



Deep Learning para el Reconocimiento de Texto Manuscrito

Trabajo Fin de Grado

**¿Quién es este
chiquillo?**

Índice

- **Introducción**
- **Datos**
- **Redes Neuronales Artificiales**
- **Arquitecturas**
- **Pruebas y Resultados**
- **Conclusiones y Trabajo Futuro**

Introducción

Introducción

Motivación

- Introducción
- Datos
- Redes Neuronales Artificiales
- Arquitecturas
- Pruebas y Resultados
- Conclusiones y Trabajo Futuro

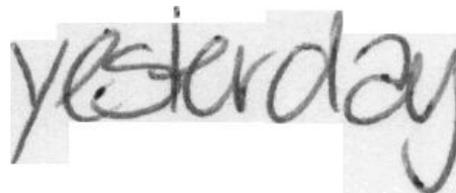
Reconocimiento Óptico de Caracteres (OCR)

- Caracteres definidos e invariantes.
- Segmentación y transcripción.

Aa Bb Cc Dd Ee Ff Gg Hh Ii Jj
Kk Ll Mm Nn Ññ Oo Pp Qq Rr
Ss Tt Uu Vv Ww Xx Yy Zz
< [“ (& i * # \$ % & / = ’ ?) } ”] >
0 1 2 3 4 5 6 7 8 9

Reconocimiento Óptico de Manuscritos (OHR)

- Caligrafía arbitraria y variante.
- Difícil segmentación.



Introducción

Objetivo

- Introducción
- Datos
- Redes Neuronales Artificiales
- Arquitecturas
- Pruebas y Resultados
- Conclusiones y Trabajo Futuro

Sistema OHR basado en Deep Learning.



Mediante entrenamiento, el sistema “aprende a leer” palabras completas.

Datos

Datos

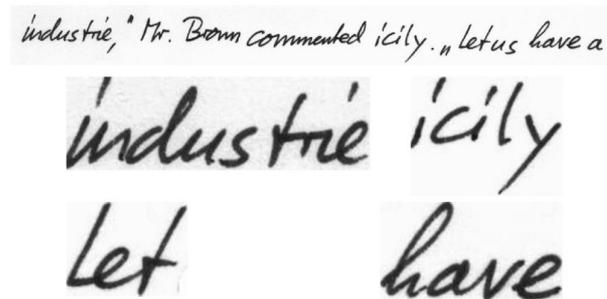
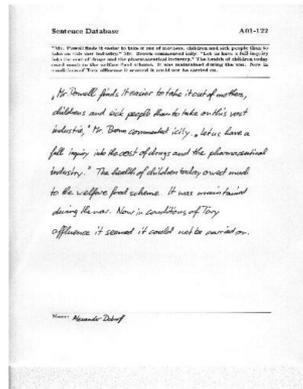
Origen

IAM Handwriting Database de la Universidad de Berna

- 1.539 páginas manuscritas en inglés.
- 657 escritores.
- 87.107 muestras.

Caracteres (Etiquetas)

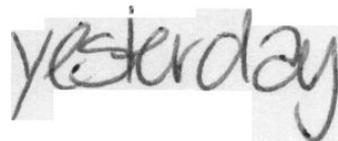
- Números
- Mayúsculas
- Minúsculas
- 62 símbolos en total.



Datos

Tratamiento

- **Input: Tensores orden 3 que representan las imágenes.**
- **Output: Secuencias de enteros que representan los símbolos.**

A handwritten word "yesterday" in black ink on a white background. The word is written in a cursive style. A light gray grid is overlaid on the word, with each cell containing a small number, representing a tensor representation of the image.

```
255 255 255 255 255 255 255 251 251 251 145 251 251 251 251 255 255 255
255 255 255 255 255 255 255 250 250 88 251 251 251 251 251 255 255 255
255 255 255 255 255 255 255 251 248 224 251 249 251 251 251 255 255 255
255 255 255 255 255 255 255 250 174 249 251 251 251 251 251 251 183 251
251 250 106 250 62 251 57 250 70 241 225 251 156 136 247 214 248 251
166 251 57 248 101 242 152 231 227 51 248 228 165 65 251 49 241 72
73 250 54 88 251 111 247 67 69 43 247 236 62 243 160 70 52 63
141 84 119 238 130 217 251 54 230 222 46 69 55 50 234 59 240 65
229 245 245 74 240 248 251 255 255 255 255 255 255 255 250 235 221
```

yesterday



'60', '40', '54', '55', '40', '53', '39', '36', '60'

Datos

Preprocesado para las Imágenes

- Invierten tonos.
- Reescalado guardando las proporciones.
- Relleno tras la palabra.
- Normalización del tensor asociado.

- Introducción
- Datos**
- Redes Neuronales Artificiales
- Arquitecturas
- Pruebas y Resultados
- Conclusiones y Trabajo Futuro

Datos

Preprocesado para las Imágenes

- Introducción
- Datos**
- Redes Neuronales Artificiales
- Arquitecturas
- Pruebas y Resultados
- Conclusiones y Trabajo Futuro



```
245 240 247 232 72 245 245 241 246 241 249 249 255 255 255 255 255 255
255 255 255 255 255 255 255 255 255 247 241 241 235 53 231 247 248 243
191 242 243 255 237 247 240 246 250 247 240 38 243 255 255 255 255 255
219 62 73 67 44 47 240 178 92 170 32 37 255 59 236 237 68 249
243 234 46 134 78 249 240 96 250 66 238 236 117 93 76 42 107 245
234 81 36 255 96 94 237 62 243 242 245 235 37 60 224 222 35 236
65 168 52 224 230 228 204 37 233 237 235 34 255 234 52 54 117 58
237 243 88 210 236 51 68 189 36 230 223 240 237 241 242 231 209 46
45 114 237 255 239 241 241 238 223 51 54 235 233 239 236 246 244 35
```

