

20/06/2018

Arbres de decisió per a anàlisis de dades sobre qualitat 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

Actualment, amb l'ús de noves tecnologies hi ha una gran oportunitat de recopilar dades. Tanmateix l'anàlisi d'aquestes dades és difícil de fer de forma manual i és aquí on són útils les eines de descobriment de coneixement (*knowledge discovery* en anglès). Aquestes eines permeten obtenir patrons satisfets per subconjunts de dades, de manera que donen una caracterització de les regularitats d'aquest subconjunt.

Una manera d'obtenir aquests patrons és fent servir arbres de decisió (*decision trees*) que és un mètode d'aprenentatge inductiu. Normalment, l'objectiu de construir un arbre de decisió és el de classificar nous objectes, però nosaltres proposem utilitzar-los per analitzar les dades. Bàsicament, un arbre de decisió fa una partició de les dades inicials fins a aconseguir que cada subconjunt de la partició tingui només exemples d'una de les classes. La forma que tingui l'arbre de decisió ens pot donar una idea de com és la base de dades o de si ens falten exemples d'alguna de les classes.

Hem utilitzat aquesta forma d'anàlisi sobre dues bases de dades: una que consta de descripcions de pigues que poden ser o no melanomes; i una altra on hi ha descripcions de diferents dimensions referents a la qualitat de vida de persones amb discapacitat intel·lectual.

En la base de melanomes hem vist que hi ha algunes parts del domini on falta informació, ja que, basant-nos en els atributs que els experts consideren rellevants, l'arbre dona *overfitting* i és molt profund. En aquest cas, doncs, cal una revisió de com s'ha fet la descripció del domini.

En la base referent a la qualitat de vida de les persones amb discapacitat intel·lectual la situació és diferent. Les dades han estat obtingudes a partir de la puntuació d'unes enquestes per part d'educadors socials. Aquesta puntuació ha estat discretitzada i, pels resultats obtinguts sembla que els intervals de discretització no han estat els adients. En aquest cas la tècnica ens ha servit per veure que cal una anàlisi en profunditat de les dades i de com interpretar-les per a poder-les discretitzar.

Així, un arbre de decisió de poca profunditat i poca amplada, implica que les classes són ben representades i que poden separar-se bé perquè les seves característiques són diferents. El fet que tingui poca profunditat vol dir que amb pocs atributs podem ser capaços de dir a quina classe pertany un objecte. D'altra banda, un arbre molt profund vol dir que per discriminar bé entre les classes són necessaris molts atributs, caldrà doncs tenir molts exemples per a poder fer una bona caracterització. Si les fulles de l'arbre tenen pocs elements (1 o 2) vol dir que hi ha objectes molt similars que pertanyen a classes diferents. Això pot ser degut o a un error en la base (en descriure algú d'aquests objectes) o bé que amb la descripció que hem triat per descriure els objectes del domini no podem separar bé les classes.

El nostre treball ha estat motivat pel treball interdisciplinar que estem fent amb professionals de l'educació social i la medicina. L'anàlisi de dades fent servir intel·ligència artificial ha de tenir en compte sempre que el darrer objectiu és la millora efectiva de la qualitat de vida de les persones. Aquest fet enforteix la recerca, on nous problemes tant de naturalesa pràctica com teòrica sorgeixen d'aquesta interacció.

Pilar Dellunde

Institut d'Investigació en Intel·ligència Artificial (IIIA-CSIC)
Departament de Filosofia
Universitat Autònoma de Barcelona

Eva Armengol

Institut d'Investigació en Intel·ligència Artificial (IIIA-CSIC)

Àngel García-Cerdaña

Institut d'Investigació en Intel·ligència Artificial (IIIA-CSIC)
Universitat Pompeu Fabra (UPF)

Referències

[View low-bandwidth version](#)