

Segundo Examen Parcial

Y. Sarahi Gracia González



Instrucciones Para CADA presentación/material deberás elaborar lo siguiente:

Un mini resumen en PDF con la siguiente estructura para cada presentación/sesión/clase/podcast:

- MOTIVATION: al menos 40 palabras explicando la motivación y lo que se presenta en la sesión.
- HOW IT WORKS: al menos 70 palabras de como se abordó y en que consistió la presentación, práctica/experimentación y/o casos de estudio.
- RESULTS: al menos 70 palabras de los resultados. Funciona? no funciona? funciona más o menos? con cuanta diferencia? vale la pena? etc? o bien principales conclusiones.
- WHY IT MATTERS?: al menos 40 palabras de porque deberíamos estar interesados en el tema?
- TAKE AWAYS AND WHAT WE ARE THINKING?: al menos 50 palabras...

Sesgo Algorítmico <https://youtu.be/8insq1Si98I>

■ MOTIVATION

Durante esta sesión se presenta un problema crucial en los algoritmos de inteligencia artificial, particularmente con los modelos de procesamiento del lenguaje, este problema es el sesgo y comienza desde los datos en que se entrenan estos algoritmos pues éstos contienen sesgos ya que toda persona los tiene. En el video se presentan varias ideas, una de ellas es que si ya existen sesgos en los datos, un primer objetivo es que durante e entrenamiento éstos no se amplifiquen.

■ HOW IT WORKS

Se exponen varios casos sobre los grandes sesgos en empresas y algoritmos importantes. El primero es el de un algoritmo de Amazon, que te mostraba precios distintos basado en tu localización. Otro es sobre los videos falsos. Y uno más sobre qué tipos de datos pueden o no venderse a las empresas, en el video se comenta el caso hipotético de mostrar un producto a un grupo antisemita.

Se abordan cuatro problemáticas: Derechos y libertades: ¿Qué pasa con los seguros?, Trabajo y automatización: Registro biométrico, Seguridad e infraestructura: Sistemas hackeables y Sesgo algorítmico.

■ RESULTS



En este video no se mencionan como tal resultados pues no se proponen relaciones lo que sí se menciona es que hay algunos factores que pueden mejorarse para lidiar con este problema:

1. Corregir problemas metodológicos relacionados con los datos
2. Corregir problemas de representación (particularmente a las minorías)
3. Corregir el sesgo inducido o cultural
4. Corregir sesgos de contraste

Algunos ejemplos que se mencionan sobre esto son: los sesgos de género relacionados con la cantidad de documentos escritos por hombres vs los que son escritos por mujeres. El sesgo que se produce en un momento preciso como al hacer reclutamiento. En cuanto a la poca representación se menciona el caso de una estudiante mujer y de raza negra que notó que el reconocimiento facial no funcionaba en ella.

■ WHY IT MATTERS?

Creo que es muy obvio porque esto importa, el sesgo es algo que afecta a toda la población y principalmente a aquellas personas que pertenecen a diversas minorías o a grupos que han sido históricamente discriminados. Los sesgos con los que estamos más familiarizados en algorítmicos computacionales son los de raza, género, y estado socioeconómico. Sin embargo me parece importante también mostrar los sesgos que no son tan familiares como la baja representación que hay de ciertos idiomas, así como lenguas (ej. lengua de señas).

■ TAKE AWAYS AND WHAT WE ARE THINKING?

Creo que algunas maneras de abordar estos problemas son la diversificación de Datos, es decir, asegurar que los datos de entrenamiento sean variados y representen a todos los grupos. Hay que realizar auditorías y evaluaciones. También hay que hacer diseños inclusivos, hay que involucrar a diversos grupos en el diseño y desarrollo de algoritmos. Los algoritmos y sus decisiones deberían ser transparentes y explicables. Y También me parece muy importante que existan leyes que regulen la ética.

