



Tarea 3: aprendizaje semi-supervisado de percepción urbana en grafos

Introducción

En esta tarea tendrá la oportunidad de experimentar con el uso de redes neuronales para grafos y aprendizaje semi-supervisado, con el fin de aprender a predecir atributos de percepción urbana, esta vez para la red vial de Santiago. En particular, dado un grafo que representa la red vial de Santiago, para los cuales solo algunos nodos poseen información perceptual, deberá entrenar GNNs para estimar estos atributos en otros nodos y arcos del grafo, ya sea de manera inductiva o transductiva. Una vez realizado esto, deberá estimar estimar rutas óptimas en la red, usando como métrica alguno de los atributos perceptuales estimados.

Para el desarrollo se utilizará el framework Pytorch y la librería Pytorch Geometric (PyG) sobre la plataforma [Colab](#) de Google. Dado que existe abundante material disponible en línea para colaborar con el desarrollo de la tarea, se espera que todo recurso externo utilizado, sea este código, librerías o papers, esté debidamente indicado.

Set de datos

En esta tarea trabajará con dos fuentes primarias de información. La primera será el archivo csv utilizado en la Tarea 2 (`vars_perceptuales_santiago.csv`), que contiene más de 100.000 registros georreferenciados en la ciudad de Santiago, cada uno indicando su ubicación espacial (latitud, longitud) y un vector de valores (puntajes/utilidad) para indicadores visuales asociados a estas ubicaciones. Específicamente, estos datos capturan variables continuas asociadas a la percepción visual de barrios urbanos, especificada a partir de 6 atributos distintos: bello, deprimente, aburrido, seguro, adinerado, animado. Más detalles sobre la metodología utilizada para construir esta base de datos pueden leerse [aquí](#).

La segunda fuente primaria será la red vial de la ciudad de Santiago, que puede ser descargada utilizando la librería *OpenStreetMap* (`osmnx`), como un grafo en formato *NetworkX*. El grafo descargado debe limitarse

al *bounding box* definido por las ubicaciones que posean los datos perceptuales y debe contener al menos las vías primarias y secundarias.

Además de estos datos, deberá utilizar como fuente secundaria de datos una serie de archivos de tipo *shape* de la ciudad de Santiago. Estos archivos poseen información sociodemográfica y de transporte georreferenciada, que deberá ser usada para caracterizar los distintos nodos y arcos.

Si requiere material adicional para estudiar sobre el uso de los archivos **shape**, uso de grafos en Python y de la librería *OpenStreetMap*, se recomienda revisar los recursos disponibles [acá](#).

Modelos

Para esta tarea, debe utilizar modelos de redes para grafos como los descritos en el capítulo 4 del curso, sin limitante en relación al tipo de capas que se puede utilizar (convolucional, de atención, *dropout*, *batchnorm*, etc). Se recomienda revisar bibliografía relacionada para esto. Considere además preprocesar las entradas de acuerdo a lo utilizado en la literatura (por ej., restar la media, normalizar en el rango $[0,1]$, etc.). No hay problema en basarse completamente en modelos propuestos previamente en la literatura o en tutoriales. En cualquier caso, debe justificar sus elecciones.

Tareas a realizar

Para la tarea se espera se espera que realice al menos las siguientes actividades:

- Combine los atributos perceptuales georreferenciados y el grafo de la red vial, imputando la información perceptual de cada registro del archivo csv, al nodo más cercano en el grafo que representa la red vial (y no al revés). Los valores perceptuales deben ser imputados considerando sus valores continuos y también cuantizados 5 niveles distintos (muy bajo, bajo, medio, alto, muy alto) de acuerdo a algún criterio definido por ud. Al finalizar este proceso, debe obtener un grafo de la red vial, que contenga nodos con y sin información perceptual.
- Aprendizaje transductivo: entrene un modelo de GNN utilizando la red vial de Santiago generada en el ítem anterior, que permita estimar los atributos perceptuales para cada nodo del grafo. Utilice como *features* a la matriz identidad (es decir, $X = I$), de modo que la única información disponible sea la propagada por los nodos. Realice el proceso para estimar inicialmente los valores cuantizados de percepción (clasificación de nodos) y luego repita lo mismo para los valores continuos (regresión), analizando las diferencias en los resultados. Considere que los grafos en formato de *NetworkX* pueden

ser fácilmente convertidos para ser utilizado en PyG, a través de métodos propios de esta última librería.

