

ALGEBRAI SZÁMÍTÁSOK

Polinomok véges ponthalmazokon

Tanulmányoztuk a véges ponthalmazok lexikografikus standard monomjainak szerkezetét. A standard monomok -- mivel gyakran áttekinthető bázisát alkotják a ponthalmazon értelmezett polinomfüggvények terének -- egy sor probléma kapcsán hasznosak -- az elméleti alkalmazásoktól a konkrét interpolációs kérdésekig. A lexikografikus standard monomokat egy kombinatorikai játék, az ún. lex játék segítségével sikerült jellemeznünk. A jellemzésből egyfelől egy új algoritmust nyertünk a standard monomok számítására. A módszer gyorsabb, mint a korábban publikált eljárások. Több érdekes esetben lineáris idejű, tehát aszimptotikus értelemben optimális sebességű. Másfelől a játék segítségével néhány kombinatorikai szempontból érdekes halmazrendszer standard monomjaira sikerült áttekinthető leírást adni. A leírás rácspolygonok segítségével fogalmazható meg kényelmesen. Ez utóbbi eredményt, illetve megközelítést extrémális kombinatorikai problémákra is alkalmaztuk.

Az n elemű alaphalmaz részhalmazaiából álló halmazcsalád teljes l -széles, ha valamilyen k -ra pontosan a $k, k+1, \dots, k+l-1$ elemű részhalmazokból áll. Sikerült leírunk az ilyen rendszerek karakterisztikus vektoraihoz tartozó Gröbner-bázist és bizonyos kapcsolódó struktúrákat. Az eredmény kombinatorikai alkalmazását is adjuk. Az előző eredmények egy részét általánosítjuk arra az esetre, amikor a részhalmazok elemszámára egyenlőség helyett modulo q kongruenciafeltételeink vannak (q egy p prím hatványa), és az alaptest p -karakterisztikájú test. Itt is adódnak érdekes kombinatorikai alkalmazások.

A standard monomok játékkal (ez a fentebb említett Lex játék) való jellemzését a radikál-ideálok esetéről ki tudtuk terjeszteni az ún. lokálisan monomiális ideálok esetére.

T. Harima egy szép tétele szerint erős kapcsolat van a test feletti projektív térbeli X, Y diszjunkt, véges ponthalmazok Hilbert-függvénye között, ha a Z uniójuk teljes metszet. Ennek az eredménynek egy moduláris és kombinatorikus változatát bizonyítottuk. Itt X, Y nulla-egy vektorokból állnak, Z a teljes Boole-kocka, és test helyett általánosabban D -gyűrű felett érvényesek a formuláink.

Olyan többváltozós polinom egyenletrendszereket vizsgáltunk, amelyek perturbációi többszörös gyököket tartalmazó (nulla dimenziós) rendszereknek. Ilyenek sokszor előfordulnak a gyakorlatban, amikor az igazi egyenletrendszernek csupán egy közelítését tudjuk kinyerni a mérési adatainkból. Azt vizsgáltuk elsősorban, hogy az eredeti rendszer milyen jellemzői kaphatók meg globális algoritmikus módszerekkel. Megmutattuk, hogy a rendszer ún. multiplikációs mátrixából hagyományos numerikus lineáris algebrai módszerekkel sok fontos információt kaphatunk meg az eredeti egyenletrendszerről.

Kvantumszámítások

Franciaországi együttműködő partnerünkkel közös munkánkban megmutattuk, hogy véges kommutatív csoportok szorzótáblája hatékonyan tesztelhető randomizált algoritmussal. Olyan polylog(n) idejű módszert adunk, amely egy n -szer n -es táblázatot mindenképpen elfogad, ha

az egy n elemű kommutatív csoport szorzótáblája, és jó eséllyel elutasít, ha az (relatív edit-távolságban mérve) nem áll közel egy csoport szorzótáblájához. Az algoritmus akkor is működik, ha az n számot nem ismerjük pontosan, de ekkor a csoportelemek inverzére is szükség van. Eljárásunk exponenciálisan gyorsabb az eddig ismert módszereknél.

