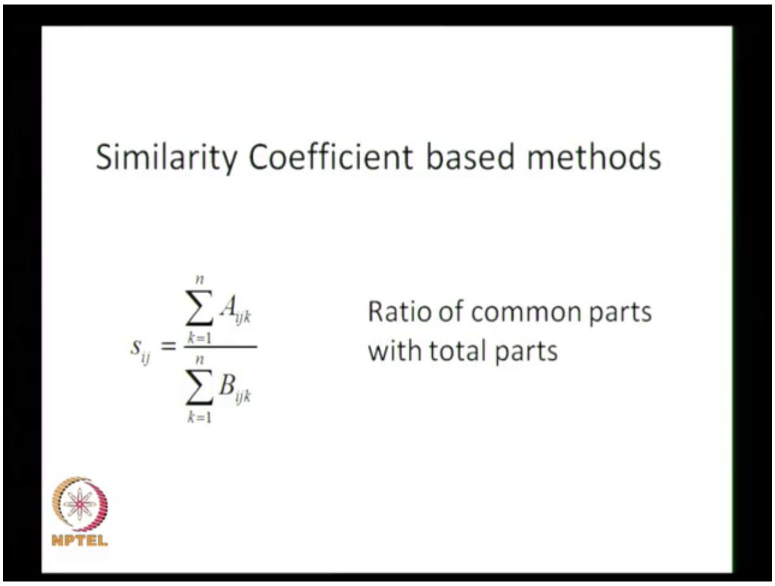
Manufacturing Systems Management

Prof. G. Srinivasan
Department of Management Studies
Indian Institute of Technology, Madras

Lecture - 09
Similarity Coefficient based clustering algorithm

In this lecture we continue the discussion on similarity coefficient based methods for cell formation. In the last lecture we started the discussion by introducing a similarity coefficient S ij given by this equation.

(Refer Slide Time: 00:31)



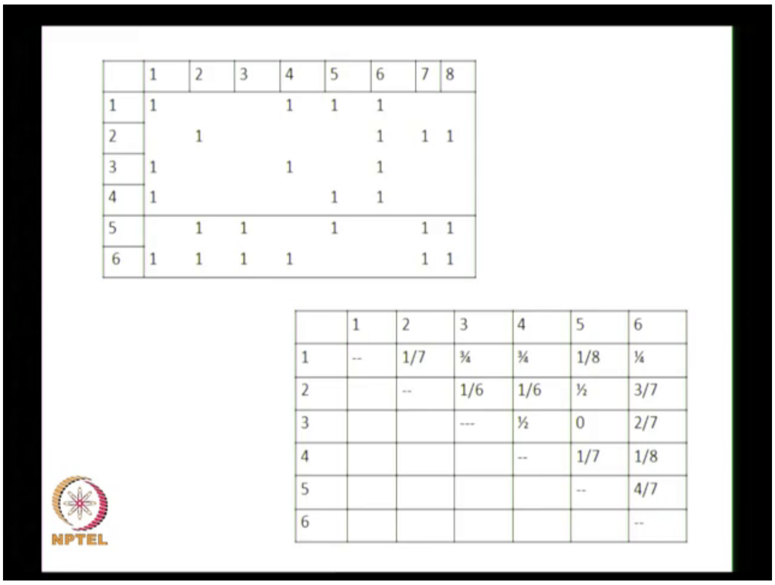
Similarity Coefficient based methods

$$s_{ij} = \frac{\sum_{k=1}^{n} A_{ijk}}{\sum_{k=1}^{n} B_{ijk}}$$

Ratio of common parts with total parts

And this similarity coefficient is called the Jaccard's similarity coefficient. Now, the similarity coefficient is calculated as follows using the example.

(Refer Slide Time: 00:41)



	1	2	3	4	5	6	7	8
1	1			1	1	1		
2		1				1	1	1
3	1			1		1		
4	1				1	1		
5		1	1		1		1	1
6	1	1	1	1			1	1

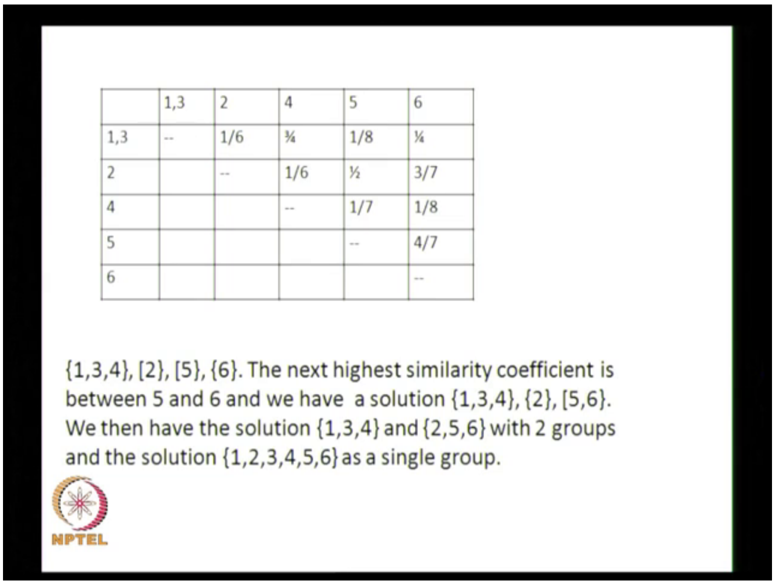
	1	2	3	4	5	6
1	--	1/7	¾	¾	1/8	¼
2		--	1/6	1/6	½	3/7
3			---	½	0	2/7
4				--	1/7	1/8
5					--	4/7
6						--

If we consider a machine component incidence matrix which is shown here on the top with 6 rows representing 6 machines and 8 columns representing the 8 parts or components, the similarity coefficient matrix can be calculated for both machines or rows as well as for components or columns. The similarity matrix shown in the bottom is calculated for rows or machines and since there are 6 rows this is a 6 by 6 matrix. We also explained how these numbers are obtained, a very quick recap if we consider the similarity between 1 and 2 we go back to rows 1 and 2 of this matrix and find out the number of common ones there is only 1 place where the ones are common. So, the numerator is the number of common ones and the denominator is the union of ones.

