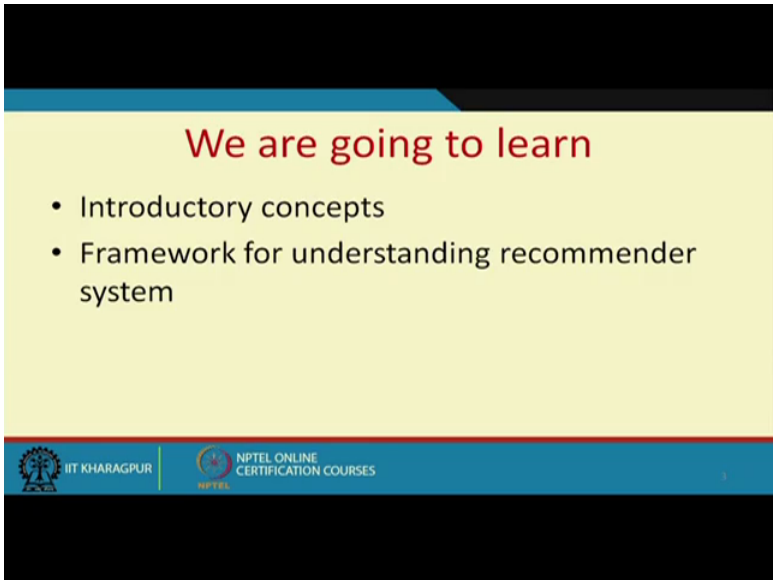
E Business
Professor Mamata Jenamani
Department of Industrial and Systems Engineering
Indian Institute of Technology Kharagpur
Lecture-52
Introduction to Recommender System

So today we are going to look at a new topic that is Recommender system. This is a part of a series of lectures we are going through on online decision support, in this context we have already seen about 1 data source that is your log file, which gets generated from the server which is captured by the server which captures the client-server interaction and the file is there in the server and from that server we saw that how we can using that data source we saw 2 things. First, how to create a customer behaviour model and how that customer behaviour model can help in various situations that also we discussed, then we also saw this capacity planning situation. Now we are going to look at a new data source and let us try to figure out how this data source gets generated.

(Refer Slide Time: 1:38)

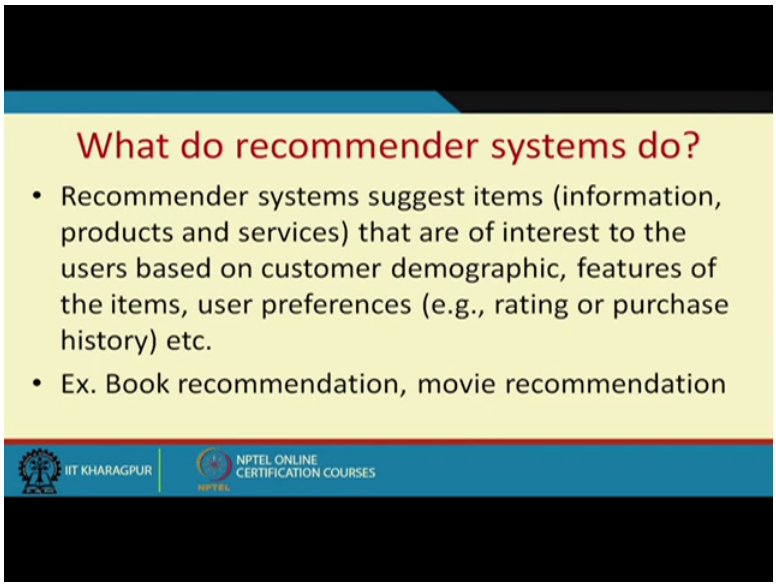


Look, recommender system is almost there in every website every commercial website now. So whenever you go to buy a book from Amazon or in that case any other website, if you search for a book you will get the details of the book but you are going to get a number of other recommendations which other books you should buy. You are going to see a movie in certain website, enter the movie name if you are subscribed you will be getting access to the movie and you will also get a number of other suggestions about the movie. Similarly, consider about social networking site, you always get suggestions for a number of friends, so

whenever there is any suggestion or any recommendations for you, we say there is a recommender system implemented in the website.

So in this lecture in the first lecture we are going to learn the few introductory concepts related to recommender system then we are going to understand a generic framework for our further studies on recommender system.

