

Одеський національний університет імені І. І. Мечникова
Факультет математики, фізики та інформаційних технологій
Кафедра оптимального керування і економічної кібернетики

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття ступеня вищої освіти «магістр»

«Аналіз ефективності алгоритмів випадкового пошуку в машинному навчанні»

«On the effectiveness analysis of Random search optimization algorithms in machine learning»

Виконала: здобувачка денної форми навчання

спеціальності 113 Прикладна математика
Освітня програма «Прикладна математика»

Висторобська Лоліта Вячеславівна

Керівник канд. фіз.-мат. наук, доц. Страхов Є. М.

Рецензент канд. техн. наук, доц. Мороз В. В.

Рекомендовано до захисту:

Протокол засідання кафедри

№ ____ від _____ р.

Завідувач кафедри

Захищено на засіданні ЕК № _____

протокол № ____ від _____ р.

Оцінка _____ / _____ / _____
(за національною шкалою, шкалою ECTS, бали)

Голова ЕК

(підпис)

(прізвище, ім'я)

(підпис)

(прізвище, ім'я)

Odesa I. I. Mechnikov National University
Faculty of Mathematics, Physics and Information Technology
Department of Optimal Control and Economic Cybernetics

QUALIFICATION WORK

for obtaining the degree of higher education «master»

«On the effectiveness analysis of Random search optimization algorithms in machine learning»

Fulfilled by: full-time student

specialty 113 Applied Mathematics

Lolita Vystorobska

Supervisor: Associate Prof. Y. Strakhov

Reviewer: Associate Prof. V. Moroz

CONTENTS

INTRODUCTION	5
ВСТУП	8
CHAPTER 1 LITERATURE ANALYSIS AND APPROACHES INVESTIGATION	11
1.1. Classical methods limitations and formal description of HP optimization	11
1.2. Random search.....	12
1.3. Genetic search.....	13
1.4. Harmony search	15
1.6. Artificial Bee Colony search.....	18
1.7. Comparison study attempt	19
CHAPTER 2 SEARCH ALGORITHMS IMPLEMENTATION	21
CHAPTER 3 EXPERIMENT SETUP AND RESULTS	24
3.1 Task and dataset description	24
3.2 Data preprocessing.....	24
3.3 DNN implementation	25
3.4 GRU implementation.....	25
3.5 Computational experiment	26
CONCLUSIONS	30
ВИСНОВКИ	33
APPENDIX A	38
APPENDIX B	56

INTRODUCTION

Optimization is a frequent goal in many studies, and here optimization in the context of neural networks will be discussed as well, namely the optimization of hyper parameters. Here, a set of methods is to be evaluated and compared with significant emphasis on random search and natural computing algorithms.

So let us introduce the first base term referring to this work, namely as from [1] Stochastic optimization (SO) methods are optimization methods that generate and use random variables. For stochastic problems, the random variables appear in the formulation of the optimization problem itself, which involves random objective functions or random constraints. Stochastic optimization methods also include methods with random iterates. More specific sub-type of such methods introduced here are Natural Computing (NC) ones.

Beginning with definition of Natural Computing term as mentioned in [2], [it] refers to computational processes observed in nature, and human-designed computing inspired by nature. When complex natural phenomena are analyzed in terms of computational processes, our understanding of both nature and the essence of computation is enhanced. Characteristic for human-designed computing inspired by nature is the metaphorical use of concepts, principles and mechanisms underlying natural systems.

Beneficial cooperation among organisms and entities has suggested new ideas for search and control engineering. The look at highly interconnected networks of simple biological processing unit, that can learn and adapt, has introduced the way for our development of computational systems that can differentiate between complex patterns, and improve themselves over time.

By studying complex biological organisms, looking at nature designed systems, industry and science was brought ways to explore innovative design approaches and develop even new products. It is worth to add note from [3], where 10 real-world successful implementations of NC algorithms were presented, that NC methods are more than substitute approach to the challenges faced in various domains. In many

fields, nature-inspired methods have overcome barriers in the prior achievements and capabilities of classical computing.

