

決定木学習を用いた学習履歴データからの退学者予兆の試み

Trial Study on Dropouts Detection based on Learning Log Data Using the Decision Tree Learning

松居 辰則^{*1}, 吉田 自由児^{*2}

Tatsunori MATSUI^{*1}, Jiyuuji YOSHIDA^{*2}

^{*1}早稲田大学人間科学学術院

^{*1}Faculty of Human Sciences, Waseda University

^{*2}株式会社デジタル・ナレッジ

^{*2}Digital Knowledge co.,ltd.

Email: Matsui-t@waseda.jp

あらまし：本研究の目的は、学習履歴データからの退学者予兆の可能性を検討することにある。具体的には八洲学園大学（2004年開学の通信制大学）の過去12年分の学習履歴データを対象に、決定木学習を用いて退学傾向にある学習者の特徴の抽出を試みた。その結果、いくつかの興味深い特徴を抽出することができた。本発表においては、1）比較的大規模な学習履歴を構成する属性の中から高い説明能力を有する属性の選定、2）決定木学習を用いた特徴抽出の方法と結果（解釈）を中心に報告を行う。

キーワード：Learning Analytics, 学習履歴データ, 決定木学習, 退学者予兆

1. はじめに

ICT (Information Communication Technology) の進展に伴い、学習・教育の文脈では学習過程を説明可能な多種多様なデータの取得・蓄積（学習履歴データ）、そして大規模データの解析が可能になったことから、Learning Analytics (LA) への期待は大きい。特に、学習評価の観点からは学習者の知識・理解等の認知的側面に関する評価のみならず、学習者の心的状態や興味・関心等の情動的側面の評価の可能性を有することが LA の本質的興味である。従来までも、学習履歴データを用いた学習者の心的状態の推定や分析の研究は行われてきており(1)(2)、その可能性が実証的に示されている。さらに、生体情報を用いた学習者の心的状態の推定も試みられており(3)、今後の進展が期待される。一方、教育機関においては、成績不振者や退学の可能性のある者を早期に検出して適切なフォローを行うことは、教育の質保証の観点からも重要である。そういった問題意識から LA の目的を「ドロップアウト、成績不振者の検出」に焦点化した研究も行われてきている(4)(5)。しかし、教育機関の経年データから「退学者」の予兆を目的とした研究は行われていない。そこで、本研究では、過去12年分の学習履歴データから「退学者予兆」の可能性を検討する。

2. 本研究における「退学」の考え方

本研究においては、「退学」を次のように考える。【「退学」は「本人の意志が固く退学した人」である。しかし、「退学」に至らなくても、学校からの案内に応じて「休学」したり「履修登録をしない」状態が続いたりする人もいる。これら「休学」や「履修登録をしない人」も「退学」と同じ位置付けで扱うものとする。】また、「退学」を厳密に規定するために【退学した人のデータをすべて退学者のデータとして扱うのではなく、在籍期ごとに評価し、退学直前期のみを退学者のデータとして扱う】こととする。

また、直前期履修登録ゼロだった場合などは「かくれ退学」と定義し、学籍上は「在籍」しているものの退学相当とみなす条件を定め、退学同等のデータ処理を行うこととした。

3. 「退学者予兆」の試み。

本研究における「退学者予兆」の基本的な流れは

以下のとおりである。

1. 退学者のデータを分析しそこに非退学者と異なる顕著な行動（退学前兆行動）のモデル化。
2. 在校生のデータを分析し退学前兆行動の有無を検証。退学前兆行動がある者は退学可能性が高いと判断。

また、本研究においては、退学に至る様相を、【何らかの「原因」（経済的理由、興味の低下、体調不良など）により学習に対する「姿勢」（学習に向かう意欲や態度など）が変化し、その結果、「退学」に繋がる「事実」（成績不振、履修科目減少、欠席増加など）が生じるもの】として捉えることとした。対面授業では教師は経験則的に『姿勢の変化』を認識しており、適切な学習支援への結びつけている。すなわち、本研究では、LMS に記録・蓄積された学習履歴・活動履歴といった教育ビッグデータから、教師が認識するような『姿勢の変化』を紡ぎ出し、ノウハウとして顕在化させることが目標となる。なお、「姿勢」の裏には「姿勢」に関連した「行動」があるものと考えられる（例えば、アクセス頻度、授業入退室時刻、質問回数など）。本研究では、これらの【「行動」の変化で「姿勢の変化」を検出可能である】との仮定を設定している。

