Machine Learning on Minimizing Irrigation Water for Lawns

Weiqing Gu\textsuperscript{1}, Zhaocheng Yi\textsuperscript{2}

\textsuperscript{1} Professor of Mathematics and Director of the Harvey Mudd College Mathematics Clinic, Harvey Mudd College, 301 Platt Boulevard, Claremont, CA 91711, USA e-mail: gu@g.hmc.edu
\textsuperscript{2} Pitzer College, 1050 N Mills Ave, Claremont, CA 91711, USA e-mail: zhaoyi@students.pitzer.edu

Cite as: Gu, W., Yi, Z., Machine Learning on Minimizing Irrigation Water for Lawns. J. sustain. dev. energy water environ. syst., DOI: https://doi.org/10.13044/j.sdewes.d7.0304

ABSTRACT

Conserving water has always been very important for California especially during drought seasons, due to the fact that California geographically consists most of the hot and dry deserts. The goal of this project is to create an automatic recommendation irrigation system for the purpose of minimizing water use but keeping lawn grass still green. In this project, a mathematical model has been developed and big data analytic techniques have been applied to achieve the goal using the lawns of Harvey Mudd College as an example. The major results of this project include a smart irrigation algorithm by taking weather conditions and human's induced irrigation patterns and an application to notify the user the irrigation rate calculated automatically by the newly developed algorithm. The method here is scalable to different lawns, whether owned by an individual or by an organization as long as there is historical irrigation data available. In conclusion, this method saves money, minimizes water pollution, and preserves water resources especially for drought regions in the world.

KEYWORDS

Energy saving, Smart irrigation system, machine learning, data analytics, emerging needs, saving water, App

INTRODUCTION

Water sustainability is always an important issue for California, USA. California had been experiencing severe drought since 2012. The situation had become even worse as Governor Jerry Brown proclaimed a drought state of emergency on January 17\textsuperscript{th}, 2014, when California was facing water shortfall in the driest years in the recorded state history [1]. Though currently Governor Jerry Brown officially has ended the drought state of emergency on April 7\textsuperscript{th}, 2017, water is still the most precious resource for California State. Saving water is still one of the most challenging tasks.

The goal of this research is to meet the above challenge of conserving water specifically focusing on smart lawn irrigation by using data-to-decision machine learning (ML) techniques. This fills the gap of the current ML methods mostly for smart farms using farm-specific data and control systems, which are not familiar or suitable for ordinary lawn owners or managers. The approach presented here is practical and effective since it only uses online data rather than any advanced data by technical machines or sensors. And it takes advantage of the hypothesis that most lawn owners have mobile phones to receive the irrigation rates predicted by the proposed algorithms instead of any sophisticated control system. This research only assumes that there is historical irrigation data available. Then, an algorithm was developed and implemented to collect online weather-related data since the analysis based on ML showed that the Harvey Mudd College
(HMC) weather station data is highly correlated to the online weather data. The fundamental reason why this proposed method works is that key features were identified, parameters were filtered and then ML was applied to achieve the goal of minimizing irrigation water while still keeping the lawn green. The results of this paper yield a smart irrigation algorithm by taking online weather data and human’s induced irrigation patterns. An Application (APP) was also developed to notify the irrigation rate calculated by this algorithm to users automatically.

Many existing studies in the literature have examined different components of irrigation. Traditionally, irrigation was calculated based on the evapotranspiration (ET) formula. As suggested by the word evapotranspiration itself, it is a sum of evaporation and plant transpiration, counting the irrigation water needed for any plant. Earlier researchers, Monteith and Penman developed an equation to approximate the net ET using parameters such as water volume, air density etc. [2,3]. Zotarelli conducted a step-by-step calculation of ET to determine the water loss and the amount of irrigation needed for plants [4], since ET was believed to be essential in determining the irrigation rates. It was newly found that in fact there is a high correlation between ET and temperature by this study.

