ISSN: 2393-8528

IJICSE
Open Access

Contents lists available at www.ijicse.in

International Journal of Innovative Computer Science & Engineering

Volume 4 Issue 3; May-June-2017; Page No. 80-84

A Survey on Air quality Forecasting and Agricultural Intelligence using Data mining Techniques

A.P Christopher Arokiaraj[1], P.Kalaikannan[2]

[1] Assistant Professor, Department of Computer Science,

KG College of Arts & Science, Coimbatore-35,

christoperarokiaraj.ap@kgcas.com

[2]Assistant Professor, Department of M.Sc. Software Systems,

KG College of Arts & Science, Coimbatore-35,

kalaikannan.p16@gmail.com

Received 06 April 2017; Accepted 09 May 2017

ABSTRACT

Air pollution is gradually becoming a global environmental threat as a result of industrialization, globalization and urbanization. The quality of the air we breathe is becoming essential both for the environment as well to the society. There are numerous tools categorized as numerical and statistical for the prediction and analysis of quality of the air. In this problem area, Artificial Neural Network is considered to be an exceptional predictive and data analysis tool for Air quality forecasting. Agricultural intelligence is a definite field associated with techniques involved to an increased understanding of cultivation, productivity of crop, and reduced risk associated agriculture. Crop prediction is a vital agricultural drawback. To address this drawback, crop prediction technique is used. It is the one of the most typically used intelligence technique supported by data processing (DM) ideas to predict the crop yield for maximising the crop productivity. Hence, this paper focuses on a comprehensive review on existing air quality forecasting techniques through soft computing and studies and records the varied data processing techniques on the market within the literature for higher crop productivity.

Keywords: *Air quality, Artificial Neural Network, Fuzzy logic, Air pollutants, Software tools for ANN. Data mining, Crop prediction, k-means, k-nearest neighbor, Fuzzy sets, Regression, Classification, Neural network Association Rule.*

INTRODUCTION

Air Pollution as a whole is a global term and it can be defined as the existence of substances in atmosphere insufficient concentrations so that they are threaten to be harmful to human being, plant, animal life, and the most drastic damage to the environment and the climatic conditions. Further, it can be said that, air pollution destroys the comfortable delight of life, property and environment. In the current scenario of industrialization, the adverse effect on environmental problems is considered as an immense problem to both developing and developed countries. The vision of meeting world demands for the increasing the crop yield for population throughout the world is turning into a lot of necessary within the recent years. Crop model and call tools square measure progressively used in agricultural fields to improve production potency. The combination of advanced technology and agriculture used to improve the assembly of crop yield. The information mining techniques like classification, neural networks and regression square measure needed to apply on the realistic information sets for analyses and build the prediction on the agriculture crop yield.

The information mining tasks may be classified into 2 classes.

- Descriptive data processing.
- Predictive data processing.

Descriptive information mining:

Descriptive modelling is a mathematical method that describes universe events and relation-ships between factors accountable for them. The method is used by shopper driven by organizations to assist them target their selling and advertising efforts.

The main perspective of descriptive modeling includes:

**Corresponding author: A.P Christopher Arokiaraj |*

- Customer segmentation: Partitions a customer base into number of groups with various impacts on marketing and service.
- Value-based segmentation: It indicates and estimates the value of a customer to the organization.
- Behavior-based segmentation: It evaluates customer's product usage and purchasing patterns.
- Needs-based segmentation: It visualizes the ways to capitalize on motives that force customer behavior.

Predictive data mining: Predictive analytics is the procedure of extracting information from bulky sets in order to make forecast and analyses accurate results about future outcomes. The list of yield prediction models that more of them have been generally classified in two groups methods.

Traditional Approach

- Artificial Intelligence

II. UTILIZATION OF DATA MINING TECHNIQUES IN AGRICULTURE

Crop prediction is one of the modern agriculture intelligence techniques. It can be developed by using following methodologies.

Association Rule: Association rule mining technique is one of the most proficient techniques of data mining to explore hidden or preferred pattern among the enormous amount of data. In this method, the focus is on finding relationships between the dissimilar items in a database that deals with transactions. Association rules are used to find out elements that co-occur frequently inside a dataset consisting of many independent selections of elements, and to discover rules that occur within them. The simple problem statement is: Given a set of transactions, where each transaction is a set of literals, an association rule is a phrase of the form X => Y, where X and Y are sets of objects. The spontaneous meaning of such a rule is that transactions of the database which contain X tend to contain Y. An application of the association rules mining is the market basket analysis, customer segmentation, store layout, catalog design, and telecommunication alarm prediction.

