HYBRID MACHINE LEARNING AND NUMERICAL METHODS FOR HIGH-DIMENSIONAL OPTIMIZATION PROBLEMS

SURESHREDDY KOMSANI
Lecturer in Computer Science
Governmnet Degree College,Yellareddy Dist: Kamareddy-503122
PEDDAKAPU DASHARATHAM
Ph.D Research Scholar
SJJTU Vidyanagri,,Jhunjhunu,Bisau Road, ChudelaDist- Jhunjhunu, Rajasthan -333010
YENDRU REVATHI
Ph.D Research Scholar
ARNI University, (Kathgarh), Tehsil Indora, Distt. Kangra, Himachal Pradesh – 176401

ABSTRACT

Accompanying exact methods are numerical methods that include a deterministic synthesis of drift terms and a stochastic interpolation of diffusion terms with the purpose of increasing precision and stability and optimizing used computing time. The focus of this work is to consider composite numerical techniques for the approximation of SDEs with nonlinear coefficients in the drift and diffusion terms.While discretising and solving differential equations can be accomplished with the help of traditional numerical methods like finite difference, finite element, and spectral methods, neural networks, especially deep learning models, excel at learning complicated patterns and estimating solutions in high-dimensional spaces. In situations when precise analytical answers are unattainable or very difficult to get, hybrid techniques may improve the solution process in terms of accuracy, efficiency, and generalisability by combining various methods. Combining these two approaches improves the management of singularities, boundary conditions, and irregular geometries, all of which are common in real-world applications such as biological modelling and fluid dynamics.Ultimately, when the amount of samples in the data set is reduced, the generalization of the physics-related corrective model outperforms the purely data-driven model, which also demonstrates efficient applicability of the proposed hybrid modelling approach to problems where data is scarce.

Keywords: numerical methods, hybrid modelling, numerical techniques, high-dimensional, fluid dynamics.

INTRODUCTION

Mathematical models are present in almost every area of science, as they play a vital role in problem-solving. These models provide a simplified representation of reality from mathematical formulations. These formulations make it possible to understand complex systems, solve problems, and obtain essential information to support intelligent decision-making. Mathematical models use algorithms to find the most appropriate solution to the problem described. Numerical Optimization is a well-known area of Mathematical Sciences that aims to identify extreme points of a function, whether maximum or minimum points. Optimization methods have become a crucial tool for management, decision-making, technology improvement, and development in the last two decades, providing competitive advantages to various

systems. Thus, optimization models and algorithms gained visibility in several areas, such as industry, disease diagnoses, professional and resources scheduling and allocation, financial area with capital management and scenarios forecast, among others. Another well-known area focused on solving problems using mathematical models and algorithms is Machine Learning. In some real-world problems, much information (data) should be processed. This amount of data usually requires computational assistance to transform the data information into relevant knowledge for problem solving. In this context, machine learning algorithms are extremely useful. These models and algorithms intend to generate a mathematical model that describes the data set and generalizes the knowledge to unknown data samples. Machine learning models and algorithms also have application in several domains named: industry, health, financial, education.

LITERATURE REVIEW

Manar Abd-ElRahman et al., (2025) Feature selection (FS) is critical for datasets with multiple variables and features, as it helps eliminate irrelevant elements, thereby improving classification accuracy. Numerous classification strategies are effective in selecting key features from datasets with a high number of variables. In this study, experiments were conducted using three well-known datasets: the Wisconsin Breast Cancer Diagnostic dataset, the Sonar dataset, and the Differentiated Thyroid Cancer dataset. FS is particularly relevant for four key reasons: reducing model complexity by minimizing the number of parameters, decreasing training time, enhancing the generalization capabilities of models, and avoiding the curse of dimensionality.

Kyoon-Ho Cha et al., (2025) This study introduces novel hybrid machine learning (ML) models that integrate six state-of-the-art ML algorithms with the Harris Hawks Optimization (HHO) algorithm to enhance the prediction of the incubation dose in irradiated metals. A comprehensive database comprising 305 experimental samples with 24 input features is used to develop the models, with hyperparameters optimized through a combination of cross-validation method and HHO. Performance evaluation across various metrics identifies the hybrid model combining HHO and categorical gradient boosting (CGB), named HHO-CGB, as the most accurate and stable for predicting the incubation dose. To gain further insights, the Shapley Additive Explanations method is employed to assess the global and local contributions of input variables, revealing Fe (wt.%), temperature (K), dose rate (dpa/s), and V (wt.%) as the most influential factors.

