ԵՐԵՎԱՆԻ ՊԵՏԱԿԱՆ ՀԱՄԱԼՍԱՐԱՆ

Դանոյան Հայկազ Էդվարդի

ԴԻՍԿՐԵՏ ԷՔՍՏՐԵՄԱԼ ԽՆԴԻՐՆԵՐԻ ՀԵՏԱԶՈՏՈՒՄ

Ա.01.09-«Մաթեմատիկական կիբեռնետիկա և մաթեմատիկական տրամաբանություն» մասնագիտությամբ ֆիզիկամաթեմատիկական գիտությունների թեկնածուի գիտական աստիճանի հայցման ատենախոսության

ՍԵՂՄԱԳԻՐ

ԵՐԵՎԱՆ 2013

ЕРЕВАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Даноян Айказ Эдвардович

ИССЛЕДОВАНИЕ ДИСКРЕТНЫХ ЭКСТРЕМАЛЬНЫХ ЗАДАЧ

АВТОРЕФЕРАТ

Диссертации на соискание ученой степени кандидата
физико-математических наук по специальности
01.01.09-«Математическая кибернетика и математическая логика»

ЕРЕВАН 2013

Ատենախոսության թեման հաստատվել է Երևանի պետական համալսարանում:

Գիտական ղեկավար՝ ֆ.մ.գ.դ. Լ. Հ. Ասլանյան
Պաշտոնական ընդդիմախոսներ՝ տ.գ.դ. Գ. Խաչատրյան
ֆ.մ.գ.թ. Է. Եղիազարյան
Առաջատար կազմակերպություն՝ Երևանի պետական մանկավարժական համալսարան

Պաշտպանությունը կայանալու է 2013 թ., հունիսի 24-ին ժ. 14:00-ին ԵՊՀ–ում գործող ԲՈՀ-ի 044 «Մաթեմատիկական կիբեռնետիկա» մասնագիտական խորհրդի նիստում, հետևյալ հասցեով՝0025, Երևան, Ալեք Մանուկյան 1:

Ատենախոսությանը կարելի է ծանոթանալ Երևանի պետական համալսարանի գրադարանում:
Սեղմագիրն առաքված է 2013թ. մայիսի 23-ին

Մասնագիտական խորհրդի
գիտական քարտուղար,

ֆիզ.-մաթ. գիտ. դոկտոր՝ Վ. Ժ. Դումանյան

Тема диссертации утверждена в Ереванском государственном университете.

Научный руководитель: д.ф.м.н. Л. А. Асланян
Официальные оппоненты: д.т.н. Г. Хачатрян
к.ф.м.н. Э. Егиазарян
Ведущая организация: Ереванский государственный педагогический университет

Защита состоится 24 июня в 14:00 на заседании специализированного совета 044 «Математическая кибернетика» ВАК при ЕГУ по адресу: 0025, г. Ереван, ул. Алека Манукяна 1.

С диссертацией можно ознакомиться в библиотеке Ереванского государственного университета.

Автореферат разослан 23-го мая 2013 г.

Ученый секретарь
специализированного совета

доктор физ.-мат. наук В. Ж. Думанян

Թեմայի արդիականությունը: Ներկա աշխատանքը նվիրված է կոմբինատոր փնտրման խնդիրների մի մասնավոր դասի և դրա հետ կապված որոշ դիսկրետ օպտիմիզացիոն խնդիրների ուսումնասիրմանը: Համակագչային գիտությունը և ինֆորմացիոն տեխնոլոգիաները, որոնք միտված են լուծելու ինֆորմացիոն արդյունաբերության զարգացման և ինֆորմացիոն հասարակության ստեղծման խնիրները, գործ ունեն ինֆորմացիայի գերմեծ և աճող ծավալների և դրանց տարաբնույթ ալգորիթմական մշակումների հետ: Այսօր նշվում է օրինակ, որ պատմականորեն, մինչև 2003 թվականը մարդկությունը կուտակել է շուրջ $5 \cdot 10^{18}$ (five exabyte) ինֆորմացիա այն պարագայում, երբ ներկայումս այդ նույն ծավալի ինֆորմացիա արտադրվում է յուրաքանչյուր երկու օրվա ընթացքում: Պարզ է, որ նախկինում ստեղծված ալգորիթմները կարող են բավարար արդյունավետ չլինել մուտքային տվյալների նման ծավալների և դրաց աճի նման արագությունների դեպքում: Ըստ այդմ՝ անհրաժեշտություն է առաջանում մշակել նոր ալգորիթմներ և խնդիրների դիտարկման նոր մոդելներ, որոնք իրենց մեջ կներառեն մուտքային տվյալների մեծ և դինամիկ աճող քանակության իրողությունը: Օրինակ՝ որպես այդպիսի մոդել այսօր հաճախ քննարկվում է տվյալների հոսքի մոդելը (data stream model):

Տվյալների հավաքածուների/պարունակության հետ իրականացվող հիմնական տեխնիկական գործողությունները ընդգրկում են տվյալների գրանցման, խմբագրման, հեռացման և փնտրման գործողությունները: Սակայն վերջնական արդյունքում պահանջը որոշակի գիտելիքի կորզումն է, որը կարող է հանդիսանալ փնտրումից հետո իրականացվող մեկ առանձին փուլի աշխատանքի արդյունք, բայց և այն կարող է ինտեգրված լինել փնտրման/ընտրման փուլի հետ:

Գերմեծ ծավալի ինֆորմացիայի մեջ տվյալների հասանելիության ապահովման համար տվյալները պետք է համապատասխան ձևով կազմակերպված լինեն: Սովորաբար դա իրականացվում է տվյալների հենքերի ղեկավարման համակարգերի միջոցներով: Սակայն փնտրման խնդիրը ինքը ավելի լայն գաղափար է, քան փնտրումը տվյալների հենքերում և դա կապված է տարբեր հարակից պայմանների առկայության հետ, ինչպես օրինակ հիշողության սարքերի յուրահատկությունները, որոնց վրա գրառված է տվյալների պարունակությունը, հաճախ կրկնվող գործողությունները, որոնք իրականացվում են տվյալների նույն բազմության վրա և այլն: Օրինակ՝ որոշ համակարգերում տվյալները պահվում են մագնիսական ժապավենի վրա՝ լուծելով գրանցված տվյալների ընտրման և դրանց փնտրման/կարգավորման խնդիրները: Ցույց է տրվել, որ, օրինակ n երկարության հաջորդականության մեդիանը հաշվելիս, երբ օգտագործվում են ընդամենը երկու անցումներ ժապավենի վրայով, անհրաժեշտ է առնվազն $\Omega(\sqrt{n})$ հիշողություն, և որ $O(\sqrt{n} \cdot log\, n)$ հիշողությունը բավարար է[1]: Տվյալների մշակման դասական մոդելում ալգորիթմները գործ ունեն տվյալների մեծածավալ հենքերի/հաջորդականությունների հետ, որոնք մշակվում են դրանց վրայով p անցումների (passes) միջոցով և սահմանափակ s ծավալի օպերատիվ հիշողության օգտագործմամբ: Պարզ է՝ ցանկալի է, որ իրականացվի միայն մեկ անցում, և որ

[1] J. Munro, M. Paterson, Selection and sorting with limited storage, Theoretical Computer Science, Vol. 12, 1980, pp. 315-323

հիշողության ծավալը լինի հնարավորին չափ փոքր (հաճախ՝ առավելագույնը բազմանդամալոգարիթմային (polylogarithmic)), սակայն սա ալգորիթմի նկատմամբ խիստ պահանջ չէ և սոսկ ընդհանուր ցանկություն է: Երբեմն սա նաև անհնար է: Ուսումնասիրությունների սկզբնական փուլում որոշ խնդիրների համար պարզվել է, թե p և s պարամետրերի որ արժեքների դեպքում են այդ խնդիրները լուծելի (կամ մոտարկելի):

Անդրադառնանք փնտրման այնպիսի խնդիրների, որոնք իրականացնում են բազմակի փնտրում տվյալների միևնույն կառուցվածքի վրա: Դասական օրինակը փնտրումն է ծառերում [2]: Ներկա աշխատանքի ուսումնասիրման առարկան որոշ տեսակետով նման է ծառերում փնտրմանը: Ինֆորմացիան հարկ է արտապատկերել մեքենայական հիշողության հաջորդական հատվածի վրա այնպես, որ, ըստ մուտքային բառի, հնարավոր լինի իրականացնել նրա բոլոր մոտակա հարևանների արդյունավետ փնտրումը: Ծառատիպ փնտրման կառուցվածքում առաջանում և դիտարկվում են ծառի վերակառուցման խնդիրներ՝ ըստ հիմնական ինֆորմացիայի փոփոխման: Նմանատիպ հարցեր առաջանում են նաև դիտարկվող մոդելում, սակայն այստեղ տվյալները պահպանվում են ցանկերի բազմության միջոցով, և խնդիրը հանգում է այդ ցանկերի վերակազմավորմանը՝ ըստ մուտքային ինֆորմացիայի փոփոխման:

