



UVODNA RAZMATRANJA O EKSPERIMENTALNOJ ANALIZI STOHAŠTIČKIH OPTIMIZACIONIH ALGORITAMA

INTRODUCTORY CONSIDERATIONS ABOUT EXPERIMENTAL ANALYSIS OF STOCHASTIC OPTIMIZATION ALGORITHMS

TATJANA JAKŠIĆ KRÜGER¹, TATJANA DAVIDOVIĆ¹

¹ Matematički institut Srpske akademije nauka i umetnosti, Beograd, {tatjana,tanjad}@mi.sanu.ac.rs

Rezime: Ovaj rad je posvećen ispitivanju performansi (meta)heurističkih metoda i njihovom međusobnom poređenju. Poslednjih godina u velikom broju publikovanih radova u oblasti operacionih istraživanja su implementirane preporuke o pravilnom sprovođenju eksperimenata radi utvrđivanja efikasnosti predloženih (meta)heurističkih metoda. Najveći uticaj došao je iz oblasti primenjene statistike i mašinskog učenja, te danas imamo listu pravila koje istraživači treba da poštuju pre publikovanja radova. Istraživači su danas suočeni sa rigoroznijom metodologijom rada, a eksperimentalni radovi su sve zahtevniji. Ovim radom želimo da doprinesemo daljem usmeravanju pažnje na aktuelnosti iz oblasti eksperimentalne analize metaheurističkih metoda.

Ključne reči: Stohastički algoritmi, metodologija istraživanja, statistika, planiranje eksperimenata

Abstract: The paper describes standard methodology for investigation of (meta)heuristic methods and their comparisons. Recently, a large body of papers in OR has implemented some of the recommendations for valid experimental analysis of presented (meta)heuristic methods. The largest influence comes from the applied statistics and machine learning, thus, we are able to infer a list of rules which should be followed before publishing. Researches are faced with rigorous requirements related to the experimental methodology. We aim to point out the importance of comprehensive methodology and its implementation in the research work.

Keywords: Stochastic algorithms, research methodology, statistics, design of experiments.

1. UVOD

Tema ovog rada je pregled literature i uvod u sistematizaciju izvršavanja eksperimentalne analize performansi metaheurističkih metoda i njihovog međusobnog poređenja (*eng.* benchmarking). Računarski eksperimenti su postali neophodni kod istraživanja algoritama. Naime, izučavanje najgoreg i očekivanog slučaja u teoriji složenosti izračunavanja nije dovoljno da se izvedu predikcije za različite klase problema, pa čak i različite instance problema [Ognjanović and Krdžavac, 2004]. Moramo se suočiti sa time da je izučavanje formalnih sistema često poželjno sprovesti empirijskim (eksperimentalnim) putem [Hooker, 1994]. Zahvaljujući razvoju različitih grana matematike i računarstva (pre svega mašinskom učenju), istraživači u oblasti operacionih istraživanja imaju na raspolaganju bogatu literaturu o pravilnoj primeni strategija eksperimentalne analize na (meta)heurističke (MH) metode. U novije vreme se od istraživača očekuje da opišu i razjasne strukture predloženih implementacija metoda, počev od izabranih vrednosti parametara. Svaka konkretna implementacija MH metode predstavlja jedan stohastički algoritam čiju efikasnost je potrebno utvrditi i uporediti sa nekim drugim algoritmima.

Sve češće se susrećemo sa zahtevima za primenu metode *planiranja eksperimenata* (Design of Experiments-DOE) [Montgomery, 2001]. Osnova ovog zahteva je da se konkretan stohastički algoritam, predložen za rešavanje odgovarajućeg problema optimizacije, ispita tako da se rezultati mogu reprodukovati na istom ili drugačijem računarskom sistemu, od strane istih ili drugih istraživača. Implementacije MH metoda uglavnom zahtevaju nameštanje većeg ili manjeg broja parametara radi što boljih performansi tokom izvršavanja. Nameštanje (*tuning*) parametara dobijenog stohastičkog algoritma se može sprovesti raznim metodama, pri čemu su DOE postale standard u literaturi.

