Multidiszciplináris tudományok, 10. kötet. (2020) 4 sz. pp. 214-218 **https://doi.org/10.35925/j.multi.2020.4.25**

FITNESS LANDSCAPE ELEMZÉSI TECHNIKÁK ÁTTEKINTÉSE

Agárdi Anita
tanársegéd, Miskolci Egyetem, Informatika Intézet
3515 Miskolc, Miskolc-Egyetemváros, e-mail: agardianita@iit.uni-miskolc.hu

Kovács László
egyetemi tanár, Miskolci Egyetem, Informatika Intézet
3515 Miskolc, Miskolc-Egyetemváros, e-mail: kovacs@iit.uni-miskolc.hu

Bányai Tamás
egyetemi docens, Miskolci Egyetem, Informatika Intézet
3515 Miskolc, Miskolc-Egyetemváros, e-mail: alttamas@uni-miskolc.hu

Absztrakt

Ebben a cikkben a fitness landscape elemzési technikákat mutatjuk be. A keresési tér elemzése fontos lehet az optimalizáló algoritmusok során. Amikor a kutatók kifejlesztenek egy új algoritmust, azt ismert benchmark adatsoron tesztelik. Ha az eredmények elérik az ismert legjobb megoldást, vagy közel jutnak hozzá, akkor publikálják a megoldásukat. Azonban fontos lehet az optimalizáló algoritmusok elemzése is, hogy valójában miért is bizonyultak jobbnak, mint a meglévő módszerek. A fitness landscape elemzéssel bizonyíthatjuk, hogy miért is hatékonyak az egyes algoritmusok.

Kulcsszavak: *fitness landscape elemzés, keresési tér, optimalizáció*

Abstract

In this article, we present fitness landscape analysis techniques. Fitness landscape analysis can be important in optimization algorithms. When researchers develop a new algorithm, it is tested on benchmark data. If the results reach or are close to the best known solution, their solution is published. However, it may also be important to analyze the optimization algorithms as to why they actually proved to be better than the existing methods. With fitness landscape analysis, we can prove why each algorithm is effective.

Keywords: *fitness landscape analysis, search space, optimization*

1. Bevezetés

A fitness landscape elemzés az optimalizációs algoritmusok hatékonyságának, a keresési tér elemzésének tárgya. A keresési tér elemzésével kevés kutató foglakozott eddig. A kutatók többségében csak teszt eredményekkel bizonyítják azt, hogy az ő algoritmusok hatékony. Azonban találunk néhány publikációt, melyek a fitness landscape elemzéssel foglalkoznak. A következőkben ezeket mutatjuk be.

A [2]-ben a Job Shop Scheduling Problem feladatot oldják meg a szerzők. Statisztikai fitness landscape elemzést készítenek, az autokorreláció kiszámításáról számolnak be.

A [3]-ban a járatszervezési feladat (Vehicle Routing Problem) fitness landscape elemzése történik, információ elméleti szemszögből. A genetikus algoritmus elemzését készítik el. A szerzők olyan fontos fogalmakról számolnak be, mint a fitness landscape, szomszédság, neutrality. Ezután információelméleti szemszögből tárgyalják a keresési tér elemzést, mint az információ tartalom, részleges információtartalom, sűrűség alapú információ. A genetikus algoritmusnak az alábbi operátorokat alkalmazták: PMX, UOX, CX, Swap, Inversion, Insertion, Displacement.

A [4]-ben szintén a járatszervezési feladat keresési tér elemzése történik. Párhuzamosított szimulált lehűtéssel oldják meg a feladatot a szerzők.

A [6]-ben a Differenciális Evolúciós algoritmus hatékonyságát elemzik. A fitness landscape fontosabb fogalmaival, mint a fitness distance correlation, correlation length, epistasis, neutrality.

A [7]-ben a szerzők a keresési tér, a fitnesz függvény, szomszédság, navigációs stratégia, lokális és globális optimum, basin of attraction fogalmakat mutatják be.

A [8]-ban a szerzők fogalmak bemutatása után konkrét feladatokat elemeznek. Az alábbi fogalmakat mutatják be: lokális és globális optimum, landscape walks, korrelációs hossz, autokorreláció, basin of attraction, fitness barrier. Az alábbi problémákat elemzik: Quadratic Assignment Problem, Traveling Salesman Problem.

