中学生の学習能力向上を目指した授業後振り返り文に基づく成績推定

新谷 一朗[1,a)]　峯 恒憲[2]

概要：学生の成績を推定することは成績不良の学生の特定に非常に有用である．本研究では学習塾の中学生が記述した授業後の振り返り文から成績推定モデルの構築を行う．各学生が生徒自身が受講している科目の授業後に書いた自分の理解したことや分からなかったことなどの振り返り文を単語ベクトルに変換して，そのベクトルに各生徒の成績の良し悪しのラベル付けを行った．その後次元圧縮を行い，様々な機械学習器より生徒の成績推定を行った．その結果データセット，質問項目，コメント取得月の全てに対して安定し，かつ高精度な成績推定の手法はなかった．また時期や質問によって成績推定の精度にばらつきが生じるといったことが見受けられた．

キーワード：機械学習，テキストマイニング，中学生のコメント，成績推定

Comment Mining to Estimate Junior High-school Student Performance toward Improvement of Student Learning

Niiya Ichiro[1,a)]　Mine Tsunenori[2]

Abstract: Estimating student learning performance is very useful for identifying students who will take poor grades. This research proposes an approach to build learning performance estimation models of junior high school students using their self-evaluated sentences written after every lesson. They wrote their self-evaluated sentences as answers to questions about what they understood or did not understand to subjects they learned, and what they want to do till the next lesson and what they do not understand currently. Our approach first transforms the sentences into word vectors and applies dimension reduction techniques such as Principal Component Analysis (PCA) or Latent Dirichlet Allocation (LDA). Then we build student learning performance estimation models from them using machine learning techniques: Support Vector Machines (SVM), Decision Tree (DT), and Random Forest (RF). The models estimate student learning performance and classify them into two: Good and Bad learning performance. Experimental results were evaluated using F-measure. The results show F-measure scores vary monthly and a method having the highest score of F-measure also varies per month or by data set.

Keywords: Machine learning, Text mining, style files, Junior high school student comments, Student performance estimation

1. はじめに

1.1 研究の背景

Web や SNS などのレビュー記事のマイニングによって

[1] 九州大学工学部電気情報工学科
Kyushu University
[2] 九州大学大学院システム情報科学研究院
Faculty of Information Science and Electrical Engineering
Kyushu University
[a)] niiya@ma.ait.kyushu-u.ac.jp

表 1　質問

質問 1	今日の範囲で分かった事，今日の学習で新しく覚えたこと，理解できたことを教えてください．
質問 2	今日の範囲で分からなかった事，今日の学習で分からなかったことや，質問したいことを教えてください．
質問 3	次回やりたいこと，今分からないことを教えてください．

.

表 2　使用したデータ

	データセット 1	データセット 2
生徒数	13 人	43 人
日付	2015/7/27〜2016/3/31	2016/4/4〜2016/9/27
科目	英語 9 人分，数学 10 人分，社会 3 人分	英語 26 人分，数学 35 人分，社会 6 人分，理科 3 人分，国語 1 人分

商品の良し悪しに関する意見内容を抽出する研究が盛んに行われている．教育分野における成績推定の研究では同じくマイニングの技術が使われているが，学習者の授業内容の評価ではなく学習状況や能力を推定に焦点が当てられている．単に解析精度を上げるだけでなく，機械学習に詳しくない教育者に解析結果，解析過程，判断基準がわかるような成績推定モデルの構築が必要である．成績推定は教師が成績不良者を特定して，彼らがより良い点数を取れるようにすることに役に立つという利点もある．

本研究の目的は対象の生徒（中学生）の学習能力及び学習態度の向上のための成績推定モデルの構築である．振り返り文から成績の良い生徒と悪い生徒の特徴的な記述を抽出して成績推定を行う．今回の研究は先行研究と違い，中学生を対象とする．中学生の場合，大学生よりも質の悪いコメントを書くことが予想されるため，学生に合った成績推定を行うことはより困難な課題と想定される．

本研究では学習塾の中学生たちが受講している授業終了後に書いた授業の振り返り文をもとに，

- 成績の良い人と悪い人のコメントの特徴の調査
- 機械学習を使って成績推定

の 2 点を実施した．その結果，コメント収集を行った全期間を通して常にほかの手法よりも高い精度を得る手法がなかったなど大学生と同じような結果は得られなかった

以下では，2 章は関連研究について述べ，3 章は今回使用したデータについての説明をする．4 章は分析に用いたデータの準備方法を述べ，5 章で分析手法について述べる．6 章で評価指標について述べ，7 章で分析結果と考察を述べる．最後に，8 章で全体のまとめと，今後の課題を述べる．

