6M-06

レンタル用ユニットハウスの返却時期予測における共同共進化を用いたデータ前処理手法

舟久保龍成† スティーヴェン・クレイネス§ 大谷紀子‡

東京都市大学大学院環境情報学研究科† 東京都市大学メディア情報学部‡ 三協フロンテア株式会社§

1. 研究の背景と目的

現在，三協フロンテア株式会社ではユニットハウスのレンタル事業を行っている．ユニットハウスとは再利用可能な組み立て式の建築物で，仮設住宅や仮設事務所等に使用されている．貸出し時にユニットハウスの返却予定日は申告されるものの，状況に応じて使用期間を借り手が変更することができる．ユニットハウスの生産は在庫数によって決める必要があるが，返却予定日が変更される可能性があるため在庫を予測して生産計画を立てることが難しい．先行研究として，2018 年より決定木と遺伝的アルゴリズムを用いた返却時期予測の研究[1]が進められているが，高い予測精度が得られていない．決定木とはツリー構造で表された分類規則であり，分類基準が明確に示されるため，各クラスを特徴づける要素を知ることができる点が長所である．ユニットハウスの返却時期予測においては，予測結果だけでなく返却時期を左右する要因を知ることも重要であるため，決定木の使用が有効であると考えられる．

先行研究では，生成された決定木の正解率や可読性の低さが問題として挙げられる．決定木生成に使用したデータを調査した結果，データの属性の調整，および属性値の変換に問題があることがわかった．決定木生成アルゴリズムには，C4.5 を採用している．C4.5 では，属性値の種類数により各ノードの分岐数が決まるため，属性値をカテゴリ化して種類数を減らす必要がある．また，決定木生成に使用する属性の選定に問題があることも考えられる．決定木の精度を上げるには重要な属性を組み合わせてデータセットを作成する必要がある．決定木に使用するデータの適切な属性及び属性値の調整ができれば，予測精度の向上が見込める．予測精度向上を達成することによって，ユニットハウスの返却時期と，返却時期を左右する要因を知るという目標が達成可能である．

本研究では，正解率が高く，可読性の高い決定木の生成を目的として，共同共進化[2]に基づく適切なデータ前処理手法を提案する．データの前処理では，特徴選択と属性値のカテゴリ化を対象とする．

2. 共同共進化に基づくデータ前処理手法

共同共進化は，複雑で動的な問題を扱うことを目的として提案された遺伝的アルゴリズムである．統治分割法に基づき，問題に対する解がいくつかの部分の組み合わせにより構成されると考える．各部分に対する解の候補を個体として保存し，各個体を協力的な相互作用によって進化させることで最適解を構成するために必要な部分を獲得する．解の一部を部分解，部分解の組合せによって形成される完全な解を全体解と呼ぶ．本手法では，部分解で各属性の特徴選択と属性値のカテゴリ化を表す．使用するデータに含まれる属性数を N，属性 A_i の属性値の種類数を $n(A_i)$ とすると，全体解は N 個の部分解へのポインタで構成される．部分解の長さは $n(A_i)+1$ で，先頭遺伝子で属性 A_i の特徴選択を表し，以降の遺伝子で属性値のカテゴリ化を表す．先頭遺伝子は 0 か 1 で表現され，0 の場合その属性をデータセットに使用，1 の場合その属性をデータセットに使用しない，とする．先頭以降の遺伝子は，0 以上属性値の種類数未満の値で表現され，同じ値のものを同カテゴリとして使用する．全体解は各属性に対応する部分解の組み合わせによって構成され，1 個体で決定木生成に適したデータの前処理を表現する．各個体を評価するために，全体解ごとに決定木を作成する．訓練データやテストデータとは別に決定木の精度を求めるデータを精度測定用データとして使用する．得られた決定木の正解率を x，ノード数を y，正しく分類された事例の偏り度合いを z として適応度 $f(x, y, z)$ を式(1)より算出する．

Data preprocessing method using cooperative coevolution in predicting return date of rental unit houses

† Ryusei Funakubo, Graduate School of Environmental and Information Studies, Tokyo City University

‡ Noriko Otani, Faculty of Informatics, Tokyo City University

§ Steven Kraines, SANKYOFRONTIER CO.,LTD

Copyright ©2020 Information Processing Society of Japan.
All Rights Reserved.

