NYUx - Theories of Media and Technology
Philip Hughson - philipfhughson@gmail.com
January 2021

Reseach 3 - Post-human possibility mappings

In my curation, I looked at the idea that the cyborg could be understood as an artificial extension of human capacities. Since this is a very broad conception - presumably covering stone age technology as well as smartphones, and all that lies between, I decided to narrow my focus to one - still broad - category, which is the human capacity to imagine and consider alternative possibilities in a given situation.

(This capacity has been pithily summarised as our ability to have our ideas die in our stead (karl popper https://www.informationphilosopher.com/solutions/philosophers/popper/natural_selection_and_the_emergence_of_mind.html): since we have imaginations, we can run alternative scenarios in our heads, and choose the one that works best, or at least doesn't involve dying).

Of the examples I looked at, I've chosen three: the designer's sketchbook, the neverending bass solo, and generative poetry (based on predictive text) to work from. These cover a scope which I find interesting: roughly stated, that we can use cyborg - or post-human - ideas to develop the creative process, somewhat as a designer might in a sketchbook, perhaps by using recurrent neural networks (or otherwise) to interate and choose between possibilities.

The ideas at work here can be broken down:

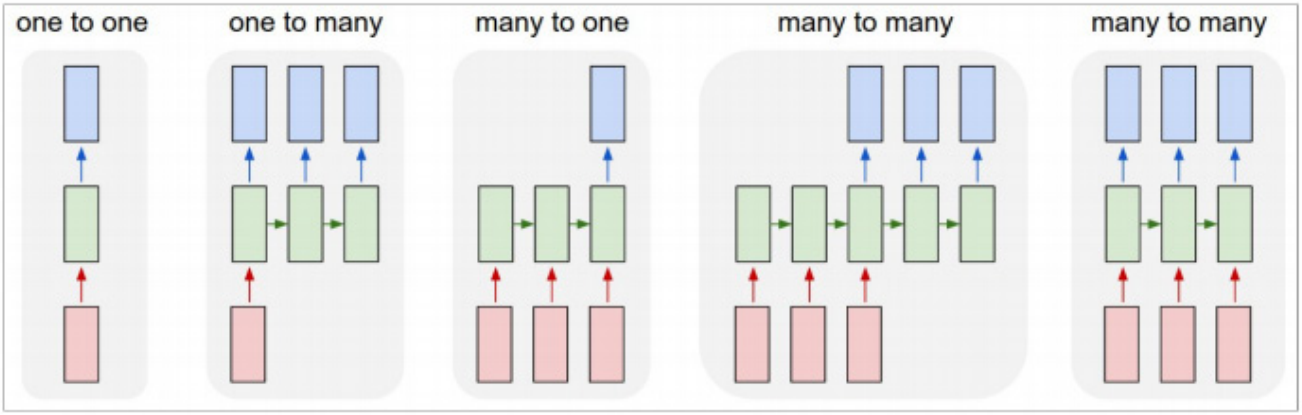
(a) Creativity involves a tension between present reality and an undefined and plural future state, and the creative process involves making a step from the former to the latter.
(b) there are often multiple possible steps that can be taken
(c) the outcomes of each step can be evaluated, and the step developed further or discarded
(d) This evaluation can be done in reality, or in the imagination, or in a simulation (on paper, on a computer), and in can be done individually or socially
(e) The process is iterated
(f) These iterations could produce a branching tree like structure.
(g) In this process, when artificial means are used, there can be alternation between human and artificial creation and / or discrimination .
(h) also, limitation can be a creative step (e.g. choosing between three words, or playing music only in one key).
(i) In more general terms, we could think of each step of such a process being done by human or artificial agents, with an interplay between different combinations of them.

So, a rough direction for my project might be to try to use Recurrent Neural Networks for the development of an idea, similar perhaps to some applications Google's Magenta (https://magenta.tensorflow.org/). But I might aim to chart the development of an idea over time or iterations, perhaps using a tree like structure. One idea might be to work with sound, similar to what I did for the first project (https://phhu.org/beat-conv), but using (stealing?) somethink more sophisticated for the variation step, and / or introducing an element of choice, similar to the three words made available by predictive text on mobile phones (as used for predictive poetry, for example). Alternatively, it might be easier to work with words, along the lines of predictive poetry, but perhaps exploring the tree of choices behind sentences a little more visually.