Čiljeve eksperimentalne analize MH metoda možemo razvrstati u dve kategorije: (1) razumevanje implementiranih algoritama i njihovih performansi, (2) poređenje sa implementacijama drugih MH metoda (*benchmarking*). Pod razumevanjem se podrazumeva eksperimentalno izučavanje osetljivosti algoritma na izmene njegovih modula ili vrednosti parametara (*parameter tuning* i *sensitivity analysis*). Metodologija ispitivanja je slična za

obe kategorije, i zato ćemo se truditi da ukažemo i na njihove razlike. U ovom radu se fokusiramo na probleme sa jednom funkcijom cilja, ali naša razmatranja se mogu iskoristiti kod višekriterijumske analize. Skrećemo pažnju da kod metaheuristika pitanje o konvergenciji ka optimalnom rešenju nije suštinsko, već se odnosi na trošenje računarskih resursa radi generisanja dovoljno dobrog (suboptimalnog) rešenja [Gutjahr, 2010].

Rad se sastoji od 5 odeljaka. Drugi odeljak daje pregled radova važnih za eksperimentalna izučavanja u računarstvu. Odeljak 3 opisuje metodologiju rada i definicije relevantne kod prijavljivanja rezultata. Odeljak 4 sadrži istraživačke korake neopohodne pre publikovanja rada. Lista koraka data u literaturi je vrlo iscrpna i potrudili smo se da apsolutnom početniku predočimo najnužnije stavke. Zaključne napomene date su u odeljku 5.

2. PREGLED LITERATURE

Istorijski gledano, heurističke metode optimizacije su predlagane da bi se što brže generisala dobra (suboptimalna) rešenja postavljenog optimizacionog problema. Za NP-teške probleme uspešno su razvijene razne MH metode. U neretkim slučajevima nije bilo bitno da li se rešenje može ponoviti, dok god je bilo "najbolje izloženo u javnosti". Kao rezultat, istraživači su najčešće prijavljivali samo najbolje rešenje i vreme za koje je ono generisano. Problem sa ovim pristupom je neponovljivost istog rezultata MH metodom, čak i na istom računarskom sistemu. Takođe, u literaturi se predložena implementacija često opiše samo kroz listu glavnih karakteristika, bez predstavljanja celog kôda. S obzirom na to da je ponašanje rezultujućeg algoritma stohastičko, čak i kada je kôd javno dostupan pitanje je za koje seme (*seed*) je generisano dato rešenje.

U prethodnih 30 godina nekoliko istraživača pokušava da skrene pažnju na postojeće principe istraživanja kod MH metoda. U nastavku izdvajamo njihove radove, kritike i predloge koji su doprineli sistematizaciji analize MH metoda, predstavljanju rezultata, ispitivanju i poređenju njihovih performansi. Pregled je dat hronološki.

Počecemo od [McGeoch, 1987, McGeoch, 1996] koja je doprinela razvijanju oblasti, poznate danas kao *eksperimentalna algoritmika* (eng. experimental algorithmics). Koncept ove discipline je da se ustanove i standardizuju statističke i eksperimentalne tehnike radi objektivne analize stohastičkih algoritama [McGeoch, 2012]. U [McGeoch, 1996] se ističe ispitivanje algoritama radi njihovog sveobuhvatnog razumevanja. Poruka rada je da analiza algoritma podrazumeva mnogo više od izbora adekvatne mere performansi ili pogodne klase instanci problema. Naime, prvi korak je modeliranje ili *simuliranje algoritma* koji će se izvršavati u realnim situacijama. Simulacija proističe kao posledica osetljivosti algoritma na karakteristike instanci problema, uticaj strukture podataka na brzinu izvršavanja, izbor računarskog sistema itd. Već sada možemo da vidimo da je eksperimentalna studija MH metoda zahtevnija od studije egzaktnih algoritama optimizacije.

U [Hooker, 1994] se naglašava važnost empirijske metodologije kod proučavanja MH metoda. Kao i mnogi drugi, autor je prepoznao da analiza najgoreg i očekivanog slučaja nisu dovoljni da se predvidi kako će se odgovarajući algoritam ponašati nad realnim test instancama. U radu se predlaže da se često zastupljeno *takmičarsko testiranje* (eng. *competitive testing*), gde jedan algoritam pobeđuje drugi generisanjem boljeg rešenja za kraće vreme, zameni sa *kontrolisanim eksperimentom* (eng. *controlled experiment*). Zapravo, metodama eksperimentalne algoritmike možemo da ukažemo na značajnu vezu između performansi algoritma i određenih karakteristika problema. Međutim, karakteristike problema koje utiču na težinu rešavanja nisu lake za uočavanje. Pored očiglednih razlika računarskih sistema, na negativne strane takmičarskog testiranja utiču i razlike u stilu programiranja, nameštanju parametara i uloženom trudu.

