

5J-10

決定木による多次元データ分類モデルの構築

堀江隆太† 光吉和哉† 中富竜一郎† 杉本典子‡

崇城大学大学院工学研究科修士課程† 崇城大学情報学部‡

1 はじめに

気象情報を活用した農業データマイニング手法の開発は、地球温暖化など社会的に問題となっている異常気象に対応し、効率的な農業経営を推進するための重要な課題となっている。本研究では、機械学習システム BONSAI [1] を活用して、日照時間や、降水量などの多次元データから、農作物の育成に影響を与える気象条件を生成する手法を提案する。BONSAI は、シンプルで精度が高い規則を生成するシステムである。しかし、BONSAI などの機械学習システムは、既存手法の殆どが、単一の構造に特化したものであるため、多様な属性から構成されるデータを直接マイニングの対象とすることは困難である。[2] そこで、1次元データから生成した規則を1つに統合することで、多次元データから農作物の育成に影響を与える気象条件の生成を行うための手法を提案する。

2 農業データマイニング

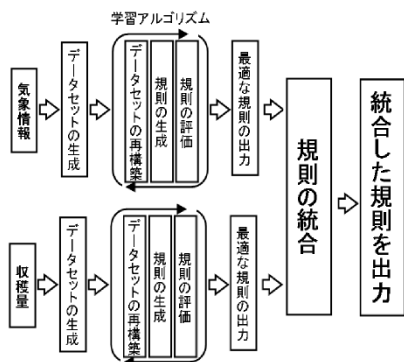


図1 農業データマイニングのプロセス

図1に、本研究における農業データマイニングのプロセスを示す。気象情報や農作物の収穫量情報など、農作

Constructing a Model for Multi Dimensional Data Classification by Decision Trees.

† R. Horie, K. Mituyoshi, R. Nakatomi · Faculty of Engineering, Sojo University.

‡ N. Sugimoto · Department of Computer and Information Science, Sojo University.

物の育成に影響を与えていると思われる情報を用意する。それらの情報を数値から記号に変換することで生成した、データセットを学習アルゴリズムに入力することで、農作物の育成に影響を与える気象条件が生成される。次に、学習アルゴリズム内で生成された気象条件の評価が行われ、精度が悪いと評価された場合は、データセットの再構築が行われ、改めて気象条件が生成され、評価が行われる。このプロセスを繰り返し、精度が高い気象条件が出力される。この気象条件は、データセットにした各情報ごとに出力される。これらの気象条件を統合していくことで、多次元データからの農業データマイニングを行う。本研究では、学習アルゴリズムとして機械学習システム BONSAI を活用し、農作物の育成に影響を与える気象条件を正則パターン (regular pattern) 上の決定木で表現する。

3 正則パターン上の決定木

Σ を有限アルファベットとし、 X を変数の帰納的集合とする。このとき、パターン (pattern) とは、 $\Sigma \cup X$ の要素からなる文字列であり、各変数が高々1回しか出現しないパターンを正則パターンと呼ぶ。正則パターン π が定義する言語 (Language defined by π) とは、 π 中の全ての変数を Σ 上の文字列に置き換えて得られる文字列全体からなる集合であり、 $L(\pi)$ で表される。

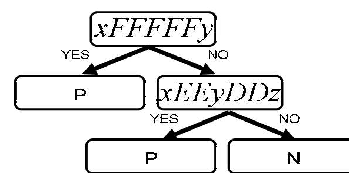


図2 正則パターン上の決定木

図2に、正則パターン上の決定木を示す。正則パターン上の決定木とは、内部ノードに正則パターンが、葉ノードにクラス名がラベル付けされた木構造のグラフである。この決定木に与えた文字列 w に対して、ルートノードから葉ノードに向かって、各内部ノードにラベル付けされている正則パターン π に対して、 w が $L(\pi)$ の要素か否かが判断される。 w が $L(\pi)$ の要素であれば左の子ノードへ、要素でなければ右の子ノードへと進む。

