Unrolling the Learning Curve: Aesthetics of Adaptive Behaviors with Deep Recurrent Nets for Text Generation

Sofian Audry
School of Computing and Information Science
University of Maine, Orono, ME, USA[1]
sofian.audry@maine.edu

Abstract

Machine learning has traditionally focused on problem-solving and optimization. But contemporary conceptions of art usually describe art as non-purposeful and non-optimizable. In this paper, I propose an alternative approach to using machine learning for artistic creation by using the training phase itself as a generative process of new aesthetic forms. Contextualizing my approach within media art history and the history of artificial intelligence, I describe a series of experiments performed using this approach using Long Short-Term Memory (LSTM) recurrent neural networks applied to text generation.

Introduction

Machine learning has recently become a popular approach for studying artistic creativity and creating new forms of art. Oftentimes, this requires framing the creative process as a problem to be solved using some form of optimization. For example, such approaches have been used to evolve new 3D creatures based on subjective preferences; [1,2] to generate music scores that "sound like" the dataset they have been trained on; [3,4] to transfer a painter's style onto another painting; [5] and even to generate images that often feel "more artistic" (at least to the layman) than those of contemporary painters. [6]

Indeed, machine learning is designed to recognize regular patterns, and when employed for generative purposes, is attuned to reproducing things that already exist. Artists, in contrast, seek to create the unexpected. Optimization is inherently dichotomic to artistic practice. Studies that try to tackle artistic production as an optimization problem are immediately faced with problems such as the existence of multiple maxima (e.g., there is no such thing as "the best movie" or "the best painting"); the possibly infinite and incommensurable domains in which artworks exist; and the fact that art is often precisely described as non-purposeful and non-optimizable. [7,8]

In this paper, I explore an approach to computational art that uses the optimization process of machine learning algorithms as a raw material. This technique unrolls the iterative steps in the training phase, thus revealing the temporal structure of the learning agent's behavior. I examine one particular set of experiments that was conducted using this technique, involving a deep learning model known as a long short-term memory (LSTM) recurrent neural network, trained on a text database. The creative artistic and technical approach is presented, as well as the outcomes. Finally, I discuss the implications of the work in the field of computational media art.

Context

Machine learning finds its origin in *cybernetics*, a disruptive science that impacted not only computer science and artificial intelligence, but also biology, neurology, sociology, anthropology, and economics. Furthermore, it had a profound impact on art in the 1960s, and

1 This research was initiated and conducted as part of my post-doctoral studies at the Comparative Media Studies/Writing, Massachusetts Institute of Technology, Cambridge, MA, USA.

foreshadowed the later development of new media art.

One of the central concepts of cybernetics was that of systems or agents, some of which, using feedback from their environment, were able to adapt over time by trial and error. [9] This very basic concept of an agent iteratively and incrementally adapting to its environment by adjusting its own structure is at the core of deep learning, which is based on layers of densely interconnected agents, called neurons, which work together to achieve a greater, more complex level of agency at the global scope. In current deep learning applications, these millions of agents are force-fed gigabytes of data, resulting after several iterations in the *foie gras* of the deep learning revolution: fully optimized models often performing above human level.

Since the 1950s, many artists have exploited the adaptive features of cybernetics systems and other learning agents, not by applying optimized models, but by exploding the learning process itself, often running it in real time. Consider, for example, Hungarian artist Nicolas Schöffer's piece *CYSP I*, which was directly inspired by Norbert Wiener's theory of control and communication. [10, p. 472] Or Karl Sims' *Galápagos* (1997), in which visitors are asked to select their favorite artificial 3D creatures in a virtual environment, and where the selected creatures' genetic code is then used to create the next generation using genetic algorithms. *Performative Ecologies* (2008—2010), by architect Ruairi Glynn, is another example. Inspired by the work of Gordon Pask, especially his 1968 installation *Colloquy of Mobiles*, Glynn's installation creates a conversational space in which dancing robots evolve in constant interaction with one another and with the public.

Most of my own work over the past decade has focused on the design of computational artificial agents, and documenting the performance behavior of these agents in the real world. For example, in my series of site-specific interventions *Absences* (2008-2011), I created small, autonomous, ephemeral agents that acted within natural environments, such as forests and mountains. My robotics installation *Vessels* (2010-2015), created in collaboration with Samuel St-Aubin and Stephen Kelly, involves a group of autonomous, water-dwelling robots that react collectively to their environment through an emerging group behavior. Through this earlier research I developed an interest in how self-organizing and adaptive processes impact both artistic practice and the viewer's experience. Hence, in *Vessels*, a genetic algorithm procedure is used to allow robots to collectively converge to a common group behavior. A similar mechanism has been explored by Stephen Kelly in his work *Open Ended Ensembles* (2016), in which two agents use genetic programming (GP) to move along a fluorescent tube.

Artist and media theorist Simon Penny calls these kinds of works "embodied cultural agents" or "agents as artworks" and integrates them within the larger framework of an "aesthetic of behavior", a "new aesthetic field opened up by the possibility of cultural interaction with machine systems". [11] These works are distinct from so-called generative art, which uses computer algorithms to produce stabilized morphologies, such as images and sound: their aesthetics are about the performance of a program as it unfolds in real-time in the world through a situated artificial body.

