Orissa Journal of Physics
ISSN 0974-8202

Vol. 27, No.2 –
Vol. 29, No. 1
Aug. 2020 – Feb. 2022
pp. 29-38

Artificial Intelligence Enhanced by Physics

R R RATH[1] and B R MOHAPATRA[2]

[1]Asst. Prof.,Department of Electrotins & Communication Engineering, NM Institute of Engineering and Technology,Bhubaneswar,India
rrretc1@gmail.com

[2]Faculty of physics, NM Institute of Engineering and Technology, Bhubaneswar, India
brmohapatra2@rediffmail.com

Received 18, 11. 2021 , Accepted 117.12. 2021

Abstract. We propose that intelligently combining models from the domains of Artificial Intelligence or Machine Learning with Physical and Expert models will yield a more "trustworthy" model than any one model from a single domain, given a complex and narrow enough problem. Based on mean-variance portfolio theory and bias-variance trade-off analysis, we prove combining models from various domains produces a model that has lower risk, increasing user trust. We call such combined models - physics enhanced artificial intelligence (PEAI), and suggest use cases for PEAI

Key Words: Narrow Artificial Intelligence, Machine Learning, Portfolio Analysis, Risk Management, Design Optimization, Knowledge Fusion

1. Introduction

1.1 Motivation

Recent theoretical developments in machine learning (ML), complemented by the astounding growth of computational power and the genesis of large data sets, have contributed to the rapid development of artificial intelligent (AI) systems. Even though the key findings required for a general AI system (strong AI) are considered a distant endeavor [1], AI systems designed to solve narrowly defined yet challenging enough problems (weak AI or narrow AI) are often comparable to or exceeding the performance of average humans [2, 3], and in many cases human experts, at these same tasks [4, 5, 6, 7]. These narrow AI solutions offer a great potential for industry to automate, improve, and surpass

unaided human productivity. In the rest of this paper, the term AI will refer to this narrow AI, unless stated otherwise. In spite of the great performance potential AI encompasses, user adoption has always been challenging. In many cases user trust becomes a bottleneck towards industry-wide adoption, especially in aerospace, safety, and defence, to name a few. New ways to enhance user trust in AI can directly affect user adoption at a large scale. In addition, if new ways to enhance user trust in AI can take advantage of existing solutions that were developed prior to AI solutions, it is likely to be more resource-friendly and even more attractive to industry.

1.2 Background and Related Work:

In pursuit of industry-wide adoption of AI, new areas of research that focus on the trustworthiness of AI have emerged. Trust is a topic of rich content deeply rooted in many historical and philosophical discussions, and is often tied with the study of risk in philosophical research [8]. As we are primarily interested in user adoption of new technology, we are not approaching trust from a philosophical research viewpoint and focus on the aspect of user trust.

To many ordinary users, the lack of trust in AI may have originated from the perception of the technology as a 'black box'. This perception reflects several other profound issues between human users and AI, including lack of understanding of the scientific principles of how the AI is constructed, lack of understanding of the functionality and limitations of ML based systems, and lack of transparency in the AI design process. Even for experts, the lack of straightforward ways of explanation of AI action using domain knowledge can lead to a 'black box' perception. Recently there are considerable research efforts on developing AI systems that are easily interpreted by humans, resulting in an emerging research field of eXplainable Artificial Intelligence, or XAI [9, 10, 11, 12]. XAI aims to bridge the gap of trust between AI system and its users by providing explanation of AI systems with the intentions to justify, control, and improve AI actions. Since explanations are subjective to the human observer, this area has also expanded to include psychology, philosophy, and cognitive science of people [13].

2. Physics Enhanced Artificial Intelligence (PEAI)

2.1 Definitions and Assumptions

In order to discuss the mathematical description of PEAI, we first discuss the basic concepts behind the idea of trust and risk in this context. In general, a user is more likely to trust and adopt a new technology - presented in the form of a model, when it is explainable and has good performance. We therefore assume

that user trust, for any given individual, is composed of three properties of the model:

1. Interpretability or Explainablity (E)
2. Performance Accuracy (A)
3. Performance Consistency (C)

In many practical applications where the task at hand is complex, AI models learned from data have lower interpretability compared with physics-based or expert-based models, and their consistency can be unknown or poor depending on how they are trained and the training data provided. However, they tend to be more accurate. Physics-based models are often quite interpretable and consistent, but are often not as accurate as the AI derived models. Expert models, while often accurate, may not be as consistent or explainable as physics-based models. We argue that combining models or knowledge from different domains of AI, physics, and expert for narrowly defined tasks, will yield a more trustworthy model than any one model or knowledge base they are composed from. Interpretability is subjective, and it is beyond the scope of our discussion of PEAI. By assuming that the interpretability of the models in question is constant based upon a given individual, we maximize user trust by maximizing the accuracy for a given model consistency.

