R & D Project Report-2

Academic Year- 2019-20

On

INTELLIGENT ANALYSIS ON SATELLITE IMAGERY

Submitted by

1. Anjali Mishra BT17GCS157 CSE
2. Arundhati Das BT17GCS016 CSE
3. Manaal Soni BT17GCS052 CSE
4. Yash Kejriwal BT17GCS123 CSE
5. Himanshu Nikhare BT17GCS036 CSE

Supervisor: Dr. Sudip Sanyal

Introduction:
Nowadays, satellite imagery is widely used in remote sensing and other research areas for various applications. Intelligent analysis of satellite images is a process of collecting and interpreting the satellite images from multiple satellites such as Sentinel-2B, Landsat-8, etc.The aim of the study is high resolution satellite imagery land cover classification with higher accuracy. Land cover classification is a well-studied problem in the domain of remote sensing.In this study we will compare different algorithms on the basis of their accuracy and find the most efficient one.The study includes comparison of different single class algorithms for single class classification and predicts the higher accurate results implementing the best defined approach.We'll begin by comparing different unsupervised learning techniques & then accordingly choosing the best technique for the classification. Future scope of the study is to extend the optimization criteria for multiple classes. We will also try to introduce several methods that will make the system more robust so that the same model can be used for a larger region.

Problem statement:
High Resolution Satellite Imagery land cover classification which includes identification of higher accuracy classification algorithm.

Literature review:
Riese et. al 2019,the present study introduces Self Organising Map-based framework, the Supervised Self-organizing maps (SuSi) framework, which can be used for supervised and semi-supervised regression and classification tasks as well as for unsupervised clustering and visualization.To evaluate the SuSi framework is applied on two exemplary hyperspectral datasets and compared with random forest (RF) model for the respective regression and classification results.the results of semi-supervised regression SOM clearly outperforms the RF baseline classifier and achieves satisfying results.

Hemant Kumar Aggarwal & Sonajharia Minz 2015,this paper presents the three types of unsupervised learning techniques used for the change detection in water, vegetation and built-up land cover classes of Delhi region in India. Eight images were taken from satellite Landsat TM and Landsat ETM+ from the year 1998-2011 for preprocessing. Three features namely Soil Adjusted Vegetation Index(SAVI), Modified Normalized Difference Water Index (MNDWI), and Built-up from Normalized Difference Built-up Index (NDBI) were extracted at the preprocessing stage. These features corresponds to vegetation, water, and built-up classes respectively.The three clustering algorithms k means, fuzzy c mean and expectation maximization were selected to represent the partition based, fuzzy, and probability based technique respectively.The three algorithms were implemented to cluster the pixels of all the eight images using the above three features namely SAVI, MNDWI and NDBI.Those features enabled a 50% reduction in the dimensionality of data & the three algorithms also yielded good results.Based on the results obtained from the features & algorithms, it had been seen that vegetation has decreased every year whereas urban area has increased. Based on the measure of silhouette coefficients,partition based clustering algorithm is more effective in comparison to probabilistic and fuzzy based clustering techniques and thus for change detection.

Mohsen Gholoobi & Lalit Kumar 2015, this paper proposes a hierarchical land cover classification based on the image objects that are created from multiresolution segmentation.A rule-based strategy is used to implement object-based land cover classification on a IKONOS image of Shiraz, Iran.As the urban classes are presented as a set of adjacent pixels, an object-based classification technique has an advantage over the traditional pixel-based classification.These procedures utilize the spectral, geometrical, textural, and conceptual features of an image.To improve the accuracy, a new spatial geometrical analysis for the reclassification of unclassified land cover objects is also utilized. After this, an object-based land use classification is implemented based on the land cover results. Overall classification accuracy was 89% and 87% for the land cover and land use approaches, respectively. In the best unclassified object analysis, ~70% of unclassified objects were reclassified correctly.

Usman 2013,the study mainly focuses on the use of K-means clustering algorithm to classify satellite imagery into specific objects within it.Segmentation and classification of high resolution satellite imagery is a challenging problem due to the fact that it is no longer meaningful to carry out this task on a pixel-by-pixel basis.K-means clustering algorithm is a better method of classifying high resolution satellite imagery. The extracted regions are classified using a minimum distance decision rule.The procedure significantly reduces the mixed pixel problem suffered by most pixel based methods and getting better classification accuracy with the overall performance for accuracy percentage as 88.889%.

