卒業論文

評価の割り当て方や評価者の誤りの量を変えた時の相互評価

寺床 秀太

2022 年 2 月 9 日

岐阜大学 工学部 電気電子・情報工学科 情報コース
鈴木研究室

本論文は岐阜大学工学部に
学士（工学）授与の要件として提出した卒業論文である.

寺床 秀太

指導教員：
鈴木 優 准教授

評価の割り当て方や評価者の誤りの量を変えた時の相互評価*

寺床 秀太

内容梗概

講義を担当する教員が多くの学生のレポート課題を評価することは，大きな負担となっている．本研究では学生同士の相互評価を用いて，レポート課題を評価する際の教員の負担を軽減する手法の提案を行う．本研究の最終目標は，学生の負担を減らしつつ，教員の評価と同等のレポート課題の評価を得ることである．最終目標を達成するには，教員の評価と同等か否かに関係しそうなパラメータを考え，それを変更した時の結果を調べようと考えた．考えたパラメータの 1 つ目は各学生に割り当てる課題の個数，つまり学生の負担量である．この値を最適な値にすることで，最終目標に近づくことができる．2 つ目は各学生の評価の正確さを調べる時に影響を与える，学生同士が共通に見る課題個数である．信頼度の高い学生の評価をより重視することで，学生の負担を増やすことなく教員の評価に近づくことができる．実験では，パラメータを変更した際の結果の変動を観察し，最適なパラメータを明らかにする．また学生同士の相互評価に含まれる誤りの量がどれほど多くても，教員の評価に近い課題全体の評価が得られるか調べる．そのために各学生の評価結果を集計し，課題全体について順位予測を行うシミュレーションを行った．

キーワード

相互評価，順位付け，信頼度

*岐阜大学 工学部 電気電子・情報工学科 情報コース 卒業論文, 学籍番号: 1183033102, 2022 年 2 月 9 日.

目次

図目次

表目次

第 1 章　はじめに

学生同士の相互評価とは，学生が自分以外の学生の課題を評価することである．全学生が課題を提出した後，各学生に匿名の課題を複数割り当てる．各学生は，割り当てられた課題に順位を付ける．ある課題 1 つの最終的な評価は，この課題を割り当てられていた複数の学生による評価を集約する．これに対し課題の評価は，一般的に担当の教員によって行われる．しかし，課題を評価する際の教員の負担は大きい．教員は，課題 1 つ 1 つを閲読するために時間を割き，課題の評価を行う．このとき，教員が担当する講義に 200 人の学生が受講しており，課題 1 つを閲読するのに 3 分かかるとする．すると教員は，全ての課題を閲読するまでに単純計算で 10 時間もの時間を要する．したがって教員には，多大な労力が要求される．この教員が課題を評価する際の負担を減らすために，学生同士の相互評価を用いることを検討する．このとき，学生が課題を評価する負担は，学生の数が多いために分散される．よって教員が課題を評価していたときの負担と比べると，学生 1 人当たりの負担は比較的小さい．また学生同士の相互評価では，1 つの課題は複数の人の目でチェックされることになり，教員が行ったときよりも評価に厚みがある．これらのことから，学生同士の相互評価を用いることは有効であると考えた．しかし学生同士の相互評価を集計することで課題全体を評価した結果は，教員の評価と同等であるかわからない．そこで本研究は，学生同士の相互評価を用いて課題全体の評価を行い，教員の評価と同等の結果を得ることを最終目標とする．

学生同士の相互評価を集約した結果が教員による評価に近づく要因を考えたとき，学生が課題を評価する際の負担量に注目した．学生 1 人あたり割り当てられる課題の個数を増やすと，学生の負担は増える．すると，各学生が課題の優劣を比較する量は増える．よって 1 つの課題はより多くの他の課題と比較されるため，より良い結果が出る．逆に負担を減らすと，得られる課題の優劣の情報量が減り，結果は悪くなる．しかし良い結果を得るために，ただ学生の負担を増やせば良い訳ではない．増やしすぎた負担は，非現実的である可能性がある．また，学生のモチベーションを下げて結果を悪くする原因になる可能性がある．よって最終目標を達成する上での問題は，学生の負担を減らしつつ，良い結果を保つことである．

上の段落で述べた問題を考慮しながら最終目標を達成するため，2 つの方針を考

えた．1 つ目は，最適な学生の負担量を見つけることである．学生の負担量が多い状態では，より良い結果，つまり教員の評価に近い課題全体の評価が得られることが期待される．このとき課題全体の評価は，学生の負担量を徐々に減らしても，教員の評価から遠ざからない，または遠ざかり方が緩やかである可能性がある．このことから学生の負担量を，教員の評価とは乖離しない程度まで減らすことで，学生の負担量を最小限に減らすことと，より良い課題全体の評価の維持することの両立ができると考えた．よって，学生の負担を最適な量に設定することで，最終目標を達成できると考えた．2 つ目は，課題全体の評価をする際，より正確な学生の評価をより重視することである．最終的な課題全体の評価は，各学生の課題に対する評価を集約することで行われる．そして各学生の評価は，学生の評価特性や能力，モチベーションに左右され，良し悪しのばらつきが生じると考えられる．このばらついた評価のうち良い評価，つまり正確さの高い学生の評価を他の学生より重視することで，課題全体の評価はより教員の評価に近づくと考えられる．これを実現させるためには，各学生について，評価の正確さを数値化する必要がある．正確さを数値化するために，どのような学生の評価であるほど正確であるか考えた．その結果，ある 1 人の学生は，この学生と同じ課題を割り当てられていた他の複数の学生と評価が一致しているほど，正確な評価をしているのではないかと考えた．これに基づき各学生の評価の正確さの指標を，学生の信頼度として求めることとした．学生の信頼度は，他の学生の評価との一致度合が高い学生ほど，大きい値になるよう計算した．この学生の信頼度がより正しく計算されることにより，課題全体の評価は，正確さの高い学生の評価がより正当に重視されると考えられる．そして，教員の評価に近づくことが期待される．つまりこの方針の目的は，学生の信頼度をより正しく計算することである．

それぞれの方針に従って最終目標を達成するため，学生同士の相互評価を行う過程において，達成する方針に沿ったパラメータを設定する．そしてパラメータの値を変えることで結果の変動を観察し，最終目標に近づいているか確認すればよいと考えた．そのために提案手法において，パラメータを次のように導入する．1 つ目は，各学生に割り当てる課題の個数である．これは，学生の負担量を意味する．この値を増やすほど，学生はより多くの課題を読むことになり，学生の負担量は増える．その一方，結果はより良いものになることが期待される．このパラメータの値

を増やしていった時の変化に注目することで，学生の負担を減らしつつ，教員の評価に近い結果を得ることができると考えた．2 つ目は，各学生同士が共通に見る課題の最小個数である．このパラメータの意味を説明するため，各学生の順位付けの正確さを計算するときに注目する．このとき，ある学生は他の複数の学生と課題の順位付けが一致しているかどうか調べる．そしてある学生と他の学生が，共通して読んだ課題に注目して計算する．このパラメータは，この課題の個数や，ある学生と一致度合を調べる他の学生の人数を変化させる．つまり，各学生の順位付けの正確さを計算する際影響を与える．各学生の順位付けの正確さは，課題全体についての順位予測をする際利用される．そのうちある 1 つの課題の順位を予測するとき，その課題を割り当てられていた全ての学生に注目する．そして，その各学生が課題に付けた順位を集約する．このとき，計算した順位付けの正確さがより高い学生の付けた順位ほど，集約結果により強い影響を与えるようにする．これにより，課題の順位予測はより正確さの高い学生が付けた順位ほど，より重視される．ここで，このパラメータの値を変えることにより，各学生の順位付けの正確さがより正しく計算される場合が存在することが期待される．そして学生の負担を増やすことなく，教員の評価に近い結果を得ることができると考えた．

