Remote Sensing and GIS for Rural Development Professor. Pennan Chinnasamy Centre for Technology Alternatives for Rural Areas (CTARA) Indian Institute of Technology, Bombay Lecture 41 Different types of LULC classifications

Hello everyone, welcome to the NPTEL course on remote sensing and GIS for rural development. This is week 9 lecture 1. This is an important course for understanding or getting an introduction to the potential of remote sensing data for rural development.

(Refer Slide Time: 0:44)



Initially, we had looked at remote sensing applications for rural development by defining what is rural development and then we had gone through each and every specifics of the remote sensing and GIS aspects. Now, we have jumped into the application part. Let us do a recap of week 8 and then see how it is linked to week 9.

In week 8, we had discussed LULC, where LU is the land use and LC is the land cover. We define land cover as the cover on the bare skin of the earth what is on top of the Earth, which is the bare like rocks and lithology on top of it, what is there? That is the land cover. Suppose, there is nothing else, barren land is also used as land cover in some classification methods.

Then we looked at land use, where you have a land cover how has it been used? For example, one may have a forest and converted to a conservation area, another person can convert it into a school, Monastery, etcetera. And the last person can convert into a resort as usual. So, there is one land cover, but the land use is different.

And then the last point was we discussed land use land cover change, how does it change for a particular time period, let us say a decade 10 years, once 5 years once etcetera, etcetera. This is important because we need to understand what is the land and resources available for rural development. And since it is at large scale and a large time span, you cannot keep collecting data. Rather, you will be using remote sensing and GIS data.

Then we looked at how LULC contributes to water and land management. We looked at what are the data issues specifically for estimating land use and land cover. And then especially the change. As I explained the spatial coverage is big. And temporal analysis is there. There are two types of temporal analysis. One is over a period of 10 years. And it is not once every year, you have to take during the seasons. There is Rabi, Kharif, Zaid season.

So, at least 3 times a year you take, and then 10 years to look at some quantitative data for assessing LULC change. So, data is expensive and a lot of issues. And then we looked at a lot of proxy data that can be used for LULC, some sources were given, of which Bhuvan, the ISRO Indian government's data base was used. And we showed a lot of maps, a lot of attributes and how to use it, how to document the land use land cover, and add overlay options.

We also looked at on the fly remote sensing data for LULC using Google Earth Pro. So, we did a couple of things in one timeframe like we looked at the Bhuvan 2015-2016. We looked at USGS 2015-2016 which is the NASA and we also looked at Google Earth 2015-2016 which is our College of data from Sentinel ESA, which is European Space Agency and NASA. So, so much of data. What we have done is, we have compared and then beautifully we saw that how the forest has been converted to rubber plantation in one part of India. And then the other part we saw the Ganges water, a lot of erosion, land being submerged and not the land was not treated properly because of extensive agricultural activities.

So, now we are continuing week 9 as I promised, we will jump into the hands on tutorial this week. In the next class, we will add tutorials so that you can download a particular dataset and do very minimal land use land cover classification. Again, these are like masters projects 10 years ago.

Now, because of the click of button and quick download of data, you can do it within 30 minutes, so we will show you how to do it. And then with good accuracy and stuff, you jump back and forth between multiple datasets. So, in this course, I am not using one, only one

single data set, we jump back and forth to different datasets, because the beauty of addressing these complex issues is you have access to multi source and multi-dimensional data.

You just do not trust only observation data, you do not just focus only on satellite data, we merge them together. And then somewhere we make a data product. So, in week 9, what we will be doing is, we will look at the types of LULC classification, we now know that land use land cover classification is very important. However, we do not know like what are the different types. So, let us go through the types. And depending on your time, your data availability and cost, you can pick and choose which scheme you want to do.

I will mostly focus on open source data, open source software, so that you can do it, and everyone can do it just with access of internet and a computer. So, this is what we call as giving everyone equal advantage initially, those who had the software only could make these maps. But now through this NPTEL course, we are propagating into the idea that everyone can map everyone can contribute to remote sensing and GIS analysis, especially for rural development.

