

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
MATEMATINĖS INFORMATIKOS KATEDRA
BIOINFORMATIKOS SPECIALYBĖ

Biogeografija paremti optimizavimo metodai

Biogeography based optimization methods

Bakalauro baigiamasis darbas

Atliko:	Bioinformatikos 4 kurso 1 grupės studentas Simonas Suprinavičius	(parašas)
Darbo vadovas:	Lekt. Irus Grinis	(parašas)
Recenzentas:	J. M. Darbuot. Karolis Koncevičius	(parašas)

Vilnius – 2021

TURINYS

IVADAS	2
1. SKIRTINGI EVOLIUCINIAI ALGORITMAI	4
1.1. Genetiniai algoritmai	4
1.2. Skruzdžių kolonijos optimizacija	4
1.3. Dalelių spiečiaus optimizacija	5
1.4. Diferencialinės evoliucijos optimizacijos metodas	6
2. BIOGEOGRAFIJA PAREMTI ALGORITMAI	7
2.1. Trumpai apie biogeografiją	7
2.2. Optimizacijos principas	8
2.3. Biogeografinio metodo pagrindas	9
3. PRAKTINIS PRITAIKYMAS	10
3.1. Paieškos rezultatas	10
3.2. Pilnasis perrinkimas	12
3.3. Naudingos funkcijos	12
3.4. Failų nuskaitymas	12
3.5. Salų kūrimas	13
3.6. Algoritmas	13
3.6.1. Ploto skaičiavimas	13
3.6.2. „Bausmės“ pritaikymas	13
3.6.3. Tinkamumo įverčiai	14
3.6.4. Individų keitimas	14
3.6.5. Migracija	14
3.6.6. Rezultatų tikrinimas	14
3.7. Parametrai	14
3.8. Biogeografinio metodo modifikacijos	16
3.8.1. Tinkamumo įvertis	16
3.8.2. Mutacijos	17
3.8.3. Nutraukimas	17
3.8.4. Individų perkūrimas	18
3.8.5. Iteracijų ir individų skaičius	18
3.8.6. Blogiausių individų pakeitimas	19
3.9. Biogeografinio metodo hibridizavimas	19
4. REZULTATŲ ANALIZĖ	20
4.1. Analizės apibendrinimas ir pastebėjimai	26
REZULTATAI IR IŠVADOS	27
LITERATŪRA	28
BIOGEOGRAFIJA PAREMTI OPTIMIZAVIMO METODAI SIMONAS SUPRINAVIČIUS SANTRAUKA	29
BIOGEOGRAPHY BASED METHODS OF OPTIMIZATION SIMONAS SUPRINAVIČIUS SUMMARY	30
PRIEDAI	31

Įvadas

Didelė dalis sudėtingų skaičiavimų vyksta naudojant vienokius ar kitokius metodus. Yra sukurta begalė skirtingų algoritimų specifiniams skaičiavimams atlikti. Tokie algoritmai kuriami būtent tam, kad sprendimo paieška užtruktų kiek galima trumpiau ir vyktų sklandžiau. Algoritmu galima laikyti tam tikrą metodą, kuris turi specifines gaires ir apribojimus, vykdant jo žingsnius pa-eiliui bus pasiektas norimas rezultatas. Taip pat algoritmus galima ir optimizuoti, siekiant pagerinti jų veikimą bendrai arba specifiniams uždaviniams, kadangi tam tikri parametrai gali ir palėtinti kai kuriuos skaičiavimus. Biogeografija paremti skaičiavimo algoritmai yra maža įvairių evoliucinių skaičiavimo metodų dalis. Reikėtų pradėti nuo to, kas yra evoliuciniai skaičiavimai. Tai metodika, naudojama sprendžiant sunkius uždavinius, siekiant paspartinti rezultatų radimą. Evoliuciniai skaičiavimai simuliuoja gamtoje vykstančius mechanizmus ir apribojimus, tokius kaip paveldimumas, migracija, natūrali atranka, mutacijos, realo talpa, rekombinacija¹ ir t.t.

Biogeografiniai algoritmai simuliuoja individus, tarp kurių vyksta migracijos. Keičiantis kartoms, nauji individai, su geresnėmis savybėmis, pakeičia senuosius. Tokiu būdu artėjama prie geriausio sprendimo. Šiame metode atsiranda salos konceptas, kuris yra būtinas, norint suprasti optimizacijos veikimą. Salose esanti informacija yra perduodama iš kartos į kartą. Tokiu būdu vyksta evoliucija.

Kaip ir kiti evoliuciniai metodai, taip ir šis, naudoja didelį kiekį individų paieškos erdvėje, kurie migruodami, keičiantis kartoms artėja link geriausio sprendimo. Evoliuciniai skaičiavimai naudojami ne tik įvairiuose gamtos mokslų uždaviniuose. Šie metodai yra naudojami plačiai ir kitose srityse, kadangi sprendžiant problemas, neretai tinkamų variantų pasitaiko daug. Žinoma, ne visuomet skaičiavimų pabaigoje gautas rezultatas bus pats geriausias, tačiau laiko atžvilgiu, greitas šių sprendimų radimas gali kompensuoti dalinį tikslumą.

Darbo tikslas – ištirti biogeografinio metodo varianto naudą konkrečiai problemai spręsti. Konkreti problema – turint 64×64 pikselių paveikslėlį, konkrečia spalva nuspalvintuose plotuose rasti didžiausią nudažytą kvadratą, kurio kiekvienas pikselis yra tos pačios spalvos. Darbo tikslai:

1. Sukurti duomenų rinkinį, kuriame yra 64×64 pikselių paveikslėliai, kuriuose yra atsitiktinai nuspalvinti plotai.
2. Parašyti pilnojo perrinkimo programą, kuri rastų didžiausią juodą kvadratą kiekviename paveikslėlyje.
3. Sukurti biogeografiniu principu veikiančią programą.
4. Rasti tinkamą parametrų derinį, kuris ganėtinai greitai ir tiksliai rastų kiek galima geresnius rezultatus.
5. Atlikti skaičiavimus visame duomenų rinkinyje, išsaugoti laiko ir surasto ploto rezultatus.
6. Atlikti skaičiavimų rezultatų analizę.

¹Genetinė rekombinacija, naujų genų derinių susidarymas palikuoniams (DNR molekulėse, chromosomose, dukterinėse ląstelėse), kai įvairiais būdais keičiamasi genais, genetinėmis struktūromis arba jų dalimis.

7. Padaryti išvadas.

1. Skirtingi evoliuciniai algoritmai

1.1. Genetiniai algoritmai

Genetiniai algoritmai buvo pirmieji pradėti tyrinėti evoliuciniai algoritmai. Jie buvo plačiai žinomi ir naudojami dažniau, nei kiti evoliuciniai algoritmai. Genetiniai algoritmai simuliuoja natūralią atranką taip sprendami optimizavimo problemas skaičiavimuose. Taip pat žinoma, kad pirminė idėja buvo skirta nagrinėti adaptyvias sistemas, o ne optimizuoti funkcijas. Genetiniai algoritmai simuliuoja šias ypatybes:

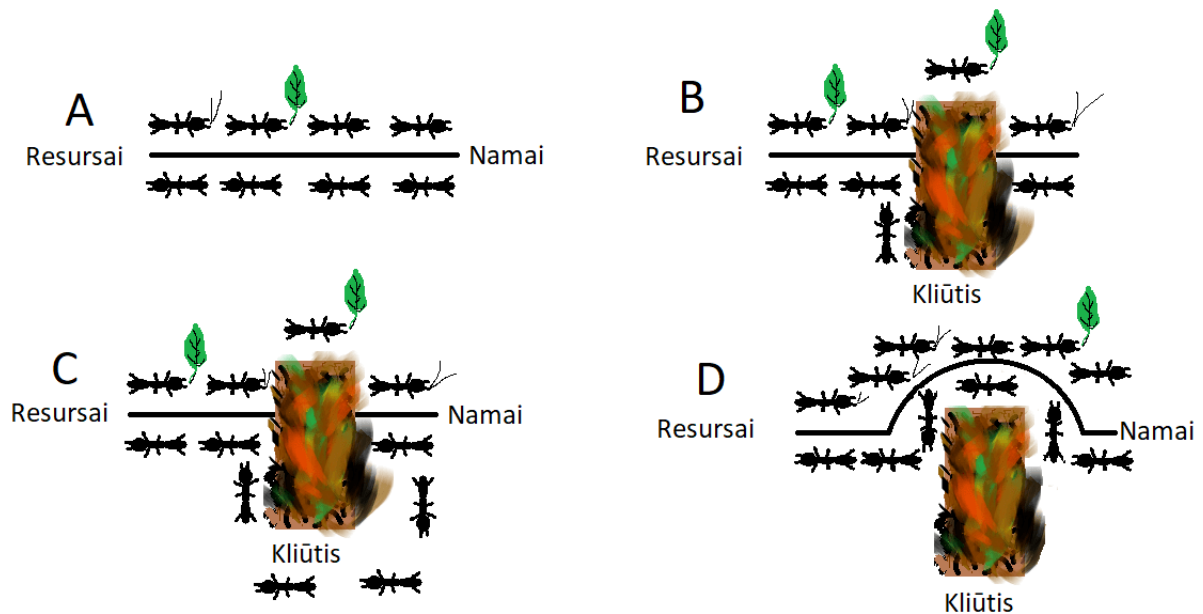
- Biologinėje sistemoje yra individų populiacija, kurioje dauguma jų turi reprodukcinės savybes.
- Individai turi ribotą gyvenimo trukmę.
- Vyksta populiacijos kaita.
- Išgyvenimo galimybės teigiamai koreliuoja su reprodukcine galimybe.

Pritaikant šias ypatybes, aprašoma optimizavimo problema, kurioje sprendimai (kandidatai) prilyginami individams. Dalis sprendimų yra geri, dalis netenkina poreikių. Tie individai (sprendimai), kurie yra geri, yra linkę daugintis. Palikuonių tėvai atmetami (bet neatmetama galimybė likti), o patys palikuonys tampa geresni. Taip keičiantis kartoms išlieka tik tie individai, kurie geriausiai prilįgsta siekiamai optimizacijai [Sim13].

1.2. Skruzdžių kolonijos optimizacija

Skruzdėlės pavieniui negeba atlikti sudėtingų darbų ir atskirai neišgyventų, tačiau gyvendamos kolonijoje jos geba dirbti vieningai, todėl kartais laikomos kaip vienas didelis organizmas, turintis kolektyvinį mąstymą. Vienas iš pagrindinių skruzdžių užduočių yra pastovios maisto ir kitų resursų paieška. Kuomet maisto šaltinis ištuštėja, ieškoma kito. Tai daro didelė dalis skruzdžių kolonijos. Tokiu būdu padidėja paieškos plotas ir greičiau randamas maisto šaltinis. Tam, kad skruzdės radusios maistą gebėtų jį vėl pasiekti, keliaudamos, skruzdėlės išskiria feromonus². Ten, kur kvapas intensyvesnis, yra rodiklis, kad keliaujant tuo maršrutu bus pasiektas tikslas. Tačiau ne visi keliai yra optimalūs. Bet tam yra išeitis – su laiku dalis skruzdėlių vis tiek ima klajoti, ir taip gali rasti greitesnį maršrutą. Judant mažesniais atstumais, dažnėja ir kelionių kiekis, todėl dažniau išskiriami feromonai, kurie tik intensyvina taku sklindantį kvapą. Tokiu būdu laikui bėgant, optimaliame kelyje kvapas stiprėja, o nepalankiame kelyje – blykšta. Galiausiai, skruzdės ima sekti stipresnį kvapą turintį taką. Šis metodas gali būti pritaikomas trumpiausio kelio paieškose (Pav. 1).

