ՀՀ ԳԱԱ ԻՆՖՈՐՄԱՏԻԿԱՅԻ ԵՎ ԱՎՏՈՄԱՏԱՑՄԱՆ
ՊՐՈԲԼԵՄՆԵՐԻ ԻՆՍՏԻՏՈՒՏ

Տիգրան Արշակի Գրիգորյան

Կանոնավոր արտահայտություններ՝ բազմաժապավեն վերջավոր ավտոմատների լեզուների համար

Ե.13.04 «Հաշվողական մեքենաների, համալիրների, համակարգերի և ցանցերի մաթեմատիկական և ծրագրային ապահովում» մասնագիտությամբ ֆիզիկամաթեմատիկական գիտությունների թեկնածուի գիտական աստիճանի հայցման ատենախոսության

ՍԵՂՄԱԳԻՐ

ԵՐԵՎԱՆ 2021

INSTITUTE FOR INFORMATICS AND AUTOMATION
PROBLEMS OF NAS RA

Tigran Grigoryan

Regular Expressions for Languages of Multitape Finite Automata

SYNOPSIS

of the thesis for the degree of candidate of physical and mathematical sciences in the specialty 05.13.04 – “Mathematical and Software Support of Computers, Complexes and Computer Networks”

Yerevan 2021

Ատենախոսության թեման հաստատվել է Երևանի պետական համալսարանում

Գիտական ղեկավար՝	ֆիզ.-մաթ. գիտ. դոկտոր Ս. Կ. Շուքուրյան
Պաշտոնական ընդդիմախոսներ՝	ֆիզ.-մաթ. գիտ. դոկտոր Լ. Հ. Ասլանյան, ֆիզ.-մաթ. գիտ. թեկնածու Ա. Հ. Կոստանյան
Առաջատար կազմակերպություն՝	Հայ-ռուսական համալսարան

Պաշտպանությունը կկայանա 2021թ. դեկտեմբերի 17-ին, ժ. 13^{00}-ին ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտում գործող ԲՈԿ-ի 037 մասնագիտական խորհրդի նիստում (հասցե՝ 0014, Երևան, Պ. Սևակի փ. 1):

Ատենախոսությանը կարելի է ծանոթանալ ՀՀ ԳԱԱ ԻԱՊԻ-ի գրադարանում:
Սեղմագիրն առաքված է 2021թ. նոյեմբերի 8-ին:

037 մասնագիտական խորհրդի գիտական քարտուղար,
ֆիզ.-մաթ. գիտ. դոկտոր Հ. Գ. Սարուխանյան

The topic of the thesis was approved in Yerevan State University

Scientific advisor	Doctor of Phys.-math. Sciences S. K. Shoukourian
Official opponents	Doctor of Phys.-math. Sciences L. H. Aslanyan, Candidate of Phys.-math. Sciences A. H. Kostanyan
Leading organization	Russian-Armenian University

The defense will be held on December 17, 2021 at 13:00 at a meeting of the specialized council 037, operating at the Institute for Informatics and Automation Problems of NAS RA (address: 0014, 1 P. Sevak St., Yerevan).

The thesis can be found at the IIAP NAS RA library.
The synopsis was sent on November 8, 2021.

Scientific secretary of specialized council 037
Doctor. of Phys.-math. Sciences H. G. Sarukhanyan

Աշխատանքի ընդհանուր բնութագիրը

Թեմայի արդիականությունը

Կանոնավոր արտահայտությունները առկա են բազում ծրագրավորման լեզուներում, ընդ որում ոմանց մեջ ներդրված են շարահյուսական մակարդակում: Դրանք լայնորեն կիրառվում են տվյալների որոնման և նմուշի հետ համապատասխանեցման խնդիրներում: Սակայն, մի շարք խնդիրներում կարիք կա դիտարկելու տառերի կամ բառերի միջև տեղափոխելիության հնարավորությունը: Բառերի տեղափոխելիությունը կարելի է բերել տառերի տեղափոխելիությանը՝ բառերը դիտարկելով որպես այբուբենի սիմվոլներ: Ստորև ներկայացված են մի շարք խնդիրներ, որոնցում կարիք կա դիտարկելու տառերի կամ բառերի միջև տեղափոխելիություն.

Բիոինֆորմատիկայում գենոմային շղթաների անալիզը 1990-ական թվականներից սկսած, հասանելի տվյալների քանակին զուգահեռ սրընթաց աճ է գրանցել: Գենոմային շղթաների անալիզում կիրառվող հավասարեցումով մեթոդներում հաշվի չեն առնվում շղթայում տեղափոխելի սեգմենտների առկայությունը: Վերջիններիս գոյությունը ապացուցված և հիմնավորված փաստ է: Այդ իսկ պատճառով կիրառվում է մեթոդների մեկ այլ դաս՝ հավասարեցումից անկախ մեթոդները, որոնք հաշվի են առնում տեղափոխելի սեգմենտները և որոնց հիմքում ընկած է շղթաների միջև հեռավորության հաշվումը[1]:

Եվս մեկ նման խնդիր է բիոինֆորմատիկայում հանդիսանում սպիտակուցների կառուցվածքում ամինաթթուների շրջանաձև կարգավորությունների դիտարկումը: Հետազոտությունները ցույց են տվել, որ մի շարք նման կարգավորությունների արդյունքում սպիտակուցը պահպանում է իր կենսաբանական հատկությունները: Անհրաժեշտություն է առաջանում տվյալների բազաներում որոնման ընթացքում հաշվի առնել շրջանաձև կարգավորությունների առակայությունը[2]:

Հաջորդ խնդիրը կապված է գրագողության բացահայտման հետ: Հայտնի են գրագողության մի շարք տեխնիկաներ, որոնցից է նախադասությունում բառերի կամ բառակապակցությունների հերթականության փոփոխությունը[3]: Որոշ լեզուներում, մասնավորապես՝ սեմական, մի շարք խոսքի մասեր նախադության մեջ կարող են հայտնվել տարբեր դիրքերում[4]: Այսպիսի լեզուների համար տվյալների որոնման և նմուշի հետ համապատասխանեցման խնդիրներ դիտարկելիս կարևոր է հաշվի առնել նախադասությունների

---

[1]Susana Vinga and Jonas S. Almeida. Alignment-free sequence comparison-a review. Bioinform. 19(4), pp. 513–523, 2003.

[2]Wei-Cheng Lo and Ping-Chiang Lyu. Cpsarst: an efficient circular permutation search tool applied to the detection of novel protein structural relationships. Genome Biology 9(1), pp. 1-16, 2008.

[3]Salha Alzahrani, Naomie Salim, and Ajith Abraham. Understanding plagiarism linguistic patterns, textual features, and detection methods. IEEE Trans. Systems, Man, and Cybernetics, Part C, 42(2), pp. 133–149, 2012.

[4]Shuly Wintner. Morphological processing of semitic languages. In Imed Zitouni, editor, Natural Language Processing of Semitic Languages, Theory and Applications of Natural Language Processing, pages 43–66. Springer, 2014.

ավելի ազատ կազմության հանգամանքը:

Դիտարկված խնդիրների լուծման համար որպես մոդել կարելի է դիտարկել բազմաժապավեն վերջավոր ավտոմատները, որոնց տարբեր ժապավենների տառերը տեղափոխելի են: Այդ իսկ պատճառով, կարիք կա բազմաժապավեն վերջավոր ավտոմատների համար նշանագրություն ներմուծելու: Եթե որպես նշանագրություն հնարավոր լինի կիրառել կանոնավոր արտահայտությունների արդեն իսկ հայտնի գրելաձևը, ապա մեկ ժապավենանի ավտոմատների և կանոնավոր արտահայտությունների մի շարք հատկություններ կարելի կլինի ընդլայնել բազմաժապավեն ավտոմատների դեպքի համար:

Վերը նշված խնդիրներում կարևոր է դիտարկվող բառերի կամ բառերի բազմությունների միջև հեռավորության հաշվումը, այդ իսկ պատճառով, բազմաժապավեն վերջավոր ավտոմատների կողմից ճանաչվող լեզուների համար մետրիկական տարածություն սահմանելու կարիք կա:

Նշենք բազմաժապավեն վերջավոր ավտոմատների համար նշանագրություն կամ կանոնավոր արտահայտություն սահմանելու խնդրին վերաբերող մի շարք արդյունքներ.

