Talkative Chatbot

FINAL PROJECT OF ARTIFICIAL INTELLIGENCE FALL 2017
JANUARY 08, 2018

Team Members

Ange CRESPO A06922201
Valentine JALOUX A06922103
Léo NUGRAHA R05922151

DATA MINING

1. How to get the data ?

The **dataset** we used comes from **Reddit**. This name is a play-on-words with the phrase "read it". It is a social new aggregation, web content rating and dicussion website. This site is a collection of entries submitted by its users.

There we can find a lot of data which are stored since 2004.
For our project, we downloaded 2 years of data, from 2014 to 2015.

2. How to storage the data ?

Using **Python** and **MySQL**, we created a **new database** containing only usefull data for our project. Then we had a set of Questions/Answers.

Then this new database is devided in two text files, one contains Questions, the other contains Answers.

3, How do we prepare the data ?

We used tokenization to change questions and answers into usefull words. Tokenization is the process of segmenting running text into words and sentences.

The most important steps of tokenization are :

1. Segmenting text into words
2. Handling abbreviations
3. Handling Hyphenated words

At the end of the tokenization process, we get two sets of tokens that will be sort accordind to them frequency of occurance.

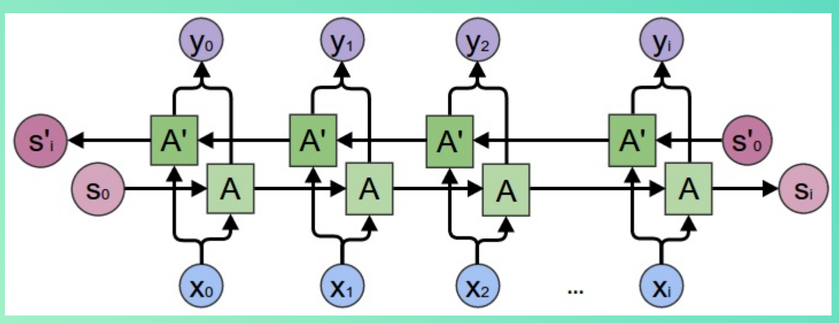
TRAINING

ARCHITECTURE:

By having a left and right accumulating map zipped altogether, the neural network can make predictions over a sequence with both past and future context.

PRINCIPLES:

Split the neurons of a regular RNN into 2 directions; one is for positive direction (forward states) and the other one is for negative direction (backward states).
The 2 outputs from 2 states are not connected to inputs of the opposite direction states.



TRAINING:

Training consists of 3 parts: Forward Pass, Backward Pass, and Weight Update, performed consecutively.

FORWARD PASS	BACKWARD PASS	WEIGHT UPDATE
Run input data through RNN (1) for both forward and backward states (2) for output neurons	Calculate the part of the objective function derivative (1) for output neurons (2) for both forward and backward states	After each epoch, update all weights to ensure that the obtained solution minimizes errors

TESTING

2 bots created
2 hyperparameters
2 behaviours
TESTING

SCORING

SCORING
SCORING
SCORING

REFERENCES:

[1] Kinsley, Harrison. "Creating Chatbot from Deep Learning, Python, and Tensorflow". October 2017
[2] Kukiela, Daniel. "Neural Machine Translation". https://github.com/daniel-kukiela/nmt-chatbot
[3] Olah, Christopher. "Neural Networks, Types, and Functional Programming". September 2015. olah.github.io
[4] Schuster, Mike and Paliwal, Kuldip K. "Bidirectional Neural Network". IEEE Transactions on Signal Processing, Vol 45, No 11. Nov 1997