Machine Learning for Soil and Crop Management
Professor. Somsubhra Chakraborty
Agriculture and Food Engineering Department
Indian Institute of Technology, Kharagpur
Lecture 27
Use of ML for Portable Proximal Soil and Crop Sensors (Contd.)

Welcome friends to this second lecture of week 6 of this NPTEL online certification course of Machine Learning for Soil and Crop Management. And in this week, we are talking about the use of machine learning for portable proximal soil and crop sensors.

So, in our first lecture of this week, we have discussed what is the requirement of different types of advanced sensors from the point of view of soil survey, as well as for site specific nutrient management and what is the interrelationship between the smart soil sensing and site-specific nutrient management we have also discussed.

And then I showed you the classification of proximal sensors, specifically proximal soil sensors. And we started discussing about portable XRF. Remember portable XRF is an x-ray fluorescence-based sensor and this field filled portable and it requires minimum cal consumables and they are non-destructive.

That means, they are not consuming the soil samples while analysing and they can measure 25 to 30 elements starting from Magnesium to Uranium within only 60 to 90 seconds. So, we discussed about the working principle of XRF. And also, we have seen the how PXRF elemental, PXRF base elemental contents of soil can be utilized for exploring different soil properties.

So, today we are going to discuss about the evolution of statistical approaches for PXRF and also side by side we will be seeing some PXRF application for soil and crop. Particularly in this lecture we will be focusing on soil. And in upcoming lectures, we will also see the PXRF application for crop also. And the major emphasis in this lecture will be given to describe how the application of RXRF utilized simple form of statistical analysis to advance machine learning based analysis for exploring different soil properties.

So, these are the some of the important keywords for this lecture. We will be talking about PXRF, compost salinity, then soil horizons, soil salinity and also soil heavy metal contamination.

So, 10 to 12 years back when the initial application of PXRF started in soil for exploring different types of soil elements, then we did an experiment in Southern United States and we used the portable exertive instrument to gather the information of elemental content in the peri urban areas that means, surrounding some industrial sites. So, soil samples were collected as you can see these soil samples were collected around two places industrial places of the city of Baton Rouge, Louisiana USA.

And these samples were analyzed using the standard laboratory measurement. They are digested and measured through inductively coupled plasma and atomic emission spectrophotometer. And then, we measured these heavy metals, arsenic, chromium, manganese, lead, zinc, iron, copper, cobalt and barium. Side by side we took the reading PXRF and then we try to correlate both the results.

And we have seen that for most of the elements, there is a good correlation between the PXRF elements as well as the standard elemental analysis result. So, that shows the applicability of PXRF for exploring the soil elements. Not only that using geo statistics, we have mapped the enrichment factors or in other words, we mapped the enrichment of a heavy metal in that soil using geo statistical approaches or pre interpolation.

So, that shows that not only this type of equipment is helpful for measurement of elements in the soil, but also for the environmental contamination point of view, they can be helpful for mapping the soil, special distribution of soil heavy metals. So, this was one of the first application of PXRF.

And then the PXRF was applied for measurement or prediction of soil properties through modeling through regression modeling. So, here we have used the simple linear regression and then multiple linear regression. Now, you know that PXRF cannot measure sodium and that is why chlorine which can be measured by PXRF was used as a proxy for predicting the electrical conductivity of the soil.

So, here, you can see, this is the log of electrical conductivity values and this is the predicted electrical conductivity values using the chlorine, using the Cl measured by PXRF. And the calibration R square in this case was 0.83 and validation R squared was 0.77 and that is using the simple linear regression. Simultaneously, we did also the multiple linear regression and where we combined the chlorine, sulphur, potassium, calcium, apart from that, we have also included some of the easily accessible or easily measurable soil properties like sand, clay and LOI.