In my past work, I developed an ontological framework of behaviors by looking at the distinctive way behavior morphologies unfold over time. [12] While existing taxonomies of cybernetics systems have focused mainly on their relational and structural aspects, I look at the temporal dimension of agent behaviors and its aesthetic potential. [13,9] In particular, I hypothesize that adaptive behaviors are distinguished from non-adaptive behaviors by their ability to change over time and therefore belong to a "second order" of behaviors – those whose behavior evolves over time. With that in mind, we can start considering how the shape of a behavior emerges from randomness (morphogenesis), transforms over time (metamorphosis), or remains stable (morphostasis).

Using this framework, we can establish that most learning algorithms go through a phase of morphogenesis, during which their behavior changes, until they eventually stabilize in a final stage of morphostasis. I posit that this process of transformation and stabilization is artistically relevant and can be harnessed as a creative method.

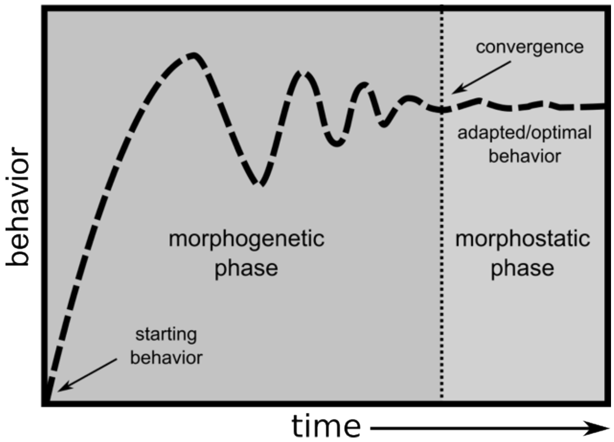

Fig. 1: Schematization of the temporal evolution of an adaptive behavior. Distance along the vertical axis represents difference in the form of observable events produced by the agent. The graphic shows how second-order, adaptive behaviors iteratively change over time through a process of morphogenesis, until they stabilize into an optimal first-order behavior, thus entering the phase of morphostasis.

Approach

In this research, machine learning is used to generate new forms of *behavior*. Following cybernetician Gordon Pask, we define a behavior as a stable form of events caused by an agent, as perceived by an external observer. [14, p. 18] This work fits within the larger artistic discipline of agent-based art – what artist Simon Penny calls "behavior aesthetics". These works engage the performance of one or many synthetic agents as they unfold temporally in the world through situated artificial bodies. [11, 398] Such works are distinct from so-called "generative art" or "algorithmic art", which use algorithmic processes not as an end, but as a means to produce stabilized morphologies, such as images, sound, and text. [12]

This study involves a series of artworks in which LSTM recurrent neural networks were trained on a single text corpus: a version of Emily Brontë's novel *Wuthering Heights*, adapted from the Gutenberg online library. [2] Snapshots of the trained models were saved on disk at different steps in the learning process, resulting in a set of increasingly optimal models. These models were then used as part of a generative process to create a new text.[3]

The first artistic output of that approach, *for the sleepers in that quiet earth*, takes the form of an artbook printed as a series of 31 unique copies,[4] each of which has 642,746 characters – the same length as the version of *Wuthering Heights* that was used for training the neural network. Each copy is generated by a deep learning agent, known as LSTM, trained on the book. LSTM recurrent neural networks are a kind of artificial neural network with recurrent connections, which can "learn" from sequences of data, such as words and characters. They are used in state-of-the-art language processing applications, such as speech recognition and automated translation.

The result is a unique record of the agent as it reads the book and learns the probability distribution of characters, thus somehow becoming increasingly "familiar" with its syntax and style, while at the same time becoming more and more complex in its generative features. This unicity is important, because I see the work less as a trace of the agent's behavior than as a way to experience its behavior as if it were happening in real time.

Like many other deep learning systems, LSTM agents are both predictive and generative. In most scientific applications, it is their predictive capabilities that people are interested in. For example, in machine translation, deep learning systems of the LSTM type are used to compare the probability of different candidate translations and keep the one that is more likely.

Another unique feature of deep-learning systems is that unlike other AI approaches, they improve iteratively. Starting from nothing, as they become more and more exposed to data, they improve and become better at prediction, which also directly impacts their generative capabilities, if they have any.

2 http://www.gutenberg.org/cache/epub/768/pg768.txt
3 The source code used in this project is available here: https://github.com/sofian/readings
4 The work is published at Bad Quarto. Editor: Nick Montfort.

These two ideas – generation and adaptation – form the basis of *for the sleepers in that quiet earth*. My intention in this work was not so much to produce an accurate "optimal" system that could generate rich, human-level, grammar-correct sentences. Instead, I sought to allow the hesitant, strenuous learning process of the system to reveal itself as it goes through all of its sub-optimal states of being.