Transformer Implementation: <https://youtu.be/pgR-5oRKAag>

■ MOTIVATION

Esta sesión se centra en entender a fondo la arquitectura del transformer y su aplicación en tareas de procesamiento de lenguaje natural. Este análisis se realiza a través de la revisión del paper "Attention is All You Need" y la implementación de BERT en el contexto del dataset MMIMDb. El objetivo es proporcionar una comprensión detallada de cómo funcionan los transformers y su relevancia en la traducción automática y otras aplicaciones de NLP. La clase es presentada por Diego Moreno.

■ HOW IT WORKS

La presentación se estructuró en dos partes principales:

1. la explicación teórica del modelo transformer

Aquí se describió cómo el transformer utiliza un mecanismo de atención para procesar secuencias de tokens de un idioma a otro, enmascarando tokens no utilizados y empleando varias capas de encoders y decoders. Se detallaron componentes clave como la self-attention y las capas de feed-forward



2. la implementación práctica en un notebook

Aquí se mostró cómo implementar el modelo utilizando PyTorch, explicando cada paso del código, desde la clonación de capas hasta la normalización y la aplicación de funciones de atención y feed-forward.

■ RESULTS

Los resultados de la implementación del transformer y BERT en el dataset MMIMDb mostraron que estos modelos son altamente efectivos para tareas de clasificación y traducción de lenguaje.

En la revisión del código y la experimentación práctica se confirmó que los transformers, con su capacidad para paralelizar operaciones y manejar MUY grandes corpus de texto, ofrecen mejoras significativas en precisión y eficiencia comparados con modelos previos a estos. Creo que los resultados enmarcan la importancia de utilizar arquitecturas basadas en atención para aplicaciones complejas de NLP.

■ WHY IT MATTERS?

Me parece que entender y aplicar los transformers es crucial porque representan el estado del arte en modelos de NLP actualmente, estos superan técnicas anteriores en términos de precisión y versatilidad. Desde luego se encuentra el problema de lo costoso que es computacionalmente, aún así me parece de suma importancia entender a fondo el funcionamiento de éstos para así poder proponer versiones asequibles. La capacidad de estos modelos para manejar grandes volúmenes de datos y aprender representaciones contextuales ricas es fundamental para avanzar en aplicaciones como traducción automática, análisis de sentimientos y otras áreas críticas de la inteligencia artificial.

■ TAKE AWAYS AND WHAT WE ARE THINKING?

Los aspectos de este video que me parecen más relevante incluyen la importancia de una implementación cuidadosa y detallada para aprovechar al máximo las capacidades de los transformers. También práctica con el notebook y la revisión del código fue esencial para mí para desarrollar una mejor comprensión y más profundidad de estos modelos. Me parece que es fundamental continuar explorando y experimentando con transformers en diferentes contextos y tareas para descubrir nuevas aplicaciones y mejorar las existentes. Sobre todo teniendo en cuenta que nos encontramos en un momento crucial del desarrollo de estas arquitecturas.

BERT y GMMU: <https://www.youtube.com/watch?v=q8rI21Y3pyU>

■ MOTIVATION

La sesión se centra en explorar y comprender en detalle el modelo BERT y su aplicación en la clasificación multimodal utilizando el modelo GMMU (Gated Multimodal Unit). Se explica cómo BERT puede ser adaptado y extendido para tareas que involucren múltiples modalidades de datos con un simple finetuning en particular se muestra un ejemplo usando texto, imágenes y audio, en el contexto de clasificación de películas.

■ HOW IT WORKS

La presentación comenzó con una explicación de BERT, un modelo basado en transformers que se entrena en dos fases:



1. preentrenamiento
2. ajuste fino (fine-tuning)

BERT utiliza técnicas como el enmascaramiento de palabras con el token especial [MASK] usando como referencia los modelos MML y la predicción de la siguiente oración (Next Sentence Prediction) para aprender representaciones contextuales del lenguaje.

Por otro lado, el modelo GMU se implementa para combinar diferentes modalidades de datos mediante compuertas que controlan la contribución de cada modalidad. El notebook provee una implementación detallada de BERT para la clasificación de géneros de películas en el dataset MMIMDb, donde se combina texto con pósters y otros datos visuales utilizando la arquitectura GMU.

