for the sleepers in that quiet earth.: Experiencing the Behavior of a Deep Learning Neural Network Agent through a Generative Artbook

Sofian Audry
School of Computing and Information Science
University of Maine, Orono, ME*
sofian.audry@maine.edu

Abstract

This paper describes an artbook created through the adaptive behavior of a deep learning neural network computational agent as it "reads" a novel. Through this process, the agent builds a model of the syntactic and stylistic principles behind the original text and uses this model to generate new, unforeseen content. A limited set of unique printed copies of the artbook are generated through this process. Each unique edition of the work thus embodies the learning process of the agent as it goes through the adaptive process, one wherein the agent begins from a state of randomness and gradually refines its output as it reads the novel. I examine the text through an analysis of generated excerpts, discussing how they reveal the behavior of the system as an adaptive agent. Practical and theoretical implications are discussed in the context of generative literature, machine learning, and behavior aesthetics.

Keywords

Agent-based Art, Artbooks, Behavior Aesthetics, Deep Learning, Generative Literature, Long Short-Term Memory, Machine Learning, Natural Language Processing, Neural Networks.

Introduction

for the sleepers in that quiet earth. is a computer-generated artbook produced through means of a deep learning, recurrent neural network agent known as a *long short term memory* (LSTM) [3] as it processes or "reads" *Wuthering Heights* – the only novel penned by Emily Brontë, published one year before her death [2]. The work is published in a limited edition of thirty-one (31) unique copies (Bad Quattro, editor: Nick Montfort). The agent encounters the text character by character, familiarizing itself with these syntactical and styllistic materialities of the original text. As it reads and re-reads the book, it slowly reinterprets Brontë's style within the constraints of its own artificial "body", hence finding its own unique "style" or "voice", so to speak.[1] The resulting work unfolds a new narrativity where the reader encounters and experiences directly the adaptive behavior of a learning agent.

The project attempts to subvert how deep learning systems such as LSTMs are used in usual scientific and engineering practice by conflating it with the words and imagination of a 19th century female author.[2] Normally, such systems are aimed at optimization: they iterate over a database (for example, a huge corpus of text extracted from diverse sources) until they converge to an optimal state. The trained neural network can then be used in text-related applications such as speech recognition and automated translation.

In the arts, deep learning recurrent neural networks have largely been used as a means of exploring the aesthetic fruits and limitations of the machine learning process: for example, consider *Sunspring* (2016), a science-fiction movie based on a screenplay generated by a LSTM agent named "Benjamin" which was trained on a databse of sci-fi screenplays.[3] Or 2017 project *Rerites* by David Jhave Johnston, a collaboration between the author and a quasi-recurrent neural network (Q-RNN) to create a series of print-on-demand poetry books.[4]

for the sleepers in that quiet earth. takes an alternative approach to these artistic works by engaging in a literary embodiment of the cognitive gestures of the agent, focusing on the agent's learning *behavior* in all of its deficiencies and unpredictability. Through this work I explore how computation allows for the emergence of a new artistic practice which Simon Penny has described as "behavior aesthetics" [6]. This new domain considers the behavior of artifical agents as a medium for artistic creation – comparable to media such as video, photography, etc. which come with their own sets of effects, affects, materialities and affordances. In general, behavior aesthetics presupposes the presence of an embodied agent interacting in real-time in the real world.

The project, generated by a disembodied agent, yet printed on unique paper copies, resides in a liminal space, as it borrows certain elements from behavior aesthetics, but also from generative poetry and algorithmic art. In this article, I explore the project through an examination of the

*This research was initiated and conducted as part of my postdoctoral studies at the Comparative Media Studies/Writing, Massachusetts Institute of Technology, Cambridge, MA.