Another key conceptual dimension of the work resides in the ability of the agent to be both a reader and a writer. If we picture the text of *Wuthering Heights* as the "world" in which the agent dwells and tries to make sense of by "reading" sequences of characters, then as it becomes more familiar with its environment, it is also able to "write" new sequences, which can give an insight into the agent's understanding of its world. The performance trace of this agent is made concrete in the archetypal object of authorship: a book.

I decided to distribute only a printed version of this book, not a digital version. This aspect of the work is crucial, as it lends a physical materiality to the agent and confers an identity beyond its abstract virtual existence. The artbook format contributes to the hybrid nature of the work, combining visual arts, electronic arts, and electronic literature.

The second output of the project is a series of two sound-art pieces and one performance realized in collaboration with Erin Gee[5]. These works explore different modes of revoicing texts generated by the algorithm, using a technique known as Autonomous Sensory Meridian Response (ASMR), which involves the use of sonic "triggers", such as gentle whispering, or fingers scratching or tapping, to induce tingling sensations and pleasurable auditory-tactile synaesthesia in the user. The phrases *of the soone* and *to the sooe* are variations on the incremental learning process used in *for the sleepers in that quiet earth*, but using a shorter text generated by a simpler model. Finally, the work *Machine Unlearning* reverses the process as part of a live performance, in which Gee reads a generative text that starts with the fully trained neural network and slowly regresses to randomness.

Preprocessing

Wuthering Heights contains a few more than 600,000 characters, which is rather small compared to state-of-the-art language modelling datasets, which usually contain several million characters. [6] Starting with an open-access version of *Wuthering Heights*. [15] I slightly reduced the complexity of the learning task by reducing the number of different characters encountered, by (1) making all the letters lowercase (so that the agent does not need to distinguish between uppercase and lowercase letters); and (2) removing low-frequency characters such as parentheses, which appeared only a few times in the text and would only confuse the agent.[7]

Training

To produce the work, an LSTM was trained on the complete text of *Wuthering Heights*[8] over many iterations. Snapshots of the agent's weights were saved at different steps in the learning process, from the beginning, where it is initialized randomly, to the end, after it has read the book 150 times.

Learning was asymptotic, with many changes happening during the first steps of training. This resulted in the system appearing already "overly trained" after the first epoch. To compensate for this, I saved 200 snapshots during this first run-through using mini-batches of different sizes (Fig. 2).

5 https://eringee.net

6 As a point of comparison, consider the difficulty of learning how to write a book in an language unknown to you, with the only information being a single book written in the language.

7 The preprocessed version of the text which was used as the training set is available here: https://github.com/sofian/readings/blob/master/data/wuthering.txt

8 Some basic preprocessing was done to the text, as I explain later.

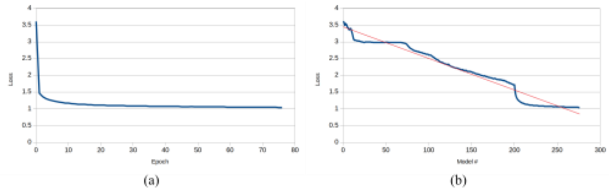

Fig. 2: Training loss (categorical cross-entropy) plotted against (a) the training epoch for the first 75 epochs, and (b) the saved model number up to the first 75 epochs. These graphs show how the process of saving models during the first epoch flattened the learning curve, allowing for more fine-grained evolutions during the generative step. Notice that the first 200 saved models happened during the first epoch alone.

These 351 snapshots – one in the starting state, 200 during the first epoch[9], and 150 (one per epoch) for the rest of the process – were then used in a generative fashion to produce each version of the work. Each snapshot was used to generate an approximately equal portion of the 642,746 characters in the book.

The way the LSTM was trained helps understand its behavior during the generative phase. The network modelled the distribution of sequential text patterns by estimating the conditional probability of the next character x_i given the past N characters $h_i = x_{i-N} \ldots x_{i-1}$:

$$P(x_i|h_i)$$

This probability distribution is represented by a function that produces one probability value for each possible character. For example, let us say that the $N=10$ previous characters seen by the agent are “wutherin”. After training, we would expect the agent to emit a high probability $P(g|wutherin)$ for the letter *g* (wutherin*g*), a lower probability $P('|wutherin)$ for a single quote (*’*) (wutherin*’*), and near-zero probability for every other character.

The network can then be used to generate new sequences, simply by sampling randomly using the distribution and repeating the procedure. To get back to our previous example, after choosing the letter *g*, the agent would sample a new character, this time using the input “uthering” – in which case we would likely expect a high probability of *s*, a white space (_), and other punctuation marks (*.,?!*).

This kind of statistical approach, which looks at the previous *N* units in a sequence, is known as the Markovian process, which is very common in natural language processing. [16] One of its limitations is that it makes the assumption that the closest elements in the past are the most important for predicting the future, which is an imperfect premise to say the least, especially when it comes to language, where there are often very long-term dependencies. This explains to a large extent why the sentences generated by the agent, even in the later stages of training, are somehow detached from one another, as the neural network fails to grasp long-term dependencies between sentences.

