Applications of Machine Learning in Environmental Engineering

Catherine Wilcox
Wei Lee Woon
Zeyar Aung

Technical Report DNA #2013-03

December 2013

Data & Network Analytics Research Group (DNA)
Computing and Information Science Program,
Masdar Institute of Science and Technology,
PO Box 54224, Abu Dhabi, UAE.

Applications of machine learning in environmental engineering

Catherine Wilcox[1], Wei Lee Woon[2], Zeyar Aung[2]
[1]Water and Environmental Engineering &
[2]Computing and Information Science programs
Masdar Institute of Science and Technology
Abu Dhabi, UAE
Email: {cwilcox,wwoon,zaung}@masdar.ac.ae

***Abstract*—Environmental engineering is the study of interactions between humans and the surrounding environment. In recent times this endeavor has become increasingly data intensive, a challenge that has been met by rapid progress in the development of computational techniques for handling large data sets. Such techniques are collectively known as machine learning. In this report, a survey of recent applications of machine learning in environmental engineering is presented. In particular, the focus is on the topics of water resources management, flood prediction, rainfall-runoff modeling, wastewater treatment and environmental assessment. Approaches covered include support vector machines (SVM), neural networks, the genetic algorithm and k-means. The relative novelty of machine learning in this field means that many additional applications can be developed in the future, making this a rich and exciting area of research.**

I. Introduction

The discpline of Machine Learning focuses on the use of specialized algorithms for extracting patterns and information from large and complex data sets. The falling costs of data storage and processing power have seen the fields of data mining and machine learning deployed in an increasingly broad range of applications [1]. One such area is Environmental Engineering, broadly defined as the study of human-nature interactions from the perspective of science and engineering. Of particular interest are the areas of drinking water and wastewater treatment, air pollution, environmental impact assessment, water resources management, and hazardous waste.

The field of environmental engineering often demands large amounts of data collection. Types of data collected include records of rainfall, streamflow, temperature, and concentration of pollutants. Data may be collected over various time intervals, including instantaneous, hourly, daily, or annually. Commonly, data is reported in daily amounts.

Although care is taken in data collection, the information often suffers from incompleteness, measuring errors (from the human or the measuring device itself), and discrepencies in data publishing. For instance, daily quantities could be an average of many instantaneous values, a sum of all values over the course of the day, or one instantaneous value collected at some point during the day. This reduces the clarity of the data and may not provide the interpreter with a complete picture of environmental phenomenon. [2]

The result of environmental engineering fieldwork is the generation of databases which feature tens of thousands of data points. This vast quantity of data would be night unto impossible for human beings to analyze via traditional environmental engineering methods; to say the least, the process would be highly time consuming. In recent years researchers have begun to apply machine learning techniques to such databases in order to more efficiently generate new understandings of environmental phenomenon, and to aid in environmental design and decision making. The size of the databases, coupled with robust data mining algorithms, have the potential of leading to rapid advances in environmental engineering. Machine learning algorithms also have the potential to overcome the challenges presented by data quality issues in the field. [1]

The objective of this paper is to provide an overview of machine learning approches within the discipline of environmental engineering. In Section II the relevant algorithms are presented and discussed. Methods described include genetic programming, neural networks, support vector machine, and k-means. Case studies which provide representative applications of these methods are then presented in Section III.

II. Machine learning algorithms

In this section several machine learning algorithms are explained. All of these have been utilized in the field of environmental engineering, as will be discussed in the next section.

These methods are fairly diverse and provide a wide range of capabilities. To organize this discussion, we divide them into the following three classes:

1) *Supervised learning* - these are methods where a given set of independent variables are to be matched to one or more dependent variables. The aim is to learn this matching and this is most often done using "training data", i.e. data for which the correct dependent variables are already known. The learning algorithms then adjust the parameters of a particular model such that the outputs of the system match the known outputs as closely as possible.
2) *Unsupervised learning* - in contrast, with unsupervised methods there is no prior "correct" data. In most cases there isn't even a pre-specified dependent variable as the

aim is simply to extract underlying structure and patterns in the data.

3) *Optimization* - techniques for finding the optimal set of parameters which minimize a pre-defined cost function.

