Learning Analytics Tools

Professor Ramkumar Rajendran

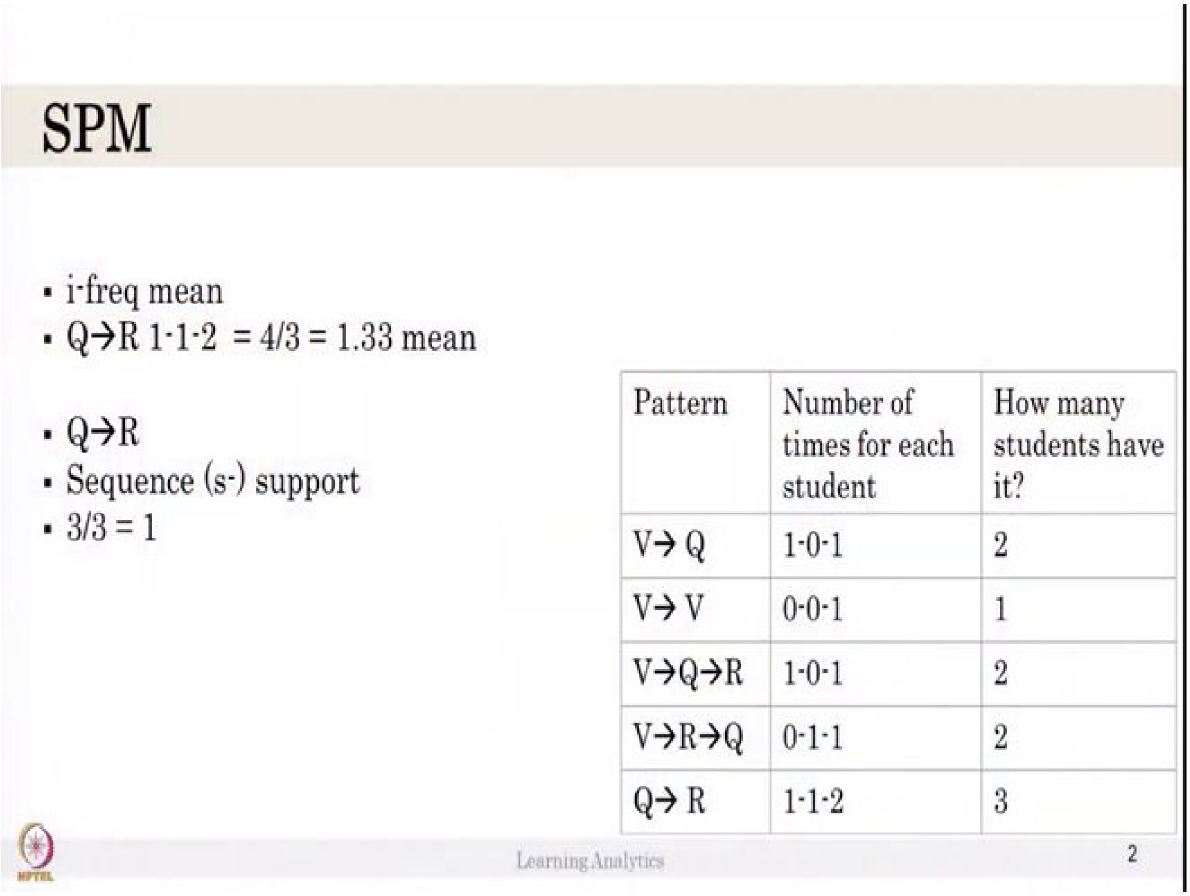
Educational Technology

Indian Institute of Technology, Bombay

Lecture 33

SPM II

(Refer Slide Time: 0:24)



SPM

- i-freq mean
- Q→R 1-1-2 = 4/3 = 1.33 mean

- Q→R
- Sequence (s-) support
- 3/3 = 1

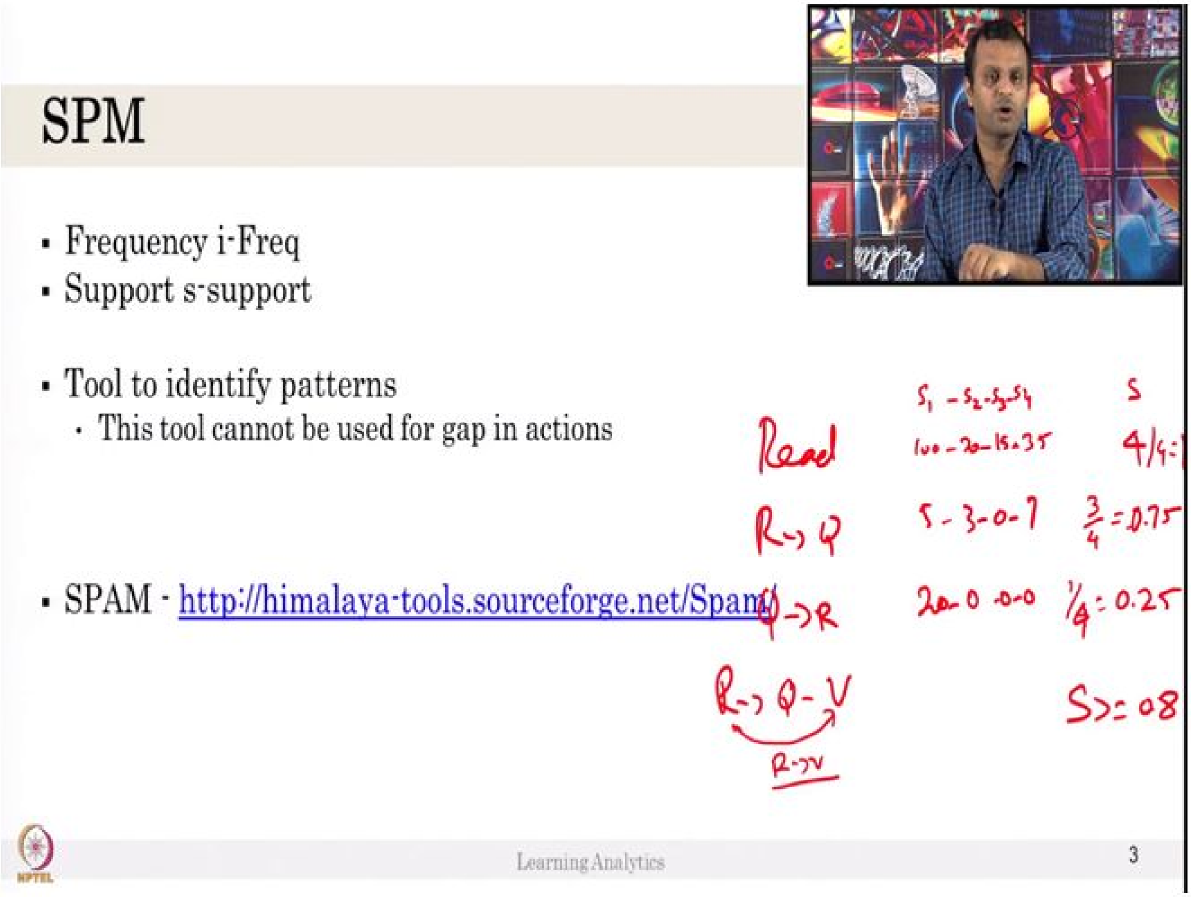
Pattern	Number of times for each student	How many students have it?
V→ Q	1-0-1	2
V→ V	0-0-1	1
V→Q→R	1-0-1	2
V→R→Q	0-1-1	2
Q→ R	1-1-2	3

