The Application of Machine Learning in Chemical Engineering: A Literature Review

Baiyu Lu(✉)

Department of Chemical Engineering, Guangdong Technion - Israel Institute of Technology, Shantou, China
baiyu.lu@gtiit.edu.cn

Abstract. In chemical engineering research, to develop a Physico-chemical model is expensive. Also, it's very difficult for a computationally tractable model to perfectly describe many complex phenomena. Nowadays, machine learning can learn complicated behavior. Also, the development of its model is more economical. Machine learning offers significant advantages over traditional modeling techniques, including flexibility, accuracy, and speed of execution. This paper conducted a literature review in the application of machine learning (ML) in chemical engineering (CE). Four questions were being researched. What are the CE fields that ML applied mostly to? For what purposes is ML used in CE research? What are the algorithms that are mostly used in CE research? What are the difficulties of these applications? After examining 48 papers, the conclusion was that chemical process engineering was the field ML was applied mostly, the primary purposes for applying ML algorithms in CE were prediction and optimization, Artificial Neural Network was the most used algorithm in CE research, and the limited application field was the main existing difficulty. This paper can be a groundwork for future CE and ML research efforts as a result of these findings.

Keywords: Chemical engineering · Machine Learning

1 Introduction

Chemical engineering (CE) is the field of engineering concerned with the operation and design of chemical plants as well as the improvement of production methods. A CE process unit usually is designed to efficiently use, produce, design, transport and transform energy and materials. And it will always be modeled and programmed for control before construction. Due to many reasons, for example, the economic reason needs the process to be shorter, the environmental reason needs the process to be recyclable and so on, the CE process requires dynamic operation and a new paradigm for identifying new process routes and the design of flexible plants. In order to optimize in complex environments, traditional models are developed based on mechanical understanding and optimization. However, developing a Physico-chemical model is expensive. Also, it's very difficult for a computationally tractable model to perfectly describe many complex phenomena [1].

As a research field, machine learning (ML) is a subset of artificial intelligence (AI). AI refers to the ability of a machine to perform assignments that are often related to

R. B. B. M. Hussain et al. (Eds.): ICHSSR 2023, ASSEHR 765, pp. 57–66, 2023.
https://doi.org/10.2991/978-2-38476-092-3_9

the behavior of intelligent beings, for example, human beings. In the 1980s, the field of machine learning had begun to develop. However, regardless of some exceptions like Artificial Neural Network (ANN) and Genetic Algorithms (GA), a gap of about 10 years was experienced in the growth of machine learning in the chemical engineering research. However, the past decade has marked a breakthrough in deep learning. Deep learning is a subset of ML that builds an Artificial Neural Network that we have mentioned above to simulate the human beings' brains. But what makes deep learning different is that it provides the computational means to train multiple layers of neural networks, which are called deep neural networks. These new advances stimulated the interest of chemical engineers, as evidenced by the exponential growth in publications on the subject [2]. In the past, AI techniques could never become a standard tool in chemical engineering.

Nowadays, ML can learn complicated behavior. Also, the development of its model is more economical. As a result of this, ML has the potential to overcome the limitations of traditional mechanistic modeling [3]. Aiming to ensure the optimal operation and shorten the development cycles for new processes, more and more researchers apply ML in CE process modelling [4]. For these reasons, CE is currently transitioning to digitalization and full automation of industry and research. As a result, we face a technology-driven and industry-pulled situation, and ML creates new opportunities to get over the impending challenges of CE [5]. To get an overview of the research on applying ML in CE research, we conducted a literature review.

In this literature review, we came up with 4 questions in Sect. 2 and examined 48 papers to present the findings of the above questions in Sect. 3, before concluding the paper in Sect. 4.

2 Literature Review

The main purpose of this research was to better understand the application of ML in CE. In this study, 48 existing CE papers for which ML was applied were selected. We analyzed them by addressing 4 research questions and examining the literature. Finally, the results of the findings were documented.

2.1 Research Questions

- RQ#1. What are the CE fields that ML applied mostly?
  In the beginning, we wanted to understand a wide range of application scenarios using ML in CE research projects.
- RQ#2. For what purposes are ML used in CE research?
  Next, we wanted to analyze that the different motivation of applying ML in CE.
- RQ#3. What are the algorithms that are mostly used in CE research?
  Through this question, we gained an overview of specific practical algorithms in the application of ML in CE.
- RQ#4. What are the difficulties of these applications?
  Finally, the problems and difficulties in existing applications were analyzed. These findings indicated research directions for further improvement [6].