A. Supervised Learning Algorithms

1) Support Vector Machine: Support vector machines (SVMs) are a set of supervised learning methods for classification. It seeks to draw lines through the location with the largest gap between data points. If necessary, it first transforms the data into a more separable form, usually into higher-dimensional space. The method then establishes vectors which divide the points. The results are binomial classifications for the series of data points. Although limited to binomial situations (without enhancement), SVM is a robust and powerful method of distinguishing data qualities. [3]

2) Decision Trees: Decision trees are a classification method which is readily interpretable by the user. It consists of splitting data sets at nodes, with each point assigned to a branch of the node based on some criteria or probability. The tree continues to split until either all points are classified or some stop criterion has been met. Unlike other methods such as neural networks which may be considered a ”black box,” the human interpreter can easily determine how classifications were made in decision trees. [2]

3) Artificial Neural Networks: Artificial neural networks (ANN) are models inspired by the workings of the central nervous system. It consists of an artificial network of interconnected neurons. In most instances, ANN functions by adapting its structure based on internal or external information which goes through the system during the learning phase of development. [4] ANN approaches are predominantly non-linear statistical or signal processing-based modeling tools which are adept at finding complex relations and patterns in the data. [5] ANNs are commonly referred to as a ”black box” models, meaning that most of the machine learning is not readily interpreted by the recipient of the results. ANN's ability to process complex non-linear systems indicate much promise for its appplication in new fields. [6]

4) Case-based reasoning: Case-based reasoning (CBR) is the broad process of solving problems in new situations based on its similarity to previous situations. Such an algorithm begins by first retrieving the most similar case(s) in the database to the case it hands. It then decides to either reuse (if the case is similar enough) or adapt (if there are significant discrepencies) the information in the selected case in order to solve the new case. The proposed solution is then evaluated, either via a simulated model or via input from a human agent. Finally, the information learned from solving the new case is added to the database of cases, as either a success or an example of failure of the solution does not work. [7]

B. Unsupervised Learning Algorithms

1) k-means: The k-means algorithm comes from a class of unsupervised clustering methods. The goal of the algorithm is to identify the weighted centers (means) of ”k” number of clusters, and assign labels to data points or vectors according to an evaluated distance from the mean. It is performed by first initializing k center points at random, assigning each data point to the nearest cluster mean, calculating the mean position of the newly formed cluster, reassigning the mean to that point, and then repeating the process. K-means continues until convergence is reached. [2]

2) Association rules: Association rule mining is a well-developed technique for machine learning. It consists of the identification of patterns and associations which could guide classification. For instance, in an online marketplace database, this method could discover that people who purchased a certain item also purchased the same other item with a high rate of consistency. These generated association rules are highly useful in making predictions and correlations. [2]

C. Optimization Algorithms

1) Genetic Algorithm: The genetic algorithm (GA) is based on principles of natural evolution and Darwinian science. It is particularly useful in search and optimization problems. It functions by first randomly generating a ”parent generation” of data points or parameters. The parent generation recombines and produces ”offspring” whose features consist of mixtures of the parent generation's features, as occurs in biological reproduction. To introduce more variability, many GA algorithms incorporate some sort of ”mutation” or ”crossover” of the features. Mutation occurs when one random piece of formula is replaced by another randomly generated piece of formulal. Crossover is when characteristics or a piece of formula from one place is replaced by formula from another location. These processes allow for increased possible combinations, or ”genetic diversity,” in the offspring generation. [1]

After reproduction, the offspring are tested for their ”fitness” or performance according to the task's end criteria. The algorithm may, for example, only permit the most fit offspring to further reproduce in the next iteration. The iterations usually cease after a certain number of generations, or after some stop criterion is met. The generation with the highest performance is presented as the model result. An overview of GA can be found in Figure 1. As seen in Figure 2, in which the GA was tested on whether it could discover the well-proved Bernoulli equation, the error in the results decreases with each generation. [8]

D. Learning apprentice systems

In addition to the three classes mention, the involvement of an “expert” may be used to augment the efficacy of any given algorithm. Machine learning generally occurs with the computer performing its classification or analysis task without the regular guidance of a human being. In these cases it may assume that uncorrelated items are actually unrelated, or otherwise create results that contradict the reality of the situation.

