Journal of Knowledge Information Technology and Systems(JKITS), Vol. 14, No. 5, pp. 535~542, October 2019
https://doi.org/10.34163/jkits.2019.14.5.010

Journal of Knowledge Information Technology and Systems

ISSN 1975-7700

http://www.kkits.or.kr

A Study on Machine Learning-based Grass Demand Forecasting

Wansik An*

Department of Physical Educations, SungKyul University

ABSTRACT

The origin of the grass varies between regions and countries, but in the West, crops, which have been widely used for feed, have long been adapted to livestock grazing, and are derived from plants and perennials with good land cover capacity. In Korea, the origin of grass is different from that of the West. It has been used to decorate and cover the tomb's ground. Thus, Grass is one of the essential elements in our life. Grass is the major resource of various ecosystems and it also provides a space to relax. Nevertheless, Recently, Korea is recognized as a recession due to the reduction of new golf course construction and the slowdown of construction industry. However, since the 5-day system was implemented due to economic development and national income improvement after the Olympic and World Cup, the demand for grass as a green space for recreation and sports Is increasing. In particular, the use of new towns, the West Coast Saemangeum project, neighborhood parks, school grounds, and general residential gardens is increasing. Grass is expected to increase the value added of social indirect capital such as highways, the increase of golf population, the greening of urban and national lands using grass such as the increase of recreational activities and urban grass parks. In addition, the grass industry is a comprehensive field that includes the development, production, composition and management of garden, slope and sports. However, the grass industry is limited to production. This situation is lacking, and there is also a lack of basic data on the system or industry that can support the grass industry. Accordingly, we are necessary to have a research how we improve to use of grass and suggest newly methods with water demand.

KEYWORDS: Machine Learning, Water demand, Forecasting, Irrigation Schedule

ARTICLE INFO: Received 21 September 2019, Revised 11 October 2019, Accepted 11 October 2019.

*Corresponding author is with the Department of Physical Education, Sungkyul University, Sungkyul University-ro 53, Manan-gu Aanyang-si, Gyeonggi-do, 14097, KOREA.
E-mail address: aws119@sungkyul.ac.kr

1. Introduction

Grass is one of the majorally used sport, playground in a world. Major share of available freshwater is used to irrigate grass[1]. Lowland irrigated grass is inefficient measurement because a large proportion of it is lost to seepage and deep percolation[2]. The water-use efficiency of grass is lower than other crops. On an average, 2500 liters of water is used, ranging from 800 liters to more than 5000 liters to produce 1 kg of grass. An increase of 10% in irrigation efficiency can help extent irrigation to an additional 14 million. Solely reducing water use in puddled transplanted grass resulted in proportional reduction in yield, hence various management practices of grass cultivation have to change simultaneously to improve water productivity, without compromising on the productivity of yield. Soaring population is resulting in a huge pressure on the limited fresh water resources of the world[3]. Growing population is also contributing in an increased demand for water from domestic as well as industrial sectors. Consequently, increasing scarcity of available water demands efficient use of irrigation water[4]. Since the available water sources are already almost completely exploited, there is need for achieving increased water productivity through efficient irrigation[5]. Usually 15% of the water delivered is lost in conveyance, another 15% is lost during on-farm supply through field channels and 25% is lost due to inefficient water use practices, thus leaving only 45% of the water supplied to be utilized[6]. The efficiency of the irrigation water use depends on the timing, duration and method of the irrigation employed. To efficiently manage irrigation requirement in crops, information from multiple sources such as soil, plant and atmosphere is required. There are two different approaches for estimating the crop water requirement; one, a conceptual approach based on various factors[7]. Soil moisture, seepage and evapotranspiration, and the other, a theoretical approach based on training a model using the available data. The theoretical approach has been found to be more accurate than the conceptual approach[8]. This developed a framework for controlling real-time irrigation scheduling that took into account some of the most common constraints such as restricted water availability or the maximum or minimum quantity of water applied. It is desired for a effective irrigation scheduling that the water supply matches the water demand as closely as possible. grass has been generally grown under puddled transplanted conditions over large tracts in irrigated areas[9]. However, for improving water use efficiency in irrigated grass, experiments are being conducted to raise the crop under drip irrigation system. In this the water slowly drips out of an emitter, either into the root zone or onto the surface[10]. This minimizes fertilizer and water use. Data available on water use in drip irrigated grass and its irrigation schedule is limited. The study attempts to estimate crop and irrigation water requirement of grass. In our work, the authors have integrated decision support system and a wireless sensor network to achieve an autonomous flexible zone-specific closed loop irrigation system[11].

