



Dale, dale DATA dale

Fernando Villanueva Arce

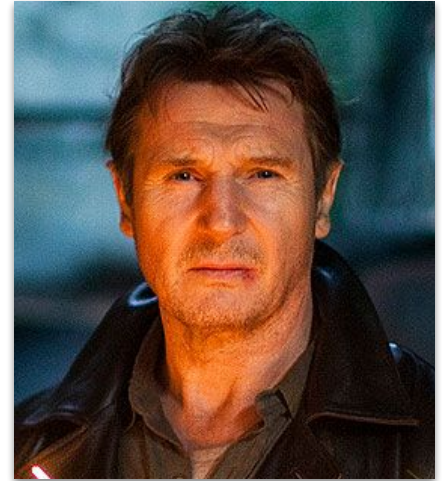
Pero algo ocurrió en la guerra de katas...

```
def __init__(self, *args, **kwargs):
    self.file = None
    self.fingerprints = set()
    self.logdupes = True
    self.debug = debug
    self.logger = logging.getLogger(__name__)
    if path:
        self.file = open(os.path.join(path,
            self.file_name), 'a')
        self.fingerprints.update([self.file_name])

    @classmethod
    def from_settings(cls, settings):
        debug = settings.getbool('debug', False)
        return cls(job_dir(settings), debug)

    def request_seen(self, request):
        fp = self.request_fingerprint(request)
        if fp in self.fingerprints:
            return True
        self.fingerprints.add(fp)
        if self.file:
            self.file.write(fp + os.linesep)

    def request_fingerprint(self, request):
        return request_fingerprint(request)
```



El generador de reggaeton

Generación de letras de reggaeton mediante una red neuronal recurrente (RNN), en concreto LSTM (Long Short-Term Memory)

¿Y por qué?

So cool!

- Texto predictivo en Whatsapp
- Autocompletado en Gmail

Mensaje nuevo

Destinatarios

Asunto

Hola,

Gracias por tu respuesta



¿Y por qué?

≡ EL PAÍS

TECNOLOGÍA

SUSCRÍBETE

Cómo puede la inteligencia artificial detectar y crear noticias falsas

Sistemas como los algoritmos de aprendizaje automático pueden evaluar si un texto o un titular coinciden con el contenido real de un artículo



¿Y por qué?



Porque si ellos pueden, mi red neuronal también.

¿Y cómo se hace
ehto brodel?

Keras

The Python
Deep Learning library

- **import** keras libraries
 - **load** the dataset
 - **split** into X and y variables
 - **define** the model
 - **compile** the model
 - **fit** the model
 - **predict** the lyrics
-

El dataset

Fuente: <https://www.letras.com/mais-acessadas/reggaeton/>

Faltas ortográficas

No puedo seguir haci
Lying to you, you said you'd do it for me
Sigues preguntantdo porque

Spanglish

But I still feel cold por adentro
En mis brazos, while you're sleeping

English???

From yuh neva lie wid annada man inna bedâ€
(tell dem mi seh nuttin no go so)

Otros idiomas

Si facile d'approcher mais
Difficile de t'atteindre
On est si différent

Gyothe issodo nege sarangur marhedo
gude maumun nege iji anhajyo
haruharu ga ibyorur hyanghe godnun
surphun shigan duriojyo

Que agora vai sobrar então (o quê? O quê?)
Un pedacito a cada nena

Split X, y

Raw

```
<p>Se acostó temprano, mañana hay que estudiar<br>Pero llamó a la amiga  
diciendo pa' janglear<br/>Tiene un culito ahí que le acabó de
```

Arrays
de strings

sentence	next_char
"se acostó temprano mañana hay que estudi"	"a"
"e acostó temprano mañana hay que estudia"	"r"

Cada char se
puede vectorizar

Vectores

X

Array 3D booleano:

- Posición de sentence
- Posición del char
- N° equivalente al char

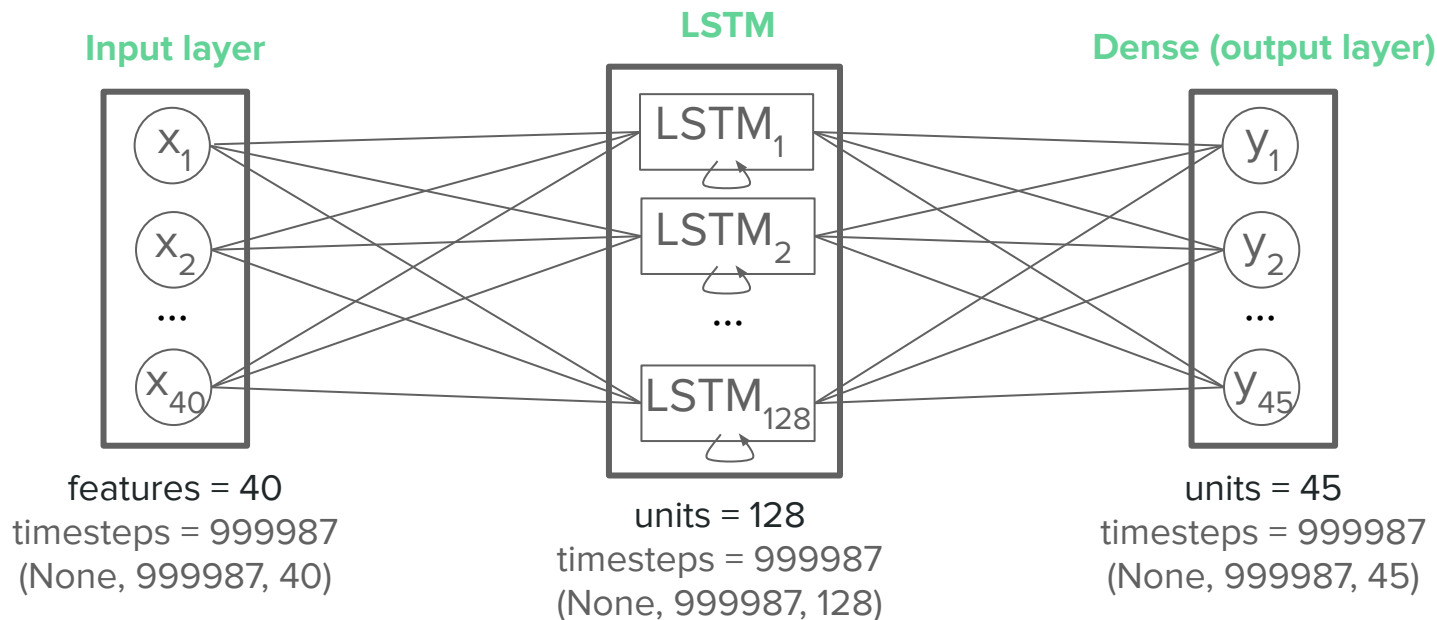
y

Array 2D booleano:

- Posición de sentence
- N° equivalente al char

Estructura del modelo (RNN)

En Keras, un modelo se define como una secuencia de layers, que vamos añadiendo una a una hasta conseguir nuestra red neuronal.



Entrenamiento del modelo

Mínimos*

- Epochs > 20
- Corpus > 100.000 caracteres
- GPU

* Recomendados por Keras en su documentación para Text Generation

Proyecto

- Epochs = 40
- Corpus = 3.000.000 caracteres**
- GPU (Google Colab)

** De 15.000.000 de caracteres posibles

Entrenamiento del modelo

Coherence

Epoch 1/40

999987/999987 [=====] - 765s 765us/step - loss: 1.9979 - acc: 0.4011

Epoch 2/40

999987/999987 [=====] - 752s 752us/step - loss: 1.6977 - acc: 0.4867

Epoch 3/40

999987/999987 [=====] - 746s 746us/step - loss: 1.5958 - acc: 0.5169

...