Ebben a munkában a háromszög triangulációinak színezésével kapcsolatos Sperner-lemma bonyolultságát vizsgáltuk. Algebrai topológiai eszközökkel több-dimenziós sokaságokra is általánosítható determinisztikus algoritmust adtunk, amely egy konstans faktor erejéig csak annyi lekérdezést végez, amennyi a korábbról ismert alsó korlát. Randomizált, illetve kvantum-algoritmusok lekérdezéseinek a számára valamivel gyengébb, de még a felső becslésünkkel inverz polinomiális kapcsolatban levő (négyzetgyök, illetve negyedik gyök nagyságrendű) alsó korlátot sikerült igazolni.

Brüsszeli együttműködő partnerünkkel közösen sikerült megmutatnunk egy fontos és ígéretes kvantumszámítógép-modellről, hogy képes a hagyományos értelemben vett kvantumszámítógép hatékony szimulációjára. A szóban forgó modellben a 2 kvantumbites kapuknak megfelelő műveletek a kvantumbit-pár eltoljtaira "egyszerre", pontosabb megfogalmazásban "logaritmusok" szintjén átlagolva hajtódnak végre. A szimuláció alapötlete az, hogy bizonyos kvantum-biteket rögzített állapotban tartva a megfelelő unitér Lie-algebrában kétszeres kommutátor-képzéssel a több pozícióra átlagolt műveletből kinyerhetők a kívánt helyeken végrehajtandó műveletek.

Bebizonyítottuk, hogy véges kvantum-kapukészletek univerzalitása algoritmikusan eldönthető, mégpedig a legtöbb szokásos számítási modellben polinom idejű algoritmussal. A bizonyításban rávilágítottunk az univerzalitás problémájának kapcsolatára véges lineáris csoportok invariánsaival, valamint polinomideálok Hilbert-függvényével.

Lie-algebrák és véges csoportok algoritmusai

A klasszikus Lie-algebrák hosszú gyökei geometriájának, valamint a Timmesfeld által tárgyalt gyökrészcsoporthalmozatok geometriájának egy közös axiomatikus jellemzését adtuk meg. Lie-algebrák esetén az axiómák természetesen adódnak a hosszú gyökökhöz tartozó filtrációkból. Eredményeink klasszikus Lie-algebrák explicit izomorfizmusproblémáját megoldó hatékony algoritmusok alapjául szolgálhatnak.

Több dolgozatban is foglalkoztunk azon kváziprimitív és innately tranzitív csoportokkal, amelyek előfordulnak primitív koszorúszorzatok részcsoporthalmozataiként. Osztályoztuk azon innately tranzitív csoportokat, amelyek úgy ágyazhatók be primitív koszorúszorzatokba, hogy a felső csoportra vetített képük tranzitív.

Egy permutáció-csoportot k -csillagnak mondunk, ha az k -alaphalmaz minden k elemű részalmozatainak a stabilizátora nem-triviálisan hat a k -elemű halmazon. Becsléseket adunk a k -csillag csoportok orbitjainak számára.

Foglalkoztunk az olyan p-csoportok szerkezetével, amelyekben a kommutátorlánc két szomszédos tagja között a hányadoscsoport kicsi (nevezetesen eléri a Philip Hall által 1934-ben igazolt alsó korlátot). Új eredményeket sikerült igazolni ezekről a csoportokról.

Sims jól ismert módszere segítségével, egy véges csoport elegendően finom részcsoporthálózatát felhasználva, a csoport elemeit fel tudjuk írni adott generátorok szorzataként. Az általunk kidolgozott általánosítás részalalmaz-láncokat használ, így olyan esetekben is alkalmazható, amikor a maximális részcsoporthálózatok indexe rendkívül nagy. Módszerünket a sporadikus egyszerű csoportok felismerési problémájának megoldására használtuk.

KOMBINATORIKUS ALGORITMUSOK

Csúcsösszefüggőség növelése

Algoritmikusan megválasztottuk azt a kérdést, hogy egy adott irányított (aszimmetrikus) hálózatot legkevesebb hány él hozzáadásával lehet bővíteni úgy, hogy egy előírt számú csúcs együttes meghibásodása esetén is működőképes maradjon. A feladat algoritmikusan igen nehéz; nehézségét az is jelzi, hogy a szimmetrikus (irányítatlan) változata máig megoldatlan. Az elkészült dolgozat a korábbi, ellipszoid módszeren alapuló és így a gyakorlatban nem kivitelezhető algoritmust egy hatékonyabbal helyettesíti. Az eredmény előzetes fogadtatását jelzi, hogy konferenciaváltozatának elfogadásával egyidőben lehetőséget kapott az ACM Transaction on Algorithms különszámában való megjelenésre.

