ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 9(33), DOI: 10.17485/ijst/2016/v9i33/84153, September 2016* ISSN (Online) : 0974-5645

A Novel Fuzzy-Genetic Hybrid Classification Algorithm for Soil Profile Data

D. Ashok Kumar[1*] and N. Kannathasan[2]

[1]Department of Computer Science, Government Arts College, Tiruchirapalli - 620 022, Tamil Nadu, India; akudaiyar@yahoo.com

[2]Department of Computer Science, Kanchi Mamunivar Centre for Post Graduate Studies Puducherry – 605 008, U.T. of Puducherry, India; nkthasan@gmail.com

Abstract

Background/Objectives: Fuzzy Logic is derived from Fuzzy Set that deals with reasoning that is accurate rather than precisely based on degrees of membership, and are well-liked tools in the application of prediction, classification and recognition based problems. **Methods/Statistical Analysis:** This work projected a novel Fuzzy Classification Algorithm to train Fuzzy Inference System to classify and predict soil profile data with soil Total Porosity and proposed a fuzzy-genetic hybrid based clustering algorithm namely GENetic algorithm for SOIL (GENSOIL) profile data. Soil samples were collected from International Soil Reference and Information Centre. **Findings:** Fuzzy Rule Base has been developed and using in a proposed Fuzzy Classification Algorithm to train Fuzzy Inference System for soil profile data classification and prediction. The proposed algorithm classified the soil samples based on Fuzzy Rule Base using with Fuzzy Membership Function and hybrid with a randomized and optimized Genetic algorithm. A Novel Fuzzy-Genetic Hybrid Classification Algorithm compared with k-Means and Fuzzy C-Means are assessed on the basis of the time complexity of clustering. **Applications/ Improvements:** The advantage of novel Fuzzy-Genetic Hybrid Classification Algorithm is their applicability in different types of related optimization problems with a superior speed of calculation and found solution very close to the best one.

Keywords: Classification, Fuzzy C-Means, Fuzzy Logic, Fuzzy Rule Base, Genetic, K-Means, Soil Profile

1. Introduction

In developing countries, hunger is forcing people to cultivate land that is not suitable for farming. With the growing pressure of urbanization, it is going to be a challenge for farmers to produce food for the whole humanity with less land and water resources. In addition, growth of the plant depends upon multiple factors such as soil type, crop type, and weather. Due to lack of information about the growth of the plant and expert advice, most of the farmers fail to get good yields. Most knowledge about soil comes from soil survey efforts. Initially, the knowledge extraction was computed and evaluated manually using statistical techniques. Subsequently, semi-automated data mining techniques emerged because of the advancement in the technology. Such advancement was also in the form of storage that increases the demands for analysis. In such case, semi-automated techniques have become inefficient. Therefore, Soft Computing approaches were introduced to synthesis knowledge efficiently. The benefits of a greater understanding of soil pattern could improve productivity in farming to maintaining biodiversity, to reduce reliance on fertilizers and create a better integrated soil management system for both private and public sectors[7].

Soil supports all forms of life on the planet and plays a vital role for their existence. Soil quality influences crop-yield. Soil provides minerals and water to the plants. Characteristics of soil such as fertility, depth, texture and

**Author for correspondence*

structure govern the type of plants and crops that can be grown in any region[7]. The objectives of soil classification are: 1. Providing information needed for developing plans for effective usage of lands bringing new areas under irrigation and evaluating the suitability of soils for irrigation of agricultural crops. 2. Demarcating the soils with problems such as alkali-soils, waterlogged soils, eroded soils, and waste lands and suggesting soil and water conservation measures to overcome these problems. 3. Providing basic data for land settlement issues rehabilitation, tax appraisal, to locate and design highways, airports and other engineering structures, and public sanitation works.

The basic aim of Soft Computing is to build intelligence into the machine. In order to build intelligent machine some elements or attributes of intelligence must be identified. Soft Computing is not a single methodology. Rather, it is a coalition (or) consortium of distinct methodologies. The purpose of new approaches is to reduce the complexity of the system while maintaining its full utility and hence to reduce the modeling time on computational cost. Soft Computing does not perform much symbolic manipulation, so it is a complimentary approach to conventional Artificial Intelligence approaches[1].