Kaul, H.A. & Ingle, Sopan. (2012) in Land Use Land Cover Classification and Change Detection Using High Resolution Temporal Satellite Data proposed a survey of land use/land cover of Jalgaon District. In their study they tried to compare their proposed supervised approach from the one proposed by Rao & Narendra (2006) and Boakye et al (2008), both these researchers had mapped the land use/ land cover changes using unsupervised classification.
The researcher used ERDAS imagine software for classification and created a spatial digital database. They reported land use/ land cover classification with max. Classification accuracies were acquired by using maximum likelihood classification decision rule. The remote sensing data was collected from IRS-P6 , LISS-III images for march and data for november was collected from NRSC, Hyderabad. The data has 4 spectral bands of 23.5m spatial resolution. Overall accuracy of classification is 89% and 91.02% and Kappa statistics is 0.86 and 0.88 respectively.

Gonçalves et. al 2011, this particular paper is on Land-Cover Classification Using Self-Organizing Maps Clustered with Spectral and Spatial Information. Using unsupervised learning is a good approach when there is very less prior information about the data. Unsupervised learning uses clustering methods for classifications. These methods examine the unknown pixels in an image and cluster them into different classes. The basis of selecting a method depends on image size and feature dimension. Self Organising Maps **(SOM)** is an unsupervised and competitive learning Artificial Neural Network model. SOM is a method which is used **(i)** When data to be clustered is unlabelled i.e., the number of classes are unknown and **(ii)** For dimension reduction of the data. This paper presents a 2 level SOM based clustering approach. The first level is SOM Training → map original patterns of image to a reduced set of prototypes arranged in a 2D rectangular grid. Here two factors are kept in mind: Image sampling process and determination of the SOM training parameters. The second level is SOM Segmentation → segmenting the SOM output map using an additional clustering method. This approach reduces the computational load of the classification process. After training the SOM the second level will automatically, without user interference, reduce the dimensions as per the need. Only two parameters that need to be defined by the user: size of samples and number of SOM neurons.

Workflow:

- Collecting Data
- Data Preprocessing
- Classification of Data(unsupervised)
- Post classification(error reduction)
- Accuracy assessment
- Result ,analysis & conclusion

Proposed methodology:

A high resolution satellite imagery illustrating various types of land use and land cover will be used as the test image for classification.Initially the supplied image will be extracted from its compressed format of around 10 m resolution for the bands 5, 4, 3, and 2 of NIR, Red, Green and Blue and then the pixels of the image is clipped out and saved. A composite of Bands 4,3,2 will then be performed that will give an output and be saved as TIFF/GEOTIFF format. Now,first an unsupervised classification will be performed on the image using different algorithms to classify the image into the desired classes . A raster layer will be generated using the unsupervised algorithm while running QGIS/ENVI.The pixels will be identified for each of the categories and then they will be grouped into land cover categories.

Technology

1. **Software Used** :
   i. Sentinel-2 Satellite Image (Corpeicus)
   ii. QGIS 3.12.1
   iii. ENVI 5.0
2. **Hardware Used**
   i. Laptop

Algorithms Studied:

We have analyzed some of the unsupervised learning algorithms like K-means,DBSCAN,SOM, ISODATA.

A.] K-means Clustering:

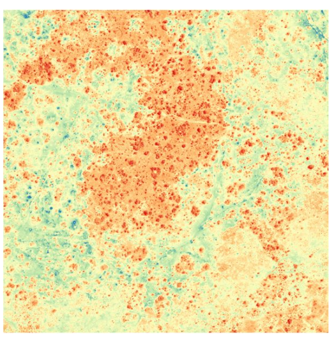
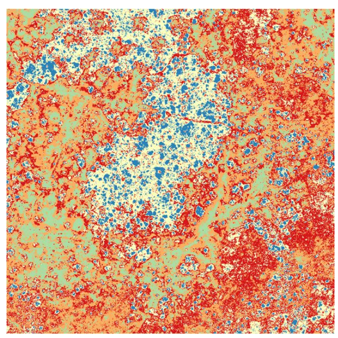
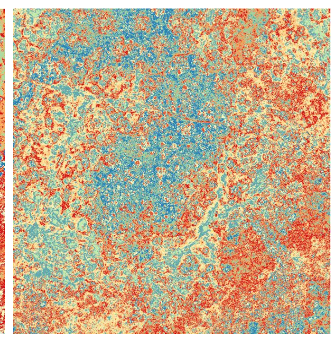
It is one of the most common unsupervised learning clustering algorithm used to get an intuition about the structure of the data.It's an iterative algorithm which clusters, or partitions the given data into K-clusters or parts based on the K-centroids.It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid is at the minimum. The less variation we have within clusters, the more similar the data points are within the same cluster.

Algorithm:

1. Select the number of clusters K.
2. Choose randomly K points, the centroids.
3. Assign each data point such that it is assigned to the closest centroid
4. Compute and place the new centroid of each cluster.
5. Reassign each data point to the new closest centroid.
6. If any further reassignment of data point took place, go to step 4, otherwise, the model is ready.