4. 「退学者予兆」の試み

本研究では、①データ収集、②クレンジング（データ加工）、③分析、④可視化、自動化、の手順で「退学者予兆」を試みた。

4.1 データの収集

八洲学園大学（2004年開学の通信制大学）で運用されている LMS に蓄積された過去12年分（2004年春～2015年春）、延べ人数（1期の在籍を1とカウント）221076名（学生総数：2395名、退学者：430名、卒業生：785名）を分析の対象とした。

4.2 クレンジング（データ加工）

八洲学園大学に蓄積されている成績情報、学生情報等の分散管理された多種多様なデータベースから、八洲学園大学の教職員への「退学者予兆」に関するアンケートやインタビューを通して、データ数が十分であり、かつ「退学」に関係があると判断された以下の23個の属性を選択してレコードを構成した。

さらに、各属性（データ）の基本統計量、「退学（withdraw）」との相関係数から以下の10種類の特

微量を定義した。そして、学生・在籍期ごとに全23項目に整理し検証用データを作成し、各項目に対し10種類の特徴量を算出し、全230パターンからなるデータセットを構成した。

属性	意味
#num	当該期の履修登録数
#season_span	開始期からの経過期数
#new_status	補正後の退学状況
#last_season_span	最終 season_span の番号
#is_last_season	当該期が最終期かどうか
#LIVE.attend_num	該当コマで当該受講者が出席したライブ授業数
#LIVE.attend_rate	該当コマの当該受講者の出席率 (%)
#LIVE.enter_diff_mean	該当コマの当該受講者の入室時刻のズレの平均値 (分)
#LIVE.exit_diff_mean	該当コマの当該受講者の退室時刻のズレの平均値 (分)
#LIVE.regist_num	当該期のライブ授業の履修登録数
#LOGIN.times	該当コマのシステムへのログイン回数
#LOGIN.days	該当コマでシステムにログインした日数
#LOGIN.login_per_day	該当コマの1日あたりのログイン回数
#LOGIN.day_login	該当コマの日次ログイン率
#ENQUETE.Q2_answer	授業内容に対する満足度
#ENQUETE.Q1_1_answer	機能の安定性
#ENQUETE.Q1_2_answer	音声や映像により教室の雰囲気
#ENQUETE.Q1_3_answer	板書エリアに表示される背景教材や教員の板書の有効性
#ENQUETE.Q1_4_answer	チャット発言、理解度確認やクイズの効果的利用
#ENQUETE.Q3_1_answer	授業に「夢中」で取り組めたスキルや知識が「身に付いた」
#ENQUETE.Q3_2_answer	新しい視点を獲得「考えが豊かに」なった
#ENQUETE.Q3_4_answer	実生活で「実践」できそう
#QA.times	該当コマ期間で当該受講者が質問した回数
#QA.span_mean	該当コマ期間で当該質問者の質問から回答までに要した時間の平均値
#QA.times_per_day	該当コマ期間で当該受講者が1日あたりに質問した回数
#BBS.times_per_day	該当コマ期間で当該受講者が1日あたりにBBSに投稿した件数
特徴量	
1. 入力ベクトルの平均値	
2. 入力ベクトルの標準偏差	
3. 入力ベクトルの増減数 (増加)	
4. 入力ベクトルの増減数 (減少)	
5. ゼロ/NAを含む入力ベクトルの要素出現率	
6. 前半と後半の値の変化量	
7. 特異点の数 (箱ひげ図のヒゲ範囲外)	
8. 特異点の数 (範囲拡大: 箱ひげ図のボックス外)	
9. 特異点の数 (範囲拡大・超過: 箱ひげ図のボックス外&オーバー)	
10. 特異点の数 (範囲拡大・減少: 箱ひげ図のボックス外&アンダー)	

