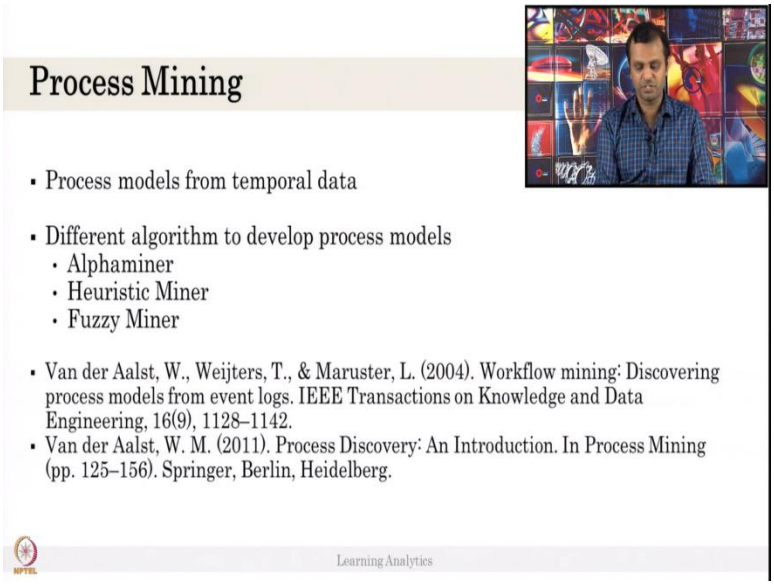
**Learning Analytics Tools**
**Professor Ramkumar Rajendran**
**Educational Technology**
**Indian Institute of Technology, Bombay**
**Lecture 35**
**Process Mining**

This is the last video in this week - Process Mining. So, process mining is like getting processed models from temporal data.

(Refer Slide Time: 00:25)



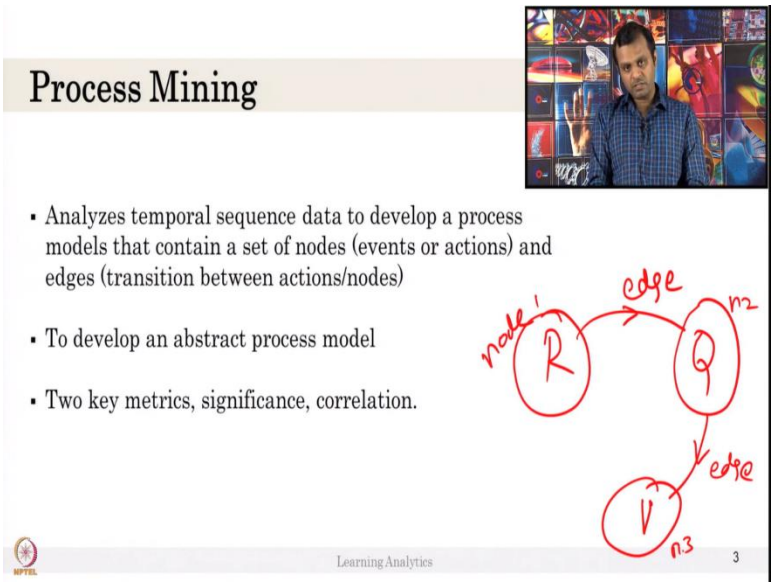

It is commonly used in customer care centres or banking or in the health care sectors where, suppose a call centre person calls you first, okay? Based on your responses, your profile responses, what is the next process to do? If there are, branched out into three or four responses, what is the next process? Should we send an email, should we call him back again and what is the process of applying for a credit card, to the end of credit (card closing)?

What is the process of taking a loan - start to end? So, this is associated with start to end, what is the process. And this has been used heavily in bank sector or customer care sectors. But this also can be applied to education data; very few researchers actually use process models for education data. It is interesting actually if you know what process model is and you can also apply, and you can understand how the student's interaction behaviour in a complete process.

So, there are different algorithms to develop the process model - Fuzzy Miner, Alpha miner or Heuristic Miner. You can go ahead and read about this, but let's look at one of these models in this course.

So, the process model, if you want to know the process mining software, more about the process models or Fuzzy Miner, I recommend you to read these two papers. These two papers are not for your assignment or anything; this is all extra course, extra work for you, if you are interested to understand what is the process model and it gives you examples how it is created and everything. So, especially the authors, the Van der Aalst is actually one who created the tool which we are going to use in this course.

(Refer Slide Time: 02:16)



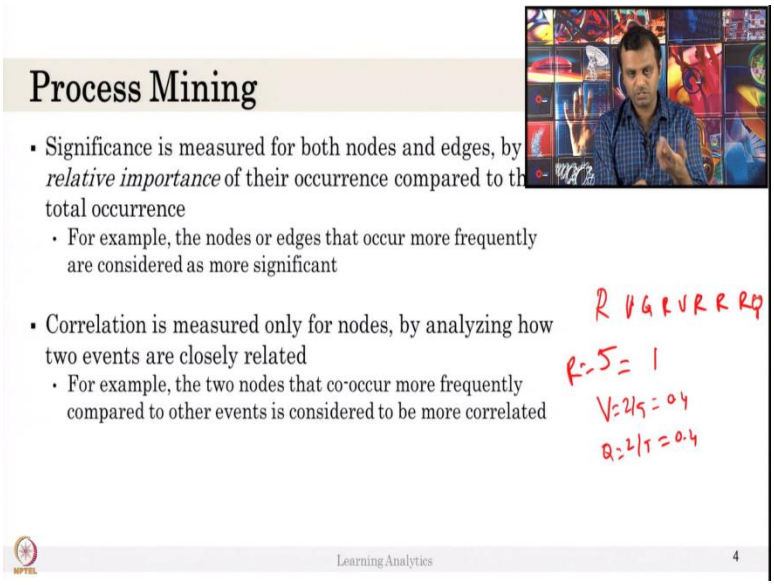

So, process mining analyses temporal sequence data to develop a process model that contains set of nodes (events or actions) and edges (transitions between actions or nodes). If you see nodes and edges, you might have remembered that last week video, we computed a state transition diagram. We mentioned there is a node and edge kind of things. Exactly the same kind of patterns we will use for the process model.

