for even more effective applications in various domains.

3. EXISTING SYSTEM

Khosla et al. [1] used the image content and the user context to predict the image popularity based on millions of images. They methodically analyzed the impact of low-level, middle-level, and high-level features on prediction accuracy. Wu et al. [2] merged multiple time-scale dynamics into a sequential prediction of popularity. In [3], Van Zwol studied the characteristics of users' social behavior on Flickr. He revealed that photos received the majority of their views within the first two days of being uploaded. Moreover, the popularity of images was influenced by the owners' contacts and social groups to which he or she belonged. There are also several works studied on other platforms. Hessel et al. [4] analyzed that the combination of visual and textual modalities generally leads to the best accuracies for predicting relative popularity on Reddit. Mazloom et al. [5] proposed that there are several important features, called engagement parameters, such as sentiment, vividness, and entertainment. They used these parameters for predicting

the popularity of brand-related posts on Instagram.

Many researchers predicted social media popularity based on ACM Multimedia Challenge 2019 or earlier [29, 30, 31, 35] . For example, Hsu et al. [7] employed word-to-vector models to encode the text information and image semantic features extracted by image caption. Ding et al. [15] fused textural and numerical data with deep neural network techniques to predict the popularity score. Li et al. [19] presented a Doc2Vec model and an effective text-based feature fusion engineering, but these works only concatenated the different types of features then fed them to the regression model, they did not consider the correlation between different features. Hsu et al. [21] proposed an iterative refinement method to compensate for prediction error and [22] computed the view count of a post by residual learning. However, this works only adopted limited types of social media data, there are still a lot of useful data that can improve the performance of prediction.

With the rapid development of machine learning or deep learning, many works present vision-based applications, for example, Lin et al. [37] employed multiple residual dense blocks to perform pattern removal. Yeh et al. [38] proposed a visual attention module to enhance image classification capability. Ortis et al. [40] considered visual and textual information to perform sentiment analysis through the SVM classifier, and Katsurai et al. [41] exploited the SentiWordNet to retrieve sentiment information and fused the visual and textual views to classify the post belongs positive or negative via SVM as well, however, the SVM model cannot afford the large-scale dataset, and it is hard to apply to high dimensional data.

In 2016, He et al. [10] proposed a novel deep learning architecture, Residual Network (ResNet), generally, the deeper network will get better performance, however, there exists a degradation problem: when the number of layers increases, the accuracy will decrease. ResNet has an identity mapping mechanism

to solve problems of gradient vanishing and explosion.

4 PROPOSED SYSTEM

In this paper, we proposed a network that exploits semantic (text) and numerical (number) modalities to estimate the popularity of a social media post based on the self-attention mechanism. Due to the data type discrepancy, we divided the data into semantic and numerical branches. In the semantic branch, the image contents are transferred to caption texts and tags, all of the textual features are converted into tokens, each token has an associated with word embedding [23], since the attention mechanism [9] is shown effective to extract contextual information, to better aggregate the sequence of embedding, we also develop a feature attention mechanism for the purpose, which can deal with dispensing recurrence, and convolutions entirely. Using only the semantic features modality is not sufficient for some types of social media posts, so we used the numerical features as well which can be easily converted into scalars, such as timestamps, geolocation. After preprocessing, we extracted and fused the features in both modalities respectively, and assemble two models to calculate 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 analyzed the influence of semantic features on the model performance. Moreover, we generated additional numerical features, the result indicates the derived features are beneficial to improve our network performance.

We demonstrated that our method outperforms the other state-of-the-art methods in Social Media Popularity Dataset.

5 SYSTEM ARCHITECTURE

System Architecture mainly consists of 2 modules and database to store all the data. Those are:

a) Service provider

b) Remote user

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

This modules are the key features of the project. The User uploads racist contained tweets datasets in the upload datasets submodule which is available in the user module.The User module can perform the following operation : Register and Login, My Profile, Upload Datasets, View all uploaded datasets, Find sentiment type by hashcode, Find Sentiment Type.