So, numerator has 1 which is common here denominator has 1 2 3 4 5 6 and 7. Component number 3 has a 0 0 pair and therefore, does not contribute to the denominator, the rest of them have either a 1 one pair or a 1 0 pair or a 0 1 pair and they contribute to the denominator. So, denominator is 7 numerator is 1, the ratio is 1 by 7 and the similarity coefficient is 1 by 7.

In the previous lecture we saw how to make groups from this similarity coefficient. We first said that we take the maximum similarity coefficient which happens to be 4, 1 and 3 or 4 1 and 4 and then we form the first group containing 1 and 3.

(Refer Slide Time: 02:43)



	1,3	2	4	5	6
1,3	--	1/6	¾	1/8	¼
2		--	1/6	½	3/7
4			--	1/7	1/8
5				--	4/7
6					--

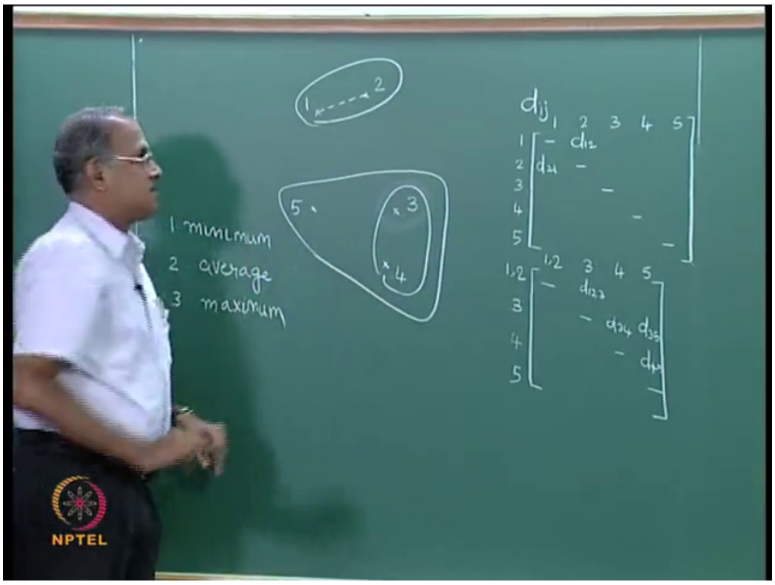
{1,3,4}, [2}, [5}, {6}. The next highest similarity coefficient is between 5 and 6 and we have a solution {1,3,4}, {2}, [5,6}. We then have the solution {1,3,4} and {2,5,6} with 2 groups and the solution {1,2,3,4,5,6} as a single group.

We redefined the similarity coefficient matrix because we have formed a group 1 and 3 and we said that the similarity of a group with another entity will be the maximum of the similarities and we introduced this method as a single linkage cluster analysis or hierarchical cluster analysis.

So, let me explain the basic principles of hierarchical cluster analysis and also explain to you what is meant by single linkage cluster analysis through an example and then show how it is used for cell formation both in the context of the algorithm that we have seen now, as well as in the context of several other algorithms. So, let me explain some basic ideas in cluster analysis as well as explain what is hierarchical clustering and what is non hierarchical clustering. So, we will also be seeing a couple of algorithms that use the ideas from non hierarchical cluster analysis. So, first let us look at what hierarchical cluster analysis is and related to single linkage clustering, we do it through an example.

Now, let us for a moment leave the cell formation problem behind and start looking at a very general grouping problem.

(Refer Slide Time: 04:20)



Now, let us assume that there are 5 points here, let us say these are the 5 points and let us call these points 1 to 5. Now, let us say we are interested in grouping them. Now, when we look at these 5 points and ask a question how many groups are there many times our answer would be two groups with 1 and 2 as one group 3 4 and 5 as another group is a very common answer that one would get if we post this question. But let us assume that there are 5 points here and each of these 5 is a group for example, if I said can you identify 5 groups from this we would say this is 1 group this is another 3rd, 4th and 5th.

Now, we can pose a question and say give me 4 groups out of it we will get an answer, if we say we want, 3 groups we will get an answer, 2 groups we will get an answer and say 1 group then we will say all 5 can be put into 1 group. So, how do we group these 5 points in considering 2 extreme solutions where one extreme is each one of them is a group, which means there are 5 groups the other extreme is all of them are in one group. So, let us assume that we also have a distance matrix d ij between or amongst the points where the distance between 1 and 2 is actually this Euclidean distance from between 1 and 2 similarly 1 and 3 1 and 4 etcetera.

So, let us assume that we have this distance matrix between 1 2 3 4 5 and 1 2 3 4 5. Now, distance between a point and itself is 0, but we do not consider it as 0 let us say we put a dash saying that we are not interested in finding the distance between a point and itself we are interested in finding distance between the points. So, this will be d 12 this will be d 21 which is the same as d 12. So, this will be d 12 this will be d 21 which is equal to d 12 distance between points 1 and 2 is the same as the distance between points 2 and 1. Now, if we want to group them the most logical thing to do is to look at 2 points which are close to each other which have the smallest distance.

