Bírálat Miklós István Leszámlálások és mintavételezések bonyolultságelmélete a bioinformatikában című az MTA doktora cím elnyerésére benyújtott dolgozatára

May 5, 2024

Miklós István (továbbiakban MI) dolgozata a matematikai bonyolultságelmélet, leszámlálás és mintavételezés alkalmazásáról, fejlesztéséről szól a bioinformatika területe számára. Ezzel a meghatározással is azt akarjuk hangsúlyozni, hogy MI munkássága egyértelműen a matematika területére tartozik és Rényi Alfréd szellemében merít kérdéseket egy más tudományterület a genomika és annak alkalmazásának kérdéseiből. Konkrétan az ott szükséges leszámlálási illetve elsősorban mintavételezési feladatok bonyolultságát vizsgálja, gyorsításuk lehetőségeire ad módszereket illetve negatív eredményeket bizonyít amikor ez nem lehetséges.

Elmondható, hogy e kutatás időszerű és MI eredményei komoly előrelépést jelentenek egyes mintavételezési algoritmusok tekintetében. A génszekvenlás, a génterápia a személyre szabott gyógyászat forradalma zajlik. A kidolgozott módszerek az elméleti kutatás fontos eszközeiként szolgálhatnak.

A dolgozat felépítése és tartalma.

A dolgozat négy fejezetből áll. Az első megismerteti az olvasót a bonyolultság elmélet alapfogalmaival illetve a szükséges speciális fogalmakkal. Bevezeti a Markov láncokat és azok gyors keverése igazolásának technikáját. Ugyanítt bemutatja a bioinformatikaban használatos alapvető matematikai modelleket, gráfokat, biológiai sorozatokat és azok átrendezését. Utóbbiaknak különösen nagy jelentősége van a carcinogén mutációk vonatkozásában. Ez a fejezet nagyban hozzájárul a szakterületet kevésé ismerő számára a későbbiek megértéséhez. Jól szerkesztett, példákkal is kellően ilusztrált.

A dolgozat első része polinomiális időben megoldható függvény-problémákat

tárgyal. E fejezetben egy félgyűrű kerül bevezetésre illetve azon egy algebrai dinamikus programozási feladat. Ezek segítségével sikerül gyors számolási algoritmust adni Zucker-Tinoco energiamodellben az RNS szekvencia térszerkezetei Boltzmann eloszlásának a k-adik momentumaira. Második tézise memória hatékony algoritmust ad véletlen reguláris nyelvtanok Baum-Welch tanitására. Harmadik tézise dinamikus program segítségével két leszámolás FP (függvény polinomiális osztály) beliségét igazolja: egyrészt a legtakarékosabb SCJ átrendezésekét illetve a legtakarékosabb mediánokét. Az első két tézis 2005-ben került publikálásra és 30 illetve 27 idézettel rendelkezik. Megjegyezem, hogy az értekezés bevezetője mernöki szintűként aposztrofálja az 2005-s eredményeket, de véleményem szerint, minkét esetben ügyes matematikai megoldásról van szó. Szellemes az eredeti feladat átvitele egy matematikai strukturára majd algebrai dinamikus program alkalmazása a félgyűrű struktuán.

A dolgozat második része 4. fejezete gyorsan keverő Markov láncok egy osztályát vizsgálja. Igazolja, hogy páros fokszámsorozatok esetén a gyors keverés öröklődik Tyshkevich decompozió elemeiről a dekomponált sorozatra. Ezután P-stabil illetve egyes nem P-stabil fokszám sorozatokra igazol gyors keverést és megadja P-stabil fokszamsorozatok egy elég széles osztályát, a lineárisan korlátozott fokszámsorozatokét. Igazolja, hogy bármely együttes fokszám-mátrix kiegyensúlyozott realizációin a switch Markov lánc irreducibilis és gyorsan konverő. A fejezet végén pedig egy korábbi Tannier-el közös eredményt ismertet, nevezetesen, hogy van gyors approximáló véletlen függvény feladat megoldó (FRPAS) illetve teljesen polinomiális majdnem egyenletes gyors fügvény feladat megoldó (FPAUS) a DCJ modellben a legtakarékosabb szcenáriókra.

Úgy vélem, hogy ennek a fejezetnek a tézisei és a genom átrendezésre vonatkozó eredmények a dolgozat legfontosabb, megjelenésükkor áttörést jelentő eredményei. Ezt jelzi magas idézettségük, a tézisek sorrendjében 30,18,36,34,13.