1966թ.-ին Բ. Միրկինի կողմից դիտարկվել է բազմաժապավեն վերջավոր ավտոմատների համար հատուկ կոդավորում[1]: Փորձ է արվել այդ կոդավորման միջոցով լուծելու բազմաժապավեն վերջավոր ավտոմատների անալիզի, սինթեզի և այլ խնդիրներ: Սակայն, իր աշխատությունում Միրկինն ապացույցները բերում է միայն 2 ժապավենի դեպքում առանց ցույց տալու, որ դրանք կարելի է ընդհանրացնել 2-ից ավել ժապավենների համար: Ավելին՝ հետագա հետազոտությունները ցույց են տվել, որ 2-ից մեծ թվով ժապավենների դեպքում կոդավորումը չի հանգեցնում ստացված արդյունքին:

Մեկ այլ նշանագրություն դիտարկվել է Պ. Ստարկեի կողմից: Իր 1975թ.-ի հոդվածում ստացվել է հետևյալ արդյունքը. «$W(X)$-ի վրա n չափանի R հարաբերությունը բերելի է վերջավոր դետերմինացված n-ժապավենանի ավտոմատի, եթե գոյություն ունի T թույլատրելի կանոնավոր արտահայտություն, այնպես որ $R = Val_n(T)$»[2]: Սակայն հենց նույն հոդվածում հեղինակը նշում է, որ գոյություն ունեն ոչ թույլատրելի T կանոնավոր արտահայտություններ, այնպիսիք որ $Val_n(T)$-ն կարելի է ներկայացնել դետերմինացված ավտոմատով:

1980թ.-ին Ա. Գոդլևսկու, Ա. Լետիչևսկու և Ս. Շուքուրյանի կողմից առաջարկվել է ազատ մասնակի տեղափոխական կիսախմբերի հատուկ երկուական ներկայացմամբ կոդավորում[3] չվերասերված օպերատորներով ծրագրերի սխեմաների համարժեքության

---

[1]B.G. Mirkin, On the theory of multitape automata, Cybernetics 2(5), pp. 9–14, 1966.

[2]P.H. Starke, On the representability of relations by deterministic and nondeterministic multi-tape automata, In Jirí Becvár, editor, Mathematical Foundations of Computer Science 1975, 4th Symposium, Mariánské Lázne, Czechoslovakia, September 1-5, 1975, Proceedings, volume 32 of Lecture Notes in Computer Science, pp. 114–124. Springer, 1975.

[3]A. B. Godlevskii, A. A. Letichevskii, and Samvel K. Shukuryan. Reducibility of program-scheme functional equivalence on a nondegenerate basis of rank unity to the equivalence of automata with multidimensional tapes. Cybernetics 16(6),

խնդրի լուծման համար: Հետագայում այս կոդավորումն օգտագործվել է բազմաժապավեն վերջավոր ավտոմատների մի շարք չլուծված խնդիրների լուծման համար[1][2]:

2020թ.-ին դետերմինացված բազմաժապավեն վերջավոր ավտոմատների համարժեքությունը ստուգող բազմանդամային արագությամբ ալգորիթմի[3] առաջարկումից հետո, բազմաժապավեն ավտոմատների համար կանոնավոր արտահայտությունների ներմուծումը դառնում է ավելի արդիական:

Աշխատանքի նպատակը

Աշխատանքի նպատակն է բազմաժապավեն վերջավոր ավտոմատների կողմից ճանաչվող լեզուների համար ներմուծել կանոնավոր արտահայտության և կանոնավոր պատահույթի գաղափարները՝ պահպանելով դասական կանոնավոր արտահայտությունների գրելաձևը և մի շարք հատկություններ: Հետազոտել ներմուծված կանոնավոր արտահայտությունները և պատահույթները: Ներմուծված կանոնավոր պատահույթների համար սահմանել մետրիկա և մետրիկական տարածություն, առաջարկել մետրիկան հաշվարկող ալգորիթմ:

Հետազոտման առարկան

Հետազոտման առարկա են հանդիսանում բազմաժապավեն վերջավոր ավտոմատները, նրանց համար սահմանված կանոնավոր արտահայտությունները և կանոնավոր պատահույթները:

Հետազոտման մեթոդները

Աշխատանքում օգտագործվել են ավտոմատների տեսության և հանրահաշվական (կիսախմբերի տեսության) մեթոդներ:

Գիտական նորույթը

Բոլոր հիմնական արդյունքները նոր են: Դրանք են՝

pp. 793–799, 1980.

[1] Alexander A. Letichevsky, Arsen S. Shoukourian, and Samvel K. Shoukourian. The Equivalence Problem of Deterministic Multitape Finite Automata: A New Proof of Solvability Using a Multidimensional Tape. In Adrian-Horia Dediu, Henning Fernau, and Carlos Martín-Vide, editors, Language and Automata Theory and Applications, 4th International Conference, LATA 2010, Trier, Germany, May 24-28, 2010. Proceedings, volume 6031 of Lecture Notes in Computer Science, pp. 392–402, Springer, 2010.

[2] Hayk A. Grigoryan and Arsen S. Shoukourian. Equivalence of Processes in Partially Commutative Object-Oriented Environments. Fundamenta Informaticae 105(4), pp. 417–434, 2010.

[3] H.A. Grigoryan, S.K. Shoukourian. Polynomial Algorithm for Equivalence Problem of Deterministic Multitape Finite Automata, Theor. Comput. Sci. 833, pp. 120–132, 2020.

- Ներմուծվել են կանոնավոր պատահույթ և կանոնավոր արտահայտություն գաղափարները բազմաժապավեն վերջավոր ավտոմատների համար:
- Ապացուցվել է, որ սահմանված կանոնավոր պատահույթների հանրահաշիվը Քլինիի հանրահաշիվ է:
- Ապացուցվել է, որ եթե այդ պատահույթների հանրահաշվում գծային հավասարումների համակարգն ունի միակ լուծում, ապա այն կարելի է գտնել փոփոխականների արտաքսման եղանակով:
- Ապացուցվել է, որ օգտագործվող կոդավորումից բխում է հետքերի մոնոիդների նոր բնութագրիչ բազմություն:
- Ապացուցվել է, որ որոշակի պայմանների դեպքում նոր բնութագրիչի կիրառությունից բխում է հետքերի համարժեքության ավելի արագ ալգորիթմ:
- Լուծվել են բազմաժապավեն վերջավոր ավտոմատների սինթեզի և անալիզի խնդիրները:
- Կանոնավոր պատահույթների մի շարք մետրիկաներ ընդլայնվել են բազմաժապավեն վերջավոր ավտոմատների կանոնավոր պատահույթների վրա: Ցույց է տրվել, որ որոշակի դեպքերում ընդլայնված մետրիկաները տարածությունն ադեկվատ չեն ներկայացնում:
- Ներմուծվել են բազմաժապավեն վերջավոր ավտոմատների համար կանոնավոր պատահույթների երկու նոր մետրիկաներ: Մշակվել է այդ մետրիկաները հաշվարկող երկու մոտավոր ալգորիթմ:

Տեսական և կիրառական արժեքը

Ստացված արդյունքները ներկայացնում են տեսական հետաքրքրություն բազմաժապավեն ավտոմատների հետագա հետազոտության համար, իսկ այլընտրանքային բնութագրիչը և առաջարկված ալգորիթմները կարող են կիրառվել տվյալների որոնման և նմուշի հետ համապատասխանեցման խնդիրներում, երբ անհրաժեշտ է դիտարկել որոշ տառերի կամ բառերի տեղափոխականությունը:

Ապրոբացիա և հրապարակումներ

Ատենախոսության հիմնական արդյունքներն ու դրույթները քննարկվել և զեկուցվել են ՏՏ կրթական և հետազոտական կենտրոնի ընդհանուր սեմինարում (2021թ.) և ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտի ընդհանուր սեմինարում (2021թ.):

Աշխատանքի հիմնական արդյունքները հրապարակված են 4 հոդվածներում, որոնց ցանկը բերված է սեղմագրի վերջում:

Աշխատանքի կառուցվածքն ու ծավալը

Ատենախոսությունը բաղկացած է ներածությունից, երեք գլխից, ամփոփումից, երկու հավելվածից և օգտագործված գրականության ցանկից: Աշխատանքի ծավալը 82 էջ է: Օգտագործված գրականության ցանկն ընդգրկում է 47 անուն: Աշխատանքը ներառում է 7 նկար և 2 աղյուսակ:

Աշխատանքի բովանդակությունը

Ներածությունում հիմնավորված է արդիականությունը, ձևակերպված է աշխատանքի նպատակը: Բերված է աշխատանքի համառոտ նկարագրությունն ըստ գլուխների:

Առաջին գլխում դիտարկված են ազատ մասնակի տեղափոխական մոնոիդները և դրանց կապը բազմաժապավեն վերջավոր ավտոմատների հետ: Սահմանված են կանոնավոր արտահայտությունների և կանոնավոր պատահույթների գաղափարները ազատ մասնակի տեղափոխական մոնոիդների համար: Դիտարկված և լուծված են հետևյալ խնդիրները. կանոնավոր պատահույթների հանրահաշվի վրա գծային հավասարումների համակարգերի լուծումը, հետքերի մոնոիդների այլընտրանքային բնութագրիչ բազմության ներմուծումը և բազմաժապավեն վերջավոր ավտոմատների սինթեզի և անալիզի խնդիրները:

1.1-ում սահմանվում են ազատ մասնակի տեղափոխական մոնոիդի գաղափարը և նրա տարրերի հատուկ՝ երկուական կոդավորմամբ ներկայացումը: Ներկայացվում են այդ կոդավորումից բխող մի շարք հայտնի արդյունքներ: Դիտարկվում է ազատ մասնակի տեղափոխական մոնոիդների կապը բազմաժապավեն վերջավոր ավտոմատների հետ:

Դիցուք Y-ն այբուբեն է: Y այբուբենում բոլոր բառերի բազմությունը, ներառյալ դատարկ բառը, կնշանակենք Y^*-ով, իսկ բառերի բոլոր n էլեմենտանոց կորտեժների բազմությունը՝ $(Y^*)^n$-ով:

Դիցուք $G = \langle Y, R \rangle$-ը $Y = \{y_1, y_2, \dots, y_n\}$ ծնիչներով և R առնչություններով ծնված մոնոիդ է: G-ն կոչվում է ազատ մասնակի տեղափոխական մոնոիդ, եթե R-ը վերջավոր է և կազմված $y_i y_j = y_j y_i \;\; (i \neq j)$ տիպի առնչություններից:

Նշանակենք $I_Y \subseteq Y \times Y$-ով բոլոր $(y, y'), y \neq y'$ տեղափոխելի տառերի զույգերի բազմությունը, և $D_Y \subseteq Y \times Y$-ով՝ բոլոր $(y, y'), y \neq y'$ ոչ տեղափոխելի տառերի զույգերի բազմությունը:

Կասենք, որ $w_1 \in Y^*$ և $w_2 \in Y^*$ բառերն իրար համարժեք են՝ $w_1 \rho w_2$, եթե $w_1 = w_2$ որպես G մոնոիդի էլեմենտներ: Սահմանված ρ համարժեքության հարաբերությունը Y^*-

ը բաժանում է համարժեքության դասերի, որոնք կանվանենք տեղափոխականության դասեր: w բառի համարժեքության դասը կնշանակենք $[w]$-ով:

Հետքերի տեսության[1] մեջ տեղափոխականության դասերը կոչվում են հետքեր, իսկ ρ հարաբերությունից բխող քանորդ մոնոիդը՝ հետքերի մոնոիդ: Վերջինս նշանակվում է $M(Y, I_Y)$-ով:

$K : Y^* \to (\{0,1\}^*)^n$ հոմոմորֆիզմն, որն Y^* բազմության բառերն արտապատկերում է $\{0,1\}$ բինար այբուբենի n էլեմենտանոց կորտեժների վրա, սահմանվում է հետևյալ կերպ.

$$K(y_i) = (a_{1i}, \ldots, a_{ni}),\ \text{where } a_{ij} = \begin{cases} 1, & i = j \\ e, & y_i y_j = y_j y_i \\ 0, & y_i y_j \neq y_j y_i \end{cases}$$

$K(e) = (e, \ldots, e)$, որտեղ e-ն դատարկ բառն է, իսկ $K(y_i y_j)$-ն սահմանվում է որպես ձախ կոնկատենացիա (այսուհետ ձախ կոնկատենացիան կանվանենք ուղղակի կոնկատենացիա).

$$K(y_i)K(y_j) = K(y_i y_j) = (a_{1j}a_{1i}, \ldots, a_{nj}a_{ni}):$$

Լեմմա 1 [2] Կամայական n ծնիչից ծնված ազատ մասնակի տեղափոխական մոնոիդ իզոմորֆ է n հատ 2 ծնիչից ծնված ազատ մոնոիդների Դեկարտյան արտադրյալի ինչ-որ ենթամոնոիդի:

Լեմմա 1-ը թույլ է տալիս մոնոիդի տարրերի փոխարեն օգտագործել իրենց երկուական կոդավորումները: Ստացված երկուական կոդավորմամբ բառերը կարելի է դիտարկել նաև որպես թվեր, իսկ որպես այդ թվերի կոնկատենացիա՝ դիտարկել դրանց երկուական կոդավորումների կոնկատենացիայի թվային արժեքը:

Լեմմա 2 [3] Կամայական n ծնիչից ծնված ազատ մասնակի տեղափոխական մոնոիդ իզոմորֆ է $\mathbb{N}^n$-ում ինչ-որ մոնոիդի, որտեղ մոնոիդային գործողությունը թվային կորտեժների կոնկատենացիան է ($\mathbb{N}$-ը բոլոր բնական թվերն ու 0-ն պարունակող բազմությունն է):

Աշխատանքում դիտարկվում է բազմաժապավեն վերջավոր ավտոմատի երկու մոդել.

[1] Volker Diekert and Yves Métivier. Partial commutation and traces. In Grzegorz Rozenbergand Arto Salomaa, editors, Handbook of Formal Languages, Volume 3: Beyond Words,pages 457–533. Springer, 1997.

[2] A. B. Godlevskii, A.A. Letichevskii, S.K. Shukuryan, Reducibility of program scheme functional equivalence on a nondegenerate basis of rank unity to the equivalence of automata with multidimensional tapes, Cybern. Syst. Anal. 16(6), pp. 793–799, 1980.

[3] H. A. Grigoryan, S. K. Shoukourian, Polynomial algorithm for equivalence problem of deterministic multitape fnite automata, Theor. Comput. Sci., 833, pp. 120-132, 2020.

Ռաբին-Սքոթի մոդելը[1] (ՌՍՄ) և խառը վիճակներով մոդելը[2] (ԽՎՄ)։ ԽՎՄ-ը ի տարբերություն ՌՍՄ-ի ավտոմատի յուրաքանչյուր վիճակի հետ չի համապատասխանեցնում որևէ ժապավեն։

Դիցուք $G = \langle Y, R\rangle$ և R-ն այնպիսին է, որ Y-ը կարելի է ներկայացնել չհատվող բազմությունների տեսքով՝ $Y = Y_1 \cup Y_2 \cup \ldots \cup Y_m$ այնպես, որ $\forall y, y' \in Y$, եթե $y, y' \in Y_i$, ապա $yy' \neq y'y$, իսկ եթե $y \in Y_i, y' \in Y_j, i \neq j$, ապա $yy' = y'y$։ Այս պայմանը թույլ է տալիս Y-ը դիտարկել որպես այբուբեն բազմաժապավեն վերջավոր ավտոմատների համար։

Դիցուք Q-ն վիճակների վերջավոր բազմություն է, Y-ը այբուբեն է, որը բավարարում է վերը նշված պայմանին, $\Delta : Q \times Y \to 2^Q$ անցումների ֆունկցիան է, $q_0 \in Q$ սկզբնական վիճակն է, իսկ $F \subseteq Q$ վերջնական վիճակների բազմությունն է։

m ժապավենանի ավտոմատ ՌՍՄ-ում (ԲՎԱ-ՌՍ) կոչվում է $A = (Q, T, Y, \Delta, q_0, F)$ վեցնյակը, որտեղ $T : Q \to \{1, \ldots, m\}$ ֆունկցիա է, որը յուրաքանչյուր $q \in Q$ վիճակ համապատասխանեցնում է որոշակի ժապավենի, իսկ $Q = \bigcup_{i=1}^{m} Q_i$, այնպես որ $Q_i = \{q | q \in Q, T(q) = i\}$ $\forall i = 1, \ldots, m$։

m ժապավենանի ավտոմատ ԽՎՄ-ում (ԲՎԱ-ԽՎ) կոչվում է $A = (Q, Y, \Delta, q_0, F)$ հնգյակը։

Երկու մոդելների համար ε-բազմաժապավեն վերջավոր ավտոմատը սահմանվում է նույն կերպ՝ Y-ի փոխարեն վերցնելով $Y \cup \{\varepsilon\}$ այբուբենը, որտեղ ε-ը դատարկ բառն է։

1.2-ում սահմանված է ազատ մասնակի տեղափոխական մոնոիդի համար կանոնավոր պատահույթի, կանոնավոր արտահայտության և պատահույթների հանրահաշվի գաղափարները։ Ապացուցված է, որ սահմանված պատահույթների հանրահաշիվը Քլինիի հանրահաշիվ է։

$K(Y)$-ում կանոնավոր պատահույթը կսահմանենք հետևյալ կերպ.