2.2 Method

Our approach is based on the research process proposed by R.L. Glass et al. and comprises three steps which we describe in the subsequent sections [7].

Selection of Papers. In order to obtain an important and current snapshot of the ML field in CE research, we decided to conduct this study over the period 2017–2023. We first chose the key journals that have been used over the years for other study about the Top Scholars and Top Institutions in the fields of ML and CE [7]. Then, papers were selected in key journals of the artificial intelligence (AI) and CE literature during that period (Table 4).

The search string contained keywords that were relevant to our research questions and basically, the search string used terms of the two fields of ML and CE. Synonyms were also extracted for them from different papers. At last, we got the search strings (Table 1).

And we adjusted the search string according to the specific syntax of the data bases. Then, we searched the strings in the title, abstract, or the main text of the article.

Inclusion and Exclusion Criteria. We evaluated each article based on its title, inclusion (Table 2), and exclusion criteria (Table 3). We first applied exclusion criteria to each article we found. If none of the exclusions are correct for the publication, we determine inclusion by applying the inclusion criteria [6]. A publication was included if it met at least one of the inclusion criteria.

There were 1697 articles originally. After initial evaluation and removal of duplication, the number of articles was reduced to 183. We excluded articles that did not address our research questions and ended up with 48 articles for review (Table 4).

Categories and Classification Schemes. In classifying the CE fields and the purposes, we thought that it was necessary to define in advance the categories into which we would classify the papers. We viewed this as a top-down approach. However, since there are too many algorithms and the existing difficulties are unknown, we applied a bottom-up classification driven by the papers themselves [8–10].

Table 1. The search string

Chemical Engineering	Machine Learning
Artificial Intelligence	Algorithms
Deep Learning	Digital Chemical

Table 2. The inclusion criteria

1	The paper contains an approach of the application of ML in CE.
2	The paper contains an approach to developing an ML algorithm in CE.
3	The paper contains a comparison to some ML algorithms in CE.
4	The paper contains a study from a ML algorithm which applies in the CE.

Table 3. The exclusion criteria

1	The paper that doesn't address the ML and CE concepts.
2	The paper that is not written in English.
3	The full text of paper is not available.
4	Conference articles, book chapters, and review papers.

Classifying CE Fields and Purposes. To classify the branches of CE, we used the classification from Wikipedia (Table 5). And there were some papers that included several CE fields. We classified them into all fields that it involved in. When investigating the motivation of applying ML in CE of the papers, we distinguished between nine types of papers (Table 6).

Classifying Algorithms and Difficulties. For algorithms, we found them while we were examining the papers and recorded the mainly used algorithms mentioned in the papers and can just simply classify them by their names. We classified the papers that mentioned more than 1 algorithm independently. That means if one paper mentioned 2 algorithms, it will present in 2 categories (Table 7). Similarly, the mentioned difficulties were recorded and classified (Table 8).

Examining the Papers. Once the duration of the study, the journal to be studied, and the classification scheme to be used were decided, we began to study the paper itself. We first tried to classify a paper on the basis of its title, abstract, introduction, and keywords. If we can't classify them in that way, then the full paper had to be examined.

3 Findings

After finishing examining the papers, we collated the data into 4 tables respect to our 4 research questions. The tables contained the classification of each problem and the number of papers that fit in that classification. Also, the percentage weight of papers were calculated.

3.1 CE Fields that ML Applied Mostly

ML as a field addresses itself to the solutions in any possible application domain, the diversity of those domains becomes far-reaching. We found that 17 CE fields that ML have been applied in our selected papers, which have almost covered all branches of chemical engineering (Table 5). The mostly applied field was chemical process engineering with 35.4% in the papers we choose. Biochemical engineering and chemical reaction engineering also were the ML mainly applied fields. Their weight were 22.9% and 20.8%, respectfully. The second major applying fields were fluid dynamics and thermodynamics. Here, there were 16.7% for fluid and 10.4% for thermodynamics. The rest of fields were only include 1–3 papers, note that this was a significant drop from the number in chemical process engineering. Clearly, the interests of ML application field

