

The Mathematics of Complex Streamed Data

Terry Lyons

Multimodal streamed data is essentially different to unimodal streamed data. Consider this:

‘A commuter arrives at a bus stop before the bus’ – they catch it;
‘The bus arrives first’ – they miss it.

These are the same two events, but the order changes everything. Yet most models treat these as identical: ‘A bus and a person arrived’ They ignore timing and relationships.

This simplification isn’t harmless. A timed series gives no information about order within sampling intervals. As a result, the sampling rate has to come from the bottom up if it is to preserve this order information. Rough path theory makes a radical change and describes the stream over an interval using a group element. According to the choice of group it is possible to capture order information and to allow a top down description of the data stream without using essential information about the order of events.

This approach to describing streamed data is important to data science because it reduces the dimension needed for descriptive feature sets and so reduces the size of the data set needed to train. There are numerous prize winning illustrations of the methodology in use and the impact can be measured in the hundreds of millions of US dollars.