So, as I said, this week, we will have one hands on land use land cover tutorial, please go through it carefully and then do it yourself. One recommendation is please first do what was discussed in class and then you can do extra data and analysis always this is the way you should learn you should first reproduce what is produced in class, because we spend a lot of time, we make sure that the data is correct, we make sure that the steps are correct. So, please reproduce our results, then you can do some data now you know how to do it and you will be able to do it by yourself.

Issues in water availability for crop irrigation, we will discuss a little bit especially, remote sensing per be irrigation. As I said one of the dominant features in this course is that it is for rural development and rural development is around mostly agriculture. India still an agrarian nation, most of the population depends on agriculture for livelihood. So, if we give them good access to improve the agricultural productivity and lessen the losses, then rural development will occur.

So, on that note, we will look at some issues and availability of data and then how remote sensing can help. Remote sensing for crop growth, acreage, health and some indicator like NDVI we will use we will look at Bhuvan NDVI. And then we will also look at the new need directions future, research directions that are needed to identify crop type, crop acreage and crop yield. Slowly we will be getting more into applications. Because the tools are now introduced to you. It is the basic tools that we will be using, but what kind of applications we can do, what kind of research we can do is of utmost importance.



(Refer Slide Time: 9:06)

So, let us discuss the first types of land use land cover classification. Mostly, there are 3 types, depends on which book or reference you use, but most mostly it is 3 types. In some books that will be just 2 types, which is unsupervised classification, and supervised classification. We will be teaching the 3 types. So, unsupervised means no classification details are there, which means you have an imagery and now you have to classify. Let us draw an example. You have a 4 by 4, 2 rows and 2 columns, 4 pixels. So, suppose you have 4 pixels. That is the image and now you have to classify each pixel, each grid. So, this is what we are going to look at how do you classify?

So, the basic methodology can be divided into 3 types one is unsupervised where you do not know what green is? What yellow is? What blue is? What red is? We do not know look at the ocean, river, lake or even in Mumbai if you fly during the monsoon, you will see half of Mumbai having blue sheets on top, it is not water, what is it? It is a tarp that prevents water from entering the house. So, most of the houses even IIT Bombay we have a lot of these blue sheets on top. So, these have to be not marked as water it is not water correct.

So, then we have the green color, it could be golf court, it could be an irrigation fiel,d sugarcane or it can also be a park. So, all these are not the same for agricultural development and demand you need to identify what it is. So, what is the step is as the diagram says there is generate ISO clusters which means ISO clusters is same color ISO is same. So, depending on the color, you will differentiate it.

So, now I have given 4 colors, suppose my computer can only understand the 4 colors, then it will say this is green, blue, yellow, red, but it does not know which what these are because I am not telling it and the computer cannot miracally find it as green as vegetation or not. So, this is unsupervised classification, no classification detail is available, I do not know what green is green could be..., let us say green is plant but what plant is classification?

Please remember in the previous class, where I share that there is multiple greens and one is broadleaf trees, one is pine trees, one is plantation, one is rubber plantations. So, all these are green, but you should tell like which green exactly represents the deciduous forests and rubber plantation.

So, then pure clustering approach so ISO same colors are clustered by the computer once you click the tool for this, what it does, the computer, it will take all the green of a particular wavelength a particular color, it is not green, it is just green, there are at least 200 300 shades of green. So, it will take each shade and then club it or it will tell you the histogram the frequency of colors, and based on your telling the computer how many classes you want, it will give you 4 so let us say here I say 4, it will give me 4 classifications, if I say 3 what it will do is because every color is a function of the primary colors, it will mix and match the pixel and then give you 3. Suppose 2 black and white.

So, pure clustering approach is applied by the computer, it will take all the greens and cluster into one group, take all the blues clusters into one group and then give you the output. Properties/labels are superimposed. So, the labels we do not know once the color comes, then we think about it and say there cannot be for example as I said the Bombay or Mumbai sky view application, we definitely know that it is not water, then I would say that blue is tarp, because that is my field experience, a personal experience from the field you can share.

