Classification of Remotely Sensed data using Gravitational Symbolic Clustering

D.S.Vinod,
Research scholar,
Department of Computer Science and Engineering,
Sri Jayachamarajendra College of Engineering,
Manasagangothri, Mysore 570 006,
Karnataka, India.
Email:-ds_vinod@mailcity.com,
vinod@sjce.ac.in .
Tel:-0821-512568(college), 481314(Res.).

T. N. Nagabhushana,
Asst. Prof.
Department of Computer Science and Engineering,
Sri Jayachamarajendra College of Engineering,
Manasagangothri, Mysore 570 006,
Karnataka, India.
Email:-tnn@sjce.ac.in
Tel:-0821-512568(college).

Abstract
Remotely sensed data is obtained through artificial satellites from various geographic and astronomical sources. Remotely sensed image data are well adopted for monitoring and analyzing the behavior of different covers and also the temporal changes occurring on the target. Observations are made by periodically collecting the data and then comparing previous data. Using various sensors, data is collected and further analyzed to obtain information about objects and areas under investigation. The number of features in each sample of the 'Remotely sensed data' or 'multispectral image' depends upon the number of channels from which the information is collected. Data from different channels are combined to obtain the detailed information about the body or area under observation. We propose an improvised method of classification of such data using gravitational symbolic clustering. Symbolic data is a special case of conventional data, which is closer to real life interpretation and analysis. In conventional data, the objects are individualized, whereas in symbolic data sets they are more unified by means of relationships. The multispectral image is a quantitative interval type of symbolic data. It requires lot of time and memory space to analyze or classify the data. In order to overcome such limitations a data reduction technique is proposed. The data reduction technique uses storage bin arrays to store useful information of the reduced image data. The major idea is to present a clustering algorithm based on gravitational symbolic approach. The procedure is based on the physical phenomenon in which a system of particles in space converges to the centroid of the system due to the gravitational attraction between the particles. Both Agglomerative and Disaggregative Gravitational Symbolic approaches are considered. The concept of mutual pairs is used to merge the samples. The process of merging reduces the number of samples each time they are available for consideration. The process terminates at a stage when no more mutual pairs are available for merging. A detailed study has been carried out on different sets of multispectral images. We have also compared our results using cluster indices.

Keywords: Cluster Coglomerative strength, Clustering, Global coglomerative strength, Gravitational Clustering, Symbolic objects, Cluster Index, Composite symbolic objects, Mutual Pairs.

1.0 Introduction

Symbolic objects are extensions of classical data types. In conventional data sets, the objects are individualized, whereas in symbolic data sets they are more unified by means of relationships [(2-4,16-17)].

Gowda and Diday[(2-3)] have presented an agglomerative clustering algorithm clustering algorithm for symbolic objects. They form composite symbolic objects (CSO) using a Cartesian join operator whenever a mutual pairs of symbolic objects are selected for agglomeration based on minimum dissimilarity and maximum similarity.

The combined usage of similarity and dissimilarity measures for agglomerative and divisive clustering of symbolic objects is presented by Gowda and Ravi[(5-6)].

To manage time and memory requirements, the clustering method involves the use of data reduction techniques[11-12]. Here the data reduction as proposed by K.C.Gowda and S.K.Prakash[15] is modified to accommodate the task without dimensionality reduction.

We have also incorporated the gravitational clustering techniques to cluster multispectral images of satellites[7,10]. We have also compared our results using cluster indices[4]. Both Agglomerative and Disaggregative Gravitational Symbolic approaches are considered. The concept of mutual pairs is used to merge or split the samples.

2.0 Data Reduction Techniques

Data reduction techniques are used before clustering the data. The data reduction technique uses bin arrays to store useful information of the reduced image. Here the **m** samples of d-dimensions are mapped on to bin arrays. Two cases are:

1) When dimension $d \geq 2$ and
2) $d = 1$.

The results of the reduced data must be mapped on to the original data after clustering the reduced data. This requires a set of labels of all the m samples, which identifies the location of each sample in the nonempty bins. This is used for making the classification-output map after clustering the reduced data.

2.1 When dimension $d \geq 2$

This works by assigning d-dimensional m number of samples to any one of the 2-dimensional arrays. The d 2-dimensional $P \times P$ bin arrays (P is a positive integer as required) are used to store the useful information of all the features. The data before reduction requires normalization between 1 to P. The normalized features are assigned to the $P \times P$ arrays.

The idea that the first feature is used to determine the column position of the bin to which the sample is to be assigned. The combination of the remaining feature values is used to determine the row position of the bin.

