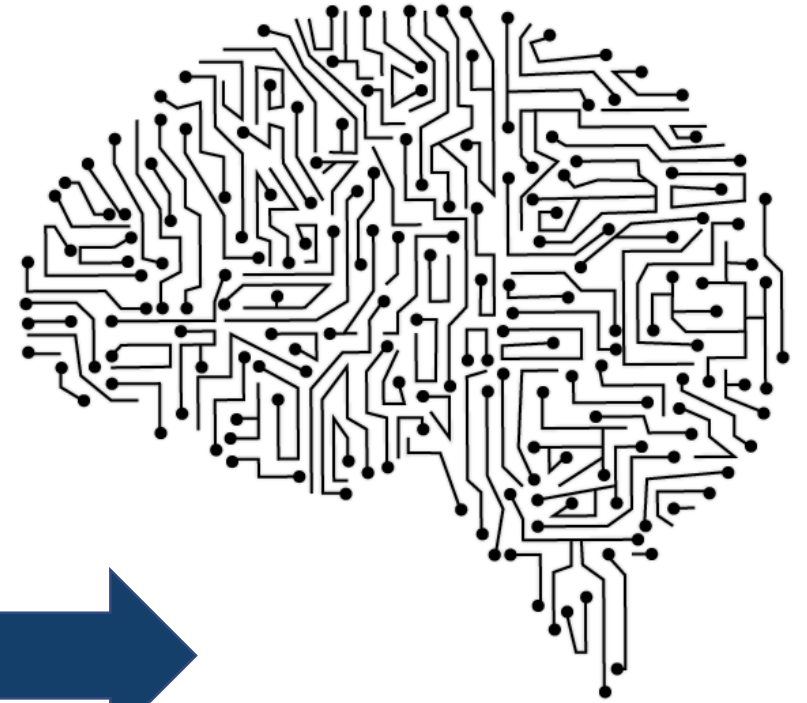


Simple AI

Movie Genre Classification
Deep Learning



Sergio Castelblanco
Jesús Solano
August 2018

Speakers Bio

Jesús Solano



Educación

- ✓ Físico, Universidad de los Andes.
- ✓ Ingeniero Industrial, Universidad de los Andes.

Experiencia

- ✓ Redes neuronales para detección de partículas en el colisionador CERN.
- ✓ Deep Learning para la detección del cáncer de Mama en mamografías.
- ✓ Modelos de Machine Learning para valorar derivados financieros climáticos.

Speakers Bio

Sergio Castelblanco



Educación

- ✓ Ingeniero Industrial,
- ✓ Especialista en Gerencia de Proyectos, Universidad el Bosque.
- ✓ Magister Analytics Cand., Universidad de los Andes.

Experiencia

- ✓ Deep Learning para la predicción de propiedades físicas de hidrocarburo a partir de información geoquímica.
- ✓ Machine learning en firmas espectrales y sensores remotos.
- ✓ Machine Learning para clasificar el estrés de la vegetación por contaminación de hidrocarburos en cuencas colombianas.
- ✓ Desarrollo de algoritmos DL & ML en geociencias y yacimientos.

**“La inteligencia artificial
es el eje transformador
por el cual estamos
repensando cómo estamos
haciendo todo”**

Sundair Pichai
CEO Google

Descripción del Problema



“As the Avengers and their allies have continued to protect the world from threats too large for any one hero to handle, a new danger has emerged from the cosmic shadows: Thanos. A despot of intergalactic infamy, his goal is to collect all six Infinity Stones, artifacts of unimaginable power, and use them to inflict his twisted will on all of reality. Everything the Avengers have fought for has led up to this moment - the fate of Earth and existence itself has never been more uncertain.”

¿Qué genero es la película?

Retos del Problema

¿Qué caracteriza el poster de un genero determinado?

¿Cómo determinar el contexto o sentido de una sinopsis?

¿Es posible reducir la complejidad del problema?

¿Cómo combinar las predicciones de texto e imágenes?

¿Qué enfoques podemos usar?

1. Machine Learning

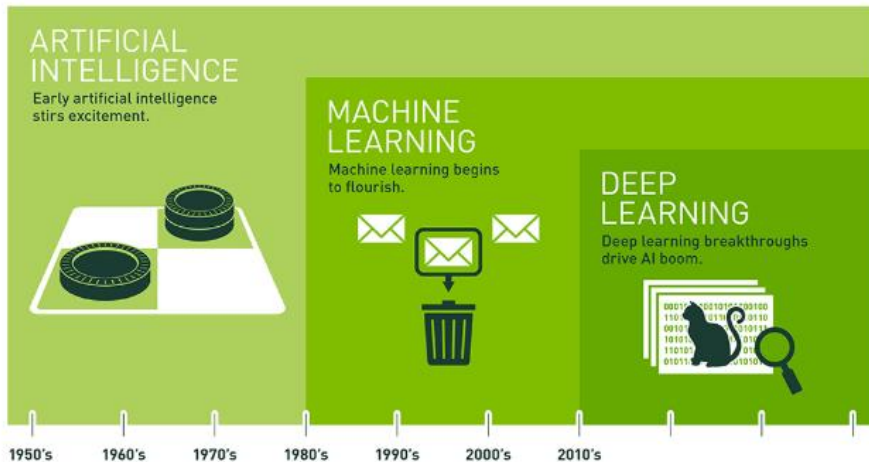
- ✓ Random Forest
- ✓ Gradient Boosting Classifier
- ✓ Decision Trees.