This paper is organized as follows. In Section 2 details methods and feature measurements. Then methodology are presented and discussed in Section 3. Finally, Section 4 concludes the paper by briefly summarizing the main points.

2. Methods and Feature Measurements

The weather data at site was recorded using an automated weather station. The incorporated sensors are an air temperature and humidity sensor, wind sensors, thermopile pyrometer and automatic rain gauge[12]. All variables are sampled every hour and recorded on a daily basis. The site is a reclaimed alkali loam soil. <Table 1> lists the soil characteristics of the experimental site. Laser-leveler was used to level the field which was then divided into 12 permanent plots.

Table 1. Soil Properties

Soil Property	Soil Sampling Depth (Mean +- SE)	Unit
Clay	19.89 +- 0.50	(%)
Silt	46.07 +- 0.76	(%)
Sand	34.03 +- 0.77	(%)
pH	8.00 ± 0.02	(1:1 soil:water)
EC	0.37 +- 0.02	(dS m-1) (1:1 soil:water)
Total Carbon	0.56 +- 0.01	(%)
Available P	5.74 +- 0.29	(mg kg-1)
Exchangeable K	130 +- 1.73	(mg kg-1)
TN	0.06 +-0.002	(%)
Particle Density	2.57 +- 0.01	(g cm-3)

These plots were separated by 1.0 m wide and 0.20 m high earthen bunds. The crop stand was uniform across the entire area. Drip irrigation lines (67.5 cm apart with emitter spacing of 30 cm) were used for each row.

Multiple parameters effect crop water demand. Some of these are weather based, some crop based, other soil based and still some more. Soil parameters such as salinity, texture, structure, drainage and fertility have not been accounted for because the data has been collected from sub plots of a farm, hence not providing variation that may have otherwise been modeled[13]. However, soil temperature at different levels and at different times during the day has been used to somehow capture the effect of soil parameters. Crop parameters such as crop type, crop variety, crop height and plant density have also not been explicitly included. Crop growth stage and number of days since sowing have been engineered to partially reflect the crop parameters. Other parameters such as type of irrigation system, type of farm, residue management practice and water quality have not been accounted for, as there was no source of variability that may have helped the model to learn better[14]. Weather based (such as solar radiation, rainfall, air humidity, wind direction, wind speed, evapotranspiration and air temperature) parameters have been used in this study[15].

3. Methodology

3.1 Data Collection and Processing

The first step in any machine learning approach is data collection and processing. As data about the weather and the data about irrigation came from two different sources, it had mismatched sampling intervals. This resulted in need of compiling the data from the weather station and the flow meter into one data set. This was done with the guidance of an agriculture expert. This resulted in a data set that contained a one dimensional array containing all the values corresponding the input variables for a day. Five data sub-sets (corresponding to the year 2013, 2014, 2015, 2016 and 2017) were created (having variable season length of 127, 98, 92 and 102 entries each) and complied into one. Since the data set represents replication of the experimental plot for the same set of input parameters, we obtained different quantities of water as target value. The average was considered as the modified target and the standard deviation as the range. Also, wind direction was a categorical variable, containing different wind directions. One-hot encoding was applied to the categorical variable wind direction and it was converted to a Boolean one. would be directly delivered to the application layer.

3.2 Feature and Model Selection

The next step applied was feature selection, which was accomplished using correlation. Since the variables are continuous in nature, Pearson Correlation Coefficients was used to create a matrix depicting the dependencies in the data. A subset of features was obtained by cutting off the variables with values at fixed thresholds and discarding those that had a value above the threshold. Models chosen were Liner regression, Support vector Regressor, Decision Tree regressor, Random forest regressor, Extra tree regressor, adaboost regressor and gradient boosting regressor. It was used to implement these models in Python. Linear model is based on an assumption that the output variable can expressed as a linear combination of the input variables. Given the complex interaction between the input variables, linear model was unable to capture the behavior of the input features for this regression exercise. Hence, it was found to be highly unsuitable for use. A Support Vector Regressor (SVR) finds the hyper-plane that differentiates the data plotted as a point in n-dimensional space where the value on the co-ordinate is equal to the feature value. A decision trees (DT) uses a recursively partitioning approach where growth of the tree is achieved by splitting at each attributes iteratively. This model is computationally cheap, easy to understand-implement-use, straightforward to train and easy to interpret. Over fitting is prevented by pruning the tree. Random forest is an accurate and efficient method which grows out of many trees. Each tree returns a value for the target and decision is made based on the most votes. A random forest regressor (RFR) fits a number of decision trees on sub-samples of the data (the size of which is the same as that of the original input sample). An extra tree regressor (ETR) fits numerous randomized decision trees on variable sized sub-samples of the data. Both RFR and ETR improve accuracy by averaging the result. A gradient booster regressor (GBR) fits a regression tree on the negative gradient of a loss function. At every iteration it optimizes an arbitrary differentiable of the loss function. An AdaBoost regressor (ABR) iteratively fits a regressor with adjusted weights on original data, and we shown by <Table 2>.

