Next, I’d like to talk about what we have done with NIR spectrum. So first I ‘d like to introduce about the NIR spectrum. NIR is short for Near-infrared Spectroscopy. It is based on molecular overtone and combination vibrations. The molecular can only absorb the radiation when the infrared radiation has the same frequency as the vibration frequency of the molecular. So in the picture we can see that x axis represent the wavelength, and y axis represent the Reflection ratio. The wavelength region where the curve suddenly drops down, means that there are Bond-Absorption.

So base on this character, we use NIR spectrum to achieve two targets. Here is our working flow. First is to predict the chemical Composite content, in our project is Kaolin, then we also did a Classification project basic on the feature from the curve. For content prediction, we select the a typical wavelength region, which is character of Kaolin with Hydrogen-oxygen bond. Then we get data from the experiment with selected wavelength and train the model with different method and use this model to predict the Kaolin content in new Dataset

For Classification we also try to use different method to reduce the feature, because the feature of NIR spectrum is too big. And get a new Feature Matrix in order to train the model

I will introduce the process of Feature Extraction and my college will in next Chapter show what we have done in the Step Machine Learning.

First I’d like to show you the detailed working flow about the NIR feature extraction. First we implement a pre-process method to remove the noise and smoothie the data. And besides, the standard normal variate can help us to reduce the influence from physical value. For Content prediction, we use the 1. And 2. Order of the curve. then we cut the experiment data and reserve the typical wavelength, in our project, is 1350nm to 1500nm. Finally we get the handled dataset and it prepare for future Regression

Since there are too much feature in a spectrum, in our project it is totally 224 wavelengths, that means the original Feature matrix is too big and redundancy. Some feature is not necessary for classification. So we need to find out which feature or which dimension has the most important feature. From literature we find that the most common method to handle the NIR spectrum is LDA linear discriminant analysis and PCA, Principle component analysis. Both of the method transfer the dimension of the feature to find the significant direction which can represent the feature best. Besides we also use the Matlab feature extractions function to find which feature is most relevant for classification

Here is our result of the Matlab feature extraction, the y axis indicates the score of the wavelength, The higher the score , more important this wavelength. And we select the top six wavelength as feature.

As I have mentioned before, we also use LDA and PCA to achieve the dimensionality reduction, and from literature and our compare we find that LDA has better performance for classification problem than PCA. And here is our result, It shows the scatter plot between the first two main component calculated by LDA. We can see that some class the white paper can be already good separated in this dimension.

And my college will show the Work about the machine learning.