Rad [Barr et al., 1995] je možda najuticajniji na temu eksperimentalne analize MH metoda. U radu su date smernice za ispitivanje i prijavljivanje rezultata računarskih eksperimenata sprovedenih nad implementacijama MH metoda. Uspostavljena je terminologija koju je usvojila šira naučna zajednica.

Kritički osvrt na eksperimentalno istraživanje u evolutivnom računarstvu (*evolutionary computation*, EC), podjednako je primenljiv na sve MH metode dat je u radu [Eiben and Jelasity, 2002]. Autori analiziraju poređenje implemetiranih stohastičkih algoritama na čitavim klasama problema i ističu značaj pravilnog izbora skupa test instanci. Naime, klasa problema ne bi trebalo da je previše uopštena ili specifična. Opštost klase problema mora da odgovara cilju istraživanja, odnosno, da omogućiti uočavanje i opisivanje razlika između uporednih algoritama ili varijanti jednog algoritma. Tako na primer, ukoliko istaknemo da je problem NP-težak, odrednica je isuviše uopštena i nedostaju informacije o strukturi samog problema. Slično je ako navedemo da se bavimo problemom raspoređivanja jer je ovo izuzetno široka klasa problema. Takođe, instance jedne klase problema bi trebalo da se znatno razlikuju, ali da sadrže neku zajedničku karakteristiku. Na ovaj način već možemo da okarakterišemo kao kvalitetan algoritam onaj koji ne pokazuje značajno različito ponašanje na primerima unutar razmatrane klase. U [Eiben and Jelasity, 2002] autori predlažu da se kod nameštanja parametara algoritma koristi pristup sličan kao kod mašinskog učenja. Naime, da bismo generisali povoljnu konfiguraciju parametara razmatranog algoritma, iz klase problema *C* izaberemo skup instanci za *treniranje*. Da bismo uporedili različite algoritme koristi se *skup za testiranje*. U radu možemo da nađemo i vrlo korisnu diskusiju oko uticaja ciljeva

istraživanja na izbor mere performansi MH metode. Naime, često u praksi imamo jedan od dva scenarija: 1. dato je dovoljno vremena da ponovimo izvršavanja (dane ili nedelje) ali nas interesuje najbolje rešenje, 2. možemo algoritam da izvršimo samo jednom. U prvom slučaju performanse algoritma procenjujemo upotrebom najboljih rešenja generisanih za predefinisano vreme ili broj evaluacija. Međutim, ukoliko nam je potreban algoritam koji ćemo u praksi pustiti jednom, preporučljivo je upotrebiti srednji kvalitet rešenja. Zato je interesantno da, dok se prvi scenario češće dešava u praksi, istraživači češće prijavljuju aritmetičku sredinu. Međutim, skrećemo pažnju da kod nameštanja parametara algoritma imamo daleko veći izbor za mere performansi.

U radu [Czarn et al., 2004] autori su identifikovali tri bitne faze kod izvođenja empirijske analize GA: kontrolisanje slučajnog semena (eng. blocking for seed), određivanje snage testa (eng. calculating power) i analiza krive rezultata. Predložene faze su nezavisne od GA strukture i lako se izvode za sve tipove MH metoda. Zaključak rada je snažna preporuka za primenu statistike kod razumevanja prirode parametara MH metode i njihovih interakcija, pre svega između karakteristika instanci problema i kvaliteta rezultata. Zbog bogatstva literature i nedostatka prostora, pregled završavamo ovde, a zainteresovane čitaoce upućujemo na noviju literaturu [Eiben and Smit, 2011, Kendall et al., 2016, Beiranvand et al., 2017, Bartz-Beielstein et al., 2020].