Classification: Classification and forecasting are two forms of data analysis that can be used to extract models portraying important data classes or to predict future data trends. It is a formula in which a model understands to forecast a class label from a set of training data which can then be used to expect discrete class labels on new samples. To exaggerate the predictive accuracy obtained by the classification model while segregating examples in the test set hidden during training is one of the major goals of classification algorithm. Data mining classification algorithms can follow three different learning approaches namely:

Semi-supervised learning, supervised learning and unsupervised learning. The different classification techniques for discovering knowledge are Rule Based Classifiers, Bayesian Networks (BN), Decision Tree (DT), Nearest Neighbor (NN), Artificial Neural Network (ANN), Support Vector Machine (SVM), Rough Sets, Fuzzy Logic, and Genetic Algorithms.

Clustering: In clustering, the emphasis is on discovery of a partition of data records into clusters such that the points contained by each cluster are close to one another. Clustering groups the data occurrences into subsets in such a manner that analogous instances are assembled collectively, while contradictory instances belong to diverse groups. Since the aim of clustering is to discover out a new set of categories, the latest groups are of interest in themselves, and their evaluation is inherent. The different clustering methods namely Hierarchical Methods(HM), Partitioning Methods (PM), Density-based Methods(DBM), Model-based Clustering Methods(MBCM), Grid-based Methods and Soft computing Methods, Squared Error Based Clustering, network data and Clustering graph.

Regression: Regression is attainment of a function that relates data item to a real-valued prediction variable. The different applications of regression can be predicting the amount of minerals present in a forest, analyzing the chance of patient will survive or not on the set of his diagnostic tests, predicting consumer's demand for a new product that has been introduced. Here in this model it is trained to envisage an uninterrupted target. Regression tasks are often treated as classification tasks with quantitative class tag. The methods for prediction are Nonlinear Regression (NLR) and Linear Regression (LR).

Neural networks: It focuses the information about weather and are observed and stored. The recorded parameters are used to forecast weather. If there is a change in any one of the recorded parameters like wind speed, wind direction, temperature, rainfall, humidity, then the upcoming climatic condition can be predicted using artificial neural networks, back propagation techniques. The increase in signal range will work in large areas as well.

Fuzzy set: It describes its utilization in agriculture associated areas namely yield prediction is a very important agricultural problem. Any farmer might be interested in knowing how much yield is expected. In the past, yield prediction was achieved by considering

farmer's experience on particular field, crop and climate condition. We have discussed additional information about data like probability in probability theory, grade of membership in fuzzy set theory.

Decision tree and Bayesian classification:

The findings of the study reveals that the choice tree analysis indicated that the productivity of barley crop was principally influenced by comparative wetness followed by temperature and precipitation. the choice tree analysis shows that the productivity of paddy crop was principally inclined by precipitation followed by comparative Evaporation and wetness. the principles fashioned from the choice tree area unit helpful for characteristic the conditions supposed for top or low crop productivity. Bayesian network could be a powerful tool and broadly speaking employed in agriculture datasets. The model developed for agriculture application supported the Bayesian network learning methodology. The results show that Bayesian Networks area unit possible and economical.

III.AIRPOLLUTION

The recent advancements in technology and fast amendment publicly requirements/needs ends up in manufacture, ends up in pollution, becomes major concern to be foretold and several other management methods/strategies ought to be undertaken to deal with the adverse effects of pollution. The scarce quantity of drugs that square measure side to the atmosphere either through artificial or natural processes, referred to as air pollutants. the subsequent square measure the main originators of air pollutants:

•Indoor air pollutants
•Out-of-doors air pollutants

A. Indoor pollution

In Asian country and alternative developingcountries, giant population still depends on bio fuels for cookery and heating, ensuing indoor pollution. Indoor air pollutants causes' major health disorders to kinsmen United Nations agency live for a chronic amount of your time in a much engorged setting.

B.Out-of-doorpollution

The out of doors pollution is formed primarily owing to cars and numerous industries. The out of doors air pollutants square measure accountable for the climatical result alongside the result on the environmental inequality to social life. the subsequent square measure the classes of air pollutants:

•PrimaryPollutants.
•SecondaryPollutants.

1.Primarypollutants:

Primary Pollutants results from the combustion of fuels and industrial operations.

2.Secondarypollutants:

Secondary pollutants square measure those that square measure made owing to the reaction of primary waste matter within the atmosphere.