²Tai – lakiosios medžiagos, kurias gamina gyvūnų specialios liaukos ir išskiria į aplinką. Jos turi signalinę reikšmę, joms veikiant gali keistis gyvūnų elgesys, medžiagų apykaita ir kt.



Pav. 1. Skruzdės randa greičiausią maršrutą atsiradus kliūtim: A. Standartinis maršrutas. B. Atsiranda kliūtis (nuvirsta medis). C. Skruzdės vaikšto aplink ir ieško kelio. D. Randamas optimalus kelias.

Dalis mokslininkų teigia, kad tai nėra evoliucinis algoritmas, nes sprendimų kandidatai tiesiogiai nesikeičia sprendinių informacija. Tačiau tai vis tiek efektyvus, biologija ir populiacija besiremiantis algoritmas [Sim13].

1.3. Dalelių spiečiaus optimizacija

Panašiai kaip ir pastaroji skruzdžių kolonijos optimizacija, dalelių spiečius remiasi kolektyvine sąmone ir darbų pasiskirstymu. Panašūs pavyzdžiai yra maisto ir kitų resursų paieškos grupėje, padedančios išplėsti ir spartinti paiešką, susibūrimas į vieną grupę mažina šansus individui būti surastam bei maitinimosi atvejais, kuomet dalis individų stebi aplinką, kol grupė pasisotins. Dalelių spiečiaus optimizacija remiasi stebėjimais, jog individai ne tik dirba kartu tam, kad paspartintų grupės darbą, tačiau ir kiekvieno individo darbingumą. Šie pastebėjimai matomi ir žmonių aplinkoje. Paprastumo dėlei išskiriamos šios idėjos:

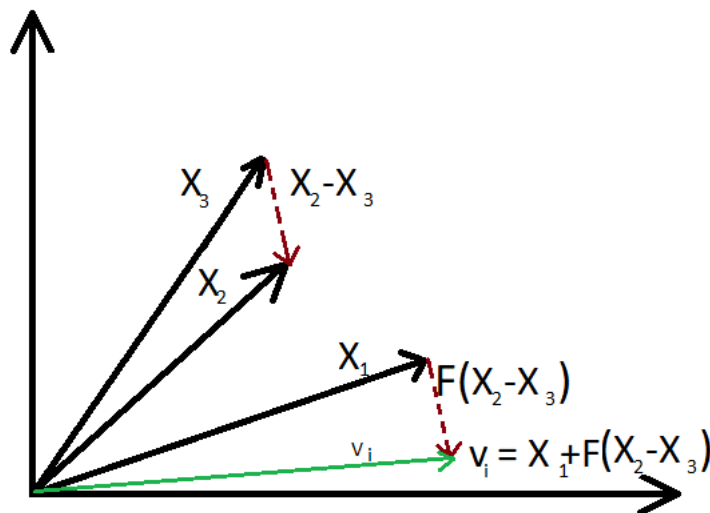
- Inercija – tęsiame tai, kas mums iki šiol sekėsi gerai ir keisti nieko nereikia.
- Visuomenės įtaka – remiamės žinomomis sėkmės istorijomis ir sekame sėkmingųjų pėdomis.
- Artimųjų ratas – didesnę įtaką padaro mus supantys artimieji [Sim13].

Šie algoritmai skiriasi nuo kitų evoliucinių algoritmų tuo, kad standartiškai evoliuciniai algoritmai yra statiški, nes perduoda sprendinius iš kartos į kartą, tačiau nemodeliuoja dinaminio sprendinių judėjimo per paieškos erdvę. Šiuose algoritmuose individas keliauja per paieškos erdvę su inercija, taigi turi ir greitį, o turimas pagreitis gali kisti dėl kelių priežasčių:

- Individas prisimena geriausią poziciją praeityje, todėl nori keisti greitį siekdamas grįžti atgal. Atitikmuo yra žmogaus prisiminimas senais gerais laikais ir bandymas atkurti to prisiminimo patirtį. Šiuo atveju individas keliauja per paieškos erdvę ir jo pozicija kinta per kartas, tačiau individas prisimena buvusių kartų pozicijas ir lokaciją, kur buvo geriausias darbo atlikimas praeityje.
- Taip pat individas žino šios kartos artimųjų pozicijas erdvėje, tad reikalingas artimųjų apibrėžimas ir artimieji turi komunikuoti tarpusavyje siekdami spartinti optimizacijos problemos sprendimą [Sim13].

1.4. Diferencialinės evoliucijos optimizacijos metodas

Šis metodas yra paprasčiau aprašomas programavimo kalba, nei dauguma kitų metodų. Metodas sukurtas visai neseniai, daugmaž prieš 26 metus [Sim13]. Per ganėtinai trumpą laiką buvo sukurta įvairių alternatyvų pritaikomumui. Taip pat šis metodas buvo genetinio algoritmo variacija. Metodas kurtas su tikslu optimizuoti n -matį tolydžiojo srityje. Kiekvienas individas yra n -matis vektorius reprezentuojantis problemos sprendimą. Pats metodo principas yra ganėtinai paprastas – iš jau esamų dviejų vektorių yra apskaičiuojamas skirtumas, kuris yra pridodamas prie trečiojo jau egzistuojančio vektoriaus ir taip sukuriamas naujas vektorius (Pav. 2), atspindintis naują sprendimą (kandidatą).



Pav. 2. Paprastas pavyzdys ieškant sprendimo dvimatėje erdvėje. x_1 , x_2 ir x_3 yra sprendiniai. Apskaičiuotas atstumas tarp x_2 ir x_3 yra pridodamas prie x_1 ir gaunamas naujas mutuotas vektorius v_i , kuris yra naujas sprendinys kur n – populiacijos dydis, o $i \in [1, n]$.

2. Biogeografija paremti algoritmai

2.1. Trumpai apie biogeografiją

Dar visai nesenai informacijos apie šiuos metodus buvo ganėtinai mažai, „Google Scholar“ paieška 2008 metais rodė vos vieną straipsnį, o šiemet (2021 metais) jau apie 671000. Remiantis tokiais skaičiais galima pastebėti, kad ši sritis sparčiai auga ir turi daug potencialo. Jau iki šiol šis metodas yra pritaikytas daugelyje sričių, keletas iš jų yra robotika, antenų dizainas, navigacija, planavimas, orlaivių varikliai ir t.t. [Sima; Simb]. Biogeografija imta tyrinėti jau XIX–ame amžiuje [Wal76], tačiau tai tebuvo stebėjimai. Matematiniai modeliai imti kurti jau apie 1960–uosius metus. Modeliai buvo fokusuoti į rūšių pasiskirstymus kaimyninėse salose, bei išnykimą ir migraciją [Rob67].

Biogeografiniai algoritmai remiasi populiacijos dydžiu ir migracija. Ten, kur populiacija yra maža, individai yra linkę atkelti (imigruoti) ir pasilikti, tačiau didelėse populiacijose individai iškeliauja (emigruoja) dėl maisto stokos, padidėjusio plėšrūnų kiekio ir t.t.. Buveinės, izoliuotos nuo kitų buveinių ir kuriose individai migruoja, vadinamos salomis. Paprastai salos yra izoliuotos vandens, tačiau šiuo atveju tai gali būti ir kalnai, ir plėšrūnai, ir tvoros ar netgi dykumos.

Toliau naudojamas trumpinys HSI yra buveinės tinkamumo indeksas (*angl. Habitat suitability index*). HSI nurodo salos tinkamumą apsigyvendinimui. HSI sudaro vegetacinė ir topografinė įvairovė, krituliai, temperatūra ir jos pokyčiai, esamas plotas ir kiti panašūs faktoriai. Kiekvienas iš šių faktorių yra vadinamas tinkamumo indekso kintamasis (*angl. Suitability index variable*), trumpinys yra SIV. SIV yra nepriklausomas, o HSI priklauso nuo SIV. Didelį HSI turinčiose gyvenvietėse egzistuoja daug skirtingų rūšių, analogiškai mažas HSI lemia mažą rūšių kiekį. Dėl didelio kiekio skirtingų gyventojų, populiacijos salose yra linkusios emigruoti, nes toje vietoje visko per daug, o dėl maisto tenka ir pakovoti. Tad galima teigti, kad didelis HSI lemia emigracijos padidėjimą. Ten, kur HSI mažas didesnė yra imigracija. Taip pat, rūšiai emigruojant, dalis populiacijos pasilieka senoje vietoje, o dalis palieka salą. Šis reiškinys yra svarbus mūsų taikomam metodui. Kaip ir kituose metoduose, mes turime individus, kurie atitinka kandidatus į sprendinius. Biogeografijos metode šie individai yra salos, kur individai gali atlikti užduotį gerai arba ne taip gerai ieškodami sprendimo. Galimybė gerai spręsti užduotį individui, mūsų atveju yra salos su dideliu HSI, kuris nurodo tinkamumą gyvenimui saloje. Geri individai yra atsparesni pokyčiams, todėl yra labiau linkę palaikyti homeostazę ³ ir gebėti atsistatyti. Didesniam supratimui imigracija laikoma pokyčiu, todėl mažas HSI lemia didelius pokyčius, o salose, kuriose HSI yra didelis, pokyčiai maži ir individai yra linkę dalintis geromis savybėmis, todėl emigruoja į kitas salas perduodami geras savybes ir tuo pačiu saugosi nuo blogų pokyčių neleidami imigracijos į didelį HSI turinčią salą [Sim13].

³Tai yra išskirtinė gyvųjų organizmų savybė – gebėjimas palaikyti vidinės terpės pastovumą kintančių sąlygų metu.

2.2. Optimizacijos principas

Sakykime, kad turime N dydžio populiaciją, x_k yra k -tasis individas (objektas) populiacijoje X . Problemos dimensija yra n , o $x_k(s)$ yra s -tasis nepriklausomas x_k kintamasis, kur $k \in [1, N]$ ir $s \in [1, n]$. Kiekvienoje kartoje ir kiekvienoje sprendimo savybėje yra imigracijos tikimybė λ_k , kad ši savybė bus pakeista: $\lambda_k =$ tikimybė, kad s -toji x_k savybė bus pakeista. Jei sprendimo savybė turi būti keičiama, tuomet parenkamas emigruojantis sprendimas su proporcinga tikimybe keičiamajai savybei μ . Taip pat kiekviename individe pokyčiai kartoje įvyksta prieš pasikeičiant bet kuriam individui. Tam reikalinga laikina populiacija Z [Sim13].

Toliau ruletės metodas, kuris bus aptartas vėliau, šioje vietoje vadinsis migracijos kriterijumi. Trumpai tariant, šis metodas vykdo individo pasirinkimą pagal tinkamumą lyginant su visa populiacija. Sakykime, kad turime dvi n dydžio populiacijas X ir Z , kuri yra X kopija, kiekvienas individas populiacijose turi s savybių. Pavyzdžiui: individas (sala) yra funkcijos x^3 kintamasis, kuris yra dvejetainis skaičius, sudarytas iš keturių bitų. Kiekvienas bitas yra šio individo savybė. Tikslas – surasti didžiausią funkcijos rezultatą intervale $[0; 15]$. Sakykime, kad X populiacijos individai yra: 1010 1101 0000 0111 ir migracija imta vykdyti pirmajame individe:

1. Tarkime, kad imigracija neįvyksta pirmoje savybėje, todėl x_1 nepakinta.
Individai (salos): 1010 1101 0000 0111
2. Tarkime, kad imigracija vyksta antroje savybėje, migracijos kriterijus nutaria, kad x_2 yra emigruojantis individas, todėl x_1 antroji savybė pakeičiama x_2 antrąja savybe.
Individai: 1110 1101 0000 0111
3. Tarkime, kad imigracija vyksta trečioje savybėje, migracijos kriterijus nutaria, kad x_4 yra emigruojantis individas, todėl x_1 trečioji savybė pakeičiama x_4 trečiąja savybe.
Individai: 1110 1101 0000 0111
4. Tarkime, kad imigracija vyksta ketvirtoje savybėje, migracijos kriterijus nutaria, kad x_4 vėl yra emigruojantis individas, todėl x_1 ketvirtoji savybė pakeičiama x_4 ketvirtąja savybe.
Individai: 1111 1101 0000 0111

Ši individo modifikacija išsaugoma Z populiacijoje, tuomet vykdomi panašūs veiksmai su kitais individais. Po šių įvykių vykdomos atsitiktinės mutacijos. X populiacija pakeičiama Z populiacija ir veiksmai atliekami iš naujo. Tarkime, kad veiksmų sekos nebekartojame, X populiacijos individai susidarė tokie: 1111 1100 0110 1111. Palyginę individus, pastebime, kad du iš jų yra vienodi ir yra tinkamiausia, kadangi panaudoję juos funkcijoje, gausime geriausią rezultatą, todėl galime laikyti, kad radome didžiausią rezultatą.

