Gesture Recognition with Recurrent Neural Networks

Justin Fu
University of California, Berkeley
justinfu@berkeley.edu

Siddhartho Bhattacharya
University of California, Berkeley
siddartho_b@hotmail.com

Abstract

Gesture recognition is typically done through images, but it requires extensive effort to process image data into a form useful for gesture recognition. Our project uses accelerometer data from smartphones, so no special hardware is required. Recurrent neural networks provide a natural way to model time sequences, and the LSTM architecture in particular provides a way to model long-range dependencies that are prevalent in the gestures we tested. We imlemented X gestures, and achieve an accuracy of X percent on the test set we generated.

1 Introduction

Intro goes here.

2 Data

We collected accelerometer data streamed from a smartphone.

3 Model

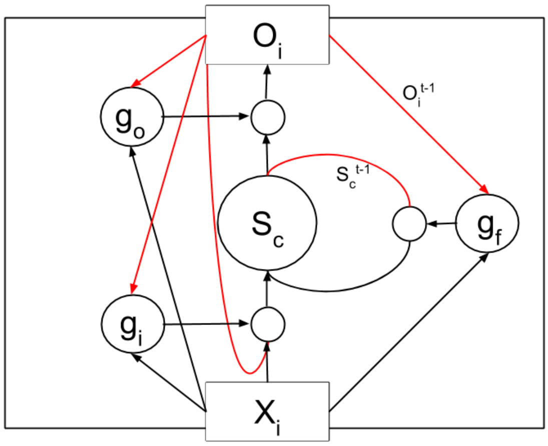
We implemented an LSTM (long short term memory) network based on the architectures proposed by Graves [1]. Specifically, our network contains the forget, input, and output gates, but lacks the peephole connections that connect the internal cell state to these gates (as done in Vinyals et al. [2]).

We then interleave layers of LSTM units with normal network layers to create a multilayer network.

3.1 Equations

We used squared error as our objective function. O is the network output and Y is the teacher value. n sums over each training example, t sums over time, and i sums over the dimensions of each training example at a given timestep. Note that t varies with each training example.

Figure 1: A single LSTM cell.



$$L(Y,O) = \sum_{n}^{N} \sum_{t}^{len(O^n)} \sum_{i}^{I} \frac{1}{2}(o_i^{n,t} - y_i^{n,t})^2$$

Most of the equations on the forward and backward passes are the same as those in Graves [1], but some were missing so we specify all for completeness.

3.2 Forward Dynamics

The forward dynamics of the network are specified as follows. Equations are specified for a whole layer (rather than per unit). Bolded values represent vectors. Note that $*$ represents a point-wise multiply.

Gates

$$\mathbf{a}_i = W_{i,x}\mathbf{x}^t + W_{i,o}\mathbf{o}^{t-1}$$

$$\mathbf{g}_i = \sigma(\mathbf{a}_i)$$

$$\mathbf{a}_f = W_{f,x}\mathbf{x}^t + W_{f,o}\mathbf{o}^{t-1}$$

$$\mathbf{g}_f = \sigma(\mathbf{a}_f)$$

$$\mathbf{a}_o = W_{o,x}\mathbf{x}^t + W_{o,o}\mathbf{o}^{t-1}$$

$$\mathbf{g}_o = \sigma(\mathbf{a}_o)$$

Cell State

$$\mathbf{a}_c = W_{c,x}\mathbf{x}^t + W_{c,o}\mathbf{o}^{t-1}$$

$$\mathbf{s}_c = \mathbf{g}_i * \sigma(\mathbf{a}_c) + \mathbf{g}_f * \mathbf{s}_c^{t-1}$$

Cell Output (Hidden State)

$$\mathbf{o} = \mathbf{g}_i * \tanh(\mathbf{s}_c)$$

Normal Layer Output

$$\mathbf{a}_n = W_n\mathbf{o}$$

$$\mathbf{n} = \sigma(\mathbf{a}_n)$$

3.3 Backpropagation

δ_k^{L+1} represents the error backpropogating from the above layer. As with before, $*$ represents a pointwise multiplication.

Cell Output

$$\delta_n = \delta_k^{L+1} * \tanh'(\mathbf{a}_n)$$

$$\delta_h = W_{i,h}^T\delta_i^{t+1} + W_{f,h}^T\delta_f^{t+1} + W_{o,h}^T\delta_o^{t+1} + W_{c,h}^T\delta_s^{t+1}$$

$$\frac{\partial L}{\partial \mathbf{o}} = W_n^T\delta_n + \delta_h$$

Gates

$$\delta_o = \sigma'(\mathbf{a}_o) * \tanh(\mathbf{s}_c) * \frac{\partial L}{\partial \mathbf{o}}$$

BLAH

4 Training

We trained our network using BPTT (backpropogation through time) and gradient descent with momentum.

5 Decoding

6 Example LaTeXstuff

- Item Example 1
- Item Example 2

Noindent

6.1 subsection

```
\usepackage{times}
\usepackage{latexsym}
```

"(Graves, 2008) Quote"

Column	**Col**	**Col**
row blah	blah	blah
row blah	blah	blah

Table 1: Example table!

6.2 Sections

Citations: Citations within the text appear in parentheses as (**?**) or, if the author's name appears in the text itself, as Gusfield (**?**). Append lowercase letters to the year in cases of ambiguity. Treat double authors as in (**?**), but write as in (**?**) when more than two authors are involved. Collapse multiple citations as in (**?**; **?**). Also refrain from using full citations as sentence constituents. We suggest that instead of

you use

"Gusfield (**?**) showed that ..."

If you are using the provided LaTeX and BibTeX style files, you can use the command `\newcite` to get "author (year)" citations.

As reviewing will be double-blind, the submitted version of the papers should not include the authors' names and affiliations. Furthermore, self-references that reveal the author's identity, e.g.,

"We previously showed (**?**) ..."

should be avoided. Instead, use citations such as

"Gusfield (**?**) previously showed ... "

6.3 Footnotes

Footnotes: Put footnotes at the bottom of the page and use 9 points text. They may be numbered or referred to by asterisks or other symbols.[1] Footnotes should be separated from the text by a line.[2]

Acknowledgments

Brian is our savior.

References

[1] Alex Graves. 2008. *Supervised Sequence Labelling with Recurrent Neural Networks*. PhD Thesis.

[2] Orial Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan. 2014. *Show and Tell: A Neural Image Caption Generator*.

[1]This is how a footnote should appear.

[2]Note the line separating the footnotes from the text.