Առավել ճշգրտումներից առաջ բերենք մի պարզ օրինակ, որը բնութագրում է դիտարկվող հիմնական խնդիրը: Դիցուք տրված է որևէ լեզվի ուղղագրական մի բառարան: Տրվում է մուտքային բառ, որը ընդհանուր առմամբ կարող է գրառման թերություն ունենալ: Բնական է, որ մենք կձգտենք գտնել բառարանից այն ուղղագրական ճիշտ գրառումները, որոնք առավել նման են մուտքային բառին: Դրանք շատ են: Որպես մուտքային բառի ճշգրտում կարելի է վերցնել ստացված նմանակներից որևէ մեկը: Բառարանը, որը այս խնդրում ծավալուն է, մեկ անգամ գրանցվում է հիշողության մեջ, բայց այն շարունակաբար օգտագործվում է շատ մուտքային բառերի համար, և ստացվում է հաճախ կրկնվող գործողությամբ փնտրման խնդիր: Մոտակա հարևանների փնտրման խնդիրը ունի լայն կիրառություն, և բավարար է այդ դասից նշել կերպարների վերծանման խնդիրը:

Խոսելով խնդրի լուծման ալգորիթմների և դրանց արդյունավետության մասին՝ կարելի է նշել ալգորիթմների երկու դաս՝ ծառատիպ և հաշավորման (hashing) տիպի ալգորիթմներ: Ծառատիպ ալգորիթմները լայնորեն ուսումնասիրված են և ունեն շատ կիրառություններ՝ օրինակ՝ պատկերների GIF-ֆորմատում, տվյալների սեղմման LZW մոդելում: Այս մոդելի կոմբինատոր ալգորիթմական վերլուծությունները ևս հայտնի են:

Ներկա աշխատանքը ուղղակիորեն նվիրված է հաշավորման տիպի ալգորիթմների ուսումնասիրմանը: Այս ալգորիթմները հայտնի են վաղուց, նրանք ուսումնասիվել են R. Rivest-ի կողմից (RSA կրիպտոհամակարգի համահեղինակ), որը էմպիրիկ ճանապարհով ստացել է ալգորիթմի արդյունավետության գնահատական: Նա նաև առաջադրել է մի պնդում, ըստ որի ալգորիթմական կառուցվածքների օպտիմալ

[2] A. V. Aho, J. E. Hopcroft and J. D. Ullman, The Design and Analysis of Computer Algorithms, Addison-Wesley, 470p. 1974

ընտրությունը գտնվում է դիսկրետ իզոպերիմետրիկ խնդրի տիրույթում[3]: Ներկա աշխատանքի հիմնական նպատակը ուսումնասիրումների այս ոլորտի խորացումն է: Տիրույթի այլ ուսումնասիրություններ չեն անդրադարձել նշված խնդիրներին, և աշխատանքի դերը կարող է կարևոր լինել տիրույթի վերջնական ձևավորման իմաստով:

Լավագույն համընկնման փնտրման որոշակի ալգորիթմներ տարածական տիրույթի համար կիրառում են փնտրման առաջին-լավագույնը (best-first) սկզբունքը, ինչպես, օրինակ, ընդլայնումներով ալգորիթմները[4]: Ուսումնասիրման նախնական փուլի ալգորիթմներից է լավագույն համընկնման փնտրման Էլիասի ալգորիթմը, որը տրոհում է տարածությունը և նրանում գրանցված տվյալները որոշակի ցանցային (grid) տրոհման համապատասխան: Ալգորիթմը վերլուծում է մուտքային բառին համապատասխան նրա մոտակա տիրույթները ցանցից՝ սկսելով մուտքային բառի տիրույթից և աճեցնելով այն մոտքային բառից ունեցած հեռավորությանը համապատասխան: Ցանցի տիրույթի նվազագույն հեռավորությունները մուտքային բառից օգտագործվում են որպես շեմք ցանցի որոշակի մասերի փնտրման տիրույթից դուրս թողնելու համար: Այս մոտեցման մի տարբերակ, որը դիտարկում է բազմաչափ տվյալներ, դիտարկվել է Վեբերի, Շեկի և Բլոտտի կողմից[5]: Մեկ այլ ալգորիթմ[6] արտապատկերման վրա հիմնված մի մոտեցում է, որը փորձում է իրականացնել չափողականության նվազեցում:

Այսպիսով «լավագույն համընկման» կամ «մոտակա հարևանների» որոնման խնդիրը իմաստալից կայանում է հետևյալում. դիցուք ունենք բազմություն, տրված է դրա որևէ ենթաբազմություն և որևէ տարր: Պահանջվում է՝

ա) Պարզել՝ արդյո՞ք տրված տարրը պատկանում է տրված ենթաբազմությանը:

բ) Գտնել տրված ենթաբազմությունից տրված տարրին «ամենամոտ» կետ:

գ) Գտնել տրված ենթաբազմությունից տրված տարրին «ամենամոտ» կետերի բազմությունը:

Այստեղ «մոտ» կամ «ամենամոտ» հարևան արտահայտությունները պետք է հասկանալ Հեմմինգյան հեռավորության իմաստով: Աշխատանքում դիտարկվում է գ) խնդիրը: Պարզ է, որ հատարկման եղանակով հնարավոր է հաշվել տրված կետի և ենթաբազմության բոլոր կետերի հեռավորությունները և դրանից ելնելով գտնել պահանջվող տարրերը: Որոշ դեպքերում դա կարող է հանգեցնել հաշվումների ահռելի մեծ քանակության, ինչը խնդրի լուծումը կդարձնի գործնականորեն անհնար: Աշխատանքում դիտարկվում է այն դեպքը, երբ որոնման տիրույթը բինար n-չափանի միավոր խորանարդի որևէ ոչ-դատարկ ենթաբազմություն (ֆայլ) է: Երկու կետերի միջև հեռավորություն ասելով կհասկանանք նրանց միջև եղած Հեմմինգյան հեռավորությունը:

---

[3] R. L. Rivest, On the optimality of Elias's algorithm for performing best-match searches, Information Processing, pp. 678–681, 1974

[4] G. R. Hjaltason and H. Samet, Distance browsing in spatial databases, ACM Transactions on database Systems, 24(2):265–318, 1999

[5] R. Weber, H.-J. Schek, and S. Blott, A quantitative analysis and performance study for similarity search methods in high-dimensional spaces, In Proceedings of the 24th International Conference on Very Large Data Bases (VLDB), pp. 194–205, New York, 1998.

[6] J. H. Friedman, F. Baskett, and L. J. Shustek, An algorithm for finding nearest neighbors. IEEE Transactions on Computers, 24(10):1000-1006, October1975

Կենթադրենք նաև որ ֆայլը ներկայացված է մի քանի կապակցված ցուցակների միջոցով, այնպես որ յուրաքանչյուր ցուցակի էլեմենտներ պատկանում են միավոր խորանարդի ինչ-որ հատկությունների բավարարող բլոկի: Ենթադրվում է, որ նախապես հայտնի է միավոր խորանարդի՝ նշված բլոկների տրոհումը: 1971թ. Փ. Էլիասի կողմից առաջարկվեց նշված խնդրի լուծման դինամիկ ծրագրավորման տիպի հաշ-կոդավորման ալգորիթմ: Հետագայում Ր. Րայվեսթի աշխատանքները հիմք հանդիսացան ենթադրելու, որ Էլիասի ալգորիթմը չհատվող հավասար հզորությամբ բլոկներով տրոհման դեպքում ինչ-որ իմաստով օպտիմալ է, երբ դրանցից յուրաքանչյուրը հանդիսանում է դիսկրետ իզոպերիմետրիկ խնդրի լուծում՝ մասնավորապես գունդ: 70-ականներին Ա. Տիետավաինենի և զուգահեռաբար Վ. Զինովևի և Վ. Լեոնտևի կողմից ապացուցվեց, որ կատարյալ կոդեր կարող են գոյություն ունենալ միայն խիստ սահմանափակ դեպքերում, կամ որ նույնն է՝ միավոր խորանարդի տրոհումներ չհատվող հավասար շառավղով գնդերի գոյություն ունեն շատ «նեղ» պարամետրերի դասերի համար: Ելնելով խնդրի կիրառական բնույթից՝ նպատակահարմար է դիտարկել կատարյալ կոդերին «մոտ» կոդերի միջոցով ստացվող տրոհումներ և ընդհանրացնել Էլիասի ալգորիթմը նմանատիպ բլոկների համար: Որպես այդպիսիք կարող են հանդես գալ՝