3. DEFINICIJE I METODOLOGIJA

Kod *računarskog testiranja* algoritma, *eksperiment* se sastoji od rešavanja niza instanci problema (primera) pod kontrolisanim uslovima. Faktori su bilo koje promenljive koje možemo kontrolisati, a koje utiču na ishod eksperimenta. Kod MH metoda, to su izbor problema, implementacija, računarsko okruženje, izbor mere performansi, izbor parametara i prijavljivanje rezultata. Svaki od ovih faktora može da ima značajan uticaj na ishod eksperimenta. Zbog toga je bitno da se uspostavi što objektivniji način analize, što je omogućeno ukoliko se koristi planiranje eksperimenata.

MH metode su popularne kao alat koji brzo generiše rešenja visokog kvaliteta. Njihova prednost je mogućnost primene na različite klase problema. Zato je važno da se prilikom eksperimentalne analize pokažu osobine kao što su: preciznost, jednostavnost, inovativnost, brzina, robusnost i uopštenitost [Barr et al., 1995]. U novijoj literaturi sve je više radova o izučavanju različitih aspekata performansi implementacija ponuđene ili postojeće MH metode. Ovo je nastalo kao posledica loše prakse gde se za poređenje koristi isključivo jedna mera performansi, kao što je *najbolje rešenje* (eng. best solution, x_b), *najgore rešenje* (eng. worst solution, x_w), ili *usrednjeno rešenje* (eng. average solution, \bar{x}). Međutim, postoje i drugi parametri koji su podjednako bitni da bismo ustanovili objektivan prikaz implementacije neke metode ili izvršili fer poređenje. Iz literature izdvajamo tri bitna pitanja:

- Izbor mere performansi,
- Izbor i opis problema i instanci problema,
- Ispitivanje metode i njenih implementacija.

Ova tri pitanja su međusobno povezana tokom naše analize. Kod ispitivanja rezultujućeg algoritma, jednostavnost i preciznost su prioritet [McGeoch, 2012]. Krajnji cilj je razviti algoritam koji će efikasno rešavati instance jedne klase problema, što nas upućuje na njihov pravilan izbor, kao i izbor parametara, pa time i na odgovarajuće mere performansi. Takođe je važno da instance problema budu reprezentativne za svoju klasu problema, kako bismo osnažili jačinu naših argumenata. Drugi važan cilj je *robustnost* algoritma, pri čemu mislimo na što manju varijabilnost rezultata izazvanu promenama: vrednosti za seme, parametara i izbora instance problema [Eiben and Smit, 2011]. Ukoliko algoritam pokazuje veliku varijabilnost za različite vrednosti semena, onda je preporučljivo korigovati odgovarajuće segmente kôda.

Mere performansi svrstavaju se u dve grupe, one koje utiču na *efektivnost* i na *efikasnost* algoritma [Jakšić Krüger, 2017]. Pod efektivnošću se podrazumeva sakupljanje podataka o kvalitetu rešenja u zadatim vremenskim okvirima (ili za maksimalni broj evaluacija). U tom slučaju se opisu performansi algoritma pridodaje statistika kao što su mera centralne tendencije i varijabiliteta. Poznate mere efektivnosti su aritmetička sredina i standardna devijacija, odnosno, medijana i razmak varijacije.

Ukoliko želimo da saznamo koliko je vremena ili iteracija potrebno da algoritam pronađe neko unapred zadato (sub)optimalno rešenje, tada izučavamo njegovu efikasnost. Mera efikasnosti koja se najčešće koristi je *dužina rada* (eng. run-length). Ona predstavlja dužinu izvršavanja, mereno vremenom ili brojem diskretnih koraka (kao što su iteracija ili evaluacija), potrebnog da se generiše rešenje jednako ili kvalitetnije od nekog unapred zadatog rešenja [Hoos and Stützle, 2007]. Međutim, analiza dužinom rada nije uvek moguća. Naime, ukoliko je problem takav da algoritam ne može da ispuni dato ograničenje onda je preporučljivo koristiti mere centralne tendencije [Eiben and Jeliasity, 2002]. U tom slučaju se preporučuje statistička deskriptivna analiza, odnosno, pored vrednosti funkcije cilja za x_b , x_w i \bar{x} , koriste se grafički prikazi kao što su box-plot, histogram ili kriva rezultata. Za to nam mogu pomoći gotovi softverski paketi za analizu podataka (R, SPSS, Excel).