Gráfszínezési feladatok

A listás él-multiszínezés problémában minden élhez tartozik egy engedélyezett színlista, és az, hogy az él hány színt kér. Marcotte és Seymour polinom idejű algoritmust adtak a problémára fák esetén. Ezt sikerült kiterjeszteni kevés körrel rendelkező gráfokra.

A "Graph coloring problems and their applications in scheduling" c. dolgozatban áttekintettünk néhány gráfszínezési problémát (listás színezés, minimális összegű színezés, stb.), és példákat mutattunk arra, hogy ezek miként használhatók ütemezési feladatok modellezésére.

A "Minimum sum multicoloring on the edges of trees" c. dolgozatban polinom idejű közelítő sémát javasoltunk fák minimális összegű él-multiszínezésére. Hasonló séma már létezett a megfelelő csúcsszínezési kérdésre.

Paraméteres bonyolultság

Gráfok szétvágásával kapcsolatos problémákat vizsgálunk a paraméteres bonyolultság szemszögéből. Többek között uniform polinomiális algoritmust adunk arra a feladatra, amelyben kijelölt terminálokat kell elválasztani egymástól k csúcs törlésével.

Különböző gráfosztályokon vizsgáltuk a csúcsszínezés és a színezésbővítés feladatainak paraméteres bonyolultsági vonatkozásait. Bonyolultsági eredményeket igazoltunk arra az esetre, amikor a gráf csúcs/él törlésével merev körű gráf/intervallum gráf lesz.

Igazoltuk Schaefer Dichotómia Tételének egy paraméteres bonyolultsági megfelelőjét. A feladat az, hogy egy pontosan k súlyú behelyettesítést találjunk, amely kielégíti az összes megadott Boole-feltételt. Kiderül, hogy a feltételek típusától függően ez a probléma vagy uniform polinom időben megoldható, vagy pedig $W[1]$ -teljes.

SZÁMÍTÁSELMÉLET

Részben a jelen OTKA-pályázat támogatásával készült a következő kötet:

Formal methods in computing. (Editors: Ferenczi Miklós, Pataricza András, Rónyai Lajos), Akadémiai Kiadó, Budapest, 2005, 1-425.

A kötet angol nyelvű áttekintő tanulmányok formájában nyújt bevezetést, illetve első tájékoztatást néhány, a számítások világában használatos alkalmazott matematikai területről. Ezek a feldolgozás sorrendjében: algoritmusok és bonyolultságelmélet, gráfelméleti módszerek és alkalmazásai a számítógépes tudományokban, klasszikus és nem klasszikus logikák, válogatott fejezetek az automaták elméletéből, bevezetés a fatranszformációk elméletébe, a logika alkalmazásai a számítástudományban, véges forrású tömegkiszolgálási rendszerek és alkalmazásai, metamodellek és modelltranszformációk.

A kötet első tanulmánya a jelen pályázat támogatásával (is) készült. Rövid áttekintő jellegű bevezetést nyújt az algoritmusok elméletének a gyakorlati szempontból hatékony számításokkal kapcsolatos részeibe. A tárgyalt témák a következők: számítási modellek, a Turing-gép és a RAM-modell; a polinomiális idő, nevezetes és hasznos P-beli és FP-beli feladatok; az NP-nyelvosztály, nemdeterminisztikus számítások, NP-teljesség; néhány nevezetes algoritmus-tervezési technika (mohó módszer, dinamikus programozás, elágazás és korlátozás, közelítő algoritmusok, véletlent használó módszerek, néhány heurisztikus technika). A hangsúlyt az algoritmika konstruktív, a gyakorlatban is jól használható vonatkozásaira helyeztük.