Learning Analytics

2

In this video let us continue sequential pattern mining with some examples. So, if you remember this table from the last video we computed “i-frequency mean” and “s-support”, lets try to understand this in detail.

(Refer Slide Time: 0:32)



SPM

- Frequency i-Freq
- Support s-support

- Tool to identify patterns
  - This tool cannot be used for gap in actions

- SPAM - http://himalaya-tools.sourceforge.net/Spam/

	$S_1 - S_2 - S_3 - S_4$	S
Read	100 - 20 - 15 - 35	4/4 =
R -> Q	5 - 3 - 0 - 7	$\frac{3}{4}$ = 0.75
Q -> R	20 - 0 - 0 - 0	1/4 = 0.25
R -> Q - V		S >= 08
R -> V		

Learning Analytics

3

So, what is “i-frequency”, what is “s-support”, it is very important to understand these two values. Identified patterns the patterns can be just a single action. Suppose we have only 4 students and this single action read would have occurred 100 times, 20 times 15 and say 35 times for each of the students. So, the s support will be 1, because

$$\frac{4}{4} = 1$$

So, s support value is 1.

A single item can be also a pattern. But this is just an action distribution, you can just plot it, each action along with the frequency it occurred in descriptive analytics. So, do not consider the single action as a pattern for analysis, but it can be a pattern. Do not avoid single actions. When you apply any pattern mining algorithm on the sequence of actions on your data, it tries to give you the single-action pattern also, but you can ignore them because these are the simple distribution of how many times this particular action occurred on each student data which can be computed in descriptive analytics, if you know how to count the individual actions in each sequence.

Let us see their second action say "read to quiz". It occurred

"5- 3 -0- 7", some values.

So, it occurred for 3 students, which means

$\frac{3}{4} = 0.75$

So, you have to understand, it occurred to 3 students. So, it is

$\frac{3}{4} = 0.75$

And this value will indicate how many students have this particular pattern whether this pattern is important or not.

For example, if I identified a pattern say "quiz to read", if I have a quiz to read, and it occurred only for student1 20 times, 0 for others. It is very interesting, but only 1 student has it. Do you want to consider this pattern "quiz to read" in your analysis that will be defined by the "s-support". So, when we had to run a pattern mining algorithm on a sequence of data, it tries to give all the possible combination of patterns from a single action, action compared to the other action.

Which pattern we should use for our inferences that will be decided by these two metrics, why these two metrics just to pick the right pattern to make the inferences, this tells you more about the patterns, the number of times it occurred, how many students got it, that is what these two metrics are.

Now, the idea is, if you want to consider the patterns which occurred at least 80 per cent of students, then you can pick “s-support” value is greater than or equal to 0.8. Suppose you have 60 students and you want to consider the patterns that are occurred for more than 80 per cent of students in your class.

You do not care about the other patterns because there will be too many patterns that are coming out from your pattern mining algorithm, which means you have to pick the right value. So, we usually keep”s-support” 0.6 or 0.5 based on the number of students. If the number of students are more, the number of combination of patterns are too high, we need to reduce it. We will not keep 0.5 will keep 0.8, but if the students are only 30 in number and the combination of actions are not too much, you are not making any inference. We can go down to 0.6 or 0.5 And this is what “s-support” tells you.

What I forgot to tell you is how to use this metric identify whether this pattern is evenly distributed across all students or not. Depending upon this, one should consider this particular pattern for analysis or inferences. Combination of “i-frequency” and “s-value” will tell you which pattern to pick. Let us look at some activities to make you understand these two. There are lot of tools to identify patterns, we will give you one tool, just a python script, one of our TA created it, we will share the tool, this will tell you how to use the tool to identify patterns if your sequence of actions arranged in a certain format (the format we saw in a previous example).

