EMLÉKEZTETŐ

Az MTA Statisztikai és Jövőkutatási Tudományos Bizottság (SJTB) Statisztikai Tudományos Albizottsága (STAB) 2021. szeptember 16-án általános, nyílt ülést tartott, amelynek keretében

***Bin Yu*, a Berkeley Egyetem professzora, az amerikai Nemzeti Tudományos Akadémia tagja tartott előadást „Trustworthy AI through veridical data science and interpretable machine learning"** (Megbízható AI[1] a hiteles adattudományon és az értelmezhető gépi tanuláson keresztül) **címmel.**

A jelenlevők száma 42 fő, ebből STAB-tag 13 fő volt.

Az online rendezvényt *Kovács Péter*, a STAB elnöke nyitotta meg, köszöntve a meghívott előadót és a hallgatóságot. Majd *Bin Yu* munkásságának rövid bemutatása után átadta a szót a professzornak.

A prezentáció *Bill Gates* híres figyelmeztetésével indult: „Az AI olyan, mint az atomenergia: ígéretes és egyben veszélyes." Az AI-nak három pillére van: *1*. az adattudomány, *2*. a matematika és a statisztika, valamint *3*. egy adott szakterület magas szintű ismerete. Az adattudomány célja az adatok és egy-egy szakterület ismereteinek összekapcsolása algoritmusok használatával a döntéshozatal megkönnyítése és új ismeretek szerzése érdekében. A gépi tanulás a számítástechnika/számítástudomány, illetve a matematika és a statisztika közötti interfész. A megbízható AI-jal kapcsolatban két egymást kiegészítő megközelítés létezik: „legjobb gyakorlatok" megosztása az előrejelzés megbízhatóságának maximalizálására (elővigyázatosság) és kárelhárítás a veszélyek, kockázatok csökkentése érdekében (beavatkozás). A professzor és csapatának meghatározása szerint a hiteles adattudomány feladata, hogy megbízható és reprodukálható információkat vonjon ki az adatokból egy bővített, ember-ember, valamint ember-AI közötti kommunikációra alkalmas technikai nyelv segítségével empirikus tapasztalatok, ismeretek értékelése céljából.

A tudományos gépi tanulás, amely a megbízható AI részét képezi, és tudományos célokat szolgál, első lépésben az adatokból tudást generál, a másodikban pedig

[1] Mesterséges intelligencia (artificial intelligence, AI).

algoritmusaiban tudományos szabályszerűségeket/elméleteket határoz meg; és e két lépést sorozatosan ismétli. Az így kapott eredmények megfelelnek a tudományos elvárásoknak. A tudományos gépi tanulásra példaként a professzor egy olyan multidiszicplináris projektet (Chan Zuckerberg Biohub) említett, amely egy szív- és érrendszeri betegség, a HCM (hiptertrofikus kardiomiopátia) kialakulásáért felelős géninterakciók feltárását célozta. Ennek munkafolyamata több részfeladatból állt (megfelelő génpaneladatok kijelölése, az adattisztítás módjának meghatározása, a modell kidolgozása előtt feltáró adatelemzés végzése, nemlineáris interakciók feltárására képes algoritmusok megadása és az eredmények értelmezése, értékelése), melyek végrehajtásakor számos szubjektív analitikai döntést kellett hozni. Ha a rendszer robusztus, az egymásra épülő lépések és döntések garantálják az adatok megfelelő minőségét. Az adattudományi problémamegoldás tehát végső soron egy minőségbiztosítási folyamat, amelynek „életciklusa" a következő: problémameghatározás, adatgyűjtés, adattisztítás, adatfeltárás és vizualizáció, modellezés, utólagos elemzés, eredmények értelmezése és közlése, majd olyan, valóságot tükröző következtetések levonása, amelyek kiállják az „emberi kérdések próbáját". A folyamat multidiszciplináris jellegére tekintettel tehát a szubjektív emberi döntésnek minden lépésnél jelen kell lennie. Mivel a korlátozottan transzparens döntések helytelen megállapításokhoz vezethetnek, az adattudományban minőségellenőrzésre és az empirikus gyakorlat által irányított standardizációra van szükség.