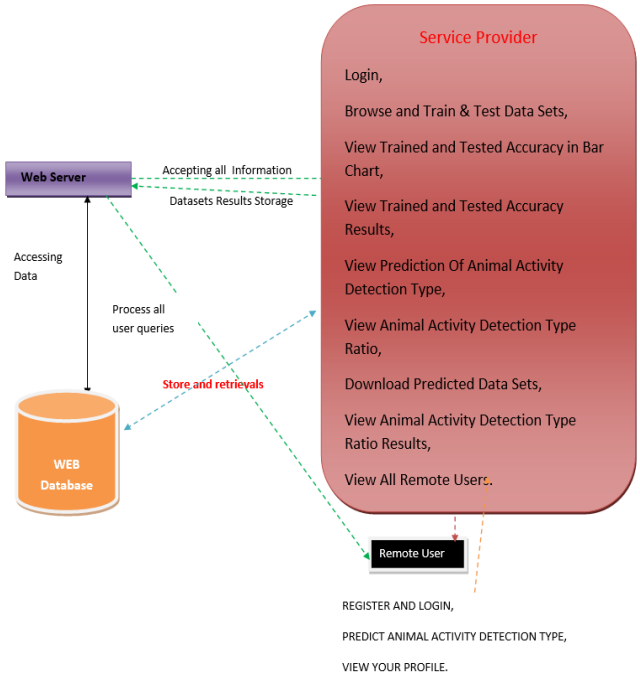
The Admin module which authorizes the users and can see all the user uploaded datasets. The admin module can perform the following operation such as Login, View All users and authorize, view all datasets, View All datasets by GRU, View sentiment type

6 MODULES INVOLVED

- **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 Browse and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of Animal Activity Detection Type, View Animal Activity Detection Type Ratio,

Architecture Diagram



Download Predicted Data Sets, View Animal Activity Detection Type Ratio Results, View All Remote Users.

- **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.

Remote User

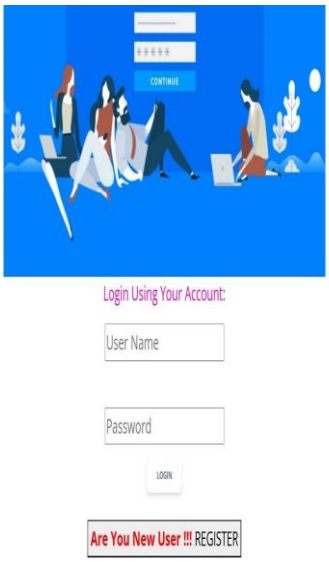
In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

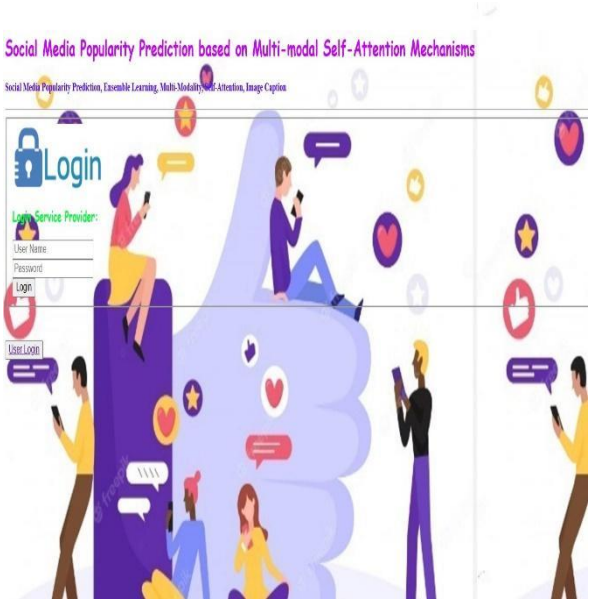
login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT ANIMAL ACTIVITY DETECTION TYPE, VIEW YOUR PROFILE.