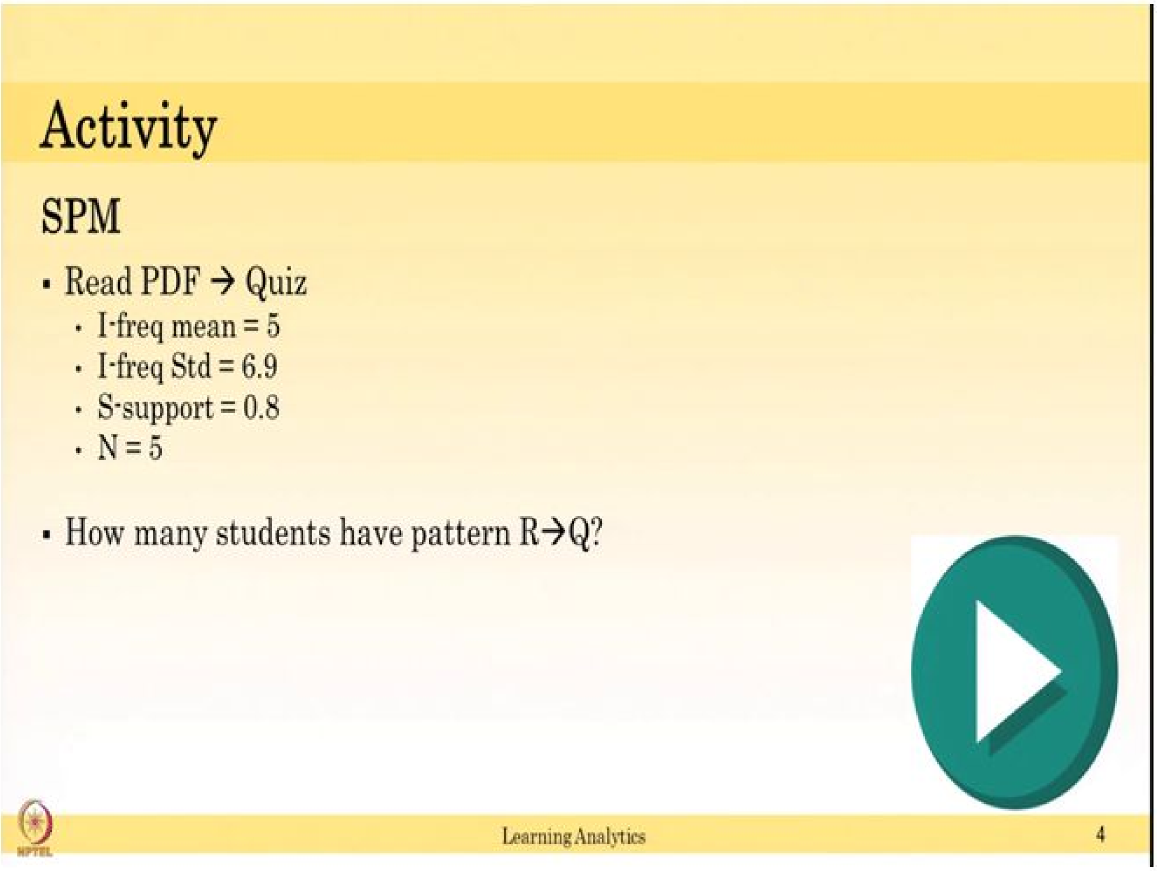
So, this tool cannot be used for the gaps in the actions. For example, some might be interested to identify the patterns with the one gap, “read quiz video”. I am interested in patterns which are occurring immediately like “read quiz”, also I am interested in a pattern which occurs with a gap of one action like how many times “read to video” occur. It should include both immediate occurrence also with a gap of one more action in between. This is just to avoid mistakes/slip i.e. a student really wanted to watch video, but have clicked something else say, went to the quiz but immediately went back to video.

If you want to consider this kind of gap also, this tool will not help. We might have to create a new tool something like that. But there are other tools available, which we use in our research. So, if you are interested, I will discuss a paper and you can look at the paper and use that tool.

So, this pattern mining is computationally really costly. In a sense, it takes a lot of computational power. If you want to identify all the patterns, it \ take the first action read it, combine it all other actions and deep search happens, so I do not want to go into that instead, the spam algorithm you created by this, in the particular slide will actually helping you to identify the pattern mining in less computational costly. So, if you want to know about sequential pattern mining, how the algorithm works, you can use check this particular page and read about it.

But in this course, we will give you the tool to identify patterns from the sequence of actions. We want you to understand what this particular tool gives you? What "i-frequency" means or whether you can compute the "i-frequency" or "i-frequency mean" or "i-standard deviation" or "s-support"? What these two metric means we want you to understand that is why I was trying to teach. What is the input to this pattern mining tool and what is the output metrics which can be displayed and which can be used for analysis? If you want to know how this tool works go and read this page.

(Refer Slide Time: 8:05)



Activity

SPM

- Read PDF → Quiz
  - I-freq mean = 5
  - I-freq Std = 6.9
  - S-support = 0.8
  - N = 5
- How many students have pattern R→Q?

