

Generating Sequences with Recurrent Neural Networks

Alex Graves

Presenter: Ryan Wolter

Goal

“to demonstrate that LSTMs can use memory to generate complex realistic sequences containing long-range structure”

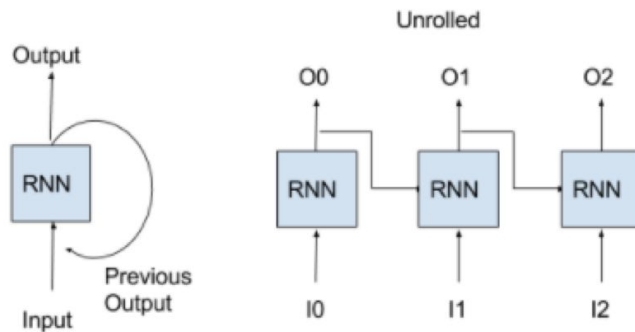
```
<page>
  <title>Italo-II article</title>
  <id>14848</id>
  <revision>
    <id>15912447</id>
    <timestamp>2004-11-19T19:47:39Z</timestamp>
    <contributor>
      <username>El andres</username>
      <id>5099991</id>
    </contributor>
    <minor />
    <comment>
      disambiguation from [[List of ISBN]] newsletters, see [[ISO site]]
    </comment>
    <text xml:space="preserve">#REDIRECT [[Quartic war]]
  </text>
</revision>
</page>
<?xml version="1.0" encoding="utf-8"?>
```

Main Contributions

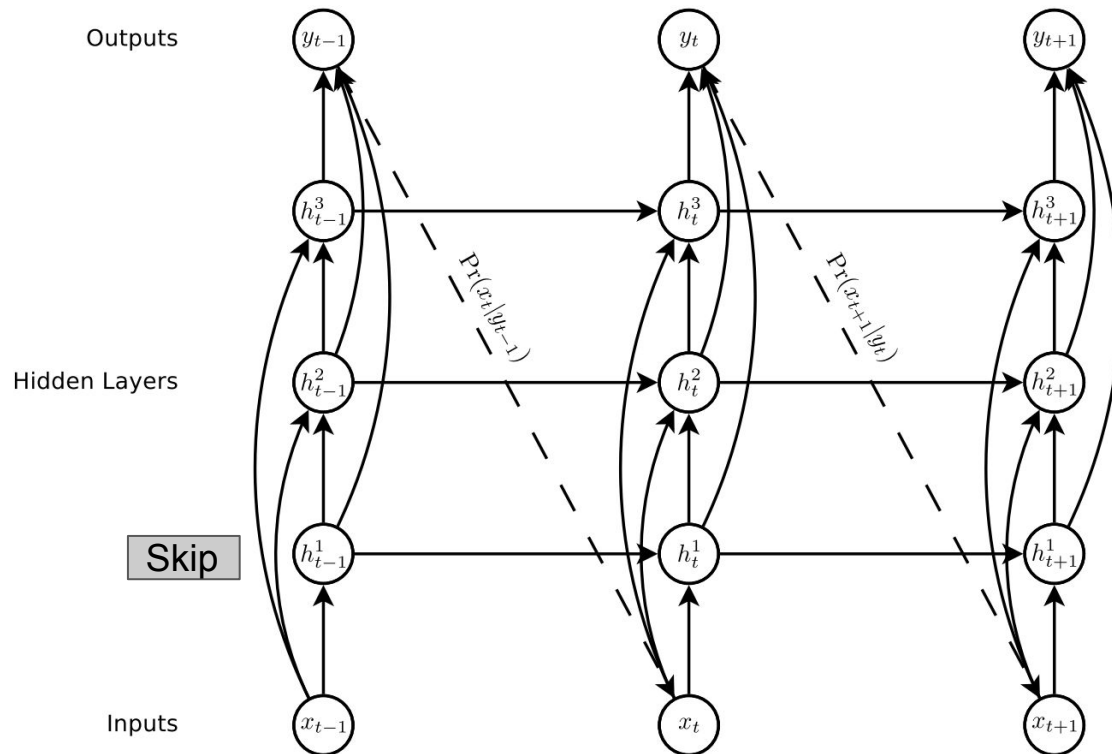
- Proved LSTMs could be used for sequence generation
 - Discrete text generation
 - Real-spaced handwriting generation

Intuition

1. Build a predictive model
2. Condition next prediction on all prior predictions
3. Then treat each prediction as if it were real, and repeat the cycle.