The expert can interact in one of two modes: proactive or reactive. In the reactive mode, the expert corrects incomplete or incorrect rules by providing correct examples to be used in

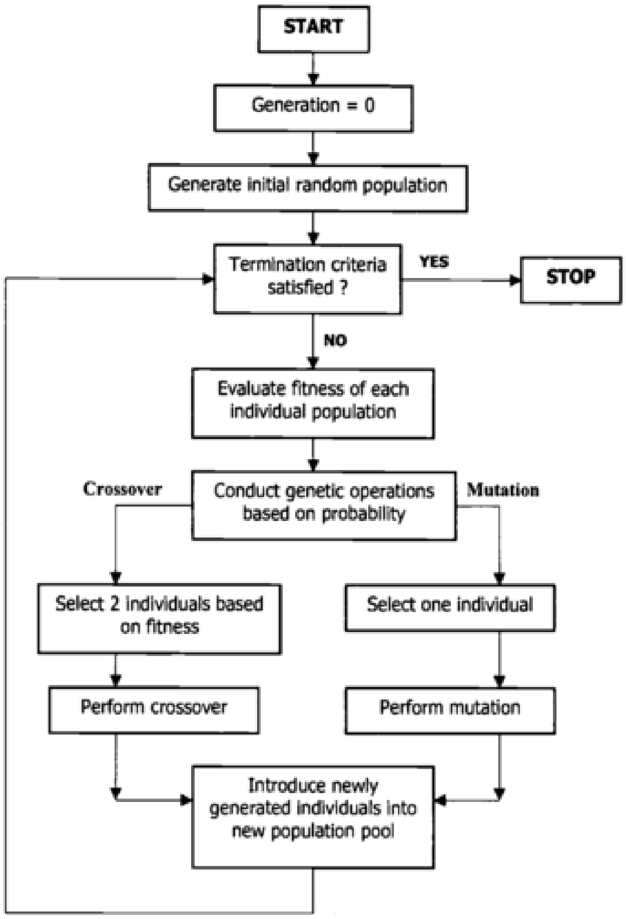

Fig. 1. Flowchart for the genetic algorithm

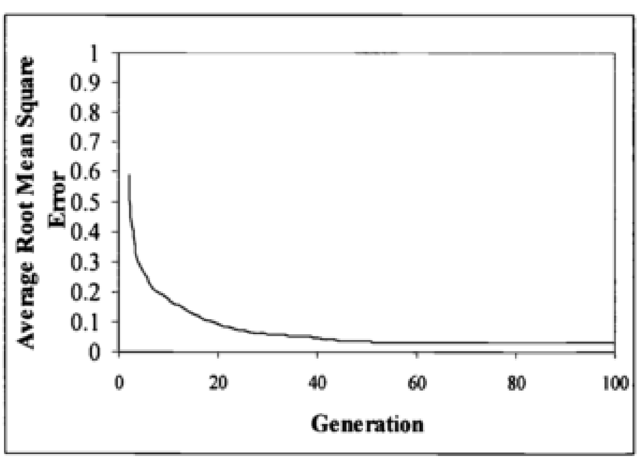

Fig. 2. Rapid convergence of the genetic algorithm when performing a test search for the Bernoulli equation

the next iteration. This prevents those examples and similar data points from being further misclassified. In the proactive mode, the expert instead provides correct examples or guidance relations before the machine learning begins. This prevents mistakes preemptively. Depending on the weight of each expert-defined rule, the system is guided to a more ideal set of results. [9]

III. Case studies of Machine learning in Environmental Engineering

The following sections highlight specific examples of algorithms applied to environmental engineering situations.

A. *Wastewater Treatment*

Wastewater Treatment Plants (WWTP) are necessary for environmental protection, buffering nature from human and industrial waste. However, WWTPs consist of a wide variety of operations, including mechanical, electrical, chemical, biological, and physical. If any one of these fail, the plant may not clean the water to sufficient levels, which risks discharging polluted water into the environment.

One method of evaluating, preventing, and solving these failures is by using CBR. In one study, researchers set up system to handle continuous domains, where the knowledge base is readily increased as more problems are solved. Such problems are traditionally handled via a mathematical approach, but this does not incorporate years of expert experience, case studies, or incomplete yet potentially useful information.