1. $\emptyset$-ը (դատարկ բազմությունը) կանոնավոր պատահույթ է։
2. $\tilde{E} = \{[(e, \ldots, e)]\}$ բազմությունը կանոնավոր պատահույթ է։
3. $\forall y \in Y, \{[K(y)]\}$ բազմությունը կանոնավոր պատահույթ է։
4. Եթե P-ն ու Q-ն կանոնավոր պատահույթներ են, ապա կանոնավոր պատահույթներ են նաև հետևյալ բազմությունները.
   a) $P + Q = P \cup Q$,
   b) $PQ = \{[s] \mid s = pq, [p] \in P, [q] \in Q\}$, որտեղ pq-ն p-ի և q-ի վերը սահմանված ձախ կոնկատենացիա գործողությունն է,
   c) $P^* = \bigcup_{n\geq 0} P^n$, որտեղ $P^0 = \tilde{E}$, $P^n = PP^{n-1}, n \geq 1$։
5. $K(Y)$-ում այլ կանոնավոր պատահույթներ գոյություն չունեն։

[1] M.O. Rabin, D.S. Scott, Finite Automata and Their Decision Problems, IBM J. Res. Dev. 3(2), pp. 114–125, 1959.

[2] H. Tamm, On Minimality and Size Reduction of One-Tape and Multitape Finite Automata, Ph.D. thesis, Department of Computer Science, University of Helsinki, Finland, 2004.

$K(Y)$-ում կանոնավոր արտահայտությունը կսահմանենք հետևյալ կերպ.

1. $\emptyset$-ը կանոնավոր արտահայտություն է, որով նշանակվում է $\emptyset$ կանոնավոր պատահույթը:
2. $K(e) = (e, \ldots, e)$-ը կանոնավոր արտահայտություն է, որով նշանակվում է $\tilde{E}$ կանոնավոր պատահույթը:
3. $\forall y \in Y, K(y)$-ը կանոնավոր արտահայտություն է, որով նշանակվում է $\{[K(y)]\}$ կանոնավոր պատահույթը:
4. Եթե p-ն և q-ն կանոնավոր արտահայտություններ են, որոնցով նշանակվում են համապատասխանաբար P և Q կանոնավոր պատահույթները, ապա կանոնավոր արտահայտություններ են նաև հետևյալ արտահայտությունները.
   a) $p + q$, որով նշանակվում է $P + Q$ կանոնավոր պատահույթը,
   b) pq, որով նշանակվում է PQ կանոնավոր պատահույթը,
   c) $p^* = \bigcup_{n\geq 0} p^n$, որտեղ $p^0 = K(e)$, $p^n = pp^{n-1}, n \geq 1$, որով նշանակվում է P^* կանոնավոր պատահույթը:
5. $K(Y)$-ում այլ կանոնավոր արտահայտություններ գոյություն չունեն:

Կանոնավոր պատահույթ կամ կանոնավոր արտահայտություն ասելով, այսուհետև կհասկանանք ազատ մասնակի տեղափոխական մոնոիդի համար սահմանված կանոնավոր պատահույթներն ու կանոնավոր արտահայտությունները:

«p բառը պատկանում է P կանոնավոր պատահույթին» ասելով, կհասկանանք՝ $[p]$-ն պատկանում է P-ին:

Պարզության համար, կանոնավոր արտահայտությունների գրելաձևում $K(Y)$-ի էլեմենտների փոխարեն երբեմն կօգտագործենք Y այբուբենի համապատասխան տառերը:

Օրինակ 1 Դիցուք $G = \langle Y, R\rangle$, $Y = \{y_1, y_2, y_3\}$ և $R = \{y_1y_2 = y_2y_1, y_1y_3 = y_3y_1\}$: Այս դեպքում $K(Y)$-ի տարրերն են $(1, e, e), (e, 1, 0)$ և $(e, 0, 1)$ կորտեժները: Դիտարկենք $(b_1, b_2, b_3) \in K(Y)^*$ կորտեժները: Հետևյալ բոլոր երեք կանոնավոր արտահայտություններին համապատասխանում է նույն կանոնավոր պատահույթը՝ բոլոր այն կորտեժների բազմությունը, որոնցում b_2-ը վերջանում է 1-ով.

$$(1, e, e)^* \left((e, 1, 0) + (e, 0, 1)\right)^* (e, 1, 0),$$

$$\left((1, e, e) + (e, 1, 0) + (e, 0, 1)\right)^* (e, 1, 0),$$

$$\left((e, 1, 0) + (e, 0, 1)\right)^* (e, 1, 0)(1, e, e)^*:$$

Պարզության համար, այս երեք կանոնավոր արտահայտությունների փոխարեն կարող ենք օգտագործել համապատասխանաբար հետևյալ գրելաձևերը. $y_1^*(y_2 + y_3)^*y_2$, $(y_1 + y_2 + y_3)^*y_2$ և $(y_2 + y_3)^*y_2y_1^*$:

Կանոնավոր պատահույթների համար սահմանվում է մասնակի կարգ. P և Q կանոնավոր պատահույթների համար, կասենք $P \leq Q$, եթե որպես բազմություններ $P \subseteq Q$: $K(Y)$-ում բոլոր պատահույթների բազմությունը նշանակենք $A_{K(Y)}$-ով: $(A_{K(Y)}, +, \cdot, ^*, \emptyset, \tilde{E})$ վեցնյակը կանվանենք պատահույթների հանրահաշիվ:

Թեորեմ 1 $(A_{K(Y)}, +, \cdot, ^*, \emptyset, \tilde{E})$ պատահույթների հանրահաշիվը Քլինիի հանրահաշիվ է:

Հայտնի է, որ դասական պատահույթների հանրահաշիվը լրիվ է Քլինիի աքսիոմների նկատմամբ: Այս փաստը և Թեորեմ 1-ը թույլ են տալիս պնդել, որ դասական պատահույթների հանրահաշվի նույնությունները տեղի ունեն նաև ազատ մասնակի տեղափոխական մոնոիդի համար սահմանված պատահույթների հանրահաշվում:

1.3-ում դիտարկված են ազատ մասնակի տեղափոխական մոնոիդի համար սահմանված պատահույթների հանրահաշվում գծային հավասարումների համակարգերը: Ապացուցված է, որ եթե նման գծային հավասարումների համակարգն ունի միակ լուծում, ապա այն կարելի գտնել փոփոխականների արտաքսման եղանակով:

Դիտարկենք գծային հավասարումների հետևյալ համակարգը.

$$\begin{cases} X_1 = X_1S_{11} + X_2S_{21} + \ldots + X_kS_{k1} + P_1 \\ X_2 = X_1S_{12} + X_2S_{22} + \ldots + X_kS_{k2} + P_2 \\ \vdots \\ X_k = X_1S_{1k} + X_2S_{2k} + \ldots + X_kS_{kk} + P_k \end{cases} \tag{1}$$

որտեղ S_{ij}-երը և P_i-երը տրված կանոնավոր պատահույթներ են, իսկ X_i-երն՝ անհայտ:

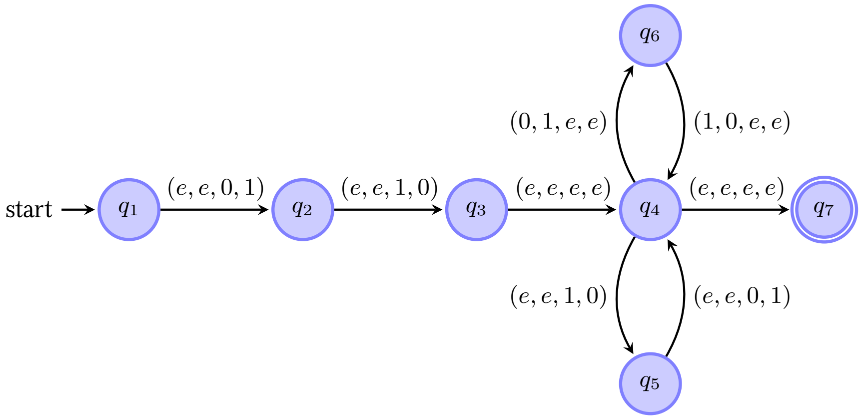
Պնդում 1 Եթե $[\{(e, \ldots, e)\}] \notin S$, ապա $X = XS + P$ հավասարումն ունի միակ լուծում, որը կարելի արտահայտել $X = P(S)^*$ տեսքով:

Պնդում 2 Եթե (1) հավասարումների համակարգն ունի միակ լուծում, ապա այն կարելի է գտնել փոփոխականների արտաքսման եղանակով:

Դիտարկենք Թեորեմ 2-ի կիրառման օրինակ.