4.3 分析

上記のデータセットに対して様々な統計量を算出した (紙面の都合で省略)。その結果、有意な属性を

選出し決定木学習を行った。決定木学習の結果を図1に示す。

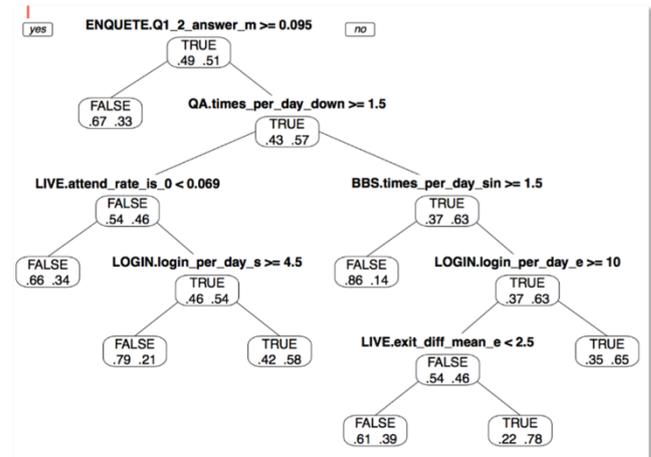


図1 決定木学習の結果 (Rによる)

その結果、退学につながる姿勢の変化と関係性の高い行動が以下のように抽出された。

- (1) 「授業終了後の退室時間、授業の回数ごとに増加もしくは減少する傾向に要注意」(授業終了後も教室にダラダラと残ったり逆にサッサと早く退室する傾向が続くと、「姿勢」が悪化している傾向が見られる)
- (2) 「会議室での発言が突然増減するのは要注意」(普段会議室へ積極的に発言している人の発言数が突如減少したり、普段あまり発言しない人が突如発言数が増加した場合、「姿勢」が悪化している変化が見られる)
- (3) 「質問数の増減には要注意」(質問件数が徐々に増加もしくは減少することが「姿勢」の悪化につながる傾向が見られる)
- (4) 「授業後アンケートの特定質問と退学の相関関係が強い」(授業後毎回収集する授業後アンケートのうち、「姿勢」を大きく反映する有効な項目があった。質問:「音声や映像により教室の雰囲気が十分伝わった」)

5. まとめと今後の課題

本研究では、LMSに記録・蓄積された大規模データから「退学者予兆」の可能性の検証を行った。その結果、いくつかの興味深い特徴を抽出することができた。今回は、分析結果の説明能力が高いこと、データに分布等の前提を必要としないことから決定木学習を用いた。今後も他の分析手法を適用して、「退学者予兆」の自動化、および学習適応化に向けての検討を行う予定である。

参考文献

- (1) 中村, 角所, 村上, 美濃, “e-learningにおける学習者の顔動作観測に基づく主観的難易度の推定”, 電子情報通信学会論文誌 D, Vol. 93, No.5, pp.568-578 (2010)
- (2) 堀口, 小島, 松居, “MRAを用いた学習者の Low-Level Interaction 特徴からの行き詰まりの推定手法”, 第58回先進的学習科学と工学研究会資料(SIG-ALST-A903), pp.1-6(2010)
- (3) 竹花, 松居, “インタラクションに関する多面的データと学習者の心的状態の関係の分析”, 教育システム情報学会, 2016年度第1回研究会資料, pp.1-8 (2016)
- (4) 岩崎, 荻木, 宇佐川, “e-Learningにおけるアクセス履歴解析による学習者の分類とドロップアウト検知”, 教育システム情報学会, 2015年度第6回研究会資料 (2016)
- (5) 長谷川, 新村, 鈴木, 不破, 今井, 小松川, “学習行動の特徴分析による成績不振者の早期発見法の検討”, 教育システム情報学会, 2015年度第6回研究会資料 (2016)