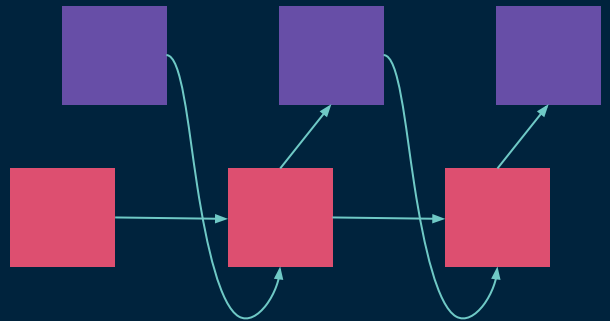
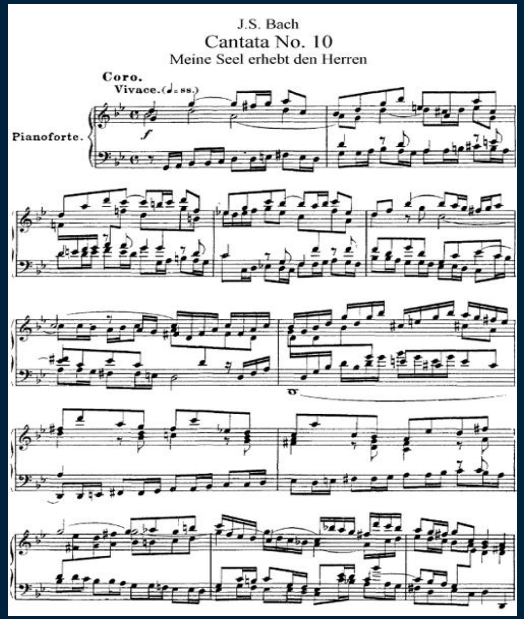
Generating Event Sequences with RNNs

Kyle Kastner (kkastner@)
Intern from Université de Montréal - MILA
with Mike Schuster (schuster@)
and Douglas Eck (deck@)



Basics: Representations

Sheet Music



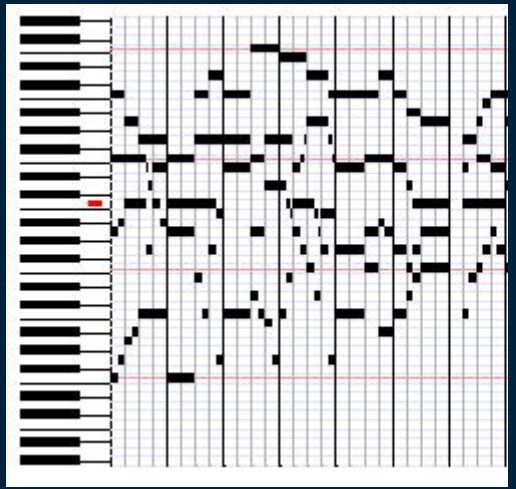
http://www.all-music-sheets.com/images/BachChurchCantataBWV10.jpg

MIDI

```
midi.Track(\
 [midi.TextMetaEvent(tick=0, text='Melody',
data=[77, 101, 108, 111, 100, 121]),
  midi.ProgramChangeEvent(tick=0, channel=0,
data=[57]),
  midi.NoteOnEvent(tick=720, channel=0,
data=[62, 126]),
  midi.NoteOffEvent(tick=242, channel=0,
data=[62, 64]),
  midi.NoteOnEvent(tick=0, channel=0, data=[62,
126]),
  midi.NoteOffEvent(tick=360, channel=0,
data=[62, 64]),
  ...
```