As an attempt to qualify the relationship between trust and a given model, we express user trust as $\text{User Trust} \propto \text{UT}(a, c).$ (1)

where $a \in A$, $c \in C$, and UT is a function of performance qualities that are assumed to belong to partially ordered domains. Further, we assume that

$$\sup\{\text{UT}(a, c)\} = \text{UT}(\sup\{A\},\sup\{C\}), \quad (2)$$

$$\inf\{\text{UT}(a, c)\} = \text{UT}(\inf\{A\}, \inf\{C\}), \quad (3)$$

such that we can optimize UT. PEAI aims to develop a model to solve a narrow problem which consists of a set of rules and requirements that is described by task T . To avoid the discussion of trivial and unreasonable situations, we assume that T is sufficiently complex that the optimal solution is not known, and will require near infinite resources to identify. In order to solve T , a model, or a solution, $f : X \mapsto YT$ is constructed, where X consists of inputs from sensor measurements and YT its outputs. Let $U 6= \emptyset$ be the universal set of solutions. As models can be constructed using different methods, we define the following:

$A = \{fA \in U : fA \text{ constructed using only AI based methods }\}$ (4)

$P = \{fP \in U : fP \text{ constructed using only physics based methods }\}$ (5)

$$E = \{fE \in U : fE \text{ constructed using only expert based methods } \} \quad (6)$$

$$AP = \{fAP \in U : fAP \text{ constructed using AI and physics based methods } \} \quad (7)$$

$$AE = \{fAE \in U : fAE \text{ constructed using AI and expert based methods } \} \quad (8)$$

$$PE = \{fPE \in U : fPE \text{ constructed using physics and expert based methods } \} \quad (9)$$

$$APE = \{fAPE \in U : fAPE \text{ constructed using AI, physics, and expert methods}\} \quad (10)$$

We also assume that there exist multiple competing models from each of the above defined sets. A PEAI system (or model), fP EAI , belongs to the set

$$S = APE \cup AP \cup AE \cup (A \cap AP) \cup (A \cap AE) \cup (A \cap PE) \cup (A \cap P) \cup (A \cap E) \quad (11)$$

$$= APE \cup AP \cup AE \cup (A \cap PE) \cup (A \cap P) \cup (A \cap E). \quad (12)$$

For a complex task T , it is highly unlikely that one arrives at exactly the same mathematical model when using different modeling methods, therefore the intersections between any two sets of model sets are expected to be empty, i.e, $A \cap P$, $A \cap E$, $A \cap PE = \emptyset$. Under these situations, S reduces to

$$S* = APE \cup AP \cup AE. \quad (13)$$

We will examine fP EAI $\in S*$ for the remainder of this paper and discuss strategies for constructing fP EAI. For a complex problem, models are expensive to make, and no one model is perfect. We consider composing N finite number of models. Finally we assume that all models that are examined in U are constructed in good faith and aim to provide the best results possible given their application and method.

2.2 Construction of PEAI

There are two strategies to make a PEAI algorithm. The first is a composite model output approach - take models from the sets A, P, and E, and combine their outputs to form a new composite model in $S*$. The composite model approach will be analyzed using an analogy to classical mean-variance portfolio theory. The second is a hybridization model approach - modify the form of constructing the model by applying an intelligent constraint using information from another domain, generating a model in $S*$. We will analyze this hybridized model by using classical bias-variance trade-off analysis. We show that composed models using the above strategies yield a more consistent model for a desired accuracy.

2.2.1 Composite PEAI using Mean-Variance Portfolio Theory

The 1990 Nobel prize was awarded to Harry Markowitz for his 1952 'Portfolio Selection' essay[19]. His work laid the mathematical foundation of diversification by demonstrating that the combination of risky assets is less risky than any single asset. By treating the available models in U as risky assets, we can maximize user trust by minimizing the variance (risk) of the composite model. While this is conceptually a simple idea, it has profound impact on the understanding of ML ensembles and composite model techniques.