Now, let us look at this and let us say the 2 points close to each other let us assume that it is 1 2. So, say these this is let us assume that d 12 is smaller than d 34. So, the smallest distance from here will be d 12. So, we can bring points 1 and 2 together. Now, let us say we have grouped them together and say this is this is one group, this is one group. So, now, there are 4 groups, how did we get the first group? We had 5 points initially each point represented a group and we had a distance matrix amongst the points which is actually the distance matrix amongst the groups and then we picked that distance which is the smallest and then we brought those 2 points together or we brought those 2 groups together. So, we said we have got this.

Now, we want to find out the next group. So, there are 4 groups and therefore, we need a 4 by 4 distance matrix. So, the 4 by 4 distance matrix will be like this 1 and 2 is one group, 3 is another, 4 is another, 5 is another 1 and 2 3 4 5. Now, the distance between a group and itself is a dash because we are not interested in this, distance between 3 and 4 3 is here 4 is here. So, it does not change. So, this will be d 34, this will be d 35, this will be d 45 they do not change. But what is the distance between the group 1 2 and the group 3 earlier 1 and 2 were 2 distinct points with say known coordinates 3 also coordinates are known. So, d 12 and d 13 could be formed because the coordinates are known.

But now, when we have brought 1 and 2 together how do we find the distance between 1 and 2 which is a group and 3 because. Now, 1 and 2 has to be represented as an equivalent point. Normally what we do is we try and take the centroid of this 1 and 2 here and we could say that this is the equivalent point and then say that distance between the group 1 and 2 and 3 is actually the distance between this point and 3 similarly this point and 4 and this point, and 5 and we can complete these 3 numbers here.

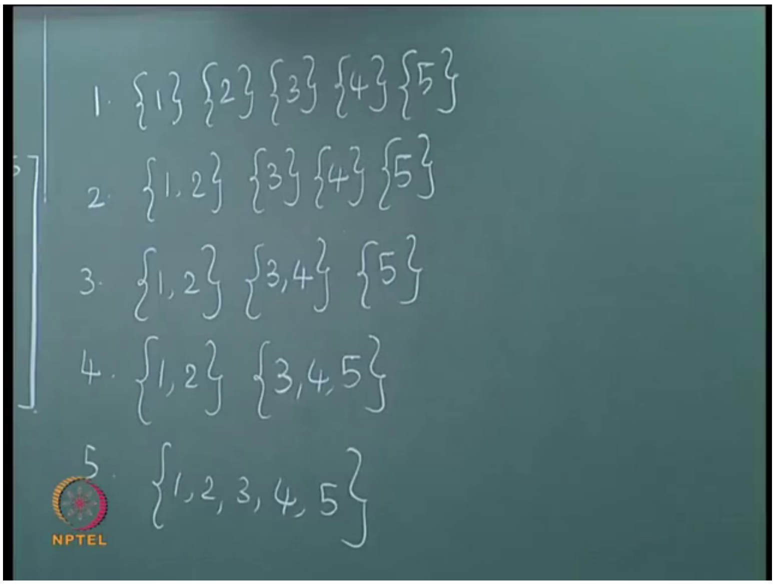
But there are several ways of actually computing this equivalent point or equivalent the distance between groups. Now, what we normally do is if we assume that we want to find out the distance between the group 1 2 and the group 3 we actually find out the distance between 1 and 3, 2 and 3 and whichever is smaller we take that as the distance between the group 1 2 and the group 3. So, in this case it will become 2 3 assuming that 2 3 is smaller than one thing it will not be something like 1 3 plus 2 3 by 2 is another way.

So, we will now, make an assumption that distance between this group and distance between this group, this is another group, represented by this point this is the minimum distance amongst the points. So, it is minimum of d 13, d 23 and that way we can complete this value. So, d 12 comma 3 similarly we can write 1 2 comma 4 5 and so on. As I said this equivalent distance can actually be done in more than one way one of which is to take the minimum distance the other is to take the average distance the third is to take maximum distance.

Now, we have taken the minimum distance and what we have done. So, now we know this matrix and let us say we are now, interested in finding the smallest distance in this matrix and let us assume the smallest distance happens to be for 3 and 4. So, now 3 and 4 becomes another group and in the similar manner we now, have to construct a 3 by 3 where the distance between this group and this group is the minimum of these 4 distances the distance between this and this is the minimum of these 2 distances and the distance between 5 and this can be taken from the previous one.

So, we have a 3 by 3 matrix and then a 2 by 2 matrix and all of them together. So, when we do the next iteration it is quite likely that this will be the smallest distance 5 to 3 versus 5 to 1, 5 to 2. So, the next group will comprise of this and the other one.

(Refer Slide Time: 13:24)

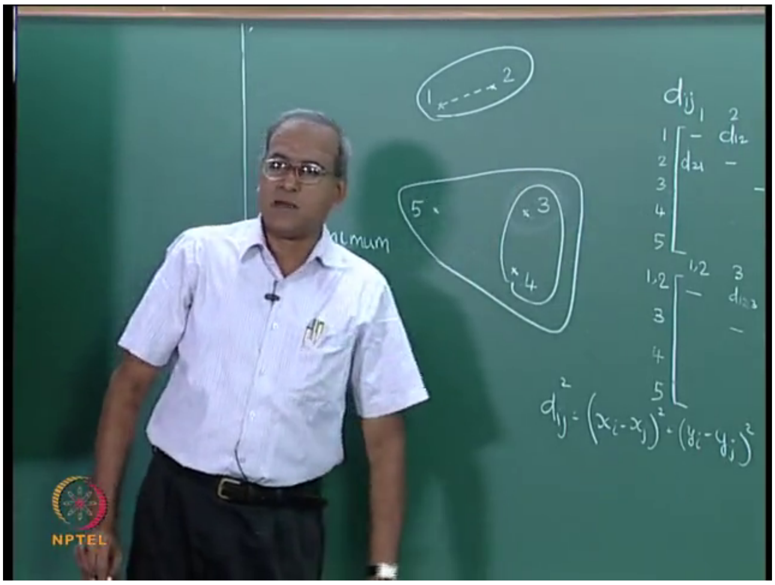


1. $\{1\} \{2\} \{3\} \{4\} \{5\}$
2. $\{1, 2\} \{3\} \{4\} \{5\}$
3. $\{1, 2\} \{3, 4\} \{5\}$
4. $\{1, 2\} \{3, 4, 5\}$
5. $\{1, 2, 3, 4, 5\}$