**Imagen con la palabra “down” y su tensor asociado.
(Representación simbólica, no exacta)**

Datos

Mapeo de caracteres

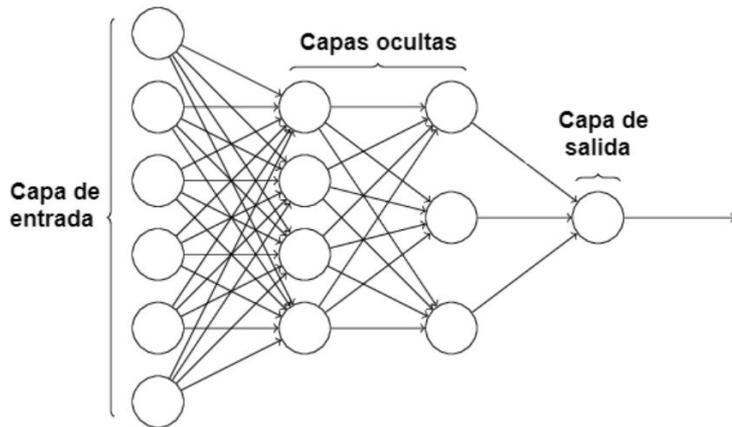
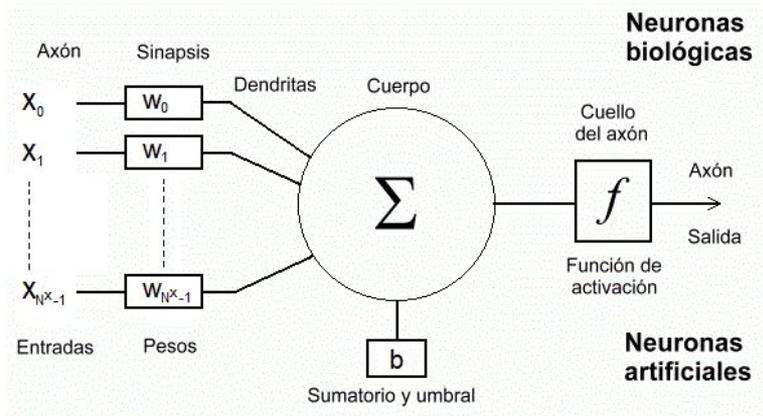
0	0		<i>A</i>	10		<i>K</i>	20		<i>U</i>	30		<i>e</i>	40		<i>o</i>	50		<i>y</i>	60
1	1		<i>B</i>	11		<i>L</i>	21		<i>V</i>	31		<i>f</i>	41		<i>p</i>	51		<i>z</i>	61
2	2		<i>C</i>	12		<i>M</i>	22		<i>W</i>	32		<i>g</i>	42		<i>q</i>	52			
3	3		<i>D</i>	13		<i>N</i>	23		<i>X</i>	33		<i>h</i>	43		<i>r</i>	53			
4	4		<i>E</i>	14		<i>O</i>	24		<i>Y</i>	34		<i>i</i>	44		<i>s</i>	54			
5	5		<i>F</i>	15		<i>P</i>	25		<i>Z</i>	35		<i>j</i>	45		<i>t</i>	55			
6	6		<i>G</i>	16		<i>Q</i>	26		<i>a</i>	36		<i>k</i>	46		<i>u</i>	56			
7	7		<i>H</i>	17		<i>R</i>	27		<i>b</i>	37		<i>l</i>	47		<i>v</i>	57			
8	8		<i>I</i>	18		<i>S</i>	28		<i>c</i>	38		<i>m</i>	48		<i>w</i>	58			
9	9		<i>J</i>	19		<i>T</i>	29		<i>d</i>	39		<i>n</i>	49		<i>x</i>	59			

Redes Neuronales Artificiales

Redes Neuronales Artificiales

Conceptos básicos.

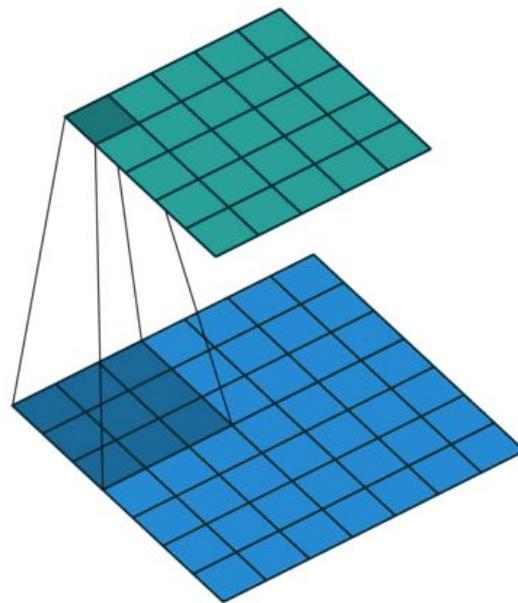
- Entradas.
- Umbral y Pesos ajustables.
- Función de activación.
- Salida propia.
- Estructuradas en capas.
- Transmiten la señal de forma dirigida.
- Descenso del Gradiente
- Retropropagación del Error



Redes Neuronales Artificiales

Red Neuronal Convolutacional

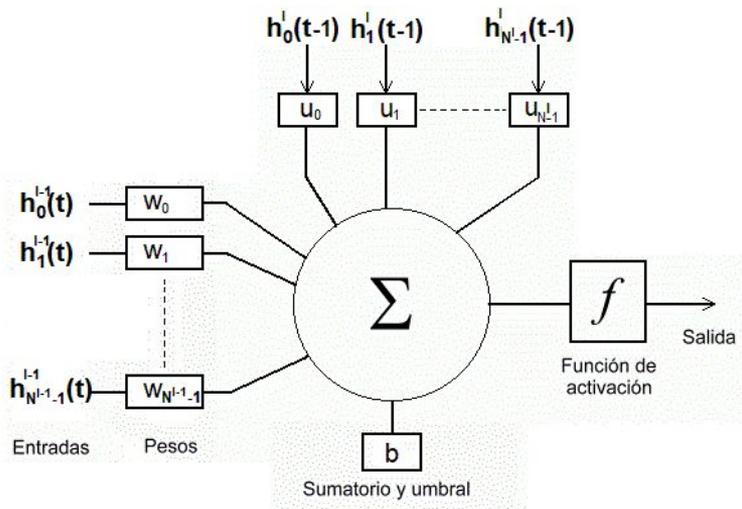
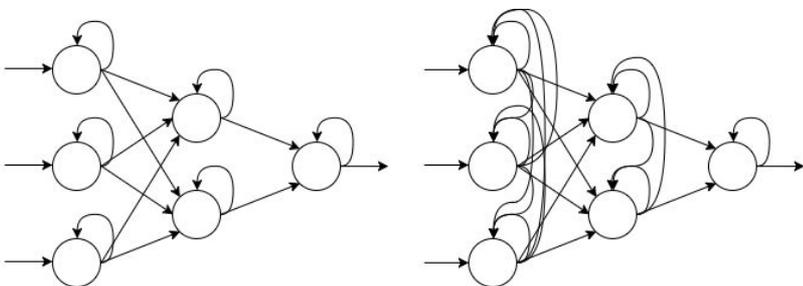
- **Convolución de tensores en dos dimensiones.**
- **Valor, posición y vecindad.**
- **Filtros actúan como núcleos de convolución.**
- **Todos los filtros operan sobre la misma entrada.**
- **Se concatenan las salidas generadas por cada filtro.**