Assume there exists a function m: YT 7→ R + that can be evaluated on each of the models that gives a meaningful representation of the model performance. Each model i ∈ {1, 2, ..., N}, has the output Yi . Without a loss of generality, we further assume larger value of m(Y) indicate better performance. For the combined model, we have m(YC) = Σwim(Yi), (14)

where wi is the relative weight given to model i with P i wi = 1, and wi ≥ 0, ∀i. Therefore C is composed of all N models. Using the distributive property of the expectation denoted by E[·], we solve for the first moment of YC ,

$$\mu C = E\,[m(YC)] = \Sigma E[m(Yi)]wi = \mu\, T\, w, \quad (15)$$

where μk = E[m(Yk)], w = [w1 w2 ... wN], and μ = [μ1 μ2 ... μN]. We will assume that each element of μ is not equal to each-other, as the models are expected to give different expected values. The second moment of

$$m(YC),\ \sigma\, 2\, C = Var\,[m(YC)] = wT\, \Omega w, \quad (16)$$

Finally by substituting the above results into equation 16,

$$\sigma\, 2\, C = w{*}T\, \Omega w{*} = w{*}T\, \Omega\, \Omega\, {-}1\, (\lambda 11 + \lambda 2\mu)$$

$$= \lambda 1 + \lambda 2\mu C = (\alpha \mu 2\, C - 2\beta \mu C + \gamma)/\, \delta \quad (17)$$

It is important to note a few properties about α, γ, and δ. Since Ω is positive definite, α > 0 and γ > 0, and by the Cauchy-Schwarz inequality δ > 0.

Equation 17 allows us to solve for the minimal risk for a desired mean value. For each μi being unique, the models not being perfectly correlated, and N ≥ 3, the feasible region for portfolio theory can be shown to be a two-dimensional surface that is convex to the left, and is represented on the σ 2 C vs. μC plane, see Figure 1. In this figure, the class of optimal combined models lines on the thick black line between points P1 and P2. A sub-optimal model found by combining model outputs from various domains, as shown by P3, lies within the region with a dotted line boundary, and an unfeasible model lies outside this boundary, as shown by point P4.

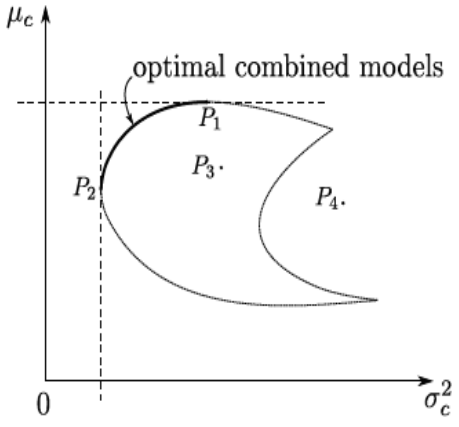

Figure 1

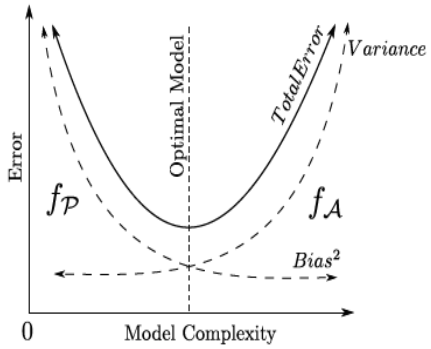

Figure 2

Fig. 1: Example of a feasible region of composite models from portfolio analysis

Fig 2: Model complexity vs. Total Error. The optimal models are likely to belong to the class of PEAI models.

The optimal combined model available exists on the curve between points P1 and P2 and is marked by the thicker line.

An example of the feasible point P3 could be constructed and interpreted under the following conditions: If one were to make a model using a linear combination of the outputs of all N models, and assign each model the weights of 1N, then one would construct the equivalent of an ensemble of models using majority vote to make a decision. In practice this has shown to increase the accuracy and reduce the variance of the model [20]; however, here we show that there exists a set of weights for each of these models that would minimize the risk of the prediction. Therefore, this ensemble, majority vote model is likely to be a sub-optimal solution for a given performance.