Table 2. Model and its optimized parameter value

<table>
<tr><th>Model</th><th>Parameter</th><th>Value (Set1)</th><th>Value (Set2)</th><th>Value (Set3)</th><th>Value (Set4)</th></tr>
<tr><td rowspan="2">Support vector regressor</td><td>C</td><td>2</td><td colspan="2">256</td><td>128</td></tr>
<tr><td>Gamma</td><td>0.000244140625</td><td colspan="2">0.0039063</td><td>0.0078125</td></tr>
<tr><td rowspan="2">Decision tree regressor</td><td>Maximum depth</td><td>2</td><td>7</td><td>3</td><td>3</td></tr>
<tr><td>Minimum samples per split</td><td>32</td><td>35</td><td colspan="2">60</td></tr>
<tr><td rowspan="3">Adaboost regressor</td><td>Learning rate</td><td>0.001</td><td colspan="3">0.0001</td></tr>
<tr><td>Loss function</td><td>Square</td><td colspan="2">Exponential</td><td>Linear</td></tr>
<tr><td>Number of estimators</td><td>500</td><td colspan="2">400</td><td>800</td></tr>
<tr><td rowspan="3">Random forest regressor</td><td>Maximum depth</td><td>2</td><td colspan="2">3</td><td>4</td></tr>
<tr><td>Minimum samples per split</td><td>32</td><td>42</td><td colspan="2">49</td></tr>
<tr><td>Number of estimators</td><td colspan="2">125</td><td>25</td><td>10</td></tr>
<tr><td rowspan="3">Extra tree regressor</td><td>Maximum depth</td><td>10</td><td>7</td><td>13</td><td>9</td></tr>
<tr><td>Minimum samples per split</td><td>35</td><td>41</td><td>35</td><td>33</td></tr>
<tr><td>Number of estimators</td><td>10</td><td>800</td><td>10</td><td></td></tr>
<tr><td rowspan="4">Gradient boosting regressor</td><td>Maximum depth</td><td>3</td><td>2</td><td colspan="2">3</td></tr>
<tr><td>Minimum samples per split</td><td>111</td><td>68</td><td>108</td><td>112</td></tr>
<tr><td>Number of estimators</td><td>10</td><td>25</td><td colspan="2">10</td></tr>
<tr><td>Learning rate</td><td>0.4</td><td>0.2</td><td colspan="2">0.5</td></tr>
</table>

3.3 Model parameter optimization

Of the numerous parameters used for model building only 6 were optimized using GridSearchCV viz: maximum depth (md), maximum features (mf), minimum samples per split (mss), minimum samples per leaf (msl), number of estimators (ne), learning rate (lr), gamma (G) and C, where applicable. Support vector regressor, decision tree and adaboost had limited number of parameters that needed to be optimized. Hence, their optimization was done in one go. Random forest regressor, Extra tree regressor and gradient boosting regressor however have multiple parameters that can be candidate for optimization. To reduce the time complexity of the exhaustive grid search on the limited computing resources available, this was done in a two stage approach. Maximum depth, minimum samples per leaf and minimum samples per split were optimized at stage one. The values received were then added to the values being sent with maximum features, learning rate and number of estimators. The parameter values received from this stage were used to train/test and validate the models.

3.4 Model performance evaluation

Performance of the model can be measured using multiple evaluators, some more suitable than others. The models used have been evaluated against mean square error (calculated using eq 1), coefficient of determination (calculated using eq 4), mean absolute error (calculated using eq 2), estimated variance explained (calculated using eq 3) and accuracy with respect to a range. 3 -fold cross validation is used to measure the performance of the predictive models.

3.4.1 Mean Square Error (MSE)

It is the average of the square of errors. This parameter is similar to MAE, but is sensitive to outliers. Where a = actual value of the target variable, p = predicted value of the target variable and n = no.of instances

$$MSE=\sum_{i=1}^{n}(p_i-a_i)^2/n \tag{1}$$

3.4.2 Mean Absolute Error (MAE)

Mean Absolute Error (MAE) It is the average of how much does the predicted value deviate from the real value.

$$MAE = \sum_{i=1}^{n} (p_i - a_i)^2 / n \tag{2}$$

Where a = actual value of the target variable, p = predicted value of the target variable and n = no.of instances.