■ RESULTS

Se mostró que la combinación de BERT con GMU para clasificación multimodal mejora significativamente el rendimiento en comparación con el uso de BERT solo. El modelo GMU original propuesto por Arévalo logró una mejora en el score micro F1 al integrar múltiples modalidades, alcanzando un 50% en comparación con el 40% de BERT solo. Además, la implementación jerárquica de GMU a pesar de ser más compleja no mostró mejoras realmente significativas sobre el GMU original. Lo que sugiere que la simplicidad del modelo original es suficiente, al menos para esta tarea específica.

■ WHY IT MATTERS?

Este tema es crucial porque muestra cómo los modelos de procesamiento de lenguaje natural como BERT pueden ser extendidos para tareas multimodales, que son comunes en aplicaciones del mundo real como la clasificación de contenido multimedia. La capacidad de integrar diferentes tipos de datos y mejorar la precisión de las predicciones tiene implicaciones significativas para el desarrollo de sistemas de inteligencia artificial más robustos y versátiles.

■ TAKE AWAYS AND WHAT WE ARE THINKING?

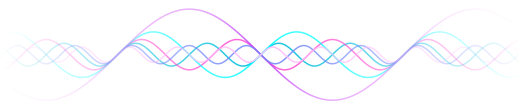
Las conclusiones que me parecen más importantes son la efectividad de BERT en combinación con GMU para tareas de clasificación multimodal. La implementación práctica en el notebook demuestra cómo adaptar y extender modelos preentrenados para aplicaciones específicas, resaltando la importancia de utilizar arquitecturas adecuadas para cada tipo de tarea.

Creo que es esencial continuar explorando y optimizando estas combinaciones para mejorar aún más el rendimiento y adaptabilidad no sólo en los modelos de lenguaje sino en general de los modelos de inteligencia artificial en contextos complejos y variados.

Music Generation <https://youtu.be/oPmOgWDG1BU>

■ MOTIVATION:

La motivación de esta sesión fue desarrollar un modelo capaz de generar música utilizando capas recurrentes y transformers. Este proyecto se realizó en la tesis de maestría de Abdiel Beltrán con el Doctor Pastor, el objetivo de ésta fue aplicar técnicas avanzadas de procesamiento de secuencias para la generación automática de música. Se buscó explorar modelos como GRU y transformers que pueden aprender patrones musicales a partir de datos MIDI y generar nuevas secuencias coherentes. Abdiel mencionó que no tiene conocimientos musicales



profesionales lo que hace muy interesante que pueda abordar este tema únicamente desde las técnicas de inteligencia artificial.

■ HOW IT WORKS

La implementación se centra en el uso de la librería music21 para manejar datos musicales y archivos MIDI. El proceso comienza con la carga y preprocesamiento de datos MIDI, estos se transforman en secuencias de notas. Luego, estas secuencias se convierten en representaciones numéricas y ya pueden ser utilizadas por modelos de aprendizaje profundo. Se utilizan modelos GRU y transformers para entrenar con estas secuencias, estos aprenden a predecir la siguiente nota dada una secuencia de entrada.

Se implementan tres modelos:

1. con una capa GRU
2. con una capa GRU y atención global
3. con un encoder de transformer

Cada modelo se entrena y se evalúa en su capacidad para generar música coherente.

■ RESULTS

Se mostró que la capa GRU es capaz de generar secuencias musicales coherentes, aunque no muy complejas. La implementación con atención global mejora ligeramente la coherencia de las secuencias generadas, mientras que el encoder de transformer tiende a reciclar notas y no genera secuencias tan variadas.

También se observó que el tamaño de la secuencia de entrada influye en el rendimiento del modelo, siendo el GRU más adecuado para secuencias más cortas, mientras que el transformer podría beneficiarse de secuencias más largas. La evaluación de los modelos incluyó la generación de nuevas secuencias musicales y la comparación de la coherencia y la variedad de las notas generadas.