In the case where models come from different domains, it is more likely that the models are going to have different _i, be less correlated with each other, and have different properties of their predictions. For example, the models in A are more likely to be more accurate than those in P, but have a higher variance. Therefore when these models are combined the composite model is able to obtain a more optimal performance than any one given model.

2.2.2 Hybridization PEAI using Bias-Variance Trade-Off Analysis:

Hybridized PEAI models can be shown to have lower risk and enhanced user trust. First we derive the expressions of bias and variance in a model. Here we will assume that the model can be represented as a function of the inputs plus some error: y = f(x) + E: (18)

where E is noise with zero mean and variance _2, such that E[€] = 0 and E[€2] = Var[y] = _2. Also let ^ f(x) be a deterministic approximation the function f(x), F = f(x) and ^ F = ^ f(x).

Therefore we can express the mean squared error as a function of the Bias[^ F], _2, and Var[^ F]:

$$E\,[(Y - F\hat{})^2] = = \text{Bias}[\,\hat{}\,F]^2 + \delta^2 + \text{Var}[\,\hat{}\,F] \quad (19)$$

Normally, a physics models tend to have higher bias, but low variance. On the other hand, AI models tend to have a high variance and low bias. By placing physics constraints on the AI model during learning or during run-time, one place a bias on the new hybrid model, limiting the output space. This will cause the model to have a larger bias, and if done correctly will dramatically reduce the variance of the hybrid model. This concept can be shown graphically by model complexity vs. error diagram in Figure 2. By intelligently introducing physics based constraints to AI models or vice versa, we can arrive at models that have lower total error.

3. Implications of PEAI

Human learning builds on human observations and empirical evidence of the surrounding world, and this accumulated and learned knowledge is passed on, resulting in a systems design or model that is physics-based, as shown by the architecture in Figure 3. Similarly, data driven AI approaches use ML to arrive at a design or model based on the collected data. The combined model allows solutions of AI to be constrained by physical solution and expert knowledge, which enhances performance and trust. In addition, user trust in PEAI can be further enhanced by having a human supervisor in the loop, where the supervisor monitors the AI and provides feedback to the system that can improve its performance.

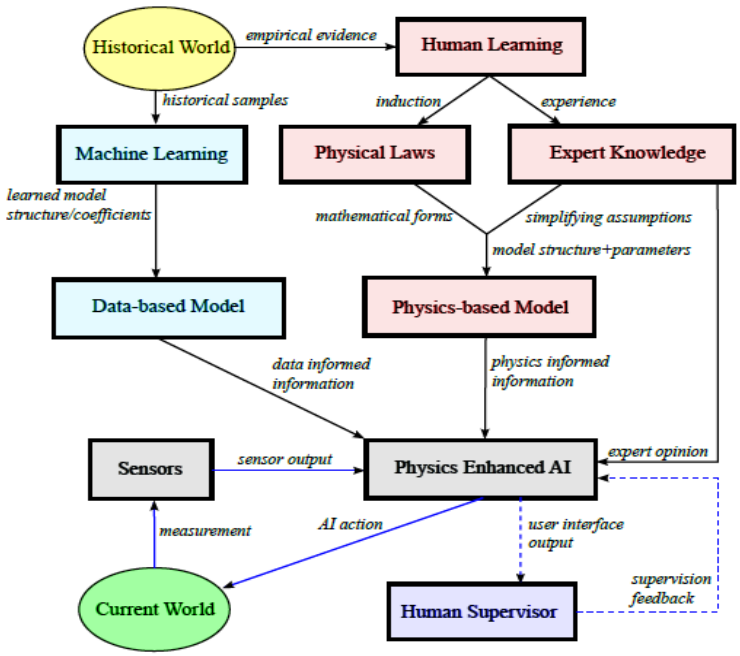

Fig. 3: Physics enhanced AI and its relationships with machine learning and human learning

It is interesting to point out that many have already worked on the class of PEAI systems, though the direct increase in user trust was not the motivation. For example, AI system with physical constrains is shown best by the work Physics informed deep learning [21, 22], where differential equation constraints that can represent a physical model is combined with a neural network to form a PEAI. In the field of predictive turbulence modeling, Physics-informed ML framework has been proposed [23], where the functional form of the Reynolds stress discrepancy in Reynolds-averaged Navier-Stokes (RANS) is learned directly based on available data. In the area of real-time vision-based event monitoring at industrial sites, Physical constraints are added to AI models in order to improve performance and enhance user trust. Being able to quickly identify and design AI solutions that have potential for wide-range industrial adoption is challenging. The AI solutions that are more likely to receive wider adoption are ones that can earn trust from users, and deliver an improved productivity at the same time. By explicitly pointing out the connection between enhanced user trust and combining models from different domains, it is suggested that one should always seek to combine AI models with prior models, if available.

