NSGA-II trained Artificial Neural Network in Soil Moisture Quantity Prediction

Project report in partial fulfilment of the requirement for the award of the degree of

Bachelor of Technology

In

COMPUTER SCIENCE & ENGINEERING

Submitted By

Supriyo Jana	Department of Computer Science & Engineering
University Roll No. 12016009001079	UEM, Kolkata
Sagnik Sengupta	Department of Computer Science & Engineering
University Roll No. 12016009001093	UEM, Kolkata
Sagar Das	Department of Computer Science & Engineering
University Roll No. 12016009001069	UEM, Kolkata

Under the guidance of

Prof. Sankhadeep Chatterjee

Prof. Somarpita Dutta

Department of Computer Science & Engineering

UEM, Kolkata

UNIVERSITY OF ENGINEERING & MANAGEMENT, KOLKATA

University Area, Plot No. III – B/5, New Town, Action Area – III, Kolkata – 700 156

CERTIFICATE

This is to certify that the project entitled **NSGA-II trained Artificial Neural Network in Soil Moisture Quantity Prediction** submitted by **Supriyo Jana (University Roll No. 12016009001079), Sagnik Sengupta (University Roll No. 12016009001093)** and **Sagar Das (University Roll No. 12016009001069)** Students of UNIVERSITY OF ENGINEERING & MANAGEMENT, KOLKATA, in fulfilment of requirement for the degree of Bachelor of Computer Science & Engineering is a bona fide work carried out by them under the supervision and guidance of Prof. Sankhadeep Chatterjee and Prof. Somarpita Dutta during 8^{th} Semester of academic session of 2018-2019. The content of this report has not been submitted to any other university or institute for the award of any other degree.

I am glad to inform that the work is entirely original and its performance is found to be quite satisfactory.

Prof. Somarpita Dutta	Prof. Sukalyan Goswami
Assistant Professor	Head of the Department
CSE Department	CSE Department
UEM, Kolkata	UEM, Kolkata

ACKNOWLEDGEMENT

We would like to take this opportunity to thank everyone whose cooperation and encouragement throughout the ongoing course of this project remains invaluable to us.

We are sincerely grateful to our guide Prof. Somarpita Dutta and Prof. Sankhadeep Chatterjee of the Department of Computer Science & Engineering, UEM, Kolkata, for his wisdom, guidance and inspiration that helped us to go through with this project and take it to where it stands now.

We would also like to express our sincere gratitude to Prof. Sukalyan Goswami, HOD, Computer Science & engineering, UEM, Kolkata and all other departmental faculties for their ever-present assistant and encouragement.

Last but not the least, we would like to extend our warm regards to our families and peers who have kept supporting us and always had faith in our work.

Supriyo Jana

Enrollment No. 12016009001079

Sagnik Sengupta

Enrollment No. 12016009001093

Sagar Das

Enrollment No. 12016009001069

TABLE OF CONTENTS

ABSTRACT

Soil moisture is one of the main factors in agricultural production and hydrological cycles, and its precise prediction is important for the rational use and management of water resources. However, soil moisture involves complex structural characteristics and meteorological factors, and it is difficult to establish an ideal mathematical model for soil moisture prediction. Existing prediction models have problems such as prediction accuracy, generalization, and multi-feature processing capability, and prediction performance must improve.

Based on this, taking the Environmental Datasets of University of Toronto Mississagua Campus which present measurements taken in 3 distinct environmental settings, the NSGA-II with big data fitting capability was proposed to construct a soil moisture prediction model. The data are separated into pond, field and forest data. Data are collected at these sites using HOBO U30 data loggers equipped with sensors monitoring soil moisture, soil temperature, air temperature and relative humidity. Data are collected hourly and downloaded monthly.

By integrating the dataset, analysing the time series of the predictive variables, and clarifying the relationship between features and predictive variables selected meteorological parameters can provide effective weights for moisture prediction. Test results prove that the deep learning model is feasible and effective for soil moisture prediction. Its' good data fitting and generalization capability can enrich the input characteristics while ensuring high accuracy in predicting the trends and values of soil moisture data and provides an effective theoretical basis for water-saving irrigation and drought control.

