講義に関する自由記述アンケートと成績の関連性調査

小池 正基[†1,a)]　平川 翼[†1]　山下 隆義[†1]　藤吉 弘亘[†1]

概要：教育環境のデジタル化に伴い，学生のデジタル教材に対する操作ログデータから講義理解度を把握し，学習支援を提供することが期待されている．しかし，ログデータを用いる手法では学生が講義内容について理解しているかといった曖昧な情報を加味した判断が不可能である．そこで我々は，学生が講義をどのように解釈しているかといった情報を含む，自由記述型の講義後アンケートに注目し，どの要素が学生の講義理解度に影響しているかを説明可能な機械学習モデルを構築する．具体的には，各講義回でのアンケートに対する回答文およびその回答文の時系列変化を距離情報として捉え，成績との関係性を調査する．そして，得られた距離情報から探索木モデルを用いた講義理解度予測を行い，予測精度の向上と予測根拠の提示を図る．提案手法を用いた評価実験では，文章の埋め込み表現をそのまま入力するモデルと比較して予測精度が 9.9pt 改善した．また，予測モデルは判断根拠として学生の講義に対する姿勢を重要視していることが示された．

An Investigation of the Relationship between Open-Ended Questionnaires and Lectures

1. はじめに

学習行動を記録したログデータを活用することで，講義を十分に理解できていない学生を早期発見し，講義や学習に対する取り組み方の改善や教育サポートを提供しようとする試みが活発である [1]．既存の研究では，学生がデジタル教材に対して行った操作や，過去の成績，出席率，そして宿題などのデータから，サポートベクタマシン (SVM)，ニューラルネットワーク，決定木などのモデルによって成績を予測する [2], [3]．また，Stephen ら [4] は，学生がデジタル教材に対して行ったビデオ視聴行動，演習，課題の回答，クイズの成績などのデータから重回帰（MLR）と主成分分析（PCA）を組み合わせて予測を行っている．

しかし，これらの手法は学生が行った動作を記録したものであり，学生が講義の内容についてどう解釈し，どの程度まで理解しているかといった概念を考慮することができず，適切な学習支援を提案することが難しい．学生の主観に寄り添い，より最適な学習支援を提供するための方法として，我々は講義についての自由記述アンケートに注目する．自由記述形式のアンケートは，教科書の操作内容などを記録するログデータよりも個人の理解が文章として反映されるため，主観的な要素を多く含み，学生の理解度という概念的な要素を具体的に考慮することができる．

しかし，自由記述形式のアンケートは学生間における表現の違いが大きく，単純な分類は非常に困難であるうえ，予測した成績の根拠を示すことが難しい．そこで，各学生が提出したアンケートの意味的な変化を時系列的に分析し，その変化量データを探索木モデルで分類することで，予測の根拠を示すことができるより精度の高い機械学習モデルの構築を図る．

2. 関連研究

教育分野において学生の成績の要因を調査する研究が活発である．

Stephanie ら [5] は，工学専攻の大学生に対して 3 問からなる自由記述形式のアンケートを実施し，アンケートの回答内容と GPA に相関があるかを調査した．調査には使用される単語の頻度，t 検定，z 検定，回答の長さと単語

[1] 情報処理学会
IPSJ, Chiyoda, Tokyo 101–0062, Japan
[†1] 現在，中部大学
Presently with Chubu Uniersity
[a)] masa1357@mprg.cs.chubu.ac.jp

数から分析を行った．分析の結果，成績の高い学生と低い学生の間で用いる単語に明確な差があることが示された．特に，アンケートにおける“In your own words, what do engineers do?”については，“why”や“Test（動詞）”などの単語と成績が相関関係にあることが確認された．

Sukrit ら [6] は，講義理解に対してリスクのある学生を特定し，適切なサポートを提供するためのリスク分類モデルである，Attn-ANN を提案した．リスク分類には学生が電子教科書に対して行った操作のイベントログデータを用い，講義理解度の低い生徒 (At-risk) 及び講義を理解している生徒 (No-risk) を予測することで，支援が必要な学生かどうかを判別する．Attn-ANN は RNN モデルによる講義の重みづけ，Attention によるログデータへの重みの適用，全結合層による成績予測で構成される．RNN を用いて学生の学習ログと講義回数の差分係数に準じて重みを付加することで，重要と考えられる講義回を重視するように学習を行った結果，モデルは時間要素，電子教科書のハイライト操作やメモ（ノート）の回数，出席回数や総アクション数に強く注目し，他モデル (MLP,GRU,LSTM) と比較して，講義が終盤に近付くほど高い精度を示した．