IV. METHODS FOR PREDICTING POLLUTION

The pathways for the transport and transformation of matter inside four categorical aras that compose planet Earth are known as as Bio-geo-chemical cycle. These cycles governs the functioning of the planet. the planet is receptive radiation from the sun and space, however may be a just about closed system with respect to matter. Thus, the matter self-enclosed inside the planet from the time of its birth is remodeled and circulated geographically. Bio-geochemical cycles ar natural cycles however as a result of differing types of physical likewise as chemical world effects, these cycles ar affected. With the speedy changes in these cycles the atmosphere close the planet, liable for the existence of life on the earth is affected that reciprocally affects the life on Earth. as a result of such adverse effects, forecasting, AN application of Science and Technology is employed to predict the state of the atmosphere for various locations. Air quality models play an important role all told aspects of pollution management and air quality coming up with, wherever prediction may be a major element. Air quality forecasts give the general public with air quality data that permits individuals to require preventative measures to avoid or limit their exposure to unhealthy levels of pollution.

A.EFFECTS OF POLLUTION

As we know, the occurring life existence on Earth is in peril owing to fast increase of varied effects like warming, several health disorders that square measure found in physique in addition in animals. Even the food, fruits and vegetables we tend to eat isn't safe, it conjointly get full of numerous air pollutants. Thus, the study shows that the kinsfolk, animals, plants and therefore the setting encompassing them square measure all full of the pollution and its harmful pollutants

V. LITERATURE REVIEW

The prediction of air quality is turning into crucial for minimizing the environmental imbalances additional effectively addresses the pollution. There area unit completely different kind of numerical also as applied math tools for the prediction and analysis of pollution. The emergence of advanced computing/analysis techniques from ancient computing strategies to recent soft computing techniques area unit effectively

addresses the air quality prediction. The traditional approach for air quality prediction uses mathematical and applied math techniques. In these techniques, ab initio a physical model was designed then knowledge is coded with mathematical differential equations. however such strategies suffers from disadvantages like they supply restricted accuracy as they were unable to predict the acute points i.e. the pollution most and minimum cut-offs can not be determined exploitation such approach. Also, such strategies were drawn-out and inefficient approach for higher output prediction. however with the advancement in technology and analysis, another to ancient strategies has been planned i.e. AI (AI) techniques are often used for prediction functions. Among varied styles of soft computing techniques, the subsequent area unit the main pollution prognosticative model techniques.

A. Artificial Neural Networks (ANN).
B. Support Vector Machines (SVM).
C. Fuzzy Logic (FL)D.Hidden mathematician Model (HMM).
E. Genetic formula.
F. Particle Swarm Intelligence.
G. Hybrid soft computing techniques.
A. ARTIFICIAL NEURAL NETWORKS
With the pioneering work of McCulloch & Pitts, Artificial Neural Networks (ANN) has its roots in wide knowledge domain history from the first 1940's. ANN raised as a mechanism to mimic the human's brain processes. ANN is Associate in Nursing intelligent system that has the capability to be told, hit the books and make relationships among the information. ANN is formed up by the easy process units, the neurons, that area unit connected in an exceedingly network by an outsized variety of weighted links wherever the noninheritable data is keep and over that signals or data will pass. The prediction of air quality, effectively self-addressed by the prediction of assorted air pollutants like Sulphur, Nitrogen, CO, ozone, suspended stuff (SPM) by divided the information set into coaching , validation and verification more simulation victimization ANN. ANN was effectively addresses the prediction of Sulphur oxide distribution and therefore the future concentration within the air by modeling the Sulphur oxide concentration and its distribution from the pollution station.

B.SUPPORT VECTOR MACHINES (SVM)

Support vector machines (SVMs, additionally support vector networks ar supervised learning models with associated learning algorithms that analyze knowledge and acknowledge patterns, used for classification and multivariate analysis. the essential SVM takes a group of computer file and predicts, for every given input, that of 2 potential categories forms the output, creating it a non-probabilistic binary linear classifier. The SVM model provides a promising various and advantageous in times series knowledge analysis for predicting the extent of air pollutants. The potential of applying SVM model in close air waste matter prediction studied and projected as a most promising approach in prediction of PM10 waste matter.

C. FUZZY LOGIC

The term "fuzzy logic" was introduced with the proposal of fuzzy pure mathematics may be a type of many-valued logic, deals with reasoning that's approximate instead of mounted and actual. mathematical logic deals with reasoning and provides a more robust summary within the type of rules that defines all the conditions that area unit needed for predicting the pollution prediction. In sugarcane process business, mathematical logic may be accustomed classify and quantify levels of pollution as poor, ordinary, superb and glorious. The Mamdame fuzzy reasoning system provides the results for prediction of the air quality in and round the sugarcane business.

D. HIDDEN-MARKOV MODEL (HMM)

A Hidden Andre Markoff model (HMM) may be a classical approach for statistic analysis and prediction. A HMM is predicated on the relationships between the attributes of specific information things and an information set. Hidden Andre Markoff Model (HMM), a probabilistic operate of a Markoff chain, permits the prediction of PM2.5, victimisation the meteorologic measurements and its observation levels.