さらに 2 つ目の方針に従い各学生の順位付けの正確さをより正しく計算するため，学生の信頼度を繰り返し計算し更新することを検討した．各学生の信頼度は，学生 1 人につき課題に対する順位付け作業について周囲の他の学生と一致度合を計算し，一致度合の集計が大きいほど高い値となる．このとき他の学生の作業には，学生の評価特性や能力，モチベーションなどの要因により正確さのばらつきが生じると考えられる．したがって他の学生との一致度合を比較する際，その学生の順位付けの正確さを加味して計算することを考えた．．このとき一致した他の学生の順位付けの正確さが低いならば，一致度合を意図的に小さく計算するようにした．これにより，不適切な作業との一致度合が高くなっている学生の信頼度を，低く計算できると考えられる．このように学生の信頼度を用いて一致度合を求めることと，一致度合を集計して学生の信頼度を求めることを繰り返すことで，学生の信頼度はより正しく計算されることが期待される．

実験では，シミュレーション上で教員による課題全体の順位付けと，その課題全体に対する提案手法の順位予測を得る．2 つの順位を用いて順位相関係数を計算

し，提案手法を評価するための結果とする．これにより強い正の相関が得られるか調べ，最終目標を達成できているか確認した．このシミュレーションを，各学生に割り当てる課題の個数のみ変える，学生同士の相互評価に含まれる誤りの量のみ変える，各学生同士が共通に見る課題の最小個数のみ変える，という 3 つの種類の実験に分けて繰り返し試行する．各種実験では，変えた値による結果の変動を観察する．これにより，最終目標に最も近づく最適なパラメータを調べた．または，割り当てられた課題に学生同士の相互評価に含まれる誤りの量がどれだけ多く存在しても，良い結果を維持できるか調べた．実験の設定場面は，教員の負担が大きいと考えられる，講義を受けている学生 200 人が全員課題を提出した場合である．このとき，学生 1 人当たり 4 つの課題を評価するよう負担を設定しても，順位相関が 0.8 以上という強い正の相関が得られた．つまり提案手法による課題の順位予測は，相互評価数が比較的小さい値でも，教員の評価に近い可能性があることがわかった．また，シミュレーションにおいて再現した評価が正確ではない学生が 200 人中 30 人ほど存在しても，強い相関が維持できることがわかった．その理由は，1 回更新した学生の信頼度を用いた順位予測が，初期値の信頼度のときよりも正の相関が強くなったためであることがわかった．しかし，学生の信頼度を 2 回以上更新することによる順位相関の変動は，あまり確認できなかった．また，ずらし数を変えることによる順位相関の変動は，あまり確認できなかった．

本研究の貢献は，学生の負担を減らしつつ教員の評価に近い課題全体の評価を得る方針をもとに，学生同士の相互評価を用いる方法を提案したことである．またこの方針に従ったパラメータを学生同士の相互評価を用いる過程に導入し，パラメータを適切に設定することで，教員の評価と同等の結果が得られる可能性を示したことである．

本論文の構成は以下の通りである．2 章では，基本的な事項ついて述べる．3 章では，学生同士の相互評価について関連する研究を紹介する．4 章では，本提案手法について述べる．5 章では，提案手法を用いた実験について述べる．最後に 6 章では，本論文のまとめと今後の課題について述べる．

第 2 章　基本的事項

2.1　スピアマンの順位相関係数

統計学において順位データから求められる相関の指標．チャールズ・スピアマン (Charles Spearman) によって提唱された [1]．集めたデータのうち，ある 2 つの変数の直線的な関係を，数値で記述する．ノンパラメトリックな指標であり，データが正規分布に従っていなくてもよい．尺度水準が比率、間隔尺度、順序尺度のデータを用いることができる．スピアマンの順位相関係数 ρ は-1 から 1 の間の値で，絶対値が大きいほど相関関係は高い．ρ を求めるには，変数の値を順位に変換し，各ペアにおける 2 つの変数の順位の差 D を計算する．そして値のペア数 N を用いて，次のように計算する．

$$\rho = 1 - \frac{6 \sum D^2}{N^3 - N} \tag{2.1.1}$$

これを利用する例として，学生が提出した課題 200 個全体に対する教員の順位付け 1 位から 200 位と，学生同士の相互評価を集約した順位付けとの間の直線的な関係性がどの程度か調べたい場面を挙げる．このとき 2 つの変数は，教員による順位付けと，学生同士の相互評価による順位付けである．そして，各課題に付けられた教員による順位と相互評価による順位の差を計算することで，各課題の順位の差 D を計算する．D と課題の個数 $N = 200$ を上記の式に代入することで，スピアマンの順位相関係数を計算する．この係数が 0.8 以上になっていれば，2 つの順位の間に強い正の相関があることが分かる．これにより学生同士の相互評価による課題全体の評価は，教員による評価に近い結果が出ていることが確認できる．

2.2　ケンドールの順位相関係数

統計学において順位データから求められる相関の指標．1938 年にモーリス・ケンドール (Maurice Kendall) によって開発された [2]．ノンパラメトリックな指標であり，順位相関を計測する別の方法としてスピアマンの順位相関係数がある．どちらの方法を用いても，値が示す相関関係はほぼ同じ傾向がある．それぞれの方法に

より計算したお互いの値には，直接的な関係がない．スピアマンの順位相関係数との違いは定義の差であり，計算方法が異なる．その方法を，n 個の項目に関して 2 つの順位データ $x = (x_1, \cdots, x_n)$ と $y = (y_1, \cdots, y_n)$ がある場合を用いて説明する．i 番目の項目は，変数 x に関しては x_i 位，変数 y に関しては y_i 位であることを意味する．ここで $\{i, j\} \in \binom{n}{2}$ となる全ての項目のペア i, j において，次のことを満たすのペア数 K, L を数える．このとき，n 項目の中から 2 項目 i, j を選んだ際，その順位の大小関係を x と y の 2 つに関して調べる．そして K を 2 つの大小関係が一致するペア数とし，「$x_i > x_j$ かつ $y_i > y_j$」または「$x_i < x_j$ かつ $y_i < y_j$」となる i, j の数とする．また L を 2 つの大小関係が一致しないペア数とし，「$x_i > x_j$ かつ $y_i < y_j$」または「$x_i < x_j$ かつ $y_i > y_j$」となる i, j の数とする．このときケンドールの順位相関係数 τ は，K, L を用いて次のように計算する．

$$\tau = \frac{K - L}{\frac{n(n-1)}{2}} \tag{2.2.1}$$

本研究では，順位相関係数を求める両者の方法がほぼ同じ順位相関の傾向を示すため，スピアマンの順位相関係数を使う．

第 3 章　関連研究

学生同士の相互評価に関連するさまざまな研究が行われている．角田 [3] の研究では，授業支援のための相互レビューシステムを提案している．この研究では，試験的に実現された「旧投票システム」[4] を改良している．記述者情報からのバイアスを避けることができる匿名化や，学生の労力を抑える「グループ化」という機能を継承している．提案手法では，Web 上で学生が課題レポートを提出し，それらを匿名で別の学生にランダムに再配布する．各学生は，配布された複数の他人のレポートの中から良いものと判断した 1 位，2 位を選んで投票する．レポートの評価は，レポートが獲得した票数を得点に換算し合計することで行う．そして Web 上のシステムに蓄積された数年間のデータを調査し，添削する課題を効率よく選択できているか考察している．

石橋 [5] の研究では，学生に評価者の視点を育てるための相互評価法を提案している．学生の評価する個数のレポートのランダムな組み合わせを授業を受けている学生数分作り，各学生に提示する．このとき，互いに顔と名前が一致する関係の中で，相互評価を行っている [6] と，人間関係の質によって評価が大きく左右されてしまう．そこで，匿名性を保証した中で行う．学生は，受け取ったレポートに順位を付ける．各レポートの評価は，レポートの平均得点を計算することで行う．そしてこの方法を組み込んで実践した著者の授業を例にして，学生の判断基準を考察している．