Against many existing methods of calculation of ET rates including the estimation of crop ET using satellite remote sensing-based vegetation index [5], the product of this research avoids using evapotranspiration, but only temperature data. This discovery based on data analytics is helpful to many users, at least to those who are non-farmers, to save time, effort and resources to collect ET data or employ software to calculate ET rates.

Another component of irrigation focuses on precision irrigation system development. Sadler et al. suggested the idea that precision irrigation would significantly help with the water shortage especially in the drought seasons [6]. Also, it was demonstrated that precision irrigation is technically feasible but economically costly. During the same time, there were researchers following the precision irrigation concept. For example, Unami, Kawachi and Yangyuoru examined the optimal storage for an irrigation tank system using traditional optimization methods [7]. It was presented in the framework of variational calculus by its minimum principle for an optimal management strategy in small-scale tank irrigation systems. Instead of using the minimum principle for optimal decision using control theory, this study is heavily based on data analytics and ML.

Next, the developments of wireless sensor networks, smart agriculture and online communication tools have led many irrigation managers and farmers to collect relevant data. Often those data were collected by certain paid hardware or software for the owners or farmers. For example, HMC had been annually paying a software company to collect various irrigation data including temperature, soil moisture, humidity, wind speed, and evapotranspiration for the sole purpose of irrigating the HMC lawns [8]. For farmers, the usual purpose of collecting data are mainly for monitoring and controlling crops, such as recording water consumption, identifying, and monitoring soil moisture levels or disease symptoms. For examples, certain decision support system was employed for irrigation scheme management [9]; similarly, a centralized remote-control irrigation system was developed for irrigating large areas of land with a distance of tens of kilometers from the nodes to the control center [10]; a site-specific irrigation control system was built by using soil water potential measurements to control the amount of water applied to each specific zone of a field [11]; relationships between cropping sequences and irrigation frequency under self-propelled irrigation systems in the northern great plains were investigated [12]; remote sensing and control of an irrigation system using a distributed wireless sensor network were developed [13]. Even though many researchers collected large data from different sources, the data sets were often analyzed separately, and the potential of information extracted from existing data is not fully integrated or implemented [14].

In contrast, this paper presents that no data collection effort is needed for an individual lawn owner or an institutional irrigation manager since the results obtained shows that there are strong
correlations between online weather data and the collected data. In this context, the product of this research becomes realistic to use for many individuals who would not be able to afford to use various machines or sensors to collect data or who would not have time to collect the variety of data by themselves.

Recently, the concept of remote irrigation using Internet of Things (IoT) becomes increasingly prevalent. Bo and Wang focused on the combination using cloud computing and IoT on the field of agriculture and forestry [15]. In addition to collecting data, researchers started to employ ML and modern optimization methods for irrigation automation mostly for smart farms, including but not limited to online decision support for irrigation of farmer and optimal scheduling [16-18]. Kissoon, Deerpaul and Mungur proposed a smart irrigation and monitoring system for a large farm by employing Microsoft Azure ML to process data received from sensors and weather forecasting venues to recommend farmers the appropriate moment of irrigation [19]. For another example, Liu et al. focused on optimizing water allocation for a single crop by using a genetic algorithm under the constraints including water cost, irrigations, crop costs etc. to resolve under-irrigation or over-irrigation problem [20]. They employed techniques of a decision support system for precision irrigation. These models, however, mostly focus on the smart farming, which is not suitable for lawn irrigation.

In addition, there are other kinds of approaches by using irrigation models and control systems in literature for determining irrigation needs based on plot-specific data. Hedley et al. used ML to predict soil water status and water table depth by studying electromagnetic mapping [21]. Smith and Peng classified soil textural composition using ML on soil-sensor data to extract data features for a deficit irrigation control system [22]. Goldstein et al. investigated an irrigation process that is capable of predicting and recommending for a specific agriculture plant species, Jojoba [23]. They used as many collected data as possible to reveal insights and relationships between different variables including soil, weather and irrigation characteristics, and obtained results by applying and comparing various ML methods to identify the most effective method.

