

Los algoritmos de agrupación

Existen diversos algoritmos de agrupación que podemos aplicar para resolver problemas de búsqueda de patrones y encontrar y caracterizar estructuras en los datos. En la Fig. 1 se presentan algunos de ellos:

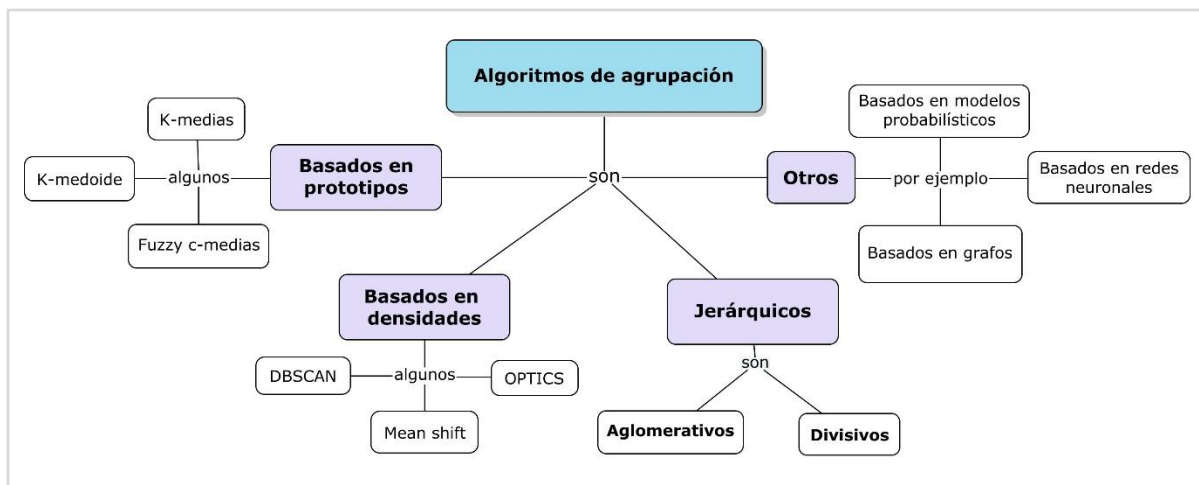


Fig. 1. Algunos tipos de algoritmos de agrupación.

Los algoritmos basados en prototipos o en particiones construyen los grupos con base en puntos promedios o representativos de los datos. Los grupos se encuentran al mismo nivel y pueden ser exclusivos o no. Estos algoritmos requieren que les especifiquemos a priori el número de grupos que van a formar a partir de los datos y utilizan una medida de similitud o de distancia para construir la agrupación.

Los algoritmos basados en densidad parten de la premisa de que un grupo es una región densa de objetos, que se separa de otros grupos por zonas de baja densidad. En general, se basan en un criterio de separabilidad exclusivo y los grupos se encuentran al mismo nivel. Para construir la agrupación es necesario determinar la densidad en el entorno de cada dato para decidir a qué grupo será asignado. No requieren que les especifiquemos el número de grupos, pero sí

parámetros que le suministren información de la vecindad sobre la cual basar sus decisiones y para determinar la contigüidad entre instancias.

Por su parte, los algoritmos jerárquicos utilizan un criterio de división que habilita que los grupos se organicen en forma de una jerarquía (lo cual induce un solapamiento entre grupos). Esto se puede realizar dividiendo o fusionando grupos, lo que da lugar a dos grandes clases de algoritmos: los divisivos y los aglomerativos. En general, no requieren que les especifiquemos el número de grupos a formar y utilizan métricas de enlace para determinar la similitud entre grupos.

Otros ejemplos de algoritmos que podemos emplear son los basados en modelos, que asumen que los datos siguen una distribución de probabilidad para derivar los grupos, como el caso del agrupamiento de mezclas Gaussianas (*Gaussian Mixture Clustering*); los basados en redes neuronales, como los mapas autoorganizados (*Self-Organizing Maps, SOM*), que utilizan un tipo de aprendizaje no supervisado para generar un espacio de menor dimensión, el mapa, que mantiene las características topológicas del espacio de entrada en la cual se pueden visualizar los grupos; y los basados en grafos, como el agrupamiento espectral (*Spectral Clustering*), que realizan una reducción de la dimensionalidad para construir un nuevo espacio de características en el cual se encuentran los grupos.

Como puedes ver, son muy variados los algoritmos de agrupación que puedes emplear. Cada tipo de algoritmo tiene sus fortalezas y debilidades, por lo que será importante comprender las características de cada uno y cómo se pueden emplear para obtener resultados significativos en diferentes contextos de aplicación. También será crucial la elección de la métrica y parámetros para determinar la similitud o contigüidad entre objetos, con base en el tipo de datos.

Bibliografía

Patel, A. (2019). *Hands-On Unsupervised Learning Using Python. How to Build Applied Machine Learning Solutions from Unlabeled Data*. O'Reilly.

Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly.

© - **Derechos Reservados:** la presente obra, y en general todos sus contenidos, se encuentran protegidos por las normas internacionales y nacionales vigentes sobre propiedad intelectual, por lo tanto su utilización parcial o total, reproducción, comunicación pública, transformación, distribución, alquiler, préstamo público e importación, total o parcial, en todo o en parte, en formato impreso o digital y en cualquier formato conocido o por conocer, se encuentran prohibidos, y solo serán lícitos en la medida en que se cuente con la autorización previa y expresa por escrito de la Universidad de los Andes.

De igual manera, la utilización de la imagen de las personas, docentes o estudiantes, sin su previa autorización está expresamente prohibida. En caso de incumplirse con lo mencionado, se procederá de conformidad con los reglamentos y políticas de la universidad, sin perjuicio de las demás acciones legales aplicables.
