Learning Analytics Tools
Professor Ramkumar Rajendran
Educational Technology
Indian Institute of Technology, Bombay
Lecture 5.5
Data Mining

In this video we will see other techniques for the diagnostic analytics. It is , Mining the data and finding the state transition between the data and how to plot it. We will see more of diagnostic analytics in coming weeks.

(Refer Slide Time: 00:39)

Hope you remember the MettLE environment, shown in a week 2. We told you, that what type of data can be collected and how the log file might look like. So, you know that, different action are stored in a Mongo DB and you converted it into CSV. We try to find out how many different kind of actions a student can do in the learning environment.

We found out that around 100 plus different actions can be done in the MettLE environment, But we grouped them because 100 plus actions will be too complex and it will not make any good. So, we grouped the actions based on which part of the problem they are working on. If you remember MettLE environment, the problem map is classified into four major maps and two calculation and evaluation. So, the problem map is divided into three major subtask, like is Functional map, quantitative model and qualitative model, then it also have a calculation and

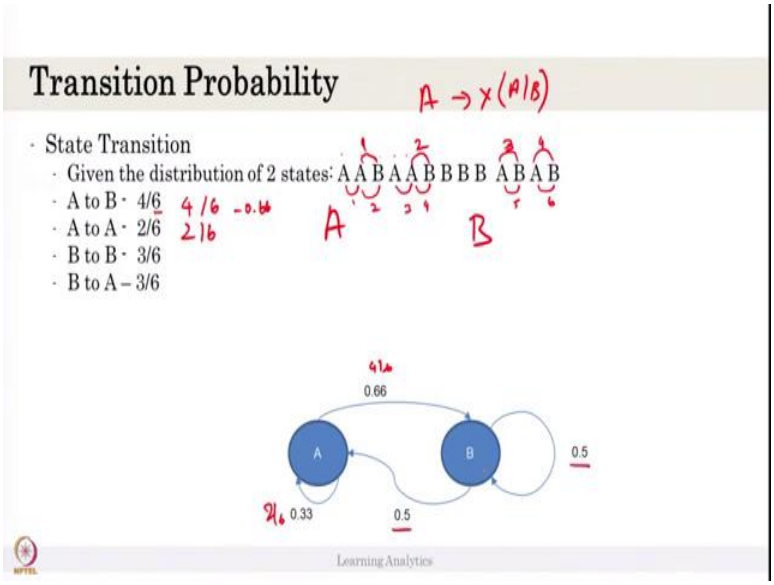
evaluations. Also the menu had simulator, info and other part of tabs available for the students to use it. So what we did, we grouped all the planning actions in the functional model and all the execution or contextualisation values in functional models FMEC.

Similarly, for Quantitative model, we name it as QM QMP or QMEC, Similarly for Qualitative model we have planning and Contextualisation and Calculations we target as Cal and evaluations we just label it has Eval. So FM is a one value, FMP and FMEC the three variables here. Similarly, QnMP, QnMEC, QIM, IMP QIMEC and Cal and Eval also. We group all this actions into 11 eleven different set of actions. Also, we have simulator action and in infographics, that is are they using hints or they logging in or log-off action. So other than 11, we are another few more actions.

Let us consider these actions, can we create process model of student's interaction behaviour in the MettLE environment,. Here we do not know what we are predicting, we are not predicting the students maths score or students performance in the final test or post-test, instead we want to see how students is doing, what is the student's interaction with the data?

We might have out of say 5 students 2 students would have done good in the post-test after the intervention with the MettLE environment, 4 would have not done it. We want to see what, why happened, why student is not able to score. What are the student's interaction behaviour in that particular learning environment? Consider we have a learning environment such a time stamp and all this actions are captured and you having a time series data. Let is see how to compute the transition between these variables.

(Refer Slide Time: 04:29)



Before we go into looking at MettLE environment transition. Let us see the example to explain what is transition probability. Given the distribution of two states say A and B, this is arranged in the time series data and it happened in a time series manner. So, first action A happened action B A B A A B... something like that, we have arrange this in the action series.

