Learning Analytics Tools

Professor Ramkumar Rajendran

Educational Technology

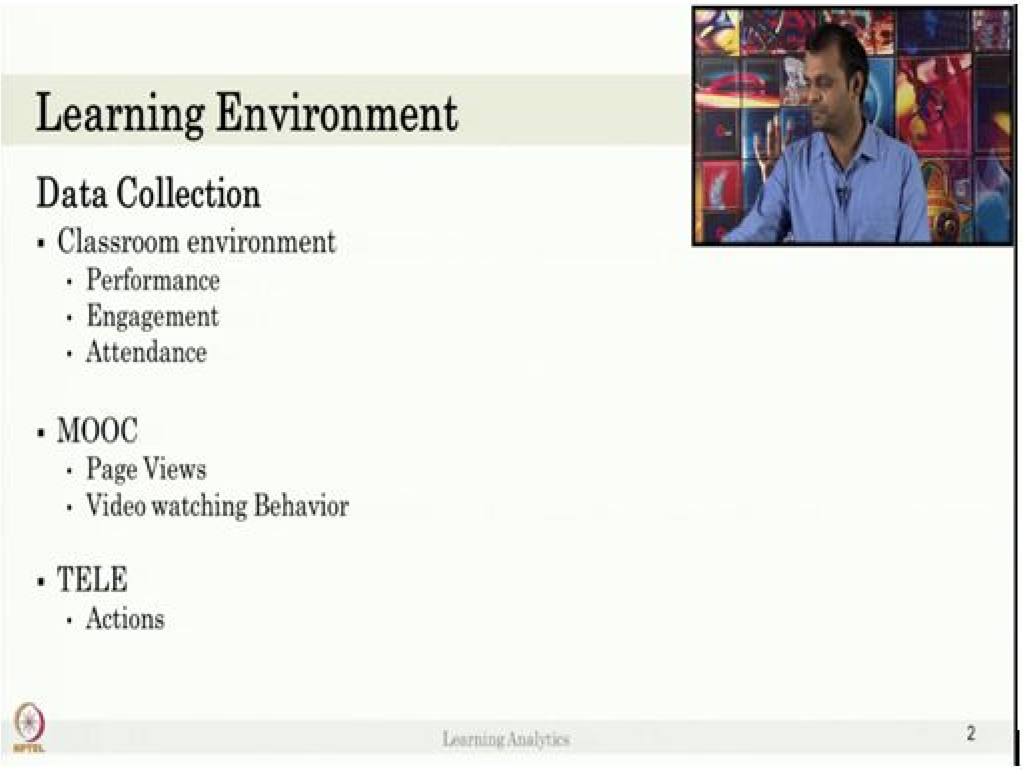
Indian Institute of Technology, Bombay

Lecture – 2.3

Data Preprocessing

Welcome to Learning Analytics Tools course. In the last video, we saw how to collect data from the open-ended learning environment. There, the data collected is clickstream data, we collected too much data like we collect every click of the students. So, we call it as raw data.

(Refer Slide Time: 00:36)



