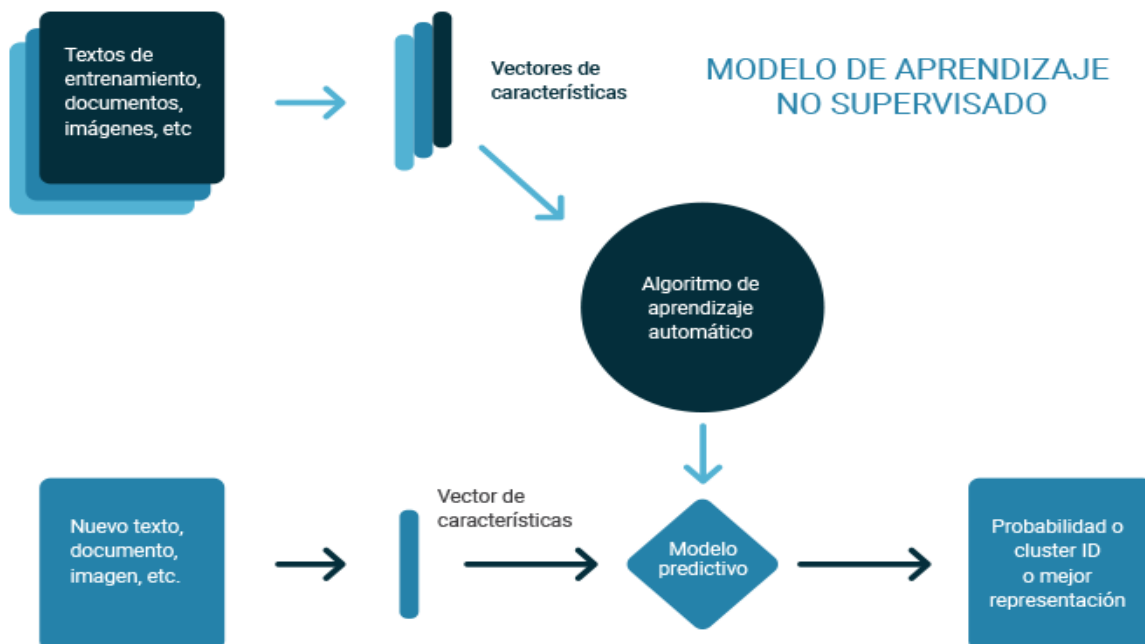


Lección 7: Aprendizaje no supervisado: Agrupamiento y reglas de asociación

Dentro del aprendizaje automático, además del aprendizaje supervisado, se encuentra también el **aprendizaje no supervisado**. En este caso, los algoritmos trabajan de forma parecida a los supervisados, con diferencia de que estos solo ajustan su modelo predictivo a través de la toma de datos de entrada, dejando de lado los datos de salida. En otras palabras, a diferencia del supervisado, los datos correspondientes a la entrada no se encuentran clasificados ni etiquetados, por lo que no son necesarias estas características para entrenar el modelo.

Si no se cuenta con un conjunto de entrenamiento, es decir, no existe comprensión acerca de las etiquetas correspondientes a los patrones, se pueden entonces clasificar objetos a través de un proceso previo de análisis de los datos conocido como clasificación no supervisada, aprendizaje no supervisado, **técnicas de agrupamiento** o clustering, que permite conocer la estructura de los datos.

deseables.



Fuente: Medium

Dentro de esta variedad de algoritmo, el agrupamiento o clustering, es el más utilizado, puesto que divide los datos en grupos que tengan características similares entre sí. El agrupamiento busca clasificar un conjunto de objetos en grupos, de tal manera que los objetos que se encuentren dentro de cada grupo sean altamente semejantes. Teniendo en cuenta la gran cantidad de algoritmos de agrupamiento que existen actualmente, es importante que estos logren satisfacer ciertas propiedades.

El agrupamiento resulta ser una herramienta bastante utilizada en diversos contextos como:

- La recuperación de información y minería de textos
- Procesamiento de secuencias descriptoras de genes y proteínas
- Seguimiento y detección de sucesos en flujo continuo de noticias
- Segmentación de imágenes
- Comprensión de datos
- Procesamiento de bases de datos especiales
- Clasificación de zonas geográficas
- Comprensión de imágenes de satélites
- Visualización de datos
- Prospección geológica
- Organización de documentos en bibliotecas
- Otras aplicaciones a grandes volúmenes de datos

A raíz de la gran variedad de algoritmos de agrupamiento, se han generado diferentes clasificaciones de los mismos, pero teniendo en cuenta diversos aspectos.