Ana I. Pereira et al., (2024) Notably, real problems are increasingly complex and require sophisticated models and algorithms capable of quickly dealing with large data sets and finding optimal solutions. However, there is no perfect method or algorithm; all of them have some limitations that can be mitigated or eliminated by combining the skills of different methodologies. In this way, it is expected to develop hybrid algorithms that can take advantage of the potential and particularities of each method (optimization and machine learning) to integrate methodologies and make them more efficient. This paper presents an extensive systematic and bibliometric literature review on hybrid methods involving optimization and machine learning techniques for clustering and classification. It aims to identify the potential of methods and algorithms to overcome the difficulties of one or both methodologies when combined.

Zhang, L., et al., (2019) In this study, we combine the theory of stochastic process and techniques of machine learning with the regression analysis, first proposed by to solve for American option prices, and apply the new methodologies on financial derivatives pricing. Rigorous convergence proofs are provided for some of the methods we propose. Numerical examples show good applicability of the algorithms. More applications in finance are discussed in the Appendices.

Li, H., et al., (2017) Modern manufacturing systems are expected to undertake multiple tasks, flexible for extensive customization, and that trends make production systems become more and more complicated. The advantage of a complex production system is a capability to fulfill more intensive goods production and to adapt to various parameters in different conditions. The disadvantage of a complex system, on the other hand, with the pace of the increase of complexity, lies in the control difficulties rising dramatically. Moreover, classical methods are reluctant to control a complex system, and searching for the appropriate control policy tends to become more complicated. Thanks to the development of machine learning technology, this problem is provided with more possibilities for the solutions. The objective of this paper is to cut down the makespan and the due date in the manufacturing system.

Conventional Numerical Techniques for Differential Nonlinear Equations

It is necessary to be familiar with the traditional numerical approaches that have been extensively used to estimate solutions for non-linear equations (NDEs) before exploring hybrid methods. The problem domain is often discretised and the differential equations are transformed into algebraic equations that may be solved repeatedly using numerical techniques like the finite difference method (FDM), finite element method (FEM), and spectral approaches. One of the most popular numerical algorithms for solving ordinary and partial differential equations is the finite difference method (FDM). It requires dividing the continuous domain into a grid of points and then using finite differences to approximate the derivatives in the equations. While FDM excels at solving issues involving basic geometries and boundary conditions, it fails miserably when faced with nonlinearities and complicated boundary requirements. One more popular numerical approach, particularly for issues involving complicated geometries, is the finite element method (FEM). To solve the governing equations, FEM discretises the domain into smaller subdomains, called elements. However, using FEM to very nonlinear systems may be computationally costly and complex, despite FEM's great flexibility and ability to provide more accurate solutions in irregular domains. Expanding the differential equation solution into a sequence of basis functions, usually trigonometric polynomials or Fourier series, is what spectral techniques do. Spectral techniques are great for solving smooth problems because of their fast convergence features, but they have difficulties with nonlinearities or discontinuities and aren't always the best choice for nonlinear differential equations. Although these approaches have their merits, they all encounter difficulties when dealing with nonlinear equations, complicated boundary conditions, high-dimensional issues, and irregular geometries. Here, hybrid approaches that use both numerical and neural network methodologies shine.

Solving Nonlinear Differential Equations using Neural Networks

The capacity of neural networks, particularly deep neural networks (DNNs), to understand underlying patterns in massive datasets and approximate complicated functions has drawn a lot of interest. Neural networks' primary benefit is its capacity to approximate very nonlinear functions without requiring a

thorough comprehension of the system's underlying mathematical structure. It is possible to train neural networks to approximate the solution of nonlinear differential equations across a specified domain. In order to teach the neural network to learn the mapping between input variables (such spatial and temporal coordinates) and the appropriate solution values, the NDE solution is framed as a regression problem. In practice, a neural network is trained to minimise a loss function associated with the differential equation, such that the network's output accurately satisfies the equation. The Physics-Informed Neural Network (PINN) technique is a prominent neural network application for NDEs. PINNs use the physics of the differential equation as a constraint during training in addition to the data to train the neural network.