Learning Environment

Data Collection

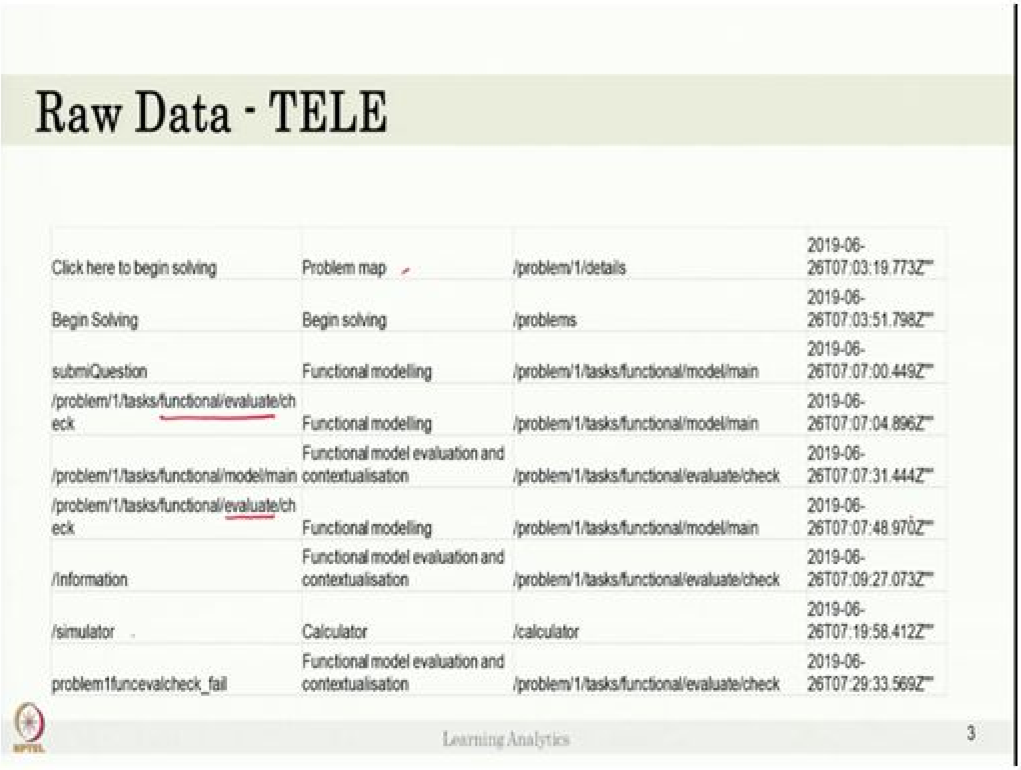
- Classroom environment
  - Performance
  - Engagement
  - Attendance
- MOOC
  - Page Views
  - Video watching Behavior
- TELE
  - Actions

Learning Analytics 2

So we saw that data we collect from a classroom environment, MOOC or TELE, in classroom environment we collected performance, engagement, attendance, that data we can collect, it is not in a raw format; already the data is in kind of a feature format, you can create new features from the data.

Or in the MOOC, we talked about looking at the page views, video watching behaviour, that is still again a raw data format. We have to extract features or construct feature from that raw data. In TELE also we showed that there is a lot of clickstream or user interaction data can be collected. How do we create features from this kind of raw data? We will look at that in this video.

(Refer Slide Time: 01:12)



Raw Data - TELE

Click here to begin solving	Problem map	/problem/1/details	2019-06-26T07:03:19.773Z"
Begin Solving	Begin solving	/problems	2019-06-26T07:03:51.798Z"
submiQuestion	Functional modelling	/problem/1/tasks/functional/model/main	2019-06-26T07:07:00.449Z"
/problem/1/tasks/functional/evaluate/check	Functional modelling	/problem/1/tasks/functional/model/main	2019-06-26T07:07:04.896Z"
/problem/1/tasks/functional/model/main	Functional model evaluation and contextualisation	/problem/1/tasks/functional/evaluate/check	2019-06-26T07:07:31.444Z"
/problem/1/tasks/functional/evaluate/check	Functional modelling	/problem/1/tasks/functional/model/main	2019-06-26T07:07:48.970Z"
/Information	Functional model evaluation and contextualisation	/problem/1/tasks/functional/evaluate/check	2019-06-26T07:09:27.073Z"
/simulator	Calculator	/calculator	2019-06-26T07:19:58.412Z"
problem1funcevalcheck_fail	Functional model evaluation and contextualisation	/problem/1/tasks/functional/evaluate/check	2019-06-26T07:29:33.569Z"