For each new WWTP problem, the CBR system goes through a case library of previous problems and selects cases with the most relevant parameters. If the retrieved case is ”close enough” to the case at hand, only a few small parameter adjustments are required in order to reach a solution. If the cases do not display enough similarity, linear interpolation is performed between similar past cases in order to obtain a more similar representations.

The CBR system was also trained to learn from its own results. If the proposed solution was a success, then the case is added to the case library for comparison with future WWTP issues.The study showed that CBR performance rapidly improved as its case library increased. [7]

B. *Environmental Assessment*

An approach based on the integration of learning apprentice techniques was applied to the discipline of environmental assessment. In many nations, laws have been enacted to ensure environmental protection. Often times this calls for an environmental impact assessments, an in-depth evaluation of the effects a certain human action may have on the environment. This task is subjective and demands human interpretation; however, at the same time a wealth of previous environmental assessments could provide assistance and guidance in carrying out new evaluations.

One such study sought to use a learning apprentice system which allowed for the input of an expert and use of a weak and flexible domain theory during the process of mining previous

Table 3. Accuracy of impact significance predictions

Training set size	Test set size	Predictive accuracy Total	90% Conf. Int.	Number of conditions	Number of rules
6 (26%)	17 (74%)	12 (71%)	48–88%	5	5
12 (52%)	11 (48%)	9 (82%)	53–97%	7	6
17 (74%)	6 (26%)	6 (100%)	61–100%	9	8

Table 4. Accuracy of impact duration predictions

Training set size	Test set size	Predictive accuracy Total	90% Conf. Int.	Number of conditions	Number of rules
6 (26%)	17 (74%)	6 (35%)	17–58%	12	12
12 (52%)	11 (48%)	3 (27%)	8–56%	16	16
17 (74%)	6 (26%)	1 (17%)	1–58%	20	21

Fig. 3. Results from the learning apprentice system approach to environmental assessment

environmental assessments to induce rules. The induction algorithm works by calculating the probability of occurrence of a certain class given a certain attribute. The label with the max probability is assigned to the point, and a subset of the whole dataset is created with this new information. The process is repeated until the probability of that label is equal to one. Those instances covered by the induced rule are moved from the training set, and the process is repeated until all examples have been removed.

An expert is called on to improve this process via the proactive development of guidance relationships. These relationships include such rules as ”include point in class,” ”exclude,” or ”free” (allowing the machine to determine the relation). The strength of each rule is determined; for example, a legal and thus mandatory item would always be classed as ”strong” in importance, whereas other less significant attributes may be defined as ”weak” and thus guide the results less. The expert also has the option of performing in a reactive mode, correcting false positives and introducing more accurate examples to a class.

Results (shown in Figure 3) were mixed on the effectiveness of this method. Although it has promise for the utilization of past assessments in future ones - as past assessments are often the best guideline for present decision-making - the researchers in this case did not recommend the use of the learning apprentice technique for relatively inexperienced environmental practitioners. The method still requires the hand of an expert in order to produce high-quality results about environmental decision-making. [9]

C. Water resources management

1) Reservoir decision making: GA was utilized in a water quality conflict resolution model for a reservoir management system. In particular, a varying chromosome length GA (VLGA) was used. The system model incorporated both physical parameters, such as pollutants, water demand, and water availability, in addition to human factors which influence decision making. Results determined that the use of GA optimization improved operational strategies in this system. [10]

Statistical measures of SVM and RVM ability to predict GSL volumes

Statistics	RVM Training	RVM Testing	RVM Confidence limits (95%) Lower	RVM Confidence limits (95%) Upper	SVM Training	SVM Testing	SVM Confidence limits (95%) Lower	SVM Confidence limits (95%) Upper
Correlation coefficient	0.990	0.982	0.973	0.993	0.989	0.982	0.973	0.992
Coefficient of efficiency	0.979	0.965	0.955	0.975	0.977	0.924	0.915	0.934
Bias	0.000	0.451	0.441	0.461	−0.035	0.809	0.799	0.819
RMSE	0.749	0.869	0.859	0.879	0.785	1.290	1.280	1.300
Mean absolute error	0.567	0.696	0.686	0.705	0.595	1.052	1.042	1.061
Index of agreement	0.995	0.991	0.982	1.000	0.994	0.978	0.969	0.988

Fig. 4. Statistical measures of SVM and RVM ability to predict GSL volumes.

