Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/4/20230345

Review on the coupling and promotion of machine learning methods to CFD

Haocheng Zheng

Hohai university, Nanjing, Jiangsu Province, China

Zhenghaocheng2001@outlook.com

Abstract. As the computer level advances in 1990s, CFD (the computational fluid dynamics, for short) has promoted fast. The process of using CFD to analyze complex or ideal conditions has a very wide application background. However, confronted by cases with higher accuracy as well as larger number of samples, CFD needs to spend much time and money for the sake of resolving the issues. ML (Machine learning, for short) approach provides a promising choice for CFD. This paper reviews the coupling of ML and CFD and the progress in promoting the application of CFD. This paper briefly introduces CFD along with ML approaches, such as supervised learning, unsupervised learning and reinforcement learning. This article also discusses challenges and issues with the aim of being resolved in the research of ML model based on CFD, such as using multiple machine learning models or hybrid models to solve problems and quantifying the uncertainty of machine learning models. If these problems are solved, ML method can provide a promising development prospect for CFD.

Keywords: Machine Learning, CFD, Promotion.

1. Introduction

As the computer technique along with numerical analysis technique advances continuously, CFD has developed rapidly in the past two decades. Its cost is lower than the practical experiment. Besides it can encourage advantages of complex or perfect situations make it become backbone of many fields of research, and make virtual modeling become a reasonable substitute for physical development and testing. But in some cases computational fluid dynamics is not the preferred solution. By the virtue of the perplexity along with highly accurateness demands of the system, computational fluid dynamics needs to call vast quantities of functions. Simultaneously, the iterative optimization course requires much time cost, which is often not affordable for researchers.

In this case, the data-driven ML model is able to resolve the issue well. In recent years, with computing ability improves continuously, machine learning has slowly become one of the pivotal ways to deal with issues of complex engineering systems. It is committed to studying how to improve the performance of the system itself by means of calculation and experience. Machine learning method can effectively use big data for nonlinear regression, classification and prediction, opening up a variety of new fields and generating revolutionary new progress. At the same time, machine learning allows data sets to have certain errors, and has a high fault tolerance rate. It can adapt to a wide range of input conditions, and can continuously optimize to obtain reasonable and better learning results. The learned

DOI: 10.54254/2755-2721/4/20230345

machine learning model can attain the direct nonlinear mapping correlation between specific input and output conditions. Besides, it can achieve real-time prediction and early warning for problems studied.

A recent study summarizes the progress of machine learning methods in gasification research, and points out that machine learning can become a promising choice for traditional modeling techniques in the field of gasification [1]. However, as far as author knows, there is no literature on the coupling and optimization of machine learning for traditional CFD technology. Therefore, this paper will discuss and analyze mixing and optimization of traditional CFD technique based on ML model in current world. Also, the paper discusses limitations of coupling optimization of the two, wishing to offer a few references for researchers in this domain.

2. Machine learning method and introduction

ML is based on data with the intention of setting up one probabilistic statistical model, and then use this model in order to analyze and predict. Machine learning is an inevitable product of the development of artificial intelligence to a certain extent. From the 1950s to the early 1970s, AI (the artificial intelligence, for short) study was still in the early development stage. During this stage, the " Logic Theorist " program of A. Newell and H. Simonde1 and the " General Problem Solving " program all produced encouraging results. Since the mid-1970s, vast quantities of expert systems have emerged, and have achieved remarkable results for many fields. However, at the same time of development, people also realize that expert system faces the bottleneck of knowledge engineering. People realize that it is difficult for human beings to summarize knowledge and then teach computers, so scholars begin to seek ways for machines to learn knowledge themselves. In the summer of 1980, the initial ML seminar (IWML) was successfully conducted in Carnegie Mellon University in America. During 1986, the first ML professional journal Machine leaning was founded. Since 1980s, machine learning has gradually become an independent discipline, which is also together with the appearance of diverse machine learning technologies. In the mid-1990s, statistical learning occupied the mainstream of research, its representative technology is support vector machine and more general kernel method. Nowadays, machine learning is well adapted to the needs of this era for the analysis and utilization of large amounts of data, so the field naturally has great development. At the same time, machine learning also provides important technical support for many interdisciplinary subjects, and is closely related to the lives of ordinary people, such as commercial marketing, weather forecasting, etc. ML generally comprises supervised and unsupervised learning as well as learning of the reinforcement.

