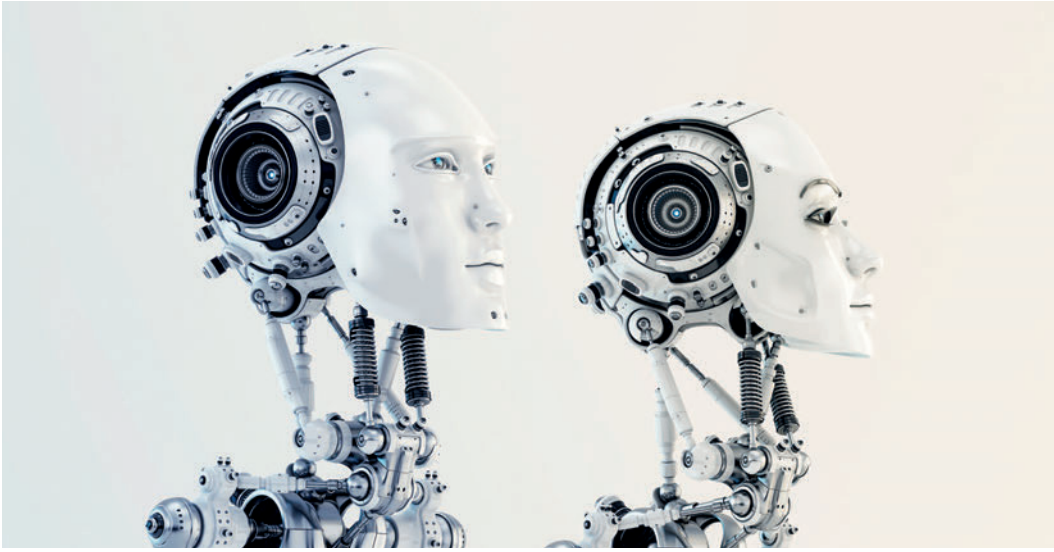
Generoak bereizteko joera adimen artifizialean

— *1. irudian bi robot adimendun daude: mutil ala neska? Zein da zein?*

— *Baina, zertaz ari gara? Nola da posible adimen artifizialak generoak bereizteko joera izatea? Makinek ez dute generorik, ez da hala?"*



1. irudia: Neska-mutil robotikoak. ARG.: Ociacia/Shutterstock.

Halaxe da. Makinek duten adimena da adimen artifiziala (AA) edo makina-adimena. Ez da berria; izan ere, urte asko joan dira 1956. urtetik hona, Samuelek, Simmonsek, Newellek, MacArthyk eta Minskyk adimen artifiziala ikerkuntza-lerro izendatu zutenetik. Orduz geroztik, teknologia ikaragarri aurreratu da. Lengoaia naturala ulertzea, estrategia konplexuak behar diren jokoetan arrakastaz aritzea (xakea, Go jokoa), gidatzeko sistema autonomoak garatzea eta puntu anitzeko banaketa-sareetan bideak modu optimoan planifikatzea dira, besteak beste, AAri esker posible izan diren aurrerapenak. Horretarako, metodo estatistikoak eta formalismo sinbolikoak erabiltzen dira, besteak beste. Izan ere, ikasketa automatikoak sostengatzen du gaur egungo AAren iraultza. Datu-base marduletatik ezagutza eratortzen duten metodoak dira ikasketa horiek, eta konpainia zein gobernuetako agentziek, ospitaleek eta bestelako enpresek [2] erabiltzen dituzte askotariko aplikazioetarako. Esaterako, bideo-sekuentzietan objektuak identifikatzeko, mailegu-eskaeren sinesgarritasuna balioztatzeko, legezko kontratuetan akatsak bilatzeko edo minbizia duen pertsona batentzat tratamendu egokiena topatzeko [1] [2].