2) Lake volume prediction: One study sought to apply SVM techniques to the Great Salt Lake drainage basin. The researchers were especially interested in SVM's ability to function in sparse, chaotic systems. The Great Salt Lake, like most hydrologic systems, are highly complex and variable. However, chaos theory states that such if one assumes it is a deterministic chaotic system, it will behave in a similar (albeit chaotic) manner as it has in the past. If one can generate a full knowledge of the system with proper estimates of the time delay and state-space scenario, one can predict future behavior. In order for this to be effective the time series in question must be sampled at sufficient resolution, over a sufficient period of time, and without excessive noise corruption. However, the analyzer must be aware that even solutions from high-quality data and estimations may be unstable and dynamic due to the chaos itself. If the underlying system is not identified it further reduces the effectiveness of the analysis. [3]

The researchers sought to increase the robustness of the state-space reconstronction method by combining chaos theory with SVM. SVM's cousin algorithm, the relevance vector machine (RVM), was also applied in this study for comparison, and functions similarly to SVM except with a Bayesian probabilistic separation instead of a firm linear separation. The goal of the study was to develop techniques for identifying model parameters, with the example scenario of predicting biweekly volumes in the Great Salt Lake.

The Great Salt Lake receives water from three sources: three inflowing rivers (about 66% of contributed volume), direct precipitation (about 31%), and groundwater recharge (about 3%). Data on lake volume has been collected since 1843, and the historic record shows a series of dramatic rises and falls. What is most interesting and challenging is that the lake volume does not directly correlate to nearby rainfall and streamflow time series according to traditional hydrologic analysis. The lack of an ability to predict lake levels has led to sudden and great expenses and difficulties for Salt Lake City during unexpected drought and flooding periods.

In the study the researchers reproject the lake volume data into multidimensional state-space in order to develop forecasts of different lead times. SVMs and RVMs are then applied to the state-space reconstruction. They evaluated the success of their results via the root mean squared error (RMSE), mean absolute error, and the index of agreement in order to present a holistic view on the variance of the results.

Results of the study are presented in Figure 4. They show that both SVM and RVM provide accurate precition results which could be utilized in developing water resources management strategies for the Great Salt lake. They found that both models behaved well in the sparse and chaotic system. SVM tended to minimize structural risk in reaching a sparse solution, whereas RVM tended to better capture examples of the underlying distribution. [3]

D. Flood frequency analysis and flow characterization

ANN was applied to a set of hydrologic basins in Quebec, Canada with the aim of regionalizing the conclusions to nearby ungauged but comparable sites. The method was developed by making a canonical correlation analysis (CCA) model of the relevant physiographical and meteorological characteristics. As this information is typically readily available, one could presumedly expand models at sites with similar characteristics but also with hydrological data collection sites to locations which have no active measurement system.

The study compared a variety of methods under the same task, including ANN alone, ANN ensemble, and each of these methdos combined with a CCA space projection of the data. Results showed that while the ANN ensemble performed better than the standard ANN, both produced better results when applied to the CCA-space projection of site variables. However, all of the methods underestimated flood magnitudes at ungauged sites with very high specific quantiles. [11]

1) Peak classification: A clustering algorithm such as k-means may be used in order to analyze and classify streamflow peaks. This classification problem presents a few challenges: although the peaks may have comparable shape, they may be slightly out of phase, different in magnitude, or start from different base levels. The algorithm can eliminate these discrepencies in order to discover similar flow behavior by normalizing the mean and other variables to zero for each peak, thus allowing for clustering under equivalent conditions. [2]

This method could similarly be applied to digital elevation models (DEM) with the aim of identifying regions with similar geomorphological features. This would allow for the potential application of models in one region to regions with similar features. The same benefit could be reached by clustering entire time series to one another, thus enhancing regionalization analysis potential. [2]

E. Rainfall Analysis

Rainfall data may be processed by machine learning in a variety of methods. One such method is via a decision tree. If the user wishes to identify only ”intense” rainfall periods for analysis, they could establish a decision tree algorithm under certain criteria. The resulting output could indicate the presense of an underlying phenomenon depending on how they are clustered, such as in summer storms. The effectiveness of the algorithm could be further improved with higher resolution than daily data, but this data is usually difficult to come by with rainfall and expensive to collect. [2]

