**Data Analysis and Decision Making – II**
**Prof. Raghu Nandan Sengupta**
**Department of Industrial & Management Engineering**
**Indian Institute of Technology, Kanpur**

**Lecture – 31**
**TOPSIS**

Welcome back my dear friends a very good morning, good afternoon, good evening to wherever you are in this part of this world, may be you are in India or outside India. And as you know this is the DADM which is Data Analysis and Decision Making II course under the NPTEL MOOC series. And this total course duration is for 12 weeks which is 30 hours and each week we have which would to totally be 60 lectures. The reason being each lecture is for half an hour.

And the total duration as I said was 12 weeks and each week we had 5 lectures and after each week we have an assignments. And as you can see my name is Raghunandan Sengupta from IME department at IIT Kanpur. And we are going to start today the week, we are already com completed 30 classes we are going to start the 31st class; that means, we are going to the 7th week.

Now, if you remember in the last whole 1 week we had in this which was in the 6th week we had the all the discussions on what ELECTRE method and what was the basic concepts of ELECTRE, the normal ELECTRE method the epsilon ELECTRE method or normalization could be done along the rows or along the columns. And keeping in mind the utility function and also keeping in mind that in the direction where you are doing the normalisation the sum should add up to 1.

And technically you had two different matrices, one was the waited weightage one and we also consider initially I should mention that that they were m number of such alternatives; i is equal to 1 to m small m and there were n number of criteria or decision what you call or characteristics based on which you will take the decision for any alternative. And that change that with that was nomenclature through j, j is equal to 1 to n.

Now, once you are doing the normalization, remember the nomination would be based on the fact that what type of utility function which you have and we had already discussed in quite earlier, that if your utility function is quadratic your returns for the investment whatever the decision is would be normal in nature and vice versa. And we will try to basically that maximize your expected value of the utility or minimise the variance of the utility or whatever combination you want to do. Like you want to take the ratio of the mac expected value divided by the variance and rank them from the highest to the lowest or take the ratio of the variance to the expected value and write them in the lowest to the highest.

You can take different course of action. And in the ELECTRE method we if you remember we have we had the concept in general ELECTRE the cordon set and the discordon set concordant sets are those values of j's; j's means basically 1 to n those values of the criteria which give you a positive benefit for taking the decision a k with respect to a l. K and l are some values of n and a being the as the nomenclatures a was given was the alternative and you will basically club them in the con cordon set in the discordon set.

Then find out the concordant indices, discordance indices then found out the values of the mattresses which was capital C and capital D. And then you found out the corresponding values of g and f which was when you compare the values of using the concordance set and compare the values using the discordance set and then one once g n f are given you multiply each element of g that mean g i j. Here i and j are does not have any such conversation with respect to m and n. Multiply that value of g with respect to value of f, g n f as I said; f i j and you basically get the corresponding cell value of e.

E was basically the combined concordance, discordance metrics value and then you can basically find out which decision alternative helps you with respect to the combination of the alternates. And the weights if you remember or predefined depending on the overall weightages or overall importance of the person decision maker used to give on each and every alternative base for each and every criteria.

Now, in the epsilon case we consider the whole set of criteria being divided into 3 mutually exclusive and exhaustive sets, one was basically definitely positive when a k s value with respect to that criteria was greater than a l.

In another case the a k value was less than a l and another case when it was in between we basically had the in different set i such that you are indifferent whether you take a k.

Then when you reverse the decision in the case when you are taking a l and obviously, they would be corresponding value of concordance discordance and this concept of concordance and discordance need not be always positive and negative going hand in hand. It may be possible and we will see that later also in the TOPSIS method. I am sorry I am going a little bit more into the depth of the explanation.

In the TOPSIS method you will also see as you we saw in the ELECTRE method that if I like or if a decision maker likes alternative k a k with respect to a l for any of the criteria see for example, i 1 being suffix 1 it can and it gives see for example, some positive value some quantum see for example, 10. It does not mean that when the person takes the decision a l with respect to a k the value would be something minus 10 it can be see for example, plus 5 also; that means, whether I take k or whether take i a l I will get a benefit in both the cases.