[1] Brontë's work was originally published under a gender neutral pseudonym, and years later was deemed "monstrous" by some critics, seeming to dole out unreasonable punishment to its characters who fell victim to their burning passions. This juxtaposition of computational agent with a 19th century author invokes yet another monstrous perversion of consciousness, or perhaps subconsciousness, as the new author learns nothing else of the world but the syntactic patterns of the original tale.

[2] The status of agent as object and objectified resonates with the sometimes questionable status given to non white-male authors in the past.

[3] `http://www.imdb.com/title/tt5794766/`

[4] `http://glia.ca/2017/rerites/`

work itself and a review of the creative process of producing the work using machine learning neural nets. This is followed by a discussion about the nature of the work in relation to behavior aesthetics.

The Work

Most of my work over the past decade has been 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 have developed an interest in how self-organizing and adaptive processes impact both artistic practice and viewer's experience. Hence, in *Vessels*, a genetic algorithm procedure is used to allow the robots to collectively converge to a common group behavior. A similar mechanism has been explored by Stephen Kelly – also a collaborator on *Vessels* – in his work *Open Ended Ensembles* (2016), where two agents use genetic programming (GP) to move along a fluorescent tube.

In my current research, I am interested in using machine learning as an approach to generate novel real-time processes which I refer to as *behaviors*. Following cybernetician Gordon Pask, I define a behavior as a stable form of events caused by an agent, as perceived by an external observer. [5, p. 18] I fit my own work within a larger artistic discipline which I call agent-based art, or "behavior aesthetics" – an expression coined by artist Simon Penny. which implies the performance of a synthetic agent as it unfolds temporally in the world through a situated artificial body [6, 398]. Such works are distinct from so-called "generative art" or "algorithmic art" which use computer algorithms to produce stabilized morphologies such as images, sound, and text.[1]

The work *for the sleepers in that quiet earth.* was created during my postdoctoral studies at The Trope Tank[5] at MIT in 2017, takes the form of an artbook, with a restricted number (31) of unique copies which each counts 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 a long short term memory (LSTM) trained on the book by Emily Brontë. LSTMs – which are used in state-of-the-art language processing applications such as speech recognition and automated translation – are a kind of artificial neural network with recurrent connections, which are able to "learn" from sequences of data such as words and characters.

To produce the work, I first trained a LSTM on the complete text of *Wuthering Heights*[6] over many iterations. I saved snapshots of the agent at different steps of the learning process, from the beginning where it is initialized randomly, to the end where it has read the book 150 times. In particular, there are many changes happening in the agent during its first reading of the book; thus I saved 200 snapshots during this first run-through alone. These 351 snapshots – one in the starting state, 200 during the first epoch[7], and then 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 of the book.

The result is a unique record of the agent as it reads the book and becomes more and more familiar with its syntax and style – and at the same time becomes more and more complex in its generative features. This unicity is important, because I see the work less as the trace of the agent's behavior than as a way to experience the behavior as if it was 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 for comparing the probability of different candidate translations and keep the one that is more likely.

Another unique feature of these systems is that, contrarily to other AI approaches, deep learning systems 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.

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. On the opposite, I seeked 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 being 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 possible sequences which can give an insight on the agent's understanding of its world. The performance trace of this agent is made concrete in the archetypal object of authorship par excellence: the book.

I made the decision to distribute this book only as a paper-based object, rather than in a digital version. This aspect of the work is crucial, as it lends a physical materiality to the agent and confers to it an identity beyond it's abstracted, virtual existence. As such, the artbook format contributes to the hybridity of the work, which in my view and intention lie between visual arts, electronic arts, and electronic literature.

[5] `http://nickm.com/trope_tank/`

[6] Some basic preprocessing was done to the text, as I explain later.

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

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 atune to the “world” it lives in – that is, the text it is reading. As is traditionnally done, the neural network is initialized with random weights, representing a neutral state: at this point the agent has not been subjected to any observations and has thus no understanding of the world. Accordingly, in the first few pages of the book, the agent behaves completely randomly, as it has been initialized with random weights.