■ WHY IT MATTERS?

Este proyecto es relevante porque explora el uso de técnicas avanzadas de aprendizaje profundo para la generación automática de música, una aplicación creativa del procesamiento de secuencias. La capacidad de generar música de manera autónoma tiene implicaciones significativas en la industria de la música, la educación y el entretenimiento, ofreciendo nuevas herramientas para compositores y artistas, así como nuevas formas de interacción con la música.

■ TAKE AWAYS AND WHAT WE ARE THINKING?

Lo que me queda de esta etapa es la efectividad de las capas GRU para la generación de música y las limitaciones actuales de los transformers en este contexto específico. Aunque la implementación práctica mostró que, aunque los transformers tienen potencial, su rendimiento puede mejorar con secuencias de entrada más largas y posiblemente con ajustes en la arquitectura.

Me parece que adelante, es importante explorar más a fondo cómo adaptar estos modelos para mejorar la generación de secuencias musicales, experimentando con diferentes configuraciones y técnicas de preprocesamiento. La comunidad de investigación en IA y música puede beneficiarse de estos hallazgos para desarrollar modelos más sofisticados y eficientes para la creación musical autónoma.



UDA https://youtu.be/_CfFm6fyn3s

- **MOTIVATION** En esta clase que dio Ivonne Monter, la motivación fue aplicar el esquema de Unsupervised Data Augmentation (UDA) para la clasificación multimodal, utilizando tanto texto como imágenes. EL objetivo principal fue el crecimiento en el volumen de información multimodal en los últimos años y la necesidad de mejorar la capacidad de generalización de los modelos en estas tareas.

También busca reducir el sobreajuste que ocurre debido a la mayor complejidad (por la naturaleza de los datos) y el número de parámetros en modelos multimodales, aprovechando técnicas de aumento de datos no supervisado para mejorar la precisión y sobre todo la robustez del modelo.

- **HOW IT WORKS**

La implementación de UDA se realiza en el dataset de películas MovieLens, donde se trabaja con sinopsis de películas (texto) y sus pósters (imágenes).

Se utilizó BERT para el procesamiento de texto y una red convolucional ResNet-152 para el procesamiento de imágenes. UDA aplica técnicas de aumento de datos tanto para texto (back-translation) como para imágenes (RandAugment), creando versiones perturbadas de los datos originales para entrenar el modelo de manera más robusta.

La pérdida de consistencia se calcula usando la divergencia de Kullback-Leibler entre la distribución del dato original y el aumentado, y se combina con la pérdida supervisada para optimizar el modelo. Se configuran diferentes modelos para texto, imágenes y multimodal, y se entrena y evalúa cada uno con y sin UDA.

- **RESULTS**

Los resultados muestran que aplicar UDA mejora la precisión y reduce el sobreajuste en los modelos. En el modelo de texto (BERT), la precisión macro y micro mejora ligeramente al aplicar UDA, alcanzando 7% y 73% respectivamente. En el modelo de imágenes (ResNet-152), la mejora es más significativa, con un incremento en la precisión macro de 43% a 44% y en la precisión micro de 51% a 53%. El modelo multimodal también muestra mejoras, con un incremento en la precisión macro de 71.5% a 72.3% y en la precisión micro de 74.1% a 74.9%. Estos resultados indican que UDA es más efectivo en imágenes, pero también ofrece beneficios en

- **WHY IT MATTERS?**

Este trabajo es relevante porque aborda uno de los desafíos más importantes en el aprendizaje multimodal: el sobreajuste. Al aplicar UDA, se mejora la capacidad de los modelos para generalizar a datos no vistos, lo cual es crucial en aplicaciones del mundo real como la recomendación de contenido, la visión por computadora y la interpretación de lenguaje natural. Mejorar la precisión y la robustez de los modelos multimodales tiene implicaciones significativas en diversas industrias, incluyendo el entretenimiento, la seguridad y la conducción autónoma.

