УДК 004.63

Гарасівка А. В., Лупенко А. М., д. т. н.
(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

ВИКОРИСТАННЯ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ ДЛЯ ОПТИМІЗАЦІЇ РЕЗЕРВНОГО КОПІЮВАННЯ ДАНИХ

Harasivka A. V., Lupenko A. M., Dr.

USING MACHINE LEARNING APPLICATIONS FOR DATA BACKUP OPTIMIZATION

Data backup plays an important role in ensuring data integrity, security, and availability in the era of exponential data growth. Traditional backup strategies rely on static configurations or manual operations, often failing to adapt to dynamic environments or customers' evolving needs. Recent developments in machine learning (ML) and artificial intelligence show promising opportunities to optimize data backup systems.

ML offers new approaches to solve long-standing challenges in data backup area, usually, it requires a model, which is a set of algorithms and environment, which include training data. ML models can be broadly categorized into few categories. Supervised learning, which trains models using labeled datasets, can aid in predicting future requirements of the system or failure probabilities based on historical behavior, it could solve classification or regression problems. Unsupervised learning, which focuses on discovering patterns or anomalies in unlabeled data, can be used for detecting inefficiencies or unusual patterns in backup workflows via regression. Reinforcement learning (RL), a paradigm where models learn optimal policies through trial and error under developer control, is particularly promising for adaptive backup optimization. Features of different ML types shown in fig. 1.

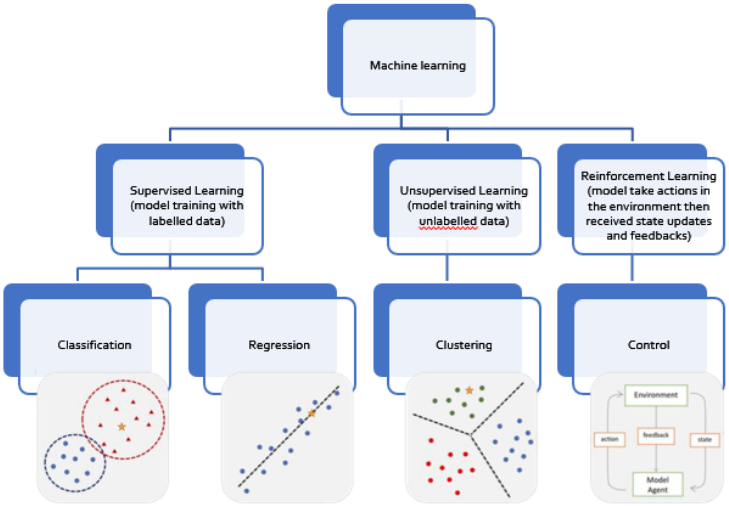

Figure 1. types of calculations that can be done using machine learning

RL algorithms can dynamically optimize backup schedules, types (e.g., select needed type of backup: incremental, full, or differential), and compression parameters based on system feedback, reducing resource usage while maintaining robustness. Supervised and unsupervised learning models can predict failure points or anomalies in backup processes, enabling proactive

interventions. Other ML techniques, like natural language processing techniques can assist in intelligent indexing and search within backup archives, improving accessibility.

Applying ML to modern data backup solutions has multiple benefits:

- Dynamic scheduling: ML models can analyze historical backup tasks to determine the optimal frequency and timing, balancing system load with recovery time objectives (RTOs).
- Compression optimization: models can predict compression ratios for various datasets and select configurations that maximize storage efficiency while minimizing compute costs.
- Anomaly detection: using historical data, a model can identify deviations in backup size, time, or frequency, marking potential risks like ransomware attacks or system misconfigurations.
- Storage forecasting: time-series models can predict storage consumption trends, enabling preemptive capacity planning.

Aside of benefits, there are a few challenges encountering during ML.

Despite its potential, deploying ML in backup systems faces issues such as data sparsity, heterogeneity of the system, and the computational cost of real-time inference. Privacy and security concerns, especially in handling sensitive datasets for training, require subtle solutions. Furthermore, the interpretation of guided ML decisions in critical backup operations remains a key challenge.

Future research should focus on developing lightweight ML models suitable for real-time predictions, integrating more complex learning techniques for secure and distributed model training, and combining ML with emerging technologies such as blockchain for immutable and transparent backup solutions. Collaborative efforts between the open-source community and resources from huge businesses can further accelerate the adoption of ML in backup systems, bridging the gap between theoretical advancements and practical implementations.

The integration of machine learning into data backup systems could cause a huge change, which will create smarter, more adaptive, and more reliable backup solutions. By addressing the outlined challenges, ML can transform backup processes, offering significant benefits to businesses and customers in managing their data effectively.

Література

1. Sutton, R. S., Barto, A. G., 2018. Reinforcement Learning, second edition, ISBN: 9780262039246.

2. Kumar S.. 2024. ENHANCING CLOUD STORAGE EFFICIENCY AND ACCESSIBILITY WITH ARTIFICIAL INTELLIGENCE: A COMPREHENSIVE REVIEW, https://iaeme-library.com/index.php/IJARET/issue/view/IJCET_2024_15_03

3. Züfle M., Moog F., Lesch V., Krupitzer C., Kounev S., 2021. A machine learning-based workflow for automatic detection of anomalies in machine tools, ISA Transactions. http://dx.doi.org/10.1016/j.isatra.2021.07.010

4. Ghadge N., 2024. MACHINE LEARNING: ENHANCING INTELLIGENT SEARCH AND INFORMATION DISCOVERY, Computer Science & Information Technology (CS & IT), 10.5121/csit.2024.141021.

5. Shahid N., Sheikh N., 2021. Impact of Big Data on Innovation, Competitive Advantage, Productivity, and Decision Making: Literature Review, Open Journal of Business and Management, 10.4236/ojbm.2021.92032.