Learning Analytics

3

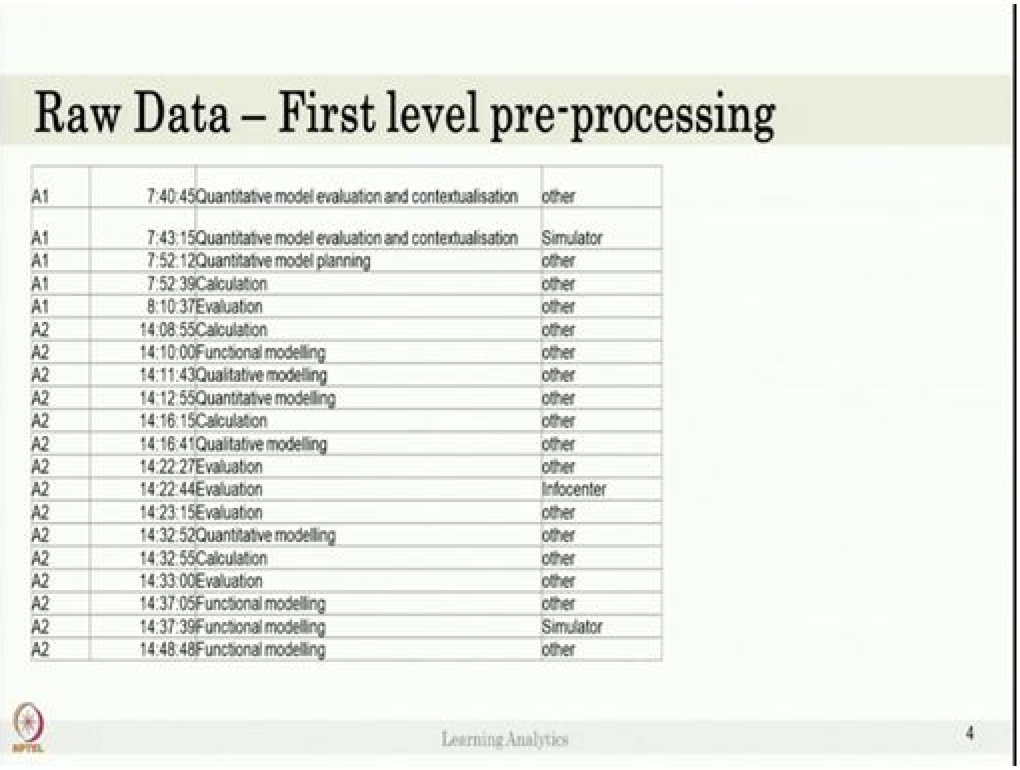
So, this is the raw data we collected from “MettLe”. Here is the raw data. The first action is, it is the user-readable action like this user started solving the problem, then he started solving the problem, then he is submitting some questions.

Then he is in a functional model is checking something, after checking the functional model the student moved to the main model, that is, the main problem page. After that, he went to the evaluate sub-problem, then event information, simulator, the student is moving towards different sub-task in the main problem page.

What is he doing in that particular problem page? For example, this(context) tells you that the student is in problem map, he began solving the problem using a functional model. So, where the student is currently- can be captured here.

Then the student's detail like what is he doing in that particular page can be given from this task, like is checking or is doing the calculation in the simulator or he is checking the other actions. So, the student's detailed action can be captured here and also with the timestamp. This data is obtained directly from MongoDB by converting to CSV. This kind of data we can get it if we write a logging mechanism(how to log a student's interaction from a learning environment).

(Refer Slide Time: 02:46)



Raw Data – First level pre-processing

A1	7:40:45	Quantitative model evaluation and contextualisation	other
A1	7:43:15	Quantitative model evaluation and contextualisation	Simulator
A1	7:52:12	Quantitative model planning	other
A1	7:52:39	Calculation	other
A1	8:10:37	Evaluation	other
A2	14:08:55	Calculation	other
A2	14:10:00	Functional modelling	other
A2	14:11:43	Qualitative modelling	other
A2	14:12:55	Quantitative modelling	other
A2	14:16:15	Calculation	other
A2	14:16:41	Qualitative modelling	other
A2	14:22:27	Evaluation	other
A2	14:22:44	Evaluation	Infocenter
A2	14:23:15	Evaluation	other
A2	14:32:52	Quantitative modelling	other
A2	14:32:55	Calculation	other
A2	14:33:00	Evaluation	other
A2	14:37:05	Functional modelling	other
A2	14:37:39	Functional modelling	Simulator
A2	14:48:48	Functional modelling	other

But I will show you the better data from the CSV(of MongoDB download). We can convert with very minimal effect, assign student ID A1, A2 or some student ID. The timestamp can be converted to the readable format instead of UTC timestamp, Unix timestamp we saw in the last table, we can convert into a timestamp in a different format. Then we can make readable actions like the student is in functional modelling.