The agent then proceeds to read the book, one character at a time, in an attempt 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 learns more and more about the author’s style as it reads, starting with building a comprehension of sequences at the character level and incrementally building from it to groups of two, three, four characters forming syllables, on to building words, and finally complete sentences.[8]

The following is a case study of a particular unpublished “reading” of the book, and thus construction of an LMST agent. Here is an excerpt of the first “sentence” generated by the agent in an unpublished version of the work:

```
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 starts to utter erratically some of the characters it has 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

```
  iiiiii    i   iiiiiiiii   ii   i   iii   iiiiiii   iii   iiiii   ii
i  i   iiiiiiiii    iii    i  i  iii  i  iiiiiil  ii  iii  i  ii    i
 iiii  ii  i   iii  iii   iiiit   iiii  ii  ii  ii  ii  iiii   iiiiiii
iii    iit  i   iiiiiii   ii  iiiiii  i  iiii  i  i  i  ii   iiiiiiiii
  iiiiiiiiiii        iii  i  ii  i  iii  ii   iii  i  iii  i   i
 iiiiiii    ii   iiiiiiiiiiiii   ii   i   i  iii  i  ii  i  i  iii
 i  i  i  ii  iii   iiiii  i  i  iiii  iiiiii   iiiiiiiiii   i  iii  ii
 ii  i  ii  iiii  iii  i  i  i  iii  iiii  iiiii  ii  ii  iiii  i  i  i  i
```

Excerpt at t=130

Later on, as it has seen more, it becomes 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 more chances of making the right guess if you chose a white space than any other character.

After reading a few hundred characters the letters produced by the neural net become more condensed and we see appearing some duplications of characters. These are the early steps of the agent moving from merely counting the frequency of characters as a predictive measurement.

```
      e          e     e e       e eee
            e        e e    e        e e
   e      e     e e                 e      e
                  e   e             ee   e
 e  e             e e       e     e        e
ee          e  e    e      e e    e       e
                    e    e    e    e     e
         e    ee      e e                e  e
```

Excerpt at t=1,810

The Glitch

Around having read about 5% of the book, the letters become more condensed and the agent even starts to tentatively concatenate frequent letters:

[8] In this paper, I describe the progress of the agent as it runs through the reading in terms of time. Here, time is to be understood in terms of character position. 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.

```
 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

To my own surprize, not long after this point, the agent seems to regress to an earlier stage and starts behaving erratically for a while. I have not been able to replicate, let alone explain, the reasons of this glitch, despite several attempts.

```
jj      jj : jj :: jjj : jjjj  :: j : j  e      :          :::: j : jj         o
 t            : j :: j : j :  o                  e  e      ee        e  t  e
               te                               e       e                 o
 e   :! j                                     jjjj  ::  jjj         t      jz
: j : jjj : jj :          “j        :: jj :: j”                    j : j :: jj : j :
 jzjj :” jj           a      e        t                          e  a  e
    a              “?j               e   e               t    t    j
      jj         j ::   jjjzjjjj    o           e       jjzj     a   j :           o
```

Excerpt at t=36,370

It soon becomes even worse:

```
“”“j” zjjjj ” jjj ”“” jj ” jj ”“ jjjjjj ” 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 is 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 moment in the sequence – as opposed to learning about syllables which involve looking backwards to the one or two previous characters. The fact that the glitch happens around a position in the text where the agent encounters one of the first double-quote character, more specifically in the form of a quote inside a quote:

‘ “Maister Hindley!” shouted our chaplain. [...]’[2]

This, as well as the presence of tentative sequences of double-quotes in the next few learning steps, hint into that direction – although I was not able to verify it with certainty. Importantly, whereas I ran several training procedures in order to produce the work, tuning the model and the training procedure, this “glitch” only appeared in one of these experiments. Even a slight modification in the training data – for example, at some point I tried to remove the chapter titles – prevented the appearance of the glitch. Since I thought this was such a fascinating accident, I had no choice but to work with the specific experiment that produced it.