Same thing I work a lot in Tamil Nadu and Maharashtra, the Ganges regions I know for example, in a particular area that cannot be blue suddenly it cannot be just water suddenly, it could be a farm pond. Same thing the green color, extra green color that comes cannot be in Chennai it cannot be agriculture it is mostly golf or a park.

So, moving on, we have the next part is the unsupervised classification is done now we will go to the next part which is supervised classification. So, in supervised classification, we give the labeled input data for example, I say these greens, clump it together and call it agriculture, especially paddy. So, once I tell that these greens scrub together, all these greens will be clumped together and they will be used as green for paddy in the classification. So, what will be the output?

The output will be generation of clusters with some labels, which are the same as I have given. So, suppose I will find labels, only those 5 labels will come not more than that sometimes it will be less, because suppose that color is not available, it will not come. For example, you have water river flowing, I give blue for the classification and classifying, classifying suddenly the non-monsoon, summer peak, summertime, there is no water in the river, so it will be brown, so there will be no water in that class.

So, you can always be below the number of classifications not above. So, here as I said, we start with create a training set, which is I will say green is equal to agriculture, blue is equal to water, red is equal to damage crop, so I create that bank or training set, then I give it to the computer, or the program GIS program QGIS, and then develop the signature file. So, this is called a signature file, or a file that has what color represents what.

See in the first aspect, I am surprised that is not there, because we do not know, we do not know the labels, we do not know the colorings. So, we did not give it. But, in the second step, we will give the color plus the name as a code and give it to the computer which is developing signature file.

Now, using that file, the computer will run the classification. Basically, all the greens, it will cluster and then go to this file, the signature file and say what is the green, it is agriculture, it will label in agriculture very very accurately. So, this is uses label input data generates cluster clusters with same labels.

One example is you are able to give the coloring by yourself to the computer or you can go to the image and then say I know that this is barren urban land in the barren class under the barren classification, so, what you could say is you can take this color so, what you are doing you are saying I am creating a training sample, I am taking this color.

So, within that box, whatever color comes, all the colors with that with that small box, the average color and the wavelengths will be used as urban and for that urban give red color. So, now the paint is going to be applied all those black and white you could see which is roads or a urban cement coverage, all that will be classified as urban and then coloring will be given and 278. So, number of pixels having that color is 278. Everything will be color coded. So, this is supervised.

Now, the semi-supervised comes where there is a mix of labels and unlabeled classes. For example, you do not know all the pixels color in your image. You know, sure that I know in Chennai, this is cement roll, this is barren land, this is the golf field, but I do not know some green colors because Chennai, Bangalore, Mumbai still have a lot of trees. So, the tree could be a green or it could be a paint that is applied on top of a roof top. So, what will happen is you will apply some labels and some labels you will not touch, let the computer run and give it to you as a cluster, unlabeled cluster and you will label it. So, this is mix of labels and unlabeled classes is semi-supervised.

So, now, the question comes is which one is more accurate? Yes, the more accurate one will be the surprise version. However, it is also time consuming because you need to have high level of confidence in your data, collect the data for all the labels and give the computer there is also a cost involved you have to go collect. So, the collecting cost the training cost everything is there. So, all this has to be input. So, always the most expensive is the high accurate in this case.

Then the semi-supervised is good also because at least half of the labels is correct the others are not but once the clusters happen, then you will apply it or go to the field and then correct it and say this is not plantation it is forest or suppose it is a forest it is a rubber. So, there is a difference between all these 3 types of classification. Which one is accurate is more supervised classification. But it also means that all the classes have to be labeled, and your computer may take longer time for running the analysis. However, in rural development scenarios and real-life applications, you do not have to be very, very accurate because accurate data takes time and cost.

So, some errors are... so that is why you see a lot of papers, scientific articles coming with semi-supervised classification and unsupervised classification. You could say that semi-supervised is part of supervised where some labels are missing. So, it is data gaps. So, that is where in some books, it is only 2 types unsupervised and supervised, in the supervised there is semi-supervised where not all labels are known, only some labels are not.