表1　決定木の精度

	決定木A	決定木B
正解率(%)	53.6	41.2
ノード数	157	6867
属性数	4	7

表2　データセットの属性

決定木A	決定木B
住所,用途,業種,現場区分	引取予定日,出庫日,数量,住所,用途,業種

$$f(x,y,z) = x \times \frac{1}{1+e^{-(-1*y+1000)*0.01}} \times z \qquad (1)$$

正しく分類された事例の偏り度合い z は,各クラスの事例数を C_i，正しく分類できた事例数を C_a として式(2)より算出する．

$$z = \prod \frac{C_a}{C_i} \qquad (2)$$

共同共進化の全体解集団の世代交代は，親の選出をランキング選択，交叉方法を一様交叉，突然変異確率を 1%とした．部分解の世代交代はHSモデルにより行う．

3.　評価実験

提案手法により得られたデータセットAと，人手によって前処理を行ったデータセットBを用いて決定木を作成し，精度を比較する．使用するデータは2011年度から2017年度までのユニットハウスのデータである．2011年度から2015年度までのデータを用いて決定木を作成し，提案手法の適応度算出には 2016 年度のデータを用いる．2017 年度のデータを用いて決定木の精度の比較を行う．提案手法で使用するデータは，先行研究で使用するデータに新しく3属性を追加したものを使用する．提案手法の共同共進化の設定は，全体解の集団サイズは 400，部分解の集団サイズは400，世代交代数は500回とする．

データセットAを用いて生成した決定木を決定木A，データセットBを用いて生成した決定木を決定木Bとし，それぞれの結果を表1に表す．正解率は決定木Bより約12%向上した．ノード数は6867から157まで削減され，可読性が大幅に向上した．それぞれのデータセットで使用した属性を表2に表す．決定木Aで使用された属性は4個で，新しく追加した属性からは，現場区分が選ばれた．カテゴリ化による属性値数の変化を表 3 に示す．カテゴリ化により，属性値が大きいものでも4カテゴリに分類された．提案手法によるカテゴリ化の結果として，属性「住所」の例を表4に示す．7つの県が属性値として存在していたが3つのカテゴリに分けられた．生成された決定木の根ノードの属性は，決定木Bが業種だったが，決定木Aでは現場区分に変わった．

表3　カテゴリ化による属性値数の変化

属性	元データ	処理後
住所	7	3
クライアントの業種	7	4
使用用途	5	2
現場区分	5	4

表4　属性「住所」のカテゴリ化の結果

カテゴリA	カテゴリB	カテゴリC
宮城	東京,神奈川,埼玉	茨城,千葉,愛知

4.　考察

評価実験の結果，提案手法を用いてデータセットを作成することで，人手によって作成されたデータセットよりも正解率，可読性ともに良い決定木を生成することができた．ノード数を大幅に削減することができたため決定木から返却時期を左右する要因を把握することが容易となった．しかし，正解率が実用的とはいえないため改善が必要である．ノード数の減少は，属性数が減ったことと，カテゴリ化によって属性値の種類数が減ったことが主な要因である．正解率が上がっていることから，提案手法によって返却時期に影響を及ぼす属性を選定することができたと考えられる．表4より，提案手法のカテゴリ化によって属性「住所」の属性値は3カテゴリに分類された．地域などでカテゴリ分けされているようにみえるが，カテゴリCでは茨城，千葉と愛知が同じに分類されているため，別の類似点がある可能性も考えられる．機械的にカテゴリ化を行っているため概要の把握が難しい．提案手法では，各個体の適応度の算出に精度測定用データを毎回使用しているため，過適合が懸念されるが，評価実験では比較した決定木Bよりも高い正解率を獲得することができているため汎化性能も十分保たれると考える．

参考文献

[1] 舟久保龍成，土屋直樹，スティーヴェンクレイネス，大谷紀子，″モジュラー建築ユニットの返却時期を予測する決定木の GA による精度向上″，情報処理学会第80回全国大会予稿集，Vol.1，pp.299-300，2018.

[2] Potter，M.A. and De Jong，K.A.，″A cooperative coevolutionary approach to function optimization ″，in Proc. of the 3rd Parallel Problem Solving from Nature，pp.249-257，1994.

Copyright ©2020 Information Processing Society of Japan.
All Rights Reserved.