2. Deep Learning

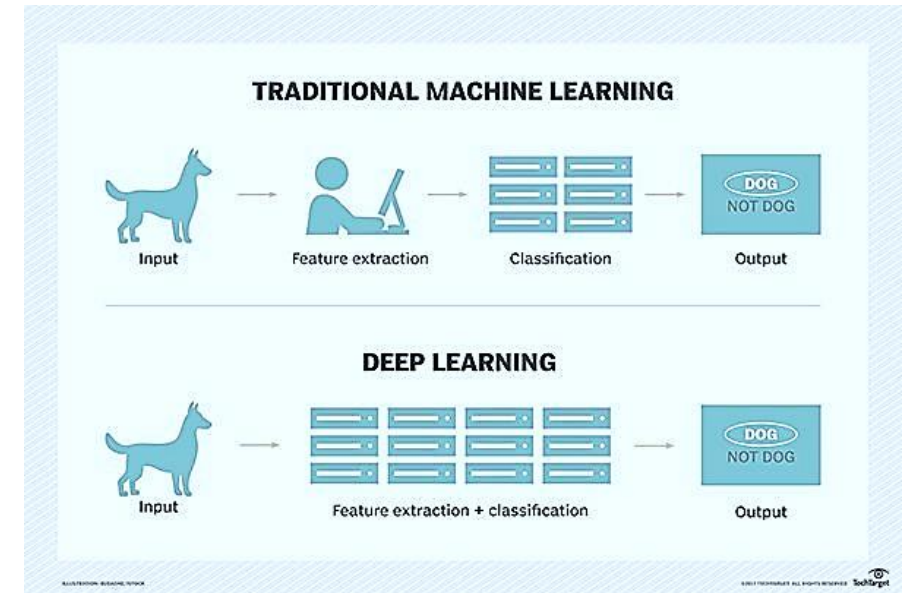
- ✓ Convolutional Neural Networks.
- ✓ Long Short Term Memory(LSTM).
- ✓ Word Embeddings.

Deep Learning

¿Qué es Deep Learning?



- Inteligencia Artificial
- Modelos Bio-inspirados
- Extracción de características + Aprendizaje
- Conexiones densas Imitar comportamiento Sistema nervioso



Deep Learning - Imágenes

- ✓ Multimodal IMDb Dataset.
- ✓ 25959 Imágenes
- ✓ 160 x 256 pixeles.



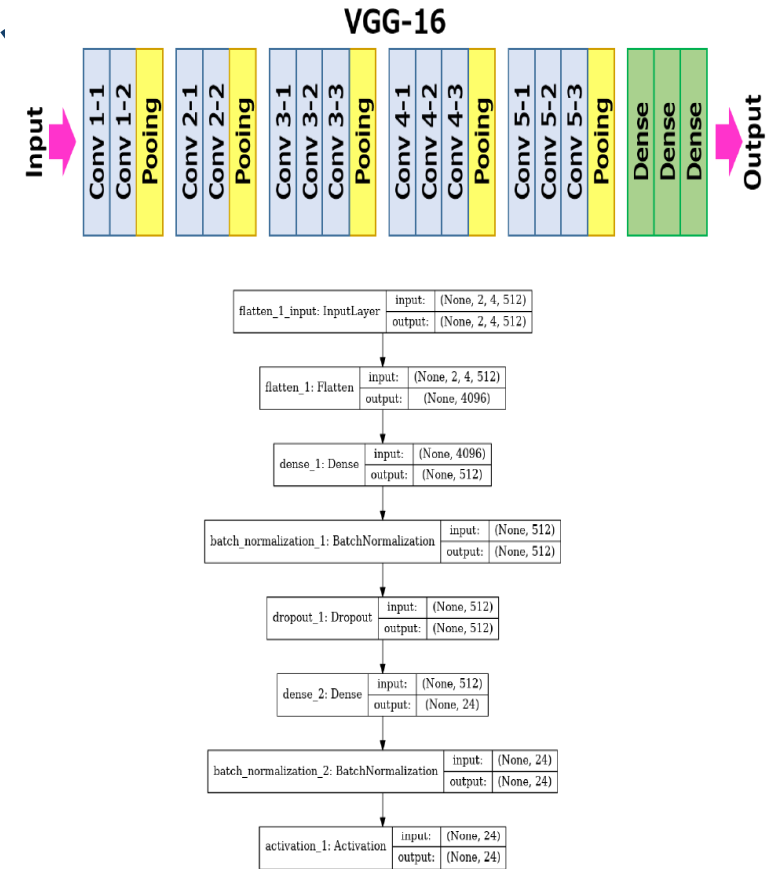
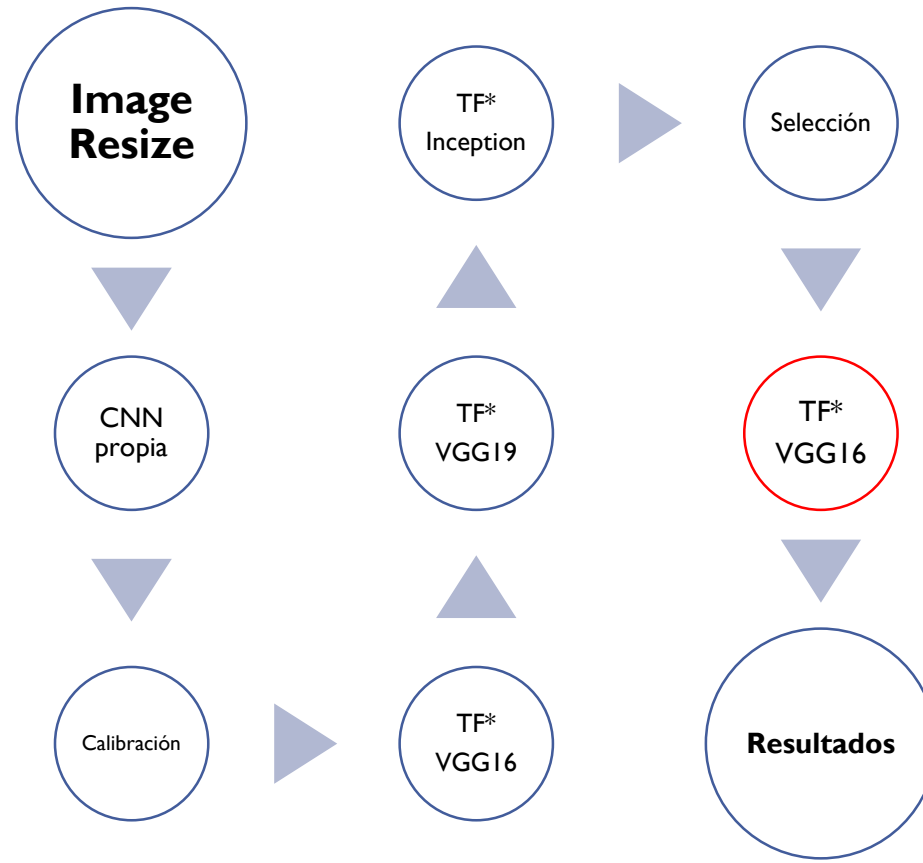
Convolutional Neural Networks