この過程を繰り返し，到達した葉ノードにラベル付けされているクラス名で，与えた文字列 w を分類する．

4 決定木の統合

機械学習システム BONSAI は，降水量や風速など1つの属性から決定木を生成するシステムである．しかし，農作物の育成には気象情報以外にも土地情報など，様々な属性が複雑に影響を与え合っていると考えられる．そこで，BONSAI を用いて，1つの属性ごとのデータセットから生成した，農作物の育成に影響を与える気象条件の決定木を1つに統合する．これにより，多次元データから，気象条件を決定木で生成する手法を提案する．属性ごとの決定木の中から等価な決定木を見つけ，それらの決定木を1つに統合していく．等価な決定木とは，各決定木に分類のために与えた文字列に対して同じ分類を行う決定木である．

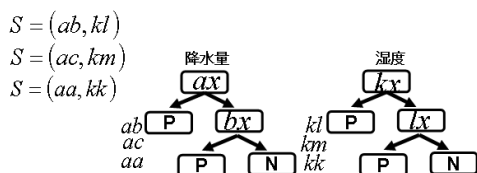


図3 等価な決定木

図3に，等価な決定木の例を示す． S を，生成した決定木に与えるテストデータとする．降水量情報から生成した決定木と，湿度情報から生成した決定木に対して S を与え分類させる．このとき， S の文字列が，降水量の決定木において，ルートノードから左の子ノードへ分類された場合，必ず，湿度から生成した決定木でも，同じ位置の葉ノードに， S の同じ組の文字列が分類される．このような分類を行う決定木を等価な決定木とする．

等価な決定木の統合は，等価な分類を行う葉ノードの深さがより浅い決定木で，もう一方の決定木を包含することで統合する．

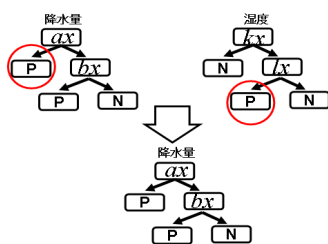


図4 等価な決定木の統合

葉ノードを，与えた文字列に対して同じ分類を行ったノードとする．降水量の決定木と同じ分類を行う葉ノードの深さが，湿度の決定木と同じ分類を行う葉ノードよりも浅いため，湿度の決定木が降水量の決定木に包含されている．もし深さが同じ場合は，葉ノードの分類の精度がより高い決定木で低い決定木を包含する．

この統合手法の計算時間を考える．この手法では，決定木の生成において決定木の深さを限定しなかった場合，計算時間が爆発してしまう．しかし，本研究で活用した BONSAI は深さが高々2～3の小さな決定木を生成するため，計算時間は爆発しない．

5 まとめ

本研究では，農作物の育成に影響を与える気象条件を，平均気温や降水量などの多次元データから決定木で見つける手法を提案した．この手法は，稲や根菜など様々な農作物に適用できる汎用性が高い手法であると考えられる．本研究で提案した手法は，農業経験が浅い農家の方々の，農作物の育成における意思決定を支援するための手法の1つになりうる．今後は，農作物の育成における意思決定をより強力に支援するために，気象情報のみならず土地情報や肥料などの農業用資材に関する情報なども活用して，農作物の育成に与える影響を検出する手法を開発することを考えている．本研究で提案した手法で，気象情報だけでなく土地情報などの他の情報も組み合わせ，農作物に影響を与える気象条件が出せるかを，実験を通じて検証を行っていく予定である．

参考文献

- [1] Shimoazono, S., Shinohara, A., Shinohara, T., Miyano, S., Kuhara, S., Arikawa, S., Knowledge Acquisition from Amino Acid Sequences by Machine Learning System BONSAI, Transactions of Information Processing Society Japan : 35, pp 2009-2018, 1994 .
- [2] Ozaki, T., Watanuma, T., Ohkawa, T., Mining Classification Rules from Multidimensional Structured Databases, Transactions of the Japanese Society for Artificial Intelligence : 22, pp 173-182, 2007.

図4に，等価な決定木の統合を示す． で囲っている