3.4.3 Explained Variance Score (EVS)

It measures how well the model has explained the variations in the dataset. This gives the ratio between the variance of error and variance of true values. The highest value for this parameter can be 1.

$$EVS = 1 - Var(p-a) / Var(a) \tag{3}$$

Where a = actual value of the target variable and p = predicted value of the target variable.

3.4.4 Coefficient of Determination (R2 score)

It describes the amount of variance explained by the regression model. One is the desired value. R^2 equal to zero means that the model has failed.

$$R^2 = SSR / SST = 1 - (SSE / SST)$$

Where : Sum of Squares Total:

SST = $\sum (p - \bar{p})^2$, sum of squares regression:

SSR = $\sum (\acute{p} - \bar{\acute{p}})^2$, sum of squares Error:

SSE = $\sum (p - \bar{\acute{p}})^2$, p = predicted value of the target variable.

4. Conclusions

In this paper, we explored traditional machine learning methods for estimating the irrigation schedule of drip irrigated grass. The performance of the models is expressed using the common regression performance parameters such as mean absolute error, median absolute error, explained variance score, coefficient of determination, mean square error and accuracy. The models are validated using <Figure 3> fold iterative process in which the previously used dataset becomes the training set and the data of the next time quantum becomes a test set.

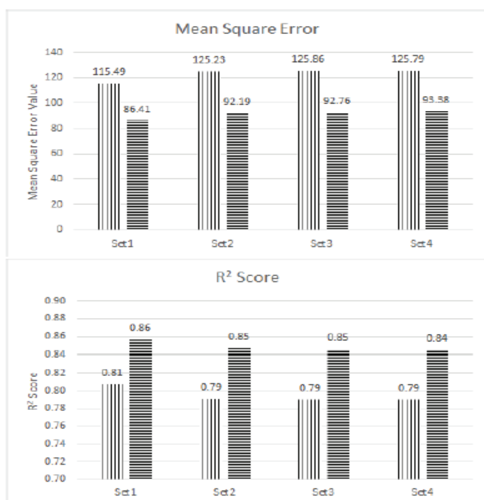

Figure 3. Comparison of Results

The performance of the models has been improved by using fine-tuning hyper - parameters. Given the widespread application of neural networks in the field of irrigation scheduling, we can extend this current work using the concepts of deep learning. In the coming future, water availability is likely to be a bigger constraint than land so it is time to likely change mind set from improving agricultural productivity from per unit land to per unit of water. Hence, the use of a computationally intelligent approach becomes more relevant in the present scenario.

References

[1] J. Timsina and D. J. Connor, *Productivity and management of grass¬wheat cropping systems: issues and challenges,* Field crops research, Vol. 69, No. 2, pp. 93-132, 2001.

[2] S. Kukal, G. Hira, and A. Sidhu, *Soil matric potential-based irrigation scheduling to grass (oryza sativa)*, Irrigation Science, Vol. 23, No. 4, pp. 153-159, 2005.

[3] P. Belder, B. Bouman, R. Cabangon, L. Guoan, E. Quilang, L. Yuanhua, J. Spiertz, and T. Tuong, *Effect of watersaving irrigation on grass yield and water use in typical lowland conditions in asia,* Agricultural water management, Vol. 65, No. 3, pp. 193-210, 2004.

[4] Z. Mimi and M. Smith, *Statistical domestic water demand model for the west bank,* Water International, Vol. 25, No. 3, pp. 464-468, 2000.

[5] I. Pulido-Calvo, and J. C. Gutierrez-Estrada, *Improved irrigation water demand forecasting using a soft-computing hybrid model,* Biosystems Engineering, Vol. 102, No. 2, pp. 202-218, 2009.

[6] S. K. Saleem, D. Delgoda, S. K. Ooi, K. B. Dassanayake, L. Liu, M. Halgamuge, and H. Malano, *Model predictive control for real-time irrigation scheduling,* IFAC Proceedings Volumes, Vol. 46, No. 18, pp. 299-304, 2013.

[7] K. Djaman, V. Mel, A. B. Balde, V. B. Bado, B. Manneh, L. Diop, D. Mutiibwa, D. Rudnick, S. Irmak, and K. Futakuchi, *Evapotranspiration, irrigation water requirement, and water productivity of grass (oryza sativa l.) in the sahelian environment,* Paddy and water environment, Vol. 15, No. 3, pp. 469-482, 2017.