```
“”“”v”“ovn”ob”o ojo   ae    e  t    ee  ee     aae    e  t  e  ee      e  e
e      t   tt   ae    e    a  :b”“!b!!”“v”!v”“  v    e  aei  e  e       ett  e
e       ett        ee       t      nt  et     et “  ja    “”j !”bbb”“”!j”oj”ov”
bx e        tte          t  tee     ee      ea    eee    e     e       e  ta    e    ta”e
;”“”“”“”v”“bj”j”“x!j”“     t        e  ee    ea    e       ee  e         e          ne
 “    eet    aq!”“b”vbv’oo ooonroooo e  e      eeea        te     atee
e      e    e      e  eaee      eet     i       eete       eivo”!boox   xoee       ei
```

Excerpt at t=44,050

Morphemes and Proto-Words

Not long after resolving the “glitch”, the agent eventually relaxes its generation of spaces. It seems to finally have learned one of the most basic principles of English language: the separation of groups of letters using individual spaces. From this point on, it starts to tentatively build morphemes of increased length separated by a single space. Sequences are first limited to 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 starts combining more diverse groups of letters. We even start to see the appearance of short words.

```
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

```
anten a set tean an mor ander and ho sant and the mor
heatheren an thar an ter hand tor hir a tous hor tamle tor
 hearh an ho anter and sos mhe merte and hoat he mam ho
anten a sise ter hat hean has and that in ser at an tou
sared an sorer at a tos tor marles the a sate hor mhan a
at the and samh hor i tor hashen and i samt i sore tor hor
 a sor tor he mer tout af heus a mas a ther a and tor
heurt in a hase on the sote an has thor and the hor heur
```

Excerpt at t=184,850

This is shortly followed by early attempts to build short sequences of words, some of which are even correct English such as “in the”, “that is”, as well as “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 inder the he

Excerpt at t=215,570

Punctuation and Sentences

After having read about a third of the book, the agent starts using punctuation. This excerpt contains the first usage 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

Periods start appearing sporadically, such as in this excerpt:

ind he andersed he he the her to he the sase of her he contered as he he hre hat the hand sore he her i here the mat sor and the porter at a sase to ansered . and i was the herper of the sose he haster and the her i chanded the handser he he hand the cand to the pandt he sortelling

Excerpt at t=280,850

At about two thirds through the book the agent is able to construct sentences of varying length, making syntactically appropriate use of periods, commas, and quotes. These sentences are 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's voice; the abundance of dialogs; and the construction of long sentences with many complementary clauses as is common in 19th century English literature. Above all, it is the rythmical qualities of the text produced by the artificial agent that bears the most 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

As a point of 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!"' [2]

Improvements

This is an excerpt after one epoch of training – that is, after the agent has read the book once. At this point the agent has 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 can 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 is trained for several epochs, having re-read the novel up to 150 times. Changes in the agent's outputs 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 marks. In the following iterations over the novel, the agent seems to expand these basic building blocks by (1) polishing grammar;

(2) expanding vocabulary; and (3) diversifying the lengths and structures of sentences – including producing dialogic constructs which are common in the original text.

To get a sense of this evolution, here are some sample sentences from epochs 5, 20, 80, and 150, which may give a sense of the transformation in the agent's behavior.

> he was a sure that i should never tell me to her heart in the same on the grange and countenance and could not be a strange of my little constant to bear him a chorg and companion and sure the sits that i was a sudden and sense of a moon of the door; i supposed me to receive the subject of a place to his father's any range.