Against the above background, this project focuses on lawn irrigation not for agriculture farms. It also just utilizes mobile communication where ML methods applied automatically without having a working control room for any system. Moreover, this approach selects data types, which is most simple and understandable to ordinary users by applying triage data techniques for achieving the goal of the project. In other words, the proposed techniques uncover various data correlations and patterns so that as few key data features as possible are used as variables; then the features are fed into some modern ML algorithms to predict optimal irrigation rates. The analysis selects temperature over ET as a feature. Online temperature data is much easier to be obtained for lawn owners. The goal is to make this approach to be realistic for many homeowners or organizations without having to collect many types of the data. Moreover, this study presents a complete data-to-decision process by starting from data collection, feature selection, and data analytics, all the way to make an APP to notify irrigation rates to users. This method suggests people to use online data instead of collecting data for saving water while keeping the lawn grass green. This method saves money for an organization to purchase or maintain any software. Moreover, this method is scalable to different lawns whether owned by an individual or by an organization as long as there is historical irrigation data available. The approach presented in this paper is the novel one.

**TECHNICAL APPROACHES**

The major methods include the following data acquisition/process, mathematical modeling, ML, and App creation and result validation.

1. Obtained raw irrigation rate data, which was controlled by the lawn manager from HMC Facility and Management (F&M) Department.
2. Processed the irrigation data by cleaning, appropriately filling missing data, and organizing the data.
3. Wrote Python scripts to get the time corresponding weather data online including air temperatures, evapotranspiration rates, precipitation, and humidity.
4. Modeled mathematical relations between irrigation patterns and the weather data, identified the irrigation patterns and anomalies by following ML methods, including Maximum Likelihood Estimation, K-Mean Cluster, Gaussian Mixture Model, and Principal Component Analysis.
5. Designed predictive methods to use the patterns in the data and predict the next day irrigation rate including the rate for special events of HMC’s needs such as alumni weekends.
6. Developed an application (APP) implementing above data analytics algorithm which can send a Short Message Service (SMS) to automatically notify the owner or lawn manager the exact amount of water for irrigation at early morning each day.
7. Validated the results by comparing the predicted irrigation rate with the rate being set by the lawn manager based on his experience.

DATA ACQUISITION AND PROCESSING

This research was kicked off by obtaining the data from HMC and understand the details of the data, then working with the subject matter experts (SMEs), HMC lawn managers and his supervisor, Senior Director of Facilities and Management to understand the current status of HMC lawn irrigation.

Interview with the subject matter experts

There are six meter-stations for the whole HMC campus. And there are four flow sensors in the main lawn area monitoring how much water flows each day. The SMEs also showed us how the data was organized by clock number, channel number, zone number, and station connection. Detailed notes were taken during the interviews for purpose of developing a mathematical model. Figure 1 next page provides a part of site plan which associates locations of flow and moisture sensors.

Moreover, it was learned from the SMEs that current irrigation system partly depends on underground sensors, which are usually broken by ground animals. Due to the lack of water resources in California, ground rodents such as mice, rabbits, and squirrels bite the underground irrigation pipes for drinking water. They cause a great waste of the water without being noticed by lawn managers. Then, some of the current irrigation system depending on ground sensors no longer works effectively due to the damages of the sensors and underground pipeline leaking.
Figure 1. A part of site plan of HMC associated locations of flow and moisture sensors

For instance, when the lawn manager was on holiday, the sensor indicated the irrigation is on with the rate as he expected. However, one of the most important lawns of HMC turned brown, because the underground sensor and water pipelines were damaged by ground animals. Water leaked to the sands ground, meaning the water did not retain on the surface of the ground and therefore the grass was not irrigated. Figure 2 below demonstrates the condition of the lawn when he came back from his vacation. Therefore, he emphasized that some new irrigation technology, which suggests the correct daily irrigation rate without depending on underground sensors, is highly desirable. In fact, as he pointed out, this new technology will not only save water usage but also save money for the installation and replacement of pipelines and sensors. Further, it helps with the management time of irrigation.