ա) համարյա կատարյալ կոդերը,

բ) հավասարաչափ փաթեթավորված կոդերը,

գ) քվազիկատարյալ կոդեր,

դ) միավոր խորանարդի ծածկույթներ տարբեր շառավիղներով չհատվող գնդերով,

ե) միավոր խորանարդի ծածկույթներ չհատվող «գնդանման» բազմություններով:

Աշխատանքում ուսումնասիրվում է Էլիասի ալգորիթմի օպտիմալությունը կատարյալ կոդերի համար, հաշ-կոդավորման սխեմաների ընդհանրացումը որոշ քվազիկատարյալ և հավասարաչափ փաթեթավորված կոդերին համապատասխանող տրոհումների համար: Նշված դեպքերում հետազոտվում է ալգորիթմի աշխատանքի արդյունավետությունը: Ուսումնասիրվում է նաև ալգորիթմի էֆֆեկտիվությունը այնպիսի չհատվող բլոկների դեպքում, որոնցից յուրաքանչյուրը հանդիսանում է ավելի փոքր չափանի գնդերի դեկարտյան արտադրյալ:

Ատենախոսության նպատակը և խնդիրները: Ըստ կատարյալ, քվազիկատարյալ և համարյա կատարյալ կոդերի՝ ուսումնասիրել մոտակա հարևանների փնտրման խնդրի Էլիասի մոդելը և ալգորիթմը, ստանալ ալգորիթմի արդյունավետության հետազոտում, արդյունավետության անալիտիկ արտահայտություններ և գնահատականներ: Ուսումնասիրել նշված մոդելի օպտիմալությունը՝ կախված դիտարկվող բլոկների երկրաչափությունից: Տալ հաշ-կոդավորման սխեմաների ընդլայնում՝ ելնելով կատարյալ կոդերի տարատեսակ ընդհանրացումներից և դրանց դեկարտյան արտադրյալների կիրառումից:

Հետազոտման օբյեկտը: Աշխատանքում դիտարկվում է մետրիկական տարածության մեջ տվյալ տարրի մոտակա հարևանների փնտրման Էլիասի հաշ-կոդավորման մոդելը և ալգորիթմը, ուսումնասիրվում է դրա օպտիմալությունը՝ կախված բլոկների երկրաչափական հատկություններից, դիտարկվում են նաև հաշ-կոդավորման սխեմաներ, որոնք համապատասխանում են կատարյալ կոդերին և տարատեսակ այլ ընդհանրացումներին: Հետազոտման առարկան մոդելի հնարավոր ընդհանրացումն է, օպտիմալության տիրույթի ուսումնասիրումն է և արդյունավետության ճշգրիտ գնահատումը:

Հետազոտման մեթոդները: Հետազոտման մեթոդները են՝ սխալ ուղղող կոդերի տեսության մեթոդները, ինչպես նաև դիսկրետ օպտիմալացման որոշ մեթոդներ և արդյունքներ՝ ինչպես, օրինակ, դիսկրետ իզոպերիմետրիայի հատկությունը:

Գիտական նորությունը:

● Ստացվել են կոմբինատոր փնտրման Էլիասի ալգորիթմի բարդության անալիտիկ արտահայտություններ այն դեպքի համար, երբ ալգորիթմը օգտագործում է փնտրման տիրույթի տրոհում՝ ըստ սխալներ ուղղող կատարյալ կոդերի:

● Ալգորիթմը ընդհանրացվել է որպես հատվող բլոկներով հաշ-կոդավորման սխեմաներ (այստեղ բլոկները գնդեր են), և դուրս են բերվել ալգորիթմի բարդության բանաձևերը քվազիկատարյալ և համարյա կատարյալ կոդերի օգտագործման դեպքերում:

● Ուսումնասիրվել է միավոր խորանարդի ենթաբազմության հարևան գագաթների բազմության ծավալի աճի ֆունկցիան և այդ հիման վրա ապացուցվել է, որ բալանսավորված հաշ-կոդավորման սխեմաների համար Էլիասի ալգորիթմը օպտիմալ է, երբ համապատասխան բլոկները ունեն գնդային կառուցվածք:

● Տրվել է ալգորիթմի ընդհանրացում բալանսավորված հաշ-կոդավորման սխեմաների տեսքով, որոնց դեպքում բլոկները հանդիսանում են ավելի ցածր չափի գնդերի դեկարտյան արտադրյալ:

Կիրառական նշանակությունը: Մոտակա հարևանների փնտրման օգնությամբ սովորաբար լուծվում են կլաստեր անալիզի և ճանաչողության տիպի խնդիրներ, սակայն անհրաժեշտ են նոր ալգորիթմներ և ընդհանրացումներ տվյալների մեծ ծավալների համար, անհրաժեշտ են ալգորիթմներ՝ հարմարեցված մեքենայական բառի փոփոխվող երկարությանը, և որ կարևոր է՝ որոնք թույլ են տալիս կարգավորել փնտրման ցուցակների ծավալները կախված փնտրման տիրույթի ծավալից:

Ստացված արդյունքների ապրոբացիան և հրապարակումները: Ատենախոսության արդյունքները զեկուցվել են ԵՊՀ Ինֆորմատիկայի և կիրառական մաթեմատիկայի ֆակուլտետի և նրա Դիսկրետ մաթեմատիկայի և տեսական ինֆորմատիկայի ամբիոնի գիտական սեմինարներում, ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտի սեմինարում և Հայ-ռուսական (սլավոնական) համալսարանի յոթերորդ տարեկան գիտաժողովում (3-7 դեկտեմբեր

2012թ.): Ատենախոսության հիմնական արդյունքները տպագրված են երեք գիտական հոդվածներում:

Ատենախոսության կառուցվածքը և ծավալը: Ատենախոսությունը բաղկացած է ներածությունից, երեք գլուխներից, եզրահանգումից և գրականության ցանկից, որը ներառում է 68 աշխատանք; Ատենախոսության ծավալը 81 էջ է:

Աշխատանքի բովանդակությունը:

Ներածության մեջ հիմնավորված է հետազոտական թեմայի արդիականությունը, ձևակերպված են աշխատանքի նպատակները, գիտական նորույթը և հիմնական դրույթները, որոնք ներկայացվում են պաշտպանության:

ԳԼՈՒԽ 1. ԿՈՄԲԻՆԱՏՈՐ ՓՆՏՐՄԱՆ ԷԼԻԱՍԻ ԱԼԳՈՐԻԹՄԸ ԵՎ ԴՐԱ ԲԱՐԴՈՒԹՅԱՆ ԲԱՆԱՁԵՎԵՐԸ ԿԱՏԱՐՅԱԼ ԿՈԴԵՐԻ ՀԱՄԱՐ

1.1 Հիմնական սահմանումներ: Դիցուք $E = \{0,1\}$: E^n-ով նշանակենք n-չափանի միավոր խորանարդի գագաթների բազմությունը: $d(x,y)$-ով կնշանակենք x և y ($x, y \in E^n$) վեկտորների միջև հեմմինգյան հեռավորությունը: Կամայական $x \in E^n$ վեկտորի համար x կենտրոնով և r շառավղով գունդը կնշանակենք $S_r^n(x)$-ով:

Սահմանում 1.1. Կոդ ասելով կհասկանանանք E^n-ի կամայական ոչ դատարկ ենթաբազմություն: Սովորաբար դիտարկվում են այնպիսի կոդեր, որոնք բավարարում են նաև այլ պայմանների (օրինակ գծային են, ցիկլիկ են և այլն):

Սահմանում 1.2: C կոդի մինիմալ հեռավորություն կոչվում է հետևյալ թիվը՝

$$d_C = min_{c_1,c_2 \in C, c_1 \neq c_2} \{d(c_1, c_2)\}: \quad (1.1)$$

Կոդի փաթեթավորման շառավիղ կանվանենք $r_C = \left\lfloor \frac{d_C - 1}{2} \right\rfloor$ թիվը:

Սահմանում 1.3: C կոդի ծածկման շառավիղ կոչվում է հետևյալը՝

$$R_C = \max_{x \in E^n} \min_{c \in C} d(x,c) \quad (1.2)$$

Սահմանում 1.4: C կոդը կանվանենք կատարյալ, եթե նրա համար տեղի ունի $r_C = R_C$ հավասարությունը:

Սահմանում 1.5: C կոդի հարակից դաս ըստ $\forall x \in E^n$ վեկտորի կոչվում է $x + C = \{x + c / c \in C\}$ բազմությունը:

Սահմանում 1.6: C $[n,k,d]R$ կոդի ծնող մատրիցա է կոչվում այն $n \times k$ մատրիցան, որի տողերը հանդիսանում են C կոդի բազիս: C կոդի ծնող մատրիցան կնշանակենք G_C-ով:

Սահմանում 1.7: H_C մատրիցան կոչվում է C $[n,k,d]R$ կոդի ստուգող մատրիցա, եթե նրա համար տեղի ունի

$$\mathrm{c} \in \mathrm{C} \Leftrightarrow H_C \mathrm{c}^{\mathrm{T}} = 0 \quad (1.3)$$

Սահմանում 1.9: C կոդի կշռային սպեկտոր կանվանենք $A_0^C, A_1^C, \dots, A_n^C$ թվերը, որտեղ $A_i^C = \{c \in C \mid wt(c) = i\}$:

Սահմանում 1.10: C կոդի կշռային ֆունկցիա կանվանենք հետևյալ երկու փոփոխականներից կախված համասեռ բազմանդամը՝

$$W_C(x,y) = \sum_{i=0}^{n} A_i x^{n-i} y^i: \tag{1.4}$$

Դիտարկվում են նաև մեկ կամ երկուսից ավել փոփոխականներից կշռային ֆունկցիաներ: Մեկ փոփոխականից կշռային ֆունկցիան կնշանակենք $W_C(x) = \sum_{i=0}^{n} A_i x^i$: Նշանակենք $K_j^n(x)$-ով j-րդ աստիճանի Կրավչուկի բազմանդամը, ինչպես հայտնի է՝ $K_j^n(x) = \sum_{l=0}^{j} (-1)^l \binom{n-x}{j-l}\binom{x}{l}$, որտեղ $\binom{x}{m} = \frac{x(x-1)\dots(x-m+1)}{m!}$:

1.2. Էլիասի ալգորիթմը: Դիցուք $F \subseteq E^n$ որևէ ենթաբազմություն (ֆայլ) է, և $x \in E^n$ որը կոչվում է հարցում: Դիտարկենք F-ից x-ի «ամենամոտ հարևանների» գտնելու ալգորիթմական խնդիրը, որը F_x բազմությունը գտնելու խնդիրն է, որտեղ

$$F_x = \{y \in F / d(x,y) = min_{z \in F}\, d(x,z)\}: \tag{1.8}$$

Սահմանում 1.11: Հաշ-ֆունկցիա կանվանենք հետևյալ տիպի արտապատկերումը՝

$$\mathrm{h}: \mathrm{E}^{\mathrm{n}} \to V, \tag{1.9}$$

որտեղ V-ն վերջավոր N էլեմենտանոց բազմություն է՝ $V = \{v_1, \dots, v_N\}$: Հաճախ դիտարկվում են այնպիսի դեպքեր, երբ $V = E^k$, որտեղ $k \le n$: Որոշ դեպքերում հնարավոր է, որ $h(x_1) = h(x_2)$ երբ $\mathrm{x}_1 \neq \mathrm{x}_2$: Նման իրավիճակները կանվանենք հակասություններ (collision): Նշանակենք $B_i = \{x \in E^n \mid h(x) = v_i\}$ և թող $L_i = \{x \in F \mid h(x) = v_i\}$: ակնհայտ է, որ $\cup_{i=0}^{N} B_i = E^n$: Հաշ-կոդավորման սխեման կանվանենք բալանսավորված եթե $|B_i| = \frac{2^n}{N}$, $i = 1, \dots, N$: Հակասություններից խուսափելու համար կիրառվում է շղթաների մեթոդը, որի էությունը կայանում է հետևյալում. պահվում է N էլեմենտանոց զանգված (որի էլեմենտները համապատասխանում են հաշ ֆունկցիայի արժեքներին), որի յուրաքանչյուր էլեմենտ հիշողության ցուցիչ է (pointer) L_i բազմության վրա, որն էլ իր հերթին ներկայացված է որպես գծային ցուցակ: Ստորև բերված է ալգորիթմի կոդը՝ պայմանական լեզվով:

Elias Algorithm: comment: n is the word length, N is the number of blocks
input x, F, comment: $F \neq \emptyset$
integer $\delta = \infty$, comment: the current best match distance
set $S = \emptyset$, comment: S-is the current set of vectors of F located at distance δ from x
integer $j = -1$, comment: current distance of blocks under consideration from x
while($j < \delta$)
{
$j++$,
if(s(j) $\neq$ 0) comment: s(j) is the number of blocks in distance j from x
for($integer\ i = 0;\ i < s(j);\ i++$)
{

if($L_{j_i} \neq \emptyset$) comment: start examine the list L_{j_i}, i-th list with j distance block
if($\delta \leq d(x, L_{j_i})$ comment: δ is unchanged
$S = S \cup \left(O_\delta^n(x) \cap L_{j_i}\right)$ comment: $O_\delta^n(x)$ is the δ neighbourhood of x
else
{
$S = O_\delta^n(x) \cap L_{j_i}$, comment: δ is changed
$\delta = d(x, L_{j_i})$
}
}
}
return S, comment: F_x, $\delta = d(x, F)$

Որպես ալգորիթմի բարդության չափ կընդունենք դիտարկվող բլոկների միջին քանակը ըստ բոլոր x հարցումների և F ֆայլերի այն ենթադրությամբ, որ ֆայլերը կարող են հանդես գալ հավասարաչափ հավանականային բաշխմամբ:

1.3. Հեմմինգի կոդը և դրա հարակից դասերի կշռային սպեկտորները: $\mathcal{H}_m$ –ով նշանակենք $2^m - 1$ երկարությամբ Հեմմինգի կոդը: $\mathcal{H}_m$ կոդը ունի երկու տիպի հարակից դաս, որոնց սպեկտորներն են՝

$$A_j^{\mathcal{H}_m} = \frac{1}{2^m}\left(K_j^n(0) + (2^m - 1)K_j^n(2^{m-1})\right): \quad (1.11)$$

$$A_j^{e_i+\mathcal{H}_m} = \frac{1}{2^m}\left(\binom{2^m - 1}{j} - K_j^n(2^{m-1})\right), \text{ որտեղ } w(e_i) = 1:$$

1.4. Գոլեյի կոդի հարակից դասերի կշռային սպեկտորները: Սահմանենք Գոլեյի ընդլայնված կոդը, որը կնշանակենք Γ_{24}-ով: Դիցուք A_{11}-ը Պելիի տիպի Հադամարի մատրիցա է[7]: I_m-ով կնշանակենք $m \times m$ չափի միավոր մատրիցան: Ընդլայնված Գոլեյի կոդը՝ Γ_{24}-ը, կսահմանենք իր ծնող մատրիցայի միջոցով հետևյալ կերպ.

$$G_{\Gamma_{24}} = \begin{pmatrix} 1_{11}^T & I_{11} & 0_{11}^T & H_{11} \\ 0 & 0_{11} & 1 & 1_{11} \end{pmatrix},$$

որտեղ 0_m-ով և 1_m-ով նշանակված են m երկարության համապատասխանաբար լրիվ զրոներից և մեկերից կազմված վեկտորները: Գոլեյի կոդի հարակից դասերի կշռային սպեկտորները կգրենք՝ օգտվելով Լյոյդի թեորեմից:

1.5 Էլիասի ալգորիթմի բարդությունը կատարյալ կոդերի համար: Այժմ դիցուք ունենք C $[n, k, d]R$ կատարյալ կոդ: Կսահմանենք C-կոդի հետ ասոցիացված հաշ-ֆունկցիա հետևյալ կերպ՝

$$h_C(x) = y, \; d(x, y) = d(x, C) \quad (1.20)$$

(1.20)-ի կոռեկտությունը հետևում է C կոդի կատարյալ լինելուց. նշված y վեկտորը միշտ գոյություն ունի և միակն է: Նշանակենք

[7] F. J. Mac-Williams, N. J. Sloane, The theory of error-correcting codes, Amsterdam: N.-H. Mathematical Library, 762 p., 1977

$$V(j) = (1-(1-p)^{\binom{n}{j}})(1-p)^{\sum_{l=0}^{j-1}\binom{n}{l}}. \quad (1.24)$$

