A SELF-OPTIMIZING SYSTEM FOR MEASURING AND PREDICTING SOIL MOISTURE CONTENT AND LEAF WETNESS LEVEL IN CROP FIELDS

[1]THOMAS TRUONG, [2]PHAM SON, ANH DINH

[1]Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, Canada
Email: thomas.truong@usask.ca

Abstract- Global crop irrigation consumes a substantial amount of the world's freshwater withdrawals. Optimizing water usage efficiency in agricultural practices becomes a priority in ensuring global water and food security. This paper presents the design of a system which will assist crop field managers in determining the optimal irrigation schedule for their crops by providing real-time information on the soil moisture content and leaf wetness levels of their crop fields. Additionally, the system predicts the next day soil moisture content and leaf wetness levels using a self-optimizing support vector machine regression algorithm. Ultimately, the results of this system will assist crop field managers in optimizing the water usage efficiency of their crops.

Keywords- Machine Learning; Weather Prediction; Precision Agriculture;Local Weather Station; Internet Of Things

I. INTRODUCTION

Global crop irrigation consumes an alarming 70% of the global freshwater withdrawals[1]. Moreover, irrigation consumption is expected to increase substantially with global crop demand being estimated to double by 2050[2]. Optimizing water usage efficiency in agricultural uses becomes essential for ensuring global water and food security. Proper optimization of water usage requires thorough and accurate data on environmental conditions such as soil moisture content and leaf wetness level.

Soil moisture content is a measure of the water in the soil and thus is a rough measure of the water available for the crop. The significance of soil moisture content information available to the crop is obvious as it informs crop field managers if the crops have access to sufficient water resources to optimally grow. Too much or too little soil moisture content is harmful for water usage efficiency in crop production.

Leaf wetness level is a measure of water present on the leaves of a crop and is a rough measure of the dew present on the crop leaves. While the contributions of dew to providing crops water is negligible, dew has been found to be a major ecological factor in improving the water usage efficiency of crops[3]. As such, leaf wetness becomes an important factor for determining the quantity of water available for optimal corp production.

The significance of soil moisture content and leaf wetness in optimizing water usage efficiency creates a need for accurate information on soil moisture content and leaf wetness level. In the past systems have been designed to estimate, record, and predict future soil moisture content levels using neural networks [4][5] and support vector machines[6]; however, these past works are limited to very short term (one hour in advance) predictions and are not designed to improve as new measurements are acquired.Moreover, the accuracy of recently popular machine learning methods on leaf wetness have not been well documented in literature.

This paper presents a self-optimizing system which provides knowledge of past, current, and future measurements of soil moisture content and leaf wetness level. Predictions are made using support vector machine regression. An automatic optimization algorithm that improves results as new data is acquired is also designed.The information which this system provides will help crop field managers optimally determine their water management schemes in order to maximize water usage efficiency for crop production. This system is a modification of an existing system which has already been successfully implemented by the authors to assist crop field managers in predicting high risk environments for harmful fungal diseases in the field [7].

II. HARDWARE DESIGN

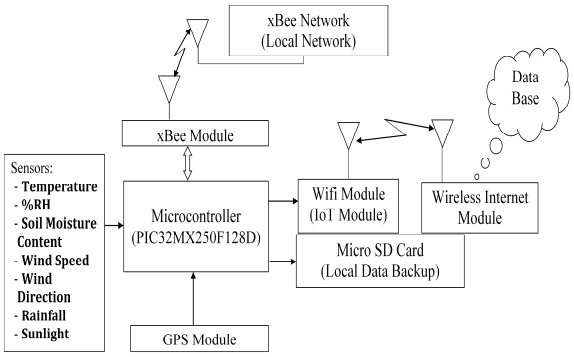


Fig. 1. The high-level hardware block diagram for a single data collection device.

Fig. 1 is the high-level hardware block diagram for the device which provides past and current values of soil moisture content and leaf wetness. Additionally, this device is responsible for managing the communications between various sensors and sending the information to a cloud data base.