So I went to the first level of action I do not want to go to second level, like in functional modelling what is the student doing? That depends on your research question. Then you can say what is he doing in that model, he is doing something else, is at info centre, simulator(some contextual information). You can add another column to add more contextual information. It depends on the research question we are asking.

So my suggestion is, capture all the data in a raw format, then create/construct your own features based on what is your research question and what is your aim and how do you want to construct the features. So this is the data we can obtain from the MongoDB with very little effort.

(Refer Slide Time: 04:03)

Activity:

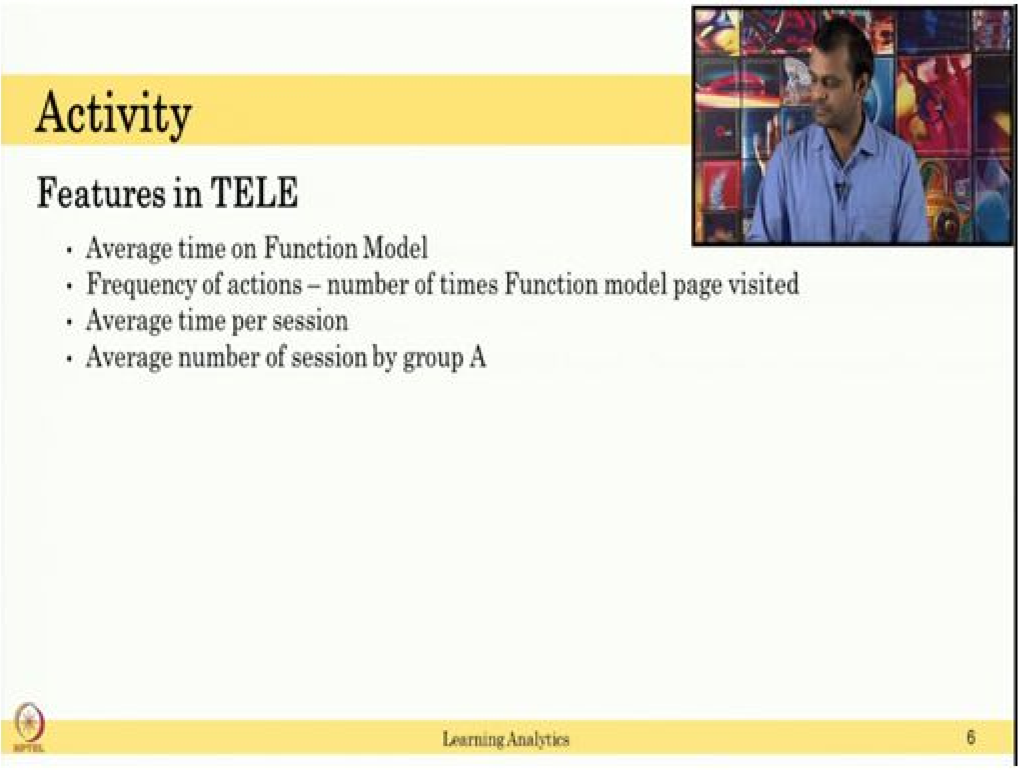
Data Collection

- Given the raw data, list down at least three features

Learning Analytics 5

So let us think about it. So, we have this data, let us do a small activity. Given the raw data you saw in the last two slides, can you list on at least three features? Please pause this video and write down your answers and resume to continue.

