Análisis de Poemas con Glove y SOM

Esquivel Guzmán

Análisis de Poemas con Glove y SOM

Karla Adriana Esquivel Guzmán

Redes Neuronales

15/Junio/2020

¿Qué es Glove?

Análisis de Poemas con Glove y SOM

Esquivel Guzmán

Glove es un algoritmo utilizado para obtener una representación vectorial de las palabras, esto se logra mapeando las palabras en un espacio dónde la distancia esta relacionada con respecto a su similitud semantica.

Ejemplo:

Análisis de Poemas con Glove y SOM

Tenemos como ejemplo que las palabras más cercanas a la palabra Gato serían:





León



Tigre



Puma





Más Sobre Glove

Análisis de Poemas con Glove y SOM

Karla Adriar Esquivel Guzmán

En este proyecto no se va a generar un Glove, sino más bien se va a utilizar un recurso ya existente que nos proporciona la universidad de Stanford, el cual nos ofrece vectores pre-entrenados de palabras, con tamaños de 50, 100, 200 y 300, este se encuentra en la página como Glove.6b. Que de acuerdo a la información está entrenado utilizando la wikipedia de 2014 y Gigaword5

¿Qué es SOM?

Análisis de Poemas con Glove y SOM

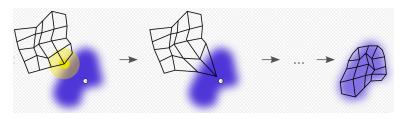
Esquivel Guzmán

Es una Red Neuronal autoasociativa que reprenta vectores de dimensión alta a dimensión baja. A diferencia de otras redes neuronales estas usan competitive learning, esta es una forma de aprendizaje no supervisado en dónde los nodos compiten para responder a un subconjunto de datos de entrada.

Ejemplo:

Análisis de Poemas con Glove y SOM

Karla Adriar Esquivel Guzmán



clic para ver el GIF

MiniSOM

Análisis de Poemas con Glove y SOM

Karla Adriai Esquivel Guzmán

En este proyecto se utiliza MiniSOM https://pypi.org/project/MiniSom/

Explicación de Análisis de Poemas

Análisis de Poemas con Glove y SOM

Karla Adrian Esquivel Guzmán Se obtuvieron los poemas de varios autores del sitio www.poemhunter.com, el proceso para convertir un poema a su valor vectorial es el siguiente: Como primer paso para ejemplificar tomaré el poema

The Appeal

It I have given you delight
By aught that I have done,
Let me lie quiet in that night
Which shall be yours anon:
And for the little, little, span
The dead are born in mind,
Seek not to question other than
The books I leave behind.

Segundo Paso

Análisis de Poemas con Glove y SOM

Karla Adriai Esquivel Guzmán

Hay que tokenizar los poemas para ello se deben eliminar las palabras vacias (stop words) dentro del poema, así que nos quedará la siguiente lista de palabras:

```
['given', 'delight', 'aught', 'done', 'let', 'lie', 'quiet', 'night', 'shall', 'anon', 'little', 'little', 'span', 'dead', 'born', 'mind', 'seek', 'question', 'books', 'leave', 'behind']
```

Segundo Paso

Análisis de Poemas con Glove y SOM

Esquivel Guzmán

```
También debemos eliminar palabras repetidas así que obtendremos la siguiente lista:
['anon', 'aught', 'behind', 'books', 'born', 'dead', 'delight','done', 'given', 'leave', 'let', 'lie', 'little', 'mind', 'night', 'question', 'quiet', 'seek', 'shall', 'span']
```

Tercer Paso

Análisis de Poemas con Glove y SOM

Karla Adrian Esquivel Guzmán

Una vez tokenizado el poema obtendremos el vector asociado a cada una de las palabras del Poema por ejemplo para la palabra books el vector asociado es el siguiente:

```
(Pdb) glove['books']
array([-0.03345 , 0.80877 , -0.20659 , -0.92842 , 0.27433 , 0.17549 ,
-1.5501 , -1.9302 , 0.42939 , 0.15984 , -0.9263 , 1.1073 ,
-0.15518 , -0.3959 , 1.431 , -0.74583 , -0.33683 , -0.21206 ,
0.15386 , -0.022504 , 1.127 , 0.31919 , 0.83898 , 0.48734 ,
0.2966 , -0.81942 , -1.5519 , -0.77488 , -0.28272 , -0.71643 ,
2.6664 , -0.57047 , 0.10581 , -0.10116 , -0.48662 , 0.63422 ,
-0.60514 , 0.19788 , -0.25741 , -0.20731 , 0.95538 , -0.10275 ,
0.41266 , 0.7293 , -0.29704 , 0.60264 , 0.29637 , -0.021069 ,
-0.65104 , -0.58676 ])
```