Az Informatikai algoritmusok II. (alkotó szerkesztő Iványi Antal) Algebra c. fejezetében elsősorban a véges testek feletti, illetve racionális együtthatós polinomok felbontására használt módszerekbe adtunk betekintést. A fejezetben részletezett anyag egy algebrai algoritmusok kurzus anyagának több, mint felét teszi ki, és tartalmazza a véges testekről szóló alapvető tételeket, melyek például egy kódelméleti tárgyhoz szükségesek.

ADATBÁZIS-ELMÉLET

Foglalkoztunk a relációs adatbázisok világában fellépő funkcionális függésekkel kapcsolatos teendők és műveletek vizualizációs kérdéseivel. A függések tulajdonságainak elemzése, megértése fontos fázisa az információs rendszerek tervezésének. Az itt felmerülő feladatok megoldása során hasznosak lehetnek a képernyő-adta megjelenítési lehetőségek. Az általunk javasolt grafikus technikák alkalmazásával a tervező könnyebben és gyorsabban boldogul az

adatbázismodell felállításával kapcsolatos feladatokkal. Az általános grafikus módszerek mellett a táblázatkezelő programok ötleteiből is merítettünk.

A "Relációs adatmodell tervezés" című könyvfejezetben rövid bevezetést nyújtunk a relációs sémák tervezésének az elméletébe, különös figyelmet szentelve a kérdéskör algoritmikus vonatkozásainak (a fejezet az "Informatikai algoritmusok" c. kötet részét képezi). Az anyagot a funkcionális függések részletes tárgyalására építettük, innen jutottunk el a lezárások, kulcsok, illesztések, felbontások és normálformák fogalmaihoz. Ezt a klasszikusnak mondható anyagot a többértékű és általános függések (utóbbin belül összekapcsolási és elágazó függések) áttekintése egészíti ki. Reményeink szerint a munka egyetemi tankönyv(részlet)ként használható. Ezt a kifejtés részletességén túl példák, gyakorlatok és feladatok is segítik.

Megjelent a témában a következő könyvünk:

Békéssy András, Demetrovics János: Adatbázis-szerkezetek. Akadémiai Kiadó, 2005. (1--481. old.)

A munka tekinthető egyfelől egyetemi tankönyvnek, amely korszerű bevezetést nyújt az adatbázis kezelés alapvető témáiba. Másfelől a relációs modell feldolgozása monografikus igényű, így a munkát haszonnal forgathatják azok is, akik mélyebben kívánnak foglalkozni a relációs adatmodellel. A tárgyalt témák a következők: Adatbázis-alapfogalmak, tárgykapcsolat modell, hálós és hierarchikus adatbázisok, fizikai szervezés, relációs adatmodell. Ez utóbbi téma mintegy 270 oldalt teszi ki a kötetből. A különböző függőségek, normálformák és alkalmazásaik igen alapos tárgyalása miatt a munka a relációs tervezéselmélet alpművének tekinthető. A könyv elnyerte az Akadémiai Kiadó 2005-ös nívódíját.

Informatikai algoritmusok II. (alkotó szerkesztő Iványi Antal) kötetben a relációs adatkezelésben fontos szerepet játszó lekérdezés-átírásról írtunk egy tanulmányt. Az adatbázis kezelés központi feladata a felhasználói kérdések hatékony megválaszolása. Ennek az egyik legerősebb eszköze a lekérdezések átírása olyan alakba, amelyik kedvezőbb végrehajtást tesz lehetővé. A tárgyalásban központi szerephez jutnak a nézetek. Az átírás bonyolultsági vonatkozásai után a szerzők tárgyalják a legfontosabb gyakorlati megoldásokat is.

INTERNETES ALGORITMUSOK

Keresés a Weben

A webkereső rendszerekben alkalmazott rangsorok fontos komponense a hiperhivatkozások által meghatározott webgráf elemzésén alapuló, nagy számítási igényű PageRank eljárás. Továbbfejlesztése, a személyre szabott PageRank rögzített webgráf esetén különböző bemeneti, ún. személyre szabott eloszlásokhoz különböző rangsorokat rendel. Kidolgoztunk a személyre szabott PageRank megvalósítására egy hatékony kétfázisú módszert: az előfeldolgozás során véletlen séták szimulációjával előállított, a csúcsok számában lineáris méretű adatbázisból tetszőleges személyre szabó eloszláshoz tartozó személyre szabott PageRank gyorsan és pontosan közelíthető. Ellentétben a korábban ismert módszerekkel, a lehetséges személyre szabó eloszlásokat algoritmusunk nem szorítja meg. Cikkünkben

