

Bírálat Miklós István "Computational complexity of counting and sampling problems in bioinformatics" című MTA doktori értekezéséről

Az értekezés fő témája leszámplálási folyamatok bonyolultságának elemzése és ehhez kapcsolódóan különböző Markov-láncok keverési idejének vizsgálata. A vizsgált leszámplálási feladatok többek között a gének szekvenciájának meghatározásából és előrejelzéséből származnak, így az értekezés a bioinformatikában fontos, számos alkalmazással rendelkező kérdéseket tanulmányoz. Ugyanakkor a bemutatott eredmények eléréséhez széles matematikai eszköztárra illetve új módszerek kidolgozására is szükség volt, a disszertáció tehát elméleti matematikai szempontból is értékes munka.

Az összességében majdnem 300 oldalas mű egy előkészítő részből és négy fő részből épül fel. A bevezetőben a szerző részletesen, az alapoktól kezdve ismerteti a felhasznált fogalmakat elsősorban a számítástudomány (bonyolultsági osztályok), illetve valószínűségelmélet (Markov-láncok, keveredés) területéről. A pontos definíciók nagyban segítik a következő fejezetek megértését.

Az első rész algebrai módszerekből indul ki, a génszekvenciákból adódó sorozatok itt félgűrűk segítségével reprezentálja a szerző, ennek alapján az algebrai dinamikus programozás eszköztárát is felhasználva igazolja bizonyos leszámplálási feladatok polinomiális idejű megoldhatóságát. A bizonyításhoz ezen kívül rejtett Markov-lánc modellek felépítésére, elemzésére is szükség van. A kidolgozott módszerekre két alkalmazást is ismertet a szerző, a disszertációban több szempontból vizsgált "Single Cut or Join" modellről, mely a génszekvenciák átrendezésére vonatkozó modellek közé tartozik.

A második részben a szerző áttér adott fokszámeloszláshoz tartozó páros gráfokon értelmezett Markov-láncok vizsgálatára. Itt a Markov-lánc alaplépése az, hogy négy csúcs között futó két él egy-egy végpontját felcseréljük, vagyis a (u, v) , (u', v') éleket töröljük, és az (u, v') , (u', v) éleket adjuk hozzá a gráfhoz (fontos, hogy ezek a párok előtte ne legyenek összekötve egymással). Ebben a részben a szerző különböző elégséges feltételt ad arra, hogy az így definiált Markov-lánc gyorsan keverő legyen, további kapcsolódó Markov-láncokat is elemez, majd mindezek segítségével bizonyítja, hogy "Double Cut and Join" konfigurációk számának polinomiális idejű meghatározhatóságát.

A harmadik rész középpontjában olyan eredmények állnak, ahol a vizsgált leszámplálási probléma nem oldható meg polinomiális időben, sőt az eldöntéshez szükséges tanúk száma is NP-teljes. Itt a bizonyítás fő eszköze a problémák adott szerkezetű fákra való visszavezetése, és ezek leszámplálása, szükség esetén esetek szétválasztásával, összegyűjtésével.

A negyedik részben egy, gráfok színezésével kapcsolatos kérdés megoldása után a szerző visszatér a Markov-láncoknak a gének konfigurációira vonatkozó problémákra való alkalmazására. Pontosabban, részletesen elemzi a "Double Cut and Join" modellnél pontosabbnak mondható "reversal" modellt, illetve belátja, hogy az előbbiből kis perturbációkkal is eljuthatunk az utóbbihoz. Végül a Gibbs-féle mintavételezés és egyéb, Markov-láncok keverését használó módszerek szerepét vizsgálja ezekben a feladatokban.

A doktori munka I–IV. részeiben ismertett, a szerző cikkeiből, illetve könyvéből származó

tételeket értékes, új eredménynek fogadom el. A dolgozat erősségének tartom, hogy a bioinformatikából származó problémák megoldására a Markov-láncok elméletének nagyon széles eszköztárát használja fel (kiegészítve néhány egyéb, például algebrai vagy kombinatorikai módszerrel), és számos saját ötlet segítségével fejleszti tovább ezt a kutatási területet. Az itt bemutatott eredményeket a szerző több, magas színvonalú folyóiratban publikálta. A dolgozat jól felépített, és bár a megadott területen belül változatos problémákat vizsgál, az egyes nagyobb részekben belül vizsgált kérdések sok szempontból összetartoznak és így vezetnek a vizsgált kérdés mélyebb megértéséhez. A hivatkozások listájában a szerző 25 saját cikke mellett 102 hivatkozást találunk. Ezzel együtt, a dolgozat több helyen kevésbé részletesen mutatja be, hogy az elért eredmények a szakirodalomban található cikkekhez képest miben jelent újdonságot, illetve hogy az adott problémakörrel kapcsolatban mik a legújabb ismert eredmények (a hivatkozásoknak is csak egy kisebb része vonatkozik az utóbbi 10 évben megjelent cikkekre).

Összességében Miklós István doktori munkája a bonyolultságelmélet és a valószínűségelmélet területén számos új, érdekes eredményt mutat be, értékes módszereket fejleszt és a bioinformatikai alkalmazások során is több fontos példát tartalmaz.

A fentiek alapján a disszertációt az MTA doktori fokozat követelményeinek megfelelőnek tartom, és javaslom a disszertáció nyilvános vitára bocsátását.

Kérdések

1. A II. részben leírt "switch" műveletnek milyen általánosításai ismertek (esetleg más, vagy speciális szerkezetű gráfokra), és ezek közül melyekre lehetnek esetleg általánosíthatók a dolgozatban bemutatott módszerek?
2. A 12. fejezet eredményei közül van-e olyan, mely általánosítható arra az esetre, ha páros gráf helyett egy több osztályból (részből) álló gráfot, például Turán-gráfot tekintünk?

Budapest, 2024. május 10.

Backhausz Ágnes

Eötvös Loránd Tudományegyetem

Természettudományi Kar

Matematikai Intézet

Valószínűségelméleti és Statisztika Tanszék

Rényi Alfréd Matematikai Kutatóintézet

Mesterséges Intelligencia Osztály