So, nodes can be the actions and transition between actions is called as edges. It aims to develop the abstract process. I have two key metrics, only two things to remember in each algorithm. Let

us keep, see the two key metrics in this model or so, like a process mining significance and correlation.

So, what is node - suppose there is an action read, there is an action quiz. This is considered to be a node. Node A is Read, node B is quiz, okay? The transition between these nodes, take V to quiz transition is the edge. Okay? Maybe this goes to watch the video this another edge. This is node 1, node 2, node 3. We will see what is two metrics for these nodes and edges.

(Refer Slide Time: 03:56)




So, significance is a measure for both nodes and edges by the relative importance of their occurrence compared to the total occurrence. Simply the node or edge that occur more frequently is considered to be more significant compared to other nodes and other edges. For example, it is simple. We will see a detailed example for nodes and edges computed in some sequence. But I just want to give you an example of nodes and edges and significance here.

For example, there is an action 'Read', 'Watch video', 'Quiz', 'Read' and 'Quiz',

so what is the most significant action here? Its Read which occurred - 1, 2, 3, 4, 5 times. So, more significant action is 'read' - it occurred 5 times so we will be given first value for significance. Relative to these particular actions, significance of other actions are considered. For example, Quiz occurred twice - 2 by 5, it will be 0.4. Significance of watch video is 0.4, the significance of Quiz is 0.4, relative to the most significant action ( that is Read), Read is 1. Okay? That is how this significance is computed.

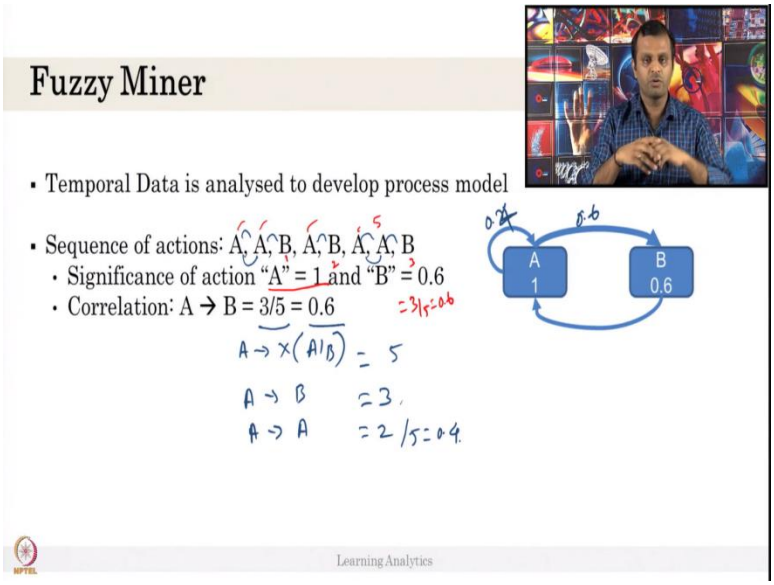
Significance is not only computed for the nodes, also for edges. Which edge occurs more repeatedly compared to the other edges? If you are more Read to Quiz, compared to other edges say Quiz to Watch Video or Quiz to Simulate or Read to Watch video, then this will be, Read to Quiz will the most significant, that will be 1. Other's significance will be computed based on that value.

And what is correlation? Correlation is how closely each two actions occur always - like which pattern is more frequent? Basically, it is a pattern of two actions - like Read and Quiz occurs most relatively, most occurred patterns, then that will be more correlated compared to the other things. So, correlation is simple, measure for, only for nodes not for edges by analyzing how two events are closely related. If Read occurs, always Quiz occurs with that. So, correlation is computed. This is exactly what we did in state transition diagram, that is kind of a correlation here.

(Refer Slide Time: 06:47)

Let us see a simple example, there is a temporal data - so there is a sequence of fractions - A, A, B, A, B, A, A, B. So can you compute the state transition diagram for this particular sequence? We computed last week video, but can you repeat it here in this week? After you create a state transition diagram, resume the video to continue.

(Refer Slide Time: 07:12)




So, I am not going to show the state transition diagram that is, you might have done it because we discussed that last week. Let us look at what is process model for that particular sequence of data and we will use a Fuzzy Miner algorithm. There are others like Alpha miner or Heuristic Miner, each has a different behaviour but let's look at the Fuzzy Miner in this course.

So, let us see. A occurred 5 times, okay? That is the most significant action, so A is equal to 1 because A occurred 1, 2, 3, 4, 5 times. And B occurred 1, 2 and 3 times. Which means B occurred 3 by 5, the significant of action B or the node B is 0.6 in this particular sequence relative to the most significant action (that is A).

