福岡工業大学　令和７年度　卒業研究論文

授業評価の数値に表れない学生の本音
―マルチタスク学習とSHAP分析による満足度要因の解明―

指導教員　佐藤　大輔

福岡工業大学情報工学部
システムマネジメント学科

学籍番号　22M11178
氏名　蘭牟田　晃弘

目 次

図 目 次

表 目 次

第1章　はじめに

1.1　研究背景

1.1.1　高等教育における授業評価の位置づけ

高等教育では教育の質保証が重要な課題であり，学生による授業評価（Student Evaluation of Teaching: SET）は教育改善の基盤として広く実施されている [8, 9]．授業評価は教員へのフィードバックや全学的な教育改善に用いられる．

授業評価アンケートは多段階の評価スコアと自由記述から構成されることが多い．評価スコアは授業間比較や経年変化の把握に適している一方，自由記述は評価理由や具体的な要望を把握できる．

1.1.2　自由記述分析と感情分析の課題

自由記述は非構造データであり，学期ごとに大量に収集されるため，人手による読解だけで全体傾向を把握することは困難である．また，読解者の主観により解釈がばらつく可能性がある．

教育データマイニングや学習分析の文脈では，テキスト分析を含む大規模データの自動分析が進められている [10]．感情分析はテキストに含まれる肯定的・否定的・中立的な感情を推定する技術であり [7]，自由記述を数値化して評価スコアとの関係を検討する手段として有用である．

1.1.3　本研究の対象データ

本研究では，福岡工業大学における 2018 年度から 2023 年度までの 6 年間の授業評価データを分析対象とする．対象データの規模を表 1.1 に示す．

本研究の規模は，統計的分析と機械学習モデルの検討に十分な量である．

表 1.1: 本研究の対象データ

項目	値
対象期間	2018 年度〜2023 年度（6 年間）
対象学科数	9 学科
授業数	3,268 件
自由記述総件数	83,851 件
平均自由記述数/授業	25.2 件

1.2 課題の整理

授業評価の活用には以下の課題がある．第一に，評価スコアは総合判断であり，どの要因がどの程度影響したかを直接把握できない．第二に，自由記述は非構造データであり，83,851 件の記述を人手で一貫して分析することは現実的ではない．第三に，限られた教育改善資源の中で，改善優先度を客観的に判断するための定量的根拠が不足している．

1.3 研究目的

本研究の目的は，授業評価アンケートの自由記述から感情スコアを推定し，授業評価スコアとの関係性を分析することで，授業評価に影響する要因を定量的に特定することである．

具体的には，以下の 3 点を目的とする．

1. **関係性の把握**: 自由記述の感情スコアと授業評価スコアの関係を統計的に検討する．
2. **共通要因と特化要因の分離**: 感情スコアと授業評価スコアを同時に予測するモデルにより，共通要因とタスク特化要因を分離する．
3. **要因の定量化**: SHAP 分析により語彙レベルの寄与度を可視化し，改善の示唆を得る．

1.4 研究仮説

本研究では以下の仮説を設定する．

仮説 1: 授業単位で集約した感情スコアと授業評価スコアには正の相関関係がある．

仮説 2: 感情スコアと授業評価スコアの両方に影響する共通要因が存在する．

仮説 3: マルチタスク学習により，共通要因と特化要因を分離できる．

1.5 研究のアプローチと特徴

本研究では，日本語の事前学習済み BERT を基盤とした感情分類モデルを構築し [1]，83,851 件の自由記述から感情スコアを推定する．さらに，感情スコアと授業評価スコアを同時に予測するマルチタスク学習モデルを構築し [5]，両タスクに共通する特徴と各タスクに固有の特徴の分離を図る．モデル解釈には SHAP 分析を用い [3]，語彙レベルの寄与度を定量化することで，授業改善に直結しうる要因の抽出を目指す．

1.6 研究の意義

本研究の学術的意義は，感情と評価を統合的に扱う分析枠組みを提示し，授業評価における要因構造の理解を深める点にある．マルチタスク学習と SHAP 分析を組み合わせることで，複数指標を同時に解釈可能な形で扱えることを示す．