2. 関連研究

本研究の関連研究 [1,2,3,4,5] としては，Sorour らは [1] 大学生が授業後に授業で分かったことや分からなかったことなどを記した振り返り文を利用し，成績推定を行った．手順としては生徒たちの振り返り文から単語ベクトルを作成して，そこから機械学習を用いて成績推定を行った．その結果 LDA[6] と SVM を使った方法が最もより良い成績推定ができたことが分かった．またコメント文の質が悪いと推定精度が落ちることも分かった．単語ベクトルを作成方法として，LDA，pLSA，LSA を用い，機械学習方法として ANN，SVM，DT，RF を用いた．

また Bhardwaj らは [2] 生徒のデータとして前回の試験の点数や前の授業のテストの点数，授業中の態度，出席や宿題を使って成績推定を行った．結果として成績は生徒の努力に依存するとは限らないと述べている．

他の研究では Yadav ら [3] が対象とする大学生が期末試験の合否を推定する決定木のプログラムを作成し，67.78 %という精度を示している．

生徒の振る舞いについての研究では Qiu ら [4] は応答から 1 日以上が経過した学生の反応に関する予測を Knowledge Tracings 予測を使い，調べた．そしていくつかのデータを学生データにフィッティングさせるより良い Knowledg Tracings 予測モデルを提示した．

教育用のテキストマイニングに関する研究として，Tane らは [5] トピック別に e-learning の文章をグループ化するためにクラスタリングを使用したと述べている．

しかしこれらの先行研究では，大学生を対象としており，中学生に対しての成績推定を行う研究は行っていない．そのため大学生の成績推定に利用されたモデルをそのまま中学生に適用できるかどうかは不明である．また今までの大学生の成績推定の研究では各授業の最終成績のみを成績推定に利用しているが，本研究では，中学生の学期ごとの中間試験，期末試験の結果を利用して，時期を分けた成績推定を行った．

3. データ説明

以下で生徒たちが記述した振り返り文の質問を表 5 示す．

3.1 使用したデータの説明

今回用いたデータは以下の通りである．またデータは振り返り文だけでなく 1 学期期末，2 学期中間，2 学期期末，3 学期期末の 4 つの試験結果が与えられている．表 6 にデータの内容を示し，表 7 と表 8 に月別のコメントの総

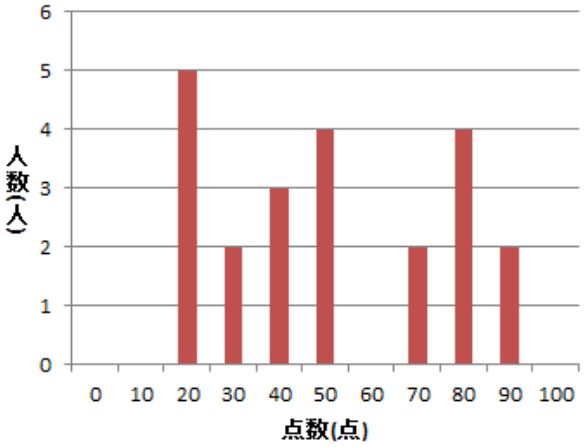
表 3　質問

質問 1	今日の範囲で分かった事，今日の学習で新しく覚えたこと，理解できたことを教えてください．
質問 2	今日の範囲で分からなかった事，今日の学習で分からなかったことや，質問したいことを教えてください．
質問 3	次回やりたいこと，今分からないことを教えてください．

.

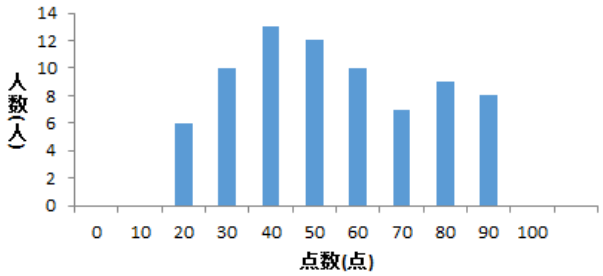
表 4　使用したデータ

	データセット 1	データセット 2
生徒数	13 人	43 人
日付	2015/7/27〜2016/3/31	2016/4/4〜2016/9/27
科目	英語 9 人分，数学 10 人分，社会 3 人分	英語 26 人分，数学 35 人分，社会 6 人分，理科 3 人分，国語 1 人分



.