Ankstesniuose darbuose bandymai buvo atlikti ieškant didžiausio ar mažiausio funkcijos rezultato, didžiausių įvairių figūrų nuspalvintų plotų. Šių bandymų metu buvo nustatyta, kad neatsižvelgiant į laiko limitą, biogeografinis metodas galiausiai pateikia geriausią sprendimą. Tačiau jei nėra atsižvelgiama į laiką, yra ir paprastesnių metodų, galinčių pateikti rezultatą, pavyzdžiui, pilnojo perrinkimo metodas. Šis algoritmas išbando visus įmanomus variantus. Pilnojo perrinkimo metodu

keliaujant per paieškos erdvę pakanka palyginti naują kandidatą į sprendinį su senuoju ir jei naujas kandidatas geresnis, senąjį reikia pakeisti geresniu. Kyla klausimas – kam tuomet reikalingas evoliucinis metodas, jei yra paprastesni būdai sprendimams rasti? Paprastam klausimui egzistuoja ir paprastas atsakymas. Pilnojo perrinkimo algoritmas, tikrindamas kiekvieną galimą variantą gali užtrukti labai ilgai, jeigu paieškos erdvė yra didelė. Tinkamai optimizuotas evoliucinis metodas, turėdamas gerus parametrus gali dideliu tikslumu rasti norimą rezultatą per žymiai trumpesnę laiką. Taip pat, jeigu nėra svarbus 100 % tikslumas, paieška gali vykti dar greičiau. Tačiau norint efektyviai įgyvendinti tokius metodus, reikia algoritmus optimizuoti ir pritaikyti konkrečioms užduotims.

2.3. Biogeografinio metodo pagrindas

Biogeografinis metodas remiasi dviem faktoriais: migracija ir mutacija. Migracija naudojama informacijos mainymuisi tarp individų, problemų sprendimui. Mutacija įveda papildomą įvairovę, galimybę atrasti naujus sprendimus. Metodo įgyvendinimui svarbu tinkamai apjungti šiuos faktorus, kad jų bendras veikimas duotų geriausius rezultatus [Zha19]. Kiekvienos iteracijos pabaigoje individai apsieičia savo informacija ir pritraukia kitus individus geresnio rezultato link. Ne visa turima informacija yra perduodama mažiau tinkamiems individams, todėl mutacijos, įnešdamos įvairovę gali padėti nukreipti juos tinkama linkme. Papildomai, jeigu subjektai yra susitelkę ties ne pačiais geriausiais sprendimais, mutacija padėtų paskirstyti susitelkimą link kitų paieškos grupių. Egzistuoja ir įvairūs metodai, skirti optimizuoti algoritmą. Vienas iš jų yra mišrus metodas, kuris naudoja dviejų individų SIV kombinaciją, kuri perduodama trečiajam subjektui [H P11].

3. Praktinis pritaikymas

Skaitytojui primenama, kad sprendžiama problema yra didžiausio juodo kvadrato paieška 64×64 matmenų dvispalviame (juodai baltame) paveikslėlyje. Šiame tyrime biogeografinis metodas bus naudojamas paieškos optimizavimui. Žinoma, kad pilnojo perrinkimo metodas geba rasti geriausią sprendimą. Su mažu kiekiu duomenų šis skaičiavimo metodas gali susidoroti su užduotimi per gan trumpą laiką. Tačiau davus didelį duomenų kiekį, kiekvieno įmanomo varianto patikra užims begalę laiko. Todėl evoliucinis metodas gali turėti pranašumą, jei bus tinkamai įgyvendintas. Tyrimo metu bus kuriama programa, kuri analizuos paveikslėlių vaizdus. Paveikslėliai bus sudaryti iš dviejų spalvų – baltos ir juodos. Programa turi surasti didžiausią įmanomą vien tik iš juodos spalvos sudarytą kvadratą. Kvadratas gali būti paverstas⁴, tačiau negali savyje turėti nei vieno balto pikselio. Paieškos rezultatas yra skaičius, nusakantis kiek pikselių užima kvadratas.

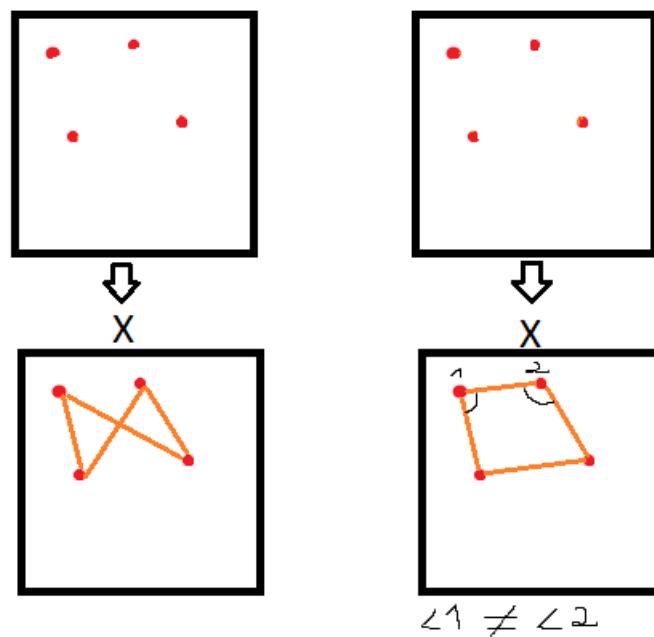
Rezultatų palyginimui ir metodo tikslumui nustatyti bus panaudotas pilnojo perrinkimo metodas. Bus lyginamas laikas, per kurį programos apskaičiuos rezultatus kiekviename paveikslėlyje atskirai ir bus palygintas paieškos tikslumas. Laikoma, kad pilnasis perrinkimas visada skaičiuoja 100% tikslumu, o biogeografinio metodo rezultatas padalinamas iš pilnojo perrinkimo rezultato ir gaunama procentinė dalis parodo, koku tikslumu veikia evoliucinis metodas.

3.1. Paieškos rezultatas

Kadangi ieškomas didžiausias įmanomas kvadratas, reikia apibrėžti šią figūrą. Kvadratas yra taisyklingas keturkampis, kurio visi kampai yra vienodo dydžio ir kraštinių ilgiai taip pat yra vienodi. Kvadratas plokštumoje gali būti orientuotas bet kaip, tačiau, kadangi plokštuma turi ribas, nei viena viršūnė negali išeiti iš ribų.

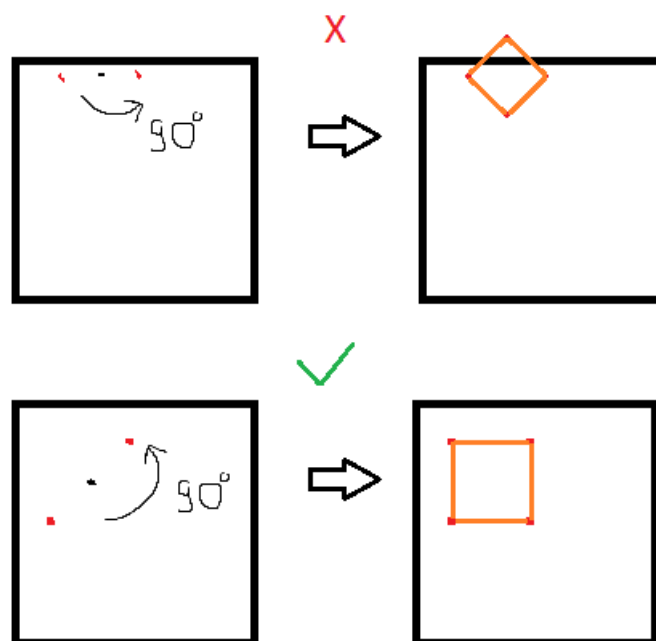
Norint apibrėžti kvadratą, galima būtų apibrėžti visas jo viršūnes, tuomet reikėtų papildomai mėginti tinkamai sujungti kraštines. Šiuo būdu nei viena viršūnė neišeitų už ribų. Tačiau problema yra ta, kad sukūrus keturias kraštines papildomai reikėtų tikrinti ar tikrai susidarė kvadratas, kadangi plokštumoje išdėstytos viršūnės gali atsirasti netinkamose vietose (Pav. 3).

⁴Paverstu kvadratu laikome tokį kvadratą, kurio nei viena kraštinė nėra lygiagreti paveikslėlio sienoms.



Pav. 3. Kuriant keturis kampus, tenka susidurti su visų viršūnių apjungimu ir kampų tikrinimu.

Kitas variantas yra apibrėžti dvi viršūnes, laikyti, kad jos yra ant vienos įstrižainės. Tokiu būdu galima sukurti kitas dvi viršūnes pavertus pastarąsias 90 laipsnių kampu per centro tašką. Šiuo metodu bus užtikrinta, kad kvadratas yra tinkamai sukurtas. Tačiau kuriant šią figūrą arti plokštumos ribų, kraštinės gali nebetilpti į paveikslėlio ribas (Pav. 4).



Pav. 4. Kuriant kvadratus iš dviejų viršūnių, pakanka patikrinti ar kitos dvi viršūnės neišeina už ribų.

Tenka rinktis – kuris metodas yra geresnis. Paprastesnis yra antras metodas, kadangi nereikia bandyti sieti visų viršūnių ir galvoti, kaip jas jungti, o patikrinti ar viršūnė neišeina už ribų yra labai paprasta, todėl priimtas sprendimas pasinaudoti šiuo metodu.

3.2. Pilnasis perrinkimas

Kaip jau anksčiau buvo paminėta, pilnasis perrinkimas išbando kiekvieną įmanomą variantą. Kadangi paveikslėlio dimensijos yra 64×64 pikselių, mėginant kiekvieną variantą, susidaro $(64 \times 64) \times (64 \times 64) = 64^4 = 16777216$ tikrinimų. Taip yra todėl, kad abiemis viršūnėms pradedant nuo koordinatų pradžios taško, pirmiausia slenkama antroji viršūnė per visą plokštumą, tikrindama kiekvieną variantą, tuomet, viršūnė grįžta į pradžią, pirmoji viršūnė pasislenka per vieną pikselį ir vėl antroji viršūnė keliauja per plokštumą. Taip keliaujama tol, kol pirmoji ir antroji viršūnės pasieks plokštumos paskutinį tašką. Žinant, kad tai tik 64×64 paveikslėlis, o variantų jau yra daug, galima suprasti, kad padidėjus išmatavimams, skaičiavimų trukmė prailgs labai didelį kiekį kartų. Tiksliau paskaičiavus, plokštumos matmenims padvigubėjus skaičiavimai prailgsta $2^4 = 16$ kartų. Didinant plokštumos matmenis ne vieną kartą šis skaičius auga eksponentiškai.