Analysis:

The original image is being displayed below. On that we performed K-means clustering by first using 5 clusters & then with 10 clusters.A significant change has been observed after increasing the cluster to 10. Increasing the number of clusters actually helps us to optimize the model in terms of how good it actually separates from other clusters.

Original image **Cluster=5** **Cluster=10**

B.] Self Organising Map:

Self Organising Maps **(SOM)** is an unsupervised and competitive learning Artificial Neural Network model. SOM is a method which is used **(i)** When data to be clustered is unlabelled i.e., the number of classes are unknown and **(ii)** For dimension reduction of the data.

Algorithm:

1. Initialize the SOM.
2. Get input pattern in vector x.
3. Initialize all the neurons i and a weight vector w_i associated to a neuron i.
4. Repeat steps 5 and 6 for all nodes.
5. Calculate the euclidean distance between the input vector and the weight vector.
6. Find best matching unit (BMU) i.e., the node with the minimum euclidean distance
7. Calculate learning rate and neighborhood function
8. Calculate neighborhood distance weight matrix
9. Modify SOM weight matrix

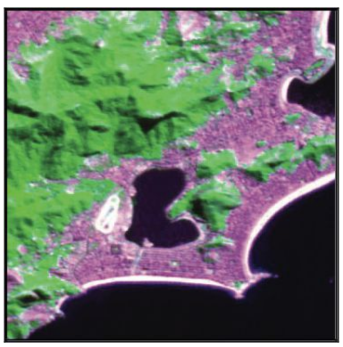
10. Repeat from step 4 until the maximum number of iterations is reached.

SOM v/s K Means and ISODATA clustering algorithms

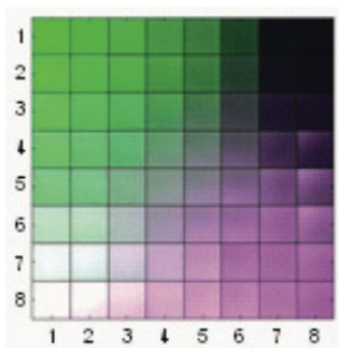
1. To begin with K Means and ISODATA, the number of classes need to be known beforehand.
2. Users need to manually define many parameters to which K Means and ISODATA algorithms are very sensitive.
3. High computational cost for K Means and ISODATA, when data to be analyzed is very large.

Analysis:

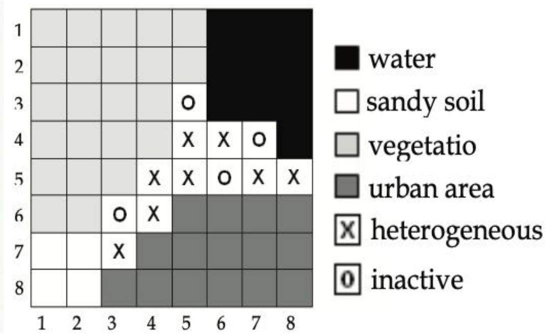
Image: Rio de Janeiro
Sample Window: 7x7
Samples: 1521
Of the samples obtained an 8x8 rectangular grid is obtained



Input



SOM grid



Classified SOM grid

Image classified by SOM

Image Classified by K Means for K=4

C.] ISODATA

The ISODATA classifier is a modified form of the K-means classifier, with the ability to split classes with too much variance and merge classes that are too similar between each iteration.

ISODATA Algorithm:

1) ISODATA computes class means consistently circulated in the data space before iteratively clusters the continuing pixels utilizing least distance approaches.
2) Every iteration recalculates means as well as reclassifies pixels through respect to the new means.
3) It may turn out later that more or fewer clusters would fit the data better.

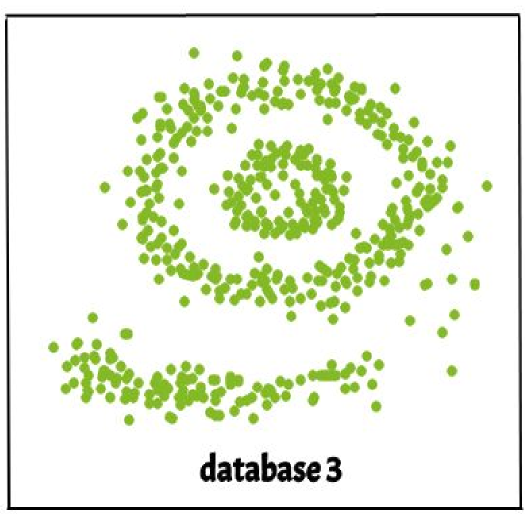
Analysis:
ISODATA Algorithm, allows the number of clusters to be adjusted automatically during the iteration by merging similar clusters and splitting clusters with large standard deviations unlike K-Means Algorithm.