So, these three sand, clay and LOI were incorporated as auxiliary soil predictors for prediction of soil electrical conductivity. And using the multiple linear regression, the calibration R squared, we got 0.90. And validation was 0.70. So that shows the importance of PXRF or how PXRF can be useful for measurement of soil salinity.

Soil salinity is an important indicator of soil fertility, not only that, using the standardized coefficient, we were able to identify which element were more influential for prediction of soil salinity and we have found apart from loss on ignition organic matter and sand, the calcium content and potassium content are also very much important when we consider this multiple linear regression.

So, this was the application and also that shows the presence of chloride and potassium ions and calcium ion in the soil, calcium salts in the soil, which may be the reason for soil salinity. So, that shows the applicability of PXRF for me predicting soil salinity.

Subsequently, the PXRF was used to predict the soil pH. Soil pH, you know it is very important soil property. And several hundreds of soil samples were collected from multiple states of US and we utilized the PXRF elemental content to predict the pH with an R square value 0.77.

Similarly, PXRF was used to predict the catenation capacity, the catenation capacity is a very important indicator of soil fertility. So, catenation capacity has also been utilized as an important property which can be predicted by PXRF. So, using the PXRF elemental content, we can predict the catenation capacity of the soil with good accuracy.

Then another problem we face during a pH measurement is destructive. That means, you have to sample the soil, you have to take the sample and then you have to process the sample before you can make a saturate paste or you can do a 1 is to 2 soil water or 1 is to 2.5 soil water suspension or if you want to extract saturation extract. These are time consuming, but what happens when the soil is frozen and this type of condition is mainly seen in Tundra Region and were the soil is sometimes shows the presence of ice flakes.

So, in that case it is very difficult to take those soil samples and dissolve in the water. Also, sometime it is difficult to do any measurement in monoliths in preserved monoliths. So, in this type of condition to resolve the issue of measurement of soil pH, we extended our application of PXRF for measurement of soil pH. So, we collected several soil samples from Alaska which are permafrost soil.

And these permafrost soils pH were measured via elemental contents measured using PXRF. So, here you can see these plot shows the measure pH versus predicted pH in Alaska. In Alaska soil sample using PXRF reported elemental content. So, this shows the relationship between the predicted pH and the measured pH by a portable XRF with a field Geochem Mode. So, there are two modes we have used one is Geochem Mode, another is Soil Mode, and we scan the soil both in the field as well as in the lab.

So, this shows the results for regression results for field Geochem Mode and this is the laboratory Geochem Mode. And this is the, third one here you can see the field soil mode and the fourth one is the laboratory soil mode. And in all these conditions you can see more or less, of course, the laboratory applications are more precise. However, in case of field applications also we are getting decent R squared values. So, that shows the applicability of PXRF for measurement of frozen soil or permafrost soil pH.

Not only we predicted the pH of this permafrost soil, but also using principal component analysis, we are able to segregate the soil samples coming from different sites. So, here you can see the scree plot of the first principal, 11 principal components of, and then we can see the PC1 versus PC2 plot to qualitatively separate samples from five different sampling sites using in the laboratory Geochem Mode.

So, based on and we can see that these clustering of the samples were due to some of their similarity in the parent material as well as dominant vegetation and also management strategies which impacted the elemental content of those sites. So, the elemental content variation due to the variation of parent material as well also geology and also different variation of the soil properties can be easily identified using the PXRF.

Similarly, here shows the scree plot showing the first principal, 9 principal component and this is in the laboratory soil mode. And similarly, here also we are able to cluster the soil sample based on their relative similarity in geology, as well as other soil properties, which left an imprint on the elemental content on these sites.

Then, we apply this PXRF for base saturation percentage prediction, base saturation percentage is another indicator of soil fertility. And in this research in 2018 Rawal et al in our group, we have also proved that we can use the elemental content from PXRF for predicting the base saturation percentage. And we have compared different types of model, one is GAM model or generalized additive models, and then multiple linear regression models, regression tree model and random forest model.

