Generation of Memes to Engage Audience in Social Media

Andrew Kurochkin[1] and Kostiantyn Bokhan[2]

[1] Ukrainian Catholic University, Applied Sciences Faculty, Lviv, Ukraine
kurochkin@ucu.edu.ua
[2] Whirl Software, Kyiv, Ukraine
k.bokhan@whirl.sg

Abstract. In digital marketing, memes have become an attractive tool for engaging online audience. Memes have an impact on buyers and sellers online behavior and information spreading processes. Thus, the technology of generating memes is a significant tool for social media engagement. The primary purpose of the project is to develop a new approach and compare it to the existing baselines in the field of social media content generation, more precisely - meme generation. A meme is an image superimposed with text, which has humoristic or sarcastic sense; a meme is just another type of visual online content. This project is aimed at applying state of the art Deep Learning techniques as Transformer architecture to the meme generation problem. To achieve project objectives, we are going to collect dataset; create a model for generation of memes and titles based on the input text; create a model for defining optimal time to make a post; measure and analyze system performance in terms of social network audience engagement.

Keywords: Meme generation · Social network · Computational social science · Social media interaction · Memetics · Content generation · Reddit

1 Introduction

Social networks are mass media; they are information hubs. In 2018, digital consumers spent an average of 2 hours 22 minutes per day on social networks and messaging [1]. People get information about the news and events from across the world on social networks every day. Being present on social media is crucially important for an organization, which provides all kinds of services, products, and information. Organizations put in great effort to be properly presented in social networks and to run massive information campaigns.

One of the primary purposes of this activity is to engage their audience. Different kinds and forms of information spread in social networks. Information can be in the form of text, video, audio or image. Image superimposed with sarcastic or humoristic text is one of the most common form of the internet meme [2]. A simple form of the internet meme is called image macro shown in Fig. 1.

In: Proceedings of the 1st Masters Symposium on Advances in Data Mining, Machine Learning, and Computer Vision (MS-AMLV 2019), Lviv, Ukraine, November 15-16, 2019, pp. 10–20

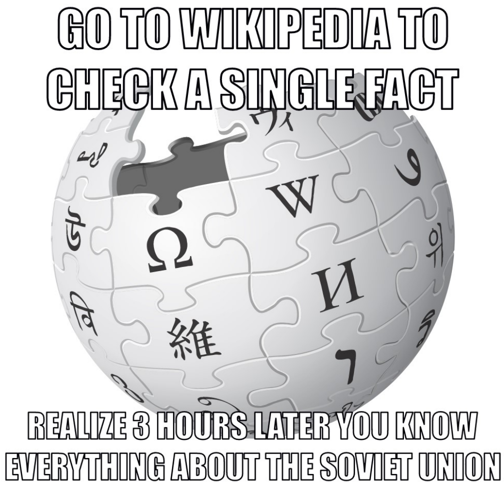

Fig. 1. An example of an image macro – a common form of an Internet meme [3]

People, who are involved in social media management (SMM), track trending topics on a regular basis. Keeping an eye on the trends is only part of work; another part is audience engagement by posting, including meme posts.

To create a meme on the relevant topic, an author of the meme has to come up with a caption, which will cause emotions in the audience, as well as select the image to supplement the meme. Once meme has been composed, an author creates a title or description (it depends on the specifics of a social network). On top of that, when the post is ready to be published the right choice of the posting time is essential. This whole process is time-consuming.

Thus, a solution, which is designed to automate the creation of posts with the image macro to engage the audience, is important. Even though the solution is limited to generation of only one type of content, it can be applied to generate other types of internet memes with some modification.

Engagement is a widely used metric of success for content in social media. Different actions can be used for the measurement of people engagement and its power [4]: views, likes, comments, shares, and reposts. In the scope of this research, we use a number of comments and score (upvotes - downvotes) to measure engagement. We use Reddit as the target social network in this research. Our motivation for this choice is introduced in Section 4.