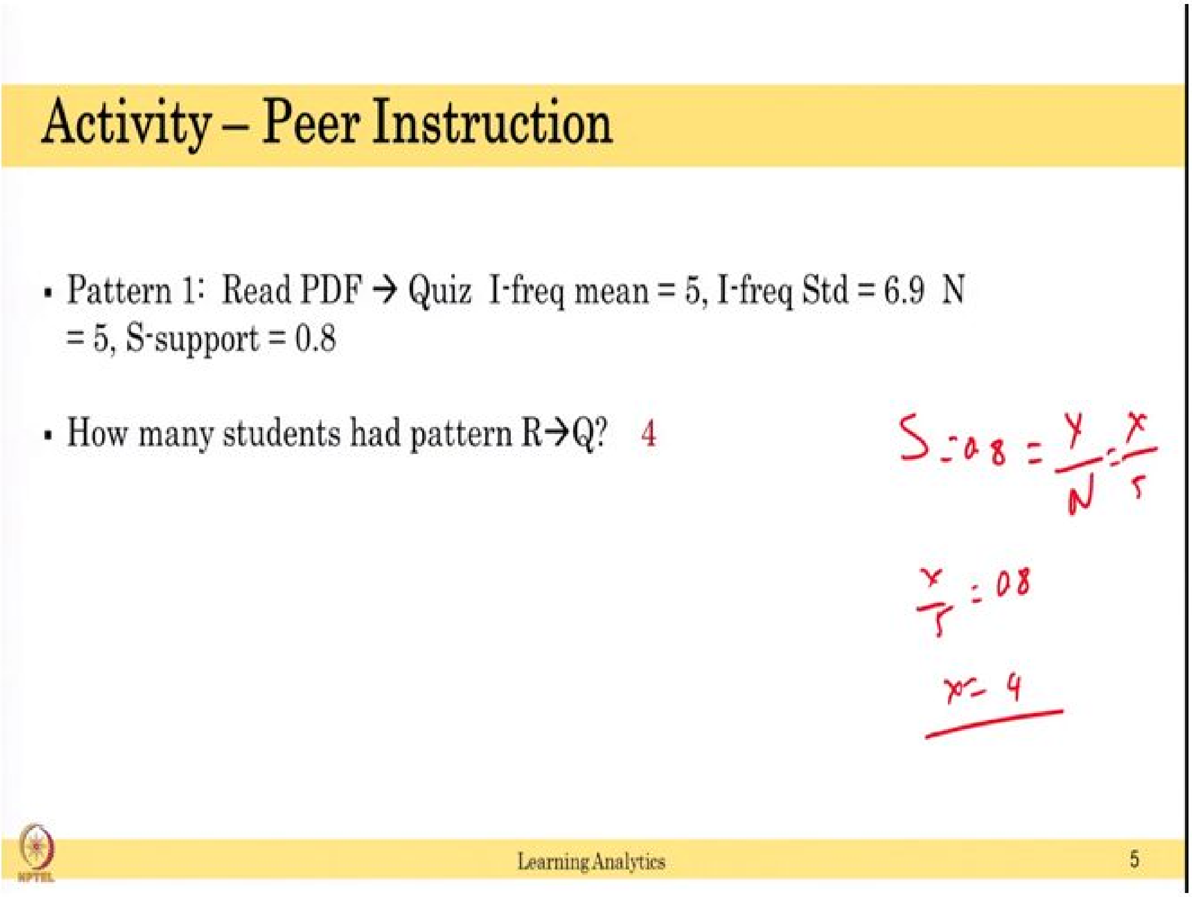
Learning Analytics 4

So, simple activity, in this activity, consider there are 5 students,

$N = 5$

and “i-frequency mean” (that is average) is 5 and “i-frequency standard deviation” is 6.9 and “s-support” is 0.8. Given these two metrics like “i-frequency mean” and “s-support” and “standard deviation”, N equal to 5, how many students have pattern “read to quiz”, the pattern is “read to quiz” how many students have pattern “read to quiz”, pause the video answer this question then continue.

(Refer Slide Time: 8:49)



Activity – Peer Instruction

- Pattern 1: Read PDF → Quiz I-freq mean = 5, I-freq Std = 6.9 N = 5, S-support = 0.8
- How many students had pattern R→Q? 4

$S = 0.8 = \frac{Y}{N} = \frac{X}{5}$

$\frac{X}{5} = 0.8$

$\underline{x = 4}$

Learning Analytics 5

So, it is simple. Because I said s-support equal to 0.8, which actually tells how many students have it, s support is the column, we said how many students have it how many students have the pattern is basically 0.8, since

$N = 5$

s support = 0.8

which means number of students having this patter (say x) divided by total N will be 0.8

$\frac{x}{N} = 0.8$

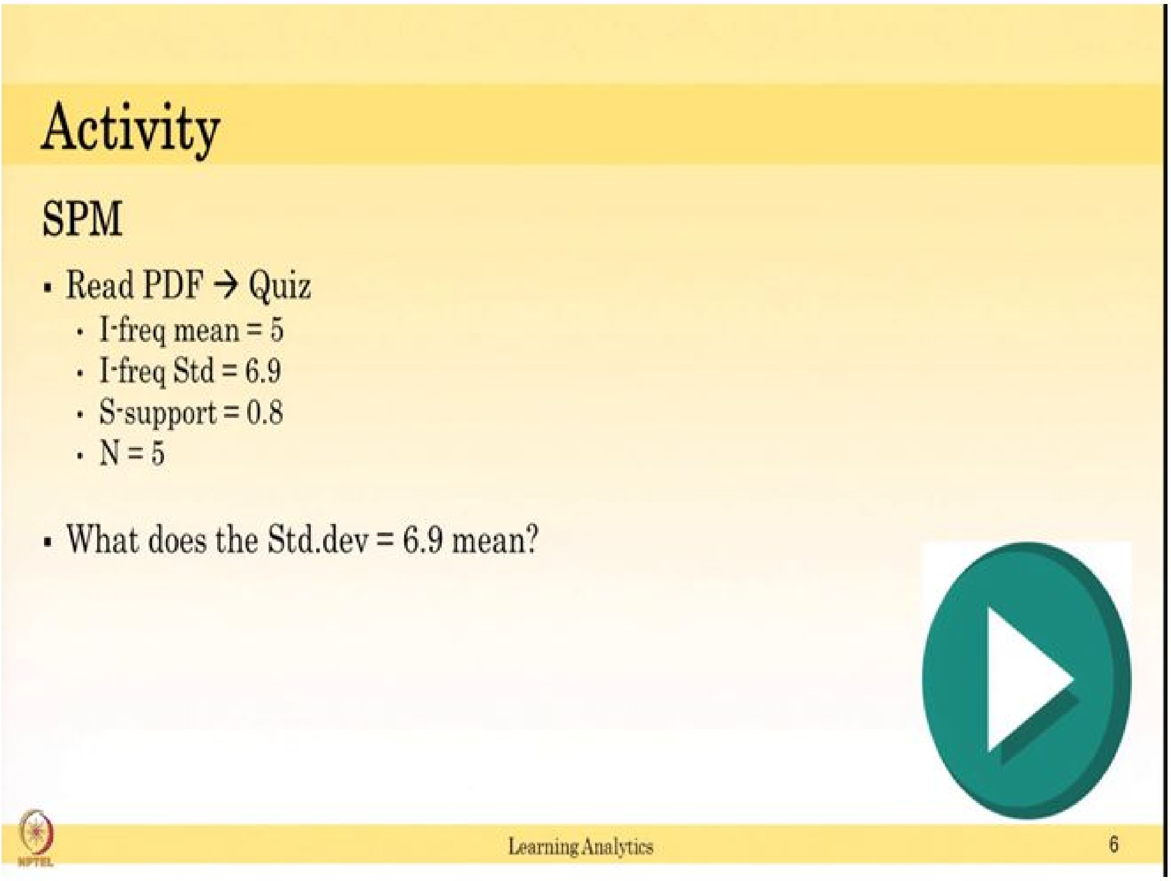
What is the X?

So,

$$\frac{x}{5} = 0.8$$

so X equal to 4, so 4 students have this particular thing that it is simple.