MODEL STRUCTURE		**TRAINING/CALIBRATION PHASE**			
		COE R^2	**MSE (cumecs)**	**MAE (cumecs)**	**MRE (%)**
	*18-20-1	0.745	0.965	0.716	49.20
MLP	*18-20-4-1	0.677	0.961	0.719	49.07
MLR	15 input	0.604	0.910	0.677	53.37
	18 input nodes	0.936	0.927	0.654	47.73
RBF	15 input nodes	0.902	0.884	0.627	54.54
HEC-HMS		0.483	3.421	2.564	98.65
MODEL STRUCTURE		**TESTING/VERIFICATION PHASE**			
		COE R^2	**MSE (cumecs)**	**MAE (cumecs)**	**MRE (%)**
MLP	*18-20-1	0.723	1.164	0.860	53.04
	*18-20-4-1	0.674	1.126	0.840	52.12
MLR	15 input	0.587	0.963	0.748	55.61
RBF	18 input nodes	0.811	1.260	0.930	58.45
	15 input nodes	0.782	1.122	0.848	65.28
HEC-HMS		0.037	4.822	3.657	256.87

Fig. 5. Comparative results of a study of ANN models for rainfall-runoff simulation. Evaluation measures include the coefficient of efficiency (R2), mean square error (MSE), mean absolute error (MAE), and mean relative error (MRE)

F. Rainfall-runoff modelling

The prediction of runoff based on the quantity of precipitation received in an area has long been an aim of environmental researchers. However, many studies have determined that the relationship between rainfall and runoff is non-linear and complex. The assistance of certain machine learning which excel in such an environment can help to overcome these challenges - notablyl, ANN and GA. Association also has potential for a more general application in this field.

1) ANN: Researchers applied ANN to a semi-developed catchment area near Selangor, Malaysia in order to develop a rainfall-runoff model for the basin. They utilized two different versions of ANN: The multilayer perceptron (MLP) model, and the radial basis function (RBF) model. They further compared the ANN models to rainfall-runoff models based on multiple linear regression and HEC-HMS, a standard hydrologic simulation tool. They evaluated the success of each model using evaluated using the coefficient of efficiency (R2), mean square error (MSE), mean absolute error (MAE), and mean relative error (MRE). Results (Figure 5) showed that that the neural network model performed better than the HEC-HMS model. The RBF ANN had better performance than the MLP ANN. The number of hidden layer neurons significantly influenced results from the ANN models; having too few would under fit the data and may not lead to high levels of accuracy, while too many nodes meant overfitting and excessive training time. [6]

2) GA: In addition, GA has been successfully applied in the area of rainfall-runoff modeling as a sort of symbolic regression. The algorithm leads a search for both the optimum model structure and optimum coefficients. The user first defines basic ”building blocks” for the end model, and uses GA to determine which combination produces the best results. [1]

	Average RMSE						
Lead-Time (hours)	Storm S11	Storm S12	Storm S13	Storm S14	Storm S15	Storm S16	Average for Six Storms
1	0.248	0.163	0.188	0.679	0.635	0.596	0.418
(NAM)	(0.878)	(2.675)	(1.810)	(4.515)	(3.199)	(1.657)	(2.456)
2	0.561	0.409	0.377	1.624	1.190	0.960	0.853
3	0.847	0.679	0.607	2.609	1.546	1.274	1.260
4	1.046	0.943	0.838	3.557	1.749	1.554	1.614
5	1.116	1.190	1.072	4.373	1.798	1.746	1.882
6	1.075	1.415	1.270	4.991	1.821	1.874	2.074
7	1.024	1.617	1.440	5.274	1.971	1.903	2.205
8	0.994	1.789	1.573	5.331	2.251	1.805	2.290
9	0.991	1.923	1.686	5.641	2.504	1.656	2.400

Fig. 6. Root mean squared error of testing storms for different prediction lead times in realtime rainfall-runoff predictions with GA

One study utilized genetic symbolic regression (GSR) to improve real-time rainfall-runoff predictions. It compared a basic rainfall-runoff method (NAM) to the same method with the addition of a GSR error correction/updating scheme. The GSR component of the model was calibrated by a set of ten storm events, all with a relatively high flow (for this basin) of 4 m3/s. The verification storms, however, ranged from 10-29 m3/s.