https://gist.github.com/tleyden/80964d7a541dd7e96cd0

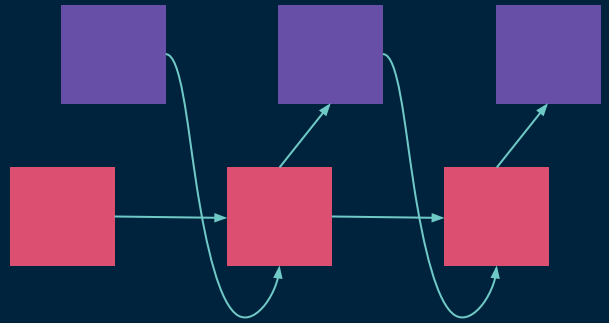
Piano Roll



http://www.umsl.edu/~fraundorfp/rworld/piano_roll.png

For example purposes only, music is NOT the same piece in each image

Basic Approach: RNN Language Model



Classic RNNLM

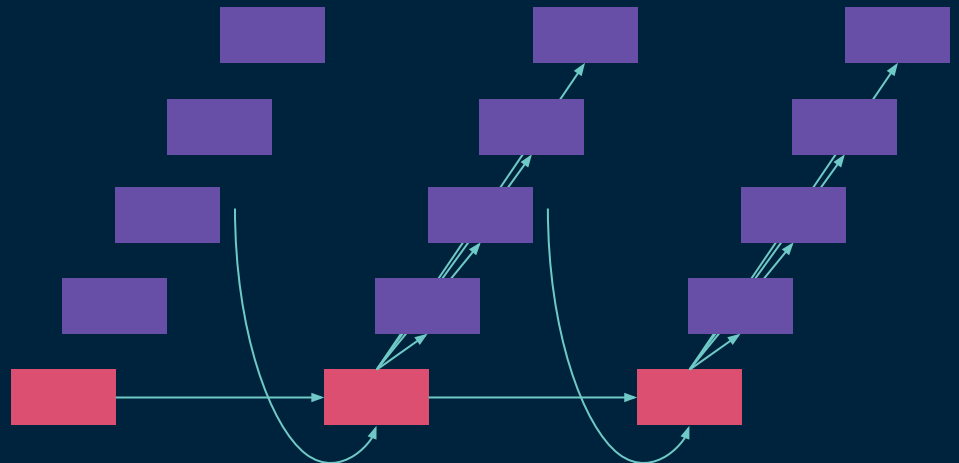
- <START>The cat eats food
- The cat eats food<EOS>

Finding chords is a learning problem itself

- Use notes directly

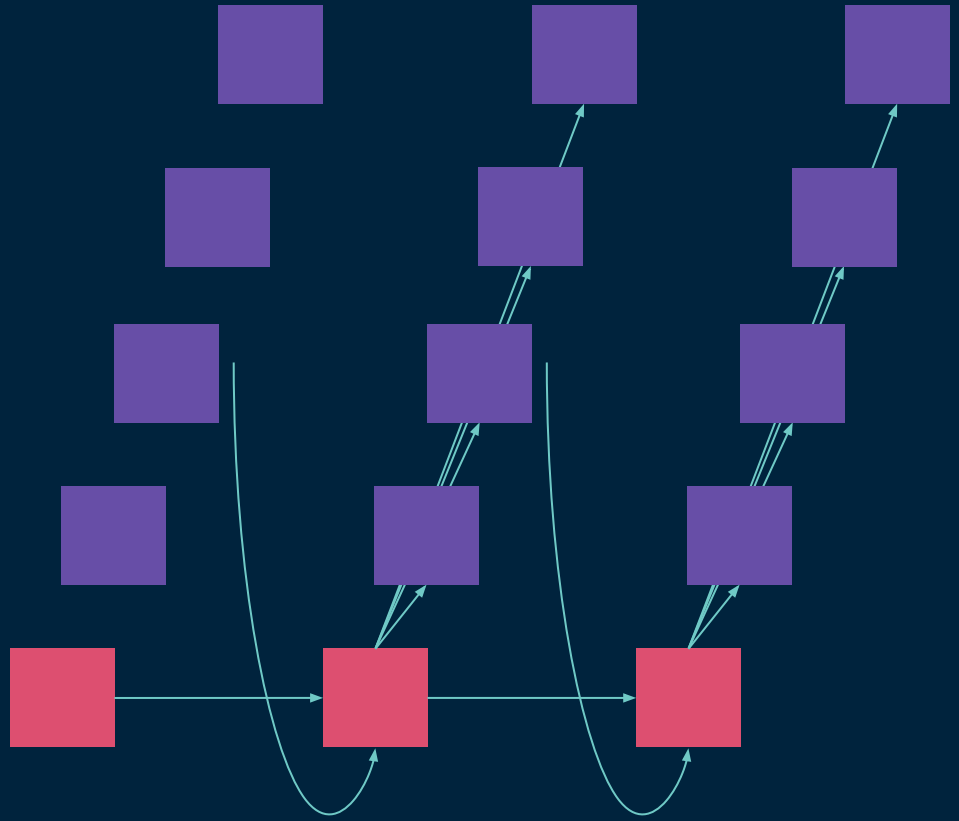
A big problem… how to do multiple outputs per step?