図 1　データセット 1 のヒストグラム



.

図 2　データセット 2 のヒストグラム

数を示す．

3.1.1　振り返り文の質問内容及び説明

表 5 から振り返り文に記述されている内容は以下のとおりである．

- 生徒 ID:生徒を一意に識別するための ID
- 日付:コメントを記述した日にち
- 科目:受講した科目
- 試験結果:日付が 4〜6 月ならば 1 学期期末試験結果，7〜9 月ならば 2 学期中間試験結果，10〜12 月ならば 2 学期期末試験結果，1〜3 月ならば 3 学期期末試験結果が記載されている．
- 質問 1，2，3:それぞれの質問に対するコメント

生徒たちの質問の回答例及び成績データ

生徒たちの質問の回答例を表 9 に，成績データのヒストグラムを図 1，図 2 に示す．

3.1.2　生徒たちのコメントデータの特徴

本研究で分析で利用したデータはデータの取得時期に応じてデータセット 1，データセット 2 に分けられている．表 10 にデータセット 1 とデータセット 2 のすべての生徒の平均文字数とその分散，平均空欄率とその分散，平均点数，及び平均点数と平均文字数，平均点数と平均空欄率との相関係数を示した．

表 4 の分散の値から今回の生徒のコメントのデータには，データセットによって空欄の割合や文字数に差があり，人によって文字数にばらつきがあることがわかった．またデータセット 1 に関しては，表 4 の相関係数の値から文字数と空欄率は点数と相関が強いがデータセット 2 では文字数と空欄率はデータセット 1 と比べて点数と強い相関がなくなっていることが分かった．

4.　データと成績ラベルの作成

この章では，成績推定に利用する機械学習手法の入力となる学習用データと成績ラベルを振り返り文をもとに作成する方法について述べる．

4.1　コメントから単語を抽出

生徒の振り返り文から名詞，動詞，形容詞，助動詞を抽出した．

4.2　単語を重みづけ

抽出した単語をすべて TF(TermFrequency)-IDF(InverseDocumentFrequency) を用いて，重みづけを行った．

4.2.1　TF-IDF

TF-IDF[8] とは文書中にある単語の重みを求める手法の一つである．

TF 値は文書内の特定の単語の頻度を表す値であり，(1) の式で表される．文書中に頻繁に出現する単語に対して大きな値が与えられる．

表 5　質問

質問 1	今日の範囲で分かった事，今日の学習で新しく覚えたこと，理解できたことを教えてください．
質問 2	今日の範囲で分からなかった事，今日の学習で分からなかったことや，質問したいことを教えてください．
質問 3	次回やりたいこと，今分からないことを教えてください．

.

表 6　使用したデータ

	データセット 1	データセット 2
生徒数	13 人	43 人
日付	2015/7/27〜2016/3/31	2016/4/4〜2016/9/27
科目	英語 9 人分，数学 10 人分，社会 3 人分	英語 26 人分，数学 35 人分，社会 6 人分，理科 3 人分，国語 1 人分

.

表 7　データセット 1 の月別コメントの総数

月	質問 1	質問 2	質問 3
7	17	16	0
8	29	28	1
9	64	62	1
10	67	65	41
11	56	53	51
12	60	57	56
1	57	56	58
2	58	56	58
3	70	59	61

.

表 8　データセット 2 の月別コメントの総数

月	質問 1	質問 2	質問 3
4	155	143	142
5	155	141	140
6	210	190	190
7	214	177	169
8	275	217	190
9	223	185	170

