

データセット作りのポイントから、
知っておきたい「精度」と「課題設定」まで

機械学習の良し悪しを 決める「学習用データ」

土井 伸洋

本章では、機械学習（深層学習も含む）にとって、なぜ、学習用のデータが重要なのかを考えてみます。

ここでは、他文献にある数学的な原理や公式は一切取り上げず、教師データや学習について、感覚的にどういうものかをつかむことを最優先に話を進めます^{注1}。

データを使って学習するとは？

● 果樹収穫支援ロボットを作るとしたら

収穫ロボットを作る…こんな想定で話を始めることにします（図1）。近年のアグリテック^{注2}の進化は素晴らしく、トマトやアスパラガスを自動で収穫するロボットも出現しています。

さて、果樹収穫を支援してくれるロボットを構成する要素としては何が必要でしょうか。果樹園内を移動するための足や、果樹をつかみ取る手、これらを動かす動力や制御部など、さまざまな要素が必要です。そしてなにより「カメラ画像を使って果樹を認識する機能」も欠かせないでしょう^{注3}。果樹を収穫するならば、その種類や熟れ具合、対象とする果樹と葉や枝との区別をしなくてはなりません。このような状況で実現に向けて解決すべき課題について具体例で考えてみます。

注1：そのため、専門家からすると語弊があると思われる部分があるかもしれません。

注2：農業（Agriculture）と技術（Technology）を掛け合わせた造語です。農業におけるさまざまな課題をAIやIoTを活用して解決しようとする取り組みのことです。

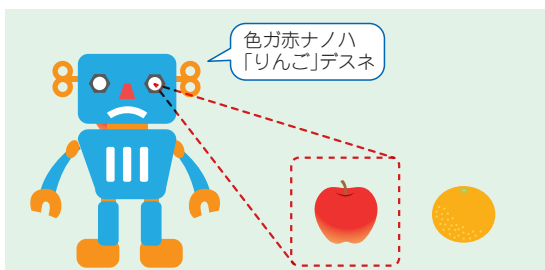


図2 色違いの果樹を区別するなら…色を手がかりにする

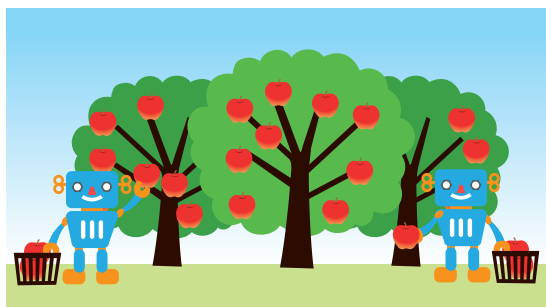


図1 果樹収穫支援ロボットのイメージ

● 第1問：りんごとみかんがある場合

例えば、果樹園の中にりんごの木とみかんの木がある場合を考えてみましょう（図2）^{注4}。カメラ画像からりんごとみかんを区別するにはどうしたらよいでしょうか。多くの方は「色を基準に区別すればよい」と考えるはずですが。

りんごは赤、みかんは橙、という判断基準をロボットにインプットしてやれば実現できそうです。

注3：カメラ以外のセンサを利用する方法もあり得ますが、話の都合上、カメラ画像を利用する構成とします。

注4：現実のアグリテックでは、多種を対象にする事例は筆者が知る限りはありません。認識対象は種別ではなく、単一の果樹に対する育成具合、大きさ、形状、色づき、病気の有無などになります。本章での課題は、あくまで説明上の課題とさせていただきます。

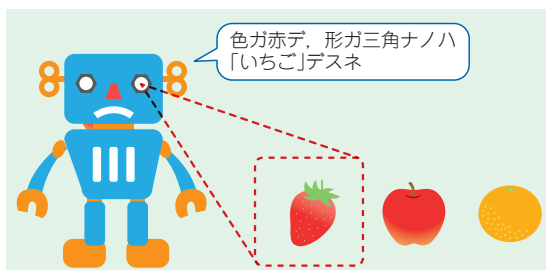


図3 同じ色の果樹を区別するなら…色と形を手がかりにする