So, when we are trying to basically find out the cap that matrix c or matrix d it basically that will give you the information that what is the cumulative concordance values and discordance values I am going to get by taken either the k th 1 or the l th 1 of the alternative based on all the corresponding criterias which I have.

And similarly in the case of those of this epsilon concept of ELECTRE I have basically divided the whole set of the criteria into 3 sets that is c, d and i. That means, c as usual the concordance matrix d as usual the discordant matrix and i would be the in different matrix based on which I will try to combine, when I combine I have basically try to find out what is the value of that alternative with respect to the criteria when I am considering the combination of c to i; that means, concordance with respect to indifference then concordance with respect to concordance with respect and in discordance.

Then basically again I go into the second set of comparison where I compare in difference to the concordance in difference to in difference and in difference to discordance and finally, I come to the facts where I compare discordance to concordance, discordance to in different and discordant to discordance. Based on that again we have those values of g and f, g would now we corresponding to c.

And f would be corresponding to c, then similarly g n f would be corresponding to i and g n f would be corresponding to the discordance. And then we combine them to find out the overall matrix e which will give me the comparison that count, cumulatively when I compare the concordance discordance and indifference sets and the values. So, what is the overall ranking of the alternative based on the cumulative criteria.

Now, in this method, in the TOPSIS method the basic concept is almost the same here also we will group them into 3 sets or 2 sets depending on whichever policy we are following. But the over and also the fact remains that the utility function will be used normalisation con concept would be used. So, when a person is trying to basically make a comparison of so they would definitely be m number of alternatives, i is equal to 1 to m. They would be n number of criterias, j is equal to 1 to n and will compare a k to a l a suffix k; that means, the kth alternative corresponding to the fact that I am comparing with the l th alternative based on each and every criteria.

But here the fact is that I will consider hypothetical the best set of alternative which will give me the best benefit some point, some point means some decision set. Similarly I will have the worst sets or the non ideal set, these are all theoretical values; that means, the best set would never may maybe utilized.

But you will try to find out in order to compare, similarly we will find out the worst set or the non ideal one and later on also we will find try to find out the comparison of the best ideal and the worst ideal in order to compare and what is the concept of distances. Now this distance functions will be utilised in such a way that will consider the Euclidean distance.

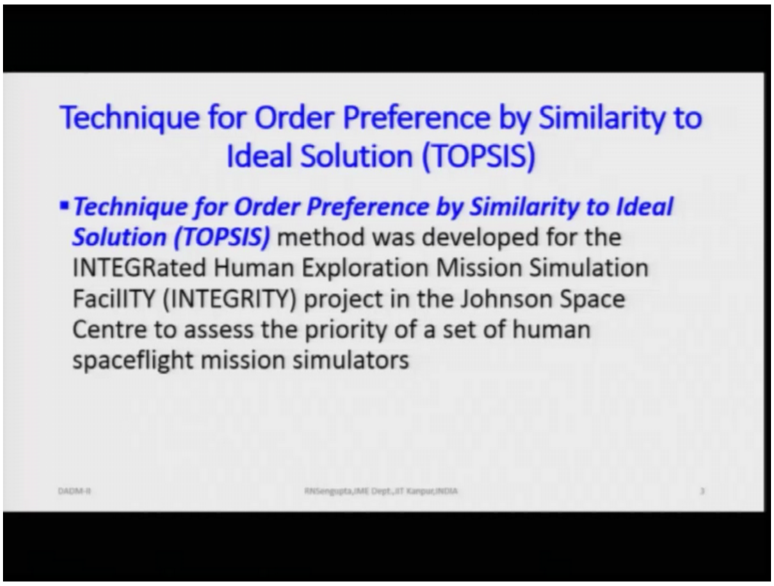
Euclidean distance means see for example, I am at x 1 y 1 point and I want to find out what is the distance from that co ordinate in the two dimensional one to at from x 1 y 1 to x 2 y 2 I will basically find out the difference between these points; that means, x 2 minus x 1 whole square plus y 2 minus whole y 2 minus y 1 whole square and square root of that.

