Project DeepToons: Synopsis

Vasanth Sarathy, Daniel Kasenberg

March 2016

Abstract

This project is an effort to automate the creation of single-panel New Yorker-style cartoons. In this short synopsis, we will outline the basic motivation, problem, hypothesis and approach.

1 Motivation

Many consider creativity to be central, and possibly even unique to human intelligence. Artistic creativity (as opposed to creative reasoning or creative problem solving) is particularly special as it distinguishes us humans from animals, who are known to show a limited degree of creative thinking and tool making, but almost no artistic ability [1]. Therefore, not surprisingly, many who are interested in understanding human cognition are deeply interested in understanding of the process of creativity and humor. Even more special is our human ability to generate and appreciate humor, a specific form of artistic creativity. In this project we seek to understand creative cognitive processes, by focusing on a specific flavor of humor, namely cartoons. To do so, we aim to construct an AI agent capable of generating "humorous" single-panel New Yorker-style cartoons (i.e., single image and short caption).

A single-panel cartoon (also known as a "gag") is a completed artwork comprising a **sketch** or drawing of a **scene** accompanied by a **caption**. In combination, the sketch and the caption together form the **joke** that provides the basis for the underlying humor. The New Yorker is the premier forum for cartoonists to publish their work and has been in circulation since 1925, with these cartoons sprinkled throughout each issue. Besides the New Yorker, single-panel cartoons can be found in Newspapers, other magazines, online blogs and many other places. Generally, these cartoons can not only serve to entertain, but sometimes can be a great tool to educate serving as a medium to deliver subject matter in a humorous and memorable way. Thus, an AI cartoonist can help people communicate their own ideas in new and interesting ways, even if they cannot sketch or generate jokes effectively.

---

[1]Some consider the elaborate structures created by Termites to be art.

2 Hypothesis/Problem statement

There are many challenges in generating such cartoons: generating sketches, recognizing the content of sketches, generating captions based on the sketches, weaving the sketch and caption together to represent a coherent idea, and choosing an idea that is inherently funny. An AI cartoonist would need to do all of these in an automated way.

Although there have been many efforts in automated joke generation (Lit Review Needed), there are currently no automated gag-cartoon generators. There is a tight coupling between a sketch and a caption in a cartoon and therefore, purely text-based joke generators are not quite enough.

Certain machine learning techniques have also been applied (Shahaf2015) to explore what makes a cartoon funny by analyzing captions. The idea there was to use the large corpus of New Yorker cartoons to learn common features and then use the trained network to predict if a caption in a set of captions will be funny.

There has been a lot of recent work in the area of "Deep Learning," the general term used to describe techniques in modeling artificial systems as multi-layered convolutional neural networks. Deep Learning has shown a lot of promise in image captioning and even automated game-playing. However, factual statements about the objects present in a scene are not enough to generate a cartoon.

We believe what is needed to construct an AI cartoonist are (1) a cognitive theory of idea generation and (2) a computational architecture for cartoon generation that can learn and improve over time. We hypothesize that a Deep Neural Net framework coupled with Q-Learning

2.1 Training

We plan to use existing Deep Learning techniques in training our AI cartoonist. The basic idea is to train the network to recognize **objects** present in photographs and images from existing databases (e.g., ImageNet) using Deep learning techniques (Fei-Fei Li). This is an $f : IMAGE \rightarrow NL$ process. A similar Deep Learning training approach will be used to learn dialogues present in movie scenes using scene descriptions in dialogue databases and book databases. This, however, is an $f : NL \rightarrow NL$ process. We will aim to use a concept hierarchy similar to WordNet or ImageNet. In addition, we will also use Sketch Datasets to learn objects present in hand-drawn sketches in images using existing approaches (Forbus' CogSketch, and Eitz' Sketch).

Our concept hierarchy allows us to define a network or graph representing the relationships between concepts as it relates to their **level of abstractness**.

We hypothesize that using existing, proven machine learning techniques, will be able to train our AI Cartoonist to recognize concepts present in images, sketches and written dialogue. We do, however, expect there to be some degree of conflict between these different types of data and will aim to address these challenges as they arise.

2.2 Cartoon Generation - Approach 1

Our first most simple hypothesis to generating "funny" cartoon is that a funny idea comes into existence when two unrelated somewhat connected ideas are combined.

Using this simple hypothesis, our first approach would be take as our 1st idea any image or sketch at random and recognize the objects and concepts present in this first image using our trained network. Then, we propose using a "scrambler" to modify the set of recognized objects and concepts with other concepts or objects from our concept hierarchy. Our scrambler can be trained (via reinforcement learning) to combine concepts in a "funny" way (reinforced by human input or some humor reward function).

Once scrambled, the objects and concepts can then be run through our trained dialogue network to generate dialogue bits relevant to the scrambled set of objects.

We also plan to use various cartoon databases to help train our AI cartoonist, and more particularly, our scrambler as to what does it mean to be funny. Specifically, we can recognize objects and concepts in images, use this to compute dialogue and then determine the difference between the synsets (cognitive synonyms) in these bits of dialogue. The idea would be to train our scrambler generate the missing (possibly scrambled) humor connection between the images and captions.

2.3 Cartoon Generation - Approach 2

In this approach, we look to start with a question "what if X were Y?" and explore funny ideas that emerge from this question. Each concept X and Y evokes certain images and dialogues, and our approach to creative cartoon generation would be to combine these.

3 Method

We will conduct a literature review to analyze existing methods in deep learning for vision and NL, computational humor, and cognitive theories for humor. We will need to collect datasets, train our networks as prescribed above and iterate and refine our approach with data. Once we begin generating cartoons, we intend to start a blog/twitter account and allow our readership to determine if cartoons are actually funny.

Our plan will primarily focus on Approach 1, and is as follows:

1.

A short summary of how you are going to confirm/falsify the hypothesis: what prototypes do you expect to build, how will you evaluate and measure them, what techniques and tools are you going to use, which people will you interview, how will you document processes and products, how will you record you progress, how will you analyze your work, how will each phase contribute to validating the hypothesis?

4   (Expected) Analyses and Results

This the main body of your report where you outline what you have done, how it contributes to analysing the problem, the results of your work, argue why your results are correct, relate them to theory, and document what has been achieved.

Remember that designs are often best described and supported by diagrams and central algorithms are best expressed in small fragments of real or pseudo code. Text like "the server calls the 'update' method which next calls the 'IAmBored' method in the ..." are terrible and next to impossible to read.

Avoid narrative writing styles like "then we did X but it did not work, so we tried Y and it worked better". Rather, use a (problem, analysis, solution) format: "Problem: (describe short and precisely), Analysis: (describe a set of problem solutions), Solution: (describe and argue why a given solution was chosen)."

Remember that the primary objective with your report is to demonstrate that you master the theory introduced in the course and your analytical skills to "think clever thoughts" and related and discuss your work. It is not to program a polished product ready for shipment.

5   Related work

[This section may be put in front of the hypothesis section or integrated into the method section, if it make the flow of text more natural.]

In this section you outline what literature and other work your project build upon: papers, books, links to webpages, tutorials, etc. All references should be resolved in the reference section, that is do not use footnotes, or put the reference directly in the text. For an example, look how references are cited in Bardram et al.[**?**].

It is good to address how your work extends, use, or build upon the cited work.

6   Conclusion

Your synopsis should clearly state: abstract, motivation, hypothesis, method, and (expected) analyses and results.

References