- Add more (softmax) layers



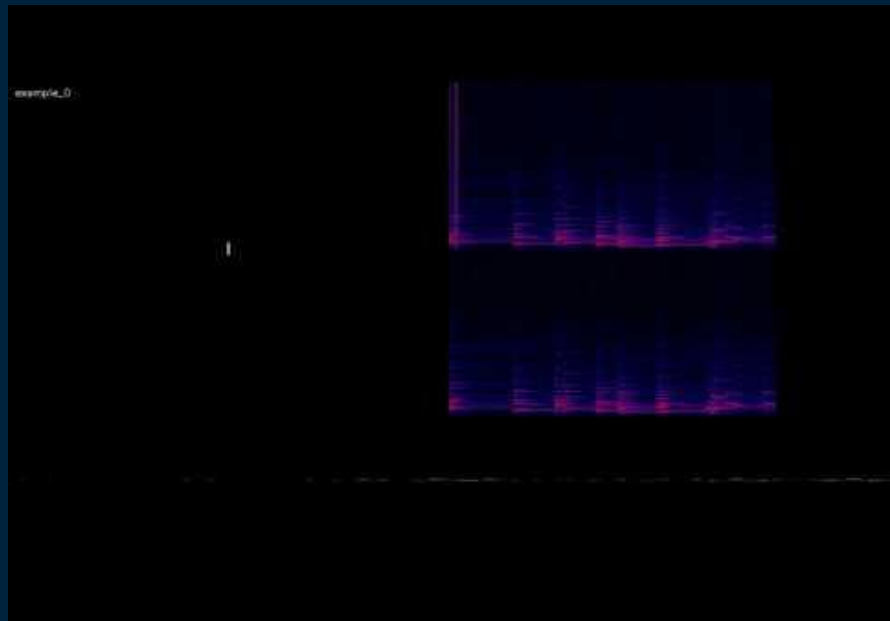
Goal: Generative modeling of polyphonic music

Approach (Week 1-4): GRUs on piano roll



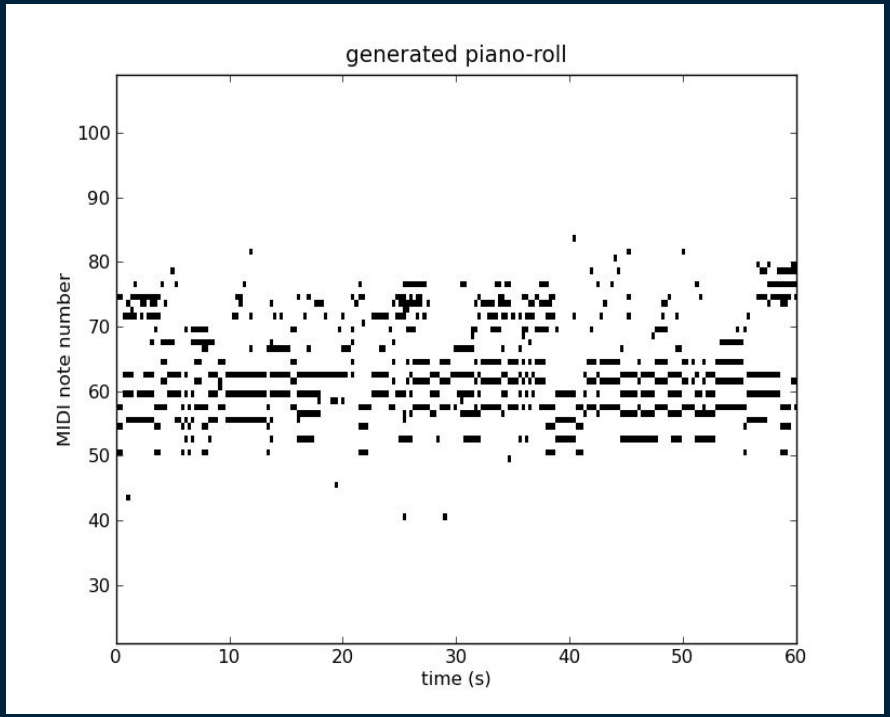
NOTES
(1 of 89)

