

農作物の灌水制御最適化に向けた強化学習手法の検討

難波 脩人

辻 順平

能登 正人

神奈川大学大学院工学研究科電気電子情報工学専攻

1 はじめに

近年、センサやIoT技術の発展によりそれらの技術を農業に生かしたスマート農業が注目されている。この分野では従来経験と勘によって行われていた農作業の方法を再現するために機械学習が用いられている。特に植物の育成データや環境データを収集し、教師あり学習によって熟練農家の知見を再現しようとする研究は盛んに行われている。しかしながら、教師あり学習を行うために収集するデータには熟練農家の知見が大きく影響し、こうした知見が真の最適解かどうかの判断は困難である。熟練農家の経験と勘に依存しない学習手法として、環境から与えられる報酬に基づいて学習を実行する強化学習が挙げられる。実際に植物の栽培に強化学習を適用した研究は存在するものの事例は少なく、その有効性は十分明らかにされていない。

本研究では、植物の栽培に強化学習を適用するための準備として、簡易的なQ学習手法に基づく小松菜の栽培を実践し、その有効性や課題を検討する。

2 関連研究

植物と強化学習を組み合わせた研究は、シミュレーションによる収穫量の最適化や実際に栽培を繰り返すことによる養液供給の最適化を目的としたものがある。実際の植物栽培に強化学習を適用した研究に焦点を当てた場合、栽培期間の長さや状態の時間変化など植物のもつ特徴によってロボティクスで用いられるような手法をそのまま適用することが困難であることが知られている。Wakaharaらの研究[1]は植物の栽培に強化学習を適用した際に生じるこれらの問題を植物の状態遷移や報酬の与え方から定義することで解消している。しかしながら、このように強化学習を植物に適用した例はあるもののその手法が植物栽培に適した手法かどうか十分な議論はされていない。

3 植物の灌水制御最適化のためのQ学習

本研究では実際に小松菜の栽培に強化学習を適用することでその有効性や課題の検討を目的とする。強化学習によって植物栽培の最適化を行うにあたり重要な点として状態の定義が挙げられる。植物の状態は時間とともに変化することから同じ状態は1エピソードの中に1度しか現れず、全ての状態に対してQ関数を作成することは現実的ではない。さらに、植物の状態は連続量であることから状態をQ関数で表すことが困難である。このような場合においてパラメータ W_{kj} を用いてQ関数の作成に関数近似を用いる手法が知られている。本研究におけるQ関数導出の概要を図1に示す。Q関数は \tanh 関数を用いて近似するため、 $f(u_j) = \tanh(u_j)$ としたとき、以下の関係式が成り立つ。

$$Q(s, a^{(j)}) = f(u_j) \quad (1)$$

$$u_j = \sum_{k=0}^n W_{kj} s^{(k)} \quad (2)$$

ここで、時刻 t の状態 s_t は時刻 t 、 $t-1$ 、 $t-2$ の各時刻における植物の背丈と葉の広がりを表した2次元ベクトルを3つ並べた6次元ベクトルとして $s_t = (s_t^{(1)}, s_t^{(2)}, s_t^{(3)}, s_t^{(4)}, s_t^{(5)}, s_t^{(6)})$ で表される。

学習時におけるパラメータ W_{kj} の更新式は $\delta_t = r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)$ とすると以下の式(3)のようになる。

$$W_{kj} \leftarrow W_{kj} + \alpha \delta_t \frac{f(u_j)}{\partial W_{kj}} \quad (3)$$

また、報酬は時刻 t と $t-1$ における植物の背丈を比較し、以下のように定義する。

$$r_t = \begin{cases} +0.01 & (\text{時刻 } t \text{ で成長している}) \\ 0 & (\text{変化なし}) \\ -0.01 & (\text{時刻 } t \text{ でしおれている}) \end{cases}$$

行動(action)は $a^{(1)} = "0 \text{ ml}"$ 、 $a^{(2)} = "10 \text{ ml}"$ 、 $a^{(3)} = "15 \text{ ml}"$ 、 $a^{(4)} = "30 \text{ ml}"$ の灌水の4パターンを用意し、状態 s_t でactionを起こした場合に次の状態 s_{t+1} に遷移するのは2日後と間隔を開けるよう設定し