Օրինակ 2 Դիցուք տրված է Նկար 1-ում ներկայացված վերջավոր բազմաժապավեն ավտոմատը.

Դիտարկվող ավտոմատի անցումների ֆունկցիան կարելի է բերել Աղյուսակ 1-ում ներկայացված տեսքի, որն էլ իր հերթին կարելի է ներկայացնել հավասարումների (2) համակարգով:



Նկար 1: $(e,e,10,01)((e,e,01,10)+(10,01,e,e))^*$ կանոնավոր արտահայտությանն համապատասխանող բազմաժապավեն ավտոմատ

	q_1	q_2	q_3	q_4	q_5	q_6	q_7
q_1	$\emptyset$	$(e,e,1,0)$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
q_2	$\emptyset$	$\emptyset$	$(e,e,1,0)$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
q_3	$\emptyset$	$\emptyset$	$\emptyset$	(e,e,e,e)	$\emptyset$	$\emptyset$	$\emptyset$
q_4	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$(e,e,1,0)$	$(0,1,e,e)$	(e,e,e,e)
q_5	$\emptyset$	$\emptyset$	$\emptyset$	$(e,e,0,1)$	$\emptyset$	$\emptyset$	$\emptyset$
q_6	$\emptyset$	$\emptyset$	$\emptyset$	$(1,0,e,e)$	$\emptyset$	$\emptyset$	$\emptyset$
q_7	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

Աղյուսակ 1: Անցումների ֆունկցիայի ներկայացում

$$\begin{cases} X_1 = X_1\emptyset + X_2\emptyset + X_3\emptyset + X_4\emptyset + X_5\emptyset + X_6\emptyset + X_7\emptyset + \tilde{e} \\ X_2 = X_1(e,e,0,1) + X_2\emptyset + X_3\emptyset + X_4\emptyset + X_5\emptyset + X_6\emptyset + X_7\emptyset \\ X_3 = X_1\emptyset + X_2(e,e,1,0) + X_3\emptyset + X_4\emptyset + X_5\emptyset + X_6\emptyset + X_7\emptyset \\ X_4 = X_1\emptyset + X_2\emptyset + X_3\tilde{e} + X_4\emptyset + X_5(e,e,0,1) + X_6(1,0,e,e) + X_7\emptyset \\ X_5 = X_1\emptyset + X_2\emptyset + X_3\emptyset + X_4(e,e,1,0) + X_5\emptyset + X_6\emptyset + X_7\emptyset \\ X_6 = X_1\emptyset + X_2\emptyset + X_3\emptyset + X_4(0,1,e,e) + X_5\emptyset + X_6\emptyset + X_7\emptyset \\ X_7 = X_1\emptyset + X_2\emptyset + X_3\emptyset + X_4\tilde{e} + X_5\emptyset + X_6\emptyset + X_7\emptyset \end{cases} \tag{2}$$

Անհայտ փոփոխականների արտաքսման եղանակով լուծելով հավասարումների (2) համակարգը, կստանանք $X_7 = (e, e, 10, 01)((e, e, 01, 10) + (10, 01, e, e))^*$, որը հանդիսանում է տրված ավտոմատին համապատասխանող կանոնավոր արտահայտություն:

1.4-ում ներկայացված է հետքերի մոնոիդների այլընտրանքային բնութագրիչ բազմություն: Հետազոտված է ներմուծված բնութագրիչների օգնությամբ հետքերի համարժեքության խնդրի լուծման արագությունը:

Լեմմա 3 (Պրոյեկցիայի լեմմա[1]) Դիցուք u-ն ու v-ն $M(Y, I_Y)$ մոնոիդի հետքեր են: $u = v$ այն և միայն այն դեպքում, երբ

$$\pi_{y,y'}(u) = \pi_{\pi_{y,y'}}(v)\ \forall(y, y') \in D_Y,$$

որտեղ π_A արտապատկերումը հետքից ջնջում է A բազմությանը չպատկանող բոլոր տառերը:

Դիցուք $G = \langle Y, R\rangle$, $w \in Y^*$ և $y \in Y$: Նշանակենք $Ocr_y(w)$-ով y սիմվոլի w բառում հանդիպելու քանակը: $Z \subseteq Y$ ենթաբազմության համար նշանակենք $Ocr_Z(w) = \sum_{y\in Z} Ocr_y(w)$:

$\sigma_{y,i,j}(w)$-ով, որտեղ $i < j$, նշանակենք w-ի այն ենթաբառը, որի առաջին տառը w բառում y-ի i-րդ պատահելն է, իսկ վերջին տառը՝ j-րդ պատահելը: $i = 0$ դեպքում $\sigma_{y,i,j}(w)$-ն w բառի նախածանցն է, իսկ $j > Ocr_y(w)$ դեպքում, $\sigma_{y,i,j}(w)$-ն w-ի վերջածանցն է:

Սահմանենք $\delta_y(w) = (\beta_0, \beta_1, \ldots, \beta_m)$ այնպես, որ $\beta_i = Ocr_{\{y'\in Y|yy'\neq y'y\}}(\sigma_{y,i,i+1}(w))$, իսկ $k = Ocr_y(w)$ և $i = 0, \ldots, k$:

Թեորեմ 2 Դիցուք g_1-ն ու g_2-ը $\{y_1, y_2, \ldots, y_n\}$ ծնիչների բազմությամբ ծնված G ազատ մասնակի տեղափոխական մոնոիդի տարրեր են: $g_1 = g_2$ այն և միայն այն դեպքում, երբ

$$\delta_{y_i}(g_1) = \delta_{y_i}(g_2)\ \forall i = 1, \ldots, n:$$

Դիտարկենք հետքերի համարժեքության խնդիրը: Դիցուք հետքերը տրված են իրենց կամայական ներկայացուցիչ բառով: Խնդիրը փոխակերպվում է տրված բառերի համարժեքությունը որոշելուն: Այն կարելի է բաժանել երկու ենթախնդրի՝ ա) տրված բառից կառուցել իր բնութագրիչ բազմությունը, բ) ստուգել թե տրված բնութագրիչ բազմությունները համընկնում են թե ոչ: Եթե խնդիրը դիտարկում ենք ֆիքսված Y այբուբենի և I_Y տեղափոխելի տառերի զույգերի համար, ապա երկու բնութագրիչ բազմությունների կիրառության պարագայում էլ և՛ ա), և՛ բ) ենթախնդիրները լուծելի են գծային ժամանակում: Սակայն, այբուբենի տառերի և ոչ տեղափոխելի տառերի զույգերի քանակից արագության կախումը դիտարկելիս պատկերը փոխվում է: Աղյուսակ 2-ում ներկայացված է երկու խնդիրների բարդությունների համեմատությունը $\{\pi_{y,y'} \mid (y, y') \in D_Y\}$ և $\{\delta_y \mid y \in Y\}$ բնութագրիչների համար, որտեղ n-ը Y այբուբենի տառերի քանակն է, l-ը

[1] Volker Diekert and Yves Métivier. Partial commutation and traces. In Grzegorz Rozenbergand Arto Salomaa, editors, Handbook of Formal Languages, Volume 3: Beyond Words,pages 457–533. Springer, 1997.