So, we did two things we started with a solution with which is like this 1 2 3 4 5 as 5 groups, then we did 1 2 3 4 5, then we did 1 2 3 4 and 5, then we did 1 2 3 4 5 and then we have, so we had 5 groups with all 5 then we have 4 groups then we have these as the 4 groups these as 3 groups these as 2 groups and this as 1 group. So now, we go back and tell the user if you want 5 groups here is the answer if you want 4 groups this is the answer and so on.

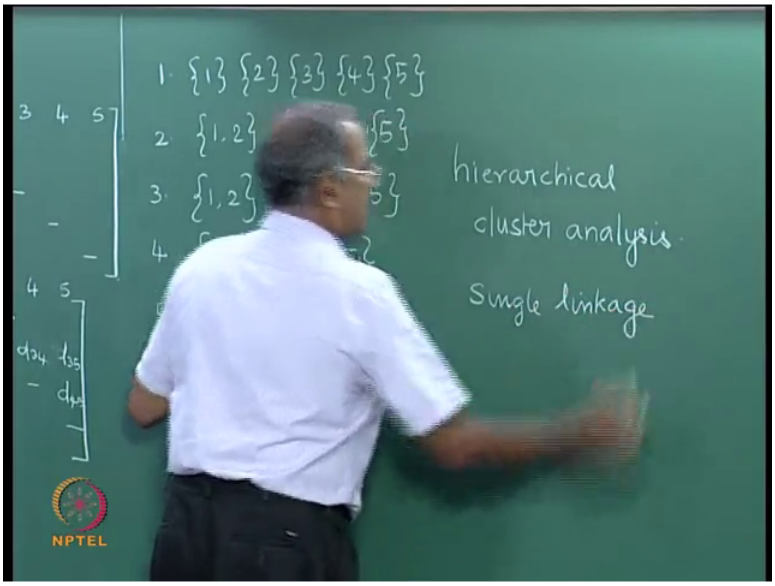
Now, there are two other considerations one of course, is we used the minimum here that is one we did not use the average or the maximum and these distances.

(Refer Slide Time: 14:52)



They are calculated in general using the formula d ij is equal to x i minus x j the whole square plus y i minus y j the whole square equal to d ij square. So, given a way to calculate the distance and given a way to relate group distance versus individual distance we are able to provide a set of solutions starting from n groups if there are n points to 1 group.

(Refer Slide Time: 15:46)



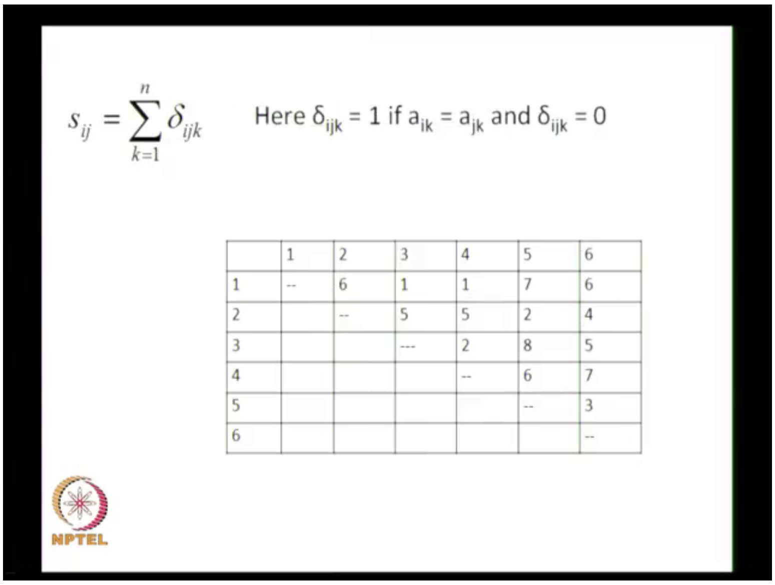
So, we can provide the solution is actually in hierarchy starting from n n minus 1 n minus 2 etcetera up to 1 and this process is called hierarchical cluster analysis, hierarchical. So, this is the basis of hierarchical cluster analysis. There are two aspects based on which the solutions can be little different and those two aspects are the way the distance is computed and secondly, the way the group distance is related to the individual distance. So, here we used minimum here we used distance.

Now, in this type of hierarchical cluster analysis the groups are essentially formed by linking or linkages through linkages and it is called single linkage cluster analysis, if we use the minimum criteria. So, this becomes single linkage cluster analysis if we use the minimum. It is called average linkage cluster analysis when we take the average and complete linkage when we take the maximum distance, many times single linkage cluster analysis is used.

