

Olatz Arregi Uriarte

Donostiako Informatika Fakultateko irakasle eta ikertzailea

Datu-meategien altxorra

Meatzaritza: Lurretik mea erauzteko jarduerak. Erauzteko lan hutssez gain, esplorazio eta bereizte-lanak ere meatzaritzaren alorrekoak dira. [ZT Hiztegia].

Lurretik mea erauzten den bezalaxe, datuetatik ezagutza edo informazioa erauztea da datu-meatzaritzaren xedea. Erauzitako informazioa egituratu egin behar da, forma eman behar zaio erabilgarri izango bada.

Datu-meatzaritzak datu multzo handien analisi automatikoa edo erdi-automatikoa egiten du, datu horietatik aurrez ezezagunak diren portaerak edo ezagutza ateratzeko. Baina, nola lor daiteke informazioa datu multzo gordinetik? Kasu bakoitza ezberdin tratatu arren, oro har, lau dira jarraitu beharreko urratsak: erabaki zer informazio atera nahi den datuetatik; datuak garbitu eta prestatu; programa bat erabili ezagutza eta informazioa erauzteko; eta, azkenik, emaitzak aztertu. Batzuen ustetan, hirugarren urratsa da datu-meatzaritza benetan, eta besteak datu-base-etako ezagutzaren aurkikuntza (Knowledge Discovery in Databases) izeneko prozesu zabalago baten barne-urratsak dira.

Meategietan mea ateratzeko tresna ezberdinak erabiltzen diren bezalaxe, arlo honetan ere oinarri oso ezberdineko teknikak erabil daitezke datuetatik informazioa eskuratzeko. Estatistika, konputazio grafiko edota adimen artifizialaren alorretatik elikatzen da datu-meatzaritza, eta, oinarrian, datuak prest daudenean, ezagutza ateratzeko ikasketa automatikoko programak erabili ohi dira. Programa horiek izaera ezberdina izan dezakete, hala nola probabilitateetan oinarritzen direnak, neuronen funtzionamendua imitatzen saiatzen diren neurona-sareak, edota erabaki-erregelak definitzen dituztenak, besteak beste.

Hainbat eta hainbat esparrutan aplikatu izan da eta aplikatzen da datu-meatzaritza: zientzia arloan, medikuntzaren domeinuan, enpresa munduan, komunikazioetan... baina ikus ditzagun adibide simple batzuk hobeto ulertzeko.

ENPRESA

Kreditu-txartela erabiltzen dugun bakoitzean, gure informazioa datu-base-ren batean biltzen da, eta horrela, informazio hori eskura duenak, datuak ustiatuz, gure gustuak, beharrak edota nahiak jakin ditzake. AEBetako Walmart multinazionalak, adibidez, 20 milioi salerosketaren informazioa jasotzen du egunero datu-base batean. Informazio horrek guztiak ez luke ezertarako balio, ez balego software bat, hortik abiatuta, marketin-kanpainak diseinatzeko irizpideak edota bezeroen leialtasunaren informazioa emango lukeena.

Espainiako telefono-konpainia batek jakin nahi izan zuen zerk eramaten zituen bezeroak konpainiaz aldatzera. Horretarako, batetik, bezeroen datu pertsonalak aztertu zituzten, adina, sexua, egoera zibila... eta, bestetik, bezero horiek zerbitzua noiz eta nola erabiltzen zuten, zein ordutan egiten zituzten deiak, nora, zein maiztasunarekin... eta datu horietatik atera zuten ondorioa izan zen baja ematen zuten bezeroek, besteekin alderatuta, promozio eta deskontu gutxi eta intzidentzia gehiago izaten zituztela.

Enpresa baten giza-baliabideen sailak, datu-meatzaritzari esker, jakin dezake

zein ezaugarri dituzten bere langile finenek, non ikasi duten, zein zaletasun dituzten... eta horrek lagun diezaioke langile berriak kontratatzeke orduan.

KOMUNIKAZIOA

Telebistetako audientzien inguruan iragarpenak egiteko ere erabili izan da datu-meatzaritza. Hain ezaguna den British Broadcasting Corporation-ek (BBC) sistema bat erabiltzen du programa baten iraupen egokia erabakitzeke urteetan bildutako datuak erabiliz.

MEDIKUNTZA

Gero eta informazio mediko gehiago dago bilduta datu-baseetan, eta horri esker, tresna ezberdinak garatu dira arlo horretan, gaixotasunen tratamenduetan laguntzeko. Oracle Data Miner (ODM) softwarea adibidez, diabetesaren tratamenduan laguntzeko erabiltzen da. Gaur egun gaixoeekin probatzen diren tratamenduetatik, botikak, dieta, intsulina, ariketa fisikoa... zein da egokiena paziente batentzat? ODM tresnak lagun dezake ditugun datuetatik portiera-patroiak ateratzen, eta, ondorioz, iragarpenak egiteko balioko digu, besteak beste, paziente baten aurrean, zein den tratamendu egokiena aholkatuko digu.

KIROLA

Luze jarrai dezakegu adibideak jartzen, baina, bukatzeko, kirol munduko adibide bat aipatuko dugu. IBMn garatutako Advanced Scout softwarea erabili da NBAko taldeen jokoa aztertzeke. Tresnak, partidetakoe errebotek, paseak, kanastak... dena gordetzen du. Honi esker, New Yorkeko Knicks taldearen entrenatzaileek markaje bikoitzetan taldeak zuen gabezia bat antzeman zuten eta arazoari aurre egin ahal izan zioten.

Argi dago gero eta datu gehiago gordetzeko ahalmena dugula, eta ahalmen hau, gainera, urtetik urtera hazten doa, baina datu horiek prozesatzea eta informazio baliagarria ateratzea ez da lan erraza. Gero eta programa gehiago dago lan honetan lagungarri bilakatu dena, baina asko dago oraindik egiteko. Datu-meategiek dituzten altxorrak, oraindik ere, ezkutuan daude. •



Kreditu txarteletan informazio asko metatzen da.