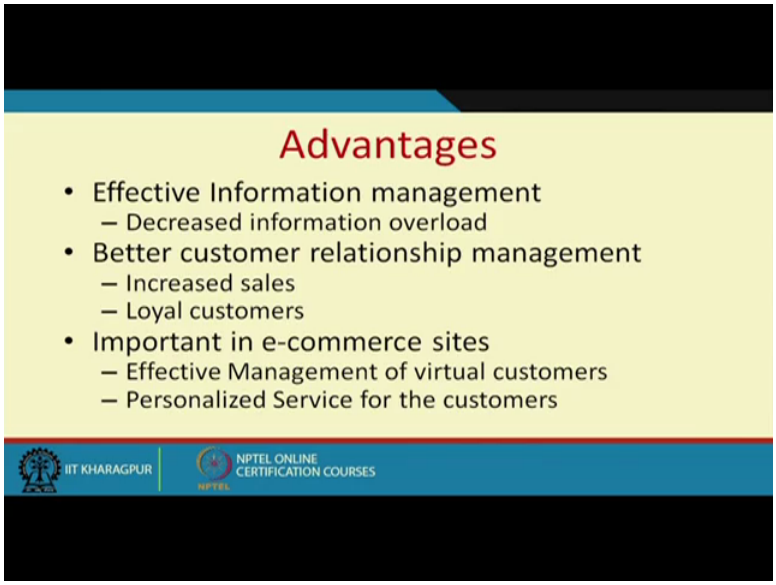
(Refer Slide Time: 3:09)



Now what do these recommender systems do? Recommender systems suggest items that is you can have news recommender system where you are not suggested with any physical product such, you are suggested with certain new information that you would like to see, there are news recommender system. So these recommender systems suggest items that its information, products or services that are of interest to the user based on the customer's demographic, features of the item, user preference that is rating provided for the product and so on.

Just now I saw and just now we discussed, there are movie recommender system, there are booklet recommender systems and so on. You can actually this list of examples is never ending, now almost every site in some form or other they implement this recommender system, whenever you see any suggestion which means you are, site is executing certain recommender systems that is why you are getting the suggestions. Now what is the benefit of recommender system?

(Refer Slide Time: 4:33)



First benefit is effective information management, so it decreases the information overload on the users. If you simply provide the list of products, beyond certain limit the user is not going to search further, but based on certain other activities like the kind of product you are you are looking for and another product with similar or maybe complementary properties. If it is suggested then it is good for you, without much search you will be knowing and you will also be knowing. So in this situation even if you do not know the detail of the product then also you are getting some information about the product, so this is the benefit. So in this situation you must be thinking why should I not use a search engine?

Okay, if you use a search engine you need to give the search parameters, search engine on its own will not give you any suggestion. So if you are searching for a book on recommender system, you will be getting books on recommender system but think of Amazon, if you are searching for recommender system book, are you getting only recommender system book? No, you might be getting try this, you might be getting the books which are of related topics where the other users who are searching for recommender system also search for those books, so that benefit you cannot get from a search engine, okay.

So it also helps in better customer relationship management, it helps increasing the sales and it also increases your loyal customers because the customers know that they are going to get many suggestions many more suggestions which may be surprising and important for them and the plus is your sales is going to increase, so for e-commerce sites it is extremely important for managing virtual customers. Look, in a physical store you have sales person to

guide a user, let us say some new book which is becoming popular has come to the store. Even if you do not know about the book, the sales person over there may suggest you that book, but if you are using a search engine to find something that something some other book that search engine is not going to tell you about the most popular book today.

So therefore but that benefit you can get from a recommender system, so if you have recommender system, if the e-commerce site implements the recommender system then it can manage its virtual customers. Even if sales person is absent there is something there is some system which can assist the virtual buyers then it helps in providing personalised services to the customer. The personalised we mean, if the site is looking at your interest in this particular session or in earlier session, how do it store your interest in earlier session? When you are going to search for something on Amazon, definitely you are not registering yourself and making your identity known, but still if you mark closely it remembers your past purchasing history, how does it do?