To model this probability distribution, I used an LSTM network with two layers of fully interconnected hidden units with 200 neurons each. Input streams were sent by chunks of 100 characters using a sliding window ($N=100$). Input characters were represented using embeddings, a technique in which each symbol is represented by a vector, which is itself trained. For example, in this work, I used embeddings of size 5, which means that each character is represented by 5 different values. These values can be seen as a representation of different characteristics of each character that can be useful for the system to make better predictions over sequences. For example, the first value might represent whether the letter is a vowel, and the second value whether it is a punctuation mark. [14]

Generating

After the training, I obtained a series of probability distributions at different stages of the evolution of the model, which were then used to generate each book.

Let $f(x|h, \theta)$ be the output of the LSTM for character x, given the N past characters h and the set of weights θ.

The probability distribution is represented by the LSTM using the following *softmax* function:

9 In machine learning jargon, an *epoch* corresponds to one full iteration over the training dataset – in this case, the complete novel.

$$P_\theta(x_i|h_i) = \frac{e^{f_\theta(x_i|h_i)/\tau}}{\sum_{x\in V} e^{f_\theta(x|h_i)/\tau}}$$

where V is the set of all possible characters (i.e., the vocabulary).

Here the hyper-parameter $\tau \in [0,\infty]$ is called the *temperature* and is typically set to 1. Raising the temperature spreads out the probabilities, making them more uniform, while lowering it makes the distribution peakier, thus making the agent even more likely to choose the letter with highest probability.

Temperature Adjustment

After some experiments, I noticed that the probability distributions in the early stages were "spread" too much across the characters (i.e., there were not too many differences between each probability) and that the agent would thus generate text that appeared "too random" for my taste. I therefore decided to slightly adjust the probability distribution to make it more "peaky" by decreasing the temperature τ – thus effectively heightening the probability of the most probable elements and decreasing the probability of the others.

However, this approach seemed too "greedy" in later stages, in which the agent became complex enough to consider different sequences of construction and completion. Thus, as the agent's training progressed, I adjusted the probability distribution to be more "spread-out" to encourage diversity (Fig. 3).

Shortlist

Still, since no character had zero probability, there were always cases in which the agent would accidentally generate a completely arbitrary character. To limit this phenomenon while allowing variety, I forced the agent to choose among only a shortlist of the n most probable characters. So the final probability distribution is as follows:

$$P'_\theta(x_i|h_i) = \begin{cases} \frac{e^{f_\theta(x_i|h_i)/\tau}}{\sum_{x\in V_n} e^{f_\theta(x|h_i)/\tau}}, & \text{if } x_i \in V_n \\ 0, & \text{otherwise} \end{cases}$$

where $V_n = V_n(h,\theta)$ denotes the set of n characters x with the largest value $f_\theta(x|h)$.

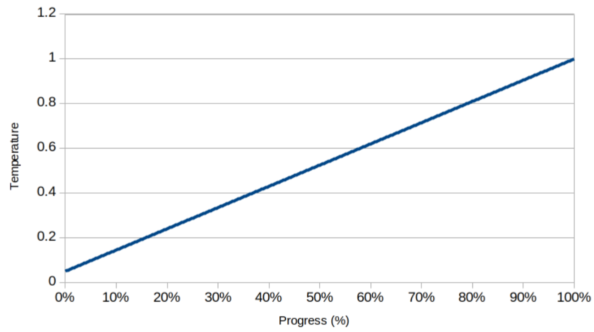

Fig. 3: Evolution of temperature (τ) throughout the book-generation process.

Transitions between Models

Finally, to allow for smooth transitions between each block of text generated by each model, in the last part of each section, I interpolated the probability distributions of the current model and the next model to generate each character. This was parameterized by a transition factor $\alpha \in [0, 1]$, representing the point of transition in each block at which I start interpolating. To generate *for the sleepers in that quiet earth*, we used $\alpha=0.2$; therefore, the last 20% of each of the 351 blocks of text (each averaging 1833 characters) was obtained by linearly interpolating the current probability distribution and the one of the next trained model.

Postprocessing

The final production of the artbooks *for the sleepers in that quiet earth* involved an additional step. Through discussions with editor Nick Montfort, we implemented a few minor changes to convert the raw generated text into book format. For instance, we interpreted the appearance of the word "chapter" followed by roman letters in the generative text (eg. "chapter xix")[10] as an indication of a new chapter, which we therefore formatted differently with a page break and bold typeface.

Results

This section discusses the results of the generative process through an in-depth

10 Notice that these appear randomly. For example, "chapter xi" might appear before "chapter iii".

examination of an unpublished version of *for the sleepers in that quiet earth*. In this section, I describe the progress of the agent as it runs through the reading in terms of time. Here "time" is understood in terms of character position and is represented by the symbol *t*. There are 642,746 individual characters in the original text. So for example, at time *t=64,274* the agent is about 10% into the book, and at time *t=321,373* it is halfway through.

Morphogenesis

The behavior of the writing agent throughout the learning process manifests itself in a number of different ways, corresponding to the state of the agent as it becomes more and more attuned to the "world" it lives in – that is, the text it is reading. As is traditionally done, the neural network was initialized with random weights, representing a neutral state. At this point, the agent had not been subjected to any observations and therefore, had no understanding of the world. Accordingly, in the first few pages of the book, the agent behaved completely randomly, as it had been initialized with random weights.