IDF 値は文書頻度の逆数であり，(2) の式で表される．特定の文書に集中して出現する単語に対して大きな値が与えられる．

TF-IDF 値は TF 値に IDF 値を掛け合わせた数値であり，その値が大きいほど各文書を特徴づける単語であるいうことを示している．(3) の式で表される．

$$tf(w,d) = \frac{n_{w,d}}{\sum_{w_i \in d} n_{w_i,d}} \quad (1)$$

$n_{w_i,d}$:ある単語 w の文書 d 内での出現回数

$\sum_{w_i \in d} n_{w_i,d}$:文書 d における単語数

$$idf(w) = \log \frac{N}{df(w)+1} \quad (2)$$

N:全文書数

$df(w)$:ある単語 w が全文書中で出現する文書の数

$$tf\text{-}idf(w,d) = tf(w,d) * idf(w) \quad (3)$$

以上の手順で文ごとに単語数が次元数となる単語ベクトルを作成した．

4.3　全ての単語ベクトルを行列化

コメントごとに TF-IDF による単語の重みを要素とする単語ベクトルを作成した．また単語ベクトルのサイズは月別のすべての文書から抽出された総単語数を n、月別の総コメント数を m として，n 次元 m 行のベクトルである．

4.4　点数によるラベル作成

図 1，図 2 のヒストグラムより成績の良し悪しの閾値を 60 点と設定した．そこで 60 点より高い点の場合成績の良い生徒とし，60 点以下の場合を成績の悪い生徒としてラベルを付けた．

5.　成績推定

この章では，成績推定手法について述べる．

5.1　行列の次元を圧縮

計算量の削減を目的として，次元圧縮によって高次元のデータを低次元のデータに変換する．次元圧縮の方法として PCA，LDA を用いた．PCA[7] とは多数の変数の情報を主成分得点と呼ばれる少数の合成変数に変換する手法である．LDA[6] とは一つの文書には複数の潜在トピックが存在すると仮定して，そのトピック分布を離散分布としてモデル化する手法である．今回 PCA では次元を 50 次元とした．同様に LDA ではトピック数を 10 にして行った．

5.2　機械学習を用い，成績推定

次元圧縮した行列データをテスト用データと訓練用データに分けて SVM，決定木，ランダムフォレストを用いて成績推定を行った．成績推定の精度に影響が出るため，訓練用のデータにテスト用データの生徒のコメントが入らないようにした．

5.3　推定結果の評価

対象の生徒をテスト用と訓練用に分けて推定を行い，その後テスト用データと訓練用データを入れ替えて繰り返し検証を行う交差検定 [10] を実施した．本研究では中学生を

.

表 9　使用した振り返り文の例

生徒 ID	日付	教科	試験点数	質問 1	質問 2	質問 3
id1	2016/1/11	数学	45	合同条件が分かった	とくにない	つづき
id1	2016/1/18	数学	45	合同条件が覚え	とくにない	つづき
id1	2016/1/25	数学	45	証明の書き方	ありません	次のところ
id1	2015/12/3	数学	45	証明の順序	ありません	つづき

表 10　データの平均文字数，平均空欄率，平均点，点数と文字数の相関，点数と空欄率の相関

	平均文字数	文字数の分散	平均空欄率	空欄率の分散	平均点	点数と文字数の相関係数	点数と空欄率の相関
データセット 1	11.9	64.3	0.29	0.06	46.6	0.73	-0.71
データセット 2	24.6	185.9	0.14	0.04	50.9	0.27	0.05

ランダムに 2 つのグループに分けて 2-fold クロスバリデーションを行い推定結果の平均をとった．

6. 評価指標

この章では，成績推定の精度の評価指標について述べる．

データに対して実際に割り当てられているラベルと予測によってデータに割り当てられたラベルが 1 対 1 に対応する場合，*Precision* , *Recall* , *F-score* は表 11 の *TP*, *FP*, *TN* を用いて次のように計算される．また *TP*, *FP*, *TN* とはそれぞれに対応する実際と予測の事例数である．

***Precision*(適合率)**

Precision とは正例と判断したもののうち，どれだけが正しかったか示す指標である．(4) の式で計算できる．

$Precision = \frac{TP}{TP+FP}$ (4)

***Recall*(再現率)**

Recall とは正例のうち，正しく正例と判断された割合を示す指標である．(5) の式で計算される．

$Recall = \frac{TP}{TP+TN}$ (5)