- **TAKE AWAYS AND WHAT WE ARE THINKING?**

Las principales lecciones incluyen la efectividad de UDA para mejorar la precisión y reducir el sobreajuste en modelos multimodales. Los resultados sugieren que el aumento de datos no supervisado es una técnica poderosa que puede aplicarse en múltiples dominios para



mejorar la robustez de los modelos. En adelante, es crucial explorar más a fondo diferentes técnicas de aumento de datos y su impacto en otras modalidades y tareas. También sería beneficioso investigar cómo combinar mejor las modalidades y ajustar dinámicamente los parámetros de regularización para optimizar el rendimiento del modelo. La comunidad de aprendizaje automático puede beneficiarse significativamente de estos hallazgos, promoviendo el desarrollo de modelos más precisos y generalizables.

José Cañete <https://open.spotify.com/episode/54TMM8W01fjjL404MybEtV?si=5b5894ae823a48e2>

- **MOTIVATION:**

La motivación de esta sesión radica en la creación y el impacto de BETO, un modelo basado en BERT entrenado específicamente con datos en español. José Cañete, co-creador de BETO, nos cuenta cómo este proyecto fue una oportunidad para llenar un vacío significativo en los recursos de Machine Learning en español, facilitando así la accesibilidad de herramientas de inteligencia artificial para la comunidad hispanohablante. También mencionó y recomendó un podcast de mujeres chilenas en el marco de la conmemoración del 8M.

- **HOW IT WORKS**

La creación de BETO comenzó como un pequeño proyecto de investigación con un doctor de su universidad y lo primero que hicieron fue llevar a cabo una recopilación exhaustiva de textos en español para formar un dataset robusto. Sus profesores y algunos investigadores jugaron un papel crucial en este proceso, desarrollando recursos en Machine Learning en español. La parte técnica del proyecto involucró la utilización del modelo BERT y su técnica de atención, que permitió la paralelización de las operaciones y el manejo adecuado corpus gigantes de textos.

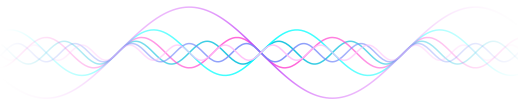
El entrenamiento del modelo se llevó a cabo utilizando hardware de alto rendimiento como GPUs, y se emplearon herramientas tecnológicas como Python, PyTorch, Jupyter Notebooks, y plataformas como Hugging Face.

- **RESULTS**

BETO ha demostrado ser un modelo efectivo en el procesamiento de lenguaje natural para el español. Ha cerrado una brecha significativa en la disponibilidad de modelos de NLP en nuestro idioma, proporcionando resultados sólidos en diversas tareas concretas. El modelo ha facilitado mejores evaluaciones y aplicaciones de NLP en español, destacando su importancia y eficacia. Las conclusiones indican que BETO ha sido un éxito y que la metodología utilizada para su creación puede ser replicada para otros idiomas y contextos.

- **WHY IT MATTERS?**

Este tema es crucial porque aborda la falta de modelos de NLP en español, uno de los cinco idiomas más hablados en el mundo. La creación de BETO representa un avance significativo en la democratización y diversificación de la inteligencia artificial, haciendo que herramientas poderosas sean accesibles para la comunidad hispanohablante, lo que potencialmente mejora diversas aplicaciones tecnológicas en esta lengua. Me parece que a futuro sería importante expandir esto a otras lenguas con aún menos recursos.



- TAKE AWAYS AND WHAT WE ARE THINKING?

Aquí me parece que destaca importancia de la colaboración interdisciplinaria y la necesidad de salir de la zona de confort para innovar. La creación de BETO demuestra que es posible desarrollar modelos de NLP efectivos y específicos para idiomas menos representados, lo que abre la puerta a futuros proyectos similares.

Me parece imprescindible seguir explorando y expandiendo estos esfuerzos para llevar a mejoras significativas en la accesibilidad y eficacia de las herramientas de inteligencia artificial en diversos idiomas.

