Heterogeneous Data and Large Representation Models in Science

Contribution ID: **4**   Type: **Oral presentation**

Learning how to design biomolecules using a neuro-symbolic architecture

Tuesday 1 October 2024 14:50 (35 minutes)

Designing requires to mix physical knowledge, experience accumulated from past designs and constraints defining design objectives.

Proteins are large biomolecules that play crucial roles in all living organisms. They are linear polymers which can be described as a sequence in a 20 letter alphabet (one for each amino acid). They can therefore be represented as discrete objects. In water, most proteins fold in a 3D structure, defining continuous atomic coordinates.

To design new proteins, we introduced an hybrid architecture that combines all above elements in a joint pairwise decomposable function over amino acid identities. Physics is represented as a force field, experience is extracted by Deep Learning from Nature's designs and design objectives represented as constraints. The resulting model is then passed to an automated reasoning prover to identify the most suitable chemical composition. The same architecture can learn how to play Sudoku from examples, w/o knowing the rules.

Contribution length

Middle

Primary author: Ms DEFRESNE, Marianne

Co-authors: Ms BARBE, Sophie (INRAE/INSA/TBI); SCHIEX, Thomas (Université fédérale de Toulouse, ANITI, INRAE)

Presenter: SCHIEX, Thomas (Université fédérale de Toulouse, ANITI, INRAE)