Excerpt at epoch 5

> '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

Practical Considerations

This work follows up from my previous research on adaptive systems. My robotic work *Vessels* (2015) mentionned before involves groups of water-based robots react to their environment and with one another by evolving a group behavior which is realized using genetic algorithms. In *N-Polytope: Behaviors in Light and Sound After Iannis Xenakis*[9] (2012) I designed different kinds of adaptive behavioral patterns actuated through sound and light using self-organizing integrate-and-fire neural nets, reinforcement learning, and genetic algorithms.

for the sleepers in that quiet earth. was sparked by an interest in working with deep learning neural networks. Such systems are now widely used in natural language processing (NLP) applications. In any computer-based work, there is a tension between authorship and autonomy. As in my preference as an artist, I chose to leave as much space as possible for the system to act independently. Yet, I had to make some choices to facilitate the agent's performance. These choices are described hereby.

Preprocessing

Wuthering Heights contains a little more than 600,000 characters, which is rather low when compared to state-of-the-art language modelling datasets which usually contain several millions of characters.[10] Starting with an open-access version of *Wuthering Heights* [2] I slightly reduced the complexity of the learning task by reducing the number of different characters encountered through (1) turning all letters to lowercase (so that the agent does not need to distinguish between uppercase and lowercase letters); and (2) remove low-frequency characters such as parenthesis – which appeared only a few times in the text and would only confuse the agent.

Training

The way the neural network is being trained can help understand how it behaves. The network attempts to model the distribution of patterns in the text. It does so by estimating the conditional probability of the next character x_t given the past N characters $x_{t-N} \cdots x_{t-1}$:

$$P(x_t | x_{t-N} \cdots x_{t-1})$$

This probability distribution is represented as a very complex function that produces one probability value for each possible character. For example, let's say that $N = 10$ and that the 10 previous characters seen by the agent are "wutherin"; then after training we would expect the agent to emit a high probability $P(g|wutherin)$ for the letter *g* (wuthering), a lower probability $P(\text{'}|wutherin)$ for a single quote (') (wutherin'), and near-zero probabilities 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 high probabilities for *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 a markovian process. Markovian processes are very common in natural language processing.[4] One of their limitations is that they make the assumption that the closest elements in the past are the most important to predict the future, which is an imperfect premisse to say the least – especially when it comes to lan-

[9] Directed by Chris Salter.

[10] As a point of comparison, consider the difficulty it would represent to learn how to write a book in an language unknown to you, with the only information being a single book written in said language.

guage 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 that probability distribution, I used a 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 where 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, the second value whether it is a punctuation mark, etc.[11][12]

The total number of free parameters – i.e., the "weights" or "synapses" in the neural network that will be adjusted during training to make the network perform better over time – in the architecture is 493,222, which is almost as high as the size of the database – that is, the number of characters in the book. This suggests that the system is likely to overfit[13] as it is trained. Overfitting is a dreaded problem in machine learning, as it prevents the system from getting good generalizations: as the system becomes too much attuned to the training set and starts remembering it "by heart", so to speak, its performance drops when subjected to new examples. However, in this project, I purposefully chose a high number in order to give enough power to the system to learn to produce interesting and diverse results – an example where scientific and artistic objectives diverge.

Generating

After the training, I thus obtained a series of probability distributions at different stages of the evolution of the model, which were then used to generate each book. After some experiments, I noticed that the probability ditributions in the early stages were too much "spread" accross the characters (i.e., there were not too much differences between each probability) and that the agent would thus generate text that appeared "too random" to my own taste. I thus decided to slightly adjust the probability distribution to make it more "peaky" by increasing the probabilities of the most probable elements and decreasing the probability of the others.

However, this approach seemed too "greedy" in later stages where the agent would become complex enough to be able to consider different sequence construction and completion. Thus, as the agent became more and more trained, I adjusted the probability distribution to be more "spread-out" to encourage diversity.[14]

Still, since no character ever had a zero probability, there were always cases were the agent would accidentally generate a completely arbitrary character. In order to limit this phenomenon while allowing variety, I forced the agent to choose among only the top 5 most probable characters.