The device is powered by a 12V lead-acid battery with a battery autonomy ranging from 3 days up to 7 days depending on the user specified sleep duration of the device. The device is recharged during operation in the field via a solar panel charger. The device is easily interfaced with up to 13 analog sensor devices. Analog sensors measuring conditions such as temperature, relative air humidity, wind speed, rain fall, and solar luminous emittance are also present alongside soil moisture content and leaf wetness sensors.The soil moisture content sensor provides a percentage value with 0% being dry and 100% being submerged in water. The leaf wetness level sensor provides a value from 0-15 with 0 being dry and 15 being submerged in water.

Internet access is provided to the device using an LTE HUA8372 WiFi dongle distributed by Bell. The xBee network acts as a data relay device for other data collection devices in the local vicinity. It allows for devices which may have a poor cellular network connection to conserve power and send its information to a master device with the best cellular or WiFi connection.

III. HARDWARE OPERATION

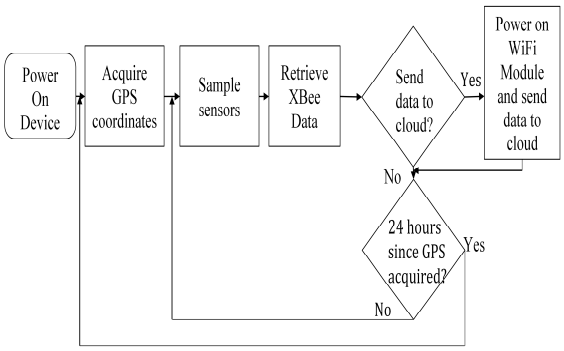


Fig. 2. The software flow diagram for a single data collection device.

Fig. 2 is the flow diagram for the data collection device. The microcontroller on the device is programmed and configured using MPLAB XC32, an integrated development environment built by Microchip Technology. The microcontroller handles all communications between various pieces of hardware on the device. On power up, the device collects the GPS coordinates of the device and then periodically samples and records environmental conditions on a Micro SD Card at a user specified interval. These environmental conditions are also sent to an online data base which can be accessed through a website built with basic HyperText Markup Language and Javascript. The website contains a Google Maps application which displays the locations of the devices in real-time and contains the hyperlinks to the data bases for each respective device. Additionally, the device will retrieve and relay any data from devices connected via the xBee network.

Once set-up in the field, the device operates discreetly and requires no maintenance. Furthermore, it is not required to go out in the field to collect the locally stored data as all the sampled information is stored on an online data base which is easily accessible from any location.

IV. SUPPORT VECTOR MACHINE REGRESSION DESIGN

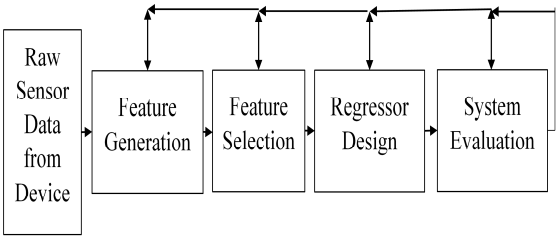


Fig. 3. The high-level software block diagram for the next day soil moisture content and leaf wetness prediction system.

Machine learning based approaches have been successful in predicting other environmental conditions such as wind speed[8], solar irradiance[9], daily potential evapotranspiration [10], atmospheric temperature[11], andsoil moisture content [4][5][6]. More specifically, support vector machine regression has been a successful technique in weather prediction and is the regressor of choice for this project.