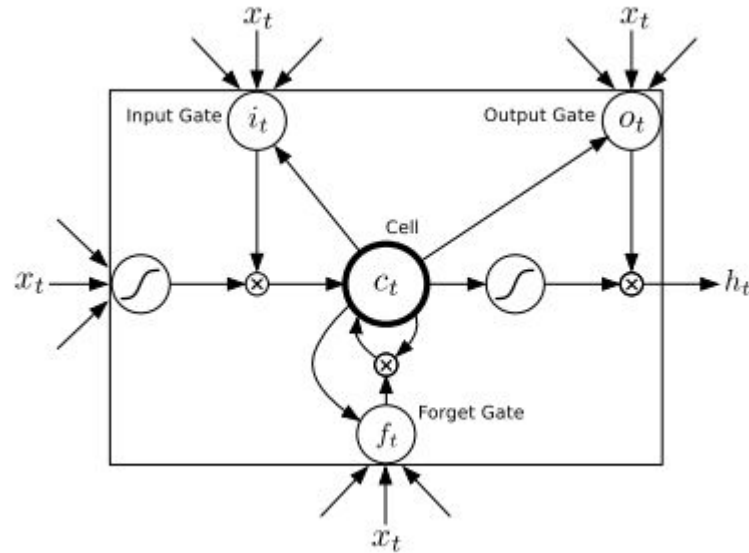


Network Architecture



LSTM: An Enhancement

- Store a sort of compressed memory of many different prior steps
- Less prone to unrecoverable error
- Remembers longer range patterns



Text Generation

- Wikipedia data - 100M bytes
- Demonstrate long-range memory
 - Balanced parentheses, quotes
 - Nested XML tags with indentation
- Model of size 700 hidden units, 7 layers deep, 21.3M weights
- LSTM cells reset only every 100 sequences, training data is not shuffled, resulting in about 10k characters memory length.
- Word-level prediction vs character-level prediction



Text Generation

```
<page>
  <title>Italo-II article</title>
  <id>14848</id>
  <revision>
    <id>15912447</id>
    <timestamp>2004-11-19T19:47:39Z</timestamp>
    <contributor>
      <username>El andres</username>
      <id>509991</id>
    </contributor>
    <minor />
    <comment>
      disambiguation from [[List of ISBN]] newsletters, see [[ISO site]
    ]
    </comment>
    <text xml:space="preserve">#REDIRECT [[Quartic war]]
  </text>
</revision>
</page>
<?xml version="1.0" encoding="utf-8"?>
```


Handwriting Generation

- IAM Online handwriting database (around 10k lines of handwritten text)
- Mixture density outputs
- Network inputs are x_1 , x_2 , real-valued pair, x_3 is binary for stroke end.
- Model of size 400×3 and 900×1 (both overfit)

would find the bus safe and sound
As for Mark, unless it were a
course at the age of fifty-five

