N-011

累積自己情報量に基づく GPA の補正

吉永 悠人 †　山岸 祐己 †§　谷川 邦 ‡　足立 智子 †
† 静岡理工科大学 情報学部 ‡ 静岡理工科大学 学務課 § 理化学研究所 革新知能統合研究センター

1 はじめに

GPA は基本的に段階評価から算出されるが，その評価プロセスは教員の主観的なものであることが多く，評価の分布も様々である．さらに，不可となった成績も計算に含まれるような GPA は，余剰単位の履修が不利益に繋がる可能性が高く，学生の多様な分野への興味関心を抑制している恐れがある．よって，段階評価を客観的評価に変換し，余剰単位の履修が不利に働くことがないような新たな評価指標の提案は重要であると言える．この目的における評価方法としては，段階反応モデルなどの IRT (Item Response Theory) モデルが挙げられるが，小規模なサンプルサイズを考慮したモデルでも最低 300 程度のサンプルが必要とされているため [1]，多様な履修者数が想定される大学の講義などでは適応が困難である．また，ほとんどの教育現場では単純な集計システムや表計算ソフトが用いられているため，各講義のパラメータ推定は，単純な計算のみで即時的かつ一意的に求められることが理想である．提案評価手法は，各講義の成績を累積自己情報量 [2] に変換することによって，どのような評価分布であっても共通の情報量の形で扱えることを特徴とし，基本的に余剰単位は有利に働くように設計されている．また，提案モデルは表計算ソフト上でマクロなどを使用せずに実装することを想定しており，データの追加や更新があっても，即時的かつ一意的に結果を再計算することが可能である．

2 提案手法

複数の講義における成績データセットを

$$\mathcal{D} = \{(c_1, s_{u_1,1}, t_1), \cdots, (c_N, s_{u_N,N}, t_N)\} \tag{1}$$

とする．c_n と u_n と $s_{k,n}$ と t_n は，n 番目に確定した成績の，対象学生 $c_n \in \{1, \cdots, i, \cdots, I\}$ と，対象講義 $u_n \in \{1, \cdots, k, \cdots, K\}$ と，成績 $s_{k,n} \in \{1, \cdots, j, \cdots, J\} = \mathcal{J}$ と，観測時刻 $t_1 \leq \cdots \leq t_n \leq \cdots \leq t_N$ をそれぞれ表す．ここで，成績は例えば，不可・可・良・優・秀や，D・C・B・A・S のような段階評価を考え，それらを順に $\{1, \cdots, j, \cdots, J\}$ のように置き換えるものとする．このとき，段階評価の仕方が複数存在する場合は，講義ごとの $\mathcal{J}_k$ を設定することも可能である．n はタイムステップとし，$\mathcal{N} = \{1, 2, \cdots, N\}$ をタイムステップ集合とする．便宜上，$s_{k,n}$ は J-次元ベクトルダミー変数として

$$s_{k,n,j} = \begin{cases} 1 & \text{if } s_{k,n} = j; \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

のように変換する．このとき，講義 k における m 番目の成績 j の評価が付与される確率 $p_{k,m,j}$ が多項分布に従っていると仮定すると，$p_{k,m,j}$ の最尤推定量は

$$\hat{p}_{k,m,j} = \frac{\sum_{\{n \in \mathcal{N} \mid n \leq m\}} s_{k,n,j}}{\sum_{\{n \in \mathcal{N} \mid n \leq m\}} \sum_{a \in \mathcal{J}} s_{k,n,a}} \tag{3}$$

のように与えられる．さらに，最尤推定量を用いた講義 k の成績確率分布を

$$\hat{\boldsymbol{\theta}}_{k,m} = \{\hat{p}_{k,m,1}, \cdots, \hat{p}_{k,m,j}, \cdots, \hat{p}_{k,m,J}\} \tag{4}$$

