“Shannon game for predicting next word”

Venali Sonone

About me

MIT Media Labs
Junior Research Scientist
(April'14 – May'15)

Barclays
Data Analyst
(May'15 – July'15)

Rakuten
Machine Learning Engineer
(June'16 – Feb'18)

UIBS
MBA Grad (Present)

Philip Morris
Data Scientist
(March'18 – Present)

SCIKIT-LEARN

Scikit-Learn is beautiful! It’s just wonderful…..

Everything is a well designed workflow

1. Transform
2. Fit
3. Predict

KERAS

KERAS is very similar to scikit ! It’s just wonderful too…..

Keras API is like lego.

AND

Everything done in scikit can be transferred to keras.

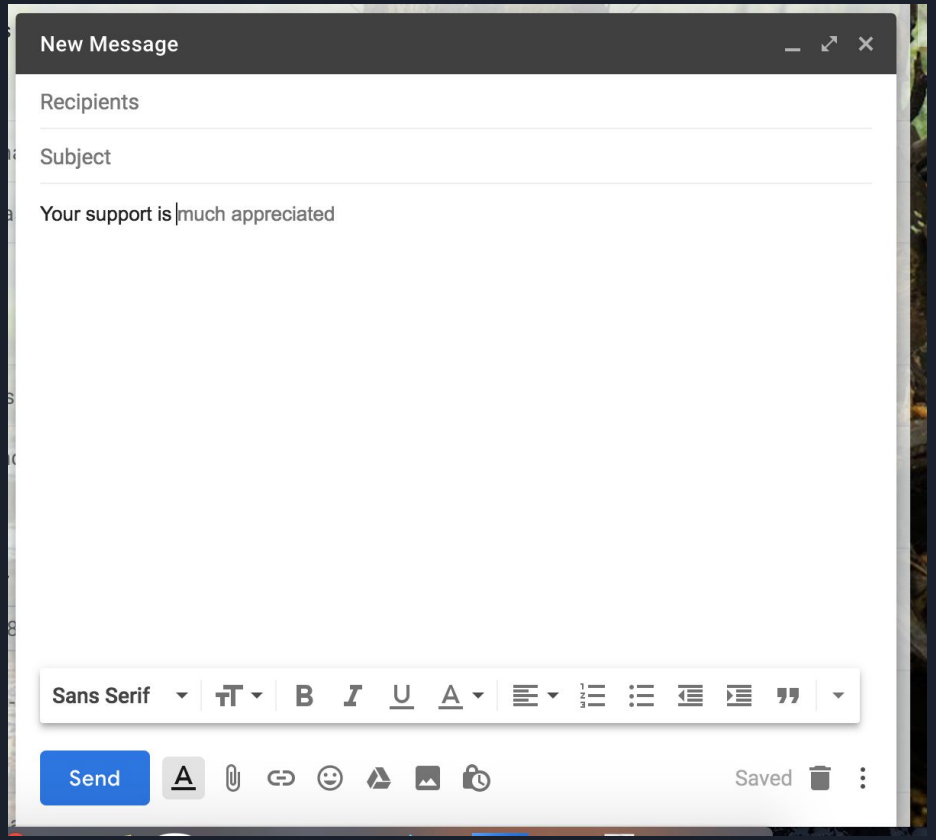
What is Shannon game?

Shannon (1950) estimated the entropy of written English to be between 0.6 and 1.3 bits per character (bpc), based on the ability of human subjects to guess successive characters in text.

So we predict next word!

Why? for fun ;p

Consider recent update in gmail auto word suggestion.



What is sequence-to-sequence learning?

Sequence-to-sequence learning (Seq2Seq) is about training models to convert sequences from one domain to sequences in another domain.