Handwriting Generation

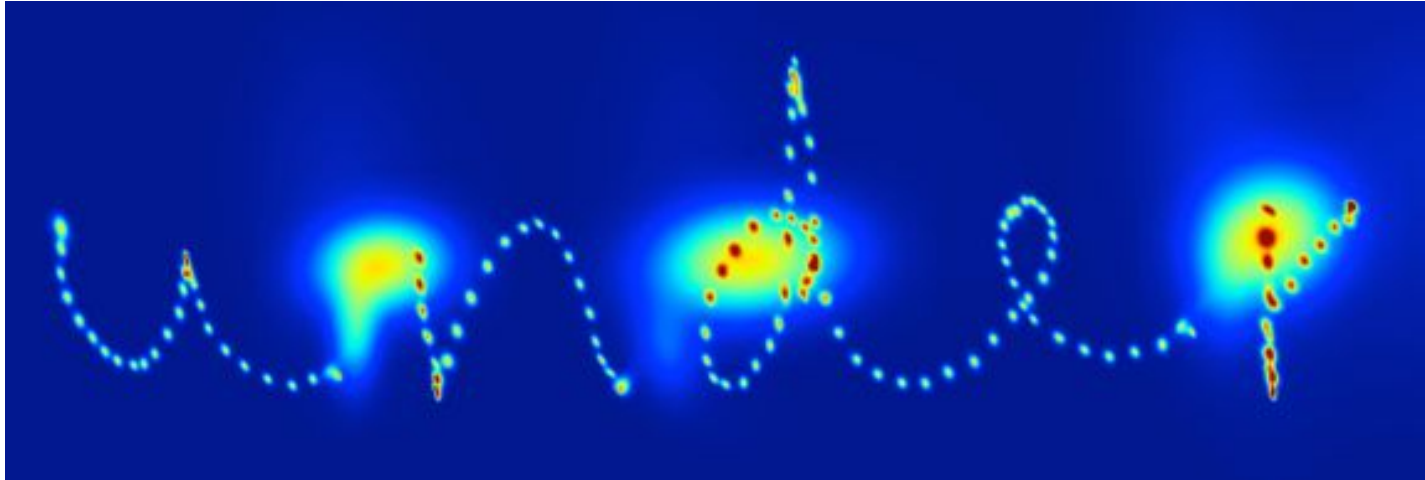
When my children grow up I hope they will

be 'very much as they are.' I hope they will be

the same gentle of high credit.

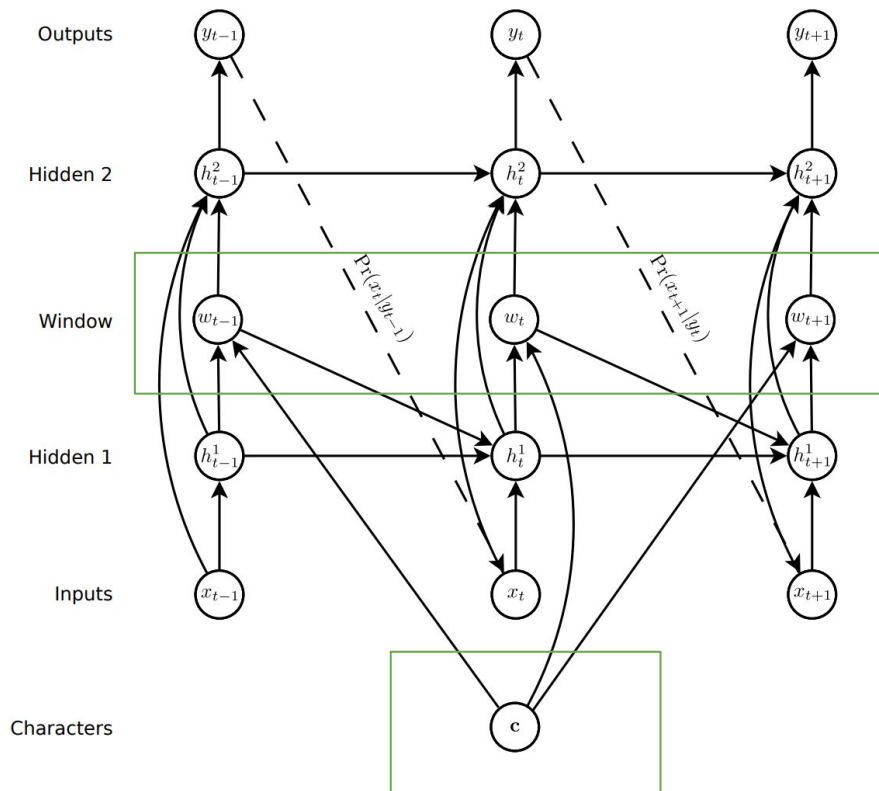
See Boulogne. The architecture is

Handwriting Generation: Visualization



Handwriting Synthesis

- Added character sequence to input, window layer.
 - Window layer: helps the network understand which character it is currently writing (Attention)
- Biased Sampling
 - Choose higher probability predictions to improve handwriting (recalculate StdDev and mixture weights)