In abstract terms, this could represent a more cyborgian notion of the notebook disucssed by Andy Clark in his 1998 article "the extended mind": e.g. can we make a neural net notebook which offers up alternative next steps for ideas (such as musical phrases, or poems.... anything that is relatively easy to digitise and make discrete.)

The sources I've included here are ones that I've found useful while researching around this topic. A recurring theme has been the use of Recurrent Neural Networks, so I've started with an introduction to there.

The unreasonable effectiveness of recurrent neural networks



Each rectangle is a vector and arrows represent functions (e.g. matrix multiply). Input vectors are in red, output vectors are in blue and green vectors hold the RNN's state (more on this soon). From left to right: **(1)** Vanilla mode of processing without RNN, from fixed-sized input to fixed-sized output (e.g. image classification). **(2)** Sequence output (e.g. image captioning takes an image and outputs a sentence of words). **(3)** Sequence input (e.g. sentiment analysis where a given sentence is classified as expressing positive or negative sentiment). **(4)** Sequence input and sequence output (e.g. Machine Translation: an RNN reads a sentence in English and then outputs a sentence in French). **(5)** Synced sequence input and output (e.g. video classification where we wish to label each frame of the video). Notice that in every case are no pre-specified constraints on the lengths sequences because the recurrent transformation (green) is fixed and can be applied as many times as we like.

Andrej Karpathy's 2015 blog entry on "The Unreasonable Effectiveness of Recurrent Neural Networks" provides a good overview of the functionality of Recurrent Neural Networks (RNNs), which are central to the operation of two of the example I looked at: the endless bass solo, and predictive word / sentence completion. His essential poit is that that RNNs are more effective because their API (programming interface) is more flexible that that of more conventional neural networks, which typically use fixed length vectors for input and output. RNNs on the other hand allow a sequence of inputs to be applied (perhaps a sequence of characters, in the case of a sentence, or a sequence of images, for a video), and probability distributions of categories or next items to be output. The screenshot above illustrates some of these possible formats.

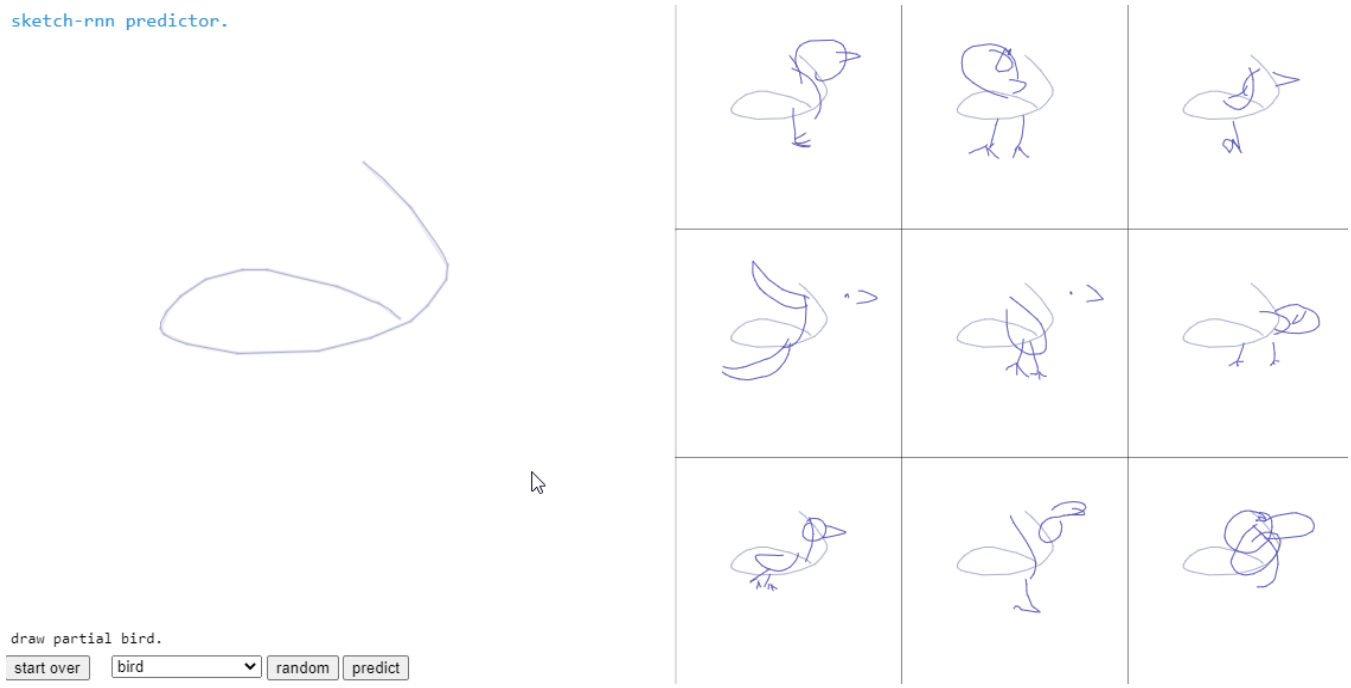
The article goes on to give a simple example of this API and an simple example of a network being trained to spell the world "hello". A distinction is also made between Long Short Term Memory (LSTM) networks and RNNs, though the two are similar, and RNN is often used to refer to both. There are then various examples of RNNs being used to generate or complete text samples based on training on a specific text corpus (say the text of Wikipedia, or the complete works of Shakespeare, or computer code). With adequate training time, these produce excellent results, even if long term sense is somewhat missing. After this there is some illustration of how RNNs work, including illustration of the improvement in quality with training time, and the development of specialised "neurons" for certain features (such as URLs or quotation marks). The article finishes with some links to code, and comments about further directions.