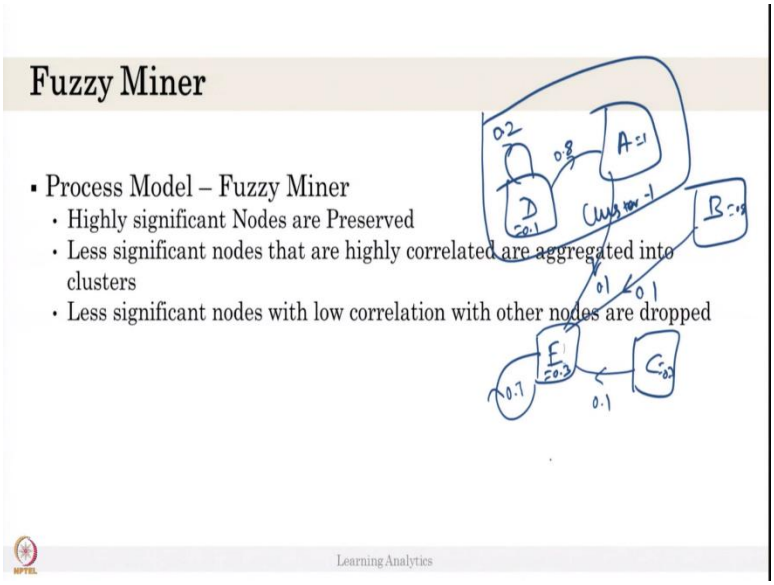
Similarly, what is the correlation that we saw in the state transition diagram, what is the transition property of A to B, significance of edges can be also computed, you can compute it. Let's look at the correlation for this example. A - there are how many actions, so let's look at the correlation in detail.

How many A to X  transactions are there? (X can be A or B), There are five actions, there are 5 transitions from state A. In that, A to A is 2. So, which means A to B will be 3, right? 3 plus 2 equal to 5. So the correlation of A to B is 3 divided by 5, it is 0.6. So, the correlation of that is 0.6. So, this is 0.6, okay? And correlation between B to A might be different. That will be 1, so this will be high. There is no self-correlation.

So, this is the self-correlation - A to A will be 0.4( 2 by 5). So, you can compute significance and correlation. It is not that you have to compute significance and correlation, that is enough for Fuzzy Miner. It is important you have to apply a set of algorithms to create the process model on this set of metrics. Consider you have these two metrics - significance and correlation and you have seven actions in your learning environment. You computed this process, this kind of process model - that will look like spaghetti. It has a lot of axis going from one node to each node. How to abstract this? The proceed model is creating the abstract model form the sequence data. How to abstract this?

For Fuzzy Miner, we apply three rules. That is the rules we are going to talk about for Fuzzy Miner. For other algorithms, rules might change, okay? The basic is significance and correlation. Let us look at the rules for Fuzzy Miner.

(Refer Slide Time: 10:46)




Highly significant nodes are preserved, okay? The nodes which are more significant are preserved. We will not remove those nodes. So, consider you have a node A, node B, node C, D.

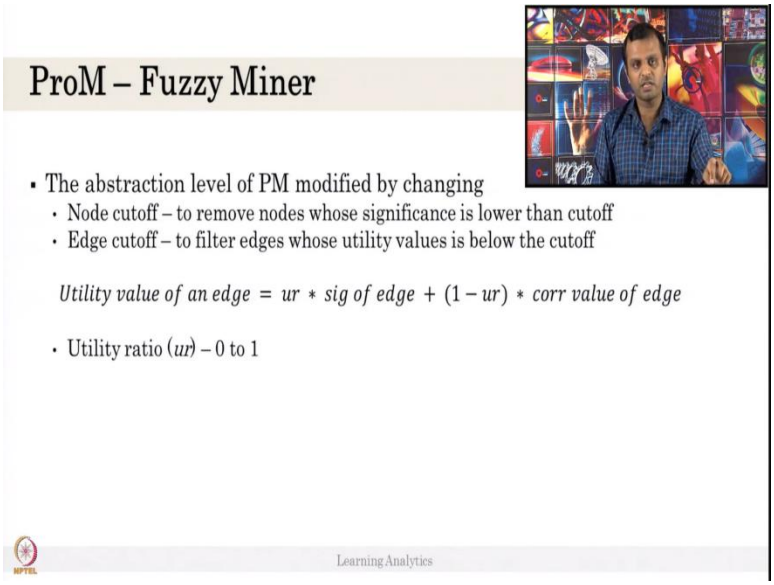
Highly significant nodes are reserved. Suppose consider the node significance of A, C point, 0.8, 0.7, 0.1, 0.3. Highly significance node that is A, B, C will be preserved. We may remove these two nodes. What will happen to these nodes, let us leave the second node. Less significant nodes that are highly correlated will be aggregated into clusters. Consider these two nodes are less correlated. Say this significance 0.1 and 0.3 - it is their significance, but have a very high correlation. For example, the relation between these two is say 0.2. So this is 0.2.

So, this particular node is less significant but it always occurs. If it occurs, less significant means it occurs few time, in a sense, how do you say - it occurs only three times compared to the A. Say A is 30, it will be 3 times. It occurs very less time but it always co-occur with node A. If that is the case, then we can combine these, aggregate these two, create a cluster. This node can be combined into a cluster 1 or something like that, that is what this particular thing says.

Unless significant nodes, the node which is less significant with low correlation with others are dropped. For example, this node is less significant. It has self-occurrence say 0.7, it has co-occurrence only 0.1 and from here, if it has a very less correlation with any of the node but it occurs very very few times - 9 times out of 30 actions of A occurring 30 times, this occurs only 9 times, then you can drop this particular node, okay? This node is not correlated with any significant node or this node is not significant at all. So it can be removed to reduce the complexity of the process model. That is the idea of Fuzzy Miner.

