90.3k #PleaseUpVote: Virality Prediction with Title and Thumbnail Image on Reddit

Xiaowen Lin, Zijian Wang, Chaonan Ye

{veralin, zijwang, yec0214}@stanford.edu

Stanford

Introduction

Reddit is the 5th most visited website in the United States. A campaign or advertisement post can potentially reach millions of target audience. However, there is little work on predicting the virality, or popularity, of Reddit posts. This is indeed a hard task because even humans could hardly distinguish between two posts with different popularity. Here, we propose to train a multimodal neural network based on titles and thumbnails. Our results show that our model could capture the nuance signal of virality, and joining image and text information yields the best result.

Data

We extracted four subreddits from Reddit dump from 2015 to 2018. Titles are included in the dump while thumbnails were crawled and resized or padded to 224 x 224. Labels are derived from the upvotes of the posts. We split the dataset as 90% training, 5% dev, and 5% test.

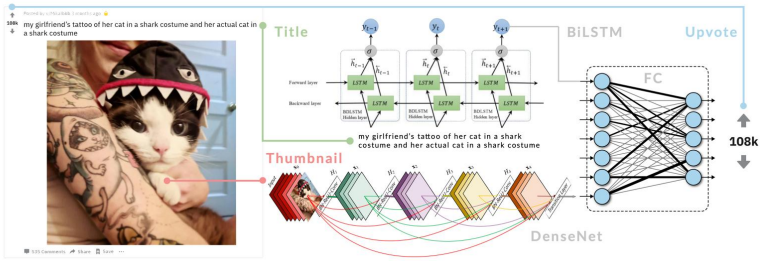
Subreddit	Number of Entries	Number of Avail. Images
aww	1,720,414	1,076,371
politics	1,432,923	545,644
The_Donald	4,573,934	1,858,266
EarthPorn	280,745	129,251

Table 1: Sizes of datasets with numbers of available images

Hyperparameters

- Data: **Oversample** or not
- Image: **DenseNet** / ResNet; w/wo pre-training
- Text embedding: Word2Vec / Glove / Char / **Non**
- Optimizer: **SGD w/ momentum**, Adam or AMSGrad

Models



We investigate different state-of-the-art models in this project as follows:

- Image: DenseNet and ResNet with different sizes.
- Text: two-stack bi-directional LSTMs, with character-level and word-level embeddings.
- Multimodal: joining the feature layers of the pretrained models with fully-connected layers on top.

Results

For the following section, we focus on `aww` as i) it has a large amount of data, and ii) it is less temporal. We formulate the task as a two-class classification problem, where we take the posts less than *mean - std* upvotes as negative, and more than *mean + std* as positive.

	Method	Macro-F1
Image	ResNet + P + SGD	0.678
	DenseNet + NP + SGD	0.653
	DenseNet + P + SGD	**0.702**
	DenseNet + P + Adam	0.681
Text	LSTM + NP + SGD	**0.618**
	LSTM + W2V + SGD	0.603
	LSTM + Glove + SGD	0.617
	LSTM + Char + SGD	0.609
	Multimodal	**0.738**

Table 2: Performance comparisons between different models.

Discussions

- We demonstrate that there are sufficient cues to distinguish low and high-virality posts on some subreddits.
- The multimodal model performs the best, and images give stronger signals than textual data.
- Small vision models could capture virality signals well, while complicated ones do better.
- Using pretrained word embeddings does not lead to the better performance. The reason might be that the vocabulary on Reddit is different from the general text (Google News for Word2Vec; Wikipedia for Glove).

Future Work

- Collect datasets with high resolution images: Thumbnails are small, but people may click into the post and view the original image.
- Model temporal subreddits: our investigations show that it is very hard to predict virality in time-sensitive and news-related subreddits, e.g., `politics` and `The_Donalds`.
- Formulate the problem as a regression task: *coming in the report*.

Acknowledgements

We thank Pedro Garzon, Ahmadreza Momeni, and all CS230 staff for their suggestions and support through this project.

References

1. Huang, Gao, et al. "Densely connected convolutional networks." CVPR. Vol. 1. No. 2. 2017.
2. He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
3. Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." Advances in neural information processing systems. 2013.
4. Pennington, Jeffrey, Richard Socher, and Christopher Manning. "Glove: Global vectors for word representation." Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014.
5. Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980(2014).
6. Reddi, Sashank J., Satyen Kale, and Sanjiv Kumar. "On the convergence of adam and beyond." (2018).
7. Mazloom, Masoud, et al. "Multimodal popularity prediction of brand-related social media posts." *Proceedings of the 2016 ACM on Multimedia Conference*. ACM, 2016.
8. Dynamic Multimodal Regression." *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2018.