դիտարկվող բառերից կարճագույնի երկարությունն է, իսկ $2k$-ն՝ D_Y բազմության զույգերի քանակը:

Բնութագրիչ	Բնութագրիչ բազմության կառուցում		Համարժեքության ստուգում	
	Միջին բարդություն	Վատագույն բարդություն	Միջին բարդություն	Վատագույն բարդություն
$\pi_{y,y'}$	$\Theta\left(\frac{kl}{n}\right)$	$\Theta(nl)$	$\Theta\left(\frac{kl}{n}\right)$	$\Theta(nl)$
δ_y	$\Theta\left(\frac{kl}{n}\right)$	$\Theta(nl)$	$\Theta(l+n)$	$\Theta(l+n)$

Աղյուսակ 2: Բնութագրիչ բազմության կառուցման և բնութագրիչ բազմությունների համարժեքության ստուգման խնդիրների բարդությունների համեմատությունը $\{\pi_{y,y'} \mid (y,y') \in D_Y\}$ և $\{\delta_y \mid y \in Y\}$ բնութագրիչների համար

Որպես հետևանք, ստացվում է, որ եթե $k > n/2$, ապա $l > cn^2/(2k-n)$ երկարությամբ բառերի համար ավելի արագ է δ_y բնութագրիչների կիրառումը:

1.5-ում լուծված են բազմաժապավեն վերջավոր ավտոմատների անալիզի և սինթեզի խնդիրները:

Այս գլխում դիտարկվող կանոնավոր արտահայտությունները սահմանված են միայն այնպիսի այբուբենների վրա, որոնք կիրառելի են բազմաժապավեն վերջավոր ավտոմատների համար՝ $Y = Y_1 \cup Y_2 \cup \ldots \cup Y_m$ այնպես, որ $\forall y, y' \in Y$, եթե $y, y' \in Y_i$, ապա $yy' \neq y'y$, իսկ եթե $y \in Y_i, y' \in Y_j, i \neq j$, ապա $yy' = y'y$:

Թեորեմ 3 (Բազմաժապավեն ավտոմատի սինթեզի խնդիր) Գոյություն ունի ալգորիթմ, որը տրված կանոնավոր արտահայտությունից սինթեզում է ε-ԲՎԱ, այնպիսին, որ դրանց երկուսի կողմից ճանաչվող լեզուները նույնն են:

Թեորեմ 4 (Բազմաժապավեն ավտոմատի անալիզի խնդիր) Գոյություն ունի ալգորիթմ, որը տրված ԲՎԱ-ից կառուցում է կանոնավոր արտահայտություն, այնպիսին, որ դրանց երկուսի կողմից ճանաչվող լեզուները նույնն են:

Լեմմա 4 Յուրաքանչյուր ε-ԲՎԱ-ՌՍ-ի (կամ ε-ԲՎԱ-ԽՎ-ի) համար գոյություն ունի դրան համարժեք չդետերմինացված ԲՎԱ-ԽՎ:

Երկրորդ գլուխը նվիրված է ազատ մասնակի տեղափոխական մոնոիդների վրա սահմանված կանոնավոր պատահույթների մետրիկական տարածություններին:

2.1-ում դիտարկված են դասական կանոնավոր պատահույթների համար հայտնի մետրիկաների ընդլայնումները ազատ մասնակի տեղափոխական մոնոիդի համար սահմանված կանոնավոր պատահույթների վրա:

Թեորեմ 1-ը թույլ է տալիս Բոդնառչուկի մետրիկան[1] ընդլայնել մեր կողմից սահման-

[1] V.G. Bodnarchuk, The Metrical Space of Events, Part I, Cybernetics 1(1), pp. 20–24, 1965.

ված կանոնավոր պատահույթների վրա: Այն սահմանվում է հետևյալ կերպ. P և Q կանոնավոր պատահույթների միջև Բողնաոչուկի հեռավորություն է կոչվում

$$d_B(P,Q) = 2^{-d(P\triangle Q)}$$

թիվը, որտեղ $d(S) = \min\{|s| \mid [s] \in S\}$, $|s|$-ը s բառի երկարությունն է, իսկ $P\triangle Q$-ն P և Q պատահույթների որպես բազմություններ սիմետրիկ տարբերությունը:

Դիցուք $G = \langle Y, R\rangle$ և R-ն այնպիսին է, որ Y-ը կարելի ներկայացնել չհատվող բազմությունների տեսքով՝ $Y = Y_1 \cup Y_2 \cup \ldots \cup Y_m$ այնպես, որ $\forall y, y' \in Y$, եթե y և y' տառերն պատկանում են նույն Y_i բազմությանն, ապա $yy' \neq y'y$, իսկ եթե պատկանում են տարբեր Y_i և Y_j բազմությունների, ապա $yy' = y'y$: $w_1, w_2 \in Y^*$ բառերի համար Լեվենշտեյնի հեռավորությունը կսահմանենք հետևյալ կերպ՝

$$d_{\hat{e}}(w_1, w_2) = \sqrt{\mu_1^2 + \mu_2^2 + \ldots + \mu_m^2},$$

որտեղ μ_i-ն $\pi_{Y_i}(w_1)$ և $\pi_{Y_i}(w_2)$ բառերի միջև Լեվենշտեյնի հեռավորությունն[1] է:

$K(Y)$-ում P և Q կանոնավոր պատահույթների միջև Լեվենշտեյնի հեռավորություն է կոչվում

$$d_{\hat{E}}(P,Q) = \min\{d_{\hat{e}}(p,q) | p \in P, q \in Q\}$$

թիվը:

2.2-ում և **2.3-ում** ներմուծված են նոր մետրիկաներ: Նոր մետրիկաների անհրաժեշտությունն հիմնավորվում է նրանով, որ ընդլայնված մետրիկաները չեն արտահայտում կանոնավոր պատահույթների միջև հեռավորության ինֆորմատիվ արժեքներ:

Առաջին մետրիկան սահմանվում է Էվկլիդյան մետրիկայի հիման վրա: Դիցուք $G = \langle Y, R\rangle$ ազատ մասնակի տեղափոխական մոնոիդ է: Դիտարկենք $K \circ Num : G \to \mathbb{N}^n$ արտապատկերումը, որտեղ Num-ը երկուական ներկայացման կորտեժները արտապատկերում է թվային կորտեժների: Նշանակենք Num_i'-ով $K \circ Num$ արտապատկերման պրոյեկցիան $\mathbb{N}^n$-ի i-րդ առանցքի վրա:

$g_1, g_2 \in G$ տարրերի համար լոգարիթմական հեռավորություն կանվանենք

$$d_l(g_1, g_2) = \sqrt{\lg^2 \frac{Num_1'(g_1)+1}{Num_1'(g_2)+1} + \cdots + \lg^2 \frac{Num_n'(g_1)+1}{Num_n'(g_2)+1}}:$$

Լեմմա 5 d_l հեռավորության ֆունկցիան մետրիկա է G-ում:

Սահմանված d_l մետրիկան կընդլայնենք կանոնավոր պատահույթների վրա՝ օգտագործելով Հաուսդորֆյան մետրիկան:

[1]V.I. Levenshtein, Binary Codes Capable of Correcting Deletions, Insertions, and Reversals, Sov. Phys. Dokl. 163(4), pp. 845–848, 1965.

Դիցուք P-ն ու Q-ն կանոնավոր պատահույթներ են $K(Y)$-ում: Նրանց միջև d_L հեռավորությունը սահմանվում է հետևյալ կերպ՝

$$d_L(P,Q) = \max\left\{\sup_{p\in P}\inf_{q\in Q} d_l(p,q), \sup_{q\in Q}\inf_{p\in P} d_l(p,q)\right\}:$$

$K \circ Num$ արտապատկերման առանձնահատկության պատճառով, որոշ դեպքերում d_L մետրիկան այնքան էլ ճիշտ չի արտահայտում կանոնավոր պատահույթների միջև հեռավորությունը: Այդ իսկ պատճառով ներմուծվում է շտկիչ, որը կոմբինացվում է d_l-ի հետ:

$g \in G$ տարրի համար, նշանակենք $d_1^n(g)$-ով այն n-տեղանի վեկտորը, որի i-րդ տեղում գրված է g-ում y_i ծնիչի պատահելու քանակը: $g_1, g_2 \in G$ տարրերի միջև շտկիչ հեռավորություն կանվանենք $d_1^n(g_1)$ և $d_1^n(g_2)$ վեկտորների միջև Էվկլիդյան հեռավորությունը և կնշանակենք $d_a(g_1, g_2)$-ով:

Լեմմա 6 d_a հեռավորության ֆունկցիան փսեվդոմետրիկա է G-ում:

d_l և d_a ֆունկցիաների կոմբինացիայի միջոցով սահմանվում է նոր հեռավորության ֆունկցիա, որը կկոչենք շտկված լոգարիթմական հեռավորություն.