また, 筆者の以前の研究 [7] では, 学生から収集した講義後アンケート内の単語と成績の関連性を調査した．具体的には，各成績の学生が頻繁に用いる単語を Term Frequency-Inverse Document Frequency(TF-IDF) によって算出し，その単語が入力内に存在する場合は Transformer[8] モデルの Attention に対してバイアスを加えた予測を行った．結果，最終的な成績予測精度は 2.5pt 向上し，中間にあたる成績の予測精度が向上した．一方で，アンケートの時系列性を考慮できていない点，予測の判断根拠を十分に提示できていない点が課題である．

3. 提案手法

本研究では，学生が記録するアンケートの持つ時系列変化に着目し，アンケート文が持つ意味情報の変化量を算出し，探索木モデルを用いて講義理解度予測を行う．図 1 に提案手法の概要を示す．

3.1 Word2vec による文章埋め込み

アンケート文の持つ意味情報をベクトル形式に変換するため，Word2vec[9] を用いて使用単語の意味関係をベクトル表現に変換する．Word2vec は自然言語を数値ベクトルで表現するように学習されたニューラルネットワークモデルである．

Word2Vec は，「同じ文脈に現れる単語は似た意味を持つ傾向がある」という分布仮説に基づき，1 つの隠れ層を持つニューラルネットワークを使用して学習を行う手法の総称を指す．本研究では，skip-gram によって訓練されたモデルを用いる．図 2 に skip-gram の概要を示す．

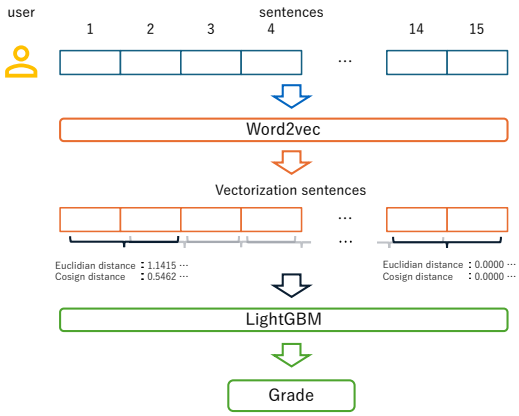


図 1: 提案手法の概要図

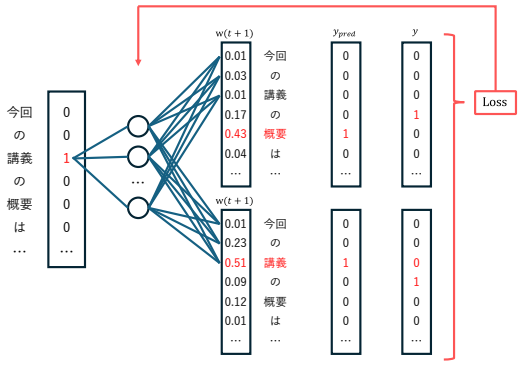


図 2: skip-gram の概要図

skip-gram は，文脈上のある単語に対して，周辺に現れる単語を予測することで学習を行う．式 (1) に skip-gram の目的関数を示す．ここで，$t = 1 \ldots T$ は単語列，w_t は t 番目の単語，m は予測にする単語の範囲を表す．

$$J = -\sum_{t=1}^{T} \sum_{-m \leq j \leq m} \log(P(w_{t+j} \mid w_t)) \qquad (1)$$

本研究では，文章内にある全ての単語ベクトルの平均値を文全体の意味ベクトルと捉えるものとする．

3.2 ベクトル間距離によるアンケート変化量の取得

Skip-gram によって得られた文章埋め込みを用いて，学生アンケートの時系列に沿った変化量を算出する．文章埋め込みを用いた変化量の算出方法を本節で述べる．

3.2.1 ユークリッド距離

ユークリッド距離はユークリッド空間における 2 点の直線距離を求める計算手法である．文章ベクトル同士のユークリッド距離を求める場合，両者の意味的な類似度を示す指標として表すことができる．式 (2) にユークリッド距離の計算式を示す．このとき，x, y はそれぞれ文章ベクトル $x = (x_1, x_2, \ldots, x_n)$，$y = (y_1, y_2, \ldots, y_n)$ である．

