Course Name - Recommender Systems
Professor Name - Prof. Mamata Jenamani
Department Name - Industrial and Systems Engineering
Institute Name - Indian Institute of Technology Kharagpur
Week - 04
Lecture - 16

Lecture 16: Association rule based model

Hello everyone. We are now starting a new module. The module is on model based collaborative filtering. And in this particular lecture that is lecture 15, we are going to talk about association rule based models. So, this is the concepts covered. So, we will be talking in general what a model based collaborative filtering is and we will be talking about the association rule based model.

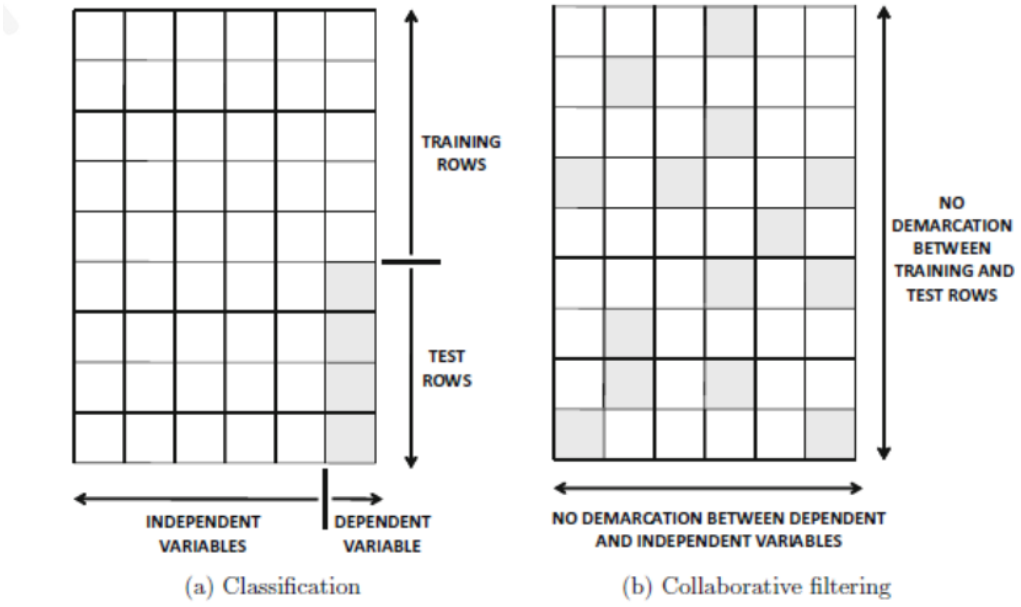
Collaborative filtering as we know has 2 approaches, one is neighborhood based another is model based. Now, when it we talk about the neighborhood based method, if we look closely they are actually some kind of lazy learner. Now, what is a lazy learner? In while talking about machine learning foundations, we know that this k nearest neighbor like classifiers, where we do not have a predefined model and at the time of prediction only we run the model and then and there we find out the class of a particular ah object. However, in case of neighborhood method though they are kind of lazy learner, we do some of the activities so that this similarity matrix is stored in the ah is created in the offline setting.

So, which means we in some sense this neighborhood based methods also utilize the concept of machine learning. Now, in case of model based approach, we have to create models so that they are no morelazy learners. So, which means we will be train the model and decide the model parameters. So, when a new new ah object comes you just put it in the model and do the mathematical operations with the model parameters and you get the output. So, most of these traditional classification and regression problems are special cases of matrix completion problem, which is not I mean your collaborative filtering is a matrix completion problem.

And most of these classifiers and regression problems can be connected to this setting. Now, look at this here we compare we visually compare typical machine learning supervised machine learning setting in this case we are providing a classification problem and compare it with case of collaborative filtering setting. In case of classification or regression problem, we have the data in which the last element the last element is actually the last element is actually the response variable. These will be the features and this will be the response. And you will have a set of training data with which you build the model and there will be a set of test rows.

But the problem is different here. Now, look at this one this one. So, these are the places where the rating is known and in rest of the places rating is not known. So, which means now this known values what are the known values in case this case these are the known values this part is this training part is the known where these values are known. So, with respect to this known and taking these attributes you were building a model.