藤原ら [7] の研究では，評価者の割り当て方によって生じる不公平を問題としている．不公平が生じる要因として，学習者間で評価をする時に，お互いに評価しあうかお互いに評価しあわないかの違いに注目している．そして 2 つの場合では，どちらが適切な評価をするか実験している．そして実験結果をもとに，合理的に評価すべき相手を選択できる相互評価支援システムを開発している．

布施ら [8] の研究では，独自の多段階相互評価を提案している．学習者が，複数他者の評価をする相互評価と，その評価内容の確認を行うことは，藤原ら [9] によって有効性が確立されている．提案手法ではこれに加え，確認した相互評価結果に対して，各学習者が妥当性を再評価する．そしてこれを用いた学習を，100 人以上の多人数一斉授業の授業時間外課題として実践している．授業は大人数であるため ICT

を用いて実践しており，多段階相互評価を実装した学習支援システムを使用している．知識定着，意見の明確化，成果物の質的向上をそれぞれ目的としている 3 つの実践結果をもとに，その学習効果について議論している．

本研究が関連研究と異なる点は，学生同士の相互評価を行う際に調節できるパラメータの影響や，パラメータの最適な値について，主眼を置いている点である．提案手法に導入したパラメータを適切に設定することで，学生の負担を減らしつつ，教員が行った課題全体の評価に近い結果を出せると考えた．

第 4 章　提案手法

この手法が想定している場面は，授業や講義を担当している教員が学生に対してレポート課題を課し，課題の提出を求めているときである．提案したシステムは，学生が提出した課題に対して学生同士の相互評価を行わせる．最後に学生の評価を集約し，課題全体についての順位予測を行う．このシステムの入力は，授業や講義を受けている学生と，各学生が提出した課題である．システムではこの入力を用いて，各学生に匿名の課題を複数個割り当てる．またシステムの出力は，学生が提出した課題全体についての順位予測である．

このシステムは，そのまま各学生の成績を出力するわけではない．講義や授業において提出した課題は，学生の成績の決定材料となる要因の中の 1 つにすぎないからである．またこのシステムの出力は，学生が提出した課題を相対的に比較し，優秀である順に順位を付けたものだからである．そこでこの出力の意義を確認するため，教員による課題への一般的な評価方法に注目する．そして提案手法では，これをどの程度まで補助できるのか説明する．

ここで，学生同士の相互評価を用いる対象となる課題の内容は，どのようなものであるか確認する．このとき課題は，正誤問題により単一の得点が出るものではなく，明確な正解がないものや，記述式のものを想定している．次に，教員が全ての課題を閲読する一般的な場合の，各課題に対する教員の評価に注目する．教員は，課題内容に関する採点項目を設け，課題内容に対して各項の達成度合を段階的に評価する．このとき課題の間には，達成度合の差により優劣が生まれる．そして課題全体には，一般的に教員が調べたり学生に公表されたりすることはないような，内部的な順位が存在すると考えられる．

このシステムでは，各学生対して，割り当てた複数の課題の優劣を比較させている．つまり学生は，1 つ 1 つの課題に対して相対評価を行っている．そして最終的な出力は，提出した課題が優秀である順番に沿った，学生の課題の順位である．このとき教員は，提案手法が出力した順位が近い課題の集合に注目することで，教員が考える評価項目に関して似通った達成度合である課題をまとめて閲読することが可能である．また教員が考える評価項目に関して，重点的に閲読したい段階の達成度合である課題の集合を，選択することが可能である．

提案手法は次の手順に従う．

1. 学生に課題を割り当てる．
2. 学生に課題の順位付け作業を行ってもらう．
3. 学生 1 人につき，周囲の学生との順位付けの一致度合を計算する．
4. 手順 3 の結果を用いて，各学生の順位付けの正確さを計算する．
5. 手順 4 の結果を用いて，手順 3 を再び行う．手順 3 と 4 を交互に繰り返すことで，各学生の順位付けの正確さを更新していく．
6. 各学生の順位付けを集計し，課題全体についての順位予測をする．

手順 5 では，各学生の順位付けの正確さを繰り返し計算している．この理由を説明するため，手順 3 に注目する．手順 3 において，一致度合の計算対象である学生は，他の複数の学生ごとに，課題に対する順位付け作業について比較を行っている．そして各学生の順位付けの正確さは，他の各学生との一致度合が大きいほど，高くなるという考えに基づき計算する．しかし学生の作業内容には，学生の評価特性や能力，モチベーションなどの要因により正確さのばらつきが生じると考えられる．このことを考慮するため，手順 3 において他の学生との一致度合を比較する際，他の学生の順位付けの正確さを加味する必要があると考えた．これを実現させるため，他の学生 1 人との一致度合が大きいときについて，2 つの場合に分けて考える．他の学生の正確さが低い場合，他の学生は不適切な作業をした可能性が高い．したがって手順 3 における対象の学生の作業は，不適切な作業との一致度合が高くなっていると言える．そのため他の学生の順位付けの正確さの低さに応じて，一致度合を意図的に低く計算するようにした．逆に他の学生の正確さが高い場合，他の学生の順位付けの正確さの高さに応じて，一致度合を意図的に高く計算するようにした．ここで学生の順位付けの正確さは，初期値のとき全員同じ定数である．しかし初期値の状態で手順 3 を行った後，手順 4 によって更新される．これにより学生の順位付けの正確さは，ばらつきが存在し，計算された状態になる．この状態で手順 3 を行い，再び手順 4 によって更新することを繰り返すことで，より正しく計算されることが期待される．そのため，手順 5 において学生の順位付けの正確さの計算を繰り返す．

4.1　学生への課題割当

提出された学生の課題を，学生へ割り当てる．その様子を，表 4.1 を用いて説明する．表の列名は，課題 A〜H やその他全ての提出された課題である．表の行名は，学生 1〜5 など全ての学生である．表は各学生に，行中に示されている ✓ の付いた課題をそれぞれ割り当てる様子を示す．ここで，学生に自分自身の課題を割り当てないようにする．次に前述した 2 つのパラメータの意味を，表を用いて確認する．

1 つ目は，各学生に割り当てる課題の個数である．今後は，相互評価数と呼ぶ．表中の相互評価数は 4 である．割り当てられた課題の個数は，学生 1〜5 ともに 4 つであることが確認できる．

2 つ目は，各学生同士が共通に見る課題の最小個数である．表中の ✓ は，最上行である学生 1 から順に n 行下がるたび，n 個右へずらすようにしている．今後は，n のことをずらし数と呼ぶ．ずらし数はこの割り当て方法を，パラメータとして数字で表現するために使われる．表中のずらし数は 1 である．ずらし数を 2 に変えた場合を例に挙げ，ずらし数による課題の割り当て方の変化を確認する．学生 1 から 1 行下がった学生 2 は，レ点がずれていない課題 A〜D が割り当てられる．また学生 1 から 2 行下がった学生 3 とその下行の学生 4 は，2 個 ✓ が右にずれた課題 C〜F が割り当てられる．また学生 1 から 4 行下がった学生 5 とその下行の学生 6 は，2×2 個 ✓ が右にずれた課題 E〜H が割り当てられる．

このような割り当てを行うと各課題は，課題 1 つあたり何人の学生によって読まれたか，つまり各課題における他の学生に読まれる人数が 1 人より多くなる．これに反する場合として，ある課題 2 つについて，この 2 つの優劣を比較する学生が 1 人のみであったときを考える．するとこの 2 つの課題の間の順位付けは，この学生 1 人のみに委ねられる．この学生がいい加減な作業を行っていた場合，課題に対する不当な評価がそのまま通ってしまう．しかし提案手法では 1 つ 1 つの課題に対し，1 人以上の学生による複数回の順位付けがなされることで，評価に厚みが増す．これにより，不当な学生による順位付けは相対的に弱まる．そして課題に対する複数の学生による順位付けは，より正確なものになると期待される．

表 4.1 を作ることによって，どの学生にどの課題を割り当てるか決定する．このような表を実際に作成するために，その一般化した手順を説明する．まず，表の列

と行を作成する．表の列名は，提出された全ての課題とする．表の行名は，課題を提出した全ての学生とする．このとき表に記載する列内の課題や，行内の学生は，ランダムな順番とする．