Hope you understood these three step - it is very important, you know a highly significant node preserved less significant and less correlated is dropped. Less significant and highly correlated nodes are combined to create a cluster. That is a very basic step in Fuzzy Miner.

(Refer Slide Time: 14:18)




So, to change the abstraction level, it is not that the system comes up with own model, you might say no no, there are some thing I really want. Sometimes significance is more important, sometimes correlation is important, then you can use some particular formula. We use two value that is node cut off and edge cut off. You can say all the nodes which are significant less than 0.4 should be removed, so can put a node cut off say 0.4.

So, node cut off we saw in the last slide, 0.3 is low significant, we do not know. But you can define the threshold to say it is 0.3 is less than 0.5, my threshold is 0.5, whichever node which has significance less than 0.5 should be removed. So, we can put the node cut off as 0.5. An edge cut off to filter out the edges which utility value is below the cut off. What is utility value? Utility value is a combination of significance of edge also the correlation value of the edge. I said that - edge also can have a significant value.

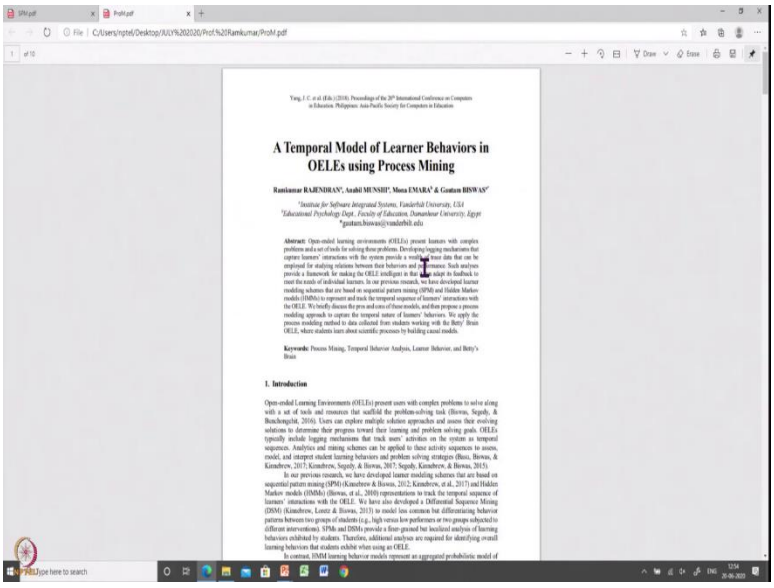
So that as a combination, weighted combination of significance of edge and the correlation edge the weight you are depends on you. If you want to have mode weightage to the significance, keep 1. If you want less weightage to significance, then keep 0.

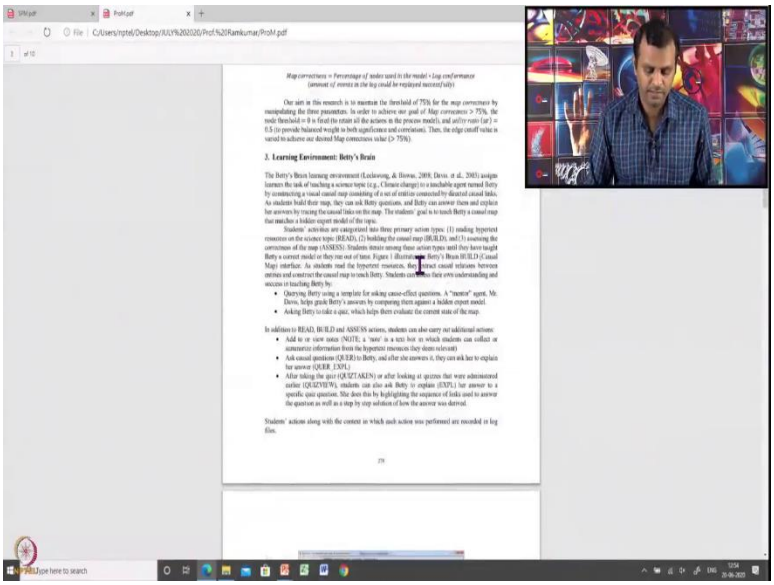
Applying this particular formula, you can modify the process model. The Fuzzy Miner may try to give you the abstract. It is not that any generic model can be applicable for you. Then you can say, no I want to keep all the nodes, I do not want to remove any nodes but I may remove some edges which has less correlation value. Then if it is less correlation value, then put “ur” equal to

1, and the correlation goes off only the significance edges are stayed back. This changes the complexity of the process model.

Please look at this particular equation. It is very simple, just you have to understand how the abstraction is happening using node cut off and edge cut off only two things. In the process model software, you can actually vary this node cut off and edge cut off in the software. I will show the demo actually, you will understand this thing.