INTRODUCTION

Water is the primary resource that determines the survival and development of the Earth's inhabitants. Soil moisture not only plays an important role in maintaining plant growth but also is a key link in the water cycle of soil-plant-atmosphere continuum systems. However, as human activities intensify, groundwater resources deteriorate in water quality, and the amount of excavation is significantly exceeded. The continuous decline of groundwater levels leads to a decrease in soil water content and reduces the effective water storage capacity of the soil. Especially in dry areas, the lack of precipitation causes the soil water to not replenish in sufficient time, which negatively affects the normal growth of crops. In this case, it is particularly important to develop an appropriate irrigation system at the right time. The growth and regression of soil moisture directly affects water consumption and growth of crops. It is an important indicator for drought resistance, flood control, and precision irrigation decisions in agricultural production. It is important to achieve accurate prediction of soil water regression regular patterns to properly manage agricultural water resources and promote crop yield increases.

Artificial Neural Networks (ANNs) are system composed of neurons organized in input, output, and hidden layers. The neurons are connected to each other by a set of synaptic weights. An ANN is a powerful tool that has been applied in a broad range of problems such as pattern recognition, forecasting, and regression. During the learning process, the ANN continuously changes their synaptic values until the acquired knowledge is sufficient (until a specific number of iterations is reached or until a goal error value is achieved). When the learning process or the training stage has finished, it is mandatory to evaluate the generalization capabilities of the ANN using samples of the problem, different to those used during the training stage. Finally, it is expected that the ANN can classify with an acceptable accuracy the patterns from a particular problem during the training and testing stage.

Several classic algorithms to train an ANN have been proposed and developed in the last years. However, many of them can stay trapped in no desirable solutions; that is, they will be far from the optimum or the best solution. Moreover, most of these algorithms cannot explore multimodal and noncontinuous surfaces. Therefore, other kinds of techniques, such as bioinspired algorithms (BIAs), Genetic Algorithm like PSO, NSGA, NSGA-II are necessary for training an ANN.

With the rapid development of artificial intelligence in recent years, in 2006, Hinton proposed Deep Learning (DL), which uses a multiple hidden layer structure to increase the classification and fitting capability to big data and multi-feature data. Compared with traditional neural networks, it shows strong computing power and has been successfully applied in image recognition, search engines, stock price predictions, and other fields. Owing to the nonlinear and extremely complex nature of soil, some scholars have introduced DL into soil particle size and soil texture analysis in recent years, overcoming the problems of low prediction accuracy. Based on this, our aim is to construct and optimize a soil moisture prediction model through deep learning and its powerful data processing capabilities to achieve high-precision prediction of soil moisture.

LITERATURE AND SURVEY

2.1 Data acquisition and overview

The Environmental Datasets represent measurements taken in 3 distinct environmental settings at the University of Toronto Mississauga campus. The data are separated into **pond**, **field** and **forest** data. Data are collected at these sites using HOBO U30 data loggers equipped with sensors monitoring soil moisture, soil temperature, air temperature and relative humidity. Data are collected hourly and downloaded monthly.

Column Name:	Description:
`excel_datetime_code`	Data type: Float | The Excel date and time code corresponding to the recorded measrurement. Format this is as a date or time in Excel or use as an arbitrary timescale in other software packages.
`excel_day`	Data type: Integer | The Excel code corresponding to the day of observation, format as a date in Excel or convert to other formats for use in other packages.
`excel_time`	Data type: Float | The Excel code corresponding to the time of observation (in EST).
`field_soil_temp_c`	Data type: Float | Soil temperature at the field site in degrees Celcius from a sensor buried 30cm below surface.
`field_air_temp_c`	Data type: Float | Air temperature at the field site in degrees Celcius.
`field_rh`	Data type: Float | Relative humidity in (%) at the field site.
`field_soil_wc`	Data type: Float | Soil water content in m3/m3 recorded at the field site by a sensor 30cm below soil surface.
`forest_soil_temp_c`	Data type: Float | Soil temperature at the forest site in degrees Celcius from a sensor buried 30cm below surface.
`forest_air_temp_c`	Data type: Float | Air temperature at the forest site in degrees Celcius.
`forest_rh`	Data type: Float | Relative humidity in (%) at the forest site.
`pond_soil_temp_c`	Data type: Float | Soil temperature at the pond site in degrees Celcius from a sensor buried 30cm below surface.
`pond_air_temp_c`	Data type: Float | Air temperature at the pond site in degrees Celcius.

Column Name:	Description:
pond_rh	Data type: Float | Relative humidity in (%) at the pond site.
pond_soil_wc	Data type: Float | Soil water content in m3/m3 recorded at the pond site by a sensor 30cm below soil surface.