次に，パラメータの値を設定する．相互評価数は，設定者が各学生に割り当てたいと考える，適当な課題の個数とする．ずらし数は，相互評価数未満とする．またずらし数は，相互評価数を割り切れることが望ましい．これを満たしていなければ，各課題における他の学生に読まれる人数が，平等ではなくなるからである．またずらし数は，課題を提出した全ての学生数を割り切れることが望ましい．これを満たしていなければ，ずらし数が相互評価数を割り切れる状態でも，一部の課題に関して他の学生に読まれる人数にばらつきが生じるからである．一部の課題とは，表の左端や右端の列周辺のいくつかの課題が該当する．

各パラメータについて，相互評価数を m，ずらし数を k のように設定した後，各学生が読む課題に対して印を付ける．このとき，m, k を用いて次のように ✓ を表に付ける．ただし，学生に自分自身の課題を割り当てることのないよう注意する．

1. 作成した空白の表のうち最上行を，✓ を付け始める行とする．また表のうち左端の列を，✓ を付け始める列とする．
2. ✓ を左列から順に，相互評価数 m の数だけ付ける．
3. ずらし数 k の回数分，上の手順を 1 行下がりながら繰り返す．
4. ずらし数 k の列数分，✓ を付け始める列を右にずらす．
5. 手順 2 に戻る．最下行が埋まるまで行う．

上記の手順 2 を繰り返していると，✓ が表の右端の列に達する行が発生する．このままでは，この行の ✓ の数は，m 未満になってしまう．これに対応するため，手順 2 に従い表の右端まで ✓ を付けた後，✓ を付ける位置を表の左端の列へ折り返す．そしてこの行の ✓ の数が m になるまで，✓ を左の列から順に付ける．

学生同士の相互評価を行わせるために，実際に学生へ再配布する課題は，各学生の行に示されている ✓ に従う．このとき，k が m を割り切ることができ，かつ k が列名として記載した全ての課題の数を割り切ることができるならば，各課題における他の学生に読まれる人数が等しくなることが保証される．

4.2　学生による課題の順位付け作業

各学生が割り当てられた課題に順位を付ける．これらの課題は匿名となっており，各学生が誰の課題を見ているのかは分からない．付ける順位は，重複がないようにする．例えば 4 つの課題を割り当てられた学生は，最上位 1 位から最下位 4 位までを重複なく各課題につける．

4.3　学生間における順位付けの一致度合の計算

各学生の順位付けの正確さを計算するため，計算対象の学生と周囲の他の学生との一致度合を計算する．本研究では，各学生の順位付けの正確さを数値化した値を，学生の信頼度と呼ぶ．

まず学生の信頼度を計算する理由を説明する．学生の信頼度は，課題全体についての順位予測を行う際に使う．このとき，より信頼度が高い学生が行った課題の順位付けをより重視する．作業内容がより正確である学生の順位付けを重視することで，より良い結果が得られることを期待する．

次に学生の信頼度は，どういう学生であるほど高くなるべきか考えた．その結果，ある 1 人の学生は，他の複数の学生と課題の順位付けがより一致しているほど，正確な順位付けをしているのではないかと考えた．順位付けが一致しているか調べる例として，ある 2 人の学生それぞれの順位付けを比較する場面を挙げる．そして課

表 4.1　学生 1〜5 に課題 A〜H を割り当てる様子

	学生が提出した課題								
	A	B	C	D	E	F	G	H	...
学生 1	✓	✓	✓	✓					
学生 2		✓	✓	✓	✓				
学生 3			✓	✓	✓	✓			
学生 4				✓	✓	✓	✓		
学生 5					✓	✓	✓	✓	
...									

題 A,B は，両者へ共に割り当てられていた全ての課題であったとする．両者共に課題 A は課題 B より優れていると評価していた場合，課題 A,B に関する両者の順位付けは一致していると言える．この場合に反し，一方の学生が課題 B は課題 A より優れていると評価していた場合，一致していない．ここで，一方の学生による課題 A,B の優劣関係の順番は，課題 A,B の順番を 1 回入れ替えることにより他方の学生と一致する．この入れ替え回数は，両者へ共に割り当てられていた課題が 2 個以上であった場合も，課題の優劣の順番のうち隣同士を入れ替えることで，同様に数えられる．このとき，両者共に割り当てられていた全ての課題に対して，両者が異なる順番で良い順位を付けているほど，入れ替え回数は増える．なぜならば，一方の学生による課題の順番を他方に一致させるために，より多くの入れ替え回数を要するからである．そこで，この入れ替え回数に注目し，順位付けの正確さを計算したい学生と他の複数の学生それぞれとを比較する．このとき，正確な順位付けをしている学生であるほど，どの他の学生と入れ替え回数を調べても，0 に近い値になると考えられる．ただし学生のうち，正確な順位付けをする学生が大多数を占めているとする．逆に誤った順位付けをしている学生は，他の学生とは課題に付けた優劣関係が一致せず，入れ替え回数が大きい値になると考えられる．この考えに基づき，他の複数の学生と順位付けが一致している学生であるほど，信頼度がより高くなるよう計算することにした．

学生の信頼度を計算し更新する対象である学生を s_r とする．ここで，s_r の信頼度を求めるために必要となる，s_r と周囲の他の学生との一致度合を計算する手順を説明する．まず s_r が読んだ課題のうち，2 つ以上の課題を読んでいる全ての他の学生に注目する．それらの学生の集合を S とする．このとき，s_r と $s \in S$ となる s がそれぞれ 4.2 節で行った順位付けの不一致度合 $w(s_r, s)$ を求める．そのために，s_r, s が共に読んだ全ての課題に注目する．そして s_r, s それぞれ 4.2 節で行っていた作業をもとに，この課題の順番について比較する．このとき s_r による課題の順番は，順番内で隣同士の課題を何回入れ替えれば，s のものと一致するか数える．この入れ替え回数の最小値を，$w(s_r, s)$ とする．$w(s_r, s)$ の値は，対象の 2 者それぞれの課題に対する順位付けを比較したとき，順位付けが一致しているほど 0 に近くなる．また $w(s_r, s)$ の最大値を，$x(s_r, s)$ と表す．$w(s_r, s) = x(s_r, s)$ となるのは，$w(s_r, s)$ を数えるために対象の 2 者各々による課題の順番を比較したとき，

真逆であった場合である．$x(s_r, s)$ は，s_r と s が共に読んだ全ての課題の個数が多いほど，大きな値になる．そして，$s \in S$ となる s と s_r との順位付けの一致度合 $M_n(s_r, s)$ を求める．$M_n(s_r, s)$ は，w, x と，n 回更新した学生の信頼度 r_n を用いて，次のように計算する．

$$M_n(s_r, s) = \{x(s_r, s) - w(s_r, s)\} \cdot r_n(s) \tag{4.3.1}$$

$M_n(s_r, s)$ は，s_r と s それぞれの順位付けが一致しているほど，大きな値になる．さらに一致度合を調べる相手である s の信頼度が高いほど，大きな値になる．