3.3. Naudingos funkcijos

Dalis funkcijų yra naudojamos tiek pilnojo perrinkimo metode, tiek biogeografiniame. Pirmą funkcija yra visų kvadrato viršūnių nustatymas, žinant dvi viršūnes, kurios yra ant kvadrato įstrižainės. Ši jau minėta funkcija pirmiausia randa vidurio tašką, tuomet pasuka turimas dvi viršūnes 90° per vidurio tašką. Taip gaunamos kitos dvi viršūnių pozicijos. Kita funkcija reikalinga nustatyti ar taškas yra kvadrato viduryje. Turint visas keturias viršūnes, galima apibrėžti tam tikras nelygybes, kurios pagal taško koordinates nustato ar jis yra kvadrato viduryje. Paskutinė perpanaudojama funkcija yra atsakinga už ploto skaičiavimą. Ši funkcija radialiniu rekursiniu būdu, nuo kvadrato centro tikrina visus jame esančius taškus.

Toliau sukuriamos dar kelios pagalbinės funkcijos – dvejetainio skaičiaus vertimas dešimtainiu, dvejetainio skaičiaus kūrimas iš pasirinkto kiekio bitų ir „Roulette-wheel selection“ metodas, kurį vadinsime rulete. Ruletės metodas išsirenka individą iš populiacijos, atsižvelgdamas į jo tinkamumą. Ši funkcija labai svarbi biogeografiniam algoritmui tuo, kad padeda išrinkti individus, kurių tinkamumas yra didžiausias.

3.4. Failų nuskaitymas

Dirbti su grafiniais failais yra sudėtingiau, nei tekstiniais, nes reikia skaidyti informaciją į mažesnius lygmenis, todėl siekiant paspartinti skaičiavimus, grafiniai failai pirmiausia konvertuojami į tekstinius failus, kuriuose talpinama tik svarbiausia informacija. Žinant, kad paveikslėliai sudaromi iš dviejų spalvų, pakanka baltus pikselius žymėti „0“, o juodus pikselius pažymėti „1“. Tokiu būdu tekstiniame faile lieka tik informacija apie pikselio poziciją ir spalvą. Pikselio pozicijos atskirai apibrėžti nereikia, kadangi kiekvienas grafinio failo taškas paeiliui ir analogiškai talpinamas į tekstinį failą.

3.5. Salų kūrimas

Metodo įgyvendinimui kritiškai svarbu yra tinkamai apibrėžti kvadratą. Skirtingai nei pilnojo perrinkimo metode, nepakanka naudoti natūraliuosius skaičius, kadangi perduodama tik dalinė informacija. Todėl labai pravartu, kad turime paveikslėlio matmenis, kuriuos nesunkiai galime apibrėžti dvejetainėje sistemoje ir pakanka vienodo ilgio bitų sekų. Kiekvienas taškas plokštumoje gali būti nusakytas dviem skaičiais. Pirmasis nusako x ašį, antrasis – y . Bet kurą x ar y poziciją nusakyti pakanka $\log_2(64) = 6$ bitų. Kadangi kvadratą apibrėžiame dviem viršūnėmis, visą kvadratą apibrėžti užtenka 24 bitų. Šie bitai yra salos tinkamumo parametrai, kurie migracijos metu bus perduodami kitiems individams.

3.6. Algoritmas

Pradžioje nuskaitomi failai, ir kuriamas pasirinktas kiekis individų, kurie vykdys migracijas (skiltis nr. 3.5.). Sekantis žingsnis yra pradėti pagrindinį ciklą, iš pasirinkto kiekio iteracijų, kuris atstos pasirinktą kiekį kartų⁵ pasikeitimą. Ciklas susideda iš kelių esminių veiksmų:

1. Nustatyti visas individo viršūnes ir paskaičiuoti individo užimamą plotą.
2. Individams, kurių užimamuose plotuose yra nenuspalvintų vietų, pritaikyti „bausmes“.
3. Apskaičiuoti individo tinkamumą.
4. Keisti blogiausius individus.
5. Atlikti migraciją.
6. Patikrinti ar rezultatas pagerėjo.

Toliau paaiškinami ciklo veiksmai.

3.6.1. Ploto skaičiavimas

Ploto paieškos metodas šiek tiek išsiskiria tuo, kad skaičiuojant individo plotą, ieškomi nuspaldinti ir nenuspalvinti pikseliai. Tuomet, jei visas plotas buvo nuspaldintas, šis individas pretenduoja būti sprendinys, jo suskaičiuotas ploto dydis nekeičiamas, tačiau kitu atveju jis nėra kandidatas būti sprendiniu. Tokiu atveju apskaičiuotas kvadrato ploto dydis sumažinamas sekančiame veiksmo.

3.6.2. „Bausmės“ pritaikymas

Plotas laikomas ne pilnai nuspaldintu, jeigu kvadrato yra vietų, kuriose yra kitos spalvos pikselių. Tokiais atvejais vykdoma „bausmė“, kurios metu sumažinamas gautas ploto dydis. Ploto mažinimui pasirinkti keli intervalai:

⁵Nuo žodžio karta.

1. Nuspalvintas plotas užima $[95; 100)\%$ individo ploto, apskaičiuotas dydis sumažinamas 90%.
2. Nuspalvintas plotas užima $[90; 95)\%$ individo ploto, apskaičiuotas dydis sumažinamas 20%.
3. Nuspalvintas plotas užima $[85; 90)\%$ individo ploto, apskaičiuotas dydis sumažinamas 25%.
4. Nuspalvintas plotas užima $< 85\%$, individo ploto, apskaičiuotas dydis prilyginamas 0.

3.6.3. Tinkamumo įverčiai

Apskaičiuoti dydžiai lyginami tarpusavyje ir nustatomi kvadratų tinkamumai. Lyginimas vyksta taip: individai suskirstomi į eilę pagal dydį didėjimo tvarka, jiems priskiriamas tinkamumas. Tinkamiausias individas yra tas, kurio ploto dydis didžiausias. Tuomet sukuriamos individų kopijos, kuriose vyks permainos.

3.6.4. Individų keitimas

Dalis blogiausių individų gali būti iš karto pakeičiami geriausiais, o migracijoje atliekami papildomi pakeitimai. Taip daroma todėl, kad patys blogiausi individai ilgai artės prie geresnių sprendimų vietų. Migracijos metu, šie individai bus papildomai keičiami, todėl tikėtina, kad bus išvengta individų identiškumo.

3.6.5. Migracija

Pirmiausia atliekami blogiausių individų pokyčiai – pagal tinkamumą skaičiuojama tikimybė, kad individas patirs imigraciją, tuomet ieškomas emigruojantis individas ruletės pagalba, atliekami pokyčiai bitų pozicijose, atliekamos atsitiktinės mutacijos, kurios aprašomos žemiau. Po visų permainų, kai individai pakeičiami jų kopijomis.

3.6.6. Rezultatų tikrinimas

Patikrinama ar naujoje kartoje atsirado geresnis kandidatas į sprendinius, jei taip – šis rezultatas išsaugomas. Ciklui pasibaigus paskutinį kartą paskaičiuojami plotai, patikrinama ar neatsirado didesnis nei maksimalus surastas plotas. Išvedami rezultatai.

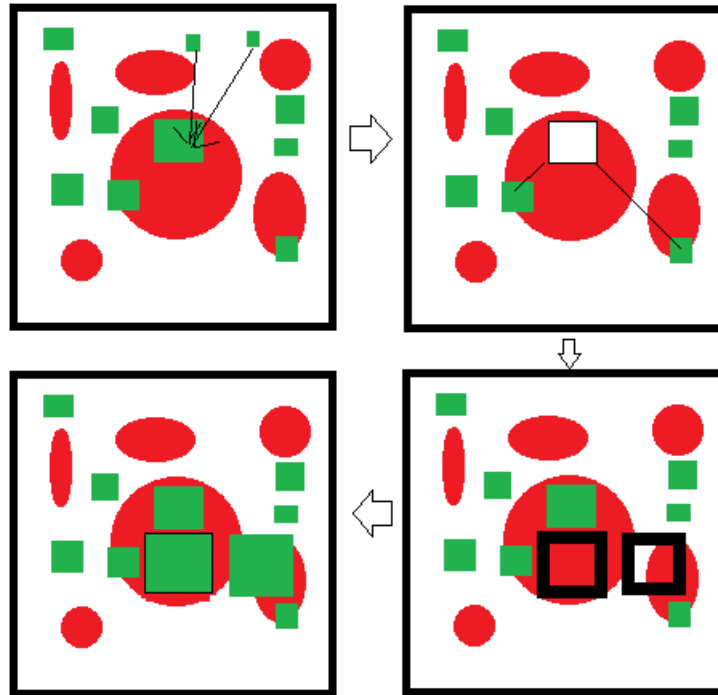
3.7. Parametrai

Dirbant su sąlyginai mažų matmenų paveikslėliais, buvo pasirinkta naudoti 25 individus, kurie keliaus per paieškos erdvę. Ciklo, reprezentuojančio kartų pasikeitimą, dydis buvo 2500 iteracijų. Bandymų metodu, tikrinant kelis rankiniu būdu sukurtus failus, buvo bandomos įvairios parametų kombinacijos, ir būtent ši kombinacija rodė potencialą.

Papildomai, metodo įgyvendinimui, buvo įvestos kelios modifikacijos. Viena iš jų – individų perkūrimas, kuris vyko dviem etapais. Pirmasis etapas buvo patikrinti naujai sukurtą individą. Tikrinamos buvo visos viršūnės ir žiūrima ar nei viena iš viršūnių neišeina už ribų. Jeigu taip –

individas perkuriamas. Antrasis etapas vyksta kartos pasikeitimo iteracijos pabaigoje (skiltis 3.6.6). Jeigu 200 kartų neatsiranda naujas geresnis rezultatas – perkuriami visi individai.

Antra panaudota modifikacija yra individų kopijavimas. 25% blogiausių individų yra pakeičiami geriausiais, tačiau jų tinkamumo įvertis nekinta. To pasekoje šie individai yra greičiau priartinami prie geriausio rezultato ir kartu modifikuojami, todėl jie pasiskirsto aplink geriausią rezultatą, kur gali būti naujas, geresnis rezultatas (Pav. 5).



Pav. 5. Šiame paveikslėlyje vaizduojama, kaip du kvadratai yra perkeliami į geriausio kvadrato vietą, tuomet yra modifikuojami kitų gerų rezultatų ir galiausiai šie individai atsiranda kitose vietose, kur vienas iš jų papuolė į dar geresnę vietą (žalias keturkampis apvestas plona juoda linija).

Tinkamumo įverčiui nustatyti buvo pasitelkta „S“ formos kreivė, kuri yra aprašyta žemiau ir turimomis žiniomis veikia geriau, nei tiesinio tinkamumo variantas [Sim13]. Kadangi dalis blogiausių variantų jau ir taip yra stipriai modifikuojami, tai padeda vienodai išskirstyti šiuos individus aplink geriausius rezultatus.

