

# 社会学習エージェント系におけるダイナミクスとナッシュ均衡への収束

信大理<sup>A</sup>, 北里大理<sup>B</sup>  
中山一昭<sup>A</sup>, 守真太郎<sup>B</sup>

<sup>A</sup>*Shinshu Univ.*, <sup>B</sup>*Kitasato Univ.*

K. Nakayama<sup>A</sup> and S. Mori<sup>B</sup>

社会的学習を行う  $N$  エージェント系。社会的学習とは他者の行動を参考に自分の意思を決定する方法。これに対して個人的学習は自力で情報を探して意思を決定する方法。

各エージェントは非定常多腕バンディット (rMAB) の「レバー」を引くことで得られる「利得」を最大化するように行動する。rMAB は利得 0 のレバーと 1 のレバーを 1 本ずつ持つ。無作為にエージェントを一人選ぶ。彼は 1 のレバーを知っていればレバーを引いて利得を得る。知らなければ確率  $r$  で他のエージェントの知っているレバーを選択し、確率  $1-r$  で自力で 1 のレバーを探索。前者の成功確率は 1 とし、後者のそれは  $q_I$  とする。次に rMAB は確率  $p_C/N$  で 1 のレバーが 0 に変化し、新たに 1 のレバーが出現する。これを繰り返すとき、エージェントの利得は一定値に収束することが示される。

さて、いまその収束が十分速いと仮定。つまり各エージェントが  $r$  の値を変更したとき、各エージェントの利得は直ちにその一定値で表されるとする。このとき「各エージェントは自分の利得が増大する方向へ  $r$  を変更する」という力学系を自然に考えることが出来る。

一つの力学系として次のようなものを考えてみる。エージェント  $n$  の  $r$  の値を  $r_n$  で表すと、その最適応答は  $r_k, k \neq n$  たちの関数。それを  $f_n$  とするとき

$$\frac{dr_n}{dt} = f_n - r_n, \quad n = 1, \dots, N$$

このとき  $t \rightarrow \infty$  にて  $(r_1, \dots, r_N)$  は Nash 均衡点へ収束することが示される。

より現実的には、各エージェントは自分の最適応答値  $f_n$  を知ることは出来ないと思われる。例えば、各エージェントは  $r_n$  を微小変更したときの利得の変化、つまり利得の微分係数しか分からない場合、状況はかなり複雑になるが、現在までに判明していることについては報告したい。