Fig. 3 shows the block diagram for the support vector machine regression system for next day soil moisture content and leaf wetness prediction. The raw sensor data is retrieved from the online data base from a remote computer in a lab and the data is used to create a feature vector of the form:

$$\mathbf{x} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1M} \\ x_{21} & x_{22} & \cdots & x_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ x_{(L-1)1} & x_{(L-1)2} & \cdots & x_{(L-1)M} \end{bmatrix} \quad (1)$$

Each of the L rows corresponds to each of the days collected, ie. row 4 corresponds to the data collected on day 4. Each of the M columns corresponds to a raw value for a specific sensor at a specific time for that day, ie. column 1 will correspond to the soil moisture content measurement at 00:00. Feature vector **x** is appended once a full set of M measurements are made for that day. As such, day L is omitted from the feature vector until all M measurements are made for day L.

Next, these feature vectors are used to construct truth vectors of the form:

$$\mathbf{Y_T} = [Y_{T2} \quad Y_{T3} \quad \cdots \quad Y_{T(L-1)}] \quad (2)$$

In total there are two distinct truth vectors$\mathbf{Y_T}$, one for soil moisture content and one average leaf wetness. Each vector is used to train two distinct support vector machine regression systems, one predicting next day soil moisture content and the other predicting next day leaf wetness level. Each element in these truth vectors contain the daily average measurement for that specific day. The values are calculated based on the M

elements for each row in **x**. Note that Y_{T2} is the first element and is calculated based on the M elements in row 2 (from day 2) in **x**; however, Y_{T2} is treated by the machine learning algorithm as the truth value (the result of) measurements made on day 1. Ie., the first element in the vector $\mathbf{Y_T}$for leaf wetness level contains the average leaf wetness level for day 2 and is the truth value for the measurements made on day 1. The structure of these truth vectors allows for training machine learning algorithms to predict future values based on past and present data.

Feature vector **x** and truth vector $\mathbf{Y_T}$are then used as inputs to a support vector machine regression machine learning algorithm using a radial basis kernel provided by the Scikit-Learn library[12]. A prediction for day L is made and is placed in a prediction vector of the form:

$$\mathbf{Y_P} = [Y_{PL}] \qquad (3)$$

Note that there are also two $\mathbf{Y_P}$ vectors corresponding to each $\mathbf{Y_T}$vector.

Once a complete set of sensor data for day L is collected, **x** is appended with row L containing M elements and $\mathbf{Y_T}$is appended with the column L containing the value Y_{TL}. The performance of the system is evaluated through calculation of an absolute error.

$$\text{Absolute Error} = |Y_{PL} - Y_{TL}| \qquad (4)$$

The support vector machine regression algorithm is then retrained with the newly updated vectors **x** and $\mathbf{Y_T}$and then new predictions are made for day L+1 so that $\mathbf{Y_P}$is appended to become:

$$\mathbf{Y_P} = [Y_{PL} \quad Y_{P(L+1)}] \qquad (5)$$

This process of iteratively updating vectors **x** and $\mathbf{Y_T}$ and training the machine learning algorithm to predict elements in $\mathbf{Y_P}$forms a next day prediction system for soil moisture content and leaf wetness.

Next, the feature vector **x** is optimized to each of the truth vectors $\mathbf{Y_T}$ by calculating the cross correlation between elements in **x** and elements in $\mathbf{Y_T}$. The results of these calculations can then be used to select the most statistically significant elements in **x** to train the machine learning algorithm. A new feature vector is created of the form:

$$\mathbf{x_{opt}} \subseteq \mathbf{x} \qquad (6)$$

This optimized feature vector $\mathbf{x_{opt}}$is used in place of **x** to train the algorithm. The chosen optimal number of elements to retain is determined by choosing the number of most statistically significant elements which minimizes the error in the past predictions made. This optimization is done after each new day of data is collected to continually improve prediction results.

Furthermore, a grid search of parameters C (the penalty parameter of the error term) and ϵ ("slack" parameter associated with the training loss function) is done. The optimal parameters are chosen by minimizing error in past predictions made. This optimization also done after each new day of data is collected to continually improve prediction results. A detailed formulation of these parameters and the support vector machine regression algorithm can be found in the documentation of the Scikit-Learn library[12].