- Transfer Learning.
- CNN propia (3 Conv 2D)

ANN– Imágenes blanco y negro

Pre-procesamiento y Selección del modelo Final

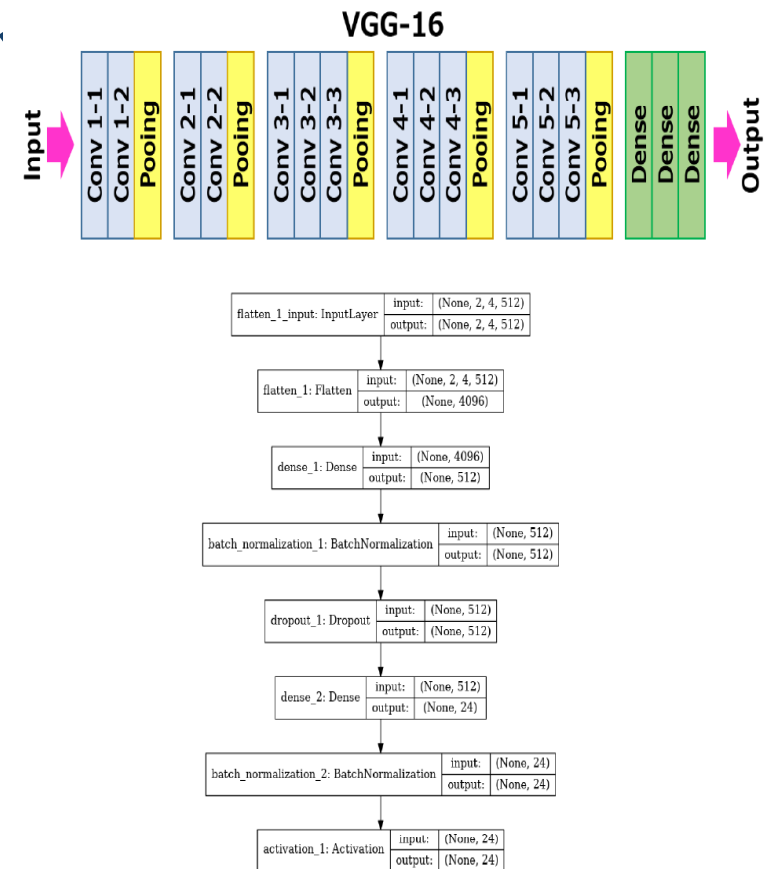
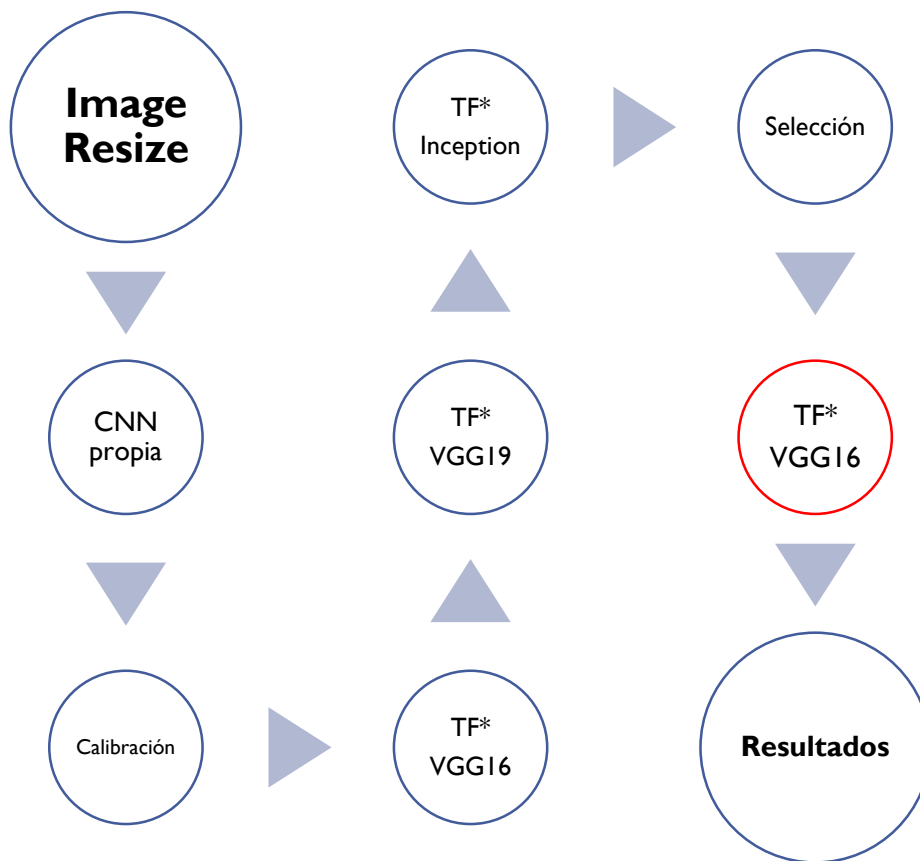


