Discovering Knowledge from Consecutive Data Streams Using Unsupervised Learning

Prof Saman K. Halgamuge, FIEEE

Real-world data are often obtained consecutively or sequentially, e.g., information from sensors deployed in a bushfire-prone country, a gas sensor network to measure air quality deployed in a city or data acquired from an epidemiological study of a virus outbreak. Extracting information from data often cannot wait until the complete data set is available. The often-made assumption that the data available at the time of analysis is self-sufficient, i.e., it fully describes the underlying process or system being examined, may not be true. Thus, data analysis models must be capable of Incremental learning, i.e., to learn, interpret and build the model using the available data first and incorporate new knowledge as new data are acquired incrementally.

In such scenarios, learning algorithms that can find models –underlying structures or distinct patterns within data – without relying on labels (i.e. using Unsupervised Learning), have made great progress toward answering these sorts of questions. How well do the existing data analysis tools cope with learning tailored to take the advantage of consecutive data streams coming from sensors or lab experiments in biology, astronomy etc.? Do these data streams coming from different experiments or sensors observing the same phenomenon have the same knowledge? Our new research shows that only some methods are capable of analysing such data. Here I refer to our recent work jointly published with Dr Damith Senanayake, Dr Wei Wang and Dr Shalin Naik.