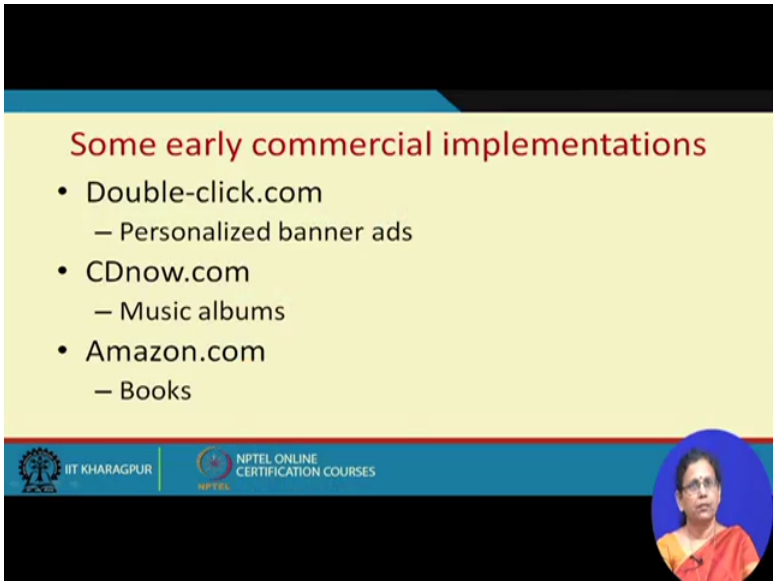
In fact, we have discussed it several times, it is maybe the site is using cookies and cookies looks at your information so even if http is a (())(9:07) protocol, through the user cookies once you visit the site and if you are using the same computer of course, nowadays though everything is related with your email ID as well. From your email ID you can I mean from your certain social networking account or your email ID you can connect to various places I am not talking about that I am talking about a much simpler situation where maybe you have your past data stored in the cookies and that is sent to the server, so server also shows you some appropriate information.

So whenever so which means depending on your interest you will be getting certain suggestions and depending on my interest I will be getting certain suggestions, so that way you are providing personalised assistance, personalised services for each customer. Okay, so in this regard let us look at few early initiatives that happen in the field.

(Refer Slide Time: 10:19)

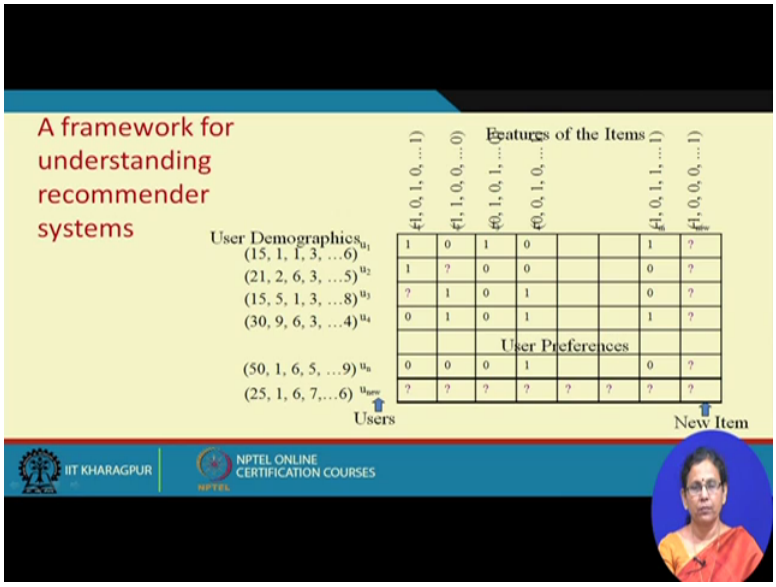
In fact, the very first recommender system around 90 or so probably 94-95 came out as a result from MITs media lab. This paper this particular paper on automating this word of mouth was recommending was in the lab it was designed to recommend music albums. Then there were certain other initiatives for providing news items that was from GroupLens another research group which had this group lens. Then Citeseer you might have seen that whenever you are searching for a research paper you get this sometimes you get it recommendations from the Citeseer, it is it recommends relevant research articles in a specific field.

(Refer Slide Time: 11:32)



There are certain commercial initiatives as well, Amazon has been the pioneer from very beginning, this Double-click.com it used to provide personalised ads services based on your I mean it used to recommend personalised banner ads, then there was one CDnow which used to suggest music albums, these are some of the early initiatives but nowadays almost in every site you can find out a recommender system, so we are going to see now how this recommender system works. So to understand the working principle of recommender system let us first try to understand what kind of data you require for recommender system and how this data gets generated and what exactly is the recommender is coming or what kind of recommender is coming out of this data.