This paper proposes an innovative model named as GENetic Algorithm for SOIL profile data. It (GENSOIL) is a randomized search and optimization technique that is guided by the principles of genetics and evolution, which has a large amount of implicit parallelism. On the basis of the result drawn by this experiment, it may be stated that the proposed GENSOIL algorithm is less time consuming that k-Means and Fuzzy C-Means algorithm and therefore it is superior and suitable for classification of soil profile data. One of the advantages of GENSOIL is its applicability to different types of related optimization problems with a high calculation speed and getting optimal values. The test dataset used in this research paper consists of 705 samples of particular region as in Table 1. The soil profile dataset used in this work was collected from World Soil Information-International Soil Reference and Information Centre[8].

Table 1. List of soil variables, their abbreviations and units of measurement

Abbreviation	Description	Units
PHH_2O	Soil reaction in water	pH units
SAND	Sand	%(mass)
SILT	Silt	%(mass)
CLAY	Clay	%(mass)
TOTPOR	Total Porosity	%

2. Proposed Fuzzy Classification Algorithm

The proposed Fuzzy Classification algorithm for soil profile data takes the soil major textural classes and relative percentage of total porosity as input in Fuzzy Inference System (FIS). The soil textural class includes sand, silt, and clay. The proposed system fuzzifies these values based on fuzzy membership functions. The fuzzified values are passed to the fuzzy inference engine where IF-THEN fuzzy rules are applied to get a fuzzy output. The output is passed through defuzzifer to get crisp value. A new surface is generated which provides us the pattern of the relative contaminated percentage of the soil area.

Algorithm 1: Fuzzy Classification for Soil Profile Data.
Input: Create FIS, The training sample of Sand, Silt, Clay, pH, Total Porosity.
Output: A FIS trained to classify the soil samples with Total Porosity.
Step 1. FIS=Create Fuzzy Inference System
Step 2. TexturalClass=Read Soil texture with pH, Total Porosity
Step 3. TotPoro=Totper(TexturalClass)
Step 4. FuzzyPoro=Fuzzify(TotPoro)
Step 5. Call Fuzzy Fitness Rules
Step 6. FuzzyOutputPoro=FuzzyInferenceEngine(FuzzyPoro)
Step 7. CrispOutput=Defuzzify(FuzzyOutputPoro)
Step 8. TotPoro(TexturalClass)=CrispOutput
Step 9. Display TotPoro

2.1 Fuzzy Fitness Rules

- If (Sand is high) ^ (Clay is low) ^ (Silt is low) ^ (totpor is low) ^ (pH is low) → (SoilType is Sand)(Field is Sunflower)
- If (Sand is high) ^ (Clay is low) ^ (Silt is low) ^ (totpor is medium) ^ (pH is low) → (SoilType is Sand)(Field is Potato)
- If (Sand is high) ^ (Silt is low) ^ (Clay is low) ^ (totpor is medium) ^ (pH is medium) → (SoilType is Sand)(Field is Groundnut)