Պնդում 1.1: Էլիասի ալգորիթմի բարդության համար $[2^m-1, 2^m-m-1, 3]1$ $\mathcal{H}_m$ Հեմմինգի կոդի կոդի դեպքում տեղի ունի հետևյալը.

$$\alpha(h_{\mathcal{H}_m}) = \frac{1}{2^m}\sum_{0\le j\le 2^m-1} V(j)\left(\sum_{i=0}^{j+1}\left(A_i^{\mathcal{H}_m} + (2^m-1)A_i^{e_i+\mathcal{H}_m}\right)\right).$$

Պնդում 1.2: Էլիասի ալգորիթմի բարդության համար $[23,12,7]3$ Գոլեյի կոդի դեպքում տեղի ունի հետևյալը.

$$\alpha(h_{\Gamma_{23}}) = \sum_{0\le j\le 23} V(j)\sum_{i=0}^{j+3}\left(\frac{1}{2^{11}}A_i^0 + \frac{23}{2^{11}}A_i^1 + \frac{253}{2^{11}}A_i^2 + \frac{5819}{2^{11}}A_i^3\right),$$

որտեղ $A_0^j, A_1^j, \dots, A_n^j$-ով նշանակված են j մինիմալ կշիռ ունեցող հարակից դասերի կշռային սպեկտորները:

ԳԼՈՒԽ 2. ԷԼԻԱՍԻ ԱԼԳՈՐԻԹՄԻ ԳՈՐԾԱԾՈՒԹՅԱՆ ՏԻՐՈՒՅԹԻ ՀՆԱՐԱՎՈՐ ԸՆԴԼԱՅՆՈՒՄՆԵՐ

2.1 Կատարյալ կոդերի որոշ ընդհանրացումներ և հատվող բլոկներով հաշ կոդավորման սխեմաներ: Առաջին գլխում դիտարկեցինք մոտակա հարևանների փնտրման ալգորիթմ կատարյալ կոդերի հետ ասոցիացված հաշ ֆունկցիաների համար: Հայտնի է, որ կատարյալ կոդեր կարող են գոյություն ունենալ միայն պարամետրերի շատ սահմանափակ արժեքների դեպքում, որոնք են.

ա/ $\left(2^m-1, 2^{2^m-m-1}, 3\right)1$,

բ/ $(23, 2^{12}, 7)3$,

որտեղ ա/ պարամետրերի բազմությունը համապատասխանում է Հեմմինգի կոդի պարամետրերին, իսկ բ/-ն՝ Գոլեյի: Նկատի ունենալով սա և մոտակա հարևանների խնդրի կիրառական բնույթը՝ կդիտարկենք կատարյալ կոդերի որոշ ընդհանրացումների (որոնք ինչ-որ իմաստով մոտ են կատարյալ կոդերին) հետ ասոցիացված հաշ կոդավորման սխեմաներ և կհետազոտենք ալգորիթմի արդյունավետության խնդիրը նշված դեպքերում: Կատարյալ կոդերի հայտնի ընդհանրացումներ են համարյա կատարյալ կոդերը, հավասարաչափ փաթեթավորված կոդերը, քվազիկատարյալ կոդերը և այսպես կոչված p-շառավղով կատարյալ կոդերը:

Սահմանում 2.3: $C\ (n,k,d)R$ կոդը կանվանենք հավասարաչափ փաթեթավորված, եթե գոյություն ունեն $a_0, \dots, a_R$ թվեր այնպես, որ $\forall x \in E^n$ համար տեղի ունենա

$$\sum_{i=0}^{R} a_i A_i(x) = 1 \quad (2.12)$$

հավասարությունը:

Քվազիկատարյալ կոդեր: Համարյա կատարյալ կոդերի համար դրանց կոդային բառերի շուրջ կոդի փաթեթավորման շառավղով գնդերը չեն հատվում, և միևնույն

ժամանակ մեկով ավել շառավղով գնդերը հատվում են և ծածկում են միավոր խորանարդը: Սակայն նշված գնդերի հատումը կամայական չէ, ինչ-որ իմաստով ռեգուլյար է դա, և հնարավոր է դարձնում տալ դրա կոմբինատոր նկարագրությունը: Քանի որ դիտարկված կոդերի դասերը ևս գոյություն ունեն պարամետրերի ֆիքսած դասերի դեպքում, կդիտարկենք ևս մեկ ընդհանրացում:

Սահմանում 2.3: Դիցուք տրված է C $(n, M, d)R$ կոդը: Կասենք, որ C-ն քվազիկատարյալ կոդ է, եթե տեղի ունի $R_C = r_C + 1$ հավասարությունը:

Հատվող բլոկներով հաշ-կոդավորման սխեմաներ: Նախորդ գլխում դիտարկեցինք հաշ-կոդավորման սխեմաներ, որոնց բլոկները հանդիսանում էին միևնույն հզորությամբ չհատվող գնդեր: Այս գլխում դիտարկում ենք հաշ-կոդավորման սխեմաներ, որոնց դեպքում բլոկները հատվող, միևնույն հզորությամբ գնդեր են: Այդ դեպքում կատարյալ կոդերի վերը նշված ընդհանրացումները հնարավոր կլինի կիրառել Էլիասի ալգորիթմում: Ստորև կբերենք կոդի (հնարավոր է ոչ կատարյալ) հետ կապված հաշ-ֆունկցիայի սահմանում, որը ընդհանրացնում է սահմանում 1.20-ը:

Սահմանում 2.5: Դիցուք ունենք C $(n.M, d)R$ կոդ: Այդ դեպքում C կոդի միջոցով կսահմանենք հաշ-ֆունկցիա հետևյալ կերպ՝

$$h_C(x) = \{y/d(x, C) = d(x, y)\}: \tag{2.13}$$

2.2.2 Ընդլայնված Հեմմինգի կոդը և դրա հարակից դասերի կշռային սպեկտորները: Դիտարկենք ընդլայնված Հեմմինգի կոդը, որը կնշանակենք $\widehat{\mathcal{H}_m}$-ով: Հայտնի է, որ այն հանդիսանում է $[2^m, 2^m - m - 1,4]2$ քվազիկատարյալ հավասարաչափ փաթեթավորված կոդ:

$$A_j^{\widehat{\mathcal{H}_m}} = \frac{1}{2^{m+1}}\left(K_j^n(0) + (2^{m+1} - 2)K_j^n(2^{m-1}) + K_j^n(2^m)\right): \tag{2.17}$$

Այժմ դիտարկենք $y + \widehat{\mathcal{H}_\mathrm{m}}$, $y \in E^n \backslash \widehat{\mathcal{H}_\mathrm{m}}$ հարակից դասերի կշռային սպեկտորները գտնելու խնդիրը: Օգտվելով Մակ-Վիլյամսի թեորեմից՝ կստացվեն հետևյալ արդյունքները՝

$$A_j^{e_i + \widehat{\mathcal{H}_m}} = \frac{1}{2^{m+1}}\binom{2^m}{j}\left(1 - (-1)^j\right): \tag{2.18}$$

$$A_j^{l_i + \widehat{\mathcal{H}_m}} = \frac{1}{2^{m+1}}\left(\binom{2^m}{j}\left(1 - (-1)^j\right) - 2K_j^{2^m}(2^{m-1})\right) \tag{2.19}$$

Նկատի ունենալով 2.17 2.18, 2.19 բանաձևերը, ինչպես նաև հաշվելով համապատասխան հարակից դասերի քանակը՝ կստանանք՝

Պնդում 2.1. Էլիասի ալգորիթմի բարդության համար $[2^m, 2^m - m - 1,4]2$ $\widehat{\mathcal{H}_\mathrm{m}}$ ընդլայնված Հեմմինգի կոդի կոդի դեպքում տեղի ունի հետևյալը.

$$\alpha\left(h_{\widehat{\mathcal{H}_m}}\right) = \sum_{0 \le j \le 2^m} V(j)\left(\sum_{i=0}^{j+2}\left(\frac{1}{2^{m+1}}A_i^{\widehat{\mathcal{H}_m}} + \frac{1}{2}A_i^{e_i + \widehat{\mathcal{H}_m}} + \frac{2^m - 1}{2^{m+1}}A_i^{l_i + \widehat{\mathcal{H}_m}}\right)\right): \tag{2.20}$$