4. Conclusions

Physics enhanced AI (PEAI) is a class of model that is formed by intelligently combining models from the domains of artificial intelligence or machine learning with physical and expert models. It was shown that by doing so, model risk is reduced, resulting in a more "trustworthy" model than any one model from a single domain. PEAI is shown as a solution to improve user adoption of an AI which solves a complex yet narrow enough problems.

References

[1] V. C. Müller and N. Bostrom, Future Progress in Artificial Intelligence: A Survey of Expert Opinion, pp. 555–572.Cham: Springer International Publishing, 2016.

[2] Y. Qian, Y. Yongxin, L. Feng, S. Yi-Zhe, and X. Tao, "Sketch-a-net: A deep neural network that beats humans,"Int J Comput Vis, vol. 122, pp. 411–425, 2017.

[3] C. Lu and X. Tang, "Surpassing human-level face verification performance on lfw with gaussianface," in AAAI, 2015.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in The IEEE International Conference on Computer Vision (ICCV), December 2015.

[5] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou,V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap,M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of go with deep neural networks and tree search," Nature, vol. 529, p. 484–489, 2016.

[6] Y. M. Assael, B. Shillingford, S. Whiteson, and N. de Freitas, "Lipnet: Sentence-level lipreading," CoRR,vol. abs/1611.01599, 2016.

[7] N. Brown and T. Sandholm, "Superhuman ai for heads-up no-limit poker: Libratus beats top professionals,"Science, vol. 359, no. 6374, pp. 418–424, 2018.

[8] S. Roeser, R. Hillerbrand, P. Sandin, and M. Peterson, eds., Risk and Trust, pp. 858–873. Dordrecht Heidelberg London New York: Springer, 2012.

[9] A. Amina and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (xai),"IEEE Access, vol. 6, pp. 52138–52160, 2018.

[10] S. Mohseni, N. Zarei, and E. D. Ragan, "A survey of evaluation methods and measures for interpretable machine learning," CoRR, vol. abs/1811.11839, 2018.

[11] F. K. Došilovi´c, M. Br˘ci´c, and N. Hlupi´c, "Explainable artificial intelligence: A survey," in 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 0210–0215, May 2018.

[12] O. Biran and C. V. Cotton, "Explanation and justification in machine learning : A survey or," in IJCAI-17 Workshop on Explainable AI (XAI) Proceedings, 2017.

[13] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," CoRR, vol. abs/1706.07269,2017.

[14] D. Sculley, J. Snoek, A. Wiltschko, and A. Rahimi, "Winner's curse? on pace, progress, and empirical rigor,"ICLR 2018 Workshop, 2018.

[15] D. Castelvecchi, "Can we open the black box of AI?," Nature, vol. 538, pp. 20–23, Oct. 2016.

[16] H. Jiang, B. Kim, and M. R. Gupta, "To trust or not to trust a classifier," in NeurIPS, 2018.

[17] J. Heinrich and N. Ofir, "Identifying and correcting label bias in machine learning," arXiv, vol. 1901.04966, 2019.

[18] M. Hind, S. Mehta, A. Mojsilovic, R. Nair, K. N. Ramamurthy, A. Olteanu, and K. R. Varshney, "Increasing trust in ai services through supplier's declarations of conformity," arXiv preprint arXiv:1808.07261, 2018.

[19] H. Markowitz, "Portfolio selection," The journal of finance, vol. 7, no. 1, pp. 77–91, 1952.

[20] F. Chollet, Deep Learning with Python. Greenwich, CT, USA: Manning Publications Co., 1st ed., 2017.

[21] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics informed deep learning (part ii): Data-driven discovery of nonlinear partial differential equations," arXiv, vol. 1711.10566, 2017.

[22] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations," arXiv, vol. 1711.10561, 2017.

[23] J.-X. Wang, J. Wu, J. Ling, G. Iaccarino, and H. Xiao, "A comprehensive physics-informed machine learning framework for predictive turbulence modeling," arXiv preprint arXiv:1701.07102, 2017.