Parámetros: 138,357,544 – Profundidad: 25 – Tiempo: 18 min GPU

*TF: Transfer Learning

ANN– Imágenes a color

Pre-procesamiento y Selección del modelo Final

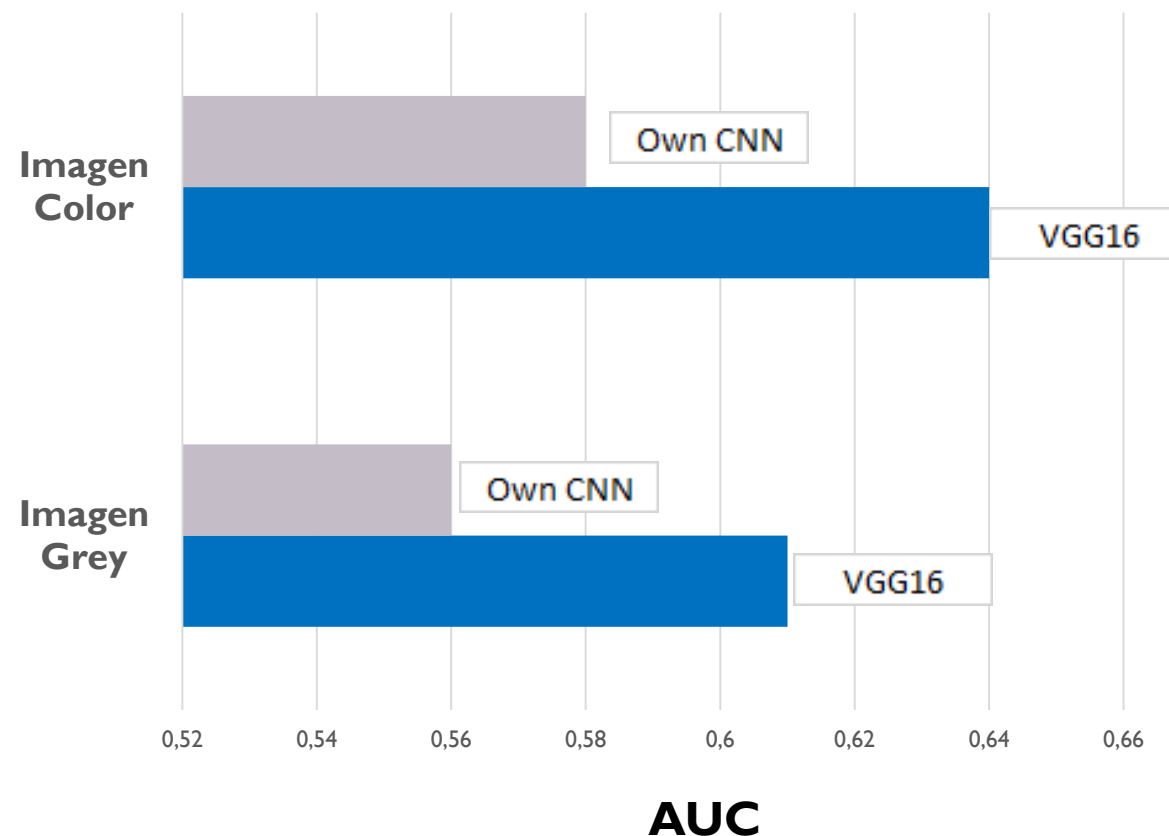


Parámetros: 138,357,544 – Profundidad: 25 – Tiempo: 29 min GPU



Resultados ANN Imágenes

- Las imágenes en escalas de grises no generan un alto desempeño dado que los colores pueden describir patrones de interés.
- Las imágenes a color tienen un mejor desempeño pero tienen mayor costo computacional.
- Para aumentar el AUC es necesario + GPU+ RAM + Imágenes
- el rendimiento de VGG16 es superior a VGG19 e InceptionV3



Deep Learning - Texto

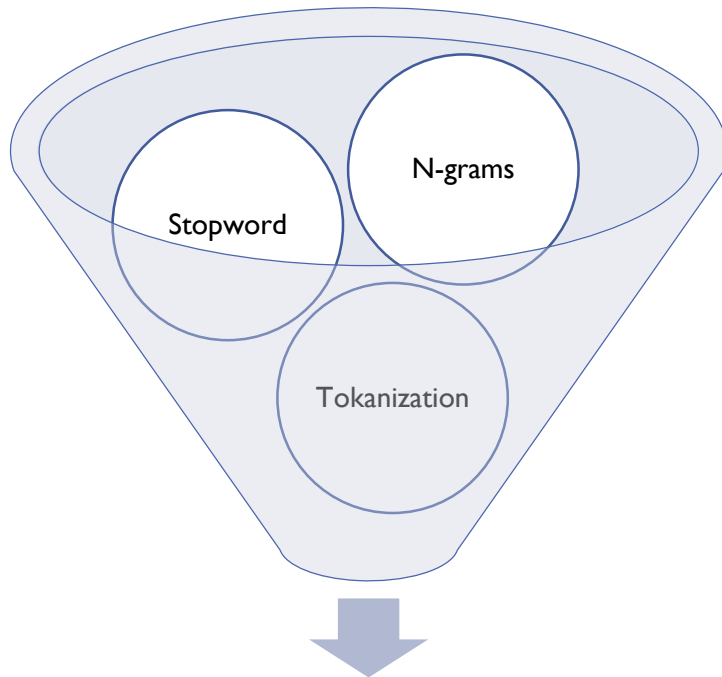
“As the Avengers and their allies have continued to protect the world from threats too large for any one hero to handle, a new danger has emerged from the cosmic shadows: Thanos. A despot of intergalactic infamy, his goal is to collect all six Infinity Stones, artifacts of unimaginable power, and use them to inflict his twisted will on all of reality. Everything the Avengers have fought for has led up to this moment - the fate of Earth and existence itself has never been more uncertain.”

- ✓ Multimodal IMDb Dataset.
- ✓ 25959 Imágenes
- ✓ 160 x 256 pixeles.



