

学部4年生時の 卒業研究について

名古屋工業大学大学院舟橋研究室

福田 太一

自己紹介

生年月日	1996年8月27日愛知県生まれ
所属	名古屋工業大学大学院・博士課程前期1年 舟橋研究室
趣味	料理, サイクリング, TVゲーム, ストレッチ
メールアドレス	f-taitai@i.softbank.jp , guhaabeshi@gmail.com
一言	データサイエンスの知識を深めるべく日々邁進中です. (現在はPythonに力をいれています.)

習得したプログラミング言語

言語	経験年数	習熟度 (5段階)	経験の具体例
R	2年間	★ ★ ★	<ul style="list-style-type: none">ロジスティック回帰主成分分析数理モデルの実装時系列分析 (初歩)
Python	8ヶ月	★ ★ ★	<ul style="list-style-type: none">主成分分析xgboostを用いたGBDT 回帰モデルの実装など

これまでの 研究の成果物

1. ロジスティック回帰を用いた要注意学生の推定
(学部4年次の研究テーマ)
2. COVID-19の流行を予測するモデルの作成
(博士課程前期の研究テーマ)

1.ロジスティック回帰を用いた 要注意学生の推定

問題提起と実践した内容

何が起きているのか

教育現場での要注意学生
(将来留年・退学する学生)
[1] の存在

➡ **教員の負担が増える**

指導量

時間

解決のために何をするのか

学生のデータを
データマイニング

➡
要注意学生の
早期の抽出

➡
教員の負担の軽減

[1]伊藤圭佑:“データマイニングによる『要注意学生』の発見に関する研究”、
平成 25 年度名古屋工業大学修士 論文、2013.

プログラムに用いたデータ (対象: 4年次学生110人)

目的変数 (分類したいデータ)

留年判定 : 学生の留年判定を記録したデータ

(0: 留年している, 1: 留年していない)

説明変数 (分類のために必要なデータ)

GPA: 学生の成績の指標となるデータ

学生生活実態調査: 学生生活に関するアンケート調査の結果

学生生活実態調査のデータ

学生生活実態調査の結果からは
次のものが成績に影響すると仮説を立て、採用した。

- 睡眠データ

平日就寝時間, 平日起床時間, 平日睡眠時間
休日就寝時間, 休日起床時間, 休日睡眠時間

- 住居・通学データ

出身校所在地, 住所, 通学手段, 入構手段, 同居人

データマイニングの手法

これらの手法をプログラミング言語Rで実装.

ソースコード: [bond2580/lo \(github.com\)](https://github.com/bond2580/lo)

- ロジスティック回帰モデル L

$$L = \frac{\exp(\alpha + \sum_{i=1}^n \beta_i x_i)}{1 + \exp(\alpha + \sum_{i=1}^n \beta_i x_i)}$$

x : 説明変数 α, β : 定数

- 主成分分析

$$Z_n = a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n$$

Z_n : 主成分得点, a : 係数, x : 説明変数

要注意学生の推定実験

1. ロジスティック回帰モデルの出力が閾値を超えた場合に要注意学生と推定.
2. 1.を採用データ, モデルへの入力, 変数選択法を変えて繰り返す.