2.2.3 Կենտ m-ի համար 2^m-1 երկարությամբ երկու սխալ ուղղող ԲՉՀ կոդերի հարակից դասերի կշռային սպեկտորները: Նշանակենք F_q-ով (q-ն պարզ թվի աստիճան է) q էլեմենտ պարունակող վերջավոր դաշտը: Այսուհետև կդիտարկվեն միայն 2 բնութագրիչ ունեցող վերջավոր դաշտերը: Հիշեցնենք, որ դաշտի բնութագրիչ կոչվում է ամենափոքր p թիվը, այնպես որ $\forall a \in F_q$ տեղի ունի $\underbrace{(a+\cdots+a)}_{p\ անգամ}=0$: F_q դաշտի պրիմիտիվ էլեմենտը կնշանակենք α-ով: Նշանակենք F_q դաշտից գործակիցներով ֆորմալ բազմանդամների բազմությունը $F_q[x]$-ով: Դրա հետ կապված կդիտարկենք $R[x]=F_q[x]/(x^n-1)$ ֆակտոր օղակը, որը ինչպես հայտնի է, հանդիսանում է գլխավոր իդեալների օղակ՝ այսինքն յուրաքանչյուր I իդեալի համար գոյություն ունի $g(x)\in I$ այնպես, որ կամայական $a(x)\in I$ տարրի համար գոյություն ունի $f(x)\in R[x]$ այնպես, որ $g(x)=a(x)f(x)$: Այլ կերպ ասած՝ գլխավոր իդեալը հավասար է իր տարրերից մեկի և $R[x]$-ի տարրերի արտադրյալների բազմությանը: $(n,M,d)R$ պարամետրերով C-կոդը կանվանենք ցիկլիկ, եթե

1. C-ն գծային է,

2. $c=(c_1,\dots,c_n)\in C$ պայմանից հետևում է, որ $(c_n,c_1,\dots,c_{n-1})\in C$:

$(a_1,a_2,\dots,a_n)\in E$ վեկտորին կարելի է համապատասխանության մեջ դնել $a(x)=a_1+a_2x+\cdots+a_nx^{n-1}$ բազմանդամը, և այսպիսով յուրաքանչյուր կոդ հնարավոր կլինի դիտարկել որպես $R[x]$-ի ենթաբազմություն: Հայտնի է, որ ցիկլիկ կոդ լինելը համարժեք է $R[x]$-ի իդեալ լինելուն, և հետևաբար կամայական C $[n,k,d]R$ ցիկլիկ կոդի համար գոյություն ունի $g(x)$ ամենափոքր աստիճանի նորմավորված բազմանդամ, այնպես որ C-կամայական տարր՝ ներկայացված որպես $R[x]$-ի տարր, կարող է ներկայացվել $g(x)f(x)$ տեսքով որևէ $f(x)\in R[x]$-ի համար:

$n=2^m-1$ երկարությամբ երկու սխալ ուղղող ԲՉՀ կոդը կնշանակենք $\mathfrak{B}_m$-ով կսահմանենք ծնող բազմանդամի միջոցով, հետևյալ կերպ՝

$$g(x)=scm\{M_\alpha(x),M_{\alpha^3}(x)\}, \tag{2.21}$$

Որտեղ $M_{\alpha^i}(x)$-ով նշանակված է α^i տարրի մինիմալ բազմանդամը:

Հայտնի է, որ $n=2^m-1$ երկարությամբ երկու սխալ ուղղող ԲՉՀ կոդը հանդիսանում է $[2^m-1,2^m-2m-1,5]3$ կոդ, այսինքն՝ այն պատկանում է քվազիկատարյալ կոդերի դասին: Հայտնի է նաև, որ կենտ m-ի դեպքում $\mathfrak{B}_m$-ը նաև հավասարաչափ փաթեթավորված կոդ է, $a_0=a_1=1$ և $a_2=a_3=\frac{6}{n-1}$, իսկ դրա Լյոյդի բազմանդամի արմատներն են՝ $\xi_1=\frac{n+1}{2}-\sqrt{\frac{n+1}{2}}$, $\xi_2=\sqrt{\frac{n+1}{2}}$ և $\xi_3=\frac{n+1}{2}+\sqrt{\frac{n+1}{2}}$:

Պնդում 2.2. Էլիասի ալգորիթմի բարդության համար $[2^m-1,2^m-2m-1,5]3$ $\mathfrak{B}_m$ կոդի դեպքում տեղի ունի հետևյալը.

$$\alpha\left(h_{\mathfrak{B}_m}\right)= \tag{2.22}$$
$$=\sum_{0\le j\le 2^m-1}V(j)\sum_{i=0}^{j+3}\left(\frac{1}{2^{2m}}A_i^0+\frac{2^m-1}{2^{2m}}A_i^1+\frac{(2^m-1)(2^{m-1}-1)}{2^{2m}}A_i^2+\frac{2^{2m-1}+2^{m-1}-1}{2^{2m}}A_i^3\right),$$

ԳԼՈՒԽ 3. ԷԼԻԱՍԻ ԱԼԳՈՐԻԹՄԻ ԱՐԴՅՈՒՆԱՎԵՏՈՒԹՅԱՆ ՀԱՐՑԵՐ

3.1 Գնդերի հատումը Հեմմինգի մետրիկայում

3.1.2. Մի քանի գնդերի հատումը: Երկու գնդերի հատումը միավոր խորանարդում ունի պարզ նկարագիր, և դրա համար դուրս են գրվել հատումների ծավալների բանաձևերը: Հետաքրքրություն է ներկայացնում դիտարկել նաև երեք և ավելի գնդերի հատման և միավորման և դրանց հզորությունների աճման խնդիրները: Կդիտարկենք երեք և ավելի գնդերի հատման դեպքը, երբ դրանց կենտրոնները հանդիսանում են քվազիկատարյալ կոդի կոդային բառեր:

Պնդում 3.1. Կամայական R_C ծածկման շառավղով C քվազիկատարյալ կոդի k(k ≥ 3) կոդային բառերի շուրջ գծված R_C շառավղով գնդերի հատման հզորությունը միշտ ≤ 1:

3.2. Գնդերի միավորման՝ ըստ շառավղի աճման հզորության աճի ֆունկցիան

Դիցուք ունենք որևէ $C \in E^n$ ոչ դատարկ ենթաբազմություն: Նշանակենք.

$$f_C(i) = |\cup_{c\in C} S^n_{i+1}(c) \setminus \cup_{c\in C} S^n_i(c)|: \tag{3.12}$$

Մեզ անհրաժեշտ կլինի հետազոտել $f_C(i)$ ֆունկցիայի վարքը, մասնավորապես պարզել դրա մաքսիմումների քանակը: Նախ դիտարկենք այն դեպքը, երբ $C = \{\alpha, \beta\}$:

Պնդում 3.2: $f_{\{\alpha,\beta\}}(i)$ ֆունկցիան ունի մեկ կամ երկու մաքսիմում:

Պնդում 3.3: $f_C(i)$ ֆունկցիան ունի մեկ կամ երկու մաքսիմումի կետ ընդ որում այն դեպքում, երբ մաքսիմումի կետերը երկուսն են, ապա f_C-ն այն ընդունում է հարևան կետերի վրա:

3.3 Դիսկրետ իզոպերիմետրիկ խնդրի դրվածքը և լուծումը

Սահմանում 3.1: Դիցուք ունենք $A \subset E^n$ ենթաբազմությունը: $a \in A$ կետը կանվանենք A բազմության ներքին կետ, եթե նրա համար տեղի ունի $S^n_1(a) \in A$ պայմանը: Հակառակ դեպքում a կետը կկոչվի եզրային:

Սահմանում 3.2. Կասենք, որ A բազմությունը բերված է, եթե նրա համար տեղի ունի

$$A^0(i) \geq A^1(i) \tag{3.15}$$

անհավասարությունը կամայական $1 \leq i \leq n$ համար:

Լեմմ 3.3. Դիցուք A-ն E^n-ի կամայական ոչ-դատարկ, բերված ենթաբազմություն է, և

$$|A| = a = \sum_{i=0}^{k} \binom{n}{i} + \delta, \text{ որտեղ } 0 \leq k \leq n \text{ և } 0 \leq \delta < \binom{n}{k+1} \tag{3.16}$$

Այդ դեպքում գոյություն ունի j համար $1 \leq j \leq n$, այնպիսին, որ

$$|A^1(j)| \geq \sum_{i=0}^{k-1} \binom{n-1}{i} + \delta \frac{k+1}{n}: \tag{3.17}$$

Այժմ դիտարկենք կամայական a թիվ և այն ներկայացնենք (3.16) տեսքով: n-ից կախված ինդուկցիոն եղանակով կսահմանենք a հզորությամբ A բազմություն, որը կոչվում է ստանդարտ տեղաբաշխում հետևյալ կերպ՝

Սահմանում 3.2: Եթե գոյություն ունի $(x_1, \dots, x_n)$ վեկտոր, և փոփոխականների այնպիսի $x_{i_1}, \dots, x_{i_n}$ հաջորդականություն, որ.

