Optimization Techniques in Artificial Intelligence: A Short Survey

Sanjay Goswami
Institute of Computing and Analytics
NSHM Knowledge Campus
Kolkata, India
sanjay.goswami@nshm.com
ORCID: 0000-0002-4141-4934

Prithwish Sarkar
Institute of Computing and Analytics
NSHM Knowledge Campus
Kolkata, India
prithwishsarkar.22@nshm.edu.in
ORCID: 0009-0009-0160-8943

***Abstract* -- The field of Artificial Intelligence (AI) is fundamentally the problem of optimizing the learning models. An AI system is expected to learn autonomously from the data, encoding the real life scenarios, and optimize the learning models to reduce prediction errors. This paper offers a concise exploration of prominent optimization techniques employed across the diverse landscape of AI. We explore both the gradient-based and derivative-free methods, illuminating their strengths and limitations within the context of various AI applications. Additionally, we explore recent advancements in optimization techniques tailored for specific AI subfields and showcase real-world applications that leverage these techniques to achieve intelligent behavior.**

I. Introduction

A. Role of Optimization in AI:

Artificial Intelligence (AI) has revolutionized numerous aspects of our world, from enabling self-driving cars to powering intelligent virtual assistants. At the core of these advancements lies a fundamental concept: optimization. AI systems, despite their impressive capabilities, are essentially algorithms designed to navigate vast search spaces and identify the most desirable solutions or models. This section delves into the intricate relationship between optimization and core AI functionalities like learning, pattern recognition, and decision-making, highlighting the crucial role optimization plays in driving AI's success.

B. Learning as an Optimization Problem:

One of the defining characteristics of AI is its ability to learn from data. Machine learning, a powerful subfield of AI, relies heavily on optimization techniques. During the training process, machine learning models are presented with vast volumes of data. The goal is to adjust the model's internal parameters (weights and biases) in a way that minimizes the difference between the model's predictions and the actual data points [1]. This process can be framed as an optimization problem, where the objective function represents the prediction error, and the optimization algorithm iteratively searches for the parameter configuration that minimizes this error [2].

C. Pattern Recognition through Optimized Feature Selection:

Pattern recognition is another cornerstone of AI, empowering systems to identify patterns and relationships within data. Optimization plays a crucial role in feature selection, a critical step in pattern recognition tasks. By selecting the most informative features from the data, the model can focus on the aspects that best distinguish between different patterns [3]. Feature selection techniques often leverage optimization algorithms to identify the optimal subset of features that maximizes the model's performance while minimizing its complexity [4].

D. Decision-Making Optimized for Optimal Outcomes:

AI systems are increasingly tasked with making complex decisions in dynamic environments. Reinforcement learning, a powerful AI technique, tackles these challenges by training an agent to learn through

ISBN : 978-81-978522-2-0 https://doi.org/10.37285/bsp.oaisem2025.22

trial and error [5]. The core idea involves optimizing an agent's policy, which determines its actions in different situations, to maximize its long-term rewards. Optimization algorithms play a vital role in this process, guiding the agent towards policies that lead to the most favorable outcomes [6].

E. Motivation for the Survey: The Growing Complexity of AI and Data

The field of AI is witnessing a continuous surge in model complexity. Deep learning architectures, with millions or even billions of parameters, are increasingly employed for tasks like image recognition and natural language processing [7]. This complexity presents a significant challenge. Training such models requires vast amounts of data, often measured in terabytes or petabytes [8]. However, with this ever-growing data, comes the crucial need for efficient and robust optimization techniques, where traditional methods may struggle to handle the sheer volume of data and the intricate parameter landscapes of modern AI models.

II. BACKGROUND: OPTIMIZATION PROBLEMS IN AI

A. General Optimization Formulation:

In mathematical optimization, we formulate problems to maximize or minimize an objective function while adhering to certain constraints. Here are the key components [9]:

1) *Decision Variables:* Denoted as ($x_1, x_2, x_3, \ldots$), these represent the quantities we can control or manipulate. For instance, in production planning, x_1 might represent the number of units produced, and x_2 could be the raw material used, etc.
2) *Objective Function:* Denoted as $f(x)$, this function quantifies what we want to optimize. It could be *profit*, *cost*, *efficiency*, or any other measurable outcome. For example, if we aim to maximize profit, our objective function might be $f(x) = 3x_1 + 2x_2$.

3) *Constraints:* These are equations or inequalities that limit the feasible solutions. Constraints ensure that our decisions adhere to specific requirements. For instance, if we are allocating resources, we might have constraints like $2x_1 + 3x_2 \leq 100$ (resource availability) and $x_1, x_2 \geq 0$ (non-negativity).