2.2 Data processing and analysis

The most important elements to design and improve the accuracy of an ANN are the architecture (or topology), the set of transfer functions (TF), and the set of synaptic weights and bias. These elements should be codified into the individual that represents the solution of our problem. The solutions generated by the bioinspired algorithms will be measured by the fitness function with the aim to select the best individual which represents the best ANN. The three bioinspired algorithms (basic PSO, SGPSO, and NMPSO) are going to lead the evolutionary learning process until finding the best ANN by using one of the eight fitness functions proposed in this paper. It is important to remark that only pattern classification problems will be solved by the proposed methodology.

In Figure 1, a diagram of the proposed methodology is shown. During the training stage, it is necessary to define the individual and the fitness functions to evaluate each individual. The size of the individual depends on the size of the input patterns as well as the desire patterns. The individual will be evolved during a certain time to obtain the best solution (with a minimum error). At the end of the learning process, it is expected that the ANN provides an acceptable accuracy during the training and testing stage.

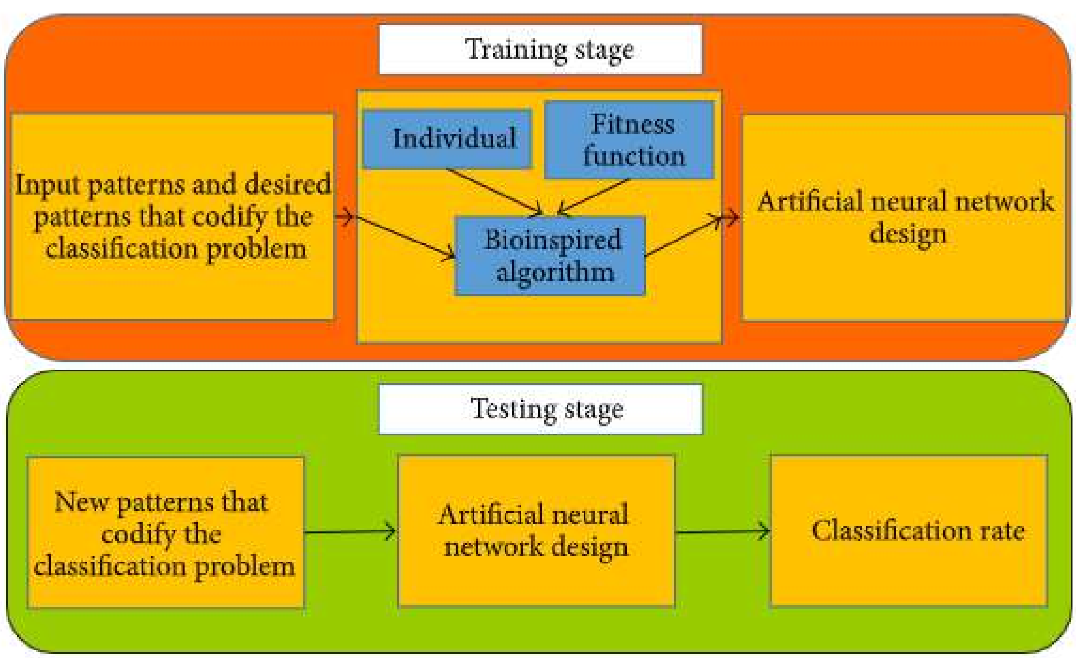


Fig 1

PROBLEM STATEMENT & DISCUSSION

2.1 Problem Statement

NSGA-II trained Artificial Neural Network in Soil Moisture Quantity Prediction

The basic regression model with different sets of more readily observed environmental variables (land use, topographic and meteorological factors) were developed for the prediction of soil moisture content in space and time. The model performances were evaluated in the University of Toronto Mississagua Campus, with soil moisture content measurements.

It was indicated that the regression models could describe the relationships of soil moisture content with environmental attributes. It was found that the regression model showed the best goodness of fit since it explained the greatest fraction of soil moisture variation in both space and time, and the predicted mean, standard deviation, minimum and maximum soil moisture were closest to the observed values. This model was also either the most precise or the most economical in prediction of soil moisture content in space and time since it gives the lowest values in explained variance score (EVS), max error (ME), mean absolute error (MAE) and mean squared log error (MSLE).

2.2 Discussion

Four evaluation measures were selected to indicate the performance of the different models.

