Course Name - Recommender Systems
Professor Name - Prof. Mamata Jenamani
Department Name - Industrial and Systems Engineering
Institute Name - Indian Institute of Technology Kharagpur
Week - 01
Lecture - 01

Lecture 1: Recommender system and its business value

Hello everyone. Welcome to the first lecture of the Recommender System course. In this module, we are going to learn about the business value of recommender system. Let us start. To start with, recommender systems are one of the most successful applications of artificial intelligence and they have become an integral part of our daily internet experience. And why this system is developed? As internet became to the public domain, there were large number of large amount of information appearing on the internet on various websites.

As a result, there were information overload on the individuals. So, while identifying the right item of their interest, people wanted some kind of tool. Similarly, the various commercial agencies like online e-commerce platforms, they also wanted to deliver value by providing right kind of information to the customers and also capturing them. So, these such systems actually help websites to improve their user engagement and they impact the user behavior and business.

There are many examples and I need not tell in detail and as we go ahead, we will see few such examples. Now, look at this. This is a recommender system handbook. During our intro lecture, I told you about one book that is recommender system which is a textbook. This is another book, Recommender System Handbook.

This is also is going to be one of the very important key learning material for you. So, when I tried looking at this book, I got various other suggestions about what other customers bought while buying this book. Similarly, when I tried searching for the recommender system in YouTube, I got many such suggestions after I saw one lecture from certain individual. So, look at this. This is the friend suggestion from the Facebook.

Now, here I would like to tell you a very nice story. This story is about two books. First book which with title Touching the Void which came in 1998 was actually about the experience of somebody who went for hiking Peruvian Andes and how he faced almost the near-death experience. And during that time after its publication, it was not very popular and soon people tried forgetting. Almost a decade later, another book came into, similar book came into picture.

The name of the book was into Thin Air. And this book was also on mountain climbing tragedy. When this book was published, after some time suddenly once again Touching the Void which was almost out of print, came into existence. People started ordering for Touching the Void book and the publisher because the book was out of print, the publisher tried printing it and tried promoting the book again. And you would not believe that book became the bestseller and it spent almost 14 weeks in the New York Times bestseller list.

Now, this is a very strange phenomenon. After 10 years a book becoming suddenly very popular. What happened? Around that time Amazon tried implementing its recommender system. And because of the similarity in the content, because Touching the Void and Into the Thin Air both were on mountain climbing tragedy experience. So, this second, the older book that is Touching the Void came into the knowledge of the users.

So, as a result people tried ordering it and because it was a good book, it became popular and stayed in the bestseller list. Why such a thing happened? Look at this particular phenomena which the book experienced is because of the long tail of the items which are stored by the retailer. Look at this in a typical retail environment, we have certain constraints. We have the constraints of the size of the bookstore and how many physical copies of which type of book we will keep that depends on the popularity of that book or the newness of that book at that point of time. So, if we look at this long tail, the first part which is the head part contains most popular books and such books actually appear in the store.

Because store has its limit in keeping the books, but in a retail store which is online there is no such limit. You do not have to keep a physical copy. You simply have to keep the details of the book. So, those other books which may not be new or may not be popular at a particular time, all of them can be there in an online store. So, they appear in the tail part.

So, all these products when people try searching along with the popularity, along with the popular products, the products which are also there in the stores list, they also appear if they are of similar types based on their content and based on how users try purchasing them together. So, because of this, this into thin air which was somewhere here and the torching of void the second the first book which was somewhere here as a list enlisted item, suddenly both of them become popular. Had it been only the brick and mortar store that is a offline store, this would not have been possible because this tail end products nobody keeps in the store. But they suddenly become important because they are similar in content and that the second the first one which was almost inactive for almost 10 years suddenly become active. So, let us look at the formal definition of recommender system.

These recommender systems are software tools and techniques that provide suggestions for items that are most likely be of interest to a particular user. The second definition could be the recommender system or the or they are also sometimes called recommendation system. They are the platforms or engines. They are a subclass of information filtering systems that seek to predict the rating or preferences that the user would give to an item and based on this suggested or predicted rating, the top few items are suggested to a particular user. So, these are typical features of a recommender system.

As you know in a store there will be millions of users. So, also there will be thousands of items. So, the data set which typically is for the users as well as the items are very large. And we also consider the rating data which the user gives on the items. So, which means the rating itself makes a very sparse data set.

Why do you why does this data set become sparse? So, just imagine you are you are a member of let us say Amazon Prime and you are ordering items. Amazon has thousands and thousands in fact millions of items. How many of these you are ordering? Very few. How many of the you are rating? Almost you know very meager amount you are rating. So, a very small fraction if you imagine this data to be a matrix where users are the rows and the items are the columns only few ratings appear.

So, it is a large and a sparse data. Similarly, when you talk at the users, users may not be willing to give their details. So, that data also becomes sparse. And as this data size increase or we get more data, the algorithm that is used for giving suggestions must learn continuously from this data. So, these systems are learning systems.

