EXperimental Learning

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Big Data and Social Analytics certificate course

MODULE 5 UNIT 1
Video 2 Transcript

MIT BDA Module 5 Unit 1 Video 2 Transcript

Speaker key

YM: Yves-Alexandre de Montjoye

HY: Hapyak

YM: Hi, I'm Yves-Alexandre. I hope you've enjoyed this course so far. This lesson is going to be about bandicoot. A tool we developed at the MIT Media Lab to allow you to derive insights for mobile phone metadata.

What you learn in this lesson is however applicable to any kind of big data. Internet logs, purchases data and others. A lot of you are probably used to work with data that looks like this. You know, nicely tabulated customer data in Excel. One line per customer, including the usual demographics, age, gender, zip code. The demographics you use to do customer segmentation, market analysis, or design pricing strategies.

But that's not what big data is. A large fraction of what we call big data, are actually metadata recording how people behave. For example, the data generated by our mobile phones; the data recorded every time we search for something online; the data recorded every time we swipe our credit or loyalty card. And this data is a lot richer, but also a lot messier than the usual Excel-type data.

00:01:15

And this is why we developed bandicoot to help data scientists use and make sense of large-scale mobile phone data. Bandicoot focuses on mobile phone metadata. The same data that is collected by virtually any carrier in the world.

Before we look at bandicoot, let's look at what this data looks like. One line is generated into database every time someone makes a phone call or receives a text. And that line contains who calls, or text, who, at what time and from where. Furthermore if this is a phone call, you also have the duration of the phone call.

In the United States, billions of these record are generated every year. So the full thing bandicoot allows you to do is to explore new data. You can visualize what a user's social network looks like. How much he called, how much he texts a certain person. You can also look at how his behavior changed over time. Visualization is an essential part of data science. It give you a sense of what's in the data and often it's the best way to spot potential issues in your data.

And once we're sure that the data is okay, we can start playing with it. At its core, what bandicoot allows you to do is to extract information on how a user behaves from this metadata.

HY: How often do you use visualization when working with data?

a. Always
b. Sometimes
c. Never

Thank you

YM: For that we've designed algorithm that will take the data and compute more than a hundred behavioral indicators for you. These behavioral indicators range from simple ones, like how much one calls or texts, to more complex ones. For instance, bandicoot takes a look at how much time a user who just received a text to actually answer it.

00:02:56

Let's imagine you just receive a text so what bandicoot is going to do is basically look at how much time it takes you on average to actually answer that text and send a text back.

Actually, under the hood, bandicoot does a bit more like that. It computes all the behavioral indicators, but also the way they might vary across time. What bandicoot does is it divides user records by week, compute the indicators each week, for example the mean duration of calls that week, and then returns to you the mean and the standard deviation of that behavioral indicator across weeks.

Okay, now why don't we look at a quick example? Now let's imagine that we have the duration of the calls for one user across three weeks. What bandicoot does is it will compute the mean duration of the calls for that user on week one, week two and week three. And then it will return to you the mean of these calls and the standard deviation of these calls. Once bandicoot has computed all the indicators for you, which usually takes less than a minute, you can then export them into two different formats, .json or .csv.

And if you don't have data to play with, either bandicoot provides you test data, or you can download the bandicoot app

HY: Bandicoot App

YM: on any android device and with this app you can export your own mobile phone metadata; email the file to yourself; open the file using the bandicoot software package on your laptop and then visualize and analyze how you behave.

Bandicoot is already used by numerous organizations: Orange, Telenor, the World Bank and others. I hope this video gives you a good understanding of what bandicoot is and what it does. In the next video we'll see how to use these behavioral indicators in different ways.