- If (Sand is low) ^ (Silt is low) ^ (Clay is medium) ^ (totpor is low) ^ (pH is high) → (SoilType is Clay) (Field is Paddy)
- If (Sand is low) ^ (Silt is low) ^ (Clay is high) ^ (totpor is high) ^ (pH is high) → (SoilType is Clay)(Field is Paddy)
- If (Sand is medium) ^ (Silt is low) ^ (Clay is high) ^ (totpor is medium) ^ (pH is high) → (SoilType is Clay)(Field is Paddy)
- If (Sand is low) ^ (Silt is low) ^ (Clay is high) ^ (totpor is high) ^ (pH is medium) → (SoilType is Clay)(Field is Paddy)
- If (Sand is low) ^ (Silt is high) ^ (Clay is low) ^ (totpor is medium) ^ (pH is medium) → (SoilType is Silt) (Field is Wheat)
- If (Sand is low) ^ (Silt is high) ^ (Clay is low) ^ (totpor is high) ^ (pH is high) → (SoilType is Silt)(Field is Wheat)
- If (Sand is low) ^ (Silt is medium) ^ (Clay is low) ^ (totpor is medium) ^ (pH is medium) → (SoilType is Loam)(Field is Maize)
- If (Sand is low) ^ (Silt is high) ^ (Clay is low) ^ (totpor is medium) ^ (pH is medium) → (SoilType is Loam) (Field is Wheat)
- If (Sand is low) ^ (Silt is high) ^ (Clay is medium) ^ (totpor is medium) ^ (pH is medium) → (SoilType is Loam)(Field is Soyabean)
- If (Sand is medium) ^ (Silt is low) ^ (Clay is medium) ^ (totpor is medium) ^ (pH is medium) → (SoilType is ClayLoam)(Field is Chilli)
- If (Sand is medium) ^ (Silt is low) ^ (Clay is low) ^ (totpor is low) ^ (pH is medium) → (SoilType is SandyLoam)(Field is Cotton)
- If (Sand is medium) ^ (Silt is low) ^ (Clay is low) ^ (totpor is medium) ^ (pH is medium) → (SoilType is SandyLoam)(Field is Groundnut)
- If (Sand is high) ^ (Silt is medium) ^ (Clay is low) ^ (totpor is medium) ^ (pH is medium) → (SoilType is SandyLoam)(Field is Cotton)
- If (Sand is high) ^ (Silt is low) ^ (Clay is low) ^ (totpor is low) ^ (pH is low) → (SoilType is LoamySand) (Field is Cotton)
- If (Sand is high) ^ (Silt is low) ^ (Clay is low) ^ (totpor is low) ^ (pH is low) → (SoilType is Sand)(Field is Pigconpea)
- If (Sand is high) ^ (Silt is low) ^ (Clay is low) ^ (totpor is medium) ^ (pH is medium) → (SoilType is Sand) (Field is Pigconpea)
- If (Sand is high) ^ (Silt is low) ^ (Clay is low) ^ (totpor is high) ^ (pH is high) → (SoilType is Sand)(Field is Pigconpea)
- If (Sand is low) ^ (Silt is low) ^ (Clay is high) ^ (totpor is high) ^ (pH is high) → (SoilType is Clay)(Field is Pigconpea)

Fuzzy Fitness Rules are derived based on USDA Particle Size and Porosity Ranges for Sand, Silt and Clay. The proposed Fuzzy Classification Algorithm for Soil Profile Data predicts the suitable crop for the classified soil that develop the agriculture in an efficient way to establish the human development. Fuzzy Rule base in a Fuzzy Inference System for soil classification in data mining takes the soil major textural classes and relative percentage of total porosity as input. The soil textural class includes sand, silt, clay, total porosity and pH value. This classification system fuzzifies these values based on fuzzy membership functions. The fuzzified values are passed to the fuzzy inference engine where the IF-THEN fuzzy rules are applied to get a fuzzy output. The output is passed through defuzzifier to get crisp value[2]. Then it displays the classified soil and its suitable product for the soil.

Soil classification system is developed using mamdani fuzzy model. It consists of five inputs, sand, silt, clay, total porosity and pH value and it provides the soil to cultivate the suitable crop yield according to the FIS rules. The system has two crisp output that controls the crop yield and soil type. The soil is taken to be in ranges of 0 to 100 respectively. The membership functions have three fuzzy values and they are low membership functions, medium membership functions, and high membership functions and these functions are find out the solutions using fuzzy based rules to yield the suitable crop shown in Figure 1. The proposed system fuzzifies these values based on fuzzy membership functions and has a range of 0 to 1. They have low, medium, high and it have a ranges of soil classification. It classified the soil profile in the membership functions based on the rule based system shown in Figure 2. Fuzzy rules are also called rule-based model according to the rule model, inserted the five attributes they are sand, silt, clay, total porosity and pH values to find the rule base for crop yield it have (0,1) fuzzy logic. Construction of rule base is done to control the output variable. It is a simple IF-THEN rule with a condition is used soil classification and sample fuzzy rules for the soil classification.

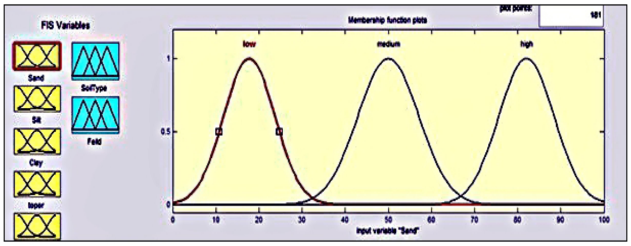

Figure 1. Membership functions for input values.