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \qquad (2)$$

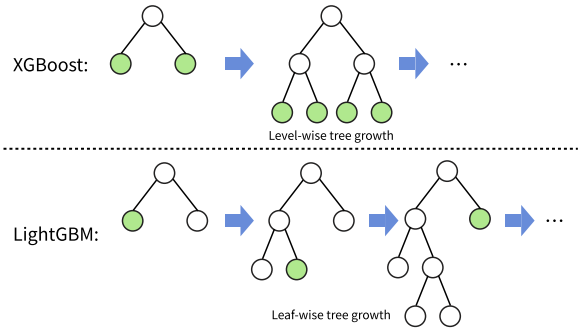


図 3: LightGBM の学習フロー

3.2.2 コサイン距離

コサイン距離は 2 ベクトル間の角度に基づいた距離を求める計算手法である．式 (3) にコサイン距離の計算式を示す．このとき，x, y はそれぞれ文章ベクトル $x = (x_1, x_2, \ldots, x_n)$，$y = (y_1, y_2, \ldots, y_n)$ である．

$$c(\mathbf{x}, \mathbf{y}) = 1 - \frac{\mathbf{x} \cdot \mathbf{y}}{|\mathbf{x}| \cdot |\mathbf{y}|} \tag{3}$$

これらのベクトル間距離は，小さいほど単語同士の意味的な類似度が高い．

3.3 探索木モデルによる学生の講義理解度予測

探索木モデルの 1 つである Light Gradient Boosting Machine(LightGBM)[10] を用いて変化量データでの学習を行い，講義理解度予測精度を検証する．LightGBM とは，弱学習機である決定木モデルを勾配ブースティングによりアンサンブル学習するアルゴリズムである．LightGBM の学習の流れを図 3 に示す．図 3 のように，XGBoost などの従来の探索木モデルは深さ方向に拡張しながら学習を行うが，LightGBM は葉ごとに学習を行うため，より効率的な学習を行うことが可能である．

探索木は式 (4) のような損失関数を最小化する計算によって学習を行う．

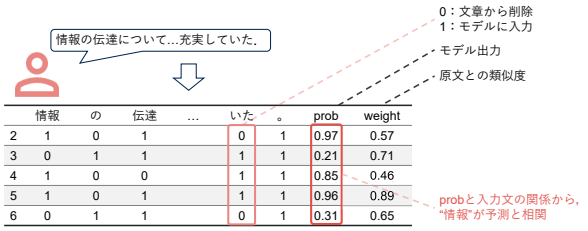
$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{N} \Omega(f_k) \tag{4}$$

ここで，$l(y_i, \hat{y}_i)$ は損失関数，$\hat{y}_i$ は i の予測値，$\Omega(f_k)$ はモデルの複雑さを制御するための正則化項，N は木の総数，n はインスタンス数を表す．

本研究では，学生の最終成績を講義理解度とみなし，LightGBM モデルによって予測することで講義理解度予測を行う．

3.4 SHapley Additive exPlanation(SHAP) による各特徴量の寄与度分析

SHapley Additive exPlanation(SHAP)[11] とは，協力ゲーム理論のシャープレイ値（Shapley Value）を機械学習に応用することで，入力されたデータのどの部分が結果に寄与しているかを数値化する手法である．SHAP の簡単な概要を図 4 に示す．



	情報	の	伝達	…	いた	。	prob	weight
2	1	0	1		0	1	0.97	0.57
3	0	1	1		1	1	0.21	0.71
4	1	0	0		1	1	0.85	0.46
5	1	0	1		1	1	0.96	0.89
6	0	1	1		0	1	0.31	0.65

図 4: SHAP の概要

図 4 にあるように，入力されるデータの一部を消去して機械学習モデルに入力することで変化する出力の傾向を算出し，入力と出力結果の関係性を求める．特徴 j における shap 値 ϕ_i は式 (5) によって求められる．

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [k(S \cup \{i\}) - k(S)] \tag{5}$$

ここで，S は i 以外の入力，N は全ての特徴の集合，$k(S)$ は特徴量集合 S に基づいて予測モデルが出力する予測値を示す．

4. 実験

得られた講義毎の変化量を可視化し，成績との関連性を調査する．また，変化量データを用いて探索木モデルを訓練し，成績予測および特徴量重要度からどの要素が最も成績と相関があるかを分析する．