7. SCREEN

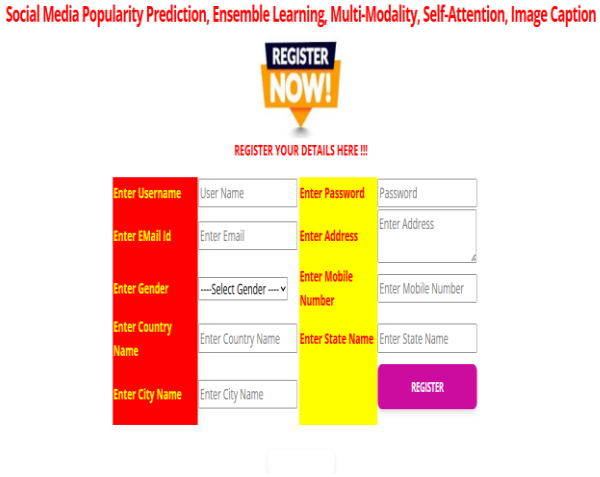
Login Page



User Login

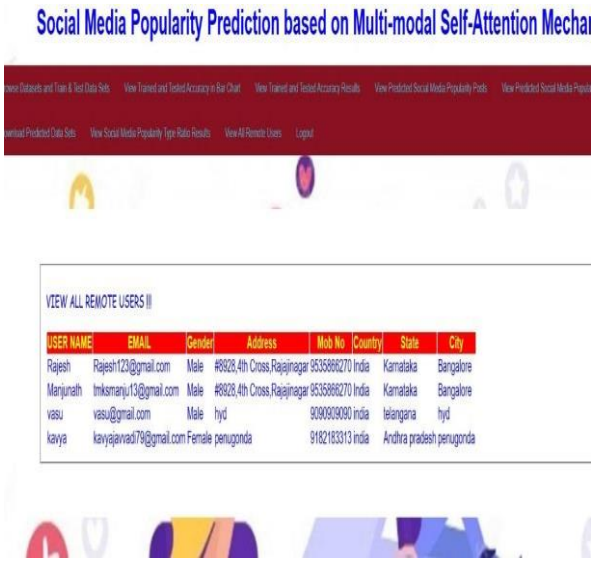


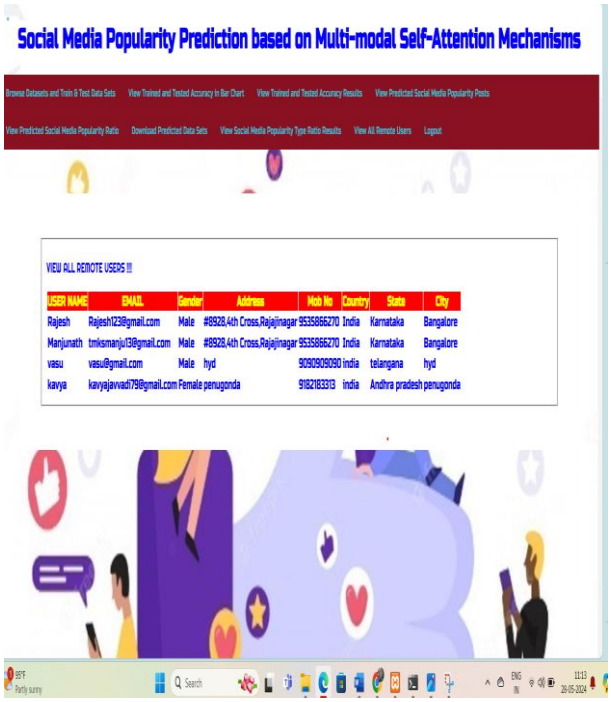
User Registration



Remote users

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

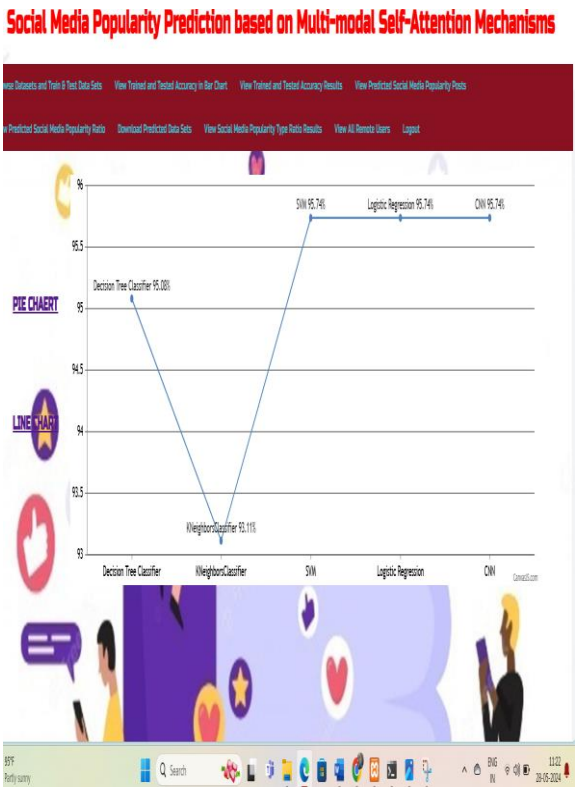


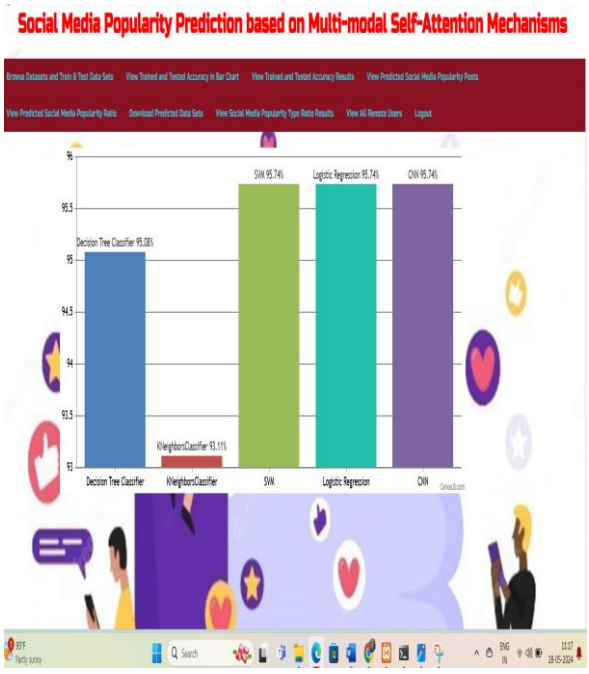


ISSN2321-2152

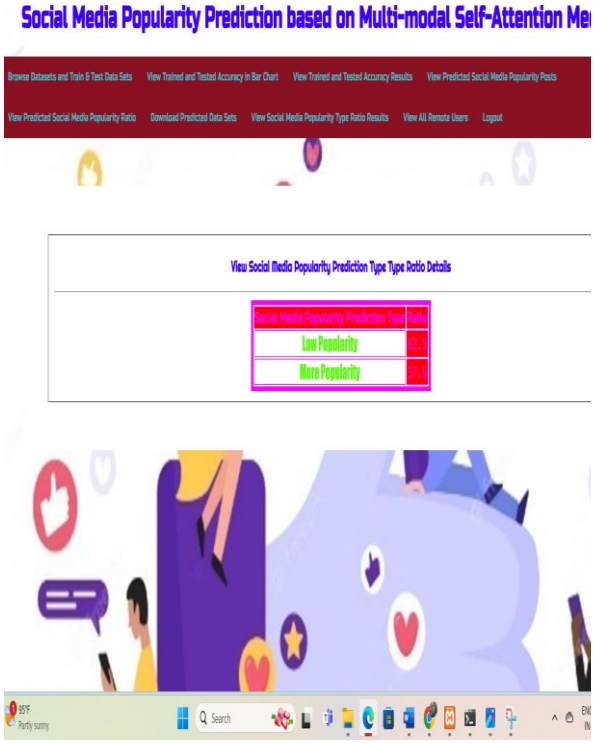
www.ijmece .com

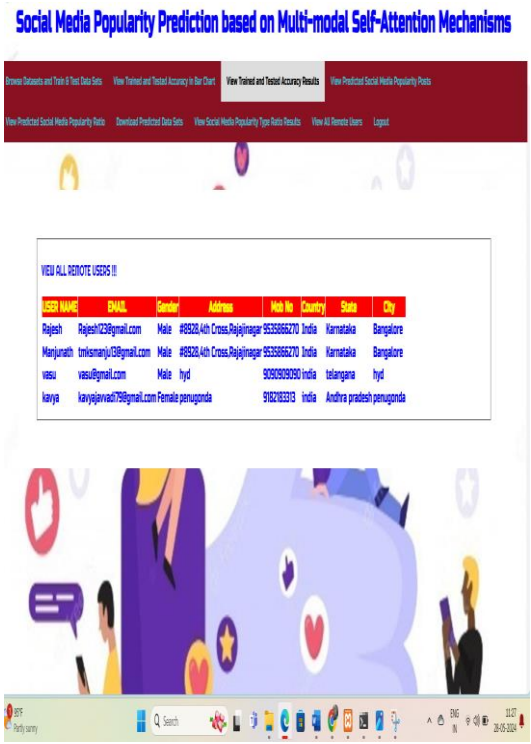
Vol 12, Issue 2, 2024

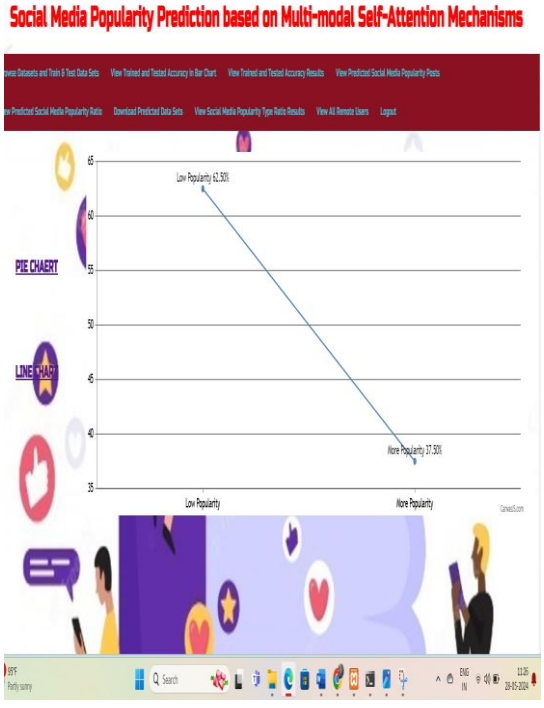




ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024







CONCLUSION

In this paper, we proposed a social media popularity prediction method with multi-modal input and attention-based mechanisms. Specifically, our method uses semantic and numerical features to compute the popularity score. Semantic features are text-based and sequential hence attention-based networks (i.e. Transformer) have good synergy with this task. We also converted images to semantic features using existing image captioning algorithms. Furthermore, we augmented the existing numerical features to increase the performance of our model. We showcased that our method performs reasonably well against other state-of-the-art methods.

9 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, "Time matters: 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, MarcelWorring, 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.

[12] Douwe Kiela, Suvrat Bhooshan, Hamed Firooz and Davide Testuggine, “Supervised multi-modal Bitransformers for Classifying Images and Text,” arXiv preprint arxiv.org/abs/1909.02950., 2019.

