Kopulové EDA algoritmy

Martin Hyrš
3. ročník, prezenční forma
Školitel: Josef Schwarz

Fakulta informačních technologií, Vysoké učení technické v Brně
Božetěchova 1/2, 612 66 Brno
ihyrs@fit.vutbr.cz

***Abstrakt*—EDA algoritmy (Estimation of Distribution Algorithm) patří mezi stochastické optimalizační techniky, založené na vytváření a vzorkování pravděpodobnostní modelu. V tomto článku prezentuji CEDA (Copula-EDA) algoritmus, ve kterém je použita kopule k vytvoření modelu a následné migraci parametrů pravděpodobnostního modelu. Experimentální výsledky potvrzují, že tento koncept vede ke zlepšení konvergence (ve srovnání se sekvenční variantou navrženého algoritmu nebo s výsledky CEDA algoritmů jiných autorů).**

***Klíčová slova*—EDA, kopule, migrace modelů.**

I. Úvod

EDA algoritmy (Estimation of Distribution Algorithms) patří do třídy evolučních optimalizačních metod, které prozkoumávají stavový prostor pomocí odhadu/aproximace a vzorkování explicitního pravděpodobnostního modelu nad množinou řešení.

Běžné EDA algoritmy používají grafický pravděpodobnostní model, např. řetězec, stromy nebo Bayesovskou síť, tyto algoritmy jsou známé pod názvy MIMIC [1], BMDA [2], BOA [3], UMDA [4]. Obsáhlý přehled těchto algoritmů je uveden v [5] nebo [6].

V posledních pěti letech začal být používán nový přístup ke konstrukci pravděpodobnostního modelu – postup založený na teorii kopulí [7], [8], [9]. Kopule jsou speciální funkce rozdělení pravděpodobnosti. Díky jejich vlastnostem je lze použít k modelování korelací ve vícerozměrných problémech – sdružená distribuční funkce je rozdělena na jednorozměrná marginální rozdělení a na korelaci, která je popsána kopulí.

II. Teorie kopulí

Definice. Kopule C je vícerozměrná distribuční funkce, jejíž marginální rozdělení jsou rovnoměrná na intervalu $\langle 0; 1\rangle$.

Teorém (Sklarův teorém [10])**.** *Nechť F je d-rozměrná sdružená distribuční funkce s marginálními rozděleními $F_1, \ldots, F_d$. Potom existuje d-rozměrná kopule C taková, že pro všechna $(x_1, \ldots, x_d) \in \mathbb{R}^d$ platí $F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$.*

Zabýváme se dvěma rodinami kopulí – Archimedovskými a eliptickými:

Archimedovské kopule jsou poměrně populární, protože popisují různé vzory závislostí, a mají relativně jednoduchý funkcionální zápis $C(u_1, \ldots, u_d) = \varphi_\theta\left(\varphi_\theta^{-1}(u_1) + \ldots + \varphi_\theta^{-1}(u_d)\right)$. Jejich definice je založena na generátoru φ, tato funkce má typicky jeden parametr θ, který vyjadřuje míru závislosti. To, že i vícerozměrné Archimedovské kopule mívají jen jeden parametr, je do určité míry nevýhoda, protože jejich schopnosti popsat vícerozměrné závislosti jsou výrazně omezené. Toto omezení lze kompenzovat např. pomocí hierarchických Archimedovských kopulí (HAC).

Eliptické kopule jsou odvozeny od příslušného eliptického rozdělení, příkladem je Gaussova kopule $C(u_1, \ldots, u_d) = \Phi_R\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right)$, kde $\Phi_R(x_1, \ldots, x_d)$ je sdružené normální rozdělení s pozitivně-semidefinitní korelační maticí R, Φ je standardní normální rozdělení, Φ^{-1} je jeho kvantilová (inverzní) funkce. Obdobným příklad je t-kopule, odvozená od Studentova t-rozdělení. Na rozdíl od Archimedovských kopulí není počet dimenzí nijak limitován, eliptické kopule se typicky používají jako vícerozměrné kopule, parametrem je korelační matice R.

III. mCEDA: EDA, kopule a migrace

V této kapitole stručně popíšeme algoritmus mCEDA, který je podrobně popsán v našem článku [11]. Jedná se o EDA algoritmus, který používá několik typů kopulí, a který je paralelizován s využitím ostrovního komunikačního modelu.