Single linkage cluster analysis also has some relationships with minimum spanning trees in case you are familiar with minimum spanning trees. So, a single linkage cluster analysis provides a hierarchy of solutions, it uses minimum as the criterion to relate individual distance versus group distance. The advantage of using hierarchical cluster analysis is that given the number of groups you can actually get the solution, but the disadvantage is that once a group is formed between 1 and 2 at this stage 1 and 2 will continue to be in the same group you may add some more, but they will not at a later point go up to different groups. For example 3 and 4 have come here after that 3 4 are together in this group and 3 4 again together in this single group. So, they do not go to different groups once they are linked which is a disadvantage of the hierarchical cluster analysis.

Now, what is the relationship between the algorithm that we have seen here to group these 5 points versus the algorithm that we used to start with this solution to start with the given incidence matrix compute similarity coefficients and finally, say that we have a solution with 2 groups like this. So, if we compare what we have seen. Now, with what we have seen in the previous class there are lots of things that are common.

(Refer Slide Time: 19:01)



$$s_{ij} = \sum_{k=1}^{n} \delta_{ijk}$$ Here $\delta_{ijk} = 1$ if $a_{ik} = a_{jk}$ and $\delta_{ijk} = 0$

	1	2	3	4	5	6
1	--	6	1	1	7	6
2		--	5	5	2	4
3			---	2	8	5
4				--	6	7
5					--	3
6						--

The commonality is a matrix except that this is a similarity matrix whereas; the matrix that we used here is a distance matrix which is a difference.

The common thing between this algorithm and the algorithm to group machines is the fact that there is a matrix which is a similarity matrix in this case. We have also seen things which are common you see that there are two machines are in one group and then we computed the equivalent similarity and we took the maximum similarity as the equivalent similarity which is comparable to choosing minimum distance in the other algorithm. So, it is a form of a single link clustering algorithm that we have seen.

We also started with 6 groups individually. Now, there is a solution with 5 groups. Now, here is a solution with 4 groups and then we have a solution with 3 groups, then we have a solution with 2 groups and then we have a solution as a single group. So, it is a hierarchical clustering algorithm. So, the algorithm that we saw in the previous lecture where we have used similarities instead of distance where we have taken the maximum to represent maximum similarity as equivalent of minimum distance and provides us with a hierarchy of solutions is indeed a single linkage clustering algorithm it is a non hierarchical clustering algorithm that uses similarities instead of distances.

The other difference is here we start with a machine component incidence matrix from which the similarity matrix is formed or computed whereas, here we start with points which means these points are given as x y coordinates. One way is to say that this is computed if the location of these points are known this could be computed either by using a scale if it is written on a blackboard or if the coordinates are given as x i, x j then we can use this formula.

So, instead of saying that these points are actually drawn and written on the board one could say the coordinates of these points are known. So, once coordinates of these points are known the distance matrix can be computed. Now, we have to relate the similarity matrix that we used in the cell formation algorithm and we have to relate it to the distance matrix here and we have to relate the machine component incidence matrix which is a binary matrix to the coordinates of these points.

Now, let us do these two a little carefully let us do this relationship. Now, let us assume here the coordinates are known and based on the coordinates we have got a distance matrix. Now, the distance represents how far a point is from some other point since we want groups to be packed we want groups or we want points in the group such that their distance is minimized whereas, in the cell formation example that you see there on the screen you want rows or machines to be grouped such that the similarity is maximized.

Now, if we go back to this if we go back to this we have constructed the similarity matrix for the machines and we first started with this 3 by 4 because it is the maximum similarity. Now, let us go back to 3 and 4. Now, you look at 3 and 4 a similarity coefficient of 3 by 4 has been calculated, how did this 3 by 4 come? It came because if we see this there is a 1 1 pair sorry, if we take 1 and 3 take 1 and 3 this the maximum similarity of 3 by 4 is coming for 1 and 3. So, let us look at rows 1 and 3. Now, there is a 1 1 pair here, there is a 1 1 pair here, there is a 1 1 pair here

So, there are 3 components that use these 2 machines and then we also observe that there is a 1 0 pair. So, the numerator is 3, the denominator is 4 and the similarity is 3 by 4. So, if we take machines 1 and 3, there are 3 components that are visiting machines 1 and 3 with a similarity of 3 by 4. If we take 1 and 2 there is only 1 component that it is 1 and 2. So, would be group 1 and 2 or would be group 1 and 3 the obvious answer is we would group 1 and 3 because more components visit 1 and 3. So, higher the similarity they will become into the come into the same group. Now, this similarity coefficient is a ratio of between 0 and 1 and our cluster analysis here essentially grouped rows that have maximum similarity.

Suppose from this similarity matrix let us say we create another matrix another matrix where the entries are 1 minus similarity matrix remember the maximum value you can get in this individually is 1 that is the maximum value you can get in this matrix. Suppose you create another matrix with 1 minus similarity then we can call that as a dissimilarity matrix, matrix that represents 1 minus similarity. So, larger the similarity here smaller the number will be there. So, performing a cluster analysis with maximizing similarities in a similarity matrix would be the same as performing clustering by minimizing the dissimilarity in the other case. So, dissimilarity can be thought of as distance. So, maximum similarity gives rise to minimum dissimilarity which gives rise to minimum distance.