bizonyítjuk továbbá, hogy a közelítés szükségszerű: a pontos személyre szabott PageRank értékek (bizonyos gráfokon) csak túlságosan nagy, a csúcsok számában négyzetes méretű adatbázisban tárolhatók. Nagyméretű, valódi webgráfokon végzett kísérleteink igazolták, hogy a javasolt Monte Carlo-módszer mérsékelt erőforrásigény mellett a webkereső rangsorok számára elégséges pontosságot ér el.

Kapcsolódó eredményként a személyre szabott PageRank feladatra javasoltunk egy közelítő számítási eljárást. Algoritmusunk a nagyméretű személyre szabott PageRank feladatot a webgráf adjacencia mátrixának szinguláris érték felbontására (SVD) és egy kisméretű mátrixinverzióra vezeti vissza. Az SVD-t Drineas és Kannan randomizált eljárásával közelítettük. Gyakorlati tapasztalatok alapján az SVD alapú módszer bizonyos esetekben jól alkalmazható, de pontossága általában elmarad a fentebbi Monte Carlo módszertől.

Új, Monte Carlo alapú számítási algoritmust adtunk a webkeresők „Hasonló lapok keresése” feladatának egyik lehetséges matematikai modelljére. Ez a modell a lapok hasonlóságát a hiperlinkek grájában vett tág környezetük segítségével definiálja. Korábban nem volt ismert egyszerű erő-forrásigényű megoldó algoritmus ilyen hasonlósági mértékek számítására. A gyakorlatban alkalmazható hasonlósági mértékek vagy a lapok szövege alapján kerestek hasonló lapokat, ami a nyelvek és megfogalmazások változatossága miatt csak korlátozott eredményre vezethetett, vagy pedig a hiperlinkek grájában csak kis környezeteket vettek figyelembe, és ezért kaptak túlságosan kevés hasznos információt.

A hasonlóságkeresésével és spektrálfelbontással kapcsolatban eredmények elismertségét jelzi, hogy ez irányú kutatásaink Yahoo Faculty Research Grant támogatásban részesültek, illetve meghívott összefoglaló cikkben ismertettük a Magyar Tudomány hálózatokkal foglalkozó különszámában.

A hasonlóságkeresési és személyre szabott rangsorolási feladatok közös jellemzője, hogy a hasonlósági mátrix kvadrátikus méretű, amelynek tárolása külső tárban sem megoldható gyakorlati feladatméret (százmillió vagy egymilliárd csúcs és tízszer ennyi él) esetén. A kimenet méretében szublineáris -- és ezért valós méretekben is működő -- algoritmusokat adunk a feladatok megoldására. Mind elméleti, mind gyakorlat-orientált eredményeink a legrangosabb konferencián (IEEE FOCS és WWW) jelentek meg, ahol egy cikk elfogadása igen nagy presztízszt jelent.

A webhelyek látogatottsága - és így üzemeltetőjük bevétele - jelentős mértékben attól függ, hogy a hely milyen előkelő helyezést ér el a különböző keresők rangsorában. Utóbbi rangsort nem csak a weboldalak szöveges tartalma, hanem a hiperlinkek által kialakított webgráfban elfoglalt pozíció, a hivatkozások száma és minősége is befolyásolja. Emiatt egyes webhelyek nem megengedett eszközöket alkalmaznak, hogy a keresőrangsorbeli helyüket mesterségesen javítsák azáltal, hogy sűrű szövésű hálóként kölcsönösen hivatkoznak egymásra, vagy pedig értelmetlen tartalom nélküli weboldalak százezreit kizárólag azzal a céllal hozzák létre, hogy azok a kívánt webhelyre hivatkozzanak. Kidolgoztunk egy olyan automatikus, mindennemű emberi beavatkozást nélkülöző módszert, amely képes az előbb említett link spam jelenségek nagy hányadának felderítésére. Algoritmusunk kizárólag a webgráf linkstruktúrájában észlelhető devianciák statisztikai alapú szűrésén alapul, így nyelvtől és tartalomtól függetlenül alkalmazható. Gyakorlati felhasználhatóságát alátámasztja, hogy a .de címtartományból származó valódi részgráfra alkalmazva a kapott eredmények (spam-e vagy sem) döntően egyeztek a kiértékelést végző szakemberek ítéletével. A kézi tanítás nélkül szűrő algoritmusunk leírását, amelynek referált folyóirat változatát 2006-ban fogadták el, már közel

