EXperimental Learning

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Big Data and Social Analytics certificate course

MODULE 2 UNIT 1
Video 2 Transcript

MIT BDA Module 2 Unit 1 Video 2 Transcript

Speaker key

AP: Alex Pentland

HY: Hapyak

AP: OK, well let's talk about sources of data, so in the last module, we talked about this design exercise I did 20 years ago and I mentioned that the surprising thing was the amount of data people put off in day to day. So there's the browsing and email you do on your phone; there's who calls you; who you call; do you text them or call them; where you are; how you move; where you go next; what you spend; where you spend it; health data; who else is around you. So we can talk about all these types of data that everybody's putting off and what we find is there's a lot of rich relationships between them and the relationships give us greater insight into human behavior.

They're not all created equal, though. For instance, browsing, email and apps tend to be things that are your face to the public, so what you say on social media is really what you want people to see more than what you really actually believe. You may say you want to do this but you may not actually do that thing.

00:01:16

Browsing tends to be a little bit more private. Apps tend to be more, well, keeping up with the Jones's and what the other guys do at work. Another type of data which is more powerful is where you are, how you move, where you spend and what you spend. Those are types of data that are really your commitment. You may not admit them but they really are the thing that defines how your life is and, along with that of course, comes health data, which is another thing you can't control and things like what you eat, what you buy, what you do, where you go, have rich relationships with health and help you interpret health data and look at risks for various sorts of diseases, for instance.

A final type of thing is the context you're in, the social context, so who calls you, who do you call, who do you message, those talk about your human relationships, and remember – most of what you do is a function of the other people and what they're doing. It's not your internal processes. Similarly, who else is around? Who else do you see doing things? Who else do you move to spend time with and just in general, as I said earlier, social is generally more powerful than an individual. That's remarkable because most data analytics to date have been about individuals. Now that we begin to have network data, we can get much more powerful by looking at social context and what the people around you do.

HY: Data that is generated has numerous uses. GPS data for example helps the user find their desired location, but by aggregating the data from many users it can also be used by city planners to map transport systems. What other uses can you think of for GPS data?

It can map person-to-person interactions and thus predict the spread of disease and best areas to vaccinate. It can help with planning how to maintain or expand the electric grid. Aggregate patterns of interaction also indicate poverty rates within neighborhoods.