(Refer Slide Time: 04:17)



Activity

Features in TELE

- Average time on Function Model
- Frequency of actions – number of times Function model page visited
- Average time per session
- Average number of session by group A

Learning Analytics 6

So the features in TELE can be average time on the functional model. For example, a student will be using this material for multiple sessions and he might be moving around the functional model, qualitative model, quantitative model multiple times. What is the average time students spent on the functional model? That can be a feature.

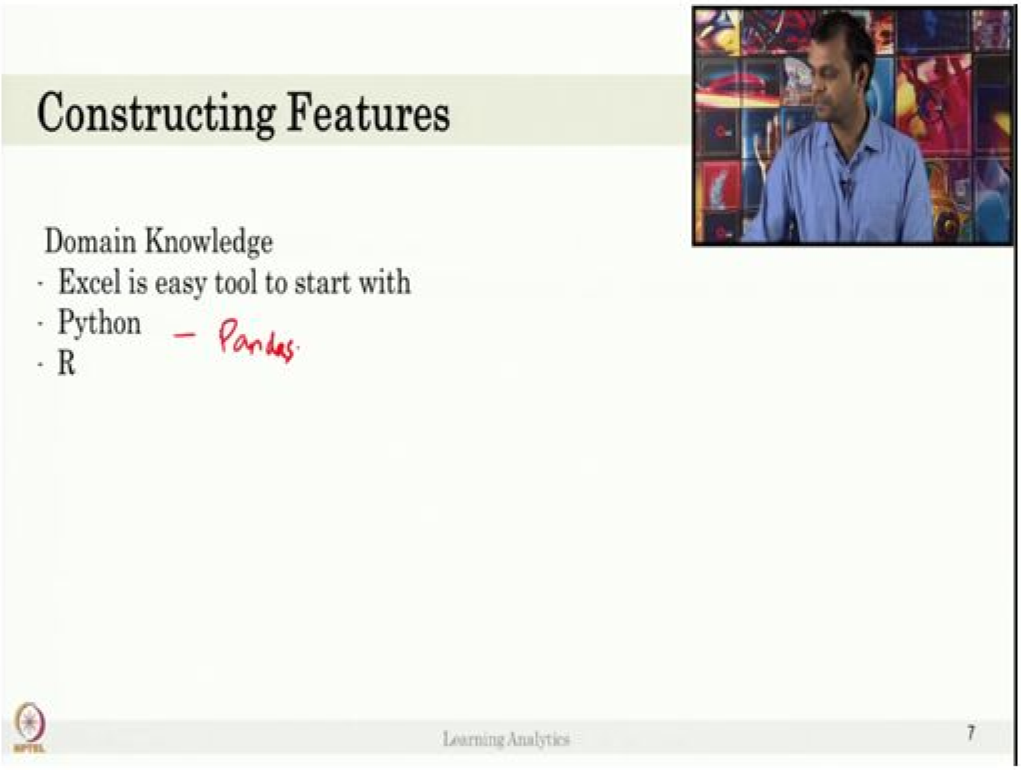
So, how do you compute it? You have to use Excel sheet, simple excel sheet tools tricks, you have to define what is the student ID, for a given student ID and you can compute when and all students use the functional model and you have to use the timestamp data to compute the average. So, you might know how to use the Excel as if you have watched the Week 2 video.

Now, can you consider the frequency of actions? So, the first one is time, the second one is the frequency of actions. A number of times a student use a functional model in a session or number of times a student used a functional model in a week. So you can come up with a different set of features.

So there are two basic common features we should consider - time and frequency. Like functional model you can create features for a quantitative model, information centre or calculator, simulator, you can create a lot of features from the log data. Also, you can compute features like average time per session, what is the average time a student is spending per session and is it different in Group A and Group B?

Suppose you have two set of groups. One is control group or experimental group and you might be introducing a new way intervention in the OLE for Group B, then you want to see that whether this recommendation or intervention improves student's average time spent on the OLE. So you can compute these kinds of features from the log data in a raw format.

