

機械学習による植物の病気の自動判別システムの開発

史 中超 研究室

1561089 野村 祐貴

1561094 二見 大揮

1. 研究背景・目的

近年、新規就農人口者数が減少傾向にある。

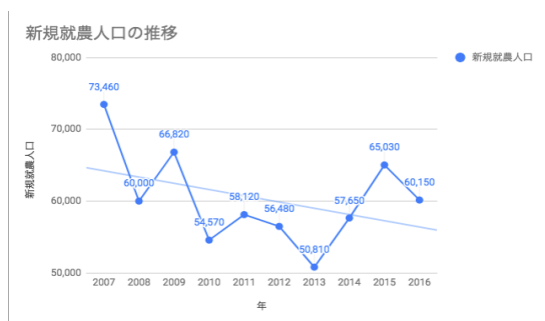


図1 農林水産省 新規就農人口者数推移[1]

図1に示すように、2007年に7万人以上だった新規就農人口者数は翌年に6万人に減少している。以降7万人台に回復した年はない。また、農業全体で高齢化が進んでおり、農業における就農人口の減少に拍車をかけている[2]。そんな中で近年、農業界では効率的な支援や方法が行われている。その中の一つにドローンの農薬散布がある。農薬の散布をドローンにすることで時間やコストの削減に役立てることができる。しかし、農薬を散布するには植物の病気に関する知識を有している者が必要であり、そのような知識を有している者が減っているのが現状である。そこで、本研究では機械学習を用いて植物の病気の知識を持っていない者でも病気を判別することができるモデルを作成し、農業に関する知識を保っていくとともに、効率的な農業に応用していくことを目的とする。また、今回は数多くある植物の中で国内の圃場規模と実用性を考慮し、イネの病気を判別するモデルを作成する。

2. 機械学習とは

機械学習は主に大量のデータから数値を予測する回帰問題と、データからクラスを分ける分類問題に大別することができる。本研究では大量の画像からイネの病気を判別するため分類問題を題材として機械学習を行う。一般的に画像認識における機械学習方法でもっともよく利用されるアルゴリズムは畳み込みニューラルネットワーク (Convolutional Neural Network) である。本研究ではこの畳み込みニューラルネットワークを利用した Inception V3 と呼ばれる機械学習モデルを使用して、イネの画像の学習を行う。イネの病気はいもち病、縞葉枯病、ごま葉枯病の3種類である。

3. システムの開発

イネの病気を判別する機械学習モデルを作成する際に大量の画像データが必要になる。画像の収集にはスクレイピングという技術を用いる。本研究ではスクレイピングツールとして、icrawler と呼ばれるオープンソースライブラリと Google Custom Search API を使用する。言語は Python を使用する。

機械学習のモデルの開発を行う。言語は Python を使用する。また、機械学習を行う際にはかなり強力なマシンパワーが必要となる。本研究では Google 社の提供する Google Colaboratory を採用している。Google Colaboratory が公開されたのは 2018 年であり研究前例がないことや GPU を無料で利用することができることが採用理由である。

図2のように Google Drive 上からプロジェクトを作成する。ディレクトリ配置は Colab Notebook のソースコードをルートディレクトリに置き、train ディレクトリに学習用、validation ディレクトリに検証用を置いている。

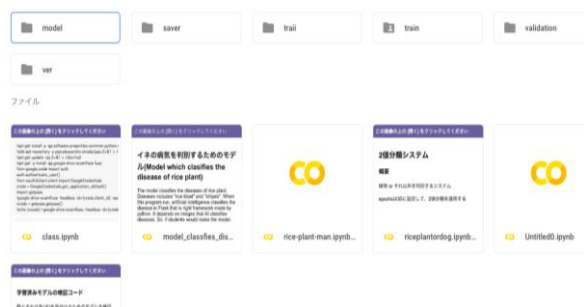


図2 Collaboratory プロジェクト

train ディレクトリ内には3種類の病気がファイルごとに分かれている。validation ディレクトリ内も同様の分け方をしており、Keras[3]で畳み込みニューラルネットワークを利用してそれぞれの特徴を学習させていく。また、Web アプリケーション化を行うことで実用性を高めている。

4. 検証

まず、イネの病気が写っている画像を収集する。図3にスクレイピング手法で収集した画像の一部を示す。



図3 収集した画像

収集した画像の枚数はそれぞれ300枚程度であった。中にはイネの病気とまったく関係のない画像も含まれている。本研究では、目視で収集した画像からイネの病気画像かどうかを選別し、それぞれの病気を約100枚程度選んだ。

次に、イネの病気の判別システムの検証を行う。図4は作成した Web アプリケーションのトップページを示したものである。図5に示すように、

ファイルを選択するボタンをクリックすると画像をアップロードできる。アップロードされた画像は機械学習モデルに適用され畳み込みニューラルネットワークで解析される。画像の特徴量から出力層に閾値が出力され、もっとも近い閾値だった病気に当てはまる。

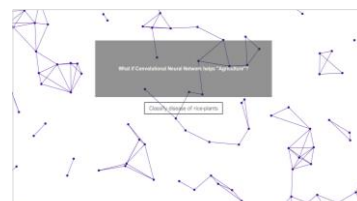


図4 Web アプリケーション



図5 病気の判別

図5の Prediction の後に病名が表示され、その確率と当てはまり度が Confidence となる。

5. まとめ

本研究では新規就農者がイネの病気を発見した際に病気判別の一助として機械学習を利用したシステムを提案した。この機械学習モデルは実用性に長けている。イネ以外の植物の病気に画像を入れ替えられるという汎用性も兼ね備えている。また、Google 社は機械学習の実用化に向けて GPU を無料で開放しており、機械学習を実生活に活用する環境は整ってきている。一方、病気の画像を集める際には大量で高品質なデータが必要となる。学習用データ自体の品質が低い場合には判別の精度が低くなってしまふ恐れがある。本研究においてもサンプルが少なかったため、認識の確率があまり高くないものもあった。今後、サンプルを増やし、より多くの種類の病気をより正確に判別できるような機械学習モデルへの改善が課題として残っている。

6. 参考文献

- [1]e-Stat 政府統計の総合窓口
<https://www.e-stat.go.jp/>
- [2] 農林水産省 農業労働力に関する統計
<http://www.maff.go.jp/j/tokei/sihyo/data/08.html>
- [3]Keras
<https://keras.io/ja/>