2.1. Supervisory learning

Supervisory learning denotes ML issue of studying the forecast pattern from corresponding information of given input along with output. The essence from the supervised learning is just analyzing mapping law from the input to the output. Supervisory learning is one of the important branches of ML approaches. Also it is currently the most extensively adopted method for various fields. Categorization and regression are the representatives of supervised learning, including linear regression, decision tree, SVM (the support vector machine, for short), ANN (the artificial neural network, for short),Bayesian categorization algorithm. ANN is a widely parallel interconnected network composed of adaptive simple units, whose organization can imitate the interaction of biological nervous system to objects. The support vector machine is the foundational pattern of linear classifier with the biggest interval in characteristic space. This method can find an optimal hyperplane for local disturbance tolerance of training samples, diminish the experienced mistake as well as the perplexity of model, for the purpose of obtaining one generalized regression problem. ANN and SVM are able to resolve the complex issues. Nevertheless, the sample size which SVM could handle is small, and if the sample size is too large, classification accuracy will decrease. These two models are widely used in the coupling model of CFD and machine learning.

DOI: 10.54254/2755-2721/4/20230345

2.2. Unsupervised learning

The unsupervised learning is a crucial branch of ML, too. Unsupervised learning is mainly used to analyze natural data, namely unclassified data. Mainly includes clustering, dimensionality reduction, probability statistics. Unsupervised learning can be used for data analysis or pre-processing of supervised learning. Unsupervised learning mainly includes convolution neural network, principal component analysis (PCA), k-means clustering, hierarchical clustering and autoencoder.

2.3. Other machine learning methods

Besides the supervised learning together with the unsupervised learning, ML contains semi-supervised learning, reinforcement learning as well as active learning. Semi-supervised learning is a machine learning model that uses labeled data along with unlabeled data with the intention of studying the forecast models. It can use rules of unlabeled data to help with the supervised learning of the labeled data together with the save costs. Reinforcement learning refers to the machine learning problem that machines gradually learn and obtain the optimal strategy in the interaction with the environment. Active learning stands for ML problem which the machine gives examples continuously and actively to obtain annotations, and then uses annotation data with a view to learning the prediction model.

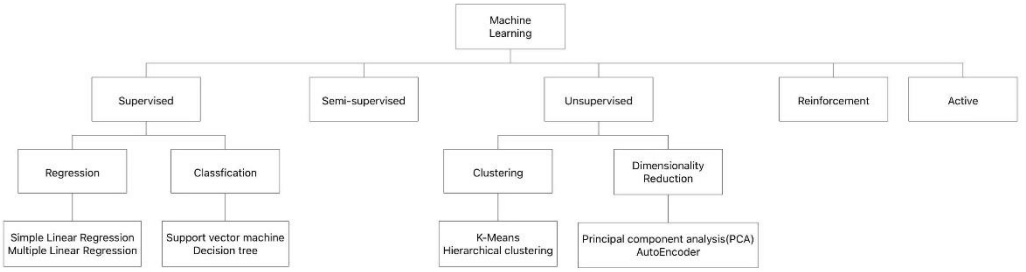

Figure 1. Machine Learning Classification.

3. Introduction of CFD

CFD is a branch of fluid mechanics, which is an interdisciplinary subject combining fluid mechanics. numerical mathematics as well as the computer science. CFD principle depicts characteristics of fluid flow through differential equations comprising much conservation, momentum conservation, energy safeguard as well as definite solution conditions, and then solves these discrete differential equations in space and time by computer to obtain numerical solutions. The CFD approach is just resolving governing equations which describe flow characteristics on the grid nodes. General solving process is as follows :(1) establishing the governing equations (2) establishing the initial conditions and boundary conditions (3) dividing the computational grids and generate computational nodes (4) establishing discrete equations (5) separating the primary and boundary conditions (6) giving control parameters (7) solving discrete equations (8) if the equations converge, then display and output the calculation results [2].

