Log Book for Robopoet (aka C.R. Maxwell)

4/2/16

      Created large poetry corpus of
-50000 lines of the Norton Anthology of Poetry
-Wordsworth
-Shelley
-Tennyson
-Rossetti
-Dickinson
-Blake
-Milton
-Frost
-Keats
-Yeats
-Coleridge

      Total is 1.5 million words, 225000 lines.

      Punctuation was retained. The end of every poem was marked ENDPOEM.

      The text files for individual poets came from poemhunter as pdfs, then were converted to text with extraneous markings removed. I estimate less than 1% corruption of the input. It's called cleanpoems.txt.

      No changing of case was done.

The best model trained so far was

th train.lua -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 10000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 2 -batch_size 50 -seq_length 75 -learning_rate 0.8e-3 -max_epochs 10 -dropout 0.5 -seed 77

th train.lua -init_from cv/${model} -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 10000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 2 -batch_size 50 -seq_length 75 -learning_rate 0.8e-3 -max_epochs 3 -dropout 0.5 -seed 77

th train.lua -init_from cv/${model} -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 10000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 2 -batch_size 50 -seq_length 75 -learning_rate 0.8e-3 -max_epochs 10 -dropout 0.5 -seed 87

th train.lua -init_from cv/${model} -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 10000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 2 -batch_size 50 -seq_length 75 -learning_rate 0.8e-3 -max_epochs 10 -dropout 0.5 -seed 97

(The seeds ensure mixing of the training sets. I'm actually not certain that's a good idea because it also mixes the training into the test. But I want to keep mixing up the training without changing the test.)

I continued training thereafter; the test sets got slowly worse as the training sets got better, so it becomes optimistic. The best test loss (lm_lstm_epoch10.00_4.7658.t7) was 4.77 but the training losses eventaully went as low as 4.2.

With a temperature of 0.8, the poems are good and metrical and somewhat sensible, but still a little incoherent and occasionally ungrammatical.

When sampled with temperature of 0.5, the following came out:

numerous days
 and the sweet world may be a maiden ,
 And the night , and the moon , and the stars
 Of the earth , in the higher night ,
 Is the voice of the world , and the wind
 Of the sky , and the stars of the moon ,
 The light of the sky , the air of the earth ,
 The mountain - - and the air
 Of the sun - green waters of the air ,
 And those that know the truth ,
 And the new - born world , that they have seen
 In the dark woods of the sea ,
 With the dead - born winds to pray
 To the storm , and the winds ,
 And the stars - - the music goes
 To the sky and the sky .
 ENDPOEM

 The Lady of a New Lady
 I thought that I am dead ,
 When I had heard a day
 That I had been a man ,
 And heard a thing , a man ' s true voice ,
 And then I was a thing .
 ENDPOEM

 A Boy
 I
 I am a man ,
 I am but a thing ,
 And I ' m not a boy ,
 But I was not the old man
 I had not a man of the house .
 ENDPOEM

 The World
 I
 I
 O .

The lot of the aged man
Is not to be built ,
And the upper sword ,
The old man ,
That the man and the old
The man ' s dust ,
And the old man
And the old man .
ENDPOEM

The Song
I
There was a bird
Under the sun of the sea ,
And the pale eyes
Of the heart that is gone ,
And the birds were the sea
And the wild moon ,
And the winds of the moon
Made the air too .
ENDPOEM

The Grave
The man is all of all ,
And the man of a man ,
And the moon and the moon ,
And the old hour came
That the old man gave ,
And his heart was dead ,
And he had heard her place ,
And the heart had been so long .
It was not a thing to do
And the wind is old ,
But she stood to her feet
And in a golden dream
She had been the heart of life .
ENDPOEM

These are not highly atypical selections. These are 6 of the 16 poems it generated. This quality is typical.

I consider these generally superior to anything I would write. Not Yeats, but not embarrassing.

Next Directions
Better mixing and aggregation
      -Aggregation: only considers 50 words at a time
      -Mixing: I want to re-mix the training batches every so often, but hold out the same test batch
Longer sequences
Stress encoding

Rhyme/syllable encoding
       -Should probably remove Milton
Larger training sets
"In the style of"
Function of temperature
50-dimension word vectors
200-dimension word vectors
Stemming and casing
GRU

1815 4/1/16
Lowercased the bigpoems file
mydir=data/lowerpoems
thold=2
epochs=10

echo "Starting 50/75"
th train.lua -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 10000 -print_e
very 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 2 -batch_size 50 -
seq_length 75 -learning_rate 0.8e-3 -max_epochs 3 -dropout 0.5 -seed 77

echo "Starting 50/75 ii"
model=`ls -rt cv | tail -1`
th train.lua -init_from cv/${model} -gpuid 0 -num_layers 3 -rnn_size 256 -eval_v
al_every 10000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -thres
hold 2 -batch_size 50 -seq_length 75 -learning_rate 1e-3 -max_epochs ${epochs} -
dropout 0.5 -seed 1007

echo "Starting 50/75 iii"
model=`ls -rt cv | tail -1`
th train.lua -init_from cv/${model} -gpuid 0 -num_layers 3 -rnn_size 256 -eval_v
al_every 10000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -thres
hold 2 -batch_size 50 -seq_length 75 -learning_rate 1e-3 -max_epochs ${epochs} -
dropout 0.5 -seed 1017

echo "Starting 50/75 iv"
model=`ls -rt cv | tail -1`
th train.lua -init_from cv/${model} -gpuid 0 -num_layers 3 -rnn_size 256 -eval_v
al_every 10000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -thres
hold 2 -batch_size 50 -seq_length 75 -learning_rate 1e-3 -max_epochs ${epochs} -
dropout 0.5 -seed 1027

Led to
lm_lstm_epoch10.00_4.6846.t7
with excellent optimism (nearly none)

Temperature 0.5:

 the child

i
i am a place ,
and i am little to the land ,
and i shall hear my way ,
and i have heard my heart and me ,
and i have loved you ,
and i am all the world and all ,
and i may give you more .
ENDPOEM

After 10 more epochs, th sample.lua cv/lm_lstm_epoch10.00_4.7013.t7 -temperature 0.6

recollecting
if it were given us to be done
to the river , in the burthen
of the that place of the memory
of some single - a friend ,
at least there was an end
of the world , and with its own
the less of his own eyes .
ENDPOEM

the daffodil
the moon is close , the earth is green ,
the south - field is a rose ,
the the moon ' s edge of winter
sees the snow - grey cloud ,
and the clouds are like the stars ,
and the green water
in the morning
with the big flowers ,
with the light of the sun ,
the wisdom of the earth ,
and the white air ,
the long ocean of the sun ,
that is not to hear
the sun - white line .
the sun is on the sun ,
and the garden are gone .
ENDPOEM

Two Structural Questions
Can I shuffle training batchs every epoch?
How to extend the glove embedding? (Related question: how are new tokens embedded?)

Notes from Reading

train.lua
122

local loader = CharSplitLMMinibatchLoader.create(opt.data_dir, opt.batch_size, opt.seq_length, split_sizes, opt.word_level == 1, opt.threshold)

245
function eval_split(split_index, max_batches)

256
local x, y = loader:next_batch(split_index)

charsplitlmminibatchloader.lua

43
CharSplitLMMinibatchLoader.text_to_tensor(word_level, threshold, input_file, vocab_file, tensor_file)

69-75:

local ydata = data:clone()
ydata:sub(1,-2):copy(data:sub(2,-1))
ydata[-1] = data[1]
self.x_batches = data:view(batch_size, -1):split(seq_length, 2) -- #rows = #batches
self.nbatches = #self.x_batches
self.y_batches = ydata:view(batch_size, -1):split(seq_length, 2) -- #rows = #batches
assert(#self.x_batches == #self.y_batches)

Interesting question: it seems like it should be seq_length+1? Or is that the sub/clone monkey business? Even so it seems not right.

[result] split([result,] tensor, size, [dim])

Splits Tensor `tensor` along dimension `dim` into a `result` table of Tensors of size `size` (a number) or less (in the case of the last Tensor). The sizes of the non-`dim` dimensions remain unchanged. Internally, a series of [narrows](#) are performed along dimensions `dim`. Argument `dim` defaults to 1.

If `result` is not passed, then a new table is returned, otherwise it is emptied and reused.

Example:

```
x = torch.randn(3,4,5)

> x:split(2,1)
{
  1 : DoubleTensor - size: 2x4x5
  2 : DoubleTensor - size: 1x4x5
}

> x:split(3,2)
```

```
{
  1 : DoubleTensor - size: 3x3x5
  2 : DoubleTensor - size: 3x1x5
}

> x:split(2,3)
{
  1 : DoubleTensor - size: 3x4x2
  2 : DoubleTensor - size: 3x4x2
  3 : DoubleTensor - size: 3x4x1
}
```

Conclusion (!)
(!)
Torch does not randomly mix up the batches!
It's completely sequential. Always the same order.
It's in the order in the input file.
There's no easy way to randomize this. Tensor:split is used in a single line, 69. Enormous
complications would ensue to reorder that.

Note on data

The data is divided into a number of Batches.
Each Batch is batch_size*seq_length long (as measured in tokens).
nBatch*batch_size*seq_length is the number of used tokens. It is nominally 95% of the datafile tokens
but can be slightly less due to cutting off the end to get an integer number of chunks that are each
Batchsize*seqlength long.

Glove
Unrelated question: how are new tokens handled?

Glovemebedding.lua
53-55
 for word, idx in pairs(word2idx) do
word_lower2idx[word:lower()] = idx
 end

It appears that words are effectively lowercased.
This slightly contradicts the observation that there is different vocab size for the lowercased document
(20800 vs 24800)

word2idx is passed in as vocab in line 163 of train.lua, and ultimately traces to back to
CharSplitLMMinbatchloader

163
embedding = GloVeEmbedding(vocab, 100, opt.data_dir) --GloVeEmbeddingFixed(vocab, 200,
opt.data_dir)

Line 72

```
for word, idx in pairs(word2idx) do
if not loaded[word:lower()] then
print("Not loaded: " .. word:lower())
for i=1, file_embedding_size do
weight[idx][i] = torch.normal(0, 0.9) --better way to do this?
end
end
end
```

Words not in the vocabulary are set to a random vector, apparently with rms 0.9.
I'd probably set them to the nearest word by edit distance.

This also shows us how to add dimensions. We'll have to break the file_embedding_size into two pieces, a glove piece and an additional piece. Then modify weight[idx][glovedim+i] for I in the additional piece.

GUTENBERG

http://web.eecs.umich.edu/~lahiri/gutenberg_dataset.html

```
@ARTICLE{2013arXiv1310.5111L,
   author = {Lahiri, Shibamouli},
    title = {{Complexity of Word Collocation Networks: A Preliminary Structural
Analysis}},
  journal = {ArXiv e-prints},
archivePrefix = "arXiv",
   eprint = {1310.5111},
 keywords = {Computer Science - Social and Information Networks, Physics - Physics
and Society},
     year = 2013,
    month = oct,
   adsurl = {http://adsabs.harvard.edu/abs/2013arXiv1310.5111L},
  adsnote = {Provided by the SAO/NASA Astrophysics Data System}
}
```

cat txt/*F* > giantF.txt
34M words

Plan: train a net of the same size and see what comes out.
Using threshhold of 5.
If this works well, append the poems 3 times (so that anytthing that appears 2 times in poems appears 5 times in this set) and rerun.

Use the resulting model as a starting point to train only on poems for a while. (Use the touch trick if needed.)

echo "Starting 50/75"
th train.lua -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 1000 -print_ev

ery 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 5 -batch_size 50 -s
eq_length 75 -learning_rate 2.0e-3 -max_epochs 3 -dropout 0.5 -seed 77

Vocab was 85000 words

Threshhold 5 runs out of memory, set to 15
(will have to include 8 copies of the poems…)
Vocab is 45000 words, still dies out of memory

Threshhold to 35
Vocab 31494 , memory is OK
Need 18 copies of poem!

After 3 epochs it's very slowly still improving, gets to cv/lm_lstm_epoch3.00_4.4565.t7 almost no
optimism.

Model has 12.6M parameters (cf 8M for the poetry) and 34M words.

Some output:

" We would marry the thing you don ' t keen . "

" Yes . "

*" To be sure all a few times ago ; and here I don ' t think there was a
very good room , and now I am going to do . I thought it would be a
matter that might be so much . I will have it possible to say that they
were doing anything . "*

*" Yes , " said Mervyn ; " that you shall be of a very little weak
head . "*

*" I wish you are to have a word to tell you what I was , " said old Jane .
" What is this ? "*

*" Why , " said he , " I think , as I think , you are the next of my innocent -
days . It ' s quite fault . There is something better than the " - - " and
" I will do you at my notions , " you said . " I hope you are certainly a
whole man . "*

Quesion: is 75 time steps overkill?

Next experiment: 50 time steps
        So it appears off-the-cuff that 75 converges faster, BUT 75 gets more backprops per cycle by
50%. So one ought really to compare 900 50's to 600 75.
75 loss @ 600: 5.1
50 loss @ 900: 5.1
50 loss @ 600: 5.3

So it appears in the early going that 50 is not so different from 75.
1048: still true at 3000 50 vs 2000 75 both at 4.8
1148: 50/10000: 4.6 75/6700: 4.6
10000 checkpoint for 75: 4.48
15000 checkpoint for 50: 4.63
1248 50/20000 (epoch 1.1): 4.63 (!)
1416 50/30000 (epoch 1.65) 4.70 and diverging (without optimism?!); 70 was 4.53 at same epoch, difference is certainly meaningful

If true, I'd drop down to 40 and double the width (for memory reasons.) I'd like to increase the batch size, but I'm not certain it would help much. The batches locally are all in the same style.

The more I think about the batching, the more I'm troubled. It makes perfect sense when you have a long single text, or many short unrelated texts. Having many long related texts, it's trouble. Basically, if correlations are strong on the scale of seq_length*batch_size words, then you aren't mixing very well. Here that's 2500 words.   In both giantF and cleanpoems, that's not large enough to move you out of a style.

You want sense correlation that is large (measured in word toekns) compared to seq_length and small compared to seq_elength*batch size.

Cleanpoems is easily fixed. Shuffle by ENDPOEM. Or maybe every 5th ENDPOEM.
giantF needs more work. Need to preserve on the level

Early on, differences between 50 and 75 are negligible. Around 1 epoch, they are barely noticeable but probably real. We'll see at 2.

50 seems to hit a wall already at the end of 1 epoch.

About giantF
3.06M nonblank lines
3.71M blank lines
34.3M words

To get about 400 words, take 40 lines

Conclusions after 20000 iterations of seq_length 50

50 hits a wall and has no optimism.

75 has no visible optimism.

Suggests I can widen model.

So lower batch size and length, increase width, let it bottom out.

Once it has bottomed out, then raise length and decrease batch size.

If I can mix well the batch size limitations might be only annoying.

MIXED giantF to bitsF.txt

Chose 40 lines, go until next cr/lf then write, and shuffle all those bits

Leave at 50 but lower dropout to 0.25. Training at 600 iterations is noiceably better (5.05 vs 5.3), still ahead at 900 also by about 0.25

lm_lstm_epoch0.27_4.4953.t7 is even better than 75 at epoch 0.27!

cv/lm_lstm_epoch0.55_4.4519.t7 is the best loss encountered so far.

I'll let it go through one epoch.
lm_lstm_epoch0.82_4.4388.t7
Best so far
cv/lm_lstm_epoch1.10_4.3923.t7
cv/lm_lstm_epoch1.37_4.4471.t7
cv/lm_lstm_epoch1.65_4.4653.t7
cv/lm_lstm_epoch1.92_4.4529.t7
cv/lm_lstm_epoch2.20_4.4590.t7

No noticeable optimism (if anything, pessimism)

Bouncing around at the 0.01 level. Empircally, it's hard to notice differences in output quality at that level.

Using th4 4.39:

" You may be able to remember that I do not know how to do it . "

" I have a kind of changing , " said the young man , " and I should have a
great deal of the most important and friendly little man ; but I
have not seen a man in her mind , in the same manner . "

" It is the same . "

" The great people , " said she , " I am not so much inclined to
give my blood to my own dear friend , who will be a fair person of
the world , and I hope you would be an old man . "

" The man is not a very good little man , " said the girl , " but I have
made those sprang on the street . "

" I have been a very large person , " said the boy . " It is all I have
been in a mind with the latter , and I am not not in the least as
much as I can think . It is a little more than I be , and I sleep ,
and I ' ll throw it back from you . I ought to have finished my best to
take a good deal of your life , and I feel so much as that , and I
shall go to a small City , and I will tell you that you have not
seen him in the way . "

" Yes , " said the man , " but I was not to be able to go to the pound .

" You see ; I believe that you are going to take it in a few days , "
said she , " and I am now heard of it . "

" Thank you , " said the girl . " We are in their way to the house ,
a few miles of the hill , and the boat will be as well as you ' ve been
.

Comment: the grammar is all formally correct, though the semantics are dicey. The managing of quotation marks with speakers is notable.

Conclusions:
-12.6M parameters will accurately learn grammar.
-Long-term sense is missing
-Short term sense is fleeting

Because even short-term sense is fleeting, I'm not inclined to muck around with long seq_lengths.

Next experiment:
        -seq_length=40
        -Double the depth (!)
                -Claim: for fixed number of parameters, depth buys you more than width
        -Retain full batch size if memory allows

th train.lua -gpuid 0 -num_layers 6 -rnn_size 256 -eval_val_every 5000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 35 -batch_size 50 -seq_length 40 -learning_rate 2.0e-3 -max_epochs 10 -dropout 0.25 -seed 77

Number of parameters in the model: 14.2 M
??? Only about 50% more parameters?
Embedding: 100 x 30k parameters: 3M parameters
Final Softmax: 256x30k parameters: 8M parameters
Each layer: 3x256x256: 0.2M per layer (I believe each RNN layer has 3 matrices)

This does give roughly the right answer. THE MODEL IS MOSTLY LEARNING MEANING AND SOFTMAX.

Which is idiotic, of course. It will never understand meaning better than Glove did, and it effectively has one layer.

Width dominates now.

Interestingly, the 6-layer model is VERY slow out of the gate.
        By 1100 iterations it looks a lot like the 50, maybe a little slower. So it's probably ok.

At 4800 iterations it's definitely lagging the 3-layer rnn, by about 0.15

cv/lm_lstm_epoch0.22_4.6770.t7
As expected (cf 4.49 for 50 / 3)
At 6300, it's about 4.55 so about 0.05 worse than the 50/3
cv/lm_lstm_epoch0.44_4.6102.t7
lm_lstm_epoch0.66_4.6307.t7

Not looking good for the home team

Going back to 75!
And batches of 40:

th train.lua -gpuid 0 -num_layers 6 -rnn_size 256 -eval_val_every 5000 -print_ev
ery 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 35 -batch_size 40 -
seq_length 75 -learning_rate 2.0e-3 -max_epochs 10 -dropout 0.25 -seed 77

lm_lstm_epoch0.33_4.5824.t7
lm_lstm_epoch0.66_4.5010.t7
lm_lstm_epoch0.99_4.4085.t7
lm_lstm_epoch1.32_4.4805.t7
lm_lstm_epoch1.65_4.4989.t7
lm_lstm_epoch1.98_4.4345.t7

Hmmm, dubious. I doubled the depth and it didn't help.

Let it run all day, see if it bottoms out here.

Next Directions, 4/4 0807 (strikethrough have been addressed, red are new, green are results)
~~Better mixing and aggregation~~
        -Aggregation: only considers 50 words at a time
        -Mixing: I want to re-mix the training batches every so often, but hold out the same test batch
        -Mixed batches, results improved
~~Longer sequences~~
        -Modest impact
Stress encoding
Rhyme/syllable encoding
        -Should probably remove Milton
~~Larger training sets~~
        -Modest score improvement but clear grammar learning even for complex grammar
        -Clean romantic without milton or frost or norton and some additional poets
"In the style of"
~~Function of temperature~~
        -Best results typically around 0.5-0.65. Up to 0.8 viable. Lower leads to a lot of repetitiveness,
higher tends to be too trippy.
50-dimension word vectors
200-dimension word vectors
~~Stemming and casing~~
        -Investigated casing, casing is accounted for, stemming is not
        -Stemming may be valuable for new words
GRU

Better new-word embedding
        -Currently new words are randomly initialized—should stem/overlap/edit distance to guess nearest word and initialize with that
        -There are both Levenshtein Distance and Porter Stemming implementations in lua
        -Should only update new words, not old ones
Independent word list
        -Should be able to independently provide a word list to append to current list.

4/4 1949

lm_lstm_epoch0.33_4.5824.t7
lm_lstm_epoch0.66_4.5010.t7
lm_lstm_epoch0.99_4.4085.t7
lm_lstm_epoch1.32_4.4805.t7
lm_lstm_epoch1.65_4.4989.t7
lm_lstm_epoch1.98_4.4345.t7
lm_lstm_epoch2.31_4.5056.t7
lm_lstm_epoch2.64_4.5140.t7
lm_lstm_epoch2.97_4.4440.t7
lm_lstm_epoch3.30_4.4930.t7
lm_lstm_epoch3.63_4.5263.t7
lm_lstm_epoch3.95_4.4472.t7
lm_lstm_epoch4.28_4.5007.t7
lm_lstm_epoch4.61_4.5186.t7
lm_lstm_epoch4.94_4.4467.t7

It's certainly not appearing to improve any more. 5 epochs after the best, it seems to limit cycle at best.

Optimism is near zero.

Called, time of death 1949

Better initial embeddings

Let's try this:
        -Baseline 6/256: 14.2M parameters
        -Raise dimension from 256 to 384, 6/384: 21.9M parameters
        -Raise from 6 layers of 256 to 16 layers, 16/256: 19.5M parameters
        -Raise from 6 layers of 384 to 16, 16/384: 33.7M parameters

        -530k parameters per network layer @ 256; 1.18M parameters per network layer@384
        -3.1M parameters for embeddings

        -6x384: 10.2M layer + embed; 11.7 M parameters for 384x31k softmax (cs calc 11.9M)
        -16x256: 11.6M layer + embed; 8.1M parameters for 256x31k softmax (vs. calc 7.9M)
        -6x256: 6.3M layer+embed; 7.9M parameters for 256x31k softmax (calc 7.9M)

Conclusion: understand parameter count.
Prediction: 3/256: 12.6M parameters

Doing carefully

Output Vocab	31494	31494	31494	31494	31494	31494	31494
Input Vocab	28000	28000	28000	28000	28000	28000	28000
Layers	6	16	6	16	3	3	14
Glove Dim	100	100	100	100	100	100	100
RNN Size	256	256	384	384	256	384	256
RNN Matrices	8	8	8	8	8	8	8
RNN Pars	3145728	8388608	7077888	18874368	1572864	3538944	7340032
Embed Pars	2800000	2800000	2800000	2800000	2800000	2800000	2800000
Softmax Pars	8062464	8062464	12093696	12093696	8062464	12093696	8062464
Total	14008192	19251072	21971584	33768064	12435328	18432640	18202496
Obs.	14.2	19.5	21.9	33.7	12.7		

This gets the parameter count right to about 1%. Note that input and output vocabulary differ because of the casing; they are tied together on input (same embedding) but not on output (separate softmax outputs in the one-hot.) I believe that the output vocabulary is the vocabulary number reported in the log.

Note that raising the width to 384 from the 3-layer baseline is the same as increasing the depth to 14!

ONLY these very deep architectures have parameter counts rivaling the final softmax!

16/384 has similar parameter count to the word count, with 25% dropout should be reasonable. Let's try that! But may only get seq_length of 40.

~2230
th train.lua -gpuid 0 -num_layers 16 -rnn_size 384 -eval_val_every 5000 -print_e
very 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 35 -batch_size 30
-seq_length 40 -learning_rate 2.0e-3 -max_epochs 10 -dropout 0.25 -seed 77

Interestingly, it seems to hit a wall IMMEDIATELY at loss of 6.4. I'm assuming this is because deep architectures need a lot of slow baking.

4/5/16 0716
cv/lm_lstm_epoch0.53_6.6400.t7

Too much slow baking.
saving checkpoint to cv/lm_lstm_epoch0.13_6.4010.t7
saving checkpoint to cv/lm_lstm_epoch0.26_6.4524.t7
saving checkpoint to cv/lm_lstm_epoch0.40_6.5551.t7

saving checkpoint to cv/lm_lstm_epoch0.53_6.6400.t7
saving checkpoint to cv/lm_lstm_epoch0.66_9.1730.t7
saving checkpoint to cv/lm_lstm_epoch0.79_10.0355.t7
saving checkpoint to cv/lm_lstm_epoch0.92_7.1595.t7
saving checkpoint to cv/lm_lstm_epoch1.05_13.3089.t7
saving checkpoint to cv/lm_lstm_epoch1.19_10.1773.t7
saving checkpoint to cv/lm_lstm_epoch1.32_7.6467.t7

Not looking good for the home team.

Called, time of death 1937.

Next!

512x3
Expected Pars: 25.3M
Actual Pars: 24.8M
Increase seq_length to 50

th train.lua -gpuid 0 -num_layers 3 -rnn_size 512 -eval_val_every 5000 -print_ev
ery 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 35 -batch_size 30 -
seq_length 50 -learning_rate 2.0e-3 -max_epochs 10 -dropout 0.25 -seed 77

There are 5.73 characers per word in the giantF data.

Entropy estimates per character run in the range of 0.6-1.4 bits per character.

Assuming a model of equiprobability, this means the average word has 3.5-8.02 bits per word, which
would correspond to an (equiprobable) equivalent log likelihood of 2.4-5.52 .

Which is to say, finding LL in the neighborhood of 4.4-4.6 without finding ways of significant
improvement may really mean we're up against the limits of English entropy.