(Refer Slide Time: 20:37)



So, let us look at a schematic of the land use land cover types. So, types of value is the classification we have unsupervised, supervised and semi-supervised in hyperspectral data which has multiple spectral very, very fine wavelengths. So, for example, a green will be divided into 200 greens, that is hyperspectral and you have unsupervised and then you run the unsupervised tool, it comes clusters. So, that is one classification.

In a supervised, you give abundant references, references are labels, you let the computer run and give you the clusters, then you apply the labels or you apply the labels and then a transit, then you have a supervised classification with labels that is called supervised. In the first part, you do not know the labels. So, you will have to figure out what the label is after the clustering. Here, the clustering comes with labels.

In the semi-supervised, few references are there. So, half of the labels are being populated, whereas half are not populated, there will be some errors. So, this may be full of errors. This may be least of errors, this is the least, least of errors because it is all with labels, it is more expensive. I will say unsurprised is not errors, I will say it as no labels. So, data gaps, so you will have to put the data later. And then say what class it is.

(Refer Slide Time: 22:00)



Let us look at some examples. In this study, we see that Ismail has taken some. So, the grid is their images, their image has been graded each pixel and each pixel you do not take ground routing. So, this is the references as I said, you will go to the field, collect spectral signatures. So, they will go to that single grid in the center, they will stand and then in their phone or an app, they will say what is present in that box. So, in that grid, they know the size 30 by 30 meters, for example. So, they will go to a 30 by 30-meter grid, after they print this.

First step is they print this image grid and then go to the same location using the map stand in the center of the grid. And then like my point was in the center, and then take a note of what color and what is there. For example, that is green color, I will say green is some type of vegetation, which is vegetables, let us say that is vegetables. And here is a forest. So, they are growing vegetables on the side because the colors are the same, you have the dark green and then you have light green there, some type of cropping. Then you take number 20, it is on the border number 28 is more urban, you could see that it is urban land.

So, this is how you randomize the grids. It is not a particular pattern, you cannot say that 1 after 2 another 1, and then have to do another one. No, it is random based on the spectral signatures. Based on your data, you have to say how many samples you need to collect. Suppose this is only black and white. And you just have 2 colors, you just go to the most black and then most white, take 2 samples, and then somewhere in the gray 3 third sample you are done. But here there is multiple colors.

And within the colors, there are different shades. So, you will have to pick each shade pick each color. So, just look at how many samples that has been taken. The best is taking each

grid sample, but then it is too expensive time consuming. It is not needed. That is where you have the computer to do the classification. So, you take representative samples, for example, this grid will be given the colors of all the sides, this 8 will 8, 9 and 14 can give colors of all this forest, you can take duplicates, also. So, 8 and 9 will be the same. But you need multiple samples, you can clearly see in this forest, they have taken a central value, they have taken a middle value and then the boundary values to document the forest.

So, using this data, they have the study has done a supervised classification, which you see on the left and also ran an unsupervised classification, which you see on the right. They have been very particular about the number of classes only 5 classes have been given, of which vegetation, so vegetation could include forests, crops, vegetables, anything, all green plants or vegetation, even golf come somewhere under as vegetation because you put in the golf field.

Then you have urban buildings, grassland, water bodies and then barren. So, the grassland can include the natural grass or artificial turf or golf. So, you have all these different classes and what has happened here is you can see that your unsupervised classification is almost perfect, it has classified all the colors together it did not have the labels.

So, this label will not come in this output, but looking at this and going to the field quickly you can put the labels so, for example, now you know all the green you just go to one green so, instead of going to all these 17 20 blocks, you can just go to one of the block because all of green same green just go to one of the block, write it as vegetation; go to one of the pink and then write it as one thing is representative, all the things you say that is urban, you can take another sample if you want grassland water bodies and barren.

