Learning Analytics Tools

Professor Ramkumar Rajendran

Educational Technology

Indian Institute of Technology Bombay

Lecture 32

Diagnostic Analytics - SPM

Welcome back to learning analytics tools course. In this week we will continue diagnostic analytics, we will see sequential pattern mining, differential sequencing mining and process mining in this week. You might have noticed that the first two-three weeks we are talking about data collection, different environments, how to collect data, that is to give you motivation, what data to collect and what environment to consider.

From last week we started talking about algorithms or the tools we can use to apply. So, whenever we give a demo of a tool, please try to explore with different algorithms, think about what we are learning here, so that will help you to understand how these tools are used for education data.

(Refer Slide Time: 1:00)

So, what is sequential pattern mining? Consider your sequence of actions, say you have collected student's behaviour in a MOOC, or in TELE, like “MettLE”, or some other environment or moodle. So, you have a sequence of actions of students interacting with your system. So, a student asked to log into a system and they are doing sequence of actions, say actions such as read, go and watch the video and answer the questions. And, each action is different. A student can read a page, again, he can read another page, he is not on the same page but he is reading, he is not continuing the same page, but he is reading(another action).

So, read can be continued by another read action, but each reader is different. It is not on the same context, or context varies. So, let us consider that as two actions read content one, read content two, we can talk about how to merge everything later, but let us consider that each action

is different. So, you need to identify a unique set of actions. So, in the last week, we saw that in "MettLE" we classified the group, the action into six or seven sets.

So, like a functional model planning, qualitative model, quantitative model like that you need to come up with a unique set of actions. In a MOOC, it is simple, MOOC we have only few set of actions possible like reading, watching the video, interacting with the forum, or answering the questions, assignments, so that kind of simple set of actions. So, let us create such a set of actions the unique set of actions, arrange it in a sequential manner as it appeared based on the timestamp you captured.

The example action sequence is here, for example, this is in MOOC, a student is actually watching a video after watching the video is adding some post, so the student is adding, going to the comment in the forums and adding a new post or is going to upvote someone's already added post. And he is reading a PDF. We do not know, what is he reading? It might be in the context of information.

Let us see the action is reading then is taking the quiz. We do not know which question he answered, what is the response? Let us see, he is in a question just looking at the page again reading, again watch the video, add a post, quiz, read, upvote, read, doing this kind of sequence of actions.

Please understand this sequence of actions clearly a student enters to the MOOC and watches a video after that he is adding a post and upvoting then he goes back to some other page where he is reading the navigation is not given but the navigations can be understood indirectly from this kind of sequence of actions.

He is reading a new page in an online resource you provided, reading particular page PDF, after that he is again going back to the quiz, to answer some question but he is not able to answer, going back read again is not able to understand that. They might be watching the video to understand, if they still do not understand, they go and ads a post saying that particular question is not discussed in the class or something.

So, then he posted it, again going to quiz and trying to understand, read, upvote. There might be something he was looking for in the forum then again read so this is a set of sequence of actions

a person is doing in a MOOC consider that. How do you say sequence is based on a timestamp we captured.

But, how long a student was watching the video? How much time he spent on like reading the PDF, or how many questions he answered in a quiz that is not captured yet? Let us consider just the sequence of action this will capture the frequency, last week we saw just the process model, just this transition between each actions now we are considering the frequency, consider that we will consider frequency in pattern mining.

So, what are the patterns we might have from this "watch video, add post"? So, see this has repeated twice this interesting right? This is the action, a student does twice, it happened twice, There is a sequence of actions "upvote and read" - this sequence of action also appeared twice, The action "quiz and read" -This also appeared twice that might be other actions which might be appearing twice because I did not check it but let us see - "read and upvote" it is not occurring twice, a lot of combinations can be possible but only these set of actions occur twice other sequences have occurred only once. So. the patterns from this particular set of sequence of action from one student can be these three patterns possible like

- "watch the video, add post"
- "upvote, read"
- "quiz, read"

(Refer Slide Time: 6:11)

Let us do a small activity, the sequence of actions, you have example sequence of actions given from the last slide and there are three students actions, you have to find patterns. So, that 3 students actions, the sequence of actions given here when they interact with a MOOC, you trying to find out the patterns in this three sequence of actions. Please pause the video, after you find out the patterns, kindly resume.

Please pause the video, try to find out the sequence of actions, we will discuss a few of them.

(Refer Slide Time: 6:47)

So, let us look at the three students sequence of actions S1, S2, and S3. And our aim is to identify which action sequence occurs in which student interaction, so instead we try to create this kind of table. Let us see this is a table and I will explain this table in a minute. Let us see

- V is watching the video,
- Q is the quiz.

How many times the sequence of action/pattern - "video followed by a quiz action" occurred?

- So, video followed by quiz occurred once for student 1
- It did not happen, this pattern did not occur for student 2
- This pattern occurred for student 3 video.

So, 1 time for student 1 and 1 time for student 3, 0 time for student 2.