If so, we hit those limits reasonably well with smallish models! Yay. I guess.

3/512:

-rw-r--r-- 1 andrew andrew 397449294 Apr 5 21:19 lm_lstm_epoch0.33_4.5755.t7
-rw-r--r-- 1 andrew andrew 397569318 Apr 5 22:06 lm_lstm_epoch0.49_4.5814.t7
-rw-r--r-- 1 andrew andrew 397689342 Apr 5 22:53 lm_lstm_epoch0.66_4.6040.t7

3/192

lm_lstm_epoch0.16_4.6737.t7
lm_lstm_epoch0.33_4.6942.t7
lm_lstm_epoch0.49_4.7193.t7
lm_lstm_epoch0.66_4.7541.t7
lm_lstm_epoch0.82_4.7900.t7
lm_lstm_epoch0.99_4.7248.t7

```
lm_lstm_epoch1.15_4.7791.t7
lm_lstm_epoch1.32_4.8162.t7
lm_lstm_epoch1.48_4.8259.t7
lm_lstm_epoch1.65_4.8384.t7
lm_lstm_epoch1.81_4.8801.t7
lm_lstm_epoch1.98_4.8389.t7
lm_lstm_epoch2.14_4.8364.t7
lm_lstm_epoch2.31_4.8958.t7
lm_lstm_epoch2.47_4.8919.t7
lm_lstm_epoch2.64_4.9011.t7
```

Diverging almost immediately.

The best so far was 256/3, which is surprising! It was simply the leftover of what was there.

Make sure I can recover 256/3:

```
echo "Starting 50/75"
th train.lua -gpuid 0 -num_layers 3 -rnn_size 192 -eval_val_every 5000 -print_ev
ery 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 35 -batch_size 30 -
seq_length 50 -learning_rate 2.0e-3 -max_epochs 10 -dropout 0.25 -seed 77
```

```
lm_lstm_epoch0.27_4.4884.t7
lm_lstm_epoch0.55_4.4506.t7
lm_lstm_epoch0.82_4.4400.t7
```
<mark>lm_lstm_epoch1.10_4.3851.t7</mark>
```
lm_lstm_epoch1.37_4.4453.t7
lm_lstm_epoch1.65_4.4653.t7
lm_lstm_epoch1.92_4.4598.t7
lm_lstm_epoch2.20_4.4527.t7
lm_lstm_epoch2.47_4.4987.t7
lm_lstm_epoch2.75_4.4950.t7
lm_lstm_epoch3.00_4.4093.t7
```

Starting to diverge.

Highlighted is lowest to date.

" It is a great thing to have brought away , " said Mrs .
Palliser , when she paused .

" The lock is on the ground , " said the little woman , and ,
armed with the man , made a turn over his arm , and the young man ,
as he had been in the eaten , broke out the favorite part of the
light , and sat down in a wide style , and our steps were relieved .

" Why should I give him a letter , " she said , " that , indeed ,
I had no chance to take the little lady to the study . He has
no need to be sure to stand , . "

*" Why , the Captain , " said he , " it ' s a sharp stranger , and I have
not known anything of the matter . "*

*" I suppose I expect you are to see a man of free , " he said . " And
you have never . "*

She looked at the door .

*" Yes , " said Mr . Elmore , " I am going to do it , and I ' m in the
way , and that ' s what you say of it . "*

*" Don ' t you , " said the girl , " and if you do not know what of
mine I ' ll tell you , Miss Effingham . "*

" I do not think it a very serious thing . "

" That will be a good deal . "

" I could not say it . "

" I suppose it was a great thing to you . "

*" I think it is just a little difficult to have any time about the
servants . "*

Excellent grammar, moderate sense.

Let's try two layers

th train.lua -gpuid 0 -num_layers 2 -rnn_size 256 -eval_val_every 5000 -print_ev
ery 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 35 -batch_size 50 -
seq_length 50 -learning_rate 2.0e-3 -max_epochs 3 -dropout 0.25 -seed 77

lm_lstm_epoch0.27_4.4472.t7
lm_lstm_epoch0.55_4.4188.t7
lm_lstm_epoch0.82_4.4292.t7
lm_lstm_epoch1.10_4.4004.t7

Let's decay this thing!

th train.lua -learning_rate_decay 0.94 -learning_rate_decay_after 0.75 -gpuid 0 -num_layers 3
-rnn_size 256 -eval_val_every 5000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir}
-threshold 20 -batch_size 50 -seq_length 50 -learning_rate 2.0e-3 -max_epochs ${epochs} -dropout
0.25 -seed 77

Also: lower the threshhold
Vocab size: 42426

If vocabulary were evenly distributed this would lead to a penalty of 0.27 to the minimum. So 4.67 should be considered a good place to be .

lm_lstm_epoch0.27_5.2284.t7
lm_lstm_epoch0.55_5.1866.t7
lm_lstm_epoch0.82_5.1916.t7
lm_lstm_epoch1.10_5.2194.t7
lm_lstm_epoch1.37_5.2873.t7
lm_lstm_epoch1.65_5.3193.t7
lm_lstm_epoch1.92_5.3573.t7
lm_lstm_epoch2.20_785.6955.t7

      Aborted for exploding loss

      The output at 5.2194 is gobbledygook.

lm_lstm_epoch0.27_5.2235.t7
lm_lstm_epoch0.55_5.1771.t7
lm_lstm_epoch0.82_5.1987.t7
lm_lstm_epoch1.10_5.2707.t7
lm_lstm_epoch1.37_5.3039.t7
lm_lstm_epoch1.65_5.3624.t7
lm_lstm_epoch1.92_5.3705.t7
lm_lstm_epoch2.20_5.3631.t7
lm_lstm_epoch2.47_152.3261.t7
lm_lstm_epoch2.75_142.7818.t7
lm_lstm_epoch3.02_491.6938.t7
lm_lstm_epoch3.30_73.6116.t7

saving checkpoint to cv/lm_lstm_epoch3.30_73.6116.t7
60000/182060 (epoch 3.296), train_loss = 5.60705689, grad/param norm = 9.5423e-04, time/batch = 0.4890s
60100/182060 (epoch 3.301), train_loss = 21.35040150, grad/param norm = 1.4427e-02, time/batch = 0.4832s
loss is exploding, aborting. (10.62 vs 4187.44)

      Aborted for exploding loss

andrew@LinuxMint ~/torch/word-rnn-adf-glove100 $ diff train_bits_256_3.sh train_bits_256_3_lowt.sh
6c6
< th train.lua -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 5000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 35 -batch_size 50 -seq_length 50 -learning_rate 2.0e-3 -max_epochs 3 -dropout 0.25 -seed 77

> th train.lua -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 5000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 20 -batch_size 50 -seq_length 50 -learning_rate 2.0e-3 -max_epochs ${epochs} -dropout 0.25 -seed 77

Only difference is word threshhold!
That leads to parameter count of 16.6M instead of 12.4M

Not at all obvious why it now blows up.

Go back to threshhold 35 and reduce learning rate 4x

th train.lua -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 5000 -print_ev
ery 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 35 -batch_size 50 -
seq_length 50 -learning_rate 5.0e-4 -max_epochs 3 -dropout 0.25 -seed 77

lm_lstm_epoch0.27_4.6686.t7
lm_lstm_epoch0.55_4.4967.t7
lm_lstm_epoch0.82_4.4125.t7
lm_lstm_epoch1.10_4.3317.t7
lm_lstm_epoch1.37_4.3375.t7

In early iterations, it learns about half as fast as 2E-3. This is already impressive and important—it's
not 1/4th as fast. Note that is has overtaken the learning rate 2.0E-3 by 0.82 epochs.

I think ideally, you'd be hitting your minima after 5 or so epochs. Suggests that another factor of 10 in
learning rate may not be out of the question.

Literary Aside: On regulariazation of RNNs

http://arxiv.org/pdf/1409.2329v4.pdf

the meaning of life is
that only if an end would be of the whole supplier. widespread
rules are re-
garded as the companies of refuses to deliver. in balance of t
he nation 's information and loan
growth associated with the carrier thrifts are in the proces
s of slowing the seed and commercial paper.

the meaning of life is
nearly in the first several months before the government was a
ddressing such a move as
president and chief executive of the nation past from a natio
nal commitment to curb grounds. meanwhile the
government invests overcapacity that criticism and in the o
uter reversal of small-town america.
Figure 4: Some interesting samples drawn from a large regula
rized model conditioned on "The
meaning of life is". We have removed "unk", "N", "$" from the set of permissible
words

These are considered good results. Mine are much better.

Literary Aside: http://arxiv.org/abs/1312.6120

Initialization

If the matrix is NxN, the norm-preserving random matrix has standard eviation 1/Sqrt(N)

But the claim is that an orthognal matrix is better.

https://github.com/IoannisAntonoglou/optimBench/blob/master/factories/torchExtensions.lua


```
function torch.orth(dim)
assert(dim>0, '0x0 orthogonal matrix cannot be defined')
local a = torch.randn(dim, dim)
local u, s, v = torch.svd(a)
return u
end
```

It is clear to me this generates orthogonal matrices with "random" entries. However, it's not obvious to me that it randomly fills the space of NxN orthogonal matrices. (Nor is it obvious that it does not.)

Apparently the canoncinal way is to generate a random matrix X and do QR decomposition. Q is then orthogonal.

Back to my nets

Have achieved the best loss yet, 4.33 after 1 epoch with very small learning weight. Temperature 0.6:

I will not have been a man of the old man , and I have no position to me . Do you know that I have been finished by my own time as I have been done to me . You know that I should have been a good deal of my own and more than the money of my life . "

" It is a very good thing , " said the girl , " and I would have been ready to go from the land . "

" What do you mean ? " asked the girl .

" It is very great . "

" You are not a little known , " said the girl .

" We are in the country , " said the girl .

" You must have been a somewhat serious way , " he said . " I am sure I am not to go back to the rest , and I suppose that I have a great deal of a long time , but I have not forgotten that I could not have to be the truth . "

A paper on optimizing GPU use makes it sound like powers of two are a good thing. So, reduce learning rate 10x and make powers of 2:

th train.lua -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 5000 -print_ev
ery 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 35 -batch_size 50 -
seq_length 64 -learning_rate 5.0e-5 -max_epochs ${epochs} -dropout 0.25 -seed 77

lm_lstm_epoch0.70_5.0834.t7
lm_lstm_epoch1.05_4.8724.t7
lm_lstm_epoch1.41_4.7736.t7
lm_lstm_epoch1.76_4.7086.t7
lm_lstm_epoch2.11_4.6295.t7
lm_lstm_epoch2.46_4.6051.t7
lm_lstm_epoch2.81_4.5729.t7
lm_lstm_epoch3.16_4.5190.t7
lm_lstm_epoch3.52_4.5144.t7
lm_lstm_epoch3.87_4.4849.t7
lm_lstm_epoch4.22_4.4599.t7

The 4.4599 output is trippy at tmperature 0.8:

*But as the island , is that a dying man has been taken to the permanent ,
but the seven hours is more small . But this we have money , and it was
like the loyal form . . The other is grown and herself
There are pleasant all as soon . A man must allow that some are the
direction of the government was the further . At least that you had not
made a great deal of our part , and in the use of the full hills - - the*

At temperature 0.6:

*The friend is a good deal of a boy , and the most smell of such
love of the little and devoted , with the old man - - whom he is
made to be his secret to her .*

*" You have got the same , ' said he , " and is a man who is an family of
an hour . "*

*" What is you , " replied the girl , " I shall have seen - - too , I know ,
for I must have a lady ' s work . "*

*" I am sure , " said she , " and you are to take a little in my hand ,
but you are just when the other ' s would be a great use . "*

*" I was once to ask me , " she said , " I am not sure of my Lord
of me . "*

" That ' s you this day ? " said the young man . " There ' s a man ! "

Not quite as good as 4.33 but acceptable.

Convergence is about 8x slower than 5e-4. I'd guess that when convergence is linear in the step size, then the step size is not a very important factor in finding the right minimum. Right now, when I reduce stepsize, it does not reduce time as much. To me that says it should still be kept small. Large step sizes act like momentum, so they have some extra advantage. If the time decrease is not even as bad as linear, the small steps are doing something "qualitatively good."

Large step size, 4.49 (cv/lm_lstm_epoch0.55_4.4967.t7), temperature 0.6

" What will not be ? "

" Mrs . Budd and my mother , " said the old man , " for this things , as
I have a secret of the King , and I am not going to let me the same
family . "

" Oh , I am many of the favour , " said the aged woman , " and I shall be
prepared to think you will be able to take your interests . "

Comparable.

lm_lstm_epoch0.35_5.4310.t7
lm_lstm_epoch0.70_5.0834.t7
lm_lstm_epoch1.05_4.8724.t7
lm_lstm_epoch1.41_4.7736.t7
lm_lstm_epoch1.76_4.7086.t7
lm_lstm_epoch2.11_4.6295.t7
lm_lstm_epoch2.46_4.6051.t7
lm_lstm_epoch2.81_4.5729.t7
lm_lstm_epoch3.16_4.5190.t7
lm_lstm_epoch3.52_4.5144.t7
lm_lstm_epoch3.87_4.4849.t7
lm_lstm_epoch4.22_4.4599.t7
lm_lstm_epoch4.57_4.4507.t7
lm_lstm_epoch4.92_4.4208.t7
lm_lstm_epoch5.27_4.4182.t7
lm_lstm_epoch5.62_4.4061.t7
lm_lstm_epoch5.98_4.3743.t7
lm_lstm_epoch6.33_4.3805.t7
lm_lstm_epoch6.68_4.3733.t7
lm_lstm_epoch7.03_4.3265.t7
lm_lstm_epoch7.38_4.3502.t7
lm_lstm_epoch7.73_4.3411.t7
lm_lstm_epoch8.08_4.3035.t7
lm_lstm_epoch8.44_4.3241.t7
lm_lstm_epoch8.79_4.3150.t7

Temperature 0.8

*If they ' ve not been compelled to take your hand from the place -
stock would be generally as much as I can take out of the place ,
and to find it our best to take the money for the people , or too
well to keep them . They will sit at a discovered fortnight , and
here we might not admit that we shall be ready to mother , and
to keep away that I should wish to greet his part of a rumour .
But if I did not difference for your friend myself did I know that
I had always been sure to be buried with the other , and you may
have to do this sinking . Here is a good way that you are for
acquainted your father . "*

" I ' ll make by the field . "

*" Then you will not twelve hundred times . It is very very true ,
for , I can tell you my mind , except that I have heard you might
have come , and you ever become in of what I will may state . "*

*" I will tell you that I am going to talk to you so , " continued she
. " I have been running by for the present place . But when I can
see it the more ships I have to listen to you , and they there ' ll
take my hands as I could . I have certainly to get up , but I
am able to think that I should be able to suit the estates to
give your work . "*

*" I am very glad to keep my happiness into my advice . I am not
afraid that Mr . swindling has a few times to have been in the
first way , and he has not brought his younger wife to let my end ,
and I don ' t think I shall very well be ! "*

*" I think I will have myself done to me . I will have it to do
that Lady hankering for yours , and I am in the room , who shall be
here - - or that I have never seen it ? "*

Temperature 0.6

*The solitude of the question , as the most consideration of the old ,
stood by the rock , and the others were in place .*

" Had you loved him ? " asked the man .

*" You are the very brother , " he answered , " and that will be a whole
thing , and you must be alone in the world . "*

The servant was a little longer , who had been already accounted by last his friend ' s knife , and he pointed at the top of the stranger . The officer was mistaken ; and the captain was standing on his head , and came into the line of the door .

" It was my father ! " said the King , " that I will not tell you that he is quite fond of you . "

" Yes , " said the captain , " I am not in the forest . "

" My mother , " said the captain , " that I have sent to the mind of the public , and I will not enter the field . "

" Yes , " said the King , " what is the matter of my life ? "

" You have no doubt , " said the lady , with a smile . " I should have a strange deal of a different time , but we have a good deal to do . "

" I have come to see me with a man , " said he , " but I am not even the poor woman that I ' ve seen you . "

" I am not the most sentiment of this , " said the young man , " and you have been a good deal of a young man , and I am not sure that I have not been sure that I have not given him to the better of my own .

" Then I am afraid of the moment , " said the young man . " You may not learn that they are not to be brought out in our power . "

" You will not know that , " replied the captain , who , after his study , " I will not have put the matter . "

" I would not escape , " said the King , " himself has been a little Miss old friend , and that is a word for the other time . "

Temperature 0.5

The most of the most important mode of the country which is not to be found in the same time . It is not the most thing , for that I dare see that the people of the people are not in the state of the family , and that they would be sure to get down the river , and the whole of the world are the infant of the world .

CHAPTER XV .

The next day , I have left it , for the first time he had been on the ground , and I shall not have been able to make a great deal of the business of my own life .

I have been a man of the English , or that I have been most of the same day , and I have told you that I shall not be able to give the money in the world . I will learn to speak for the last time , and I will not be able to learn that , if you can , I will have to make a man in my way . I am not sure I would have been a good deal of the law , and I am sure I should be sure , and of course I will not have been the more than I should have made . I am sure , I am sure , but I know that I have been too much to do it , and that I am not now to take the way of the life . I shall have to take up the river , and I will not leave you to do it . "

" I am not sure , " said the young man , " but I should not have been to do that I can get to the end of the world . "

" I am sure , " observed the old man , " and there is no one else , and it is to be a very well - looking French . "

" I can ' t go to the ground , " said the boy , " and I will go at a little time . "

" I don ' t know , sir , " he said , " but I have a good deal of other things in the world ! I think you are not a good man , but I don ' t think that you are in the same way . "

" I am not sure , " said the boy , " and I must not go to the house . "

" I am sure , " said the man , " I may have to give a chance to get my home to the town . "

Observation: a difference of 0.1 seems to correspond to a perceptible improvement in output.

Let's keep going from here, with lower learning rate, and start decaying it

th train.lua -init_from 'cv/lm_lstm_epoch10.00_4.2673.t7' -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 5000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 35 -batch_size 50 -seq_length 64 -learning_rate 2.0e-5 -max_epochs ${epochs} -dropout 0.25 -seed 77 -learning_rate_decay_after 1 -learning_rate_decay 0.94

lm_lstm_epoch0.35_4.2869.t7
lm_lstm_epoch0.35_4.2869.t7
lm_lstm_epoch0.70_4.2914.t7
<mark>lm_lstm_epoch1.05_4.2565.t7</mark>
lm_lstm_epoch1.41_4.2822.t7
lm_lstm_epoch1.76_4.2837.t7

The primetext seems not to work.

Revisit the plan.

Next Directions, 4/9 0748 (strikethrough have been addressed, red are new, green are results)

~~Better mixing and aggregation~~
    -Aggregation: only considers 50 words at a time
    -Mixing: I want to re-mix the training batches every so often, but hold out the same test batch
    -Mixed batches, results improved

~~Longer sequences~~
    -Modest impact

Stress encoding

Rhyme/syllable encoding
    -Should probably remove Milton

~~Larger training sets~~
    -Modest score improvement but clear grammar learning even for complex grammar
    -Clean romantic without milton or frost or norton and some additional poets

"In the style of"

~~Function of temperature~~
    -Best results typically around 0.5-0.65. Up to 0.8 viable. Lower leads to a lot of repetitiveness, higher tends to be too trippy.

50-dimension word vectors

200-dimension word vectors

~~Stemming and casing~~
    -Investigated casing, casing is accounted for, stemming is not
    -Stemming may be valuable for new words

GRU

Better new-word embedding
    -Currently new words are randomly initialized—should stem/overlap/edit distance to guess nearest word and initialize with that
    -There are both Levenshtein Distance and Porter Stemming implementations in lua
    -Should only update new words, not old ones

Independent word list
    -Should be able to independently provide a word list to append to current list.

Primetext
    -Doesn't work

Learning rates

Orthogonal matrix initialization
    -Probably better initialization, maybe not so hard

Syllabification

https://webdocs.cs.ualberta.ca/~kondrak/cmudict.html

The following file contains the CMU Pronouncing Dictionary version 0.6 augmented with syllable boundaries (syllabified CMU). The boundaries were generated automatically using a structured SVM approach, which was shown to correctly syllabify over 98% words in the CELEX dictionary. If you use this data in your work, please cite the following paper:

Susan Bartlett, Grzegorz Kondrak and Colin Cherry.

On the Syllabification of Phonemes.
NAACL-HLT 2009.

cmudict.0.6d.syl

Last modified: October 2009

Did the following:
 -Created vocab with romantics.mixed together with Gutenberg Fbits.
 -Will now train on Gutenberg only.
 -At 10 epochs, pull off gutenberg and train on romantics.mixed.

lm_lstm_epoch0.33_5.4465.t7
lm_lstm_epoch0.67_5.0960.t7
lm_lstm_epoch1.00_4.9218.t7
lm_lstm_epoch1.33_4.8206.t7
lm_lstm_epoch1.67_4.7428.t7
lm_lstm_epoch2.00_4.6867.t7
lm_lstm_epoch2.33_4.6433.t7
lm_lstm_epoch2.67_4.6059.t7
lm_lstm_epoch3.00_4.5740.t7
lm_lstm_epoch3.33_4.5461.t7
lm_lstm_epoch3.66_4.5176.t7
lm_lstm_epoch4.00_4.4933.t7
lm_lstm_epoch4.33_4.4732.t7
lm_lstm_epoch4.66_4.4564.t7
lm_lstm_epoch5.00_4.4400.t7
lm_lstm_epoch5.33_4.4260.t7
lm_lstm_epoch5.66_4.4123.t7
lm_lstm_epoch6.00_4.3973.t7
lm_lstm_epoch6.33_4.3865.t7
lm_lstm_epoch6.66_4.3766.t7
lm_lstm_epoch7.00_4.3663.t7
lm_lstm_epoch7.33_4.3575.t7
lm_lstm_epoch7.66_4.3486.t7
lm_lstm_epoch8.00_4.3420.t7
lm_lstm_epoch8.33_4.3327.t7
lm_lstm_epoch8.66_4.3246.t7
lm_lstm_epoch8.99_4.3192.t7
lm_lstm_epoch9.33_4.3113.t7
lm_lstm_epoch9.66_4.3049.t7
lm_lstm_epoch9.99_4.2996.t7
lm_lstm_epoch10.00_4.3005.t7

Attempt to switch to Gutenberg
 The vocab and the data file are tied together. I've tried to split/replace them but without luck so far. Not sure how to deal with it.

There should be a way to pull in the model and vocab separately from the data, but it's really not obvious how to do that within the structure.

```
print('saving checkpoint to ' .. savefile)
local checkpoint = {}
checkpoint.protos = protos
checkpoint.opt = opt
checkpoint.train_losses = train_losses
checkpoint.val_loss = val_loss
checkpoint.val_losses = val_losses
checkpoint.i = i
checkpoint.epoch = epoch
checkpoint.vocab = loader.vocab_mapping
checkpoint.optim_state = optim_state
checkpoint.optimizer = opt.optimizer
torch.save(savefile, checkpoint)
```

Protos is the flattened model, but not unrolled in time. The checkpoint is 212MB, but there are only 13M parameters. If they're doubles, that's 52MB. Not sure where the 4x comes from. Vocab mapping is 9MB.

So for now, simply apply the learning-rate knowledge, training only on poems

th train.lua -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 5000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 2 -batch_size 50 -seq_length 64 -learning_rate 5.0e-5 -max_epochs ${epochs} -dropout 0.25 -seed 77 -learning_rate_decay_after 5 -learning_rate_decay 0.97

OK, will run it overnight.

lm_lstm_epoch6.39_5.5681.t7
lm_lstm_epoch12.77_5.3285.t7
lm_lstm_epoch19.16_5.2158.t7
lm_lstm_epoch25.54_5.1399.t7

The 5.1339 with temperature 0.7 is not that bad.