E. GENETIC FORMULA

Genetic algorithmic program is predicated on Darwin's Theory. It begins with whimsical created individual population so fitness is evaluated and fogeys ar selected from the people. Genetic Algorithms effectively addresses the modification of the buildup of the encompassing atmosphere and prediction of the thickness of the air pollutants. Genetic Algorithms ar expeditiously applied to extract the optimum feature set of an outsized info containing waste concentration measurements, and feeds to a nearest neighbor algorithmic program so as to predict the daily most concentration for pollutants.

F. PARTICLE SWARM INTELLIGENCE

Particle Swarm Intelligence may be a inhabited search methodology that resembles a faculty of flying birds. Particles square measure candidate solutions to

drawback in hand. Every particle adjusts its flying per its own flying expertise and its companion's flying expertise. A PSO adapted to coach multi-layer perception to predict air quality parameters additional effectively. Particle swam improvement algorithmic program devised with the characters of pellucidprinciple and physical explication, to guage the grade of atmospherical pollution and multi pollutants.

G. HYBRID SOFT COMPUTING TECHNIQUES

The combination of over one soft computing techniques forms Hybrid soft computing technique. variety of hybrid soft computing techniques applied in assessment of air quality prediction expeditiously. associate degree hybrid soft computing technique with the mixture of ANN along side mathematical logic or with HMM and even with alternative soft computing techniques will be terribly effective for pollution prediction and statistic analysis. A Particle Swarm improvement based mostly neural network takes a singular quite improvement algorithmic rule i.e. PSO algorithmic rule to coach the multi-layer perceptrons. The complicated neural network with biological process algorithmic rule devised, rtNEAT (real-time Neuro-Evolution of Augmenting Topologies), to handle the air quality prediction additional effectively.

VI.CONCLUSION

This paper has been conferred with analysis potential for the appliance of knowledge mining methodologies or approaches to the issue of yield prediction in agriculture. From the study, it's been determined that there area unit a large variety of applications of knowledge mining techniques in agriculture. it's clearly notable that there area unit many data processing approaches within the literature to enhance the crop productivity within the areas involved. For this, several researchers have enforced data processing approaches like K-Means algorithmic program, K Nearest Neighbor, Support Vector Machines, Neural Network, Regression techniques like liner and multiple correlation techniques. it's positive that international development through the facility of agricultural intelligence is feasible by reducing the knowledge gap in agriculture. A study was additionally been meted out on numerous Air Pollution/quality prediction techniques with most rising soft computing techniques. it's determined the most causes for pollution and therefore the factors which will be accountable to reduce it. In Pollution prediction, soft computing techniques plays important role. more we have a tendency to area unit about to designate Hybrid soft computing technique supported ANN with formal logic for predicting the pollution for a specific town.

REFERENCES

1. Yethiraj N G, Applying Data Mining Techniques In The Field Of Agriculture And Allied Sciences, International Journal of Business Intelligents ISSN: 2278-2400, Vol 01, Issue 02, December 2012.
2. Ramesh D, Vishnu Vardhan B., Data Mining Techniques and Applications to Agricultural Yield Data , IJARCCE, Vol. 2, Issue 9, September 2013.
3. Mucherino, A., Papajorgji, P., & Pardalos, P. (2009), —Data mining in agriculture (Vol. 34), Springer.
4. Srikant, R V Q & Agrawal, R (1997, August), Mining Association Rules with Item Constraints. In KDD (Vol. 97, pp. 67-73).
5. Xu, R &Wunsch, D (2005), Survey of clustering algorithms , Neural Networks, IEEE Transactions on, 16(3), 645-678.
6. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (2011), from data mining to knowledge discovery in databases . AI magazine, 17(3), 37.
7. K. Verheyen, D. Adriaens, M. Hermy, and S. Deckers, High resolution continuous soil classification using morphological soil profile descriptions , Geoderma, vol. 101, pp. 31-48, 2011. International Journal of Computer Science and Technology IJCST Vol. 2, Issue 1, March 2011.
8. BindhuLal, Sanjaya Shankar Tripathi, "Prediction Of Dust concentration in open cast coal mineusing artificial neural network." Atmospheric Pollution Research 3 (2012) 211-218.
9. Sun W, Zhang H, "Prediction of 24-hour-average PM (2.5) concentrations using a hidden Markov model with different emission distributions in Northern California." Sci Total Environment, 2013.
10. Limin Wang,A PSO-based Model for Assessment of Atmospheric Quality, IJCSNS International Journal of Computer Science and
11. Network Security, VOL.6 No.3A