(Refer Slide Time: 16:53)


SRM.pdf
ProM.pdf
File | C:/Users/nptel/Desktop/JULY%202020/Prof.%20Ramkumar/ProM.pdf
1 of 10
Map correctness = Percentage of nodes used in the model + Log conformance
(amount of events in the log could be replayed success fully)
Our aim in this research is to maintain the threshold of 75% for the map correctness by manipulating the three parameters. In order to achieve our goal of Map correctness > 75%, the node threshold = 0 is fixed (to retain all the actions in the process model), and all/tr ratio (arc) = 0.5 (to provide balanced weight to both significance and correctness). Then, the edge cutoff value is varied to achieve our desired Map correctness value (> 75%).
3. Learning Environment: Betty's Brain
The Betty's Brain learning environment (Leelawong, & Biswas, 2008; Davis, et al., 2003) assigns learners the task of teaching a science topic (e.g., Climate change) to a teachable agent named Betty by constructing a visual causal map consisting of a set of entities connected by directed causal links. As students build their map, they can ask Betty questions, and Betty can answer them and explain her answers by tracing the causal links on the map. The students' goal is to teach Betty a causal map that matches a hidden expert model of the topic.
Students' activities are categorized into three primary action types: (1) reading hypertext resources on the science topic (READ), (2) building the causal map (BUILD), and (3) assessing the correctness of the map (ASSESS). Students iterate among these action types until they have taught Betty a correct model or they run out of time. Figure 1 illustrates the Betty's Brain BUILD (Causal Map) interface. As students read the hypertext resources, they construct causal relations between entities and construct the causal map to teach Betty. Students can express their own understanding and success in teaching Betty by:
Querying Betty using a template for asking cause-effect questions. A "mentor" agent, Mr. Davis, helps grade Betty's answers by comparing them against a hidden expert model.
Asking Betty to take a quiz, which helps them evaluate the current state of the map.
In addition to READ, BUILD and ASSESS actions, students can also carry out additional actions:
Add to or view notes (NOTE; a 'note' is a text box in which students can collect or summarize information from the hypertext resources they deem relevant)
Ask causal questions (QUERY) to Betty, and after she answers it, they can ask her to explain her answer (QUER_EXPL)
After taking the quiz (QUIZTAKEN) or after looking at quizzes that were administered earlier (QUIZVIEW), students can also ask Betty to explain (EXPL) her answer to a specific quiz question. She does this by highlighting the sequence of links used to answer the question as well as a step by step solution of how the answer was derived.
Students' actions along with the content in which each action was performed are recorded in log files.
Type here to search
ENG
12:54
21-04-2020

the question as well as a step by step solution of how the answer was derived.

Students' actions along with the context in which each action was performed are recorded in log files.

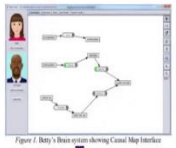


*Figure 1.* Betty's Brain system showing Causal Map Interface

Student performance in the Betty's Brain environment is measured by their current "map score", which is computed as the difference between the number of correct and incorrect links present in the student's map at any point of time. Students' learning behaviors in Betty's Brain are derived from a cognitive metacognitive task model (Kinnebrew, Segedy, & Biswas, 2017). Their interactions with the system are mapped to particular skills (for example, reading hypertext resources is mapped to the information acquisition skill), which are then interpreted in terms of the overall learning objectives. A sequential combination of primitive skills, performed in a context, is interpreted as a problem solving strategy. Researchers have employed a combination of analytics methods (Kinnebrew, Segedy, & Biswas, 2017) and exploratory sequence mining techniques for detecting and characterizing students' metacognitive processes in Betty's Brain environment. Betty's Brain has been shown to significantly improve student learning, as measured by gains observed from pre- to post-tests (Kinnebrew, et al., 2017, Segedy, et al., 2015).

For analyzing learners' behaviors, their logged actions are further classified based on time taken to perform each action and/or its impact on the current map score. For example, READ actions are characterized as READ-SHRT (3 seconds or less spent on reading a page) and READ-LONG (more than 3 seconds spent on a page). A linkedit (adding, modifying or deleting a causal link in the Causal Map Interface) action that increases the map score is logged with the suffix –EFF (for 'effective'), and a linkedit action that decreases the map score is logged with the suffix –INEFF (for 'ineffective'). A second qualifier for link edit actions is the suffix –SUP/UNSUP, which indicates that editing action performed is supported by information acquired from a previous READ or QUIZ action that occurred within the last 5 minutes (Segedy, et al., 2013). The descriptions of the actions



completed a post-test that was identical to the pre-test. The pre- and post-tests tested the students' knowledge on the science domain concepts and understanding of causal relations using a combination of multiple choice (*maximum score* = 7) and short-answer (*maximum score* = 9) questions.

Table 1

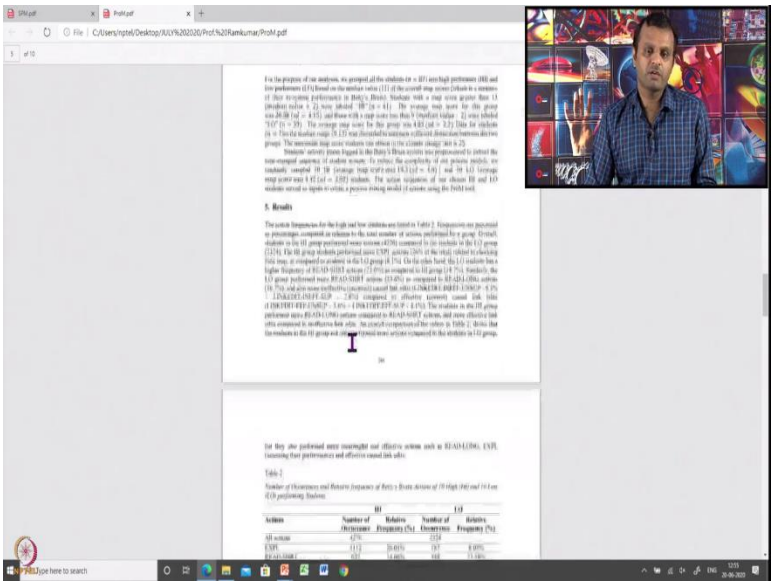
*List of Betty's Brain Actions with descriptions used for deriving the Process Model*