But run the lush chain , the green , on the blue
Of the sky , and the war of the bright song ,
Of a cry and sunny wished ;
Thou the dainty far , and them him ,
Where such the morning , and the gentle bed
A price with the hour of the gold .
The night of the earth of the wilderness ,
A lyric shore at a emblem ,

The Wi , of the noon of the old man ,
And others and her , as the eye of night
And the saint of the way .
 ENDPOEM

Run 100 epochs, starting the learning rate at 2.5E-5 and degrading from there

th train.lua -init_from 'cv/lm_lstm_epoch25.54_5.1399.t7' -gpuid 0 -num_layers 3
 -rnn_size 256 -eval_val_every 5000 -print_every 100 -word_level 1 -glove 1 -dat
a_dir ${mydir} -threshold 2 -batch_size 50 -seq_length 64 -learning_rate 2.5e-5
-max_epochs ${epochs} -dropout 0.25 -seed 77 -learning_rate_decay_after 5 -learn
ing_rate_decay 0.97

lm_lstm_epoch1.65_4.4653.t7
lm_lstm_epoch9.14_4.2878.t7
lm_lstm_epoch6.39_4.9708.t7
lm_lstm_epoch12.77_4.7670.t7
lm_lstm_epoch6.39_5.5681.t7
lm_lstm_epoch12.77_5.3285.t7
lm_lstm_epoch19.16_5.2158.t7
lm_lstm_epoch25.54_5.1399.t7
lm_lstm_epoch6.39_5.0928.t7
lm_lstm_epoch12.77_5.0581.t7
lm_lstm_epoch19.16_5.0352.t7
lm_lstm_epoch25.54_5.0186.t7
lm_lstm_epoch31.93_5.0032.t7
lm_lstm_epoch38.31_4.9940.t7
lm_lstm_epoch44.70_4.9883.t7
lm_lstm_epoch51.09_4.9791.t7
lm_lstm_epoch57.47_4.9729.t7
lm_lstm_epoch63.86_4.9695.t7
lm_lstm_epoch70.24_4.9670.t7
lm_lstm_epoch76.63_4.9602.t7
lm_lstm_epoch83.01_4.9575.t7
lm_lstm_epoch89.40_4.9565.t7
lm_lstm_epoch95.79_4.9534.t7
lm_lstm_epoch100.00_4.9526.t7

4.9602, tempewrature 0.7

ENDPOEM
 The Poor Blow
 The hill of the grateful rock ,
 And the holy fall !
 And the sun of slow chill ,
 And the sand of night .
 I hear the waters
 About the river ' s side ,
 And the walk of the door ,

Till is a sun
The night of their dews .
I have heard the sun of May ,
I ' ve seen the heavy village ;
And the day I never sever
The old ' s bed and come .
ENDPOEM
The stars
The bed of the We are long ,
But the morning - door ,
With the sea of a living sky
Filled the earth ;
The cry that a want is done
And that might believe .
ENDPOEM
The Land
She , and the awe of a husband ,
The sound of his skin , and there
With a hour of whirling .
Then with a wigwam - moon
I saw a tale of the world ,
And her far , so - - I ,
And now we know a best
Of a tragedy - -
ENDPOEM
The Summer Of the)
Come , at a black stream .
I sat with the shore of a storm ,
I ' ll know the city of the hills ,
And the golden way my soul ,
And the moon of her times I know ,
When the long hills are wild down ,
She ' er the harp of the Wild sea
The roots of the dew ;
O ' s , I ' ll have be my joy
Through I know my sweet ,
And I am , I ' s know my hand
That I ' ve come to think .
She ' s I ' t work the ' take - -
I ' ll could make a way .
For the - glory , you ' ll fell - - -
And my heart are man ,
And as a heart I they ' ll die ,
And me she ' s a man .
ENDPOEM

The final epoch 100,

Temperature 0.7

With a house
I have see and bright ,
Have spring , nor fair ,
I wish not in my face ;
And I saw my book ,
I not be to be .
I am a tale upon me ,
And I am my heart
My life ' s - pain .
But I ' ll laugh ; " I heard me ,
A master came ,
And they will have not seen
I ' ll not you hear ,
And I ' m a I ' s a wife ,
And better I gives . "
And bold is I would die
In my heart I ' ll die .
I ' ve been a father ,
I ' m I ' ll make the think .
I ' m perish ' s good to me ,
I ' m a wife .
ENDPOEM

Temperature 0.6
th sample.lua cv/lm_lstm_epoch100.00_4.9526.t7 -temperature 0.6 -seed 55

The Him .

.

What sees the letter of the morning ,
The wind of the wind - tree ,
All all the desert of a door ,
And the sun is a year .

ENDPOEM

The birds of the wings , the church
Of dark beside the river hills .
He is a little happy ,
And the stars of a wolf ' s breast ,
And the faint of the wind
Of our sweet windy eyes ,
And , like the white of a snow ,
And the distance of the sun ,
And the stars of childhood ' s dawn ,
And the wave of the night .
The fields of the summer falls

Of the dew of the sun ,
And the cold o ' er the sky ,
And the sea of the sky !
The waters of the thunder ,
And the sun of the life .
The dark , the flowers of a moon ,
The human friend was about ;
And the good of the slave
Of the morning ' s side ;
And the gray bleeding lost
In the tree ' s pledge
ENDPOEM

The Rises
The Day of its world was dry ,
And the moment ' s green song .
In the dark of the night
Before the bloom of day ,
And the dark ' s soft sighing
Once the sad ' s head !
Beside the water ' s moon ,
The mountain hill ,
And the leaves of the night

The Rises did not complete…

Temperature of 0.5, seeded with ENDPOEM, yields only Newlines and random punctuation.
Temperature of 0.6 does the same!
Temperature 0.7 does generate a poem, though followed by a lot of lines and punctuation

Modified CharSplitLMMinibatchLoader.lua to separately make vocab and data .t7 files
Successfully tested that in base case (neither present), it correctly generates them (identical to what was generated before)

OK, it seems to have done what I wanted. Just one worry, the data file is exactly the same size as before. I'm not sure it's wrong, but I'm not sure that's right, either.

Somehow miscoordinated which vocab is in the checkpoint! The checkpoint has 31494 which is bitsF at threshhold 35. (This is too bad. I wanted Fromantics at 20. I can come back to that.)

Seems to work, nothing crashes now.

Ran it. Start at 100 with loss of 5.35. Huge improvement in start!

Does very rapidly hit a wall around 4.95.

epochs=50

th train.lua -init_from 'cv/lm_lstm_epoch9.14_4.2878.t7' -gpuid 0 -num_layers 3

-rnn_size 256 -eval_val_every 5000 -print_every 100 -word_level 1 -glove 1 -data _dir ${mydir} -threshold 2 -batch_size 50 -seq_length 64 -learning_rate 2.0e-5 -max_epochs ${epochs} -dropout 0.25 -seed 77 -learning_rate_decay_after 5 -learni ng_rate_decay 0.97

Note added 4/14: it seems I did not save this to a Logfile! Doh!

Slowly breaking through around 5 epochs, at least in training; possibly optimistic?

About 4.82 in training around 6 epochs; validation is...4.6867
!!!!!!!!!!!!!!!
cv/lm_lstm_epoch6.39_4.6867.t7

At epoch 11, training is about 4.75

The best poem training to date appears to have been lm_lstm_epoch10.00_4.6846.t7

Now I get
cv/lm_lstm_epoch12.77_4.6196.t7

At epoch 18, training is around 4.65

cv/lm_lstm_epoch19.16_4.5845.t7

lm_lstm_epoch6.39_4.6867.t7
lm_lstm_epoch12.77_4.6196.t7
lm_lstm_epoch19.16_4.5845.t7
lm_lstm_epoch25.54_4.5622.t7
lm_lstm_epoch31.93_4.5464.t7
lm_lstm_epoch38.31_4.5350.t7
lm_lstm_epoch44.70_4.5264.t7
lm_lstm_epoch50.00_4.5197.t7

Training at epoch 50 has about 4.58

From last, temperature 0.8 (all these are the default seed)

 And thus we ' ll be every hemlock ,
 And this is the Death .
 The first of the plains
 To the sea - wind and afar ,
 Through the green rocks and leaves - -
 The doom of earth .
 And one with corpse playing them
 On the earth ' s bank ,
 And I am wide and much
 Of with a dead and white ,
 The feast of the green ,

In the welcome ,
And yet and in its blue
I kept it away
For that I know ,
A third , a sense ,
With the proper wish
To have to be to shut .
...
And so when I are dead and those
With the lovely wind ,
The old and the golden hue
Which is the early years .
And while we loved
In the wave of love ,
And which to buy a prize !
But yet , such as you come ,
My beautiful ' s heart ,
To leave the heart so .
O , whose things be my eyes
On school and heart ,
And sit to thy eyes ,
I ' m wrong in love ;
Nor I who might am in bring
The spirit on earth ,
And with human heart ,
And fill the while of death ,
Those of the past ;
In vain to the decay
Of the echoing song
The little sun are light .

Temperatur e0.7:

Who is not all his own ?
Though the man ' s found it still by the same ,
But he knows , but his life is dared .
His heart is from his head he brought ,
And life ' s leave in his hand .
But it is not that in the bells
In earth and scant things .
" The faith and the prayers is not ,
The soul is done , - - years .
Still that I have done , " What I am
They shall never see ,
And I will not know that I
You will go and live or will .

The sky was just a drifting thing ,
And when I turned the heart with the sighing ,

And I drove after , " Though I ' ll go !
Then , wish I make a bit of you
And make to your eyes If you come . "

Temperature 0.6 is GOOD

The letter of the old Lord , and the ancient
 Of the mountains , in the black of the moss ,
 The echo of the summer - - and the water ,
 And the purple - green horror of the sea ;
 The moving of the moss , and the dead of the seas ,
 The clouds are seen , and the sea of the sky
 Of the green - bell , and the light of the air
 Of the glory which is the barren sea ,
 And the beams of the waters , the murmurs of the road ,
 And the night and the dying ' s sight , and the sun is heard ,
 And the earth is the wind , and the wind is a bright stone ,
 The brown - green river , the snowy wood ,
 That the green above the sea , and the sea , and the skies ,
 And the sun ' s sea , the moon ' s roof , the smooth grain of rock ,
 The waters , the sky , and the clouds , and white and green ,
 Where the stream of the sky , the shore , the rock , the gay .

 And the pale - brown , rivers , the red , and the dark ,
 The wild , and the white - green blue , and the white sea ;
 The white - green , with a gray - hearted , and the heart
 And the marble , and the hare ' s white - trees , and the sun ,
 The snow - trees , to the sea , and the bleak ,
 The dim and forest - wide , and the waves of the sky ;
 The Northern waters of the sky , and the silent sky ,
 The very heavy valley , the long - wall of the sea ,
 Or the sea - pearl is in the noon , and the day
 The sun is a bright and deep , and the waters
 Of the mountains , the waters of the heavens ;
 The broad hills of the mountain ' s lake of green ,
 And the green straggling blast of the sea ;
 The moon - tree ' s voice are a deep , and the sun
 Was the long skies that wide and its blood ,
 And the sate ' s heart is ever a voice of eye
 Of the dark and wind , on the mountains on the shore
 The sun - lake , the gray - flowers , and the green sea ,
 With the mountain ' s green , and the red - silent flame
 Of the blue wind of the Alone of the sun - -
That the thundering valley is growing open ;
 The stone of the sun - tower on the hill ,
 With a clear light , and the heat , with the waves ,
 And the gray hill , like a wild weather , and lies
 In the forest - shore , the leaves of the moon .

I never saw the green - grey leaves , and now ,
Nor long in the winds , the pale and light of air ,
And the water - roar in the river ' s sleep ,
And the dark glow of the ancient sky ,
And then , and the white stars , with the ancient wave ,
Or the beautiful scene of the sea and sea ,
With the dark - trees of the moon , the sky ,
And the sky of the mountains , by the sky
Where the deep leaves the deep leaves of the grain
Of the sap - stone , in the sky , and the sun .
The same , from the cold light of the world ,
The village of the hills of the sea is done ,
By the trees of the river , and a tall stream ,
And the twenty waters , and the eighty - white ,
And the leaves of the apart in the snow ,
The sky of the forest , the the air ' s lids ,
The sound that the sky is the quick light ;
And the stone , from the winds of the mountain ,
The sun was suddenly .

The grass lights from the plain and the mountains ,
The mountain boughs that came in their fields .
The mountain was not as a child of old
By the wind , and as the wind were bare on its feet ,
And on the light of the mountain - wind came
A waters in the song of a tide .

She saw the light that still the summer leaves
In the green - air , and the wild green above
The day of a day , in the dark and stream ,
And in the twilight , they made a soft ball
And then , like the great sky , the blue birds ,
Until the cruel eye of the darkness ,
And the light that felt the whole of a cloud ,
And , as the sun of the sea - lights midnight ,
And all the dark woods of the dark hide ,
The white as the light of the deep stars ,
And the cold , being sweet , the shadow of the sea ,
And the white - eyed - clouds , and the green and bright ,
The sky of the green - rock , and the moon ' s brain ;
And , from the stream , the plain of the sun
…

Even temperature 0.4 has its moments

The sun is a golden cloud ,
The sun and the hills ,

The wind of the sea ,
And the green hills and the sky
And the sun in the sky ,
And the sun , the sand ,
And the wild sea and the sand ,
And the sky and the sea ,
And the night , the gleaming trees ,
And the waters of the sea ,
And the waves the sun
Of the earth , and the sun ,
And the hills of the sun .

All of them have lost ENDPOEM.
Because it's not in the vocabulary. The vocabulary here was trained off bitsF only (see above).

Kepp going;

echo "Starting 50/64"
th train.lua -init_from 'cv/lm_lstm_epoch50.00_4.5197.t7' -gpuid 0 -num_layers 3
 -rnn_size 256 -eval_val_every 5000 -print_every 100 -word_level 1 -glove 1 -dat
a_dir ${mydir} -threshold 2 -batch_size 50 -seq_length 64 -learning_rate 1.0e-5
-max_epochs ${epochs} -dropout 0.25 -seed 77 -learning_rate_decay_after 5 -learn
ing_rate_decay 0.97

lm_lstm_epoch6.39_4.5084.t7
lm_lstm_epoch12.77_4.4996.t7
lm_lstm_epoch19.16_4.4921.t7
lm_lstm_epoch25.54_4.4865.t7
lm_lstm_epoch31.93_4.4823.t7
lm_lstm_epoch38.31_4.4789.t7
lm_lstm_epoch44.70_4.4762.t7
lm_lstm_epoch51.09_4.4732.t7
lm_lstm_epoch57.47_4.4718.t7
lm_lstm_epoch63.86_4.4704.t7
lm_lstm_epoch70.24_4.4690.t7
lm_lstm_epoch76.63_4.4680.t7
lm_lstm_epoch83.01_4.4670.t7
lm_lstm_epoch89.40_4.4665.t7
lm_lstm_epoch95.79_4.4659.t7
lm_lstm_epoch100.00_4.4656.t7

Still getting somewhere! Last learning rate was 5E-7

Next Directions, 4/12 2016 (strikethrough have been addressed, red are new, green are results)
~~Better mixing and aggregation~~
        -Aggregation: only considers 50 words at a time
        -Mixing: I want to re-mix the training batches every so often, but hold out the same test batch
        ~~-Mixed batches, results improved~~
~~Longer sequences~~

~~-Modest impact~~
Stress encoding
Rhyme/syllable encoding
　　　~~-Should probably remove Milton~~
~~Larger training sets~~
　　　~~-Modest score improvement but clear grammar learning even for complex grammar~~
　　　　　~~-Clean romantic without milton or frost or norton and some additional poets~~
"In the style of"
~~Function of temperature~~
　　　　　~~-Best results typically around 0.5-0.65. Up to 0.8 viable. Lower leads to a lot of repetitiveness,~~
~~higher tends to be too trippy.~~
50-dimension word vectors
200-dimension word vectors
~~Casing and~~ Stemming
　　　-Investigated casing, casing is accounted for, stemming is not
　　　-Stemming may be valuable for new words
GRU
Better new-word embedding
　　　-Currently new words are randomly initialized—should stem/overlap/edit distance to guess
nearest word and initialize with that
　　　-There are both Levenshtein Distance and Porter Stemming implementations in lua
　　　-Should only update new words, not old ones
~~Independent word list~~
　　　　　~~-Should be able to independently provide a word list to append to current list.~~
Primetext
　　　-Doesn't work
~~Learning rates~~
　　　　　~~-Valuable! They were set way too high to start and limited progress~~
Orthogonal matrix initialization
　　　-Probably better initialization, maybe not so hard
Bias initialization
　　　-Final softmax can be initialized to give raw word count frequencies
Capitalization bit
Post-processing
　　　-Punctuation space-removal
　　　-Duplicate words in a row removal

Test orthogonal initialization in LSTM

Link to Fromantics.txt, let it create vocab map, data, glove, and initialize

th train.lua -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 5000 -print_ev
ery 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 20 -batch_size 50 -
seq_length 64 -learning_rate 2.0e-5 -max_epochs ${epochs} -dropout 0.25 -seed 77
 -learning_rate_decay_after 5 -learning_rate_decay 0.97

Everything works, no errors. Orthogonally initialized now!

Next: link in bitsF.txt. Delete data. Retain glove and vocab. Touch them after relinking bitsF.

th train.lua -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 5000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 20 -batch_size 50 -seq_length 64 -learning_rate 5.0e-5 -max_epochs ${epochs} -dropout 0.25 -seed 77 -learning_rate_decay_after 1 -learning_rate_decay 0.97

Oddly, this crashes, out of memory. It's odd because bitsF is smaller than Fromantics, and Fromantics had sufficient memory (that's how the orthogonalization was tested.) It's not insane that it would crash when the earlier runs did not, because it has a larger vocabulary (44k words, 17M parameters).   But I don't know why Fromantics worked, then.
Do batches of 40

th train.lua -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 5000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 20 -batch_size 40 -seq_length 64 -learning_rate 5.0e-5 -max_epochs ${epochs} -dropout 0.25 -seed 77 -learning_rate_decay_after 1 -learning_rate_decay 0.97

First 100:
cloning criterion
100/1778000 (epoch 0.006), train_loss = 8.26734713, grad/param norm = 1.0438e-01, time/batch = 0.6055s

Then reran to put in log, Never seen this:

loss is NaN. This usually indicates a bug. Please check the issues page for existing issues, or create a new issue, if none exist. Ideally, please state: your operating system, 32-bit/64-bit, your blas version, cpu/cuda/cl?

Then reran again, got:
100/1778000 (epoch 0.006), train_loss = 8.23235862, grad/param norm = 1.0438e-01, time/batch = 0.6097s

This is peculiar. There is a default seed of 123. It should be the same every time. ???

Note on orthogonl initialization
The initial loss was larger than when running last time (the vocab is larger this time.) By the 400th iteration, the loss is better than last time. After that it begins the long, slow descent.

Scale the Local Summit! Or valley, if you prefer.

Set it to learn for 100 epochs. That's 2 weeks of time! Last time I stopped after 10 epochs and I think that seems right to me. But this gives me the option to just let it run free if it's getting somewhere.

At epoch 0.33 the training is now significantly behind where it was before (smaller vocab, uniform initialization). The long slow descent is now slower, by approximately a factor of two.

lm_lstm_epoch0.28_5.9510.t7

It was 5.4465 last time, though at slight advance. But I think at 0.33 this will still only be at about 5.93.

lm_lstm_epoch0.28_5.9510.t7
lm_lstm_epoch0.56_5.9062.t7
lm_lstm_epoch0.84_5.9060.t7
lm_lstm_epoch1.12_5.9561.t7
lm_lstm_epoch1.41_6.0357.t7
lm_lstm_epoch1.69_6.1067.t7
lm_lstm_epoch1.97_6.1986.t7
lm_lstm_epoch2.25_6.2686.t7
lm_lstm_epoch2.53_6.3501.t7
lm_lstm_epoch2.81_6.4160.t7

This is bad.

OK, try a uniform initialization. Sad face.

At first, not looking any better. Hmm.

Backup plan if this goes awry: vocab threshhold of 35. But that will be mysterious.

lm_lstm_epoch0.28_5.9199.t7
lm_lstm_epoch0.56_5.9062.t7
lm_lstm_epoch0.84_5.9060.t7
lm_lstm_epoch1.12_5.9561.t7
lm_lstm_epoch1.41_6.0357.t7
lm_lstm_epoch1.69_6.1067.t7
lm_lstm_epoch1.97_6.1986.t7
lm_lstm_epoch2.25_6.2686.t7
lm_lstm_epoch2.53_6.3501.t7
lm_lstm_epoch2.81_6.4160.t7
lm_lstm_epoch0.28_5.9199.t7
lm_lstm_epoch0.56_5.8994.t7
lm_lstm_epoch0.84_5.9053.t7
lm_lstm_epoch1.12_5.9782.t7
lm_lstm_epoch1.41_6.0194.t7
lm_lstm_epoch1.69_6.0598.t7
lm_lstm_epoch1.97_6.0940.t7
lm_lstm_epoch2.25_6.1395.t7
lm_lstm_epoch2.53_6.1680.t7
lm_lstm_epoch2.81_6.1895.t7
lm_lstm_epoch3.09_6.2044.t7
lm_lstm_epoch3.37_6.2259.t7
lm_lstm_epoch3.66_6.2762.t7

Not looking good for the home team.

Let's back up and try 35.

Re-link Fromantics. Regenerate vocab and glove.

Move data, relink bitsF, touhch glove and vocab

So, oddly, that worked. It's at 5.67 already at 3000 iterations.

Something about the large vocab is killing it?

Strategy Recapitulation: set up vocab on both poems and prose. Train the hell out of the net on the (giant) prose dataset. Then train on the poem data with that as initialization.

cv/lm_lstm_epoch0.35_5.4057.t7

This is actually a hair better than last time: lm_lstm_epoch0.35_5.4310.t7

This is currently a uniform initialization. I'm going to save the 10000 checkpoint:

lm_lstm_epoch0.70_5.0531.t7   (a hair worse than last time)

Now restart with orthogonal initialization:

In the first 1000 batches it's not at all obvious there's any difference.

Around 2000 it's not impossible it's slowing down worse than uniform

Around 4000, pretty similar, ortho may be about 0.07 worse

lm_lstm_epoch0.35_5.4737.t7

Right on the money, 0.06 worse.

Acceptable for now, for a chance to try something new and press on. Keep the orthog.

lm_lstm_epoch0.70_5.1170.t7
cv/lm_lstm_epoch1.05_4.9095.t7

Keeping up, good for the night

lm_lstm_epoch0.35_5.4057.t7
lm_lstm_epoch0.70_5.0531.t7
lm_lstm_epoch0.35_5.4737.t7
lm_lstm_epoch0.70_5.1170.t7
lm_lstm_epoch1.05_4.9095.t7
lm_lstm_epoch1.41_4.8054.t7
lm_lstm_epoch1.76_4.7265.t7
lm_lstm_epoch2.11_4.6695.t7
lm_lstm_epoch2.46_4.6197.t7
lm_lstm_epoch2.81_4.5800.t7

lm_lstm_epoch3.16_4.5528.t7
lm_lstm_epoch3.52_4.5220.t7

Pretty much spot on with before

Observation on really deep nets: the uniform was fixed to be [-0.08, 0.08], which is (roughly) right for 256. But that roughness would make you pay, the deeper it is. So maybe not surprising the 16-deep nets failed. Less clear why the 6-deep would fail. Would be worth revisiting with orthogonal initialization.

Everything known about nets suggest that a size 192 net 5 layers deep would be superior to 256 3 layers deep.

Note added 4/22 2231: I believe the following is simply a restart of the last run, orthog. Init and training on bitsF.

lm_lstm_epoch0.35_5.4737.t7
lm_lstm_epoch0.70_5.1170.t7
lm_lstm_epoch1.05_4.9095.t7
lm_lstm_epoch1.41_4.8054.t7
lm_lstm_epoch1.76_4.7265.t7
lm_lstm_epoch2.11_4.6695.t7
lm_lstm_epoch2.46_4.6197.t7
lm_lstm_epoch2.81_4.5800.t7
lm_lstm_epoch3.16_4.5528.t7
lm_lstm_epoch3.52_4.5220.t7
lm_lstm_epoch3.87_4.4983.t7
lm_lstm_epoch4.22_4.4762.t7
lm_lstm_epoch4.57_4.4586.t7
lm_lstm_epoch4.92_4.4410.t7
lm_lstm_epoch5.27_4.4268.t7
lm_lstm_epoch5.62_4.4106.t7
lm_lstm_epoch5.98_4.4011.t7
lm_lstm_epoch6.33_4.3863.t7
lm_lstm_epoch6.68_4.3728.t7
lm_lstm_epoch7.03_4.3627.t7
lm_lstm_epoch7.38_4.3576.t7
lm_lstm_epoch7.73_4.3450.t7
lm_lstm_epoch8.08_4.3352.t7
lm_lstm_epoch8.44_4.3278.t7
lm_lstm_epoch8.79_4.3204.t7
lm_lstm_epoch9.14_4.3133.t7
lm_lstm_epoch9.49_4.3055.t7

Then the disk filled.

It's OK. Will start from here with the poems.
Move the data, relink romantics.mixed, touch vocab and glove

It's running now. Loss is 5.48 at 100. The last time appears not to have been saved to a log, but the notation in this logbook is that the loss then was 5.35 at 100 iterations. Yay. Looks like last time was 4.6867 at 5000 iterations. If I'm even in the ballpark I'll be happy.
cv/lm_lstm_epoch6.39_4.8033.t7

In the ballpark at least!

lm_lstm_epoch6.39_4.8033.t7
lm_lstm_epoch12.77_4.7333.t7
lm_lstm_epoch19.16_4.6951.t7
lm_lstm_epoch25.54_4.6716.t7
lm_lstm_epoch31.93_4.6555.t7
lm_lstm_epoch38.31_4.6419.t7
lm_lstm_epoch44.70_4.6331.t7
lm_lstm_epoch51.09_4.6248.t7
lm_lstm_epoch57.47_4.6194.t7
lm_lstm_epoch63.86_4.6145.t7
lm_lstm_epoch70.24_4.6112.t7
lm_lstm_epoch75.00_4.6084.t7

Last time it was at about 4.48, so consistently running about 0.12 behind. Last time the limit was 4.465. The below are generated from 4.6084.

Temperature 0.6

" My love , " said she , " I love thee !
I know you were my old child ,
And all in the world ' s day ;
And you may have when you come ,
And I do not know you . "
When she had been to the land
And the stones of the young ,
Was the little child of the sea ,
And she was very pale .
She felt , too , like a small dream ,
She could not keep her to sleep ,
And now she knew she turned ;
And so she went by his arm ,
And with her eyes she saw her .
She came to her , and she was ,
And now the boy , for she has done ,
She had a little place .
She was at the two times in her ,
And she was a moment .
She stood in the shadow of her hand ,
She heard her huge hair ,
And with her lips she died ,
She was silent , and she had made
Her arms in her face .