Redes Neuronales Artificiales

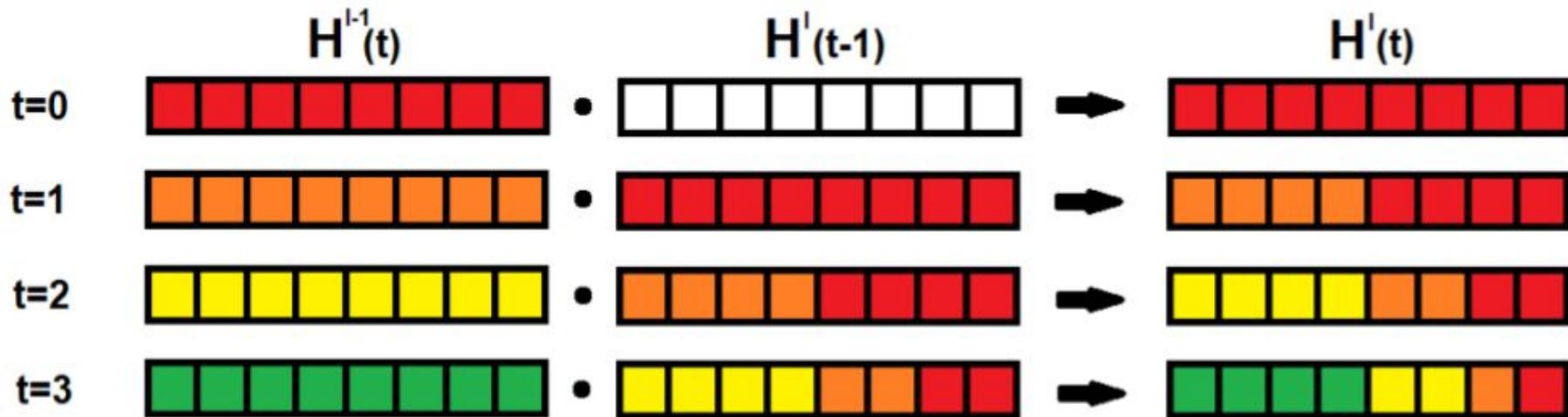
Red Neuronal Recurrente

- Ciclos en las interconexiones de la red.
- Tienen “memoria”.
- Retropropagación a través del tiempo
- Trabajan con secuencias de una limitada longitud.



Redes Neuronales Artificiales

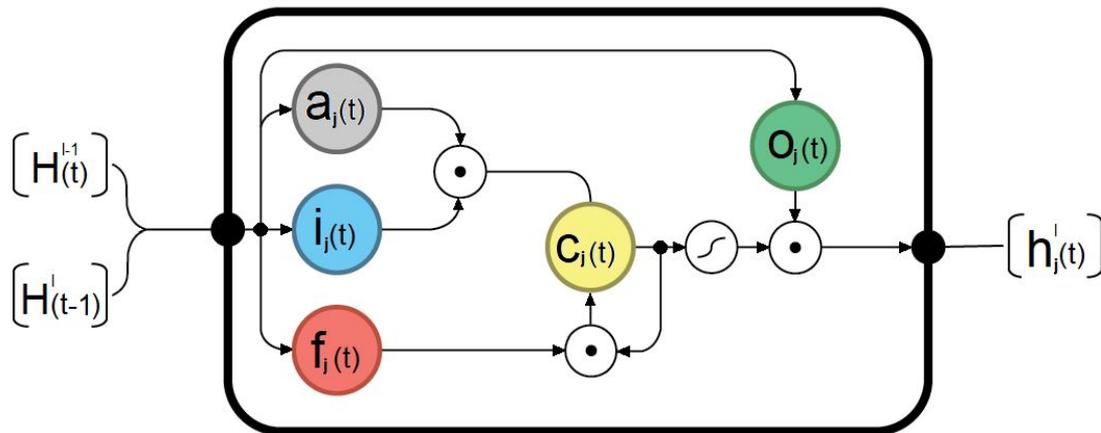
Desvanecimiento del Gradiente



Redes Neuronales Artificiales

Long Short Term Memory

- Permiten trabajar con secuencias de mayor longitud.
- Neuronas sustituidas por celdas LSTM, más complejas.
- Cada celda tiene un estado interno y unas determinadas puertas.



Redes Neuronales Artificiales

Connectionist Temporal Classification

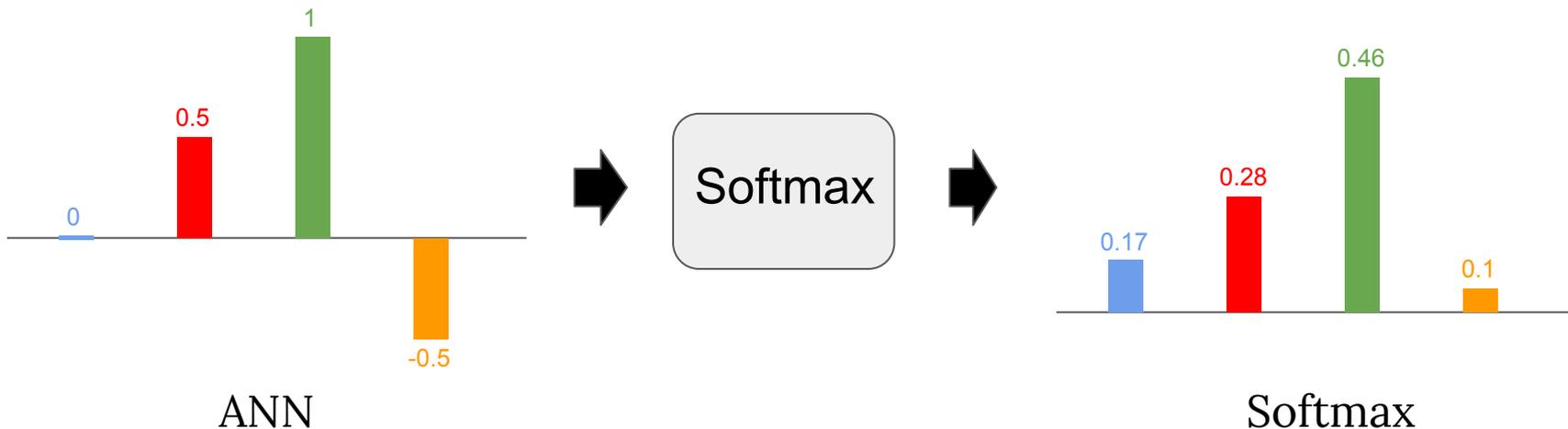
- Patrones entre secuencias de distinta longitud.
- ANN previa.
- Alfabeto = etiquetas + “blanco”.



Redes Neuronales Artificiales

Softmax

- Distribuciones de probabilidad condicionadas.
- Tantas salidas como entradas.