[13] Bo Wu, Wen-Huang Cheng, Yongdong Zhang, Qiushi Huang, Jintao Li, and Tao Mei, “Sequential Prediction of Social Media Popularity with Deep Temporal Context Networks,” International Joint Conference on Artificial Intelligence., p.p. 3062–3068. 2017.

[14] Zehang Lin and Feitao Huang and Yukun Li and Zhenguo Yang and W. Liu, “A layer-wise deep stacking model for social image popularity prediction,” International Conference on World Wide Web., p.p. 1639-1655. 2018.

[15] Keyan Ding, Ronggang Wang, and Shiqi Wang, “Social Media Popularity Prediction: A Multiple Feature Fusion Approach with Deep Neural Networks,” ACM International Conference on Multimedia., p.p. 2682–2686. 2019.

[16] Ziliang He, Zijian He, Jiahong Wu, and Zhenguo Yang, “Feature Construction for Posts and Users Combined with LightGBM for Social Media Popularity Prediction,” ACM International Conference on Multimedia., p.p. 2672–2676. 2019.

[17] Junhong Chen, Dayong Liang, Zhanmo Zhu, Xiaojing Zhou, Zihan Ye, and

Xiuyun Mo, “Social Media Popularity Prediction Based on Visual-Textual Features with XGBoost,” ACM International Conference on Multimedia., p.p. 2692–2696. 2019.

[18] Peipei Kang, Zehang Lin, Shaohua Teng, Guipeng Zhang, Lingni Guo, and Wei Zhang, “Catboost-based Framework with Additional User Information for Social Media Popularity Prediction,” ACM International Conference on Multimedia., p.p. 2677–2681. 2019.