How many actions are A to B? That is count 1, 2, 3 and 4. Some of you might know what is what I am doing, I am going to draw the state transition table or state diagrams. If you know, it is good we can skip this video, but for others who do not know the state transition, I am just trying to explain the basics.

So, there are A to B actions, these countsare 4. What is 6? 6 is indicating how many times A to other actions that is X happen, X can be A or B. So, if you look at A to other actions, if you look at it, this is 1, 2 see this is a 1 transition from A, this is second transition from A, this is third transition from A, this is the fourth transition from A, this is the fifth, this is sixth.

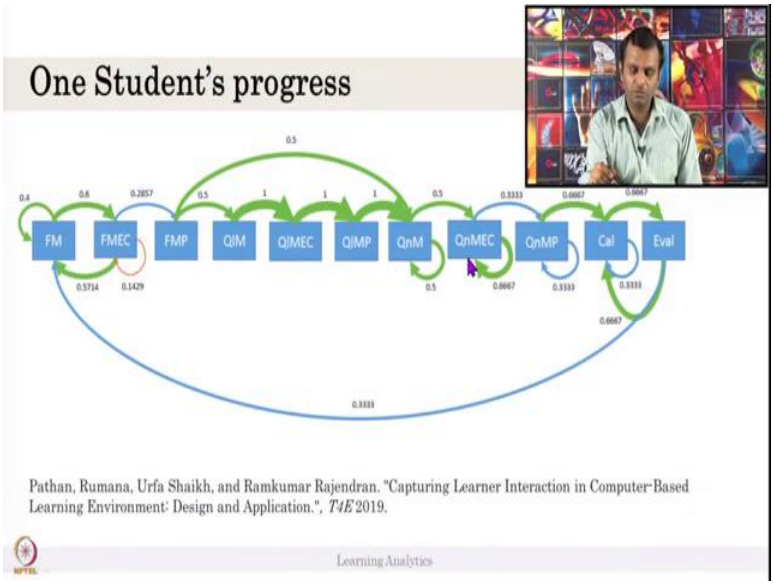
There are 4 of them are A to B only, so 4 by 6. Also, A to A will be 2 by 4 because there is 1 and 2. If there are 6 transition from A, 4 of them are to B which means 6 minus 4 be 2, so 2 by 6. Similarly you can compute what is the transition from B to B and B to A you can compute it.

So, after you compute it, we can draw the state transition diagram. This is the state transition diagram or state diagram. Let us not go into the details of what is edges, notch all these things. Let us see a simple state transition diagram. This is two states A and B, from A to B the value is

0.66. So, what is the probability of you will get the next action B if you are current action is A that is 4 by 6. Like out of six actions out of A four times B occurred so we have high probability and is 33 percentage of time is possible that you will come back to the A action only.

Similarly, we also calculated it for B. Hope you understand this figure. So, this figure indicates basically the in a time series data, what is the transition between two actions A and B? Let us consider we have, 9 action related to models and two actions calculate and simulator are two other actions. So, you have like 11 actions and what will be the transition diagram based on students interaction with the system.

(Refer Slide Time: 08:34)



This is the transition diagram based on student interaction with the system, its one student's progress. And this is from this particular paper, if you want to know more about this transition, you can go and check this paper in internet. Let us look at this transition diagram in detail. It is not indicating where the student start and end. Let us assume that we start from the functional model. The process are color coded, this green colour indicates above 0.4, this colour indicates less than 0.4 but above 0.2 like the blue colour and this indicates less than 0.2. This is one student interaction behaviour in a MettLE. The student started with functional model, as did the context relations, then planning a bit but it most of them goes back to functional model see the 0.57 compared 0.8. So, that is that compare to 0.285 is the most of them is going back to the functional model. There is a self-loop and the student is kind of going around. If we move to the functional model planning, there is high chance a student goes to qualitative model or quantitative model as both has 0.5 and 0.5 and you might see sometimes it is not adding up to 1. It is because we are not showing all the values is less than 0.1 in this figure to reduce the complexity of the figure.