実践的意義として，大規模自由記述の自動分析により，授業改善の優先順位付けに資する定量的根拠を提供できる点が挙げられる．

1.7 本研究の構成

本研究は全 5 章からなる．第 2 章では関連研究を整理し，本研究の位置づけを明確にする．第 3 章ではデータセット，前処理，モデル構成，評価指標を説明する．第 4 章では実験結果と考察を示す．第 5 章では結論と今後の課題を述べる．

第2章　関連研究

本章では，本研究に関連する先行研究を整理する．授業評価研究，自由記述分析と感情分析，BERT，マルチタスク学習，解釈可能 AI，順序回帰について概観し，本研究の位置づけを明確にする．

2.1　授業評価研究

2.1.1　授業評価の意義と活用

授業評価（SET）は教育の質保証と改善に活用される指標であり，大学教育で広く実施されている [8, 9]．評価スコアは定量比較に適する一方，評価理由の把握には自由記述が重要となる．

2.1.2　信頼性・妥当性

授業評価は多面的な構造を持つことが示されており，信頼性・妥当性に関する議論が蓄積されている [8, 9]．一方で，評価が授業条件や学生側の要因に影響される可能性も指摘されているため，評価結果の解釈には注意が必要である．

2.2　自由記述分析と感情分析

2.2.1　自由記述の特性

自由記述は非構造テキストであり，人手による読解は大規模データでは困難である．教育分野では学習分析・教育データマイニングの文脈でテキスト分析が行われている [10]．

2.2.2 感情分析の基礎と教育応用

感情分析はテキストの感情極性を推定する技術であり，レビュー分析などで広く利用されている [7]．教育分野でも，学習者の自由記述から満足度や不満の傾向を把握する手段として位置づけられる [10].

2.3 感情分析手法の発展

2.3.1 辞書・特徴量ベースの手法

辞書ベースや特徴量設計に依存する手法は実装が容易で解釈しやすいが，文脈依存表現への対応に限界がある [7].

2.3.2 深層学習と事前学習モデル

深層学習は文脈情報を自動的に捉える利点を持つ．Transformer に基づく事前学習モデルの登場により，少量の教師データでも高精度な分類が可能になった [2, 1].

2.4 BERT と事前学習言語モデル

2.4.1 BERT の特徴

BERT は双方向の文脈情報を同時に考慮できる言語モデルであり，Masked Language Model 等の事前学習により汎用表現を獲得する [1].

2.4.2 日本語モデルとドメイン適応

日本語テキストに対しても事前学習済みモデルが利用可能であり，教育分野の自由記述に対してはドメイン適応を意識した微調整が求められる．

2.5　マルチタスク学習

2.5.1　原理と利点

マルチタスク学習は複数タスクを同時に学習し，共通表現を獲得することで性能向上と正則化効果を得る手法である [5].

2.5.2　本研究への適用

感情スコアと授業評価スコアは自由記述に基づく関連タスクであるため，マルチタスク学習により共通要因と特化要因を分離できる可能性がある [5].

2.6　解釈可能 AI と SHAP

2.6.1　XAI の背景

モデルの予測根拠を説明するために解釈可能 AI（XAI）が重視されており，教育分野でも説明可能性は重要である [4].

2.6.2　SHAP の特徴

SHAP は Shapley 値に基づいて特徴量の寄与度を算出する手法であり，局所・大域の両方の説明が可能である [3]．テキスト分類では語彙レベルの寄与度を提示できる．

2.7　順序回帰

2.7.1　順序尺度への適用

授業評価スコアは順序尺度であり，順序性を考慮した回帰手法が求められる [6].

2.7.2　ニューラルネットワークとの統合

ニューラルネットワークと順序回帰を統合する手法により，順序一貫性を保った予測が可能となる [6].

表 2.1: 関連研究の焦点と本研究の位置づけ

研究タイプ	評価スコア分析	自由記述分析	統合分析	要因の定量化
評価スコア中心研究	○	−	−	△
自由記述中心研究	−	○	−	△
統合分析研究	○	○	△	△
本研究	○	○	○	○

2.8 教育分野での分析枠組み

2.8.1 学習分析・教育データマイニング

教育分野では学習ログやアンケートを用いた分析が進展しており，テキスト分析もその重要な要素である [10].