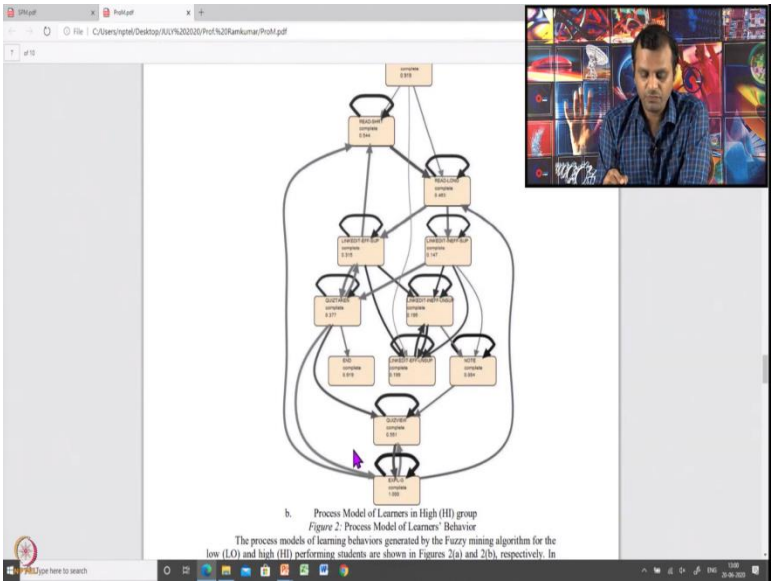
| Primary Action Category | Actions | Description |
|---|---|---|
| READ | *READ-SHRT* | Reading resources in science page for < 3 seconds |
| | *READ-LONG* | Reading resources in science page for > 3 seconds |
| BUILD | *LINKEDIT-EFF-SUP* | Add or edit a causal link that is correct and supported by previous READ or QUIZ actions |
| | *LINKEDIT-INEFF-SUP* | Add or edit a causal link that is incorrect and supported by previous READ or QUIZ actions |
| | *LINKEDIT-EFF-UNSUP* | Add or edit a causal link that is correct but not supported by previous READ or QUIZ actions |
| | *LINKEDIT-INEFF-UNSUP* | Add or edit a causal link that is incorrect and not supported by previous READ or QUIZ actions |
| ASSESS | *QUIZTAKEN* | Betty takes a quiz when asked to do so. |
| | *QUIZVIEW* | Viewing results of quizzes (either the current one or those that were taken earlier) |
| | *EXPL* | Asking Betty to explain her answer to a quiz question, and viewing the highlighted map that she generates |
| | *NOTE* | Adding or viewing notes in the note taking interface |

*4.2. Data Preprocessing and Process Mining Analysis*

For the purpose of our analysis, we grouped all the students (n = 87) into high performers (HI) and low performers (LO) based on the median value (11) of the overall map scores (which is a measure of their in-system performance in Betty's Brain). Students with a map score greater than 13 (median value + 2) were labeled "HI" (n = 41). The average map score for this group was 20.58 (sd = 4.15), and those with a map score less than 9 (median value – 2) were labeled "LO" (n = 39). The average map score for this group was 4.03 (sd = 2.2). Data for students (n = 7) in the median range (9,13) was discarded to maintain sufficient distinction between the two groups. The maximum map score students can obtain in the climate change unit is 25.

Students' activity traces logged in the Betty's Brain system was preprocessed to extract the time-stamped sequence of student actions. To reduce the complexity of our process models, we randomly sampled 10 HI (average *map score was* 19.3 (sd = 4.8) ) and 10 LO (average






So, let us look at an example process model applied for one such paper. So, in this paper, we apply process mining on learner behaviour from one of the OELE that OELE is again is Betty's Brain, we saw in SPM and DSM, this is Betty's Brain and its set of actions. Let's consider the actions here are a bit different from last paper. The action says Read short, Read long. Reading a list of science for less than 3 seconds, I will consider as Read short because in 3 seconds, you might be looking for 1 or 2 words, not really reading.

More than 3 seconds is looking is called Read Long. But also Read Long was upper 100 thresholds. If you are reading more than say 1 minute, it is not actually reading anything, it is just kept the slide open and it went somewhere off. So, we are thinking cut off between Read short,

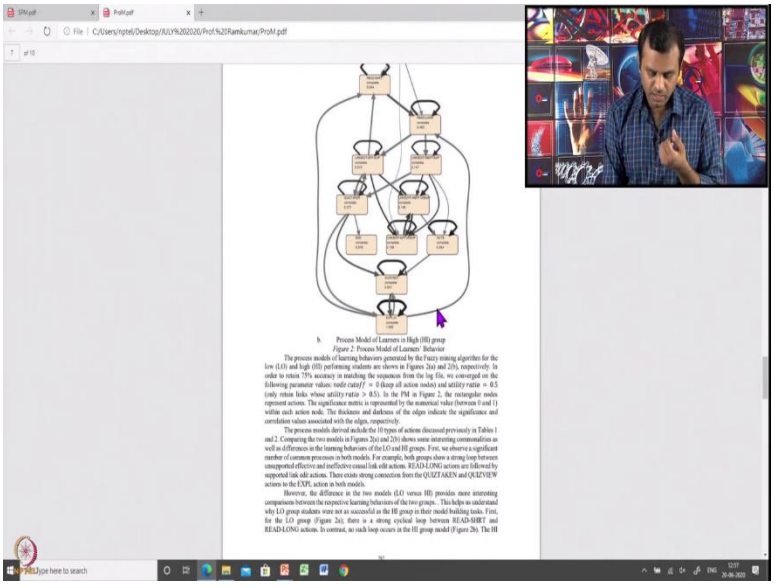
Read long. Then we had a link edit support, link edit ineffective, effective link edit. We added a link, this is correct, effective, if it is incorrect, it is ineffective that is what we try to say.