A statisztikus *Leo Breiman* szerint a statisztikai modellezésben két „kultúra" létezik. Az egyik az adatmodellezés, amely sztochasztikus adatmodellek segítségével von le következtetéseket, a másik pedig az algoritmusmodellezés, amely algoritmikus modelleket használ, és az adatmechanizmust ismeretlenként kezeli. A professzor és munkatársai a két irányzat integrálására hozták létre az ún. előrejelző képességi, (számítástechnikai értelemben vett) számíthatósági és stabilitási (predictability, computability, stability, PCS) keretet, amely új módszertanok kidolgozása, új adatokra épülő következtetések levonása és a már alkalmazott modellek értékelése esetén egyaránt hasznos. E három kulcsfontosságú alapelv (PCS) a hiteles adattudomány minimumkövetelményei. Az előrejelző képesség vizsgálatakor az eredményeket a valósággal vetjük össze. A számíthatóság nemcsak a számítástechnika alkalmazására utal az alapadatok gyűjtésétől az eredmények értékeléséig, hanem az algoritmusok, a modellépítés és a szimulációk számítástechnikai megvalósíthatóságára is. A stabilitás, amely az eredmények elfogadható konzisztenciáját jelenti az adat- vagy modellperturbációkkal szemben, a reprodukálhatóság és az értelmezhetőség alapfeltétele. Elemzése az adattudományi életciklus minden lépése során a már említett szubjektív emberi analitikai döntések útján történik a rendszer robusztusságának tesztelése érdekében.

A PCS-keret fontos részét képezi a dokumentáció (a GitHub-on), amely összeköti a valóságot a modellekkel. Minden lépést dokumentálni kell kvantitatív és kvalitatív leírások, indoklások formájában, hogy más kutatók is ellenőrizhessék és eldönthessék a következtetések helytállóságát, valósághűségét.

Adatperturbációt a következők okozhatnak:

– adatok előfeldolgozása, adattisztítás

  Erre *Reinhart, C. M.* és *Rogoff, K. S.* (2010) megállapítását[2] hozta példaként; eszerint létezik egy olyan államadósság/GDP arány (90%), amelyet meghaladva csökken a gazdasági növekedés üteme. *Herndon et al.* (2014)[3] újra lefuttatva *Reinhard* és *Rogoff* tesztjeit, arra a következtetésre jutottak, hogy a szerzők adatszelekciós, kódolási és súlyozási hibákat is ejtettek, így eredményük pontatlanul tükrözi az államadósság és a GDP növekedése közötti kapcsolatot.

– adatválasztás

  Jó példa erre a HCM szívizombetegség okait kutató genetikai kísérlet, amelyben nem mindegy, hogy alacsony vagy nagy felbontásban vizsgálják a genetikai interakciókat hordozó sejtek számát és méretét.

– adatparticionálás

  A különböző adatparticionálási módszerek eltérő következtetésekhez vezethetnek. Például a Vioxx fájdalomcsillapító okozta gyomor- és bélrendszeri, illetve szívbetegségek kockázatának feltárását célzó vizsgálatban az adatkészletet 3 tanulásra szánt (keresztvalidáció), 1 validációs és 1 tesztcsoportra (jövőbeli adatokra vonatkozó proxyra) osztották véletlenszerűen és idő szerint, a kezelések és annak eredményei szerint rétegezve.

– módszer- és algoritmusválasztás.

Az előadó három ajánlást fogalmazott meg:

– fontos, hogy a modellek a különböző tudományterületek szakemberei számára ugyanazt jelentsék;
– az adattisztítás nagyban függ a szubjektív emberi döntéstől, ezért érdemes a tisztított adatok több verzióját is megőrizni;

[2] Reinhart, C. M. – Rogoff, K. S. (2010): Growth in a time of debt. *American Economic Review: Papers & Proceedings*, Vol. 100. No. 2. pp. 573–578.
[3] Herndon, T. – Ash, M. – Pollin, R. (2014): Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. *Cambridge Journal of Economics*, Vol. 38. No. 2. pp. 257–279.

– az előrejelzések pontosságának vizsgálata után érdemes több modellt is megtartani (erre a tudósok által „végeredménynek tekintett” 9 klímamodellt hozta példaként, melyek 1990-től 2100-ig 1,5 és 5,5 °C közötti hőmérsékletemelkedést prognosztizálnak).

A statisztikai következtetések levonásakor napjainkban a cél a döntéshozatalt megalapozó adatok, eredmények átláthatóságának biztosítása annak érdekében, hogy a különböző szakterületek képviselői megismerhessék azok előállításának folyamatát, és megítélhessék a megbízhatóságukat. Problematikussága miatt a pszichológiai folyóiratok megtiltották a szerzők számára a *p*-érték használatát. A professzor véleménye szerint vissza kell hozni a manapság ritkán használt modelldiagnózist. A *p*-érték jelenleg is univerzálisan használt, és segítségével értelmezhetők az eredmények, de emellett más módszerek alkalmazására is szükség van.

Az értelmezhető gépi tanulás elengedhetetlen a tudományos gépi tanuláshoz és a megbízható AI-hoz. Ez az adatokban fellelhető releváns ismeretek és a gépi tanulási modellek által megtanult összefüggések kinyerését jelenti. E tudás az emberek számára akkor tekinthető relevánsnak, ha valóban betekintést nyújt egy adott szakterület problémáiba. Az értelmezhető gépi tanulás három alappillére az előrejelzési pontosság, a leíró pontosság és a relevancia.