In the subsequent understanding of this paper, these NC algorithms in essence and in the actual sense refer to a more global concept of random search and stochastic optimization methods.

Another part of the current paper is Hyper Parameters optimization in Neural Networks (NN), where, for clarification, such a parameter is the one to be set prior the training process, cannot be deduced via model learning, and mainly, a set of hyper parameters can manage and control learning process itself. It can be a loss function, proper model layer configuration, activation function type, optimization technique etc. For now, hyper parameters will not be divided into those that belong to model and to algorithms, nevertheless such a separation could take place as well.

Thus, the main purpose of the current report is an investigation of various approaches of Hyper Parameters optimization in Neural Networks and its comparison. As Hyper Parameters optimization is showed up as a non-trivial task, especially in case of high number of the latter, it is needed to apply more sophisticated methods for resolving such a problem.

Nevertheless, there are no special literature and only a few papers that could present a broad comparison of the various methods of Hyper Parameters optimization. In most cases, researchers take some model by specific task, simple NN like Multilayer perceptron or Convolutional NN, and some Optimization method that did not used before or was modified, and eventually explore how those two perform together. Unfortunately, it is not standardized cause each time there are different models, data, search space as well as general parameters of the tests. That is why in this research the main objective is to sum up it all together, implement the selected search algorithms for specific task and perform the comparison analysis based on results.

For the following report, natural language processing (NLP) aspect-based classification task has been chosen. As models, deep neural network (DNN) and gated

recurrent unit (GRU) model were used, and for a comparison, simple random search, tree-structured Parzen estimator (TPE); simulated annealing, particle-swarm optimization, harmony search, as well as genetic algorithms are to be evaluated.

ВСТУП

Оптимізація є частою метою в багатьох дослідженнях, і тут буде також обговорюватися оптимізація в контексті нейронних мереж, а саме оптимізація гіперпараметрів. Тут необхідно оцінити та порівняти набір методів із значним акцентом на випадковому пошуку та природних обчислювальних алгоритмах.

Отже, введемо перший базовий термін, що відноситься до цієї роботи, а саме з [1] Стохастичні методи оптимізації (SO) – це методи оптимізації, які генерують і використовують випадкові величини. Для стохастичних задач випадкові величини з'являються у формулюванні самої задачі оптимізації, яка включає випадкові цільові функції або випадкові обмеження. Методи стохастичної оптимізації також включають методи з випадковими ітераціями. Більш специфічним підтипом таких методів, представлених тут, є природні обчислення (NC).

Починаючи з визначення терміну «Природні обчислення» (NC), як зазначено в [2], [він] відноситься до обчислювальних процесів, що спостерігаються в природі, і розроблених людиною обчислень, натхнених природою. Коли складні природні явища аналізуються з точки зору обчислювальних процесів, наше розуміння як природи, так і суті обчислень покращується. Характерним для розроблених людиною обчислень, натхнених природою, є метафоричне використання понять, принципів і механізмів, що лежать в основі природних систем.

Благотворна співпраця між організмами та сутностями запропонувала нові ідеї для пошукової і контрольної техніки. Погляд на дуже взаємопов'язані мережі простих біологічних процесорів, які можуть навчатися та адаптуватися, відкрив шлях до нашої розробки обчислювальних систем, які можуть розрізняти складні моделі та вдосконалюватися з часом.

Вивчаючи складні біологічні організми, дивлячись на створені природою системи, промисловість і наука знайшли шляхи дослідження інноваційних

підходів до проектування та розробки навіть нових продуктів. Варто додати примітку з [3], де було представлено 10 реальних успішних реалізацій NS-алгоритмів, що методи NS є більш ніж замінним підходом до проблем, з якими стикаються в різних областях. У багатьох галузях природні методи подолали перешкоди в попередніх досягненнях і можливостях класичних обчислень.