4.1 データセット

本研究では，九州大学で収集された，情報科目の講義に関するアンケートの回答文と，回答者の成績で構成されるデータセットを用いる．アンケートは 2020 年から 2022 年までの期間，講義直後に行われた．アンケートは一回につき 5 問収集され，各講義 14 回分，1 人に対して計 70 問回答する．アンケートの質問内容は以下の通りである．

- Q1: 今日の内容を自分なりの言葉で説明してください．
- Q2: わかったことを書いてください．
- Q3: わからなかったことを書いてください．
- Q4: 質問があれば書いてください．
- Q5: 今日の講義の反省，感想を書いてください．

本研究では，最も回答率が高く，講義に関して主観的な解釈を多く含む Q1 のみ用いる．

各受講者の成績は A，B，C，D，F が与えられる．成績 A が最もく，成績 F が与えられた生徒は本講義の単位が与えられない．成績ごとのアンケートの未提出回数を図 5 に，アンケートの平均回答単語数を図 6 に示す．

各学期のコース受講者を成績の分布とともに表 1 に示す．アンケートに協力した学生 377 名のうち，8 割の学生（298 人）の回答を学習データ，2 割の学生（75 人）の回答

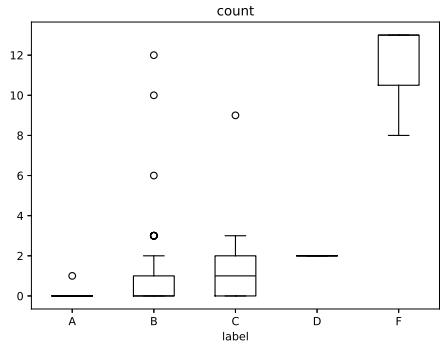


図 5: 成績ごとのアンケート未提出回数

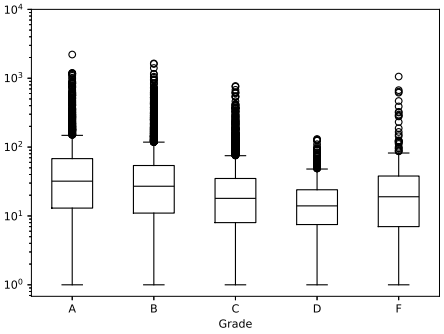


図 6: アンケートの平均回答単語数

表 1: 各学期におけるコース受講者の成績分布

Tearm	A	B	C	D	F	Total
2021-1	9	53	32	7	6	107
2021-2	15	88	37	9	25	174
2022-1	17	37	34	4	4	96
Total	41	178	103	20	35	377

表 2: Word2vec モデルのハイパーパラメータ

ベクトル次元数	200
最小単語出現回数	5
スレッド数	40
ウィンドウサイズ	10
ダウンサンプリングの閾値	1e-3
エポック数	25
ネガティブサンプリングに用いる単語数	10

を評価データとして使用する．

4.2 Word2vec による文章埋め込みの取得

アンケートの持つ時系列情報と講義理解度の相関関係を調査するため，Word2vec を用いて学習データ内の全文章を埋め込みに変換し，各埋め込みの変化量を算出する．Word2vec のハイパーパラメータを表 2 に示す．