So, we will try to follow this Euclidean distance which would have some semblance with the concept of the utility function which is quadratic because the quadratic concept will give us 2 important things. Point one that if quadratic utility function use we know that we can safely use the normality distributions for the returns point 1.

And point 2 also remember when we try to find the variance of the quadratic utility function and try to basically find out the minimum of that depending on the deviation, we will always see to the fact that will basically try to minimise the variance which is what we want. And if we remember in the quadratic utility function or the quadratic loss function which we have uses the squared error loss function this concept comes out time and again.

So, with this small introduction I will basically start the concept of the TOPSIS. So, this as I said would be the 31st lecture for and which is the starting of the 7th week.

(Refer Slide Time: 11:25)

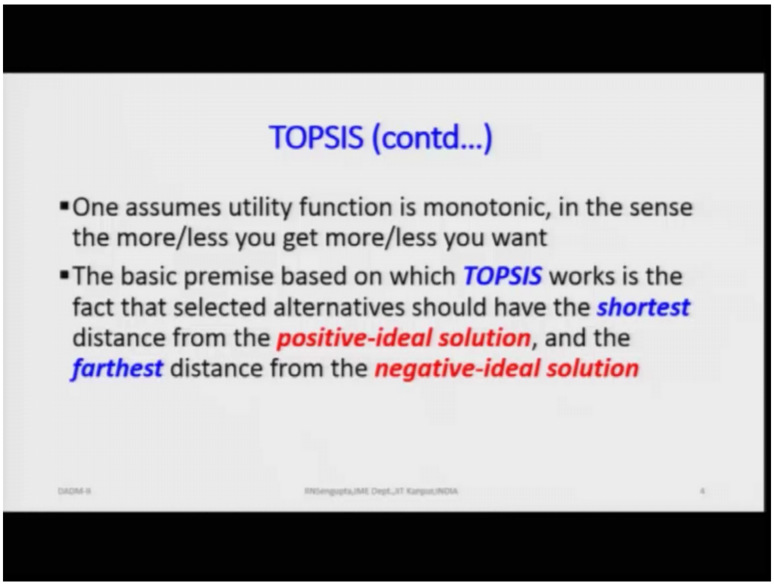



Now, the Technique for Order Preference by Similarity to Ideal Solution which is known as the TOPSIS and this method was developed for the integrated Human Exploration Mission Simulation Facility.

Which were basically utilised for the project in the John Johnson space centre to assess the priority of set of human space flight missions simulator. So, you they would simulating different type of space flight environment and based on that the concept that how you will basically try to find out the least distance and the worst distance or the best distance and the farthest distance in such a way that we can find out which are the best set of solution and which are the worst set of solution depending on the different type of criteria's' which you have.

So, you are trying to basically come compose a set of alternatives which are best, compose a set of alternative which are the worst based on the fact that you will basically find out the a found find out a cumulative overall weightages based on the criteria's. In this case of TOPSIS we will assume the utility function is monotonic it is increasing and decreasing.

(Refer Slide Time: 12:33)




In the sense that more you have it more you will want less of any positive value you will have less you will want of that. In the sense that positive less you are you I am basically putting the positive sense, but technically that if it is a positive thing more I take more I want; that means, the first derivative of that positive function would always be positive.

In the sense if you remember the concept of utility we have considered that you u which is utility function will be always increasing and that increase can happen in three ways. One is increasing at an increasing rate increasing at a constant rate and increasing at decreasing rate. But in all these cases we considered the concept that u prime would always be positive so; obviously, u double prime may be greater than 0, may be equal to 0, may be less than 0.

Now in the case when I use the word less it means that if there is some negative value, navigate value accruing due to my decision making ability then; obviously, I will try to reduce the level of negativity as far as possible; that means, positivity is more I want negatives less I want and this concept of increase and decrease will continue to which you have whichever level of the utility of the value of wealth or the value of decision I am going to take or based on the criteria.