If the first sample have the feature values $f_{11}, f_{12}, f_{13}, \ldots, f_{1d}$, which are normalized between 1 to P. This sample is assigned to a bin having a column value of f_{11}. As this is the first sample, the row value of its bin is also 1.

If the second sample have the feature values $f_{21}, f_{22}, f_{23}, \ldots, f_{2d}$. Say if $f_{11} = f_{21}$, then the bin to which this sample is assigned also has a column value of f_{11} and its bin has a row value of 1 if the below condition is satisfied:

$$|f_{12} - f_{22}| + |f_{13} - f_{23}| + \ldots + |f_{1d} - f_{2d}| < \text{threshold } T \quad \ldots\ldots\ldots(1)$$

Where **T** is a user-defined limit. If the above condition is not satisfied, then the row value of the bin is 2. Accordingly for a given column position, a new bin with a higher row value is considered only when the present sample cannot be assigned to any other bin with lower row value.

The steps are as follows:

If $S_1, S_2, S_3, \ldots S_d$ be the $P \times P$ 2-dimensional arrays to store updated information of each feature and W is another such $P \times P$ 2-dimensional such bin-weight array to update number of samples assigned to each corresponding bin.

1) Normalize the d dimensional m samples between 1 to P.
2) The first feature gives the column position of the bin.
3) The row value of the particular bin is determined as above.

4) As each sample is assigned to a bin the arrays S_1, S_2, S_3,…………S_d is updated.
5) Also the number of nonempty bins is updated in W bin-weight array.

2.2 When dimension d=1

This works by assigning 1-dimensional m number of samples to any one of the 1-dimensional arrays. The one 1-dimensional P bin arrays are used to store the useful information of all the features. The data before reduction requires normalization between 1 to P. The normalized features are assigned to the P arrays.

The idea that the first feature is used to determine the column position of the bin to which the sample is to be assigned. If the first sample has the feature values f_{11}, which is normalized between 1 to P. This sample is assigned to a bin having a column value of f_{11}.

If the second sample have the feature values $f_{21.}$ Say if $f_{11} = f_{21}$, then the bin to which this sample is assigned also has a column value of $f_{11.}$

The steps are as follows:

Let S_1 be the P 1-dimensional arrays to store updated information of the feature and W is another such P 1-dimensional bin-weight array to update number of samples assigned to each corresponding bin.

1) Normalize the d dimensional m samples between 1 to P.
2) The first feature gives the column position of the bin.
3) As each sample is assigned to a bin the arrays S_1 is updated.
Also the number of nonempty bins is updated in W bin-weight array.

3.0 Cluster Coglomerate and Global Coglomerate Strengths

The cluster coglomerate strength (CCS) of cluster C_j is :

$$\mathbf{CCS_j = (m_j \ / D_j^{\ s}) \times \sum_{X_i \in C_j} (D^r (X_i , X_m))} \text{ ……………………(2)}$$

Where m_j cluster weight of C_j ,
X_m Composite object of C_j ,
D_j $\max_{X_i \in C_j} D (X_i , X_m)$ or maximum dissimilarity.
s: a suitable positive number and
r: a suitable negative number depends on the input sample set.

The global coglomerate strength (GCS) of cluster at any level is defined as the sum of all the CCS of all the prevalent clusters at that level. Thus the GCS when cluster C_j is formed, is defined as:

$$\mathbf{GCS_j = \sum_{C_j} (CCS_i)} \text{ …………………… ……………(3)}$$

4.0 Symbolic Gravitational Clustering Algorithm

4.1 Agglomerative Method

1) Reduce the data by choosing an appropriate bin size and threshold. Let the initial number of clusters be equal to number of samples in the reduced data (N) each cluster weighted 1, and CCS =0 and GCS =0.

2) Using the similarity measure, find all the mutual pairs present in the symbolic data set.

. 3) Consider the mutual pairs X_i and X_j Merge the two objects to form a CSOm. Find the CCS ‘CCSm’ of the clusters consisting of two objects X_i and X_j and also GCS ‘GCSm’ is also found out.

4) The cluster threshold $CT_{i,j}$ is calculated

$$CT_{i,j} = ((n_i \times CCS_i) + (n_j \times CCS_j)) / (n_j + n_i) \ldots..(4)$$

and global threshold $GT_{i,j}$ is also calculated :

$$GT_{i,j} = ((n_i \times GCS_i) + (n_j \times GCS_j)) / (n_j + n_i) \ldots. (5)$$

n_i and n_j are the cluster weights of X_i and X_j respectively.

The two objects are replaced by a CSO if the below two conditions are satisfied :
a) $CCSm > CT_{i,j}$
b) $GCSm > GT_{i,j}$
Then reduce the number of clusters by one.