У подальшому розумінні цієї статті ці алгоритми NS по суті та в реальному сенсі відносяться до більш глобальної концепції методів випадкового пошуку та стохастичної оптимізації.

Іншою частиною поточної статті є оптимізація гіперпараметрів у нейронних мережах (NN), де, для уточнення, такий параметр є параметром, який встановлюється перед процесом навчання, не може бути виведений за допомогою навчання моделі, і, головним чином, множина гіперпараметрів може керувати і контролювати сам процес навчання. Це може бути функція втрат, належна конфігурація рівня моделі, тип функції активації, методика оптимізації тощо. Поки що гіперпараметри не будуть розділятися на ті, що належать моделі та алгоритмам, проте такий поділ також може мати місце.

Таким чином, основною метою даної роботи є дослідження різних підходів оптимізації гіперпараметрів у нейронних мережах та їх порівняння. Оскільки оптимізація гіперпараметрів виявляється нетривіальною задачею, особливо при великій кількості останніх, для вирішення такої проблеми необхідно застосовувати більш складні методи.

Тим не менш, немає спеціальної літератури і лише кілька статей, які могли б представити широке порівняння різних методів оптимізації гіперпараметрів. У більшості випадків дослідники беруть певну модель за конкретним завданням, просту NN, як-от багатосаровий персептрон або згорткову CNN, і деякий метод оптимізації, який раніше не використовувався або був модифікований, і в кінцевому підсумку досліджують, як вони працюють разом. На жаль, це не стандартизовано, тому що кожен раз є різні моделі, дані, простір пошуку, а також

загальні параметри тестів. Тому в даному дослідженні головною метою є підсумувати все разом, реалізувати обрані алгоритми пошуку для конкретного завдання та провести порівняльний аналіз за результатами.

Для наступної роботи було обрано завдання класифікації відносно сутностей обробки природної мови (NLP). В якості моделей використовувалися класична нейронна мережа та модель вентильного рекурентного вузла, а для порівняння — простий випадковий пошук, tree-structured Parzen estimator (TPE); алгоритми змодельованого відпалу, рою часток, гармонійного пошуку, а також генетичний алгоритм мають бути оцінені.

Мета роботи: реалізувати обрані алгоритми пошуку для конкретного завдання, класифікації відносно сутностей, та провести порівняльний аналіз за результатами.

Об'єкт дослідження: моделі машинного навчання: класична нейронна мережа та модель вентильного рекурентного вузла; алгоритми випадкового пошуку: випадковий пошук, алгоритми змодельованого відпалу, рою часток, гармонійного пошуку, а також генетичний алгоритм.

Предмет дослідження: порівняльна характеристика методів випадкового пошуку.

Методи дослідження: обчислювальні експерименти.

ВИСНОВКИ

Проведено аналіз літератури щодо стохастичної оптимізації гіперпараметрів (HP). Як видно, існує купа досліджень, що досліджують конкретний стохастичний або природний обчислювальний алгоритм для певної конкретної проблеми. А саме, у цій роботі було досліджено 6 підходів: пошук по сітці, випадковий пошук, генетичний алгоритм (GA), оптимізація роя частинок (PSO), гармонічний пошук (HS) та алгоритм штучної бджолоїної колонії (ABC). Варто зазначити, що в більшості статей методи не були реалізовані як є, через неоднорідність простору пошуку HP. Для деяких алгоритмів автори змінюють внутрішню структуру алгоритму (GA, ABC, HS), таким чином двійкові або категоріальні змінні можна перетворювати вперед і назад; для PSO навіть це було неможливо, і було додано відображення точності моделі за допомогою Байєсівської оптимізації. Була лише одна стаття, яка містить порівняння між методами, але тут більший акцент був зроблений на конкретному завданні, і не було представлено, що налаштування тесту було повністю стандартизовано, тому результати можна було б перенести на інше завдання. Більше того, багато досліджень вносять модифікації до методів такі, як умови та внутрішня оцінка під час пошуку, щоб не досліджувати потенційно тупикові напрямки або агрегація проміжних результатів, які можуть зменшити споживання ресурсів тощо. Тим не менш, досі бракує повноцінних досліджень, які могли б показати достовірну картину, щодо ефективності набору стохастичних алгоритмів та дати повний порівняльний аналіз, що визначено як основне завдання дослідження.