And there is a interaction, suppose the student goes to qualitative model mostly goes to contextualisation and planning then to quantitative model. So there is a high transition, theree is not no self-loop or no other action is happening.

And this indicates most student were on evaluation go back to calculation or they go back to functional model. Also, there might be a less probability of self-loop or less than 0.1 probability of going to other actions. So, this graph indicates how the student's progress or students interaction behaviour in the system.

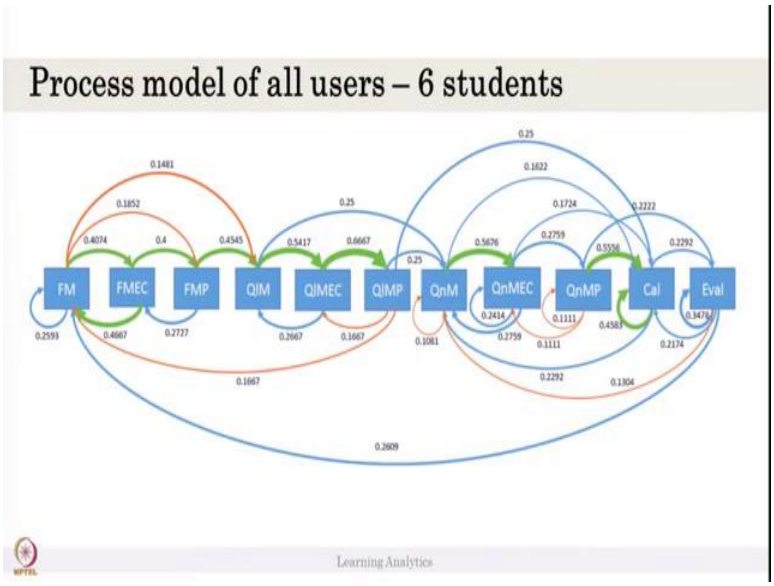
So, which way the student would have progress. First you would have started with functional model then you would have spent time on functional model, for some time before you jump into a functional model planning, once in the functional model planning, you would have directly jump to qualitative or quantitative models. In the qualitative model you have then qualitative model, contextualisation, planning, then you would have went to quantitative model. So that indicates that how the student would have been progressed in the learning environment.

So, then the student would have to spend less time on evaluation or he might moving back to functional model or he might have come back to evaluation with a less probability. So, this graph indicates the student's interaction behaviour in the learning environment. We are not able to show how much time a student spend on each of these actions or number of times the each action if you that information we will get a more complete detail. We will look at this that how to add those information's, when we discuss about a, Pattern mining and Process mining.

(Refer Slide Time: 12:57)

Can you think of, how this state transition plot will vary if you use interaction of more than one learners? Say 3 learner's data, 10 learner's data. Just think about it, if you able to think about, how the plot will be then you can resume the video to continue.

(Refer Slide Time: 13:22)



So, this is the Process model of 6 users, 6 students using lot of to and forth between all these actions. What you can do is you can group the students who are able to do well in post-test say 3 students who did good in post-test after the intervention in learning environment, 3 students did not do good, you can divide them into two groups.

Then you can compare these two, whether, which one is doing good which is not doing good why the student is the group is not able to do well. Because they were not able to do the transition properly between the models or they stuck somewhere, they are going in a loop you can find out and add information's without much effort.

So in this figure, it shows that it is not just a direct jump and QIM to QM that student would have been like directly computing QIM to QMEC instead of a lot of other loops as possible because some other students would have done a different way. So, and this also tells you that if you have a proper logging mechanism.

If n is 1 or n is 100 you can compute this kind of diagram easily because it is all it is all the same for the computation scripting all these thing of course completion time is more I am saying it will be simpler to scale. If you have logging mechanism with the proper time and action Id's.

(Refer Slide Time: 15:17)

So, in this video, we saw an example of how to look at the learner's behaviour by mining their actions by plotting a state transition diagram. In next week we will discuss in detail sequential pattern mining and the process mining and once we do that, we will go on to a clustering mechanism that will end the diagnostic analytics part of the leaner matrix course. Thank you.