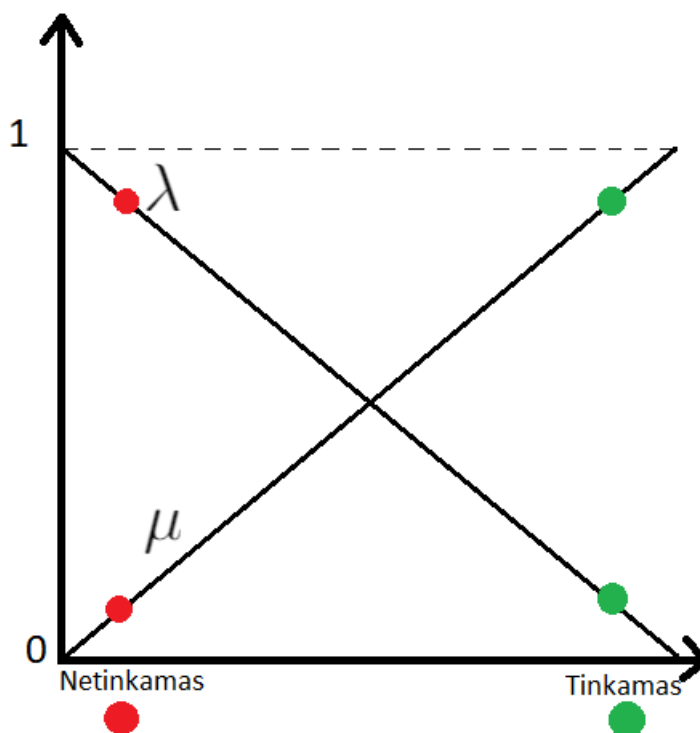
Mutacijos veikimas – kadangi kiekvienas individas yra sudarytas iš 24 savybių, kurios yra bitai, mutacija elementariai invertuoja bitą. Įvykus šiam pokyčiui, galima laikyti, kad individas mutavo. Mutacijos tikimybė palikta nedidelė – po 0,01 kiekvienai savybei. Kadangi yra pakankamai individų, o kiekvienas iš jų turi po 24 savybes, tikėtina, kad vis tiek nors vieną kartą keičiantis kartai bus įvykusi mutacija. Blogiausi individai jau yra ganėtinai smarkiai kaitaliojami, todėl papildomos mutacijos iš esmės keistų vidutinius rezultatus turinčias salas.

3.8. Biogeografinio metodo modifikacijos

3.8.1. Tinkamumo įvertis

Norint vykdyti migracinius mainus populiacijoje, reikia nustatyti kiekvieno individo tinkamumą. Kuo individas yra geresnis, tuo jo tinkamumo įvertis yra didesnis. Tinkamumo įvertis žymimas μ , jis yra atvirkščiai proporcingas λ , kuris atspindi kiekvienos salos imigracijos tikimybę, tuo tarpu μ rodo, kokia tikimybė, kad bus emigruojama iš salos. Vadinasi, kad jei sala, kaip individas, yra gerai prisitaikiusi, ji mielai pasidalins savo savybėmis, tačiau nebus linkusi priimti pokyčių, o blogai prisitaikiusi sala negalės perduoti blogų savybių, tačiau bus gretinama arčiau geriau prisitaikiusių salų.

Kadangi realių skaičių aibė yra be galo didelė, režiuose $[0; 1]$ taip pat yra be galo daug skaičių, todėl galime sutalpinti visų individų tinkamumus į šį režį tam, kad vėlesniuose veiksmuose generuojant atsitiktinį skaičių, jį galima būtų pasiimti iš šio intervalo. Individų tinkamumui palyginti, paprasčiausias metodas yra priskirti eilės numerį. Tai yra, jeigu turime N dydžio populiaciją, tai geriausias individas gaus tinkamumo įvertį $\mu = N/N = 1$, o $\lambda = 1 - \mu = 0$. Jeigu individai yra vienodai geri, jų tinkamumas bus vienodas. Grafiškai atvaizdavus šį pasiskirstymą, pastebėsime, kad susiformuoja dvi tiesės, kurios viena kitai yra atvirkščios. Kadangi šios tiesės nėra išsikraipiusios, tai šitokį tinkamumo pasiskirstymą galima vadinti tiesiniu (Pav. 6).

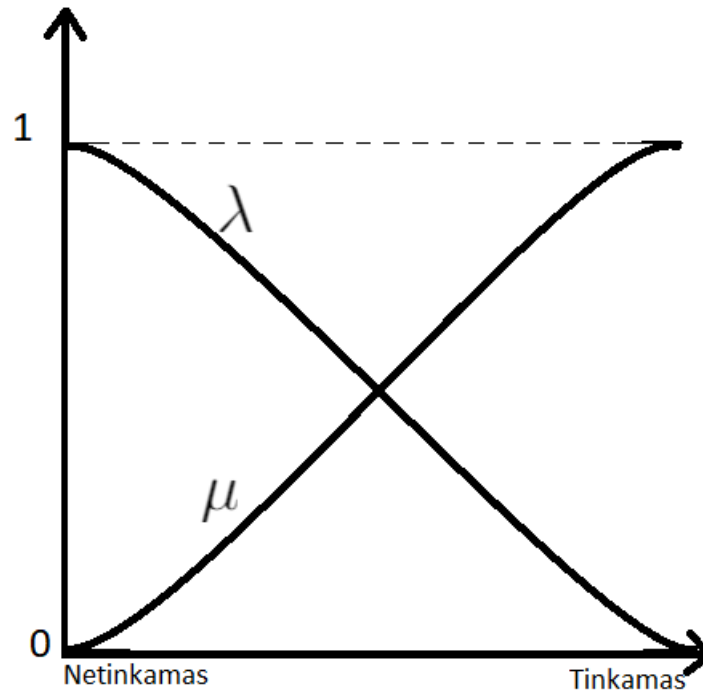


Pav. 6. Tiesinio tinkamumo pavyzdys. Raudonai pažymėtas individas nėra gerai prisitaikęs, todėl jo μ yra mažas, o λ didelis. Žaliojo individo situacija kitokia. Jis bus linkęs pasidalinti savo savybėmis, tačiau jis bus atsparesnis kitų įtakai.

Siekiant įvesti daugiau įvairovės tiesiniam tinkamumui intervalo režis nustatomas $(0; 1)$, o geriausio individo tinkamumas priskiriamas $\mu = N/(N + 1)$. Viena iš tiesinio tinkamumo alternatyvų yra „S“ (Pav. 7) formos kreivę atspindintis tinkamumas, kuris geriausiam individui priski-

riamas:

$$\mu = (1/2) \cdot (1 - \cos(\pi N/N)) \text{ arba } \mu = (1/2) \cdot (1 - \cos(\pi N/(N + 1))).$$



Pav. 7. „S“ formos tinkamumo pavyzdys.

Pastebėta, kad „S“ formos tinkamumas pagerina gerų rezultatų radimą [Sim13].

3.8.2. Mutacijos

Siekiant pajvairinti individų savybes kiekvienoje kartoje, galima atlikti mutaciją su maža tikimybe. Taip pat galima mėginti atlikti mutacijas tik tuomet, jei vyko individo migracija ir buvo bandoma keisti individo savybę. Mutacijos gali padėti pajvairinti ir gerai prisitaikiusius individus, kurie yra atsparesni pokyčiams. Tačiau reiktų nepamiršti, kad migracijos padeda pritraukti individus prie gerų rezultatų, tad mutacijos gali ir pabloginti rezultatą, todėl nereikėtų siekti įvesti didelį mutacijų dažnį.

3.8.3. Nutraukimas

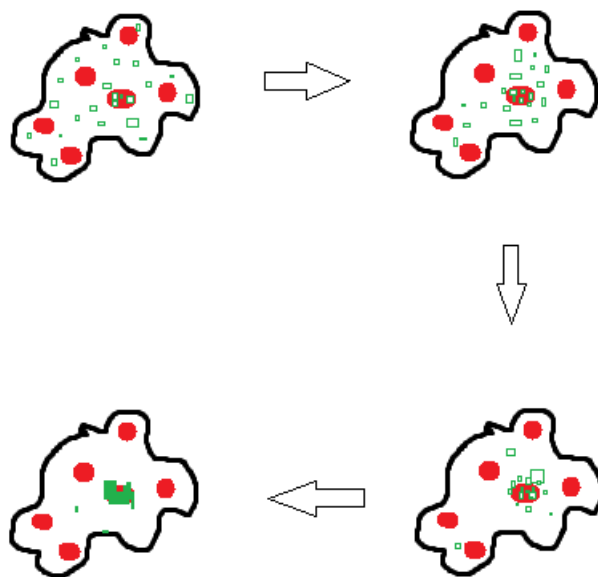
Kadangi šis evoliucinis metodas naudoja informacijos perdavimą iš kartos į kartą, gali susidaryti situacija, kad geras sprendinys jau egzistuoja ir vykdyti evoliuciją tampa nebereikalinga. Siekiant taupyti laiką galima įvesti įvairius kriterijus, kurie nutrauktų ciklą. Pavyzdžiui: įvairiaspalviame paveikslėlyje ieškoma didžiausio viena spalva nudažyto kvadrato. Vykstant 5000 iteracijų ciklui, ties 35–a iteracija buvo rastas kvadratas, kuris užima 50% paveikslėlio, galima daryti išvadą, kad neberasime kita spalva nudažyto didesnio kvadrato, todėl galima būtų nutraukti ciklą. Paieškos nutraukimas taupo išteklius ir išvengia bereikalingų skaičiavimų. Dirbant su dideliu informacijos kiekiu ši modifikacija gali būti labai naudinga.

3.8.4. Individų perkūrimas

Jeigu keičiantis kartoms neatsiranda geresnių kandidatų į sprendinius, tačiau nutraukti paieškos nesisinori, kitas variantas yra perkurti dalį arba visus individus. Jeigu individai kuriami atsitiktinai, toks metodas gali išbarstyti individus kitose paieškos erdvės vietose, kur sprendimai gali būti geresni, kadangi prieš tai buvusioje konstrukcijoje individai galėjo susitelkti į vieną vietą, kurioje gerų rezultatų nėra. Taip pat šis metodas gali pagelbėti, jeigu kuriami individai atsiranda ir už paieškos ribų. Jeigu salos užėina už ribų nors dalinai, jas reikėtų perkurti, kadangi jų rezultatai nebus naudingi.

3.8.5. Iteracijų ir individų skaičius

Keliaujant per paieškos erdvę yra svarbu turėti pakankamą kiekį individų, kurie galėtų ieškoti sprendimų skirtingose vietose, tačiau ilginiui imti telktis link optimalių sprendimų aplinkos (Pav. 8). Tačiau dideli pokyčiai nevyksta per mažą kartų pokytį, todėl turint didelį kiekį iteracijų ir individų paieška gali užtrukti. Jeigu iteracijų kiekis yra radikaliai mažas, tai labiau tampa atsitiktiniu spėjimu keliose vietose, negu evoliuciniu metodu. Analogiškas yra ir atvirkštinis variantas – turint daug iteracijų tačiau mažai individų, paieška bus siauroka ir kartų keitimasis taps bereikšmis, kadangi individai beveik nesiblašys per paieškos erdvę. Žinant iteracijų ir individų naudas galima bandymų metodu pritaikyti optimalų jų kiekį ir santykį, kuris algoritmą pavers labai efektyviu.



Pav. 8. Reprerzentacija, kaip tinkamai subalansuotas iteracijų ir individų kiekis, laikui bėgant, sutelktų individus link geriausio sprendinio. Žalia spalva pavaizduoti individai, raudona – sprendiniai.

3.8.6. Blogiausių individų pakeitimas

Dar viena palanki modifikacija gali būti dalies individų pakeitimas geriausiais individais prieš pradant vykdyti migraciją. Pasirenkamas kiekis (pvz. blogiausi 25%) individų, kurių savybės bus absoliučiai pakeistos tokiu pačiu kiekiu kitų individų. Tinkamumo įverčiai nėra keičiami, tokiu būdu galima greičiau pritraukti netinkamus individus, tačiau juos vis tiek keisti esamos iteracijos metu [Dep].

3.9. Biogeografinio metodo hibridizavimas

Evoliuciniai metodai pasižymi tuo, kad jie gali veikti pavieniui arba būti sujungti su kitomis metodikomis. Biogeografinis algoritmas ne išimtis. Šį metodą apjungiant su kitais galima gauti žymiai sparčiau veikiančią algoritmą. Šis jungimas vadinamas hibridizacija⁶.

