Séminaire ISIR

Mardi 27 novembre 2018 à 14H00

Alexis Dubreuil

Campus Jussieu, 4 place Jussieu, Paris
Salle H20

Reverse-engineering recurrent neural networks via low-rank approximations

Abstract : Recurrent Neural Networks are artificial neural networks that can be trained to perform a variety of tasks. They have been proven to be useful for neurosciences as a way to find new solutions for the implementation of cognitive tasks by neural networks. Indeed they can be trained on tasks typically used in experimental neurosciences and yield functioning neural networks that are fully accessible, contrary to biological networks. Here we take advantage of the accessibility of dynamics and connectivity of RNNs to propose a method that gives a concise mapping between cognitive mechanisms and network's structure. To do so we use recent work that allows a detailed analytical description of the link between structure and dynamics in networks with low-rank recurrent connectivity matrix, i.e. that writes as outer products of a few column vectors $\vec{m}_k$ and line vectors $\vec{n}_k^T$. It has been shown that by examining the geometrical arrangements of the vectors $\vec{m}_k$'s and $\vec{n}_k$'s, together with vectors mediating network's inputs and outputs, one can understand how recurrent connectivity processes inputs and produces outputs to perform a task. Based on these results, here we adopt the following approach to analyze trained RNNs, i) we train a RNN to perform a given task, ii) find an approximate low-rank connectivity structure that preserves network's functionality, and iii) describe the cognitive mechanisms at stake by linking network's structure and dynamics. We demonstrate the validity of this approach on two commonly used tasks in experimental neurosciences: a perceptual decision-making task (Random Dot Motion task) and a delayed discrimination task (Romo task). From this we can give a precise description, in the language of dynamical system theory, of the cognitive mechanisms involved in performing those tasks. We then discuss the application of this method to more elaborate cognitive tasks.

Short bio : Post-doctorant au "group for neural theory" à l'ENS. Intéressé à comprendre comment des réseaux de neurones implémentent des processus cognitifs. A mené des travaux théoriques sur les réseaux à attracteurs, pendant la thèse sur la mémoire de travail et pendant un travail de post-doctorat sur les réseaux de cellules de grille. Puis travail sur des données biologiques pour comprendre l'implémentation neuronale d'un comportement sensori-moteur chez le poisson zèbre.

Institut des Systèmes Intelligents et de Robotique
UPMC – CNRS / UMR 7222
Pyramide Tour 55 - Boîte courrier 173
4 Place Jussieu – 75252 PARIS cedex 05 – France
Tél. : +33 (0)1 44 27 51 31 / 51 41
Fax : +33 (0)1 44 27 51 45