So, because the machine learns it, they are basically machine learning systems. And because you are providing these algorithms on a platform where the user requires quick response, you need fast algorithms for online implementation. Search engines are also used to find out various information of our interest. A recommender system also gives us certain information on products. Then where is the difference? Is there any difference between a search engine and the recommender system? Are the recommender systems search engine? To answer this, let us look at what search engine does.

In a search engine like that of let us say Google, you have to type the keywords. So, which means you are aware of what you are searching for. Whereas in a recommender system, if you are watching a movie in Netflix, many other movies come which you may not be aware of. If you constantly keep on watching, let us say comedy movie, many new comedy movies or very popular comedy movies will come up. Okay? So, recommender systems come into play when the user do not explicitly know what they want exactly.

The users cannot explicitly use keywords to describe their need. System does it for them and system uses some algorithm for this purpose. If you look at the foundation of this recommender system algorithms, its foundation lies in statistics, linear algebra, machine learning, data mining, text mining, information retrieval, pattern recognition, convex optimization to name few. There are other things as well because this is basically one application area. Now let us look at the business value of this recommender system.

Why this recommender system was adopted? Recommender system is adopted because people wanted to know information which they can derive from the large amount of data that they are getting exposed to. And this data, compiling this data and finding relevant information slowly become difficult. So therefore, to help the users and in turn to generate value for their business, various commercial agencies like online e-commerce sites, media streaming platforms etc., they introduced recommender system.

So here are some examples. It says that study has found that 35 percent of Amazon's revenue is generated by its recommendation engine. Similarly, Netflix says 80 percent of watched content is based on algorithmic recommendation. People do not search. Whatever is recommended to them, they tried watching those movies. Now this business value of recommender system can be assessed using few metrics like click through rates, adoption and conversion rates, sales and revenue, effects on sales and distribution and user engagement.

So the first metric is click through rate. With click through rate, we measure in some form how many clicks are generated by the recommendations. So what do you mean by this? When you show the recommendation, people try clicking on the suggestions. So the underlying assumption here is more clicks are made on the recommended items, more efficient is the system.

Let us see one example. Google news personalization engine, some author has found that the personalized recommendation has led to an average increase of clicks by 38 percent. The second one that is again very important in terms of assessing business value is adoption and conversion rate. While adoption is about getting the, I mean it goes beyond getting the attention and the interest and it is about looking at whether the recommended items beyond clicking are actually seen by the user. For example, one news article you might have clicked, but did you actually read that news article? One recommended movie you might have clicked, but did you really completed watching the movie? So this kind of metrics, this adoption is the metric which goes beyond clicking and it suggests whether the clicked items are actually viewed or used or seen. So similarly, conversion rate is about finding whether that particular recommended item is actually purchased or not.

So while accommodation can be measured based on whether the user really liked the recommended news article or not that they clicked on, conversion rate can be measured based on the purchase history. So you must have seen in many news recommendations, they ask you whether you like this news or not. So, they are explicitly also you can express and implicitly they can observe how much time you are exposed to the news and based on that they can derive this. Some examples include in real life long CTRs for YouTube, then take rates in case of Netflix, bid through rate in case of eBay, link through or side through rate for research papers and so on. These are examples of adoption rate and conversion rate which is also known as purchase rate can be measured by making adequate arrangement for observing addition of the item to the shopping cart and actual purchase data.

The third one is sales and revenue. So it is difficult to measure whether overall business value in terms of sales and revenue did increase due to adoption of recommender system. It is really difficult because many times as your personal experience might have gone, you view the item recommended item on the online retail store, try to get just the details of the item, but while coming to purchase you go to a physical store and you buy it. In that case recommender system definitely affected your decision for purchase, but getting the data that it was because of the recommender system cannot be really assessed. So,

therefore, the user may be aware of the item through the recommender system, but it is purchased in a physical store. So, for online only businesses however, it is possible to assess through some programming effort in the sense you have to accommodate may be some additional lines of code to capture the data about whether the actual purchase actual sales are happening after clicking the recommended item list.

And this can be assessed in many studies in fact, has been done for assessing this during alpha and beta testing of the software. Then the next metric is understanding the effect on sales distribution. So, last matrix was on sales and revenue. Now what is this sales distribution? Sales distribution is not about how much what is the how much in terms of money how much you have sold. It is about beyond what is the consumer is targeting to buy beyond that is there anything which because of the recommendation the consumer is buying.

So, this is because of the stimulus that is given to the user because of the recommendation system. Many of the cross sale scenarios, cross sale in the sense let us say you are buying a laptop and because you are buying a laptop something like a let us say laptop bag or certain additional accessory as suggested to you. And if this is the case then recommendation is giving you opportunity giving the business the opportunity to sell the items which are from different categories. So, how diversified is your sale across your catalogue measures the sales distribution. So, this effective catalogue size that expresses the amount of catalogue exploration by the user is used by Netflix for this purpose.