とすれば，その累積分布関数

$$F(v; \hat{\boldsymbol{\theta}}_{k,m}) = \sum_{\{j \in \mathcal{J} \mid j \leq v\}} \hat{p}_{k,m,j} \tag{5}$$

を考えることができ，ある学生 i に対して付与された成績 $s_{k,n}$ を，そのときの講義 k の累積自己情報量 $-\log F(s_{k,n}; \hat{\boldsymbol{\theta}}_{k,m})$ として変換することができる．成績 1 が最低評価を意味し，J が最高評価を意味している場合，提案情報量 $-\log F(s_{k,n}; \hat{\boldsymbol{\theta}}_{k,m})$ は，m 番目の成績が確定したときの講義 k にとって，成績 $s_{k,n}$ がいかに珍しく評価が低い（ネガティブ）かを示していることになるため，以降では式 (5) を $F_{neg}(s_{k,n}; \hat{\boldsymbol{\theta}}_{k,m})$ と表す．それに対し，成績 j について逆順の累積分布関数

$$F_{pos}(v; \hat{\boldsymbol{\theta}}_{k,m}) = \sum_{\{j \in \mathcal{J} \mid v \leq j\}} \hat{p}_{k,m,j} \tag{6}$$

を考えれば，同様に成績 $s_{k,n}$ がいかに珍しく評価が高い（ポジティブ）かを示していることになる．ここで，m 番目に確定した成績までに 1 人の学生にしか成績が付与されていない，もしくは複数人の学生に対し同じ成績 j しか付与していない講義 k の提案情報量は 0 になることに注意されたい．また，提案情報量は，その加法性により，時系列データとして累積し続けることも可能であると考えられる．このとき，提案モデルは観測数 N に対して $O(N)$ で全ての学生の時系列データを生成可能であり，データの追加や更新があった場合でも，$O(N)$ でそれらを再計算することが可能である．

Correction of GPA Based on Cumulative Self-Information
†Yuto YOSHINAGA　†§Yuki YAMAGISHI　‡Hiroshi TANIGAWA
†Tomoko ADACHI
†‡Shizuoka Institute of Science and Technology
§RIKEN

3 評価実験とまとめ

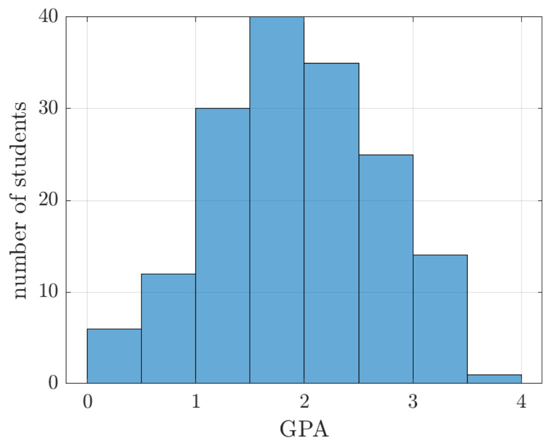
匿名化された学生 $I = 163$ が履修した講義 $K = 186$ における成績 4 年分 $N = 10886$ を評価実験の対象とした．今回は，w 番目までに学生 i に付与された成績 $\mathcal{N}_{i,w} = \{n \in \mathcal{N} \mid c_n = i, n \leq w\}$ におけるポジティブな意味での提案情報量の累積値

$$CP(i, w) = \sum_{m \in \mathcal{N}_{i,w}} -\log F_{pos}(s_{k,m}; \hat{\boldsymbol{\theta}}_{k,m}) \tag{7}$$