-
- LSTM.
 - N-grams.
 - Conv ID
 - Word2Vec

Descomposición de texto

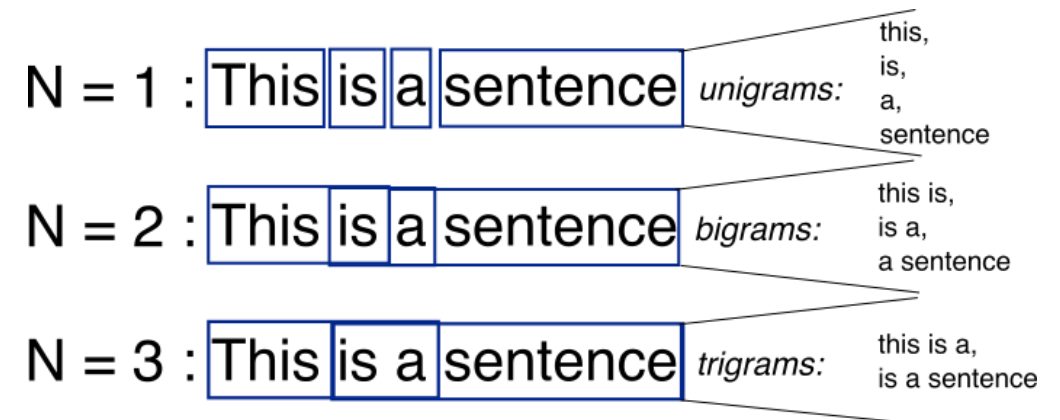


Datos Pre-procesados

N-grams



Partición de una frase en subconjuntos de palabras que recrean un contexto



Descomposición de texto

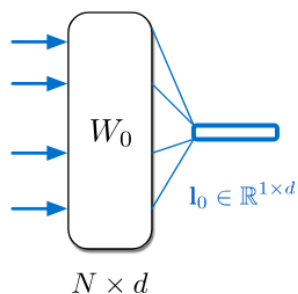
Word2Vec

Método no supervisado ML - Google

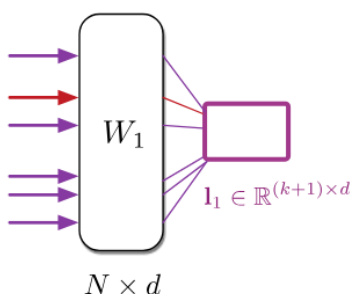
target word
The black cat slept on the bed.
context words

Training phase

1) take sum of context embeddings



2) select target and k noise weights (negative sampling)



3) compute error & backpropagate

$$\text{err} = \mathbf{t} - \sigma(\mathbf{l}_0 \cdot \mathbf{l}_1^T)$$

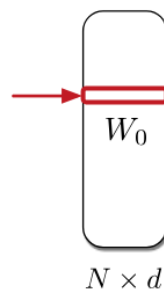
with:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

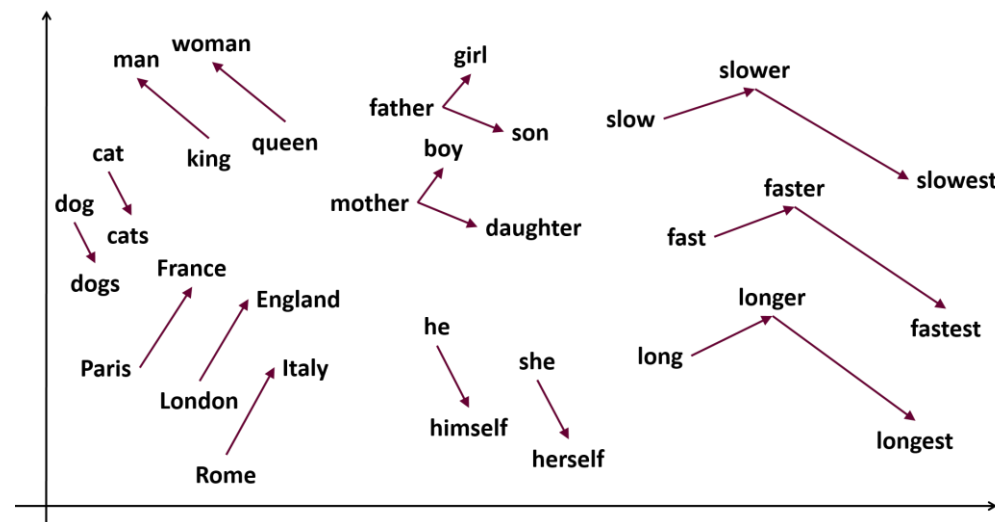
\mathbf{t} : binary label vector

After training

target embedding
 $\in \mathbb{R}^{1 \times d}$

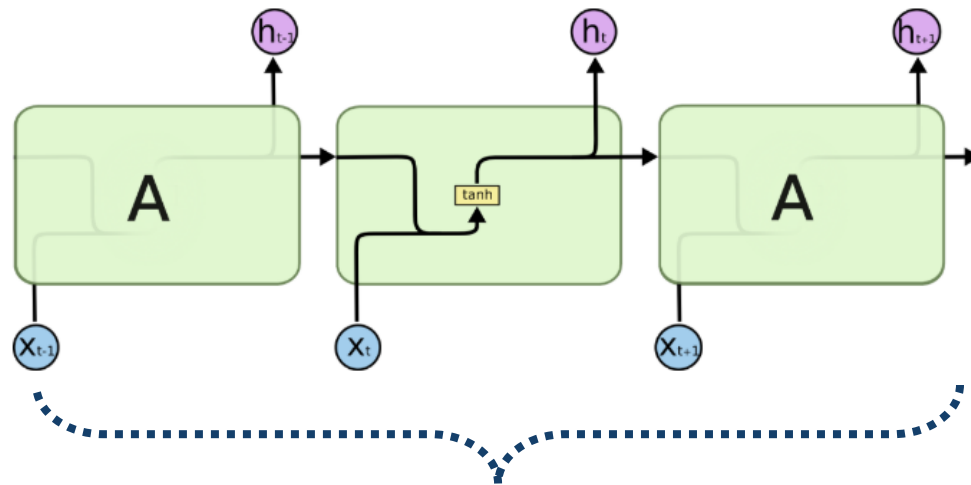


A partir del Contexto predice una palabra y de una palabra predice el contexto

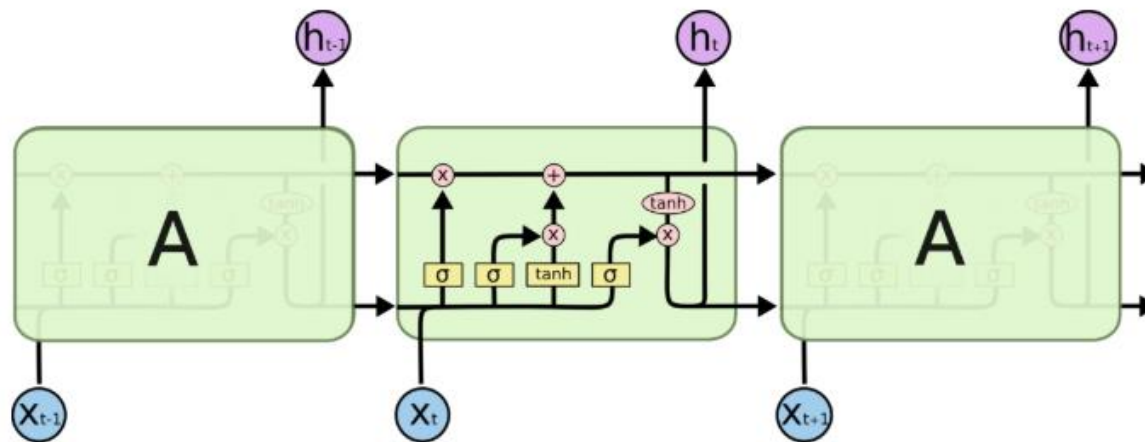


LSTM - Texto

- ✓ Red neuronal Recurrente.
- ✓ LSTM extiende memoria de RNN's
- ✓ No hay 'exploding gradients'