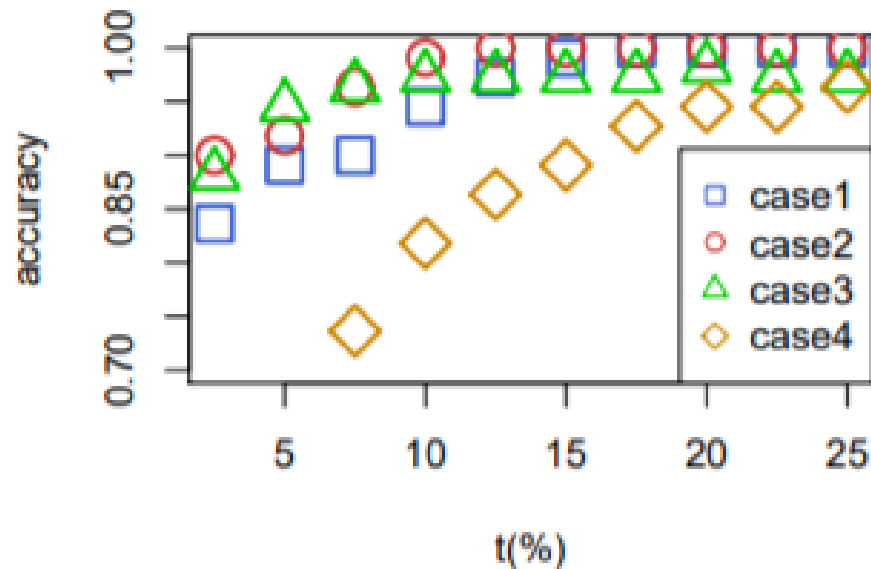
モデルへの入力	<ul style="list-style-type: none">• 実データ• 実データの主成分得点
変数選択	<ul style="list-style-type: none">• 全て投入 (強制投入法)• 徐々に減らして最良化(ステップワイズ法)

採用データ別最良モデル (ここではcaseと呼称する)

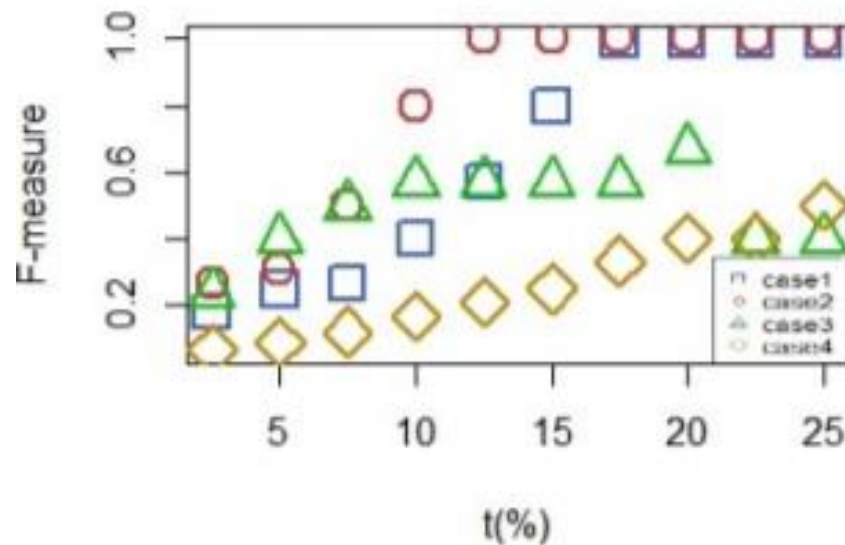
ケース	採用データ	入力	変数選択
case1	GPA, 睡眠 住居・通学	主成分得点	ステップ ワイズ法
case2	GPA, 睡眠	主成分得点	ステップ ワイズ法
case3	GPA 住居・通学	実データ	ステップ ワイズ法
case4	睡眠 住居・通学	実データ	強制 投入法

各ケースの精度の比較 (横軸: 閾値)

モデルの正答率



F値 (モデルの精度の高さ)



case2の高い推定率が確認できた.

各ケースの推定結果のまとめ

- case1がcase2より劣る
住居・通学データは推定においてノイズ
- case2が全体的に推定率が高い
 - GPAと睡眠データは推定に有効
 - 入力は主成分得点, 変数選択はステップワイズ法が有効
- case2がcase3より優れる
推定においてGPA以外のデータも必要

むすび1

- プログラム概要

GPAと学生生活実態調査の結果を用いた
ロジスティック回帰分析による要注意学生の推定

- 結果

GPAと睡眠データによるモデルの高い推定率を確認

- 課題

- 過学習の回避
- 新たな種類のデータの追加
- GPAと睡眠データの関係性の調査



最後に

最後までご覧いただきありがとうございました。