4. Application of ML to CFD Optimization

4.1. Reduce time cost

ML has been used in research to reduce the cost of CFD simulation, which is the most obvious advantage of this technology for CFD and therefore most used by researchers. In the traditional CFD simulation, the established model can be used to sort out the different effects under different input conditions. However, in the case of too many input parameter combinations, it is unrealistic and uneconomic to use CFD to simulate all parameter combinations. Then, representative data can be adopted for the purpose of drilling ML pattern, and the model accuracy can be verified by comparing the predicted CFD

DOI: 10.54254/2755-2721/4/20230345

simulation results with learning results of ML pattern. This approach is able to obviously decrease the time of the computation and facilitate the study while keeping computing resources unchanged.

The specific steps are as follows :

Firstly, collect initial data : get corresponding output situations under diverse input situations by CFD simulation.

Secondly, to instruct ML pattern: to instruct ML pattern using data obtained.

Thirdly, validation of the model : use the learned model and CFD prediction results are compared.[3]

Sarah et al. pointed out that the machine learning tools trained by CFD simulation of ideal dune data can help form a coherent flow structure. Machine learning tools can also classify and synthesize the specific configuration of parameters, which is of great significance for the formation of turbulent structure on the dune system. These achievements are often difficult to achieve by hundreds of CFD simulation results. At the same time, this method will be two-thirds less simulation than other methods, which will greatly save computing time. [4]

In another study, the machine learning model trained by 2000 samples generated by CFD allows less CFD simulation compared with the traditional CFD-GA technology after being combined with genetic algorithm (GA), but still can obtain optimization precision of advantages of traditional method. At the same time, ML-GA technology can also be extended to higher computing platforms such as supercomputers [5].

With CFD output as input and output data of ANFIS method, Xu et al found that the combination of ANFIS and CFD method can avoid complex equations in CFD, reduce the computational time of CFD method, and predict more nodes in a short time. Synchronously, ANFIS can visualize nanofluid flow in cavity. Also, it can display the heat allocation with the change of the heat source [6].

4.2. Complicated issues solving and the best solution seeking

When seeking an optimal solution, machine learning method is often a good method to evaluate the alternatives. The machine learning model can set up a value function covering multiple variables and seek to minimize the risk between the experiment and the candidate scheme.

In the study of fuel alternatives in CFD simulation, a Bayesian optimization algorithm based on customized value function is developed to determine the best alternative. The definition target of this function is just pursuing minimum Euclidean distance between alternatives and experiments. This method was applied to determine the best alternatives to four different gasolines and could better predict the performance of two hydrocarbon gasolines and one ethanol gasoline [7].

Li et al. tested the accuracy of residence time and nondimensional groups of basin geometric and dynamic similarity (Hazen, Reynolds and Schmidt numbers) in measuring the separation performance of clarifier. A 160,000CFD simulation was carried out through a wide range of basin configurations, loading conditions, and PM granulometry configurations. Based on the CFD data set, four new particle separation prediction alternative models (ANN, SR, DT, RF) were developed and trained. The prediction results of the new ML model have high accuracy, and the prediction results of particle separation are always within ± 15 % [8].

From another study, RBNN (the radial basis function neural network, for short) trained by three-dimensional computational fluid dynamics solution is used as an alternative pattern. Additionally, PSO (the particle swarm optimization, for short) algorithm is integrated to obtain optimal position of the jet. In the case of predicted optimal value, CFD solution is computed and compared. It is found that percentage error of thermal resistance along with the temperature uniformity of objective function is small. The benefit of this approach is just that the perplexity of fluid interaction in design space can be considered [9].

4.3. Construction of real-time monitoring and early warning system

ML is able to discover rules and forecast deep development trend based on one huge historical data set. This feature is often used in detection and early warning systems. In addition, machine learning has high

DOI: 10.54254/2755-2721/4/20230345

robustness, reliability and universalization capability, which is to manage complex problems greatly. This is the traditional manual experience judgment or general mathematical model is difficult to do.

Yang et al. used 240 sets of hydraulic erosion data obtained by CFD model to participate in machine learning program, and applied GBRT model to erosion prediction for the first time. This combination of CFD and ML can only enrich the prediction method to better predict the erosion degree of elbow under high internal pressure, which has high reliability and can be further applied to practice [10].

