Fuzzy C-Means Classifier for Soil Data

P.Bhargavi,M.Sc.,M.Tech
Associate Professor
Department of CSE
Madanapalle Institute of Technology and Science,
Madanapalle

Dr.S.Jyothi, M.Sc.,M.S.,PhD.
Associate Professor & Head,
Department of Computer Science
Sri Padmavathi Mahila Visva Vidyalayam, Tirupati.
(Womens University)

ABSTRACT:

The distribution of soil classes is an important factor in agricultural soils. In order to generate the soil classification, fuzzy soil classifications were developed to provide the means to characterize and quantify the soil classes. This paper presents an index of fuzzy soil classification generated by Fuzzy C-means classification. The ability of classification of the soils is tested with a Soil database. Fuzzy c-means approach is also capable of handling the uncertainty existing in soil parameters. As a result, fuzzy c-means clustering can be successfully applied to classify soils.

Key words:

Soil Classification, Fuzzy C-Means, and Soil database.

1. INTRODUCTION

Soil characteristics are single parameters which are observable or measurable in the field or laboratory, or can be analyzed using microscope techniques. They include such characteristics as color, texture and structure of the soil, features of biological activity, arrangement of voids and pedogenic concentrations as well as analytical determinations.

Soil properties are combinations of soil characteristics which are known to occur in soils and which are considered to be indicative of present or past soil-forming processes.

Soil horizons are three-dimensional pedological bodies which are more or less parallel to the earth's surface. Each horizon contains one or more property, occurring over a certain depth, which characterizes it.

Soils are defined by the vertical combination of horizons, occurring within a defined depth, and by the lateral organization of the soil horizons, or by the lack of them, at a scale reflecting the relief or a land unit.

Soil texture is an important property for agricultural soil. Soil texture is a permanent feature of a soil, unless subjected to erosional depositions or removals. It influences fertility, drainage, water holding capacity, aeration, tillage, and strength of soils.

To classify uncertainty data, Fuzzy C-means classifier is very much useful. Since Soil data may have uncertainty values to classify soils, Fuzzy C-means classifier is very much useful.

In this paper we are applying Fuzzy C-means classifier to classify soil texture based on the properties of soils. Results of their study indicated that, Fuzzy C-means algorithm is capable of accurate real time soil profile classification. This produces clear membership patterns.

2. FUZZY C-MEANS ALGORITHM

The Fuzzy *C*-means (FCM) classification algorithm is proposed by Bezdek [2] aims to find fuzzy partitioning of a given training set, by minimizing the basic *c*-means objective functional:

$$J(Z;U,V)=\sum_{i=1}^{c}\sum_{k=1}^{N}(\Phi_{ik})^{m}\|Z_{k} - V_{i}\|^{2} \qquad A$$

where:

− $U=[\Phi_{ik}] \in M_{fc}$ is a fuzzy partition matrix of $\mathbf{Z}$

− $V=[v_{1}, v_{2}\ldots v_{c}]$, $v_{i} \in \mathbf{R}^{n}$ is a vector of cluster prototypes, to be determined

− $\|Z_{k} - V_{i}\|^{2}$ is dissimilarity measure between the sample $\mathbf{Z}_{k}$ and the center $\mathbf{V_{i}}$ of the specific cluster i (Euclidean distance)

− m $\in(1,\infty)$ is a parameter, that determines the fuzziness of the resulting clusters

The minimization of $J(\mathbf{Z;U,V})$, under the constraint

$\sum_{i=1}^{c}(\Phi_{ik}) = 1$ leads to the iteration of the following steps.

$$V_{i}= \frac{\sum_{i=1}^{k}(\Phi_{ik})^{m}\mathbf{Z}_{k}}{\sum_{k=1}^{N}(\Phi_{ik})^{m}} \qquad , 1 \le i \le c \qquad (1)$$

and

$$\Phi_{ik} = \begin{cases} \dfrac{1}{\sum\limits_{j=1}^{c} 3(D_{ik}/D_{jk})^{2/(m-1)}} \\ \\ \Phi_{ik}=0 \end{cases}$$

If $D_{ik} > 0$ and $\Phi_{ik}0 < 0,1>$,
$$\sum_{i=1}^{c} (\Phi_{ik}) = 1$$

The iteration stops when the difference between the fuzzy partition matrices in two following iterations is lower than å.