Mean Absolute Error (MAE) is: $\frac{1}{m}\sum_{i=1}^{m}\left|\left(y_i - \hat{y}_i\right)\right|$ **(1)**

Mean Squared Error (MSE) is: $\frac{1}{m}\sum_{i=1}^{m}\left(y_i - \hat{y}_i\right)^2$ **(2)**

In the above formula, $\hat{y}_i$ is the predicted value, y_i is the true value, and $\bar{y}_i$ is the average value. MAE is the average of absolute errors; it can reflect the actual situation of the predicted value error. MSE is the expected value of the square of the difference between the parameter estimate and the parameter true value, it can evaluate the degree of the data change, and the smaller value of the MSE, the better accuracy of the prediction model. RMSE is the arithmetic square root of MSE. R^2 can eliminate the influence of dimension on evaluation measure.

Explained variance : The **explained_variance_score** computes the explained variance regression score.

If y^ is the estimated target output, y the corresponding (correct) target output, and Var is Variance, the square of the standard deviation, then the explained variance is estimated as follow:

explained_variance (y, y^) =1− (Var{y−y^} Var{y})

The best possible score is 1.0, lower values are worse.

Max error : The **max_error** function computes the maximum residual error , a metric that captures the worst case error between the predicted value and the true value. In a perfectly fitted single output regression model, max_error would be 0 on the training set and though this would be highly unlikely in the real world, this metric shows the extent of error that the model had when it was fitted.

If y^i is the predicted value of the i-th sample, and yi is the corresponding true value, then the max error is defined as

Max Error (y, y^) =max(|yi−y^i|)

PROPOSED SOLUTION & RESULT ANALYSIS

Solution

NSGA-II is one of the most popular multi objective optimization algorithms with three special characteristics, fast non-dominated sorting approach, fast crowded distance estimation procedure and simple crowded comparison operator [2]. Deb et al. [7] simulated several test problems from previous study using NSGA-II optimization technique, and it is claimed that this technique outperformed PAES and SPEA in terms of finding a diverse set of solutions [3,5]. Generally, NSGA-II can be roughly detailed as following steps.

Step 1: Population initialization Initialize the population based on the problem range and constraint.

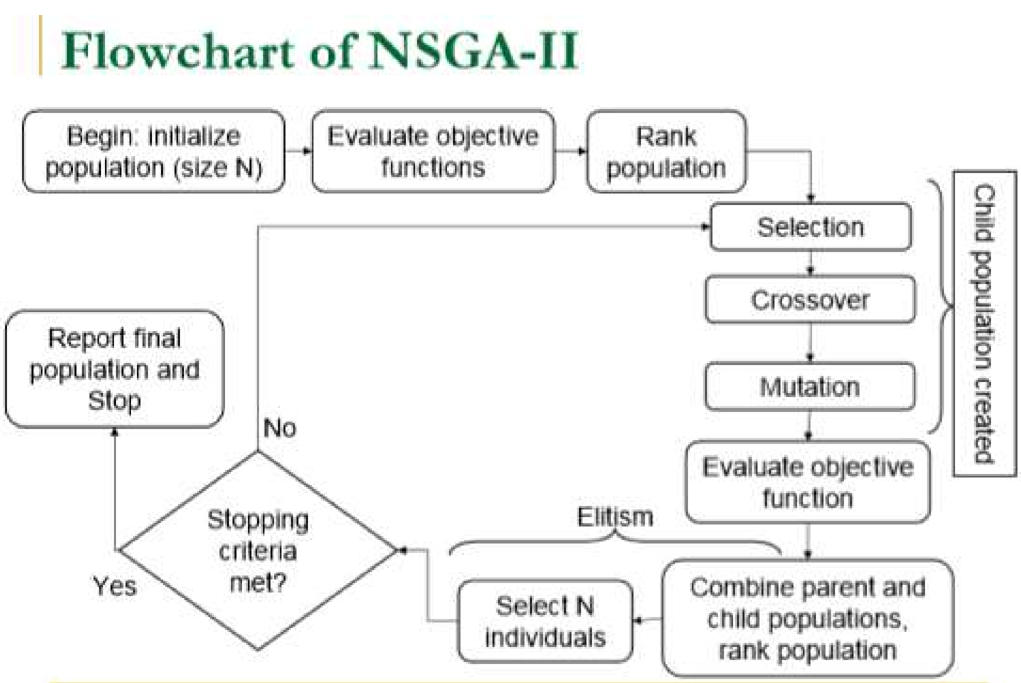
Step 2: Non dominated sort Sorting process based on non-domination criteria of the population that has been initialized.