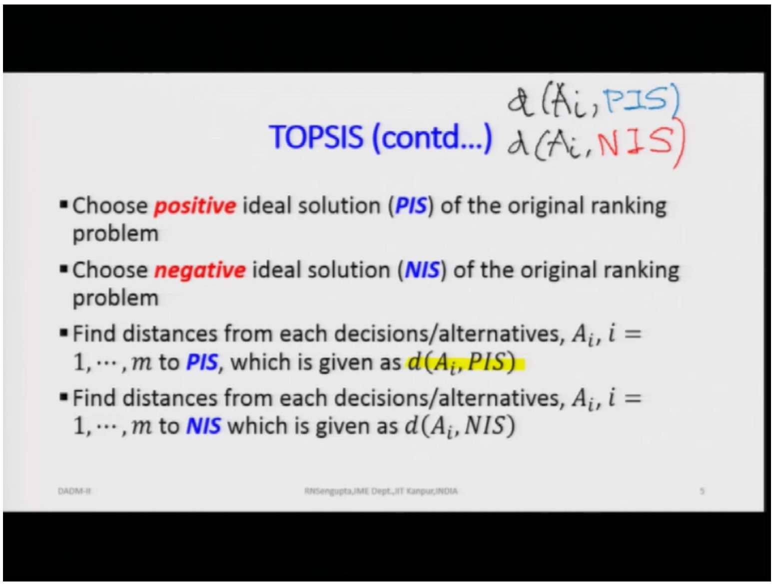
That means if I increase the value of the criteria and if it is giving me positive values I will continue increasing it. If I increase the value of the criteria and the value of that overall decision is giving me negative values; that means, it is giving me not positive or decreasing positive values then I will; obviously, I will try to decrease it as fast as possible. That means, make the values 0 in the negative sense. The basic premise based on which TOPSIS method works is the fact that selected alternatives; that means, a k and a l which you are taking and these are the sets of all combined together all alternatives which are basically from 1 to m.

So, hence the selected alternative should have the shortest distance from the positive ideal solutions and the farthest distance from the negative idle solution. So, what we want is that we will have a theoretical value, the positive set by taking a combination of the criteria and the alternatives when are there when they are be in combined. And we will take the negative value in the sense that for some sets of criteria's we will find out some set of alternative which will give us the negative value, the worst so called value.

So, we find try to find out the theoretical positive value theoretical negative value and try to compare the distances of each and every alternative from those theoretical positive and theoretical negative values and try to find out how far we are.

So, closer we are to the positive ideal solution better the alternative is, further we are from the negative ideal solution better we are so; obviously, it means that closer I should be to the positive sense further I should be from the negative sense. So, when I basically tried to compare the other way round that if I further away from the positive one and more closer to the negative one; obviously, it will mean a bad decision or bad alternative based on the cumulative criteria set which you which I have.

(Refer Slide Time: 15:56)




We will choose a positive ideal solution which would basically be nomenclature PIS, Positive Ideal Solution of the original ranking problem which we have. So obviously, they would be ranking set and ranking set in the sense their alternative based on which we have we would have given some weightages to the alternatives and we also would have different weightages for the criterias which are there in each and every alternative. So, based on that positive whole set which we have will find out say for example, the positive ideal solution PIS. Similarly we if we have the negative set of solutions.

For all the alternatives for all the criterias we will have a theoretical Negative Ideal Solution which will be known as the NIS and which will be based on the original ranking problem. Once this is done we will find out the distances from each decisions on alternatives. So, each decisions on alternative which is A i, i is equal to 1 to m. So, A 1, so distance from PIS will be calculated A 2's distance from PIS will be calculated, similarly we will find out all the distances still we find out A m.

So, distance A m is basically the alternatives distance from the PIS, which will be which we have. Now this distances if you remember I said that we will use a function or distance function which will give as distance between the two values which is A i and PIS so; obviously, you will have the values d A 1 from PIS. Similarly you will have d A 2, PIS and we will continue finding out till the m th one.