Now, here this known values are distributed throughout this matrix. So, such situations are called matrix completion problems. So, now we are supposed to if you would like to use the traditional machine learning models we have to create a situation like this where you will have a number of features and some kind of dependent or the response variable. How do we make it here? So, one of the way to look at it is as we already know that we can make some u v we can represent this matrix M as u v transpose ok and this u v are to be determined somehow. So, how it will be determined that we will discuss later that again can be determined using some kind of regression based approach only.



(a) Classification (b) Collaborative filtering

So, now once we have these values oh determine these values of u and v we can fill up this matrix and specifically for u taking u as a feature we can construct some data set right that with a latent feature user 1, user 2 up to user M and latent feature 1, latent feature 2 then rating. And in one of the earlier lecture we already know that this latent features that we see actually connect users to certain concepts. So, each user will be related what is his latent rating with respect to concept 1, concept 2 and sorry latent features with respect to concept 1, concept 2 and what is the corresponding rating. So, we can build a model. So, the various ways it can be done this is just one example

situation, but it can be directly matrix completion can happen using this u v decomposition.

Now whatever may be the case this approach which is pretty straight forward in case of such situations demand little modification, but the conceptually starting from decision tree and other techniques and even regression based techniques can be used for this purpose. So, one particular method we have decided to demonstrate demonstrate on how to make a model from the available ratings or available user feedback. So, this is a rule based method. What is a rule based method? We have already discussed. In a rule based method you have something called antecedent if something happen then something has to be done.

So, then if that condition part is the antecedent and next part is the consequence. So, we have to construct the rule base. So, one of the approach for constructing the rule base in machine learning is association rule mining. So, now this association rule mining how it can be used as a model for collaborative filtering kind of situation we are going to look at. So, this association rule mining is based on something called a frequent pattern analysis.

Now what is a frequent pattern? A pattern which can be set of items, subsequences, substructures etcetera that occur very frequently in a data set. Now this frequent pattern analysis is the basis of association rule mining. So, there are many applications for this and one of the application that we can we are particularly now interested here is if a particular user has seen has done many transactions together or has rated many items together. Based on that can we decide which items are correlated hence to be recommended. So, we start with the basic concept of frequent pattern and association rule.

In this particular example we have some 5 transactions with different IDs and in under each transactions a number of items are bought. So, if you look at this which items are bought together. If we let us say consider buying A and B together how many times they have bought they have been bought. So, here A D sorry A and D here also A D here A D here of course, A D is not there. So, 3 times A and D are bought together.

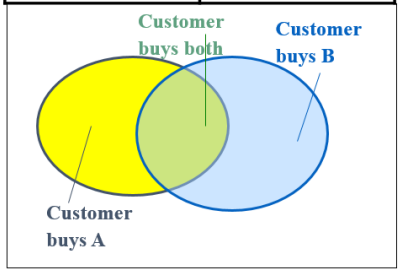
So, there are 2 matrix using which we decide. So, because A and D bought together can we decide a rule if somebody is putting A in his shopping cart let us recommend item D as well. So, to do so and to decide a rule many criteria's you can adopt 2 such very important criteria are support and confidence. In case of support it is the probability that a transaction contains both items X and Y and confidence on the other hand tells that if a transaction contains X whether it contains Y or not. So, it is a kind of conditional probability.

Now, consider the case of A and D items together. So, what is the support and what is the confidence? Now, how many times A and B D are bought together? 3 times and out of how many transaction groups out of 5 out of 5 transactions there is evidence of A and D occurring together in 3. So, which means we have 60 percent support for occurring A and D occurring together. Now, what about the confidence? It is the conditional probability that a transaction having X also contains Y. So, if I find the form the rule that if A is there D is also there true whenever A is there D is also there.

So, my confidence is 100 percent, but if I make this rule when D is there A is there or not. See this particular one has D, but A is not there. So, out of 4 transactions which has D only 3 contains A. So, the confidence is 75 percent. Now, considering the support minimum suppose minimum support is 50 percent and minimum confidence is 50 percent what are the frequent patterns? A in the so, while finding frequent patterns we have to take first single items, one item pattern, two item patterns, three item patterns and so on.