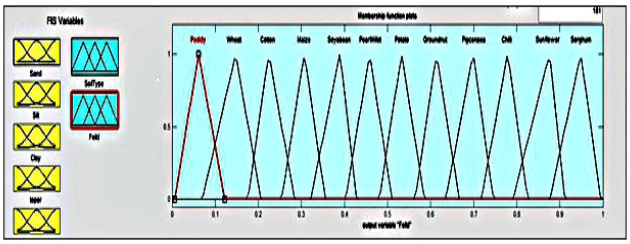

Figure 2. Membership function for output values.

3. Proposed Fuzzy-Genetic Hybrid Classification Algorithm and Experimental Analysis

The basic genetic algorithm is described by Goldberg[3] and is used in Fuzzy-Genetic hybrid algorithm to illustrate the basic components.

GENSOIL (**GEN**etic algorithm for **SOIL** Profile Data):

Step 1. The percentage of Sand, Silt, Clay and Total Porosity (totpor) and pH are represented as the chromosome.

Step 2. Under supervised mode the chromosome resembles a classifier structure.

Step 3. Select the better individual with filter solutions which are more likely to be selected.

Step 4. Apply selection method for selection.

Step 5. Build the next generation that is yielded using crossover or mutation by making changes in the characteristics of the chromosome.

Step 6. The Classifier classifies the chromosome on rule-based event with respect to sand, silt, clay, totpor and pH.

Step 7. The new off string is tried with the Fuzzy Inference System and checks whether it yields Best Soil Solution.

Step 8. Find the worst case that yields the population which newer should occur.

Step 9. Terminate with good population results.

3. 1 Implementation of K-Means Clustering Algorithm with Silhouette Plot

The Silhouette plot of a clustering with k=12 of the 705 numbers of chemical properties of soil profile data shown in Figure 3 displays a measure of how close each point in one cluster is to points in the neighboring clusters. k-Means, partitioning based clustering algorithm is applied to study soil surface data properties using k-means with Euclidean Distance. All the clusters are represented by silhouettes, that indicating objects are well-separated from neighboring clusters[4]. k-Means repeats the clustering process starting from different randomly selected centroids for each replicated[6] tabulated in Table 2. The final solution that k-Means returns is the one with the least total sum of distances, among all the replicates. The sum of distances within each cluster for the best solution is 13566.1. The total elapsed time is 0.281008 seconds.

3.2 Implementation of Fuzzy C Means Algorithm

Hard computing k-Means algorithm considers the objects of partition matrix using two-valued logic. Thus,

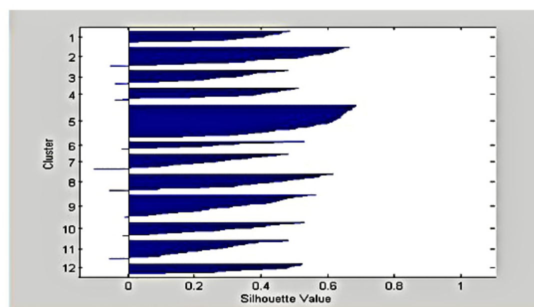

Figure 3. Silhouette plot.

Table 2. Sum of distances for different randomly selected centroids

No. of Iterations	Sum of Distances
7	13574.5
8	13572.4
13	13566.2
14	13566.1
19	13563.2
21	13566.1
26	13566.1
15	13566.1
18	13566.1

it causes a data object crisply either belonging to a cluster or not. Fuzzy C-Means algorithm was presented by Bezdeck as the extension of k-Means algorithm with the advantages of Fuzzy Set concepts. Fuzzy C-Means algorithm permits a data object's belonging to one or more clusters utilizing degrees of membership concept[5]. The Fuzzy C-Means takes a data set and a desired number of clusters and returns optimal cluster centers and membership grades for each data object is shown in Figure 4. Fuzzy C-Means results a total objective function equals to 46322.211442. The total elapsed time is 0.745535 seconds.

3.3 Implementation of Fuzzy-Genetic Hybrid Classification Algorithm (GENSOIL)

Genetic Algorithms motivated by natural evolution[6], make use of evolutionary operators and a population of solutions to obtain the globally optimal partition of the data objects. The output of SOIL illustrates in Figure 5. Each of these replicates began from a different set of initial iterations. The sum of distances within each cluster for that Best Solution is 0.0064539 and the total elapsed time is 0.0529.