Analiza performansi algoritamske implementacije MH metode se može vršiti na dva načina, grafički i statistički. Grafička analiza se oslanja na vizualne prikaze kojima dobijenu reprezentativnu vrednost (ili krivu) možemo uporediti sa drugom da bi se ustanovila dominacija jedne implementacije u odnosu na drugu [Hoos and Stützle, 2007]. Statistička analiza uglavnom podrazumeva primenu nekog statističkog testa (parametarskog ili neparametarskog) nad uzorcima dobijenim izvršavanjima odgovarajućeg algoritma za različita semena [Bartz-Beielstein et al., 2020]. Ukoliko ti uzorci imaju Gausovu raspodelu, primenjuju se parametarski testovi, u suprotnom treba koristiti neparametarske. Pravi izbor statističkog testa zavisi i od veličine uzoraka, zavisnosti između podataka, tipa podataka, itd. Vizualizacija rezultata se može iskoristiti i da bi se utvrdilo da li važi normalna raspodela rezultata izvršavanja, što pomaže prilikom izbora pravilnog statističkog testa. Literatura je danas bogata radovima o automatizaciji nameštanja parametara. Izdvojicemo PARAMLS, *F-Race (irace)*, SPO i SMAC kao gotova softverska rešenja koja se često primenjuju [Eiben and Smit, 2011].

Iste mere se mogu primeniti i prilikom poređenja implementacija više različitih metoda, ali se tada može pojaviti problem različitih računarskih sistema, nedostajanja adekvatnih podataka za poređenje, itd. Prema savetima iz eksperimentalne algoritmičke, poređenja bi trebalo izvršavati na istom računarskom sistemu [McGeoch, 2012]. U najgorem slučaju, ukoliko nije moguće reprodukovati program, očekuje se poređenje sa nekom od standardnih tehnika kao što su greedy search ili local search [Barr et al., 1995]. Prilikom implementacije MH metoda postoji više stvari o kojima moramo voditi računa. Počnimo od primedbe da razni implementacioni trikovi mogu da utiču na performanse više nego same algoritamske modifikacije [McGeoch, 1996]. Zapravo, govorimo o određenim nivoima specifikacije neke metode (eng. instantiation). Na primer, VNS metoda je minimalno specifikiran algoritam, te mnogi detalji mogu uticati na performanse implementacije, npr. tipovi struktura podataka. Zato se preporučuje da se, kad god je to moguće, kôd učini javno dostupnim ili navede što više specifikacija.

Osobine koje analiza nekog stohastičkog algoritma mora da zadovolji su *postojanost*, *ponovljivost* i *reproducibilnost* izvršavanja [Bartz-Beielstein et al., 2020]. Sve tri reči na srpskom su sinonimi, ali ćemo ih iskoristiti radi uvođenja tri važna koncepta istraživanja. Ovi izrazi su izvedeni od engleskih izraza *repeatability*, *replicability* i *reproducibility* i predstavila ih je ACM zajednica (Association for Computing Machinery) (<https://www.acm.org/publications/policies/artifact-review-badging>). Postojanost rezultata kod izvršavanja algoritma podrazumeva smernicu "isti tim, isto eksperimentalno okruženje". Pod ovim smatramo da istraživač može uvek da generiše identične rezultate i zaključke svog istraživanja, pod uslovom da koristi jedan računarski sistem. Ukoliko drugi istraživač može da ponovi rezultate i zaključke na istom računarskom sistemu, to nazivamo ponovljivošću. Konačno, cilj svakog istraživanja i opisa rezultata je da bilo ko može da proizvede identične rezultate potpuno nezavisno od računarskog okruženja. Ovo nazivamo reproducibilnost, a naziv je rezervisan za najopštiju grupu ciljeva istraživanja jer je dugo u upotrebi.