Predicting diffusion boundary of the gas divulgation accidents has always become a tough issue in the chemical industry. At present, the most pervasively adopted approach is just using CFD for predicting. But workplace facilities are often easy to change, using CFD means a lot of computing costs. Therefore, Dooguen et al. employed a neural net with the objective to studying the correlation among leakage gas diffusion, velocity field and facility layout. The results of this study can facilitate staff to formulate emergency plans and can be used for the development of CFD surrogate models [11].

5. The existing problems

In the author ' s data collection and collation, also found that the field still needs to be addressed and worthy of attention :

(1) When using ML pattern, a new origin of indetermination is also introduced. Therefore, quantifying this uncertainty is an interesting area to explore.

(2) For specific machine learning model applications, using multiple models together or using hybrid models can avoid the inherent shortcomings and vulnerabilities of individual machine learning models. This will become a mainstream trend in machine learning model applications.

(3) For the CFD dataset learned by machine learning, higher dimensional, higher precision and more representative data should be selected.

(4) CFD users in many fields lack relevant knowledge of ML. They demand to carry out interdisciplinary training for the purpose of understanding ML algorithms as well as the related theories.

6. Conclusion

At present, the coupling between machine learning and CFD is becoming more and more closely and widely. At the same time, the machine learning model based on CFD is gradually replacing the traditional CFD because of its lower cost requirements and the same high accuracy. Machine learning model has good real-time prediction and early warning ability, and is also favored by many disaster management personnel. Due to the author's limited vision and knowledge, this article inevitably exists a certain omission and summary of the loss. However, it can be imagined that the CFD-based machine learning model has a very broad application prospect.

References

[1] Umenweke Great C.,Afolabi Inioluwa Christianah,Epelle Emmanuel I.,Okolie Jude A.. Machine learning methods for modeling conventional and hydrothermal gasification of waste biomass: A review[J]. Bioresource Technology Reports,2022,17.

[2] Wang J J. Study on optimization of slagging prevention furnace for waste incinerator based on CFD numerical simulation [D]. Zhejiang university,2022. DOI:10.27461/d.cnki.gzjdx.2022.000154.

[3] Qiu Fang, Zhe Li,Yaonan Wang,Mengxuan Song,Jun Wang. A neural-network enhanced modeling method for real-time evaluation of the temperature distribution in a data center[J]. Neural Computing and Applications,2019,31(12).

[4] Wakes Sarah J.,Bauer Bernard O.,Mayo Michael. A preliminary assessment of machine learning algorithms for predicting CFD-simulated wind flow patterns over idealised foredunes[J]. Journal of the Royal Society of New Zealand,2021,51(2).

[5] Moiz A A, Pal P, Probst D, et al. A machine learning-genetic algorithm (ML-GA) approach for rapid optimization using high-performance computing[J]. Society of Automotive Engineers Technical Paper Series, 2018, 11(5).

DOI: 10.54254/2755-2721/4/20230345

[6] Xu P, Babanezhad M, Yarmand H, et al. Flow visualization and analysis of thermal distribution for the nanofluid by the integration of fuzzy c-means clustering ANFIS structure and CFD methods[J]. Journal of Visualization, 2020, 23(1): 97-110.

[7] Singh Samrendra K.,Braginsky Daniel,Tamamidis Panos,Gennaro Monacelli. Designing Next Generation Exhaust Aftertreatment Systems Using Machine Learning[J]. SAE International Journal of Commercial Vehicles,2020,13(3).

[8] Li Haochen,Sansalone John. A CFD-ML augmented alternative to residence time for clarification basin scaling and design.[J]. Water research,2021,209.

[9] Qidwai Mohammad Owais,Badruddin Irfan Anjum,Khan Noor Zaman,Khan Mohammad Anas,Alshahrani Saad. Optimization of Microjet Location Using Surrogate Model Coupled with Particle Swarm Optimization Algorithm[J]. Mathematics,2021,9(17).

[10] Yang Siqi,Fan Jianchun,Zhang Laibin,Sun Bingcai. Performance prediction of erosion in elbows for slurry flow under high internal pressure[J]. Tribology International,2021(prepublish).

[11] Song Dooguen,Lee Kwangho,Phark Chuntak,Jung Seungho. Spatiotemporal and Layout-adaptive Prediction of Leak Gas Dispersion by Encoding-Prediction Neural Network[J]. Process Safety and Environmental Protection,2021(prepublish).