So, according to one item pattern there are total 5 unique items. So, A occurs in 50 percent of the transactions, B also occurs in 50 percent of the transactions, D, E what about C? C is occurring here and here. So, out of 5 transactions only in 2 transactions it is occurring. So, it is not a frequent pattern D and E also turns out to be frequent pattern. Now, comes to 2 item are A B frequent pattern? No, A B occurs only once only A D turns out to be a frequent pattern because we have decided our support and confidence to be 50 percent or above.

Transaction-id	Items bought
10	A, B, D
20	A, C, D
30	A, D, E
40	B, E, F
50	B, C, D, E, F



- Itemset X = $\{x_1, \ldots, x_k\}$
- Find all the rules $X \rightarrow Y$ with minimum support and confidence
  - support, *s*, probability that a transaction contains $X \cup Y$
  - confidence, *c*, conditional probability that a transaction having X also contains *Y*

Let sup_{min} = 50%, $conf_{min}$ = 50%
Freq. Pat.: {A:3, B:3, D:4, E:3, AD:3}
Association rules:
A → D (60%, 100%)
D → A (60%, 75%)

So, this about these 2 I was talking these 2 are the interestingness measures support and confidence. So, this support is the probability that a transaction contains both the items x and y and the confidence is the conditional probability that I have already explained. So, this is how it is to be computed and that computation procedure also I have shown with respect to that particular example. So, now, there are many algorithms for this association

rule mining. Choice of this algorithm depends on the size of the data set because while doing this computation computation steps many times you require huge memory.

So, therefore, you may be partitioning your problem and keeping it in the secondary memory then bring the data to primary memory do computation send it back and get another chunk of data. So, each process is going to take a lot of time. So, based on such considerations there are many other algorithms as well, but one of the very basic algorithm which may not be very computational efficient, but easier to explain that we are now demonstrating just for the explanation purpose. So, this apriori algorithm is based on this apriori principle. So, this apriori principle says that suppose an item set is not frequent which means it does not have minimum support.

So, if an item A is added to this set then the resulting set cannot be cannot occur more frequently. For example, let us go back to here B is a frequent pattern B has occurred in 3. Now, suppose to B we add A, A B. So, A B cannot occur more than 3 times. So, A B is occurring only once here.

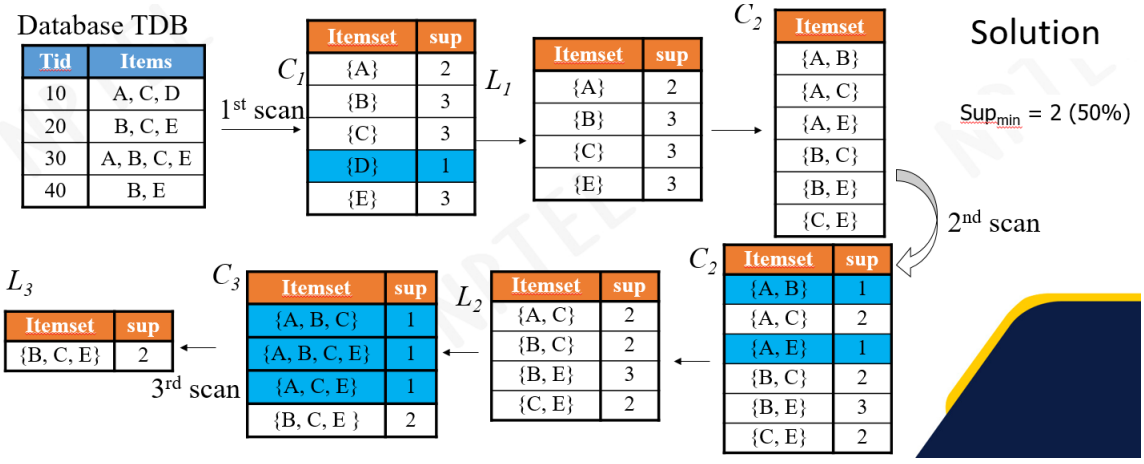
Tid	Items
10	A, C, D
20	B, C, E
30	A, B, C, E
40	B, E

Similarly, C is not a frequent pattern C is occurring in 2 places. Now, adding another item C another item A to C can it occur more times it will occur either same number of times if C occurs everywhere or it will be occurring less number of times. So, now, such properties are called anti monotone property which says that if a set cannot pass a test then all its supersets will will also fail the test that is what just now we saw. There are 2 steps in this algorithm join and prune. This is the algorithm, but we will not be going through the steps of the algorithm just like that we will be first looking at an example and based on that example we have we can come back and look at the algorithm again.