She rose to the face of the hall ,
And her the head of her eyes .
" I have told me not , " she said , "
" I ' ll be what I could it . "
And in a foreign look ,
She did not look into her face ;
Then she turned , and she died :
" I ' ll be all that I ' ve lived ;
I ' ve seen my heart on the ours ,
And I ' ve got a straight ;
And you ' re too much to be a man . "
The old man said ,
" I ' ll have a woods in my bed ,
" And the father ' s little maid ,
" Let me hear my face ,
I ' ve got a bit of my . "
" I ' m glad to know it ,
I ' ll hold me in the ground ,
And I ' ll die with a dream
Of the old . "

" But I ' m going , " said she , "
" The Spring was mother ,
" I am not a dream ;
" I would have a leaf ,
For I ' ve been a friend .
I ' ve sat a little long
Till I ' ll be a bullet ,
The ' a ' a thing .
" And I ' ve done your house , I ' ll be ,
I ' ll tell you my sister .
I ' ll never be gaze ,
I ' m sure I ' ll go to the wall ,
And the day I ' ll make it ,
It ' s a bad thing ,
I ' m glad to say you ' m the worst .
ENDPOEM

Temperature 0.7

The heat of life

The little winds of the river
Is fire , are a grave or to be
When we are well for the whole ;
The old , all of the earth of men ,
Yet that a gay walls shall break
In the vast hand of the sky ,
The last and the sky of death ;

Our limbs were not like a single ,
What is not in the last ;
The hum of the world that was
And all our men are pure .
Yet , still the world , the Place of Heaven ,
The proud of the simple ,
The motion of that , and the tale
Of the very world ,
The pale of the loud sleep .
ENDPOEM

Temperature 0.8

Who have done just an hour as I do not choose ,
If swim to the occasion of our heart ,
Since I have heard one mere eye from the wind ,
And each names that I have whistled in the air .
I cannot have grown named there : but by life
Have I paid for all such a war in the rain .
But I know my own God , who thus do ,
For such you do not love the cause of good ,
I can choose it now to its sky .
ENDPOEM

Who have done just an hour as I do not choose ,
If swim to the occasion of our heart ,
Since I have heard one mere eye from the wind ,
And each names that I have whistled in the air .
I cannot have grown named there : but by life
Have I paid for all such a war in the rain .
But I know my own God , who thus do ,
For such you do not love the cause of good ,
I can choose it now to its sky .
ENDPOEM

And where these tumult are never
A - - - the folds of the dead - - -
Of this its woman ' s , love of blood
Of human soul ! and the whole day
Is its helpless rapture ,
Who used to make it break ?
V .
With light , and dusk , is one - -
The cometh and the night ' s night .
ENDPOEM

A mettle to the coldly fence ,
His cheek is sure , and that

He lies to fear the sun ,
To her I , who is one a soul .
No soul on his heart at me ,
The Judge shall tell him in the air ,
Save it
His eyes , and he can tell ,
His tender roaring body
Has thrown him before his heart ,
And sage - answered he stand ;
And the rested with a that
The years he seems to die ;
And he , he seems - -
If it comes , no more from the world ,
And along your hall - pain ,
Such that he will be sharply
That prayer of joy to book .
It is a stir of air :
If not to lack the least .
So what is the native
What should be
In the dead of its cause ,
Or the dreamy hand of him ?
A loss of God ,
That is he to his friends ?
Could you live to find
A helmet !
The poet is not truly ,
Would ' t himself know word :
When truth men filled the woe
Of the natures of his brow ,
Nor may be his faith .
How little could find his great crowd
By his cheek and star ?
ENDPOEM

And next past , why our birds are the melancholy
On that other place ;
Nature was beautiful looking - - -
And just I ' m born and shining
Your face for the done.
ENDPOEM

Temperature 0.9

The mid What
That was fare of the late .

The gloom
I enter ' d every day to hear
So to the buzz of darkness fly ,
And sits as I bright by birds ,
No waves grew bitter in the wood
Was beautiful of yore .
The edged was round a flash
Had seen her head , and returned
All stood now and to view ;
And all you heard no track ;
Since there was time - or I must see
More too child nor drop ;
And there was a spring and night ,
But things less than alien ;
As one not , were the house
My kiss and the father ' s need .
ENDPOEM

None have turned a beautiful phrase yet. Possibly luck?
The best is 0.8, then 0.7.
0.9 is useless; 0.6 is not so bad.
This suggests somewhere just north of 0.7.
Let's do 0.7, 0.73, 0.77, 0.8
It takes 16 minutes to do 2000 words. That's 7500 words an hour. Want it to finish by 5, about 8 hours.
Do 2.5 hours each. That's 18750 words each

Also did 68,66,63,60.

66 and 63 are the best of the lot.

Temperature 0.66

But I have found the Soldier
I was a little ,
And I knew my heart ' s eyes ,
So I would be
In one of the highest .
When I saw a second time
Of God ' s light ,
And the spear was early ,
That I could not express
The mysterious ones .
If I did not find it
I would have been in my way .
Yet I ' ve not to sing .
I would not keep my face ,
And though the thoughts was my own ,
And I know one of those

My voices would be :
But I would live and I ,
And I ' m sure for you .
ENDPOEM

My Father
I am not because I ' ve not seen
I ' ve loved of least that I have been in vain ;
My heart ' s by the world to be ,
So let me choose to let me hear ,
They would not be what I can hear .
I think that , when I had no other ,
So much before my own ever can be ,
I know it was the best of your own ;
And though I ' d be love for my friend ,
I would not know how I have been afraid
To see in the length of the grave .
It was the same that the king was given ;
And as he was so , to be the best
Of the real and the Lord ' s love .
A man had no care , and I know
I had no strength to die , and I
My own is a man ;
A man to be too late to live ,
And not to be a word :
But , for the true , yet that is died ,
And would I be sick of me ,
And then he will not die .
He am not in all his own
His life is life , and he is worth .
I have no question to him ,
That I will not be in a woman ,
And I need not know that it is ,
And in the world ' s sake that may be
That he will be abroad .
For yet I have not think I ' d come ,
And I do not think he will not die :
For others I should have put
To - morrow I ' ve been a friend ;
The house is so hard to think ;
But all the world was a troop
Of the man who is not his child .
And then so I I am much ,
And I must have a little pleasure
From the land of the same .
I have seen you no more than you
And tried to see the most day .
But I may have the same man
No country , that I had made

The same and your old had ,
And all the people have been alive ;
Then , in a beggar ' s face ,
I would have a farmer ' s cause ;
And the love of the world must be
With all a life to be .
ENDPOEM

I am a universe , but I must not die
And trust her by my heart , and I have been
As after that I used it to my own ;
But I ' m only one and act , and you may be
The border of the man , and some more I set
Through the grave of the world and the moon .
And so I am my son ' s twelve years
At least that I should not listen to me .
ENDPOEM

The shows of the All that is the old

The sunshine of the Old -
The Second - the siren -
The schemes of the Imagination
Of the other ,
The host is not the doubtful light -
Of the woodland - -
For a man , who never meet
We am going away -
The work of the dark -
And the Winter - in the mountain
Her flowers - -
And the sound - that is the low -
It ' s the Square - the Youth -
But Time ' s Men -
The same - the people - -
The same is what the heart -
That is the fame -
We can not keep -
The Eternity - is the of ,
But by the owl -
It ' s simple - after -
How let us be -
It ' s not something -
An Just innocent - -
ENDPOEM

The Light - - the voice of the Sun
Of the hay -

The morn of the sea -
No minute - to the hollow voice
Of the wrecks -
The Man - in the Be -
From the salt - the root -
The II ' s time -
To this the Purple -
The level of the vengeful -
And the sublime -
A And the mystery -
The hymns of the tempest ,
The rest of the Birds -
And the grief of the Past -
But , our share , and the cloud -
The little - idle man ,
Where the Weary - ,
The streaming isle -
The Sun - was the green -
For the snake - called -
The laughing stone -
As if it is not the whispered -
The worthy - that the World -
And a flower - and the red -
They scarcely leave -
And as you would not weep -
ENDPOEM

A Day

I ' m not the same in my eyes ,
The short thing that it is !
For I were not a lady
I ' ll take the place to see
My eyes but be .
ENDPOEM

Temperature 0.63

The pale patch of the shade
And the sun ' s face
When the wind were panting ,
And the sun of the lake
And the heavens were wet ,
And the rose - cry was to see ,
And they were free ,
And the sun was strong and the sky ,
As they came to me ,
And we found the sea - glass ,
And the ice was spread ,

And they were " five o ' er
" A fit to make it .
" My friend , I ' m glad to see ,
I ' ll tell the man .
And my soul is my heart .
He ' s dead and dead of your sight .
And you ' ll find my face .
ENDPOEM

Thy South
I ' m a little man like a man ,
And I ' m not the best to see ,
And if I ' m coming on the wall ,
I ' ll see you in the hill - -
But I ' ll take a moment .
I ' ll know the favourite thing ,
But the same thing I ' m glad .
ENDPOEM

In a little lap to her door
In a poor door ,
A little time , a thousand - flower
He seems to see the sweeter
That I saw a a little man
And a little woman ' s face .
The first night , and the moon ,
And the pure I died
To - day I said :
" Oh , I ' m glad to say ,
A day I ' ll see you . "
The bitterness of the new
I have watched her door
And soon and the night
My dream , I have done .
ENDPOEM

The daily
The That of the old - blooms ,
A vast and mighty ,
But not a fair - man -
And much for the man
I think of a man ' s self -
Among a house -
My face - - - to the world and me -
And I will not tell a man
For a - Enough ,
Who ' s a child - then -
I ' m willing to get my face -
So - is a - a - the -

And a - a man -
For I am so wise -
The love - - I am not at -
My face - to the mine -
Soul - be .
ENDPOEM

The think of the eastern forms is a tear
Of the best , that is a thing of either ;
And where we have just in the other ,
And like a very thing the State is ,
And loyal desire that I could not know
The rich air of Life .
I have a sigh to a droll face ;
That now I have stopped with a star
To see my mad winter .
And when they ' re a sober song ,
For the poor I Once that I
First men leave us with life ,
And fly in the tree with a priest :
And if I ' ll go to me ,
That I shall give me again .
Though I , my soul , I count
To see the lady ' s name .
So , I ' ve been a child to be
So many a little day ;
And the same is to me ,
I ' m always born to see .
The devil ' s a little game
The two of them ,
We ' ll drink your heart ' s gold ,
So thin , so and the dead .
I ' ve found him ;
But , we ' ll be a little man ,
And I ' ll find the poor -
But we ' ll go to the land .
ENDPOEM

Where the tree is changed ,
And the birds are like the wind
And the wind is dead !
The wind is there , their little sky
With the yellow pang ,
In the river ' s the snowy sun ,
The country sea has seen ,
And the sun ' s as he will take .
ENDPOEM
The Freedom
I am no more than my love ,

With all a woman ' s less a word ;
But if I have a dream of vain
That I will not be a child .
And I will not be so weak ;
I have grown a child , a woman ,
And I can never know
The truth of his child .
ENDPOEM

Down , The cloudy moon has seen the light
Of the marble , and the blue wind ,
The light of the sea , the sky , the trees ,
In the dark , the sea of the isle - - -
And the gloomy sky of the sun
Was thick to the sky , and the breeze
Is the shadow of the sun - -
The wind is the setting right ,
And the mountain - way of the sky ,
The deep of a blue - winding pines
Of the Powers ;
And where my head is the moon
Of the roof , the meadow ,
Like the wheel , the dead - - -
ENDPOEM

The sea was down - -
A shadow , the stream ,
And the windows were near
The strained - - -
How soon I saw it - - -
On the night - side - -
I gave no hope to speak ,
And the little old sea -
And I was lawless -
And when the sail - -
And ever , half - on -
A solitude -
I have heard a dreadful -
In a quiet - -
The infancy was to be
I heard that died -
And I had seen -
The eyes - -
The curtain was the first -
And so among the leaves -
And the sounds -
They know it was not the same -
And so was a breathing -
But all the waves knew -

And I had ate -
The blind -
And then I think -
That could be changed -
And all I had been so -
One - made a - an hour -
It was - - I had felt it -
And I went so -
To I ' ve kept the Table
Nor was the dear -
I was not - to -
A bear - a man -
And she ' s not -
She was not glad -
And I would give it for my sight -
And I saw a " Give -
" I ' ve heard an " I ' ll be -
I ' m just - I wish -
I ' m a girl - - I ' m - - I ' ll leave the high -
With her - - I ' m a - like the I -
I ' ve been a mile - -
And this I - - I am mad -
My bride - the voice -
Will I tell me -
I win the heart -
And I ' ll find -
I am the mother ' s -
ENDPOEM

66 and 63 are writing some GREAT stuff. In fact, I think these two are publishable as is:

In a little lap to her door
In a poor door ,
A little time , a thousand - flower
He seems to see the sweeter
That I saw a a little man
And a little woman ' s face .
The first night , and the moon ,
And the pure I died
To - day I said :
" Oh , I ' m glad to say ,
A day I ' ll see you . "
The bitterness of the new
I have watched her door
And soon and the night
My dream , I have done .
ENDPOEM

Where the tree is changed ,

And the birds are like the wind
And the wind is dead !
The wind is there , their little sky
With the yellow pang ,
In the river ' s the snowy sun ,
The country sea has seen ,
And the sun ' s as he will take .
ENDPOEM

Consolidated Next Directions

4/14/2016 2033

Capitalization bit
Post-processing
      -Punctuation space-removal
      -Duplicate words in a row removal
Stress encoding
Rhyme/syllable encoding
Primetext
      -Doesn't work
Better new-word embedding
      -Currently new words are randomly initialized—should stem/overlap/edit distance to guess
nearest word and initialize with that
      -There are both Levenshtein Distance and Porter Stemming implementations in lua
      -Should only update new words, not old ones
"In the style of"
50-dimension word vectors
200-dimension word vectors
GRU
Bias initialization
      -Final softmax can be initialized to give raw word count frequencies

Rhyming

      Create rhyme using
ebbrowning
whittier
service
keats
byron

      I also made one from Fromantics, which includes a lot of blank verse.

The Fromantics works better, just because it's large.

Qualitatively this does relatively well:

```
./word2vec -train rhyme_dirty_bow.txt -output bow_vectors61.bin -cbow 1 -size 6
-window 1 -negative 5 -min_count 0 -hs 0 -sample 0 -threads 20 -binary 1 -iter 1
5
./distance bow_vectors61.bin
```

Even better:

```
./word2vec -train rhyme_dirty_bow.txt -output bow_vectors61.bin -cbow 1 -size 6 -window 2 -negative
20 -min_count 0 -hs 0 -sample 0 -threads 20 -binary 1 -iter 15
```

Need a way to score!
This is going to be a mess. Return to it later.

Stress, glove, and Vocab

There are several issues here.
-The RNN backprops into glove. But stresses need to be protected from that.
-When glove doesn't contain a word, it's initialized in a funny way
-There's no great way to give it a word list
-No capitalization

So I want a way to input a vocab list, not update stress or capitalization, but extend vector to stress, sound, and capitalization. I want one modifiable stress bit (essentially for one-syllable words.)

Sound: first phoneme, last vowel phoneme, last two phonemes (4 dimensions) (from CMU dict), rhyme number (from ashely?)

So I probably want:
-Only words that are in glove and CMU
-6 dimensions for stress (# sylls + 5 stresses)
-4 dimensions for sound
-1 dimension for capitalization
-1 dimension for rhyme
-Fixed embedding
-Word list with everything capitalized

So, eleven additional dimensions

For the moment 4/16 2316, I have the sound and stress parsing the CMU dict correctly. Some minor cosmetics left but OK.

4/18/16 2003
Everything seems to compile and run. Did not implement capitalization (there's no way to, actually, without redoing the entire edifice of lower2idx.) Have 11 additional dimensions for sound, and 1 (unused) for cap. So 112 dimensions.

Let's train just on the poetry, for a while.

AND KEEP THE EMBEDDINGS FIXED.

This dramatically reduces the parameters but increases memory footprint (has to keep glove vectorsi n memory, 12% longer.) Go with batch size 40.

That didn't work, try 25.

25 Works, let it ride.

At 2500 iterations (epoch 3.19), very comparable to previous run of t_R (5.91 now, 5.95 then). So at least I didn't screw anything up.

At first checkpoint (epoch 6.38), a little worse than before, 5.66 vs 5.57

lm_lstm_epoch3.19_5.9128.t7
lm_lstm_epoch6.38_5.6569.t7
lm_lstm_epoch9.57_5.5846.t7
lm_lstm_epoch12.76_5.5119.t7
lm_lstm_epoch15.94_5.4755.t7
lm_lstm_epoch19.13_5.4425.t7
lm_lstm_epoch22.32_5.4191.t7
lm_lstm_epoch25.51_5.4074.t7
lm_lstm_epoch28.70_5.3893.t7
lm_lstm_epoch31.89_5.3794.t7
lm_lstm_epoch35.08_5.3675.t7
lm_lstm_epoch38.27_5.3595.t7
lm_lstm_epoch41.45_5.3518.t7
lm_lstm_epoch44.64_5.3496.t7
lm_lstm_epoch47.83_5.3462.t7
lm_lstm_epoch50.00_5.3348.t7

The last one ran to 25.54, where is was 5.1399 (c.f. 5.4074)

At tmnperature 0.7 it's really no good.

Run 75 more iterations

lm_lstm_epoch3.19_5.3314.t7
lm_lstm_epoch6.38_5.3113.t7
lm_lstm_epoch9.57_5.3147.t7
lm_lstm_epoch12.76_5.3065.t7
lm_lstm_epoch15.94_5.2983.t7
lm_lstm_epoch19.13_5.2815.t7
lm_lstm_epoch22.32_5.2740.t7
lm_lstm_epoch25.51_5.2800.t7
lm_lstm_epoch28.70_5.2758.t7

lm_lstm_epoch31.89_5.2726.t7
lm_lstm_epoch35.08_5.2629.t7
lm_lstm_epoch38.27_5.2627.t7
lm_lstm_epoch41.45_5.2611.t7
lm_lstm_epoch44.64_5.2611.t7
lm_lstm_epoch47.83_5.2546.t7
lm_lstm_epoch51.02_5.2557.t7
lm_lstm_epoch54.21_5.2524.t7
lm_lstm_epoch57.40_5.2500.t7

It was 4.95 at this juncture before, so running consistently about 0.3 worse. Hmm. Anecdotally 0.3 has been a lot.

Let's have a peek

Temperature 0.6: these are exemplars of what happens, not good ones!

The Song
 Who ; for all the of in the ,
 The train of the , and the the
 The cloud of the , the the and
 The music of the light ,
 And the the the and forms
 Of the within the place .
 And , but , from the , and eyes ,
 The sweet of of ,
 The children of the day
 With the and fall .
 When , shrine the , the , by the sun
 Of her form ,
 And the in the of of
 The steed of the .
 The duty of the the ,
 The light are the and fell ;
 The silver , the the the ,
 And over in the ,
 The dead of ' s the again
 The angel of ' s ' s .
 The One of with the space
 The light and the ;
 The cares of are before past
 The world of the ;
 How last of , and a , no ,
 The world of soul
 Of with the , and to the
 The finger of the .
 There is the ' s low ' s ,
 And , and strong ,
 On ' s the , must be than

The God is the
The world of the to its ,
They awful to the .
ENDPOEM
The

The King .

The Woman .

I have to the :
I to me
My if the
Who the the .

ENDPOEM

The sullen of so , and never been ,
But I have in the heart
And tender live to the and men ,
And I of I and wept
As breath to the , to see my tears ,
And mirth and childhood .

The height of heart with secret ,
While we in a , and there
Of and a , that the
His nature of the , nor child - - -
For in the beauty
Was is :
So all the and the
Of of , the the ,
And all as , and love of years
On the to ever ;
The sweet of with one ,
If take I to pass
The creed , and the of
The rocky breezes of .

ENDPOEM

It shows a strong, strong affinity for one-syllable words. Which is probably somewhat consistent with the syllable count of real poetry, but comes across oddly.   I'd rate the previous trainig (100 dims, not fixed) as MUCH better at T=0.6.

Increase the temp a little to 0.75

At 0.75, it's less obviously bad, just simply, never any good.

So, now I'll unfix it and run overnight to 125 epochs.

lm_lstm_epoch3.19_5.7171.t7
lm_lstm_epoch6.38_5.4669.t7
lm_lstm_epoch9.57_5.3527.t7
lm_lstm_epoch12.76_5.2685.t7
lm_lstm_epoch15.94_5.2104.t7
lm_lstm_epoch19.13_5.1666.t7
lm_lstm_epoch22.32_5.1331.t7
lm_lstm_epoch25.51_5.1115.t7
lm_lstm_epoch28.70_5.0864.t7
lm_lstm_epoch31.89_5.0778.t7
lm_lstm_epoch35.08_5.0508.t7
lm_lstm_epoch38.27_5.0377.t7
lm_lstm_epoch41.45_5.0272.t7
lm_lstm_epoch44.64_5.0162.t7
lm_lstm_epoch47.83_5.0151.t7
lm_lstm_epoch51.02_4.9971.t7
lm_lstm_epoch54.21_4.9883.t7
lm_lstm_epoch57.40_4.9847.t7
lm_lstm_epoch60.59_4.9833.t7
lm_lstm_epoch63.78_4.9773.t7
lm_lstm_epoch66.96_4.9700.t7
lm_lstm_epoch70.15_4.9607.t7
lm_lstm_epoch73.34_4.9589.t7
lm_lstm_epoch76.53_4.9558.t7

Oops, disk filled at epoch 76.

It's about 0.02 better than the last time I did this, with a 100 dimensional vector (floating)

I will not play ,
I would not come
For me I see away
With most long ;
That it would not now
That he can see
A love of bitter
If I do me :
It , in our own brave
I never might gains
For all I ?
ENDPOEM

What Be I . '
My voice is stay ,
I go to my sight ,

And at a loss ,
And a Mother with my hand
In I ' ll say .
And the red set in the sun
That many a blush
Until it would be born
In the House of his flame .
ENDPOEM

Song
The gathered

I saw the They and done ,
And the dirt of the childhood
That the wind is lay
My God , or the rest
Might the free of God .
And I had to be
The sun of the cup
An heart that lost
The old , the very eye
Of their own place .
So all great as a day ,
And I was loud
Of the world ' s way .
ENDPOEM

Oh
I have my pleasure , but my time
We showed my heart and I
But I should be my own draw
That tender both of one
And for the world I know all
My heart for the dream .
But if a simple voice I sate
My praise had not so true .
I too been to a world ' s light
With trouble of my own ,
And I ' d so from my faith I ' ll see
To how we ' ll guess the Man .
I would have face my little eyes ,
And , though they shall be away .
I ' m sure I ' re , and he was a way ,
If my guess I ' ve get .
On the to the truth that I have ,
I ' ve tell to understand .
The kiss is a partner ,
And I ' m at a war ;
I ' ve a brave - three - - - and I ' ll know

His mind to wrong .
It ' s little thing , my years I know ,
To One I ' ll be mine .
And if I ' ll never live so true .
Yet I ' ve do not what .
No shadow in the island ;
I ' re gone to be my face .
Ah , that ' s a little light to me ,
I ' m to forget ;
' Why just it ' s candles , the comfort
For the man has die .
ENDPOEM

The Sounds was the Slowly

No humble years for the face
As if had been -
And Like a false of the place
Is a Stop they -
I had , with the last - away -
And night of all between
The sake was Night -
I have a school to lost -
The Birds of the Noon of a Man -
The show of the Born -
And no many a sun -
The Faith of the story ,
And in the features -
The tiny notes
On the chase - of a golden
The Birds of the Flags -
The world of the House is guess
To life - that Would not -
Through a House was a statue -
No fat -
ENDPOEM

It clearly knows about rhyme now.

Let's try 0.6

" My child is how , " I said to see ,
And I could be my girl ; "
" But when the brain were playing me ,
And he was gathered and gone .
The lesser old and years was done
With this man ' s right to die .
" ' What never , ' said the sure , you ' m ,
I ' m planted , " the little eyes . '

He lay the hand of his face ,
And I , a little heap .
The years he seemed to the end ,
The song of wrath and day .
I ' ve loved a mother ' s heart ,
And if he had to die !
I ' ll know them to me all my own ,
And all my song were there .
ENDPOEM

There's some real meter in this! Not to mention see/me, gone/done, die/eyes.

Like many 0.6's, it loves to spit out endless CR/LF and endless punctuation.

` `When the cry , ` ` `My soul is the press ,
` `That the first will pay him , '
` `And the night is a soul to be :
` `And , at his breaking and a ease ,
` `In the eye and all it might give
` `To the field of the great hand . '
` `And the world might be ours ,
` `Not the heart of the best and praise ,
` `With the high pleasure of the grave ,
` `And the earth of the long - morrow .
` `And , the truth of being a grave ,
` `When that man of death and speak ,
` `And the sun is the time to feel
` `Can thou , who , to not give us this ,
` `For the day ' s peace , and not his eyes
` ` `Have thus , and the world ' s mind can turn ,
` `And the child , ' the word of my heart ,
` `For the great is that ' s love they die . '
` `ENDPOEM

Note that the nas drive seems to have been failing. I updated firmware, it seems better now.

I think I'd like to train the model hard on prose, without updating the embedding. Then continue from there, on the poetry, allowing updates.

I'm afraid that, as the last 11 vector entries will always be useless in prose, they will get "taken over" to express some additional dimension of meaning, and never recovered on the poetry side. But if I preserve them into the poetry, the system will at least have learned good grammar.

That's the thinking, anyway.

Also, write 17 copies of the poetry at the beginning to make vocab. Then require 35. Anything that appears in the poetry twice will appear in the vocab as long as it shows up at least once in the prose, which seems likely.

Link r17bF, containing 17 copies of poems and one of bitsF, to input. Generate vocab, data, glove. All went fine, did 300 iterations.

Link bitsF. Touch vocab and glove. Edit train to fix the embedding.

NOTE: the glove file says x111 but I believe it's really x112

th train.lua -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 5000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 35 -batch_size 25 -seq_length 64 -learning_rate 5.0e-5 -max_epochs ${epochs} -dropout 0.25 -seed 77 -learning_rate_decay_after 1 -learning_rate_decay 0.97

So it's dying on memory. Which is bizarre—it had plenty of memory for r17bF! Neither parameters count nor data size are meaningfully larger than successful past runs. Try running again.