Vienas šio metodo hibridizacijos pavyzdys – šiuolaikinis algoritmas, jungiant skruzdžių kolonijos optimizaciją su biogeografinio metodo optimizacija (*angl. Novel Hybrid Optimization Algorithm Combined with BBO and ACO*). Biogeografinis metodas panaudojamas bendrai paieškai, tuo tarpu skruzdžių kolonijos algoritmas panaudojamas pabaigoje, skaičiavimus atliekant su preliminariais sprendimais [Dai20].

Kitas pavyzdys yra biogeografijos hibridizavimas su dalelių spiečiaus metodu. Šis hibridas buvo efektyviai panaudotas auglių atpažinime [Zha15]. Šio tyrimo metu dviejų duomenų rinkinių rezultatų tikslumai buvo pasiekti 100%, o trečiojo – 99,49%. Tai parodo, kad net labai sudėtingos užduotys, pasitelkus evoliucinius skaičiavimus su gerais parametrais gali suteikti beveik tobulus rezultatus.

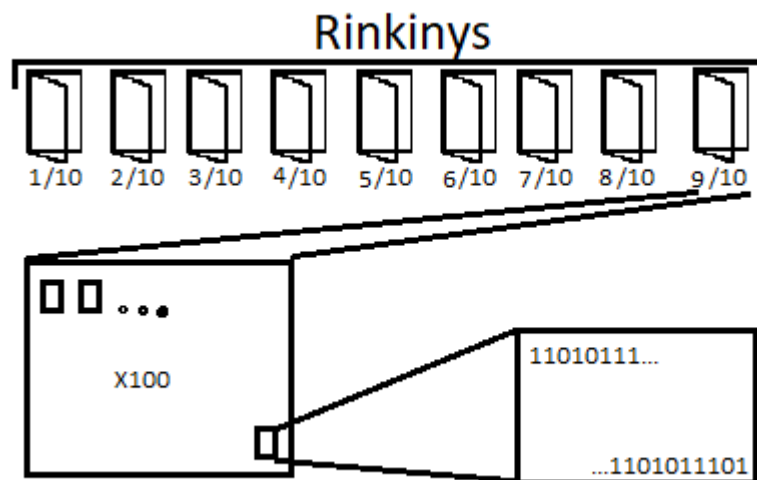
Dar vienas hibridizacijos pavyzdys yra biogeografinio metodo jungimas su diferencialinės evoliucijos algoritmu, kurio panaudojimas skirtas belaidžio sensoriaus energijos paskirstymui. Tyrimo tikslas buvo ištirti panaudojamumą ir optimizaciją visame tinkle [Bou11].

Šiame tyrime hibridizacijos metodas nėra panaudojamas, tačiau svarbu paminėti, kad algoritmas gali veikti tiek individualiai, tiek būti jungiamas su kitomis metodikomis. Kiekviena užduotis yra išskirtinė ir gali turėti ne vieną gerą sprendimų paieškos įgyvendinimą. Tačiau kadangi ne visus metodus galima išbandyti, nes jų yra begalė, tenka priimti vieną sprendimą su mintimi, kad jis gali būti geriausias pasirinkimas.

⁶Dviejų ar daugiau algoritmų suliejimas į vieną.

4. Rezultatų analizė

Analizei atlikti buvo sukurtas duomenų rinkinys. Jis buvo padalintas į 9 dalis, kur kiekvienoje iš jų buvo sukurta po 100 failų. Failai buvo kuriami iš karto tekstiniai, praleidžiant grafinį kūrimą. Visuose rinkinio failuose pikseliai buvo kuriami su tam tikra tikimybe. Pirmoje rinkinio dalyje failuose pikseliai buvo nudažomi juodai su 1/10 tikimybe, antrame – 2/10 ir taip toliau, kol paskutinėje dalyje pikseliai buvo dažomi su 9/10 tikimybe. Tai reiškia, kad pataškiui, kiekvienas failo pikselis buvo pažymimas kaip nudažytas su ta tikimybe. Pavyzdžiui, kuriant pirmojo rinkinio failus, kiekvienas faile esantis pikselis buvo nudažomas su 1/10 tikimybe, o paskutiniojo rinkinio failų pikseliai turėjo 9/10 tikimybes būti nuspalvintais (Pav. 9).

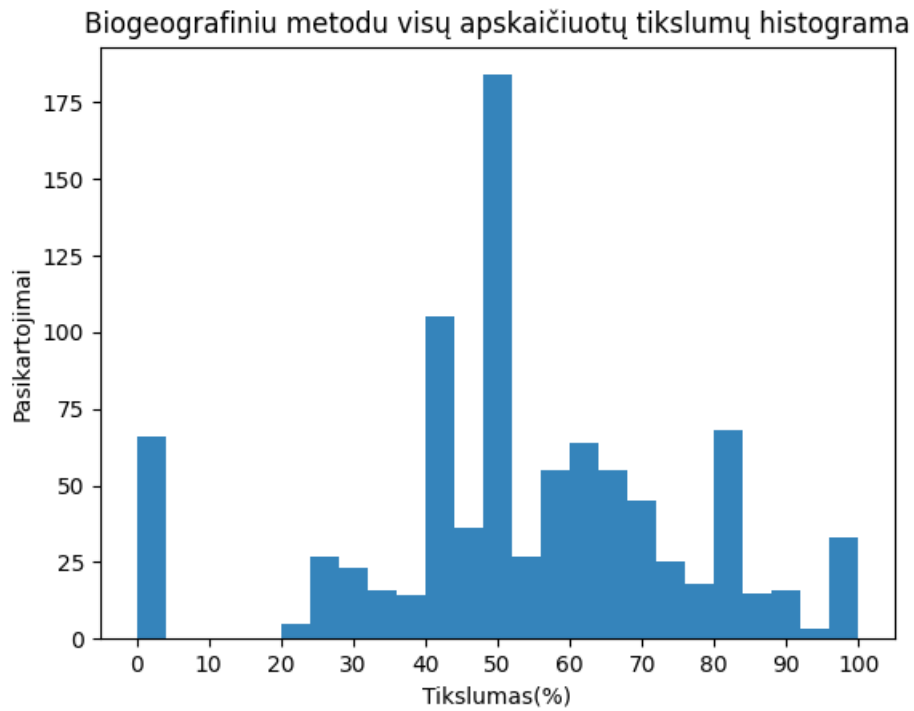


Pav. 9. Šiame paveikslėlyje vaizduojama, kaip atrodo duomenų rinkinys, sudarytas iš 9 dalių po 100 failų. Toliau parodyta kaip atrodo sukurtas failas, kur kiekvienas taškas nuspalvinamas (parašoma „1“) su tam tikra tikimybe, šiuo atveju 9/10.

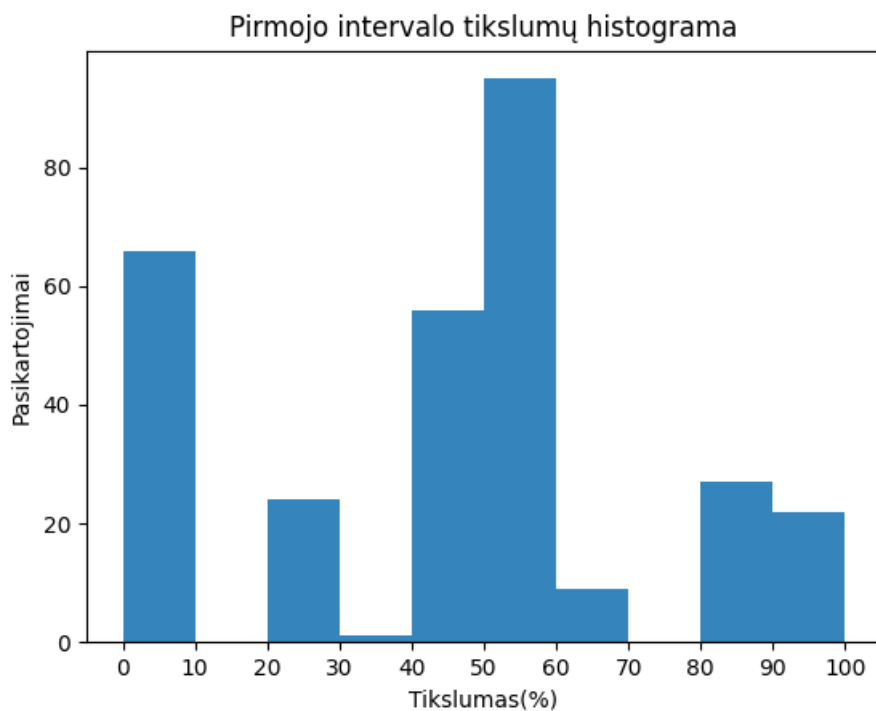
Atlikus skaičiavimus abiem metodais, buvo vykdoma analizė. Pirmiausia reikėtų pradėti nuo tikslumo. Jis gali būti apibrėžtas taip:

$$Tikslumas = \frac{Biogeografinis\ Rezultatas \cdot 100\%}{Pilnojo\ Perrinkimo\ Rezultatas}$$

Iš visų apskaičiuotų tikslumų buvo sudaryta histograma (Pav. 10), iš kurios galima pastebėti, kad didžioji dalis skaičiavimų rezultatų krypta link antrosios histogramos pusės. Paėmus visų skaičiavimų rezultatus, bendras vidurkis siekia 53,35%. Rezultatai nėra stubbinantys, tačiau reikėtų šiek tiek labiau pasigilinti į atskirus skaičiavimus. Šiam tikslui, duomenų rinkinio skaičiavimai buvo suskirstyti į 3 lygius intervalus. Kiekviename intervale yra po 300 skaičiavimo rezultatų. Pirmajame intervale yra visi rezultatai iš failų, kuriuose pikseliai buvo spalvinami su mažiausiomis tikimybėmis, o paskutiniame – su didžiausiomis. Pravartu būtų panagrinėti kiekvieną iš šių intervalų atskirai (Pav. 11–13).



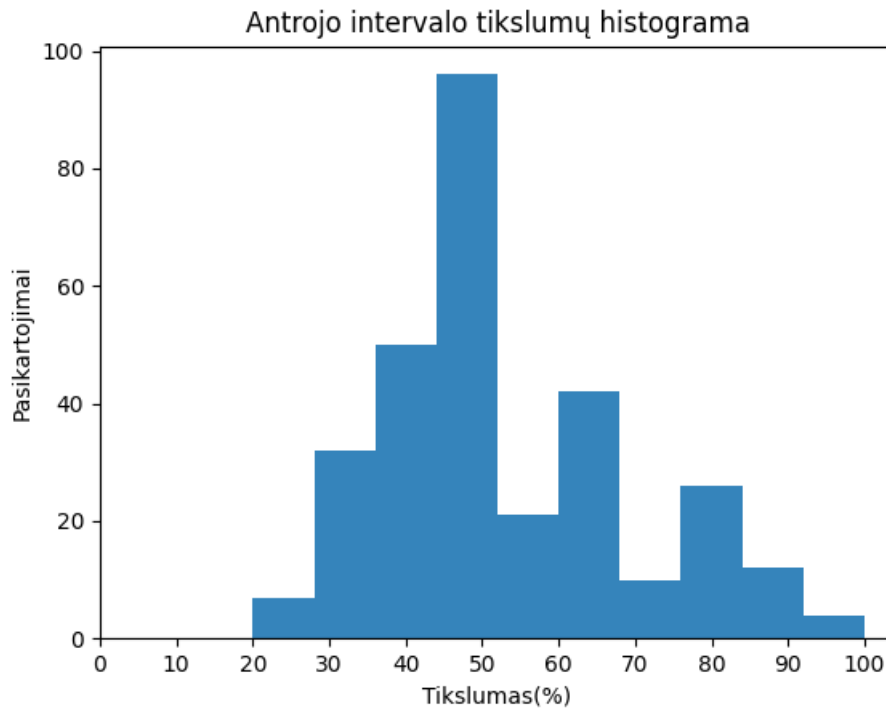
Pav. 10. Paveikslėlyje pavaizduota biogeografiniu metodu atliktų skaičiavimų tikslumo histograma. Vertikaliajoje ašyje atvaizduojami atsikartojimai, o horizontalioje ašyje atvaizduoti tikslumai išreikšti procentais.