(Refer Slide Time: 9:27)



Activity

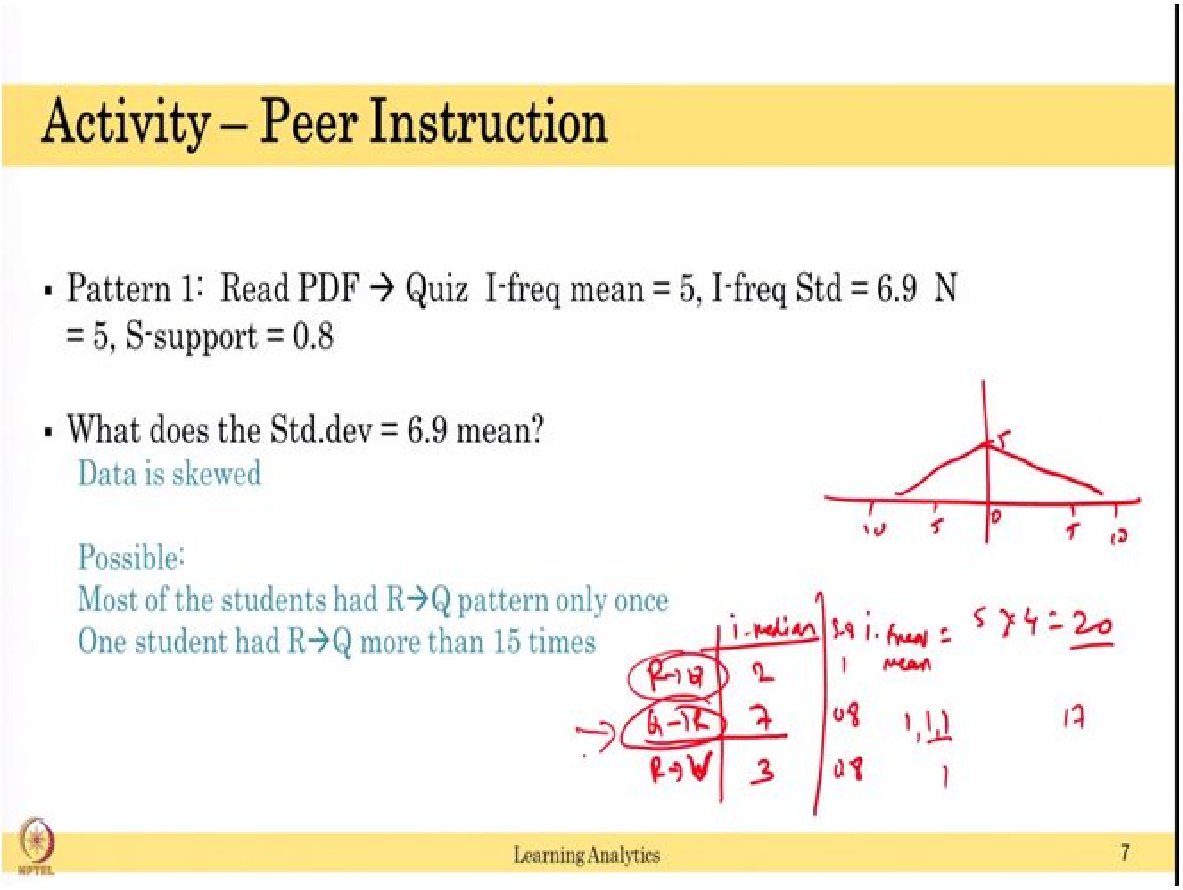
SPM

- Read PDF → Quiz
  - I-freq mean = 5
  - I-freq Std = 6.9
  - S-support = 0.8
  - N = 5
- What does the Std.dev = 6.9 mean?

Learning Analytics 6

Let us see a bit tricky activity, same metrics. And, what does standard deviation 6.9 mean? We know i-frequency mean. But, what is this 6.9 means? Is that good/bad, what is this particular thing means? Think about it. List down your answers, then resume video.

(Refer Slide Time: 9:50)



Activity – Peer Instruction

- Pattern 1: Read PDF → Quiz I-freq mean = 5, I-freq Std = 6.9 N = 5, S-support = 0.8
- What does the Std.dev = 6.9 mean?

  Data is skewed

  Possible:

  Most of the students had R→Q pattern only once

  One student had R→Q more than 15 times

Learning Analytics 7

It tells the data is skewed because the mean is 5 and 6.9 standard deviation means the data is definitely skewed. What is mean in standard deviation in a general plot? It is not the perfect curve or something like that, which means more standard deviation implies graph is kind of skewed. So that is why I said you should use median instead of for mean.

But, let us see how it goes. There are possibilities that most students have pattern only once, there are 5 students, we found out there are only 4 students have this particular pattern. So, the mean computer is for 4 students. So, the i frequency of 5, it occurred 20 times, this pattern occurred 20 times for all students together. Because

$$\frac{20}{4} = 5$$

because we have 4 students that is where frequency mean comes in.

Now, consider if the standard deviation is really high, but then there might be a chance that 3 students had pattern only once and 1 student had a pattern around 17 times, it is possible, it is just 1 possibility to get this particular standard deviation and mean. So, altogether it is a 20. So, I mean is good. So, when you look at the i-frequency, also check the standard deviation. The standard deviation tells you a better meaning that is where the data is skewed, or better use the median value. So, now if you want to rank the patterns, we have to pick the pattern which is which occur more times.