V. RESULTS

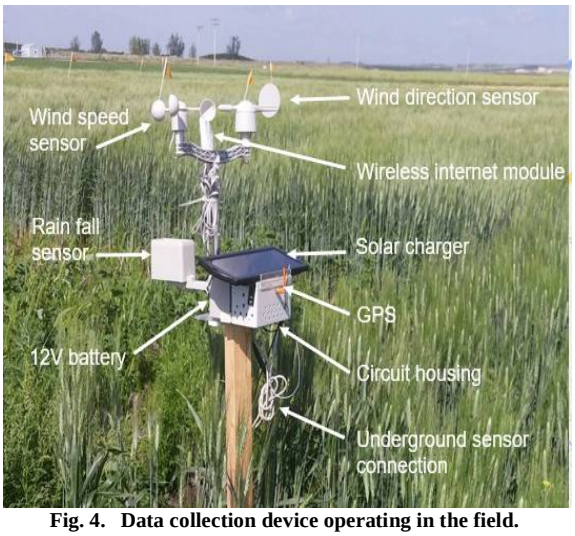


Fig. 4. Data collection device operating in the field.

Fig. 4 shows the operation of a single environmental data collection device in the field. The device sends its geographical location to a website containing a Google Maps JavaScript application program interface (API). The API takes the geographical information of the device and places a marker on the Google Map at the device location. The marker contains a hyperlink which leads to the Sparkfun database where the device measurements are retrieved.

A total of 3 devices were set-up around Saskatoon, Saskatchewan between the months of July, August, and September in 2016. The most successful device collected data from July 30th, 2016 to September 9th, 2016. The device required no maintenance during this time and was only taken down once the crops were harvested by the field managers. Data were collected from the sensors at an interval of 1 sample every 10 minutes. The machine learning algorithm outlined insection IV was developed and trained using this device's data.

Fig. 5 shows the predicted values for daily average soil moisture content having an average absolute error of 2.37%. Fig. 6 shows the predicted values for the daily average leaf wetness level having an average absolute error of 1.61. Note that the actual leaf wetness levels are very chaotic, resulting in a comparatively worse result than the soil moisture content predictions.

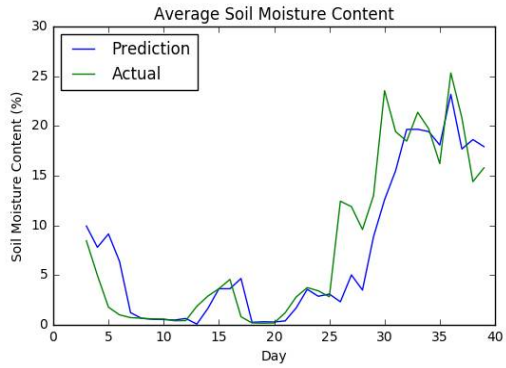


Fig. 5. The true and prediction values for average daily soil moisture content.

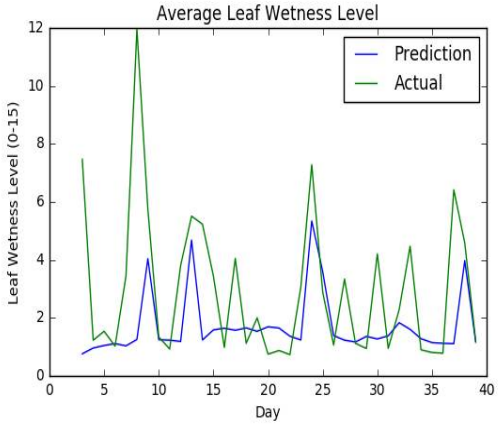


Fig. 6. The true and prediction values for average daily leaf wetness level.

VI. FUTURE WORK

More data will be collected in the 2017 growing season and will be used to further evaluate the performance of the machine learning algorithm. In addition to providing a more thorough performance evaluation of the system, the increased number of samples will ideally improve the results of the system as well.