$$d_{\tilde{l}}(g_1, g_2) = \sqrt{d_l(g_1,g_2)^2 + d_a(g_1,g_2)^2}:$$

Թեորեմ 5 $d_{\tilde{l}}$ հեռավորության ֆունկցիան մետրիկա է G-ում:

Կրկին օգտագործելով Հաուսդորֆյան մետրիկան, $d_{\tilde{l}}$ մետրիկան կընդլայնենք կանոնավոր պատահույթների վրա՝

$$d_{\tilde{L}}(P,Q) = \max\left\{\sup_{p\in P}\inf_{q\in Q} d_{\tilde{l}}(p,q), \sup_{q\in Q}\inf_{p\in P} d_{\tilde{l}}(p,q)\right\}:$$

2.3-ում ներկայացված է հետքերի մոնոիդների այլընտրանքային բնութագրիչ բազմության հիման վրա սահմանված մետրիկական տարածություն:

Դիտարկենք տեղափոխականության դասերի $\{\delta_y \mid y \in Y\}$ բնութագրիչ բազմությունը:

$g_1, g_2 \in G$ տարրերի համար $\delta'_y(g_1)$-ով կնշանակենք $\delta_y(g_1)$-ը, եթե $\delta_y(g_1)$ վեկտորի երկարությունը մեծ է կամ հավասար $\delta_y(g_2)$ վեկտորի երկարությունից: Հակառակ դեպքում $\delta_y(g_1)$ վեկտորին աջից կավելացնենք -1-եր մինչև նրա երկարությունը հավասարվի $\delta_y(g_2)$-ի երկարությանը, և ստացված վեկտորը կնշանակենք $\delta'_y(g_1)$-ով:

$g_1, g_2 \in G$ տարրերի համար սահմանենք d_c հեռավորության ֆունկցիան հետևյալ կերպ՝

$$d_c(g_1,g_2) = \left\|\left(\left\|\delta'_{y_1}(g_1) - \delta'_{y_1}(g_2)\right\|_2, \ldots, \left\|\delta'_{y_n}(g_1) - \delta'_{y_n}(g_2)\right\|_2\right)\right\|_2,$$

որտեղ $\|v\|_2$-ը v վեկտորի L_2 նորմն է:

Լեմմա 7 d_c հեռավորության ֆունկցիան մետրիկա է G-ում:

Դիցուք P-ն ու Q-ն կանոնավոր պատահույթներ են $K(Y)$-ում: Նրանց միջև d_C հեռավորությունը սահմանվում է հետևյալ կերպ՝

$$d_C(P,Q) = max \left\{ \sup_{p \in P} \inf_{q \in Q} d_c(p,q), \sup_{q \in Q} \inf_{p \in P} d_c(p,q) \right\}:$$

Երրորդ գլխում առաջարկվում են կանոնավոր պատահույթների համար սահմանված մետրիկաների հաշվման ալգորիթմներ:

3.1-ում բերվում է $d_{\tilde{L}}$ հեռավորությունը հաշվելու մոտավոր ալգորիթմ:

Հայտնի է, որ բազմաժապավեն ավտոմատների համարժեքության խնդիրն անլուծելի է, հետևաբար կանոնավոր պատահույթների միջև հեռավորությունը հաշվելու խնդիրը նույնպես անլուծելի է: Այդ իսկ պատճառով հեռավորությունը հաշվելիս հարկավոր է կիրառել մոտավոր մեթոդներ:

S կանոնավոր պատահույթի և ֆիքսված k թվի համար $W_k(S)$-ով կնշանակենք S-ի հետևյալ ենթաբազմությունը.

$$W_k(S) = \{s | s \in S, |s| \leq k\},$$

որտեղ $|s|$-ը s բառի երկարությունն է:

$d_{\tilde{L}}$ մետրիկայի արժեքի փոխարեն կհաշվենք

$$d_{\tilde{L},k}(S_1, S_2) = d_{\tilde{L}}(W_k(S_1), W_k(S_2)):$$

արժեքը: $d_{\tilde{L},k}$ հեռավորության ֆունկցիան փսեվդոմետրիկա է:

Դիցուք տրված են R_1 և R_2 կանոնավոր արտահայտությունները, որոնց համապատասխանում են S_1 և S_2 կանոնավոր պատահույթները: S_1-ի և S_2-ի միջև $d_{\tilde{L},k}$ հեռավորությունը, որտեղ k-ն նախապես ֆիքսված բնական թիվ է, կարելի է հաշվել հետևյալ ալգորիթմով.

1. R_1-ից և R_2-ից Թեորեմ 3-ի համաձայն կառուցել A_1 և A_2 բազմաժապավեն վերջավոր ավտոմատները:

2. Գտնել A_1-ի և A_2-ի կողմից ճանաչվող և k երկարությունը չգերազանցող բոլոր բառերի բազմությունները՝ $W_k(A_1)$ և $W_k(A_2)$:

3. Հաշվել $W_k(A_1)$ և $W_k(A_2)$ վերջավոր բազմությունների միջև $d_{\tilde{L}}$ հեռավորությունը:

Նկարագրված ալգորիթմի բարդությունը $O(kn(2l_1 + 2l_2)^{2k})$ է, որտեղ n-ը այբուբենում տառերի քանակն է, իսկ l_1-ն ու l_2-ը՝ համապատասխանաբար R_1 և R_2 կանոնավոր արտահայտություններում $(+, ^*, \cdot)$ գործողությունների քանակները:

3.2-ում բերվում է d_C մետրիկայի հաշվման ալգորիթմ, հիմնված դետերմինացված բազմաժապավեն ավտոմատների համարժեքությունը որոշելու ալգորիթմի վրա:

Ցուրաքանչյուր չդետերմինացված ԲՎԱ-ԽՎ կարելի է ձևափոխել դետերմինացվածի հետևյալ կերպ.

Y այբուբենին ավելացնենք $Y_{m+1} = Y_\varepsilon = \{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m\}$ սիմվոլները որպես նոր $m+1$-րդ ժապավենի սիմվոլներ: Այնուհետև տրված $A = (Q, Y, \Delta, q_0, F)$ չդետերմինացված ԲՎԱ-ԽՎ-ի վրա կատարում ենք հետևյալ ձևափոխությունները. յուրաքանչյուր $q \in Q$ վիճակի համար, եթե q-ից գոյություն ունեն տարբեր $Y_{i_1}, \ldots, Y_{i_k}$ ժապավենների տառերով անցումներ, ապա յուրաքանչյուր այդպիսի ժապավենի համար ավտոմատին ավելացնել նոր վիճակ՝ $(q_{i_1}, \ldots, q_{i_k})$: Ցուրաքանչյուր (q, y, q') q-ից անցման համար ավելացնել (q_j, y, q') անցումը, որտեղ $y \in Y_j$, և ջնջել (q, y, q') անցումը: Ցուրաքանչյուր ավելացված q_j վիճակի համար ավելացնել (q, ε_j, q_j) անցումը: Ստացված նոր ավտոմատը կնշանակենք $A^{(D)}$-ով: $A^{(D)}$ ավտոմատի կողմից ճանաչվող բառերը տարբերվում են A ավտոմատի կողմից ճանաչվող բառերից միայն ավելացրած $m+1$-րդ ժապավենի տառերով:

Դիցուք տրված են R_1 և R_2 կանոնավոր արտահայտությունները, որոնց համապատասխանում են S_1 և S_2 կանոնավոր պատահույթները: R_1-ից և R_2-ից սինթեզենք համապատասխան A_1 և A_2 ε-ԲՎԱ-ՌՍ-ները, որոնք էլ ձևափոխենք $A_1^{(D)}$ և $A_2^{(D)}$-ի: Վերջիններիս վրա կիրառենք “Congruence Builder” ալգորիթմը[1]: Եթե ալգորիթմի արդյունքում ավտոմատների սկզբնական վիճակները գտնվում են նույն համարժեքության դասի մեջ, ապա $d_C(S_1, S_2) = 0$: Հակառակ դեպքում $d_C(S_1, S_2)$-ը կհաշվենք հետևյալ մոտավոր եղանակով.