Pav. 11. Paveikslėlyje pavaizduota biogeografiniu metodu atliktų skaičiavimų pirmojo intervalo tikslumo histograma. Vertikaliajoje ašyje atvaizduojami atsikartojimai, o horizontalioje ašyje atvaizduoti tikslumai išreikšti procentais.

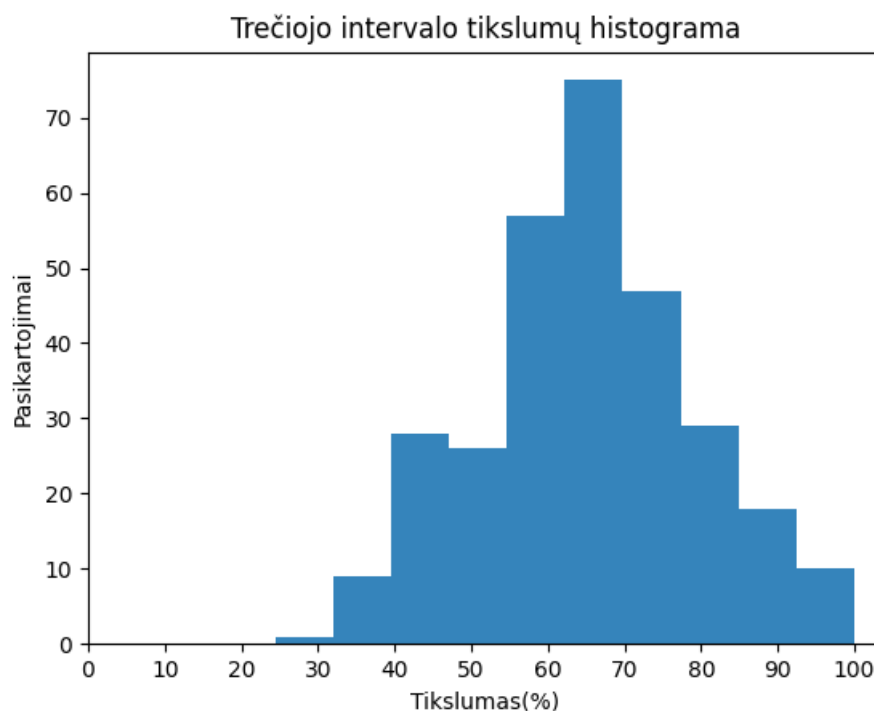
Nagrinėjant pirmąjį intervalą, aiškiai matyti, kad nemažai rezultatų tikslumų „sukrito“ ties pirmais 10%. Tačiau nemaža dalis rezultatų vis tiek viršija 50% ribą ir yra net ir 100% siekiančių

rezultatų. Paskaičiavus bendrą vidurkį, jis deja tesiekia 41.82%. Iš šių rezultatų, galima daryti išvadą, kad failuose, kuriuose yra mažai nuspalvintų taškų, kurie dar ir yra atsitiktinai išbarstyti, biogeografinis metodas didelės naudos gali ir neduoti.



Pav. 12. Paveikslėlyje pavaizduota biogeografiniu metodu atliktų skaičiavimų antrojo intervalo tikslumo histograma. Vertikaliajoje ašyje atvaizduojami atsikartojimai, o horizontalioje ašyje atvaizduoti tikslumai išreikšti procentais.

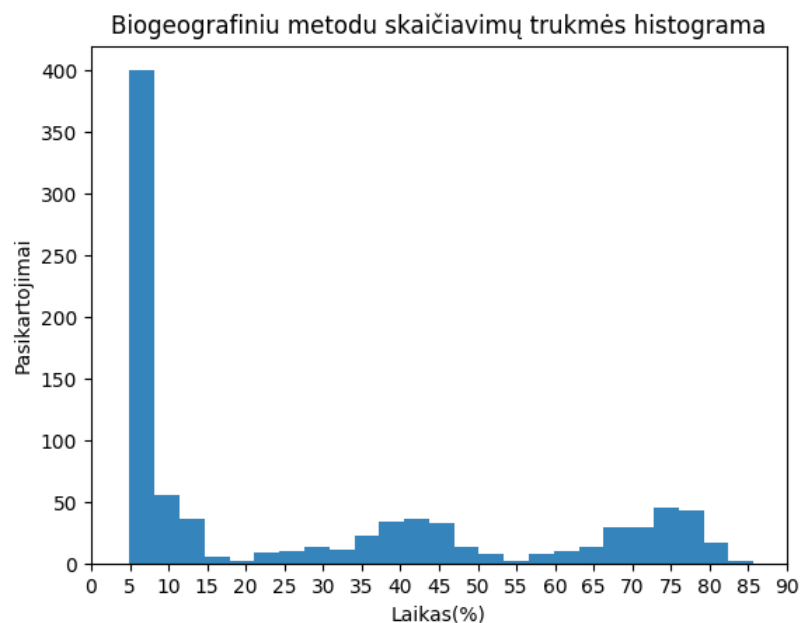
Antrojo intervalo tikslumai rodo geresnį rezultatą, iš karto matyti, kad žemiausia riba pasislinko link geresnio geresnės kokybės. Imant bendrą vidurkį, rezultatai irgi pagerėjo. Vidutinis tikslumas tapo 53.12%. Vadinasi, galima spėti, kad daugėjant atsitiktinai atsirandančių nuspalvintų taškų kiekiui, gerėja ir tikslumas.



Pav. 13. Paveikslėlyje pavaizduota biogeografiniu metodu atliktų skaičiavimų trečiojo intervalo tikslumo histograma. Vertikaliuoje ašyje atvaizduojami atsikartojimai, o horizontalioje ašyje atvaizduoti tikslumai išreikšti procentais.

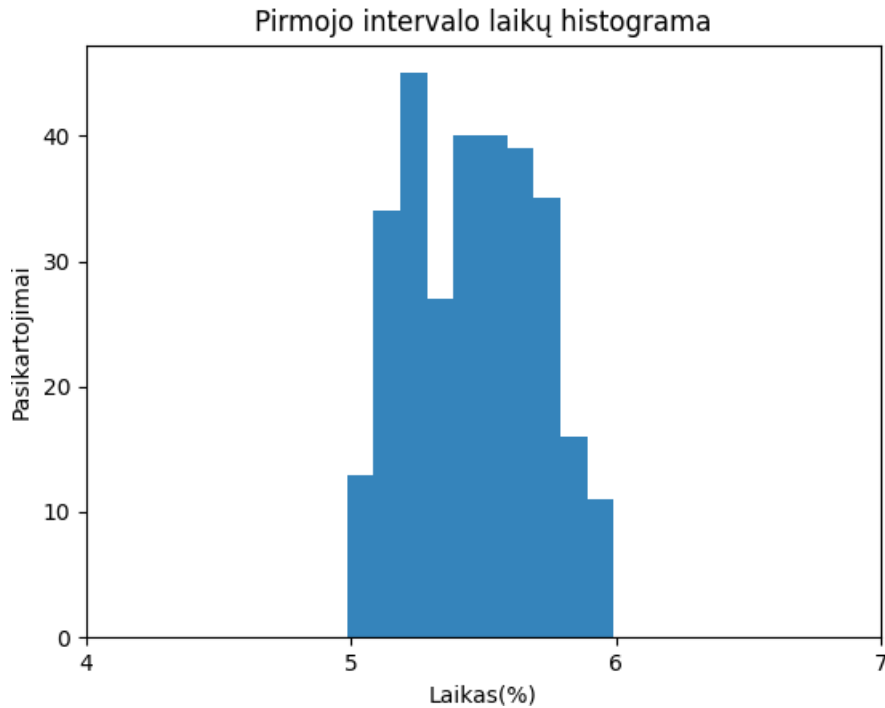
Panašu, kad anksčiau minėtas spėjimas pasiteisino. Bendras vidurkis jau viršija vidurio ribą ir pasiekia 65.10% Tikslumą. Vizualiai taip pat matosi, kad didžioji dalis rezultatų susitelkusi ties 60–70 procentų riba. Todėl tikrai galima priimti išvadą, kad esant didesnei nuspallvintų taškų koncentracijai, tikslumas taip pat gerėja.

Tačiau svarbu atsižvelgti ir į laiko dalį, kurią užima skaičiavimai (Pav. 14).



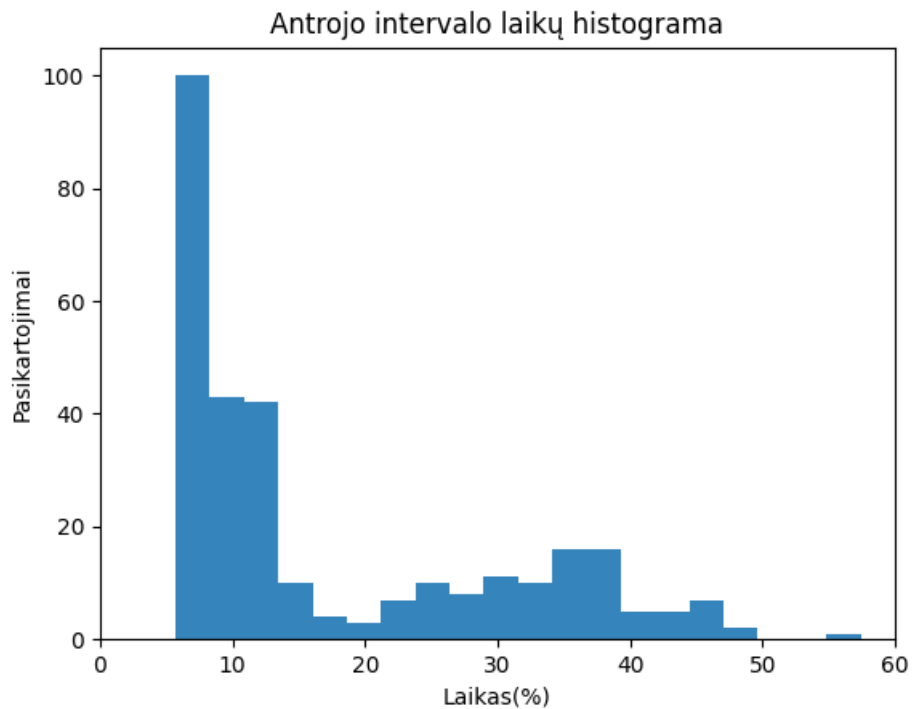
Pav. 14. Paveikslėlyje pavaizduota biogeografiniu metodu atliktų skaičiavimų laiko histograma. Vertikaliuoje ašyje atvaizduojami atsikartojimai, o horizontalioje ašyje atvaizduota laiko dalis procentais, lyginama su pilnojo perrinkimo metodu užimtu laiku.

Ganėtinau nesunku pastebėti, kad lyginant su pilnuoju perrinkimu, rezultatai tiesiog stulbina. Didžioji dalis skaičiavimų trunka iki 25% laiko. Imant tikslų vidurkį, skaičiavimų trukmė vidutiniškai tesiekia 28,56%. Tad apytikriai, vykdant tuos pačius skaičiavimus biogeografiniu metodu 3 kartus iš eilės, vis tiek rezultatą gautume greičiau. Toliau derėtų panagrinėti, kokie laiko rezultatai gaunami kiekviename iš intervalų (Pav. 15–17).