Epoch 37/40

999987/999987 [=====] - 757s 757us/step - loss: 1.3221 - acc: 0.5989

Epoch 38/40

999987/999987 [=====] - 757s 757us/step - loss: 1.3216 - acc: 0.5988

Epoch 39/40

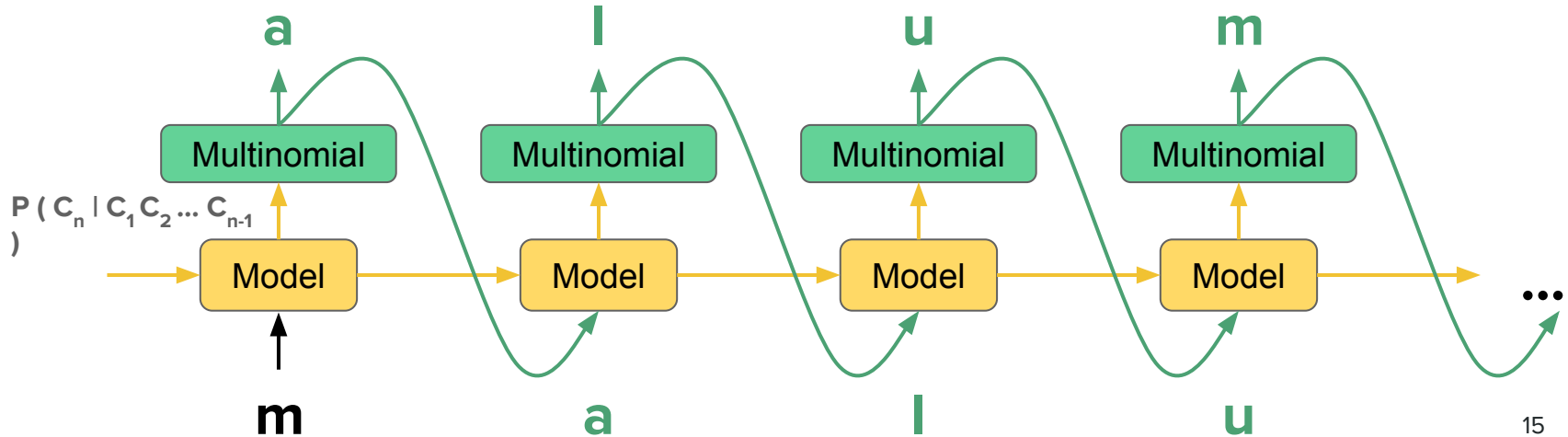
999987/999987 [=====] - 754s 754us/step - loss: 1.3205 - acc: 0.5997

Epoch 40/40

999987/999987 [=====] - 750s 750us/step - loss: 1.3196 - acc: 0.5995

Predicciones con LSTM

El modelo se retroalimenta y va obteniendo más contexto.



¿Y qué dise
el generadol?

Epoch 25

**“derrotando la clave la rumba con mi
sand”ule**

y yo soy un poco de mi

y yo soy culo

y es que estas contigo

y yo soy el que se acabe este pase

tu me pides sin ti no me desesperas

y yo soy tu cuerpo

y yo soy maluma

me da con un beso el cielo

y yo soy tu cuerpo me da control

y a tu amor si te vas a ver

tu eres mía mía

tú no me quieres que te desesperas

Epoch 31

**“a darte
chica voy azotarte
darte duro “**

tu cuerpo me llama

y yo soy un carajo el carajo

y a tu corazón de mi corazón

y si te pido que se acaba la vida

para que te pongas conmigo

y si te pongo en mi cama

y yo soy un party

si tu me muero

y yo soy tu cara de mi cara

y te vas a mí

y yo soy un party en la cama

está es la calle es la calle

Epoch 36

**"a duro buscate un buen seguro dale mas
m"ami**

que esta noche yo te voy a dar
y si tú me pones de lo que sientes que no
es pa mi
de la cama me dice que no es un poco
y es que tu eres mi vida
por qué te voy a montar
si tu quieres mami que la verdad
yo te lo da mucha control
es que tu me estás con la pista
y el corazón mi corazón
y si te vas a estar con ella
que se me vuelve en mi corazón
y es que me dejas tu cara

Epoch 40

**"rir de amor a que el amor me falte
en c"alenaza**

y se que te vas a decir
que esta noche me enamora
si se prende el corazón
el calor me dejo el paso
con esta pena de la calle
y si esta caliente
esta noche me pone mal
si te pone me prendes que te voy a dar
pero te voy a amar
que si te quiero te vas
si te vas a descontrolar
y te voy a dar

Ahora, pasito a pasito...

- Epochs > 100, 1000...
- Corpus > 15M caracteres
- Hiperparámetro *temperature* (0.3, 0.4...)
- ¿Por palabras en lugar de por caracteres?
- Añadir audio y ritmo
- GUI (Streamlit, Tkinter)



GUI (Streamlit)



Desde Pueltorico
con amol.
¡A regueteoneal!