The agent then proceeded to read the book one character at a time to build an internal representation of how character sequences are generated in Brontë's novel – in other words, by building a model of the author's style. In so doing, it learned more and more about the author's style as it read, starting with building a comprehension of sequences at the character level and incrementally building from this to groups of two, three and four characters, forming syllables, then words, and finally complete sentences.

Following is a case study of a particular unpublished "reading" of the book, and thus construction of an LSTM agent. Here is an excerpt of the first "sentence" generated by the agent:

```
k:jnjw :: j :: lc ; jc ::: c:cnqnnn−−“;x!khwwxwswvsvwvdxd’nx:
nc’i’“’‘;’n’g;pg;pguunm−nmcovo:fow:wwwjdd:nden:’‘:nn’nhk!
knhu?y?msy?yyywwowmww:fwfwwbdwdjfdnj,:jdzzr,lzrk−dqrk
”q−“k.”−−“.c”chhcixhcszzzx’
```

Excerpt at t=0

Early on in the training (after reading a few characters), the agent started to utter erratically some of the characters it had seen:

```
        t              t          t t   t t   t tt
 t             t  t   tt        t         t                t
t    t                      t       t        t        t
          t      t                      t     t             t    t  t
        tt   t    t    t t                     t         t    t      t
      tt t t        t       ttt       t      t      t
   t        t      t          t   t            tt          t         t
   t  t               t    t t                   t                  t
```

Excerpt at t=40

Later on, when it had seen more, it became obsessed with white spaces and frequent characters such as the letter "e".

```
                    e
                     e           e

                                                            e
 e                                                              e

        e                                 e
                            e
```

Excerpt at t=530

These fixations can be explained through the probabilistic approach governing the system. More frequent characters simply have a higher probability of appearing in the text. For example, imagine yourself pointing to a random character in a book and trying to guess what it is without any context; you would likely have a higher chance of making the right guess if you chose a white space than a character.

After reading a few hundred characters, the letters produced by the neural net became more condensed, and we saw appearing some character duplicates. These were the early steps of the agent moving from merely counting the frequency of characters as a predictive measurement. After it read about 5% of the book, the letters became more condensed and the agent even started to tentatively concatenate frequent letters:

```
  t e e          eea e      e     e       a   e   a    ee e    e
 ee        a             e   e                 e          e
e      e               e    e e e               a          t
   a e         t                      ee     e            e       e   e
  a e          t     e ee     a    e     e      a   eae   t       ee  e  t
  e e        e       e          e   a  e       e       ee
   e ne         n      a   t     e    e    e      e    n         e       aa
       e         e        a          a   a        ta     e           e   t  e
```

Excerpt at t=33,490

The Glitch

Surprisingly, not long after this point, the agent seemed to regress to an earlier stage and started behaving erratically for a while. This event happened in only one specific case. I have not been able to replicate this or explain the reasons for this glitch, despite several attempts.

```
“”“j” zjjjj ” jjj ”“” jj ” jj ”“  jjjjjjj  ”  jjejjjjj      “”j”j” jjjz ”“”
 jjjjj  :j”:  jjjjjj  ”“jj      t  t      ee  t  ee        jj      ee  te
ee  ea        e”“j:    e      “jj ”“j”j    e  t        e      e      e  ae:”jx:
 jjjqjj  :q”:” jjj ”   jjjqjj  :  jjjjz ”jxj   jj ”: jjz : jj   “”“ jjjjj  ”  “: jj
” jjjxqj ”“”“”j : j” jjj ”  jjjjj  ” jjje ”e     tt   i   t      e            aejjjjjx
  t       e     e     t          tt   a       jjj        j  e      aa    e  e     e     ee
ea  e     tt   e   i      j:”:    jjjjxjj  :”:j :z”“    “”“ jjjj ”   jjjjjjxxxjjxj
”“:” jjjxjzj  :”q” jqxjjj ”“j”j”   jjjqjjjjj   ”jx” jj      jejj    : zjjj :j
```

Excerpt at t=43,090

My best explanation is that this was due to an early attempt by the neural network to make sense of double-quotes (“”), which is one of the hardest mechanisms to understand for a neural network, as it involves looking backwards to a previous point in the sequence – as opposed to learning about syllables, which involves looking back only one or two characters.

This, as well as the presence of tentative sequences of double-quotes in the next few learning steps, give a hint in this direction – although I was not able to verify it with certainty. Importantly, whereas I ran several training procedures to produce the work, tuning the model and the training procedure, this “glitch” appeared in only one of these experiments. Even a slight modification in the training data, such as removing the chapter titles at one point, prevented the appearance of the glitch. Since I thought this was such a fascinating accident, I decided to work with the specific experiment that produced it.