❖ Según la forma en que procesan los objetos:

- ✓ Estáticos
- ✓ Incrementales
- ✓ Dinámicos

- ❖ Según la forma en que organizan los grupos que se van obteniendo:
 - ✓ Jerárquicos
 - ✓ Particionales
 - ✓ Basados en densidad
 - ✓ Híbridos

- ❖ Según la pertenencia de los objetos a los grupos:
 - ✓ Disjuntos
 - ✓ Solapados

- ❖ Según el mecanismo en que se basan para agrupar:
 - ✓ De pasada simple
 - ✓ Basados en grafos
 - ✓ De optimización
 - ✓ Basados en arboles

Ligado el tema de aprendizaje no supervisado, **las reglas de asociación** son utilizadas para aportar conocimiento que ayude de alguna manera a la toma de decisiones. Claro ejemplo de la utilidad de este tipo de metodologías es cuando se quieren saber las tendencias de compra de los clientes, tareas de predicción, deducir los estudios de una persona en función de su salario, minería de textos para asociar la presencia de términos en documentos, entre otros. Esta clase de modelos son usados cuando la interpretabilidad de un modelo prima, y son modelos predictivos. Es usual utilizarlo en conjuntos difusos, y, a diferencia de los árboles que parten en el espacio, las reglas cobijan parte del espacio, logrando disponer de un grado de cubrimiento y un acierto.

Las reglas de asociación funcionan como una técnica frecuentemente usada para extraer conocimiento tomando como punto de partida bases de datos grandes. Estas logran identificar relaciones existentes entre datos, que son llamados ítems.

Una regla se define como:

$X \rightarrow Y$

Donde, X e Y, son conjuntos de ítems distintos.

Por ejemplo: Pan \rightarrow Mantequilla cuando se compra pan, se compra mantequilla. Originalmente, las reglas de asociación fueron aplicadas a datos de supermercados, puesto que a estos les interesa conocer cuáles son los productos que compran con más frecuencia los clientes, para ponerlos unos cerca de los otros y mediante esta actividad, reducir el tiempo de compra del cliente. En el caso particular del supermercado, los artículos serán entonces los ítems, y el conjunto de cestas de la compra, serán las transacciones. Cada transacción es un subconjunto de ítems que se denomina itemset. Por ejemplo: Leche y pan \rightarrow Mantequilla.

Antes de aplicar reglas de asociación, es necesario determinar cuáles son los datos de la base que corresponden a ítems, cuales a transacciones. Entonces, los ítems, son los elementos por asociar (pan, aceite, mantequilla, leche...) y las transacciones son las particularidades de la relación dada entre ítems (lista de compra específica). Los ítems pueden ser de diversos tipos, cuando cada registro consiste en un listado de elementos, no existen variables, un ítem se corresponde con un producto. Pero cuando existen variables con rangos, el ítem es un par (atributo, valor) como por ejemplo una variable puesto, con valores estudiante, jefe, trabajador, debería poseer ítems como puesto-estudiante, puesto-jefe y puesto-trabajador.

Adicional al aprendizaje supervisado y aprendizaje no supervisado, existe una tercera categoría, que, aunque en muchos casos no es tomada en cuenta, es oportuno mencionar su utilidad y objetivo. **El aprendizaje por refuerzo** funciona a través de algoritmos por refuerzo, que permiten definir modelos y funciones encaminadas a maximizar una medida de recompensas basadas en acciones, y así mismo, el ambiente en el que el agente inteligente se desempeñará.

Estos algoritmos son más cercanos a la psicología conductual de los seres humanos, puesto que es un modelo que consiste en acción-recompensa, y busca que el algoritmo se ajuste a la mejor opción de recompensa existente dada por el ambiente, y que sus acciones por tomar están arraigadas a dichas recompensas. Estas metodologías pueden ser utilizadas para el aprendizaje de los robots en un sinnúmero de tareas.

MODELO DE APRENDIZAJE POR REFUERZO



Fuente: Medium

Cada enfoque posee ventajas en ciertas áreas de oportunidad. La importancia radica en saber explorar con certeza la naturaleza del problema, para de esa manera poder utilizar el algoritmo según el tipo de aprendizaje más conveniente según sea el caso.

Referencias:

- Ecured. (s.f.). Ecured. Recuperado el 7 de octubre de 2018, de https://www.ecured.cu/Algoritmos_de_clasificaci%C3%B3n_no_supervisada
- Luna González, J. (s.f.). Medium. Recuperado el 7 de octubre de 2018, de <https://medium.com/soldai/tipos-de-aprendizaje-autom%C3%A1tico-6413e3c615e2>
- Alcalde, A. 2018. Aprendizaje no supervisado y detección de anomalías: reglas de asociación. Recuperado el 7 de octubre de 2018, de <https://elbauldelprogramador.com/aprendizaje-nosupervisado-reglas/>