weekly report: amino acid seq

Jitian Zhao

University of Wisconsin Madison

October 7, 2019

Outline

- fix temperature issue
- experiments on aligned data (char-rnn)
- next move
- some lessons

temperature

- If temp too low(<0.2), model will generate same character everytime.

Figure: temp=0.0

- Set temperature around 0.4 gives us better result.

Figure: temp=0.5

features of aligned data

- all seqs are of same length
- longer seqs: 300 to 560
- all seqs are aligned: makes sense to consider reconstruction acc

experiments on aligned data

- seqs with no gap outperformed aligned data for char-rnn: generating seqs that make sense instead all '-'
- need to set high temp(>0.8) to avoid generating too many '-'
- aligned data more suitable for transformer
- an example of seqs generated by char-rnn using no gap data:
  t: IWDTFSHTPGRTVNGHTGDVACDHYHRLDSDVALMAEL
  g: IWDTFSHTPGGRVPGRVPGRVPGKGDVACDHYHRYDED

next move

- Run Transformer model on text dataset
- Run Transformer model on a.a. seqs
- Compare performance of transformer and char-rnn
- Compare Transformer's performance on different datasets(aligned and no seq)

lessons learned last week

- Go get a flu shot as soon as possible.
- Use a water-proof bag during rainy days.
- Backup your computer regularly.