



PONTIFICIA  
**UNIVERSIDAD**  
**CATÓLICA**  
DEL PERÚ

# MODELAR LA PROBABILIDAD DE PALABRAS CON UNA RED CONVOLUCIONAL UTILIZANDO PATRONES DE SONIDOS

---

INF659 - TÉCNICAS AVANZADAS DE DATA MINING Y  
SISTEMAS INTELIGENTES

JOHN EDWARD MILLER Y RICARDO JOSE LINARES

1-OCTOBRE-2019

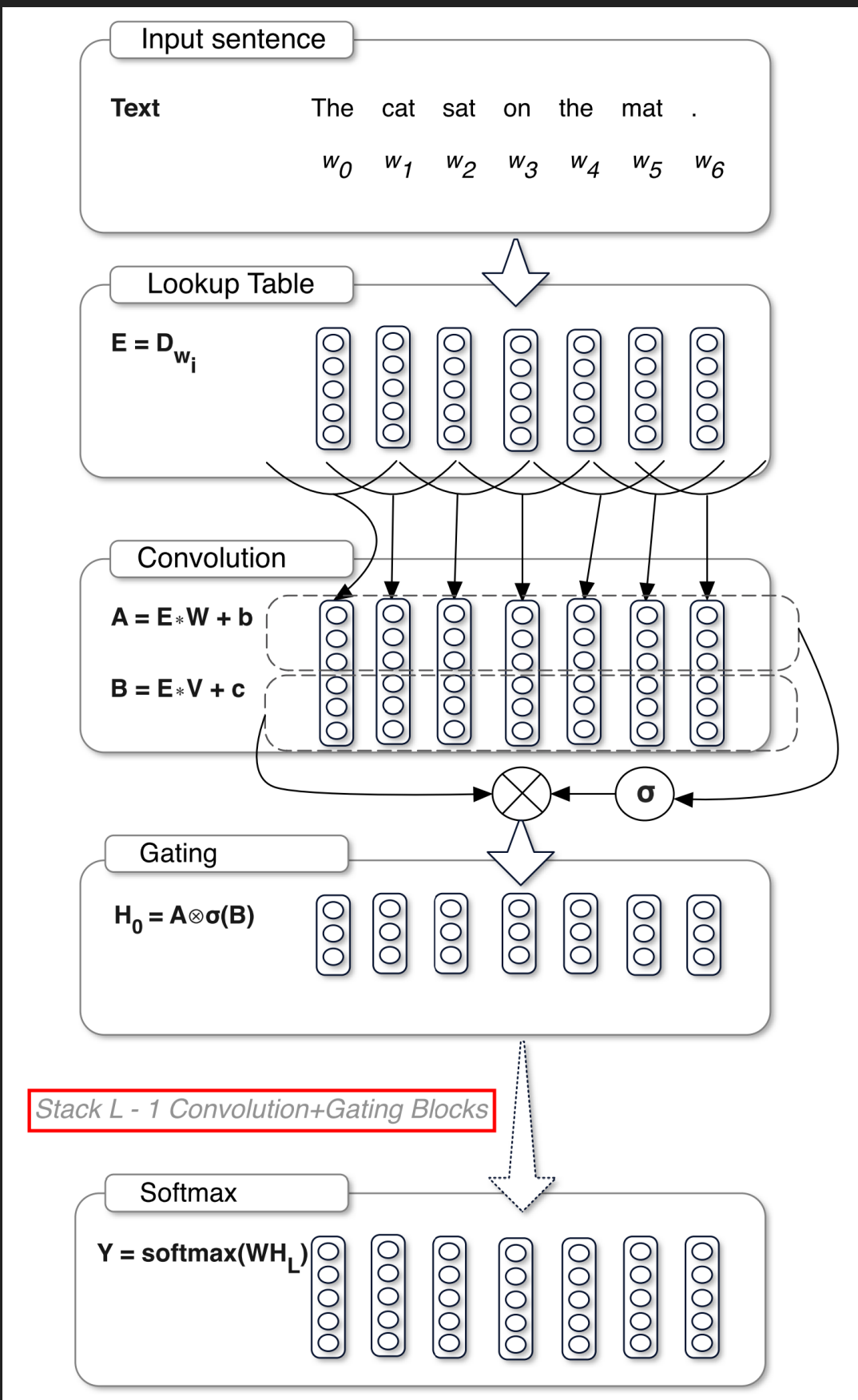
# CONTENIDO

---

- ◆ Presentación de modelo base
- ◆ Ejecución del modelo
- ◆ Propuesta de proyecto