- Aprendizaje inductivo: repita el ítem anterior, pero esta vez utilizando un modelo de predicción, es decir, utilizando como *features* para los nodos información sociodemográfica o de transporte, extraída desde los archivos *shape* (es decir, $X \neq I$). Compare y analice los resultados, en función de los obtenidos en el ítem anterior.
- Rutas óptimas: diseñe un esquema para estimar atributos perceptuales para todos los arcos de la red (puede ser neuronal o un algoritmo tradicional), de modo de poder encontrar rutas óptimas entre distintos punto de la red, utilizando las variables perceptuales como criterio (por ejemplo, el camino más seguro entre Casa Central y San Joaquín). Visualice las rutas en un mapa y analice los resultados en función de los ítems anteriores y su conocimiento de la ciudad.

Desarrollo y entrega

La tarea puede desarrollarse de manera individual o en parejas. Se recomienda utilizar la plataforma Google Colab con el fin de facilitar la instalación de librerías. Esta plataforma permite utilizar gratuitamente una GPU para el entrenamiento por intervalos de 12 horas continuos. En el *notebook* desarrollado debe ir tanto el código como un informe (preferiblemente intercalados), donde se expliquen los pasos realizados, se analicen los resultados y se planteen conclusiones. La entrega de la tarea tiene como fecha límite el viernes 24 de junio a las 23:59, a través del buzón que se habilitará en el sitio del curso. Para fines de corrección, se revisará la última versión entregada.

Política de Integridad Académica

Los alumnos de la Escuela de Ingeniería deben mantener un comportamiento acorde al Código de Honor de la Universidad:

“Como miembro de la comunidad de la Pontificia Universidad Católica de Chile me comprometo a respetar los principios y normativas que la rigen. Asimismo, prometo actuar con rectitud y honestidad en las relaciones con los demás integrantes de la comunidad y en la realización de todo trabajo, particularmente en aquellas actividades vinculadas a la docencia, el aprendizaje y la creación, difusión y transferencia del conocimiento. Además, velaré por la integridad de las personas y cuidaré los bienes de la Universidad.”

En particular, se espera que mantengan altos estándares de honestidad académica. Cualquier acto deshonesto o fraude académico está prohibido; los alumnos que incurran en este tipo de acciones se exponen a un procedimiento sumario. Ejemplos de actos deshonestos son la copia, el uso de material o equipos no permitidos en las evaluaciones, el plagio, o la falsificación de identidad, entre otros. Específicamente, para los cursos del Departamento de Ciencia de la Computación, rige obligatoriamente la siguiente política de integridad académica en relación a copia y plagio: Todo trabajo presentado por un alumno (grupo) para los efectos de la evaluación de un curso debe ser hecho individualmente por el alumno (grupo), sin apoyo en material de terceros. Si un alumno (grupo) copia un trabajo, se le calificará con nota 1.0 en dicha evaluación y dependiendo de la gravedad de sus acciones podrá tener un 1.0 en todo ese ítem de evaluaciones o un 1.1 en el curso. Además, los antecedentes serán enviados a la Dirección de Docencia de la Escuela de Ingeniería para evaluar posteriores sanciones en conjunto con la Universidad, las que pueden incluir un procedimiento sumario. Por “copia” o “plagio” se entiende incluir en el trabajo presentado como propio, partes desarrolladas por otra persona. Está permitido usar material disponible públicamente, por ejemplo, libros o contenidos tomados de Internet, siempre y cuando se incluya la cita correspondiente.