

Challenges in Automatic Meme Generation Using Deep Neural Networks

Introduction

Today’s Internet is full of **memes**: small chunks of data which represent a large proportion of a social network’s communication “protocol”. As coined by Dawkins in his 1976 book “*The Selfish Gene*” [1], the meme is the fundamental building-block of information whose behavior seems very likely to that of a gene (*better* memes tend to prevail longer in the Internet). Also, there is an evolution mechanism that allows older memes to adapt to current trends.

In this project, the problem of learning representations from a meme dataset and immediately generating new data is studied, as a basis to motivate the further study of memes using *state-of-the-art* machine learning methods. In particular, we develop a deep neural architecture which attempts to learn a language model out of a dataset of plain meme images along with several associated captions.

Much of deep learning’s success is justified by the outstanding performance of Vinyals’, *et al*, **Show-and-Tell model** [4]. We report experiments directly using such model, in comparison with a proposed model that uses a *more shallow* convolutional network.



Figure 1: Meme examples. Most of them were popular between 2009 and 2013, yet some of them have evolved! (Image taken from <https://memegenerator.net>.)

As pointed out by Figure 1, we choose to start with memes that mainly consist on an image centered on a main character and a caption. We, finally, report some evaluation measurements to analyze our architecture’s performance.

This project was mainly presented as an undergraduate thesis [2], under the supervision of Dr Ivan Vladimir Meza Ruiz. A similar work was described later in [3].

Formally, we trained our architecture with data samples consisting of (I, S) tuples, where I is an image and S is the set of its associated captions. We used 70% of our dataset for training and the remaining for testing. We use the original *Show-and-Tell* model, at first, to see how it would perform with a much smaller dataset than *ImageNet*. Then we, conjectured that a shallow convolutional network could do better; thus, we trained the model shown in Figure 2. The CNN extracted output was inserted as *initial state* to a LSTM recurrent network, whose training resembled what is defined on [4].

Lessons

The proposed neural architecture provides a baseline to work with many tasks related with learning representations from raw Internet memes. We achieved, at a high level, a significant performance when evaluating the generated captions using a perplexity measurement against a testing meme dataset¹. Furthermore, it is expected that this step, with some refinements, may be useful to train language-specific datasets.

In terms of image classification, our good results in pre-training imply the possibility of adding broader convolutional networks to the proposed pipeline. Thus, one can think of refining our classification (i.e., getting better feature vectors), prior getting into the recurrent network.

Challenges

- Any pipeline must get used to new formats, whose origin doesn't necessarily come from websites dedicated to construct a meme, but from any social network.
- Bring together a more thorough and diverse dataset that includes novel meme designs. This is especially useful (and necessary) to what was specified on the previous section. A hierarchical classification of memes, such as what ImageNet does, might be extremely useful, as well.
- Accurate evaluation metrics are still missing, for our current methods only consider language model-based similarities, through the use of perplexities. In this case, a thorough methodology must be built, including scoring by humans, considering that humour is a very ambiguous concept.
- It would rather be interesting to experiment with formats that include comic-like structures or a certain sequence. In such cases, a recurrent architecture might be helpful, to be combined with CNN's.
- Humour varies, roughly, between countries. For instance, in Mexico, the reasons a meme reaches a level of popularity depends of very different social media actors than those who influence the American "meme market". Thus, it is critical to include an attention mechanism and link it with a context representation out of actual social media trends.

Bibliography

- [1] R. Dawkins. *The selfish gene*. Oxford University Press, New York, NY, USA, 3ra ed., 2006.
- [2] A. M. Orozco Camacho and I. V. Meza Ruiz. *Generación automática de memes de internet a través de una red neuronal profunda*. 2018. URL: <http://pbidi.unam.mx:8080/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=cat02029a&AN=tes.TES01000770173&lang=es&site=eds-live>.
- [3] A. L Peirson V and E Meltem Tolunay. Dank Learning: Generating Memes Using Deep Neural Networks. *Arxiv e-prints*, June 2018. arXiv: 1806.04510 [cs.CL].
- [4] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan. Show and tell: A neural image caption generator. *Corr*, abs/1411.4555, 2014. URL: <http://arxiv.org/abs/1411.4555>.

¹ Such testing dataset was built upon cross validation splitting the crawled data, and only considering captions but not images.

Appendix A: Neural Architecture

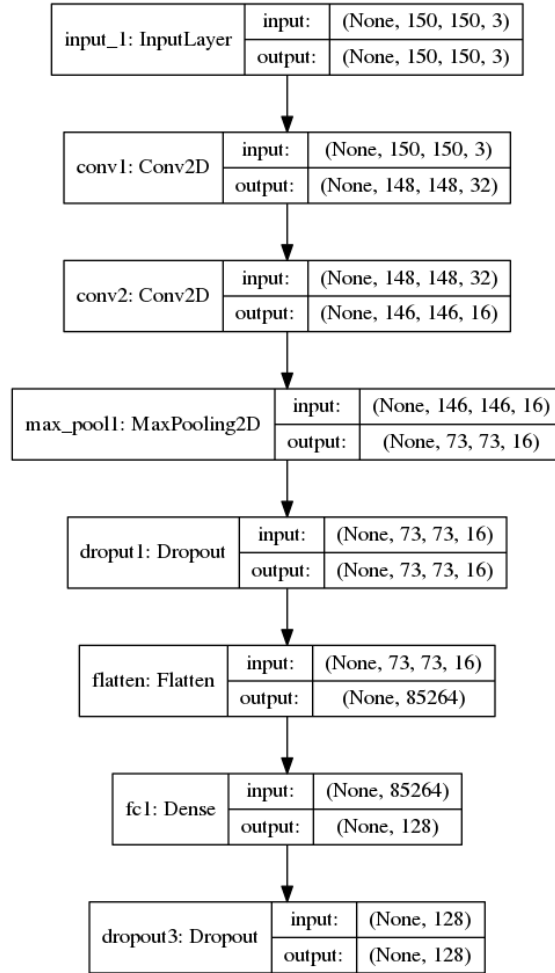


Figure 2: A shallow architecture used to extract convolutional feature maps from a crawled dataset.