This is just one example of a key metric. Then next metric is user behavior and engagement. So, what is user engagement? How much time the user spends in the site? So, more time it is perceived that more time the user spends in the site more will be your sales. So, therefore, higher engagement positively affects sales. In fact, it affects the user behavior and in turn it affects the sales. So, to measure this again is a very difficult task and during user studies performance measurement studies like user studies as well as real time applications performance evaluation during real time applications you can measure this one.

And some metrics like that of adding something to your wish list etcetera can be used for getting this metric. Now, the question is why this business value really improve? Some explanation can come from the consumer decision making process which typically is described by the marketing community. When we look at consumers decision making process it goes through a number of stages. First need recognition, then information

search, evaluation of alternative, purchase decisions and post purchase behavior. Let us go to one by one to each and how recommendation systems can affect.

First is need recommendation, it refers to understanding how the buyer can how a prospective buyer can think about purchasing certain item. I have never this used this item, but what stimulus can be given to me so that I can think of purchasing it. So, this need could arise from within me that is my internal stimuli. I am feeling cold. So, I need some kind of clothing for this, some kind of winter garment for this.

This is some kind of internal stimuli. Now somebody my friend said that this particular quote from this particular such and such company is very good. So, probably because of his recommendation I will be able to buy it, ok. So, recommender systems can act as some kind of external stimuli because even if the user is not aware of a specific need it can provide certain external stimuli to the user. Then second is information search. So, registered consumers once they identify the need for a particular product, particular item then they will be searching for further information.

So, what are various sources of this information? It can be certain personal resources by asking friends and acquaintances, family members, it can be from various commercial resources, public resources and experiential resources. So, these systems can come under this commercial resources category. Look at the next stage evaluation of alternatives. So, when it comes to evaluation of the alternatives in the online setting you will see the feedback given by the like-minded individuals. So, that can motivate you whether to buy certain product or not to buy.

The third one is making a purchase decision. So, when you make this purchase decision about a particular brand or a particular company and so on, two factors can come into picture attitude of the others and unexpected situational behaviour. When it comes to attitude of the others typically commerce sites and wherever you are getting engaged for buying the item what others are saying affects you which in turn is the outcome of a recommender system. And many times it also happens because of unexpected situational factor. For example, Netflix has recommended you a movie, you opened it and you started looking at few first few minutes then you understood that it is a very great movie and then you saw it completely. In fact, you may not be knowing the clips which is shown by the Netflix site when you open it, clips of certain clip video of certain very recent or relevant movie that is also the result of some kind of recommendation.

In fact, I would suggest there is a very interesting paper by those Netflix people themselves where they say what are different ways in which they use this recommendation engine. Now, coming to this post purchase behaviour of course, recommendation system cannot help you in providing the post purchase avenue for providing post purchase your post purchase opinion, but definitely it can provide the avenue for expressing your satisfaction or dissatisfaction. And because it is providing the expression an avenue for providing expression your expression for satisfaction and dissatisfaction and if you enter this data this is going to affect the recommendation that is coming out of the engine for the other people. So, you may not be directly spreading your frustration and happiness, but through recommendation engine recommendation engine provides you an avenue through which you can spread your post purchase the outcome of your post purchase behaviour. So, with this brief introduction, so let us go ahead and see what all books we can refer.

So, in fact, two of these books I have already shown you. One book I have shown during our intro which is recommended system by Charu Agrawal. Then second book also I saw in today's lecture that is recommended system handbook and there are two more books which I would say one is fundamental or predictive text mining even if it is not directly connected to this, but while talking about content based recommendation we are mostly going to focus on this. So, some idea on text mining is necessary. Then we have data mining concepts and techniques by Han and Kamber and in this book we are going to look at some of the fundamental concepts which otherwise are not covered in detail in recommended system books. So, even if I cover these books in cover these techniques in detail or not definitely these books are going to help you in finding more about what is being covered in the class.

Then there are two reference books which I am suggesting right now, but there are many more books and many more online materials which I am going to use during my lectures. So, as and when I use it I am going to put it in the reference list. This is the conclusion of today's lecture and we came to know that recommended system as one of the very important applications of AI and machine learning and we saw that they have become an integral part of our daily lives and they help us in dealing with information overload that we come across in online setting. And they have implemented through very powerful algorithms on large and sparse datasets and we also tried looking at the business value of recommended system which has motivated almost all the businesses which have online presence to go for recommended system implementation. So, we also saw that how from the previous studies how such metrics have improved the business value of the organizations.

And we also tried getting some details on how the consumer decision making process theory can be utilized to explain why the commercial values are improved by recommended system adoption. So, with this we wind up our lecture. Thank you everyone. See you in the next lecture.