In the last video, we saw what is s support means. Now, we learnt that i-frequency median or i-frequency mean plus standard deviation will tell you a better story. The story is, suppose there are patterns- “read to quiz”, “quiz to read” and “read to watch video”, something like that. If you have i median

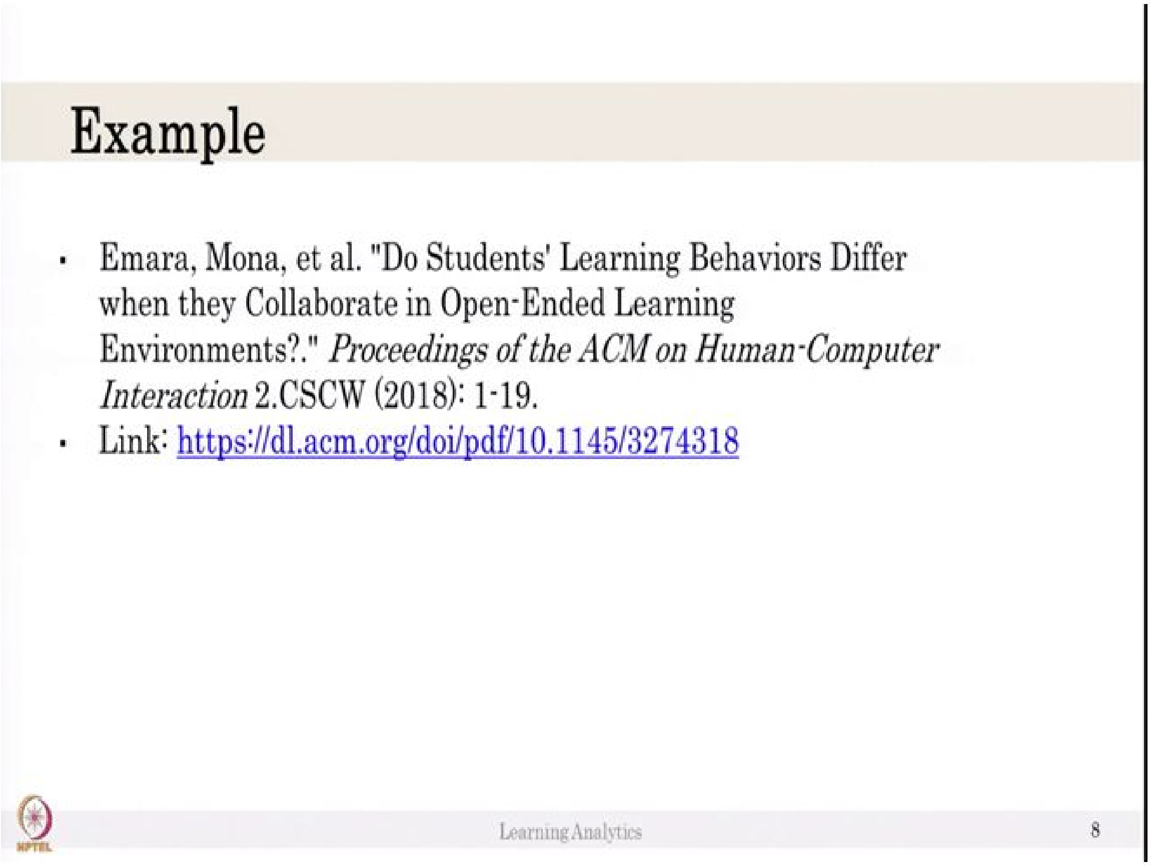
R-Q	2
Q-R	7
R-V	3

and all of them are above 0.8, i.e we just filtered all the patterns below 0.8, s support out of this, we only consider the patterns which are above 0.8 s support.

Now, if I want to order, I want to say this particular pattern is more interesting because this pattern occurs in 80 per cent of students also it occurred more number of times for each student i.e. almost all the students would have got this pattern 7 times. So, we can take the median, but if you say mean then you should consider standard deviation also.

Be careful if you take a mean, look at the standard deviation. If you consider median, you can consider this pattern might have occurred 7 times for each student. So, this pattern is more important compared to this pattern. This pattern also occurs for all students but only occur twice. So, this is maybe a strategy all your students will be trying to do your. Students might be always trying to take the quiz then read ,might be some strategies students are coming up with. So that you have to consider, hope you understood what is the meaning of s support and i-frequency.

(Refer Slide Time: 13:54)



Example

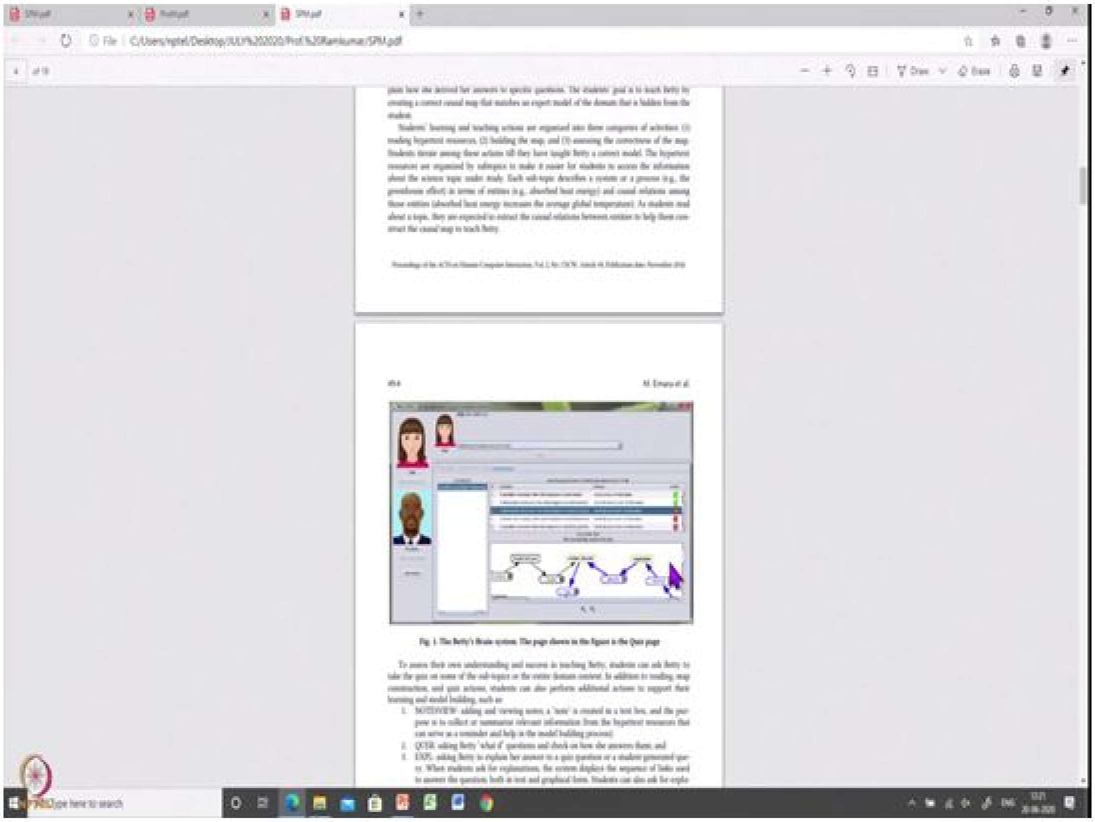
- Emara, Mona, et al. "Do Students' Learning Behaviors Differ when they Collaborate in Open-Ended Learning Environments?." *Proceedings of the ACM on Human-Computer Interaction* 2.CSCW (2018): 1-19.
- Link: https://dl.acm.org/doi/pdf/10.1145/3274318