- Bach Chorales, ~80k steps, quarter note samp.
- Training time, ~30 minutes



Major Problems

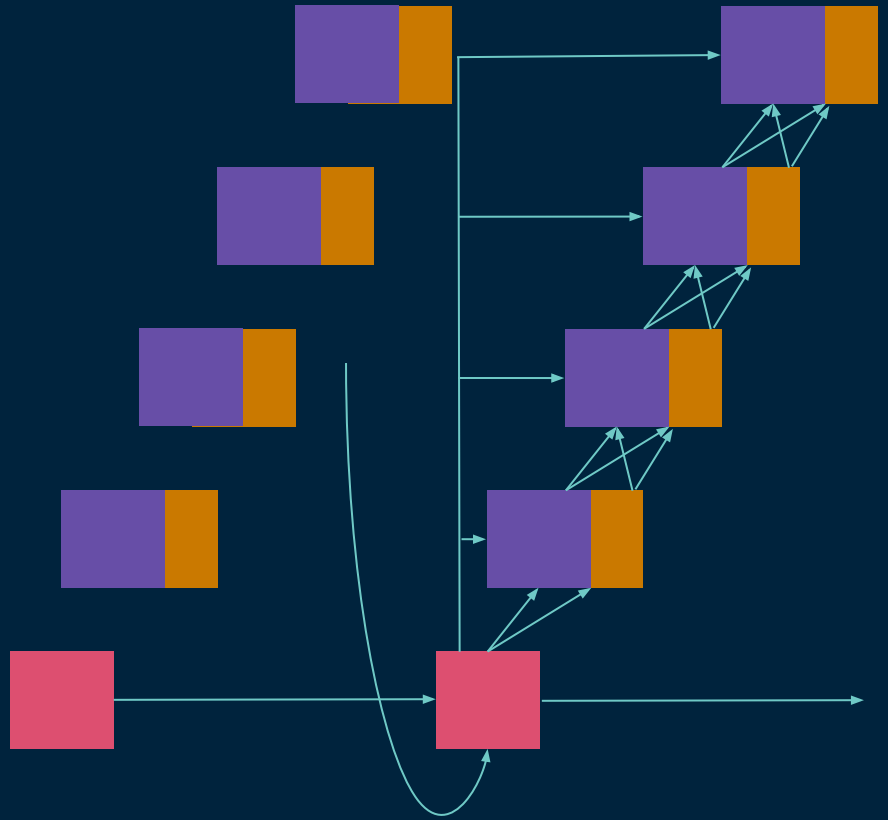
- **No output context at current time**
  - Add chain dependency on output
  - MADE style masked teacher forcing
  - Key component of Pixel RNN
- **“Piano roll” representation**
  - Continuity is *hard* in this format
  - Decisions should be all at once
  - Basic MIDI is an event stream



N Boulanger-Lewandowski, Y. Bengio and P. Vincent, Modeling Temporal Dependencies in High-Dimensional Sequences: Application to Polyphonic Music Generation and Transcription, in Proceedings of the 29th International Conference on Machine Learning (ICML), 2012.

Goal: Generative modeling of polyphonic music

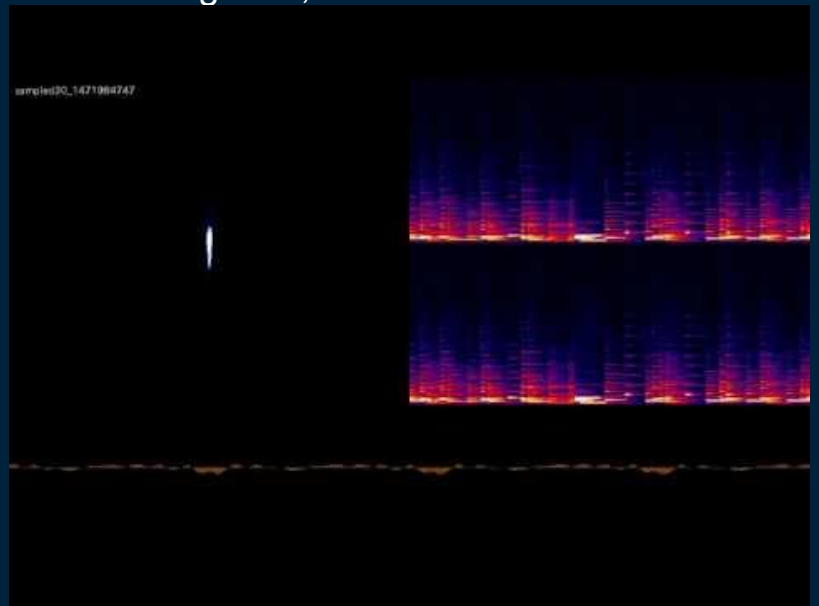
Approach (Week 5-Now): GRUs on MIDI event stream