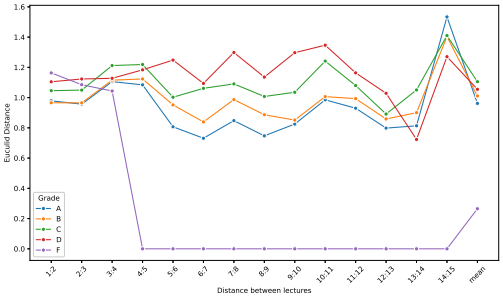


図 7: 講義間のユークリッド距離

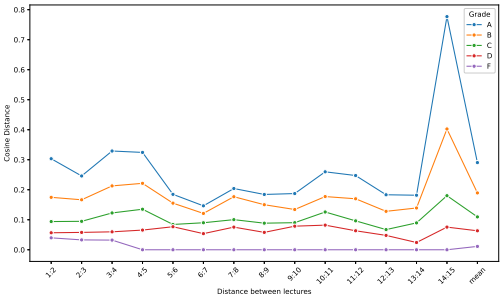


図 8: 講義間のコサイン距離

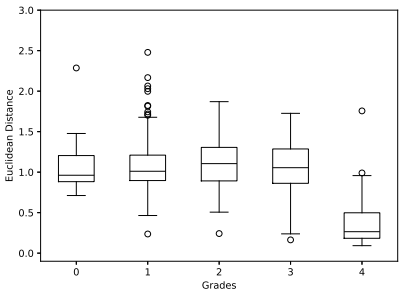


図 9: 全ユークリッド距離の平均値分布

4.3 アンケートにおける時系列変化量の分析

学生が提出したアンケートの講義間変化量を図 7，図 8 に示す．このとき，講義間の回答が同一である場合に得られるベクトル間距離は 0 となるため，学生毎の変化量の差が大きい．これを解決するため，全ての学生の変化量の中央値を用いる．図 7，図 8 より，成績 F の学生は 7 割以上の講義でアンケートの回答が変化しないことがわかる．また，成績が高い生徒ほどユークリッド距離が小さく，コサイン距離が大きいことが確認できる．つまり，ベクトルの絶対値距離は短く，方向は大きく異なると言えるので，成績の高い生徒ほど回答文の形式や文章量は類似するが，文章内で言及する内容が大きく異なると考えられる．

次に，各成績における学生のアンケート変化量についての平均値分布を図 9，図 10 に示す．図 9，図 10 を比較すると，成績 A,B,C,D のユークリッド距離は似た値をとるが，コサイン距離は成績が高いほど大きくなり，アンケートの内容の意味変化と成績には相関関係があると考えられる．

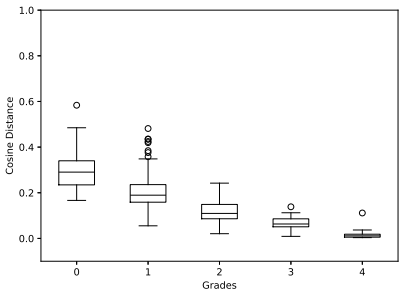


図 10: 全コサイン距離の平均値分布

表 3: 入力特徴量・名称

アンケートの未提出回数	zero-count
講義ごとの単語数	01〜15
文章のユークリッド変化量	euc_01:02〜euc_14:15,
ユークリッド変化量の平均値	euc_mean
文章のコサイン変化量	cos_01:02〜cos_14:15,
コサイン変化量の平均値	cos_mean

表 4: lightGBM モデルのハイパーパラメータ

num_class	5 or 4
metric	multi_logloss
boosting_type	gbdt
num_iterations	1000
max_depth	12
learning_late	0.01
num_leaves	31

4.4 意味的な時系列変化量を用いた学生の講義理解度予測

次に，探索木モデルを用いて，アンケートの時系列変化量を用いたモデルの有効性を検証する．比較対象として，全 15 回の講義アンケートごとに LightGBM モデルを訓練し，logit の平均値を最終的な予測確率とするアンサンブルモデルを Baseline とする．

lightGBM に入力する特徴量及び名称を表 3 に，lightGBM モデルのハイパーパラメータを表 4 に示す．

ここで，データセットの不均衡を解決するため，各クラスに対して占める割合の逆数で重み付けして学習する．評価には 5-fold クロスバリデーションを行い，Accuracy, F1-score, AUC の平均値を用いて提案手法の有無における精度の変化を検証する．

- Accuracy は正しい予測の数が全体の何割を占めるかを表す指標である．
- F1-score は精度 (Precision) と再現率 (Recall) の調和平均によって求められる評価指標である．
- AUC スコアは全ての分類閾値にわたって総合的な性能指標を測定するため，不均衡なデータセットに対して有効な評価指標である．

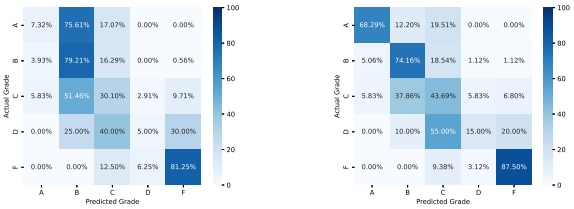
評価に際して，表 1 にあるように，成績 D の学生数が全体の 5%程度であるため，成績 C-D を同一と見なした 4 クラス分類による評価も行うものとする．