High-dimensional data

There have been extensive studies in improving the performance of LDP in applications with high-dimensional data which include the use of various techniques including user partition, binary vector encoding, and other dimension reduction techniques such as the use of hash functions or matrix transformation. Although such techniques may reduce the impact of the high dimension of the data, they cause other problems such as an increase of error in construction and an increase in decoding complexity. this introduces another challenge in the design of LDP schemes, namely finding a good balance in the performance of such schemes in these metrics. Furthermore, solutions for statistical queries on multiple attributes are typically solved by the use of sampling since the majority of LDP schemes focus on single attribute queries. Sampling requires the users to be partitioned where each group of users may only report a part of the attributes. This causes the possibility that there is an insufficient number of reports in some of those attributes, causing the estimates for such attributes to have low statistical accuracy. How to effectively tackle all these issues and produce a more accurate estimate while preserving local differential privacy remains an open problem.

Hybrid machine learning

Most learning algorithms used in ML are really good at completing one task or working with one dataset. While helpful and infinitely better than doing it manually, these algorithms won't help you realize the full potential of AI across all of your data. That's where hybrid machine learning (HML) comes in. Multiple simple algorithms work together to complement and augment each other. Together they can solve problems that alone they were not designed to solve.

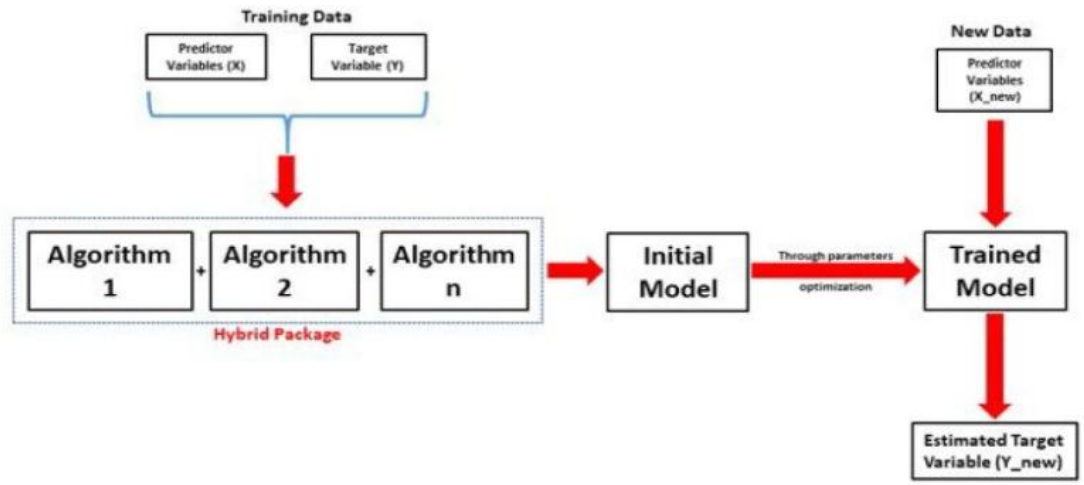

Figure 1: The HML work flow

HML Based on Architectural Integration

This type of HML seamlessly combines the architecture of two or more conventional algorithms, wholly or partly, in a complementary manner to evolve a more-robust standalone algorithm. The most commonly used example is adaptive neuro-fuzzy inference system (ANFIS).

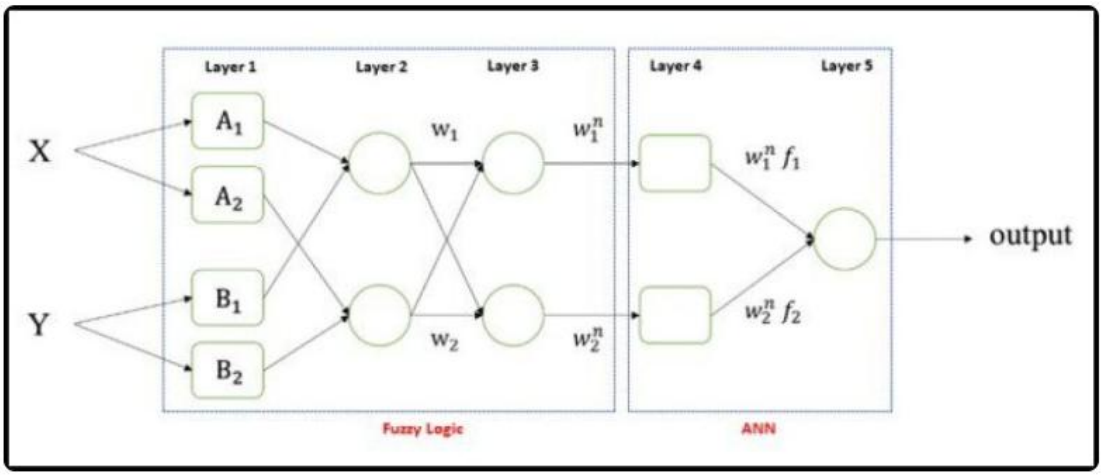

Figure2: The architecture of ANFIS

ANFIS has been used over time and is usually thought of as a standalone traditional ML method. It actually is a combination of the principles of fuzzy logic and ANN. As shown in Figure 2, the architecture of ANFIS is composed of five layers. The first three are taken from fuzzy logic, while the remaining two are from ANN. More technical details about ANFIS, and especially the role of each layer.

