

Data/text mining techniques in modelling climate change effects on crops

from data mining and knowledge discovery (in databases) to big data analytics and knowledge extraction for applications in science

“Data mining” for “knowledge discovery in databases” and associated computational processes, operations first introduced in the mid-1990s can no longer handle the so-called “big data”, a term being used to define the large volumes of diverse, dynamic, complex, longitudinal, and/or distributed data generated from instruments, sensors, Internet transactions, email, video, click streams, noisy, structured / unstructured and/or all other digital sources available today and in the future.

The big data also described using 3 Vs, volume, variety and velocity (lately with an additional 4th V for “veracity”) requires a new set of technologies i.e., exascale computing, architectures (distributed or grid), algorithms (for data clustering, generating association rules), programming languages, automated, scalable software tools, to uncover hidden patterns, unknown correlations and other useful information or “actionable knowledge” or “data products” from the raw data. Up-to-the-second information extracted using “on-the-fly” intelligent instantaneous data processing approaches, on almost every aspect of a business is demanded for critical decision making in enterprises, but more so it has ever been in human history, for scientific research, such as climate change simulation and model testing, neutron fission simulations, brain anatomic and functioning model development research.

The talk gives an introduction to the synergistic challenges in “data-intensive” science and “exascale” computing for resolving “big data analytics” and “data science” issues in four main disciplines namely, computer science, computational science, statistics and mathematics. For the realisation of vital identified foundational aspects of an effective cyber infrastructure, basic problems need to be addressed adequately in the respective disciplines and are touched upon.

The rest of the talk will take the audience through three data/text mining approaches with potentials for handling scalability issues when analysing large volumes of massive streaming dynamic data for modelling climate change effects on crops, being investigated at Geoinformatics Research Centre (GRC), Auckland University of Technology (AUT).

Bio

Dr Subana Shanmuganathan is currently a senior research fellow at GRC, AUT. After completing her PhD in artificial neural network computational modelling in ecological sciences at AUT, as a recipient of the JSPS (Japan Society for the Promotion of Science) fellowship award, for two years (2005-7), Subana was a visiting postdoctoral research fellow at Ritsumeikan Asia Pacific University in Japan. On her return to New Zealand she joined GRC and since then she has been working on several key research projects, namely, modelling climate change effects on New Zealand grapevine phenology and wine quality using modern data/text mining approaches, multi-sensor data analysis using association rule generation algorithms and spatial data mining using pixel clustering using an unsupervised neural network technique.