Generation of memes, which engage the audience in the social media, using an image superimposed with English sentences, can be treated as machine learning in computational creativity. This set of problems is not investigated as well as classification or regression tasks. The goal of computational creativity is to model, simulate, or enhance creativity using computational methods [5]. Our scientific interest is to investigate how modern Deep Learning approaches for natural language processing will cope with our task. In particularly we are going to apply a technique for natural language modeling -

Transformer architecture [6] in solving a creativity problem, which traditionally is a prerogative of a human.

We are going to create a neural network, which generates memes based on the users' comments. We use historical data for model training. For modeling of new memes, we use comments from posts related to news or events, which had not been seen by neural network (NN) before.

We were faced with a lack of justification to choose evaluation metric to measurement the humoristic text generated by NN as memes caption. There is no clear answer which metric to use. Own metrics [7], BLEU [8] or perplexity [9] are used as loss functions. However, some studies illuminate why one of the most common metric in the sequence-to-sequence (seq2seq) tasks – BLEU [10] is a wrong solution in many cases [11–14]. Due to this fact, one of the contribution of this work is the experiments to justify which metric to use for evaluation of humoristic text.

The main contributions of the proposed work should be:

1. Create a unique memes dataset based on the Reddit submissions data collected by pushshift.io [15]
2. Investigate and justify the choice of the performance metric for the humoristic or sarcastic text generation
3. Develop a pipeline for memes generation based on the input text using Transformers architecture

2 Related work

Our paper relates mainly to three research topics: story generation and image captioning, meme generation, engagement, and virality in social networks. They are briefly reviewed in this section.

2.1 Story Generation and Image Captioning

The problem of generating memes caption from input text (comments) can be approached as a task to produce a short story based on the tags, which set the storyline. In [16], the authors approached the problem of hierarchical story generation where the model first generates a premise and then transforms it into a passage of text [16]. Researchers used sequence-to-sequence (seq2seq) models [17] with the usage of a fusion mechanism [18], as it had been shown that fusion mechanisms could help seq2seq models build dependencies between their input and output [16]. In the scope of this work an open-source sequence modeling toolkit was used FAIRSEQ [19].

The problem of generating natural language descriptions from the image has been studied in [20]. The approach to encode images with Convolutional Neural Network (CNN) into vector embeddings was proposed. The decoder uses the embeddings to generate sentences based on the Long-Short Term Memory (LSTM) network. The LSTM was chosen due to its ability to deal with vanishing and exploding gradients, which are a common problem in Recurrent Neural Networks (RNN) [21].

In the other work [22], the authors concentrated on generating captions for images and videos with different styles. In this work, the authors utilized the FlickrStyle10K dataset and aimed it at the generation of humoristic or romantic image caption. Here the model architecture is also based on encoder-decoder design. The solution architecture was based on the encoder-decoder design with the modifications. The most valuable of which is the factored LSTM, it automatically distills the style factors in the monolingual text corpus [22]. A meme image can be the image with a penguin in the center, but the main message or subject of the joke can be related to the awkward social situation [23]. Since the scene presented in the image can have different meaning than the whole meme with its cultural background, an image caption does not solve our problem, as the image is not the right source of information for memes caption.

2.2 Meme Generation

The language of Internet memes was modeled in [8], where an approach which is common in the economic modeling – copula methods [24, 25] – was applied. The authors claim that the predictive power of copula models could be used for joint modeling of raw images, text descriptions, and popular votes [8]. They employed reverse image search to get text information about the input image.

In [7], the results from [20] were adopted, however, with ResNet-152 replaced CNN as a feature extraction method. In this work, authors proposed Funny Score that was used as a loss function. Funny Score metric is based on the stars from the BoketeDB, which display the degree of funniness of a caption evaluated by users of the Bokete [26].

The authors of [9] based their solution on the approach of [20]. In order to create image encoding, the system utilized a pre-trained Inception-v3 network. An important contribution of the work was a new beam search implementation in order to encourage diversity in the captions [9]. For the evaluation, perplexity and human assessment were used. Images or a combination of image and its name served as input data. The same image template can have various memes text related to it. Due to this fact, we claim that memes names have insufficient descriptive power. The authors mention that the separators between the text at the top and bottom can improve training results. Therefore, we take into account this observation in our work.

