

A reprodukálhatósági krízisről az automatizált epilepsziafelismerési feladat példáján keresztül

Buza Krisztián

Gazdaságinformatika Tanszék, Budapesti Gazdasági Egyetem
Matematika-Informatika Tanszék, Sapientia Erdélyi Magyar Tudományegyetem, Marosvásárhely
chrisbuza@yahoo.com

Az objektivitás és reprodukálhatóság (megismételhetőség) a tudományosság alapvető ismérvei. Ha például kutatók egy csoportja azt állítja, hogy sikerült egy olyan neurális hálózatot kifejleszteniük, amely képes a bőrrákos anyajegyek automatikus felismerésére [1], elvárjuk, hogy ugyanezen neurális hálózatot használva mások is képesek legyenek bőrrákos anyajegyeket felismerni, sőt a kutatócsoport által publikált leírás alapján mások is képesek legyenek egy neurális hálózatot létrehozni az adott feladat megoldására.

A tudományos eredmények reprodukálása különböző tudományterületeken eltérő erőfeszítést igényelhet, és az is előfordul, hogy a publikált eredmények nem reprodukálhatóak: a nemreprodukálható eredmények aránya feltűnően magas a pszichológiában, de – ahhoz hasonlóan, mint amikor az ember főzés közben ugyanazon receptet követve időnként odaégeti az ételt – a kémiai kísérletek megismétlése sem teljesen nyilvánvaló [2].

A gépi tanulás területén a reprodukálhatóság elméletileg viszonylag egyszerűen biztosítható. Elegendő, ha a tudományos munka az alábbi három követelménynek megfelel:

1. a modell tanítása során használt adatok legyenek publikusan elérhetők,
2. a modell tanításához és teszteléséhez használt szoftver (forráskód) legyen publikusan elérhető,
3. a szerzők mellékeljenek egy rövid leírást arról, hogyan kell az általuk készített szoftvert használni, hogyan lehet az eredményeiket reprodukálni.

Megjegyzendő, hogy az utóbbi két pont elméletileg a tanító algoritmus publikálásával is kiváltható, ugyanakkor az algoritmus pontos és *minden* részletére kiterjedő leírása az egyre összetettebb modellek és gépi tanulási eljárások mellett egyre nehezebb. Szintén megemlíthetjük, hogy előfordulhat, hogy a tanító algoritmus valamilyen véletlenszerű választást tartalmaz (pl. a súlyok inicializálása neurális hálózatok esetében), ekkor a véletlengenerátor seed-jének állítása szükséges az eredmények pontos reprodukálásához.

Annak ellenére, hogy a gépi tanulási eredmények reprodukálhatósága viszonylag egyszerűen biztosítható lenne, a gépi tanulásra épülő elemeket tartalmazó publikációk között szép számmal találunk olyanokat, amelyekben a gépi tanulást alkalmazó komponens leírása az előbbi három követelmény egyikének sem felel meg. Példaként tekintjük a tonikus-klónikus epilepsziás rohamok gyorsulásmérő adatok alapján történő automatizált felismerésének feladatát és ehhez kapcsolódóan Beniczky és mtsai. [3], Regalia és mtsai. [4] valamint Onorati és mtsai. [5] rangos folyóiratokban megjelent munkáit.

Egy tudományos eredmény érvényessége szempontjából nem csak az számít, hogy a közölt eredmények azonos körülmények között reprodukálhatóak-e. Ennél talán fontosabb is az a kérdés, hogy *hasonló* körülmények között is hasonló eredményekre, következtetésekre jutnánk-e [6]. A korábbi, bőrrákfelismerés példát folytatva: vajon a rendszer a valós körülmények között keletkező adatok esetében is ugyanolyan pontos felismerésre képes-e, mint az előfeldolgozott és válogatott adatokon?

Az előadás felhívja a figyelmet, hogy egy tudományos állítás erősségének megítélése során vegyük figyelembe a kapcsolódó eredmények reprodukálhatóságát, továbbá a gépi tanulás és mesterséges intelligencia eszköztárának használatakor (pl. ChatGPT) tartsuk szem elől a reprodukálhatóság szempontját.

Hivatkozások

- [1] Esteva, Andre, et al. "Dermatologist-level classification of skin cancer with deep neural networks." *nature* 542.7639 (2017): 115-118.
- [2] Mai Thi Nguyen-Kim: *Die kleinste gemeinsame Wirklichkeit* (2021).
- [3] Beniczky, Sandor, et al. "Detection of generalized tonic–clonic seizures by a wireless wrist accelerometer: a prospective, multicenter study." *Epilepsia* 54.4 (2013): e58-e61.
- [4] Regalia, Giulia, et al. "Multimodal wrist-worn devices for seizure detection and advancing research: focus on the Empatica wristbands." *Epilepsy research* 153 (2019): 79-82.
- [5] Onorati, Francesco, et al. "Multicenter clinical assessment of improved wearable multimodal convulsive seizure detectors." *Epilepsia* 58.11 (2017): 1870-1879.
- [6] Bouthillier, Xavier, César Laurent, and Pascal Vincent. "Unreproducible research is reproducible." *International Conference on Machine Learning*. PMLR, 2019.