CBS

Forum

Behavioral AI Models using Synthetic Language: from Fast Entropy to Information Topology

Dr. Andrew D. Back

Advance Queensland Fellow
School of Electrical Engineering and Computer Science
The University of Queensland, Australia

Monday, July 10th, 2023 13:30~14:30

CBS Central Bldg, 3F Seminar Room S305

Abstract:

Finding suitable representations of socio-behavioral systems with complex long-term dependencies and statistical recurrence properties is a challenging task for AI. Current approaches based on learning from data representations have the disadvantage of requiring massively large data, while failing to effectively provide trust, explainability, and reliability. Moreover, they require a model architecture which can learn the probabilistic dependence between objects and events. The basis for human intelligence is language which can be analysed within a probabilistic framework. This raises the question of whether language approaches can be applied to building AI for understanding non-human dynamical systems? Entropy-based models are one approach to characterize language and behaviors within an information theoretic framework. However, such models typically require a large amount of data.

Here we describe a new paradigm for AI based on considering short-term probabilistic events as words within *novel synthetic languages describing* any form of dynamic system behaviors. As a step in this approach, we derive a new fast entropy estimation algorithm for natural sequences requiring small amounts of data. The effectiveness of this algorithm is shown in some applications, including detecting dementia from conversational speech.

The challenge of this approach however, is that although synthetic language may be used to describe many forms of biological and human behaviors, for non-human languages, there are not necessarily any teachers available. We show a solution to this problem by introducing a new information theoretic approach. Short-term probabilistic curvature within an information geometric framework is extended to estimate information topology. We show how this purely probabilistic model approach enables the non-semantic behavioral analysis of synthetic language. This advantage of synthetic language is that it can be adopted to use topic analysis for understanding various sources. Examples are given to show how potential meaning can be obtained from a range of dynamical systems including birdsong, the 3-D motion of meerkats, and text documents.

Host: Hiroyuki Nakahara
Laboratory for Integrated Theoretical Neuroscience