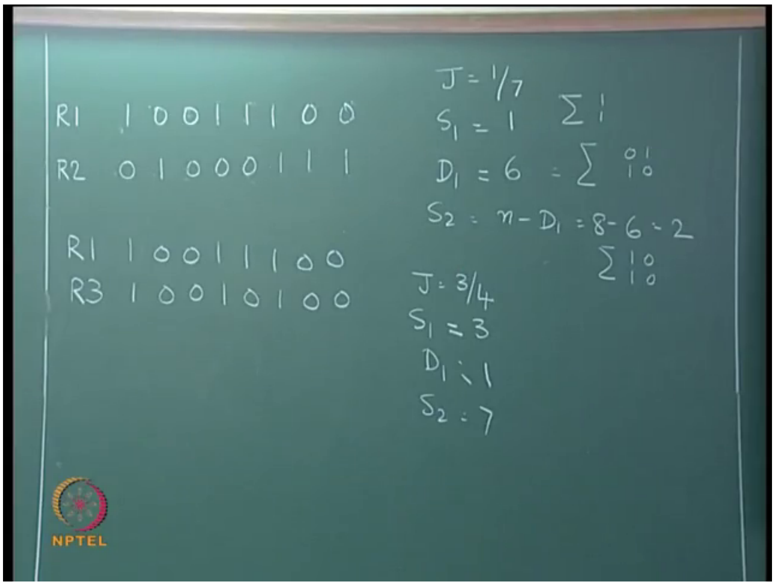
So, here in this example we have carried out cluster analysis by minimizing the distance in this example we have carried out cluster analysis by maximizing the similarity which essentially boils down to minimizing dissimilarity and dissimilarity can be thought of as distance. So, these two are equivalent.

Now, the only other thing that we have to look at is here the representation is in the form of a binary matrix in the other case the representation is in the form of coordinates. So, given the coordinates we were able to calculate the distance given this machine component incidence matrix we were able to calculate the similarity coefficient and we can proceed. So, this algorithm that we used to group machines is a hierarchical cluster analysis algorithm based on single linkage clustering because we took maximum similarity which is equivalent to taking minimum distance which is what we did that and we carried out.

Now, let us try and capture one more relationship between this and the other problem. Now, we said that. Now, we have actually calculate captured some what we captured the equivalence between data and this matrix and minimizing it versus data in a similarity matrix and maximizing it from the machine component incidence matrix we created a similarity matrix. Now, I have also said in between that it is possible to get these groups what are the inputs required to get these groups the inputs required to get these groups were, one the coordinates; two, a way to calculate distance and three, one of these.

So, let us assume that we stick to this minimum. So, that it is a single linkage algorithm. Now, let us try to see if there is there are other ways of calculating similarities or distances in the incidence matrix. Now, these 5 points are given as coordinates as x y coordinates if we go to the machine component incidence matrix we let us take rows 1 and 2 of the machine component incidence matrix and let me write these 2 rows.

(Refer Slide Time: 29:09)



R1 1 0 0 1 1 1 0 0

R2 0 1 0 0 0 1 1 1

$J = 1/7$

$S_1 = 1 \quad \sum \begin{smallmatrix}1\\1\end{smallmatrix}$

$D_1 = 6 = \sum \begin{smallmatrix}0&1\\1&0\end{smallmatrix}$

$S_2 = n - D_1 = 8 - 6 = 2 \quad \sum \begin{smallmatrix}1&0\\1&0\end{smallmatrix}$

R1 1 0 0 1 1 1 0 0

R3 1 0 0 1 0 1 0 0

$J = 3/4$

$S_1 = 3$

$D_1 = 1$

$S_2 = 7$

So, this I call as row 1 this is the second row. So, the first row is 1 0 0 1 1 1 0 this is the first row, 1 1 1 0 0 is the first row. Second row is 0 1 0 0 0 and 1 1 1 is the second row

Now, one way of calculating similarity is to find out the number of parts that visit both the machines and divide it by the number of parts that visit either or both the machines which is the Jaccard similarity coefficient that we calculated. So, the number of parts that visit both of them is this one, one part the number that visits either of them is 1 2 3 4 5 6 7 we got a 1 by 7.

If you take R 1 and R 3 together we get 1 0 0 1 1 1 0 0 is R 1. R 3 is 1 0 0 1 0 1 0 0. So, one similarity coefficient which I call as Jaccard similarity coefficient is 1 by 7 here and Jaccard similarity coefficient here is 3 by 4. Now, Jaccard similarity coefficient need not be the only way by which we compute similarity it is one of the ways by which you can compute similarity and it is an extremely popular way to compute similarity.

Now, let us look at another similarity. Another similarity is to say that if I take these two we can define similarity as the number of parts or components that visit both the machines. So, in this case similarity S 1 will be just 1 in this case similarity S 1 will be 3 the numerator alone will be the similarity. Now, if we say that two machines can come to a group or if we say that more than number of parts that visit both the machines higher their similarity coefficient and therefore, these two machines will come into the same group therefore, we are justified in defining similarity as the number of parts that are common and they are visiting both the machines.

Now, if there are parts that visits 1 machine and does not visit the other machine then that; obviously, represents dissimilarity. A 1 1 represents similarity a 0 1 and 1 0 represents dissimilarity. Now, suppose we call dissimilarity d 1 in this case only as the 1 0 or 0 1 pairs. Now, 1 2 3 4 5 6; 1 2 3 4 5 and 6 in this case either a 0 1 pair or 1 0 pair. Now, here the dissimilarity d 1 will be 1 1.

Now, here the dissimilarity is taken as 1 0 or 0 1 under the assumption that if there are many parts that visit only one machine and does not visit another machine then these two machines cannot come to the same group higher the dissimilarity less they will come into the same group. And you obviously, see a relationship between these two wherever the similarity is low the dissimilarity is high, you see similarity is high similarity is high dissimilarity is low.

Now, we have seen two ways of calculating similarity one is the Jaccard similarity coefficient, the second is the number of parts that visit both the machine the numerator of the Jaccard similarity coefficient, the third is we have defined at the similarity measure. Now, what is the maximum value this dissimilarity can take? That is 8, in the worst case I can have 1 1 1 1 0 0 0 0, 0 0 0 0 1 1 1 1 would give me a similarity of dissimilarity of 8. The maximum valued dissimilarity can take is 8, maximum value of similarity can take is 8, maximum value this Jaccard can take is 1. Now, suppose I define S 2 as n minus D 1 or 8 minus 6 which is 2 this is another form of similarity.