A dolgozat harmadik részének címe negativ eredmények. Először két tételt mutat be. A REV modell természetes Markov láncára igazolja, hogy lassan keverő. (A REV modell az egykrokoszomás lineáris genomok ezeket a tulajdonságokat nem sértő DCJ atrendezéseit tartalmazza.) Belátja, hogy az SCJ genom átrendezéseket megengedő modellben az evolúciós fán a legtakarékosabb szcenáriók #P beliek, feltéve $RP \neq NP$ akkor nem létezik FPRAS approximáció illetve a legtakarékosabb medián scenarió probléma #P-teljes . Megjegyezzük, hogy a 103. Lemma conjunktiv normal formakról pontosabban 3CNF-ről önmagában is érdekes. A 13. Tézis megad egy sok jó tulajdonsággal rendelkező Markov láncot, amiről eddig a gyors keverést nem bizonyították.

A dolgozat negyedik része először a négy reverzál sejtés igazolását fejti ki olyan előjeles permutációkra, amelyek overlay gráfja lineáris. A 114. Tétel szellemes átfogalmazása a feladatnak, a 115. Tétel pedig meglepő a laikus olvasó számára.

Ezután lehűtéses Markov lánc technikák alkalmazza, visszzavezeti a DCJ modellre a majdnem egyenletesen reverzal szortolás modell legrövidebb scenáriók mintavételezését. A 11 fejezet Gibs mintavételezést alkalmaz az SCJ modellben az evolúciós fák majdnem legtakarékosabb cimkézésére (14. Tézis). Az utolsó, 12., kissé technikai fejezet két irreducibilis Markov lánc konstrukcióját adja páros gráfok élszinezésére.

Összefoglalóan szeretém megállapítani, hogy Miklós Istvánnak a doktori cím elnyerésére benyújtott dolgozata és tézisfüzete jelentős munkásságot mutat be. Munkásságának egy része egy fontos bioinformatikai területhez a genomátrendezések leszámolásához, azok bonyolultságelméletére vonatkozik. Pozitív és negatív eredményei egyarán nagy jelentőségűnek tartjuk, amit hatásuk is igazol. Egyes bizonyításai rendkívül technikásak, ugyanakkor szellemesek. Munkáinak jelentőségét támasztja alá hogy olyan kiemelkedő kutatók támaszkodnak azokra, idézik eredményeit mint Jens Stoye, Remco van der Hofstad vagy Nicholas Wormald.

Hasonlóan, eredményes invenciózus kutatói munkáját igazolja, hogy kiemelkedő társzerzőkkel is dolgozott együtt (diákjaival közös illetve önálló munkái melett). Ezek között lehet említeni Aaron Darling-t, Eric Tannier-t, Catherine Greenhill-t, Toroczkai Zoltánt, Erdős Pétert.

Az egész dolgozat olvasmányos, jól szerkesztett, segíti az olvasót a téma megismerésében, a matematikai nehézségek és a haladás bemutatásában.

Külön kiemelném, hogy a publikációk ismertetésénél gondosan bemutatja mi volt saját hozzájárulása és mi az általa vezetett mester ill. PhD hallgatóké (esetenként a társszerzőké). Szintén kiemelendő a diákjainak adott kérdések eredményes megoldása, a jó témavezetés.

Javasolom a nyilvános vita megtartását Miklós István a doktori cim odaítélésére benyújtott pályázatáról.

A benyújtott dolgozat alapján meggyőződésem, hogy Miklós István munkássága messzemenően megfelel a doktori cím követelményeinek, annak odaítélését melegen támogatom.

Apró technikai megjegyzések.

A dolgozatban a szereplő cikkek copy-paste beillesztése csöppet sem kárhoztatható, a bíráló kifejezetten fölöslegesnek és időtrablónak gondolja a teljes

dolgozat elvárását a rövid dolgozat illetve tézisek benyújtása helyett, de a beillesztes közben három helyen észrevettük az "in this paper" utalást.

dolgozatban

p 42 l 8 valami hiányzik [...]

p 64 l 18 fölösleges "in"

p 80 és a későbbiekben számos helyen és a bibliográfiában is Erdős P.L. helyett hibásan Erdős E.P. van.

p. 84 a P-stabilitas központi szerepet játszik, jó lenne egy kis heurszitika arra, mit is jelent. Ugyanitt talán jó lett volna a multicomodity flow-ról pár szót említeni (előtörténet, jelentős eredmények, hasznosság).

p 128 -13 "markers" Valahol definiálva van? A genomikában járatlan számára szükséges volna definiálni..

p 180 e (10.2) szerencsés lenne megmagyarázni, miért tekintjük ezt energiának.

tézisekben p 23 l 13 "miden".

Kérdés:

Mi okozhatja, hogy a 13. Tézisben ismertetett Markov lánc jó keverési tulajdonságát nehéz igazolni?