У результаті цього дослідження було реалізовано два алгоритми, PSO та Harmony search, один алгоритм був модифікований, моделювання відпалу, і відповідний фреймворк Optuna було розширено, щоб задовольнити вимоги реалізації PSO. Загалом було порівняно 6 алгоритмів: випадковий пошук, tree-structured Parzen estimator (TPE); моделювання відпалу, оптимізація роїв частинок, гармонічний пошук, а також алгоритм генетичного пошуку. Крім того, усі вони були запуснені та реалізовані в одній структурі, що дозволяє проводити

достовірну оцінку між ними. Єдиним нестохастичним алгоритмом є TPE, який було взято як еталон для порівняння, оскільки він реалізований у більшості широко використовуваних систем оптимізації HP. Завданням для порівняння є класифікація почуттів відносно сутностей, а моделями, для яких оптимізувалися гіпер-параметри, є DNN і GRU. Параметри та налаштування експерименту для кожної моделі проводились окремо.

У першому експерименті з моделлю DNN імітований відпал перевершує інші алгоритми з показником точності 67,5 % і найкоротшим часом обчислення; що також важливо, найкращий результат був досягнутий під час передостаннього випробування, однак під час дослідження експеримент вийшов на плато. Другими найкращими алгоритмами з оптимізацією гіперпараметрів (HP) DNN є Harmony search і TPE з рівним результатом у 67,1%, тоді як з точки зору часу гармонійний пошук є кращим, і обидва алгоритми досягли відносного плато після 10-го випробування. Ці плато в кожному дослідженні викликані функцією чистки, яка просто припиняє всі безперспективні випробування. Таким чином, стохастичні алгоритми показали хороші результати, хоча нестохастичний також зміг потрапити до трійки лідерів. Важливо зазначити, що майже для кожного алгоритму був один HP, який мав найвищу вагу у приблизно 50%, і лише для генетичного пошуку є 2 таких гіперпараметра, кількість нейронів у шарах і швидкість навчання, з оцінкою важливості 39% і 38% відповідно. Крім того, для більшості алгоритмів найважливішими параметрами були два вже зазначених, для деяких алгоритмів грає роль і кількість епох, і що цікаво, для моделювання відпалу найважливішим параметром була саме кількість шарів.

У другому експерименті модель GRU, безсумнівно, показала набагато кращий результат через більш специфічну та складну структуру самої моделі. Тим не менш, цього разу генетичний пошук перевершує інші алгоритми з показником точності 96,3 %, але в той же час це алгоритм мав найдовший, але все ще достатньо хороший час обчислення; найкращий результат було досягнуто в середині дослідження, під час 12-го випробування, і після був вихід на плато.

Другим найкращим алгоритмом є нестохастичний TPE з результатом 96 % і майже найкращим часом обчислення. Усі інші алгоритми працювали майже однаково, їх показники коливалися в районі 95%, а деякі вийшли на плато вже з перших випробувань. Можливо, що в цій ситуації кількість проб можна просто збільшити, щоб подолати початковий поріг точності. Щодо важливості HP, у кожному алгоритмі є 2-3 основні HP, серед яких кількість епох і нейронів у шарі, ймовірність вибуття шару (dropout), швидкість навчання та іноді оптимізатор. Існує 2 алгоритми, PSO і генетичний пошук, які мають параметри, які значно переважають за важливістю інші в цих алгоритмах, близько 80%, кількість епох і швидкість навчання відповідно. Крім того, кинувся в очі розподіл важливості для імітованого відпалу, оскільки тут кількість нейронів і епох отримала однакову оцінку у 38%. Важливість параметра може бути надзвичайно корисною, якщо ресурси обмежені, модель важка, та необхідно зменшити простір пошуку, враховуючи першу обмежену кількість випробувань.