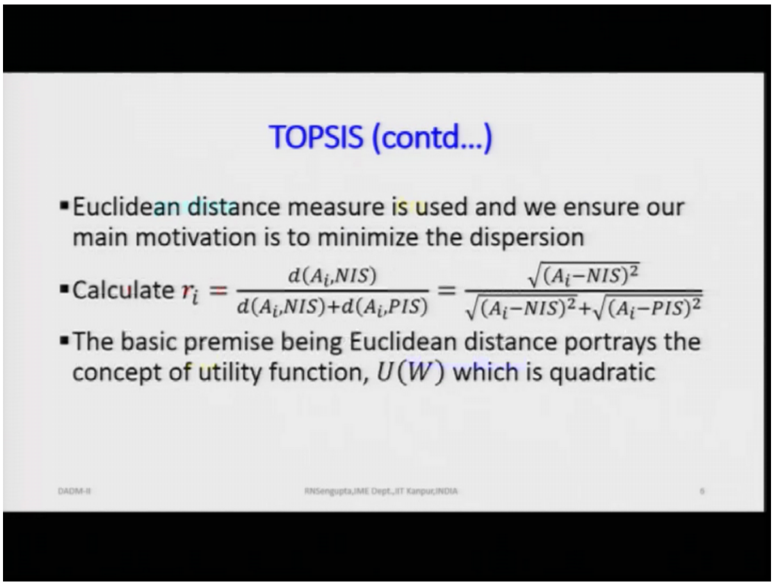
Similarly, as we have all the alternatives again from A 1 to A m, we will find out the distance functions from the negatives ideal set also; that means, the distances from A 1 to NIS, A 2 to NIS would be calculated. Now there are 2 questions; question 1 is let me remove this; the question is that the distance function which I assume for A i with PIS which is positive. Let me use different colour so it will be easy quiet let me use a different colour. So, it will easy for me to specify.

So, this is distance from A i, PIS is positive so use the blue colour, great and when I use the negative value. So, basically I would have this function, now the main question is somebody would definitely ask two things; number 1 what type of distance function we which we use for finding out the first function which is d A i with respect to PIS or d A i with respect to NIS. They can be any distance function like we can use the quadratic utility function hence the Euclidean distance, it can be either cubic distance, it can be hamming distance, it can be l 1 norm, it can be l infinity norm it can be anything.

So, we have to choose a distance function which basically gives us the best criteria, this what criteria I am using in a very general sense not with respect to the criteria as which are j is equal to 1 to m n, those are not based on the best criteria which will give us the best ranking system all the alternatives that is point 1. Point number 2 is that the distance function which a utilising for A i with respect to PIS or for A i with respect to NIS those distance function have to be same for our calculations, like this in the sense we cannot use the quadratic you utility function or the Euclidean distance to find out all the distances how far A 1 to A m are individually from PIS.

And we cannot use see for example, the n infinity norm trying to find out what is the distance function for all the A 1 to A m with respect to NIS. So, if we use a distance function it has to be same for both the sets of A i's. A i's with respect to PIS and A i's with respect to NIS so that is important to note. So, we will find out the distances from each decision alternatives similarly; that means, we will find out the distance A i with respect to NIS for i is equal to 1 to m.

(Refer Slide Time: 21:10)