ա) եթե $\delta \leq \binom{n-1}{k}$ ապա A բազմությունը բաղկացած է $E^{n-1}_{x_{j_1}}$ ենթախորանարդի $S^{n-1}_k(x)$ գնդից, և $E^{n-1}_{\bar{x}_{j_1}}$ ենթախորանարդի $\sum_{i=0}^{k-1} \binom{n-1}{i} + \delta$ հզորության ստանդարտ տեղաբաշխումից:

բ) եթե $\delta > \binom{n-1}{k}$ ապա A բազմությունը բաղկացած է $E^{n-1}_{\bar{x}_{j_1}}$ ենթախորանարդի $S^{n-1}_k(\bar{x}_{j_1})$ գնդից, և $E^{n-1}_{x_{j_1}}$ ենթախորանարդի $\sum_{i=0}^{k-1} \binom{n-1}{i} + \delta$ հզորության ստանդարտ տեղաբաշխումից:

Թեորեմ 3.2[8]: Կամայական a թվի համար, $0 \leq a \leq 2^n$, a հզորության ստանդարտ տեղաբաշխումը պարունակում է մինիմալ թվով ներքին կետեր:

Կառանձնացնենք թեորեմի մի մասնավոր դեպք՝ երբ a-ն r շառավղով գնդի հզորություն է, կամ որ նույնն է՝ ներկայացվում է $\sum_{i=0}^{r} \binom{n}{i}$ տեսքով:

Հետևանք 3.1: E^n-ում r շառավղով գունդը պարունակում է մինիմալ թվով եզրային կետեր:

3.4. Էլիասի ալգորիթմի օպտիմալությունը

Առաջին գլխում նկարագրված ալգորիթմի հետ կապված դիտարկվեցին հաշ-կողավորման սխեմաներ: Հիշեցնենք, որ դրանց հետ ասոցիացվեցին միավոր խորանարդի տրոհումներ չհատվող, հավասար հզորության բլոկների: Այս բաժնում կհետազոտենք ալգորիթմի «երկրաչափական» օպտիմալության հարցը, այսինքն կպարզենք, թե ինչ պայմանների դեպքում է ալգորիթմը օպտիմալ: Նշանակենք $\Phi(B_i) = \frac{1}{2^n}\sum_{x \in E^n} \Phi_x(B_i)$, որտեղ $\Phi_x(B_i)$-ն հավանականությունն է ըստ բոլոր ֆայլերի, որ x հարցման դեպքում B_i բլոկը կդիտարկվի: Ալգորիթմի բարդության համար ունենք, որ

$$\alpha(h) = \sum_{i=1}^{N} \Phi(B_i) \tag{3.20}$$

Թեորեմ 3.3: Եթե B_i-ն հանդիսանում է դիսկրետ իզոպերիմետրիկ խնդրի լուծում, ապա $\Phi(B_i)$ մեծությունը ընդունում է մինիմալ արժեք:

[8] Л. А. Асланян, Изопериметрическая задача и смежные экстремальные задачи для дискретных пространств, Проблемы кибернетики, Выпуск 36, с. 85-127, Москва 1979

Հիմնական արդյունքներն ու եզրահանգումները: Աշխատանքում ստացվել են հետևյալ հիմնական արդյունքները.

● կատարյալ կոդերի համար ստացվել են կոմբինատոր փնտրման Էլիասի ալգորիթմի բարդության անալիտիկ արտահայտություններ,

● ալգորիթմը դիտարկվել է հատվող բլոկներով հաշ-կոդավորման սխեմաների դեպքում (որտեղ բլոկները զնդեր են), և որոշ դեպքերի համար դուրս են բերվել ալգորիթմի բարդության բանաձևերը,

● ապացուցվել է, որ բալանսավորված հաշ-կոդավորման սխեմաների համար Էլիասի ալգորիթմը օպտիմալ է, երբ համապատասխան բլոկները դիսկրետ իզոպերիմետրիկ խնդրի լուծումներ են, մասնավորապես դրանք կարող են լինել զնդեր,

● նշված ալգորիթմը դիտարկվել է նաև այնպիսի բալանսավորված հաշ-կոդավորման սխեմաների համար, որոնց դեպքում բլոկները հանդիսանում են ավելի ցածր չափի զնդերի դեկարտյան արտադրյալներ, ինչը էականորեն ընդլայնում է ալգորիթմների դասը:

Ատենախոսության թեմայի շրջանակներում հրատարակված աշխատությունների ցանկ

L. H. Aslanyan, H. E. Danoyan, *Complexity of Elias algorithm based on codes with covering radius three,* Proceedings of the Yerevan state university, 2013 №1, pp. 44-50.

L. H. Aslanyan, H. E. Danoyan, *Complexity of Elias algorithm based on Hamming and extended Hamming codes,* Reports of NAS RA, vol. 113, no. 2, pp. 151-158, 2013.

H. E. Danoyan, *On Some Properties of Intersection and Union of Spheres in Hamming Metric*, Transactions of IIAP of NAS of RA, Mathematical Problems of Computer Science, vol. 39, pp. 119-124, 2013.

Резюме

Даноян Айказ Эдвардович

ИССЛЕДОВАНИЕ ДИСКРЕТНЫХ ЭКСТРЕМАЛЬНЫХ ЗАДАЧ

В работе рассматривается ранее предложенный алгоритм (алгоритм Элиаса), который предполагает использование второго метода. Балансированные схемы хеш-кодирования ассоциируются с покрытиями пространства Хемминга с непересекающимися подмножествами (блоками) равной мощности. В работе:

- Рассматривается эффективность данного алгоритма в случае кодов Хемминга и Голея. Получены аналитические выражения для эффективности алгоритма при использовании этих кодов.

  Поскольку совершенные коды существуют лишь в очень «узких» областях значений параметров, то:
- Рассматривается алгоритм в случаях схем хеш-кодирования, которые получаются разными обобщениями совершенных кодов, таких как квазисовершенные коды, равномерно упакованные коды, почти совершенные коды и т.д. Получены аналитические выражения для эффективности алгоритма в некоторых случаях, а именно для примитивных кодов БЧХ исправляющих двойные ошибки и для расширенных кодов Хемминга.
- Рассматривается задача оптимальности алгоритма в зависимости от геометрических свойств блоков (в случае балансированных схем хеш-кодирования). Доказано, что алгоритм является оптимальным, если отмеченные блоки изопериметрические множества, частным случаем которых являются сферы.
- Рассматривается также алгоритм в случае схем хеш-кодирования, когда блоки не пересекаются и являются декартовым произведением сфер более низких размерностей.

В работе рассматривается задача нахождения множества ближайших соседей к данному вектору из данного множества в метрике Хемминга. Эффективность решения данной задачи во многом зависит от представления множеств, в которой выполняется поиск. Одним из широко известных методов – представление данных в виде деревьев, а другой – хеширование.

Обозначим через E^n множество вершин n-мерного единичного куба. Пусть F есть некоторое подмножество (файл) E^n и $x \in E^n$. Через $d(x,y)$ обозначим расстояние Хемминга между векторами x и y. Рассмотрим задачу нахождения множества ближайших соседей F_x, то есть

$$F_x = \{y / d(x,y) = d(x,F)\},$$

где $d(x,C) = \min_{c \in C} d(x,c)$.

Определение. Хеш-функцией называется следующее отображение

$$h: E^n \to V,$$

где $V = \{v_1, \dots, v_N\}$ - некоторое конечное множество. Обычно рассматриваются случаи, когда $V = E^k$, $k < n$. Возможны ситуации, когда $u \neq v$, но $f(u) = f(v)$: Такие ситуации

называются коллизии. Коллизии решаются методом цепочек. Оно состоит в следующем: поддерживаются N связанных списков по одному на каждый возможный хеш-адрес. В каждом списке хранятся те элементы файла F, на которых хеш-функция принимает равные значения. Обозначим $B_i = \{x/h(x) = v_i\}$ и $L_i = F \cap B_i \;\; i = 1, \dots, N$. Схема хеш-кодирования называется сбалансированным, если $|B_i| = \frac{2^n}{N}$. Очевидно, что подмножества (блоки) $B_i \;\; i = 1, \dots, N$ покрывают единичный куб. В частности, такие покрытия могут получаться с помощью совершенных кодов. Предполагая, что файл представлен отмеченным образом, П. Элиас предложил алгоритм поиска множества ближайших соседей. Под эффективностью алгоритма для хеш-функции h понимается среднее число рассмотренных списков полагая, что каждый вектор z из E^n может принадлежать множесву F вероятностью p.