Still fails?!?!?!?!?

WTF, it did iterations before?!

Do seq-length 50

Got further this time, still not there. Try 40. Got further, try 35. Now it's going.

It seems to be bottoming out around 5.5. This is progress relative to the first high-vocab attempt (bottomoed out right at 6.0) but way short of the low-vocab attempt (4.6 at the same epoch.)

I'll let it go today, and then float it tonight on the poems. Maybe that can make up the difference? Otherwise Friday I'll float it on the prose to start. Sigh.

Still making progress, but still only at 5.51 at epoch 3.46. Then disk filled.

Press on to poetry.

Linked romantics to input. Touched vocab. Touched glove. Removed data.

Floated the glove embeddings

Go!

Potential issue already: parameter count is the original fixed parameter count.

After 3300 iterations the outlook is not so great. It's not getting anywhere quickly. Hovering around 6.0, which is pretty sad. 4.4-4.6 is the goal.

lm_lstm_epoch1.74_5.9654.t7
lm_lstm_epoch3.49_5.9217.t7
lm_lstm_epoch5.23_5.8960.t7
lm_lstm_epoch6.97_5.8833.t7
lm_lstm_epoch8.71_5.8745.t7

lm_lstm_epoch10.46_5.8631.t7
lm_lstm_epoch12.20_5.8543.t7
lm_lstm_epoch13.94_5.8479.t7
lm_lstm_epoch15.68_5.8428.t7
lm_lstm_epoch17.43_5.8357.t7
lm_lstm_epoch19.17_5.8280.t7
lm_lstm_epoch20.91_5.8337.t7
lm_lstm_epoch22.66_5.8232.t7
lm_lstm_epoch24.40_5.8198.t7
lm_lstm_epoch26.14_5.8208.t7
lm_lstm_epoch27.88_5.8173.t7
lm_lstm_epoch29.63_5.8169.t7
lm_lstm_epoch31.37_5.8119.t7
lm_lstm_epoch33.11_5.8101.t7
lm_lstm_epoch34.86_5.8104.t7
lm_lstm_epoch36.60_5.8086.t7
lm_lstm_epoch38.34_5.8042.t7
lm_lstm_epoch40.08_5.8061.t7

Went nowhere overnight. v/lm_lstm_epoch40.08_5.8061.t7

OK, not nowhere, but somewhere crazy-slow.

Temperature 0.7 output is useless, very trippy and apostrophe-filled

Temperature 0.6 is also useless. Very repetitive and nonsensical.

So: let the 111 float.

Save vocab, glove; float glove; relink to bitsF.

-rw-r--r-- 1 andrew andrew 310427439 Apr 22 08:52 lm_lstm_epoch0.10_6.0147.t7
-rw-r--r-- 1 andrew andrew 310547463 Apr 22 09:24 lm_lstm_epoch0.19_5.7874.t7
-rw-r--r-- 1 andrew andrew 310667487 Apr 22 09:56 lm_lstm_epoch0.29_5.7169.t7
-rw-r--r-- 1 andrew andrew 310787511 Apr 22 10:28 lm_lstm_epoch0.38_5.6622.t7
-rw-r--r-- 1 andrew andrew 310907535 Apr 22 11:00 lm_lstm_epoch0.48_5.6242.t7
-rw-r--r-- 1 andrew andrew 311027559 Apr 22 11:32 lm_lstm_epoch0.58_5.5947.t7
-rw-r--r-- 1 andrew andrew 311147583 Apr 22 12:04 lm_lstm_epoch0.67_5.5700.t7
-rw-r--r-- 1 andrew andrew 311267607 Apr 22 12:36 lm_lstm_epoch0.77_5.5547.t7
-rw-r--r-- 1 andrew andrew 311387631 Apr 22 13:08 lm_lstm_epoch0.87_5.5382.t7
-rw-r--r-- 1 andrew andrew 311507655 Apr 22 13:40 lm_lstm_epoch0.96_5.5290.t7

It's running about 0.5 behind the previous, which had smaller vocab and 100 dimensions. Hasn't died yet. So keep going.

cv/lm_lstm_epoch1.92_5.4556.t7

Progress is painfully slow but not actually zero.

It's hard to imagine this ends up below 5.3.

Question: is it the size of the vocabulary, or the 111 dimensions? (An obvious experiment is in order: 111 dimensions, old vocab., floating.)

Is the vocab a killer because of how it treats UNK? (If so, worse losses might not necessarily mean worse performance.) Basically, it used to be, that for ~50% of the vocabulary (though maybe less than 5% of words…), it only had to output "UNK" to get it right, but now it has to get exactly which UNK it is.

However, examining output with temperature of 0.7 (from llm_lstm_epoch2.11_5.4502.t7), the output is a grade-A disaster. It has fallen in love with the word "framework". And "bees".

And he could emergencies framework , but assiduous brisket - - and ,
his face , and him at framework , and it fraid gather
his head frock .

" And bees appal ' end
' claim a framing .

banjo biz . "

" I ' m - - " gather
. gather
bees framework fraid
deflection .

" framework , " framework little framework , " it is bees ! "

" Good ! banjo , " bees fired . " I know I do fraid indiscriminate . "

Then he framework at his face .

answered . " We are import kind , and heterodox fraud ,
for I had made beware . "

" And framework , " he framework .

What a mess. (It does look a little like how things looked once when the vocab mapping got confused.)

The question of what's killing it needs to be resolved. So let's do a "normal" vocab 35 (around 32000 words) with 111 floating vectors. This will tell us if the 111 vectors are killing us, or the vocab. (OR: Just realized both batch and seq are short!)

If it's the 111, the most likely answer is that its normalization is so different that it can't initialize well, and it can clear the matrices out to small enough entries in the later parts.

Two ways to address: initialize the softmax layer to the log frequencies; normalize the new entries

better.

th train.lua -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 5000 -print_ev
ery 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 35 -batch_size 25 -
seq_length 35 -learning_rate 5.0e-5 -max_epochs ${epochs} -dropout 0.25 -seed 77
 -learning_rate_decay_after 1 -learning_rate_decay 0.97

It looks like typical values for 0.35 epochs is around 5.5 for small vocab sizes and small learn rates.
(At higher learn rates, it seems to come down to 4.5 by 0.3 epochs.)

lm_lstm_epoch0.10_5.7241.t7
lm_lstm_epoch0.19_5.3267.t7
lm_lstm_epoch0.29_5.1650.t7
lm_lstm_epoch0.38_5.0499.t7
lm_lstm_epoch0.48_4.9777.t7
lm_lstm_epoch0.58_4.9227.t7
lm_lstm_epoch0.67_4.8781.t7
lm_lstm_epoch0.77_4.8495.t7
lm_lstm_epoch0.77_4.8495.t7
lm_lstm_epoch0.87_4.8133.t7
lm_lstm_epoch0.96_4.7890.t7
lm_lstm_epoch1.06_4.7563.t7
lm_lstm_epoch1.15_4.7350.t7
lm_lstm_epoch1.25_4.7179.t7
lm_lstm_epoch1.35_4.6948.t7
lm_lstm_epoch1.44_4.6753.t7
lm_lstm_epoch1.54_4.6606.t7
lm_lstm_epoch1.63_4.6494.t7
lm_lstm_epoch1.73_4.6374.t7

Conclusion: 112 seems to help, which is absolutely counter to what you'd expect. It's coming down
nearly twice as fast…

Possible answer: that could well be the batching. As long as your averaging is accurate enough to take
the right step, adding more just slows you down per unit data used. Note the batching is 2x smaller
here...So if it hits a wall, consider changing batching and seq.

Last time, it wound down very slowly from 4.90 to 4.30.

This time I'd be surprised if it gets better than 4.5 without any changes to batch or seq.

OK, I killed it at 4.637, partly because I can't stand watching, partly because I expect it to be worse
with small batches, partly because I'll fill the disk if I free-run it like this.

Increase batches to 40, seq to 75.

th train.lua -init_from 'cv/lm_lstm_epoch1.73_4.6374.t7' -gpuid 0 -num_layers 3
-rnn_size 256 -eval_val_every 10000 -print_every 100 -word_level 1 -glove 1 -dat
a_dir ${mydir} -threshold 35 -batch_size 40 -seq_length 75 -learning_rate 2.0e-5

-max_epochs ${epochs} -dropout 0.25 -seed 77 -learning_rate_decay_after 5 -lear
ning_rate_decay 0.97

Solid leap, 0.07 in 2/3 of an epoch, though perhaps just continuing the above might have done the
same. Does suggest bottom is between 4.5 and 4.25. Let it roll all night.

lm_lstm_epoch0.66_4.5632.t7
lm_lstm_epoch1.32_4.5241.t7
lm_lstm_epoch1.98_4.4950.t7
lm_lstm_epoch2.64_4.4718.t7
lm_lstm_epoch3.30_4.4502.t7
lm_lstm_epoch3.95_4.4325.t7
lm_lstm_epoch4.61_4.4185.t7
lm_lstm_epoch5.27_4.4041.t7
lm_lstm_epoch5.93_4.3912.t7
lm_lstm_epoch6.59_4.3807.t7
lm_lstm_epoch7.25_4.3707.t7
lm_lstm_epoch7.91_4.3616.t7
lm_lstm_epoch8.57_4.3545.t7
lm_lstm_epoch9.23_4.3485.t7
lm_lstm_epoch9.89_4.3419.t7
lm_lstm_epoch10.55_4.3354.t7
lm_lstm_epoch11.20_4.3285.t7
lm_lstm_epoch11.86_4.3242.t7
lm_lstm_epoch12.52_4.3172.t7
lm_lstm_epoch13.18_4.3116.t7
lm_lstm_epoch13.84_4.3067.t7
lm_lstm_epoch14.50_4.3015.t7
lm_lstm_epoch15.16_4.2973.t7
lm_lstm_epoch15.82_4.2921.t7
lm_lstm_epoch16.48_4.2880.t7
lm_lstm_epoch17.14_4.2841.t7
lm_lstm_epoch17.80_4.2802.t7

Last time a similar trajectory did get down to 4.25, at 11 epochs, so that's not impossible here. Eleven
epochs will be about midnight tonight. But while we're actually in a very similar place compared to
last time, it's a little slower now than it was at the same point. I'd guess we bottom out around 4.35,
instead of 4.25. That would be too bad.

It's worth noting that, once again, the training is actually pessimistic.

At 8.57 epochs, I'm starting to think we bottom out close to 4.28. i'll let it roll overnight.

At 16.5 epochs, a fit to the data imply bottom out around 4.264. By 22 epochs it will be around 4.274,
which is sufficiently close. 22 epochs is in the morning.

Also: reading the 0.66 and 0.63 *before* going to 112 dimensions, there is already the occasional rhyme.

Bio for C. Room Maxwell

Room is found in the northern climes of Wellesley, Massachusetts, spending time reading, writing, and hibernating. A poetic autodidact, Room finds inspiration primarily in the romantic poets and Victorian novelists, but takes insight from poetry and prose whereever it can be found to inform an idiosyncratic poetic style. Other works include a collection of short stories, and a novel currently in search of a publisher.

Also, the Chinese iambs: http://arxiv.org/abs/1604.06274

Their tricks: main one is a bidirectional LSTM, augmented by an attention mechanism. The attention mechanism is basically another internal state that reweights some input. It's not obvious how the biLSTM really operates. A Chinese speaker reports it seems natural, when not told it's a machine. They require the first sentence to be input.

How to distinguish:
-Explicit discussion of meter
-Explicit discussion of particular neurons, particularly involved in rhyme, meter, capitalization, and ending
-No input needed, just some Markov equilibration time
-Prose training

Useful ideas:
-Bidirectional. It's easy to train both a forward and a backward model. It's less obvious how to combine the models. Could presumably sum the softmax outputs. Not sure whether you should normalize before summing. Probably.

-Global context vector

CAPITALS

So how to sneak in more capitals: when you see a capital, it counts as 12 occurrences, not as 1. Easily implemented in CharSplitMiniBatchLoader. Modified it. Will see how it goes tomorrow.

Sentences are (very) approximately 20 words long. There are approximately 1.4M sentences in the bitsF dataset. So there are as many initial words there as in the entire poetry corpus. So counting 12x is essentially the same as requiring 3x in the poetry corpus anywhere. This seems very acceptable. (It's still unfortunate, in that poetry will want to capitalize in the middle of a sentence. It's hard to get a verb there even three times, so splitting there won't really be an option.)

Back to the training. Just sampled lm_lstm_epoch17.14_4.2841.t7, output looks good. I think I'll quit at next checkpoint (17.8) and run the poetry overnight.

Relinked romantics.mixed. Removed data. Touched vocab and glove.

```
th train.lua -init_from 'cv/lm_lstm_epoch17.80_4.2802.t7' -gpuid 0 -num_layers 3 -rnn_size 256
-eval_val_every 5000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 35
-batch_size 40 -seq_length 75 -learning_rate 1.0e-5 -max_epochs ${epochs} -dropout 0.25 -seed 77
```

-learning_rate_decay_after 5 -learning_rate_decay 0.97

It will do about 7 epochs an hour, so by the morning, hopefully north of 50.
(Records: looks like the best for poems has been 4.4605. The two really good poems were 4.6084.)

"Interestingly", the 100-dim (on 4/15) seems to be about 0.2 better consistently

lm_lstm_epoch5.98_4.9640.t7
lm_lstm_epoch11.96_4.8554.t7
lm_lstm_epoch17.94_4.8120.t7
lm_lstm_epoch23.92_4.7879.t7
lm_lstm_epoch29.90_4.7701.t7
lm_lstm_epoch35.89_4.7574.t7
lm_lstm_epoch41.87_4.7473.t7
lm_lstm_epoch47.85_4.7401.t7
lm_lstm_epoch53.83_4.7338.t7
lm_lstm_epoch59.81_4.7292.t7

Looking very unlikely to make below 4.65. Worth noting: the decline rate is pretty similar to the learning decay rate! Learning rate was approximately 2e-6, bumped it back up to 5e-6, see if it converges any faster. Expect to see drop of 0.01-0.03 if bumping up is safe, 0.002-0.005 if not. (So naturally I'm going to see 0.007, by the Principle of Maximum Ambiguity).

th train.lua -init_from 'cv/lm_lstm_epoch59.81_4.7292.t7' -gpuid 0 -num_layers 3
 -rnn_size 256 -eval_val_every 5000 -print_every 100 -word_level 1 -glove 1 -dat
a_dir ${mydir} -threshold 35 -batch_size 40 -seq_length 75 -learning_rate 5.0e-6
 -max_epochs ${epochs} -dropout 0.25 -seed 77 -learning_rate_decay_after 5 -lear
ning_rate_decay 0.97

lm_lstm_epoch5.98_4.7179.t7

So that helped. Drive it EVEN FASTER.

th train.lua -init_from 'cv/lm_lstm_epoch5.98_4.7179.t7' -gpuid 0 -num_layers 3
-rnn_size 256 -eval_val_every 5000 -print_every 100 -word_level 1 -glove 1 -data
_dir ${mydir} -threshold 35 -batch_size 40 -seq_length 75 -learning_rate 1.0e-5
-max_epochs ${epochs} -dropout 0.25 -seed 77 -learning_rate_decay_after 5 -learn
ing_rate_decay 0.97

lm_lstm_epoch5.98_4.6986.t7
lm_lstm_epoch11.96_4.6833.t7
lm_lstm_epoch17.94_4.6712.t7
lm_lstm_epoch23.92_4.6628.t7
lm_lstm_epoch29.90_4.6553.t7
lm_lstm_epoch35.89_4.6497.t7
lm_lstm_epoch41.87_4.6450.t7
lm_lstm_epoch47.85_4.6415.t7
lm_lstm_epoch53.83_4.6384.t7
lm_lstm_epoch59.81_4.6362.t7

lm_lstm_epoch65.79_4.6341.t7

I'm going to jack the speed back to 1E-5 and see. It's 1E-6 right now

th train.lua -init_from 'cv/lm_lstm_epoch65.79_4.6341.t7' -gpuid 0 -num_layers 3 -rnn_size 256
-eval_val_every 5000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 35
-batch_size 40 -seq_length 75 -learning_rate 1.0e-5 -max_epochs ${epochs} -dropout 0.25 -seed 77
-learning_rate_decay_after 5 -learning_rate_decay 0.97

4/26 1930 NAS Drive misbehaving again.

Without changing the rate, I'd expect the next to improve by 0.002. I kind of doubt it will leap to 0.02
(4.614 next) with a 10x learning rate.

lm_lstm_epoch5.98_4.6234.t7

So I was wrong, the improvement did go up nearly that much on the first epoch set. It's almost the
same as at the beginning! (Suggests there might be 0.07 still to be squeezed?!?!)

The previous decay is really consistent with straight 0.97 learning weight decay?!!?!? If the next one
(epoch ~12) is 4.613 that would support that learning weight decay is most of the bound.

lm_lstm_epoch11.96_4.6147.t7

F*ing nailed it. Rate-bound!

Temperature 0.65; the first poems after equilibrating

The David
The man of the old man
It was a queen and the blue ,
And the old man knew the rest
And his wife was dead .
ENDPOEM

Then ' s the new of the ancient house ,
A little thing of blood ,
And in the frost and the sun
With their all the sea ,
And one of the world
They had been all their heart
When they had grown .
Then cart and the bags
And the tide of the mountains ,
And the sun - kindly ,
And passed with a mass
The hearth , and the sea
Of the garden and the showers ;
And ' twas the way in the golden .

' What is the brute of the son ? '
But the silent throne ,
The first with a cry ,
And the great - trod it ,
I found the moon and his feet ,
And the night , in the sight
Of the great light ,
The feather of a silver mountain - tree
I saw the child ,
And she was not the spring of the sun .
So the birds pearl the flowers
Of the sea - light ,
And I saw the wind on the sky ,
And all the breeze in the air
Of the weather ' s red .
The clouds of the flap ,
And the wind heard the light of the moon
Which was the old ;
The dawn of the shade
Was the light of the sun :
When the stars was white ,
In the cold , the river ' s song
The fears of all the night
Under the lonely beach
Of the cavern made .
Yet , through the wild - room
The sunset of the sea ,
The glory of the skies
The glances of the sun .
The night - wind , all the sky ,
The sunset of the hills ,
The Sea and the morn .
And the birds are dead ;
The silent stars of the old ,
Which the light of the fire
Was a golden stream .
ENDPOEM

The Eternal blissful

A song of the rout
Of the dusky sail ,
And the bare heart is fled ,
The water grew dead and fair .
O , who , on the end of man
Is the world , is a penny ,
And to the dark is death
In the eternal sun ,
The winds of the hills ,

But the sun ' s gale are free .
The winds of the sky is long
The nearer wind is gone ,
And the golden iron is broken .
The leafy sun of the sea
The forest - boughs ,
And the grass ' s stormy flowers
Which the sorrow seems to say
The light of the sky ;
That is the storm of the dead
Of a world , and a hope
And man of the Queen ,
The fox , and the light of the night
By the cold wood .
ENDPOEM

The Maid
The lingering and a stone
The white and old ,
The glory of the forest ,
The lonely leaves of song
Of the hills of the sea ,
And the dark ' s lip and flower ,
And the last of its mother
Is the best of the lonely
Of the world .
ENDPOEM

Temperature 0.8, after equilibrating:

Age
THE doubt - (which was at last , when you found)
I saw a maid suffered with a harmless cry ,
And all these things was shown to me :
And this a hope that I never should say ,
That how I know their great truth must be ,
Then then this I ' ll give my little arm ,
And there ' s there not a shame in the shade ,
My heart , my boy , you must love ;
It ' s love and talk ; and I ' ve been as you
Who will be glad at last , and I ' m sure
I would make ready , but you are a hero ;
Poor mound , as the front of us did say
He had a worst I could do , I am dead ,
A good in a new and noble , or
And that ' s what we say to me he , who
With the steps and my heart that ' s all to do ,
And love my hand ; and when hath you in my ,
What I ' m going to - night ? and so long

I ' m going far up to my majesty ,
And I think you might that have been a fool
And never felt it . I ' ll say that I ' m to make
To small a star at the door - like complacent :
And in the state , I ' ve had a gardens ,
And there are a toes being to me .
I ' ll give you my good friend , I am sure ,
While you ' ve went my pipe and all a day ,
And you ' ll go in my usual final Hill :
But it ' s such a sinner to do ,
It might be any night to pay .
And I am de children , and your wife ,
That I ' m quite on a fairy feather ,
And now you will bet
why you will do to die .
And who has forgotten in what I might ;
But at that I will have inclined
My world , and do be easier to make
Your own , for it ' s nothing to be closer .
I know the everything she will find
So remember , such a moral dream ,
And the most thought of God ' s life for life ,
And what it should have thought of them .
So of a man that Look in my heart
My heart is much , I may not understand .
It is a human being of all
With time , and will talk an will the prime ;
But so if I am like a happy wife ,
But you must be a whiskers to see
As such as heights I can - - where we are ,
That they are won , but on you and own ,
And , will not be expect to do this day .
I cannot get what you have become ,
But a semblance of the years I know .
ENDPOEM

Obseervation: meter seems better but rhyme still nonexistent.

So here we are: we've run 100 epochs but probably the learning rates are painfully small.

Plan: do 100 epochs overnight but start at 0.01 and decay by 0.99 every time. (Likely should have started at 2.0E-5 and decayed by 0.99 for 100 epochs.)

Some math: integrated epoch*learning rate training so far: 72E-5
If you start with learning rate of 2E-5 and decay of 0.99, you reach that after 42 epochs.
After 100 epochs you have nearly twice the integrated learning rate

So that seems the thing to do. Overnight would get you to where it is right now.

Push on a fastest. Start it at 2E-5. Decay 0.99 per epoch, after 5 epochs.

th train.lua -init_from 'cv/lm_lstm_epoch11.96_4.6147.t7' -gpuid 0 -num_layers 3 -rnn_size 256
-eval_val_every 5000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 35
-batch_size 40 -seq_length 75 -learning_rate 2.0e-5 -max_epochs ${epochs} -dropout 0.25 -seed 77
-learning_rate_decay_after 5 -learning_rate_decay 0.99

Watching the results, 2 is almost certainly too big. Dial back to 1. 1 is the tiniest hair better at first,
maybe a hair worse later on. No longer convinced 2 was a disaster but ok, I won't go back and forth on
this. Let it run overnight.

lm_lstm_epoch5.98_4.6053.t7
lm_lstm_epoch11.96_4.5972.t7
lm_lstm_epoch17.94_4.5898.t7
lm_lstm_epoch23.92_4.5841.t7
lm_lstm_epoch29.90_4.5778.t7
lm_lstm_epoch35.89_4.5727.t7
lm_lstm_epoch41.87_4.5675.t7
lm_lstm_epoch47.85_4.5633.t7
lm_lstm_epoch53.83_4.5595.t7
lm_lstm_epoch59.81_4.5564.t7

The decline is STILL CONSISTENT with learning decay rate.

So, continue, with 2E-5.

th train.lua -init_from 'cv/lm_lstm_epoch59.81_4.5564.t7' -gpuid 0 -num_layers 3 -rnn_size 256
-eval_val_every 5000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 35
-batch_size 40 -seq_length 75 -learning_rate 2.0e-5 -max_epochs ${epochs} -dropout 0.25 -seed 77
-learning_rate_decay_after 5 -learning_rate_decay 0.99

lm_lstm_epoch5.98_4.5461.t7
lm_lstm_epoch11.96_4.5368.t7
lm_lstm_epoch17.94_4.5284.t7
lm_lstm_epoch23.92_4.5224.t7
lm_lstm_epoch29.90_4.5146.t7
lm_lstm_epoch35.89_4.5091.t7
lm_lstm_epoch41.87_4.5035.t7
lm_lstm_epoch47.85_4.4989.t7
lm_lstm_epoch53.83_4.4950.t7
lm_lstm_epoch59.81_4.4916.t7
lm_lstm_epoch65.79_4.4876.t7
lm_lstm_epoch71.77_4.4837.t7
lm_lstm_epoch77.75_4.4810.t7
lm_lstm_epoch83.73_4.4796.t7
lm_lstm_epoch89.71_4.4756.t7

Looks like the best was 4.4656, we're in striking range now. Learning rate right now is 1E-5, which
gives about 0.003 per 6 epochs. We'd have to go about 30 more epochs (4.5 hours). At 5E-5 (0.99

decay) could do in 12 epochs.

It will be done around 10:30. More likely than not It will end up around 4.4775.

A fit suggests 4.4705 is the limiting value. It is hard to believe that 0.0051 will ever matter. Instead, let's generate some poems.

The first is always the first equilibrated poem

Temperature 0.8

The Sweet
sang I surely saw , and hear me !
' It is the hope of wrath and pain
Than you for a life we find or sing ,
Let the bones to make us to brave ,
You ' ll come and every day of day ,
And never bless the danger :
If , alone , ' tis we to make your horn .
When I will sit to be the Lord ,
And I will go nigh to my mate . '
 ENDPOEM

Temperature 0.7

Sonnet 1 2 - The Day that doth the little stones
Of all the dead , where they are met
By some green throne of snow or rock ,
And leaves above the green green .
And if the roar of the song shall find
The spirit of the lonely springs
Of those who love the merry land ,
A pleasant day to love .
 ENDPOEM

To The Soul
A ray of bird , as the menace of man ,
In the vast rain : the lilies , a silver ignorance ,
Should there be the same place that was the spring above .
The air is seen , and the moon has made its plough .
A shadow of the west is on the frowning sea ,
And the winds a cold sea on the stream to the sea ;
And the stars are bright , and the winds are out of wind ;
And there the smitten , the stars of the dim and brave ,
We are bent like the sun , which from the outward wind
Shall glow the dark , the loud leaves of the green sea .
And the moon is very wandering , and the march is done ;
The worst is the thunder that is the hills of heaven ;
But still the soft snow - wind is no more , a mist of the desert .

