

リアル出席データを使用した GPA の予測の研究

A Study on Predict of GPA Utilizing Real Attending Data

成瀬 健一郎 山田 洋巳
Kenichiro Naruse Hiromi Yamada

〈摘要〉

大学では、高校までの評定平均に代わり、授業毎の学業成績の評価をまとめて数値化したものを GPA (Grade Point Average) と呼び、GPA が高い者は学業成績優秀者として表彰されたり、GPA が低い者は面談が行われ、改善を求められたりするなどの取り扱いが行われている。しかし、学生は、出席に関して高校までのような管理がなされないことが分かると、安易に授業を休むという事が多々見られる。授業を休めば、その部分の学習がされないために、学業成績が悪くなり、学業成績が悪くなれば、GPA が悪くなる事が予測される。

そこで、筆者らは 15 回の授業において出欠席データ及び GPA を使用して、未知の学生の出席状況から GPA の予測及び実際のデータとの比較を行い、予測が可能であることを示した。

〈キーワード〉 出席 GPA 予測 機械学習 回帰分析

I. はじめに

大学教育においては高校までの教育とは異なり、授業への出席については学生本人が責任を負うところが多くなり、自主性を求められるようになる。また、大学においては長期に欠席することを除けば、授業を欠席するからといって連絡が必要になるということはほとんどない。授業を休んだ学生に理由を聞いてみると、いろいろな理由はあるが、企業説明会、就職面接などに参加するためや、インフルエンザなどの感染症に罹患するなどして不可抗力的に授業を欠席する場合もあれば、遊び、アルバイト、体調不良、怠け、寝坊などにより授業を欠席する場合もある。

文部科学省より、大学において授業への出席を成績に加味してはならない^{[1][2]}という方針が示されている。これに従って学業成績をつけていけば、授業の出席と成績との間には直接的な相関はないものの、授業を休めば一定の程度で成績が下がることは容易に推測できる。

そこで、本研究では、学生の授業への出席状況と成績の指標である GPA を研究対象に

設定する。もちろん、多くのパラメータを用意すれば、精緻な予測データを算出する事は可能であろうが、今回は、出席データのみという、少ないパラメータから GPA を予測するという点にポイントを置いている。

II. GPA について

GPA の値は学業成績を数値化（表 1）し、その平均を取ったものである（式 1）。

表 1 評価及びその評価点

評 価	A	B	C	D	F	/
評価点	3	2	1	0	0	0

$$\text{GPA} = \frac{\Sigma(\text{科目評価点} \times \text{科目単位数})}{(\text{履修登録単位数})} \quad (1)$$

なお、成績は、A、B、C は合格、D、F、/ は不合格である。大学によっては、A 以上の成績があり、それを評価点 4 としている場合もある。

また、GPA は、大学内においていろいろと利用されている。本学の例を挙げれば、卒業時に表彰が行われる学業成績優秀学生の候補選出や、日本学生支援機構への推薦が主なものである。他大学においては、GPA の低い学生に対して、面談を行って GPA の改善を求めたり、GPA がある一定以上の数値がないと進級や卒業ができないといった制限をかけたり、退学勧告を行う大学もある。また、その逆に GPA の高い学生に対して、学期における履修単位数の上限が一定の範囲内において緩和されるといった優遇策がとられる大学もある^[3]。

GPA の利活用は学生個人にとどまることなく、現在、文部科学省から私立大学等経常費補助金の交付申請に係る算出根拠や、私立大学等改革総合支援事業のタイプ 1 に係る調査において、補助金の算出に利用されるようになった。私立大学等経常費補助金の交付申請に係る調査項目において、「GPA 制度の導入、活用」として、「成績評価において GPA 制度を導入するとともに、進級判定・卒業判定・退学勧告のいずれかの基準として用いていますか。」の調査がある。また、私立大学等改革総合支援事業のタイプ 1 の申請に係る調査項目においては、「GPA の活用」として、「成績評価において全学部等で GPA 制度を導入するとともに、進級判定、卒業判定、退学勧告のいずれか及び以下のア～ウの基準として用いていますか。ア 履修上限単位数 イ 授業科目履修者に求められる成績水準の設定 ウ 教員間もしくは授業科目間の成績評価基準の平準化」の調査がある。これらの項目に合致するように大学内の整備を行うと、補助金が増額されるなどの優遇策がとられる。これらはどちらも、GPA の利用の促進に係るものであるが、文部科学省がそれぞ