A. EDA (Estimation of Distribution Algorithm)

EDA algoritmy jsou evoluční optimalizační technika podobná genetickým algoritmům. Potenciální řešení problému je reprezentováno pomocí jedince, množina jedinců pak tvoří populaci. Co nejlepší řešení je vyvíjeno v průběhu několika generací, vhodní jedinci vybraní z jedné generace jsou použiti k vytvoření jedinců pro novou generaci. Na rozdíl od genetických algoritmů (které používají křížení a mutaci) jsou noví jedinci v EDA vytvářeni pomocí náhodného vzorkování z pravděpodobnostního modelu. Tento pravděpodobnostní model je odvozován z vybraných jedinců předešlé generace.

EDA algoritmus tedy tvoří tyto tři základní kroky:

1) Vyber vhodné jedince.
2) Vytvoř pravděpodobnostní model.
3) Navzorkuj nové jedince.

Jednotlivé EDA algoritmy se odliší druhem použitého pravděpodobnostního modelu. V případě CEDA je model tvořen kopulí a marginálními rozděleními.

Algorithm 1 Vzorkování kopulového pravděpodobnostního modelu

1) Získej náhodný vzorek kopule $(u_1, \ldots, u_d) \sim C$, kde $u_i \in \langle 0; 1 \rangle$.
2) Odvoď vektor hledaného jedince $\mathbf{x}$ za použití inverzí marginálních rozdělení $x_i = F_i^{-1}(u_i)$.

Algorithm 2 CEDA s migrací modelů (mCEDA)

1) Vygeneruj počáteční populaci.
2) **FOR** každý ostrov **DO IN PARALLEL**:
3) **WHILE** (není konec):
4) Vyber výhodná řešení
5) Vytvoř pravděpodobnostní model
6) **IF** (podmínka migrace):
7) Pošli model
8) **WHILE** (imigrantský model přijat):
9) Zkombinuj modely
10) Navzorkuj novou populaci ze získaného pravděpodobnostního modelu

B. Kopulový pravděpodobnostní model

Pravděpodobnostní model v kopulovém EDA algoritmu je specifikován ve dvou krocích. Nejdříve jsou odvozeny parametry pro jednorozměrná marginální rozdělení – používáme normální rozdělení $No(\mu, \sigma)$ s parametry μ (střední hodnota) a σ (směrodatná odchylka). Za druhé jsou vypočítány parametry pro zvolenou kopuli (R nebo θ) pomocí korelačního koeficientu Spearmanova ρ_S.

Hlavní přínos tohoto postupu spočívá v tom, že vzorkování nových jedinců (krok 10 v Alg. 2) je realizováno ve dvou jednoduchých krocích, tak jak je popsáno v Alg. 1. Tento postup umožňuje oddělit marginální rozdělení od struktury závislostí mezi dimenzemi.

C. Migrace modelů

V případě ostrovní modelu evoluční algoritmu se celková populace skládá z několik částí – ostrovů. Evoluční proces na každém ostrově běží nezávisle (viz Alg. 2). Pouze když je splněna podmínka migrace, je aktivována komunikace (v našem případě přenos parametrů modelu mezi dvojicí sousedících ostrovů).

Celý proces migrace je rozložen na interakci sousedních ostrovů – příjemce (rezidentní model) a imigrant (příchozí model). Model z imigranta je přenesen k příjemci, kde je zkombinován s původním rezidentním modelem a tak vzniká nový rezidentní model. Migrace a kombinace modelů je v Alg. 2 pokryta řádky 6–9.

Ke kombinaci modelů jsme použili následující vztahy, v souladu s [12], [13]:

- Učení středních hodnot μ_i každého jednorozměrného marginálního rozdělení $F_i(x_i)$

$$\mu_i^{new} = (1-\beta)\mu_i^R + \beta\mu_i^I \tag{1}$$

- Učení směrodatných odchylek σ_i každého jednorozměrného marginálního rozdělení $F_i(x_i)$

$$\sigma_i^{new2} = (1-\beta)\left(\left(\mu_i^{new} - \mu_i^R\right)^2 + \left(\sigma_i^R\right)^2\right) + \\ + \beta\left(\left(\mu_i^{new} - \mu_i^I\right)^2 + \left(\sigma_i^I\right)^2\right) \tag{2}$$

- Učení prvků korelační matice R_{ij}

$$R_{ij}^{new} = (1-\beta)R_{ij}^R + \beta R_{ij}^I \tag{3}$$

Koeficient β je určen vztahem

$$\beta = \begin{cases} \frac{fit^R}{fit^R + fit^I} & fit^I \leq fit^R \\ 0.1 & \text{jinak} \end{cases} \tag{4}$$

kde fit^R nebo fit^I představuje střední hodnotu fitness rezidenta respektive imigranta.