(Refer Slide Time: 12:17)



Look at this diagram, here you have 3 things; first you have user preferences, you have first of all look at this matrix, this matrix the rows here represents the users, u 1, u 2, u 3 up to u n are the users similarly, you have the items, you have the items and here the columns represent the items so item i 1, item i 2, item i 3, i 4, i 5 and i m and so on. So rows are the users, columns are the items, now what does this matrix contain? Matrix contains the preference of individual users on each of the items. So this matrix right now in this particular example situation we have put as a binary matrix, but in reality depending on the kind of recommender system you are dealing with or the kind of data that you have, this may not be binary this may contain the values as well.

Here the binary represents whether the user has seen the item or maybe he has not seen the item like in a book buying situation whether you have seen the details of the book you have not seen the details of the book, in that case we have made it Binary. But suppose you would

like to provide certain rating in the continuous scale that is also possible, it depends on the kind of data that you have. You may even ask your buyers this preference matrix, how do you generate this preference matrix? This preference matrix can be generated in an explicit way by asking the buyers, suppose you ask the buyers to rate a product so you provide a rating range may be discrete or continuous, let us say it is a discrete range between 1 to 5 depending on 1 means not so good and 5 means very good extremely good okay.

So not so good, good, average, not so good, not very good, not good , maybe average then good, very good, so this can be 5 point scale or in the scale you can design on your own may be continuous scale, – 10 to + 10, I can give let us say 9.5 which indicates it is an extremely good product or I can give – 5 which means it is not I mean it is quite bad okay, so that is how you can either give a discrete scale or a continuous scale it is up to you. But given this scale you expect your buyers to provide a value explicitly for that product. So now the question is, it is up to the up to the person who is browsing the item or who has already purchased the item. Now suppose he does not provide the rating, so which means you are not going to have the recommender system.

Look, whenever you are searching any book from a site like that of Amazon, how many times you provide a rating? Possibly many of you have not provided any rating still then also you are getting the suggestions, how does it happen? It the site the e-commerce site can implicitly if you if you do not give the value explicitly the rating value explicitly, it can derive this value in a implicit manner. Let us say simply it can implicitly assign the value 0 and 1 whether you have seen the product, not seen the product, it can give a value in a continuous scale, how much time you are exposed to the product.

Let us say you are exposed to the product depending on how do you get this value? How much time the user is exposed to a product? Possibly the whenever from the your log file data you can find out on an average how much time is spent on a particular function let us say products search function okay. So for a specific product how much time was spent, through that level you can find out or simply you can find out from your cookies, when he saw the page and when he sent the next request, how much time it was there in front of him in the browser that data you can also store in your cookie.

So whatever may be your data source, the point that I am going to make here is you can also assign a continuous value. So it depends on your data collection procedure but whether the user if the user is interested he can give the value explicitly that rating you can take or his

preference you can take or you can derive this value looking at his behaviour is browsing behaviour in an implicit manner, whatever may be the case to have recommender system you need to have this user preferences. So for all the users visiting your site, for all the items that is there in the site you can have user preference, so which means you can realise for typical site like that of Amazon or Flipkart who handle thousands and thousands of products and millions of users visit their site every day, if they try to maintain a matrix of this, how big it will be.

And if these computations have to happen online, what will be the complexity? And do you feel that it takes a lot of time? No, as soon as you enter the book, immediately you get the recommendation because they understand that user cannot stay long in the website. Anyway, we are going to deal with such issues little later but for now it has 3 different types of data in a typical recommender system. One is this matrix of user preferences, second is the list of users the list of users and their personally their demographic values or their certain characteristics. Again these demographic values which may be their age, their profession, et cetera, by asking the users you may get but user may not be interested to give this as well but other user characteristics but some of these user characteristics it is possible again to implicitly collect.