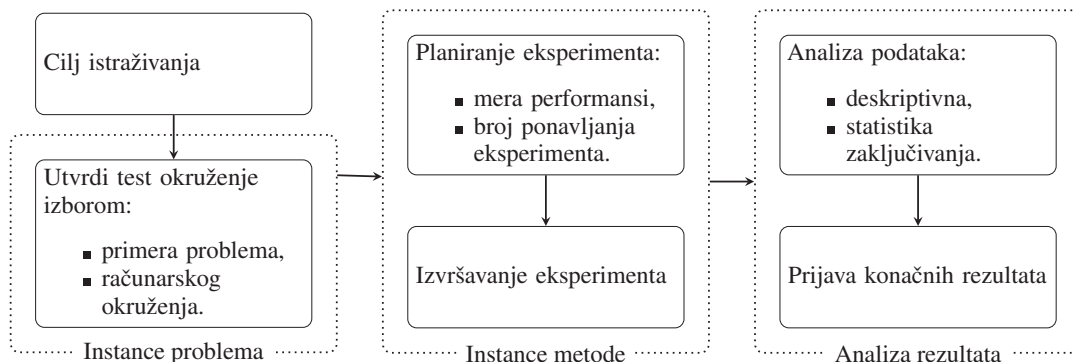
4. PLANIRANJE EKSPERIMENATA KOD MH METODA

U ovom odeljku želimo da iznesemo nekoliko predloga za objektivno istraživanje implementacija MH metoda sažetih u diagramu predstavljenom na Sl. 1. Da bismo pristupili eksperimentalnoj analizi, od izuzetnog je značaja uspostaviti pravilne ciljeve ispitivanja, a ovo je praktičnije formulisati kao istraživačko pitanje. Napominjemo da se ovde ne bavimo početnim fazama dizajniranja programa, već isključivo kasnijim, odnosno, jednom kada smo kompletirali program, šta je sledeće?

Istraživači koji zagovaraju kontrolisane eksperimente smatraju da je potrebno uložiti neko vreme u analizu instanci problema. Izbor instanci problema i njihovo analiziranje često je subjektivno. Kada smo u mogućnosti da se opredelimo za neki skup instanci problema, važno je raditi sa primerima čije su težine raznovrsne, a optimalno (najbolje poznato) rešenje dato [Beiranvand et al., 2017]. U [Bartz-Beielstein et al., 2020] je predloženo kako najbolje proceniti kvalitet skupa instanci problema.

Početna faza istraživanja je često pilot studija. Ne znajući sa čime se susreću i sa željom da sakupe što više podataka istraživači koriste ili ceo tok istraživanja opisan na Sl. 1 ili nisu toliko restriktivni. Naime, ukoliko primetimo veliku varijabilnost u kvalitetu rešenja za jednu klasu problema, poželjno je utvrditi da li i koje karakteristike kod instanci problema dovode do ovakvih rezultata. Cilj je grupisati one instance problema koje su iste težine za posmatrani algoritam.

Uglavnom se podrazumeva da se ispitivanje parametara i poređenje različitih varijanti implementacije MH metode, vrši na istom računarskom sistemu, te da sve procese možemo da kontrolišemo. Na taj način smo u mogućnosti da kontrolišemo najvažnijih faktor koji podstiče nasumičnost kod rezultujućih algoritama: seme generatora slučajnih brojeva. Ovaj parametar iziskuje da njegovom promenom ne možemo predvideti ishod izvršavanja. Kod eksperimentalne statističke analize, seme predstavlja smetnju (eng. nuisance) jer nije tema eksperimenta [Montgomery, 2001, Czarn et al., 2004]. Zato je prihvatljivo da ga kontrolišemo kako bismo doprineli postojanosti pa i reproducibilnosti rezultata [Jakšić Krüger, 2017]. U [Rardin and Uzsoy, 2001]



Slika 1 Opis toka eksperimentalne analize implementacije MH metoda.

autori preporučuju da algoritme koje poredimo izvršimo nad istim skupom instanci bez obzira na preporuke eksperimentalne statistike da se sve što se ne kontroliše eksplicitno bira nasumice.