(Refer Slide Time: 06:22)



Constructing Features

Domain Knowledge

- Excel is easy tool to start with
- Python — Pandas.
- R

Learning Analytics

7

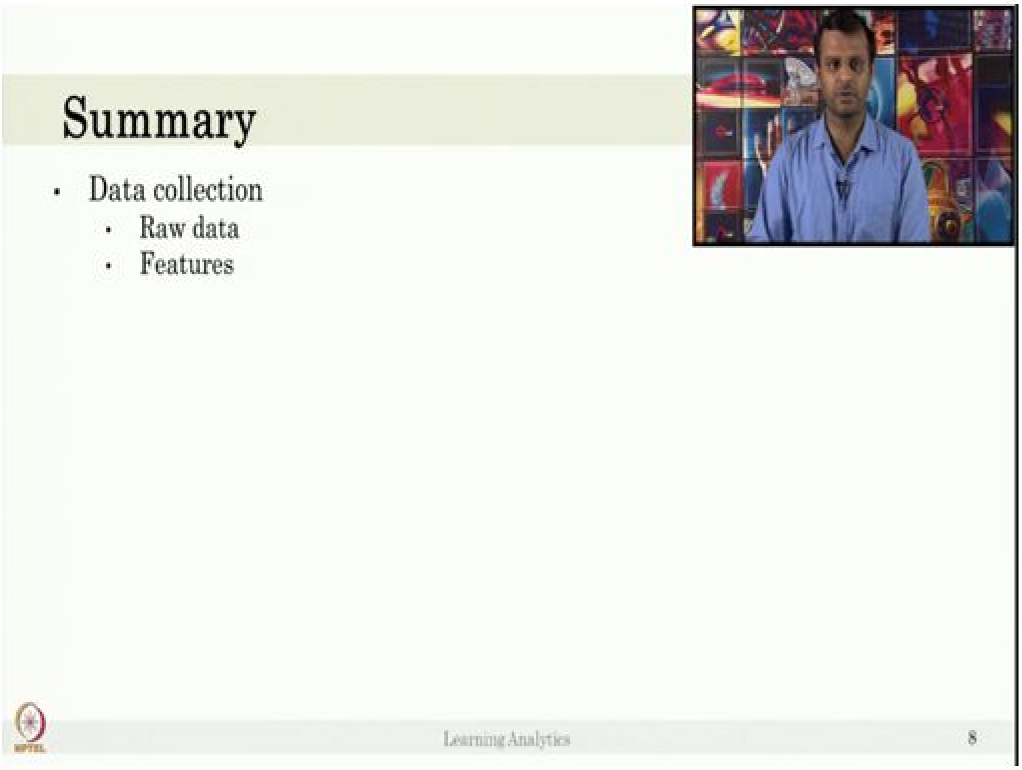
So how do you construct features? The main thing is domain knowledge. So that is why the domain expertise is more important to construct these features, the domain expertise in the problem you are handling and also you might know by your experience a student might do this actions in order to complete task Y. So that kind of domain knowledge will be helpful to create that features.

So, please read research papers which talks about features, feature construction, from that you can start your feature construction. The basic thing as I mentioned is time and frequency. Excel is an easy tool to start with, it does not require much coding or any knowledge of programming. In this week we will teach how to use the Excel sheet as a learning analytics tool.

Also, the Python or R script is really good. So I recommend even if you have not worked on any programming language, I recommend you to start learning one of these script languages, especially in Python, please use Pandas.

Especially in Python, use the library called Pandas, that helps you to process the raw data from the CSV or in a matrix format; once you have the raw data logged into the Python, you can do a lot of actions just like we do in Excel sheet. Pandas library is really good and I recommend you learning Python but if you watch the demo of how to use Excel for the LA, that is enough.

(Refer Slide Time: 07:57)



Summary

- Data collection
  - Raw data
  - Features

Learning Analytics 8

So in this video, we talked about what is raw data, raw data from MongoDB, example for matter and also how to construct features, the basic feature construction is please consider time and frequency of each action, create features from that. Thank you.