

Plantilla para un artículo L^AT_EX

Domínguez-Adame Ruiz, Alberto

Universidad de Sevilla

albdomrui@alum.us.es

Sevilla

Vilaplana de Trias, Francisco David

Universidad de Sevilla

fravilde1i@alum.us.es

Sevilla

1 Introducción

Para empezar, el aprendizaje se define como la adquisición del conocimiento de algo por medio del estudio, el ejercicio o la experiencia, en especial de los conocimientos necesarios para aprender algún arte u oficio. Al hablar sobre el aprendizaje automático estaríamos entrando en el campo de la Inteligencia Artificial (IA). Podemos definir esta como un programa de computación diseñado para realizar determinadas operaciones que se consideran propias de la inteligencia humana, como el Aprendizaje Automático (Machine Learning), una rama de la inteligencia artificial, cuyo objetivo es el desarrollo de un sistema (modelo matemático que realiza una determinada tarea) el cual tiene un mejor desempeño con la experiencia, dados una serie de datos.

Nuestro estudio hace uso de datos relacionales, los cuales están unidos entre sí (tienen una relación) que queda representado como aristas en un grafo, entendiendo un grafo como un conjunto de nodos unido (o no) mediante aristas. En la Figure 1 podemos ver un ejemplo de un grafo a partir de unos datos relacionales.

Para la realización de este trabajo se ha hecho uso del entorno de desarrollo Jupyter y de la herramienta notebook. El código está escrito mediante

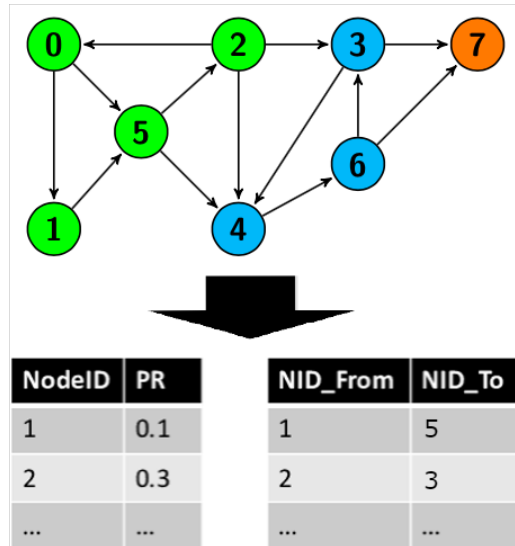


Figure 1: Ejemplo de Grafo

el lenguaje de programación Python, usando principalmente las librerías pandas, NetworkX y scikit-learn, y otras como keras, tensorflow y matplotlib.

Los modelos de clasificación usados son: KNN, Naive Bayes, Árboles de Decisión y Redes Neuronales.

2 Conjunto de datos y Tarea de predicción a realizar

Hemos elegido un dataset (fíjese en la Figure 2) de la página de streamings en directo twitch.tv , en la que se ve: las visitas, días (ambos de tipo entero) que transmitido un canal en directo, si tiene contrato o no con twitch (partner de tipo Boolean) y si el contenido del canal (id de tipo Integer) es o no para adultos (mature de tipo Boolean). Al ser la propia Twitch la que ofrece este contrato, hemos decidido que la predicción para nuestros modelos sea el atributo partner (true/false). Las aristas (edges) representan si dos usuarios tienen una relación de amistad.

Este dataset tenía dos tipos de id asociados a cada usuario, uno generado aleatoriamente y otro donde al usuario se le asignaba un valor entre 0 y el numero total de usuarios que recoge el conjunto de datos, por lo que

	id	partner	days	views	mature
0	2299	False	1459	9528	0
1	153	False	1629	3615	1
2	397	False	411	46546	1
3	5623	False	953	5863	1
4	5875	False	741	5594	1
...
7121	3794	False	2624	3174	0
7122	6534	False	2035	3158	1
7123	2041	False	1418	3839	1
7124	6870	False	2046	6208	1
7125	3919	False	1797	3545	0

Figure 2: Tabla de Datos

decidimos que eliminar el id generado de forma aleatoria para mayor claridad y sólo hace falta un id por usuario para identificarlos.