2.8.2 統合分析の課題

評価スコアと自由記述を統合的に扱った研究は多くなく，両者を同時にモデル化する枠組みの整備が課題である.

2.9 本研究の位置づけ

本研究は，BERT による感情分類とマルチタスク学習を組み合わせ，SHAP 分析によって要因を定量化する点に特徴がある．既存研究との位置づけを表 2.1 に示す.

本研究の新規性は，(1) 感情と評価を同時に学習する枠組みの導入，(2) SHAP による語彙寄与度の定量化，(3) 3,268 授業・83,851 件の自由記述に基づく大規模分析，の 3 点にある.

第3章　データと手法

本章では，本研究で使用したデータセットの概要，前処理，教師データ，モデル構成，学習設定，分析手順を述べる．

3.1　データセット

3.1.1　データ概要

本研究では，福岡工業大学の授業評価システムにおける 2018 年度から 2023 年度までの 6 年間のデータを使用した．対象は 9 学科で，授業数は 3,268 件，自由記述総件数は 83,851 件である．データセット概要を表 3.1 に示す．

3.1.2　アンケート構成と尺度

授業評価アンケートは，(1) 択一式質問の点数化による授業評価スコア，(2) 自由記述の 2 種類から構成される．自由記述は以下の 2 問である．

1. 先生に向けてこの授業の感想や学んだこと，意見や要望を記述してください
2. 次期履修者に向けて，この授業についてのアドバイスを記述してください

授業評価スコアは 4 段階（1〜4 点）であり，平均 3.459 点（標準偏差 0.216）である．自由記述は授業単位で複数件存在するため，授業単位で集約して分析する必要がある．

3.2　前処理と教師データ

3.2.1　テキスト前処理

自由記述には以下の前処理を施した．

表 3.1: データセット概要

項目	値
対象期間	2018 年度〜2023 年度（6 年間）
対象学科数	9
授業数	3,268
自由記述総件数	83,851
平均自由記述数/授業	25.2
自由記述の平均文字数	約 41 文字

1. **Unicode 正規化**: 全角・半角の統一と正規化
2. **記号除去**: 絵文字や特殊記号の除去
3. **形態素解析**: MeCab による分かち書き
4. **最大長制限**: BERT の入力長（512 トークン）に合わせて切り詰め

3.2.2 教師データの作成

感情分類モデルの構築のため，自由記述からランダムに 1,000 件を抽出し，ネガティブ（− 1），ニュートラル（0），ポジティブ（+1）の 3 クラスで手動ラベリングを行った．ラベル分布を表 3.2 に示す．

教師データは訓練用 800 件（80%）と検証用 200 件（20%）に層化分割した．

3.3 モデル構成

3.3.1 BERT の概要

BERT は Transformer に基づく事前学習済み言語モデルであり [1]，双方向の文脈情報を同時に考慮できる点が特徴である [2]．本研究では日本語の事前学習済み BERT を用いた．基本構成を表

表 3.2: 教師データのラベル分布（1,000 件）

ラベル	件数	割合
ネガティブ	191	19.1%
ニュートラル	628	62.8%
ポジティブ	180	18.0%
合計	1,000	100.0%

3.3 に示す.

3.3.2 感情分類モデル

感情分類モデルは，BERT エンコーダの [CLS] ベクトルに分類ヘッドを接続し，3 クラスの確率分布を出力する構成とした．損失関数はクロスエントロピー損失であり，クラス不均衡に対応するため重み付けを適用した.

3.3.3 マルチタスク学習モデル

感情スコア予測と授業評価スコア予測を同時に学習するマルチタスク学習モデルを構築した [5]. BERT エンコーダを共有し，感情分類ヘッドと評価スコア予測ヘッドを分岐させる構成とした. アーキテクチャを図 3.1 に示す.