(e.g. sentences in English) => (e.g. the same sentences translated to French)
"the cat sat on the mat" -> [Seq2Seq model] -> "le chat etait assis sur le tapis"

(e.g. sentences in English) => (e.g. sentences in English with some differentiation)
"the cat sat on the" -> [Seq2Seq model] -> "cat sat on the mat"

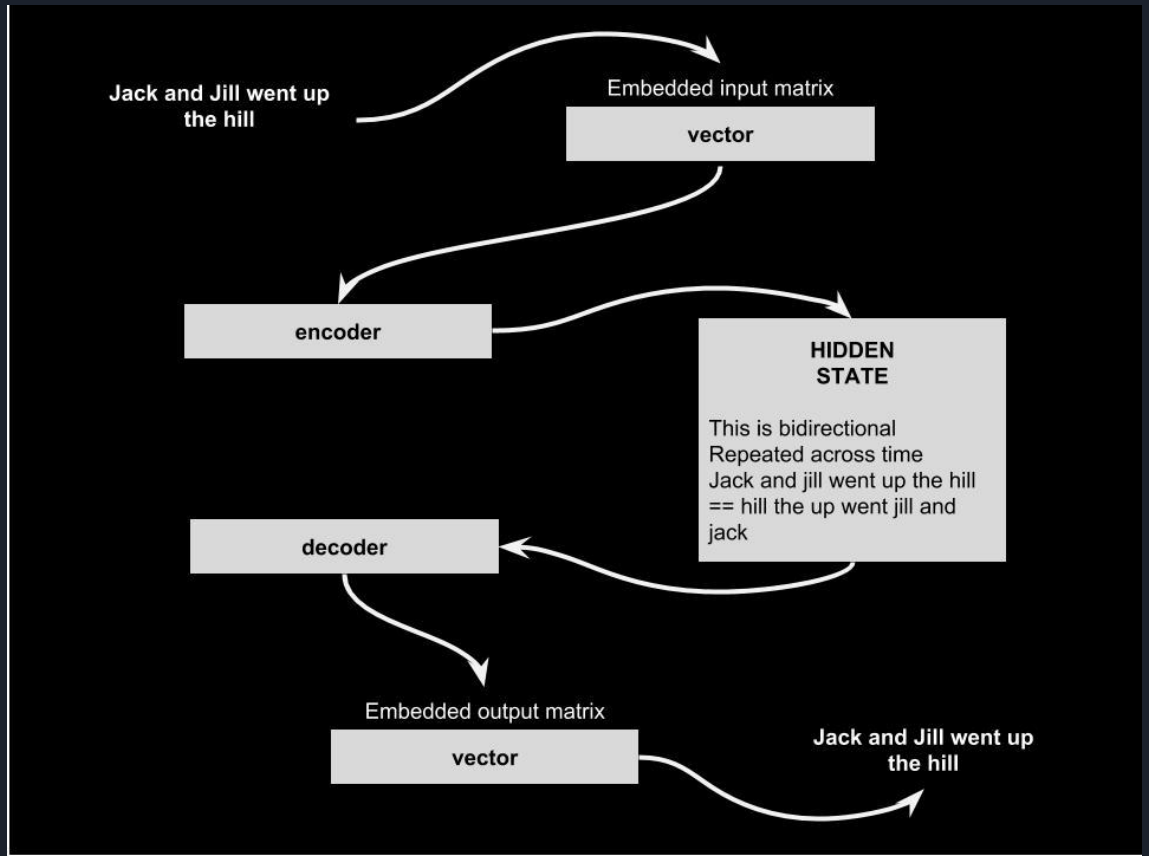
Important stuff

Keras: has 1. Computational graphs 2. Sequential
Loss: 1. Softmax 2. Sigmoid
Optimizer: e.g. adam
Regularization: l1,l2,l1l2
Epoch: mini batches for showing how much data to adjust weights
Padding: fixed length sequence.

8341000000
8347120949

Architecture

It's all about INPUT SEQUENCE <=> OUTPUT SEQUENCE as memory works :)
FUTURE <=> PAST



Links:

Today's Code!! Github: **https://github.com/venali/SquenceKeras.git**

```
In [3]: # evaluate model
        print(generate_seq(model, tokenizer, max_length-1, 'Jack and', 2))
        print(generate_seq(model, tokenizer, max_length-1, 'And Jill', 2))
        print(generate_seq(model, tokenizer, max_length-1, 'fell down', 2))
        print(generate_seq(model, tokenizer, max_length-1, 'pail of', 2))

        Jack and jill went
        And Jill went up
        fell down and broke
        pail of water jack
```

Contact: venalisono@gmail.com medium @venali

Thank you :)

Questions are welcomed !!