From this article, I would conclude that for my project it is worth considering using RNNs. It would be necessary to have a good body of training data, but given this, it should be possible to find pre-made code relatively easily.

Karpathy, A. (2015), 'The Unreasonable Effectiveness of Recurrent Neural Networks' . Available online at https://web.stanford.edu/class/cs379c/archive/2018/class_messages_listing/content/Artificial_Neural_Network_Technology_Tutorials/KarparthyUNREASONABLY-EFFECTIVE-RNN-15.pdf and http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Magenta RNNs: Sketch-RNN and others

Magenta is "An open source research project exploring the role of machine learning as a tool in the creative process." It uses the tensorflow library, and has a javascript API, meaning it could be used with the p5.js environment. Two example applications caught my eye: Recurrent Neural Networks (RNNs) used to complete sketches; and RNNs adapted to create musical structures which have better long term structure than other attempts.



The sketch-rnn demo is shown above, where a single user-inputted doodle is interpreted nine times by an RNN trained on sketches of birds, producing nine different bird-like doodles based on the user input. The accompanying article describes a few use cases: for example, taking the same input sketch and having it competed by differently trained RNNs, such that a sketched square might become, for example, a truck, or a bird, or person; or using sketch-rnn to interpolate between two sketches, producing a series of doodles that form a similarity spectrum between the two.

Considerably more detail is given in the accompanying paper "A Neural Representation of Sketch Drawings", which provides detail of the algorithms used. It distinguishes between unconditional and conditional generation: in conditional generation, we supply an input vector (in the form of an input sketch); in unconditional generation, the model is supplied with a blank canvas. This is similar to the notional of conditional probability, with which I am familiar.

Elliot Waite's article on "Generating Long-Term Structure in Songs and Stories" works in the domain of music, and deals with the problem of generating structure similar to that found in most musical composition, with ideas being repeated. "Lookback RNNs" are used for this. Exampels are given of melodies generated from a few started notes are. The article also includes a very interesting example of musicians playing and composing alongside RNN generated melodies.

Waite, Elliot (2016) "Generating Long-Term Structure in Songs and Stories". Available at https://magenta.tensorflow.org/2016/07/15/lookback-rnn-attention-rnn
Ha, David; Jongejan, Jonas & Johnson, Ian (2017) "Draw Together With a Neural Network" Available at https://magenta.tensorflow.org/sketch-rnn-demo
Ha, David & Eck, Douglas (2017) "A Neural Representation of Sketch Drawings " Avaiable at https://arxiv.org/pdf/1704.03477.pdf

Harry Yeff, aka Reaps One: “Move” solo beatbox performance, and “conversation” with “Second Self” AI

Harry Yeff, also known as Reaps One, is very good at beatboxing.... to persuade yourself, listen to the “Move” example listed below. The “Second Self” interaction, also linked below, uses the SampleRNN system, as used in the “endless bass solo” example from my curation. To get an overview of his work, it is worth watching the TEDx talk he gives, which includes descriptions of these works.

