Stanford eCorner

Using Math to Reveal the Shape of Data

Gunnar Carlsson, *Ayasdi*

February 05, 2014

Video URL: http://ecorner.stanford.edu/videos/3268/Using-Math-to-Reveal-the-Shape-of-Data

Ayasdi Co-Founder Gunnar Carlsson explains the concept of topology in mathematics and it's value as a method for understanding data through pattern recognition.

Transcript

So topology is the part of mathematics that deals with trying to describe and represent a shape. In fact it's actually a form of pattern recognition. So it was started in the 1700s by a Swiss mathematician named Leonard Euler. And it's actually thrived on the pure math side. What it does is it actually introduces in variants that allow you to measure shape. So measuring shape is sort of a, that's a funny sounding concept. Because shape to me is kind of a, it's an ill formed or a kind of vague notion. And so the idea of measuring it with numbers is a little counter intuitive. But it turns out that there are ways of doing that, actually very, very interesting and powerful ways of doing that. The second part of it though is, and I don't think it's usually talked about this way, but topology is about compressing shape.

It is about finding, if you think of a circle for example, a circle is infinitely many points and infinitely many pairwise distances between those points. Now if you're willing to sacrifice a little bit of detail, like the exact nature - curvature and so on, you can represent it by say a hexagon or an octagon which is say eight nodes and eight edges, which can be represented in a single byte. I mean so it's very very simple. And the notion of trying to sort of combinatorially compass the notion of shape into something much more understandable is the second thing that topology is about. And so in this area, what happens starting in about the year 2000, a lot of people started to have this idea that we should say these techniques in topology, which are about pattern recognition and representing shapes and sort of getting really precise about what it means by shapes, should now be transported from the pure math world where you're dealing with things where you have complete information. You have all the points or, or a description in terms of equations, to something where you only have sampled information, which is really more like real life. And so that's what's been going on in the last 15 years, porting all those techniques from understanding shapes, and I mean here even higher dimensional notions of shape, not just two and three dimensions. Porting them into what we would call the point cloud world which is, that is where data lives.