The general optimization problem can be expressed as:

$$Minimize_x: f(x)$$
$$Subject\ to\ constraints:$$
$$g_i(x) \leq 0, i = 1,2, \ldots, m$$
$$h_j(x) = 0, j = 1,2, \ldots, p$$

Where

$f: R^n \rightarrow R$is the objective function to be minimized over the n-variable vector X
$g_i(x) \leq 0$ are called the inequality constraints
$h_j(x) = 0$ are called the equality constraints
$m \geq 0, p \geq 0$ are called constants

B. AI as Optimization problem:

Artificial Intelligence (AI) tasks are often reduced to optimization problems, where the goal is to find the best solution given certain constraints. Here is an account of how different AI subfields usually frame their challenges as optimization problems, as briefed above:

1) Machine Learning (ML):

ISBN : 978-81-978522-2-0

- *Loss Function Minimization:* In ML, model training involves minimizing a loss function. For instance, in linear regression, we aim to minimize the mean squared error (MSE) between predicted and actual values [1]:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Where, $n\rightarrow$ is the sample size, y_i → is the i^{th} output, $\hat{y}_i\rightarrow$ is the i^{th} predicted output, and *MSE* → is the *Mean Square Error* between the *actual* and *predicted* output vectors.

- *Gradient Descent:* Optimization algorithms like gradient descent iteratively adjust model parameters to minimize the loss function [1].

2) *Reinforcement Learning (RL):*

- *Maximizing Long-Term Rewards:* In RL, an agent interacts with an environment to learn optimal actions. The goal is to maximize cumulative rewards over time. For example, autonomous car training involves maximizing safety and efficiency [5].
- *Q-Learning:* Q-learning algorithm estimates optimal action-values (Q-values) to maximize expected future rewards [5].

3) *Logistics and Resource Allocation:*

- *Supply Chain Management:* Companies optimize package delivery routes to minimize time or cost while satisfying certain constraints [10].

- *Electric Grid Management:* Balancing power generation and distribution minimizes costs and ensures reliability [11].

4) *Resource Allocation in Healthcare:*

- *Vaccine Distribution:* Efficiently allocating vaccines during a pandemic considers cold storage requirements and population density [12].
- *Hospital Bed Allocation:* Optimizing bed usage maximizes patient care while minimizing overcrowding [13].

5) *Ride-Hailing Services:*

- *Driver Dispatch:* Platforms optimize driver assignments to minimize wait times for passengers and maximize driver utilization [14, 15].

6) *Portfolio Optimization:*

- *Financial Investments:* Investors allocate assets to maximize returns while minimizing risk [16, 17].

7) *Objective Functions and Cost Functions:*

Objective functions play a crucial role in guiding optimization processes. They quantify the goal we want to achieve, such as maximizing profit or minimizing error. In machine learning, cost functions serve as surrogates for complex objectives. For instance, mean squared error (MSE) acts as a proxy for model accuracy, allowing efficient optimization during training [1, 9].

III. Optimization Techniques: Gradient-Based

A. *The Power of Gradients:*

Gradients play a pivotal role in optimization, guiding parameter adjustments toward optimal solutions, and providing directional information by indicating the steepest ascent or descent in a function. In machine learning, gradients guide model updates during training, enhancing performance [18].

B. *Popular Gradient-Based Methods:*

1) *Stochastic Gradient Descent (SGD):*

- SGD efficiently handles large datasets by randomly selecting a single training example (or a small batch) for gradient computation and parameter updates.
- Variants like *Adam* and *RMSprop* further enhance SGD by adapting learning rates and addressing issues like vanishing gradients [19, 20].

2) *Momentum Gradient Descent (MGD):*

- Momentum incorporates inertia from previous updates, helping escape local minima during optimization.
- It accumulates gradients over time, leading to faster convergence and better generalization [20, 21].

3) *AdaGrad Gradient Descent (AGD):*

- *AdaGrad* adapts learning rates based on historical gradients for different model parameters.
- This is effective for sparse data and non-stationary objectives [20, 21].

IV. Optimization Techniques: Derivative-Free

A. *When Gradients are Absent:*

Derivative-free optimization methods play a crucial role when gradients are unavailable. In scenarios with complex objective functions or non-differentiable models, these methods explore the parameter space without relying on gradient information. By using techniques like genetic algorithms, pattern search, and simulated annealing, they navigate challenging landscapes, making them valuable tools for optimization in diverse domains.