U drugoj fazi istraživanja potrebno je izvršiti planiranje eksperimenta (Sl. 1). Izbor mere performansi se oslanja na istraživačka pitanja. Na primer [Beiranvand et al., 2017], da li nam je važnije da algoritam rešava svaku instancu problema ili želimo najbolje srednje vreme rada? Kroz fazu nameštanja parametara moramo da prođemo i kada vršimo analizu implementacije jedne MH metode i kada poredimo različite metode. Međutim, mera performansi i broj ponavljanja eksperimenta zavise od toga šta radimo. U slučaju da nameštamo parametre, moguće je koristiti metode statističkog zaključivanja i broj ponavljanja mora biti bar 100 [Beiranvand et al., 2017]. Međutim, u slučaju da poredimo različite metode, primenjuju se mere efikasnosti, a dovoljno je 20 ponavljanja. Ukoliko podaci dolaze sa različitih računarskih sistema, potrebno je izvršiti odgovarajuće skaliranje.

U trećoj fazi pokušavamo da odgovorimo na postavljeno istraživačko pitanje. Grafički prikaz u ovoj fazi istraživanja može da nam ukaže na moguću interakciju između parametara metode ili na apsolutnu dominaciju jedne metode nad drugom sa kojom se poredi (ukoliko to postoji). Jedan od načina da izbegnemo subjektivno donošenje odluka je primena statističkih alata deskriptivne i statistike zaključivanja. Kod poređenja para ili više uzoraka rezultata, cilj je da formulišemo nultu hipotezu H_0 našeg istraživanja (npr. dve metode su generisale identične srednje vrednosti). Ideja je da se izabranim statističkim testom izneta hipoteza opovrgne, pri čemu alternativna (H_1) odražava naše očekivanje. Pre sprovođenja testa, potrebno je odrediti prag osetljivosti α (najčešće ima vrednost 0.05) koji predstavlja gornju granicu za traženu p -vrednost datog statističkog testa. Ukoliko je $p < \alpha$, H_0 može da se odbaci odnosno imamo dovoljno dokaza koji idu u prilog H_1 . U suprotnom, nemamo dovoljno dokaza da odbacimo hipotezu H_0 . Za statistički značajno različite uzorke se predlaže *post-hoc* analiza. Za višestruka poređenja ovo znači detekciju parova koji su proizveli malo p . Takođe, izuzetno je bitno iskoristiti mere *veličina efekta* i *interval pouzdanosti* [Tenjović and Smederevac, 2011]. U tom slučaju možemo se poslužiti Kohenovom tabelom da zaključimo da li su razlike male, srednje ili velike [Sheskin, 2004]. Preporučujemo upoznavanje sa novijim tumačenjem p -vrednosti u [Wasserstein and Lazar, 2016] jer se ustaljena praksa prijavljivanja isključivo p -vrednosti promenila u literaturi [Benjamin and Berger, 2019].

Kod analize rezultata (Sl. 1), zaključci našeg istraživanja treba da daju odgovore na polazna istraživačka pitanja, kao što je pronalaženje najbolje konfiguracije parametara ili identifikacija metode sa najboljim performansama. Međutim, sa stanovišta reproducibilnosti važno je da u radovima iznosimo što više informacija o sprovedenim eksperimentima sa ciljem postojanosti naših tvrdnji i rezultata. Takođe, o nekim pitanjima ne razmišljamo a u stvari su vrlo bitna da se utvrdi robustnost metode. Na primer, želimo da istaknemo performanse metode kao funkciju promene dimenzije problema.

5. ZAKLJUČAK

U radu smo analizirali korake sa kojima bi, po našem mišljenju, trebalo da se najčešće susrećemo prilikom razvoja MH metoda za rešavanja optimizacionih problema. Najčešća zamerka kod starih radova je nedovršen dokaz za dominantnost jedne metode nad drugom, kao i nedovoljno obrazložen kvalitet izabranih vrednosti parametara metode. Pod kvalitetnim vrednostima parametara metode podrazumevamo one vrednosti koje će, u srednjem, generisati dobra rešenja za različite instance razmatranog problema. S druge strane, kada se poredimo sa drugom metodom, ovde nije nužno upotrebiti veliki broj instanci. Naprotiv, često se (meta)heuristike pored

na industrijskim, odnosno, realnim primerima. Znači, vidimo da uopšten skup pravila kod eksperimentalne analize nije jednostavan i sveobuhvatan, te zato u literaturi imamo često opisane standarde za različite slučajeve. Naravno, postoje mnogi scenariji koje nismo pokrili. Smatramo da je pitanje sistematizacije eksperimentalne analize MH metoda još uvek otvoreno i da predstavlja plodno istraživačko polje za budući rad.