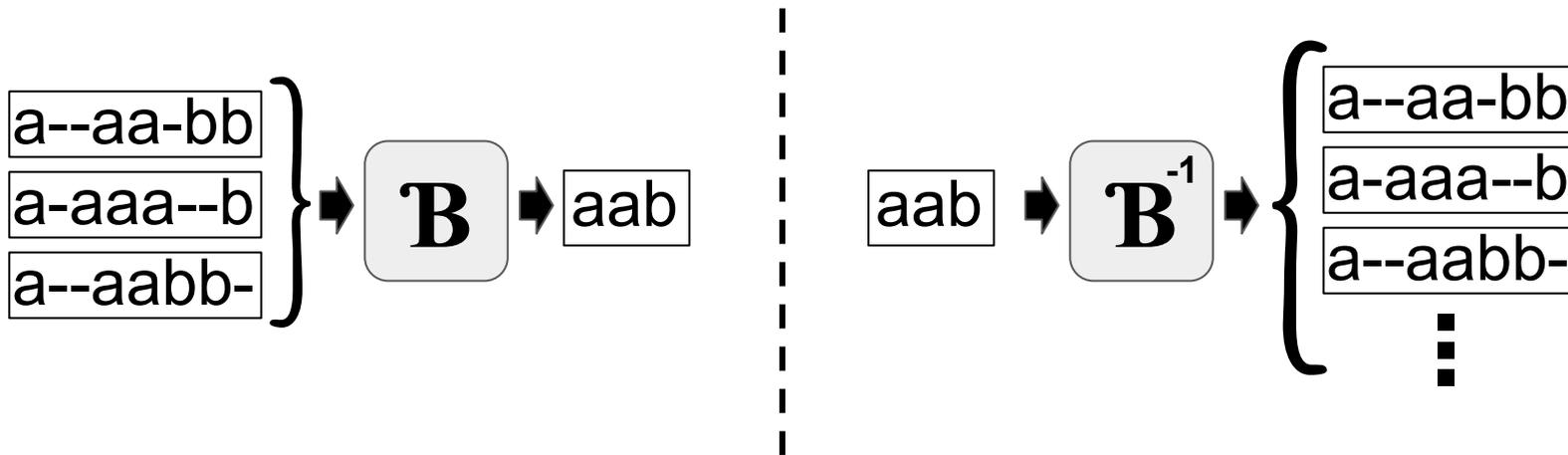


Redes Neuronales Artificiales

Función **B**: Ruta a Secuencia de Etiquetas

Mapeo *Many to one*:

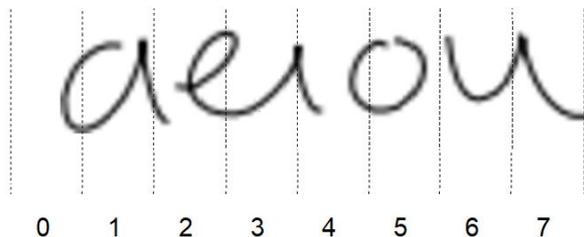
1. Eliminar las etiquetas repetidas consecutivas.
2. Eliminar las etiquetas en blanco.



Redes Neuronales Artificiales

ANN-Softmax

- Introducción
- Datos
- Redes Neuronales Artificiales**
- Arquitecturas
- Pruebas y Resultados
- Conclusiones y Trabajo Futuro



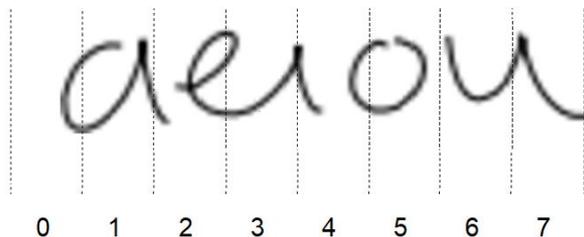
ANN
Softmax

a	0.35	0.50	0.10	0.05	0.10	0.20	0.10	0.05
e	0.10	0.05	0.55	0.20	0.10	0.10	0.05	0.05
i	0.05	0.10	0.05	0.45	0.35	0.05	0.05	0.10
o	0.30	0.20	0.05	0.10	0.15	0.50	0.05	0.05
u	0.10	0.10	0.05	0.15	0.05	0.10	0.55	0.25
-	0.10	0.05	0.20	0.05	0.25	0.05	0.20	0.50
	0	1	2	3	4	5	6	7



Redes Neuronales Artificiales

Rutas



ANN
Softmax

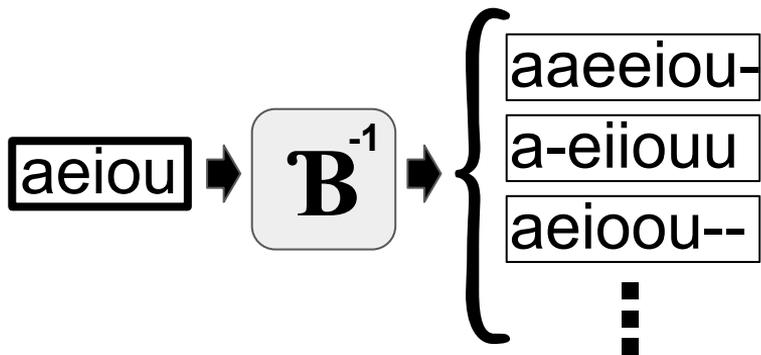
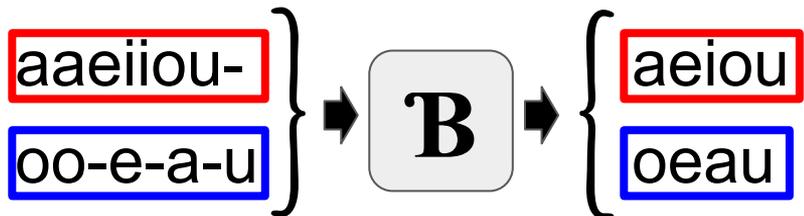
a	0.35	0.50	0.10	0.05	0.10	0.20	0.10	0.05
e	0.10	0.05	0.55	0.20	0.10	0.10	0.05	0.05
i	0.05	0.10	0.05	0.45	0.35	0.05	0.05	0.10
o	0.30	0.20	0.05	0.10	0.15	0.50	0.05	0.05
u	0.10	0.10	0.05	0.15	0.05	0.10	0.55	0.25
-	0.10	0.05	0.20	0.05	0.25	0.05	0.20	0.50
	0	1	2	3	4	5	6	7

Rutas {
aaeiou-
oo-e-a-u



Redes Neuronales Artificiales

Decodificación



Prob. de Ruta:

$$\prod \text{Prob. Símbolo}$$

Prob. Secuencia de Etiquetas:

$$\sum \text{Prob. Ruta}$$

Redes Neuronales Artificiales

Función Objetivo

Función de Verosimilitud:

Max. $Pr(\text{Sec. Correcta} \mid \text{Entrada})$

Función de Pérdidas:

Min. $-\ln(Pr)$

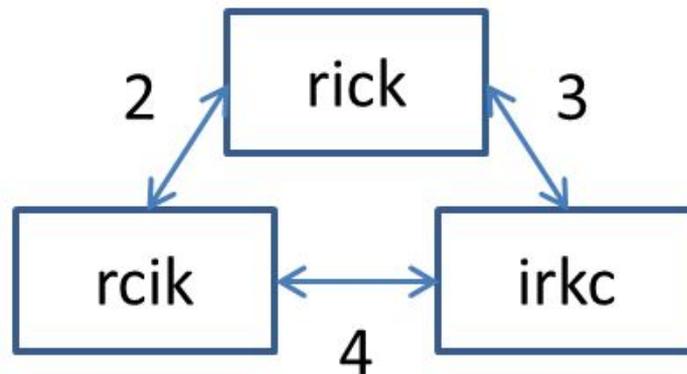
- Introducción
- Datos
- Redes Neuronales Artificiales**
- Arquitecturas
- Pruebas y Resultados
- Conclusiones y Trabajo Futuro

Redes Neuronales Artificiales

Tasa de errores de etiqueta (LER)

Distancia de edición o distancia de Levenshtein

- Inserciones
- Sustituciones
- Eliminaciones



Arquitecturas

Arquitecturas

Arquitecturas especializadas.