表 5: 5 クラス分類モデルにおける評価

指標	Baseline	Ours
Accuracy [%]	54.01	**63.10**
F1-score [%]	37.31	**54.62**
AUC	0.7266	**0.8227**

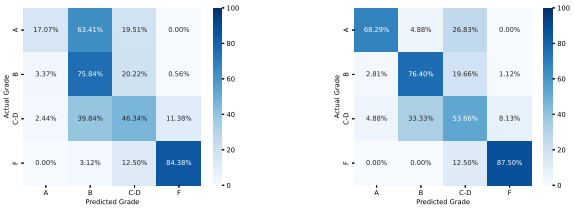
表 6: 4 クラス分類モデルにおける評価

指標	Baseline	Ours
Accuracy [%]	60.41	**68.99**
F1-score [%]	53.15	**68.84**
AUC	0.7176	**0.8237**



(a) 混同行列 (Baseline)　(b) 混同行列 (Ours)

図 11: 5 クラス分類時における混同行列の比較



(a) 混同行列 (Baseline)　(b) 混同行列 (Ours)

図 12: 4 クラス分類時における混同行列の比較

表 5，表 6 は，各手法における講義理解度予測精度の比較である．表 5，表 6 より，Baseline と比較して提案手法の F1-score がそれぞれ 17.31pt，15.69pt 向上し，各成績を正しく予測できるように学習していることが確認できる．

図 11，図 12 に，5 クラスと 4 クラス予測における混同行列を示す．混同行列から，Baseline モデルは予測が成績 B に偏っていることを確認できる．これは，表 1 にあるように，データセットの分布が成績 B に偏っているため発生する現象である．しかし，提案手法では全てのクラスに予測が分散し，各学生の特徴を正しく考慮できていると言える．

4.5 SHAP による特徴量重要度分析

今回の実験で最も高い精度を示した提案手法モデルにおいて，どの特徴量が最も成績と相関するかを特徴量重要度から分析する．図 13 に，SHAP によって算出した特徴量

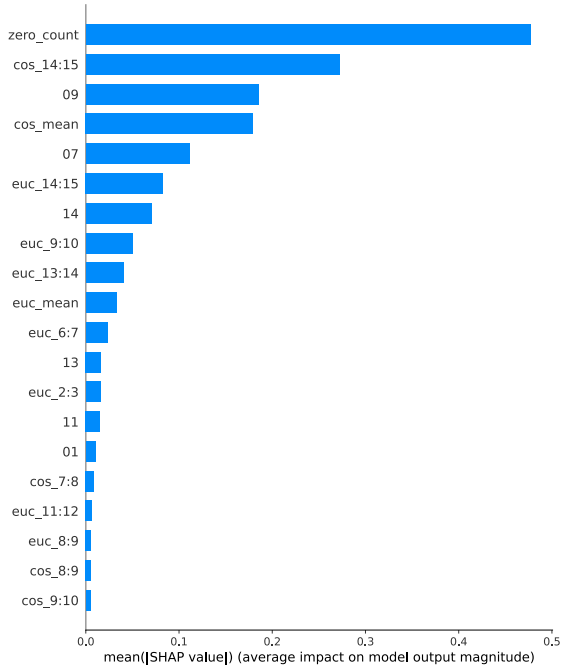


図 13: 提案手法モデルの特徴量重要度の可視化

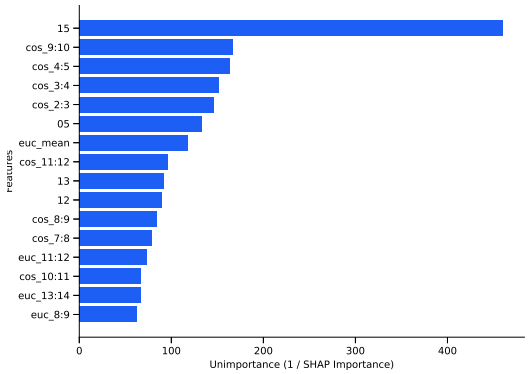


図 14: 提案手法モデルの特徴量不要度の可視化

重要度を示す．図 13 より，本モデルはアンケートの未提出回数 (zero-count)，アンケート間のコサイン変化量 (cos)，アンケート文章の単語数 (01-15) の特徴を重視していることが確認できる．