3.3.4 順序回帰モデル

授業評価スコアは順序尺度であるため，順序回帰モデルを検討した [6]．評価スコアが k 以上となる累積確率から，評価段階の確率を算出する.

表 3.3: BERT モデルの基本構成

項目	値
エンコーダ層数	12
隠れ層次元数	768
アテンションヘッド数	12
パラメータ数	約 1.1 億
最大入力トークン数	512

$$P(Y = k) = P(Y \geq k) - P(Y \geq k + 1) \tag{3.1}$$

3.4 学習設定と分析手順

3.4.1 学習設定

モデル学習のハイパーパラメータを表 3.4 に示す．最適化には AdamW を使用し [1]，早期終了を適用した．

3.4.2 授業単位集約と相関分析

感情分類モデルで推定した感情スコアを授業単位で平均し，授業評価スコアとの関係を分析した．授業 j の感情スコア平均 $\bar{S}_j$ は次式で算出する．

$$\bar{S}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} s_{ij} \tag{3.2}$$

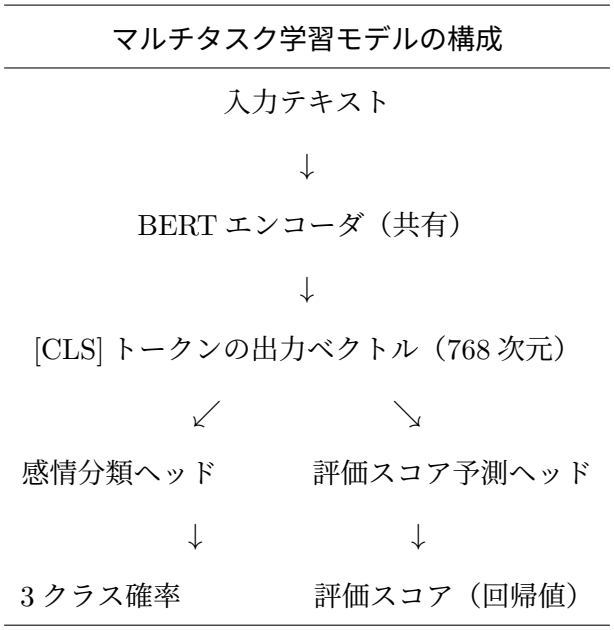

図 3.1: マルチタスク学習モデルのアーキテクチャ

相関係数としてピアソン，スピアマン，ケンドールの 3 指標を算出した．

3.4.3 SHAP 分析

モデル解釈のため，SHAP 分析を実施した [3]．計算負荷を考慮し，層化サンプリングで 5,000 件を抽出した．分析設定を表 3.5 に示す．

マルチタスクモデルの SHAP 分析では，感情重要度と評価重要度の閾値（0.0005）に基づき，共通要因・感情特化要因・評価特化要因・低重要度要因の 4 群に分類した．

3.4.4 評価指標

感情分類モデルの評価には Accuracy，Precision，Recall，F1 を用い，クラス不均衡を考慮してマクロ平均と重み付き平均を併記した．授業評価スコア予測では R^2，RMSE，MAE を使用した．

3.5 分析フローと実装環境

3.5.1 分析フロー

本研究の分析フローを図 3.2 に示す．

表 3.4: 学習のハイパーパラメータ

パラメータ	値	選定理由
バッチサイズ	16	GPU メモリ制約を考慮
学習率	5×10^{-6}	微調整に適した低学習率
エポック数	5	早期終了で過学習を抑制
最大トークン長	512	BERT の最大入力長
ドロップアウト率	0.1	過学習抑制の標準値

表 3.5: SHAP 分析の設定

項目	値
分析サンプル数	5,000 件
サンプリング手法	層化サンプリング
最小出現回数閾値	5 回
分析対象語彙数（単一タスク）	1,564 語
分析対象語彙数（マルチタスク）	3,198 語

3.5.2 実装環境

本研究の実装環境を表 3.6 に示す.

本章では，データセットの概要，前処理，モデル構成，学習設定，分析手順を説明した．次章では結果を報告する.

