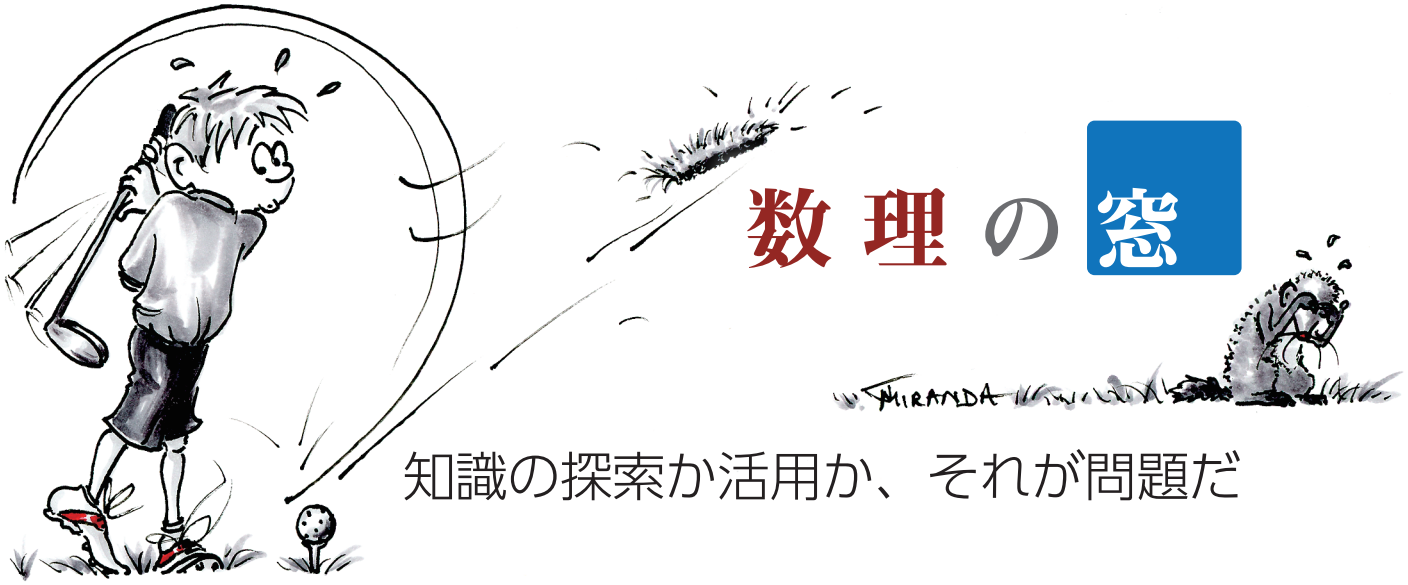

数理の窓

知識の探索か活用か、それが問題だ

ここに10台のスロットマシンがある。マシンのアームを引くと一定の確率でコインが得られる。各マシンの当たる確率はバラバラだが、時間によっては変化しない。この確率は事前に分からないとし、これから1000回引けるとして、得られるコインを最大にする戦略はあるか？

これは多腕バンディット問題と呼ばれる。ポイントは、一回一回の試行を、どのマシンが当たるかの探索に向けるか、知識を活用してベストと信じるマシンへの投資を続けるか、どちらに配分するかだ。〈探索〉と〈知識活用〉はトレードオフとなっている。

ギャンブラーにはいくつか代表的なタイプがある。タイプ1は、すでに自分内の基準を持ち、例えば30%当たり基準を上回るマシンがあったら、それにずっと賭け続ける。タイプ2は、とりあえず、全部のマシンを試行して知識を得てから、最適なマシンに投資を続ける。タイプ3は、一台に賭け続ける中で、一定の時間後に別マシンの探索も混ぜる。これらの混合も考えられる。

戦略追求のため、機械学習プログラムによるシミュレーション研究[1)]が多くある。最初に考慮すべきは、一台のマシンを何回引いたら当たり確率をフィックスさせるかと、ベストの評価基準となる確率値（これは式になる）の設定だ。さらに1台のマシンで外れたらそのマシンの確率は下げつつ、他の価値を上げる相対評価など工夫のしどころは多い。ただし、すべての確率分布に対応した最適なアルゴリズムはないとされる。

この問題は、ネット広告配信への応用が有名だ。例えば1つの新車に複数の広告がある中で、1小窓に同時には1広告しか出せない。ユーザーのクリックを報酬として、どの広告に多くの時間を割り当てるか、傾向を踏まえて動的に最適化するのだ。

また、これまで1マシンの確率は一定と仮定していたが、発展系として“敵”がギャンブラーの戦略に応じて報酬を最小とするように確率を操作するパターンもある。もちろん、公平のため敵は選択の前に確率を決めるが。これは敵対的バンディットと呼ばれ、カジノ運営に使われていたら恐ろしい。ランダム選択を混ぜることがギャンブラーに必須である。

ところで、旅先や料理の選定・恋愛においても〈探索〉と〈知識活用〉の比率は重要だ。若い時は前者を重視し、そのうち後者が多くなる。したがって、その比率から、その人の知識量や残りの試行回数が推定できそうだ。ただし、純粋な知識の増加だけに喜びを見出すタイプもいることには注意だ。

（外園 康智）

1） ε-greedyやUCB1、Thompson Samplingなどのアルゴリズムが有名だ。