Table 4. The selected articles and key journals

Journals	Number of papers
Digital Chemical Engineering	11
Chemical Engineering Journal	7
Computers and Chemical Engineering	6
Green Chemical Engineering	2
Chemical Engineering Science: X	2
Journal of Rock Mechanics and Geotechnical Engineering	1
Journal of Environmental Chemical Engineering	1
Journal of Saudi Chemical Engineering	1
Chemical Engineering and Processing - Process Intensification	1
Fuel Communications	1
Metabolic Engineering Communications	1
Chemical Engineering Science	1
Micro and Nano Engineering	1
Combustion and Flame	1
Chemical Engineering Research and Design	1
South African Journal of Chemical Engineering	1
Chem	1
Journal of Electroanalytical Chemistry	1
Metabolic Engineering	1
Chem Catalysis	1
Journal of Hazardous Materials	1
Biomedical Engineering Advances	1
Cleaner Chemical Engineering	1
Fluid Phase Equilibria	1
Food Hydrocolloids	1
TOTAL	48

were heavily focused on chemical process engineering, with a considerably diminished interest in other fields. Also, remind that since some papers were consisted of 2 or more fields, the total number of papers was greater than 48, and when we calculated the percentage weight, we also used 48 as the denominator so the total number of percentages was greater than 100%.

Table 5. The CE fields that ml applied

CE fields	Number of papers	%
Chemical process engineering	19	35.4
Biochemical engineering	11	22.9
Chemical reaction engineering	10	20.8
Fluid dynamics	8	16.7
Thermodynamics	5	10.4
Food engineering	3	6.3
Materials science	3	6.3
Biomedical engineering	2	4.2
Chemical reactor engineering	2	4.2
Green chemistry	2	4.2
Physio-chemistry	2	4.2
Unit operations	2	4.2
Plastics engineering	1	2.1
Transport phenomena	1	2.1
Environmental chemistry	1	2.1
Electro-chemistry	1	2.1
Nanotechnology	1	2.1

3.2 Purposes for ML Used in CE Research

The purposes of ML that used for CE research were also an important point to being noticed. That can help us to get an overview of what processes in CE research always apply ML, and what processes also can apply ML but not so much. We distinguished between nine types of papers. There were 14 papers that used ML algorithms for prediction and also 14 papers for optimization. These 2 purposes were the main motivations for applying ML in CE study. Each of them contributed 29.2%. Additionally, modelling and identification were the second import purposes with weight of 12.5% and 10.4%. Other purposes were mentioned less, only 1–3 papers, as shown in Table 6. Therefore, we can come to the conclusion that prediction and optimization were the leading purposes that ML used in CE research, modelling and identification were the secondary purposes, and ML can also be applied for other purposes like characterization and examination, but not often used.

3.3 Algorithms that Mostly Used in CE Research

This research question gave us some specific practical ML algorithms for CE studying. This contributed to analysis the mostly used algorithms in CE, additionally, that can be references for other researchers. We examined the papers and noted down the mainly

Table 6. The purposes for ml used in CE

Purpose	Number of papers	%
Prediction	14	29.2
Optimization	14	29.2
Modelling	6	12.5
Identification	5	10.4
Classification	3	6.3
Estimation	2	4.2
Characterization	1	2.1
Examination	1	2.1
Researching	1	2.1
Analyzing	1	2.1

used algorithms for each paper and got the following table (Table 7). The results showed that the Artificial Neural Network (ANN) was the mostly used algorithm in CE research, there were 11 papers have mentioned this algorithm with 22.9%. Following that, Support Vector Regression (SVR) and Random Forests (RF) were both being applied to 5 papers with 10.4%. Besides, there were still many algorithms that have been mentioned but used a few. Hence, we concluded that ANN was the most used algorithm in CE research, SVR and RF were also being mainly used; beyond that, there were still many algorithms used in CE but not applied frequently. Remind again that, similar to the *Sect.* 3.1, as we have mentioned above that some papers used more than 1 algorithm, so the total number of papers was bigger than 48, also, the total percentage was more than 100%.

3.4 The Existing Difficulties

Finally, the problems and difficulties in existing applications were analyzed. These findings indicated research directions for further improvement. There were 10 papers that haven't mentioned the difficulties. In addition to this, the limited application field was the mostly mentioned problems, 15 papers have talked about these difficulties with 31.3%. Then, limited samples and the expensive cost were also the existing difficulties with the percentage of 12.5% and 10.4%. There were also some difficulties caused by the data, like uncertain data, and unreliable data as shown in Table 8. In a word, the dominating existing difficulties of ML applied in CE was the limited application field. Secondly, limited samples and the expensive also caused difficulties. Besides, a few researches also influence by data problems.