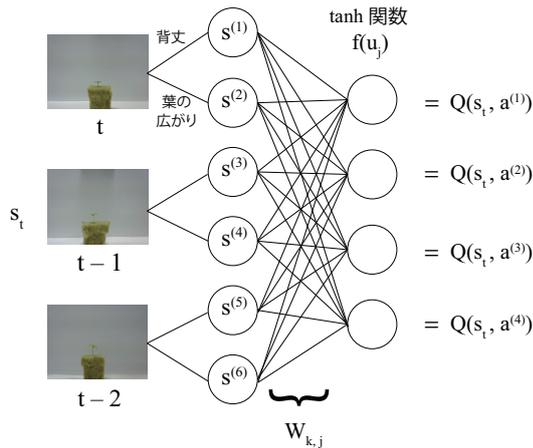


図 1: 関数近似を用いた Q 関数の作成概要

ている。さらに、農作物の栽培期間は1か月から数か月かかるものがほとんどであり、種を植えてから収穫するまでの期間で強化学習を行うことは困難であることが考えられる。そのため、種を植えてから4日後の状態を s_1 と定義し、その状態から2週間を1エピソードとして強化学習を行う。

4 実験および結果

シミュレーションに強化学習を適用する場合などと比較すると、植物はエピソード数やステップ数の少なから最適解に収束することは困難であることが考えられる。そのため、実験は experience replay (ER) [2] を採用する。つまり、人の手によって栽培することで得られるデータ $\{(s_t, a_t, r_t, s_{t+1})\}_{t=1}^T$ を保存し、パラメータ $W_{k,j}$ の更新を過去の植物の栽培データをもとに行うことを考える。データ採取は小松菜48株に対して室内栽培用ハウスキットを用いて行った。発芽しない個体や枯れてしまう個体が現れたため、最終的にデータを採取することができた個体は28株であった。1株あたり6つのデータを持つため、28エピソード168データを訓練データとしてERを行った。学習の成果を評価するために、訓練データに使用したデータを入力し、予測された action と実際の action を比較した結果の一例を図2に示す。一番良い結果は適合箇所は3つであり、悪い結果は適合箇所が1つとなっている。

また、作成した Q 関数をもとに実際に栽培することで植物が成長するかどうかの確認を行った。小松菜16株に対して実験を行った結果、9株に水の与えすぎ、3株は枯れ、4株は通常通りの成長といった様子が確認できた。

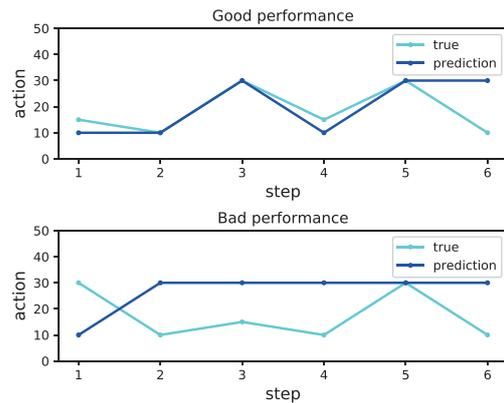


図 2: action の比較 (上が良い結果, 下が悪い結果)

5 考察

図2に示すように学習結果は良い結果と悪い結果が確認できる。本研究では学習が収束しやすいよう簡易的な Q 学習を想定している。関数近似は比較的単純な構造をしているにもかかわらず Q 関数が最適解に収束していないと考えられる。また、栽培に用いた個体数が少なく得られたデータも少ないことからERを行っている際に重要ではないデータが重複して使用されていることが考えられる。今後は植物の栽培データを増やし、重要なデータを再利用する prioritized experience replay など用いて実験結果を改善したい。

6 おわりに

本研究では農作物の灌水制御最適化に向けて実際に小松菜の栽培に強化学習を適用することからその有効性や課題の検討を目的とした。植物の特徴である状態の時間変化やエピソードの長さを考慮する手法として関数近似による Q 学習を採用し、実験を行った結果、提案手法では最適解に収束することが困難であることが確認できた。植物にとって最適な強化学習手法を明らかにするためにはより議論を深める必要があると考えられる。

参考文献

[1] Wakahara, T. and Mikami, S.: Adaptive Nutrient Water Supply Control of Plant Factory System by Reinforcement Learning, *Proc. of SCIS & ISIS 2010*, pp. 1020–1025 (2010).
 [2] Lin, L.-J.: Self-improving Reactive Agents Based on Reinforcement Learning, Planning and Teaching, *Machine Learning*, Vol. 8, No. 3-4, pp. 293–321 (1992).