Introducción

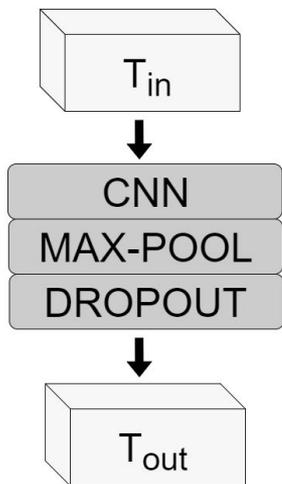
Datos

Redes Neuronales Artificiales

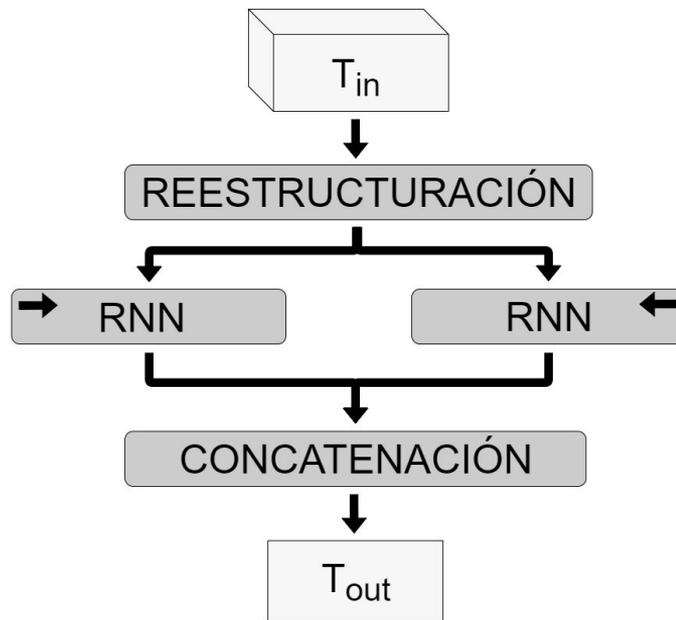
Arquitecturas

Pruebas y Resultados

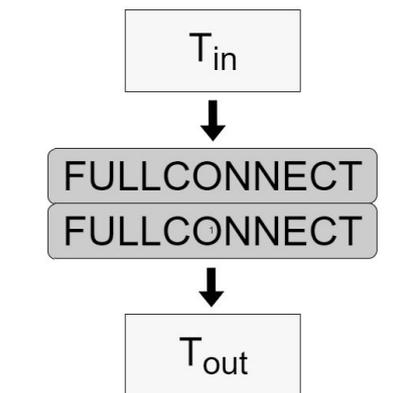
Conclusiones y Trabajo Futuro



Arquitectura para
Extracción de Características



Arquitectura Recurrente Bidireccional



Arquitectura para
Reducción de Dimensionalidad

Arquitecturas

Arq. Extracción de Características

Capa Convolutiva

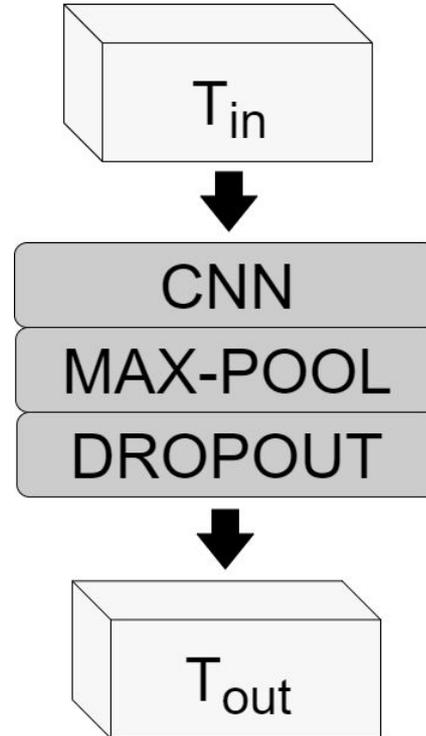
- Rel. Entrada-Salida

Max-Pool

- Red. Dimensionalidad

Dropout

- Añade Ruido



Arquitecturas

Arq. Recurrente Bidireccional

Reestructuración

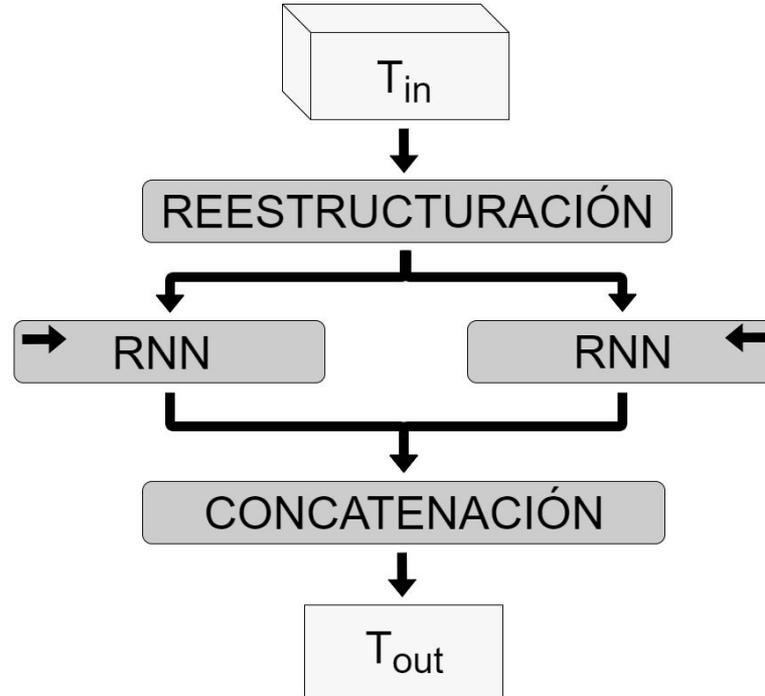
- Red. Orden
- Dim. Altura

Capas Recurrentes

- Direcciones Opuestas
- Dim. Anchura

Concatenación

- Dim. Altura

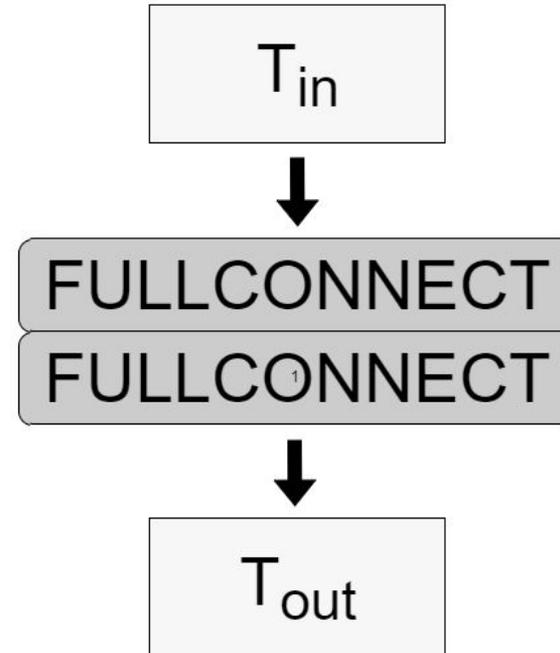


Arquitecturas

Arq. Reducción de Dimensionalidad

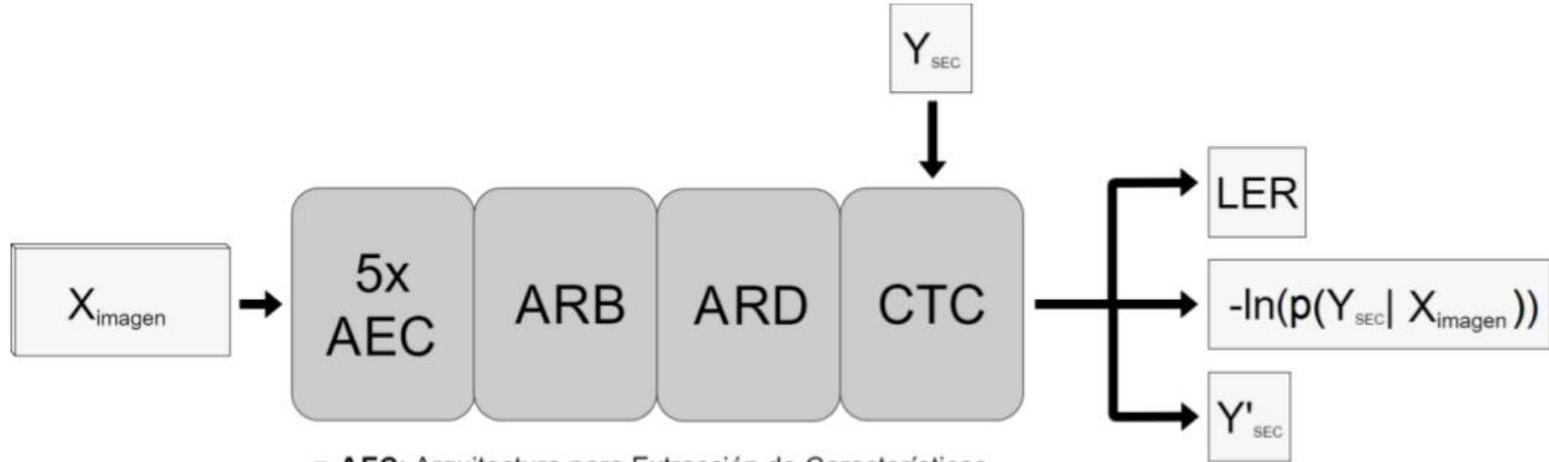
Capas Full Connect o Feedforward

- Red. Dimensionalidad
- Dim. Altura
- 63 Neuronas de salida.



Arquitecturas

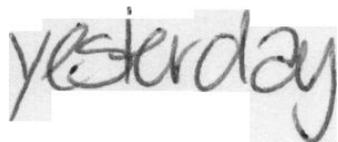
Arquitectura Global (Entrenamiento)



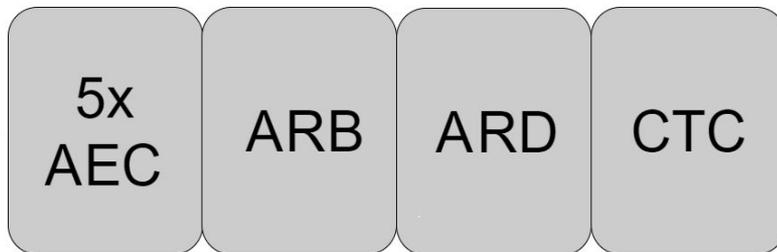
- **AEC**: Arquitectura para Extracción de Características
- **ARB**: Arquitectura Recurrente Bidireccional
- **ARD**: Arquitectura para Reducción de Dimensionalidad
- **CTC**: Connectionist Temporal Classification

Arquitecturas

Arquitectura Global (Transcripciones)



```