***F-score*(F 値)**

F-score[9] とは予測モデルの精度の指標の一つである．*F-score* が高ければ性能が良いことを意味する．(6) の式で計算される．

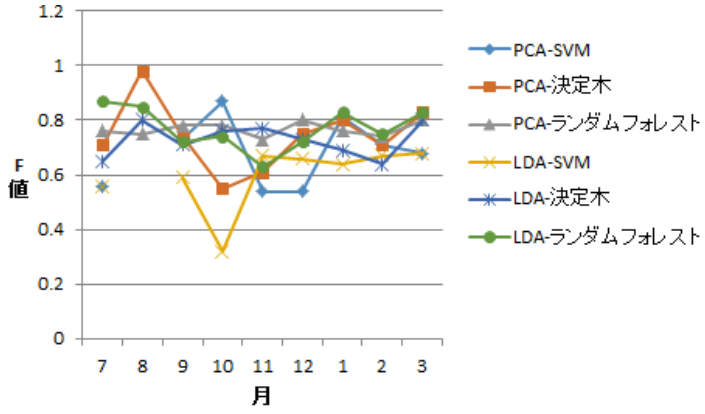
$F\text{-}score = \frac{2*Precision*Recall}{Precision+Recall}$ (6)

7. 分析結果及び考察

7.1 手法の比較

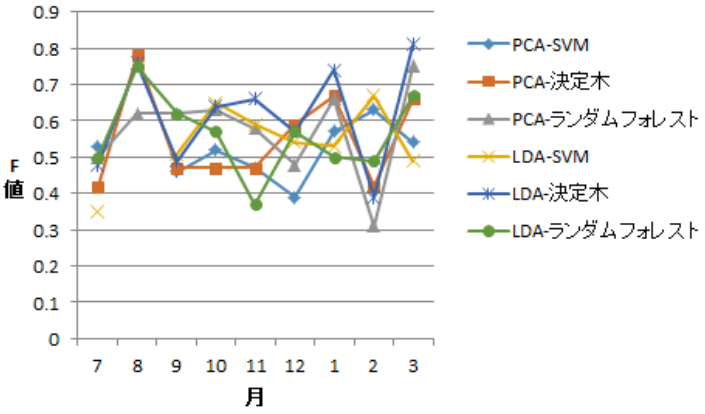
データセット別に本研究で使用した手法を比較した．質問ごとに今回使用した手法の成績推定精度について図 3〜図 8 に示す．

図 5 から 1 月の推定精度がどの手法も成績推定精度が下がり，図 6 から 8 月の成績推定精度が下がっていることがみられる．これは生徒が長期休暇前また休暇中なので生徒の学習意欲が落ちていると考えられる．また図 5 と図 8 からデータセット 1，2 の質問 3 に関する成績推定精度は大きく上下している．先行研究では質問 3 のように未来に関するコメントは成績あまり関係ないことがわかっており，質問 3 はあまり他の質問に比べて重要でないと考えられる．



.

図 3　データセット 1 の質問 1 に関しての成績推定 (F 値)



.

図 4　データセット 1 の質問 2 に関しての成績推定 (F 値)

またそれぞれの手法による成績推定精度の F 値の平均と分散を表 12〜表 15 に示す．

表 12 の結果からデータセット 1 では，質問 1 において LDA で次元圧縮してランダムフォレストで成績推定を行った結果が最も良い結果を示した．質問 1 において PCA で次元圧縮を行い決定木で成績推定を行った結果が最も悪い精度であった．

表 13 の結果からデータセット 1 では質問 3 において

.

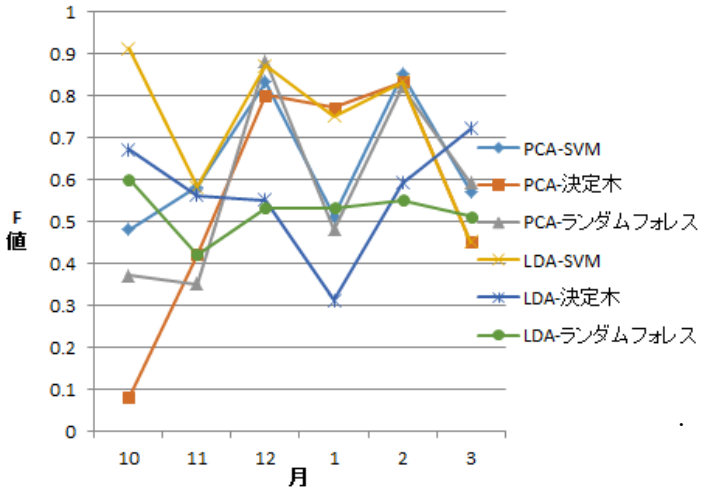
表 11　TP，FP，TN の意味

予測＼実際	正	負
正	TP	FP
負	TN	FN

.