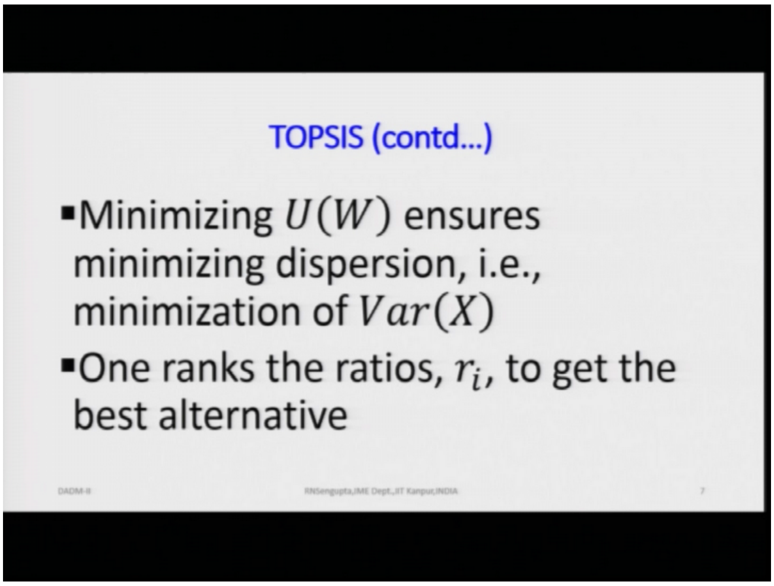



Now, we will generally use the Euclidean distance in the Cartesian co ordinate which will may it measure will be utilized and will be ensure are main motivation is to basically minimise the dispersion. Now as you are trying to minimize the dispersion the best measure of dispersion is variance. So, as you minimise the variance, minimise the sorry minimise the dispersion it automatically leads to the fact that you minimise the variance and trying to minimise the variance gives you the best measure that how you can rank the alternatives to find out the best set off ranking which will give you the best benefit for the decision maker collectively.

So, we will can calculate the index values r i. So, r i would basically be the ratio of the distance function of NIS with respect to the distance function of NIS plus PIS. So, if I basically calculate the values of r i’s it will basically be the. I am using the general formula it will be the square root of A i minus NIS square and A i minus NIS square plus a A i minus PIS square.

So, once we have the distance functions so the basic premise or basic assumptions being the Euclidean distance properties of the concept of utilities U W which is quadratic. And if you remember the quadratic utility function gives us a sqaudaral laws which is the best way trying to find out that how you can minimise the utility function.

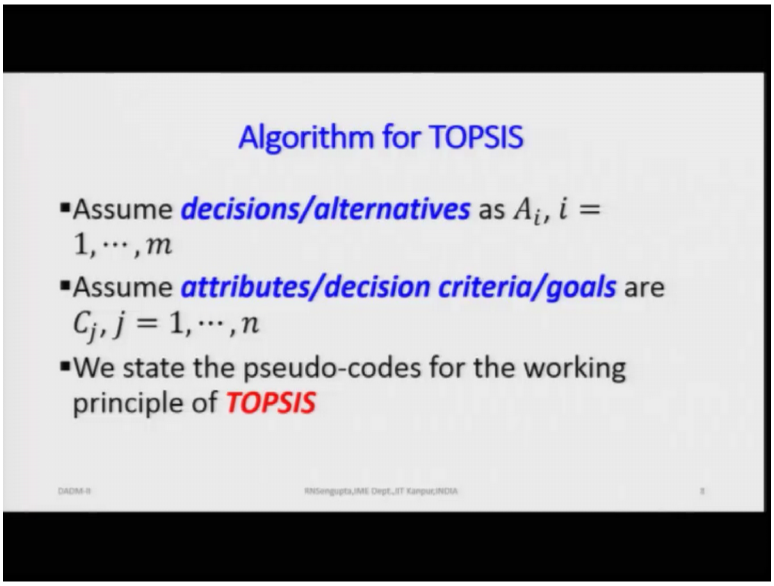
(Refer Slide Time: 22:46)




Minimising U W ensures minimising dispersion; that means, it gives us the value that when we try to minimise the dispersion we are also able to minimise the variance of that utility based on which we are trying to do the ranking. So, one ranks the ratios arise to get the best alternative and then we will basically find out the multiplication of metrics which is the decision matrix multiplied with by the weights to find out how the ranking can be done with respect to each and every alternative with respect to each and every criteria.

That means, we take one criteria try to rank A 1 to A m among themselves taking two at a times based on see for example, the first criteria. Then we take the second criteria try to rank A 1 to A m taking two at a time then we do it for the third criteria fourth criteria till the last one. And then we basically find out the cumulative overall ranking based on the cumulative course.