Napomena. Ovaj rad je finansiran od strane Ministarstva prosvete, nauke i tehnološkog razvoja Republike Srbije preko Matematičkog instituta SANU.

LITERATURA

- [1] Barr, R. S., Golden, B. L., Kelly, J. P., Resende, M. G. C., and Stewart Jr., W. R. (1995). Designing and reporting on computational experiments with heuristic methods. *Journal of Heuristics*, 1(1):9–32.
- [2] Bartz-Beielstein, T., Doerr, C., Bossek, J., Chandrasekaran, S., Eftimov, T., Fischbach, A., Kerschke, P., López-Ibáñez, M., Malan, K. M., Moore, J. H., et al. (2020). Benchmarking in optimization: Best practice and open issues. arXiv, cs.NE, 2007.03488.
- [3] Beiranvand, V., Hare, W., and Lucet, Y. (2017). Best practices for comparing optimization algorithms. *Optimization and Engineering*, 18(4):815–848.
- [4] Benjamin, D. J. and Berger, J. O. (2019). Three recommendations for improving the use of p-values. *The American Statistician*, 73(sup1):186–191.
- [5] Czarn, A., MacNish, C., Vijayan, K., Turlach, B., and Gupta, R. (2004). Statistical exploratory analysis of genetic algorithms. *Evolutionary Computation, IEEE Transactions on*, 8(4):405–421.
- [6] Eiben, A. E. and Jelasity, M. (2002). A critical note on experimental research methodology in ec. In *Proceedings of the 2002 Congress on Evolutionary Computation (CEC'2002)*, volume 1, pages 582–587.
- [7] Eiben, A. E. and Smit, S. K. (2011). Parameter tuning for configuring and analyzing evolutionary algorithms. *Swarm and Evolutionary Computation*, 1(1):19–31.
- [8] Gutjahr, W. J. (2010). Stochastic search in metaheuristics. In *Handbook of Metaheuristics*, pages 573–597. Springer.
- [9] Hooker, J. N. (1994). Needed: An empirical science of algorithms. *Operations Research*, 42(2):201–212.
- [10] Hoos, H. H. and Stützle, T. (2007). Empirical analysis of randomized algorithms. *Handbook of Approximation Algorithms and Metaheuristics*.
- [11] Jakšić Krüger, T. (2017). *Development, implementation and theoretical analysis of the bee colony optimization meta-heuristic method*. PhD thesis, Fakultet tehničkih nauke, Univerzitet u Novom Sadu. advisers prof. dr. Tatjana Davidović.
- [12] Kendall, G., Bai, R., Błazewicz, J., De Causmaecker, P., Gendreau, M., John, R., Li, J., McCollum, B., Pesch, E., Qu, R., et al. (2016). Good laboratory practice for optimization research. *Journal of the Operational Research Society*, 67(4):676–689.
- [13] McGeoch, C. C. (1987). *Experimental Analysis of Algorithms*. PhD thesis, Carnegie-Mellon University, Pittsburgh. Adviser prof. Dr. Jon Bentley.
- [14] McGeoch, C. C. (1996). Toward an experimental method for algorithm simulation. *Journal on Computing*, 8(1):1–15.
- [15] McGeoch, C. C. (2012). *A guide to experimental algorithmics*. Cambridge University Press.
- [16] Montgomery, D. C. (2001). *Design and analysis of experiments*. John Wiley & Sons.
- [17] Ognjanović, Z. and Krdžavac, N. (2004). *Uvod u teorijsko računarstvo*. Beograd-Kragujevac.
- [18] Rardin, R. L. and Uzsoy, R. (2001). Experimental evaluation of heuristic optimization algorithms: A tutorial. *Journal of Heuristics*, 7(3):261–304.
- [19] Sheskin, D. J. (2004). *Handbook of parametric and nonparametric statistical procedures*. Chapman & Hall/CRC.
- [20] Tenjović, L. and Smederevac, S. (2011). Mala reforma u statističkoj analizi podataka u psihologiji: malo p nije dovoljno, potrebna je i veličina efekta. *Primenjena psihologija*, 4(4):317–333.
- [21] Wasserstein, R. L. and Lazar, N. A. (2016). The asa statement on p-values: context, process, and purpose.