Olatz Arbelaitz Gallego
EHUko Informatika Fakultateko
irakaslea eta UEUko kidea

Elena Lazkano Ortega
EHUko Informatika
Fakultateko irakaslea

AA gure egunerokotasunean dago jada, mugikorretan daramatzagun aplikazio anitzek adimen artifizialeko teknikak (soluzioak) dituzte oinarrian, eta normaltasunez erabiltzen ditugu. Adibidez, gero eta arruntagoa da ahozko komunikazioa erabiltzea telefono eta ordenagailuekin komunikatzeko: Amazonek Alexa eskaintzen du; Googlek, Home; Applek, berriz, Siri, eta Microsoftek, Cortana. Ahots-zerbitzariak dira horiek, gure ahotsa ezagutu, prozesatu eta egindako galderaren edo eskaeraren erantzuna, Interneten bilaketa egin ondoren, ahots sintetizatu bidez helarazten diguten tresnak.

Jakin badakigu, urteetako joerak direla medio, jendartea bera ez dela neutroa, eta generoak (eta bestelakoak) bereizteko joerak nabarmenak direla, askotan. Adibidez, zaintzaren rola emakumeei atxikitzen zaie, eta indarrarekin zerikusia duten rolak, aldiz, gizonezkoei. Baina zer gertatzen da artifizialki sortutako adimenarekin? Adimen artifizialak eta ikasketa automatikoak joerarik badute, joera horiek etorkizuneko tresnetan integratzeko arriskua egongo da. Izan ere, dagoeneko gertatzen ari da.

Datuak arazo-iturri

Ikasketa automatikoko algoritmoek iragarpenak egiten dituzte, ikasi dutenaren arabera. Algoritmoek berek ez dute joerarik, ez genero aldetik, ez arraza aldetik, ez bestelakorik. Baina aipatu moduan, datuetan, matematiketan eta estatistikan oinarritzen dira. Datuetan dago joera, jendartearen isla diren eta guk hartutako erabakien ondorio diren datuetan. Pertsonok hautagaiak kontratatzeko erabiltzen ditugun irizpideetan, ikasleen txostenak ebaluatzean, mediku-diagnosiak egitean, objektuak deskribatzean, horietan guztietan ditugu kultura, arraza, hezkuntza zein bestelakoak bereizteko joerak [5], eta haien artean, noski, generoak bereizteko joerak. Laburbilduz, sortzen diren ereduak entrenamendurako erabilitako datuak bezain onak edo neutroak izango dira kasurik onenean. Guk jasotzen eta etiketatzen ditugu datuak; gure subjektibotasuna islatzen dugu datuetan, eta horrek joerak eragiten ditu ikasketa automatiko bidez garatutako sistemetan. Are gehiago, adimen artifizialak estereotipoak anplifikatzen ditu! [2]. Izan ere, ikasketa automatikoko algoritmoak erabiliz sortutako eredurik egokiena hautatzen dugunean, datuei ongien egokitzen zaiena edo asmatzetasarik handiena duena hautatu ohi dugu, bestelako irizpideak alde batera utziz.

Arazoa larria da, batez ere makinek proposatutakoa erabakiak hartzeko erabiltzen den kasuetan; gainera, maiz gertatzen da hori eguneroko bizitzan, adimen artifizialean oinarritutako tresnak oso zabalduta baitaude [4].

Generoak bereizteko joera erakusten duten zenbait adibide

Hamabost bat urte atzera egiten badugu, hizketa ezagutzeko lehenengo garapenak autoetan ezarri zirenean, sistemek ez zituzten ulertzen emakumezkoen ahotsak, gizonezkoen ahotsekin entrenatu baitziren. Azken urteetan horrelako sistemak ikaragarri aurreratu diren arren, eta emakume zein gizonen ahotsak ezagutzen dituzten arren, generoaren joera nabarmena da oraindik. Aipatu berri ditugun ahots-zerbitzarien harira, lautik hiruk emakume-izena dute, eta, areago, guztiek emakume-ahotsa dute lehenetsia. Nahiz eta gaur egun aldatzeko aukera izan, zerbitzari-lanak egiten dituzten obeditzaileak femeninoak izatea dago lehenetsita.