表 3.6: 実装環境

項目	内容
プログラミング言語	Python 3.10
深層学習フレームワーク	PyTorch 2.0
Transformers ライブラリ	Hugging Face Transformers 4.30
SHAP 分析ライブラリ	SHAP 0.42
統計分析ライブラリ	SciPy 1.11

【データ収集】 授業評価アンケート（3,268 授業，83,851 件自由記述）
↓
【前処理】 テキスト正規化，トークナイザによる分割
↓
【教師データ作成】 1,000 件の手動ラベリング（3 クラス）
↓
【モデル構築】 感情分類モデル（BERT + 分類ヘッド） マルチタスク学習モデル（BERT + 2 ヘッド）
↓
【感情スコア推定】 全自由記述に対する感情スコア推定
↓
【授業単位集約】 授業ごとの感情スコア平均を算出
↓
【相関分析】 感情スコアと授業評価スコアの相関を検証
↓
【SHAP 分析】 5,000 件サンプル（単一タスク: 1,564 語） マルチタスク: 3,198 語，4 グループ分類

図 3.2: 分析フローの概略

第4章　結果と考察

本章では，基礎統計量，感情分類モデルの性能，相関分析結果，SHAP 分析による要因抽出結果を示し，得られた知見を考察する．

4.1　基礎統計量

4.1.1　感情スコアと授業評価スコアの分布

感情スコアと授業評価スコアの基本統計量を表 4.1 に示す．感情スコアは授業単位で集約した値であり，－ 1（ネガティブ）から+1（ポジティブ）の範囲をとる．

感情スコアは平均 0.001 でほぼニュートラルに近く，授業評価スコアは平均 3.459 点である．

4.1.2　教師データのラベル分布

教師データのラベル分布を表 4.2 に示す．ニュートラルが 62.8%と大半を占め，ネガティブとポジティブは少数である．

4.2　感情分類モデルの性能

4.2.1　全体性能

感情分類モデルの性能指標を表 4.3 に示す．検証データ 200 件に対して正解率 0.770，マクロ平均 F1 スコア 0.706 を達成した．

4.2.2　クラス別傾向

クラス別の性能指標を表 4.4 に示す．ニュートラルが最も高い性能を示し，ネガティブとポジティブは相対的に低い．

表 4.1: 感情スコアと授業評価スコアの基本統計量

統計量	感情スコア	授業評価スコア
平均	0.001	3.459
標準偏差	0.260	0.216
最小値	− 1.000	2.000
第 1 四分位数（Q1）	− 0.167	3.330
中央値（Q2）	0.000	3.480
第 3 四分位数（Q3）	0.167	3.600
最大値	1.000	4.000

4.3 感情スコアと授業評価スコアの相関分析

4.3.1 相関分析の結果

授業単位で集約した感情スコアと授業評価スコアの相関分析結果を表 4.5 に示す（N=3,268）.

4.3.2 散布図と解釈

感情スコアと授業評価スコアの散布図を図 4.1 に示す．中程度の正の相関が確認され，感情スコアは評価スコアと関連するが同一概念ではない可能性が示唆される.

4.4 SHAP 分析

4.4.1 単一タスクモデルの重要語

感情分類モデルに対する SHAP 分析を行い，ポジティブ判定とネガティブ判定に寄与する重要語を抽出した．上位 10 語を表 4.6 および表 4.7 に示す.

重要語はサブワード断片を含むため，語彙の解釈は傾向として整理する必要がある．理解容易性や興味・関心に関わる表現が上位に含まれる傾向が確認される.

表 4.2: 教師データのラベル分布（1,000 件）

ラベル	件数	割合
ネガティブ	191	19.1%
ニュートラル	628	62.8%
ポジティブ	180	18.0%
合計	1,000	100.0%

4.4.2　マルチタスク学習の要因タイプ

マルチタスクモデルの SHAP 分析により，語彙を 4 つの要因タイプに分類した結果を表 4.8 に示す．

共通要因の重要語例（TOP5）を表 4.9 に示す．

共通要因は感情と評価の双方に寄与する語彙群であり，改善の優先順位付けに活用できる可能性がある．

4.5　総合考察

4.5.1　仮説検証

仮説 1 は，相関分析により中程度の正の相関が確認されたため支持された．仮説 2 は，共通要因が抽出されたことから支持された．仮説 3 は，共通要因と特化要因の分離が可能であったことから支持された．