[8] C. Goumopoulos, B. OFlynn, and A. Kameas, *Automated zone-specific irrigation with wireless sensor/actuator network and adaptable decision support*, Computers and electronics in agriculture, Vol. 105, pp. 20-33, 2014.

[9] K. Steppe, D. J. De Pauw, and R. Lemeur, *A step towards new irrigation scheduling strategies using plant-based measurements and mathematical modelling,* Irrigation Science, Vol. 26, No. 6, p. 505, 2008.

[10] P. Steduto, T. C. Hsiao, D. Raes, and E. Fereres, *Aquacropthe fao crop model to simulate yield response to water: I. concepts and underlying principles,* Agronomy Journal, Vol. 101, No. 3, pp. 426-437, 2009.

[11] I. El-Magd and T. Tanton, *Remote sensing and gis for estimation of irrigation crop water demand,* International Journal of Remote Sensing, Vol. 26, No. 11, pp.

2359-2370, 2005.

[12] A. F. Torres, W. R. Walker, and M. McKee, *Forecasting daily potential evapotranspiration using machine learning and limited climatic data,* Agricultural Water Management, Vol. 98, No. 4, pp. 553-562, 2011.

[13] H. Navarro-Hell´ın, J. Mart´ınez-del Rincon, R. Domingo-Miguel, F. Soto-Valles, and R. Torres-Sanchez, *A decision support system for managing irrigation in agriculture,* Computers and Electronics in Agriculture, Vol. 124, pp. 121-131, 2016.

[14] T. Ojha, S. Misra, and N. S. Raghuwanshi, *Wireless sensor networks for agriculture: The state-of-the-art in practice and future challenges,* Computers and Electronics in Agriculture, Vol. 118, pp. 66-84, 2015.

[15] K. H. Amer, A. A. Samak, and J. L. Hatfield, *Effect of irrigation method and non-uniformity of irrigation on potato performance and quality,* Journal of Water Resource and Protection, Vol. 8, No. 03, p. 277, 2016.

머신러닝 기반의 잔디 수요예측에 관한 연구

안완식

성결대학교 사범대학 체육학과 교수

요 약

잔디의 기원은 지역과 국가에 따라 다르지만 서부에서는 사료로 널리 사용되는 작물이 오랫동안 가축 방목에 적응해 왔으며 토지 피복 능력이 좋은 식물과 다년생 식물에서 유래되었다. 한국에서 풀의 기원은 서구의 기원과 다르다. 무덤의 바닥을 장식하고 덮는데 사용되었다. 따라서 잔디는 우리 삶의 필수 요소 중 하나이다. 잔디는 다양한 생태계의 주요 자원이며 휴식을 취할 수 있는 공간을 제공한다. 그럼에도 불구하고, 최근 한국은 새로운 골프장 건설 감소와 건설업 둔화로 경기 침체로 인식되고 있다. 그러나 올림픽과 월드컵 이후 경제 발전과 국가 소득 개선으로 레크리에이션과 스포츠를 위한 녹지 공간으로서 잔디에 대한 수요가 증가하고 있다. 특히, 새로운 도시, 서해안 새만금 프로젝트, 인근 공원, 학교 운동장 및 일반 주거 정원의 사용이 증가하고 있다. 잔디는 고속도로, 골프 인구 증가, 레크리에이션 활동의 증가와 도시 잔디 공원과 같은 잔디를 사용한 도시 및 국토의 녹화와 같은 사회적 간접 자본의 부가가치를 증가시킬 것으로 예상된다. 또한 잔디 산업은 정원, 경사면 및 스포츠의 개발, 생산, 구성 및 관리를 포함하는 포괄적 인 분야이기도 하다. 그러나 잔디 산업은 생산으로 제한된다. 이러한 이유는 잔디 산업 지원이 부족하고, 시스템 또는 산업에 대한 기본 데이터도 부족하기 때문이다. 따라서 잔디 사용을 개선하고 수요가 있는 새로운 방법을 제안하는 방법에 대한 연구가 필요하다.

Wan Sik An received the bachelor’s degree in the College of Art and Physical Education from the Hanyang University. He received the M.S. degree and the Ph.D. degree in the College of Physical Education from Dankook University, respectively. He has a professor in the Department of Physical Education at Sungkyul University in present.

E-mail address: aws119@sungkyul.ac.kr