Here S 2 will be 7 8 minus 1. Now, what is D? D is the number of 0 1 pairs or 1 0 pairs therefore, what is this S number of 1 1 pairs and 0 0 pairs they are the remaining words. This is only 1 1 pair this is 1 0 pair plus 0 1 pair this is 1 1 pair plus 0 0 pair. How do we get this 2? There is a 1 1 pair there is a 0 0 pair, so it will brings us to a very interesting question is 0 0 pair does it contribute to similarity. If 2 components or parts do not visit both the machines then can we say that these machines are similar with respect to that part. So, question which has not yet been fully answered, but one could define either this or this or these two they are one and the same because this is nothing, but a constant minus D 1. So, instead of defining a Jaccard similarity if we had defined a dissimilarity which is sum of 0 1 pairs and 1 0 pairs, so this will be equal to sigma 0 1 1 0 I am just defining it loosely here we will define it formally very soon. This will be sigma 1 1, only the 1 1 pairs this will be sigma 1 1 0 0 all 1 1 pairs and all 0 0 pairs.

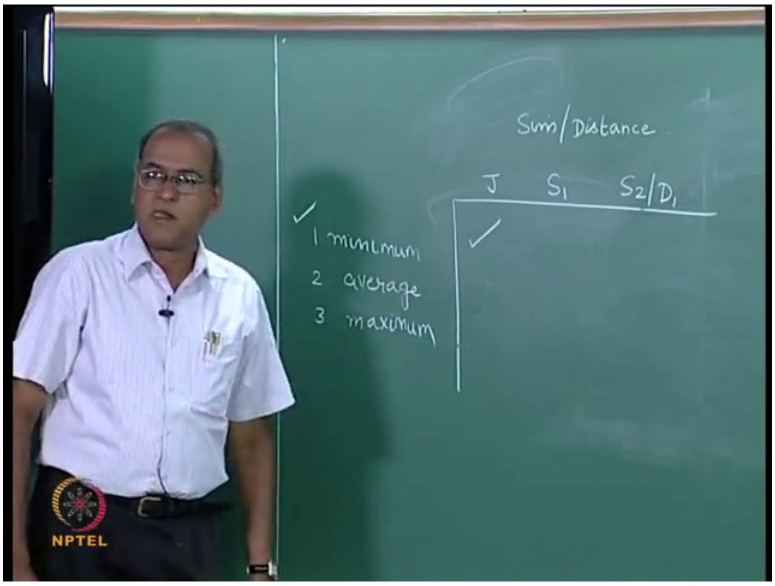
Now, these two are the same constant minus this. So now, you can see the relationship if instead of using the Jaccard similarity coefficient if we had used this similarity coefficient and then if we had taken n minus this to get this we could have used something very similar here having a distance matrix or a dissimilarity matrix or a distance matrix and getting it therefore, maximizing similarities minimizing distances are fine.

The last aspect that we have to touch upon is this. Given the coordinates x 1 x 2 or x iy x iy or x i yi and xj yj we can calculate d ij by doing x i minus x j the whole square plus y a root over. In a way we are doing exactly that except that this is not a point in 2 dimensions this is a point in 8 dimension 0 1 space, it is either a 0 or a 1 it is an 8 dimensional space this is another point. So, instead of taking you can we could even say we are taking the difference and squaring it because it is binary 1 minus 0 square is the same. So, wherever I have 1 1 pair the distance is 0 wherever I have 0 0 pair the distance is 0 when I have a 0 1 pair and a 1 0 pair the distance is 1 therefore, this becomes some of this.

So, computing this way the distance are this or similarity is very is equivalent to computing this distance given the coordinates. So, essentially each row of the machine component incidence matrix which is shown here is taken as a point in a n dimensional if there are n columns or n parts in this case 8 8 dimensional binary 0 1 space and the distance is calculated. So, we could either calculate a similarity or calculated distance we have seen different ways in this particular instance we have calculated a similarity and we have used Jaccard's similarity. We could have used instead of Jaccard similarity we could have used this similarity also or we could have used this similarity also and then we could have used a cluster analysis where we maximize the similarities.

Alternately we could have done using the incidence matrix we could have calculated this distance and instead of the Jaccard similarity coefficient matrix we could have used a distance matrix and then we could have done the clustering by minimizing the distance which is essentially the same or equivalent to doing this.

(Refer Slide Time: 41:09)



So, the algorithm that we have seen, if you see very carefully there can be different versions of the clustering algorithm depending on how the similarity is calculated, similarity or distance similarity or distance is calculated. So, we could use a jacquard similarity we could have used S 1 way of calculating the similarity that we showed here, we could use either S 2 or D 1 both are the same the only difference is if you are using S 2 you are maximizing the similarity if you are using D 1 you are minimizing the similarity.

So, the version of the algorithm in general there are now 9 ways of doing this algorithm 3 ways of defining the group distance versus the individual distance, 3 ways of calculating the similarity coefficient. So, 9 ways of doing this right. Now, the version that you see here on the screen uses Jaccards and uses minimum, so this is what it does. This is called single linkage cluster analysis and is a hierarchical clustering algorithm. Now, we could try out all of them as well we could do that.

