

Bírálat

Tóth Gergely „Adatok értelmezése szimulációkkal és a többváltozós statisztika eszközeivel a kémiában” című MTA doktori dolgozatáról

Az adatok manapság vitathatatlanul kulcsszerepet játszanak a hétköznapi életben és a tudomány számos területén. Ez különösen igaz a számítástechnika megjelenése óta, hiszen a számítógépek lehetőséget adnak nagymennyiségű adat tárolásra, sőt előállítására. Az informatika fejlődésével a hagyományos tudományágak is megváltoztak és egyre nagyobb szerepet játszanak a tudományos kutatásban a számítógépek. Így van ez a kémiával is, ami hagyományosan egy kísérleti tudomány, de a 20. század második felében az informatika fejlődésével párhuzamosan elindult a kémiai rendszerekre alkalmazható elméleti módszerek fejlesztése és alkalmazása is, illetve a kísérleti vizsgálatok is egyre gyakrabban támaszkodnak műszeres mérésekre, amelyekhez gyakran számítógépes adatgyűjtés társul. A fentiek alapján nyilvánvaló, hogy az adatok feldolgozására és értelmezésére nagy szükség van a kémiában is. Ehhez a modern témakörhöz kapcsolódik Tóth Gergely munkássága és MTA doktori dolgozata. A dolgozat számos apró témát dolgoz fel és foglal össze, amelyek saját megítélésem szerint két nagy csoportba sorolhatók. A dolgozat első fele a pályázó korai kutatási irányait mutatja be, amelyek kémiai rendszerek szimulációival foglalkoztak elsősorban a Monte Carlo módszerre támaszkodva. Ezek a munkák tartalmaztak módszerfejlesztési részeket is, valamint érdekes alkalmazásokat. Az ezen a területen elért eredmények a pályázó saját bevallása szerint általában nem keltettek fél a szakma széleskörű figyelmét, ennek ellenére elismerően tudom értékelni ezeket a munkákat is. A dolgozat második fele a pályázó egy relatíve új kutatási irányához kapcsolódik, amely a különböző, általában kémiai adatokat dolgoz fel és értelmez a kemometria módszereivel. Ezek a kutatások is tartalmaznak módszerfejlesztéseket az alkalmazások mellett.

Tóth Gergely eddig 65 cikket publikált, amelyeket felsorol a dolgozat végén. Ezekből 61 jelenik meg az MTMT szakterületi táblázatában, amelyből 57 a PhD-vá nyilvánított tudományos fokozat 1993-as megszerzése óta került publikálásra. Az MTMT-ben szereplő 61 cikkből 59 jelent meg SCI referált folyóiratban, 23 cikkben Tóth Gergely levelező szerző, 14 cikk egy szerzős közlemény és SCI publikációira ezidáig összesen 1010 hivatkozás érkezett, amelyből 829 független (MTMT adatok, 2026. január 6.) A Google Scholar alapján összesen 1281 hivatkozásnál jár (2026. január 6.). A H-indexe 20 az MTMT alapján és 23 a Google Scholar szerint. A publikációira régebben évente általában 40-50 hivatkozás érkezett, az utóbbi 5 évben emelkedő tendencia figyelhető meg és 2025-ben a korábbiakhoz képest kimagaslóan sok (126) hivatkozást kapott a Google Scholar szerint. Egy adatokról szóló dolgozat bírálata során nem kerülhető meg a fenti tudományometriai adatok kritikus értékelése sem. Sem a

publikációk száma, sem azok hivatkozottsága nem nevezhető kiemelkedőnek nemzetközi mércével mérve figyelembe véve a pályázó életkorát. Viszont az MTA doktori cím elnyeréséhez szükséges szintet bőven meghaladják ezek a publikációs adatok. Továbbá kiemelendő, hogy Tóth Gergely publikációi általában a tudományterület rangos nemzetközi folyóirataiban jelentek meg, valamint sok az egy szerzős és első szerzős közleménye, ami tükrözi, hogy a legtöbb esetben jelentős hozzájárulása volt a munkáihoz, amit a dolgozat is alátámaszt.

A dolgozat 136 számozott oldalt tartalmaz és 45 cikk eredményeit foglalja össze. A bevezetés és célkitűzések után három nagy fejezetben tárgyalja (1) a fordított Monte Carlo módszerrel kapcsolatos kérdéseket, (2) a kölcsönhatások, a statisztikus mechanikai inverz probléma, „coarse graining” témakörét és a (3) kemometriát. A dolgozat egy összefoglalással, a tézispontok felsorolásával (amit feleslegesnek érzek a dolgozatban, hiszen a téziszfűzetben úgyis szerepelnek), irodalomjegyzékkel és egy függelékkel zárul. Az irodalomjegyzékben elkülöníti a dolgozatban tárgyalt saját publikációkat, a további saját közleményeket és a külső irodalomjegyzéket. A függelék a rövidítések és szakkifejezések feloldását és néhol részletesebb értelmezését/magyarázatát tartalmazza, valamint egy táblázatot a validációs és optimalizációs paraméterek összefoglalásáról. A dolgozat az eredmények száraz bemutatása mellett tartalmaz számos személyes és tudománytörténeti megjegyzést is, amit én követendőnek tartok, hiszen így az olvasó olyan érdekes háttérinformációkhoz juthat, amelyek nem találhatók meg a kapcsolódó publikációkban. A dolgozat felépítése, stílusa és nyelvezete szerintem megfelelő, viszont a dolgozat szerkesztésén véleményem szerint lehetett volna javítani. Az ábrák minősége sok esetben gyenge, ez persze néhol talán magyarázható azzal, hogy régi cikkekből lettek átvéve. Az ábraalírások nem különülnek el jól a szövegtörzstől és az egyenletek után többször vesszővel kezdődik egy új sor. Ezen formai hibák azért nem tekinthetők súlyosnak, így az írott anyag véleményem szerint megfelel az MTA doktori dolgozattól elvárt szintnek.

