International Journal of
Advances in Scientific Research and Engineering (ijasre)
E-ISSN : 2454-8006
DOI: http://dx.doi.org/10.7324/IJASRE.2017.32524
Vol.3 (10) November-2017

Machine Learning in Chemical Industry

Matthew N. O. Sadiku[1], Sarhan M. Musa[1], and Osama M. Musa[2]

[1]Roy G. Perry College of Engineering
Prairie View A&M University
Prairie View, TX 77446
[2]Ashland Inc.
Bridgewater, NJ 08807

ABSTRACT

Machine learning (ML) is the scientific discipline dealing with the ways in which machines learn from experience. ML algorithms learn a desired input-output relation from examples in order to interpret new inputs. This is important for tasks such as drug design, industrial process, and manufacturing, with growing applications in the chemical industry. In this paper, we provide a brief introduction of machine learning in the chemical industry.

Key Words: Machine Learning, Chemical Industry.

1. INTRODUCTION

The Internet has transformed the way we live and do business. In the process, it has generated many petabytes of data. Now machine learning is revolutionizing our society again by transforming that data into useful predictions. IBM Watson is a typical example of a modern machine learning application [1].

The chemical industry is facing a great challenge: delivering profit in a hypercompetitive, global market. In view of this challenge, some chemical companies have strengthened their position through mergers and acquisitions. Given the unsettled conditions in the chemical industry, uncertainty is the new normal. The chemical industry is anticipating unprecedented opportunities by embracing maturing machine learning (ML) technology. Using ML can offer chemical industry opportunities to do more with less people and lower the cost of running the business.

Machine learning is an algorithm that estimates an unknown dependency between system inputs and its outputs from the available data. Machine learning is used ad placement, credit scoring, fraud detection, stock trading, drug design, health care, data mining, natural language processing, image recognition, expert systems, simulation, manufacturing, and many other applications [2].

Machine learning can be supervised or unsupervised. Supervised ML approaches can be used to develop empirical models that accurately classify the toxicological endpoints from large-scale data sets. Supervised algorithms come in two types: regression and classification. Unsupervised algorithms work off input data alone and the system tries to discover the hidden structure of data or associations between variables. It can be useful for tasks such as clustering, compression, feature extraction, etc.

2. MACHINE LEARNING ALGORITHMS

Common machine learning algorithms include artificial neural networks, fuzzy logic, and genetic algorithms [3,4]:

- *Artificial neural networks* (ANN): These are a computing paradigm inspired by the functioning of the human brain. They are ML techniques that consist of computing cells or neurons with massively weighted interconnections. Each neuron performs a simple operation and interacts with each other to make a decision. ANNs arose as an attempt to model brain structure and functioning.

- *Fuzzy logic* (FL): This is a method which uses a set of fuzzy rules defined by the user to map fuzzy inputs to fuzzy outputs. A fuzzy control unit performs three basic processes: fuzzification, rule, and defuzzification.

- *Genetic algorithms* (GA): These are global search and optimization techniques based on the principles of natural evolution and genetics. The algorithms work by repeatedly modifying a population of individual solutions.

Choosing an appropriate ML method for problem solving in practice is largely dictated by the problem and experience. A new development is the introduction of hybrid techniques, according to which multiple machine-learning methods are combined to improve the quality of predictions. One of the major tasks for ML researchers is to construct methods and tools for all combinations of ML tasks and model types.

Machine learning algorithms automatically learn programs from data. Once a ML algorithm is trained it can be easily implemented on a microcontroller. ML algorithms can be implemented in a variety of programming languages and software such as C++, Python, Java, R, Orange, Weka, SMILES, and MATLAB.

3. APPLICATIONS

- *Manufacturing*: This is an area where the application of machine learning is very fruitful. The chemical manufacturing industry today is facing an increasing volume of data which compromise a variety of different formats, semantics, and quality. ML techniques have been successfully utilized in various process optimization, monitoring, and control applications in manufacturing [5]. ML can help manufacturers find solutions faster. It can lead to a whole new realm of manufacturing that would be otherwise unrealizable. Machine learning is the key to speeding up the pace of sustainable, cost-effective chemical and manufacturing innovation.

- *Drug design:* Machine learning techniques, such as self-organizing maps, multilayer perceptron, bayesian neural networks, counter-propagation neural network, and support vector machines, are commonly applied as drug design tools during the last decades. This is due to the fact that the drug design is complex and requires the use of hybrid techniques. The number of studies in medicinal chemistry that employ these techniques has grown steadily [6,7].

- *Toxicity prediction*: The prediction of chemical toxicity is a significant challenge in both the environmental and drug development arenas. Machine learning methods, most notably support vector machines (SVM) and artificial neural network (ANN), have been used to analyze data sets in which in vitro bioassay data is being used to predict in vivo chemical toxicology. One begins with chemicals for which toxicology data is available (training chemicals) and develops predictive classification tools [8].