Figure 2. Inefficient irrigation when lawn manager is on holiday
Figure 3. Ground rodent [24]
Figure 4. Damage to an irrigation water pipeline [25]
Data acquisition

There are mainly 3 types of data initially obtained from the SMEs:

- Firstly, the gallons per minute for all lawns on campus, also known as irrigation rate.
- Secondly, runtime and percentage for irrigation of a specific clock, station, and channel.
- Lastly, weather data, which is generated by HMC weather station.

For the first type of data, it has a maximum irrigation rate, which monitors all sprinklers to make sure that the irrigation rate never exceeds the maximum rate for conserving water.

For the second type of data, the effort was collecting each irrigation rate corresponding to particular clock, station, and channel, which are separated by different letters and colors in Figure 5 below. HMC campus has several meter stations for the lawn irrigation.

For the last type of data, the effort was using HMC weather station data in the first place, then using free online historical weather data from California Irrigation Management Information System (CIMIS). CIMIS weather is public and accessible to everyone [26]. The historical weather data time corresponding to irrigation data obtained from HMC SMEs was downloaded. The HMC weather data collected by certain expensive software was not used since the algorithm has shown that the irrigation results would be highly similar to CIMIS weather data. This makes the algorithm does not depend on individual weather station data, also saves the time of the SMEs to maintain certain software to collect HMC weather data. Furthermore, this helps with the ultimate goal of the project, i.e., making the algorithm scalable for any lawns as long as there is available irrigation data.

![Figure 5. HMC lawn meter stations for lawn irrigation](image)

Data processing

Data visualization. After first obtaining the data, a preliminary analysis was conducted including data visualization such as in Figure 6 next page, showing a primary relationship between maximum temperature and maximum irrigation rate in April 2017, where dates are represented by various colors. From the visualization, it showed that there were eight days with no irrigation while there was an anomalous amount of irrigation on April 28, 2017. By checking those corresponding dates, an interesting thing was observed: those eight days were rainy days, and there was an important event on April 30, 2017.
Figure 6. Irrigation and maximum temperature in April 2017

**Data cleaning.** From data visualization, it was observed that the data collected was “dirty”, meaning there were missing and error data, data with noise. Then the data was processed by interpolating missing data and filtering out certain noises.

**Anomaly detection and pattern recognition.** Data Anomaly occurs when extremely high maximum temperature than historical data (Excessive Heat Warning by National Weather Service of the United States), pipelines broken by ground animals, or special events. These causes are all taken into account in the algorithm. To customize with HMC’s needs, irrigation anomalies were investigated due to special events, including HMC Alumni weekends, Clinic and Thesis Presentation, and Graduation Ceremony. For example, in Winter January 2017, the first irrigation in the entire month is the day when Spring 2017 semester started, no irrigation before the beginning of the semester. It might be reasonable to only have one irrigation before the semester started because January did have 17 occurrences of rainfall, but two more times of irrigation that took place at the end of the month demonstrated that irrigation would be necessary when the school started.

Additionally, on April 28, 2017, the irrigation rate was even greater than the normal irrigation during the summer. After looking into HMC calendar, it was noted that April 30, 2017, is the Alumni Weekend, which is considered as one of the most important events hosted by HMC to impress the alumni. Comparing monthly irrigation schedule, it was noticed that irrigation amount greatly exceeded the average monthly amount in May 2017. The reason for the high amount of irrigation was because some construction during that time, and the workers broke the pipelines. Therefore, the algorithms developed in this project were implemented into code to do anomaly detection and removed the anomaly data since it is necessary to find patterns for regular irrigation rates on a monthly basis. For example, the irrigation pattern of May 2017 is shown in Figure 7 next page. In summary, this research designed the algorithm to prepare to automatically calculate irrigation rates including finding patterns based on the weather variables, temperature, and irrigation for determining the appropriate irrigation frequencies and amounts.
MATHEMATICAL MODEL

In order to mathematically model the problem, many factors have to be considered and decided on a model, which will fit the problem well but not overfitting or underfitting.