A [9]-ben benchmark függvények elemzését készítik el a szerzők. A fitnesz távolság elemzést készítik el.

2. Keresési tér alapfogalmak

<u>Lokális és globális optimum:</u>

A keresési tér elemzésének egyik legfontosabb tényezője a lokális és globális optimumok.

Globális optimum maximalizálási feladat esetén [1]:

$$O(\mathcal{F}) \coloneqq \{x \mid x \in S, \forall (y \in S)\, f(x) \geq f(y)\} \tag{1}$$

Globális optimum minimalizálási feladat esetén [1]:

$$O(\mathcal{F}) \coloneqq \{x \mid x \in S, \forall (y \in S)\, f(x) \leq f(y)\} \tag{2}$$

(A maximalizálási feladat könnyen átalakítható minimalizálási feladattá, csak a fitnesz függvény -1-szeresét kell vennünk).

Szomszédsági alapú keresési tér esetén a lokális optimum olyan megoldás, melynek egyetlen szomszédja sem kisebb (nagyobb) mint ő maga.

Maximalizálási feladat esetén [1]:

$$\widehat{O_N}(\mathcal{F}) \coloneqq \{x \mid x \in S, \forall (y \in N(x))\, f(x) \geq f(y)\} \tag{3}$$

Minimalizálási feladat esetén [1]:

$$\widehat{O_N}(\mathcal{F}) \coloneqq \{x \mid x \in S, \forall (y \in N(x))\, f(x) \leq f(y)\} \tag{4}$$

Definiálhatjuk a lokális optimumot rekombinációs tereknél (recombination space) is.

Maximalizálási feladat esetén [1]:

$$\widehat{O_P}(\mathcal{F}) \coloneqq \{x \mid x \in S, \forall (y \in S, z \in \mathcal{R}(x,y))\, f(x) \geq f(z)\} \tag{5}$$

$$\mathcal{R}: S \times S \to P(S) \tag{6}$$

Minimalizálási feladat esetén [1]:

$$\widehat{O_P}(\mathcal{F}) \coloneqq \{x \mid x \in S, \forall (y \in S, z \in \mathcal{R}(x,y))\, f(x) \leq f(z)\} \tag{7}$$

$$\mathcal{R}: S \times S \to P(S) \tag{8}$$

A lokális és globális optimumok elemzése során az alábbi kérdések/feladatok ismeretesek [1][10]:

Lokális optimumok száma: Ha a probléma unimodal, tehát csak egy lokális optimuma van, akkor valószínűleg könnyebb megoldani, mint az olyan feladatokat, amelyeknek több lokális optimuma van.

Lokális optimumba kerülés ideje (Time to local optimum): Ez azt jelenti, hogy kiindulunk egy véletlenszerűen választott megoldásból, és lokális kereső operátort addig alkalmazzuk, míg javulást észlelünk. A lokális optimumba kerülés ideje az operátor alkalmazásának számát jelenti.

A fitness landscape elemzése és optimalizáló algoritmusok kapcsolata: Az optimalizáló algoritmusok egy jó keresési tér állapotot keresnek viszonylag alacsony (belátható, minimális) időn belül, addig a fitness landscape elemzés egy jó betekintést nyújt a problémába minimális időn belül. A betekintés „jó" (elfogadható), és „optimális" megoldásokat jelent. A keresési térbe (fitness landscape) tekintés egy jó technika lehet új optimalizáló algoritmusok kifejlesztésére. [1]

Fennsíkok (plateaus) és semleges területek (neutral areas):

A fennsík (plateau) és a semleges terület (neutral area) definíciója előtt a keresési tér összekötésének (connectivity), tehát az összekötött részhalmaznak a definícióját részletezzük [1]. Diszkrét keresési tér során szomszédsági alapú összeköttetésnél a megoldás jelöltek összekötött halmaza $S' \subseteq S$, ahol minden elem össze vannak kötve. Az ilyen halmazt a következőképpen definiálhatjuk.

Az egyelemű halmaz egyetlen x megoldással egy olyan összekötött halmaz (connected set), amely jelölése: $connected(\{x\})$.