次に，どの特徴量が最も講義理解度と相関していないかを分析する．図 14 に，SHAP 値の逆数によって各特徴量の不要度を示す．

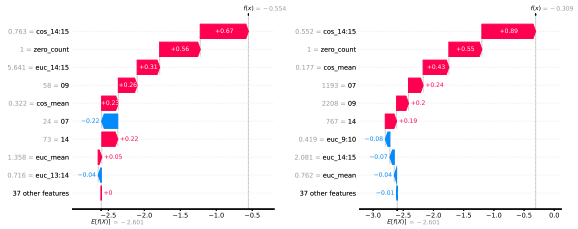
図 14 より，本モデルは 15 回目アンケートの単語数 (15) が最も成績と関係しないことが確認できる．これは，第 15 回目の講義が期末テストの週であり，アンケートの内容が講義と関係しないためであると考えられる．

次に，学生個人のアンケート文に注目し，成績が高い学生と低い学生の間でどのような違いがあるのか分析する．

評価データ内でモデルが予測に成功したもののうち，成績 A，成績 C，成績 F の学生に対してランダムに 2 人ずつサンプリングを行い，それぞれのアンケートに対してモデルがどのような判断根拠を以って予測を行っているのかを可視化して確認する．サンプリングした学生の id を表 7 に，それぞれの学生に対して，モデルがどのような根拠から成績を予測したかを図 15，図 16，図 17 に示す．

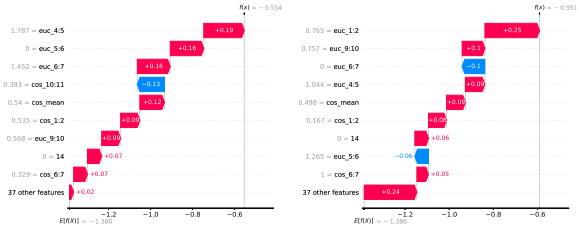
表 7: サンプリング対象

成績 A	: C-2021-2_U31
	: C-2021-2_U41
成績 C	: C-2021-2_U140
	: C-2022-1_U39
成績 F	: C-2021-2_U7
	: C-2021-2_U84



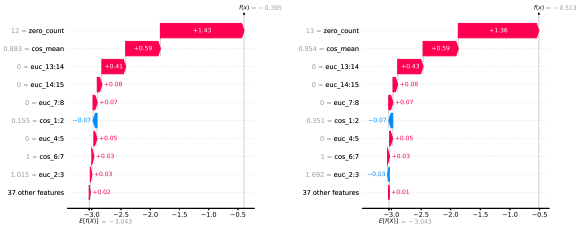
(a) C-2021-2_U31　(b) C-2021-2_U41

図 15: 成績 A の学生における判断根拠の可視化



(a) C-2021-2_U140　(b) C-2022-1_U39

図 16: 成績 C の学生における判断根拠の可視化



(a) C-2021-2_U7　(b) C-2021-2_U84

図 17: 成績 F の学生における判断根拠の可視化

図 15 より，成績 A の学生に対してモデルは講義 14，15 回間のコサイン距離，未提出数，全体的なコサイン距離の平均に注目していることがわかる．このとき，両者の 14，15 回目講義のアンケート回答文に注目すると，14 回目では「データの広がりの指標として，新たに相関について学習した」というような具体的な受講内容に言及しているが，15 回目では「テストを解きました．」などの非常にシンプ

ルな回答となっている．他の成績の学生は 14 回目で未回答であったり簡潔である場合が多く，この部分のギャップが大きな判断根拠になっていると考えられる．

図 16 より，成績 C の学生に対してモデルは講義 1，2 回目のユークリッド距離，4，5 回目のユークリッド距離など，未回答数ではなく，講義の前半から中盤部分の変化量に注目していることがわかる．このとき，学生 C-2021-2_U140 は講義 4 回目から講義 7 回目までの回答文字数が平均 50 文字となっており，これは成績 A の学生の文字数 (110.25, 1412) と比較して少ない．また，講義 4 回目と講義 5 回目のアンケート回答が一言一句同一であるため，このようないい加減な回答をモデルが根拠として提示していると考えられる．さらに，学生 C-2022-1_U39 を見ると，モデルは講義 1，2 回目に最も注目しているが，これはこの区間の回答文字数の差が 509 文字と非常に大きいためであると推測できる．

