

Interactive Social Slot Machine の開発

SP-10133 馬場 佳吾 非線形物理学講座

1. ゲームの開発目的

Rendell らの論文では、効率的な学習方法と言われるソーシャルラーニング(カンニング)について、特に最も効率的なアルゴリズムの条件について書いている。むやみにカンニングをしても良い結果は得られない。それならば、どのようにカンニングするのが一番効率的なのか?この問題を、多腕バンディッドという、多数のレバーを持ち、レバーを引いたときに出るコインの枚数が時間とともに確率的に変化するゲームでのトーナメントを行うことにより、最適なアルゴリズムの条件を明かにした。

今回は、人間 VS AI(人工知能)という形で、以下のことを明かにする実験を行うためにゲームを開発した。多腕バンディッドをプレイする人間と AI との間でどのような違いがあるのか。人間は、自分で情報を得ることと、他人の真似をすることをどのように使い分けるか、つまり、どのようにしてソーシャルラーニングとソーシャルラーニングを行うのか。

2. Social Learning とは

例えば美味しいランチを食べたいときに、適当に行ったことのない知らないお店に入るのではなく、ウェブなどを使って、どんなお店かをチェックして、評判のいいお店を選ぶ。この様に自分一人で0から進めていくのではなく、他者の実績や選択などを参考にして、物事を進めていくことがソーシャルラーニングと言える。ここでは、それぞれが他者の行った選択を見る方法があり、それを活用できる。ソーシャルラーニングを行うのと、自分の力だけで行うのとでは、それぞれ、メリットとデメリットがある。ネットの意見を参考にした方が、効率よく美味しいお店を見つけられる。一方で、ネットの情報が全く見当違いなものだった場合は行って見たが美味しくなかったとなる。自分の力で探せば、穴場の名店を見つけられるかもしれない。一方で、よいお店を見つけるのに途方もない労力を必要とする場合がある。

3. ゲームの仕様

プログラムは、PHP という言語を使って開発した。ゲームには、多腕バンディッドスロットマシンを使う。多腕バンディッドとは、それぞれ獲得枚数の異なる片腕バンディッドスロットマシンが複数並んでいるものを模したものである。100本のレバーを用意し、レバーを引くとコインを獲得できる。獲得できるコインの枚数は最小で0枚、最大で100枚。コインの枚数の出る確率は、確率的に変化する。レバーの情報、つまり、コインの枚数は一定の確率 p_c で変化して、更新される。レパートリーは、利用しているレバーのリストである。それは、コインの枚数の情報を持っている。最大5本のレバーの情報を格納する。

このスロットに3つのボタンを付けた。

In(イノベイト)

100本のレバーの中からランダムに1本選び、そのレバーを引くと何枚のコインが得られるのかを調べられる。イノベイトされたレバーの番号は、レパートリーに格納される。In を行ったレバーの情報は最新のもののなので次のターンでは変化しない。

E(エクスプロイト)

レバーを引いて、コインを得る。レパートリーに格納されているレバーでしか E は行えない。レパートリー内の最新5回のイノベイトまたは、オブザーブされたレバーの中から、1つを選択する。直前にイノベイトしたレバーを引く場合、レバー情報は更新されていないので、表示された枚数と同じだけのコインを獲得する。

Ob(オブザーブ)

直前にエクスプロイトしたプレイヤーの中でランダムに一人選んで、何番のレバーで何枚のコインを得たのかがわかる。そのレバー番号はレパートリーに格納される。Ob を行うとレバー情報は次のターンに確率 p_c で変化するので、そのレバーをエクスプロイトすると必ず同数のコインが得られるわけではない。

操作画面



プレイヤー(人間)と対戦する戦略 AI には、大越が解析したディスカウントマシンのアルゴリズムを修正したもの、ディスカウントマシンでのオブザーブの代わりにイノベイトを行うもの、最初の1回イノベイトし、その後ひたすらエクスプロイトするものなどの戦略 AI を用いる。AI の戦略を変えることでゲームの難易度を調整する。

参考文献

L.Rendell Why Copy Others? Insights from the Social Learning Strategies Tournament
Science 9 April 2010: Vol. 328 no. 5975 pp. 208-213
DOI: 10.1126/science.1184719