

Modelling and Analysis

Opening Thoughts

- What are modelling and analysis?
- One view is that modelling concerns inference but analysis concerns prediction.
- Hayashi (1998) wrote that “the method of data analysis has developed in the fields disregarded by mathematical statistics and has given useful results to solve complicated problems based on mathematico-statistical methods (which are not always based on statistical inference but rather are descriptive).”

Response to Question 1

- The main R&D challenges in modelling and analysis...

Data That Change

- How do we analyze three-way data?
- Real time analytics: when do we need it and how do we do it?
- How to produce efficiently updating models or learners?

Model Validation

- How do we deal with spurious points?
- How do we predict rare events?
- Big data vs good data
- How do we validate the models (including incorporating new labelled data)?

Challenging Data Types

- Mixed data types, natural language, video and audio, sensor data,...
- How do we handle sparse data?
- How do we handle missing data?
- How do we incorporate uncertainty from imputation?

Response to Question 2

- The main R&D challenges in modelling and analysis at the intersection with one of the other groups...

Complying With Legislation (SP)

- How do we comply with laws such as the EU's GDPR laws?
- If needed, how would we remove a person's private data from a pre-trained model?
- If needed, how could we build and deploy incrementally trained models where data can be added and removed?

Model Building (TU)

- How do we feedback the results from exploratory analysis into data cleaning?
- Can we develop a metric for trustability and usability of a model or dataset?
- When and how should we use synthetic datasets for model building or testing?

Modelling and Data Storage (MBD)

- How does storage and access impact modelling?
- How can we deal with the situation where an analysis requires different pieces in different servers/locations

The Balancing Act (DV)

- How do we handle the tradeoff between interpretability and prediction?
- How do we handle the tradeoff between model choice and visualization?

Closing Thoughts

- This has been a review of some of the questions that arose.
- Addressing these questions will require expertise in multiple areas.
- Almost every question could have been regarded as at, or near, an intersection.
- Thanks to everyone in Group 3.
- Special thanks to Dave Campbell.