Learning Analytics 8

Let us look at one example of using pattern mining for analysis. And, this is a paper published in CSCW in 2018. Let us look at the paper. So, in this paper, the authors used a system called that is “Betty’s Brain” and it has set of actions, like the reading action the student can read and they can add a note, so take a look at the notes, maybe notes action, and they can add concepts because in this particular thing, they have to create a concept map, they have to add concepts and links between it so they are saying adding concepts, adding a link or asking help from the other agent or they can take the quiz or look at the quiz they have taken, they can ask for an explanation.

So, consider they have a set of 8 or 9 actions identified from their learning environment. No need to understand all this, but let us consider they have identified eight sets of unique actions in their learning environment.

(Refer Slide Time: 15:06)



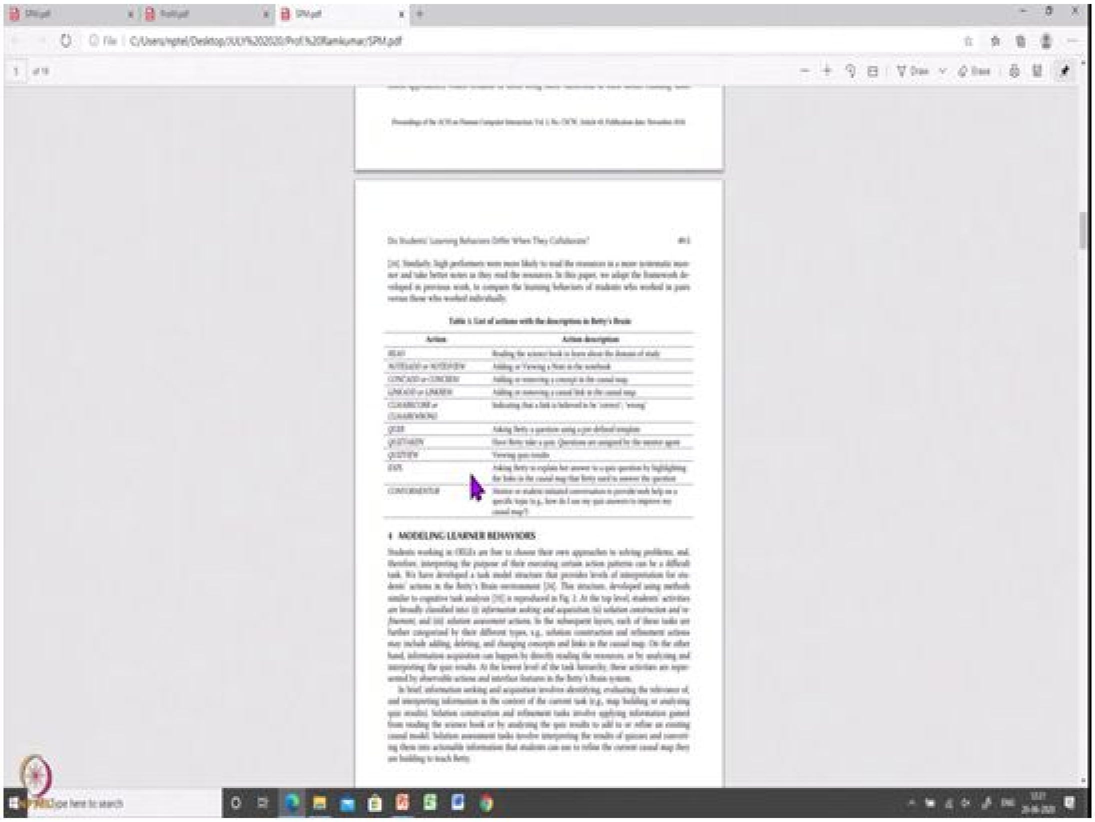
plain how she derived her answers to specific questions. The students' goal is to teach Betty by creating a correct causal map that matches an expert model of the domain that is hidden from the student.

Students' learning and teaching actions are organized into three categories of activities: (1) reading hypertext resources, (2) building the map, and (3) assessing the correctness of the map. Students iterate among these actions till they have taught Betty a correct model. The hypertext resources are organized by subtopics to make it easier for students to access the information about the science topic under study. Each sub-topic describes a system or a process (e.g., the greenhouse effect) in terms of entities (e.g., absorbed heat energy) and causal relations among those entities (absorbed heat energy increases the average global temperature). As students read about a topic, they are expected to extract the causal relations between entities to help them construct the causal map to teach Betty.

Fig. 1. The Betty's Brain system. The page shown in the figure is the Quiz page

To assess their own understanding and success in teaching Betty, students can ask Betty to take the quiz on some of the sub-topics or the entire domain content. In addition to reading, map construction, and quiz actions, students can also perform additional actions to support their learning and model building, such as:

1. NOTEVIEW: adding and viewing notes; a 'note' is created in a text box, and the purpose is to collect or summarize relevant information from the hypertext resources that can serve as a reminder and help in the model building process);
2. QUER: asking Betty 'what if' questions and check on how she answers them; and
3. EXPL: asking Betty to explain her answer to a quiz question or a student generated query. When students ask for explanations, the system displays the sequence of links used to answer the question both in text and graphical form. Students can also ask for expla-



[24]. Similarly, high performers were more likely to read the resources in a more systematic manner and take better notes as they read the resources. In this paper, we adopt the framework developed in previous work, to compare the learning behaviors of students who worked in pairs versus those who worked individually.