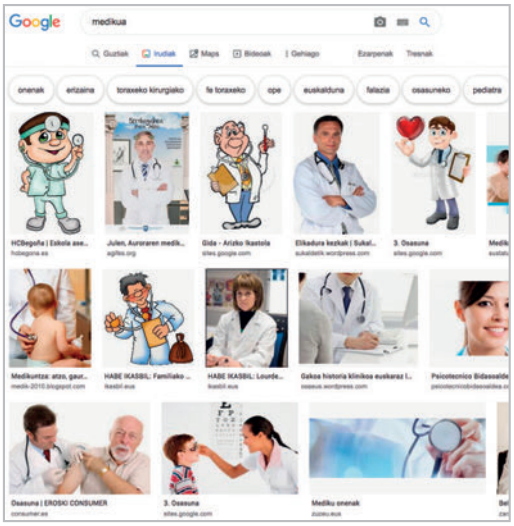
Zergatik bada? Arrazoia sinplea da, eginkizun administratiboetan eta zaintza edo zerbitzuetan emakumezkoak espero ditugulako. Inkesta bat eginez gero, ez litzateke harritzekoa gehiengoak emakume ahotsa lehenestea.

Generoak bereizteko joera ez da ahotsarekin lotutako testuinguruetan soilik agertzen. Lengoaia naturalaren prozesamenduan, adibidez, testuen

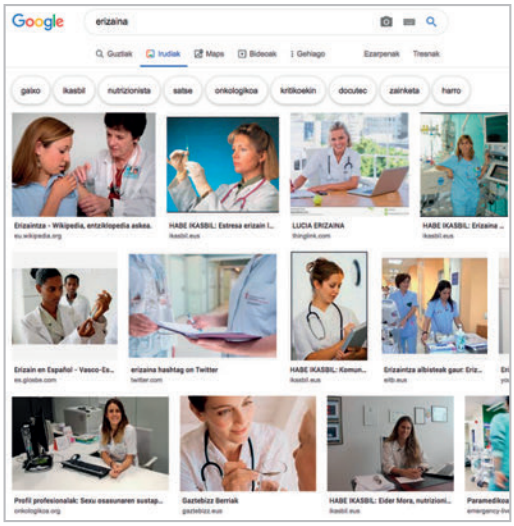
analisi semantikoak egiteko, testu-multzo handietatik erauzten diren eta hitzak elkarren artean erlazionatzen dituzten hiztegi mardulak erabiltzen dira, *word-embedding* izenekoak. Bostongo unibertsitateko ikertzaileek egindako azterketa batek [2] azaleratu duenez, hiztegi horiek erabiliz "John izenekoek Mary izenekoek baino programa hobeak idazten dituzte" erako inferentziak gerta daitezke. Haiek sortzeko erabilitako testuen ondorioa da hori, noski.

Irudia nagusitzen ari den gizarte honetan, artifizialki ikasten duten sistemak irudiz elikatzen dira, gero eta gehiago. ImSitu (http://imsitu.org/) eta COCO (http://cocodataset.org/) datu-multzo ezagunetan, adibidez, saretik jasotako 100.000 irudi baino gehiago daude, zeinetan eszena konplexuak dauden azalpenekin etiketatuta [3]. Batean zein bestean, gehiago dira gizonezkoen irudiak dituztenak, eta egiaztatu da joera nabarmena dutela objektu zein ekintzen etiketek generoak bereizteko. COCO datu-multzoan, sukaldeko tresnak (koilarak eta sardexkak) emakumeekin lotuta deskribatzen dira, eta aire libreko kiroletarako ekipamendua (elur-oholak eta teniseko erraketak), aldiz, gizonezkoekin lotzen dira. Gauza bera gertatzen da ImageNet datu-basean, zeina 2002-2004 bitartean sortu zen Yahoo News-eko irudiekin, eta oraindik arras erabilia den. Irudien % 78an gizonezkoak azaltzen dira, eta horietako % 84an azal zurikoak; izan ere, George W. Bush da gehien agertzen den pertsona. Datu-multzo horiek erabiltzen direnez irudiak ezagutzeko softwareak entrenatzeko, softwareek joera heredatzen dute. Egin proba, bilatu "medikua" Googleko irudietan eta mantal zuriz jantzitako gizonezkoz beteko zaizu pantaila. Aldiz, "erizaina" bilatuz gero, emakumezkoak nabarmenduko zaizkizu (ikus 2. irudia).