3. SOIL CLASSIFICATION

Soil classification deals with the systematic categorization of soils based on distinguishing characteristics as well as criteria that dictate choices in use. Soil classification is a dynamic subject, from the structure of the system itself, to the definitions of classes, and finally in the application in the field. Soil classification can be approached from the perspective of soil as a material and soil as a resource. Engineers, typically Geotechnical engineers, classify soils according to their engineering properties as they relate to use for foundation support or building material. Modern engineering classification systems are designed to allow an easy transition from field observations to basic predictions of soil engineering properties and behaviors.

The most common engineering classification system for soils is the Unified Soil Classification System (USCS). The USCS has three major classification groups: (1) coarse-grained soils (e.g. sands and gravels); (2) fine-grained soils (e.g. silts and clays); and (3) highly organic soils (referred to as "peat"). The USCS further subdivides the three major soil classes for clarification. A full geotechnical engineering soil description will also include other properties of the soil including color, in-situ moisture content, in-situ strength, and somewhat more detail about the material properties of the soil that is provided by the USCS code. For soil resources, experience has shown that a natural system approach to classification, i.e. grouping soils by their intrinsic property (soil morphology), behavior, or genesis, results in classes that can be interpreted for many diverse uses. Differing concepts of pedogenesis, and differences in the significance of morphological features to various land uses can affect the classification approach. Despite these differences, in a well-constructed system, classification criteria group similar concepts so that interpretations do not vary widely. This is in contrast to a technical system approach to soil classification, where soils are grouped according to their fitness for a specific use and their edaphic characteristics. Natural system approaches to soil classification, such as the French Soil Reference System (Referential pedologique français) are based on presumed soil genesis. Systems have developed, such as USDA soil taxonomy and the World Reference Base for Soil Resources, which use taxonomic criteria involving soil morphology and laboratory tests to inform and refine hierarchical classes. Another approach is numerical classification, also called, ordination, where soil individuals are grouped by multivariate statistical methods such as cluster analysis. This produces natural groupings without requiring any inference about soil genesis. In soil survey, as practiced in the United States, soil classification usually means criteria based on soil morphology in addition to characteristics developed during soil formation. Criteria are designed to guide choices in land use and soil management. As indicated, this is a hierarchical system that is a hybrid of both natural and objective criteria. USDA soil taxonomy [9] provides the core criteria for differentiating soil map units. This is a substantial revision of the 1938 USDA soil taxonomy which was a strictly natural system. Soil taxonomy based soil map units are additionally sorted into classes based on technical classification systems. Land capability Classes, hydric soil, and prime farmland are some examples. In addition to scientific soil classification systems, there are also vernacular soil classification systems. Folk taxonomies have been used for millennia, while scientifically based systems are relatively recent developments [6].

A set of soil properties are diagnostic for differentiation of pedons. The differentiating characters are the soil properties that can be observed in the field or measured in the laboratory or can be inferred in the field. Some diagnostic soil horizons, both surface and sub- surfaces, soil moisture regimes, soil temperature regimes and physical, physio-chemical and chemical properties of soils determined were used as criteria for classifying soils. The soils were classified into different orders, sub-orders, great groups, sub-groups, families and finally into series as per USDA Soil Taxonomy[8].

3.1 Soil Data

In this paper Soil data consists of attributes like (i.e., Depth, Sand, Silt, Clay, Sandbysilt, Sandbyclay, Sandbysiltclay, TextureClass). The texture of the Soil data is varied from sand to silty clay loam where as in sub-surface horizons it varied from sand to clay[2]. Table 1 shows the different soil survey symbols.