2.3 Engagement and Virality in Social Network

In [4], a 4-level system of engagement classification based on human actions was proposed: from Level 1 - views, less public and more private expressions of engagement, Level 2 is like action, Level 3 - comment or share, to Level 4 external posting, the most public level of engagement. The model for predicting Level 4 engagement was provided.

The study of memes propagation, evolution, and influence across the Web was done in the [27]. The authors used a processing pipeline based on perceptual hashing, clustering techniques, and a dataset of 160M images from 2.6B posts [27]. The researchers performed collection of the memes description based on the site Know Your Meme

[28], which gives information about the memes concepts. This information was used for the cluster analysis of memes and the creation of their embeddings.

In [29], the authors analyze how post popularity depends on the way the content is presented (the title), the community it is posted to, whether it has been seen before, and the time it was posted. The unique contribution of this work is the dataset, which contains 132K submissions, only 16.7K of which were unique, whereas the others were resubmissions. These specifics make it possible to determine the influence of the title, community, and posting time, regarding a submission. In [29], community and language models, which help target social media audience, were developed. In this paper, research focus was on viral content, in the form of republished submissions.

In [30], the phenomenon of image virality was investigated from a computer vision perspective. Virality score based on the image resubmission was proposed. The neural network for image virality prediction was created. The results show that in the task of image virality prediction, based on the high-level image description (capturing semantic information), a machine performs better than a human. The model shows 68.10% accuracy relative to 60.12% of human performance.

3 Research Objectives

The main objective of this work is to evaluate how the State-of-the-Art Deep Learning approaches perform the task of engageable content generation. On top of that, we claim the following objectives:

1. Even though a few approaches of humoristic text generation were proposed previously, such as [31–34], we aim to evaluate a data-driven approach for this task
2. To check whether the neural network trained on the memes which caused engagement (comments or votes), will be able to produce memes which trigger people engagement
3. To find out which metric should be used for memes caption evaluation

4 Approach

In this section, we define the approach to achieve objectives: collection and preparation of the dataset, model training, plan to choose optimal loss function and overall result evaluation.

4.1 Dataset

To achieve our objectives, we need a large dataset, which is unavailable. The dataset must contain unique combinations of meme templates, separated image captions with top section (also called a set-up) and bottom section (known as punch line), score and comments. We chose Reddit since, according to the official blog [35], it has 330 million of users, 850 000 communities, who generate 58 million votes and 2.8 million comments daily. It is a common practice in computational social science and social network analysis to use this platform.

Our dataset is based on the data collected by Jason Baumgartner [15] including all Reddit posts and comments since 2005. We use 3.5 years of the information, particularly the timeframe from January 2016 to August 2019. However, our pipeline can be used to extract information from the whole Reddit dataset since 2005.

Source data is split into the batches by months. To use computer storage efficiently, we are going to process data in batches by erasing all the information irrelevant to our further research - posts and metadata, which are not related to image macro. Data collection pipeline includes:

1. Filtering out memes. We extract posts with predefined characteristics from the whole batch to apply meme information retrieval techniques based on them.
2. Filtering out comments for the posts from the previous step.
3. Downloading images from the posts.
4. Optical character recognition (OCR) to extract top and bottom pieces of text from the meme. We intend to use Tesseract [30], which is one of the most common open-source tools for OCR.
5. Template recognition. We detect which template meme is based on. Each meme will be presented as an image template id and text extracted in the previous step.
6. Removing the downloaded memes images.

We plan to publish the code of the pipeline and final dataset at a public repository so these will be parts of our contribution.

