Turing-Roche Partnership Structured Missingness Workshops Report 2021

Context

In June 2021 the Alan Turing Institute and Roche launched their 5 year strategic research partnership to establish a world-leading industry and academic collaboration in advanced analytics, focused on enabling the transformative benefits of personalised healthcare to become a reality for patients around the world.

The strategic partnership will cover multiple activities with the "North Star" of developing new data science methods to investigate large, complex, clinical and healthcare datasets to better understand how and why patients respond differently to treatment, and how treatment can be improved.

One of the first projects for the partnership is in the area of structured missingness. Missing data are a common problem that arise in many fields and can significantly complicate analysis. While there are established ways to handle data that is missing at random, it is often the case that missing readings are structured in some way. Dealing with such structured missingness is substantially more challenging yet occurs commonly in healthcare (and other) contexts. For example, one cohort of patients may have been given a certain test, another may not, leading to blocks of missing readings. Simply excluding such missing (or potentially inaccurate) data can significantly bias subsequent analysis. Consequently, there is a need to develop rigorous ways to understand structured missingness and develop tools to handle it appropriately.

Purpose of the workshops

We decided to organise 3 virtual workshops on the topic of structured missingness with the aim of convening a community of researchers interested in developing new tools and methods to learn from heterogenous data with structured patterns of missing readings.

We had a short application process to apply to attend the workshops, ensuring we brought together a diverse community of researchers, including established and early career researchers.

Overall our objective for the workshops was to scope out approaches to structural missing data and formulate substantive and collaborative research questions that merit further exploration.

After the workshops, a call for research proposals will be issued to support projects working in the area of structured missingness arising from the workshop discussions.

Structure of the workshops

Workshop 1

The first workshop took place on Wednesday 10th November 2021 with around 40 attendees. We had two introductory presentations from Chris Harbron (Expert Statistical Scientist & Advanced Analytics Lead, Roche) on the Roche Clinico-Genomic Database and from Robin Mitra (Research lead for the Turing-Roche partnership) on the current state of structured missingness.

Participants then selected one of 4 breakout rooms on the topic of:

- Making **predictions** from data with structured missingness
- **Missing mechanisms**- determining principles underpinning structured missingness
- **Inference** from data with missingness
- **Imputation** of structured missingness values

These rooms were facilitated each by the research leads- Chris, Robin, Ben MacArthur (Deputy Programme Director for Health & Medical Sciences Programme, Turing) and Ryan Copping (Global Head of Data Science Acceleration, Product Development, Roche) with high level collaborative conversations on the topics and brainstorming problems that needed to be solved. All attendees then returned to the main room where key findings were shared by the research leads.

Workshop 2

The second workshop took place on the 17th November. Attendees were asked prior to the workshop to come up with a brief research idea, which the research leads then grouped into themes which formed breakout rooms. The themes were:

- Design
- Visualisation
- Mechanism
- Networks/graphical models
- Imputation
- Benchmarking
- Causal Inference/Reasoning
- Machine Learning

This time the rooms were facilitated by the attendees themselves and the majority of the workshop was spent using whiteboards to develop further potential research ideas which could form funding proposals. The group came together at the end to share-out ideas from each room.

Workshop 3

The third workshop took place on the 1st December. Attendees had collaborated via Slack and co-working sessions in between workshops to continue working on their ideas. This workshop gave them the opportunity to present these to the group as a whole. We heard from 9 groups and individuals on a range of topics with an opportunity for questions and discussion for each presentation.

Presentation titles:

- Develop and evaluate a method to comprehensively find & explain missingness in multi-modality data
- The bidirectional influence between Structured Missingness and Adaptive Designs
- Visualization to Support Investigation of Structured Missingness in Health Data
- Personalized Cancer Treatment with Missing Data: An Optimization Approach
- Characterising and Investigating Structured Missingness
- Imputation incorporating external information
- Characterising structured missingness for optimising future designs
- Quantifying and detecting (lack of) diversity
- Using networks to address structured missingness

It was great to see the breadth of ideas that had been developed through the workshops and the collaboration and synergies in and between groups.

Next steps

A funding call will be launched off the back of these workshops in early December 2021 (will be listed here when live) where we hope the ideas formed during the workshops will be submitted as proposals: all active attendees are eligible to apply for this funding. The call will also be open to researchers who were not able to attend the workshop, however priority will be given to project proposals emerging from this event, particularly collaborative proposals that combine different areas of data science and/or different academic institutions.

We also will be running some Q&A sessions about the funding call in January 2022- links to these will be posted on the website in due course.

This is the only the beginning of the partnership and we are keen to engage with as many people as possible during the 5 years. If you'd like to connect with the partnership then please join our Slack channel or email Vicky Hellon, Community Manager at vhellon@turing.ac.uk