Morphemes and Proto-Words

Not long after resolving the “glitch”, the agent eventually relaxed its generation of spaces. It seemed to have finally learned one of the most basic principles of English language: the separation of groups of letters using individual spaces. From this point on, it started to tentatively build morphemes of increased length, separated by a single space. Sequences were first limited to a series of one, two or three of the most frequent characters.

```
oe e aa oe te o oo oe e e e iea ot e ae t ae oe iat e e e
oe to ie at ee e te e ee e e e oe ee t ee ee ae e o at ee
te a ee ie oe o te te e oe e ee e e ee e ea ie oe io ee e
te oa oe ta o oa se oe t te ae e ee e e ee oe e o oat o et
 e ee e iae o see  oe oo oe te t ae ae ee t o oe te e e ee
 t e oe ie ia a te en a ao te e to e oe e ie ee ot t oe e
ea ee ie e oe e se a oe ee o te e oe t e a ea e e et ee t
e te a ot ae a e ae oe e ooe ae e a oe oe ioe e se tae e
```

Excerpt at t=59,410

Soon the agent started combining more diverse groups of letters. Short words even started appearing.

```
he to toe  site  son ae  tot  te  th  aos  tin  thr  tot  to  toe  tot
 to  te  tos  his  toot  toe  tit  tot  tat  hoe  tot  te  to  hh  te
ter   tit  hon se  te  toe  hho io  to   tit   te  tin  han  tos  hat  tot
 to  tae  tos  ioe  sos  tan  ioe  hote  ao  tat   iis   tee  to  tat  io
sot  toe  aote  ho  toe  the  tht  tot  tot  tit   tit   thre  ho  toe  th
 his  aot  toe  to  toe  toe  hoe  iho  ton  he  tis   te  hot  tot  tis
toe  aoe  toe  hhr  te  aot  tos  the  th  ais  te  ioe  toe  aoe  to  te
 tos  hos  tot  toe  ton  io  hot  tate  ih  toe  hee  ion  hit  tos  te
```

Excerpt at t=113,170

This was shortly followed by early attempts to build short sequences of words, some of which were even correct English, such as “in the”, “that is”, “the mind” and “the mister”.

```
the  sease  and  the  his  an  the  mind te  at  to  seared  the  toul
  tout  the  tis  and to  to  hhe mas is  he  the  toun  the  i
setened  the  the  me the  tor  a  sist   hit  you wos in  sos  in
 tis  to  ho the  toat  hat  hhe  seter  the  sor  ind  the  in  toe
thas  an  the  herare  the  tore  the  more a  the  the  the  mited  i
  anserting   the  hor  ho  touthed  in  a  tont  to  ceith  to  t
he  to  mise  it   teat   in  the  sorton   in  the  tor  the  that   that
 is   seres  of  at  ind  an  the  sease  the  mister  af  the  in  se
 seret  an  tout  the  to  the  ind  ander  to  se  in  seate  sis  the
heathe  i   seited  the  sant  i  sind  tho  ceatter  to  the  he
```

Excerpt at t=215,570

Punctuation and Sentences

After reading about a third of the book, the agent started using punctuation. For example, here is the first use of commas:

```
the  heer  to  chered  in  at  i  son  to  sere  the  sorter ,  and ho
merter  the  sorer  the  sand and his  the  meret  ind  the  mored
 to  me cered  the  coring  in  anter  the  mroned an the  hor
here  the  ceind  the  sere  the  sanding  in  the  carter ,  and  it
 i  seind  he  mrer and so  anter  the  the  ter  mererter  to  the
```

Excerpt at t=227,090

At about two thirds through the book, the agent could construct sentences of varying length, making syntactically appropriate use of periods, commas, and quotes. The sentences were mostly nonsensical and grammatically imperfect. Yet they seem to mirror some of the core aspects of the original text, including the use of the first person, an abundance of dialogue, and the construction of long sentences with many complementary clauses, a style that was common in 19th century English literature. Above all, it was the rhythmic qualities of the text produced by the artificial agent that bore the closest resemblance to Brontë's style.

> 'so satherine. he deat i could to she laster it the sranes and the door his sathered. 'i his lestanded to srean and hime wall at and the lister and santion.'
>
> 'you have wor trearing her an the care, and the look. i was so deat to the litter to see it i chould her a lested and to his fore the deand to her and to sathering to see the lounder her her seed to the reanten, his for had so the roster on a sould be the lose, and the had ase a meter to the leas on a mate a merared of his for shanted to me to sear the lease the dade and aspering his to sere and then i meath. i wall not he couse in the heas of the laster of him to her to mishers. i was not hear he so sann the linton his her the fide the rase her his couster the srarged a sranted the had sarle the has the loor.'