Определение. Весовым спектором кода C называются числа $A_0^C, A_1^C, \dots, A_n^C$, где $A_i^C = \{c \in C \mid wt(c) = i\}$.

Утверждение 1.1. Для $[2^m - 1, 2^m - m - 1, 3]1$ кода Хемминга $\mathcal{H}_m$ эффективность алгоритма Элиаса равно

$$\alpha\left(h_{\mathcal{H}_m}\right) = \frac{1}{2^m}\sum_{0 \le j \le 2^m - 1} V(j)\left(\sum_{i=0}^{j+1}\left(A_i^{\mathcal{H}_m} + (2^m - 1)A_i^{e_i + \mathcal{H}_m}\right)\right),$$

где e_i произвольный вектор имеющий вес 1.

Утверждение 1.2. Для $[23, 12, 7]3$ кода Голея эффективность алгоритма Элиаса равно

$$\alpha\left(h_{\Gamma_{23}}\right) = \sum_{0 \le j \le 23} V(j) \sum_{i=0}^{j+3}\left(\frac{1}{2^{11}}A_i^0 + \frac{23}{2^{11}}A_i^1 + \frac{253}{2^{11}}A_i^2 + \frac{5819}{2^{11}}A_i^3\right),$$

где $A_0^j, A_1^j, \dots, A_n^j$ весовой спектор кода Голея имеющий минимальный вес j.

Обозначим через $S_r^n(x)$ сферу радиуса r с центром x, то есть $S_r^n(x) = \{y \in E^n \,/\, d(x, y) \le r\}$.

Определение. Скажем, что точка $x \in C$ является внутренней точкой множества C, если $S_1^n(x) \in E^n$.

Теперь рассмотрим задачу нахождения множества данной мощности a, которое имеет максимальное число внутренних точек. Множества, обладающими отмеченными свойствами, назовем изопериметрическими. Л. А. Асланяном доказано, что изопериметрическими являются так называемые «стандартные размещения», частным случаем которых являются сферы.

Рассмотрим задачу оптимальности алгоритма для сбалансированных схем хеш-кодирования. Оно заключается в следующем: каким условиям должны удовлетворять блоки, чтобы среднее число рассматриваемых алгоритмом блоков среди всех файлов и запросов было минимальным? Обозначим $\Phi(B_i) = \frac{1}{2^n}\sum_{x \in E^n} \Phi_x(B_i)$, где $\Phi_x(B_i)$-средняя вероятность по всем файлам, что в случае запроса x блок B_i будет рассмотрен. Для сложности алгоритма имеется, что

$$\alpha(h) = \sum_{i=1}^{N} \Phi(B_i).$$

Теорема 3.3: Если B_i- является решением дискретной изопериметрической задачи, то величина $\Phi(B_i)$ принимает минимальное значение.

Abstract

Haykaz E. Danoyan

RESEARCH OF DISCRETE EXTREME PROBLEMS

In this dissertation the problem of finding of the set of the nearest neighbors to a given vector from a given set in Hamming metric is considered. The efficiency of the solution to this problem in many respects depends on structural representation of sets in which search is carried out. One of the widely known approaches is the data representation in the form of trees, and the other alternative one is the use of hashing technique. Dissertation considers the earlier proposed algorithm (Elias's algorithm), which assumes the second approach. The balanced hash-coding schemes are associated with the coverings of a unit cube by non-intersecting subsets (blocks) of equal power.

- The efficiency of the algorithm in case of Hamming and Golay codes is considered. Analytical expressions for efficiency of algorithm for these cases are obtained.
- As perfect codes exist only for very "narrow" domain of parameters then the algorithm is considered special hash-coding schemes obtained by different generalizations of perfect codes such as quasi-perfect codes, uniformly packed codes, nearly perfect codes, etc. Analytical expressions for efficiency of algorithm in certain cases are obtained namely for the primitive double error correcting BCH codes and for the extended Hamming codes.
- The problem of optimality of the algorithm depending on geometrical properties of blocks (in case of balanced hash-coding schemes) is considered. It is proved that the algorithm is optimum, if the mentioned blocks are isoperimetric sets, the special cases of which are spheres.
- The algorithm is considered also in case of hash coding schemes when the blocks do not intersect and they are a Cartesian product of spheres of lower dimensions.

Denote by E^n the set of vertices of the n-dimensional unit cube. Let F be a subset (or file) of E^n and $x \in E^n$ (query element). Denote by $d(x,y)$ the Hamming distance between the vectors x and y. Let us consider the problem of finding of the set of the nearest neighbors F_x, i.e.

$$F_x = \{y / d(x,y) = d(x,F)\},$$

where $d(x,C) = \min_{c \in C} d(x,c)$.

Definition. We will call a hash function the following mapping:

$$h: E^n \to V,$$

where $V = \{v_1, \dots, v_N\}$ is a finite set. Usually there are considered cases when $V = E^k$, $k < n$. There are possible situations when $u \neq v$ but $f(u) = f(v)$. Such situations are called collisions. Collisions are handled by the method of "Chainings". The method is to keep N distinct linked lists, one for each possible hash value. In each list those vectors of the file F are stored on which the hash function h takes equal values. Let us denote $B_i = \{x / h(x) = v_i\}$ and $L_i = F \cap B_i$ $i = 1, \dots, N$. The hash coding schema will be called balanced if $|B_i| = \frac{2^n}{N}$. It is obvious that, the subsets (blocks) B_i $i = 1, \dots, N$ cover the unit cube. Particularly such coverings can be obtained via perfect codes. Considering that the file is represented by the mentioned way, P. Elias proposed the algorithm of the search of the set of the nearest neighbors. Under complexity of

the algorithm for a hash function h we mean the average number of the considered lists, assuming that each vector z from E^n can belong to F by probability p.

Definition. The weight spectra of the code C are called the numbers $A_0^C, A_1^C, \dots, A_n^C$, where $A_i^C = \{c \in C \mid wt(c) = i\}$.

Proposition 1.1. For the $[2^m - 1, 2^m - m - 1, 3]1$ Hamming code $\mathcal{H}_m$ the complexity of the Elias algorithm is

$$\alpha\left(h_{\mathcal{H}_m}\right) = \frac{1}{2^m}\sum_{0\le j\le 2^m-1} V(j)\left(\sum_{i=0}^{j+1}\left(A_i^{\mathcal{H}_m} + (2^m-1)A_i^{e_i+\mathcal{H}_m}\right)\right),$$

where e_i is any vector having the weight 1.

Proposition 1.2: For the $[23,12,7]3$ Golay code Γ_{23} the complexity of the Elias algorithm is

$$\alpha\left(h_{\Gamma_{23}}\right) = \sum_{0\le j\le 23} V(j)\sum_{i=0}^{j+3}\left(\frac{1}{2^{11}}A_i^0 + \frac{23}{2^{11}}A_i^1 + \frac{253}{2^{11}}A_i^2 + \frac{5819}{2^{11}}A_i^3\right),$$

where $A_0^j, A_1^j, \dots, A_n^j$ are the weight spectra of the coset of the Golay code having a minimum weight j.

Denote by $S_r^n(x)$ the sphere of radius r centered at x, i.e. $S_r^n(x) = \{y \in E^n \,/\, d(x,y) \le r\}$.

Definition. We will say that a point $x \in C$ is a inner point of C if $S_1^n(x) \in E^n$.

Now let us consider the problem of finding of the set of a given cardinality which has the maximum number of inner points. Sets which have the mentioned property are called isoperimetric. L. H. Aslanyan proved that such sets are so called "standard placements" the separate cases of which are spheres.

Consider the problem of the optimality of the algorithm. The problem is: which conditions should satisfy the blocks that the average number of considered lists over all files and queries were minimum? Denote by $\Phi(B_i) = \frac{1}{2^n}\sum_{x\in E^n}\Phi_x(B_i)$, where $\Phi_x(B_i)$- the average probability by all files in case of query x that the block B_i will be considered.

For the complexity of the algorithm we have that

$$\alpha(h) = \sum_{i=1}^{N}\Phi(B_i).$$

Theorem 3.3: If B_i is an isoperimetric set then the $\Phi(B_i)$ takes the minimum value.

Ծավալը՝ 20 էջ: Տպաքանակը՝ 100:
ՀՀ ԳԱԱ ԻՄՊԻ կոմպյուտերային պոլիգրաֆիայի լաբորատորիա:
Երևան, Պ. Սևակի 1