Suppose, this is our example problem and these are the items which are rated by a specific users specific users. So, user ID 10 has rated this user ID 20 has rated this and so on. So, because they are rated together by some user there is some association among them. So, let us identify what are those associations and how to generate the rules. Look at this that particular database has 4 transactions and each containing few items.

So, this database will be scanning for single items at a time to find out the frequent patterns. So, we scan A is occurring twice A is occurring twice 1 1 2 B has occurred 3 B B B 3 3 times C D etcetera. So, out of this if we decide that we keep minimum support to be 50 percent then this is to be pruned. So, we have pruned this item removed this item for further consideration. Why so? Because from apriori principle know that if we keep it and add more to this set that item is not going to be frequent.

So, now, after pruning it we have A B C and E left. Now, from this we will be constructing 2 item sets possible 2 item sets. What are possible 2 item sets? A B A C A E A B A C A E then B C B E B C B E then C E. So, after doing this we have to make a second scan of the database. So, we have to go back to the database here and look for where in how many places A B is occurring A B is occurring once we saw A C 2 A twice A E once and so on.



So, we find out the frequencies. So, once we find out the frequencies based on our claim of the support which is 50 percent which means 2 items 50 percent of this is 2 items. So, these 2 are getting pruned. So, here we joined here it is a prune here it is a join here again it is a prune and here after pruning you filtered again you try making join. You join how do you make the join? The join is A C B C.

So, A B C A B C A C B E. So, A A B C E B C B E. So, B C E this then B E C E B C E this is already there. So, therefore, after this we have to go for the third scan go to the database find out the frequencies. So, it turns out that these 3 items are to be pruned. So, only thing that remains with 3 items is 3.

So, now, at the to make a rule you need one antecedent and one consequence. So, at least 2 items are necessary. So, we can make rules out of this, this, this, this and this 3 items. Had there been 4 items as well 5 items 6 items. So, from every time we have to take the items which satisfy these minimum support criteria and now form the rules.

As an example from the frequent pattern let us say B C E alone will not be considering other let us say we consider only B C E that one just for the demonstration purpose. What was frequent pattern? B C and E the last one only we are considering. How many rules we can consider? We can make taking any 2 at a time B to C C to B B to E E to B. Then taking one item one side 2 items on the other side.

So, B to C C to B C to B. So, these are some of the possible maybe I have left something some might have forgotten probably this is a total number. So, anyway assuming that these are the set of rules possible rules should we keep all of them. Now, based on our confidence only we have got this frequent pattern and these are the rules we have got based on our support. Now, we will be using this thing called confidence.

Assuming that we go for a 100 percent confidence. What is 100 percent confidence? If the antecedent occurs, then consequence is also there. So, in that case only these 4 qualify. So, finally, mean in my rule base this will be there. So, whenever a user will be giving ratings for B and C next time when he comes in we will be recommending the item E. Because our rule says that this is a rule with the pre decided support and confidence.

So, as I told you we can go back and see the algorithm once. So, what we did we scan the database once to get the frequent one item set then we pruned then we tried joining joining steps and tried getting the data from the database by scanning the database then we pruned and finally, we continue till there is no more frequent items item generation is possible. So, in case of association rule based recommendation you have to generate association rules from the transaction database to generate top end recommendation you have to find the association rule supported by the active user that is active user whose LHS. LHS is the items which he has kept in the shopping cart the item which he has already seen and so on. So, now, we are supposed to find out what are the RHS of the rule.

Now, looking at the RHS of the rule we have to choose top few looking at its confidence and the top one and based on that their order will be now making top end recommendation. So, these are the references that I have used and this is my conclusion. So, traditional classification and regression problems are special cases of matrix completion and consequently collaborative filtering systems can be addressed by those problem those those means learning algorithms. Now, we can utilize one such algorithm is association rule mining and we can utilizing this we can utilize this particular particular

algorithm to generate association rules and based on co-occurrences of items that the users frequently prefer to purchase or view together we can utilize this to form the model and based on this we give the recommendation. So, here in this particular context we are using a rule based recommendation which is getting generated from association rule mining. Thank you.