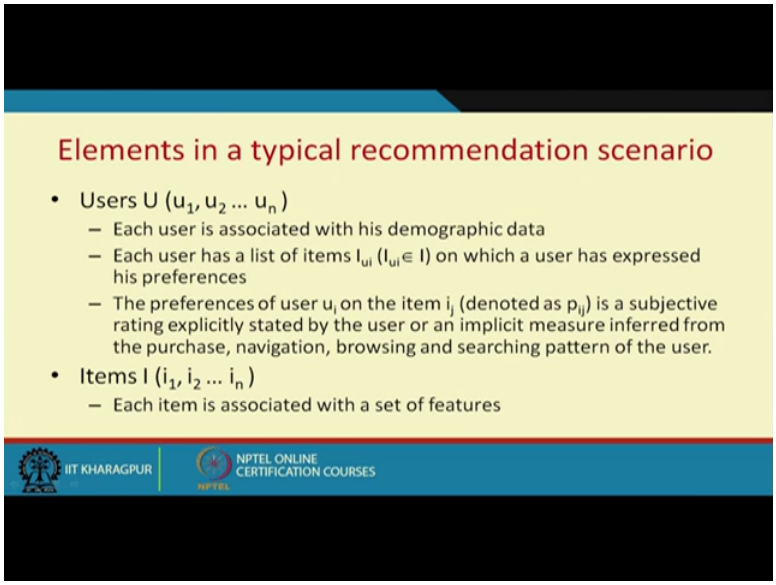
But whatever may be the case, for some of the user some of the recommender system because there are various types of recommender system, you may require this user demographics. Then next is the feature of the item, item can have many features, now what are these features? These features can be of many types again irrespective of what it is, how it is derived we have I am just telling that it has to have a each item has to have a feature vector okay. For example, let us say you are continue with our example of a book or movie or something, well these details could be name of the author, then the number of pages then the topic then the publisher and so on an category of the book. Similarly for a movie all those details, its category, is Director, its actor leading actor and so on can be some of the data which can represent the feature of the item.

Now this feature of the item can be that explicit or this can be implicit too. For example, in case of a news item or in case of even a book, these features could be the keywords, presence or absence of certain keywords or frequency of the keywords, so this is again up to the user how it defines the feature vectors of each item and feature of each user. But irrespective of data so that how it is done et cetera, which some of these we can, we will be listening at the

in the next few lectures but irrespective of that 3 things are here, one is user preference, another is user detail, user demographics and another feature of the item.

So given this situation, what does a recommender system do? Recommender system can ideally decide whether a specific item can be proposed to a specific user looking at his past or current interests okay. It can also be used to suggest out of many new items which are recently come to your store which are those users who should be targeted for marketing this particular new item, so new item has come so which all users should be targeted who has exposed to, who has shown interest in similar kind of product can be targeted. Similarly if a new user has come, which are the top few items he should be suggested okay?

(Refer Slide Time: 25:10)

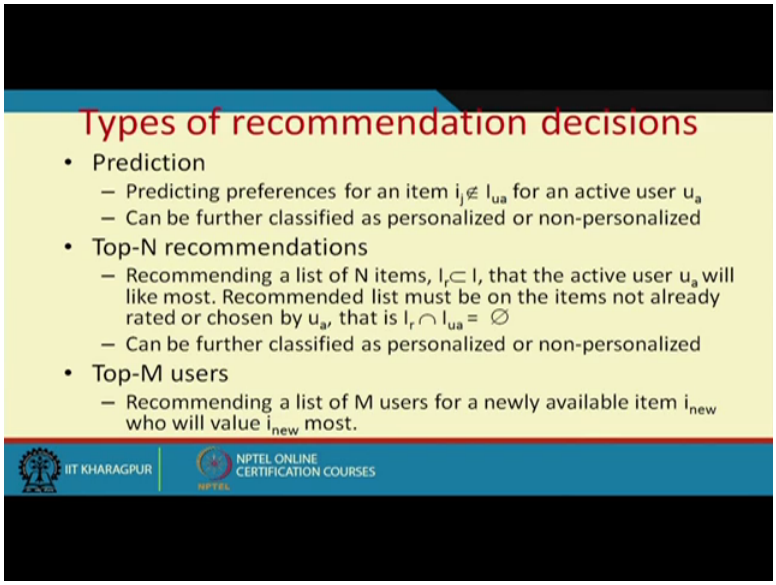


So this is what I was telling you, these are the typical elements in the recommender system recommendation scenario, the list of users U, where each user is associated with his demographic data, each user has a list of items on which he has expressed his preferences. The preferences of user u on item i j is denoted by some value, it is a subjective rating explicitly stated by the user or an implicit measure of measure inferred from the purchase, navigation, browsing and searching pattern of the user.