Excerpt at t=448,530

For comparison, consider this excerpt from Chapter VIII of *Wuthering Heights*:

> I guess she is; yet she looks bravely,' replied the girl, 'and she talks as if she thought of living to see it grow a man. She's out of her head for joy, it's such a beauty! If I were her I'm certain I should not die: I should get better at the bare sight of it, in spite of Kenneth. I was fairly mad at him. Dame Archer brought the cherub down to master, in the house, and his face just began to light up, when the old croaker steps forward, and says he—"Earnshaw, it's a blessing your wife has been spared to leave you this son. When she came, I felt convinced we shouldn't keep her long; and now, I must tell you, the winter will probably finish her. Don't take on, and fret about it too much: it can't be helped. And besides, you should have known better than to choose such a rush of a lass!" [15]

Improvements

This is an excerpt after one epoch of training – that is, after the agent had read the book once. At this point the agent had learned to generate complete sentences, with a few glitches. Many of these sentences are still grammatically incorrect and somewhat random. It is as if the agent could only "see" two or three words in the past, with usually only short sequences of two or three words making logical sense together. Consider for example the progression in the following sentence generated after the first epoch:

> i dade the cornert of her and, and he sheat it it with the deant a sood of the housh he had sather to him, and i had not the haston, and she had a contred to her to the sattle to the conder of his so stoul him the did or the seen.

Excerpt at epoch 1

From this point forward, the neural network was trained for several epochs, having re-read the novel up to 150 times. Changes in the agent's output become less perceptible over these later iterations. The first epoch allowed the agent to grow from pure randomness to building morphemes, words, and full sentences with punctuation. In the following iterations, the agent seemed to expand these basic building blocks by (1) polishing grammar, (2) expanding vocabulary, and (3) diversifying the length and structure of sentences, including producing dialogic constructs that are common in the original text.

To get a sense of this evolution, here are some sample sentences from epochs 20, 80, and 150, which may give a sense of the transformation in the agent's behavior.

> 'you have the delight is to spend them to speak to be a single things!'
>
> 'it was a grief of more truth, and the satisfaction which i was a bad contents and the house for me, and struck her features with me to the servant to and a mean and startling. he would be a state that is the case. i don't like that i shall be the door, i dare you?'

Excerpt at epoch 20

> 'were you hear the plant of his father's sort of morning? what do you stay it to my hands to me! i'm not married and desire to be always there, and you would send it out of the farm. i hope you had been hardly to have a foold at all. i can be all start and talking a minute in his senses. when i asked if i was no far which she should be sure from the house, and i could not be silent on the fire, and hid her little abode on the heights, and i have a solret associations,' interpupted stared; 'i'm nearly to do you to send them,' said the strength.

Excerpt at epoch 80

> 'they would not resign you to the danger over through to me!'
>
> 'he's both a books then,' he added, sufficiently.
>
> 'who is them insolently—spaking to him,' said catherine. 'you are a seat,' he said: 'i don't might wark his stall in that third that they are.'

Excerpt at epoch 150

Machine Unlearning

Proceeding incrementally using models of increasing accuracy is not the only way the suggested method can be used. In *Machine Unlearning*,[11] artist Erin Gee performs using a voice technique known as Autonomous Sensory Meridian Response (ASMR). She reads a text which was generated using the inverted process presented above. Here we simply regress from a fully optimized system down to an untrained model.

Following is an example of such a text, which was read by Gee during the work's premiere in May 2018:

```
seemed to tell you white deep or a look into his arm. she heard
the missive in my house, and set up–stairs, and sobbing in ten
handsome, some tranquillity of all a protention. it in twe fitch in
him her cather, i heathcing to semer to me a garing, and there, i
canted in her head, we her the sarsh, and i danded to so her and
the him.' and the master he would to that in is, he'll befingens, i
sastered, and she ware to see andittace, thall the doald and didse
of her his a sat the mrome, it his sromed insores of the gance and
 is the hid, i das net to sere than has song the salded to his a
mould and her with tourd to thing ham she fat hinds, i songing of
 his and a dor that to sent i coulled he sonns a was of the wome
fot of to crasing the minger to sorled a seolens of i mour has at
her the was her it mrot her heat i har sastering her heun her
hishen the sos to ho hande, the poon, and to a dond a to to the
here a salle a sho hor the her sardons heut aften in the cand the
sising the meer hor i sear toul an the cister the the sast an that
toul he sese the min ho hhe houl the the mas as seast af hhe hhal
hhen a ser an wish if the te mind ho andere his hor in a mh ter tor
 tha mhe to me ser teat and to sout he so see tarere that i wean on
 the int o weret tim ind won to mon tanee tor tee tor han hate tee
tan heee ho han th tat tite sor ie tat sat tin th tot ae tos hit tor
heerrtr hoe th sor sose ths tean tee to sos seoe hh aot she the tatr
toe th tiue teee sot hae hoe to hee tie co tor tad tor aer ahr se te
taat sot tae to aate tate toe ane te toe hoe the tae te tn hoae aie soe
 aoo ton hon tare he hoes oae aae ha ae ae ood he he ha te ta te ae
 hoee oa te tte o ot te te ta s oee oe ee te  eteo e   eeeo tt t o oo
oetn  eoee    eoeei    aet  te     ee  eet  e ea  esa  e       e        ee
e   et t e e e  e eate  e    oe    e      te      e       ee a    a          a  e e a
te  t e   ee          na     e e oe  e             e           oe        ea a t   et o
o       o        eo    e          ao e                e       ee     e  ie    e e e    e     ee
     e                  a           e     e          e     e     e                t    t      e
                        e          e   e     e   e    e e                 a
                  eet              a                  e         ee             e   e
                                 e                          e                  t
iiiii iiii                       t  t                 lb l lllbllllllllllll
   h   tt htt tthttttt!t thhtt nw:fe:jj:e:e
```

The generative text read in *Machine Unlearning* (2018).