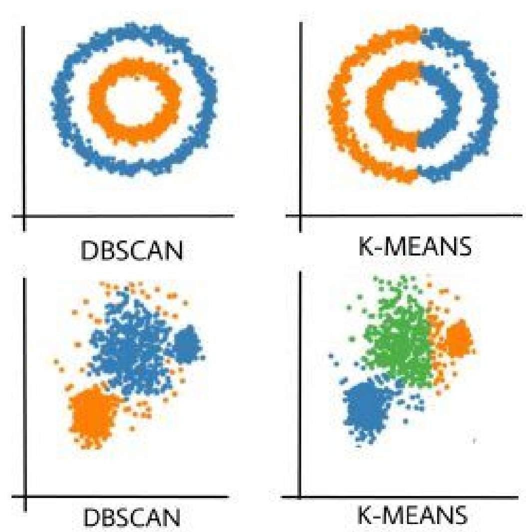
Gaps/Challenges :
One problem with the ISODATA approach is that it can be difficult to converge.

How can we overcome this Gap ?
We can start the ISODATA process with an overestimate of the number of classes and then simply merges (i.e. there is no splitting of classes).

D.] DBSCAN
Density-based spatial clustering of applications with noise (DBSCAN) is a data clustering algorithm proposed by Martin Ester, Hans-Peter Kriegel, Jörg Sander and Xiaowei Xu in 1996. DBSCAN is one of the most common clustering algorithms

Why DBSCAN ?





Partitioning methods (K-means, PAM clustering) and hierarchical clustering work for finding spherical-shaped clusters or convex clusters. In other words, they are suitable only for compact and well-separated clusters. Moreover, they are also severely affected by the presence of noise and outliers in the data.

Real life data may contain irregularities, like -
i) Clusters can be of arbitrary shape such as those shown in the figure below.
ii) Data may contain noise.

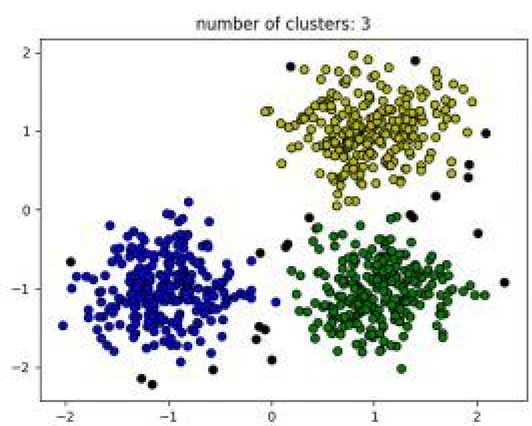
Basically, DBScan Algorithm tried to overcome most of such drawbacks of algorithms like K-Means. It identifies the dense region by grouping together data points that are closed to each other based on distance measurement.

Abstract Algorithm:
DBSCAN requires two parameters: ε (eps) and the minimum number of points required to form a dense region MinPts.

1. Find all the neighbor points within eps and identify the core points or visited with more than MinPts neighbors.
2. For each core point if it is not already assigned to a cluster, create a new cluster.
3. Find recursively all its density connected points and assign them to the same cluster as the core point.
   A point a and b are said to be density connected if there exist a point c which has a sufficient number of points in its neighbors and both the points a and b are within the eps distance. This is a chaining process. So, if b is neighbor of c, c is neighbor of d, d is neighbor of e, which in turn is neighbor of a implies that b is neighbor of a.
4. Iterate through the remaining unvisited points in the dataset. Those points that do not belong to any cluster are noise.

A naive implementation of this requires storing the neighborhoods in step 1, thus requiring substantial memory.

Output:



Complexity:

- O(n log n)
- a non-matrix based implementation of DBSCAN only needs O(n) memory.

References:

[1] Felix M. Riese, Sina Keller , Stefan Hinz (2019) Supervised and Semi-Supervised Self-Organizing Maps for Regression and Classification Focusing on Hyperspectral Data, arXiv:1903.11114

[2]Hemant Kumar Aggarwal & Sonajharia Minz (2015) Change Detection Using Unsupervised Learning Algorithms for Delhi, India

[3]Mohsen Gholoobi & Lalit Kumar (2015) Using object-based hierarchical classification to extract land use land cover classes from high-resolution satellite imagery in a complex urban area

[4]Usman Babawuro(2013) Satellite Imagery Land Cover Classification using K-Means Clustering Algorithm: Computer Vision for Environmental Information Extraction,Elixir Comp. Sci. & Engg. 63 (2013) 18671-18675

[5]Kaul, H.A. & Ingle, Sopan. (2012). Land Use Land Cover Classification and Change Detection Using High Resolution Temporal Satellite Data. The Journal of Environment. 1. 146-152.

[6] M. L. Gonçalves, J. A. F. Costa and M. L. A. Netto (2011) Land-Cover Classification Using Self-Organizing Maps Clustered with Spectral and Spatial Information