表 4.3: 感情分類モデルの性能指標（検証データ 200 件）

指標	値
正解率（Accuracy）	0.770
マクロ平均適合率（Precision）	0.707
マクロ平均再現率（Recall）	0.705
マクロ平均 F1 スコア	0.706
重み付き平均 F1 スコア	0.770

表 4.4: クラス別の性能指標

クラス	適合率	再現率	F1 スコア	サポート
ネガティブ	0.659	0.675	0.667	40
ニュートラル	0.833	0.833	0.833	132
ポジティブ	0.630	0.607	0.618	28

4.5.2 主要な示唆

理解容易性や興味・関心に関わる語彙が重要語として上位に現れる傾向は，授業評価研究の知見と整合的である [8, 9]．また，共通要因の割合が 18.0%にとどまる点は，感情スコアと評価スコアが部分的に重複しつつも異なる側面を捉えている可能性を示唆する．本結果は SHAP によって語彙寄与度を定量化した点に特徴がある [3].

表 4.5: 感情スコアと授業評価スコアの相関分析結果（N=3,268）

指標	相関係数	p 値	解釈
ピアソン相関係数	0.3097	< 0.000001	中程度の正の相関
スピアマン順位相関係数	0.2970	< 0.000001	中程度の正の相関
ケンドール順位相関係数	0.2042	< 0.000001	弱〜中程度の正の相関

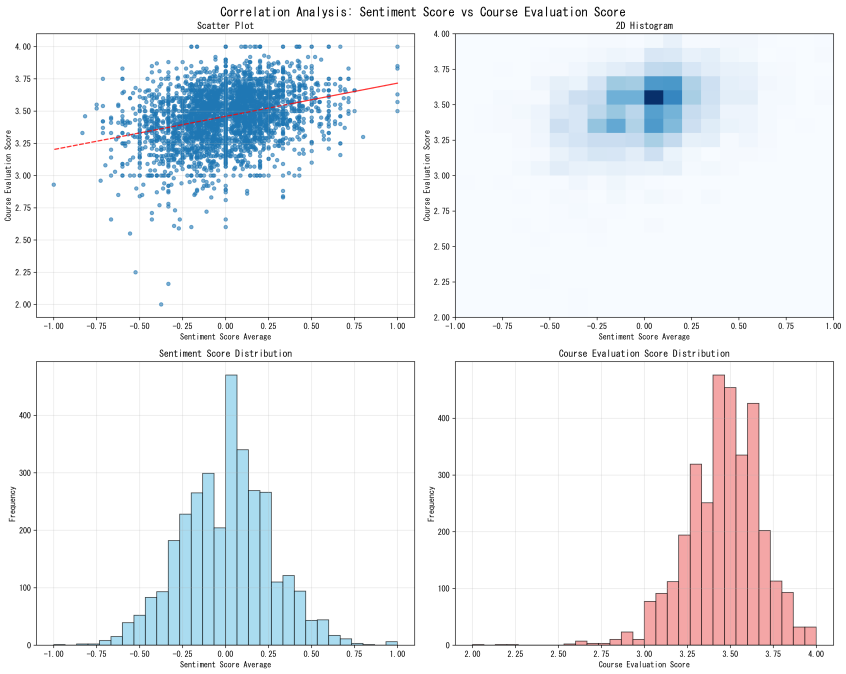

図 4.1: 感情スコアと授業評価スコアの散布図（N=3,268）

表 4.6: ポジティブ判定に寄与する重要語例（TOP10）

順位	単語	平均 SHAP 値	出現回数
1	やす	0.2660	337
2	良かっ	0.2466	207
3	おもしろ	0.2438	10
4	よかっ	0.2251	195
5	面白	0.2178	100
6	楽しい	0.1959	67
7	楽しめる	0.1876	6
8	ありが	0.1760	19
9	楽し	0.1642	192
10	面白い	0.1518	37