Initial modeling approach

The internal factors were considered first. For example, in Southern California, there are grass species including bermuda grass, buffalo grass. HMC mainly holds most kikuyu grass and some bermuda grass. They have different irrigation needs. Since one way to save water is changing grass types, this study considered whether it would be possible to change grass for HMC lawns. It turned out it would be very time consuming and expensive to remove all existing grass and install new grass species.

Different geometry of a given lawn is also a possible approach. According to its curvature and relative hill height, design a more efficient irrigation system by smartly setting up the distribution of sprinklers was planned initially. For instance, when there is a relatively high hill existing in the lawn, installation of a sprinkler on the top assists the grass down the hill get the flowing down water without any additional sprinklers. However, this approach would be site dependent. Not only any relative hill height or curvature can be hard to measure in the real-life situation, but also designing a completely new sprinkler system would be needed. The cost was estimated for this approach, and it could be very costly to replace with a whole new sprinkler system; whereas each lawn area has already had sprinklers. Therefore, a decision was made to model the problem without site dependence.

External factors were also evaluated. For instance, in order to develop the mathematical model, weather factors were taken into consideration for irrigation such as patterns of irrigation rates for summer and winter.

Feature extraction

Feature extraction is crucial for developing such a model. A large amount of time was spent to find the good features. For example, through data visualization this study showed that weather variables are essential for determining the irrigation rate, especially temperature and rainfall data. Evapotranspiration (ET) is, in fact, one of the most important characteristics in understanding the relationship between temperature and irrigation. However, this research indicated that temperature...
and evapotranspiration have strong positive correlations as shown in Figure 8 below. Therefore, one of them would suffice as a data feature for determining the irrigation rate. Temperature was chosen since people can feel and observe it directly, so it was used as one of the crucial features to focus on developing the model and the algorithms. This analysis also indicated that using temperature would be fast in computation, which justified again for the selection of the model.

![Line Chart – Average Temperature May 2017](image)

![Line Chart – Average ET Rates May 2017](image)

Figure 8: Line charts for temperature and ET rates in May 2017

Rainfall precipitation is also crucial in determining the irrigation rate. For example, the following facts were modeled: since in southern California, the climate for Claremont where HMC located is Mediterranean Climate, also known as Dry Summer Climates, so the grass needs to be watered more in order to keep it green since it is extremely dry and hot during summer; while, during the winter, irrigation water is needed much less since it often rains.

**The model**

After data analysis, the data was organized follows in a tree structure as shown Figure 9 on next page based on the physical structure of the HMC lawns according to the clock, channel, zone numbers and station connections. The model captured the needs as the HMC lawn manager required such as to develop an irrigation management system to detect precipitation dynamically and turn off if no irrigation needed. As a result, this designing did meet his need to shut down the irrigation system automatically instead of manually during rainy days. Furthermore, the historical irrigation data and online dynamical weather data were used in the model instead of that of HMC as mentioned before so that the inaccuracy sensor problem due to their damage by ground animals could be avoided. Most importantly, this research is able to mimic human experiences (in this case, the lawn manager’s irrigation experiences) to design the model using ML techniques as shown in the result section.
RESULTS

This work obtained the most results by using ML techniques including Principal Component Analysis (PCA), K-means Clustering, and Gaussian Mixture Models (GMM).

The main results of this research included a smart irrigation algorithm by taking weather conditions and human’s induced irrigation patterns to output irrigation rate which minimized the water amount while keeping the grass green. A Python script code was created to download forecast weather data online and finally an APP was developed to send a message to notify the user daily in the morning with the irrigation rate calculated automatically by the algorithm developed.