4.2 Experiment Pipeline

The key problem in our work is language modeling. Language modeling is usually framed as unsupervised distribution estimation from a set of examples $(x_1, x_2, \dots, x_n)$ each composed of variable length sequences of symbols $(s_1, s_2, \dots, s_n)$ [37]. In [37], GPT-2 – text generation model based on the Transformer architecture was presented. This NN shows the State-of- the-Art results on a few datasets without any fine-tuning; it was trained on a huge variety of Internet texts, including Reddit. Due to this fact, we use GPT-2 as NN for our approach.

We are going to use GPT-2 355M model as it is more complex than the small 124M, hence it catches text nature better and is still allows fine-tuning at the machines with GPU. We are going to do additional training where our dataset of short texts having specific nature will be used as input.

To create post titles, we use the same approach as for meme caption, but with a model pre-trained for title generation. Finally, the image (meme template) that reflects the idea of the generated text should be chosen. We embed each image in the memes space based on the encoding of the memes description. We are going to train a neural network to match the meme caption with the right meme template based on the patterns that are present in our dataset.

To define optimal time to make a post we intend to use historical information, once we achieve that we will use the information to schedule posting.

4.3 Evaluation

Evaluation includes two phases, first phase - is loss function which estimates the result of the generated text, second phase is human engagement measurement.

Even though the absolute value of the loss function could not be treated as a clear metric of how good the result is, it gives the model a tool to estimate the quality of the results. It is important to choose the right sense of humor for the neural network so it can better distinguish good memes from bad ones. We are going to train a few models based on different loss functions and generate a batch of 50 images from each of them. Estimating the quality of humor is impossible, as it is a very subjective matter. In this work, we are bounded to our target audience, so we are going to ask the English speaking audience which memes they prefer more. The loss function from the model, which outperformed the other models, will be used for the final model training.

The second part is the engagement measurement, which is the key metric in our problem. Proposed evaluation pipeline is depicted in Fig. 2. The essence of the evaluation process is to measure user interaction with the content. Metrics of engagement will be the number of comments and post overall score, which are statistics to measure the power of people engagement.

5 Research Plan

5.1 Dataset Collection

It takes weeks to process our input data to build the dataset. To minimize risks of failure, we have found a limited dataset, which can be used in our project to check the claimed hypothesis and answer the stated research questions. We intend to finish data preprocessing by the end of October.

5.2 Pipeline

We plan to use the Transformer architecture, GPT-2 model with minor modifications to adopt it to specifics of our problem. This task should be done while the data collection pipeline is working on the clusters.

During the first iteration, the models using different loss functions should be trained. At the time when the outcome of the model will be acceptable from our point of view, they will be evaluated with using independent experts (crowdsourcing service, Amazon Mturk [38]). The loss function, which yields the best results based on people opinions, will be used for future iterations.

