

Proyecto Final

Fecha límite de entrega: Domingo 9 de Marzo

Grupos de máximo 3 personas

Intr. Sistemas Inteligentes - 2024-2

Objetivo:

Desarrollar modelos de clasificación que permitan cuantificar automáticamente la celularidad tumoral en imágenes de histopatología.

Descripción

El BreastPathQ: Cancer Cellularity Challenge 2019 (<https://breastpathq.grand-challenge.org/>) fue una iniciativa organizada por SPIE, la Asociación Americana de Físicos en Medicina (AAPM) y el Instituto Nacional del Cáncer (NCI) de EE. UU. El objetivo del desafío era desarrollar biomarcadores cuantitativos para determinar la celularidad del cáncer a partir de imágenes de diapositivas completas (WSI) de muestras patológicas de cáncer de mama teñidas con hematoxilina y eosina (H&E). Los participantes debían crear métodos automatizados para analizar parches histológicos extraídos de estas imágenes y asignarles una puntuación que reflejara la celularidad tumoral.

El conjunto de datos utilizado en el desafío fue recopilado en el Sunnybrook Health Sciences Centre de Toronto e incluyó 96 imágenes de diapositivas completas de 64 pacientes con cáncer de mama invasivo residual tras terapia neoadyuvante. Estas imágenes fueron escaneadas a una magnificación de 20X. El conjunto de entrenamiento y validación constaba de 2,579 parches extraídos de 69 de estas imágenes, cada uno con una puntuación de celularidad asignada por un patólogo experto. El conjunto de prueba contenía 1,121 parches de 25 imágenes, con referencias estándar de dos patólogos. ^[108]

Aunque el desafío oficial concluyó, el sitio web continúa ofreciendo una tabla de clasificación pública para la evaluación continua de algoritmos. En total, participaron 484 equipos, con 322 envíos de soluciones.

El reto de este proyecto es construir un modelo que resuelva el problema y hacer al menos un envío al sitio de la competencia. Todo el proceso de entrenamiento, validación y prueba del modelo debe quedar debidamente documentado.

Acceso a la competencia y los datos

1. Al menos un integrante del equipo debe registrarse en la plataforma <https://grand-challenge.org/>. El nombre de usuario debe ser en lo posible el mismo del usuario de la UNAL o una modificación (adicionar un número o una letra), de manera que sea fácilmente identificable en el leaderboard de la competencia.
2. En la página de la competencia (<https://breastpathq.grand-challenge.org/>) presionar el botón **join** para unirse al mismo. Esto les permitirá unirse a la competencia y hacer envíos al mismo.
3. En la sección **data sets** pueden descargar los datos de entrenamiento, validación y pruebas. Los datos de prueba no tienen etiquetas, ustedes las deben predecir usando sus modelos.
4. En el tab **submit** pueden hacer sus envíos en la forma de un archivo csv. Las instrucciones se encuentran en la página.

Etapas del modelo

El notebook que usen para resolver el problema debe tener las siguientes secciones

A. Descripción del problema

Deben hacer una descripción clara del problema que se quiere resolver. Definir los términos técnicos y explicar la relación del problema de machine learning con el problema práctico.

B. Importación de datos

En esta sección, se importan las bibliotecas necesarias y se cargan los datos desde la fuente correspondiente, archivos csv (para las etiquetas), archivos de imágenes, etc.

C. Exploración de datos

En esta sección, se realizan análisis exploratorios de los datos para comprender mejor su estructura y distribución. Puede incluir visualizaciones y estadísticas descriptivas y la identificación de patrones interesantes en los datos.

D. Modelos y ajuste de hiper-parámetros

Aquí se seleccionan y entrenan los modelos de machine learning. Pueden probarse varios modelos, dependiendo de la naturaleza del problema. Se lleva a cabo la búsqueda y ajuste de los hiper parámetros del modelo para mejorar su rendimiento, usando los conjuntos de entrenamiento y validación.

E. Evaluación del desempeño sistemática del modelo final seleccionado

Una vez que se ha seleccionado el modelo final, se evalúa su rendimiento de manera sistemática. Además de otras métricas se debe calcular la métrica PK usada en la competencia. Deben consultar sobre la misma y explicar cómo funciona.

F. Conclusiones y resultados

A partir de los resultados obtenidos analizar los resultados de las etapas del modelo.

¿Qué se puede entender de los datos y las predicciones?

¿Cómo se comportó el modelo?

¿Qué aspectos se tuvieron en cuenta para mejorar el rendimiento del modelo?

G. Generación del archivo de envío.

Se debe aplicar el(los) modelo(s) final(es) a los datos de prueba y construir el archivo csv de envío en el formato requerido por la competencia. Se debe hacer el envío y tomar un pantallazo del puntaje y ubicación en el leaderboard.

Entregable

Además de realizar los envíos al sitio de la competencia, se debe entregar:

- **Jupyter Notebook:** con todo el código del proyecto. El notebook debe estar debidamente explicado usando celdas de texto. Todas las etapas deben estar debidamente documentadas. Asegúrese antes de hacer el envío de que el notebook esté libre de errores, se haya guardado correctamente y se visualiza apropiadamente al volver a cargarse.

Se debe enviar el notebook antes de la media noche de la fecha límite a través del siguiente link: <https://www.dropbox.com/request/E5m9YDvXi61lJoO2jls9>. El archivo debe nombrarse isi-proj--userunal1_userunal2_userunal3.ipynb, donde <userunal> es el nombre de usuario asignado por la universidad. Envíos que no sigan este formato o que se envíen después de la hora límite no se tendrán en cuenta.

Evaluación

Todos los grupos deben hacer envíos al sitio de la competencia y deben obtener al menos un 40% de exactitud (accuracy). Si esto no se cumple no se evaluará el notebook.

- **80% Notebook**
- **20% Desempeño en la tarea.** Se asignará una calificación de acuerdo con el desempeño en la competencia..