11 https://eringee.net/voice-of-echo-ii-meta-marathon-dusseldorf

Conclusion

The computational artworks described in this paper span diverse approaches, such as electronic literature, generative art, and behavior aesthetics. They make use of deep learning recurrent neural networks, not so much as a way to generate novel and creative writing by taking advantage of the system's ability to imitate human performance, but to reveal the learning process of the system. In other words, the approach explored in this study subverts the core purpose of artificial intelligence, whose aim is to reproduce or exceed human performance, in this case, by imitating the style of a well-known English author. Instead, it focuses on the behavior of the artificial agent as it tentatively tries to achieve its goals.

Rather than focusing on the literary prowess such computational systems can achieve when they are fully optimized, these works offer a unique insight into the inner workings of a machine learning algorithm by turning the experience of reading and listening into an encounter with a learning agent. While these works are certainly different in many respects from canonical forms of agent-based artworks (such as those employing situated robotic systems), it shares with them a unique focus on using behavior as an artistic form on its own – in these cases, through experiencing the learning journey of an artificial deep learning agent.

More research needs to be done to understand the relationship between the learning curve and the perception of behaviors, looking at how changes in the error rate correspond to observable changes in the agent's behavior. Furthermore, while this study is limited to the specific domain of text generators, future works should focus on applying the approach to other domains, such as robotics, sound and images.

Acknowledgements

The author would like to thank the Fonds de Recherche du Québec – Société et Culture, the Massachusetts Institute of Technology, NVIDIA Corporation, The Trope Tank, Bad Quarto, and Dr Nick Montfort for their support.

References

1. Karl Sims, "Evolving Virtual Creatures." *ACM Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques* (1994): 15–22. https://doi.org/10.1145/192161.192167.
2. Stephen Todd and William Latham, *Evolutionary Art and Computers*. (Academic Press, 1992).
3. Gaëtan Hadjieres, François Pachet, and Frank Nielsen, "DeepBach: A Steerable Model for Bach Chorales Generation," *ArXiv:1612.01010 [Cs]*, December 3, 2016. http://arxiv.org/abs/1612.01010.
4. Douglas Eck and Juergen Schmidhuber, "A First Look at Music Composition Using LSTM Recurrent Neural Networks." Technical report" Manno, Switzerland: *Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale*, March 15, 2002.http://dl.acm.org/citation.cfm?id=870511.
5. Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. "A Neural Algorithm of Artistic Style." *ArXiv:1508.06576 [Cs, q-Bio]*, August 26, 2015. http://arxiv.org/abs/1508.06576.
6. Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone, "CAN: Creative Adversarial Networks, Generating 'Art' by Learning About Styles and Deviating from Style Norms," *ArXiv:1706.07068 [Cs]*, June 21, 2017. http://arxiv.org/abs/1706.07068.
7. Simon Penny, "Agents as Artworks and Agent Design as Artistic Practice," in \ *Human Cognition and Social Agent Technology* edited by Kerstin Dautenhahn, *Advances in Consciousness Research*, 19 (2000): 395–414. https://benjamins.com/ catalog /aicr.19. 18pen.
8. Leonel Moura Pereira and Henrique Garcia, *Man + Robots: Symbiotic Art* (Villeurbanne: Institut d'art contemporain, 2004).
9. Arturo Rosenblueth, Norbert Wiener, and Julian Bigelow, "Behavior, Purpose and Teleology," *Philosophy of Science* 10, no. 1 (1943): 18–24.
10. Maria Fernández, Maria, "'Life-like': Historicizing Process and Responsiveness in Digital Art," in *The Art of Art History: A Critical Anthology*, edited by Donald Preziosi, New ed., (Oxford History of Art. Oxford; New York: Oxford University Press, 2006), 477-487.
11. Simon Penny, "Embodied Cultural Agents: At the Intersection of Robotics, Cognitive Science and Interactive Art," in *AAAI Socially Intelligent Agents: Papers from the 1997 Fall Symposium*, edited by Kerstin Dautenhahn (Menlo Park: AAAI Press, 1997), 103–105.
12. Sofian Audry, "Aesthetics of Adaptive Behaviors in Agent-Based Art," *Proceedings of the 22nd International Symposium on Electronic Art*, 2–9. Hong Kong, 2016. http://isea-archives.org/? page_id=36370.
13. Peter A. Cariani, "On the Design of Devices with Emergent Semantic Functions," State University of New York at Binghamton, 1989.
14. Gordon Pask. *An Approach to Cybernetics*. (London: Hutchinson, 1968).
15. Emily Brontë. *Wuthering Heights*, 1996. http://www.gutenberg.org/ebooks/768.
16. Christopher D. Manning and Hinrich Schütze, *Foundations of Statistical Natural Language Processing*. 1st edition. (Cambridge, Mass: The MIT Press, 1999).