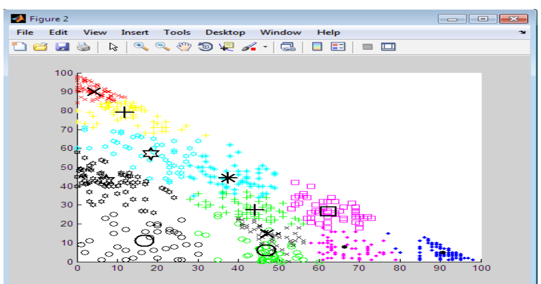

Figure 4. Fuzzy clustering with 12 clusters for 100 iterations.

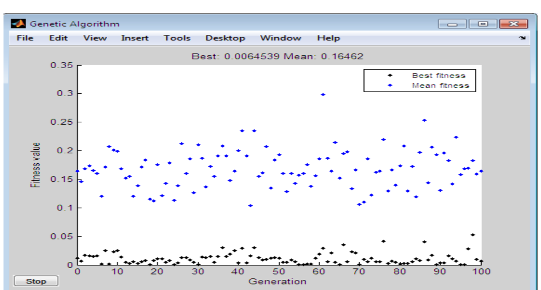

Figure 5. GENSOIL clustering fitness value for 100 iterations.

Table 3. Comparative analysis of k-Means, Fuzzy C-Means and GENSOIL algorithm

Algorithm	Elapsed Time (in Seconds)
k-Means Algorithm	0.281008
Fuzzy C-Means Algorithm	0.745535
GENSOIL (**GE**Nenetic algorithm for **SOIL** profile data)	0.0529

3.4 Comparison of Time Complexity of k-Means, Fuzzy C-Means and GENSOIL Algorithm

Hybrid systems employ more than one technology to solve a problem. A Fuzzy-Genetic Algorithm is considered as a Genetic Algorithm that uses fuzzy logic based techniques. Genetic algorithm for Soil Profile data is named as GENSOIL. This experiment brings out the fact that the GENSOIL clustering algorithm requires less elapsed time i.e. 0.0529 seconds than other two clustering algorithms such as k-Means and Fuzzy C-Means clustering algorithm that takes 0.281008 and 0.745535 seconds respectively shown in Table 3.

Based on the of the results drawn in this experiment it may be stated that the proposed GENSOIL clustering algorithm is less time consuming than k-Means and Fuzzy C-Means algorithm and therefore superior to and suitable for clustering soil profile data.

4. Conclusion

This work proposes an innovative model **GEN**etic Algorithm for **SOIL** Profile data (GENSOIL). It (GENSOIL) is a randomized search and optimization technique that is guided by the principles of genetics and evolution, which has a large amount of implicit parallelism. On the basis of the result drawn by this experiment, it may be stated that the proposed GENSOIL algorithm is less time consuming than k-Means and Fuzzy C-Means algorithm and therefore it is superior and suitable for classification of Soil Profile data. One of the advantages of GENSOIL is its applicability to different types of related optimization problems with a high calculation speed and getting optimal values.

5. References

1. Jain AK. Data clustering: 50 years beyond K-means. Pattern Recognition Letters. 2010; 31(8) :651-66.
2. Akram M, Habib S, Javed I. Intuitonistic fuzzy logic control for washing machines. Indian Journal of Science and Technology. 2014 May; 7(5).
3. Goldberg DE. Genetic algorithms in search, optimization and machine learning. Addison Wesley Publishing Company; 1989.
4. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics. 1987; 20:53-65.
5. Ghosh S, Kumar S. Comparative analysis of k-means and fuzzy c-means algorithms. International Journal of Advanced Computer Science and Applications. 2013.
6. Ashok Kumar D, Kannathasan N. A study and characterization of chemical properties of soil surface data using k-means algorithm. International Conference on Pattern Recognition Informatics and Mobile Engineering; 2013.
7. Ashok Kumar D, Kannathasan N. A novel soil profile feature reduction model using principal component analysis. Indian Journal of Science and Technology. 2015; 8(29).
8. Batjes NH. ISRIC-WISE Harmonized Global Soil Profile Dataset (Ver.3.1). A Report-2008/2.