1. "Congruence Builder" ալգորիթմի կատարման ժամանակ սկզբնական վիճակների համարժեքության դասերում պահում ենք $A_1^{(D)}$ և $A_2^{(D)}$ ավտոմատների կողմից ճանաչվող բառերի $(m+1)$ տեղանի կորտեժները (ոչ բոլոր բառերը) որպես թվային կոորդինատներ: Այդ թվային կոորդինատներից կարելի է ստանալ հենց բառերի կորտեժները: $A_1^{(D)}$-ի այդ բառերի բազմությունը նշանակենք $P_1^{(D)}$-ով, իսկ $A_2^{(D)}$-ինը՝ $P_2^{(D)}$-ով:
2. $P_1^{(D)}$ և $P_2^{(D)}$ բազմություններին պատկանող բոլոր կորտեժներում ջնջել վերջին՝ $m+1$-րդ բառը՝ ստանալով բառերի m-տեղանի կորտեժների P_1 և P_2 բազմությունները համապատասխանաբար: Նշենք, որ այս քայլում ջնջվում են միայն Y_ε ժապավենի տառերը:
3. Ցուրաքանչյուր $p \in P_1 \setminus P_2$ կորտեժի համար, եթե p-ն ճանաչվում է A_2 ավտոմատի կողմից, ապա p-ն ավելացնել P_2-ին:

[1]H.A. Grigoryan, S.K. Shoukourian. Polynomial Algorithm for Equivalence Problem of Deterministic Multitape Finite Automata, Theor. Comput. Sci. 833, pp. 120–132, 2020.

4. Յուրաքանչյուր $p \in P_2 \setminus P_1$ կորտեժի համար, եթե p-ն ճանաչվում է A_1 ավտոմատի կողմից, ապա p-ն ավելացնել P_1-ին:
5. Հաշվել $d_C(P_1, P_2)$ հեռավորությունը (P_1-ն ու P_2-ը վերջավոր բազմություններ են):

3.3-ում քննարկված են **3.2** գլխում շարադրված ալգորիթմի իրականացման դժվարությունները:

Աշխատանքի շրջանակում ալգորիթմը իրականացվել է C++ լեզվով[1]: Որպես առանձին ենթախնդիր դիտարկված է բազմաժապավեն ավտոմատի կողմից բառի ճանաչման խնդիրը: Խնդրի իրականացման արդյունավետությունը ստուգելու համար ստեղծվել և օգտագործվել է հենանիշ[2]:

Հիմնական արդյունքներն ու եզրակացությունները

Աշխատանքի հիմնական արդյունքներն են.

- Բազմաժապավեն վերջավոր ավտոմատների համար սահմանվել են կանոնավոր պատահույթ և կանոնավոր արտահայտություն հասկացությունները: Սահմանման մեջ օգտագործված հատուկ կոդավորումը թույլ է տալիս ներմուծված կանոնավոր արտահայտությունների համար կիրառել նույն նշանագրությունը ինչ օգտագործվում է դասական կանոնավոր արտահայտությունների համար, սակայն կոնկատենացիա գործողության մեկնաբանությունը տարբերվում է [1]:

- Ապացուցվել է, որ սահմանված կանոնավոր պատահույթների հանրահաշիվը Քլինիի հանրահաշիվ է [1]:

- Ապացուցվել է, որ եթե ներմուծված պատահույթների հանրահաշվում գծային հավասարումների համակարգն ունի միակ լուծում, ապա այն կարելի է գտնել փոփոխականների արտաքսման եղանակով [1]:

- Ապացուցվել է, որ օգտագործվող կոդավորումից բխում է հետքերի մոնոիդների այլընտրանքային բնութագրիչ բազմություն [3]:

- Ապացուցվել է, որ եթե դիտարկվող այբուբենում ոչ տեղափոխելի տառերի զույգերի քանակը գերազանցում է այբուբենի տառերի քանակին, ապա այլընտրանքային բնութագրիչ բազմության կիրառությամբ հետքերի համարժեքության խնդիրը լուծող ալգորիթմը ավելի արագ է քան հայտնի բնութագրիչ բազմության կիրառությամբ ալգորիթմը [3]:

[1] Իրականացման հղումը՝ https://github.com/HayMurad/multitape_regex/tree/master/src

[2] Հենանիշի (benchmark) և դրա արդյունքի հղումը՝ https://github.com/HayMurad/multitape_regex/tree/master/test

- Լուծվել են բազմաժապավեն վերջավոր ավտոմատների սինթեզի և անալիզի խնդիրները [2, 3]:
- Կանոնավոր պատահույթների մի շարք մետրիկաներ ընդլայնվել են բազմաժապավեն վերջավոր ավտոմատների կանոնավոր պատահույթների վրա: Ցույց է տրվել, որ մի շարք դեպքերում այդ մետրիկաները ադեկվատ չեն ներկայացնում կանոնավոր պատահույթների միջև տարածությունը: [4]
- Ներմուծվել են երկու նոր մետրիկաներ. առաջինը հիմնված է ազատ մասնակի տեղափոխական մոնոիդի տարրերի և $\mathbb{N}^n$ տարածության միջև կապի վրա, իսկ երկրորդը՝ հետքերի մոնոիդների ներմուծված բնութագրիչ բազմության: Մշակվել է բազմաժապավեն վերջավոր ավտոմատների համար կանոնավոր պատահույթների միջև հեռավորությունը հաշվարկող երկու մոտավոր ալգորիթմ [2, 3, 4]:

Ատենախոսության թեմայի շրջանակներում հրապարակված աշխատություններ

1. Grigoryan T. A. Some Results on Regular Expressions for Multitape Finite Automata. *Proceedings of the YSU, Physical and Mathematical Sciences*, 2019, Volume 53, Issue 2, 82-–90.
2. Grigoryan T. A. An Approximate Method for Calculating the Distance Between Regular Languages for Multitape Finite Automata. *Mathematical Problems of Computer Science*, 2020, Volume 54, 69–79.
3. Godlevsky A., Grigoryan H., Grigoryan T., Shoukourian S. Some Results on Regular Events for Multitape Finite Automata: A Preliminary Report. *Bulletin of EATCS*, 2021, Volume 133.
4. Grigoryan T. A., Hayrapetyan M. S. Measurement of Distance Between Regular Events for Multitape Automata Based on a New Characterization of Equivalence Classes. *Proceedings of the YSU, Physical and Mathematical Sciences*, 2021, Volume 55, Issue 1, 72-–80.

Abstract

The aim and main goals of the thesis are to introduce regular expressions for multitape finite automata, define and study metric spaces for languages accepted by them and propose algorithms for calculating those metrics.

The main results are:

- Regular expressions and regular events for multitape finite automata were defined using a special binary coding of the elements in free partially commutative semigroups.
- The algebra of the defined regular events was proven to be a Kleene algebra.
- It was proven, that if a system of linear equations in the algebra of events has a unique solution, then it can be found by a successive elimination of unknown variables.
- The used binary coding was proven to lead to an alternative characterization of trace monoids. The usage of the alternative characterization in the problem of checking whether two given traces are equivalent was investigated.
- The synthesis and the analysis problems were solved for multitape finite automata.
- Some metrics for regular languages for one-tape automata were extended to the multitape case. It was shown, that those metrics do not adequately measure the distance between regular events in some cases.
- Two new metrics for languages accepted by multitape finite automata were introduced. Approximate algorithms for calculating those metrics were proposed.

Резюме

Главная цель и задачи диссертации определить регулярные выражения для языков принимающиеся детерминированными многоленточными конечными автоматами, изучить метрические пространства для этих языков и описать алгоритм для вычисления значения этих метрик.

Главные результаты работы:

- Определены регулярные выражения и регулярные события для многоленточных конечных автоматов с помощью специальной бинарной кодировки элементов свободных частично коммутативных полугрупп.

- Доказано, что алгебра регулярных событий для многоленточных конечных автоматов - алгебра Клини.

- Доказано, что если система линейных уравнений в алгебре регулярных событий для многоленточных конечных автоматов имеет единственное решение, то это решение можно найти путем последовательного исключения переменных.

- Доказано, что из использованной бинарной кодировки можно получить альтернативную характеристику моноидов следа. Показано, как использование этой характеристики влияет на скорость алгоритма для определения эквивалентности двух следов.

- Решены задачи синтеза и анализа многоленточных конечных автоматов.

- Некоторые метрики регулярных языков для одноленточных автоматов расширены для случая многоленточных автоматов. Было показано, что в некоторых случаях эти метрики не позволяют адекватно измерить расстояние между регулярными событиями.

- Были введены две новые метрики для языков, распознаваемых многоленточными конечными автоматами. Предложены приблизительные алгоритмы для вычисления этих метрик.