This paper link is there, and you can download and read if you want more information. So what we did, we computed this, and we created a process model, this is the simple frequency between a high scorer versus a low scorer. High scorer, low scorer as I mentioned in last video, based on the pre and post test score, we grouped the students into high scoring and low scoring, we tried to see the process for these two groups and see if there any difference between the process.

If you look at here, there is a distribution of each action for high versus low is given here. And that is, that high scorer has then a lot of actions as compared to the low scorer also by percentage, it is more for some actions. For example, read short is more for low but Read short is not that much 14 per cent for high. So that is also given. So, you can use this table to identify the descriptive and you can plot it in a graph and see it.

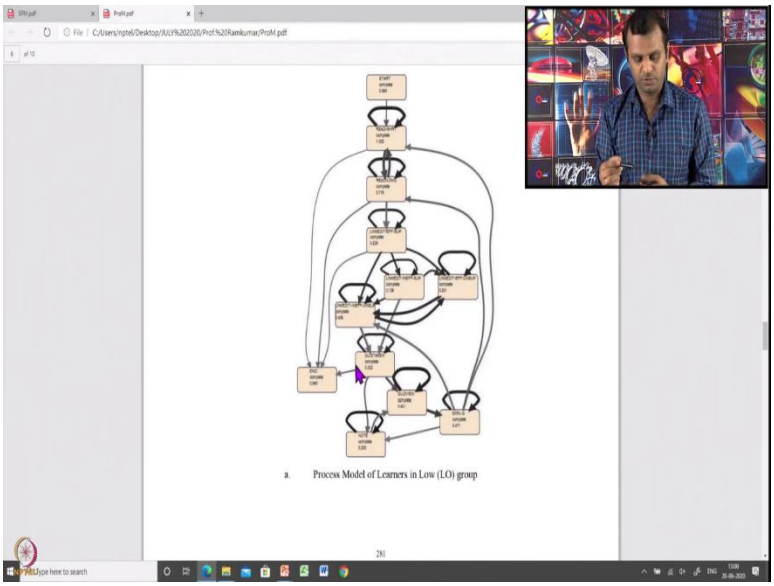
But let us look at the interesting part that is process model. So, in process model, we can add a artificial start and artificial end. So, start is here and end is here, okay? And this color, you know the thickness, the thickness of the line, this is very thin, this is kind of thick and the colors- the dark or light gray indicate significance and correlation. So, let me share what is that.

(Refer Slide Time: 19:16)



So, if you look at this, so the thickness and darkness of the edges indicate the significance and the correlation values associated with the edges. For each edge, if it is thick, okay it is thick - which means it is highly significant - this particular edge is significant. If it is dark, which means it is highly correlated with these two values.

(Refer Slide Time: 19:44)




So, let us look at the process model. Let's understand one process. So, we did one simple thing, we wanted to maintain the process model have utility value was more than 0.8 or something. So, we can give this report to understand because we adjusted. We do not want to lose any nodes so we kept all the nodes, node significance is absolutely 0.1 or 0.2, just keep all the nodes and we adjusted the edge cut off value that is utility ratio value to keep 80 percent of the sequences in the data which is 0.8 means if I want to reproduce sequence of actions from this process model, I should be able to reproduce 80 percent of original sequence data.

So I do not want to abstract too much, so that I lose all the value so when you do that, do it for a very few number of students, pick some random 10 students in one group and do it because if you keep all 30 students in a low group to create process model it will become a spaghetti and you lose a lot of information. So, decide based on what you want to do, based on your such question.

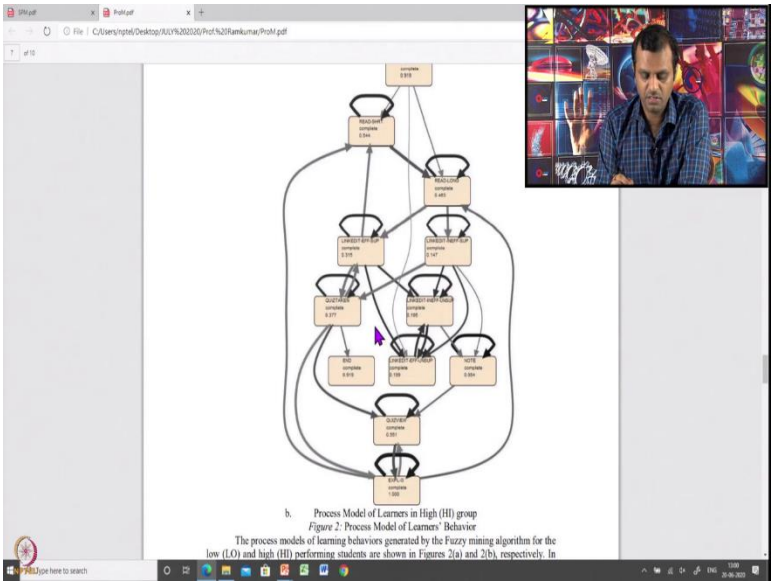
So, what happened is, the people in low group, they started reading and they are doing a lot of short Read that's why the correlation, write correlation here and after reading they go to Read

long, and there is a loop here, they read long, read short. Once they pass this particular loop of Read long, read short, if they are read long, they might go to edit the link in Betty's Brain after reading you can create a concept map - editing a link. After editing the concept map, they can remove some of the edits of add edits, they can keep on doing edits - there are three-four type of edit actions. It can be combined. There is a relation between these edit actions also. After doing that, edit actions from all these actions, they go and take the quiz.