The “Move” performance illustrates many of the unusual vocal techniques used by Yeff. Most interesting from the point of view of Cyborg theory, from my point of view, is that fact that many of the sounds he makes are very clearly inspired by electronic dance music. This is an example of digital arts feeding back into the realm of analogue human expression, such that the digital is no longer present, but rather has been subsumed into the human cortex (and vocal tract).

The “Second Self” performance shows Yeff in a beatbox exchange, or conversation, with an RNN made using SampleRNN. It illustrates it is now possible to have a very meaningful real time interaction with an RNN.

In the TedX talk, Yeff develops this idea, suggesting that we can “use technology as a mirror”: “we could use machine learning to push us in any way we want..... I could play phrase chess with myself” and “we can go beyond with how we augment”. There’s a comparison to the training of AlphaGo here, which developed its capacity by playing go against itself. In other words, RNNs could be used to train our self awareness and to promote our creative development.

Yeff, Harry (2013) “Move”. Available at https://www.youtube.com/watch?v=YH5ty3Kucz4

Yeff, Harry (2019) “Reeps One ft. A.I. ‘Second Self’”. Available at https://www.youtube.com/watch?v=q981cTdL0_Y

Yeff, Harry (2020) “How Battling A.I. Unlocked The Power of My Voice”. Available at https://www.youtube.com/watch?v=wTMMopLYJn4

Composing Poetry with Predictive Text

This blog post looks and predictive text as used for poetry composition, similar to the example I looked at in my curation. The article starts with the observation that "structure breeds inventiveness", and that structures are commonly used in poetry: consider for example the sonnet form, or the iambic pentameter that much Shakespeare is written in. The three words that phones offer are similar restrictions, and that author points out that if you look on Twitter or elsewhere on the internet, you can soon find many examples of predictive poetry.

I'm inclined to agree with the author that the format is somewhat gimmicky, but the observation that a common means of generating predictive poetry is by starting with a given word, or words, and then auto-completing, as in the example above, on "porcelain". "In the beginning" or "My life is" are other well used seeds. This idea of seeding generation is simliar to the other examples I have looked at, with music being generated for example from an initiation phrase. Perhaps with a few more operations, a richer exploration could be arrived at?

Corby, Rhiannon (2015) "Composing Poetry with Predictive Text". Available at https://www.pri.org/stories/2015-05-20/composing-poetry-predictive-text

SampleRNN: An Unconditional End-To-End Neural Audio Generation Model

This is the paper which describes the techniques behind the “endless bass solo” example in my curation, and is used by Databots (https://dadabots.com/) to generate endless examples of music within certain genres, such as death metal. I do not pretend to understand all of the paper: however I will try to summarise its ideas here.

Firstly, in this case RNNs are used to generate audio data, and this presents a particular problem: audio data has a very hgih sample rate, perhaps 16kHz for reasonable quality. This is much more than the level of detail needed to generate character data for sentences, for example. The paper presents a method of generating such high resolution data unconditionally, without significant loss of resolution. As I understand it, various methods of layering RNNs - including the use of different timespan ranges - are tested on human ears, and a 3-tier RNN arrangement is found to be the most convincing. The models are trained using three publicly avaiable data sets, covering speech, human vocalisations, and Beethoven’s music.

Based on this paper, I am not sure what conclusions I can draw for my own project. I doubt that I have time , nor yet the experience, to develop such a sophisticated model, though perhaps I could use it, as DadaBots have. Perhaps the most interesting observation is that the level of resolution required in the output of an RNN is a significant factor to be concerned with: I might be better to work in a domain where this resolution is much more restricted (in the realm of music, perhaps to Midi data, or note or harmony choice within a given musical context). Also it’s notable that unconditional generation is potentially less interactive than conditional generation - though I presume sampleRNN can be used condiionally as well. In the “endless bass solo” example, unconditional generation is used by Adam Neely as a form of accompaniment for composition, with chosen extracts used as the basis of composition and improvisation.

Mehri, Soroush et al. (2017) “SampleRNN: An Unconditional End-to-End Neural Audio Generation Model.” Available at https://arxiv.org/pdf/1612.07837.pdf