Semi-supervised learning

In semi-supervised learning, you provide the algorithm with a small set of labelled data. Then, you give it a much larger set of unlabeled data and put it to work. This type of algorithm is helpful when you need (or have) to start with a smaller batch of data upfront. It learns from all the data, not just the labelled data, and helps you organize it. This form of HML is especially helpful with data that changes over time. For example, use it to track things like the cost of supplies. As costs change, it will impact production and forecasts. You can use this method of HML in your inventory and supply chain management to forecast future costs. Or, it can be useful to track brand sentiment for customer retention. Track how current customers are engaging with or discussing your brand on social media, and use it to develop targeted mitigation strategies when customers fall below a designated threshold. Often, you can use semi-supervised learning in tandem with unsupervised and supervised learning methods. These additional models can help with grouping and training on unlabeled data.

Self-supervised learning

A self-supervised learning model combines unsupervised and supervised learning problems, then applies a supervised learning algorithm. You can create the model for the algorithm to follow, and it begins applying that to unlabeled data. This type of learning is commonly used on unlabeled images and defines actions that can be taken on those images—like rotating them, identifying color or grayscale, or distinguishing between real and fake photos.

Multi-instance learning

Multi-instance learning is a method where you are labeling groups or collections of data, rather than the individual members of the group. This is a helpful method when you're working with large sets of similar

data and have a lot of duplicates. This method uses supervised learning models to identify labels for groups of data. You train the models to recognize attributes of a few pieces of data within a group, and then it predicts labels for future groups based on attributes of some of the data within the new groups.

Benefits of the Hybrid Approach

By combining multiple algorithm types, HML offers several advantages over single-model solutions:

Higher Predictive Accuracy

Each component model in a hybrid machine learning (HML) system focuses on specific aspects of the data, combining their strengths to deliver more reliable and accurate predictions. For example, traditional machine learning can handle structured data effectively, while deep learning excels at unstructured data like images and text. By leveraging the strengths of both, HML systems significantly improve predictive performance across diverse datasets.

Greater Robustness

HML systems are designed to handle a wide variety of data types and tasks, making them highly adaptable. Whether dealing with structured numerical data, unstructured text, or multimedia content, these systems can integrate multiple approaches to ensure consistent results, even in changing conditions or when working with incomplete data. **Improved Interpretability**

Traditional machine learning components in HML systems provide valuable transparency, helping to demystify the decision-making processes of more complex deep learning models. For instance, decision trees or linear regression models can complement deep learning by offering clear, step-by-step insights into specific patterns, making it easier for stakeholders to understand results and identify potential biases.

Scalability and Flexibility

Hybrid models are designed to operate seamlessly across multiple environments, from on-premise systems to cloud platforms. This flexibility allows organizations to scale their operations efficiently, meeting growing demands without compromising performance. With the ability to integrate with modern cloud infrastructure, HML systems also support real-time processing and deployment across a range of industries and use cases.

Hybrid Approaches: Integrating Numerical Techniques with Neural Networks

Hybrid techniques aim to take use of the benefits of both traditional numerical methods and neural networks. The fundamental idea is to break the problem down into smaller parts using numerical methods and provide a first approximation of the solution. Neural networks then enhance this approximation by detecting patterns and non-linearities that traditional methods would have trouble recognising. Neural networks may be used to enhance or rectify traditional numerical methods. For example, neural networks may be used to correct the errors in an initial estimate that was created using finite difference or finite element methods.

Neural networks may also be included into the numerical framework itself. For example, neural networks may be trained to minimise the residuals of the governing differential equation and the boundary conditions, which allows them to directly approximate the solution. This method has been shown to enhance the resilience and efficiency of solving nonlinear differential equations by combining the flexibility of neural networks with the accuracy of traditional numerical approaches. Neural networks are better than traditional

methods in dealing with difficult boundary conditions, high-dimensional systems, and irregular geometries. In addition, hybrid methodologies may be used to enhance mesh production and optimisation in numerical techniques.