Table 1. List of actions with the description in Betty's Brain

Action	Action description
READ	Reading the science book to learn about the domain of study
NOTEADD or NOTEVIEW	Adding or Viewing a Note in the notebook
CONCADD or CONCREM	Adding or removing a concept in the causal map
LINKADD or LINKREM	Adding or removing a causal link in the causal map
CLMARKCORR or CLMARKWRONG	Indicating that a link is believed to be 'correct', 'wrong'
QUER	Asking Betty a question using a pre-defined template
QUIZTAKEN	Have Betty take a quiz. Questions are assigned by the mentor agent
QUIZVIEW	Viewing quiz results
EXPL	Asking Betty to explain her answer to a quiz question by highlighting the links in the causal map that Betty used to answer the question
CONVORMENTOR	Mentor or student initiated conversation to provide task help on a specific topic (e.g., how do I use my quiz answers to improve my causal map?)

4 MODELING LEARNER BEHAVIORS

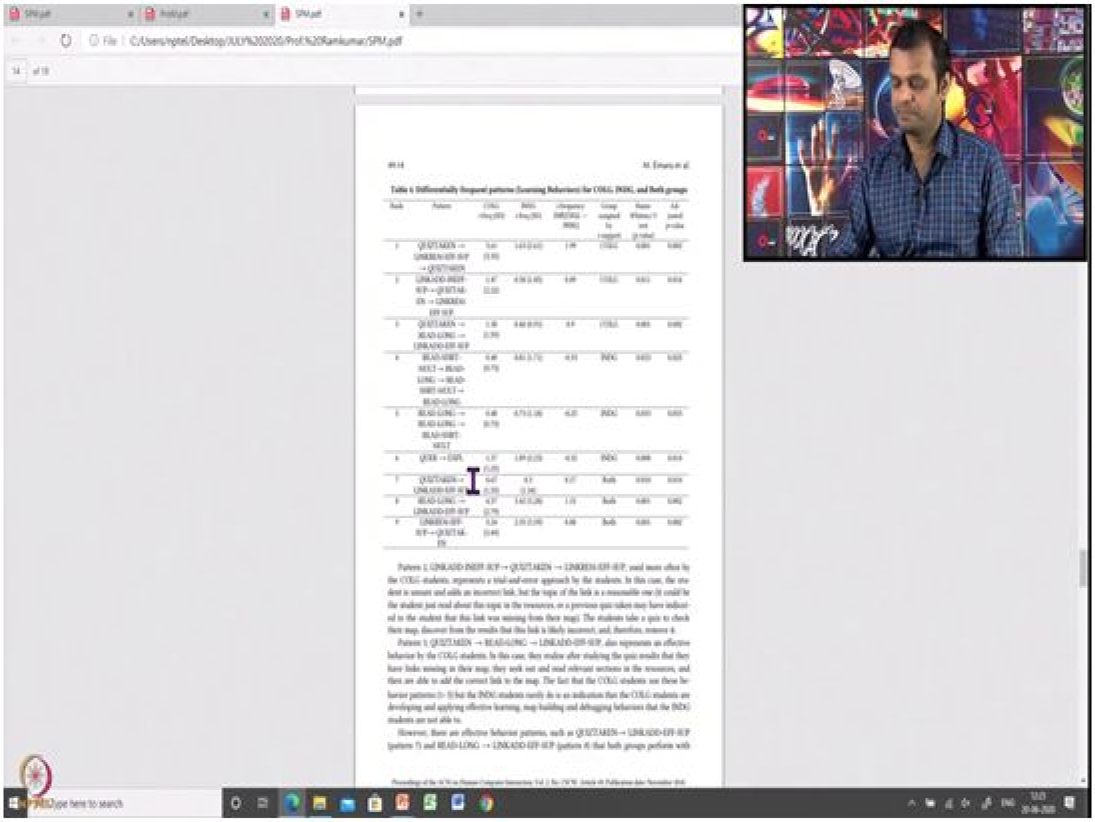
Students working in OELEs are free to choose their own approaches to solving problems, and, therefore, interpreting the purpose of their executing certain action patterns can be a difficult task. We have developed a task model structure that provides levels of interpretation for students' actions in the Betty's Brain environment [24]. This structure, developed using methods similar to cognitive task analysis [35] is reproduced in Fig. 2. At the top level, students' activities are broadly classified into (i) information seeking and acquisition, (ii) solution construction and refinement, and (iii) solution assessment actions. In the subsequent layers, each of these tasks are further categorized by their different types, e.g., solution construction and refinement actions may include adding, deleting, and changing concepts and links in the causal map. On the other hand, information acquisition can happen by directly reading the resources, or by analyzing and interpreting the quiz results. At the lowest level of the task hierarchy, these activities are represented by observable actions and interface features in the Betty's Brain system.

In brief, information seeking and acquisition involves identifying, evaluating the relevance of, and interpreting information in the context of the current task (e.g., map building or analyzing quiz results). Solution construction and refinement tasks involve applying information gained from reading the science book or by analyzing the quiz results to add to or refine an existing causal model. Solution assessment tasks involve interpreting the results of quizzes and converting them into actionable information that students can use to refine the current causal map they are building to teach Betty.

What they did is if they ran a pattern mining, so they grouped the students into two groups, they might have ran a study. So, for example, this is a group called the collaborative group, this is the individual group, what they did they identified a pattern quiz taken then they remove particular effective link and they again take quiz.

So, that particular pattern occurred i-frequency mean is 3.61 and a 3.35 standard deviation. So, like that they were trying to plot the important patterns from the data, this is the table. So, if you want to know how they use these two patterns and the values of i frequency and s support read a paper bit then you will understand it better.

(Refer Slide Time: 16:11)



So that is example paper, where you can check how they are using the pattern mining metrics to make inferences.

(Refer Slide Time: 16:20)

Summary

- I-frequency
- S-Support

Learning Analytics 9

So, in this video, we saw what is i-frequency and s-support in detail also, please check the paper if you get time. That is also kind of used for one of the questions in the assignment. So, it is not that paper reading is optional. We expect you to read the papers and understand. Why that particular metrics are used? Is it right? Think about it and this will help you to start with how to read research papers in learning analytics. So, this video is talked about i-frequency, s-support. Next video we will talk about the application of SPM. Thank you.