5) Step 3 and 4 is repeated for all the mutual pairs present at that stage.

6) Determine the Cluster Indicator value for each P^{th} merging:

$$CI = |Max_{p+1} - Max_P| / Min_p \ldots\ldots\ldots\ldots\ldots\ldots\ldots. (6)$$

Where Max_{p+1} and Max_P are the maximum similarities of $(P+1)^{th}$ and p^{th} merging respectively.

7) Iterate steps 2 to 6 until a stage is reached when no replacement of mutual pair occurs.

4.2 Divisive Method

1) Reduce the data by choosing an appropriate bin size and threshold. Let N be the initial number of samples in a sample set S. And let the number of clusters be equal to one, with N number of samples.

2) Set a threshold value TCCSm.

3) Using the similarity measure, find all the mutual pairs present in the symbolic data set.

. 4) For all the Mutual Pairs present find the CCS 'CCSm between samples. Check if CCSm>TCCSm, if so then split the Mutual pairs into two separate clusters. Increment the number of clusters by 2.

5) Step 3 and 4 is repeated for all the mutual pairs present at that stage.

6) Determine the Cluster Indicator value for each P^{th} merging as in equation 6.:

7) Decrement TCCSm in steps.Iterate steps 3 to 6 until a stage is reached when no replacement of mutual pair occurs.

5.0 RESULTS

Image of Moon taken by Galileo space craft on December 9 1990, with catalog number PIA00113 is taken for testing. Size of image is 535 * 535 with 3 features. Both Agglomerative and Divisive Gravitational Symbolic methods are considered, results are as shown in figure 1 and 2 respectively . Plot of Cluster Index Vs Number of Clusters is also shown in figure 3.

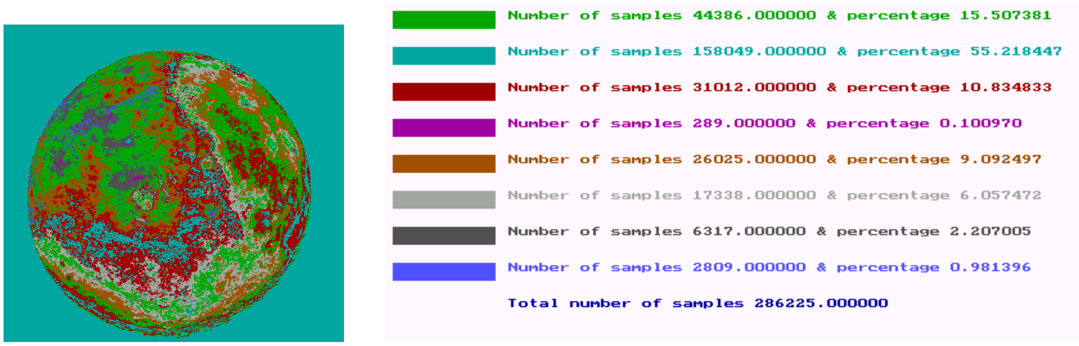

Fig 1. Classification Map and details of Classification Cover of Moon (Agglomerative method)

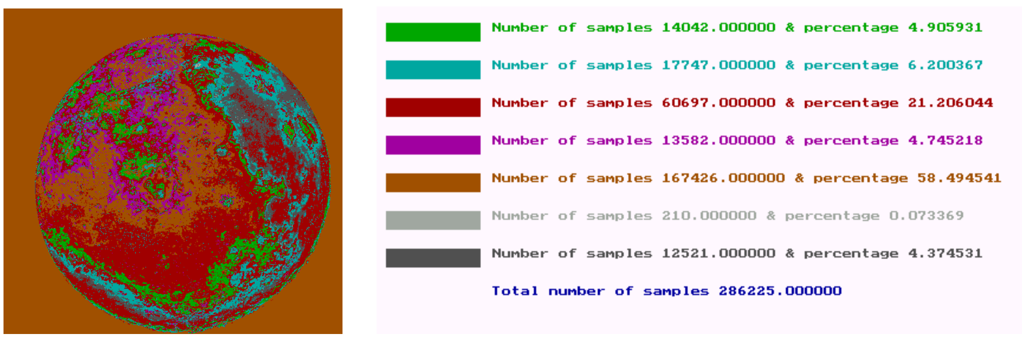

Fig 2. Classification map and details of classification cover of moon (divisive method)

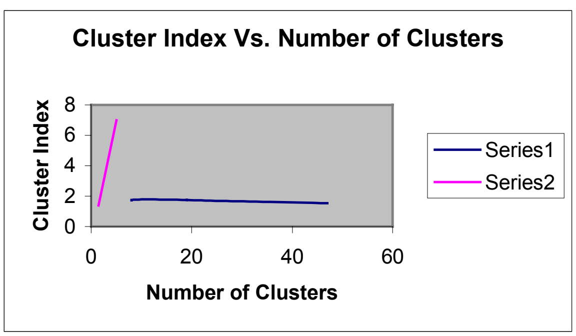

Fig 3. Variation for CI for Moon Image (Series 1 Agglomerative Method Series 2 Divisive metod.)