図 17 より，成績 F の学生に対してモデルは未提出回数と全体のコサイン変化量に注目していることがわかる．これは，図 5 にあるように，成績 F の学生が他と比べて圧倒的に未提出が多いため，その要素を判断根拠として考慮しているためであると推測できる．

以上の結果から，モデルの予測の根拠となっている部分として，アンケートの未提出数や回答文の単語数の変化，文章の使いまわしなどの要素を考慮していることが確認できる．

5. まとめ

我々は，学習行動ログデータと比較して，学生の理解度を説明するにあたって重要である主観的な要素を多分に含む自由記述形式のアンケートに対して，意味的な時系列変化量を考慮する機械学習モデルの有効性を評価した．自由記述アンケートは全 15 回行われ，その意味情報を文章ベクトル化モデルで埋め込みに変換し，時系列に沿って文章毎の変化量を算出した．結果，意味的な時系列変化量は学生の成績と相関が見られ，変化量を用いて探索木モデルを訓練することで，Accuracy が 9.9pt，F1-score が 17.31pt 向上することを確認した．また，SHapley Additive exPlanations (SHAP) によって分類モデルの判断根拠を可視化した結果，文字数や未回答率だけでなく，回答文の使いまわしなどの要素についても注目しており，学生の講義に対する姿勢を重要視していることが示された．今後は，文章間距離だけでなく，文章全体を考慮して総合的に理解度予測を行うモデルを研究し，より根拠のある説明を示しながら高い精度を示すことが可能な AI の構築を目指す．

6. 謝辞

本稿は，JST CREST グラント番号 JPMJCR22D1 により支援された．

参考文献

[1] Yağcı, M.: Educational data mining: prediction of students' academic performance using machine learning algorithms, *Smart Learning Environments* (2022).

[2] Shahiri, A. M., Husain, W., Nur' aini AbdulRashid：A Review on Predicting Student's Performance Using Data Mining Techniques, *Procedia Computer Science*, (online), DOI: https://doi.org/10.1016/j.procs.2015.12.157 (2015).

[3] Namoun, A. and Alshanqiti, A.: Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review, *Applied Sciences*, Vol. 11, No. 1 (online), DOI: 10.3390/app11010237 (2021).

[4] Yang, S. J., Lu, O. H., Huang, A. Y., Huang, J. C., Ogata, H. and Lin, A. J.: Predicting Students' Academic Performance Using Multiple Linear Regression and Principal Component Analysis, *Journal of Information Processing*, Vol. 26, pp. 170–176 (online), DOI: 10.2197/ipsjjip.26.170 (2018).

[5] Gratiano, S. M. and Palm, W. J.: Can a five minute, three question survey foretell first-year engineering student performance and retention?, *ASEE* (2016).

[6] Leelaluk, S., Tang, C., Minematsu, T., Taniguchi, Y., Okubo, F., Yamashita, T. and Shimada, A.: Attention-Based Artificial Neural Network for Student Performance Prediction Based on Learning Activities, *IEEE Access*, Vol. 12, pp. 100659–100675 (online), available from ⟨https://doi.org/10.1109/ACCESS.2024.3429554⟩ (2024).

[7] Koike, M., Kohama, H., Hirakawa, T., Yamashita, T. and Fujiyoshi, H.: Enhancing the Accuracy of Predicting Students Grades in Open-Ended Questions through Adjustments to Attention Weights, *Proceedings of the 17th International Conference on Educational Data Mining* (BenjaminPaa 歹 n, Epp, C. D., 編), Atlanta, Georgia, USA, International Educational Data Mining Society, pp. 872–876 (online), DOI: 10.5281/zenodo.12729979 (2024).

[8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I.: Attention is all you need, *Proceedings of the 31st International Conference on Neural Information Processing Systems* (2017).

[9] Rong, X.: word2vec Parameter Learning Explained (2014). cite arxiv:1411.2738.

[10] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.-Y.: LightGBM: a highly efficient gradient boosting decision tree, *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, Red Hook, NY, USA, Curran Associates Inc., p. 3149–3157 (2017).

[11] Lundberg, S. M. and Lee, S.-I.: A unified approach to interpreting model predictions, *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, Red Hook, NY, USA, Curran Associates Inc., p. 4768–4777 (2017).