れの学生の GPA の良し悪しや、大学全体の GPA の値に基づいて、大学としていくつかのアクションを求めているものである^{[4][5]}。

Ⅲ. 先行研究

成績の予測に関する先行研究は、速水、長谷川による学業成績の因果帰着について、本人の努力、能力、先生の力、運・偶然の4要因を使って評価したもの^[6]、西村、河村、桜井による学習動機付けと成績に関するプロセスについて論じているもの^[7]、田中、山内による教室における達成動機、目標志向、内発的興味、学業成績の因果モデルについて検討されたもの^[8]など、質問紙調査の結果に因子分析等を使って分析を行い、教育心理学の分野で議論がされているものの、出席状況から GPA を予測する研究は少ないと考えられる。本研究ではそれを実施し、出席状況と GPA の関係について考察する。

Ⅳ. データ予測

数値からデータの予測を行うには、回帰分析が一般的な方法である。但し、本研究の場合、GPA と関連付けてデータの算出は出席数になるため、仮に授業を満遍なく休む、授業の前半ばかりを休む、または後半ばかりを休むということがあっても、同じ出席回数であれば予測結果は同じ GPA となる。これに対して、ニューラルネットワークを用いた機械学習を使った予測を行った場合、1回から15回までの出席データを学習させた結果について予測をさせることができるため、同じ欠席数でも異なった GPA の値が予測することが可能となる。

Ⅴ. 仮説の構築

学業成績には、さまざまな要因が複雑に関係していると推測される。複雑な要因については、数限りなくあること、教科担当者によっても評価のポイントが異なるので、これを比較することは難しい。例えば、知識伝搬型の授業に限って言えば、休んだ部分の単元について十分な自習をしておくことによって、欠席した部分を補うことができる可能性もあるが、その場合、授業を受講するのに比べて何倍もの努力や時間が必要となるであろう。また、知識伝搬型の授業以外の演習、実技などの授業を欠席してしまえば、欠席した部分を自習によって補うことができる可能性は低い。従って、授業を欠席した場合、自習を行い欠席した部分の内容を補っている者を除き、授業の成績は下がることとなる。

従って、出席状況が学業成績に大きく影響を与えていると考えた。そこで、すべての教科で共通の要素である出席情報のみで判定するように単純化をして、出席情報をインプッ

トとし、GPA の予測を行うこととする。予測方法としては、回帰分析、機械学習をつかって、学業成績の予測を行い、検証を行う。

VI. 仮説の検証

出席データと GPA を表 2 の通り 2 つに分割して、1 つを予測構築用データとし、もう 1 つを検証用データとする。GPA は 787 名分のデータを使用した。

表 2 出席データ

予測構築用データ	2725 件	70%
検証用データ	1168 件	30%
合計	3893 件	100%

予測構築用データの出席データの出席回数（表 3）を元に、回帰分析を行った（図 1）。回帰式を 1 次式に求めたものを式 2、2 次式にて求めたものを式 3 に示す。式 2 を検証用データに当てはめて GPA を予測し、実際の GPA との差をヒストグラムにまとめたものを図 2 で示す。式 3 を検証用データに当てはめて GPA を予測し、実際の GPA との差をヒストグラムにまとめたものを図 3 で示す。

$$GPA_1 = 0.1619x + 0.1392 \quad (2)$$

$$GPA_2 = 0.0137x^2 - 0.058x + 0.7456 \quad (3)$$

表 3 データの例

出席数	GPA
11	1.32
15	2.13
...	...

次に機械学習の手法を使用した数値予測を行った。学習用データ（表 4）を使ってコンピュータにニューラルネットワークを用いた機械学習を行わせた後、検証用データを使って、GPA を予測させ、実際の GPA と比較した。出席データを使って機械学習で予測した値と、実際の GPA との差をヒストグラムにまとめたものを図 4 に示す。

表 4 出席データの例

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	GPA
o	o	x	o	o	o	x	x	o	o	o	o	o	o	o	1.32
.