Az S összekötött halmaz (connected set) és egy x megoldásjelölt (mely a szomszédja az S összekötött halmaznak) uniója szintén egy összekötött halmaz (connected set) [1]:

$$connected(\{x\} \cup S) :\Leftrightarrow \exists(y \in S')\ N(x,y) \wedge connected(S') \tag{9}$$

Az összekötött halmaz definíciójából már könnyen megadható a semleges terület (neutral area) és a fennsík (plateau) definíciója.

A semleges (neutral) területet az alábbi képlet definiálja [1]:

$$neutral(S') :\Leftrightarrow connected(S') \wedge \forall(x,y \in S')\ \ f(x) = f(y) \tag{10}$$

A semleges terület a keresési tér olyan összekötött részhalmaza, ahol az elemek fitnesz értéke megegyezik. A fennsík definiálásához szükség van az összekötött halmaz külső határának (outer border) definiálására is. Ezek azok a pontok, melyek az összekötött halmazt határolják [1]:

$$border(S') \ := \{x \,|x \in S, x \notin S', \exists(y \in S), x \in N(y)\} \tag{11}$$

A fennsík (plateau) definícióját az alábbiak szerint adhatjuk meg [1]:

maximalizálási feladat esetén:

$$plateau(S) :\Leftrightarrow neutral(S) \wedge \ \forall\big(x \in border(S)\big)\, f(x) < f(S) \tag{12}$$

minimalizálási feladat esetén:

$$plateau(S) :\Leftrightarrow neutral(S) \wedge \ \forall\big(x \in border(S)\big)\, f(x) > f(S) \tag{13}$$

Az egyenletekben $f(S)$ egy fitnesz érték a semleges területen (neutral area).

A fennsík (plateau) egy olyan semleges területből (neutral area) és határából áll, ahol az összekötött halmaz határán (border) nincs (maximalizálási feladat esetén) nagyobb vagy (minimalizálási feladat esetén) kisebb fitnesz érték.

A semleges területek könnyen akadályozhatják az optimalizációs stratégiákat. A lejtő (slope) hiánya miatt az optimalizáló módszerek gyakran ragadnak nagy semleges területen (neutral area). De egy gondosan kiválasztott optimalizációs séma könnyen kiaknázhatja (exploit) a semleges területeket (neutral areas), úgy, hogy az algoritmus más (semleges területen – neutral area - kívüli) megoldásjelölteket is megtalál. [1]

3. Mértékek

Keresési tér nagysága (solution space size)

A legegyszerűbb mérték a keresési tér nagysága. Véges, megszámlálható keresési térnél pl. kombinatorikus optimalizálási problémák, a keresési tér nagysága az egyik legfontosabb mérőszám a probléma nagyságának becslésére. A keresési tér nagysága lehet a dimenziók száma, lehetséges állapot pontok száma, vagy a keresési tér átmérője is. [1]

A keresési tér átmérőjét az alábbi egyenlet adja [1]:

$$diam(S) := \max\{d(x,y) \mid x,y \in S\} \tag{14}$$

ahol $d: S \times S \to \mathbb{R}$ két egyed közötti távolságot jelent.

Autokorreláció (Auto correlation)

Autokorrelációs függvény (Auto correlation function)

Az egyenetlenség (ruggedness) a szomszédos megoldások korrelációján alapszik. Ezért a fitnesz függvény autokorrelációja az egyik mértéke a keresési tér egyenetlenségének (ruggedness). Ha viszonylag sok lokális optimuma van a fitness landscape-nek, akkor egyenetlen (rugged). Ha kevés lokális optimuma van, akkor sima (smooth) vagy lapos (flat). [1]

Az egyenetlen (rugged) keresési tér (search landscape) a szomszédos fitnesz értékek nagy szórásai, az ilyen térnél nehéz a keresés

Az autokorreláció kiszámítására az alábbi képlet szolgál [1]:

$$\rho(\tau) := \frac{E[f_i \cdot f_{i+\tau}] - E[f_i]E[f_{i+\tau}]}{Var[f_i]} \tag{15}$$