Results (Figure 6) show that up to a four-hour lead time, the GSR version of the model performed better than the NAM version alone. The study concluded that although the GSR updating method performed relatively well for a lead time of up to nine hours, for optimal results the lead time should be less than or equal to the time of concentration for the basin. [8]

3) Association rules: Association rules could also be utilized in rainfall-funoff analysis by generally grouping certain rainfall patterns with certain peak streamflow responses. For example, the algorithm may discover that short but intense periods of rainfall generally lead to a particular peak size and shape. As always, the success of this analysis depends on the quantity and quality of data; if only few examples exist of certain rainfall or streamflow patterns, it will be difficult for the machine to identify any underlying rules. [2]

IV. Conclusions and Future Work

Machine learning has great potential in the field of environmental engineering. It can build both upon and beyond traditional mathematical and statestical models used in the field. Due to the large quantities of data collected in this discipline, machine learning can lead to additional insight on the structure and relations of the data points moreso than would be feasible by the alone. This leads to enhanced understanding of the environment and the engineered systems which guide human relations with the environment.

The examples discussed in this paper show that certain algorithms have particularly great potential. GA and ANN especially display the computational strength to overcome the challenges of modeling the non-linear, chaotic natural environment. Algorithms like GA and CBR can even be applied to decision-making scenarios. This is especially true when enhanced by a learning apprentice system.

As the use of machine learning and data mining in environmental engineering is still relatively new, many new applications remain. In particular, the algorithms have mainly been applied under more typical conditions. Certain challenges such as modeling extreme situations could present an opportunity for advancement in the field via machine learning, as modeling with traditional techniques infers great difficulties. Of particular interest is the modeling of streamflow and rainfall-runoff relations in desert environments, where data points are particularly sparse and varied.

References

[1] V. Babovic, "Data mining in hydrology," *Hydrological processes*, vol. 19, no. 7, pp. 1511–1515, 2005.

[2] J. Spate, B. Croke, and A. Jakeman, "Data mining in hydrology," in *MODSIM 2003: Proceedings of the 2003 International Congress on Modelling and Simulation*, 2003, pp. 422–427.

[3] A. Khalil, M. McKee, M. Kemblowski, T. Asefa, and L. Bastidas, "Multiobjective analysis of chaotic dynamic systems with sparse learning machines," *Advances in water resources*, vol. 29, no. 1, pp. 72–88, 2006.

[4] T. Ouarda and C. Shu, "Regional low-flow frequency analysis using single and ensemble artificial neural networks," *Water Resources Research*, vol. 45, no. 11, p. W11428, 2009.

[5] C. Shu and T. Ouarda, "Regional ood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system," *Journal of Hydrology*, vol. 349, pp. 31–43, 2008.

[6] S. Harun, N. Ahmat Nor, M. Kassim, and A. Hashim, "Artificial neural network model for rainfall-runoff relationship," *Jurnal Teknologi B*, no. 37B, pp. 1–12, 2002.

[7] M. Sanchez-Marre, U. Cortés, I. Roda, and M. Poch, "Sustainable case learning for continuous domains," *Environmental Modelling & Software*, vol. 14, no. 5, pp. 349–357, 1999.

[8] S. Khu, S. Liong, V. Babovic, H. Madsen, and N. Muttil, "Genetic programming and its application in real-time runoff forecasting1," *JAWRA Journal of the American Water Resources Association*, vol. 37, no. 2, pp. 439–451, 2001.

[9] B. Julien and S. Fenves, "An environmental evaluation learning apprentice system," *Engineering Applications of Artificial Intelligence*, vol. 9, no. 6, pp. 589–599, 1996.

[10] R. Kerachian and M. Karamouz, "A stochastic conict resolution model for water quality management in reservoirriver systems," *Water Resources*, vol. 30, pp. 866–882, 2007.

[11] C. Shu and T. Ouarda, "Flood frequency analysis at ungauged sites using artificial neural networks in canonical correlation analysis physiographic space," *Water Resources Research*, vol. 43, no. 7, p. W07438, 2007.