Загалом, в обох експериментах стохастичні алгоритми працювали більш ніж задовільно, де-факто випереджаючи нестохастичний алгоритм, хоча останній також показав хороші результати. Як результат цього дослідження, гармонійний пошук, генетичний пошук, а також моделювання відпалу можуть бути рекомендовані для використання на реальних, прикладних задачах.

Апробація роботи відбувалася на спільній українсько-китайській онлайн-конференції 1st Student Scientific Conference of Joint Research Cooperation between Odessa I.I. Mechnikov National University and Huaiyin Institute of Technology [19].

REFERENCES

1. Spall, J.. (2007). Introduction to Stochastic Search and Optimization. Estimation, Simulation, and Control. Neural Networks, IEEE Transactions on. 18. 964-965. 10.1109/TNN.2007.897481.
2. Natural Computing, An international Journal n.d., <https://www.springer.com/journal/11047>.
3. Corne, D., Deb, K., Knowles, J. and Yao, X., 2011. Selected applications of natural computing. *Handbook of natural computing*.
4. Kim, J.Y. and Cho, S.B., 2019, June. Evolutionary optimization of hyperparameters in deep learning models. In *2019 IEEE Congress on Evolutionary Computation (CEC)* (pp. 831-837). IEEE.
5. Bergstra, J.; Bengio, Y. Random Search for Hyper-Parameter Optimization. *J. Mach. Learn. Res.* 2012, 13, 281–305.
6. Li, C., Jiang, J., Zhao, Y., Li, R., Wang, E., Zhang, X. and Zhao, K., 2021. Genetic Algorithm based hyper-parameters optimization for transfer Convolutional Neural Network. *arXiv preprint arXiv:2103.03875*.
7. M. Suganuma, S. Shirakawa, T. Nagao. A Genetic Programming Approach to Designing Convolutional Neural Network Architectures. In: *GECCO, 2017*, pp. 497.
8. Woo-Young Lee, Seung-Min Park, Kwee-Bo Sim, Optimal hyperparameter tuning of convolutional neural networks based on the parameter-setting-free harmony search algorithm, *Optik*, Volume 172, 2018, Pages 359-367, ISSN 0030-4026.
9. Li, Y. and Zhang, Y., 2020. Hyper-parameter estimation method with particle swarm optimization. *arXiv preprint arXiv:2011.11944*.
10. Leila Zahedi, Farid Ghareh Mohammadi and M. Hadi Amini HyP-ABC: A Novel Automated Hyper-Parameter Tuning Algorithm Using Evolutionary Optimization TechRxiv, 2021

11. Zahedi, Leila & Ghareh Mohammadi, Farid & Amini, M. H.. (2021). OptABC: an Optimal Hyperparameter Tuning Approach for Machine Learning Algorithms. 10.1109/ICMLA52953.2021.00186.
12. Elgeldawi, Enas & Sayed, Awny & Galal, Ahmed & Zaki, Alaa. (2021). Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis. Informatics. 8. 10.3390/informatics8040079.
13. <https://optuna.org/>
14. K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," in IEEE Transactions on Evolutionary Computation, vol. 6, no. 2, pp. 182-197, April 2002, doi: 10.1109/4235.996017.
15. <https://spacy.io/usage/processing-pipelines/>
16. <https://pytorch.org/>
17. https://miro.medium.com/max/1400/1*UxZ0pTQW8kofL9bzPVYV1w.webp
18. Bergstra, James & Bardenet, R. & Kégl, Balázs & Bengio, Y.. (2011). Algorithms for Hyper-Parameter Optimization.
19. Vystorobska L., Strakhov Ye., On the effectiveness analysis of Random search optimization algorithms in machine learning, 1st Student Scientific Conference of Joint Research Cooperation between Odessa I.I. Mechnikov National University and Huaiyin Institute of Technology: proceedings of the conference, May 16, 2022.