An Efficient Model for Sugarcane Yield Prediction

Sumathi A, Meganathan S

Department of Computer Science, Sastra University, Srinivasa Ramanujam Centre, Kumbakonam 612001, Tamil Nadu, India.

Abstract

Objective Sugarcane is one of India's most basic renewable business crops. Yield prediction is one of the important problems in agricultural field. Data mining techniques are better choice to solve this problem. Using different data mining techniques increase the sugarcane yield performance. Method Enhanced optimization technique that is Particle Swarm Optimization (PSO) used for feature selection to increase the sugarcane performance and different classification techniques are used for early stage of yield prediction. Result Comparing the different algorithms such as K-Nearest Neighbour, Random Forest, Support Vector Machine, best model was accomplished with Support Vector Machine with elements choose by Particle Swarm Optimization (PSO). Conclusion: This model works as per the dataset collected monthly and seasonal wise. In future, use hourly wise weather data to predict the sugarcane yield in advanced stage.

Keywords: Sugarcane Yield, Data Mining, Particle Swarm Optimization.

Introduction

Data Mining is the process of extracting useful and important information from large sets of data.

The datasets are preprocessed with mean and null values to increase the classification accuracy.

In agriculture field yield prediction is very important problem to overcome this problem using data mining techniques.

In this work feature selection performed with Particle Swarm Optimization (PSO) to increase the classifier efficiency.

In the past, yield prediction was performed by considering farmers experience on particular field and crop [1].

In this system, using PSO evaluated fitness values explore the search space of a given problem to find the settings or parameters required to maximize a particular objective based on the fitness value it select the relavent features based on target data like PH, Ca , Na wind speed, sunrise, station pressure.

In any of Data Mining procedures the training data is to be collected from historical data and the gathered data is used in terms of training which has to be exploited to learn how to classify future yield predictions. Empirical models are conceptually simpler models and are based on relationships between crop outputs and its driving factors, water availability, weather conditions, and agricultural practices.

In this paper, three techniques were used to tackle this problem using Support Vector Machine (SVM), Random Forests (RF), K nearest Neighbour (KNN) [2].

Figure-1 Shows the proposed architecture of yield prediction. Yield prediction involves combination of real time weather and soil datasets.

In this work predict the early stage of sugarcane yield prediction using various data mining techniques based on feature selection and classification algorithms.
The pieces of information are chosen utilizing selected features and different classification techniques are helpful to predict the sugarcane yield performance, this may increase the performance level and compared to the final result which is the best one to increase the sugarcane yield accuracy.

Materials and Methods

Upload the Dataset


Data Preprocessing

The quality of data affects the data mining results. Real world data are also dirty, so using preprocessing to improve accuracy. In this module is used to cleaning the irrelevant, missing or noisy data from the given dataset.

Feature Selection

Feature selection is selecting the important features that are ideally necessary and sufficient to describe the target data. Feature selection can serve as a pre-processing tool of great importance before solving the classification problems. A good feature selection method can reduce the cost of feature measurement, and increase classifier efficiency and classification accuracy. In this system feature selection performed with Particle swarm Optimization (PSO) is one of the optimization technique. The particle swarm optimization extracted features using stored input data set and passed to the output is classifier generator.

Classification

In this module classifying the dataset based on high, low, medium depends on soil type, chemical properties, and water holding capacity, rainfall and so on. The classifier generates by using the Random Forest, Support Vector Machine and K Nearest Neighbour algorithms. In this module, implement classification algorithm to classify the data, finally predict the yield production [7, 8, 9].

Performance Evaluation

At last of classification predict whether the classification algorithm is more efficient to produce maximum yield prediction on sugarcane [5, 6] better than others.

Particle Swarm Optimization (PSO)

PSO adapts this behaviour and searches for the best solution in the search space. A single solution is called particle. Each particle has a fitness/cost value that is evaluated by the function to be minimized, and each particle has a velocity that directs the "flying" of the particles. The particles fly through the search space by following the optimum particles. The algorithm is initialized with particles at random positions, and then it explores the search space to find better solutions. Every iteration, each particle adjusts its velocity to follow two best solutions. The first is the cognitive part, where the particle follows its own best solution found so far. This is the solution that produces the lowest cost (has the highest fitness). This value is called pBest (particle best). The other best value is the current best solution of the swarm, i.e., the best solution by any particle in the swarm. This value is called gBest (global best).