Cuarto Paso

Análisis de Poemas con Glove y SOM

Karla Adrian Esquivel Guzmán Una vez obtenido el vector asociado para cada palabra, sacando el promedio de los vectores obtendremos el vector asociado para el poema:

```
Pdb) poemsVector[0]
rray([ 0.20987075,
                   0.22164867, -0.19477876, -0.26669845,
                                                          0.2442112 .
      0.03558175, -0.03955173,
                                0.05628565, -0.13514657,
                                                          0.21897788,
      -0.11065322, 0.00523873,
                                0.02434767, 0.08715873,
                                                         0.38328473,
      0.21781614. 0.04998812.
                                0.03060938. -0.10353592. -0.4575028 .
      -0.06806463. 0.24727543.
                                0.26713475. -0.1051461 .
                                                         0.30859751.
      -0.66750971, -0.31886267, 0.43215798, 0.50420189,
                                                         0.00996084.
      1.84621696, 0.15536525, -0.01349627, -0.24265808, -0.12012784,
      0.14951236, -0.02783506,
                                0.16087743, 0.06481896, -0.31197598,
      -0.12879216, -0.03553098, -0.03691021, 0.16485148,
                                                          0.18672237,
      0.09209333. 0.11499473. -0.23732449. -0.11618696. -0.025607351)
```

Proceso de Experimentación

Análisis de Poemas con Glove y SOM

Karla Adriana Esquivel Guzmán Le aplicamos el proceso de vectorización antes mencionado a 60 poemas de 3 autores distintos, lo que nos da un total de 180 poemas vectorizados y estos vectores se alimentaron a una implementación de SOM (Self-Organizing Map), MiniSOM.

```
#Convertimos el arreglo en un arreglo de numpy
poemsVector = np.array(poemsVector)

#Definmos el tama;o del grid al que mapeare minisom
map_dim = 16
#Definimos el tama;o de entrada de los vectores
tamEntrada = 50
#Definimos la funcion de minisom
som = Minisom(map_dim, map_dim, tamEntrada, sigma=1.0, random_seed=1)
#som.random_weights_init(poemsVector)
#Entrenamos la red neuronal con W que es el el arreglo de los revectores que
#representan los poemas
som.train_batch(poemsVector, num_iteration=len(poemsVector)*500, verbose=True)
```

Análisis de Poemas con Glove y SOM

Esquivel Guzmán

De experimentar con el conjunto de autores de una época similar:

1809-1849 Edgar Allan Poe(Amarillo), 1770-1850 William Wordsworth(Azul), 1757-1827 William Blake(Rojo)

Análisis de Poemas con Glove y SOM

Karla Adriar Esquivel Guzmán

```
| Company | Comp
```

Análisis de Poemas con Glove y SOM

Esquivel Guzmán

> Conjunto de escritores de la misma nacionalidad, en este caso Reino Unido 1608-1674 John Milton(Amarillo), 1770-1850 William Wordsworth(Azul), 1564-1616 William Shakespeare(Rojo),

Análisis de Poemas con Glove y SOM

Karla Adriar Esquivel Guzmán



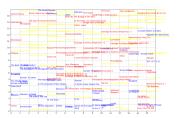
Análisis de Poemas con Glove y SOM

Esquivel Guzmán

> Conjunto de escritores del mismo pais USA 1819-1892 Walt Whitman(Amarillo), 1809-1849 Edgar Allan Poe(Rojo), 1885-1972 Ezra Pound(Azul)

Análisis de Poemas con Glove y SOM

Karla Adriana Esquivel Guzmán



Análisis de Poemas con Glove y SOM

Esquivel Guzmán

> Conjunto de escritores diferentes épocas y diferetes nacionalidades 1564-1616 William Shakespeare(Amarillo), Reino Unido, 1865-1936 Rudyard Kipling(Azul), India, 1932-1963 Sylvia Plath(Rojo), Estados Unidos

Análisis de Poemas con Glove y SOM

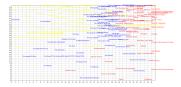
Karla Adrian Esquivel Guzmán

Utilizando un vector asociado de tamaño 50 para cada Poema y una "malla" (cuadrícula) de 16*16

Análisis de Poemas con Glove y SOM

Karla Adrian Esquivel Guzmán

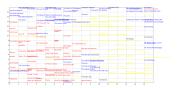
Utilizando un vector asociado de tamaño 50 para cada Poema y una "malla" de 40*40



Análisis de Poemas con Glove y SOM

Karla Adrian Esquivel Guzmán

Utilizando un vector asociado de tamaño 100 para cada Poema y una malla de 16*16



Análisis de Poemas con Glove y SOM

Karla Adrian Esquivel Guzmán

Utilizando un vector asociado de tamaño 100 para cada Poema y una malla de 40*40

Análisis de Poemas con Glove y SOM

Karla Adrian Esquivel Guzmán

Utilizando un vector asociado de Tamaño 300 para cada Poema y una malla de 16*16

```
| The state of the
```

Análisis de Poemas con Glove y SOM

Karla Adrian Esquivel Guzmán

Utilizando un vector asociado de tamaño 300 para cada Poema y una Malla de 40*40