4.4 学生の信頼度の計算

本節では，学生の信頼度を求める手順を説明する．まず学生の信頼度の初期値は，全員同じ定数 C とする．このとき信頼度は全員同じ値であり，全ての学生において順位付けの正確さに差がないとみなした状態である．そして課題全体の順位予測において，各学生の作業を等しく重視することを意味する．ここで，学生の信頼度を計算し更新する対象である学生を s_r とする．学生の信頼度は，計算を繰り返すことで更新が可能である．そのため s_r の信頼度について，n 回更新された結果を $r_n(s_r)$ と表記する．これに従い，s_r の信頼度の初期値は $r_0(s_r)$ と表記する．$r_n(s_r)$ を更新した s_r の信頼度 $r_{n+1}(s_r)$ を求めるには，s_r が読んだ課題のうち，2 つ以上の課題を読んでいる全ての他の学生に注目する．それらの学生の集合を S とする．このとき，s_r と $s \in S$ となる s がそれぞれ 4.2 節で行った順位付けの一致度合 $M_n(s_r, s)$ を求める．最終的に求める学生 s_r の信頼度 $r_{n+1}(s_r)$ は，$s \in S$ となる全ての s について計算した $M_n(s_r, s)$ の総和とする．

$$r_{n+1}(s_r) = \sum_{s \in S} M_n(s_r, s) \tag{4.4.1}$$

そして，全ての学生について信頼度を計算した後，各自の信頼度は全体で正規化したものに置き換える．

この手順を，表 4.1 の学生 3 の信頼度の例を挙げて説明する．まず最初に，学生 3 が読んだ課題のうち，2 つ以上の課題を読んでいる他の学生に注目する．学生 1,2,4,5 が該当する．そして他の学生ごとに，学生 3 による順位付けとの一致度合を計算する．その計算方法を，他の学生の 1 人である学生 2 の例を挙げて説明する．

学生 2 を s_2，学生 3 を s_3，学生 2 の信頼度を $r_0(s_2)$ と表記する．s_2 は課題 D,E,C，s_3 は課題 C,D,E，のように良い順を付けていたとする．このとき s_2 による順位付けは，隣同士の課題を 2 回入れ替えれば s_3 のものと一致する．この回数が $w(s_2, s_3)$ である．また入れ替えが最大になる場合は，s_2,s_3 共に読んだ課題は 3 個であるので 3 回である．この回数が $x(s_2, s_3)$ である．このとき，一致度合 $M_0(s_2, s_3)$ を次のように計算する．

$$\begin{aligned} M_0(s_2, s_3) &= \{x(s_2, s_3) - w(s_2, s_3)\} \cdot r_0(s_2) \\ &= (3 - 2) \cdot C = C \end{aligned} \tag{4.4.2}$$

学生 3 の信頼度は，学生 1,2,4,5 と計算した一致度合の総和とする．最後に各学生の信頼度は，学生 3 以外の全学生の信頼度も計算した後，学生全体で正規化したものに置き換える．このうち，学生 2 の信頼度 $r_1(s_2)$ が求められる過程にも注目する．$r_1(s_2)$ を求める際，学生 2 が読んだ課題のうち 2 つ以上の課題を読んでいる学生に注目している．該当者は，学生 1,3,4 と，表 4.1 では省略されているもう 1 人の学生である．このもう 1 人の学生とは，課題 A,B,C と表の右端列の課題を割り当てられていた学生である．$r_1(s_2)$ は，s_3 のときと同様，該当する学生ごとに，s_2 による順位付けとの一致度合を計算し，その総和をとったものである．この値を学生の信頼度全体で正規化した結果，$r_1(s_2) = 0.2$ のような，[0,1] の範囲の値が得られる．

4.5 学生の信頼度の更新

学生の信頼度は $s \in S$ となる全ての s について，現在の値が $r_n(s)$ であった場合，節 4.4 のように計算した値 $r_{n+1}(s)$ に置き換える．また，$r_{n+1}(s)$ を用いて節 4.3 のように学生間で一致度合を求めた後，節 4.4 に従って $r_{n+2}(s)$ を再計算することで，再び置き換えることができる．これを繰り返すことで学生の信頼度は，以前の計算結果が加味された状態の信頼度を用いて，何度も更新することが可能である．学生の信頼度の更新によって，各学生の評価の正確さはより正しく数値化されることが期待される．

ここで，節 4.4 の例で得られた各学生の信頼度を用いて，もう一度計算を繰り返

す場合の例を挙げる．これにより，2 回更新した各学生の信頼度を得る．このうち，2 回更新した学生 3 の信頼度 $r_2(s_3)$ を求める過程に注目する．再び，学生 3 が読んだ課題のうち 2 つ以上の課題を読んでいる学生 1,2,4,5 に注目する．そして学生 1,2,4,5 ごとに，学生 3 による順位付けとの一致度合を計算する．このうち，学生 2 と学生 3 との順位付けの一致度合 $M_1(s_2, s_3)$ を計算している例を挙げる．このとき $M_1(s_2, s_3)$ を求めるには，1 回計算し更新された学生 2 の信頼度 $r_1(s_2)$ を用いる．節 4.4 の例で求めた $r_1(s_2) = 0.2$ と，$w(s_2, s_3) = 2, x(s_2, s_3) = 3$ を前の計算同様に用いて，$M_1(s_2, s_3)$ を次のように計算する．

$$\begin{aligned} M_1(s_2, s_3) &= \{x(s_2, s_3) - w(s_2, s_3)\} \cdot r_1(s_2) \\ &= (3 - 2) \cdot 0.2 = 0.2 \end{aligned} \tag{4.5.1}$$

学生 2 に関する一致度合 $M_1(s_2, s_3)$ の計算を学生 1,4,5 についても同様に行い，一致度合の総和を学生 3 の信頼度 $r_2(s_3)$ とする．学生 3 以外の全学生の信頼度を同様に計算した後，信頼度を学生全体で正規化することで，最終的な $r_2(s_3)$ が求められる．$r_3(s_3)$ や，n 回更新した学生 3 の信頼度 $r_n(s_3)$ は，上記のような計算を同様に繰り返すことで求められる．

4.6 課題全体についての順位予測

課題全体についての順位予測をする．順位を予測したい全ての課題のうち，i 番目の課題を a_i とする．a_i を割り当てられた学生の集合を，S_{a_i} とする．$s \in S_{a_i}$ となる s が a_i に付けた順位を $e(s, a)$，s の信頼度を $r(s)$ と表記する．このとき，課題全体における a_i の順位を決める値 $V(a_i)$ を次のように計算する．

$$V(a_i) = \frac{\sum_{s \in S_{a_i}} \{e(s, a_i) \cdot r(s)\}}{\sum_{s \in S_{a_i}} r(s)} \tag{4.6.1}$$

この値 V は S_{a_i} に含まれる学生が課題に対して，高い順位を付けているほど小さい値になる．よって V がより小さい課題ほど，課題全体の中でより良い順位をつける．課題全体を V の値が小さい順番に並び変え，1 位から最下位まで順位を付ける．課題全体の順位には，重複が含まれないようにする．そのために V が等しい課題の順位については，ランダムに並び替えた順番とする．

V を計算する例として，表 4.1 における課題 D の場合を挙げる．課題 D を D と表記し，課題全体における D の順位を決める値 $V(D)$ を求める．D を割り当てられた学生の集合 S_D に注目する．S_D に含まれる全ての要素は，学生 1，学生 2，学生 3，学生 4 である．各学生は，各自割り当てられていた D を含む課題に対して順位付けをしている．この各学生の順位付けにおいて，D が何位に位置しているか注目する．このとき D に対して，学生 1 と学生 2 は 1 位，学生 3 は 2 位，学生 4 は 3 位をそれぞれ付けていたとする．学生の信頼度においてそれぞれ，学生 1 と学生 2 が 1，学生 3 が 0.8，学生 4 が 0.4 であった場合，$V(D)$ は次のように計算される．

$$\begin{aligned} V(D) &= \frac{\sum_{s \in S_D} \{e(s, D) \cdot r(s)\}}{\sum_{s \in S_D} r(s)} \\ &= \frac{1 \cdot 1 + 1 \cdot 1 + 2 \cdot 0.8 + 3 \cdot 0.4}{1 + 1 + 0.8 + 0.4} \\ &= 1.5 \end{aligned} \tag{4.6.2}$$

第 5 章　評価実験

本章では，教員による課題全体の順位付けと，提案手法による順位予測との間の順位相関係数を結果として出し，同等であるか調べる．このとき，パラメータである相互評価数や，ずらし数の値を変える実験を行う．この実験の目的は 2 つある．1 つ目は，相互評価数や，ずらし数と学生の信頼度が結果を変化させるか観察することである．2 つ目は，最終目標に近づくような最適なパラメータの値が存在するか調べることである．また，学生同士の相互評価に含まれる誤りの量を変える実験を行う．そして，相互評価に含まれる誤りの量や学生の信頼度が，結果を変化させるか観察する．これにより，割り当てられた課題に正しくない順位付けをする学生がどれほど多く存在しても，良い結果を保てるのか調べる．