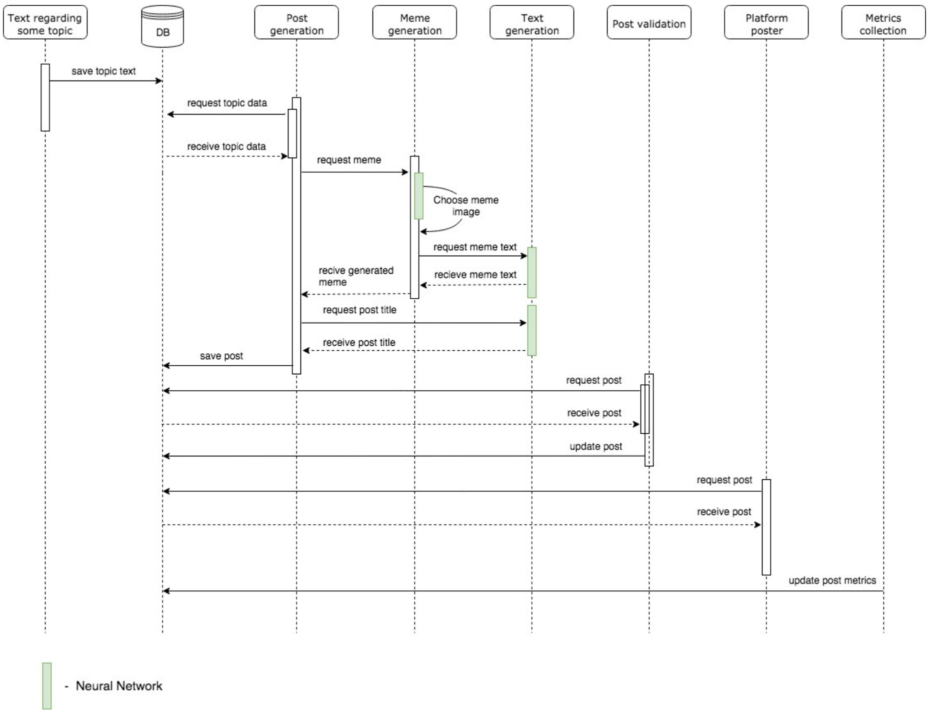

Fig.2. The pipeline for content generation and engagement measurement

5.3 Result Evaluation

During the model training process, we are going to build pipelines for posting on social networks and for feedback collection. We plan to finish this by the middle of November.

5.4 Pipeline Revision and Refinement

The pipelines used in this project are strongly dependent on a bunch of the third party applications. This is an additional risk factor, so we have planned some time to resolve potential issues. When the right metric has been found, the model training pipeline should be refined.

5.5 Supporting Activity

We use agile with weekly sprints, so we can fix all the problems and tasks in the week log. We store all ideas, hypothesis, and insights in the experiments documentation. All these materials will be used as a basis for the thesis manuscript.

6 Conclusion

The considered problem is relatively new, and it involves different disciplines and scientific areas. There have been studies on the engagement analyses, social media influence, modeling of information spreading, even memes generation already have been done, however, the combination of factors which we set as the project objectives makes this work unique.

The result of this project will be evaluation of the current progress of Deep Learning in natural language modelling. It will show how it performs for content generation task. In the future, it can be used as a base for generating more complex scenes. In the scope of the project, we aim to find a metric, which properly captures the specific nature of the memes captions based on human opinions. This knowledge is a part of our contribution as well as unique memes dataset.

We described motivation and importance of approaching meme generation problem, made an overview of works, which are related to the current study from different sides, defined clear and achievable project objectives, proposed an approach to achieve stated goals and briefly described plan of the project work.

References

1. Yavich, R., Davidovitch, N., Frenkel, Z.: Social Media and loneliness – forever connected? Higher Education Studies **9**(2), 10–21 (2019)
2. Knobel, M., Lankshear, C.: Online memes, affinities, and cultural production. A New Literacies Sampler **29**, 199–227 (2007)
3. Internet meme - Wikipedia. https://en.wikipedia.org/wiki/Internet_meme#/media/File:Wikipedia_meme_vector_version.svg
4. Aldous, K.K., An, J., Jansen, B.J.: View, like, comment, post: analyzing user engagement by topic at 4 levels across 5 Social Media platforms for 53 news organizations. In: 2019 International AAAI Conference on Web and Social Media, pp. 47–57. AAAI (2019)
5. Toivonen, H., Gross, O.: Data mining and machine learning in computational creativity. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **5**(6), 265–275 (2015)
6. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: 31st Conference on Neural Information Processing Systems, pp. 5998-6008 (2017)
7. Yoshida, K., Minoguchi, M., Wani, K., Nakamura, A., Kataoka, H.: Neural joking machine: humorous image captioning. arXiv preprint, arXiv:1805.11850 (2018)
8. Wang, W.Y., Wen, M.: I can has cheezburger? A nonparanormal approach to combining textual and visual information for predicting and generating popular meme descriptions. In: 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 355–365 (2015)
9. Vinyals, O., Toshev, A., Bengio, S., Erhan, D.: Show and tell: a neural image caption generator. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3156–3164. IEEE Press, New York (2015)
10. Bokete. https://bokete.jp/