And you can see here these are the plots of four different models and we have seen that using the random forest and also. In case of BSP, we have made, actually we model two properties one is BSP, base saturation percentage and another is CEC or catenation capacity. And if you take a look at these validation R square RMS in RPD values, you can see that in case of BSP this regression tree perform better.

Whereas, in case of CEC, the best performing model was the GAM model or generalized additive model. Now, we have also seen the variable importance of these potassium magnesium and calcium both for random forest as well as the regression tree approach. So, here we have already discussed these multiple linear regression model and also the regression tree model and random forests model.

However, generalized additive model is another very flexible model, which is an adaptation of the linear regression model. So, which allows to model the nonlinear data while maintaining the explaining power. Generally, if you want to model it very nonlinear data using a linear model, it fails miserably. So, that problem is being addressed by these GAM model or generalized additive model which can learn nonlinear feature.

So, here you can see this is the representation, linear regression or ordinary least squares regression representation where these beta 0, b1 and b10, these betas are the slope. Whereas, in case of these generalized additive model, these beta 0, beta 1 and beta n are replaced by this s0, s1, sn, which we call the smooth function which are nothing but splines. So, this by applying this smoothing function, we relax the restriction that the relationship must be simple weighted sum and instead we assume that the outcome can be modelled by a sum of arbitrary function of each feature.

So, if we mathematically represent the smoothing function, it will look like this here the k represents the weights and function per variable in the equation. So, that makes more flexible, this model more flexible and much less linear than our linear ordinary least squares regression. So, this is a very important regression, this generalized additive model and it has been used in several data mining applications for learning the nonlinear features in the data set.

Now, another application from the pedological point of view is of PXRF is most of the time the times the horizon boundaries in the soil profile are identified visually and they are described visually and qualitatively. But to reduce the dependency on qualitative description of the horizon boundaries, we have tested the PXRF elemental contents and their variation for replacing this qualitative nature of horizon delineation to with the quantitative framework.

So, what happens here you can see there are different soil profile and these soil profiles are differentiated into different horizons like AP horizon AP2, BT2, BT1 and BT3 these are different soil profile they are described in different horizons and these differentiating horizons are given by the experience pedological. And these dashed lines are showing the transition between the horizons.

So, we have calculated the elemental content variation in these transition zones, and this was denoted as DE, this was collected using the formula where these formula for calculating the formula, we have used the principal component analysis. And using the significant principal component, we have calculated that formula. In other words, we have calculated the difference in the values of one horizon from its overlying horizon.

And we have plotted that and we have seen that these DE index which is calculated by PXRF elemental content aligns very good with the qualitative description of the horizons. So, that shows that PXRF now can be used in the field for better horizon identification, it is even performing better than using the clay difference and other soil property difference also.

So, that shows that the applicability of PXRF in the domain of pedagogy, which is -- And also this is gaining more and more importance nowadays for soil property delineation. So, this was one, this type of application made some paradigm change in the field of soil pedagogy.

And then, we have also used this portable XRF for measurement of compost salinity. Now, you know compost is an important additive for maintaining the soil fertility. And in this research, we measured the compost salinity using the principal component regression. Here you can see first graph is showing the principal component scree plot. And from there we have selected the principal components and then we predicted here, we can see here we have selected two principal components for making a biplot.

And you can see here that first two principal components cumulatively show shows around 50 percent of the total variation. So, also using the principal component regression, which we have already discussed in our previous lectures. So, using principal component regression, we got an R square value of 0.80.

And also, you can see that, these are the standardized coefficients. From the standardized coefficients we can see that zinc, potassium and then chlorine and these are the major factors or major contributors for prediction of soil salinity by PXRF instrument. So, that shows the application of PXRF for compost salinity.