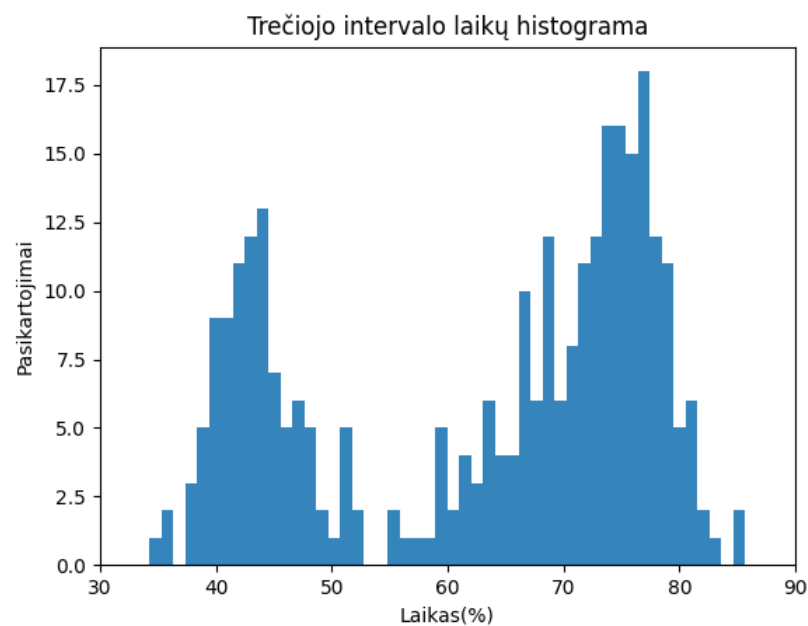
Pav. 15. Paveikslėlyje pavaizduota biogeografiniu metodu atliktų skaičiavimų pirmojo intervalo laikų histograma. Vertikalioje ašyje atvaizduojami atsikartojimai, o horizontalioje ašyje atvaizduota laiko dalis procentais, lyginama su pilnojo perrinkimo metodu užimtu laiku.

Iš aukščiau pavaizduoto grafiko, galima pastebėti, kad pirmajame intervale skaičiavimai vyksta apytikriai 20 kartų greičiau. Vidutiniškai skaičiavimai trunka 5.46% pilnojo perrinkimo užimamo laiko. Nederėtų pamiršti, kad skaičiavimų kokybė taip pat nėra pati geriausia. Žinant šią informaciją, galima daryti išvadą, kad vykdant skaičiavimus, kuomet faile yra mažai nuspalvintų pikselių, reikėtų nepagailėti padidinti individų kiekį ir galbūt padidinti ciklų skaičių. Tai reikalinga todėl, kad skaičiavimams vykstant tokį trumpą laiką, jo pailgėjimas dvigubai ar trigubai nešytų žymiai didesnę naudą, jei atitinkamai pagerėtų skaičiavimo tikslumas.



Pav. 16. Paveikslėlyje pavaizduota biogeografiniu metodu atliktų skaičiavimų antrojo intervalo laikų histograma. Vertikalioje ašyje atvaizduojami atsikartojimai, o horizontalioje ašyje atvaizduota laiko dalis procentais, lyginama su pilnojo perrinkimo metodu užimtu laiku.

Šiame grafike aiškiai matosi, kad laiko trukmė ima varijuoti, tačiau didžioji dalis laiko vis tiek užima poziciją ties 10%. Tikslus vidurkis tėra 17.32%, vadinasi ir šiuo atveju galima būtų padidinti individų kiekį ir prailginti evoliucijos ciklą, tačiau parametrus jau reiktų rinktis atsargiau. Bendrai, lyginant su tikslumu, kurį duoda šiame intervale vykdomi skaičiavimai, rezultatai yra patenkinami.



Pav. 17. Paveikslėlyje pavaizduota biogeografiniu metodu atliktų skaičiavimų Trečiojo intervalo laikų histograma. Vertikalioje ašyje atvaizduojami atsikartojimai, o horizontalioje ašyje atvaizduota laiko dalis procentais, lyginama su pilnojo perrinkimo metodu užimtu laiku.

Trečiuoju atveju, galima pastebėti, kad laiko trukmės pasiskirsto į dvi dalis. Galima tikrai spėti, kad pirmoji intervalo skaičiavimų dalis susitelkė ties 40–50 procentų riba, o likusi – link ilgesnės atkarpos. Vidutinė skaičiavimų trukmė yra 62.88%. Jau matyti, kad laiką norėtusi šiek tiek trumpinti, tačiau rezultatų tikslumas gali nukentėti.

4.1. Analizės apibendrinimas ir pastebėjimai

Pastebėta, kad biogeografinis algoritmas gali veikti ir dešimteriopai greičiau, nei kiekvieno įmanomo varianto paieška. Žinoma, rezultatai ne visuomet yra absoliučiai tikslūs, tačiau priklausomai nuo situacijos, galima laikyti, kad visgi šie rezultatai yra tenkinantys lūkesčius. Kiekvienas individualus atvejis gali pareikalaus individualių parametrų. Šių parametrų parinkimą galima būtų ir automatizuoti, tačiau tai nėra šio tyrimo esminis dalykas. Pirmajame intervale, kuris apima pirmųjų 300–ų failų rezultatus, laikas pranoksta visus lūkesčius, tačiau kokybė yra tobulintina. vidutiniškai skaičiavimai vyksta apie 5% pilnojo perrinkimo skaičiavimo laiko ir veikia apie 42% tikslumu. Antrajame intervale rezultatai yra vidutiniški, laikas vis dar stebina, taip pat ir kokybė nebėra tokia prasta. Tikslumas yra apie 53%, o laikas skirtas skaičiavimams yra daugiau nei 5 kartus greitesnis. Trečiasis intervalas jau rodo riziką laiko klausimais, tačiau vis dar rodo potencialą. kokybė taip pat yra geresnė nei vidutinė. Nors laikas jau pasiekia 63% ribą, tikslumas tuo tarpu perkopia 65%. Svarbu paminėti, kad tai nėra viršutiniai ar apatiniai rėžiai, tad priklausomai nuo užduoties kiekviename intervale, rezultatai gali būti tiek geresni, tiek blogesni. Galų gale algoritmo įgyvendinimas yra tobulintinas, taip pat ir parametrų bei modifikacijų kombinacijos. Pastebėta, kad metodas veikia itin sparčiai mažai nuspaldintuose failuose, tačiau nuo to gali nukentėti ir kokybė, tačiau galimas sprendimas, kompensuoti kokybę laiko kaina, kadangi šiuo atveju, tai tikrai galimas variantas. Kitų intervalų ribose parametrus galima tobulinti, galima ir palikti tokius, kokie jie yra. Bendrai susumavus rezultatus, Galima pastebėti, kad biogeografinis algoritmas veikia greičiau, nei pilnojo perrinkimo metodas. Algoritną optimizavus ir pritaikius kiekvienai iš užduočių, metodas gali atnešti ir dar geresnius rezultatus.

Rezultatai ir išvados

Darbe buvo apžvelgti įvairūs evoliuciniai metodai bei pilnojo perrinkimo metodas. Taip pat aprašytas biogeografinio algoritmo veikimas ir modifikacijos galimybės. Darbe iškeltas tikslas – ištirti šio evoliucinio metodo veikimą, pritaikant jį konkrečiai užduočiai. Konkreti užduotis buvo vienspalvio didžiausio įmanomo kvadrato 64×64 pikselių paveikslėlyje paieška. Išanalizavus rezultatus pastebėta, kad biogeografinis algoritmas gali veikti ir dešimteriopai greičiau, nei pilnojo perrinkimo metodu atliekama paieška. Žinoma, rezultatai ne visuomet yra absoliučiai tikslūs, tačiau, priklausomai nuo situacijos, galima laikyti, kad visgi šie rezultatai yra tenkinantys lūkesčius. Tikslumas priklauso nuo paveikslėlio nudažymo ploto – tuose paveikslėliuose, kuriuose nuspalvintų pikselių yra mažai, metodas nėra toks tikslus. Tačiau paveikslėliuose, kuriuose pikselio nudažymo tikimybė yra didelė, algoritmas teikia gerus rezultatus. Taip pat pastebėta, kad metodo įgyvendinimą galima tobulinti ir pritaikyti parametrus pagal kiekvieną individualų atvejį. Spėjama, kad atlikus šiuos veiksmus biogeografinis metodas bus dar naudingesnis.

Literatūra

- [Bou11] Ilhem Boussaid. Hybridizing biogeography-based optimization with differential evolution for optimal power allocation in wireless sensor networks, 2011.
- [Dai20] Zhuo Dai. A novel hybrid optimization algorithm combined with bbo and aco, 2020.
- [Dep] Shiyao Department of Mechanical Engineering Hubei University of Automotive Technology. Decomposition and adaptive weight adjustment method with biogeography/complex algorithm for many-objective optimization. accessed 2021-05.
- [H P11] D.Simon H. P. Ma. Blended biogeography-based optimization for constrained optimization. XXIV, 2011.
- [Rob67] Edward Osborne Wilson Robert Helmer MacArthur. The theory of island biogeography, 1967.
- [Sima] Dan Simon. Biogeography-based optimizations article https://www.researchgate.net/publication/224309395_Biogeography-Based_Optimization. accessed 2020-09.
- [Simb] Dan Simon. Biogeography-based optimizations website <http://embeddedlab.csuohio.edu/BB0/>. accessed 2020-09.
- [Sim13] Dan Simon. Evolutionary optimization algorithms, 2013.
- [Wal76] Alfred Russel Wallace. The geographical distribution of animals. I, 1876.
- [Zha15] Yu-Dong Zhang. Pathological brain detection in magnetic resonance imaging scanning by wavelet entropy and hybridization of biogeography-based optimization and particle swarm optimizationpathological brain detection in magnetic resonance imaging scanning by wavelet entropy and hybridization of biogeography-based optimization and particle swarm optimization, 2015.
- [Zha19] Xinming Zhang. Improved biogeography-based optimization algorithm and its application to clustering optimization and medical image segmentation, 2019.

Biogeografija paremti optimizavimo metodai

Simonas Suprinavičius

Santrauka

Šiame darbe atliekama vieno iš evoliucinių skaičiavimo metodų – biogeografijos – analizė. Darbo metu apžvelgiami įvairūs evoliuciniai metodai, tuomet aprašomi biogeografinis ir pilnojo perrinkimo algoritmai. Toliau atliekamas tyrimas, kurio metu, analizuojama biogeografinio optimizavimo metodo nauda konkrečiai problemai spręsti. Konkreti problema – turint 64×64 pikselių paveikslėlį, rasti, konkrečia spalva nuspalvintuose plotuose, didžiausią nudažytą kvadratą, kurio kiekvienas pikselis yra tos pačios spalvos. Rezultatai parodo, kad tiriamas metodas gali paspartinti paiešką ir rasti rezultatus greičiau, nei pilnojo perrinkimo algoritmas.

Raktiniai žodžiai: biogeografija, optimizavimas, evoliuciniai metodai, paieška.

Biogeography Based Methods of Optimization

Simonas Suprinavičius

Summary

This thesis includes biogeography – based optimization method analysis which is one of the evolutionary optimization methods. Various evolutionary methods are reviewed as well as biogeographical and full re-sampling algorithms are described. Research includes the analysis of benefits of biogeography – based optimization method for solving a specific problem. The problem is defined: to find the largest colored square in each color-shaded area with each pixel of the same color in 64×64 pixel image. The results shows that the method of the study can speed up the search and give results faster than the full re-sampling algorithm.

Keywords: biogeography, optimization, evolutionary methods, search.

Priedai

Priedas nr. 1

Programos ir tyrimo failai.

<https://github.com/semkius1913/bakalauras>

Norint atkartoti tyrimą, reikėtų failus įsikelti į „Google colab“ platformą ir atitinkamai susidėti į aplankus pagal programose nurodytus kelius.