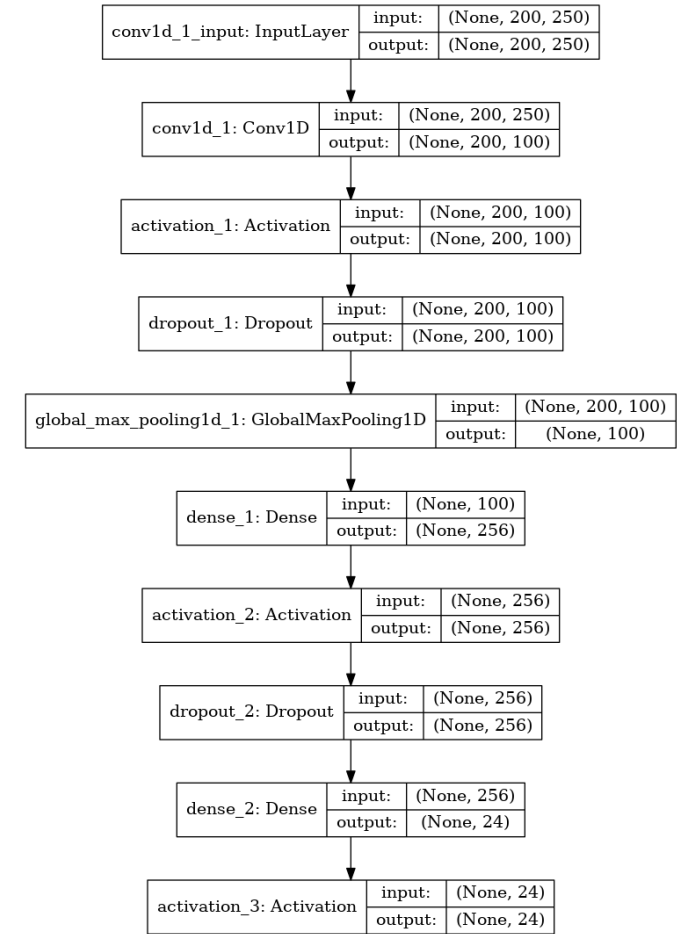
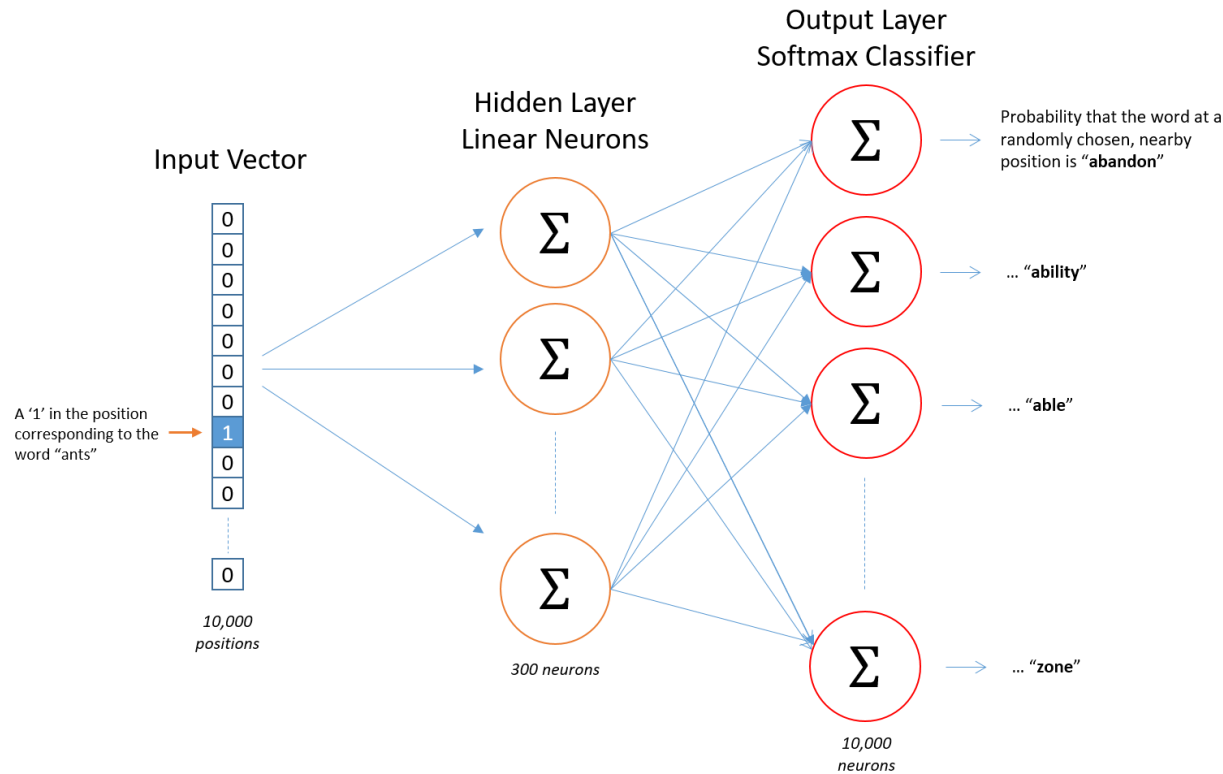