表 4.7: ネガティブ判定に寄与する重要語例（TOP10）

順位	単語	平均 SHAP 値	出現回数
1	ほし	− 0.0443	5
2	ほう	− 0.0425	98
3	大	− 0.0346	86
4	まじ	− 0.0314	5
5	難しかっ	− 0.0311	88
6	直す	− 0.0264	6
7	ほしい	− 0.0263	45
8	欲しい	− 0.0247	33
9	奥	− 0.0219	8
10	器具	− 0.0211	7

表 4.8: 語彙の要因タイプ別内訳（3,198 語）

要因タイプ	語彙数	割合	特徴
共通要因（満足度）	577	18.0%	両スコアに寄与
感情特化要因	1,200	37.5%	感情スコアのみに寄与
評価特化要因	532	16.6%	評価スコアのみに寄与
低重要度要因	889	27.8%	両スコアへの影響小
合計	3,198	100.0%	—

表 4.9: 共通要因（満足度要因）の重要語例（TOP5）

順位	単語	感情重要度	評価重要度
1	学ぶ	0.001278	0.001386
2	理解	0.001073	0.000833
3	総括	0.000974	0.000952
4	推奨	0.001132	0.000755
5	人数	0.001195	0.000704

第5章　おわりに

本章では，本研究の成果を総括し，研究の限界と今後の課題を述べる．

5.1　まとめ

5.1.1　研究目的と方法

本研究は，授業評価アンケートの自由記述から感情スコアを推定し，授業評価スコアとの関係を分析することで，授業評価に影響する要因を定量的に特定することを目的とした．BERT による感情分類とマルチタスク学習を組み合わせ，SHAP 分析により語彙寄与度を可視化した．

5.1.2　主な成果

授業単位で集約した感情スコアと授業評価スコアの相関分析では，ピアソン 0.3097，スピアマン 0.2970，ケンドール 0.2042 の正の相関が得られた．感情分類モデルは検証データ 200 件で正解率 0.770，マクロ平均 F1 スコア 0.706 を示した．マルチタスクモデルの SHAP 分析により，3,198 語を共通要因 18.0%，感情特化要因 37.5%，評価特化要因 16.6%，低重要度要因 27.8%に分類した．

5.2　研究の限界

5.2.1　データと一般化

本研究のデータは単一大学（2018 年度〜2023 年度）に限定されており，他大学や異なる教育環境への一般化可能性は検証されていない．また，本研究は相関・寄与度に基づく分析であり，因果関係の検証は行っていない．

表 5.1: 研究の限界と対応の整理

観点	内容	影響	対応の方向性
因果関係	相関・寄与度に基づく分析に限定	因果的主張は困難	介入研究・準実験デザイン
一般化	単一大学・2018–2023 年度に限定	外的妥当性が限定的	複数大学の比較分析
教師データ	1,000 件・不均衡	少数クラスの推定が不安定	追加ラベリング・複数評価者
解釈性	サブワード単位の寄与度	文脈解釈の難しさ	フレーズ単位の再分析

5.2.2 モデルと解釈

教師データは 1,000 件であり，単一評価者によるラベリングとクラス不均衡が存在する．さらに，SHAP 分析はサブワード単位であり，語彙の文脈的意味の解釈には限界がある．

研究の限界を表 5.1 に整理する．

5.3 今後の課題

5.3.1 検証研究の拡充

共通要因に基づく改善施策が実際に評価向上に寄与するかを確認するため，介入研究や準実験デザインが必要である．また，複数大学の比較や学部・学科別の分析により，知見の一般性を検証する必要がある．

5.3.2 実装・運用の発展

教師データの拡充やドメイン適応により感情分類精度の向上が期待される．順序回帰モデルの本格導入によって評価段階ごとの要因分析を深化できる可能性がある．さらに，結果を教育現場で活用するため，分析結果を提示するダッシュボード等の設計が課題である．