Table 1: Soil Survey Symbols

S	Sand
Sicl	Silty Clay Loam
Sic	Silty Clay
C	Clay

Sl	Sandy loam
Cl	Clay loam
Sil	Silty Loam
L	Loam
Ls	Loamy sand
Scl	Sand Clay Loam
Sc	Sand Clay

3.2 Uncertainty in Soil data

It is well known that soil quality evaluation has not been fully quantified, as evidenced by the ongoing debate in scientific literature. The uncertainty that is inherent in any evaluation process involves both data and model ambiguity; this ambiguity includes measurement error, inherent soil variability, soil instability, conceptual ambiguity, over-abstraction, simple ignorance of key factors that can impact soil quality. Because of the wide range of factors that make up soil quality and its inherent uncertainty, we believe that a unique approach must be taken to address soil quality. We propose that randomness and uncertainty of soil quality be dealt with by using fuzzy sets theory and fuzzy logic [19]. This theoretical approach provides the basis for analysis of systems characterized by a high degree of uncertainty, nonlinearly and complexity.

In this paper the properties of attributes in the Soil database are specified with the specified quality standards. Quality has the property of fuzziness. If same type of texture occurs with different composition of attributes of Soil data, ambiguity may emerge. All the uncertainty, in accuracy and ambiguity will bring fuzziness to the distribution of agricultural soil.
The Soil data has Fuzziness-based uncertainty.

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this paper, we have applied fuzzy C-means algorithm for Soil data which consists of 10 texture classes. The fuzzy classifiers classify each texture class by clustering them. The clusters formed for the Soil data are shown from figure 1 to figure 10.
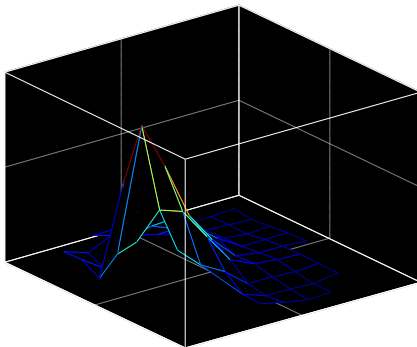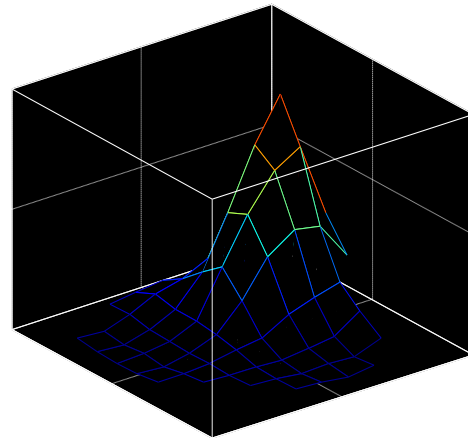

Figure 2: Cluster formed for soil texture silty clay loam

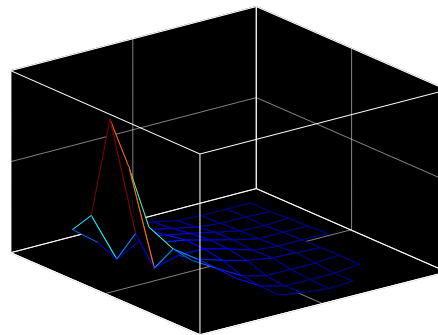**Figure 3: Cluster formed for soil texture silty clay**

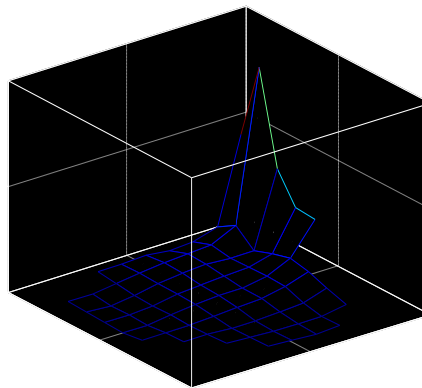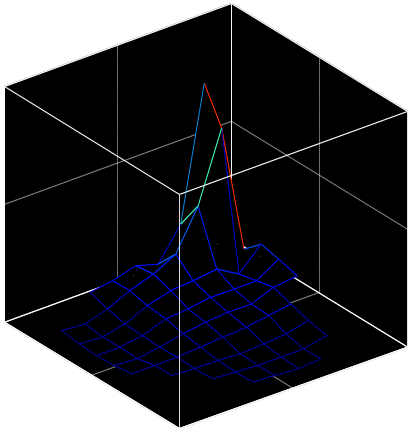**Figure 1: Cluster formed for soil texture sand**