255 255 255 255 255 255 255 251 251 145 251 251 251 251 255 255 255
255 255 255 255 255 255 255 250 88 251 251 251 251 251 255 255 255
255 255 255 255 255 255 255 251 248 224 251 249 251 251 251 255 255
255 255 255 255 255 255 255 250 174 249 251 251 251 251 251 183 251
251 250 106 250 62 251 57 250 70 241 225 251 156 136 247 214 248 251
166 251 57 248 101 242 152 231 227 51 248 228 165 65 251 49 241 72
73 250 54 88 251 111 247 67 69 43 247 236 62 243 160 70 52 63
141 84 119 238 130 217 251 54 230 222 46 69 55 50 234 59 240 65
229 245 245 74 240 248 251 255 255 255 255 255 255 255 250 235 221
```



'60', '40', '54', '55', '40', '53', '39', '36', '60'



yesterday

- **AEC**: Arquitectura para Extracción de Características
- **ARB**: Arquitectura Recurrente Bidireccional
- **ARD**: Arquitectura para Reducción de Dimensionalidad
- **CTC**: Connectionist Temporal Classification

Arquitecturas

Hiperparámetros

Hiperparámetro	Valor	Hiperparámetro	Valor	Hiperparámetro	Valor
Nº de Filtos CNN-1	20	Nº de Filtos CNN-5	400	Neuronas de las RNNs	200
Nº de Filtos CNN-2	50	Dimensión de los filtros	5x5	Neuronas primera capa de la ARD	200
Nº de Filtos CNN-3	100	Ventana Max-Pool	2x2	Número de clases	63
Nº de Filtos CNN-4	200	Probabilidad de Dropout	0.5	Longitud secuencia entrada a la CTC	32

Arquitecturas

Implementación

Python3

- TensorFlow 1.3 (GPU)
- Pandas
- Numpy
- PIL

Repositorio Público

- [Github](#)



Pruebas y Resultados

Pruebas y Resultados

Test IAM.