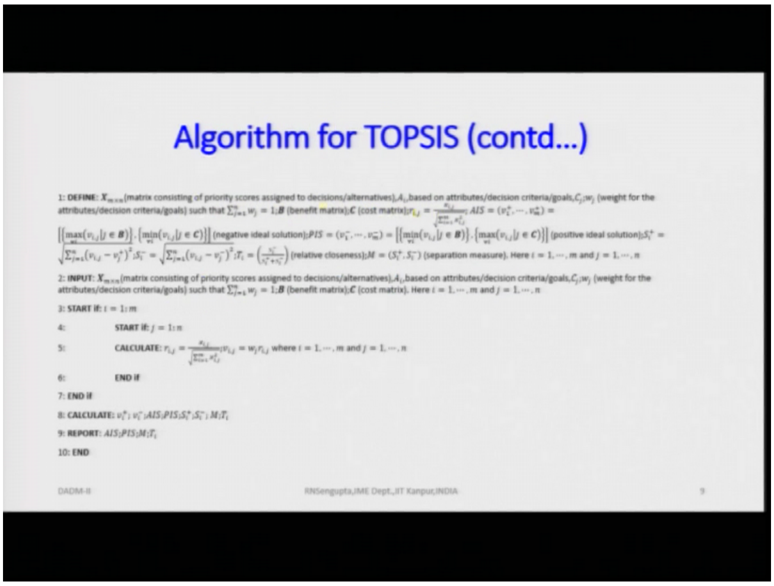
(Refer Slide Time: 23:48)




So, we will very briefly state the algorithm of the TOPSIS method here and try to follow one of the this algorithm in the very simplistic sense to solve a very simple problem we will assume the decision decisions alternatives are given as A i; i is equal to 1 to m and we will assume that attributes or decision criteria goals are given by C j where j is equal to 1 to n.

So, you will basically have A 1 to A n m as a set of alternatives and C 1 to C n as the set of criteria. So, we will basically try to find out the ranking of the alternates based on these cumulative collective group of criteria. We state the pseudo code for the working principle of the TOPSIS method and which is as follows.

(Refer Slide Time: 24:38)




Now, it looks complicated, but it is not. I will just state it and then proceed with the simple problem in a very very simplistic. So, we will basically define a set x which would basically be the matrix comprising of the so called utility, listen to this fact very carefully. So, x would be a matrix of size m cross n, m as in mango n as in Nagpur.

Why it is m and n? If you can; if you considered in each and every of this cell it will mean that if I am at the 1 comma 1 cell, it will mean that what is the overall weight level of criteria weightages or the level of utility which I am getting from criteria 1 for the alternative 1. If I consider say for example, the cell value 1 comma 2, then it will be basically be the level of criteria or the value of the criteria which I get when I am basically utilise that for the alternative 2; that means, I am going by the first row, top row.

Similarly, 1 comma 3 would be the effect or the overall value of the first criteria on alternative 3. Similarly the last one which I do will give me the overall value according to where I can find out the level of criteria 1 on see for example, the alt alternative m. Now this could be reverse also; that means, if I find out the transpose it can the overall concept will be that same. Similarly if I go to the second row it will be the level of output or the net value which am getting from the criteria 2 for each and every alternative starting from 1 to m.

Similarly, for the last value which I have I will basically have the level of the, the value of the criteria which is basically n and those values corresponding to the fact what are the weights given to the alternatives 1 to m. Now see, reverse it technically if I find it then the corresponding columns which I would have would basically be the each and every criteria and the rows would basically be the corresponding values of the alternative.

So, if I am considering m cross n it basically means all the m's. The rows are corresponding to the alternatives and the columns are corresponding to the values of the criteria. So, we will consider an m cross n and continue the with the analysis and it can be changed accordingly and we can solve the problems there is no problem in that.

So, if we consider the metric system priorities cosine to the decision alternatives where A i is based on attribute decision criterias goals and goals. Considering that time is running out for this at least for this the first class or the first lecture on the 7th week.

I will pause here end this this lecturer here and again start with the algorithm discussion so; that means, we will have much more time in trying to analyse and try to understand how the algorithm will be used. Thank you for your attention and have a nice day.

Thank you.