ENDPOEM

I think 0.7 is actually quite good.

Temperature 0.6

The Dead Of The The Angel
He who is the Lord of the King of his face ,
And the man and the God ,
And the great man ' s , and the King , who say ,
He shall be the wrong of the old ,
And the King ' s the fool , and his heart is brave ,
Or the sword of the Duke ,
And he ' s the house that his son is on .
But the king ' s to his life ,
And the sin of the world ,
And the head is not a man ,
And that is the best of the world ,
His neck , the the head of the world .
For him , who is , all that the man
Is a man , the land is there ,
And the man can not be a stranger .
He has a double place ,
The man and the world , and the craft
Of the sea , the mountain , and the moon .
A soul of the great , the stone ,
A a - - the true , my life is not ,
And it is a moment , and you
Is a man , and the same .
ENDPOEM

The Child of The Eternal
The human Prince , the good of the Lord , God is all
For the hope of God ;
And the man , who can put the place and the
,
I ' ll be a man , the best of all my yours .
And , as I thought , you saw him with the blood
Of a good , and thought , and all that I knew
And the stranger , and his hundred - forgotten ,
And I have passed the world , and I will rest .
ENDPOEM

The Rose Of The Sonnet
The other make no more of the air ,
And the child is sown , and I can not go ,
The love is a little , and it is mine
And I am far in the glory of the sky .
ENDPOEM

From the sea the ruin of the sea .
A sail of light , the moon and the sky ,
The loud , wild lightning , the deep and light ,
The wanderer , the gray , he stood , and went
The market - place and the human day .
The gray and gray - hearted , the dreary day ,
And the old , dark , cold , old , old , mild man ,
And the change of those who proud and soft ,
The very heart of all his hands were near ,
And the white heart of the men might be
With the naked face of the woods .
The river in the middle of the earth ,
The fire and the waters , the rest of all ,
And the star of the earth . - - - But the same
Of the dead is the great man ' s sake ;
And a little man , and his own son ,
Is the good and the master of God .
For the world is a man with the horn
Of the great and men , the The old , no man ,
With a thousand words , and the cloudy ,
And the old man ' s work , and I are white .
ENDPOEM

Mithrandir!!!!!!!

0.6 is good but I think 0.7 is probably better. 0.65 might be good?

In the meantime, let's do capitals.

I want anything that is capitalized twice. Set the capmult to 100!

The issue is I want to dial the vocab back down to ~30000. So 35 may not longer work.

Let's do it. Put in Fromantic and rerun. Increase learn rate.

Vocab size is 154319 with requirement of 35. This means that ~120000 unique words begin sentences in the corpus. ! Also the parsemebeddingfile crashed. Not sure if that's related or not.

Let's dial down capmult to 12, which will require being capitalized 3 times. Vocab is 75000! Pretty big. No longer crashes in the same place—crashes in new place. Dial down capmult to 8, will require 5 occurrences. (If I can get it down to 50k or so, then I can start raising the 35.)

Now vocab is 59400. Raise to vocab threshhold to 47. Crash is somewhere new again, "Loss is NaN"

Vocab is 51848. Sigh. But it seems not to crash. Will push on with it.

Now replace Fromantic with bitsF. In the past, this is where memory ran out.

th train.lua -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 5000 -print_ev

ery 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold ${thold} -batch_siz
e 25 -seq_length 35 -learning_rate 9.0e-5 -max_epochs ${epochs} -dropout 0.25 -s
eed 77 -learning_rate_decay_after 1 -learning_rate_decay 0.97

So we'll see if we get anywhere. In the past, it's just wandered around with larger vocabs.

Got NaN loss after 300 iterations.
Got NaN loss after <100 iterations
Maybe step size is too big? Drop it back down to 5E-5. (Once again: how is this not deterministic?
Maybe the random matrices don't share the seed?) Increase batch size.

Wokred ok this time. Need to increase checkpoint outputs or it will fill disk overnight. Make it 20k.

For the record: last training had 5.0499 at 20000 iterations (0.384 epochs.) Comparisons so far don't
look promising.

lm_lstm_epoch0.62_5.8858.t7
lm_lstm_epoch1.23_5.8145.t7
lm_lstm_epoch1.85_5.7850.t7

Likely to bottom out above 5.7. Killing it.

It seems there's a phase transition in there somewhere, between 32881 and 51848. Two theories: vocab
index uses 15 bits (32768); parameters index uses 24 bits (16.8M)

Let's try a vocabulary of 41500.

(There's a run with 16.5M parameters and 42426 which looks like it may have been aborted. Not a
good sign.) Raise threshhold to 63

The more I think about it, the more I neither either to lower the entire text, or tear apart the caps.

63 didn't get it low enough (close), make it 71.

Still have some room at 71, let's try 65.

It went nowhere. Wandering around at 5.8 all day:

lm_lstm_epoch0.62_5.8858.t7
lm_lstm_epoch1.23_5.8145.t7
lm_lstm_epoch1.85_5.7850.t7
lm_lstm_epoch0.62_6.1111.t7
lm_lstm_epoch1.23_6.0594.t7
lm_lstm_epoch1.85_5.9468.t7
lm_lstm_epoch2.46_5.8775.t7
lm_lstm_epoch3.08_5.8538.t7
lm_lstm_epoch3.69_5.8350.t7

OK, the zillion dollar question: when does it work????!??!?

Let's raise threshhold to ~37k: try threshhold 109 (too few, trains just fine)
Threshhold 79 gives 37332, let's try that

Thresh	V seen	V-estimated
35	59400	
47	51848	
63	43167	
65	40827	
71	39965	
73	38053	
79	37332	
81	35730	
87	35053	
89	33658	
95		33000
97		31700
101		31500
107		31000
109	30068	

Recall: 32881 is fine.

Threshhold 109, 30068 vocab, looks just fine in terms of convergence. It's going.

Threshhold 79, 37332: going absolutely nowhere.   No better than 51848.

Let's try 89, 33658. Barely more than the 32881 that runs just fine. This one seems to converge basically fine.

OK, try 87. 35053. It's converging. A little slow, but converging.

So, 81 next. 35730. Not converging.

So I think the clear conclusion is to go 87 and see if the caps matter. Somewhere between 35053 and 35730, the convergence stops.

th train.lua -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 20000 -print_e
very 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold ${thold} -batch_si
ze 40 -seq_length 35 -learning_rate 5.0e-5 -max_epochs ${epochs} -dropout 0.25 -
seed 77 -learning_rate_decay_after 1 -learning_rate_decay 0.97

lm_lstm_epoch0.58_5.3355.t7
lm_lstm_epoch1.17_5.1845.t7
lm_lstm_epoch1.75_5.1076.t7

The training is unquestionably slower than last time. It's getting about half the bang-per-epoch. That suggests a bottom around 4.8-4.9. Pretty unpleasant number.

Note:
Word rank 30000 is "zero-sum"; nearby words are overtook and sauntered. They occur at approximately 1 in a million. Words that occur at 1 per 60k (the capitals rate) are the 5000 top words: snapped, peer, faster. Words that occur at 1 per 8k (the old capitals rate) are the top 1000 (up-goer 5.) Three per million is about 18000: diver, bribery, dutifully. Lose about 3% of the vocabulary (that's a lot!)

My guess is that 33k consists of around 22k lower words and 5.5k+5.5k upper words, or 27.5k lower words and 5.5k upper words. Words around 27k, from
http://corpus.byu.edu/coca/files/100k_samples.txt:

```
27400   486     351     thong   thong   nn1
27420   485     431     curved  curve   vvd
27440   484     466     engulfed        engulf  vvn
27460   484     418     rebuilt rebuilt jj
27480   483     457     plotted plot    vvd
27500   483     254     sow     sow     nn1
27520   482     339     pretend pretend jj
27540   481     342     nurture nurture nn1
27560   480     431     screened        screen  vvd
27580   479     453     rearranging     rearrange       vvg
27600   479     360     anti-communist  anti-communist  jj
```

It's bottoming out around 4.92.

lm_lstm_epoch0.58_5.3355.t7
lm_lstm_epoch1.17_5.1845.t7
lm_lstm_epoch1.75_5.1076.t7
lm_lstm_epoch2.33_5.0582.t7
lm_lstm_epoch2.91_5.0272.t7
lm_lstm_epoch3.50_4.9951.t7
lm_lstm_epoch4.08_4.9764.t7
lm_lstm_epoch4.66_4.9625.t7
lm_lstm_epoch5.25_4.9457.t7
lm_lstm_epoch5.83_4.9353.t7
lm_lstm_epoch6.41_4.9244.t7
lm_lstm_epoch7.00_4.9212.t7

On the prose.

The prose is not so good (temperature 0.6):

 " What is all Buskirk ? " Chaldicotes asked .

 " I am not so . "

" I should have been Busy and Busy . "

" No , I don ' Busy . "

" Oh ! I ' ll be as far as I can get it . "
Busy . " It ' s a good deal of anything , " he said . " That ' s a good man ,
and ' s not a good deal of plants . "

" What may Busy ? "

" No , I don ' Cam . "

" Oh , I ' m not glad , " said Mr . m . " I don ' Compose , " he said , " I ' m going in
a place . "

" Busy , " said Mrs . gasped Busy . " I ' m going in a York . "

" I say , " said Busy , " I ' ll go and see it , Compose Cam . "

" I know , " said Cimmerian , " and I know it is , I am , she ' s a great fello . "

" I did not know , " answered CAUSE .

So right now we have two mysteries:
>   -Why is this so much worse? Missing 3% of vocabulary is something, but this much?
>   -Why is there a phase transition in training convergence?

Push the learning rate up 4x and see what happens overnight. I'm not hopeful.

Judging by the losses during iterations, that didn't help.

lm_lstm_epoch0.58_4.9232.t7
lm_lstm_epoch1.17_4.9289.t7
lm_lstm_epoch1.75_4.9330.t7
lm_lstm_epoch2.33_4.9487.t7
lm_lstm_epoch2.91_4.9695.t7

Nope.

Let's take stock. From last consolidated list, plus new information:

Consolidated Next Directions

4/14/2016 2033

Convergence phase transition
High loss convergence with capitals
Lack of improvement with sound information
Coherence: BiLSTM? Subject vector?
~~Capitalization bit~~
Post-processing
       -Punctuation space-removal
       -Duplicate words in a row removal
~~Stress encoding~~
~~Rhyme/syllable encoding~~
Primetext
       -Doesn't work
Better new-word embedding
       -Currently new words are randomly initialized—should stem/overlap/edit distance to guess
nearest word and initialize with that
       -There are both Levenshtein Distance and Porter Stemming implementations in lua
       -Should only update new words, not old ones
       -This matters more than it seems! It's half the parameters. If I can have a good start, then I can
              train with batches and sequences 2x larger.
"In the style of"
50-dimension word vectors
200-dimension word vectors
GRU
Bias initialization
       -Final softmax can be initialized to give raw word count frequencies
Added 1/17: Fixed sound vectors, better capitals handling.

I think the convergence phase transition is the most urgent problem. I hope that insight into that gives insight into the capitals problem.

Rapid progress will best be made by lowercasing the poetry and prose datasets and recovering a vocabulary of 32881. I can train and generate from that while tearing apart the rest.

Disk space needs to be rearranged.

OK. Eliminate capsmult. Use bitsF. Raise learning rate, was 5E-5 with decay of 0.97, set it to 1E-4 with decay of 0.97.

Threshhhold of 30 leads to 32620, which I love (just less than 32768).

Replaced with bitsF, touched vocab, touched glove.

th train.lua -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 20000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold ${thold} -batch_size 40 -seq_length 35 -learning_rate 1.0e-4 -max_epochs ${epochs} -dropout 0.25 -seed 77 -learning_rate_decay_after 1 -learning_rate_decay 0.98

Insight! (Maybe): why training matters
        Many words (many!) are not found in glove. "loaded" for instance (I guess, because it's the tense / participle and not the base word.) These gain tremendously from training.

lm_lstm_epoch0.62_4.7035.t7
lm_lstm_epoch1.23_4.5335.t7
lm_lstm_epoch1.85_4.4574.t7
lm_lstm_epoch2.46_4.4075.t7
lm_lstm_epoch3.08_4.3760.t7
lm_lstm_epoch3.69_4.3634.t7
lm_lstm_epoch4.31_4.3545.t7
lm_lstm_epoch4.92_4.3571.t7
lm_lstm_epoch5.54_4.3659.t7
lm_lstm_epoch6.15_4.3804.t7
lm_lstm_epoch6.77_4.3957.t7

Coming down faster than before. Was 4.91 at 0.62, 4.71 at 1.23. That sounds like it's slowing down relatively, but not: when the previous one was at 4.70, 0.6 epochs later it was at 4.61.   So at the same value of loss, this one is still coming down faster. Training is still a hair optimistic.

But it looks like it might be bottoming out around 4.5…(when epoch was near 1.23)

OK, last time, once it hit 4.5, it went in steps of about 0.01 per 0.6 epochs for another 15 epochs.

Whoa! 1.23 to 1.85...bottom looks as little more like 4.37 but with only 3 points it's hard to extrapolate. Last time it was at epoch 5 before getting to 4.45. Also, each epoch took twice as long, IIRC. Damn, look at 2.46. Was almost epoch 8 last time.

The post-processor:
-Split poems into different files
-Capitalize every word in the first line
-Capitalize every first word in a lint
-Capitalize every word following a period
-Remove spaces around punctuation.

OK, wrote postCap.sh, it does this. Also capitalizes I. Except for the first word following a period.

Last time I quit at 4.28. The trend is that I would reach that in about 3 more epochs, i.e. about 1:30 tonight. One more epoch probably gets to 4.35.

Might as well go for broke and let it train on poetry tomorrow.

So it went back up. Steps were probably too big. Dial down to 3E-5 and run.

lm_lstm_epoch0.62_4.3448.t7
lm_lstm_epoch1.23_4.3465.t7
lm_lstm_epoch1.85_4.3513.t7
lm_lstm_epoch2.46_4.3536.t7
lm_lstm_epoch3.08_4.3575.t7
lm_lstm_epoch3.69_4.3605.t7
lm_lstm_epoch4.31_4.3619.t7

Sigh. Dial down to 1E-5, start from 0.62.

th train.lua -init_from cv/lm_lstm_epoch0.62_4.3448.t7 -gpuid 0 -num_layers 3 -r
nn_size 256 -eval_val_every 20000 -print_every 100 -word_level 1 -glove 1 -data_
dir ${mydir} -threshold ${thold} -batch_size 40 -seq_length 60 -learning_rate 1.
0e-5 -max_epochs ${epochs} -dropout 0.25 -seed 77 -learning_rate_decay_after 1 -
learning_rate_decay 0.98

In the meantime: preserved a modified version of the glove100 that's completely vanilla in the code
(except for the orthogonal bit). Will want to try GRU against LSTM in it.

FWIW it's apparent that 1E-5 isn't making quick progress. But I'd settle for slow progress :(

lm_lstm_epoch0.62_4.3434.t7

I got slow progress. Actually this is roughly consistent with last time. (It's ~2x lower learning rate, and
about 2x lower gain per 0.6 epochs.)

One oddity: that last one had seq=75! I'm going to try 70 (this has a few percent more pars). 70 didn't
work. (Also: that took almost 20 minutes to clone?!)

OK, 60 worked. Starting again from 4.3448

lm_lstm_epoch0.62_4.3434.t7
lm_lstm_epoch1.05_4.3320.t7
lm_lstm_epoch2.11_4.3239.t7
lm_lstm_epoch3.16_4.3177.t7
lm_lstm_epoch4.22_4.3121.t7
lm_lstm_epoch5.27_4.3087.t7
lm_lstm_epoch6.33_4.3053.t7
lm_lstm_epoch7.38_4.3023.t7
lm_lstm_epoch8.44_4.3004.t7

At half the learning rate, this is actually faster progress at 4.34 than before. (Remember, the main
difference this time is the caps.)

It's close to 3 hours for 0.6 epochs (5 hours per epoch).

At 4.3004, this is pretty much consistent with the previous rate at 4.30, given that it has half the
learning rate.

Let's just go for it with the poetry from here.
Relink romantic lower, touch vocab and glove, move data. 50 epochs

th train.lua -init_from cv/lm_lstm_epoch8.44_4.3004.t7 -gpuid 0 -num_layers 3 -r
nn_size 256 -eval_val_every 20000 -print_every 100 -word_level 1 -glove 1 -data_
dir ${mydir} -threshold ${thold} -batch_size 40 -seq_length 60 -learning_rate 2.
0e-5 -max_epochs ${epochs} -dropout 0.25 -seed 77 -learning_rate_decay_after 1 -
learning_rate_decay 0.99

lm_lstm_epoch19.14_4.7421.t7
lm_lstm_epoch38.28_4.6577.t7
lm_lstm_epoch50.00_4.6284.t7

Fire off 50 more epochs, restarting at 2E-5. The lower case did solve the "And" problem.

lm_lstm_epoch19.14_4.5793.t7
lm_lstm_epoch38.28_4.5506.t7
lm_lstm_epoch50.00_4.5382.t7

Best ever was 4.4656. Now I'm going to go up to seq_length of 75 and decrease the batches.

th train.lua -init_from cv/lm_lstm_epoch50.00_4.5382.t7 -gpuid 0 -num_layers 3 -rnn_size 256
-eval_val_every 20000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold ${thold}
-batch_size 30 -seq_length 75 -learning_rate 2.0e-5 -max_epochs ${epochs} -dropout 0.25 -seed 77
-learning_rate_decay_after 1 -learning_rate_decay 0.99

lm_lstm_epoch17.94_4.4441.t7

BOOM!

Keep on...want it to run all night, so I'll set it to 60 epochs from now.

epochs=60

echo "Starting 40/35"
th train.lua -init_from cv/lm_lstm_epoch17.94_4.441.t7 -gpuid 0 -num_layers 3 -rnn_size 256
-eval_val_every 20000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold ${thold}
-batch_size 30 -seq_length 75 -learning_rate 2.0e-5 -max_epochs ${epochs} -dropout 0.25 -seed 77
-learning_rate_decay_after 1 -learning_rate_decay 0.99

Oops, meant to run overnight but typo in filename.

lm_lstm_epoch17.94_4.4356.t7
lm_lstm_epoch35.87_4.4269.t7
lm_lstm_epoch53.81_4.4204.t7
lm_lstm_epoch60.00_4.4182.t7

Some 0.8:

But I Lost A Thing The Lady
 I knew it was so sweet in pleasure
 The father of a hearth-
 Then lay the name of hope, and I said,
'Of a woman retired for him
 Or to swiftly deem it a man.
 But as a tale I get to pass
 To evil in the place
 I'll separate the whole and bestow
 The bounds of the day.
ENDPOEM

 The Work Of The Wrote
 The old.
 The vain of the night-side
 The's the sea-ship's feet
 Upon its burning fire,
 And the light of the cliffs
 The west is come ;
 And in the river's feet,
 A little and the wood,
 The gown is the old,
 And the snow or a pine.
 The stranger is a thing
 And a path, a adriatic,
 It's a tree, to the old.
 When the blue sudden--
 The lovely is in the air
 That I'm sure for ourselves.
 We'll come to the land,
 And the noise of the world.
ENDPOEM

 The Shows Of The All That Is The Old

 The sunshine of the old-
 The second-the siren-
 The schemes of the imagination
 Of the other,
 The host is not the doubtful light-
 Of the woodland--
 For a man, who never meet
 We am going away-
 The work of the dark-
 And the winter-in the mountain
 Her flowers--

And the sound-that is the low-
It's the square-the youth-
But time's men-
The same-the people--
The same is what the heart-
That is the fame-
We can not keep-
The eternity-is the of,
But by the owl-
It's simple-after-
How let us be-
It's not something-
An just innocent--
ENDPOEM

The Light-- The Voice Of The Sun
Of the hay-
The morn of the sea-
No minute-to the hollow voice
Of the wrecks-
The man-in the be-
From the salt-the root-
The ii's time-
To this the purple-
The level of the vengeful-
And the sublime-
A and the mystery-
The hymns of the tempest,
The rest of the birds-
And the grief of the past-
But, our share, and the cloud-
The little-idle man,
Where the weary -,
The streaming isle-
The sun-was the green-
For the snake-called-
The laughing stone-
As if it is not the whispered-
The worthy-that the world-
And a flower-and the red-
They scarcely leave-
And as you would not weep-
ENDPOEM

The Me Of The Lord-
And the monster-
The star it is not the floor-
The red-in the dark-
From the lying-the sun-

Then-and not at the old-
They know that, if the coming-
The burning ocean-
While in the hill-of the village-
The same-when, ere the light-
And will, that once
Made-to its grave-
The soul of the star-
And still the little-
For the humble of the song-
And bring a-to the grave
His secret-
The night-in the xxvi
And the fields-the world-
The the thousand-in the wind-
And all the evil-
But the afraid-morn-
And on the full-
The how who can kiss-
No depth-
No reel-the danced-
And the woods of a seer
That is the tho-
He-is a little alone-
And-for the up-
And's the door-is a gold-
The last-a living-
No more.
ENDPOEM

I'd say this is the first time that 0.8 is largely coherent.

0.6 is over-fixated, much too cold.

A hair below 0.8 might be a tiny bit better. It's never very surprising-in-a-good way, just occasionally bizarre.

Also, it seems I lost the original file of this. At least in the postCap version here, it seems very fond of dashes in place of commas.

Some more 0.8, from 4.41:

I Have To See The Stars
And as the wind may wake with me,
The summer boy is sleeping now;
And when trees are changed,
How far and still the sun is gone
With wild soft roses o'er the sea!
ENDPOEM

The Lady's Voice
 She
 Oh, I see thee from the beloved,
 And every night is gone;
 And, with its sweet, she doth clasp
 The grim and dim sweet air,
 With joy and light, with her cloud's dew,
 The tears of flowers are gone!
ENDPOEM
 The Wedding Of The Sea
 A hermit's head is tender,
 And there is no more maid;
 Like a pink heart, I know
 That man was I,
 Till my six days have gone.
 Men I have loved my lord
 As I know what I could,
 It thought the lot and know
 What to me what who went
 Alone in the mirror of the sea?
 We have a dog for me for me
 Of the days that's cold and long;
 But oh, a single man to keep
 That moon-listening from the sky.
ENDPOEM

 This Is The Faith; It Is Not Thought

 The best of the things the trees did-
 The highest of the world-
 It must be the end to be.
 And yet to set the slaves
 Of human armour!
ENDPOEM

 The Daughter Of The Quickened Sun

 I hear the mountains, the soft
 Of the sand, the lark,
 The sea plain, and the moon
 With great and low-shepherd's horn.
ENDPOEM

 On The Altar Of The Wood

 . .. My love I read
 The poet,----
 Hast thou awake--

Here---------------
A leaf--------------
Though a thought upon the sight
Of the cabin-house---
Seen a very charm---
When this is, to make it--
Since, that veil its life.
I'd look as death--
Evening----
Whom that ask so long?
---Oh! come?
ENDPOEM

Oh How One
When the cottage and the spring
They came to the land and the forest,
At the wind-tree wind;
And the summer-breeze
The sun had sung and showed and spring
In a mountain-nod.
A goodly, creature's, a little star,
A gaunt and flint;
And there was a souls on a grave
In the happier night.
Choose to wait---o'er the sky and birds,
Dreams of dew-glowing flame,
And the eglantine, that's the proud
Of the living and the good.
The smiles that dare to follow,
And the moon and bed,
The sky-forehead in the sky,
The shadow of a shade
From the harvest of the night.
Oh, what's when I'm alive?
If you're gone in my place,
You've got so much to sleep,
But it's a little life,
And I'll see the web
Comes out of my track.
There's nothing on a man that's sad,
For I think they'd have the way
To leave your shelter;
There's perhaps you're free at last.
To be most leisure
I will not stay in vain,
And think of the good sun,
And say a single spring.
ENDPOEM
Sonnet Written

On grace of truth from the young lord
Of that wild world of mine,
All blue and silver;
And, still this I am as that,
With a child's god,
And the cheek of god,
Whose eyes are full of snow;
And god's heart
My heavy arms are warm,
And if he die
In the faith of god, I know
Who shall die.
ENDPOEM

About one poem of seven is log-worthy. That's pretty high actually.