And then we have also used this instrument for predicting the compost CEC, compost CEC is an indicator of the fertility of the compost and using the PXRF elemental content, it is now possible to predict the compost CEC. Here, we have used the random forest regression and this shows the relative importance, random forests relative importance, where we also seen the presence of zinc, copper, titanium and rubidium are the most infringing parameters as far as the prediction of CEC is concerned.

And also, we have produced the biplot of CEC using the principal component analysis, these are the scores and we have seen the explored the relationship between the elements as well as the sample CEC values.

Also, the PXRF was extensively used for prediction of soil parent material. So, soil parent material is an important indicator of soil weathering pattern. And if you understand, if you know the parent material of the soil you can predict different soil properties you can assume different soil properties.

So, this research was done, was executed in Brazil in 2019, where PXRF elemental content were used for predicting the soil parent material and three types of regression machine learning approaches where used, random forests support vector machine and artificial neural network. And we have calculated the Kappa coefficient, we have already discussed the Kappa coefficient in our previous lectures. And also, user's accuracy and producer's accuracy.

Now, what is user's accuracy and producer's accuracy? The next slide will clarify these things. So, user's accuracy you can see here suppose, there are four different classes, water. forest, urban, there are three different classes water, forest and urban. So, here we can see that the these are the reference data and these are the classified data. We can see the correctly classified water sample is 21, for forest it is 31, and for urban it is total 22.

So, producer's accuracy in case of water is basically correctly classified reference site by the total number of reference sites. So, here the correctly classified reference site is 21, reference samples and the total number of reference samples is 33. So, 21 by 33, so this is the producer's accuracy. Similarly, which is 64 percent. Similarly, for forest it will be 31 by 39, so 80 percent. So, this is call producer's accuracy.

Now, what is user's accuracy? So, user's accuracy, you know here will be for water, it will be 21 correctly classified sites by the total number of classified sites 27. Similarly, forest it will be 31 by 37, so 78 percent and 84 percent, in case of urban it will be 22 by 31. So, this is the difference between producer's accuracy and user's accuracy. So, we have seen that using the

PXRF elemental content and machine learning approaches it is possible to predict the different parent materials.

And you can see here this graph shows the prediction accuracy and prediction accuracy in terms of Kappa coefficient and overall accuracy for prediction maps of soil parent material through elemental content in A, B and C horizons in Brazil. So, they have calculated, they have considered three horizons A, B and C horizon. So, this is for A horizon, this is for B horizon, this for C horizon.

So, for all three horizons we have seen that the overall accuracy is quite high. So, that shows that PXRF can be utilized for prediction of soil parent material. These graphs shows the artificial neural network, support vector machine and random forest-based comparison of, classification comparison of Ferrous, Geothitic and Hematitic parent material.

And here they have compared the producer's accuracy and user's accuracy. So, this is for neural network, this is for support vector machine and this is for random forests. They have calculated and compared the producer's accuracy and user's accuracy for all these three classes and for using these three machine learning approaches and deep learning approaches.

So, not only the PXRF based elemental contents were used for classifying the parent material, but also they have used, using this scheme they have also mapped the soil parent material, as you can see here, artificial neural network for A horizon, this is also ANN based B horizon, this is ANN based C horizon. Similarly, this is support vector machine based A horizon, B horizon and C horizon parent material distribution, then it is random forest based A horizon, B horizon, C horizon distribution maps.

So, guys, now, I hope that you are now under, you have gained enough knowledge that how the evolution of PXRF happened from very elemental statistical analysis like simple linear regression and multiple linear regression and how we went to the app higher machine learning approaches like advanced machine learning approaches like random forest, artificial neural network and support vector machine.

So, we will let us wrap our lecture here. And we will start from here and we will also see in our next lecture, what are the other application of PXRF, and then we will also see the sensor fusion between PRXF and other portable proximity sensors.

So, we will discuss some other aspects, but these are the references for this lecture. And thank you. Let us meet in our next lecture to continue from here. And then we will be discussing the other application of PXRF for soil and crop. Thank you.