Datasets Predefinidos:

- Entrenamiento (60%)
- Validación (20%)
- Test (20%)

Entrenamiento por épocas

Validaciones periódicas

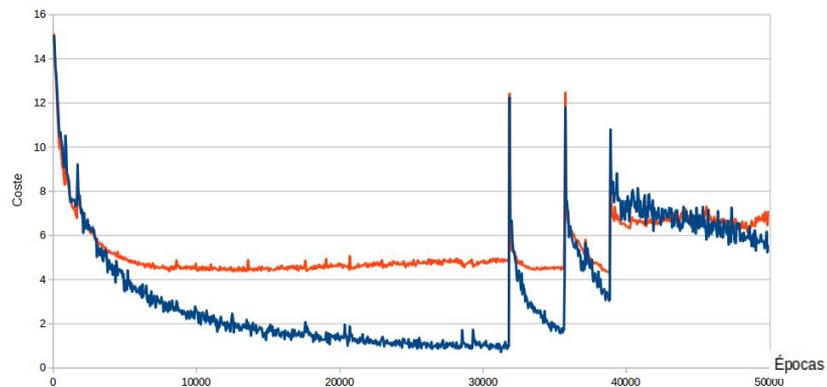
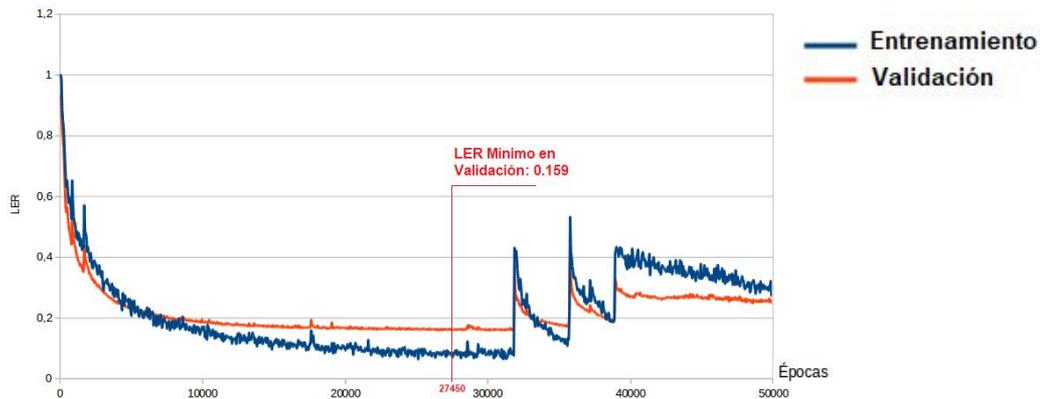
Alcanzar LER mínimo

Metricas:

- LER
- Coste

Pruebas y Resultados

Test IAM.



Época 27450

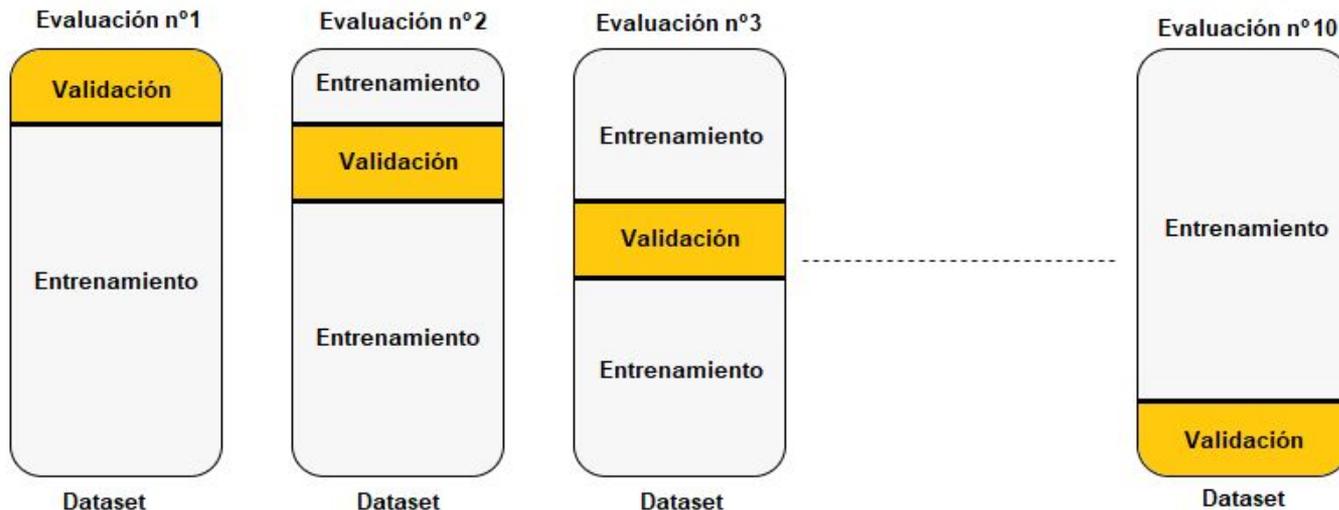
Dataset	LER	Coste
Entrenamiento	0.087	1.149
Validación	0.159	4.807
Test	0.175	3.977

Pruebas y Resultados

Validación Cruzada.

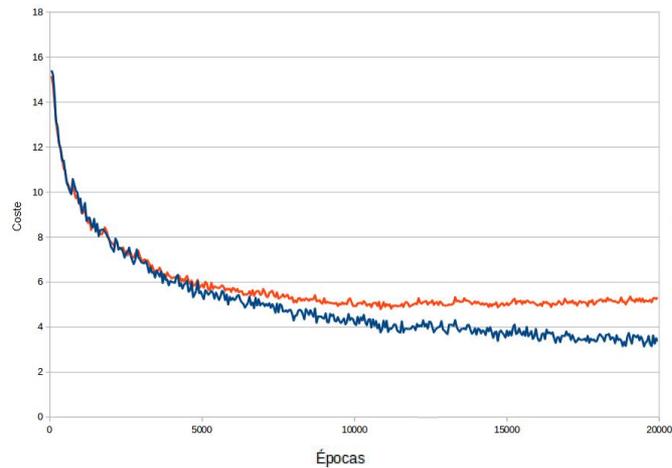
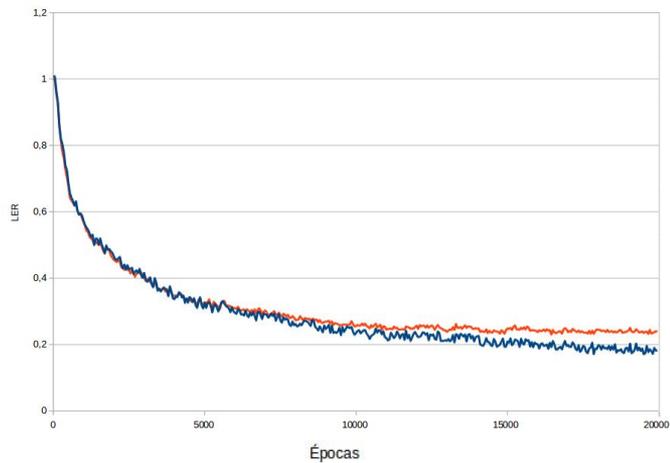
Medias Aritméticas:

- LER
- Coste



Pruebas y Resultados

Validación Cruzada.



