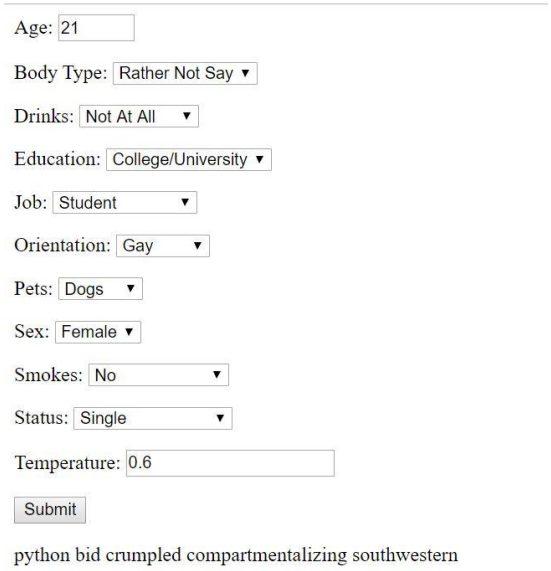
Generating Online Dating Profiles

- Pushpak Raj Gautam

Description

DateGen is a tool that generates an introduction bio for a person based on their attributes like age, sex, social habits, job, education, et cetera.

A successful example (sort of)



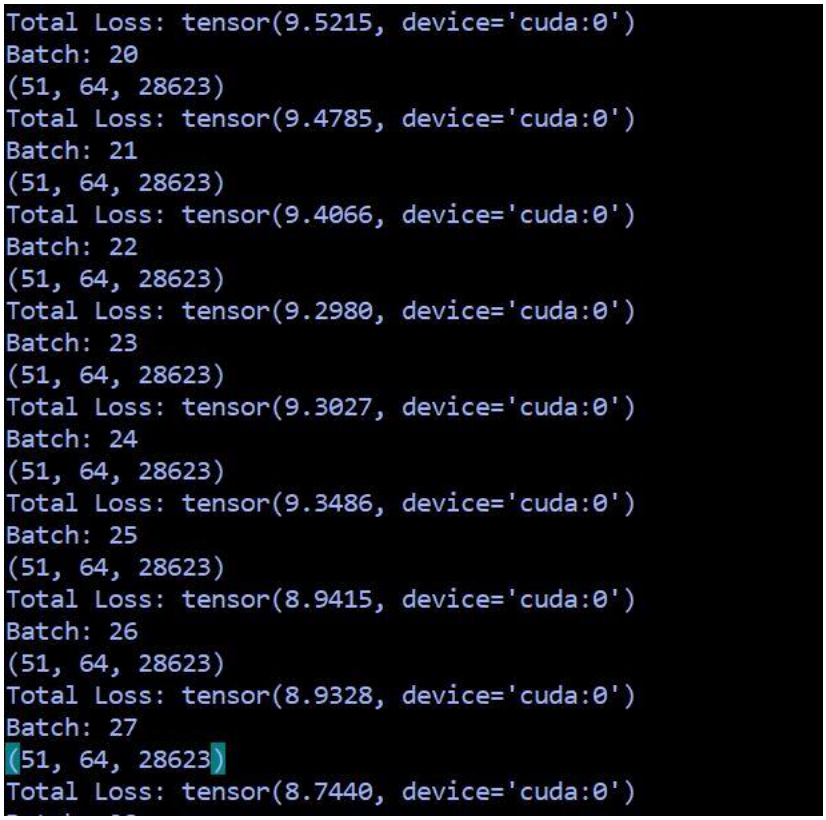
The sentence above doesn't make much sense but is a bit coherent. This example generates a lot of words and clearly gets to <End of Sentence> (word limit is 50). This is an LSTM model and for each word generation, the set of attributes along with the previous word is fed in as input. The model contains a hidden layer and a linear layer.

A failed example

The example above is generated using a temperature of 5.0 (in the softmax layer). Thus, the words are way more random and the output hits the limit of 50 words.

Creativity

The paper that I based my work on used character-level output generation. When I tried this, my output wasn't legible (due to the dataset's small size). So, I changed this to a word level output generation model. By using lower cases and removing punctuation marks, I got to a vocabulary size of around 28k words. The other attributes, when one-hot encoded, gave a 102-dimensional vector. The interesting thing is that I was able to one-hot encode everything (words and attributes) on the GPU and still get a reasonable memory consumption. In the output below (that was generated while training), one can see the shape which corresponds to a word limit of 51 words, a batch of 64 points and an input vector of 28623 elements.



Secondly, I created a web interface for my model so that it could be easier to display the generated output. I did this using Django. Here, I also added a temperature option so that I could show how the model behaves with different degrees of randomness.