B. *Beyond Gradients:*

Common derivative-free techniques include:

- *Evolutionary Algorithms (e.g., Genetic Algorithm):* Evolutionary algorithms, including the Genetic Algorithm (GA), simulate natural selection to enhance candidate solutions iteratively. By creating a population of solutions and applying genetic operators like crossover and mutation, these algorithms mimic evolution. Fitness-based selection favours better-performing individuals, leading to population improvement [22].
- *Simulated Annealing:* Simulated annealing (SA) uses a temperature parameter to explore search spaces effectively, allowing escape from local optima. At high temperatures, SA makes random uphill moves, while lowering the temperature converges toward the global optimum [23]
- *Swarm Intelligence (e.g., Particle Swarm Optimization):* Swarm Intelligence techniques, like PSO, harness the collective behaviour of simulated agents to explore search spaces effectively and converge toward optimal solutions. PSO models particles' social behavior, allowing them to communicate and adapt their positions in the search space [24].

V. Recent Advancements and Applications

A. Optimizing for Specific AI Subfields:

Emerging optimization techniques significantly enhance AI model performance. Notable methods include *AdamW*, tailored for deep learning, and *Proximal Policy Optimization (PPO)* in reinforcement learning. *AdamW* combines adaptive learning rates with weight decay, addressing overfitting. PPO balances ease of implementation and policy updates while minimizing deviation from the previous policy.

B. Real-World Applications:

Real-world examples show how optimization techniques are used in various AI domains, such as:

1) Machine Learning:

- *Model Training:* During machine learning training, models adapt their internal parameters (weights and biases) to minimize prediction errors. This optimization process involves algorithms like Stochastic Gradient Descent (SGD) and variants (*Adam*, *RMSprop*) [1,2].
- *Feature Selection:* Feature selection is the process of automatically choosing relevant features for machine learning models. It reduces noise and input data size. Optimization algorithms help identify optimal feature subsets, and techniques like L1 or L2 regularization encourage sparsity and selection [3].

2) Computer Vision:

- *Image Classification:* Classifying images into different categories (e.g., dog, cat) involves training deep neural networks. Optimization techniques are used to adjust the network's weights during training to minimize the classification error [25].
- *Object Detection:* Detecting and locating objects within images requires optimizing models to predict bounding boxes around objects and their corresponding class labels [25].
- *Image Segmentation:* Segmenting images involves dividing an image into regions corresponding to different objects or parts. Optimization helps train models to accurately assign pixels to different segments [25].

3) Natural Language Processing (NLP):

- *Machine Translation:* Machine translation involves training complex neural networks to translate text from one language to another. Optimization techniques play a crucial role in adjusting the network’s parameters to minimize translation errors and produce accurate translations [26].
- *Sentiment Analysis:* Sentiment analysis, also known as opinion mining, classifies the sentiment of text (positive, negative, neutral) using labelled data. Optimization helps adjust model parameters to minimize sentiment classification errors [27].
- *Text Summarization:* Text summarization involves compressing a piece of text while retaining crucial information. Optimization techniques are used to train models that generate concise and informative summaries [26].

4) Reinforcement Learning:

- *Agent Training:* Reinforcement learning involves training an agent to learn through trial and error in an environment. The core idea involves optimizing the agent's policy (action selection strategy) to maximize its long-term reward. Techniques like Q-learning and Policy Gradient methods leverage optimization algorithms to guide the agent towards optimal policies [5].

C. Other Applications:

1) Robotics Control: Optimizing control parameters allows robots to learn complex motor skills and adapt to their environment through reinforcement learning techniques [28].

2) *Resource Allocation:* Optimization can be used to allocate resources efficiently in various AI applications, such as scheduling tasks in a distributed system or managing network traffic [29].

3) *Planning and Decision-Making:* AI systems tackling complex planning and decision-making problems often utilize optimization techniques to find the best course of action under various constraints [30].

VI. CONCLUSION

Optimization plays a key role in advancing artificial intelligence (AI). From fine-tuning control parameters for robots to resource allocation and complex decision-making, optimization techniques enhance AI systems' efficiency and effectiveness. As we look into the future, the ongoing research aims to develop novel algorithms capable of handling increasingly large and intricate data models. Additionally, incorporating domain-specific knowledge into the optimization process remains a promising avenue for further AI research.

REFERENCES

[1] Mitchell, T. M. (1997). Machine learning. McGraw-Hill.

[2] Boyd, S., & Vandenberghe, L. (2004). Convex optimization. Cambridge University Press.

[3] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of machine learning research, 3(Mar), 1157-1182.

[4] Chandrasekaran, B., Sankaranarayanan, S., Parthasarathy, S., & Nagarajan, V. (2014). Fast greedy algorithm for optimal feature selection using mutual information. Journal of Machine Learning Research, 15(1), 1023-1050.

[5] Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.