実験は，シミュレーションにより教員による評価と提案手法を再現することで行う．そして，教員による課題全体の順位付けと，提案手法による順位予測を得るまでの流れを実験における 1 回の試行とした．シミュレーションの入力データとして [1,200] の数値を使用する．この値は，200 人の学生が全員課題を提出した後の，教員による課題全体についての順位付けを意味する．この値を正解の順位とする．シミュレーションによる 1 回の試行は，正解の順位を入力として受けた後，各正解の順位に予測順位を付けたものを出力する．実験の評価は，正解の順位と予測順位の間の正の相関が強いかどうか確認することで行う．そのために，正解の順位と予測順位を用いてスピアマンの順位相関係数を計算し，結果を出す．

各パラメータの値を変える実験や，学生同士の相互評価に含まれる誤りの量の実験は，次のように行う．

実験 1　各学生に割り当てる課題の個数に関する実験

目的　相互評価数が結果に変動を与えているか調べる．

実験 2　学生同士の相互評価に含まれる誤りの量に関する実験

目的　割り当てられた課題に正しくない順位付けをする学生がどれだけ存在しても，良い結果を維持できるか調べる．

実験 3　課題の割り当て方法に関する実験

目的　ずらし数が結果に変動を与えているか調べる．

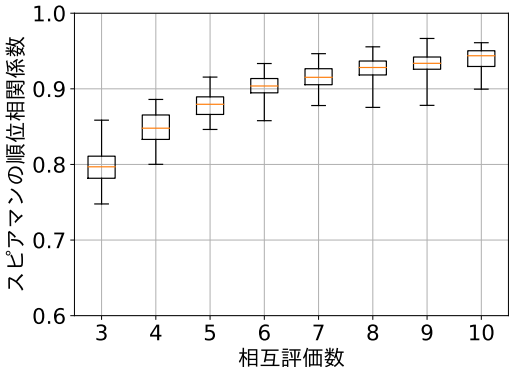

図 5.1　相互評価数ごとの試行結果

5.1　実験 1

この実験では，相互評価数が結果に変動を与えているか調べる．そのために相互評価数の値を変えるごとにシミュレーションを複数回試行する．そして正解の順位と提案手法による順位予測との間の正の相関の強さと，相関の変化を観察する．

5.1.1　実験内容

相互評価数は，[3,10] の範囲で変える．そして，相互評価数の値を変えるごとにシミュレーションによる試行を 100 回する．1 回の試行を終えた後は必ず，正解の順位と提案手法による順位予測を用いてスピアマンの順位相関係数を計算する．全ての試行において，ずらし数と学生同士の相互評価に含まれる誤りの量を固定する．ずらし数は 1 とする．学生同士の相互評価に含まれる誤りは，ないものとする．つまり全ての学生は，割り当てられた課題に対して，正解の順位が高い順に良い順位を付けられるとする．

相互評価数ごとに 100 個得られる順位相関係数は，箱ひげ図として示す．

5.1.2　実験結果

相互評価数ごとの試行結果を図 5.1 に示す．相互評価数を 1 増やしたときの，正の相関の強さの変化に注目する．このとき相互評価数が大きいほど，箱ひげ図が上部へ移動する幅が小さくなっているため，正の相関の強さの向上幅が小さい．また相互評価数が 4 の場合，順位相関係数の平均値が 0.848，悪い場合でも約 0.8 という強い正の相関が得られた．このことから，学生の負担を増やしすぎても，わずかしか教員の評価に近づけないことがわかった．そして相互評価数が比較的小さな値でも，教員の評価に近づけることがわかった．

5.2　実験 2

この実験では，学生同士の相互評価に含まれる誤りの量に注目する．

学生同士の相互評価の中に含まれる誤りの量を段階的に増やし，各段階ごとにシミュレーションを複数回試行する．このとき正解の順位と提案手法による順位予測との間の正の相関の強さは，どれほど維持できるのか観察する．

また学生の信頼度は，学生同士の相互評価の中にどのくらい多くの誤りが含まれていても，結果に良い影響を与えるか調べる．そのために，学生の信頼度が初期値の場合と，計算して更新した場合の結果を比べる．そして，正の相関が強くなっているか調べる．

5.2.1　実験内容

学生同士の相互評価に含まれる誤りの量は，割り当てられた課題に正しくない順位付けをする学生の人数を変えることで設定する．正しくない順位付けをする学生とは，割り当てられた課題に対して，正解の順位が高い順とは異なる順位付けをした学生を意味する．今後そのような学生を，順位付けを誤る学生と呼ぶ．

順位付けを誤る学生を，シミュレーション上で次のように再現した．学生全体のうち，順位付けを誤った学生の集合を S とする．$s \in S$ となる学生 s に割り当てられた課題を，正解の順位通りに順位付けした学生 s' を仮定する．このとき，s の誤りの程度の大きさを，順位付けの入れ替え回数 $w(s, s')$ のように設定する．

学生全体のうち順位付けを誤る学生の人数は，段階的に増えるように設定する．順位付けを誤る学生の人数の範囲は [0,120] とし，8 人ずつ増やす．ここで，順位付けを誤る学生全体の中では，誤りの程度の大きさ $w(s,s')$ が異なる学生がいるとする．順位付けを誤る学生のうち，$w(s,s')$ が 1，$w(s,s')$ が 2，$w(s,s')$ が 3 である学生が，それぞれ 2:1:1 の割合で存在するよう設定する．例として順位付けを誤る学生の人数を，40 人に設定した場合を挙げる．このとき $w(s,s')$ が 1，$w(s,s')$ が 2，$w(s,s')$ が 3 である学生は，それぞれ 20 人，10 人，10 人いるよう設定する．200 人のうち順位付けを誤る学生以外の学生は，全員正しい順位付けを行うものとする．

そして，順位付けを誤る学生数を変えるごとにシミュレーションによる試行を 100 回する．1 回の試行を終えた後は必ず，正解の順位と提案手法による順位予測を用いてスピアマンの順位相関係数を計算する．全ての試行において，ずらし数と学生同士の相互評価に含まれる誤りの量を固定する．相互評価数は 4 とする．ずらし数は 1 とする．

設定した順位付けを誤る学生数ごとに 100 個得られる順位相関係数は，箱ひげ図として示す．また，学生の信頼度を計算することにより，信頼度が初期値の場合と比べてどれだけ順位相関係数が向上したか調べる．シミュレーションによる各試行での理想の順位相関係数とは，順位付けを誤る学生が全員正しい順位付けをしていた場合の結果である．しかし順位付けを誤る学生が存在する場合，順位相関係数は理想よりも下がる．このとき学生の信頼度を計算し更新することで，課題全体の順位予測はより正解の順位に近づくと期待される．そして順位相関係数は，信頼度が初期値の場合よりも向上し，理想に近づくと考えられる．このとき順位相関係数の最大の向上幅は，学生の信頼度が初期値の場合の順位相関係数と理想との差である．各試行において，信頼度が初期値の場合の順位相関係数を r_{s0}，理想の順位相関係数と r_{s0} との差を d，信頼度を計算した後の順位相関係数を r_{s1} とする．このとき，順位相関係数の向上割合 i を次のように計算する．

$$i = \frac{r_{s1} - r_{s0}}{d} \tag{5.2.1}$$

1 回の試行における順位相関係数は，i が 1 のとき理想に一致したことを意味し，i が 0 のとき r_{s0} に一致したことを意味する．また i が負の値のとき，r_{s0} よりも下がったことを意味する．この値も，設定した順位付けを誤る学生数ごとに箱ひげ図