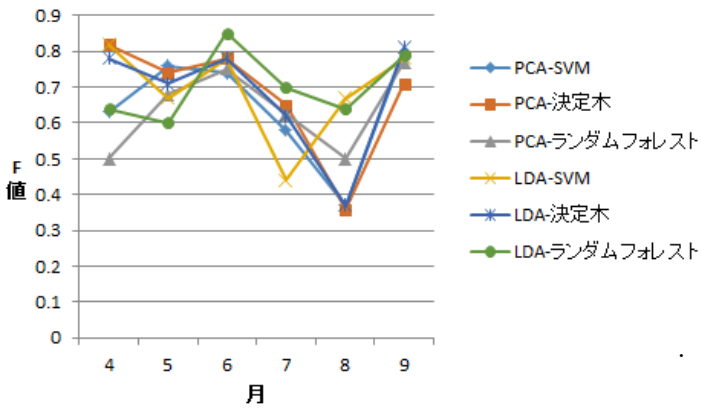
表 12　データセット 1 の成績推定精度 (F 値) の平均

	PCA-SVM	PCA-決定木	PCA-ランダムフォレスト	LDA-SVM	LDA-決定木	LDA-ランダムフォレスト
質問 1	0.68	0.74	0.77	0.60	0.73	**0.77**
質問 2	0.51	0.55	0.57	0.54	0.62	0.56
質問 3	0.52	0.49	0.74	0.56	0.75	0.55



.

図 5　データセット 1 の質問 3 に関しての成績推定 (F 値)



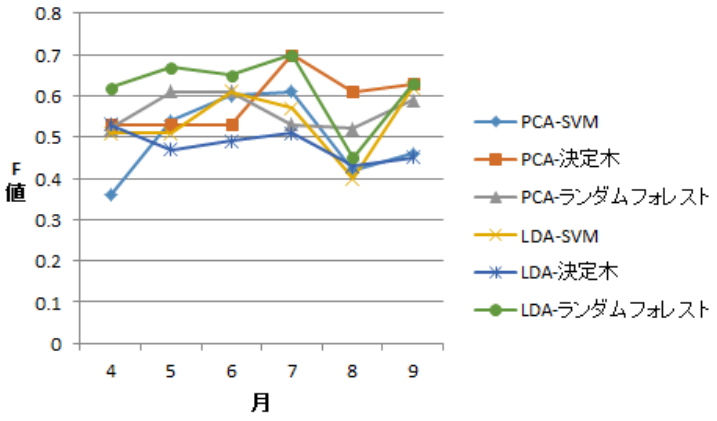
.

図 6　データセット 2 の質問 1 に関しての成績推定 (F 値)

c

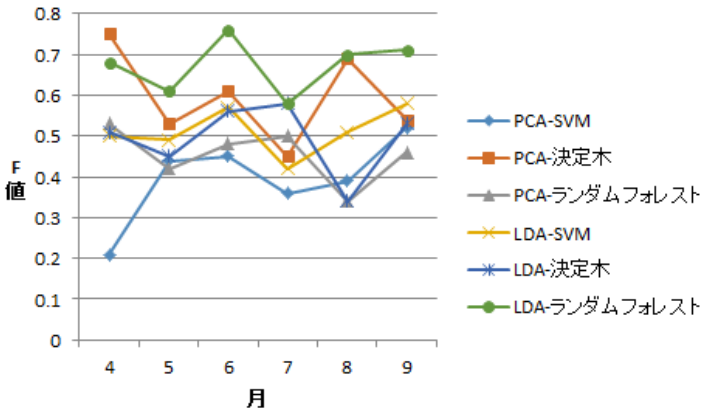
PCA で次元圧縮して SVM で成績推定を行った結果が最も大きい分散で精度にばらつきがあることを示した．質問 1 において PCA で次元圧縮を行いランダムフォレストで成績推定を行った結果が最も小さい分散でばらつきが小さく安定した精度を示した．

表 14 の結果からデータセット 2 では，質問 1 において LDA で次元圧縮してランダムフォレストで成績推定を行った結果が最も良い結果を示した．質問 3 において LDA で次元圧縮を行い SVM で成績推定を行った結果が最も悪い



.