So

$1 - 0 - 1$

The meaning is- out of three students this pattern occurred for two students. How many students have this particular pattern? 2 students, because 2 students have it, you understand this value 2, and this pattern occurred for student 1, this pattern did not occur for student 2 and this pattern occurred for student 3, 1 time each.

Let us look at the 2nd pattern -"video, followed by a video".

- "Video followed by video did not occur for student 1"
- It did not occurr for student 2
- It occurred for student 3

So, Out of 3 students only 1 student had this particular pattern.

The pattern can be not just two action sequence, it can be 3 also it can be 4 also, it depends on how much complexity you want to use. If you can make sense of 4 actions in a sequence and that makes some inference for your hypothesis, please use that.

Lets considered here, “video, quiz and read”

which means this particular sequence of actions this sequence of action is actually happening, whenever the student is watching a video, then taking the quiz, and is immediately going back to read, so this is important to compared to this(“Video-Quiz”).

So, 1 0 1,

- it occured for S1 and S3,
- not for S 2

It occurred for 2 students.

Similarly, you can compute for the others, other sequences of actions.

Now, I will show you some actions which might occur twice, so consider “read read”.

“read-read” occurred once for student1, not for student 2 and student 3. But in this case, the pattern occurs only one once for student 1.

Now consider, -“read, quiz”,

so “read, quiz” occur everywhere like “1- 1- 1” and all 3 students have it

So let us consider these two set of actions, few more patterns are possible from it. So, the pattern can be the same action repeating i.e. unary action repeating multiple times can be a pattern.

A student is reading, reading, reading, it is also a pattern. It spends a lot of time it might be a strategy for him. “Reading and taking quiz” occurring in a loop cab also be a pattern. As the student is reading immediately taking quiz, immediately reading and taking the quiz, some students do not read anything, they just take the quiz and watch videos, quiz and watch videos, it might be also pattern.

So, that is what we try to identify. I gave the examples of few patterns that can be identified from these 3 students and I want you to create this kind of table to understand this pattern occurred for student 1, student 2, student 3.

For example, here this “quiz to read” pattern occurred for student 1 once, student 2 once, student 3 twice, so “quiz and read” occurred twice for student 3, it occurred for all 3 students. You might think it just a simple sum of these number, no, it is just how many students had this pattern, like the 3 students, that is all 3 students had this particular pattern, it might have occurred in different frequency among different students, that is you have to understand there are two differences. Identifying pattern that occurred on student 1, student 2, student 3 and identifying total frequencies. This particular column, this column says, whether how many students have this pattern, student1, student2, or student3, so just to differentiate I added this particular pattern here, hope you understood this table. It is very important to understand this table, to go further.

(Refer Slide Time: 13:38)

So, now you understood this table, let us see two metrics, sequential pattern mining has only two metrics, we have to understand these two metrics. That is very basics, is enough to understand the sequential pattern mining, and talk and explore and do analysis on it. One is sequence support.

That is s support, we call it as s support that captures the number of individual action sequences for a group where that sequence of actions occurred at once or more. What it says is - “the number of individual actions sequence occurred for a group where the sequence of actions occurred at least once”.

That simply tells how many students have it that is, what is s support, sequence support, this says the number of this action sequence occurred to work group of students. Students should have it at least once. So, here it is 3, 2, 1, this is basically s support, this particular column.

Let us look at the, i support, instance support is defined as the number of times the pattern occurs within an action sequence for an individual in a particular group. The group is of 3 students, out of 3 students for each student, how many times the particular pattern, the pattern occurs within an action sequence for an individual, this is what this “i-support” like a “1 -1- 2”, this is what “i-support” means. “i support” tries to capture how many times the particular pattern occurred for an individual in a group.

The group is 3 of students here. For this group, all 3 students, how many times each pattern occurred 1 1 2, this is called “i-support”. These two are very key and important metrics for SPM, sequential pattern mining, only these two key metrics is enough to understand the sequential pattern mining.

(Refer Slide Time: 15:45)

So, let us use “i-support” and “s-support” for this particular table. Let us take the i support, from the “i support” we can compute “i frequency mean” or “i frequency standard deviation”, I frequency mean is basically finding the average of that sequence of actions. For instance “quiz to read”, occurred once 1 for S1, once for S2 and twice 2 times for S3.

So, if you add all the numbers, therefore,

$\frac{1+1+2}{3} = 1.33$

how many students it occurred, 3 students, so 1.33. So, the mean of this particular pattern is 1.33. If we have a lot of students say 10 students or 60 students and you computed their actions , you can compute the mean, median also the standard deviation.

So, if the mean is too far away from standard deviation is best to capture the median, just telling you that you can use median instead of the mean. But, i frequency mean means this value, i frequency median is another measure you can compute. S support is 1 when, all 3 students have that particular patterns.

So, $\frac{3}{3} = 1$

(Refer Slide Time: 17:06)

In this video, we introduced what is sequential pattern mining. We will give some examples in the next video. Thank you