RNN

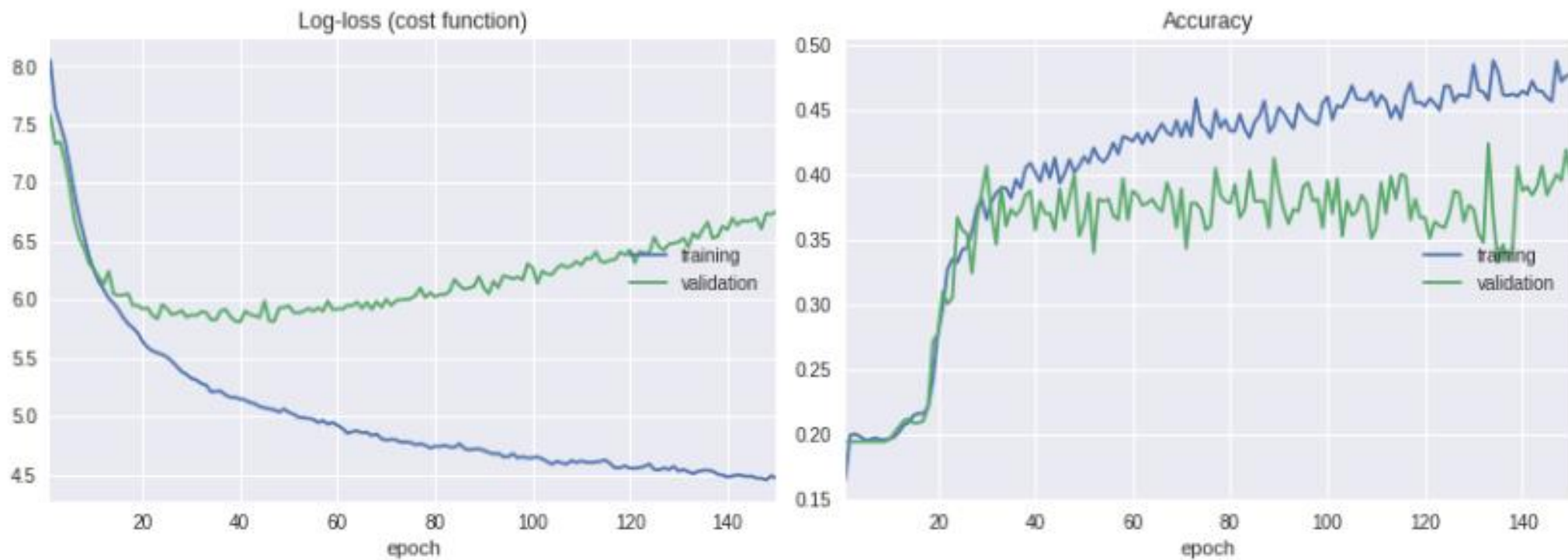


LSTM

Word2Vec + Conv 1D



Parámetros: 125.000 – Profundidad: 4 – Tiempo: 6 min GPU



Hyperparametros:

LR: 0,01

Optimizador: Adam

Dropout: 0,5

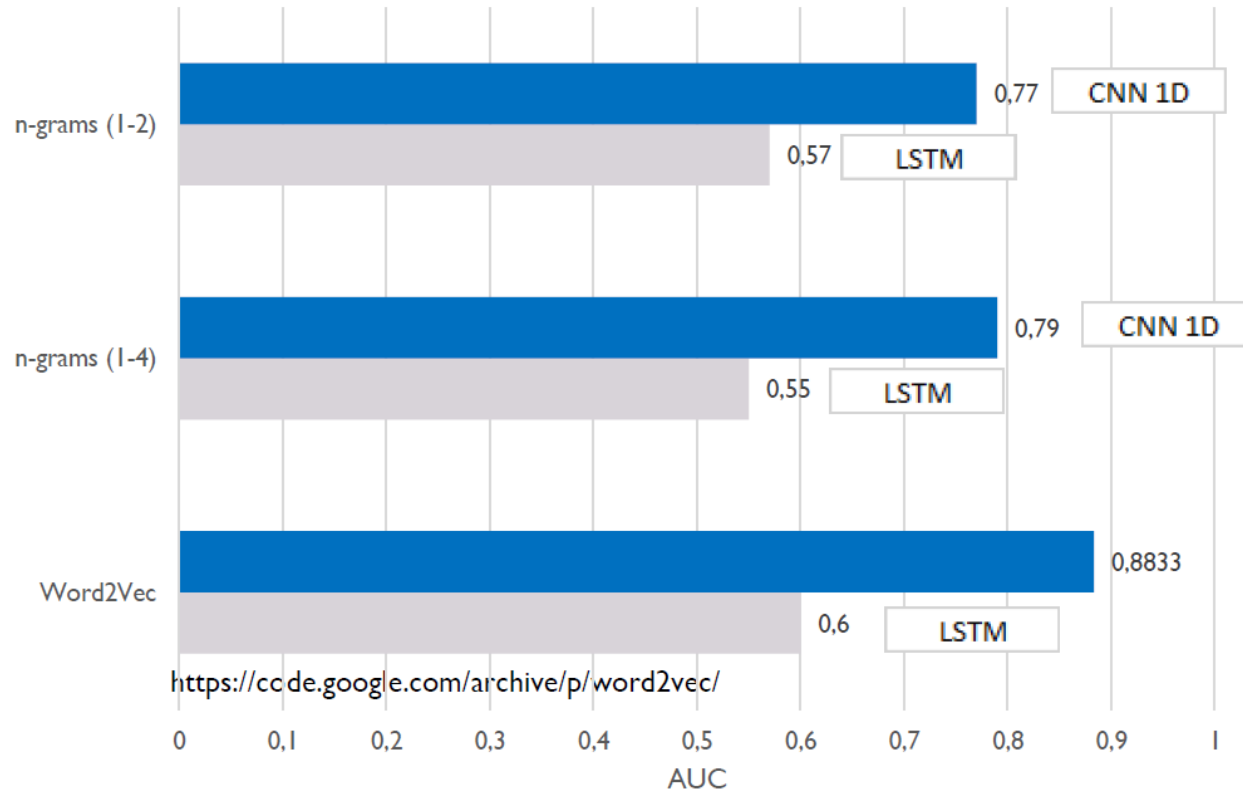
Funciones de activación: Selu, Relu, Sigmoid