[19] Liuwu Li, Sihong Huang, Ziliang He, and Wenyin Liu, “An Effective Textbased Characterization Combined with Numerical Features for Social Media Headline Prediction,” ACM International Conference on Multimedia., p.p. 2003–2007. 2018.

[20] E Ethem F. Can, Hüseyin Oktay, and R. Manmatha, “Predicting retweet count using visual cues,” ACM international conference on Information & Knowledge Management., p.p. 1481–1484. 2013.

[21] 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, “An iterative refinement approach for social media headline prediction,” ACM International Conference on Multimedia., p.p. 2008–2012. 2018.

[22] 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, “Social media prediction based on residual learning and random forest,” ACM International Conference on Multimedia., p.p. 1865–1870. 2017.

[23] Tomas Mikolov, Edouard Grave, Piotr Bojanowski, Christian Puhrsch, Armand Joulin, “Advances in pre-training distributed word representations,” International Conference on Language Resources and Evaluation., Available: http://www.lrecconf.org/proceedings/lrec2018/pdf/721.pdf. 2018.

[24] BoWu,Wen-Huang Cheng, Peiye Liu, Bei Liu, Zhaoyang Zeng, and Jiebo Luo, “SMP Challenge: An Overview of Social Media Prediction Challenge 2019,” ACM International Conference on Multimedia., p.p. 2667-2671. 2019.

[25] Qi Cao, Huawei Shen, Jinhua Gao, Bingzheng Wei, and Xueqi Cheng, “Popularity prediction on social platforms with coupled graph neural networks,” International Conference on Web Search and Data Mining., p.p. 70-78. 2020.