11. Peirson, V., Abel, L., Tolunay, E.M.: Dank learning: generating memes using deep neural networks. arXiv preprint, arXiv:1806.04510 (2018)
12. Zannettou, S., Caulfield, T., Blackburn, J., De Cristofaro, E., Sirivianos, M., Stringhini, G., Suarez-Tangil, G.: On the origins of memes by means of fringe Web communities. In: 2018 Internet Measurement Conference, pp. 188–202. ACM (2018)
13. Internet Meme Database. https://knowyourmeme.com/
14. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: BLEU: a method for automatic evaluation of machine translation. In: 40th Anual Meeting of Association for Computational Linguistics, pp. 311–318. Association for Computational Linguistics (2002)
15. Ananthakrishnan, R., Bhattacharyya, P., Sasikumar, M., Shah, R.M.: Some issues in automatic evaluation of English-Hindi MT: more blues for BLEU. In: 5th International Conference on Natural Language Processing (2008)
16. Novikova, J., Dusek, O., Curry, A.C., Rieser, V.: Why we need new evaluation metrics for NLG. arXiv preprint, arXiv:1707.06875 (2017)
17. Sulem, E., Abend, O., Rappoport, A.: BLEU is not suitable for the evaluation of text simplification. arXiv preprint, arXiv:1810.05995 (2018)
18. Reiter, E.: A structured review of the validity of BLEU. Computational Linguistics **44**(3), 393–401 (2018)
19. Reddit Statistics – pushshift.io. https://pushshift.io/
20. Fan, A., Lewis, M., Dauphin, Y.: Hierarchical neural story generation. arXiv preprint, arXiv:1805.04833 (2018)
21. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: 27th International Conference on Neural Information Processing Systems. Volume 2, pp, 3104–3112 (2014)
22. Sriram, A., Jun, H., Satheesh, S., Coates, A.: Cold fusion: training seq2seq models together with language models. arXiv preprint, arXiv:1708.06426 (2017)
23. Ott, M., Edunov, S., Baevski, A., Fan, A., Gross, S., Ng, N., Grangier, D., Auli, M.: fairseq: a fast, extensible toolkit for sequence modeling. arXiv preprint arXiv:1904.01038 (2019)
24. Bengio, Y., Simard, P., Frasconi, P.: Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks **5**(2), 157–166 (1994)
25. Gan, C., Gan, Z., He, X., Gao, J., Deng, L.: StyleNet: generating attractive visual captions with styles. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3137–3146. IEEE Press, New York (2017)
26. Socially awkward penguin – know your meme. https://knowyourmeme.com/memes/socially-awkward-penguin
27. Schweizer, B., Sklar, A.: Probabilistic Metric Spaces. Dover Publications (1983)
28. Nelsen, R.B.: An introduction to copulas. Technometrics 42(3), 317 (2000). doi: 10.2307/1271100
29. Lakkaraju, H., McAuley, J., Leskovec, J.: What's in a name? Understanding the interplay between titles, content, and communities in Social Media. In: 7th International AAAI Conference on Weblogs and Social Media, pp. 311-320. AAAI (2013)
30. Deza, A., Parikh, D.: Understanding image virality. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1818–1826. IEEE Press, New York (2015)
31. Lin, C.C., Hsu, J.Y.J.: Crowdsourced explanations for humorous internet memes. In: 28th AAAI Conference on Artificial Intelligence, 3118–3119. AAAI (2014)
32. He, H., Peng, N., Liang, P.: Pun generation with surprise. arXiv preprint, arXiv:1904.06828 (2019)

33. Kiddon, C., Brun, Y.: That's what she said: double entendre identification. In: 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2, pp. 89–94. Association for Computational Linguistics (2011)
34. Raskin, V.: Semantic mechanisms of humor. In: 1979 Annual Meeting of the Berkeley Linguistics Society. Vol. 5, pp. 325–335 (1979)
35. Upvoted. The official Reddit blog. https://ref34.com/
36. Tesseract open source OCR engine (main repository). https://github.com/tesseract-ocr/tesseract
37. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I.: Language models are unsupervised multitask learners. OpenAI Blog **1**(8) (2019)
38. Amazon Mechanical Turk. https://www.mturk.com/