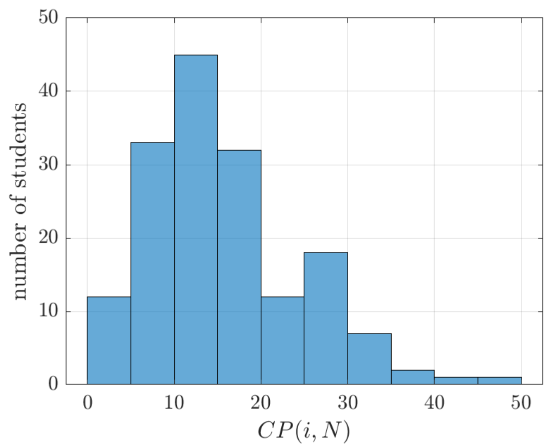
と，ネガティブな意味での提案情報量の累積値

$$CN(i, w) = \sum_{m \in \mathcal{N}_{i,w}} -\log F_{neg}(s_{k,m}; \hat{\boldsymbol{\theta}}_{k,m}) \tag{8}$$

を算出し，その最終値 $CP(i, N), CN(i, N)$ を GPA と比較した．



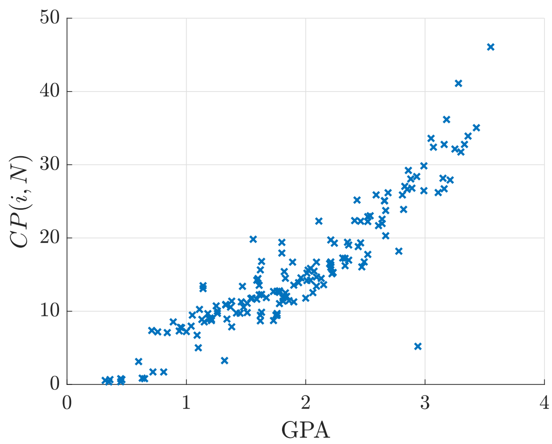
(a) GPA の度数分布



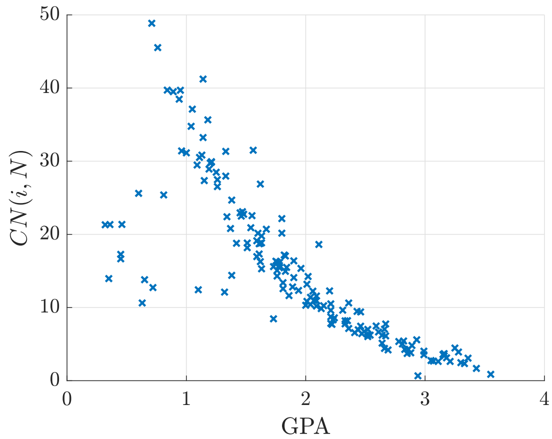
(b) 提案評価値 $CP(i, N)$ の度数分布

図 1: 度数分布による比較

図 1 より，GPA の度数分布と比較して，ポジティブな意味での提案評価値の度数分布は左寄りの正規分布に近い形となっていることが分かる．すなわち，特異に優秀な学生を見分けるという観点では，提案評価値は有用であると言える．また，図 2a より，ポジティブな意味での提案評価値は，GPA の性質を大きく崩しているわけではないことが見て取れる．実際，相関係数は 0.9160 となったが，Kendall の順位相関 [3] では 0.7936 となったため，順位構造は変化していることが分かる．また，GPA とネガティブな意味での提案評価値 $CN(i, N)$ との比較（図 2b ）において，GPA の割に $CN(i, N)$ が高い学生は，サポートが必要な可能性が高いと考えられる．



(a) GPA と提案評価値 $CP(i, N)$



(b) GPA と提案評価値 $CN(i, N)$

図 2: 散布図による比較

参考文献

[1] Shenghai Dai, Thao Thu Vo, Olasunkanmi James Kehinde, Haixia He, Yu Xue, Cihan Demir, and Xiaolin Wang. Performance of polytomous irt models with rating scale data: An investigation over sample size, instrument length, and missing data. *Frontiers in Education*, Vol. 6, , 2021.

[2] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, Vol. 27, No. 3, pp. 379–423, 1948.

[3] M. G. Kendall. *Rank Correlation Methods*. Charles Griffin, London, 1970.