- *Compound classification*: Classifying thousands of compounds manually by medicinal chemists can be a cost-intensive, time-consuming, tedious, and error-prone process. Machine learning techniques such as ANN can be used to automate this process by learning classification models from training compounds of each class [9]. It is important that the input contains all the latent feature information of the chemical compound. Such chemical information include weight, molecular formula, rings, atoms, etc. Other applications of ML in chemical industry include predicting phospholipidosis, process industry, and chemometrics.

4. CHALLENGES

Despite the enthusiasm and popularity that ML has recently gained, some experts of the field have expressed skepticism. This is justifiable given the disappointment with the previous wave of neural networks and other AI techniques.

In most of computation, the two main limited resources are time and memory. In machine learning, there is a third one: training data. Machine learning reports and papers are full of theoretical guarantees. The most common type is a bound on the number of examples needed to ensure good generalization. But one cannot make much of these guarantees. Like any discipline, ML has a lot of “folk wisdom” that can be difficult to come by [10]. Machine learning is not for the faint of heart in terms of quantitative methods. Data preparation is an important skill set. Neural networks, for example, still require preprocessing of inputs and adjustment of parameters by human experts.

5. CONCLUSION

This study has shown that machine learning techniques are promising tools for solving problems in chemical industry. Machine learning has already shown its capacity to learn and analyze, beating our best champions at complex strategy games, such as Go and Jeopardy! However, machine learning is not a silver bullet. Even with self-educating computers, the principle of "garbage in, garbage out" still applies. It is also important to understand that no one ML algorithm is right for all circumstances, and any system for helping chemists must include an armamentarium of techniques [11]. More useful information about ML can be found in Shalev-Shwartz and Ben-David [12], Alpaydm [13], and other books in Amazon.com.

REFERENCES

1. J. H. Lee, J. Shin, and M. J. Realff, "Machine learning: Overview of the recent progresses and implications for the process systems engineering field," *Computers and Chemical Engineering*, to appear soon.
2. M. N. O. Sadiku, S. M. Musa, and O. M. Musa, "Machine Learning," *International Research journal of Advanced Engineering and Science*, vol. 2, no. 4, October 2017, pp. 79-81
3. G. D. Betri et al., "Predicting copper concentrations in acid mine drainage: a comparative analysis of five machine learning techniques," *Environmental Monitoring and Assessment*, vol. 185, no. 5, May 2013, pp. 4171–4182.
4. N. O'Mahony et al., "Machine learning algorithms for process analytical technology," *World Congress on Industrial Control Systems Security*, 2016, pp. 20-26.
5. T. Wuest et al., "Machine learning in manufacturing: advantages, challenges, and applications," *Production & Manufacturing Research*, vol. 4, no. 1, 2016, pp. 23-45.
6. J. C. Gertrudes et al., "Machine learning techniques and drug design," *Current Medicinal Chemistry*, vol. 19, no. 25, September 2012, pp. 4289-4299. https://doi.org/10.2174/092986712802884259
7. A. Lavecchia, "Machine-learning approaches in drug discovery: methods and applications," *Drug Discovery Today*, vol. 20, no. 3, March 2015, pp. 318-331.
8. R. Judson et al., "A comparison of machine learning algorithms for chemical toxicity classification using a simulated multi-scale data model," *BMC Bioinformatics*, vol. 9, 2008, https://doi.org/10.1186/1471-2105-9-241
9. T. Lang et al., "Feasibility of active machine learning for multiclass compound classification," *Journal of Chemical Information and Modeling*, vol. 56, 2016, pp. 12-20.
10. P. Domingos, "Tapping into the 'folk knowledge' needed to advance machine learning applications," *Communications of the ACM*, vol. 55, no. 10, October 2012, pp. 78-87.
11. E. D. Salin and P. H. Winston, "Machine learning and artificial intelligence: an introduction," *Analytical Chemistry*, vol. 64, no.1, January 1992, pp. 49-60.
12. S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. New York, NY: Cambridge University Press, 2014.
13. E. Alpaydm, *Introduction to Machine Learning*. Cambridge, MA: The MIT Press, 3rd edition, 2014.

AUTHORS

Matthew N.O. Sadiku is a professor in the Department of Electrical and Computer Engineering at Prairie View A&M University, Prairie View, Texas. He is the author of several books and papers. His areas of research interest include computational electromagnetics and computer networks. He is a fellow of IEEE.

Sarhan M. Musa is a professor in the Department of Engineering Technology at Prairie View A&M University, Texas. He has been the director of Prairie View Networking Academy, Texas, since 2004. He is an LTD Sprint and Boeing Welliver Fellow.

Osama M. Musa is currently Vice President and Chief Technology Officer for Ashland Inc. Dr. Musa also serves as a member of the Advisory Board at Manhattan College's Department of Electrical and Computer Engineering as well as a member of the Board of Trustees at Chemists' Club of NYC. Additionally, he sits on the Advisory Board of the International Journal of Humanitarian Technology (IJHT).