Additionally, the irrigation schedule of the crop field is unknown but will be highly correlated to the measured soil moisture content and leaf wetness levels. As such, including the irrigation schedule as an input in the algorithm should also improve the results.

The system will also be further modified to incorporate other environmental predictions for different time frames to assist crop scientists and field managers in continually improving agricultural practices.

CONCLUSION

This paper presents the design of a self-optimizing system which predicts the soil moisture content and leaf wetness levels in crop fields. The devices in the system collected data for about 40 days at 10 minute intervals between the months of July, August, and September in 2016 in various crop fields around Saskatoon, Saskatchewan.Data are sent in real-time to an easily accessible online data base. The data are then retrieved by a remote computer and is processed and used to predict average soil moisture content and average leaf wetness levels on a day to day basis using a support vector machine regression algorithm. The algorithm optimizes its performance as real-time data arrives to the online data base.

With the data collected so far, the system can predict soil moisture content with an average absolute error of 2.37% and leaf wetness levels to an average absolute error of 1.61 (on a wetness scale ranging from 0 to 15). These predictions will assist crop field managers in determining the optimal irrigation schedule for their crops which optimize water usage efficiency.

ACKNOWLEDGMENT

The authors would like to acknowledge the support from the Agriculture Development Fund, Saskatchewan, Canada.

REFERENCES

[1] S. Siebert, J. Burke, J. Faures, K. Frenken, J. Hoogeveen, P. Doll and F. Portmann, "Groundwater use for irrigation - a global inventory," Hydrology and Earth System Sciences, vol. 14, pp. 1863-1880, 2010.

[2] D. Tilman, C. Balzer, J. Hill and B. Befort, "Global food demand and the sustainable intensification of agriculture," Proceedings of the National Academy of Sciences in the United States of America, vol. 108, no. 50, 2011.

[3] J. Ben-Asher, P. Alpert and A. Ben-Zvi, "Dew is a major factor affecting vegetation water use efficiency rather than a source of water in the eastern Mediterranean area," Water Resources Research, vol. 46, 2010.

[4] S. K. Lenka and A. G. Mohapatra, "Gradient descent with momentum based neural network pattern classification for the prediction of soil moisture content in precision agriculture," in IEEE International Symposium on Nanoelectronic and Information Systems, Indore, India, 2015.

[5] R. Liang, Y. Ding, X. Zhang and W. Zhang, "A real-time prediction system of soil moisture content using genetic neural network based on annealing algorithm," in International Conference on Automation and Logistics, Qingdao, China, 2008.

[6] L. Pasolli, C. Notarnicola and L. Bruzzone, "Estimating Soil Moisture With the Support Vector Regression Technique," IEEE Geoscience and Remote Sensing Letters, vol. 8, no. 06, pp. 1080-1084, 2011.

[7] T. Truong, A. Dinh and K. Wahid, "An IoT Environmental Data Collection System for Fungal Detection in Crop fields," in IEEE Canadian Conference on Electrical and Computer Engineering, Windsor, Canada, 2017.

[8] S. Salcedo-Sanz, E. G. Ortiz-Garcia, A. Perez-Bellido, A. Portilla-Figueras and L. Prieto, "Short term wind speed prediction based on evolutionary support vector regression algorithms," Expert Systems with Applications, vol. 38, 2011.

[9] F. Wang, Z. Zhen, Z. Mi, H. Sun and G. Yang, "Solar irradiance feature extraction and support vector machines based weather status pattern recognition model for short-term photovoltaic power forecasting," Energy and Buildings, 2015.

[10] Torres, W. Walker and M. McKee, "Forecasting daily potential evapotranspiration using machine learning and limited climactic data," Agricultural Water Management, 2011.

[11] Y. Radhika and M. Shashi, "Atmospheric temperature prediction using support vector machines," International Journal of Computer Theory and Engineering, 2009.

[12] Scikit-learn developers, "Documentation," 2016. [Online]. Available: http://scikit-learn.org/stable/documentation.html#. [Accessed 1 May 2017].

★★★