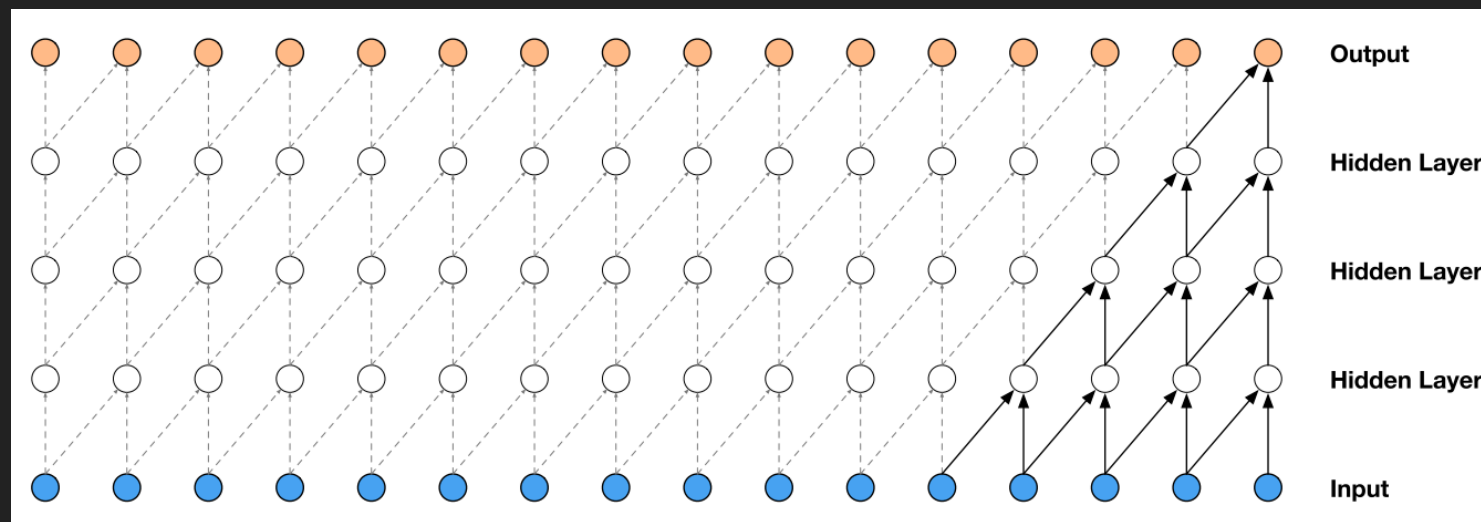
# IDEA CENTRAL Y ARQUITECTURA

[Y. N. Dauphin et. al., Language modeling with gated convolutional networks]



- ▶ Sustituir capas recurrentes por capas convolucionales en modelo de lenguaje para:
  - ▶ Aumentar paralelismo
  - ▶ Reducir parámetros
  - ▶ Igualar exactitud
- ▶ Representar entradas con «embeddings»
- ▶ Calcular convoluciones causales y «gated»:
  - ▶ «Chunk» { «Block» { «Unit» }<sup>+</sup> } ×  $m$
  - ▶ «Unit» [ancho, # filtros] – «gated»
    - ▶  $h_l(X) = (X \cdot W + b) \otimes \sigma(X \cdot V + c)$
  - ▶ «Block» – termina con «residual activation»
    - ▶  $H(X) = h_{l+b-1} \dots h_l(X) + X$
- ▶ Calcular probabilidades con  $\text{softmax}(W \cdot H_L)$

# CONVOLUCIONES CAUSALES Y CONFIGURACIONES DE MODELOS



[van den Oord, et. al., Wavenet: A generative model for raw audio]

- ▶ Causal implica direccion tal como en las capas recurrentes
- ▶ Pero sin las capas!

Name	GCNN-13	GCNN-14B	GCNN-9	GCNN-8B
Dataset	Google Billion Word			
Lookup	128			
Conv1	$[4, 1268] \times 1$	$[5, 512] \times 1$	$[4, 807] \times 1$	$[1, 512] \times 1$
Conv2.x	$\begin{bmatrix} 4, 1268 \\ 4, 1268 \end{bmatrix} \times 12$	$\begin{bmatrix} 1, 128 \\ 5, 128 \\ 1, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 4, 807 \\ 4, 807 \end{bmatrix} \times 4$	$\begin{bmatrix} 1, 128 \\ 5, 128 \\ 1, 512 \end{bmatrix} \times 3$
Conv3.x		$\begin{bmatrix} 1, 512 \\ 5, 512 \\ 1, 1024 \end{bmatrix} \times 3$		$\begin{bmatrix} 1, 256 \\ 5, 256 \\ 1, 512 \end{bmatrix} \times 3$
Conv4.x		$\begin{bmatrix} 1, 1024 \\ 5, 1024 \\ 1, 2048 \end{bmatrix} \times 6$		$\begin{bmatrix} 1, 1024 \\ 1, 1024 \\ 1, 2048 \end{bmatrix} \times 1$
Conv5.x		$\begin{bmatrix} 1, 1024 \\ 5, 1024 \\ 1, 4096 \end{bmatrix} \times 1$		
Conv6.x				
Conv7.x				
AdaSoftmax	10k,40k,200k		4k,40k,200k	