[6] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Dean, D., & Riedmiller, M. (2013). Playing games with deep neural networks. arXiv preprint arXiv:1312.5905.

[7] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

[8] Dean, J., & Ghemawat, S. (2018). MapReduce: Simplified data processing on large clusters. Communications of the ACM, 51(1), 107-113.

[9] Bertsimas, D., & Tsitsiklis, J. N. (1997). Introduction to linear optimization. Athena Scientific.

[10] Xu, X., Liu, J., & Wang, J. (2019). A N–N optimization model for logistic resources allocation with multiple logistic tasks under demand uncertainty. Transportation Research Part E: Logistics and Transportation Review, 129, 1-17.

[11] Zhang, Y., & Wang, J. (2015). Supply chain intelligence for electricity markets: A smart grid perspective. Information Systems and e-Business Management, 13(4), 679-696.

[12] Raghupathi, V., & Raghupathi, W. (2023). The association between healthcare resource allocation and health status: an empirical insight with visual analytics. Journal of Public Health, 31, 1035–10571

[13] Patel, P. C., Tsionas, M. G., & Devaraj, S. (2023). Relative bed allocation for COVID-19 patients, EHR investments, and COVID-19 mortality outcomes. PLOS ONE

[14] Mitropoulos, L., Kortsari, A., & Ayfantopoulou, G. (2021). A systematic literature review of ride-sharing platforms, user factors, and barriers. European Transport Research Review, 13, 61

[15] Jaydarifard, S., Behara, K., Baker, D., & Paz, A. (2023). Driver Fatigue in Taxi, Ride-Hailing, and Ridesharing Services: A Review of Literature. SSRN

[16] Gunjan, A., & Bhattacharyya, S. (2023). A brief review of portfolio optimization techniques. Artificial Intelligence Review, 56(9), 3847–38861

[17] Dangi, A. (2013). Financial Portfolio Optimization: Computationally guided agents to investigate, analyze, and invest. arXiv preprint arXiv:1301.41942

[18] Hoffman, C. M. (n.d.). Why Are Cells Powered by Proton Gradients? University of Texas at Austin. Retrieved from https://hoffman.cm.utexas.edu/courses/lane_atp_synthase.pdf

[19] Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. Proceedings of COMPSTAT'2010, 177-1861.

[20] Wadia, N. S., Dandi, Y., & Jordan, M. I. (2023). A Gentle Introduction to Gradient-Based Optimization and Variational Inequalities for Machine Learning. arXiv preprint arXiv:2309.04877.

[21] Maurya, M., Yadav, N. (2023). A Comparative Analysis of Gradient-Based Optimization Methods for Machine Learning Problems. In: Yadav, A., Gupta, G., Rana, P., Kim, J.H. (eds) Proceedings on International Conference on Data Analytics and Computing. ICDAC 2022. Lecture Notes on Data Engineering and Communications Technologies, vol 175. Springer, Singapore. https://doi.org/10.1007/978-981-99-3432-4_7

[22] Alhijawi, B., Awajan, A. Genetic algorithms: theory, genetic operators, solutions, and applications. Evol. Intel. 17, 1245–1256 (2024). https://doi.org/10.1007/s12065-023-00822-6

[23] Kirkpatrick, S., Gelatt Jr, C. D., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. Science, 220(4598), 671–6802

[24] Poli, R., Kennedy, J., & Blackwell, T. (2007). Particle Swarm Optimization: An Overview. Swarm Intelligence, 1, 33–572.

[25] Szeliski, R. (2022). Computer Vision: Algorithms and Applications, 2nd Ed.,Springer

[26] Yadav, D., Desai, J., & Yadav, A. K. (2022). Automatic Text Summarization Methods: A Comprehensive Review. arXiv preprint arXiv:2204.01849

[27] Sharma, N. A., Ali, A. B. M. S., & Kabir, M. A. (2024). A review of sentiment analysis: tasks, applications, and deep learning techniques. International Journal of Data Science and Analytics.

[28] Padilla-García, E. A., Cervantes-Culebro, H., Rodriguez-Angeles, A., & Cruz-Villar, C. A. (2023). Selection/control concurrent optimization of BLDC motors for industrial robots. PLoS ONE, 18(8), e0289717.

[29] Biswas, S., Belamkar, P., Sarma, D., Babaee Tirkolaee, E., & Bera, U. K. (2024). A multi-objective optimization approach for resource allocation and transportation planning in institutional quarantine centers. Annals of Operations Research.

[30] Emmerich, M., & Deutz, A.H. (2024). Multicriteria Optimization and Decision Making Principles, Algorithms and Case Studies. https://doi.org/10.48550/arXiv.2407.00359