本研究は，授業評価の自由記述を感情分析し，評価スコアとの関係を定量的に示した．得られた知見は，データに基づく教育改善の検討に資する基盤となる．

付 録A　感情ラベル定義と例

本研究で用いた感情ラベルの定義と代表例を表 A.1 に示す．ラベル付けは自由記述の感情的な表現に基づき，人手で分類した．

表 A.1: 感情ラベルの定義

ラベル	感情スコア	定義	例
ネガティブ	-1	不満・批判・否定的な感情を含む記述	この授業はつまらない
ニュートラル	0	事実記述で感情が明確でない記述	板書が多い
ポジティブ	+1	満足・肯定的な感情を含む記述	とても楽しい授業だった

付 録B　データセット詳細

本研究で使用したデータセットの概要を表 B.1 に示す.

教師データのラベル分布を表 B.2 に示す.

表 B.1: データセット概要

項目	値
対象期間	2018 年度〜2023 年度
対象学科数	9
授業数	3,268
自由記述総件数	83,851
平均自由記述数/授業	25.2

表 B.2: 教師データのラベル分布（1,000 件）

ラベル	件数
ネガティブ	191
ニュートラル	628
ポジティブ	180

謝辞

本卒業研究を無事に完了できたのは，多くの方々のご指導とご支援のおかげであり，心より感謝申し上げます．

まず，日々ご指導いただいた髙橋先生に深く感謝いたします．研究テーマの選定から論文作成に至るまで，親身にご指導いただきました．特に初期段階では，適切な助言をいただき研究の方向性を定めることができました．また，専門的な助言だけでなく，私の成長を促す課題を与えてくださり，それを乗り越える中で知識やスキルを向上させることができました．先生のご指導なくして本研究の成果は得られなかったと確信しております．

また，研究室の仲間にも感謝いたします．日々の意見交換や協力を通じて，多くの刺激を受け，視野を広げることができました．互いに研究の進捗を共有し，課題を乗り越える中で得た経験は，大学生活における貴重な財産となりました．研究活動以外でも，共に過ごした時間はかけがえのない思い出となりました．

さらに，家族の支えなしには研究に専念することはできませんでした．特に両親には，学費や生活面での支援だけでなく，精神的な支えをいただきました．日々の何気ない会話や励ましの言葉が心の支えとなり，困難に直面した際にも家族の存在が大きな励みとなりました．改めて，家族の支えがどれほど大きな力となっていたかを実感しております．

最後に，大学生活を通じて関わったすべての方々に心より感謝申し上げます．先生方や職員の皆様，そして友人たちの支えがあったからこそ，充実した学生生活を送り，多くの学びと経験を得ることができました．これからも学び続け，支えてくださった皆様への感謝を忘れずに精進してまいります．

改めて，この場を借りて深く御礼申し上げます．

参考文献

[1] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K.: "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pp. 4171–4186 (2019).

[2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I.: "Attention Is All You Need," *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)* (2017).

[3] Lundberg, S. M., and Lee, S.-I.: "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)* (2017).

[4] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., et al.: "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI," *Information Fusion*, Vol. 58, pp. 82–115 (2020).

[5] Zhang, Y., and Yang, Q.: "A Survey on Multi-Task Learning," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 34, No. 12, pp. 5586–5609 (2022).

[6] Cao, W., Mirjalili, V., and Raschka, S.: "Rank Consistent Ordinal Regression for Neural Networks with Application to Age Estimation," *Pattern Recognition Letters*, Vol. 140, pp. 325–331 (2020).

[7] Liu, B.: "Sentiment Analysis and Opinion Mining," *Synthesis Lectures on Human Language Technologies*, Vol. 5, No. 1, pp. 1–167 (2012).

[8] Marsh, H. W.: "Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Usefulness," *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, pp. 319–383, Springer (2007).

[9] Spooren, P., Brockx, B., and Mortelmans, D.: "On the Validity of Student Evaluation of Teaching: The State of the Art," *Review of Educational Research*, Vol. 83, No. 4, pp. 598–642 (2013).

[10] Romero, C., and Ventura, S.: "Educational Data Mining and Learning Analytics: An Updated Survey," *WIREs Data Mining and Knowledge Discovery*, Vol. 10, No. 3, Article e1355 (2020).