- ▶ «Chunk» { «Block» { «Unit» }<sup>+</sup> }  $\times m$
- ▶ B «bottleneck» ayuda con generalización
- ▶ Bloques terminan con ancho de 1

[Y. N. Dauphin et.al., Language modeling with gated convolutional networks]

# EXPERIMENTACIÓN Y RESULTADOS

[Y. N. Dauphin et. al., Language modeling with gated convolutional networks]

## GOOGLE BILLION WORD – PERPLEJIDAD

Model	Test PPL	Hardware
Sigmoid-RNN-2048 (Ji et al., 2015)	68.3	1 CPU
Interpolated KN 5-Gram (Chelba et al., 2013)	67.6	100 CPUs
Sparse Non-Negative Matrix LM (Shazeer et al., 2014)	52.9	-
RNN-1024 + MaxEnt 9 Gram Features (Chelba et al., 2013)	51.3	24 GPUs
LSTM-2048-512 (Jozefowicz et al., 2016)	43.7	32 GPUs
2-layer LSTM-8192-1024 (Jozefowicz et al., 2016)	30.6	32 GPUs
BIG GLSTM-G4 (Kuchaiev & Ginsburg, 2017)	23.3*	8 GPUs
LSTM-2048 (Grave et al., 2016a)	43.9	1 GPU
2-layer LSTM-2048 (Grave et al., 2016a)	39.8	1 GPU
GCNN-13	38.1	1 GPU
GCNN-14 Bottleneck	31.9	8 GPUs

## GOOGLE BILLION WORD – EFICIENCIA

	Throughput		Responsiveness
	(CPU)	(GPU)	(GPU)
LSTM-2048	169	45,622	2,282
GCNN-9	121	29,116	29,116
GCNN-8 Bottleneck	<b>179</b>	<b>45,878</b>	<b>45,878</b>

## PENN TREE BANK – PERPLEJIDAD

Modelo	Perplejidad
GCNN	108.7
LSTM	109.3

### Experimentación

- ▶ Configuraciones
  - ▶ «embedding»; capas de chunks, bloques, unidades; ancho y cantidad de filtros CNN
- ▶ Tipo de «gating»
- ▶ Tamaño de contexto
- ▶ «clipping» de gradiente
- ▶ normalización de pesos

# EJECUCIÓN DE MODELO

---

# PROPUESTA DE PROYECTO

---

## ▶ Nombre

- ▶ Modelar la probabilidad de palabras con una red convolucional utilizando patrones de sonido.

## ▶ Objetivo

- ▶ Desarrollar una red neuronal convolucional (causal y «gated») utilizando patrones de sonidos por palabra para predecir su probabilidad con el fin de discriminar entre palabras dentro o fuera el lenguaje.

## ▶ Conjuntos de datos a utilizar

- ▶ Intercontinental Dictionary Series (IDS): Tablas de palabras de idiomas Amazónicas ~1,310 palabras por idioma.
- ▶ World Loanword Database (WOLD): Tablas de palabras de idiomas variadas con indicador de «loan words» ~ 1,400 palabras por idioma.

# ARTÍCULOS CIENTÍFICOS RELEVANTES

---

- ▶ Y. Bengio, R. Ducharme, P. Vincent, and C. Janvin. A neural probabilistic language model. *J. Mach. Learn. Res.*, 3:1137-1155, Mar. 2003.
- ▶ S. F. Chen and J. Goodman. An empirical study of smoothing techniques for language modeling. *Computer Speech & Language*, 13(4):359-394, 1999.
- ▶ Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier. Language modeling with gated convolutional networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70, ICML'17*, pages 933-941, Sydney, NSW, Australia, 09 2017. JMLR, JMLR.org.
- ▶ M. Silfverberg, L. J. Mao, and M. Hulden. Sound analogies with phoneme embeddings. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2018*, pages 136-144, Salt Lake City, Utah, 01 2018. Society for Computation in Linguistics.
- ▶ T. Young, D. Hazarika, S. Poria, and E. Cambria. Recent trends in deep learning based natural language processing. *IEEE Computational intelligence magazine*, 13(3):55-75, 2018.
- ▶ J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, et al. Recent advances in convolutional neural networks. *Pattern Recognition*, 77:354-377, 2018.
- ▶ N. Tishby and N. Zaslavsky. Deep learning and the information bottleneck principle. *CoRR*, abs/1503.02406, 2015.
- ▶ A. M. Saxe, Y. Bansal, J. Dapello, M. Advani, A. Kolchinsky, B. D. Tracey, and D. D. Cox. On the information bottleneck theory of deep learning. In *International Conference on Learning Representations*, 2018.
- ▶ M. R. Key and B. Comrie, editors. *Intercontinental Dictionary Series (IDS)*. Max Planck Institute for Evolutionary Anthropology, Leipzig, 2015.
- ▶ M. Haspelmath and U. Tadmor, editors. *World Loanword Database (WOLD)*. Max Planck Institute for Evolutionary Anthropology, Leipzig, 2009.