Table 7. Mentioned algorithms

Algorithms	Number of papers	%
Artificial Neural Network	11	22.9
Support Vector Regression	5	10.4
Random Forests	5	10.4
Genetic Algorithm	3	6.3
Gradient Boosting Machine	3	6.3
Decision forest	2	4.2
Long-Short-Term-Memory Neural Network	2	4.2
K- Nearest Neighbour	2	4.2
Gaussian Process Regression	2	4.2
Recurrent Neural Network	2	4.2
Partial Least Square Regression	2	4.2
Naive Bayes	2	4.2
Regression Trees	2	4.2
Feed-Forward Neural Network	1	2.1
Kernel Density Estimation	1	2.1
Multiple Linear Regression	1	2.1
Stacking ensemble	1	2.1
Transcriptional Regulatory Network	1	2.1
Logistic regression	1	2.1
Decision Jungle	1	2.1
Voting Ensemble	1	2.1
Ensemble Trees	1	2.1
Binary Classification	1	2.1
Extra Trees	1	2.1
Deep Neural Network	1	2.1
Deep Deterministic Policy Gradient	1	2.1

Table 8. Difficulties

Difficulties	Number of papers	%
The limited application field	15	31.3
Limited samples	6	12.5
Expensive	5	10.4
Random factor present	4	8.3
Uncertainties of the data	4	8.3
Not reliable training data	3	6.3
High variability of the parameters	1	2.1
Not mentioned difficulties	10	20.1

4 Conclusions

CE is currently transitioning to digitalization and full automation of industry and research. ML creates new opportunities to get over the impending challenges of CE. In this paper, we conducted a literature review of the applications of ML in CE researches. We first came up with 4 research questions, and they were: What are the CE fields that ML applied mostly? For what purposes is ML used in CE research? What are the algorithms that are mostly used in CE research? What are the difficulties of these applications? Then we chose the papers in key journals and filtered them by inclusion and exclusion criteria. Finally, we got 48 papers for our literature review. We examined them by classifying them into four tables with respect to our 4 research questions.

By analyzing the 4 tables above, we already have the answers to our 4 research questions. First of all, the interests of ML application field were heavily focused on chemical process engineering, with a considerably diminished interest in other fields. Then, prediction and optimization were the leading purposes that ML used in CE research, modelling and identification were the secondary purposes, and ML can also be applied for other purposes like characterization and examination, but not often used. In addition, ANN was the mostly used algorithm in CE research, SVR and RF were also being mainly used, beyond that, there were still many algorithms used in CE but not applied frequently. Last but not least, the dominating existing difficulties of ML applied in CE was the limited application field. Secondly, limited samples and the expensive also caused difficulties. Besides, a few researches also influence by data problems.

References

1. Schweidtmann, Schweidtmann, Artur M., et al. "Machine Learning in Chemical Engineering: A Perspective." Chemie Ingenieur Technik, vol. 93, no. 12, 2021, pp. 2029–2039.
2. Dobbelaere, Maarten R., et al. "Machine Learning in Chemical Engineering: Strengths, Weaknesses, Opportunities, and Threats." Engineering, vol. 7, no. 9, 2021, pp. 1201–1211.
3. McBride, Kevin, and Kai Sundmacher. "Overview of Surrogate Modeling in Chemical Process Engineering." Chemie Ingenieur Technik, vol. 91, no. 3, 2019, pp. 228–239.

4. Venkatasubramanian, Venkat. "The Promise of Artificial Intelligence in Chemical Engineering: Is It Here, Finally?" AIChE Journal, vol. 65, no. 2, 2018, pp. 466–478.
5. Lee, Jay H., et al. "Machine Learning: Overview of the Recent Progresses and Implications for the Process Systems Engineering Field." Computers & Chemical Engineering, vol. 114, 2018, pp. 111–121.
6. Obaidi, Martin, and Jil Klünder. "Development and Application of Sentiment Analysis Tools in Software Engineering: A Systematic Literature Review." Evaluation and Assessment in Software Engineering, 2021.
7. Glass, Robert L., et al. "An Analysis of Research in Computing Disciplines." Communications of the ACM, vol. 47, no. 6, 2004, pp. 89–94.
8. Kitchenham, Barbara, et al. "Systematic Literature Reviews in Software Engineering – A Systematic Literature Review." Information and Software Technology, vol. 51, no. 1, 2009, pp. 7–15.
9. Babar, M. A. , & He, Z. . (2009). Systematic literature reviews in software engineering: Preliminary results from interviews with researchers. Proceedings of the Third International Symposium on Empirical Software Engineering and Measurement, ESEM 2009, October 15–16, 2009, Lake Buena Vista, Florida, USA. IEEE.
10. Technology, I. S. , Editor-In-Chief, C. W. , & Contributor, R. P. G. . (2013). Systematic literature reviews in software engineering. Information and Software Technology, 55(6), 919–920.