Temperature one:

Watch,-- Loud Curl Of Breath!
Of these old hills in sunlight's vale,
Plight with the stars of winter are,
And early crimson waters wander
Along the mountains---with a star it seem
Silver-tender from their busy heart.
Sign of showers, the glow therefore there
The safe-wide thoughts that turn to earth in,
Like the evening rays on all they sing.
For moon, on the mountain's shepherd's throne,
All the lies at last the lifted trees
The pensive moon, out of the stream,
Were wild, murmur, low as the wind
That sees, a amazed faster of the guilt
Which dread the height of love's zeal flushed,
Both not solace, but delivered on the horrors--
Him of the dead, whose glimpse is all to praise,
A elements that many an busily flutter
May look upon the sun and heaven, where
The tempest beat with the crowd and their breath
Of but shadow and silence, the wine
To her forget that blighted altogether,
Almost of dark arch and an wrath; and there
Hours aim'd on death, no plain not could they feel
On their own powers, but and the wild rose
Run in-work murmur how they found the moon
Mounts on the charge, nor tempted to get up.
For his being) ah! that cherish-pleading doubt
That to his rolls that was, rejoicing spread,
Made foreign depths of his immortal soul,
Though birds more sure than this, such as might long

Ponder the winters of his height the earth.
Xix
They met its words was soft and sad, and music
-- That learnt no golden murmur, and they awoke.
Well, I need still, the poor man works forever,
My genial queen, and-- who should cherish for my glory.
Gin in his infant land
Green glory renew'd the heart of the fair,
When images of infant season sought;
Upon the everlasting of high oar,
Through time on walter's speaking bed,
The holy! cold-flew greet her and too fair
My tables on gladness, and vanish'd by bliss,
And rapture, bending on the wept drums,
Sit whispered, when he shows within his shape,
Which, departed, and drew the faded maker
Of pain, and scarcely passion, turning wet.
Ah! youth, not when such thoughts replied the mortal name
That dust not shut into life on earth and death.
ENDPOEM

The Leaves
None, where awake those neat worth light
Were silent from the past,
A splendour of the air, and just
Give where its own our own.
Sweet friendship, glad to met thee,
Forgiven to mark their mate!
Long glass their brief lovely virtue
Visible among her doom.
The song awake that beauty tells:
The heart of sunshine befell,
Thou dream on every new eternal way!
And thou, why universal joy
Wise mercy canst tell, and renew
Thy true power for thy sad maid,
In grave and little gloom?
Plenty of fear and laughter will,
Wake me to thy noble ear;
And drop thee from the world for spring,
For thy voice is thy own tongue.
ENDPOEM

Cast
A summer flower
We called her brother winter
Three years in times,
That lies up one vast stone,
From the meadow half-bounds,

With footsteps frost, about
A glow of human sorrow
That she seems in his sight.
O friend! and when she would die
Honest bird had added,
In midnight friendship grew fair,
I put her in my arms.
Full sir didst lover's neighbour
At her wheels, and all men
She gave no heart & en, is grown,
` Dear joy-- now his door shall be;
How, if the trembling oath
Is thine, all looks your heart!
` Ay, she's demon, let the flood
Bloom up to and fro by
The golden skies, and amber,
Her van having ragged birds,--
The april-fish to drove your setting
With the change crept out.
'But why I love in memory
Name, my moonlight's ear,
And lies an angel's labyrinth,
Why you are learned to see,
By the good-kissed of the dawn,
By all that is spoken.
Viii.
To take too world-place to-night
When ride a airy tale
With a stranger, what stopped sir,
Come, bounding!)
" Tis act'd soft wine,'said he,
'At least I speak,
Or, with a score of third years
There in vain mary aid'd,
Ye've got me too quick for candle .'
And once I heard her say,
And the health rob was to wear
Speech against his sorry race,
And (vague of many, his greater)
And pick'd upon his phrase,
Indeed, in white and dark of new,
, Then with opinion of the youth
In reason's eyes on her,
Told the whole, false breeze since gold
The country of the dead,
Then what he loved, him dear, so sung,
And way from our ripe lands.
ENDPOEM

About one poem in three, at temperature one. !
This is not quite a complete poem, but:

To The Soft Language Of Spring
 The spirit called me with a blessed charm
 As cool as sun-lingering lie behind.
 The moss and stare of black trees and beneath
 The trumpet-wings a noble armour holds.
 Swift decay of hushed every war,
 I in my dear and poor kiss, and thy star:
 Ah, I too well must stare into politics;
 And for the infant of second choice?
 In vain we feel not, let me have a song,
 But are known not in the goal to please.
 Lo, I think this simple leaf, o golden greenness!
 Iron-ways---yet greet thee! I have found thee making
 Heaven, when I laugh when you lie down beside
 Your finger, and thy eyes and fling thee the feet
 That is alike for a world to vulgar thee
 And see beneath the desolate clothed-rising summer,
 And we might pass in sun and wintry rs's court,
 Happy, loving heart nor me! here thou lovest,
 Let might rest and see my full lined sunshine drop
 Blast for the watch-board, that so near thee came
 In dawn to thyself: when we were fled, if the moon.
 Is there not thus bad, as let our walls depart
 Lift the door to-night, the past to among the ripe?
 Then hated to be dead-- a once ere I saw.
 Alas! if wise it slide-- whither should we bless:
 Save that, work phoebe cannot fill me, and I see,
 Below, the where had calls her ringing men what well
 Did do his hour of day, of city or fine.
 Which clear god beside the sun o'er earth, on high
 Stop fair, at times, our house-top.
 Be old,
 I gladly would make me what the hoofs of mine
 Bends back by all another, , following it at least:
 There is no wonder for some real lot grown;
 Thou wilt not know it. our first end are roll'd,
 Youth made folly, and the by-land will be choose.

I must say, 0.7 turns out some great stuff, too.

Need to: remove double words in a row; put space after the title.

Of 0.7, 0.75, and 0.8, I selected 35 poems out of 126. About one poem in four is worth saving.

I'm starting to think 0.7 is the best. "The Birds" is just outstanding.

The capitals were done wrong by eidnes.

It works if only the capitals are present, or only lowercase. If both are in the vocab list, then you'll lose one (approx 50%) or the other.

STEMMING!!!!!

So:
    -Create a list of every word in Glove and it's stemmification
    -Create a list of words for each stem
    -When you encounter a word that's not in Glove
        -If its stemmification does not exist among Glove's words: choose word with nearest Edit Distance of words that are in Glove and start with the same letter. Tiebreak: simply choose the alphabetically nearest one out of the list of those with same EditDistance. If two are tied (before and after), choose the earlier.
        -If its stemmification does exist, do the above, but only out of the list of words that share the stem.

    The glove stemming should be done exactly once. Read in glove, stem the word, write out word/stem pair in some lua format.

    The read those pairs in, write back out as a two-D table keyed on stem for one index, # of words as the second. The second dimension contains the list of words with that stem. Basically it will look like:

```
while ~eof
do
        read new word, stem
        ! Create table of stems
        if stable[stem]
                ! This stem already exists, append to the list for this stem
                stable[stem][#stable[stem]+1] = word
        else
                ! New stem, add to table
                stable[stem][1]=word
        end
        ! Create table of first initial letters
        fletter=left(word,1)
        if ftable[fletter]
                ftable[fletter][#ftable[fletter]+1]=word
        else
                ftable[fletter][1]=word
        end
done
write(stable,sfilename)
write(ftable,ffilename)
```

So, if a word is not in Glove:

```
        stem=stem(word)
        if (stable[stem])
                chooselist=stable[stem]
        else
                chooselist=ftable[stem]
        end

        for listword in chooselist
        do
                edist[listword]=editdistance(listword,word,5)
        done
        minlist=edist[edist==minimum]
        minlist(#minlist+1)=word
        sort(minlist)
        wordidx=find(minlist,word)
        if wordidx=1
                matchword=minlist(2)
        else
                matchword=minlist(wordidx-1)
        end
        return matchword
```

1/30/17
        Essentially, need a function that, given a vocabulary list (i.e. from Glove) and a word, returns the nearest match to the word in the vocabulary list.

1/31/17
        Rewrote the capitals portion of the word indexing. Should be correct now (i.e. reads in the lower GloVe embedding for a capital).

        Inserted dummy code for nearest, currently returns "couldn't find anything" (index of zero). Have written code for finding nearest but have not implemented it yet. First testing the caps code.

        Slowly comes down toa a loss of 6.0, then bounces there for a veery long time.

        On page 81, see that the threshhold of 87 leads to painful convergence, but 109 is ok. Rerun caps test with 109.

```
mydir=data/Fpoem111
thold=109
epochs=10

echo "Starting 40/35"
th train.lua -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 20000 -print_e
very 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold ${thold} -batch_si
ze 40 -seq_length 35 -learning_rate 5.0e-5 -max_epochs ${epochs} -dropout 0.25 -
seed 77 -learning_rate_decay_after 1 -learning_rate_decay 0.97
```

The convergence honestly looks no better than thold=87.

Increased learning rate 4x:

mydir=data/Fpoem111
thold=109
epochs=10

echo "Starting 40/35"
th train.lua -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 20000 -print_e
very 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold ${thold} -batch_si
ze 40 -seq_length 35 -learning_rate 2.0e-4 -max_epochs ${epochs} -dropout 0.25 -
seed 77 -learning_rate_decay_after 1 -learning_rate_decay 0.97

Immediately comes down to 6.3 and sits there.

Try this expeirment with the old code, 109 and rate of 1E-5:

No, the unmodified goes right to 5.2. Confirmed that only difference are the Caps mods. Something
wrong…

But! Also noticed that there are a bunch of "loaded" statements in the unmodified that are not present
in the caps-modified. Hmmmmm

Looks like there is already a file for this vocab size, so it's not re-parsing. It does look as if it's starting
directly on the poems.

OK, confirm that the caps-modified now behaves pretty much like the original. The problem was that
it didn't reparse. All is well again.

NOW!

Need to restart it, using the prose first.

VERY peculiar, on prose the threshhold of 109 is useless. ?? Only 15k entries.

Set to 35. 31494 in vocab.

HITTING THE WALL AGAIN?!?!?!?!?!?!?!?!?!?
Dives straight for 6.2-6.3 and stays there.

I'm going to let it go overnight and see what we have in the morning.

Nb it's running on a vocabulary with capitals, and 111+1 dimensions.

Gets to 4.5 overnight (yay). Lesson: don't panic

lm_lstm_epoch0.62_4.9072.t7

lm_lstm_epoch1.23_4.6992.t7
lm_lstm_epoch1.85_4.5966.t7
lm_lstm_epoch2.46_4.5307.t7

These are pretty good numbers. The prose samples are a little goofy but do capitalize correctly.

Reminder, to transfer train:
1. Generate vocab and glove on Fromantics, quit after 100 batches
2. Link bitsF
3. Touch vocab, glove files, move data filename
4. Start training on prose
5. After prose is complete, link romantics
6. Touch vocab, glove files, move data filename
7. Start training using "-init_from"

train.lua -init_from cv/lm_lstm_epoch2.46_4.5307.t7

mydir=data/Fpoem111
thold=2
epochs=100

echo "Starting 40/75"
th train.lua -init_from $1 -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every
5000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold 35 -b
atch_size 40 -seq_length 75 -learning_rate 5.0e-6 -max_epochs ${epochs} -dropout
0.25 -seed 77 -learning_rate_decay_after 5 -learning_rate_decay 0.97

nb learning rate! Very low.

Hitting a wall at 4.8 (but possibly low learning rate?)

Output is semi-typical of Caps—lots of And's and I's and That and The to start sentences.

There are no ENDPOEMs though, which is bizarre.

Let's try the stemming.

So it actually mostly works. The logic of how to pick the right value out of the list is not quite right but I can manage that.

The bigger issue that this has revealed is that of the 4500 words not found in glove, approximately 4000 of them are legitimate words that should be in glove, but which have an extra space character at the end of them.   (It's actually ASCII 13.) So the tokenizer must not be handling CR/LF correctly?

During tokenization, if the last character is ASCII 13, the ASCII 13 should be removed.

I should really modify CharSplitLMMinibatchLoader word_iter, but I really don't understand how it works. So instead I have to modify what it returns each time in CharSplitLMMinibatchLoader .

This is dangerous.

Two things should happen by doing this. First, the vocab should go down (possibly dramatically). Second, mainly only proper names should be wholly unrecognized. They can be randomly normalized, or simply set to all zeros. THEN I CAN FIX THE EMBEDDINGS.

Vocabulary now is 25893 instead of 29000. And note that it includes more real, unique words.

There do seem to be instanaces of tokens that consist only of ASCII 13 that then become nil. Not sure what happens with those. Protected against those.

But remember the above about tokenizing the ASCII 13 and its dangers, if anything goes screwy in sampling.

Of the 25893 tokens, only 1100 are not found in Glove. Of these, not a single one stems to anything in Glove. Look like proper names.

Running now on lowercase bitsF. Heads right for 6.1 and hits a wall immediately. But damn does it get to the wall fast, about 500 batches (0.022 epochs).

I honestly think progress ends immediately. Can't see any evidence for advance between 500 and 1800.Still at 6.1.

Let it spin overnight.

(By 8000 it's down to 5.5, so it does come down, just slowly.)

Extrapolation: at 20k, will be in the fours. Extrapolation is dangerous.

2/2/17

lm_lstm_epoch0.62_5.1214.t7
lm_lstm_epoch1.23_4.7482.t7
lm_lstm_epoch1.85_4.6060.t7
lm_lstm_epoch2.46_4.5915.t7

A hair worse than before at 1.85, but more concerning: wall at 4.59. The last one moved down to 4.53 by 2.46. Training at 2.89 is around 4.52+-0.04.

lm_lstm_epoch3.08_4.5226.t7

      So it's doing OK actually.

The prose is not trippy at 0.6. It's just fine. At 0.8 still pretty OK.

Important insight: overall training is not much driven by vocabulary being zeros! (But I suppose it's possible that getting stuck or not cuold be.) Suggests fixed vocab will be ok.

Immediate directions:
      Lower word threshhold (from 35 to? 20?)
      Try fixed embeddings

INTERESTINGLY!

When transferring to the poetry, it picks right back up at 4.6!
(Before, it was worse for a long time.)

But again, little evidence of gain in the first 2500 iterations. Note low learning rate.

lm_lstm_epoch0.33_4.4655.t7
lm_lstm_epoch0.66_4.4571.t7
lm_lstm_epoch0.99_4.4527.t7
lm_lstm_epoch1.32_4.4484.t7
lm_lstm_epoch1.65_4.4437.t7
lm_lstm_epoch1.98_4.4399.t7
lm_lstm_epoch2.31_4.4370.t7
lm_lstm_epoch2.64_4.4325.t7
lm_lstm_epoch2.97_4.4299.t7
lm_lstm_epoch3.30_4.4274.t7
lm_lstm_epoch3.63_4.4240.t7
lm_lstm_epoch3.95_4.4210.t7
lm_lstm_epoch4.28_4.4195.t7

Yay! Within a hair of the all-time record.

lm_lstm_epoch4.61_4.4156.t7
lm_lstm_epoch4.94_4.4126.t7

This is now officially the best loss ever!

Just one problem.

It didn't run on the poetry data! It's all prose. (This means it's not really the best ever.)

-rw-r--r-- 1 andrew andrew 95828197 Feb 1 23:37 data_w35.t7
lrwxrwxrwx 1 andrew andrew   25 Feb 2 07:54 input.txt -> romantics.mixed.lower.txt
-rw-r--r-- 1 andrew andrew   705216 Feb 2 07:54 vocab_w35.t7
-rw-r--r-- 1 andrew andrew 23200255 Feb 2 07:55 glove_25893x111.t7

Notice it did not remake data_w35.t7 after linking the poems.

FWIW prose is not trippy at T=0.8. However, it IS very dialog-oriented.

```lua
    local input_file = path.join(data_dir,  'input.txt')
    local vocab_file = path.join(data_dir, word_level and
'vocab_w' .. threshold .. '.t7' or 'vocab.t7')
    local tensor_file = path.join(data_dir, word_level and
'data_w' .. threshold .. '.t7' or 'data.t7')

    -- fetch file attributes to determine if we need to rerun
preprocessing
    local run_prepro = false
    local run_dprepro = false
    if not (path.exists(vocab_file)) then
        -- prepro files do not exist, generate them
        print('vocab.t7 does not exist. Running preprocessing...')
        run_prepro = true
    else
        -- check if the input file was modified since last time we
        -- ran the prepro. if so, we have to rerun the preprocessing
     print('Vocab.t7 exists')
        local input_attr = lfs.attributes(input_file)
        local vocab_attr = lfs.attributes(vocab_file)
        local tensor_attr = lfs.attributes(tensor_file)
        if (input_attr.modification > vocab_attr.modification) then
            print('vocab.t7 detected as stale. Re-running
preprocessing...')
            run_prepro = true
     else
         print('Vocab.t7 is up to date')
         if not (path.exists(tensor_file)) then
          run_dprepro=true
          print('No data file, rewriting')
         else
             if (input_attr.modification > tensor_attr.modification)
then
              run_dprepro=true
              print('Data file is out of date')
           else
               print('Data file is up to date')
             end
         end
        end
    end
    if run_prepro then
        -- construct a tensor with all the data, and vocab file
        print('one-time setup: preprocessing input text file ' ..
input_file .. '...')
        CharSplitLMMinibatchLoader.text_to_vocab(word_level,
```

```
threshold, input_file, vocab_file, tensor_file)
        CharSplitLMMinibatchLoader.text_to_tensor(word_level,
threshold, input_file, vocab_file, tensor_file)
    else
      if run_dprepro then
         print('Processing data only, with existing vocab file')
            CharSplitLMMinibatchLoader.text_to_tensor(word_level,
threshold, input_file, vocab_file, tensor_file)
      else
         print('No need to reprocess data file')
      end
    end
```

Read the above code, here's the directory listing:

```
-rw-r--r-- 1 andrew andrew 95828197 Feb  1 23:37 data_w35.t7
lrwxrwxrwx 1 andrew andrew       25 Feb  2 07:54 input.txt ->
romantics.mixed.lower.txt
-rw-r--r-- 1 andrew andrew   705216 Feb  2 07:54 vocab_w35.t7
-rw-r--r-- 1 andrew andrew 23200255 Feb  2 07:55 glove_25893x111.t7
```

The data file is plainly not up to date! And yet...

```
andrew@LinuxMint ~/torch/word-rnn-adf-glove100c $ ./train_from_111.sh
cv/lm_lstm_epoch3.08_4.5226.t7
Starting 40/75
using CUDA on GPU 0...
Vocab.t7 exists
Vocab.t7 is up to date
Data file is up to date
No need to reprocess data file
loading data files…
```

I don't know why it does this. I will simply rename the data file.

```
andrew@LinuxMint ~/torch/word-rnn-adf-glove100c $ ./train_from_111.sh
cv/lm_lstm_epoch3.08_4.5226.t7
Starting 40/75
using CUDA on GPU 0...
Vocab.t7 exists
Vocab.t7 is up to date
No data file, rewriting
Processing data only, with existing vocab file
loading text file...
putting data into tensor...
saving data/Fpoem111/data_w35.t7
loading data files...
```

```
cutting off end of data so that the batches/sequences divide evenly
reshaping tensor...
```

So renaming it works.
      Loss goes up to 5.34, as has been seen previously.

```
vocab size: 25893
loading a model from checkpoint cv/lm_lstm_epoch3.08_4.5226.t7
overwriting rnn_size=256, num_layers=3, model=lstm based on the
checkpoint.
using optimizer rmsprop
number of parameters in the model: 10986069
```

ydir=data/Fpoem111
epochs=100

echo "Starting 40/75"
th train.lua -init_from $1 -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 5000 -print_every 100
-word_level 1 -glove 1 -data_dir ${mydir} -threshold 35 -batch_size 40 -seq_length 75 -learning_rate
2.0e-5 -max_epochs ${epochs} -dropout 0.25 -seed 77 -learning_rate_decay_after 5
-learning_rate_decay 0.97

      Increased learning rate from 5E-6 to 2E-5, now let it bake.

JUST REMEMBERED THAT I NEED TO DO THE FIRST VOCAB ON A COMBINED bitsF/romantics (Fromantics), in order to get ENDPOEM.

      Tidbit: the gap between input and output still hovers around 5% in terms of parameter count. I thought that might change with the stemming. Possibly bizarre punctuation?

While that's baking, next steps:

Consolidated Next Directions

2/2/2017 2225

Better ASCII 13 handling
Added 1/17: ~~Fixed sound vectors~~, ~~better capitals handling~~.
Larger vocab (lower threshhold) / Convergence phase transition
Fixed everything vectors
Deeper / Larger nets

~~Train on real whole poems~~ (poems/mixed was generated from whole poems actually)
Bias initialization
        -Final softmax can be initialized to give raw word count frequencies
High loss convergence with capitals
Lack of improvement with sound information
Coherence: BiLSTM? Subject vector?
~~Capitalization bit~~
Post-processing
        -Punctuation space-removal
        -Duplicate words in a row removal
~~Stress encoding~~
~~Rhyme/syllable encoding~~
Primetext
        -Doesn't work
~~Better new-word embedding~~
~~   -Currently new words are randomly initialized—should stem/overlap/edit distance to guess~~
~~nearest word and initialize with that~~
~~   -There are both Levenshtein Distance and Porter Stemming implementations in lua~~
~~   -Should only update new words, not old ones~~
~~   -This matters more than it seems! It's half the parameters. If I can have a good start, then I can~~
~~   train with batches and sequences 2x larger.~~
"In the style of"
50-dimension word vectors
200-dimension word vectors
GRU

lm_lstm_epoch5.98_4.8327.t7
lm_lstm_epoch11.96_4.7665.t7
lm_lstm_epoch17.94_4.7298.t7
lm_lstm_epoch23.92_4.7060.t7
lm_lstm_epoch29.90_4.6879.t7
lm_lstm_epoch35.89_4.6752.t7
lm_lstm_epoch41.87_4.6654.t7
lm_lstm_epoch47.85_4.6581.t7
lm_lstm_epoch53.83_4.6519.t7
lm_lstm_epoch59.81_4.6464.t7
lm_lstm_epoch65.79_4.6418.t7
lm_lstm_epoch71.77_4.6390.t7

So 4.63 isn't so hot, but let's see what it's putting out.

The 0.6 is not great but it is coherent and grammatical, 1.0 is coherent but makes grammatical errors, 0.8 is better in between. The poetry definitely inherits a tilt towards dialog. (Maybe one line in 5 is dialog.)

Retrain on the right stuff for generating the data and vocab. Use threshhold of 25. Vocab is 31520.

Touch everything, link bitsF.

lm_lstm_epoch0.62_5.1454.t7
lm_lstm_epoch1.23_4.9940.t7
lm_lstm_epoch1.85_4.6243.t7
lm_lstm_epoch2.46_4.6630.t7
lm_lstm_epoch3.08_4.5625.t7
lm_lstm_epoch3.69_4.5281.t7
lm_lstm_epoch4.31_4.4589.t7
lm_lstm_epoch4.92_4.4607.t7

Significantly worse in early epochs, but catching up slowly.

OK, now as good as before. Switch to poems.

lrwxrwxrwx 1 andrew andrew   25 Feb 3 20:47 input.txt -> romantics.mixed.lower.txt
-rw-r--r-- 1 andrew andrew   860124 Feb 3 20:47 vocab_w25.t7
-rw-r--r-- 1 andrew andrew 28242047 Feb 3 20:47 glove_31520x111.t7

./train_from_111.sh cv/lm_lstm_epoch4.92_4.4607.t7

100/83600 (epoch 0.120), train_loss = 5.43495256, grad/param norm = 4.3232e-03, time/batch = 0.5957s

Baking the poetry now overnight. Painfully slow.

lm_lstm_epoch5.98_4.9149.t7
lm_lstm_epoch11.96_4.8408.t7
lm_lstm_epoch17.94_4.8017.t7
lm_lstm_epoch23.92_4.7759.t7
lm_lstm_epoch29.90_4.7563.t7
lm_lstm_epoch35.89_4.7430.t7
lm_lstm_epoch41.87_4.7324.t7
lm_lstm_epoch47.85_4.7242.t7
lm_lstm_epoch53.83_4.7178.t7
lm_lstm_epoch59.81_4.7123.t7
lm_lstm_epoch65.79_4.7076.t7

Could be learning-rate limited. Restart at full learning rate.

But first, look at some output.

The 0.65 is very coherent and grammatical. Lines are unusually long, and no ENDPOEMS.

*she took her , with the bright a knight of the bride , and made her
in the world , and held her home , and though she had not made
her pleased , and in her note , she was not dead .*

she died and tried to tell her , but her own was the blind
and her in her eyes , and she was , and then she gave
her love , and all that she had done , but yet she said
she was a soul and glad of her father ' s sake .
when she was the boy and her father ' s children , her own
her wife , and her life with her , and the joy of my heart ,
and every wonderful one she ' s married , who ,
and all her father , as she ' s left for her , her bosom ,
and that he was true and she would have made her way
with love of beauty .

she ' d like that she was much as she was a child ,
and there was her , and the first , and her own , she say :
she said , " she ' s true me , and i ' ll find you there ;
but i ' ll be a little child ; and now i ' m no longer ,
that i ' m so glad to go to me , so if you may .
no more ! " i was in a goodly breast , and the moon
were dearest to her in my heart , and i would not
say , " i say , " she said , " you ' re just a hero , " said i ;
" i ' ll take it , please me , " i said : " let us be
and see you , and in the sun , oh , i ' ll never call
your way in a sort of beauty . i ' ll get my hand ,
and you will be
you ' ve gone to break me to get up here . "

and then the week in the night he went away ,
and with a fire and a dog stood in a wood ,
and i caught my head in a limb and often
that i am not to say , and set her own ,
and i thought , my heart was in a coffin .
i heard his voice the tears of the white song ,
and my heart was a little that i knew ,
and i read that i am one of the world .

i looked at my best , and i went , and i went ,
the poor man was the house of the dead ,
and i went in the grass with a dame ' s tears .
and then saw my father , and he went to me ,
and the day were done .

there was a little thing ,
the man in my lips came to me ;
but i heard the white beat , and i saw
the little thing with a melancholy sound ;
i ' ve been on this night , and i spoke , and i
was a maiden .

so i seemed to know what i could not say ,
and there was a hard man so more .

i saw the word that seemed to call me ;
i could not say that i saw my way .

With spacing and caps:

She took her, with the bright a knight of the bride, and made her
In the world, and held her home, and though she had not made
Her pleased, and in her note, she was not dead.
She died and tried to tell her, but her own was the blind
And her in her eyes, and she was, and then she gave
Her love, and all that she had done, but yet she said
She was a soul and glad of her father's sake.

When she was the boy and her father's children, her own
Her wife, and her life with her, and the joy of my heart,
And every wonderful one she's married, who,
And all her father, as she's left for her, her bosom,
And that he was true and she would have made her way
With love of beauty.