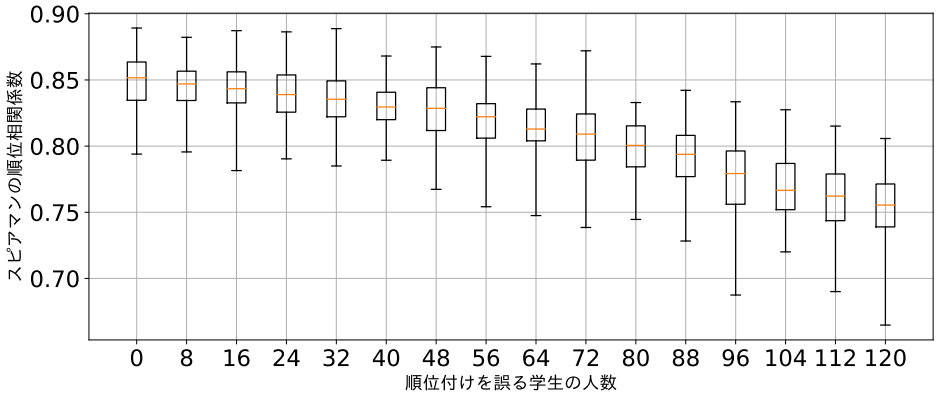

図 5.2　順位付けを誤る学生の人数ごとの試行結果

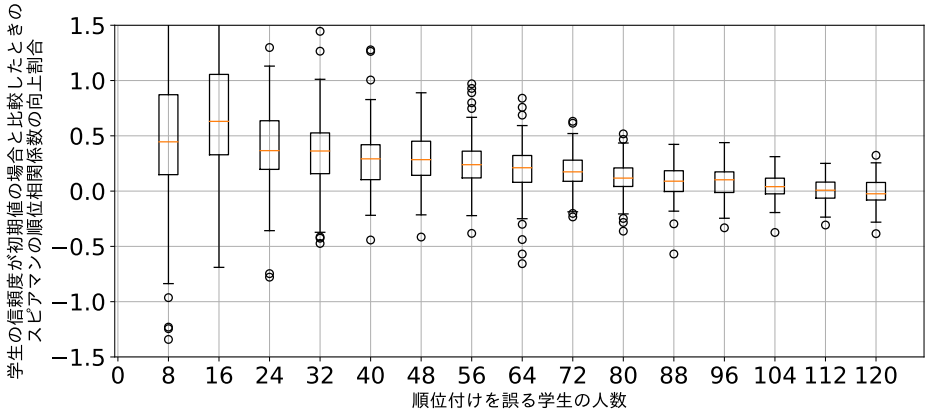

図 5.3　学生の信頼度による試行結果の向上割合

で示す．

5.2.2　実験結果

図 5.2 は各試行において，学生の信頼度を 1 回計算した場合の結果である．各試行において，学生の信頼度を 2 回以上計算し更新すると，1 回更新した場合よりも順位相関係数が低くなる場合が多い．そのため，学生の信頼度を 1 回計算した結果に注目する．また図 5.2 の試行において，順位相関係数の向上割合を計算した結果

を図 5.3 に示す.

順位付けを誤る学生の人数の範囲が [8,16] のあたりでは，全員が正しい順位付けをしている結果と同等の強い正の相関が得られた．信頼度を調べる対象の学生は，他の複数の学生と課題の順位付けが一致しているかどうか調べる．順位付けを誤る学生が少人数である場合，正しい順位付けをする学生と比較される可能性が高くなる．よって順位付けを誤る学生は，他の学生と一致していない可能性が高くなることで，信頼度が低く計算される．これにより課題全体の順位予測は，正しい順位付けをする学生の評価がより重視される．そして，より教員による課題全体の順位付けに近い結果になったと考えられる．図 5.3 において，向上割合が 1 より極端に大きい試行や，0 を下回っている試行が確認できる．これは順位付けを誤る学生が少ないため，理想の順位相関係数と r_{s0} との差 d が非常に小さい場合であると考えられる．このとき，学生の信頼度における初期値とのわずかな違いや，正しい順位付けをした学生の信頼度を低く計算してしまう影響により，たまたま理想を超える順位相関係数が出たり，向上割合が大きな負の値になる試行が出ると考えられる．

順位付けを誤る学生の人数が範囲 [24,32] のあたりでも，強い正の相関が維持されている．その範囲を図 5.3 で確認すると，比較的大きい向上割合を保っているため，学生の信頼度による効果が出ていると考えられる．

順位付けを誤る学生の人数が 56 人を超えると，順位相関係数の下がり方が顕著になっていく．順位付けを誤る学生が増えると，学生の信頼度が初期値の場合の順位相関係数も下がり，理想との差は大きくなる．さらに学生の信頼度を計算するとき，順位付けを誤る学生は正しい順位付けをする学生と比較される可能性が低くなる．この傾向は，順位付けを誤る学生が増えるほど強まる．そのため，図 5.3 において順位付けを誤る学生の人数が 120 に近づくほど，向上割合が小さくなると考えられる．

これらの結果から，順位付けを誤る学生の人数が 10 人程度で非常に少なければ，順位付けを誤る学生がいない場合と同等の結果が得られることがわかった．また順位付けを誤る学生の人数が 30 人程度に増えた場合でも，学生の信頼度を 1 回更新して順位予測をすることで，正の相関がほぼ 0.8 以上のような教員の評価に近い結果を維持できることがわかった．

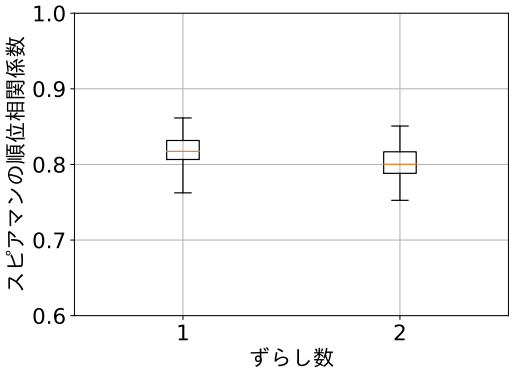

図 5.4　ずらし数ごとの試行結果

5.3　実験 3

この実験では，ずらし数が結果に変動を与えているか調べる．そのためにずらし数の値を変えるごとにシミュレーションを複数回試行する．そして正解の順位と提案手法による順位予測との間の正の相関の強さと，相関の変化を観察する．

5.3.1　実験内容

ずらし数は，[1,2] の範囲で変える．そして，ずらし数の値を変えるごとにシミュレーションによる試行を 100 回する．1 回の試行を終えた後は必ず，正解の順位と提案手法による順位予測を用いてスピアマンの順位相関係数を計算する．全ての試行において，相互評価数と学生同士の相互評価に含まれる誤りの量を固定する．相互評価数は 4 とする．学生同士の相互評価に含まれる誤りの量は，順位付けを誤る学生の人数を決めることで設定する．このとき，5.2.1 節で説明した学生の誤りの程度の大きさ $w(s, s')$ ごとに，次のように人数を決める．$w(s, s')$ が 1 の学生が 30 人，2 の学生が 15 人，3 の学生が 15 人いるとする．その他 140 人は，順位付けを誤らないものとする．

ずらし数ごとに 100 個得られる順位相関係数は，箱ひげ図として示す．

5.3.2　実験結果

図 5.4 は各試行において，学生の信頼度を 1 回計算した場合の結果である．順位相関係数の平均値はずらし数が 1 の場合 0.816，2 の場合 0.801 であった．差は 0.015 であり，ずらし数による結果の変動はほぼ確認できなかった．

また，ずらし数を変えたときの，学生の信頼度が結果に与える変動を調べる．学生の信頼度が初期値の場合に課題全体の順位予測を行った結果と，学生の信頼度を計算した場合を比較する．学生の信頼度を 1 回計算した図 5.4 の結果の場合，学生の信頼度が初期値の場合と比較したときの順位相関係数の平均値の差は，ずらし数が 1,2 どちらの場合も 0.01 未満であり，結果はほぼ変わらなかった．また学生の信頼度を 2 回以降繰り返し計算し更新した場合とも比較した．しかし学生の信頼度を 1 回計算した場合と同様に，ずらし数を 1 か 2 に変えても，順位相関係数の変化はほぼ見られなかった．