回帰分析

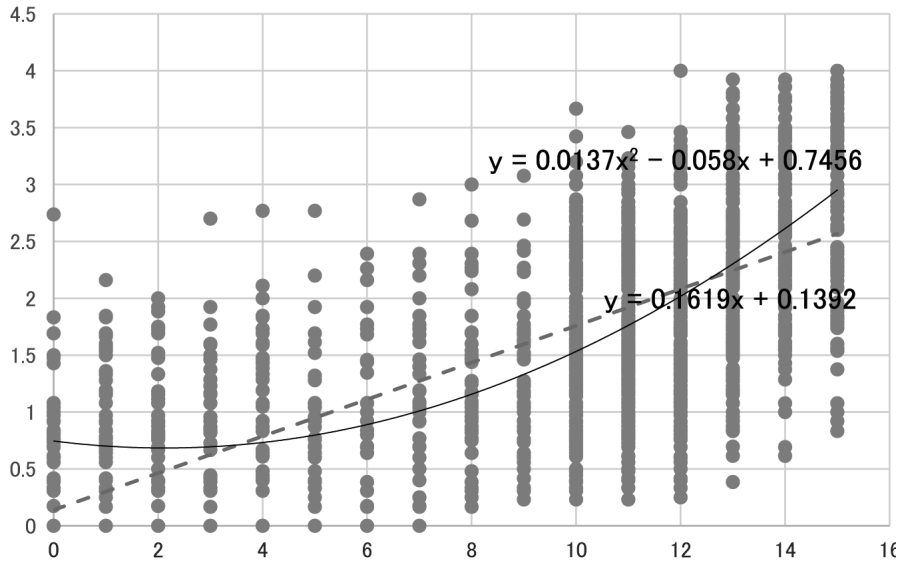


図1 データの散布図と回帰分析

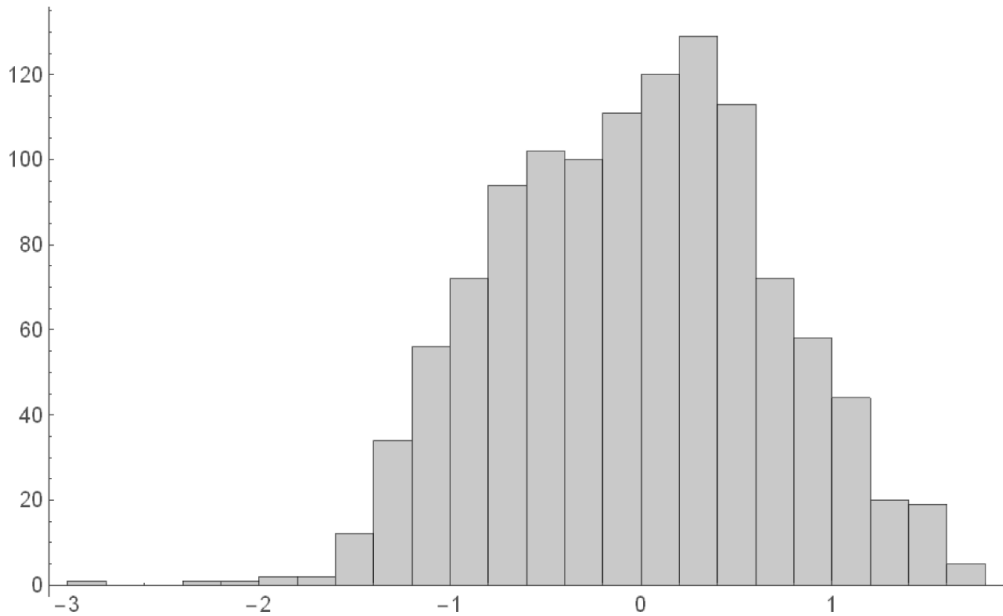


図2 回帰分析 (1次)