The next category of data item is the set of items and the features associated with them, so you have a set of items let us say some n number of items and each item is associated with a set of features and of course the third one is your user preferences so this user preference metrics, users and items. So this is the preference of each user on this is the preference of

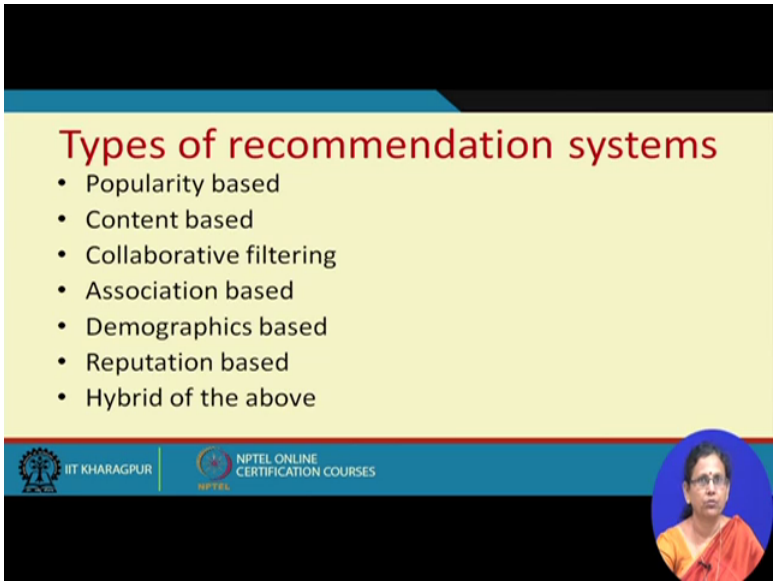
each user on the item on a specific item okay. Now what are the various types of recommender system decision that I already told you.

(Refer Slide Time: 26:49)



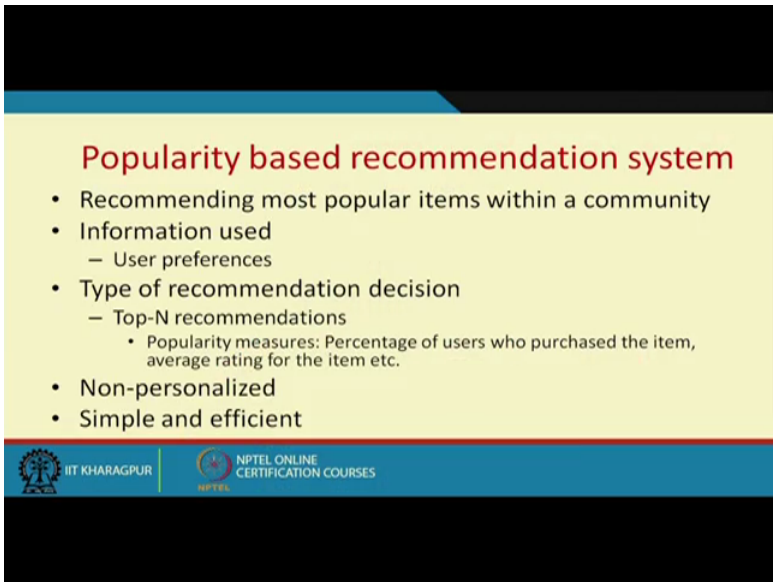
Now predicting the preference of an item for an active user and this is prediction this is what I was telling you. This is let us say this is the active user, this is the active user and this is the what is the predicted rating for this item which he has never seen for this particular user okay. Then Top-N recommendation, recommending a list of N-items that the active user u a would like most okay, so if this is the active user and he has not seen many items, which are those top few items which should be suggested to him. Then you have Top-M users, so recommending a list of M-users for a newly available item who will value that item most and hence they should be targeted for marketing activity.

(Refer Slide Time: 28:06)



Then these are few different types of recommender system and we are going to look at them in detail, but for right now let us have a look at what exactly each of them do.

(Refer Slide Time: 28:25)