20 független cikk idézi. 2006-ban két új módszert is bemutattunk: az első a hivatkozások mentén észlelt nyelvmodell eltérés alapján, a másik pedig környezetükben hasonló weboldalak alapján klasszifikálja a "spamet".

Elkészült egy nagyméretű telekommunikációs és web szerver naplóállományokat vizsgáló szoftver keretrendszer és annak bizonyos moduljai. Célunk egy olyan eszköz elkészítése volt, amely alkalmas a nagyobb internetes oldalak látogatói szokásainak vizsgálatára. Kísérleteinkben a legnagyobb magyar portál, az origo (www.origo.hu) web logjait használtuk fel. Ezen a portálon található online hírek, magazinok, szoftverbázis, fórum, ingyenes email szolgáltatás illetve kereső. A portál több mint 130.000 oldalból áll, és egy tipikus munkanapon kb. 6.500.000 HTML találatot szolgál ki, ami egy 2,5 GB méretű szerver logot jelent egy napra. Ez némi előfeldolgozás után 400 MB-ra csökkenthető, ami még mindig meghaladja egy általános rendszer tár-képességeit. A rendszer önálló értékkel rendelkező komponense a naplók tömörítését és kibontását rendkívüli hatékonysággal végző eszköz.

ADATBÁNYÁSZAT

Behatóan foglalkoztunk a gyakori elemsorozatokot kinyerő szófa alapú Apriori algoritmus több implementációs kérdésével: útvonalválasztási módok a szófa belső pontjaiban, zsákutca nyelés, equisupport nyelés, teljes nyelés hatékonysága, az adatbázis cache-elése. Munkánk során kiemelt hangsúlyt helyeztünk a gyakori elemhalmazokat és a gyakori sorozatokat kinyerő megoldások különbségeire.

Az adatbányászat egyik legaktuálisabb kérdése a létező adatbázis, ill. adattárház megoldásokkal való összekapcsolódás, az adatbányászati algoritmusok hatékony integrálása a meglévő rendszerekbe. Ebben az irányban a gyakori termékhalmozok keresésének relációs adatbázison belüli SQL alapú megoldásait vizsgáltuk. Megadtuk az FP-growth mintanövelő gyakori termékhalmoz kereső algoritmus egy a relációs adatbázisokba szorosan integrált implementációját. Ezt összevetettük a korábbi legjobb APRIORI-alapú megoldással és megmutattuk, hogy az FP-growth algoritmus számos előnye a relációs környezetben is megmarad.

Az adatbányászat területén 2006-ban elsősorban telekommunikációs ügyfelek viselkedésének modellezésével foglalkoztunk. Barabási Albert László világhírű kutatócsoportjával közös munkáink komoly PR értékűek, többek között a következő két híradásban is megjelennek: 'Life is short in online news' by P. Ball, News@Nature, 2005 'News Online Seems to Have Long Shelf Life' by N. Cohen, The New York Times, July 17, 2006 (utóbbiról az Akadémia honlapján is jelent meg tájékoztatás). A rangos VLDB konferenciához kapcsolódó workshopon bemutattuk az elemzéseink alapját képező szoftverarchitektúrát, amely extrém nagy mennyiségű adatok rugalmas feldolgozására alkalmas, és az oszlop-orientált adatbázisok, illetve a nearline megoldások ötvözeté.

Több új módszert dolgoztunk ki a szövegállományok tematizálására. Bemutattuk, hogyan lehet külső információt (lexikonokat, enciklopédiát) alkalmazva javítani a dokumentumok automatizált témameghatározásán. Kísérletünkben az egyre növekvő népszerűségű és fedettségű Wikipedia online enciklopédiát dolgoztuk fel. Ezen felül a ritka kifejezések témameghatározó erejét is demonstráltuk.