IV. Experimenty

Ukázkové experimenty byly provedeny na sadě běžných testovacích úloh (Rastrigin, Ackley, Schwefel, Rosenbrock, Sphere) a na sadě CEC 2013 [14] pro 10 rozměrů. Jsou prezentovány výsledky dvou variant mCEDA, používající Frankovu (Archimedovská rodina) kopuli (mCEDA-F) a Gaussovu (eliptická rodina) kopuli (mCEDA-G), ve srovnání se dvěma variantami sCEDA-* (sekvenční verze, tj. pouze jeden ostrov). Výsledky výše popsaných algoritmů mCEDA-* a sCEDA-* jsou znázorněny v grafech na obrázku 1.

Z grafů je patrné, že verze s migrací modelů (na grafu jako plné body) je téměř vždy lepší (nebo stejně dobrá) jako sekvenční verze (prázdné body); výjimkou je funkce č. 16 ze sady CEC13. Varianta užívající Gaussovu kopuli (čtverečky) je obvykle lepší nebo stejně dobrá jako varianta s Frankovou kopulí (kolečka). Celkově nejlepší testovanou variantou je tedy mCEDA-G (kopulová EDA s Gaussovou kopulí a migrací modelů (plné zelené čtverečky). V deseti případech ze sady CEC13 je mCEDA-G zřetelně lepší než ostatní varianty. Ve dvou případech je lepší mCEDA-F, na zbylých úlohách není patrný žádný významný rozdíl.

V Tabulkách I a II je prezentováno srovnání (střední hodnoty fitness) navrženého algoritmu mCEDA-* s dalšími publikovanými algoritmy, které nějakým způsobem používají kopule. Srovnání jsme provedli vždy pro stejný počet vyhodnocení fitness. Ve většině případů dosahuje mCEDA o několik řádů lepších výsledků.

V našem algoritmu jsme použili toto nastavení experimentů: 4 ostrovy, 250 jedinců na každém ostrově, selekce 20 %, migrace každých 10 generací, obousměrná kruhová topologie, 20 nezávislých běhů.

V. Cíl dizertace

Teorie kopulí je relativně nový koncept používaný ve statistice, ale v oblasti evolučních algoritmů se teprve začíná více rozvíjet. Moje práce zkoumá nové možnosti EDA algoritmů na bázi kopulí. Dosavadní výsledky naznačují, že CEDA

Obrázek 1. Výsledky experimentů – medián a kvartily hodnoty fitness algoritmů s Gaussovou a Frankovou kopulí; sekvenční verze a verze s migrací modelů.

Tabulka I
mCEDA VS. COPULA BAYESIAN NETWORK (CBN) [15] – 100000 VYHODNOCENÍ FITNESS.

	Rastr.	Ack.	Schw. 1.2	Rosen.
CBN	2,39e+00	3,71e-02	2,23e+01	1,05e+01
mCEDA-F	1,99e-01	1,28e-11	3,41e-21	5,90e+00
mCEDA-G	1,99e-01	1,14e-11	1,08e-20	5,66e+00

Tabulka II
mCEDA VS. CLAYTON (CL), GUMBEL (GU), SN-EDA (SN) [16] – 300000 VYHODNOCENÍ FITNESS.

	Sphere	Rastrigin's	Rosenbrock's
Cl	1,45e-07	7,00e-08	8,36e+00
Gu	3,59e-09	5,49e-09	6,62e+00
Sn	1,22e-09	9,52e-09	6,54e+00
mCEDA-F	4,05e-67	1,99e-01	5,64e+00
mCEDA-G	1,58e-67	1,99e-01	5,51e+00

umožňuje efektivní vytváření pravděpodobnostních EDA modelů, zejména v případě komplexních problémů s netriviálními vazbami mezi proměnnými.

Dalším aspektem, kterým se zabývám, je paralelizace algoritmu pomocí ostrovního modelu – tzn. použití několika samostatných populací (tzv. ostrovů), mezi kterými dochází k občasnému přenosu informací. Zejména se zaměřuji na migraci pravděpodobnostních modelů jednotlivých subpopulací mezi ostrovy.