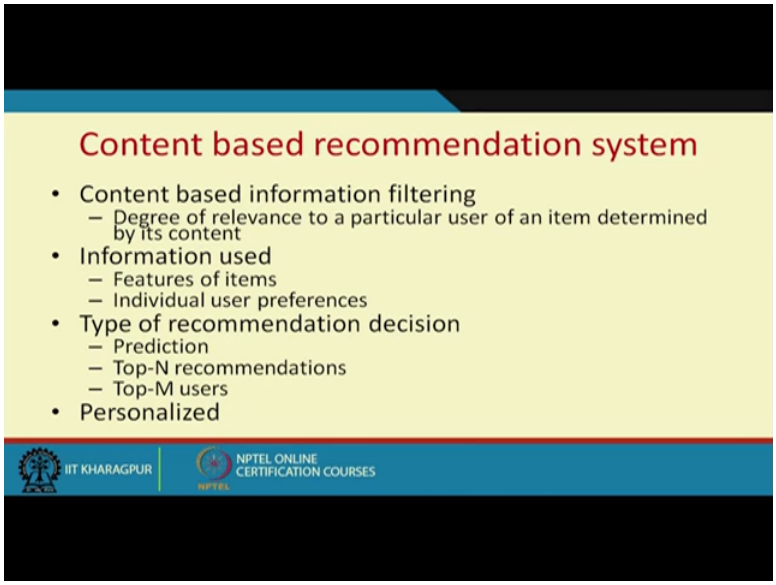


This popularity based recommender system is quite very simple. Here, the the idea is to recommend the most couple item within a community, suppose a set of people within the age group let us say 20 to 25, within this age group most of the people are downloading certain music, so that music must be suggested to others. YouTube, certain video in particular category is viewed large number of times so that should be suggested to the viewers who might be interest who might have interest in that category, okay. So here what kind of

information is used? The user preferences and what type of recommender recommendation decision can be given?

The Top-N recommendation, the popular based on the popularity measure, the percentage of users who purchased the item, using that or average rating of the item given by the users, you can give the Top-N recommendation. And it is non-personalised, what now by non-personalised we mean? If I am getting that detail, others will also get that detail, it is quite simple and efficient.

(Refer Slide Time: 29:51)



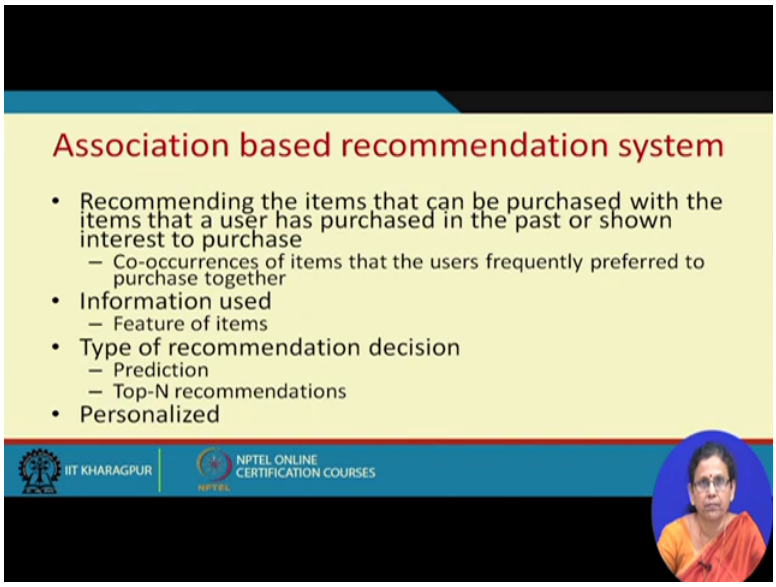
Next one is your content based recommendation system, so here the content of the item let us say in case of a book, the feature of the item could be name of the book, author, publisher, category, et cetera. Or the feature could be the keywords present in the book okay so the content based recommendation content based information filtering is the idea here for content based recommendation system. Here the degree of relevance to the particular user of an item is determined by its content by the content of the item, the information here uses the feature of the item and individual user preferences.

So out of that remember that matrix that matrix and 2 vectors 2 different types of vectors we had, the user vector, item vector and the preference matrix so here the inputs are the feature of the here the (())(30:57) the feature of the item that is those feature vectors and the preference matrix, the kind of recommendation generated, prediction, Top-N recommendation, Top-M users and it is a personalised recommendation scenario. Then you have collaborative filtering, in collaborative filtering, it can be user to user collaboration

based on test or it is item based on item-item similarity we are going to discuss both, but oldest of these 2 is your user-user correlation base.