Handwriting Synthesis - Bias

- 0 when the samples are biased
- 0.1 towards more probable sequences
- 0.5 they get easier to read
- 2 but less diverse
- 5 until they all look
- 10 exactly the same
- 10 exactly the same
- 10 exactly the same

Handwriting Synthesis - Priming

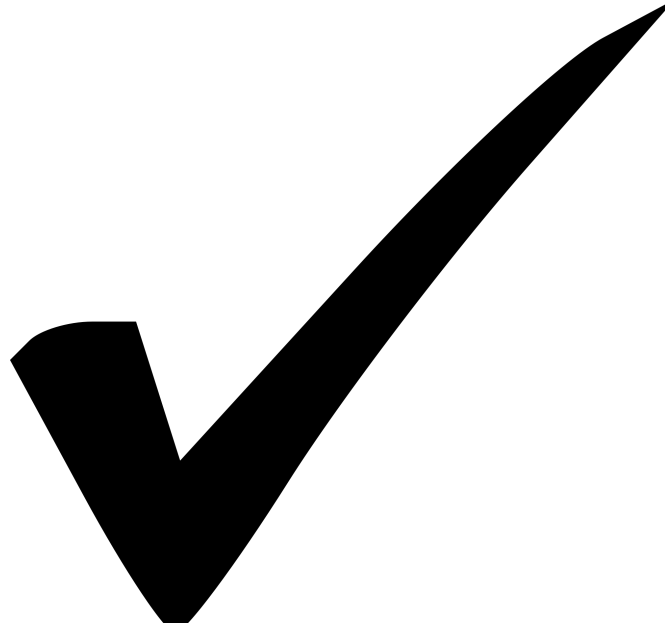
Take the breath away when they are

when the network is primed
with a real sequence

She looked closely as she

when the network is primed
with a real sequence

PROs



Goal

“to demonstrate that LSTMs can use memory to generate complex realistic sequences containing long-range structure”

Notice the balanced tags?

```
<page>
  <title>Italo-II article</title>
  <id>14848</id>
  <revision>
    <id>15912447</id>
    <timestamp>2004-11-19T19:47:39Z</timestamp>
    <contributor>
      <username>El andres</username>
      <id>509991</id>
    </contributor>
    <minor />
    <comment>
      disambiguation from [[List of ISBN]] newsletters, see [[ISO site]]
    </comment>
    <text xml:space="preserve">#REDIRECT [[Quartic war]]
  </text>
</revision>
</page>
<?xml version="1.0" encoding="utf-8"?>
```


Main Contributions - Goal met

- Proved LSTMs could be used for sequence generation
 - Discrete text generation
 - Real-spaced handwriting generation

Handwriting Synthesis - Results

' Take the breath away when they are

when the network is primed
with a real sequence

She looked closely as she

when the network is primed
with a real sequence

Thoroughness

- Network is described in detail
- Training parameters are stated clearly
- Ample examples of real data, generated data, with success and pitfalls
- Proved long-term memory of LSTMs on both continuous and discrete datasets
- Real vs Discrete - Both still language modeling
- Code (and demo) provided: <https://github.com/szcom/rnnlib>

Critiques

Limited Training Data

Lack of symbols, only 57 unique labeled characters

‘Each line was treated as a separate sequence (meaning that possible dependencies between successive lines were ignored)’ (pg 18, paragraph 3)

Limited precision, ~ 25 points per character & ~ 700 points per line

Handwriting of Wikipedia data?