6.0 Conclusions

We have proposed a clustering methodology to classify multispectral images. The algorithm is based on the physical phenomenon in which a system of particles in space converge due to the gravitational attraction between the particles. To accomplish space and time complexity requirements data reduction techniques are used. We have proposed both Agglomerative and Divisive methods of Clustering. Cluster indices for both the methods are shown in figure 3. Analyzing the plot (figure 3) it is evident that Divisive technique requires less number of iterations when compared to Agglomerative technique. The process ends at a stage when there are no more mutual pairs available for merging or splitting.

References

[1]A.K. Jain and Dubes, "Algorithms for clustering data", PHI, 1988
[2]K.C. Gowda and E.Diday, "Symbolic clustering using new dissimilarity measure", pattern recognition, vol.24, no.6,pp 567-578, 1991.
[3]K.C. Gowda and E.Diday, "Symbolic clustering using a new similarity measure", IEEE trans. Systems, man Cybernetics, vol.22, pp.368-378,1992.
[4]K.C. Gowda and E.Diday, "Unsupervised learning through symbolic clustering", Pattern Recognition Letters", vol. 12 pp 259-264., 1991.
[5]K.C.Gowda and T.V. Ravi , "Divisive Clustering of symbolic clustering using the concept of both similarity and dissimilarity", Pattern recognition, Vol.28,No. 8, PP.1277-1282,1995.
[6]K.C.Gowda and T.V. Ravi, "Divisive Clustering of symbolic clustering using the concept of both similarity and dissimilarity", Pattern recognition letter, No.16, 1995 pp. 647-652.
[7]K.C.Gowda and T.V. Ravi, "Clustering of Symbolic objects Using Gravitational approach", Vol. 29, No. 6, 1999
[8]K.C.Gowda and G.Krishna, "Agglomerative clustering using the concept of mutual nearest neighborhood", pattern recognition, vol. 10 pp 105-112, 1977.
[9]K.C.Gowda and G.Krishna, "Disaggregative clustering using the concept of mutual nearest neighborhood", IEEE transaction on systems man and cybernetics vol. SMC-8, pp.888-895, 1978.
[10]K.C.Gowda and G.Krishna, "Gravitational Clustering using the concept of Mutual Nearest Neighborhood", Joint meeting of the classification and psychometric society, Montreal, Canada, May 31-June 2, 1982.
[11]K.C.Gowda, "A feature reduction and unsupervised classification algorithm for multispectral data", Pattern recognition, vol. 17, no. 6, pp 667-676, 1984.
[12]K.C. Gowda, " A new method for mapping multidimensional data to lower dimensions", proceedings of int. Symposium. On Remote Sensing of the Environment, 2nd Thematic conference, 1982
[13]H.N Srikanta Prakash, P. Nagabhushan and K. Chidananda Gowda, "Symbolic clustering of remotely sensed data using a new dissimilarity measure", International conference on cognitive Systems-ICCS'97, New Delhi, 13-15 Dec. 1997.
[14]H.N Srikanta Prakash, P. Nagabhushan and K. Chidananda Gowda, "Symbolic Agglomerative clustering for qualitative analysis of remotely sensed data", International conference on remote sensing and GIS/GPS, India, March 1997.
[15]H.N Srikanta Prakash, P. Nagabhushan and K. Chidananda Gowda", Symbolic ISODATA clustering procedure useful for land cover classification", IGRASS-97, Singapore, 3-8 august-1997.
[16]E.Diday, "The Symbolic approach in clustering, Classification and Related Methods of Data Analysis", H.H. Bock, ed., Elesvier, North Holland, Amsterdam, 1988.
[17]E.Diday, "Data Analysis, Learning Symbolic and Numeric Knowledge", Nova Sc. Publishers,1988.
[18]D.S. Vinod and T.N. Nagabushana, "Classification of Remotely sensed data using Self –organizing Feature map", International conference on remote sensing and GIS/GPS, India, February 2001.
[19] D.S. Vinod and T.N. Nagabushana "Classification of Multispectral Images Using Symbolic Gravitational Approach", National conference on recent trends in information Technology, Coimbatore, pp 29-34, 22-24 Aug., 2002.