Az AI képes felismerni a már megismert mintázatokat, de a még ismeretlen mintázatok, struktúrák azonosítása nagyobb kihívást jelent. *Bin Yu* példaként hozta fel *Basu et al.*[4] iteratív véletlen erdő módszerét (iterative random forest, iRF). Az iRF tanulási algoritmus célja a magasabb rendű interakciók felismerése. A professzor és munkatársai ennek segítségével vizsgálták a HCM szívbetegség kialakulására utaló genetikai prediktorokat (előrejelzőket), a muslicaembriókon megfigyelhető sötét sáv elhelyezkedését és az ún. „francia zászló” modellt, amelynek alapja a sejtműködést befolyásoló molekulák eltérő reakciója a morfogén anyag koncentrációjának változására. Az iRF prediktív ereje jó, stabil interakciókat talál; például a muslica enhancer DNS-szakaszokkal kapcsolatos előrejelzési probléma esetén az algoritmus által stabilként meghatározott (és a bootstrap-ismétlések több mint fele által is annak talált) 20 páronkénti transzkripciósfaktor-kölcsönhatás közül 80%-ot már korábbi biológiai kísérletekben is igazoltak. A HCM vonatkozásában végül *Bin Yu* és csapata szív-MRI-adatokat felhasználva olyan 4 prediktív és stabil génpárt talált az iRF segítségével, amelyek felelősek lehetnek a betegség proxyjaként tekinthető bal szívkamrai izomtömeg alakulásáért.

A mélytanulási modellek számos esetben kiváló előrejelzési eredményeket produkálnak, de nem stabilak, a mély ideghálózatok értelmezésének nehézsége miatt fekete dobozként

[4] Basu, S. – Kumbier, K. – Brown, J. B. – Yu, B. (2018): Iterative random forests to discover predictive and stable high-order interactions. *PNAS*, Vol. 115. No. 8. pp. 1943–1948.

jellemezhetők, és számítási hatékonysági problémák is felmerülnek velük kapcsolatban. Egy ígéretes megoldás e problémákra az ún. adaptív „wavelet-desztilláció” (adaptive wavelet distillation, AWD), amely átviszi a tudást egy mély tanulási modellből („tanár”) egy wavelet-transzformációba („tanuló”), így az utóbbi is rendelkezik már a neurális hálózat jó előrejelző képességével, ugyanakkor javul az értelmezhetőség, a hatékonyság, illetve a tömörítés is. Az előadó az AWD használatára egy biológiai példát említett: annak előrejelzését, hogy mely esetekben megy teljesen végbe a klatrinmediált endocitózis folyamata (amely kulcsfontosságú számos molekula vezikuláris transzportjában a sejt felszínéről a citoplazmába). Ehhez először egy mesterséges rekurrens neurálishálózat-architektúra, a hosszú rövid-távú memória (long short-term memory, LSTM) modell megtanulja osztályozni, hogy mely endocitózis-események sikeresek (előrejelző változó: klatrin; célváltozó: auxilin kofaktor). Mivel az LSTM rendkívül számításigényes, és nehéz a megértése is, a professzor és munkatársai az AWD segítségével az LSTM-modellt egy nagy előrejelző erővel rendelkező wavelet-modellé „desztillálták”. Majd minden skálán csak a legnagyobb 6 wavelet-együtthatót vonták ki (az 5 skálára összesen 30 együtthatót), melyeket egy transzparens lineáris modell betanítására használtak, és csupán a legjobb hiperparamétereket választották ki a tanulásra szánt adatcsoport keresztvalidációjával.

Bin Yu legvégül a csapata által kifejlesztett, két könnyen használható és megbízható szoftvercsomagra hívta fel a hallgatóság figyelmét. Az egyik a Python programozási nyelvet használó Veridical Flow, amelynek egyszerű wrapper-jeivel lehetővé válik a legtöbb adattudományi módszer alkalmazása. A másik pedig egy PCS-szimulációs vizsgálatokat segítő R-csomag, a simChef; ez hatékony eszközöket biztosít a különböző módszerek értékeléséhez. *Bin Yu* és munkatársainak további szoftverei hozzáférhetők a https://github.com/Yu-Group oldalon, a professzor honlapjának elérhetősége pedig a következő: https://binyu.stat.berkeley.edu/

Az előadás után a hallgatóság kérdéseket tett fel az AI további alkalmazási lehetőségeiről. Majd az ülés zárásaként *Kovács Péter* elnök megköszönte *Bin Yu* professzornak az értékes előadást, a hallgatóságnak pedig az aktív részvételt.

Budapest, 2021. október 5.

Kovács Péter, a STAB elnöke

Az emlékeztetőt készítette: *Kondora Cosette*, a STAB titkára