NOTES (1 of 89)

DURATIONS (1 of 12)

- 1 step ~= 1 second of time on average
- Bach Chorales, 41,198 steps
- Training time, 2m14.780s



Takeaway Lessons / Opinions

- Softmax temperature is a crucial hyperparameter in sampling!
- Condition on ***all*** available information
- Input representation is ***key***
- Better log likelihood does ***not*** mean the samples are better
  - Early stop on validation likelihood worse. EDIT: Was a bug
- Still overparameterized, more parameter sharing could help
- May need beam search, noisy parallel approximate decoding, or other tricks

Pie in the Sky

- **Event sequences are general**
  - Alerts in a datacenter
  - Events on CAN bus (automotive)
  - Portfolio of financial timeseries
  - Appliance use in homes
- **Haven't seen others using this representation**
  - How to compare likelihoods to existing methods?

Thanks!

schuster@

deck@

The Magenta team

The Google Brain team

You, for looking at these slides

kastnerkyle@gmail.com
kastnerkyle on Twitter, GitHub

References

http://people.idsia.ch/~juergen/blues/
From Douglas Eck. LSTM Blues

https://github.com/tensorflow/magenta
From The Magenta team, led by Douglas Eck.

http://deeplearning.net/tutorial/rnnrbm.html
From Nicolas Boulanger-Lewandowski et al. RNN-RBM and RBM-NADE

http://www.hexahedria.com/2015/08/03/composing-music...
From Daniel Johnson. Results speak for themselves

https://soundcloud.com/bachbot
From Feynman Liang. A music bot for harmonization and sampling.

https://arxiv.org/abs/1506.02216
From Chung et. al (including me and Laurent Dinh). Short name VRNN

https://soundcloud.com/graphific/pyotr-lstm-tchaikovsky
From Roelof Pieters. VRNN with some more tricks

https://arxiv.org/abs/1605.07571
A followup to VRNN by Sonderby et. al., strong likelihoods

http://arxiv.org/abs/1412.7927
From Goel et. al. RNN-DBN

https://highnoongmt.wordpress.com/2015/05/22/lisls-stis-recurrent-neural-networks-for-folk-music-generation/
From Sturm et. al. RNN language model on ABC

http://arxiv.org/abs/0705.2011
From Graves et. al. Multidimensional RNN, or MDRNN

https://arxiv.org/abs/1505.00393
From Visin et. al (including me). Multidimensional RNN, called ReNet

https://youtu.be/0qnTaAz-xtQ
From Pachet et. al., Sony CSL. FlowMachines

https://www.youtube.com/watch?v=2kuY3BrmTfQ
From David Cope. EMI

Extra Slides

Previous approaches

- **Tons of Markov chain, HMM, RNN on ABC (Sturm et. al.)**
  - ABC format has limitations, but this is basically ***"cheat mode"*** - try it!
- **Sigmoid out, Bernoulli cross-entropy cost (LSTM Blues, Eck et. al.)**
  - Might need extremely deep output to get good "mixing" for density estimate
- **RNN-RBM (Boulanger-Lewandowski et. al.)**
  - Better density estimate, but need contrastive divergence training
- **RNN-NADE (Boulanger-Lewandowski et. al.)**
  - Direct gradient descent, close to my final approach
- **RNN-DBN (Goel et. al.)**
  - Even better density, need to do layerwise training
- **Multi-dimensional RNN (Graves, or Visin et. al.)**
  - Works well but naive version can be very slow during training
- **VRNN-based (Pieters, base model from Chung et. al.)**
  - Seems awesome, probably complementary to other output variations here

Approach (Week 1-2): Markov chain on ABC

- ABC is cheat mode, in a good way
- Encodes musicality
- abc2midi does automated curation!