PSUEDO Code for PSO

For each particle
{Initialize particle}

Do until maximum iterations or minimum error criteria

{For each particle

{Calculate Data fitness value

If the fitness value is better than pBest (personal best)

{Set pBest = current fitness value

If pBest is better than gBest (global best)

{Set gBest = pBest}

} For each particle

{Calculate particle Velocity

Use gBest and Velocity to update particle Data}

The procedures are,

- Initialize each particle with a random velocity and random position.
- Calculate the cost for each particle. If the current cost is lower than the best value so far, remember this position (pBest).
- Choose the particle with the lowest cost of all particles. The position of this particle is g best.
- Calculate, for each particle, the new velocity and position according to the above Equations.
- Repeat steps 2-4 until maximum iteration or minimum error criteria is not attained.

Results

K-Nearest Neighbour (KNN) Classifiers

K nearest neighbours is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). Figure-2 shows the yield prediction of k nearest neighbour classification. It works based on minimum distance from the query instance to the training samples to determine the K-nearest neighbour. The selected features passed to KNN classification it groups the features based on similarity measure and to predict the yield performance in the given dataset.

![Fig.2: Yield prediction of K nearest neighbour](image)

Based on that features it categories the yield in 3 levels like low, normal, high depends on feature selection method KNN classifies the data like sky condition, pressure, PH, Calcium (Ca), potassium (K).

Random Forest Classifiers

It is an ensemble classifier that consists of many decision trees and outputs the class that is the mode of the class's output by individual trees. Figure-3 shows the yield prediction of random forest classification. In a normal decision tree, one decision tree is built and in a random forest algorithm number of decision trees are built during the process.

A vote from each of the decision trees is considered in deciding the final class of a case or an object, this is called ensemble process. Since, many decision trees are built and used in a process of Random Forest algorithm.
To classify a new object from an input vector, put the input vector down each of the trees in the forest. Each tree gives a classification, and we say the tree "votes" for that class. The forest chooses the classification having the most votes (over all the trees in the forest). Based on that features Random forest construct many trees until get a final result and also categories the yield prediction in 3 levels like low, normal, high depends on feature selection method KNN classifies the data like sky condition, pressure, PH, Calcium (Ca), potassium (K).

**Support Vector Machine Classifiers**

It is a discriminative classifier formally defined by a separating hyperplane or given labeled training data the algorithms outputs an optimal hyperplane which categories new examples. Figure-4 shows the yield prediction of support vector machine classification.

Based on that features SVM construct hyperplane based on that margin it divides the dataset and also categories the yield prediction in 3 levels like low, normal, high depends on feature selection method KNN classifies the data like sky condition, pressure, PH, Calcium (Ca), potassium (K).

**Discussion**

Classify the dataset based on the three different algorithms such as K nearest neighbour, Random Forest, Support vector machine. At last of classification model, compare the three different classifications which is more efficient to produce maximum yield prediction on sugarcane better than others. Figure-5 shows the comparison of three classification algorithms.
Conclusion

This work shows the prediction of sugarcane yield model. Now a day, the farmers are not getting expected crop production yield due to unaware of climate changes and its conditions, rainfall variability. For getting expected crop yield, farmers need advices for predicting and analysing future crop production, so that farmers can easily take a decision before any crop production. This paper proposes how to make agriculture well organized by predicting and thus improve the crop yields by using soil and weather information. This work implement the PSO technique for feature selection and making predictions from vast amount of data with different classification algorithms to enhance the accuracy level in early stage of prediction. This work does not require any domain knowledge to the end user and easy for analyzing the result. This model works as per the dataset collected monthly and seasonal wise. In future, use hourly wise weather data to predict the sugarcane yield in advanced stage.

Abbreviations

SVM-Support Vector Machine
KNN-K nearest neighbour
RF-Random Forest

References

1. D Ramesh 1, B Vishnu Vardhan 2 Analysis of crop yield prediction using data mining techniques”, International Journal of Research in Engineering and Technology.