Figure 4: Cluster formed for soil texture clay
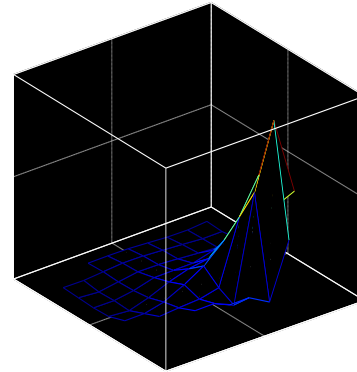
Figure 5: Cluster formed for soil texture sandy loam

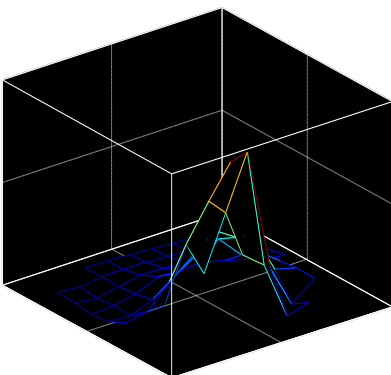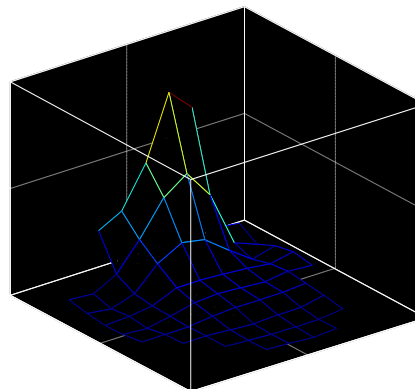

Figure 6: Cluster formed for soil texture clay loam

Figure 7: Cluster formed for soil texture silty loam

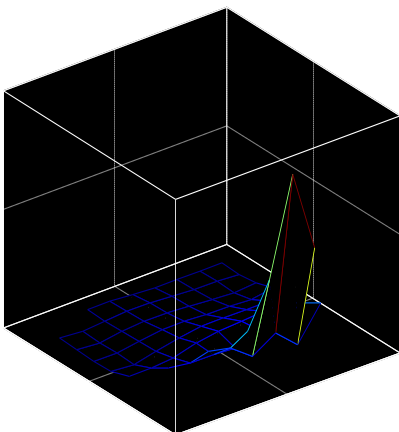**Figure 8: Cluster formed for soil texture loam**

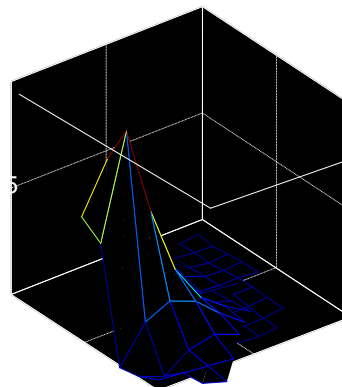**Figure 9; Cluster formed for soil texture sand clay loam**

Figure 10. Cluster formed for soil texture sand clay

Each row in the Soil dataset is a sample data point with clusters greater than one. The fuzzy clustering algorithm outputs a matrix of final cluster centers where each row provides the center coordinates shown in figure 11 and final fuzzy partition matrix and values of objective function for each iteration. The clustering process stops when the objective function improvement between two consecutive iterations is less than the minimum amount of improvement specified, with the accuracy off 0.99.
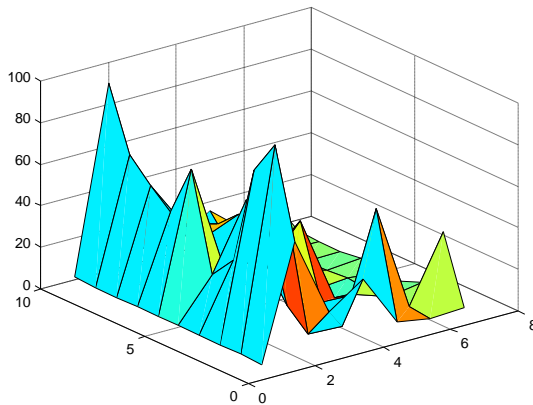

Figure 11: Final cluster centers where each row provides the center coordinates.