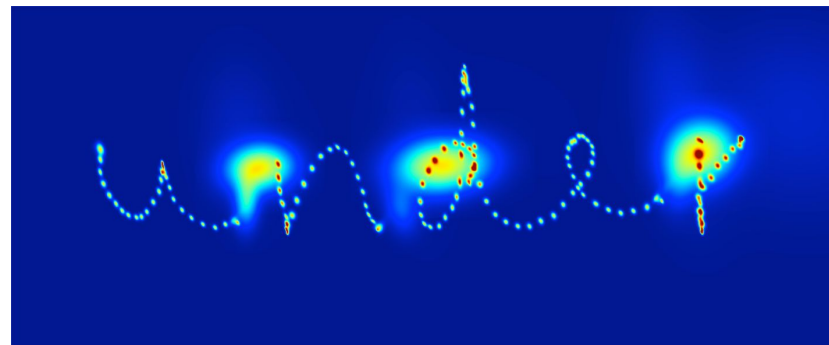


fig. 10

RNN has to be Primed

'Lastly, the network's adaptation to recent sequences during training (which allows it to benefit from dynamic evaluation) can be clearly observed in the extract' (pg 11, paragraph 2)

prison welfare Officer complement

when the network is primed
with a real sentence
the samples mimic
the writer's style

Can't Create New Handwriting Style

Most recently processed style dominates

Take the breath away when they are

when the network is primed
with a real sequence

the samples mimic

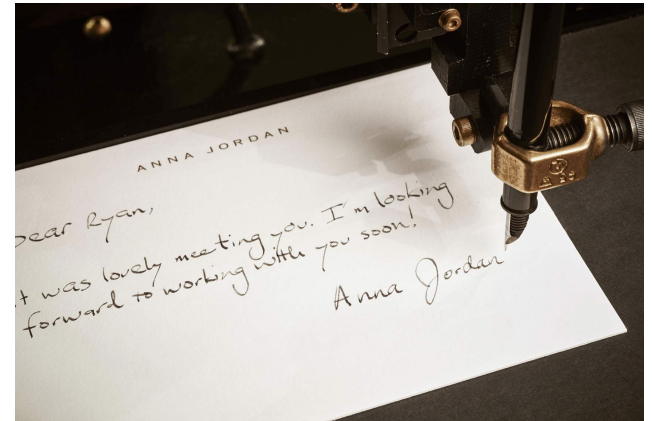
the writer's style

Real World Value

Robot mimics handwriting for personal touch

'A possible application would be artificial enhancement of poor handwriting' (pg 34, paragraph 3)

Are robot-written notes handwritten?



Illegible Generated Data

Learns structure such as word spacing, but still artifacts of illegibility

'bryoes f eald runine fs wine cureo' ?

'As with all text generated by language models, the sample does not make sense beyond the level of short phrases' (pg 10, paragraph 6)

Character window & biasing

A handwritten sample of the text 'bryoes f eald runine fs wine cureo' in a cursive script. The characters are connected and somewhat stylized, illustrating the 'illegible' nature of the generated data mentioned in the text above. The text is written in black ink on a white background.

fig. 11

Split on Discrete & Real Valued Data

Not much comparison between discrete & real data

Contrast character window with discrete characters

Contrast 25 inputs per character for real versus 1 input per character for discrete,
& burden on LSTM

Both form of language with real having control to create new symbols

The image shows the phrase "until they all look" written in a cursive, handwritten style. The letters are black on a white background. The word "until" is on the left, "they" is in the middle, and "all look" is on the right. The handwriting is fluid and somewhat informal.

fig. 16

No Comparison Metric for Handwriting Synthesis

'No benchmark results exist and the main goal was to generate convincing-looking handwriting' (pg 18, paragraph 3)

Excessive Display of Generated Data

Six pages of Wikipedia data without much analyzation

Six pages of handwriting data that splits a two paragraph conclusion

Could put in appendix