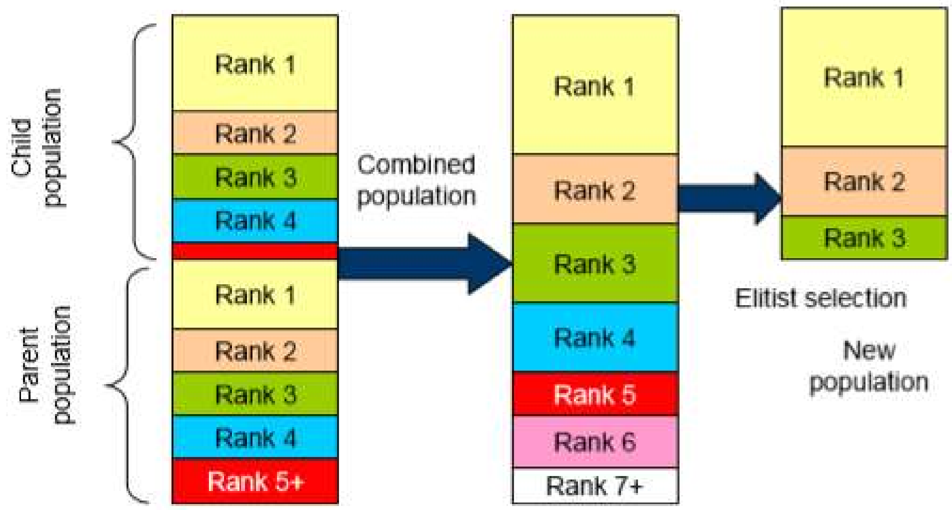
Step 3: Crowding distance Once the sorting is complete, the crowding distance value is assigning front wise. The individuals in population are selected based on rank and crowding distance.

Step 4: Selection The selection of individuals is carried out using a binary tournament selection with crowded-comparison operator.

Step 5: Genetic Operators Real coded GA using simulated binary crossover and polynomial mutation.

Step 6: Recombination and selection Offspring population and current generation population are combined and the individuals of the next generation are set by selection. The new generation is filled by each front subsequently until the population size exceeds the current population size.





Comparison:

Scoring	PSO trained ANN	NSGA-II trained ANN
Explained Variance	-2.220446049250313e-16	-11.762890
Max Error	0.045800000000000001	0.0033786286243828 18
Mean Squared Log Error	0.0009422440426689332	2.1603026001005774e-06
Mean Absolute Error	0.04534166666666667	0.0019020753898988153

CONCLUSION

1. Soil moisture data is a non-stationary time series, which presents a periodic variation regular pattern involving large fluctuations. It is known from correlation analysis that each parameter characteristic has a correlation with the moisture parameter, which affects the predicted value, and that the initial soil moisture feature has the greatest weight. Humidity and temperature are second. Although the rainfall variable directly affects the soil water content, its distribution is highly random and noisy, leading to a low weight factor that cannot be used as the only fitting parameter. Therefore, the seven input variables discussed in this paper were selected as the inputs of the prediction model.
2. Although the model has certain advantages in specific measures, the conditions are different for each model, and the composition of the data set and regional differences are difficult to eliminate.

BIBLIOGRAPHY:

1. https://www.iitk.ac.in/kangal/Deb_NSGA-II.pdf
2. https://ieeexplore.ieee.org/document/996017
3. http://www.swarmintelligence.org/tutorials.php
4. https://www.sciencedirect.com/topics/engineering/particle-swarm-optimization
5. https://www.google.com/search?q=soil+moisture+prediction+as+regression+problem&oq=soil&aqs=chrome.0.69i59j69i57j0l2.2361j0j7&client=ms-android-huawei&sourceid=chrome-mobile&ie=UTF-8
6. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0214508
7. https://www.sciencedirect.com/science/article/pii/S1877705811022466
8. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.max_error.html
9. https://scikit-learn.org/stable/modules/model_evaluation.html
10. https://towardsdatascience.com/applied-deep-learning-part-1-artificial-neural-networks-d7834f67a4f6
11. https://www.sciencedirect.com/topics/computer-science/evolutionary-algorithms
12. https://www.youtube.com/watch?v=T3804kTFhG8&list=PLbRMhDVUMnge0CU8o2gWwKXANADrpTc1q