Tóth Gergely az eredményeit nem kevés, pontosabban 20 tézispontban foglalta össze, amelyek alapján – néhány esetben rövidítve – elmondhatjuk, hogy

1. A fordított Monte Carlo módszerrel kapcsolatban megállapították, hogy a különböző szerkezeti, termodinamikai és dinamikai tulajdonságok átlagát és fluktuációját nem reprodukálja megfelelően a módszer. Az RMC a kanonikus sokaságtól eltérően mintavételezi a konfigurációs teret, valamint az általuk kifejlesztett szerkezeti alapú molekuláris dinamikai illesztés dinamikája sem felel meg a kanonikus sokaság dinamikájának.
2. Felfedezte, hogy a párkorrelációs függvény fluktuációját önmaga a függvény határozza meg. Levezette egy bizonyítását ennek, illetve egy új bizonyítását a Henderson-féle

tételnek. A szerkezeti függvény fluktuációjának számolására közelítő képletet dolgozott ki.

3. Módszereket dolgozott ki belső szerkezettel rendelkező rendszerek fordított Monte Carlo modellezésére.
4. Négy új módszert dolgozott/dolgoztak ki a statisztikus mechanika szerkezet → párkölcsönhatás inverz problémájának numerikus megoldására.
5. Coarse graining potenciálok esetére tisztázta, hogy párhuzamosan mind a szerkezet, az energetika és a dinamika reprodukciója szükséges egy új redukált modellben.
6. Alkáli ion/halogenid-higany és szisztematikus alkáli ion/halogenid-víz kölcsönhatásokat határozott meg kvantumkémiai számolások segítségével határfelületi és tömbfázisbeli molekuláris dinamikai szimulációkhoz.
7. QM/QM szemimpiikus kvantumkémiai Monte Carlo szimulációs módszert fejlesztettek ki és sikerrel alkalmaztak amorf Si-re, amorf C-re és kloroszilánokra.
8. A Car-Parrinello szimulációs módszer segítségével vizsgálta a sztratoszférikus ózon lebontásában résztvevő HCl, ClONO₂ és H₂O molekulák adszorpcióját salétromsav-monohidrát kristály felületén. A szimulációk során elsőként észlelt protonátmenetet a salétromsav-monohidrát tömbfázisában.
9. Karbamid kristálynövekedésére vonatkozó kinetikai Monte Carlo számolásaival bemutatta, hogy reális kristályosodási mikroszkopikus sebességi állandók kaphatóak pusztán a kristályok makroszkopikus morfológiájából kiinduló paraméteroptimalással.
10. Bemutatták, hogy kinetikai Monte Carlo szimulációk sebességi állandói megkaphatóak szabadenergia számoló módszerek szisztematikus alkalmazásával.
11. Definiálták a feltételes valószínűség szabályait alkalmazva a többminimumú rendszerek kommittor függvényeit és az ahhoz kapcsolódó további függvényeket, amelyek szükségesek a kinetikai Monte Carlo módszerre való áttéréshez. Informatív ábrázolási módokat vezettek be a számolt függvényekre.
12. Matematikai módszereket dolgozott ki adatok sor-oszlop rendezésére annak érdekében, hogy könnyebb legyen felismerni a két és háromdimenziós adattömbökben található mintázatokat.
13. Kifejlesztette a Neumann-féle trend teszt és a Durbin-Watson teszt egytényezős varianciaanalízisre visszavezethető változatát. Továbbfejlesztette a teszteket időeltolódásos esetre és keresztkorrelációk vizsgálatára, valamint számos példán demonstrálták a gyakorlati alkalmazhatóságát.
14. Szisztematikusan áttekintette a szórás alapú skálázás és átlag alapú centrálás hat lehetséges esetét normál eloszlású adatokra, bemutatva a mintaszámtól függően kialakuló esetek néha meglepő eloszlásait. Az egyik alesetre elsőként határozta meg a kialakuló eloszlás matematikai képletét.