So, this is the beauty of using supervised and unsupervised classification. There is, this is much, much more perfect. You could see that more grasses are here, some more green is there, the blue is bigger, etcetera. But this can also be helpful. And overall, the errors are very, very small when you convert this into soil management, water management, agricultural productivity, all these things. So, where does this data come from, you go to the field.

(Refer Slide Time: 26:40)



And I have also taught you a trick that you could use in your classes, which is your Google Earth. So, if you go to the same image satellite, go in zoom in and just look at it over timeframe, you will know that is that green and agricultural green. See this image was a one time snapshot. One time snapshot is difficult to understand what it is. But, if you go to the Google Earth, the same location go beyond particular time, let us say every month you go then you know that the crop is being put and it is growing the crop then you know it is an agricultural field.

Same way, if it does not change, it is a forest and if it is well maintained, it is a golf. So, all these things you can kind of assume by going to Google Earth and extracting the labels.

 Types of LUC:
 Marine Testinger and Approach of LUC: Constrained

 Strained Gaussi 12021
 Marine Testinger and Approach of LUC: Constrained

(Refer Slide Time: 27:28)

So, in the advanced level, we have multiple, multiple classifications let me go through some of them as per this paper. You have different types of techniques in LULC classification. There are manual classifications where you go to the field and draw it on a map like the Britishers did, that is one type. There is a numerical and digital technologies which is GIS. Hybrid technologies, which is both manual plus digital and then there is classification methods based on indicators, this is also GIS.

All the right hand side block is GIS, we will leave it out for now, we only look at the advanced levels. So, in the advanced techniques, you have artificial intelligence in LULC. So, those people who like to have big data AI ml in their profile, GIS can give you that, because all the data that comes into from remote sensing is really, really at a high, high volume and frequency. So, you get really good data at a bigger frequency.

So, then you have hard classification and soft classification. Soft is, it is empirical, it runs through models, whereas hard classification is traditional machine learning unsupervised, semi-supervised, and supervised which we discussed in the previous slide. And then the reason machine learning trends and the basic KDB of K knowledge discovery approaches. You have knowledge base classification, role-based classification, data driven products, reinforcement learning approaches and ensemble of methods.

In the recent machine learning, there are multiple, multiple developments, where hard classification is done by giving some input is kind of a semi-supervised input and the computer is smart enough it learns from the data, it trains the data and then projects it.

(Refer Slide Time: 29:29)



So, as a case study, I will discuss one paper from our group which got published in this journal. Journal of Agrometeorology, it is an Indian Journal, mostly through the association of Agrometeorology. Almost all agriculture universities are subscribing to this because it is a government journal, Association's journal and you could see that it talks a lot on the application. And one key application is the GIS application and the paper will be discussing just quickly on the advancement of data is the village level identification of sugarcane in Sangali, Maharashtra using open source data.

Why sugarcane in Maharashtra? Sugarcane belts are there in Maharashtra, a lot of and land is under sugarcane cultivation, especially in Sangali. A lot of farmer debts suicides are linked to these kinds of crop management issues in this area. And so, we would like to give some input. So, if you know the yield, if you know the acreage, you can manage the water, you can manage the soil resources and reduce the burden on the farmers. By reducing the burden on the farmers, you reduce suicide.

Farmer suicides happen because their crop is lost, they put a lot of money in the crop and the yield is not coming. So, now if we tell them that, the crop is not growing correctly, the acreage is not the yield is not going to come, they will stop investing, they can at least save a lot of money. And they can they can definitely work the next year and then come out. But if they do not, and then they just keep on pouring money in terms of fertilizers, water, labor and not get the money back then they are pushed to the brink of death. So, that is what was one of the biggest reasons farmer deaths for committing suicide.

(Refer Slide Time: 31:32)



So, in this study, what has happened is, we had a need statement, we wanted to map the farmers data using open source software, open source data so that everyone can run it. And there are a lot of advancements in classification. So, we wanted to take AI and ML methods, and then see which one is best. So, we had census data for accuracy assessment sensors, mean sensors from the ground, and also the field data where we did field visits and collected data.