— Entrenamiento
— Validación

Época	LER Train	LER Val	Coste Train	Coste Val
10850	0.231	0.245	3.147	4.804
16450	0.193	0.230	3.700	4.890

Conclusiones y Trabajo Futuro

Conclusiones y Trabajo Futuro

Conclusiones

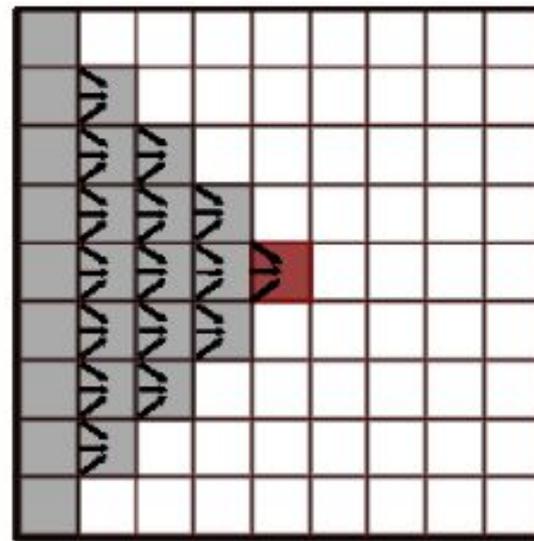
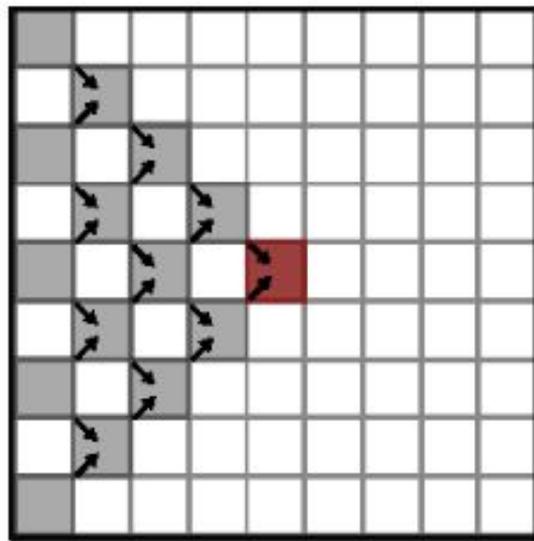
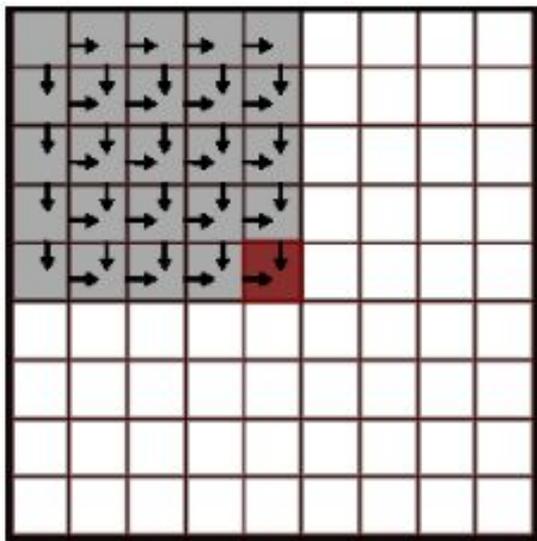
- No se ha buscado la eficiencia.
- Fundamentos matemáticos.
- Versatilidad de las Redes Neuronales Artificiales.
- Medidas poco intuitivas.
- Letras duplicadas consecutivas.
- Dataset excesivamente limpio.
- Otros campos de aplicación.

Trabajo Futuro

- Exploración de Hiperparámetros.
- Añadir nuevas capas recurrentes.
- Multi-dimensional Long Short Term Memory.
- Inclusión de etiquetas para letras duplicadas.
- Mantener las etiquetas duplicadas consecutivas.
- Añadir ruido.
- Visión Artificial.
- Sistemas dinámicos.

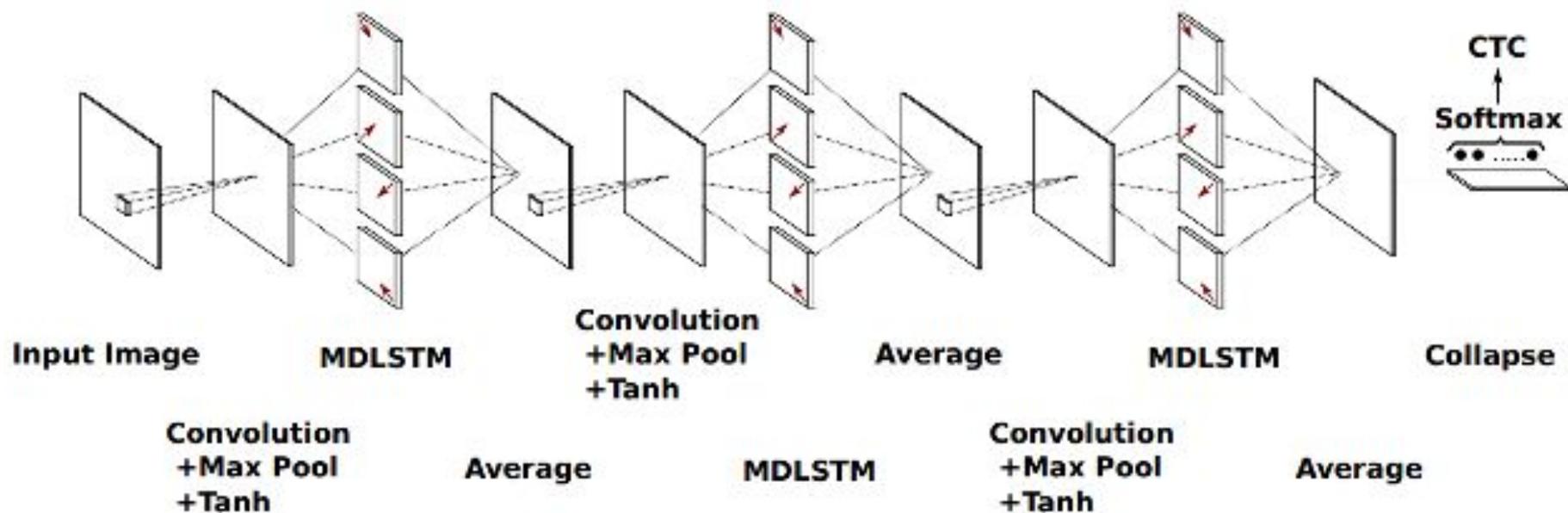
Bonus Track

Multi-Dimensional Recurrent Neural Networks (MDRNN)



Bonus Track

Multi-Dimensional Recurrent Neural Networks (MDRNN)



Gracias