So, they read, they create a concept map, then they take the quiz. After taking quiz, they check the quiz answer or they take notes, they check whether things are right or wrong or after giving the quiz, they read the explanation, they go back to read or they compute here, again take notes, go to quiz view or they do something.

So, this is the set of process of actions a student can do in the Betty's Brain. It is a set of actions the low performing group students. For example they read, read, edit, immediately take quiz, quiz view, explanation, ask for explanation, if it is wrong, go back read again, again come back this and they might end from quiz taken or from this. And this way less significant and less correlation. Just to say what are the nodes were the ending nodes for these groups.

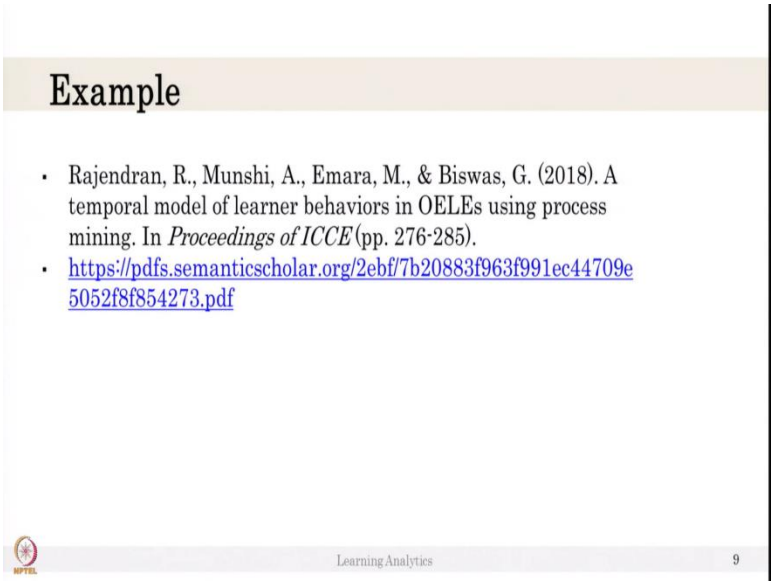
(Refer Slide Time: 23:31)




Consider the high group, they start with the same similar Read short, read long, they also do the link edit, link support most of the editing actions. They take quiz and they immediately go and look at the explanations, they also into take quiz very often. So, after that, they immediately go

to read short or they actually don't care about the quiz view, they take quiz and read explanation, they go back to read long. This set of extra things, this is not available in the read low group. So there are few significant difference, the behavioral difference that can be identified from the low group as a side group.

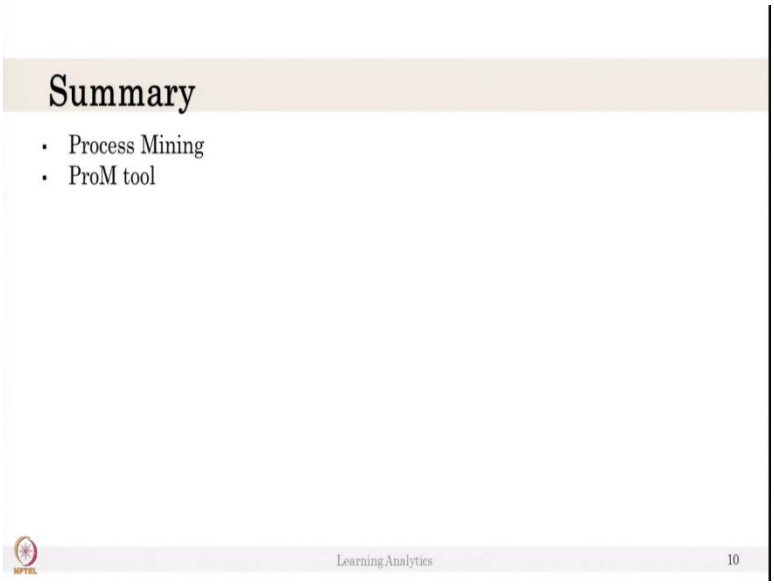
(Refer Slide Time: 23:13)




This paper is to show the process mining or the abstract difference between low group and high group. It is not to say, that high group is doing good in these actions or low group is doing good in other actions. And also here, we can compare the process model with the sequential pattern mining model. In a process model, we consider the frequency, how many times it has occurrence. But not the time they spent on each actions.

But when you use process model tool you have a option to put the time taken on each action also in time. If you give that, you can create a better model, Heuristic Miner model to look at it. I would request recommend you to go and explore the other type of models like not just Fuzzy Miner, other process mining models and see what is the difference and explore and understand, okay? So, this paper, you read it. This paper will be part of your assignments, the assignments questions means which might be part of your end exam too.

(Refer Slide Time: 24:17)




So, in this week we saw what process mining is and you might get the process mining tool demo in this week. So in this week there are two demos, three different algorithms, process mining, SPM and DSM. Let us continue talking about diagnostic analytics in one more week. Thank you