Postprocessing

Finally, through discussions with my editor Nick Montfort, we decided to implement a few minor changes to bring 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")[15] as an indication of a new chapter, which we thus formatted differently with a page break and bold typeface.

Discussion

for the sleepers in that quiet earth. is a hybrid work that borrows from different art and engineering traditions and practices. The generated text is far from being "good" and "optimal". So in that sense it does not fit within the canon of electronic literature. We do not seek here to create something that holds together in terms of English writing. On the contrary, the imperfection of the generative system's "voice" is perhaps the primary content of this work. *for the sleepers* is not an attempt in passing the Turing test of literacy, nor in creating an artificial clone of Emily Brontë. What this work affords is not so much the imitation of a style – because the agent never reaches that point – but rather, what is put forward is *the learning process itself*.

It is precisely in this sense that *for the sleepers in that quiet earth.* does not constitute a work of generative electronic literature in the traditional sense, as here the "trace" of the behavior is not differentiable from the behavior itself as the agent proceeds to evolve in its "world" delimited by the character sequences of Brontë's novel.

Through its liminality, *for the sleepers* opens a space for understanding and reinventing concepts and ideas about agents, behaviors, and deep learning. How do behaviors of adaptive systems differ from those of non-adaptive systems? What distinguishes a behavior from the trace it leaves? Does a behavior require the real-time, embodied presence of the agent that generates it?

In my past work, I have established a distinction between different categories of behaviors.[1, 7] I have referred to simple stateless mappings as nonbehaviors (or zeroth-order behaviors). Rule-based, formal systems such as those designed using traditional programming using conditions and loops produce first-order behaviors: given enough time, a perceiving entity with sufficient intellectual capacity can theoretically infer the rules that govern the actions of the

[11] In reality, we rarely see such human-based categories appear in embeddings, as the system finds its own way to represent information which is often counter-intuitive as we do not "think" like an artificial neural network does.

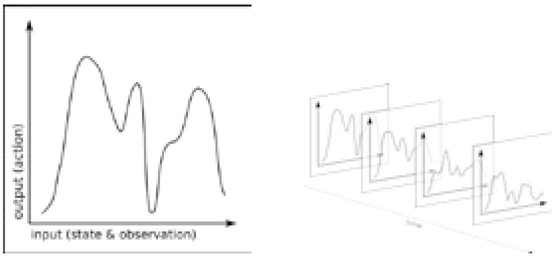
[12] The code for the system is available here: https://github.com/sofian/readings.

[13] Overfitting is a common problem in machine learning where a learning system with too much "power" (too many parameters) is trained on a database that is relatively too small. The result is that the learning system will use this power to memorize the dataset "by hearth", which will reduce its capacity to generalize its learning to new examples outside of the training data.

[14] The probability distribution is represented by a function which comprises a parameter $\tau \in [0, \inf]$ called the *temperature* which is typically set to 1. Raising the temperature spreads out the probabilities, making them more uniform, while lowering it makes the distribution peakier, hence making the agent even more greedy to choose the letter with highest probability.

[15] Notice that these appear randomly. For example, "chapter xi" might appear before "chapter iii".

agent. Following this framework, I argue that the reader who flips through the pages of *for the sleepers in that quiet earth.* observes a second-order behavior (or metabehavior); that is, a behavior that change through time according to another behavior. This kind of behavior is characteristic of self-organizing systems such as neural networks and genetic algorithms.



(a) First-order behaviors: Rule-based systems react to the world in a way that is more or less stable through time.

(b) Second-order behaviors: Adaptive and evolutionary agents modify their behavior in reaction to observations.

Fig. 1: Representation of behaviors in nonadaptive vs adaptive agents. This figure compares the response of formal systems with that of adaptive and evolutionary systems. The relationship between inputs (i.e., what the agent observes as well as its current state/memory) and outputs (ie., the actions taken by the agent) of a formal system is temporally invariant, whereas adaptive and evolutionary systems allow it to change through time.