Tools to support hybrid machine learning

The point of incorporating ML into your data science and analytics processes is to allow you to begin looking forward with your data. Rather than relying on datasets that have been cleaned and organized, you'll be able to quickly group and label data in real-time for the most accurate analysis and forecasts. When considering how to manage this data, you'll need a few key features in your business intelligence tools to support this type of advanced analysis.

Integration from all your data sources

You'll need one place to manage your data and train your ML models. Find a tool that will allow for easy integration of all your data sources.

Real-time analysis

Many of the HML models mentioned here function best as they're learning from new data. Find tools that will support real-time ingestion and analysis, and then will push that data out to workers who can use it to improve performance right then.

Automatic decisions

Find a tool that will support automatic decisions for your team, with alerts and notifications for when your data passes specific thresholds. No matter your industry, your data will continue to play an increasingly important role in how you do business. Incorporating hybrid-machine learning techniques will be one of the best ways you'll be able to create tools that will allow you to get value from your data now and as your business grows.

CONCLUSIONS

This research aimed at the synthesis and utilization of combined approximate analytical numerical methods for SDEs with nonlinear coefficients for drift and diffusion, considering major issues of accuracy, stability and computational cost. The use of both deterministic and stochastic solvers that I have described in this study presents a viable solution to the challenges in SDEs that are caused by nonlinearities which are an issue in fields such as finance, biology and engineering. Ultimately, low prediction errors were reached with respect to the desired high-fidelity solution, provided by a numerical FE simulation in the investigated use case of LSP. The high-fidelity numerical data could easily be replaced by experimental data, enabling correction towards empirical measurements. A number of prerequisites for adequately performing the correction task have been identified. Primarily, unique relationships between inputs and outputs need to exist, where redundancies in the data can be an indicator for non-unique relationships.HML methods have become common in recent applications. We have probably been using some of them without realizing it. It is, however, necessary to know about them in the context of understanding the underlying concepts of their methods and how they work.

REFERENCES

1. Ana I. Pereira et al., (2024), "Hybrid approaches to optimization and machine learning methods: a systematic literature review", Machine Learning, issn: 1573-0565, vol. 113, pages. 4055-4097.https://doi.org/10.1007/s10994-023-06467-x
2. Manar Abd-ElRahman et al., (2025), "Optimizing high dimensional data classification with a hybrid AI driven feature selection framework and machine learning schema", Scientific Reports, issn: 2045-2322, vol. 15,https://doi.org/10.1038/s41598-025-08699-4
3. Kyoon-Ho Cha et al., (2025), "Hybrid machine learning model with optimization algorithm for predicting the incubation dose of void swelling in irradiated metals", Nuclear Engineering and Technology, issn: 2234-358X, vol. 57, issue. 9,https://doi.org/10.1016/j.net.2025.103661
4. Li, H., (2017), "Improve the Performance of a Complex FMS with a Hybrid Machine Learning Algorithm", Journal of Software Engineering and Applications, issn: 1945-3124, vol.10, no. 3, pages.257-272.
5. Zhang, L., (2019), "Derivatives Pricing via Machine Learning", Journal of Mathematical Finance, issn: 2162-2442, vol.9, no. 3, pages. 561-589.
6. Wani, A.A. (2025), "Advancing Material Stability Prediction: Leveraging Machine Learning and High-Dimensional Data for Improved Accuracy", Materials Sciences and Applications, issn: 2153-1188, vol.16, no. 2, pages.79-105.
7. Vyas, O. (2014),“A Feature Subset Selection Technique for High Dimensional Data Using Symmetric Uncertainty”, Journal of Data Analysis and Information Processing,issn:2327-7203,no. 4, vol.2, pages. 95-105.
8. X. Zeng, (2013), "Evaluation and Comparison of Different Machine Learning Methods to Predict Outcome of Tuberculosis Treatment Course," Journal of Intelligent Learning Systems and Applications, Vol. 5 No. 3, 2013, pp. 184-193.
9. Wang, Y. (2019),“Variance Estimation for High-Dimensional Varying Index Coefficient Models”, Open Journal of Statistics, vol.9, pages.555-570.
10. Pasha, M. (2017),“Survey of Machine Learning Algorithms for Disease Diagnostic”, Journal of Intelligent Learning Systems and Applications, vol. 9, pages 1-16.