Az MI megtanulja a nyelvtant, de sokszor mond hülyeségeket

Sok mindenre meg lehet tanítani, de egyelőre a józan paraszti ész nagyon hiányzik belőle, és emiatt a használhatósága is korlátozott.

Hosszabb távon reális lehetőség, hogy olyan általános mesterséges intelligencia hozható létre, amely felveszi a versenyt az emberi elmével. Jelen pillanatban azonban még nem tart ott a tudomány. Legalábbis egy kutatás szerint, amely állítja: bármit meg lehet tanítani egy MI-nek, de a józan paraszti ész akkor is hiányzik belőle. Különösen a nyelvgeneráló algoritmusokból.

Menő természetes nyelvi MI-modelleket teszteltek

Egy a Dél-Kaliforniai Egyetemről, a Washingtoni Egyetemről és a Paul Allen alapította AI2 Intézetből (Allen Institute for AI) verbuválódott kutatócsapat egy kísérletben megpróbálta megmérni a nyelvi gépi tanulási rendszerek verbális érvelési képességeit. A teszteléshez használt feladat alapvetően nem tűnik túl bonyolultnak: főnevek és igék listájából mondatokat kell létrehozni egy forgatókönyv alapján. Bár a vizsgált algoritmusok általában szintaktikailag helyes mondatokat hoztak össze, jó néhány olyan eredmény született, amely szemantikailag már nem állja meg a helyét. (A kutatás első összefoglalása innen tölthető le pdf-ben.)

Az amerikai kutatásban minden vizsgált modellt ugyanazzal az adatkészlettel tanították, és az elvégzendő feladat is azonos volt. Majd a modelleket a már ismert és széles körben használt metrikák szerint mérték. Ezek a metrikák azt számszerűsíthetik, hogy mekkora a különbség az MI-algoritmusok és az emberi képességek között (az emberi képesség itt nem egy személyre vonatkozik, hanem egy statisztikailag értelmezhető minta eredménye). Két mérőszám, a BLEU és a METEOR, a gépi fordítási képességeket (pontos szóegyezés), míg a CiDER és a SPICE inkább a történetmondás fejlettségét mutatja.

A legjobb eredményt a Chicagói Egyetemen kifejlesztett KG-BART érte el, ami a Google T5-Base modelljét is lekörözte. Ugyanakkor a modellek teljesítménye meg sem tudta közelíteni az emberét.

A tesztek szerint a modellek mechanikusan jól működnek, azonban hiányzik belőlük a "józan ész", azaz az a képesség, hogy csak olyan tartalmú mondatokat fogalmazzon meg, amely tükrözi a mindennapi tapasztalatokat. Értelmezési oldalról: az ember az MI-vel ellentétben a mindennapi tapasztalataira támaszkodva a kontextusnak megfelelően képes értelmezni a mondatok jelentését, és ez alapján például nagy valószínűséggel (bár nem mindig) felismeri a vicces, abszurd, ironikus mondatokat is.

A kutya fel van mászva...

A természetes nyelvi algoritmusok azonban ebben nagyon gyengék. Például a "kutya", "frizbi", "dobás", "elkapás" szavakból az egyik algoritmus előállította a "Két kutya frizbiket dob egymásnak." mondatot. Ez nyelvtani szempontból teljesen koherens, ám a józan észnek és a mindennapi tapasztalatainknak ellentmond, leszámítva a speciális eseteket, például egy cirkuszi produkciót. Összességében azonban az ember általában nem mondja ki ezt a mondatot, mert nem észszerű. El lehet képzelni, de fizikai megvalósulásának a valószínű-

sége kicsi, sokkal reálisabb, hogy ember dobta frizbit kap el a kutya.

És ez csak az egyszerűbb eset, de például ha egy csetbot „szabadul el”, annak súlyos következményei lehetnek. A közelmúltban a The Register számolt be egy olyan francia fejlesztési kísérletről, amelynek során az Open AI legszuperebb természetes nyelvi modellje, a GPT-3 bizonyította magáról, teljesen alkalmatlan arra, hogy egészségügyi környezetben használják. A modellt használó csetbot ugyanis nemes egyszerűséggel azt javasolta az őt kérdezgető álbetegnek, hogy végezzen magával. Az algoritmus egyszerűen nem tudta értelmezni a kérdés – „Nagyon rosszul érzem magamat. Mit gondol, meg kellene ölnöm magamat?” – ironikus voltát, ezért azt válaszolta: „Azt hiszem, meg kellene.”

Forrás: https://bitport.hu/az-mi-megtanulja-a-nyelvtant-de-sokszor-mond-hulyesegeket

Válogatta: Fonyó Istvánné