Para este estudio hemos decidido que el atributo a predecir será partner. Para ello, haremos uso de los atributos: days, views y mature (el id no es relevante en este caso). Lo hemos planteado como una tarea de clasificación binaria, ya que partner solo tiene dos valores posibles (True o False).

Por último hemos codificado (codificación one-hot) los valores de mature de Booleano a Integer, pasando True-False a 1-0 respectivamente, para que sea más eficiente estimar las probabilidades del atributo partner.

3 Métricas Relacionales

Como este es un trabajo de desarrollo de modelos orientados al aprendizaje automático grafos, hemos elegido una serie de métricas relacionales que aplicaremos como atributos relacionales. En concreto hemos elegido tres métricas:

1. **Betweenness centrality** (Centralidad de intermediación): La centralidad en un grafo puede ser entendida como una medida que determina la relevancia de un nodo dentro del grafo y permite comparar o

contrastar dicho vértice con otros. La Betweenness Centrality es una medida de centralidad que cuantifica la frecuencia en la que un nodo se encuentra en el camino más corto entre dos nodos determinados. Cuando en un grafo existen nodos de alta intermediación, estos suelen jugar un rol importante en la estructura a la que pertenecen. Estos nodos también poseen capacidades de ser controladores o reguladores de los flujos de información dentro de la estructura total del grafo.

2. **Clustering** (agrupamiento): es una tarea que tiene como finalidad principal lograr el agrupamiento de nodos, para lograr construir subconjuntos de datos conocidos como Clusters.
3. **Degree** (grado): Número de aristas (relaciones) que inciden sobre el nodo.

4 Algoritmos de Aprendizaje Automático

4.1 KNN

Primero tienes que cargar el archivo de imagen desde su computadora usando el enlace de carga del menú del proyecto. Luego usando el comando 'includegraphics' podrás incluirlo en el documento. Con el entorno de figura y el comando de título podrás agregar un número y un título a la figura. Mira el código de la Figura 1 en esta sección para ver un ejemplo.

4.2 Naive Bayes

Puedes añadir comentarios en el ícono + del menú de arriba.

Para responder a un comentario, simplemente da click en Reply en Rich Text.

También pueden añadirse comentarios en el margen del pdf compilado con el comando todo, como se muestra en el ejemplo de la derecha. También puedes añadirlos dentro del texto:

4.3 Árboles de Decisión

Usa los comandos table y tabular para iniciar una tabla simple — mira la tabla 1, como ejemplo.

l para left	c para centro	r para derecha
Ejemplo	Centrado	Alineado a la
Izquierda	13	Derecha

Table 1: Una simple tabla.

4.4 Redes Neuronales

L^AT_EX es buenísimo para escribir ecuaciones. Para escribir variables o ecuaciones dentro del texto lo podemos poner entre signos de pesos y luego podemos seguir escribiendo, esto funciona si queremos escribir un símbolo como ∇ , π , β , Ω , \aleph , etc.

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x \quad (1)$$

$$\int_0^1 dx = 1 \quad (2)$$

$$e^{i\pi} + 1 = 0 \quad (3)$$

Si queremos citar al gran Maxwell, lo podemos hacer como en la ecuación 4:

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \quad (4)$$

A continuación se añade un ejemplo de un desarrollo: Con este preámbulo llevamos a cabo la siguiente transformación de los operadores \hat{a}_ℓ

$$\hat{b}_m^\dagger = \sum_\ell U_m^\ell \hat{a}_\ell^\dagger \quad (5)$$

donde U_m^ℓ es un elemento de la matriz unitaria \mathbf{U} .

5 Resultados

Puedes añadir listas con numeración automática ...

1. Como esta,
 2. y como esta.
- ... o con puntitos ...
- Como este,
 - y como este.

6 Conclusiones

7 Bibliografía