このことからずらし数は，学生の信頼度の計算結果に与える影響が小さいと考えられる．学生の信頼度を計算するためには，計算対象の学生と他の複数の学生それぞれについて，共通して読んだ課題に注目する．そして，課題の順位付けが一致しているかどうか調べる．このときずらし数は，共通して読んだ課題の個数や他の学生の人数を変化させ，学生の信頼度の計算結果に影響すると予想していた．しかしこれらの変化と学生の信頼度は，予想よりも関係が薄かったと考えられる．

第 6 章　おわりに

大人数の学生が受けている講義を，教員が担当することは珍しくない．教員は講義において課題を出したとき，提出された課題を評価するために時間をかける．この課題の数が膨大であるとき，教員には多大な時間と労力が要求される．そこで本研究では，学生が提出した課題に対して学生同士の相互評価を行わせ，教員の評価と同等の課題全体に対する評価を得ることを最終目標とした．最終目標を達成するために，相互評価するときの学生の負担量に注目した．学生の負担量として学生 1 人あたりが評価する課題の個数を増やすと，得られる課題の優劣の情報量が増え，相互評価の集約結果は教員の評価に近づくと考えられる．しかし学生の負担を増やしすぎると，学生のモチベーションが下がることにより，結果を悪くしてしまうと考えられる．そのため，学生の負担を減らしつつ，教員の評価と同等の結果を得る必要がある．これを達成するため，2 つの方針を考えた．1 つ目は，学生 1 人あたりが評価する課題の個数を，最適な値に設定することである．これを大きい値にすれば，より教員の評価に近い結果が得られると考えられる．ここで，この値を小さくしていってった場合を考える．このとき得られる課題全体の評価は，教員の評価と比較しても大きく悪化しない可能性があると考えた．2 つ目は，課題全体の評価をするために各学生の評価を集約する際，学生の評価の正確さの高い学生の評価を重視することである．学生の作業には，良し悪しのばらつきが存在すると考えられる．この作業の正確さを数値化するため，学生間における評価の一致度合を計算し，学生の信頼度を求める．この値が大きい学生の評価を重視することで，学生の負担量を増やすことなく最終目標に近づけると考えた．これらの方針に従い，提案手法において 2 つのパラメータを導入した．1 つ目のパラメータは相互評価数である．学生 1 人あたりに割り当てる課題の個数を適切に設定することで，学生の負担を減らした状態で，最終目標を達成できると考えた．2 つ目のパラメータはずらし数である．提案手法では，学生の評価の正確さを計算することで，正確さの高い学生による順位付けを重視した課題全体の順位予測を行っている．学生の評価の正確さを計算する際，正確さの計算対象の学生と他の複数の学生が，共通して読んだ課題に注目する．そして，この課題に対する順位付けの一致度合を調べる．このときの課題の個数や，一致度合を調べる他の学生の人数を変化させることで，各学生の

順位付けの正確さがより正しく計算されると考えた．そして学生の負担を増やすことなく，教員の評価に近づくことができると考えた．

実験では，シミュレーションにより 200 人の学生が課題を提出した場面を想定した．そして，教師による課題の順位付けと提案手法による課題の順位予測を再現した．最後に，2 つの順位を用いてスピアマンの順位相関係数を計算し，強い正の相関が得られるか確認する．これにより，最終目標を達成できているか評価した．この順位相関係数を計算するまでの 1 試行を，調べたいパラメータや変量の値を変えるごとに 100 回行った．相互評価数について調べた実験では，学生 1 人あたり順位をつける課題個数が 4 つであっても，順位相関が 0.8 以上という強い正の相関が確認された．学生同士の相互評価に含まれる誤りの量について調べた実験では，評価が正確ではない学生が 10 人程度で非常に少なければ，全学生が正確な評価をした場合と変わらない結果が得られた．また，評価が正確ではない学生を 30 人程度に増やしても，強い正の相関が得られた．これらの実験により，学生の負担量が比較的少ない設定で相互評価を行わせても，教員の評価に近い課題全体の順位予測が得られることが明らかとなった．また学生の信頼度を 1 回更新して課題全体の順位予測をした結果は，相互評価に誤りが含まれていても，教員の評価に近づくことが明らかとなった．しかし，学生の信頼度を 2 回以上更新した結果や，ずらし数について調べた実験では，順位相関係数の値の変化はあまり見られなかった．

今後は，学生の信頼度の更新式を見直し調整することで，2 回以上更新した信頼度によりさらに強い相関が得られると考えられる．またずらし数を変えても相関の強さに与える影響は小さかったことから，学生へ課題を割り当てる別の方法を試す必要がある．学生の信頼度は，学生同士が共に読んだ課題に注目して計算している．そのため，割り当て方法を変えることで，学生の信頼度がより正しく計算されることが期待される．また本研究の提案手法の中に，教員による一部の課題の評価作業を取り入れた，新たな手法を提案することが有効であると考える．教員の負担を抑えた設定で教員による評価を用いることにより，課題全体の順位予測はさらに教員の評価に近づくと考えられる．

謝辞

本研究を進めるときや，学会での発表，論文の作成時など，多岐にわたる面でアドバイスをくださり，支えてくださった鈴木優先生には，大変深く感謝いたします．先生のご指導がなければ，ここまで研究を完成させることはできませんでした．面倒を見ていただき，本当にありがとうございました．また，研究を進めるうえで非常に多くのアイデアやヒントを与えてくださったり，分からないことがあったらいつでも真摯に答えたり考えたりしてくださった鈴木研究室の先輩方，同期である 4 年生の皆さん，3 年生の方々，本当にありがとうございました．自身の研究でつまづいているところがあっても，研究室のメンバーの皆さんが快く支えてくださったことで，進んでいくことができたと強く感じています．優しいメンバーに恵まれて，大変幸せに思います．そして，日々の研究を続けることを陰で支え応援してくださった両親には，感謝してもしきれません．

参考文献

[1] C Spearman. The proof and measurement of association between two things. *American Journal of Psychology*, Vol. 15, No. 1, pp. 72–101, 1904.

[2] M Kendall. A new measure of rank correlation. *Biometrika*, Vol. 30, No. 1, pp. 81–89, 1938.

[3] 角田篤泰. 授業支援のための投票機能付き匿名相互レビューシステム. 情報処理学会論文誌, Vol. 50, No. 2, pp. 916–924, 2009.

[4] 養老真一. 実習成果に対する評価法―ウェッブを使ったピア・レビュー方式. 法律時報, Vol. 74, No. 3, pp. 39–42, 2002.

[5] 石橋潔. レポート相互評価法―大学における授業実践の試み. 久留米大学文学部紀要 情報社会学科編第 5 号, 2010.

[6] 安彦忠彦. 自己評価と相互評価. 辰野千壽ほか（編）, 教育評価事典. 図書文化社, 2006.

[7] 藤原康宏, 大西仁, 加藤浩. 公平な相互評価のための評価支援システムの開発と評価―学習成果物を相互評価する場合に評価者の選択で生じる「お互い様効果」―. 日本教育工学会論文誌, Vol. 31, No. 2, pp. 125–134, 2007.

[8] 布施泉, 岡部成玄. 多段階相互評価法による学習の実践と効果. 日本教育工学会論文誌, Vol. 33, No. 3, pp. 287–298, 2010.

[9] 藤原康宏, 大西仁, 加藤浩. 継続的な学習者間評価を導入した情報教育の実践. 情報処理学会論文誌, Vol. 49, No. 10, pp. 3428–3438, 2008.

発表リスト

[1] 寺床 秀太，鈴木 優，『学生の相互評価を使った学業成績の推定精度と影響を与える条件の調査』，東海関西データベースワークショップ 2021(DBWS2021)，2021 年 9 月.
[2] 寺床 秀太，鈴木 優，『評価者の負担と評価の正確さを考慮した相互評価方法』，情報処理学会第 84 回全国大会 (IPSJ2022)，2022 年 3 月.