Under which conditions does the experience of perceiving a past behavior through the trace left by an agent can be considered an experience of behavior aesthetics in itself? Here it is implied that the time in which the artificial agent evolves is distinct from human time, and that the agent lives in a very different world than the reader's. Whereas human observers have already acquired a sense of words and sequences – and attribute meaning to such sequences – the world in which the LSTM agent evolves is constituted of chains of arbitrary symbols which the system discovers sequentially. By reading the book we are thus seeing an outlandish form of acquiring language. There is therefore a disruption between these two experiences. Yet, human viewers possess the observational capacity to identify language structures (morphemes, words, sentences) that speak to us, on which we project meaning. As in the tradition of many other agent-based artistic approaches such as artificial life and robotic art, the art object instigates a process of projection and/or evocation, where the viewer attributes some form of agency to the system. As I hope to have demonstrated in this paper through a close look at the generative text process, beyond the immediate appreciation of lexical components and indeterminate literary "gems", the reading of the book reveals the ghostly presence of an artificial entity.

for the sleepers in that quiet earth. thus challenges concepts of "algorithmic art" and "behavior aesthetics". In effect, it presents itself as a stabilized object containing the trace of a computational system, and in this sense is akin to other computer-based generative artworks such as computational poetry, algorithmic music, and generative painting. While the work clearly differs from more "traditional" forms of agent-based art and behavior aesthetics such as works of artificial life and robotic art, since what is given to view to the reader's eyes is the behavior of an agent, and since the work is primarily about that behavior, I argue that it differs from other such generative artworks that attempt to produce stabilized media form using a process; in that this work's core subject and materiality is precisely the behavior of the system that generates it.

Conclusion

for the sleepers in that quiet earth. is a hybrid project that spans diverse approaches such as electronic literature, generative art, and behavior aesthetics. It uses a deep learning recurrent neural network, 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 in an effort to reveal the learning process of the system. In other words, it subverts the purpose of artificial intelligence, which aims at reproducing or exceeding human performance – in this case, by imitating the style of a well-known English author – and instead brings the focus back to the behavior of the artificial agent as it tentatively tries to achieve its goals.

As thus, rather than focusing on the kind of literary prowesses such computational systems are able to do, *for the sleepers* offers the reader a unique insight into the inner workings of a machine learning algorithm by turning the experience of reading into an encounter with a behavior. The hybridity of the artwork asks questions about its ontological position, emphasizing distinctions between generative art and behavior aesthetics. While the work is certainly different in many respects from works traditionally understood as agent-based artworks such as those employing situated robotic systems, it shares with them a unique focus on the behavior of an agent – in this case, by allowing the observer to experience the phantom journey of an artificial machine learning agent.

Acknowledgments

I would like to thank the Fonds de Recherche du Québec - Société et Culture, the Massachusetts Institute of Technology, The Trope Tank, Bad Quattro, and Dr Nick Montfort for their support.

References

1. Audry, S. 2016. Aesthetics of Adaptive Behaviors in Agent-based Art. In *Proceedings of the 22nd International Symposium on Electronic Art*.
2. Brontë, E. 1996. *Wuthering Heights*.
3. Hochreiter, S., and Schmidhuber, J. 1997. Long Short-Term Memory. *Neural Computation* 9(8):1735–1780.
4. Manning, C. D., and Schütze, H. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, Mass: The MIT Press, 1 edition edition.
5. Pask, G. 1968. *An Approach to Cybernetics*. London: Hutchinson.
6. Penny, S. 2000. Agents as Artworks and Agent Design as Artistic Practice. In Dautenhahn, K., ed., *Advances in Consciousness Research*, volume 19. Amsterdam: John Benjamins Publishing Company. 395–414.
7. Senécal, J.-S. 2016. *Machines That Learn: Aesthetics of Adaptive Behaviors in Agent-Based Art*. Phd, Concordia University.