ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 3, 2024

Social Media Popularity Prediction based on Multi modal Self Attention Mechanisms

[1]K. SUPARNA, [2]VARDHANAPU RATHNA HARIKA

[1](Assistant Professor), MSC, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM, ANDHRA PRADESH**

[2]MSC, scholar, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM, ANDHRA PRADESH**

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 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.

2. 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 mean 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, and 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.

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, and 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 pos 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.

3. SYSTEM STUDY

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 Flicker. 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. Hassel 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 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 perform 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.

Disadvantages

An existing methodology doesn't implement SEMANTIC FEATURE EXTRACTION method.

The system not implemented ENSEMBLE REGRESSOR MODEL for the datasets.

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

ISSN2454-9940
www.ijasem.org
Vol 18, Issue 3, 2024

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.

Advantages

Caption Features — Social media posts could have images or videos attached. To simplify the pipeline of our method, these attached images and videos are converted

to text using a pre-trained captioning model [7, 12] and are treated similarly as textual features.

Categorical Features — A social media post could be categorized using different systems. In this paper, a Flicker post has different levels of categorization, which are: (main) category, subcategory, concept descriptions. There are 11 classes for categories, 77 classes for subcategories, and 668 classes of concept descriptions.

Tag Features — Tag features are composed of several keywords given by the user when they are creating a post, the tags are arbitrary information, for example, the styles, location, or holiday.

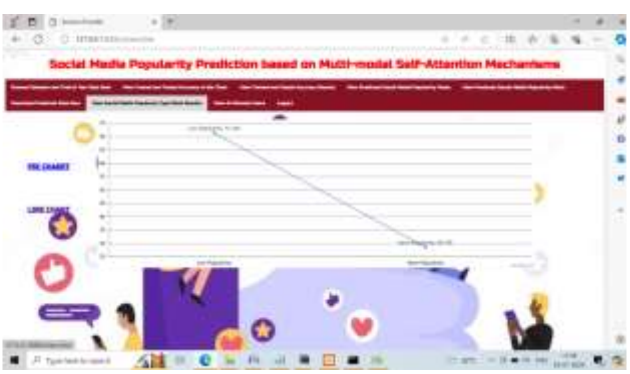
5. SCREEN SHORTS

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 3, 2024



6. 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.

7. REFERENCE

[1] Aditya Khosla, Atish Das Sarma, and Raffay Hamid, “What makesv 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.