And then we did two types of satellite data. One is Landsat 8 and Sentinel. Why because both of them have higher spatial and temporal resolutions compared to the Indian satellites. So, we use them. And then only the basic bands were used. Some indicators were made, the

classifications were done based on three semi-supervised, supervised classification, which is classification and regression tree, support vector machine and random forest.

So, in this, this is not a traditional method, as I said, these are AI and ML based models. So, it does not come under the supervisor, semi-supervised, but you have to give labels because you give a lot of data to the computer, you give a lot of labels to the computer, it will make the signature files and then classify. So, at the end of the day, the sugarcane pixel was classified. And what we found out that there are multiple models that we use CART, RF and SVM.

So, as I said, random forest support vector machine, but we use also different datasets. So, we use two datasets for each algorithm CART, RF and SVM. And we could see that everything is different. So, CART is not the same for Landsat data and Sentinel. Same, Landsat 8 for CART and RF is not the same, you can see both are horizontally they are different vertically also they are different. And the green is sugarcane that we are trying to look at. Grapes is also there in some regions in Sangali. Sangali is mostly sugarcane, a lot of sugarcane factories.

(Refer Slide Time: 33:42)





So, what happened is when we looked into the results, we saw that I could see so this is where I said we went and corrected. So, it is mostly like a supervise classification. When we did the accuracy assessment, the type of model used also plays a role and the data that fed into the model also plays a role. So, you can see the overall accuracy 78 percent was high using Sentinel to support vector machine, but support vector machine as an algorithm with Landsat 8 was not good.

So, 70 percent was the overall highest accuracy. And the producer's accuracy is that user accuracy is that so this model, this paper that I have shared is open source, please read it, and then see how this model has been created. It is an open source system and QGIS, anyone can run it, it is an advanced level, but I want to introduce the advanced level now, so that you could learn some of these techniques and then apply it for your research.

So, here we can also see that the spectral signatures and the enhanced vegetation index which is based on the bands is not the same spatially and temporally. So, temporarily, it differs because the growth period of sugarcane is different. The index is at higher value in November, October - November period, and then it is the same.

So, if you take a sample here, it is almost the same you cannot differentiate between sugarcane, wheat, grapes and banana, but here you can clearly see that the wheat is differentiated, then the grapes are differentiated the sugarcane is much, much less and so, now if I apply a filter in this 2 months or saying that only 0.2 I want or only 0, only EVI is 0 level I want all the other crops will be gone because this is 0.8, this is 0.6, this is negative 0.8 banana. So, only the green color will remain. Now, that is how classification happens. So, all the other pixels are gone. First the green is given us a value.

This is what AI and ML does. So, you could see here beautifully that the best model was as per the accuracies we had, we had both SVM Support Vector Machine model; highly accurate. And we had going back to the table is that we had better accuracies for Sentinel 2 and Landsat. So, you could see that Sentinel 2, this one was more accurate. And in this box, you could see that some browns are creeping in barren, but actually they are sugarcane. So, this is very, very important.

So, barren land does not require water, fertilizer subsidies. So, the government will think that why should I give my subsidy to farmers however, the farmers are given sugarcane in this parcel. So, if you do a small parcel, it is very small. But then if you actually accumulate all these data, it is a big data set. So, this is where advanced remote sensing techniques can help in accurately monitoring and mapping the data for land use land classification.

(Refer Slide Time: 36:55)



The other aspects of land use land classification is also forest cover biodiversity mapping, and then species mapping etcetera. All these are also done by government of India's biodiversity Information System website. I have given the website, it is also having tertiary links to rural development, because forest and the areas around the forest are mostly rural and rural communities have a very, very important engagement with the forest in terms of livelihood options, water resources, soil nutrients, medicine, herbs, everything they depend on forests, tribals, I am saying.

So, this is also a part of the land use land cover and those information can be taken from here. For this, I would like to conclude today's lecture. I will see you in the next lecture with a hands-on tutorial. Thank you.