

20/06/2018

Árboles de decisión para análisis de datos sobre calidad de vida



Investigadors de l'IIIA i la UAB han utilitzat l'anàlisi d'arbres de decisió sobre dues bases de dades de naturalesa pràctica: una amb descripcions de pigues que poden ser o no melanomes; i una altra referent a la qualitat de vida de persones amb discapacitat intel·lectual.

istockphoto/nevarpp

Investigadores del IIIA y de la UAB han utilizado el análisis de árboles de decisión sobre dos bases de datos de naturaleza práctica: una con descripciones de lunares que pueden ser o no melanomas; y otra referente a la calidad de vida de personas con discapacidad intelectual.

Actualmente, con el uso de nuevas tecnologías hay una gran oportunidad de recopilar datos. Sin embargo el análisis de estos datos es difícil de hacer de forma manual y es aquí donde son útiles las herramientas de descubrimiento de conocimiento (knowledge discovery en inglés). Estas herramientas permiten obtener patrones satisfechos por subconjuntos de datos, por lo que dan una caracterización de las regularidades de este subconjunto.

Una manera de obtener estos patrones es utilizando árboles de decisión (decision trees) que es un método de aprendizaje inductivo. Normalmente, el objetivo de construir un árbol de decisión es el de clasificar nuevos objetos, pero nosotros proponemos utilizarlos para analizar los datos. Básicamente, un árbol de decisión hace una partición de los datos iniciales hasta conseguir que cada subconjunto de la partición tenga sólo ejemplos de una de las clases. La forma que tenga

el árbol de decisión nos puede dar una idea de cómo es la base de datos o de si nos faltan ejemplos de alguna de las clases.

Hemos utilizado esta forma de análisis sobre dos bases de datos: una que consta de descripciones de lunares que pueden ser o no melanomas; y otra donde hay descripciones de diferentes dimensiones referentes a la calidad de vida de personas con discapacidad intelectual.

En la base de melanomas hemos visto que hay algunas partes del dominio donde falta información, ya que, basándonos en los atributos que los expertos consideran relevantes, el árbol da overfitting y es muy profundo. En este caso, pues, es necesaria una revisión de cómo se ha hecho la descripción del dominio.

En la base referente a la calidad de vida de las personas con discapacidad intelectual la situación es diferente. Los datos han sido obtenidos a partir de la puntuación de unas encuestas por parte de educadores sociales. Esta puntuación ha sido discretizada y, por los resultados obtenidos parece que los intervalos de discretización no han sido los adecuados. En este caso la técnica nos ha servido para ver que hay un análisis en profundidad de los datos y de cómo interpretarlas para poderlas discretizar.

Así, un árbol de decisión de poca profundidad y poca anchura, significa que las clases están bien representadas y que pueden separarse bien porque sus características son diferentes. El hecho de que tenga poca profundidad significa que con pocos atributos podemos ser capaces de decir a qué clase pertenece un objeto. Por otro lado, un árbol muy profundo significa que para discriminar bien entre las clases son necesarios muchos atributos, habrá pues tener muchos ejemplos para poder hacer una buena caracterización. Si las hojas del árbol tienen pocos elementos (1 o 2) significa que hay objetos muy similares que pertenecen a clases diferentes. Esto puede ser debido o a un error en la base (en describir alguno de estos objetos) o bien que con la descripción que hemos elegido para describir los objetos del dominio no podemos separar bien las clases.

Nuestro trabajo ha sido motivado por el trabajo interdisciplinar que estamos haciendo con profesionales de la educación social y la medicina. El análisis de datos utilizando inteligencia artificial debe tener en cuenta siempre que el último objetivo es la mejora efectiva de la calidad de vida de las personas. Este hecho fortalece la investigación, donde nuevos problemas tanto de naturaleza práctica como teórica surgen de esta interacción.

Pilar Dellunde

Artificial Intelligence Research Institute (IIIA-CSIC)
Philosophy Department
Universitat Autònoma de Barcelona

Eva Armengol

Artificial Intelligence Research Institute (IIIA-CSIC)

Àngel García-Cerdà

Artificial Intelligence Research Institute (IIIA-CSIC)
Universitat Pompeu Fabra

Referencias

Armengol E., García-Cerdàña À., Dellunde P. (2017) **Experiences Using Decision Trees for Knowledge Discovery**. In: Torra V., Dahlbom A., Narukawa Y. (eds) *Fuzzy Sets, Rough Sets, Multisets and Clustering*. Studies in Computational Intelligence, vol 671. Springer, Cham.
https://doi.org/10.1007/978-3-319-47557-8_11

[View low-bandwidth version](#)