

weekly report: amino acid seq

Jitian Zhao

University of Wisconsin Madison

October 22, 2019

Outline

- results on text data
- results on data removed all gaps

generated text

[tributed Project Gutenberg-tm eBooks with only a loose network of volunteer support. Project Gutenberg-tm eBooks are often created from several printed editions, all of which are confirmed as Public Domain in the U.S. unless a copyright notice is include]↓
d a situation to be life, and the court for the fire-trees and associated by the through the houses of the burial Paris, were brought had recovered before him that the consultancy but his streets from the part black to the stone of the world of the Doctor's except that the subject Gutenberg-tm which was at the bless with the shoemaker was no more of the courtyard, and the streets with a group--the moment of this he had more that the the streets of the light with a mind more and differed that the cause in the street, and had ready was the fire drum the mare of one of the triumph of the mailey of t↵

Figure: text generated using a tale of two cities

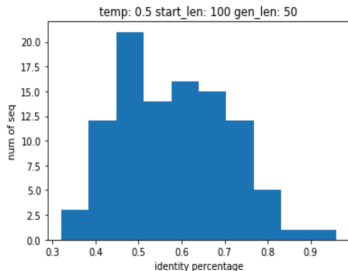
- almost spelling error
- repeatedly "of the" – degenerate problem
- parameters(context, generated length, temperature) could be tuned further

results on data removed all gaps

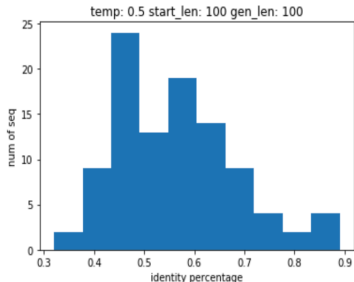
- randomly generated sequences (notebook)
- reconstruction accuracy

properties of transformer generated seqs

- error propagation not severe



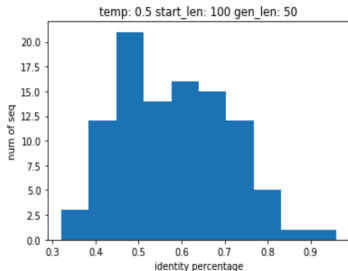
(a) anchor setting



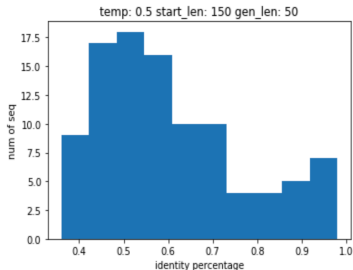
(b) generate length changed to 100

properties of transformer generated seqs

- more information, higher accuracy



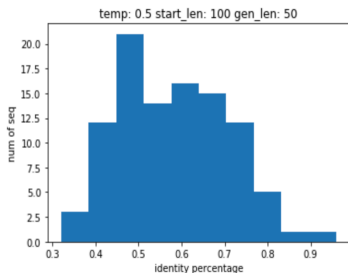
(c) anchor setting



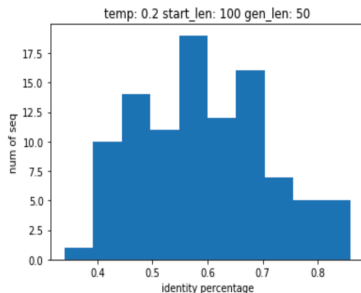
(d) given length changed to 150

properties of transformer generated seqs

- result roughly consistent with temperature (not sampling from a even distribution)
- lower temp, higher accuracy



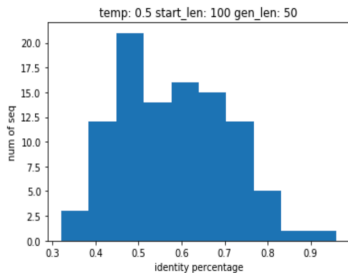
(e) anchor setting



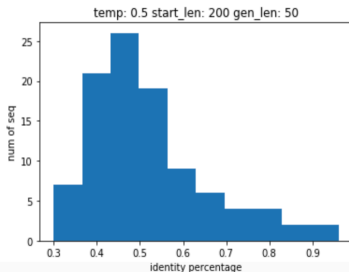
(f) temp changed to 0.2

properties of transformer generated seqs

- predicting end(200-250) not as well as core(100-150)



(g) anchor setting



(h) predicting end