Juan Manuel Pérez <https://open.spotify.com/episode/65akBnfQXJ2AamgKHbXmJx?si=a7ba984bddb145b0>

- MOTIVATION:

La presentación abordó el desarrollo y la implementación del modelo Robertuito desde sus inicios. Comenzó como una tesis de licenciatura enfocada en identificar y analizar el discurso de odio en las redes sociales. El enfoque práctico incluyó el entrenamiento de un modelo desde cero utilizando textos en español de Twitter, dada la falta de recursos en el idioma español comparado con el inglés.

La librería resultante, llamada PySentimiento, se diseñó para facilitar el análisis de opiniones y sentimientos, y para ser accesible a científicos sociales que no manejen Python. Se discutió el proceso de entrenamiento del modelo, los desafíos enfrentados y las herramientas utilizadas como PyTorch y recursos de plataformas como Hugging Face.

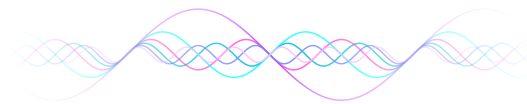
- HOW IT WORKS

La presentación consistió en pequeña entrevista a Juan Manuel Pérez. Se tocaron diversos temas, desde algunos de sus intereses personales actuales, pasando por qué y para qué sirve realizar un doctorado; hasta algunos hitos importantes en la disciplina de NLP como la llegada de Bert y sus distintas adaptaciones. De esto último precisamente se desprende el tema principal que es Robertuito, un modelo de lenguaje preentrenado para análisis de sentimiento en tweets en español. Se mencionan Beto, Roberta, y Pysentimiento así como algunos otros modelos, que son varios recursos de NLP en español. Se menciona que se basaron en el modelo Robertweet en inglés que también se entrenó en tweets pero en inglés y entran un poco en detalles sobre el proceso que realizaron para el procesamiento del texto de redes sociales.

Todo el podcast se desarrolla en un lenguaje muy casual sin muchas palabras técnicas, aunque sí se requiere de conocimiento sobre el panorama general actual de NLP para poder seguirlo.

- RESULTS

Los principales resultados presentados en este podcast fueron, primero, la presentación de Robertuito: un modelo de lenguaje previamente entrenado para texto de redes sociales (Twitter) en español. Este modelo fue entrenado en más de 500 millones de tweets. Y, segundo, la librería Pysentimiento, ésta es un pequeño conjunto de modelos pre-entrenados de transformers, en ellos se emplea la versión de BERT en Español: BETO y los datos de la versión 2020 del Taller de Análisis de Sentimiento (TASS) de la Sociedad Española de Procesamiento de Lenguaje Natural (SEPLN).



Y también

- WHY IT MATTERS?

Creo que estos temas son relevantes por varias razones, primero porque aborda la necesidad crítica de herramientas específicas para el análisis de textos en español (pero puede extenderse a cualquier idioma o lengua). Estas áreas se han estado desarrollando significativamente menos en comparación con el inglés. Robertuito y la librería PySentimiento proporcionan recursos vitales para la comunidad hispanohablante, permitiendo un mejor entendimiento y mitigación del discurso de odio en las redes sociales.

- TAKE AWAYS AND WHAT WE ARE THINKING?

Me parece que lo más importante que me llevo de aquí es la importancia de crear herramientas adaptadas a las necesidades lingüísticas específicas de diversas comunidades, en nuestro caso, el español pero también sería interesante ampliar el horizonte hacia las lenguas indígenas o de señas. La creación de Robertuito demuestra cómo un proyecto de tesis puede evolucionar para tener un impacto significativo. En adelante, es crucial seguir desarrollando y refinando estos modelos (y otros), incorporando más datos y avanzando en técnicas de procesamiento de lenguaje natural para mejorar la precisión y la eficacia. Como comunidad de NLP deberíamos continuar innovando y colaborando para enfrentar los desafíos del análisis de textos en idiomas menos representados.