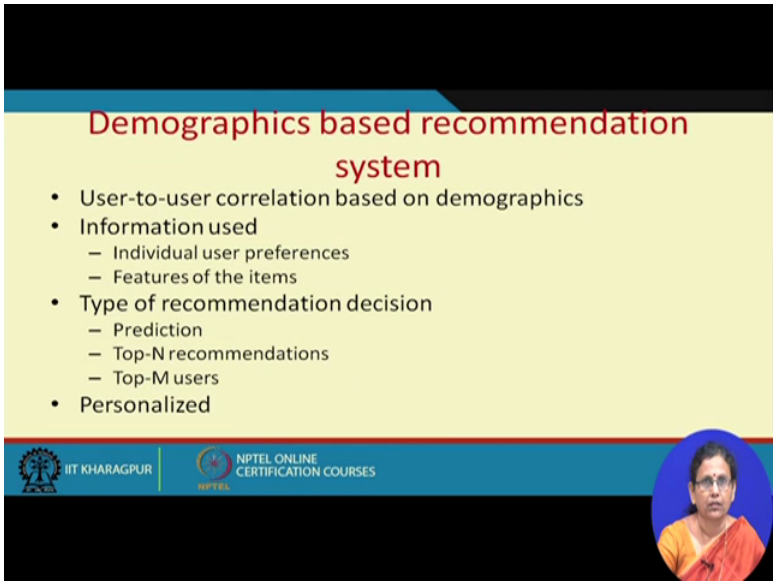
So it is also called social information filtering, here the user the information used this user preferences and type of recommender system, it can be a recommender decision which can come out of it are predictions and Top-N recommendation, it is also personalised.

(Refer Slide Time: 31:58)



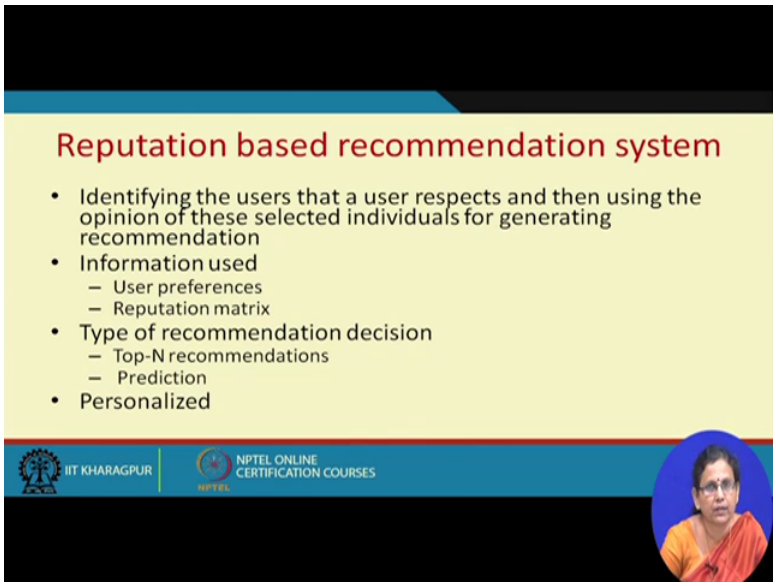
Then you have association based recommender system in which the idea is to recommend products which are often purchased together or seen together. I mean other users in the past have shown interest in purchasing them together or seeing them together at least. So it is decided using the co-occurrence of the item that users frequently prefer to purchase together. The information here used is the feature of the item and type of recommender decisions are prediction and Top-N recommendations and it is a personalised one.

(Refer Slide Time: 32:50)



Then you have your demographic based recommendation system. So here it is again user-user correlation, but in case of collaborative filtering it was the, their correlation was made based on their preferences from the preference matrix, but here the correlation has to made based on the demographics. Last one was whether they were interested in the similar products and here whether their personal data demographic data looks similar. So the idea here is you use this demographic data so it is individual user preferences and feature of the item these 2 are used to generate this demographic based recommendation. The kind of recommender decisions that can come out of it is prediction, Top-N recommendation and Top-M users, these ideas already we have discussed what are these types and all, then it is also personalized.

(Refer Slide Time: 34:07)



Then next one is reputation based recommendation system, which is about identifying the users whom a particular user community respects and based on their recommendation based on their opinion you generate recommendation. So which is the expert in the field whatever he recommends, based on that he will be getting recommendation. Here the reputation matrix one extra matrix can be used which we can call as reputation matrix which will be a user to user matrix where jointly who are the other users who do I respect can be kept or in general this will be without this matrix also you can make this computation possible by assuming that that everybody in that community respects that particular user.

In fact, you must have seen while in the books many people give their feedback and the reputation score for that particular user and reputation score can be calculated based on whether he has actually made the purchase or not, how frequently comes to the site, how loyal is to the site, what others think about that particular viewer and so on. So this book reviews you can think of this kind of reputation based recommendation can come out of the reviews of the experts. Then this can give Top-N recommendation or it can give prediction and it is personalised again personalised in nature. Thank you very much, next class again we continue with this discussion.