

Propuesta de proyecto y rúbricas

En este documento se definen los puntos que deberán considerar para la propuesta de su proyecto final, del curso de Aprendizaje Profundo. Además, se define el formato requerido del documento donde reportarán las actividades realizadas con respecto al proyecto seleccionado. Se incluye también el contenido que deben considerar en la presentación de su proyecto.

Propuesta de proyecto

El proyecto de este curso es una oportunidad para que desarrollen una aplicación (implementación) de aprendizaje profundo para un tema de su elección. Si bien este proyecto tiene cierta estructura, deberán lidiar con la ambigüedad y la toma de decisiones importantes que conforman la vida de un verdadero profesional en ciencia de datos.

Reglas

1. Los proyectos deben realizarse en grupos de 5.
2. El proyecto debe hacer uso del aprendizaje profundo, como se enseña en este curso, y el entrenamiento, validación y prueba de la arquitectura(s) de la red neuronal elegida debe formar una parte importante del proyecto.
3. El tema del proyecto es de su propia elección. Debe ser único dentro de la clase. No pueden haber dos proyectos sobre el mismo tema o usar el mismo conjunto de datos.
4. Debe haber alguna recopilación y/o limpieza de datos que es una parte significativa del proceso de entrenamiento. No podrán utilizar un conjunto de datos conocido para el mismo propósito para el que fue desarrollado ese conjunto de datos. Pueden reutilizar un conjunto de datos, combinar conjuntos de datos o recopilar sus propios datos.

Formación de equipos

Una vez que hayan formado un equipo, deben hacer una lluvia de ideas sobre cuál debería ser el tema de su proyecto. Cada equipo debe tener una idea de proyecto única y utilizar un conjunto de datos único. Para reclamar su idea/datos, envíen un correo electrónico al profesor con:

- Los nombres, números de cuenta y direcciones de correo electrónico de los miembros del equipo.
- Una breve descripción (1 o 2 párrafos) de la idea del proyecto en el que desean trabajar.
- Los conjuntos de datos que planean usar y cómo planean recopilar/limpiar/procesar los datos.

Por correo les haré saber que ustedes fueron los primeros en elegir dicho tema para su proyecto (lo que significa que nadie más ha propuesto un tema similar). O bien les comunicaré que otro equipo ya eligió ese tema y deberán proponer uno diferente.

Reporte proyecto y rúbricas

Su informe será estructurado como un trabajo de investigación y describe lo que han hecho. Para ello pueden emplear un formato definido como [Springer Lecture Notes in Computer Science](#) o [eHealth-KD 2020 Guidelines](#). También pueden utilizar Google Docs o Word. La estructura que deberá llevar su documento es la siguiente.

Introducción (5 puntos)

Una breve descripción de las motivaciones detrás de su proyecto, el objetivo de su proyecto, por qué es interesante o importante y por qué el aprendizaje profundo es un enfoque razonable.

Trabajos relacionados (5 puntos)

Una descripción de 1-2 trabajos relacionados en el campo, para proporcionar al lector una idea de lo que ya se ha hecho en esta área, por ejemplo, artículos (papers) o productos, software existentes que hacen algo relacionado, etc.

Metodología (10 puntos)

Una descripción de la metodología que propusieron para resolver el problema del tema elegido. Debe de incluir una figura o un diagrama que ilustra el modelo general o la idea de su proyecto.

Procesamiento de los datos (20 puntos)

Describan los datos que ha recopilado y limpiado. Deben ser claros y específicos al describir lo que han hecho, para que un compañero de clase pueda reproducir su trabajo. Muestren algunas estadísticas y ejemplos de sus datos.

Arquitectura propuesta (30 puntos)

Una descripción de la arquitectura del modelo de red neuronal final. No describan todos los modelos intermedios que probaron. En su lugar, presenten el modelo (o modelos) cuyos resultados mostrarán. Estos deberían ser sus modelos más interesantes. Sean tan específicos como puedan mientras sean concisos. Los lectores deberían poder reproducir un modelo lo suficientemente similar al de ustedes y obtener un rendimiento similar.

Modelo de referencia (10 puntos)

Descripción de un modelo básico con el que compararon su red neuronal. Puede ser un modelo simple de machine learning, una red neuronal básica, un modelo heurístico codificado a mano (que no usa aprendizaje automático), etc.

Resultados (10 puntos)

Una descripción de los resultados a los que llegaron. Pueden incluir Figuras y Tablas, acompañadas de una breve descripción.

Incluya algunos resultados de muestra de su modelo para ayudar al lector a comprender mejor lo que puede hacer su modelo.

Describan los resultados observados así como el comportamiento de su modelo: por ejemplo, ¿Por qué su modelo funcionó bien?, ¿Hay algún tipo de entrada en la que el modelo no funciona bien?, etc.

Discusión (10 puntos)

Discutan sus resultados. ¿Creen que su modelo está funcionando bien?, ¿Por qué sí o por qué no?, ¿Qué es inusual, sorprendente o interesante acerca de sus resultados? ¿Qué aprendieron?

Referencias

Listar los libros, papers, urls, software que mencionan en el reporte. El formato puede ser [APA](#) o alguno de los utilizados por las templates de las plantillas de Latex mencionados arriba.

Presentación

Cada presentación tendrá una duración máxima de 30 minutos. Para garantizar que todos los estudiantes tengan la oportunidad de practicar sus habilidades de presentación, cada miembro del equipo debe hablar durante al menos 5 minutos. De este modo se darán 5 minutos para preguntas.

Contenido de la presentación

La estructura de la presentación es flexible, siempre y cuando cubran suficientemente el contenido necesario sobre el desarrollo y ejecución de su proyecto, de manera que la audiencia (sus compañeros de clase) puedan entender.

Estos son los principales puntos que deben de cubrir en la presentación:

- **Problema (20 puntos):** ¿Cuál es el problema que están tratando de resolver?, ¿Por qué es importante?
- **Datos (20 puntos):** ¿Qué datos utilizaron o recopilaron?, ¿Cómo se ven?, ¿La cantidad/tipo de datos es suficiente para su proyecto?, ¿Hubo algo complicado o extraño que debieron hacer con los datos, ... con la división entrenamiento/validación/prueba?
- **Tratamiento de datos (20 puntos):** ¿Qué tratamiento o proceso realizaron sobre sus datos? ¿Por qué?
- **Modelo (20 puntos):** ¿Qué modelo(s) construyeron? ¿Cuál es la arquitectura de su(s) modelo(s) (CNN, RNN, Con Atención, Transformers, etc.) ? Pueden colocar Figuras, o Diagramas
- **Resultados (20 puntos):** ¿Cómo están midiendo y comparando el desempeño de su modelo?, ¿Qué tan bien se desempeñó su modelo?, ¿Cuáles son algunas predicciones de muestra generadas por su modelo?, ¿Tiene sentido?, ¿Seleccionó cuidadosamente los mejores resultados o está mostrando una muestra representativa?
- (Opcional) **Conclusiones:** ¿Qué aprendieron sobre el aprendizaje profundo al hacer este proyecto?

Notas finales

El reporte y la presentación junto con el notebook del proyecto deberá ser enviado por correo electrónico. El notebook puede ser compartido o enviado como archivo adjunto. La presentación y el reporte deberán estar en formato PDF.

80% Proyecto
20% Presentación

Fecha de presentación: 26 de mayo de 2022 (pendiente)
Fecha de entrega para el proyecto: 27 de mayo de 2022 (pendiente)