

1. Objetivo

El objetivo de este proyecto fue desarrollar un modelo de deep learning capaz de predecir el nivel de engagement potencial de diferentes puntos de interés (POIs) turísticos. Este modelo combina tanto información visual, obtenida de las imágenes de los POIs, como metadatos asociados.

La predicción de engagement es una tarea clave para identificar patrones visuales y contextuales que correlacionan con interacciones más altas, optimizando así la selección y priorización de contenido en plataformas turísticas.

2. Descripción del Dataset

El dataset utilizado contiene información relevante sobre puntos de interés turísticos, incluyendo:

- Imágenes principales de los POIs organizadas por identificadores únicos.
- Metadatos estructurados, como ubicación geográfica, categoría, nivel de importancia (tier) y visitas.
- Métricas de engagement, como likes, dislikes y bookmarks.

El dataset fue proporcionado en formato ZIP para las imágenes, y un archivo CSV para los metadatos.

3. Metodología

3.1 Preparación y Análisis de Datos

1. Análisis Exploratorio:

- Se exploraron las distribuciones de las variables principales, identificando desequilibrios en las clases de engagement. Esto justifica el uso de `WeightedRandomSampler` en el entrenamiento para manejar este desbalance.

2. Preprocesamiento:

- Las imágenes fueron redimensionadas a 224x224 píxeles y normalizadas.
- Los datos faltantes se manejaron generando imágenes negras para evitar interrupciones durante el entrenamiento.
- Los metadatos numéricos, como `locationLat` y `locationLon`, fueron escalados utilizando `StandardScaler`.

3. Creación de Etiquetas:

- La métrica de engagement se binarizó basándose en la proporción de likes sobre visitas ($\text{ubral} > 0.1$).

4. División del Dataset:

- Se realizó una división en train (60%), validation (20%) y test (20%), manteniendo un equilibrio en las clases.
- Se implementó una clase POIDataset personalizada para manejar tanto imágenes como metadatos.

3.2 Arquitectura del Modelo

1. Modelo Híbrido:

- Se desarrolló un modelo que combina EfficientNet-B0 y ResNet18 pre-entrenados para procesar las imágenes.
- Una red fully connected para procesar los metadatos.

2. Estrategias Arquitectónicas:

- Se congelaron las capas iniciales de las redes preentrenadas para aprovechar sus representaciones generalizadas y entrenar sólo las capas finales.
- Las salidas de EfficientNet y ResNet se concatenaron con los metadatos procesados antes de pasar por las capas finales del clasificador.
- Se utilizó una capa final de Sigmoid para la clasificación binaria.

3. Justificación:

- La combinación de arquitecturas preentrenadas permitió capturar diferentes niveles de representación visual, mientras que las capas fully connected integraron información contextual clave.

3.3 Entrenamiento y Optimización

1. Pipeline de Entrenamiento:

- Se utilizó el optimizador Adam con un aprendizaje dinámico, ajustado mediante la optimización de hiperparámetros.
- Se implementaron early stopping y dropout para mitigar el overfitting.
- Uso de WeightedRandomSampler para manejar el desbalance de clases.

2. Optimización de Hiperparámetros:

- Utilizando Optuna, se determinaron los valores óptimos de learning rate, dropout, weight decay, tamaño de batch, numero de epochs.

3. Técnicas Anti-Overfitting:

- Dropout: Aplicado en las capas fully connected.
- Early Stopping: Implementado con una paciencia de 5 épocas.

3.4 Evaluación y Análisis

1. Evaluación:

- Se calcularon métricas como Accuracy, F1-score, Precision y Recall en el conjunto de prueba.

2. Análisis de Errores:

- Se identificó un sesgo hacia la clase mayoritaria debido al desbalance de datos, ajustado mediante WeightedRandomSampler.

3. Resultados:

El mejor modelo logró un accuracy de prueba de aproximadamente 61%, con un F1-score de 0,60. Aunque estos resultados no son prometedores, hay margen de mejora.

- Accuracy: 61%
- F1-score: 0.60
- Precision: 0.73
- Recall: 0.45

4. Desafíos:

- Complejidad del problema: La variabilidad en los datos y la naturaleza multimodal del problema (imágenes + metadatos) justificó el uso de un modelo híbrido que combinara redes neuronales convolucionales (EfficientNet y ResNet) para procesar imágenes con capas fully connected para metadatos.
- Balancear la contribución de las características de imagen y metadatos fue complicado. Experimenté con diferentes tamaños de capas intermedias antes de encontrar una configuración óptima.
- El modelo mostraba signos de overfitting en las primeras iteraciones. Ajusté el dropout y el weight decay para mitigar este problema.
- A pesar de los esfuerzos de balanceo de clases, el modelo tendía a favorecer una clase sobre otra. Esto requirió ajustes adicionales en la función de pérdida y en la estrategia de muestreo.

5. Para futuros desarrollos, planeo:

- Experimentar con técnicas de data augmentation más avanzadas.
- Investigar arquitecturas de atención para mejor integración de imágenes y metadatos
- Realizar un análisis más profundo de las características más influyentes en las predicciones del modelo.

6. Conclusión:

Este proyecto representó un desafío técnico y analítico significativo, demostrando la importancia de integrar datos multimodales para resolver problemas complejos. A pesar de los desafíos, las soluciones implementadas llevaron a un modelo funcional, con potencial para optimizar el contenido en plataformas turísticas.

Las lecciones aprendidas y las áreas de mejora identificadas proporcionan una base sólida para futuros desarrollos, incluyendo el uso de arquitecturas avanzadas y optimizaciones adicionales. Este trabajo destaca la importancia de la experimentación iterativa y del aprendizaje continuo en el campo del deep learning.