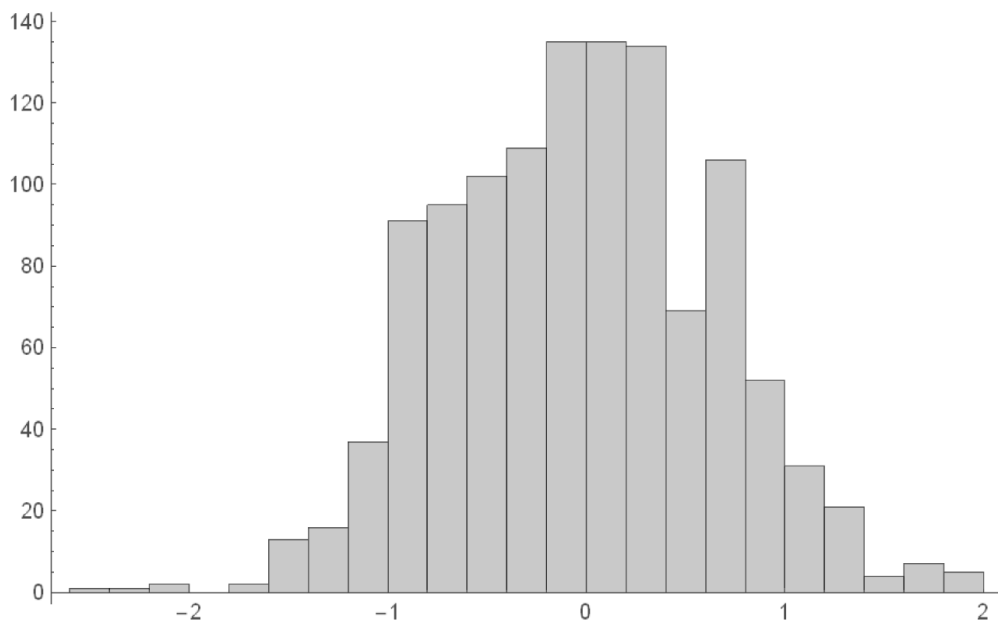


図3 回帰分析 (2次)



図4 機械学習

表5に予測データの評価値として、検証用データ、回帰分析（1次式）、回帰分析（2次式）、機械学習をつかって、GPAの予測を行った際の平均、分散、相関係数を求めた。

表5 予測データの評価値

	検証用データ	回帰分析（1次）	回帰分析（2次）	機械学習
平均	1.9224	1.8875	1.8872	2.0082
分散	0.8487	0.3259	0.3974	0.5120
相関係数	—	0.4075	0.4691	0.4181

表5によれば、平均として、最も近いものは、回帰分析（1次）、分散が最も近いものは、機械学習となった。また、相関係数が最も高いものは回帰分析（2次）となった。

VII. 結論

本研究では、出席データとGPAを使って、回帰分析を行い、1次式の近似式と、2次式の近似式を求め検証用データを使って欠席数に対応したGPAの予測を行った。さらに、出席データをコンピュータに機械学習をさせて求めた結果に検証用データを使ってGPAの予測を行った。そして予測を行ったデータの平均、分散、相関係数を求めた。

最も検証用データに近い予測をしたものは、回帰分析（2次）で予測を行ったものである。但し、これは、2次式で表されたものであるため、出席数とGPAが一对一で対応付くものであり、授業の中のどこで休んでも結果が一緒になる。

VIII. 今後の課題

今回、回帰分析を行い、2次式の近似曲線を使って予測をした数値が、最も実際のデータとの相関が高かった。これは、学習をしたからGPAが高いのか、GPAの高い者は授業を休まないのかという、逆の発想で検証を行う。また、出席データから、学生の動向が分かるので、GPA以外のものについても、予測を行う。また、数年分のデータや、授業科目ごとのデータ、講義、演習、実習など授業形態ごとに分類することにより、より精度の高い予測結果を得ることができるか検証を行う。さらに、欠席状況から、退学予備軍の学生の早期発見に利用することができないか検証を行う。

参考文献

- [1] 文部科学省初等中等教育局教職員課，教職課程認定申請の手引き（31年度開設用），pp. 53.
- [2] 文部科学省高等教育局高等教育企画課大学設置室，設置計画履行状況等調査の結果等について（平成26年度），http://www.mext.go.jp/a_menu/koutou/ninka/icsFiles/afieldfile/2015/03/06/1355057_01.pdf, pp. 14.

- [3] 武蔵野大学, 学びのサポート評価・成績, <https://www.musashino-u.ac.jp/student-life/learning/results.html>.
- [4] 平成 30 年度 教育の質に係る客観的指標調査票, 日本私立学校振興・共済事業団, 私振補第 28 号, 平成 30 年 7 月 31 日.
- [5] 平成 30 年度私立大学等改革総合支援事業に係る調査について (依頼), 文部科学省高等教育局日本私立学校振興・共済事業団, 30 文科高第 356 号, 平成 30 年 7 月 31 日.
- [6] 速水敏彦, 長谷川孝, 学業成績の因果帰着 教育心理学研究第 27 巻第 3 号, 1979, pp. 197-205.
- [7] 西村多久磨, 河村茂雄, 桜井茂雄, 自律的な学習動機付けとメタ認知的方略が学業成績を予測するプロセス—内発的な学習動機づけは学業成績を予測することができるのか?—, 教育心理学研究第 59 巻第 1 号, 2011, pp. 77-87.
- [8] 田中あゆみ, 山内弘継, 教室における達成動機, 目標志向, 内発的興味, 学業成績の因果モデルの検討, 心理学研究第 71 巻第 4 号, 2000, pp. 317-324.