Machine Learning

For example, in Figure 10, PCA was applied to two sets of the data, the maximum irrigation rates and maximum temperature, to identify data patterns. It indicates that as temperature decreases the irrigation rate also decreases with the decreasing slope and variation intervals. It was noticed that the data cloud is not bounded by an ellipse as the usual PCA result, rather a fan-like shape. The PCA was performed on minimum temperature and maximum irrigation rates; the PCA analysis for maximum temperature illustrated a much clearer pattern. Thus, the maximum temperature was chosen as one of the important indicators to build in for the algorithm.
Then the K-means Clustering was applied to let the machine learn what would be observed. Different hyperparameter k values were tested and an elbow figure was obtained to conclude which k will minimize the cost function involved. The elbow point k=5 was found which minimizes the cost. From the K-mean ML method, the variance of the irrigation rate also increases as the maximum temperature increases for each cluster. As shown in Figure 11 below, the green data around its centroid is more widely spread than that of the data around the centroid of the yellow data points.

In the meanwhile, the irrigation schedule pattern was analyzed in each month. By dividing irrigation into different months, how temperature influences irrigation amount and frequency was analyzed. The findings showed that irrigation models depend on seasons and patterns of irrigation correlated with temperature. Then, a Gaussian mixture model was used to perform further data analysis. Figure 12 shows the results of irrigation schedule output by running the algorithm.

**Auto irrigation Application development**

Finally, an APP was designed to automatically notify the user of irrigation rate daily by sending his/her a text message and programmed by utilizing Short Message Services (SMS) from Twilio.
to do so, a cloud communication platform that generates an Application Programming Interface (API) to exchange text and picture messages [27]. Specifically, the Twilio account with the phone number "+15622423632" will send a message early in the morning to the lawn manager to help him determine the optimal maximum irrigation rate. The example message is shown in Figure 13 below.

![Figure 13: Sample product result](image.png)

For validation, results were validated by sending the irrigation rates earlier in the morning to the lawn manager and comparing with the irrigation rate he set based on his irrigation human experience. Also, the results were tested for about two weeks from August 25, 2017, to September 15, 2017. From August 25, 2017, to September 12, 2017, Southern California received excessive heat warming. The average temperatures at that time exceeded normal average temperatures, jumping up to more than 100°F. During this period of time, the irrigation rate recommended is almost 15.0000%, which is the largest maximum irrigation rate during that season in the history of data recorded. Whereas, when average temperatures dropped on September 13, the recommended maximum irrigation rate changed to 13.8295%. Moreover, for September 15, it is cloudy with the maximum temperature of 78°F, the recommendation is that "No irrigation is required for today". Thus, the two-week testing periods demonstrate the algorithm works for the real-life situation. For another instance, when a meeting was held with the manager in his office on September 21, 2017, scheduled ahead randomly, it was noticed that the message sent by the APP in the morning recommends 10.0000% for irrigation on that day matched the rate the ground manager set based on his experience.

**CONCLUSION**

The advantage of this method is that it is independent of sensors and ground pipelines and wires; therefore, it avoids the problems of the most current systems in hot and dry regions with pipelines damaged by the ground animals. Instead, the algorithm can mathematically calculate the expected irrigation amount, and moreover, can perform anomaly detection to indicate whether there is a water leakage due to unexpected irrigation amount. More importantly, an adjustable irrigation scheduling was created according to the weather conditions dynamically and also according to special events managed by an organization such as HMC. In this project, techniques have been developed to optimize smart irrigation using HMC’s lawns as examples. Such techniques can be extended to different lawns regardless they are owned by homeowners or by organizations as long as there is historical irrigation data available. It is worth to point out that this method may even be extended to agricultural applications so that one can save water and money in every region in this world.
ACKNOWLEDGEMENTS

Special thanks to HMC student, Herrick Fang, for helping us to setup the APP, HMC Facility and Management Department for providing the data, and HMC Community Engagement led by Gabriela Gamiz for the funding.

REFERENCES