" I'll take it, please me, " I said: " let us be
And see you, and in the sun, oh, I'll never call
Your way in a sort of beauty. I'll get my hand,
And you will be
You've gone to break me to get up here. "

And then the week in the night he went away,
And with a fire and a dog stood in a wood,
And I caught my head in a limb and often
That I am not to say, and set her own,
And I thought, my heart was in a coffin.

I heard his voice the tears of the white song,
And my heart was a little that I knew,
And I read that I am one of the world.

I looked at my best, and i went, and i went,
The poor man was the house of the dead,
And I went in the grass with a dame's tears.
And then saw my father, and he went to me,
And the day were done.

There was a little thing,
The man in my lips came to me;
But I heard the white beat, and I saw
The little thing with a melancholy sound;
I've been on this night, and I spoke, and I
Was a maiden.

So I seemed to know what I could not say,
And there was a hard man so more.
I saw the word that seemed to call me;
I could not say that I saw my way.

Probably I was overzealous with removing the returns in the tokenizer. But I'm going to let it learn itself.

 ./train_from_111.sh cv/lm_lstm_epoch65.79_4.7076.t7
lm_lstm_epoch5.98_4.6863.t7
lm_lstm_epoch11.96_4.6692.t7

OK, so it bottomed out before because of learning rate decay, not a fundamental barrier.

lm_lstm_epoch65.79_4.7076.t7
lm_lstm_epoch5.98_4.6863.t7
lm_lstm_epoch11.96_4.6692.t7
lm_lstm_epoch17.94_4.6560.t7

Even now, the gains are simply following the learning rate.

Turn down the decrease from 0.97 to 0.99., restart.

./train_from_111.sh cv/lm_lstm_epoch17.94_4.6560.t7
lm_lstm_epoch5.98_4.6405.t7
lm_lstm_epoch11.96_4.6271.t7
lm_lstm_epoch17.94_4.6151.t7
lm_lstm_epoch23.92_4.6059.t7

Looks like a wall will be around 4.55

lm_lstm_epoch29.90_4.5962.t7
lm_lstm_epoch35.89_4.5884.t7

It might be converging a little faster than pure learning decay, but it's a very similar rate. Let's crank!

mydir=data/Fpoem111
thold=25
epochs=100

echo "Starting 40/75"
th train.lua -init_from $1 -gpuid 0 -num_layers 3 -rnn_size 256 -eval_val_every 5000 -print_every 100 -word_level 1 -glove 1 -data_dir ${mydir} -threshold ${thold} -batch_size 40 -seq_length 75 -learning_rate 5.0e-5 -max_epochs ${epochs} -dropout 0.25 -seed 77 -learning_rate_decay_after 5 -learning_rate_decay 0.99

If the learning is the limiter, the first eval should be about 4.568

lm_lstm_epoch5.98_4.5658.t7

Pow.
lm_lstm_epoch11.96_4.5470.t7

lm_lstm_epoch17.94_4.5305.t7
lm_lstm_epoch23.92_4.5172.t7
lm_lstm_epoch29.90_4.5052.t7

dimprovement / dlearning-rate is not obviously falling (but may be). So (a) not hitting a learning wall and (b) limited by learning rate

lm_lstm_epoch35.89_4.4978.t7
lm_lstm_epoch41.87_4.4883.t7

ENDPOEM just appeared!!!!!

Temperature 0.7. Slightly unfair cropping here.

 what is he ? i know that i have seen
 to the women that have come from the grave .
 the night is gone , and we have gone
 unto the sky that grows at the stars ,
 and when the wind is blue ,
 the wind is bright and cold ,
 and the deep sails are gleaming ,
 and death is the night .
 ENDPOEM

Capped

 What is he? I know that I have seen
 To the women that have come from the grave.
 The night is gone, and we have gone
 Unto the sky that grows at the stars,
 And when the wind is blue,
 The wind is bright and cold,
 And the deep sails are gleaming,
 And death is the night.
 ENDPOEM

lm_lstm_epoch5.98_4.4765.t7
lm_lstm_epoch11.96_4.4686.t7
lm_lstm_epoch17.94_4.4620.t7
lm_lstm_epoch23.92_4.4543.t7
lm_lstm_epoch29.90_4.4471.t7
lm_lstm_epoch35.89_4.4472.t7
lm_lstm_epoch41.87_4.4421.t7
lm_lstm_epoch47.85_4.4379.t7
lm_lstm_epoch53.83_4.4367.t7
lm_lstm_epoch59.81_4.4323.t7
lm_lstm_epoch65.79_4.4291.t7
lm_lstm_epoch71.77_4.4275.t7

I recall the best ever was around 4.41, so this is really good but not quite there. It doesn't look likely to get to 4.41. (Turns out 4.418) I guess that by 100, it could be close.

lm_lstm_epoch77.75_4.4260.t7
lm_lstm_epoch83.73_4.4294.t7

It bounced!!!!! :(I've never seen an actual bounce in training CRM

Also: training losses are around 4.30.

Conclusion: we've converged. The winner is 4.423, a hair worse than the best ever of 4.418.

lm_lstm_epoch89.71_4.4231.t7

The poems are generally ok to good, but even the best seem to fall short of great. A few at 0.75:

The Angel
 When the sun's keen, with die,
 And the dead and the dead
 Find the day in sweet;
 What's the beginning of us?
 Who hath set o'er the sun?
 So, this's gone before,
 The dawn is lost,
 And the sod shall be gone
 On the earth's fields;
 They shall stand in the light of the stars,
 To the grave of the dead.
 But they shall watch the eternal dawn,
 And the lambs is dead;
 And the same immortal may,
 Shall make us see
 The last sound of the world.
 ENDPOEM

 The moon
 I've heard of the wild orient-isle
 In the passing of the mountains;
 For the voice of the horizon's story
 That floods with the charm of the day,
 And the place of the wild winds are spread
 From the couch of the golden earth.
 And the birds will come in the future,
 And the stars are happy and small;
 The feast is of the old maid,
 And the windy roses are mine.

Your mother is the flower of the day,
And she's the white banks that shall be.
A tempest and a silver stone,
And many a wings of the stars!
O nature, how spend it again,
And the eyes of the cheerful night;
When a sister's smile is strong,
And the kiss of my song has turned
When the prophet whisper to sing,
And that he call'd the landscape.
ENDPOEM

 All are its began;
Runs, on the sea of night,
From the evening, gone,
And the snow is toiling.
Now, from a distant sea
Of a starry river,
The white wind goes;
The wind wilt neath itself;
A silver tear is not
The sun's more than the dead.
The sky has been white
At the sun, the sky is seen.
And the sun is over,
The sun is sure of light;
The brown green ocean come
With the sun of her red.
The fields are seven bright,
And the leaves of the sea
Is dancing, and the dew;
The birds are full of light,
And the sun's green are the
Cold and the east.
The morn is gone, the summer,
And the storm is near!
The wind hath a warm light!
That is long to come
On the wind to the sky.
ENDPOEM

Love
I.
For i have caught it, not the word;
It is not to be sure i know;
And i have lost the name no more;
So we are in a minute--
That is over me, and lift
The heart upon the heart

That none can prove the tears,
But it is dead.
ENDPOEM

The mountain to the wave
A magic sail, where the old sun has set
From its green house on the black oak;
To-morrow, to the wandered-nook,
A south-uproar, and the sea,
For the sky-wind, and the wind-tree,
Where the water, in the south,
With the rocks, and on the hill,
And the faults of the mountain-mill.
Which to be alone
Or the summer,
Till the sweet flower goes,
And the moon is so sweet.
ENDPOEM

The gleaming
The hills and the woods, the hill,
The river, the green sky,
Their leaves and their sons,
The birds, hills, and flowers,
And the summer iii.
The pines are white and green,
And the green walls shall be clung,
And the green birds gleams above,
And your spring is as a star.
ENDPOEM

The maple bells
The swift woods making the green wood,
And even the stream is come away,
But the nightingale is on the wall,
And the seasons are learnt to die;
And the sound of the wild winds are
As a stream or a moon shall be:
There is no more in the days of god
Our seasons and the slave;
And the first, the soul has no need,
And the heart of the heart is thine,
And the flower to the minds of heaven
Vainly, the earth is strong,
It dwells in the heart of love
That softly brings the door.
ENDPOEM

The cup of the rain

The wintry sun most vast and heedless,
The winds, the clouds and the winds,
The stars and the sun and the light,
The dark sea from the earth;
The dream is of death.
ENDPOEM

Song of the early sun
In the hills
Now thou art'st the last vast,
Or a notes of gold,
We come from the green sea
To the edge of the tide,
And the great to its reality
And the success of all.
There are some whose fall to the sun--
All the world's summer,
The wind is gone, but we saw on us
And we look in the air.
ENDPOEM

Such be i?
I am not old;
I'd grow as a wound or wind:
I am not blind in one.
That is no matter, i can know.
I love not less as i know.
If i have no thought, no man,
I see the little one,
And i am all because he died.
ENDPOEM

So soon i turned on the pagan tide,
I stood and saw it;
For one day of the old days, to the sea
I met what they gave,
As i had sparkled with a misty boat,
And then the deep were dead.
And the sun was as fast of the sky,
The birds were in my sight;
I knew, and i wept with the town;
I watched them in the moon.
And i knew that i was sleeping ",
And the hermit gazed on me,
And i knew how i knew that he heard
The vessel poetic night.
Then i am like a dream, and i mild
With it as i lie;
When, there i would pray, i see-- the blessing

Of my youth and me.
Then, after a golden eye i lay,
We stood and then a bright imagination;
For i heard the moon's sick in the eyes
Of the shawl of the year, and i could not have none.
For the church amid the window lay,
And i said, " you'll take and go;
And i've often you see, there's no more,
I see you gay in the moon.
" You're a little dear, " said she.
" I'll have no home. it's true--
You'll have the secret and grand-boats
When we're-- so sweetly white
I'll get the linger and away:
Some night i've been to be dead,
And i'll pay in the hour,
As the sky has been nectar.
ENDPOEM

There are 85 poems these are chosen from. So the top ~10 percentile.

Temperature 0.8

The House Of The Ocean
A winter is the intolerable hour,
And the blue becomes the dim disgrace;
With shrill eye of her eyes, like silver shade
Of the sighing of heaven,
In the sunshine of the eye of delight,
The big moon, and with the shade of heaven.
ENDPOEM

The Shepherd
The morning was the waters,
The sunset sea,
In the low air of dark
The darkness of the sky.
As the spring of summer,
Slowly the summer wind
Came by the cold sea,
The gray birds drew and loud.
In the gray spirits
The dream of the world
Came, as the sun lay,
Driving him with the fierce
Laughter of the glow.
The sun seemed bent on the plain,
And the shadowy wind grew away;

And the winds of the sun
Had fill up the dewy rain
To the sky of the snow.
Then the old birds vanish,
And the wind is gray,
And the stars that sometimes
Fell from them in the air
In the sunny land,
They passed the young moon's wild.
And the sun went on
The sunset night,
By the morn of noon,
And the living birds
In the summer day.
ENDPOEM

The Joy Of The Chill
The mask of the dead,
The stream of the hill and the sky,
When the life of our shadow,
The love of light, sweet stars.
ENDPOEM

To The Table Bound
I.
The world was late, the sweet and summer!
The sun was white, her eyes were laid,
And all the daylight was so late
And the seven days are singing.
She said, " we'll never speak, " said she,
" I've won to please me love. "
Said the lily; and here she smiled
" What is you? " on the sea,
The moon was a soft, sweet maid;
And she had had a secret laugh,
And there she was a little boy,
And she spoke for them who so saw:
But from the wide night and the moon,
And in the woe they lay,
I've loved the flowers that are dead.
ENDPOEM

Fair Peter, Looking By Me!
Dreaming of a happy night,
That from the centre
That her cheek is sweet,
And my cheek of love may meet
When her lips are pensive!
That for comfort I might find

To make a thick and old smile,
    And my heart will dismay
    The soul's end, and the day.
    When a bee, in the lamp
    Of my mother's hour,
    I feel heart, and sorrow,
    And love love and fame.
ENDPOEM

I think temperature 0.75 is better.

I'm getting out of memory errors. No idea why. Rebooting.

Ran training end-to-end. It takes about 33 hours. 3 epochs on prose, 150 on poetry. Got to 4.4786.
Thinking that should go to 5 on prose.

 the nun upon the fire
 the folded shade across the sky
 was silent , and the moon was dead :
 as on the hollow evening came ,
 the sun on the twilight feared ,
 the low one grove was shining there ,
 the clouds of light looks still .

Not too bad. But overall not so good, either. Most sets have one or two poems that are truly
great/viable, or within a single word of being so. But don't see that. Also lots of:

 the
 the
 the

 the
 the
 the
 the
 the
 the
 the
 the
 and the
 the
 the
 the
 the beggar ' s husband

But occasionally ok:

 the whole

heaven is a place to me ,
and the soul goes by ,
beyond the glory of the world ,
through a country ' s west .
there is but a man from this
who the soul is - -
god ' s - - the master ' s every word
that of the christ .
ENDPOEM

the moon
thou art not , thou dost hear thee ,
i see a smile of same ;
and , after those of gentle woods ,
shine on the glowing sky !
o thou art not a heaven of my own ;
and thou art dead with thee thou art ,
and o ' er thy bloom and lyre
conceal thee o ' er the tender hell .
so i shall love thy own delight ,
and i shall love thee in thy heart ,
that may thou sleep , for thou art gone
in thy own self and sympathy ,
and the voices of thy heart
is found for thy love o ' er thee .
thou didst not mark me ; my heart !
how he cannot win its earth ,
and love thy love , and the last heaven ,
to be decay and die .
and when thy soul is weary ,
when memory ' s self is born ,
thou must not hear the soul from thee
and in thy last and last .
here , i shall have ; the flashes
of the temple ' s breast , and thy blood
and all the world ' s c ,
and my own , and the wings of the air ,
shall yonder be the awful thought
of the living and the sky .
ENDPOEM

BUT
0.65 is working.

2/13

Been having a lot of trouble with memory errors after FIXING the embedding. This is peculiar
because the parameter count goes down. That said, I believe that in the internals, the parameter count
is the same, but the gradients of the embedding are set to zero by hand. (I'm not sure this is a good

strategy in principle, but it's probably ok in practice.)

I raised the vocab threshhold from 25 to 26 so the parameter count is strictly smaller. But it still crashes!

THCudaCheck FAIL file=/tmp/luarocks_cutorch-scm-1-7473/cutorch/lib/THC/generic/THCStorage.cu line=40 error=2 : out of memory
/home/andrew/torch/install/bin/luajit: ...rew/torch/install/share/lua/5.1/nn/ClassNLLCriterion.lua:62: cuda runtime error (2) : out of memory at /tmp/luarocks_cutorch-scm-1-7473/cutorch/lib/THC/generic/THCStorage.cu:40
stack traceback:
        [C]: in function 'resizeAs'
        ...rew/torch/install/share/lua/5.1/nn/ClassNLLCriterion.lua:62: in function 'backward'
        train_fixed.lua:312: in function 'opfunc'
        /home/andrew/torch/install/share/lua/5.1/optim/rmsprop.lua:32: in function 'optimizer'
        train_fixed.lua:359: in main chunk
        [C]: in function 'dofile'
        ...drew/torch/install/lib/luarocks/rocks/trepl/scm-1/bin/th:145: in main chunk
        [C]: at 0x00406670

I did find that truly dramatic reductions in the seq_length helped.

So I'm going to try to go ahead with seq-length of 50. That doesn't fill the memory.

But it does converge badly…

At end of prose:
lm_lstm_epoch3.00_4.6394.t7

4.55 was typical for floating embeds

lm_lstm_epoch3.99_4.8291.t7
lm_lstm_epoch7.97_4.7513.t7
lm_lstm_epoch11.96_4.7029.t7
lm_lstm_epoch15.95_4.6701.t7
lm_lstm_epoch19.94_4.6462.t7
lm_lstm_epoch23.92_4.6293.t7
lm_lstm_epoch27.91_4.6129.t7
lm_lstm_epoch31.90_4.6049.t7
lm_lstm_epoch35.89_4.5951.t7
lm_lstm_epoch39.87_4.5865.t7
lm_lstm_epoch43.86_4.5782.t7
lm_lstm_epoch47.85_4.5731.t7
lm_lstm_epoch51.83_4.5678.t7
lm_lstm_epoch55.82_4.5662.t7
lm_lstm_epoch59.81_4.5621.t7
lm_lstm_epoch63.80_4.5582.t7
lm_lstm_epoch67.78_4.5559.t7

lm_lstm_epoch71.77_4.5540.t7
lm_lstm_epoch75.76_4.5511.t7
lm_lstm_epoch79.74_4.5506.t7
lm_lstm_epoch83.73_4.5479.t7
lm_lstm_epoch87.72_4.5475.t7
lm_lstm_epoch91.71_4.5446.t7

Based on trends it would be surprising to get below 4.50.

Didn't break 4.50. best:

lm_lstm_epoch50.00_4.5295.t7

0.85 is not terrible. 0.65 is pretty bad.

The 0.65 is notable for the first-ever misuse of ENDPOEM:

for the stars of the world is gone ,
 and the waves , the night ,
 the ENDPOEM ' s year ,
 the priest of the ancient lies ,
 the old man of the eternal ;
 and the song of the sword
 of the beauty of the sky .

I have to say I feel the fixed is a failure, but a near-failure. What I am displeased with is that there's no real feeling that the CMU stuff is being taken advantage of. Now, the fixed did have a shorter sequence length but at the level of stress at least that shouldn't have mattered.

Went back and looked at 0.85 for floating
lm_lstm_epoch89.71_4.4231.t7
output is in
write_0.85_021417_2117.txt
It's not bad

 To my wife
 I laid your ruddy flowers
 I saw, the earth and through my hands,
 And i knew how the night came!
 I saw you all the morning day,
 And laid my face into my hand,
 And ran my fingers back,
 And burned it all my life to mine.
 Then i said,'ye would have been good
 All the woods from the horizon-side,
 I shall never be lost behind,
 And from ever all my will .'
 And when the moon is busy still,
 The hill and the sky is shining,

And the sun-flowers is in friends and gold,
And the image of the heart.

Come where in this night
A life is like a dream
That is old and strange;
He scorns, and is asleep
All the world thou art.
But by a shadow,
When the sun are thrown,
The summer's winter
Are beautiful and dead,
Shout and play.
But to the morning's shade
To the fields
The waving faded wind
Murmur, and the roar
Of the light and there!
And the great and dew
Be an angel of the words;
The hearts that we loves;
The winds are black,
Every sound is change,
And the withering gloom
Is no more.
The power of love
Is not proud before,
Or that, nor still,
To the grave
Shall change it.
ENDPOEM

To My Love And Hope, Life's Life

Love is the shape which made it face
On the very heart of so
Its beauty does not weep,
'Tis not for an hour of earth.
ENDPOEM

The Mountain-tree
Autumn, and the water, and the sky,
On a sweet wind and the sun,
Where the winds have the task of noon;
Winter echoes like the rain,
Life is drifted in me,
And there's nothing on the sea.
And the sweet, the solitude is
Devotion and the song,

The mountains come into the land,
And the soul is light.
ENDPOEM

The Storm
Their only marvel be in winter-place
And leave the spot below.
When every day I looked alone,
The crescent-grimly past;
Of the devils that at once I heard
Here from the hill of rain,
And the wind of the sun's pale ray,
There, no longer came,
And the morn we come, and work in vain
To follow where they follow.
ENDPOEM

The Dying
Because the ancient boyhood may be near
The duke of a venice,
Alone of the past, and the grace
Of great heart he goes.
And thou, whose moccasins, with
The world, and the weakness,
Sacred of god, thou art the end
Of the end of through.
I know how the world is as death,
To thy spirit free.
Keep in law to summer life,
Heart, my soul of john,--
And thou art worth at least,
And a faith without love.
Only, when to the rest,
If we feel if we die,
Let it understand,
Or with I will give
The soft and bitter breath
Still we may cast away.
ENDPOEM

Memory
A little, roses, the icy woods,
Whose summer gleams with light had bound;
Maybe we have seen a farewell life,
Like a great fish of death and things,
This night, a grave, a fairy child
Is all, so fair, and dead.
He knows what is it? is before?
O perhaps, no more, the little power.

Alas! the crime is only, and
The wandering heart and heart of god,
I talk, though I can make the soul
To see the world and still control.
ENDPOEM

Bewitching
With tearful eyes of woe he died;
The chance by time was done,
For music went away at last
In joy and beauty.
ENDPOEM

I Love The Life, But That I Said,
Oh is it not to feel the breath
Of the loss which my soul can do?
No doubt, to these I love I know.
My friends! each day when every day
Looked in my heart of bliss the stars,
And kiss'd with hair and mouth, my head
My blood sisterly, with shapes of tears,
Eye on it like a charm to meet
Upon me in the thunder.
And. .. to do that gentle youth I pray
My very pair, so soon as you,
Since we were lost for any man,
And turn'd from look, and seem'd to stand:
All eyes would open birth, and I
Have been no longer in my eye;
I hate not to weep, but I weep;
I cannot hope if I remember.
ENDPOEM

Tried going to 5 epcohs of prose. Very minimal improvement.

The Gloomy Choir
We may be loved: the stars which hid
The summer wind;
The word of your soul, the ocean
Come below,
We may not keep before the hills
And the sun
Together, through the world, we bring
The azure weave.

No soothing hour is not,
I never pass
Through these fair thoughts of sleep
And summer's fast.
But in our life of love we are
A pleasant life,
In the sweet night that come no more,
When the free sun
Should come to every night.
But not, when the year
Has been begun the soul
That is not friend.
ENDPOEM

The Queen Of The Lord
I
A picture
I loved my beloved:
I can not think, my love
To in the east.
I wandered in thy eyes,
No more than hell,
No tired sight with years,
The splendid stealth
Of the world could never be.
But can I awake
But because the sun
Or many a year will see,
My dark shame loitering
So dear as I have seen
A young man's hand?
Yet in my eyes I say
Bold of the divine
I thought that my father's heart
Is chance in mine.
ENDPOEM

Sonnet Silently.
Had she, fair voice of life.
And for a path of time and day the night I laid,
The extremely wild road, which would be gone,
Behind the branches, all the wheresoe foam,
And pack and skies on the ice below.
I had seen, and bore not, sense to see,
The bird-florence's safety, and the very brave,
The blossom and a light of summer's hours.
They came before the night, and then she drew,
And all her worn hair seemed to make it dry,
And I could die as if she was the past.

She said: " why should I know, you do not say? "
And what is this? one squeeze: this fiery echo
Is left that the earth cannot bending and wide.
How moment I cannot ne'er had spent for me:
The true that dropped me; but the death in my heart
I was cold---i shouldst my wish, as I was .--
Had I to such a fire shall reach the coral,
Perhaps, mine? no longer---if thou wilt not---
I thought thee, or in this end of love and woe,--
One without and bells more. I have loved thee.
ENDPOEM

The World's Lightness

'Tis the world which we meant
By ahead to our woe
And human time of nine
In the tall praise of men
By the brig's arm-
Just it to our sure-
The paradise of the moon
If it is? a child
And just the common reach
Which deemed of bands and anguish
And judge of ancient quaker.
Then in this blossom
Come back to cast a grace
And let them go and die
Away and back and now
Tell the peace of it, when lo
He suffered to the night.
ENDPOEM

The above are all from write_0.85_1487160651.log.cap
Which in turn is from cv/lm_lstm_epoch50.00_4.5127.t7 at temperature 0.85

Next Directions

2/16/2017 2147

Better ASCII 13 handling
Added 1/17: ~~Fixed sound vectors,~~ ~~better capitals handling~~.
Larger vocab (lower threshhold) / Convergence phase transition
~~Fixed everything vectors~~
Deeper / Larger nets

~~Train on real whole poems~~ (poems/mixed was generated from whole poems actually)
Bias initialization
        -Final softmax can be initialized to give raw word count frequencies
High loss convergence with capitals
~~Lack of improvement with sound information~~
Coherence: BiLSTM? Subject vector?
~~Capitalization bit~~
Post-processing
        -Punctuation space-removal
        -Duplicate words in a row removal
~~Stress encoding~~
~~Rhyme/syllable encoding~~
Primetext
        -Doesn't work
~~Better new-word embedding~~
        ~~-Currently new words are randomly initialized—should stem/overlap/edit distance to guess~~
~~nearest word and initialize with that~~
        ~~-There are both Levenshtein Distance and Porter Stemming implementations in lua~~
        ~~-Should only update new words, not old ones~~
        ~~-This matters more than it seems! It's half the parameters. If I can have a good start, then I can~~
                ~~train with batches and sequences 2x larger.~~
"In the style of"
50-dimension word vectors
200-dimension word vectors
GRU

Since last time:
        Confirmed: sound vecors aren't helping a lot. 4.51 was acheivable without them.
        Confirmed: fixing the meanings is baddish. Can't get past 4.51. Got 4.41 floating them.
Consider the 4.41 significantly superior.

        Next: Fix the 13's, then move along to deeper nets.

Conclusion: fixing the 13's (not deleting 13 off of a two-character or one-character token) didn't
meaningfully improve things. If anything there was (probably inconsequential) damage.

It's halfway through romantics1 right now, I'll see if th conclusion still holds through romatnics2.

lm_lstm_epoch50.00_4.5196.t7 is every so slightly worse than 5127

But I think the punctuation is probably better, the semantics (oddly) significantly worse.

Change length to >1 instead of >2.

If so (it is) I'll move to better nets:
1. 4 layers of 310 gives similar parameter count to the original floating embeddings
2. 4 layers of 244 gives similar parameter count (if you float the embeddings), to 3/256.

Let's do 4/310

4/310 doesn't seem any better after 57 epochs. Hmmm….