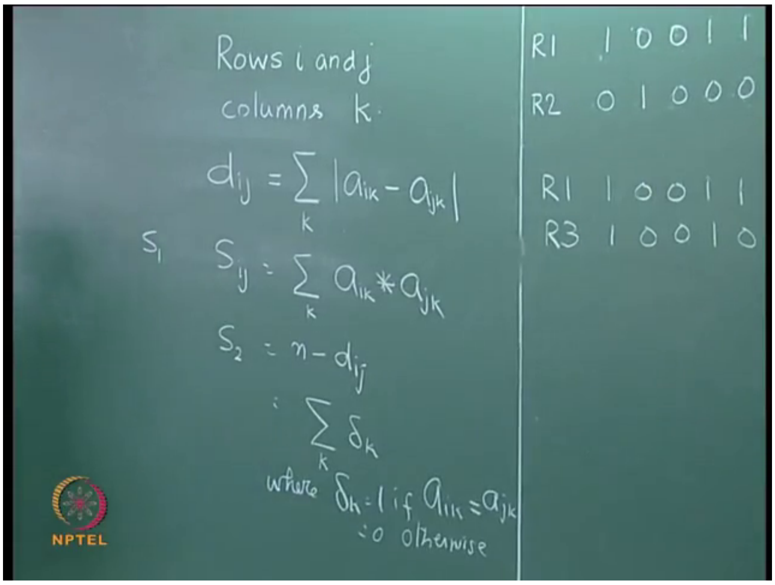
Now, many versions are available many versions of clustering algorithms for cell formation using similarity and distance methods are available and they largely differ in the way the similarity coefficient is calculated or the distance measure is calculated and the way in which whether it is a complete linkage or a single linkage or an average linkage cluster analysis all thing. So, we have represented only one of them.

Now, if we go back to the final solution here the final solution would be there are 6 rows, if you want 6 groups each row is a group, if you want 5 groups here is the solution, if you want 4 then you have this solution 1 3 4 2 5 and 6, if you want 3 groups 1 3 4 2 5 and 6 if you want 2 1 3 4 2 5 6 and if you want 1 all of this. So, one thing is to say that give me the number of groups that you want any number between 1 and 6 I will give you the solution or any number between 2 and 5, I will give you the solution because both 1 and 6 are extreme solutions.

The other is amongst these for example, 2 3 4 and 5 which is the correct one we have to answer that also to answer that you need to find out part families and then one way to do is to find out the number of intervals or there should be some other measurement criteria based on which we would say this is the best number and therefore, this is the best solution. Right now, we have not answered that question what is the correct number of groups we have said give me a number then I will give you the answer. So, that is addressed using hierarchical clustering.

Making the groups given a certain number of points or number of groups comes under non hierarchical cluster analysis as I said one of the limitations of procedure analysis once 2 elements come into a group they will not be separated whereas, in non hierarchical there is a possibility that they can be separated. So, we will see some aspect of non hierarchical clustering we will begin this in today's lecture and continue in the subsequent lectures. Before we do that let us also try and define these similarity coefficients in a more formal way. Now, I have said I have loosely defined it by saying one all sum of all 1 1 pairs sum of all 0 1 and 1 0 pairs and sum of all 1 1 and 0 0 pairs. So now, let us define these formally.

(Refer Slide Time: 46:06)



Rows i and j
columns k

$d_{ij} = \sum_{k} |a_{ik} - a_{jk}|$

S_1 $S_{ij} = \sum_{k} a_{ik} * a_{jk}$

$S_2 = n - d_{ij}$

$= \sum_{k} \delta_k$

where $\delta_k = 1$ if $a_{ik} = a_{jk}$

$= 0$ otherwise

R1 1 0 0 1 1
R2 0 1 0 0 0

R1 1 0 0 1 1
R3 1 0 0 1 0

Now, between rows i and j and columns represent k then d ij this distance, distance between 2 rows i and j is equal to summed over k summed over columns absolute value of a ik minus a jk. For example, absolute value 1 and 1 is 0 0 plus 0 plus 0 plus 1 plus 0 plus 0 plus 0, D is equal to 1. So, this will add only when you have a 1 0 pair or a 0 1 pair if you have a 1 1 pair or a 0 0 pair the absolute value will be 0 and therefore, it will not add. So, this is a very formal way of defining this.

Now, the other one S ij is equal to summation over k a ik into a jk just multiply both of them, let us see what happens here 0 0 0 0 0 1 0 0 which is 1. Let us see what happens here, 1 0 0 1 0 1 0 0 you get this. So, your S 1 that you defined there is given by this multiplication. Now, an easiest way of defining S 2 is equal to n minus d ij n minus d ij because here I am adding here I am adding 1 1 pair and 0 0 pairs. So, n minus d ij will give this, but the other way to define S 2 is equal to sigma k delta k where delta k equal to 1 if a ik is equal to a jk equal to 0 otherwise it is a slightly lengthy definition it is equal to 0 otherwise for example, go back if both are equal, its 1 2 3 4 5 6 7 3 S 2 is 7 8 minus 1 is 7. So, this is how you formally define all of these

So, we could say that any similarity stroke dissimilarity matrix using consistently either S 1 or S 2 or Jaccard's or D, D is a dissimilarity matrix or a distance matrix one can perform cluster analysis. So, from now on let us assume that we will use this d ij instead of maximizing the similarity we minimize the dissimilarity or distance. So, we will now define a distance matrix from now on which will be based on this particular expression or equation to compute the distance matrix given a machine component incidence matrix.

So, we could do a single linkage cluster analysis using D 1 and minimum also, that is another version of a single linkage cluster analysis. Whereby we compute a distance matrix using this and then we minimize the distance and we take minimum distance to represent distance amongst the groups. The other way as I said is to look at non hierarchical cluster analysis and we will see that in the next lecture.