15. A bemenő adatok allokációjának szisztematikus vizsgálatával megállapították, hogy a modellezést megelőző kísérlettervezés különböző optimálisainak és az elkészült modellek validálásakor használt paramétereknek gyenge a keresztoptimalizációs lehetősége. Szisztematikus numerikus eredményekkel megerősítették, hogy a validálási paraméterek alapján a szélső allokációknál kapunk átlagosan jobb modelleket.
16. Bemutatták, hogy periodikus változókat tartalmazó adatok vizsgálata során azt az adatablakot érdemes választani a periodikus változóra, amelyekben az adott változó varianciája minimális.
17. Szisztematikus vizsgálataikkal megerősítették, hogy az illeszkedés jóságára vonatkozó validációs paraméterek a túlillesztés miatt kis mintaméreteknél túlzottan optimisták, nagy teljesítményű modellek esetén pedig nem relevánsak. Számolásaik összegzésekként megalkottak egy átlátható validálási protokollt.
18. Felfedezte, hogy az egy és a több elem kihagyásos keresztellenőrzési validációs paraméterek egymásba skálázhatóak, így azok közül mindig csak a modell típusától függően számítási szempontból kedvezőbbet/egyszerűbbet érdemes használni.
19. Bemutatták, hogy a robusztusság vizsgálatára mind a keresztellenőrzés, mind az „out of bag” bootstrap eljárás alkalmas, de az előbbi kisebb varianciájú és könnyebben értelmezhető értékeket ad.
20. Kidolgozott egy gyors eljárást annak a detektálására, hogy ún. nagy hatású pontokat tartalmaz-e a mintánk. Számos példán bemutatták, hogy a teszt könnyen elvégezhető pusztán az R^2 és Q^2 validációs paraméterek ismeretében.

A tézispontok száma talán több a megszokottnál, de mivel a dolgozat elég sok szerteágazó kutatási projektet foglal össze, a nagyszámú tézispont szerintem rendben van és mind a 20 pontot új tudományos eredménynek tekintem és elfogadom.

A dolgozathoz kapcsolódóan a következő megjegyzéseim és érdeklődő kérdéseim vannak:

1. Az 5. oldalon olvasható, hogy „A tudományos utóélet számítógépes kutatásoknál sokszor irrelevánssá válhat, pl. egy 20-30 évvel ezelőtti Cray szuperszámítógépen, ma egy telefonon kiszámolható eredményre.” Azt hiszem értem mire gondolhatott a pályázó, de mégis vitatkoznék az idézett mondattal, hiszen hatása akkor is lehet egy régi eredménynek, ha azt manapság sokkal gyorsabban és/vagy pontosabban is ki tudnánk számolni. Mi a véleménye erről?
2. A dolgozatban szerepel az „on fly” kifejezés. Én ezt „on the fly”-ként szoktam használni, illetve így szoktam vele találkozni az irodalomban.
3. A dolgozat 55. oldalán szerepel, hogy „Mint már írtam, az inverz problémával való kutatásokat 2011-ben felfüggesztettem, mivel a területen szinte semmi visszajelzést

nem kapta, látszólag senkit se érdekelt akkor ez a terület.” Aztán később olvasható, hogy mások szép eredményeket értek el a pályázó korábbi kutatásaira is építve. Felmerül a kérdés, hogy valóban érdemes volt felfüggeszteni ezeket a kutatásokat?

4. A 64. oldalon írja a pályázó, hogy „A hagyományos $A \rightarrow B$ reakció sémában gyakran előfordul egy köztes átmeneti állapot megjelenése. Ha ez kellően stabil, akkor tekinthetjük a rendszerünket úgy is, hogy három energiaminimummal rendelkezik a rendszer.” Hogyan lehet egy átmeneti állapot stabil? Milyen értelemben használja itt az átmenti állapot kifejezést? Az én olvasatomban az átmenti állapot mindig egy nyeregponthoz tartozik, így az nem lehet stabil, illetve nem lehet minimumként értelmezni.
5. Az utóbbi időkben nagyon sikeres terület a hálózatkutatás, gondolok itt például Barabási Albert-László munkásságára. A hálózatkutatás is adatokkal és azok értelmezésével foglalkozik. Hogy látja a hálózatkutatás és a kemometria kapcsolatát?

Végül a dolgozat stílusához illeszkedve egy személyes megjegyzéssel foglalnám össze a véleményemet. Tóth Gergelyt már egyetemi hallgató korom óta ismerem, akkor még diákként vettem részt a jelen dolgozat anyagához is kapcsolódó óráin. Akkor még nem gondoltam, hogy kb. 25 évvel később a tanár úr MTA doktori dolgozatát fogom bírálni. Tóth Gergely kutatási témáit felületesen régóta ismertem, de csak most a dolgozat olvasása alkalmával tudtam részletesebben betekintést nyerni tudományos munkásságába. Igen pozitív véleményem alakult ki, szembesülve a relatíve sok és szerteágazó témájú projekt során elért eredményekkel, figyelembe véve, hogy általában egyedül vagy egy-két kollégával együttműködve dolgozott. A fentiek alapján jó szívvel javaslom a nyilvános vita kiírását és sikeres védelem esetén az MTA doktora cím odaítélését.

Szeged, 2026. január 6.

Czákó Gábor

egyetemi tanár, az MTA doktora