5. CONCLUSION

In this paper, we have applied C-means fuzzy classifier for soil data, which has several advantages such as simple and systematic structure. In the cluster generation process, a clustering algorithm based on C-means is applied to cluster the soil data such that the training data can be classified by the fuzzy classifier. In addition it has good generalization ability and can achieve classified rate in the classification problem of agriculture soil data.

6. REFERENCES

[1] R. Kruse, J. Gebhardt, and F. Klawonn, Foundations of Fuzzy Systems, Wiley, 1994.

[2] Bezdek, JC (1981). Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, New York.

[3] Zimmerman, HJ (1996). Fuzzy Sets Theory and Its Applications, 3rd ed. Kluwer Academic Publishing, Boston.

[4] S**.** Abe, and R. Thawonmas, "A Fuzzy Classifier with Ellipsoidal Regions," IEEE Trans. Fuzzy Systems, vo1.5, no. 3, pp.358-368, 1997.

[5] Jain, A.K, Murty, M.N, Flynn, P.J.: Data Clustering: A Review

[6] Bonner, R.E.: On Some Clustering Techniques. IBM, 1964.

[7] Ball, G.H., Hall, D.J.: ISODATA, A Novel Method of Data Analysis and Pattern Clasification. Standford Res. Institute, Menlo Park, 1965.

[8] Fromm, F.R., Northouse, R.A.: CLASS, A Nonparametric Clustering Algorithm. Pattern Recognition.

[9] Raja, A.,Meister, A., Martverk., P.: Fuzzy Classification Algorithms with Some Applications.

[10] Baraldi, A., Blonda. P.: A Survey of Fuzzy Clustering Algorithms for Pattern Recognition. ICSI, TR-98-038, 1998.

Ching-Chang Wong, Chia-Chong Chen, **and** Shih-Liang Yeh: K-Means-Based Fuzzy Classifier Design.

[11] Setnes, M., Kaymak, U.: Extended Fuzzy Clustering Algoritms.

[12] McBratney, A.B., De Gruijter, J.J.: A Continuum approach to soil classification by modified fuzzy k-means with extragrades.

[13] A Thesis by D.Basavaraju Characterisation and classification of soils in Chandragiri mandal of Chittoor district,Andhra Pradesh.

[14] A thesis titled "Characterization, Classification and Evaluation of Soil Resources in Sivagiri Micro-Water Shed of Pichatur Mandal, Chittoor District, Andhra Pradesh".

[15] A thesis titled "Morphology and Taxonomy of Soils in Ramachandrapuram Mandal of Chittoor District, Andhra Pradesh".

[16] George J. Klir and Bo Yan. Fuzzy Sets and Fuzzy Logic. Theory and Applications, PHI.

[17] Soil Survey Staff 1998, Keys to soil taxonomy. Eight Edition, Natural Resource Conservation Services, USDA, Blacksburg, Virginia.

[18] Soil Survey Staff 1951, Soil Survey Manual. US Department of Agricultural Hand book No. 18.

[19] Jager R., 1995. Fuzzy Logic in Control. Delft TU Press, Delft, Belgium.

Bhargavi Peyakunta is working as Associate Professor in the Department of Computer Science and Engineering, Madanapalle Institute of Technology and Science, Madanapalle, Andhra Pradesh. **Educational Qualifications:** M.Sc in Computer Science from Sri Krishnadevaraya University, Anantapur and M.Tech degree from Sri Vinayaka missions University, Salem, India. **Teaching & Research Experience:** 12 years of teaching experience & 5 years of research experience. **Current Research Interests:** Data Mining, Fuzzy Systems, Genetic Algorithms and GIS.

Jyothi Singaraju is working as Associate Professor & Head in the Department of Computer Science, Sri Padmavathi Mahila Visvavidyalayam(SPMVV), Tirupati. **Educational Qualifications:** M.Sc in Applied Mathematics from S.V.University, Tirupati , M.S in Software Systems from BITS, Pilani, & Ph.D in Theoritical Computer Science from S.V.University, Tirupati. **Teaching & Research Experience:** 16 years teaching experience & 24 years research experience. **Current Research Interests:** Fuzzy Systems, Neural Networks, Data Mining, Data Base Management Systems, Genetic Algorithms, Bioinformatics and GIS.