Cílem dizertace je ukázat, že EDA algoritmy mohou významněji profitovat z použití teorie kopulí a migrace pravděpodobnostních modelů, tj. dokázat, že pro určitý typ úloh lze použitím vhodné kopule výrazně zvýšit efektivitu EDA algoritmů vzhledem ke standardní variantě EDA.

VI. ZÁVĚR

Představili jsme použití kopulí při návrhu pravděpodobnostního modelu v EDA algoritmu a užití paralelního ostrovního modelu algoritmu s použitím migrace modelů.

Ze vzájemného srovnání jedno-ostrovní verze (sCEDA) s více-ostrovní (mCEDA) je patrný velký přínos ostrovního modelu. Toto srovnání bylo provedeno jednak na běžných testovacích úlohách, jednak na sadě úloh CEC 2013.

Výkonnost našeho mCEDA algoritmu byla porovnána s jinými publikovanými kopulovými EDA. Naše mCEDA je ve většině případů výrazně lepší.

PODĚKOVÁNÍ

This work was supported by Brno University of Technology under number FIT-S-14-2297.

REFERENCE

[1] J. S. De Bonet, C. L. Isbell, and P. A. Viola, "MIMIC: Finding optima by estimating probability densities," in *Advances in Neural Information Processing Systems*, vol. 9. The MIT Press, Cambridge, 1997, pp. 424–430.

[2] M. Pelikan and H. Mühlenbein, "The bivariate marginal distribution algorithm," in *Advances in Soft Computing*. Springer London, 1999, pp. 521–535.

[3] M. Pelikan, D. Goldberg, and E. Cantú-Paz, "BOA: The bayesian optimization algorithm," in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-99)*, vol. I, 1999, pp. 525–532 also IlliGAL Report no. 99003.

[4] M. Pelikan and H. Mühlenbein, "Marginal distributions in evolutionary algorithms," in *Proceedings of the International Conference on Genetic Algorithms Mendel 98*, 1999, pp. 90–95.

[5] P. Larrañaga and J. A. Lozano, *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation*. Norwell, MA, USA: Kluwer Academic Publishers, 2001.

[6] M. Hauschild and M. Pelikan, "An introduction and survey of estimation of distribution algorithms," *Swarm and Evolutionary Computation*, vol. 1, no. 3, pp. 111 – 128, 2011.

[7] J. Mai and M. Scherer, *Simulating Copulas: Stochastic Models, Sampling Algorithms, and Applications*, ser. Series in quantitative finance. Imperial College Press, 2012, vol. 4.

[8] R. B. Nelsen, *An Introduction to Copulas*, ser. Springer Series in Statistics. Springer New York, 2006.

[9] U. Cherubini, E. Luciano, and W. Vecchiato, *Copula Methods in Finance*. Hoboken, NJ: John Wiley & Sons, 2004.

[10] A. Sklar, "Fonctions de répartition à n dimensions et leurs marges," *Publications de l'Institut de Statistique de l'Université de Paris*, vol. 8, pp. 229–231, 1959.

[11] M. Hyrš and J. Schwarz, "Elliptical and archimedean copulas in estimation of distribution algorithm with model migration," in *Proceedings of the 7th International Joint Conference on Computational Intelligence (IJCCI 2015)*. SciTePress - Science and Technology Publications, 2015, pp. 212–219.

[12] S. Frühwirth-Schnatter, *Finite Mixture and Markov Switching Models*. New York: Springer, 2006.

[13] J. Schwarz and J. Jaroš, "Parallel bivariate marginal distribution algorithm with probability model migration," in *Linkage in Evolutionary Computation*, ser. Studies in Computational Intelligence. Springer Berlin Heidelberg, 2008, vol. 157, pp. 3–23.

[14] J. Liang, B. Qu, P. Suganthan, and A. G. Hernández-Díaz, "Problem definitions and evaluation criteria for the cec 2013 special session on real-parameter optimization," *Computational Intelligence Laboratory, Zhengzhou University, Zhengzhou, China and Nanyang Technological University, Singapore, Technical Report*, vol. 201212, 2013.

[15] M. Méndez and R. Landa, "An EDA based on bayesian networks constructed with archimedean copulas," in *2012 Fourth World Congress on Nature and Biologically Inspired Computing (NaBIC)*, 2012, pp. 188–193.

[16] B. Jia, L. Wang, and Z. Cui, "Copula for estimation of distribution algorithm based on goodness-of-fit test," in *Journal of Theoretical and Applied Information Technology*, no. 3, 2013, pp. 1128–1132.