Inicializadores: Distribución normal

Batch Size: 1000

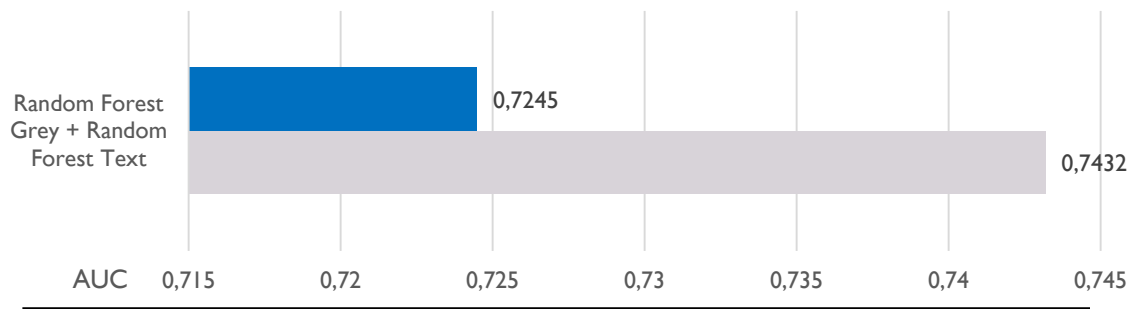
Épocas: 500

Resultados Finales Texto

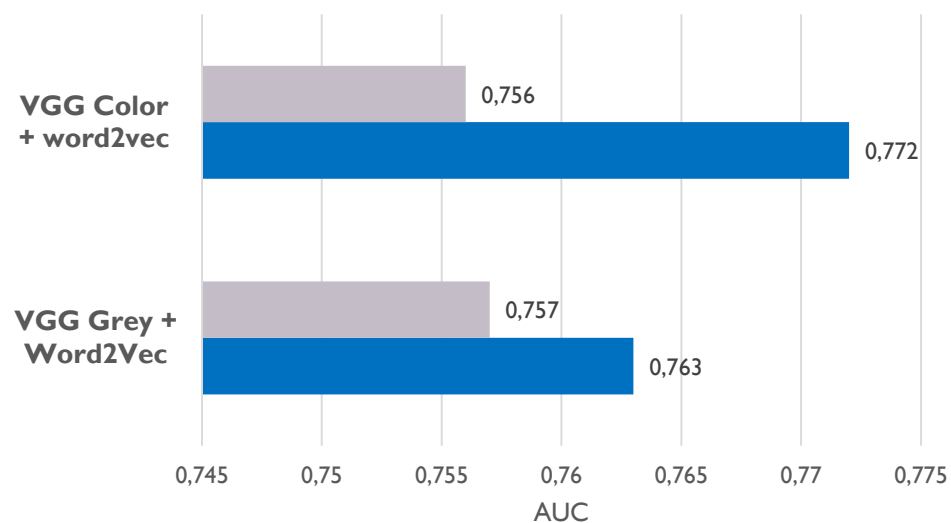


- Los mejores resultados se obtienen con un embedding de Word2Vec.
- El proceso de clasificación es mas eficiente con CNN 1D
- Para mejorar el AUC es necesario + RAM, + GPU para incrementar el vocabulario y contexto del embedding del modelo.

Text + Images Stacking

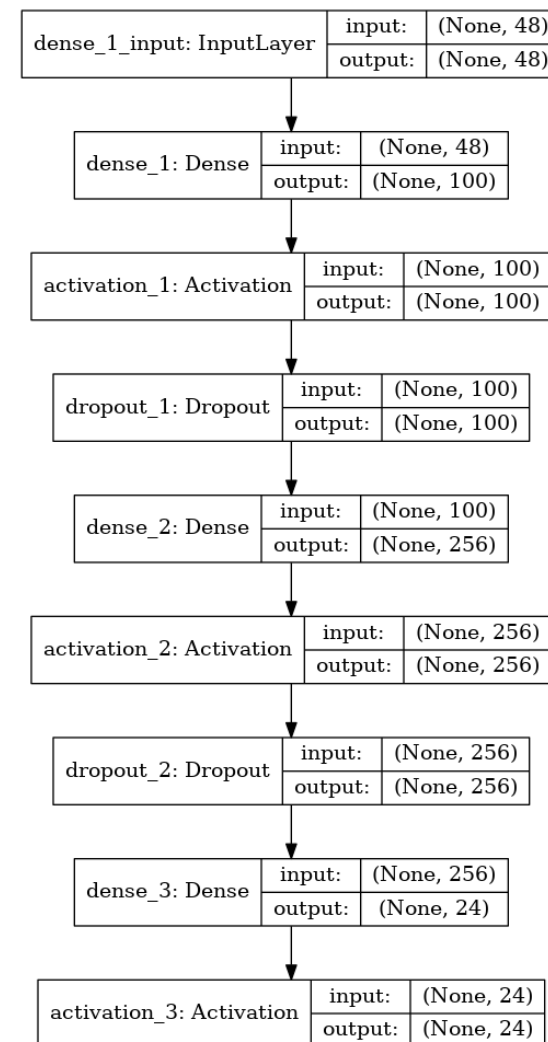


Machine Learning Stacking



Deep Learning Stacking

Topología Neuronal Stacking



Comentarios Finales

El mejor enfoque de pre procesamiento de texto en la actualidad es word2vec dada su capacidad de reconocimiento del contexto.

Si no es posible entrenar con imágenes de alta calidad es mejor clasificar solo con texto.

Es posible clasificar películas por genero con redes multimodales con un 0.89 de AUC.

Trabajo Futuro

Entrenar con imágenes de alta calidad en ambientes de mayor capacidad computacional.

Aumentar el vocabulario del contexto en word2vec.

Referencias

Arevalo, J., Solorio, T., Montes-y-Gómez, M., & González, F.A. (2017). Gated multimodal units for information fusion. arXiv preprint arXiv:1702.01992.

Zeynep Akata, Honglak Lee, and Bernt Schiele. Zero-Shot Learning with Structured Embeddings. CoRR, abs/1409.8, 2014. URL <http://arxiv.org/abs/1409.8403>.

Eric H Huang, Richard Socher, Christopher D Manning, and Andrew Ng. Improving word representations via global context and multiple word prototypes. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1, pp. 873–882. Association for Computational Linguistics, 2012

Marina Ivasic-Kos, Miran Pobar, and Luka Mikec. Movie posters classification into genres based on low-level features. In 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), volume i, pp. 1198–1203. IEEE, may 2014. ISBN 978-953-233-077-9. doi: 10.1109/MIPRO.2014.6859750. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6859750>