A képletben $E[x]$ és $Var[x]$ a várható értéke és a szórása az $\{f(x_i)\}_{i=0}^{n}$-nek. Az $\{f(x_i)\}_{i=0}^{n}$-et pedig f_i-vel rövidítjük. [1]

Autokorrelációs együttható (Auto correlation coefficient)

Ez a mérték megoldásjelöltek közötti az átlagos korrelációt adja meg a teljes keresési teren. $\rho(1)$-el jelöljük. Az aktuális autokorrelációs együttható (actual autocorrelation coefficient) az alábbi képlettel számolható [1]:

$$\lambda := \frac{1}{1-\rho(1)} \tag{16}$$

λ kicsi: egyenetlen tér (rugged landscape)
λ nagy: sima tér (smooth landscape)

4. Összefoglalás

Jelen cikkben áttekintést nyújtottunk a fitness landscape elemző technikákról. Ezen elemzés megmutatja, hogy az egyes optimalizáló algoritmusok miért hatékonyak. A cikkben a keresési tér alapfogalmait tárgyaltuk, mint a lokális és globális optimum, fennsíkok (plateaus) és semleges területek (neutral areas). Ezután a fitness landscape mértékeket mutattuk be, mint a keresési tér nagysága (solution space size), autokorreláció (auto correlation), autokorrelációs függvény (auto correlation function), autokorrelációs együttható (auto correlation coefficient). A cikkben összefoglalt elemző technikákat alkalmazva az egyes optimalimalizáló algoritmusok hatékonyságai bizonyíthatóak.

5. Köszönetnyilvánítás

A cikkben ismertetett kutató munka az EFOP-3.6.1-16-2016-00011 jelű „Fiatalodó és Megújuló Egyetem – Innovatív Tudásváros – a Miskolci Egyetem intelligens szakosodást szolgáló intézményi fejlesztése” projekt részeként – a Széchenyi 2020 keretében – az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Irodalom

[1] Erik Pitzer, E.: *Applied Fitness Landscape Analysis*, PhD dissertation, 2013, Johannes Kepler Universität Linz, Technisch-Naturwissenschaftliche Fakultät.

[2] Mattfeld, D. C., Bierwirth, C., Kopfer, H.: *A search space analysis of the job shop scheduling problem,* Annals of Operations Research, 86, (1999) pp. 441-453. **https://doi.org/10.1023/A:1018979424002**

[3] Ventresca, M., Ombuki-Berman, B., Runka, A.: *Predicting genetic algorithm performance on the vehicle routing problem using information theoretic landscape measures,* 2013 April, In European Conference on Evolutionary Computation in Combinatorial Optimization, pp. 214-225, Springer, Berlin, Heidelberg. **https://doi.org/10.1007/978-3-642-37198-1_19**

[4] Czech, Z. J.: *Statistical measures of a fitness landscape for the vehicle routing problem,* 2008 April, In 2008 IEEE International Symposium on Parallel and Distributed Processing, pp. 1-8. IEEE.

[5] Uludağ, G., Uyar, A. Ş.: *Fitness landscape analysis of differential evolution algorithms*, 2009 September, In 2009 Fifth International Conference on Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control, pp. 1-4. IEEE. **https://doi.org/10.1109/ICSCCW.2009.5379477**

[6] Watson, J. P.: *An introduction to fitness landscape analysis and cost models for local search,* In Handbook of metaheuristics, pp. 599-623. Springer, Boston, MA, 2010. **https://doi.org/10.1007/978-1-4419-1665-5_20**

[7] Pitzer, E., Affenzeller, M.: *A comprehensive survey on fitness landscape analysis,* In Recent advances in intelligent engineering systems, pp. 161-191. Springer, Berlin, Heidelberg, 2012.

[8] Müller, C. L., Sbalzarini, I. F.: *Global characterization of the CEC 2005 fitness landscapes using fitness-distance analysis,* 2011 April, In European Conference on the Applications of Evolutionary Computation, pp. 294-303. Springer, Berlin, Heidelberg. **https://doi.org/10.1007/978-3-642-20525-5_30**

[9] Tayarani-N, M. H., Prügel-Bennett, A.: *An analysis of the fitness landscape of travelling salesman problem,* Evolutionary computation, 24(2), (2016) pp. 347-384. **https://doi.org/10.1162/EVCO_a_00154**