A



B



2. irudia: Googlen *medikua* (a) eta *erizaina* (b) hitzak bilatuta lortutako irudiak.

Joera desagerrarazteko bideak

Adimen artifizialeko sistemen proposamenak fidagarri izateko, proposamenak objektiboak direla ziurtatu beharko genuke lehenengo. Algoritmoek ez dute kontzientziarik, eta, hortaz, ezin dituzte haien proposamenak aldatu.

Horri aurre egiteko modua ez da bakarra; fronte askotatik hel dakioke eta heldu behar zaio.

Alde batetik, ezinbestekoa izango da emakumezkoek parte-hartze aktiboa izatea teknologiaren garapenaren urrats guztietan. Estatistikek diotenaren arabera, aldiz, informatika-teknologiako lanpostuetan % 50a izatetik gero eta urrunago daude emakumeak. Aldaketak lagunduko duen arren, horrek soilik ez du ziurtatzen sortuko diren sistemetan generoak bereizteko joera desagertuko denik. Horren adibide da AVA (Autodesk Virtual Agent), mundu zabalean Autodesk produktuen bezeroen arreta-zerbitzua eskaintzen duen software agentea. Adimen artifiziala eta adimen emozionala konbinatzen ditu, bezeroarekin "aurrez aurre" hitz egiteko. Gehiengoa emakumezkoa duen lantalde batek sortu du AVA, baina, hala ere, emakumeahotsa eta -itxura du. Aitzakia, berriro ere, berbera da: pertsonek lagungarriago eta kolaboratzaileago sentitzen ditugula emakumezkoen ahotsak. Adibide horrek argi uzten du, beraz, ez dela nahikoa emakumeak lantaldeetan sartzea.

Ikasketa automatikoko sistemen joerak desagerrarazteko bide bat da entrenamendurako erabiltzen diren datu-baseak sakon araztea eta zuzentzea. Badira horretan diharduten ikertzaileak eta testuetatik legitimoak ez diren erlazioak ezabatu dituztenak, adibidez, gizon/ordenagailu eta emakume/etxetresna [6]. Beste ikertzaile batzuek, aldiz, proposatzen dute sailkatzaile berezituak sortzea datu-multzo batean adierazitako talde bakoitza sailkatzeko [7].

Argi dago, beraz, denok hartu behar dugula kontuan datuek joera jakin bat baldin badute joera bera izango dutela haietatik ahalegin berezirik egin gabe ikasten duten sistemek ere, eta horrek, gainera, ondorio larriak izan ditzakeela gugan. Joerak detektatu eta identifikatzeko tresnak beharko ditugu, baita datu-multzo berriak sortzeko printzipioak ere. Baliteke adituak behar izatea zeregin horretarako. Eta zergatik ez garatu adimen artifizialeko tresnak zeregin horretarako! ●

Erreferentziak

[1] The Risk of Machine-Learning Bias (and How to Prevent It) FrontiersBlog March, 2018 Chris DeBrusk https://sloanreview.mit.edu/article/the-risk-of-machine-learning-bias-and-how-to-prevent-it/

[2] Researchers Combat Gender and Racial Bias in Artificial Intelligence, Bloomberg, December 2017. Dina Bass and Ellen Huet https://www.bloomberg.com/news/articles/2017-12-04/researchers-combat-gender-and-racial-bias-in-artificial-intelligence

[3] Machines Learn a Biased View of Women . Wired. https://www.wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women/

[4] Artificial intelligence could hardwire sexism into our future. Unless we stop it.. World Economic forum. December 2017. Alison Kay .https://www.weforum.org/agenda/2017/12/sexist-bias-hardwired-by-artificial-intelligence/

[5] Bias in machine learning, and how to stop it. Tech Republic. November 2016. Hope Reese. https://www.techrepublic.com/article/bias-in-machine-learning-and-how-to-stop-it/

[6] T. Bolukbasi, K-W Chang, J. Zou, V. Saligrama, A. Kalai, Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain

[7] C. Dwork, N. Immorlica, A. T. Kalai, M. DM Leiserson. Decoupled classifiers for fair and efficient machine learning. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR 81:119-133, 2018.