図 7　データセット 2 の質問 2 に関しての成績推定 (F 値)



.

図 8　データセット 2 の質問 3 に関しての成績推定 (F 値)

精度を示した．

表 15 結果からデータセット 2 では質問 3 において PCA で次元圧縮して SVM で成績推定を行った結果が最も大きい分散で精度にばらつきがあることを示した．質問 1 において PCA で次元圧縮を行いランダムフォレストで成績推定を行った結果が最も小さい分散でばらつきが小さく安定した精度を示した．

7.2　考察

今回は次元圧縮方法として PCA と LDA を使い，機械学習の手法として SVM，決定木，ランダムフォレストを使い

.

表 13　データセット 1 の成績推定精度 (F 値) の分散

月	PCA-SVM	PCA-決定木	PCA-ランダムフォレスト	LDA-SVM	LDA-決定木	LDA-ランダムフォレスト
質問 1	0.015	0.016	0.001	0.014	0.004	0.009
質問 2	0.034	0.021	0.011	0.015	0.020	0.021
質問 3	0.079	0.010	0.026	0.030	0.019	0.010

.

表 14　データセット 2 の成績推定精度 (F 値) の平均

月	PCA-SVM	PCA-決定木	PCA-ランダムフォレスト	LDA-SVM	LDA-決定木	LDA-ランダムフォレスト
質問 1	0.70	0.69	0.78	0.60	0.49	**0.78**
質問 2	0.51	0.56	0.58	0.60	0.47	0.56
質問 3	0.53	0.49	0.57	0.48	0.50	0.56

成績推定を行ったがデータセット，質問項目，コメント取得月の全てに対して安定し，かつ高精度で成績推定できた手法はなかった．また時期によって成績推定の精度にばらつきが生じることも分かった．表 12 と表 14 より最も高い推定精度は質問 1 のコメントから LDA で次元圧縮使いランダムフォレストで成績推定を行った手法であった．本研究の実験結果では，成績の推定精度が上下した．これは，先行研究は，一つの授業で，かつ成績推定期間も，その授業が行われている期間（高々 4 か月）と短いのに対して，本研究で対象としたデータでは，複数の授業科目のデータを区別せずに利用していることと，成績推定を行う期間が長かったことから，先行研究よりも多くの要因が影響し，それらをうまく取り扱うことができていないためと考える．その一つに，成績推定の基準がある．具体的には，先行研究は最終成績のみを成績推定に使っていたが，本研究では複数の試験結果を使って成績推定している．質問ごとに見てみると質問 1 のコメントを使った成績推定の精度が最も良かった事がわかった．質問 1 はほかの質問より成績推定に寄与していると考えられる．

8. おわりに

ここでは，本論文のまとめと今後の課題を述べる．

8.1 まとめ

本研究では中学生の学習能力及び学習態度の向上のための成績推定モデルの構築の第一歩として，中学生の振り返り文から様々な手法を使い，成績推定を行った．成績推定では，1)PCA または LDA を使い次元圧縮を行い，SVM，決定木，ランダムフォレストを使った場合にでは推定実験の結果，質問の種類，推定期間など，すべてにおいて，他手法よりも安定し，高い推定精度を取る手法はなかった．2) データセット，質問，時期ごとに推定精度にばらつきが生じることが観測された，3) 高い推定精度が得られた手法はデータセット 1，2 いずれも次元圧縮で PCA を使った手法であった，4) 質問 1 のコメントを使った成績推定精度はほかの質問を使った成績推定精度より高かった，といった知見が得られた．

8.2 今後の課題

今後の課題として，

- 機械学習の方法の検討
- 次元圧縮方法の検討
- 有効な重みづけ手法の調査

などが挙げられる．

謝辞

本研究の一部は JSPS 科研費 26540183 の助成を受けて行われました．感謝をを申し上げます．

参考文献

[1] SHAYMAA EZZELARAB MOHAMED SOROUR.:A Study of Comment Data Mining to Predict Student Performance. 九州大学学位論文 p62,p163-164(2016).
[2] B.K.Bhardwaj and S.Pal.:Data Mining : A prediction for performance improvement using classification.International Journal of Computer Science and Information Security(IJCSIS),vol.9,n0.4,pp136-140,2011.
[3] S.K.Yadav and S.Pal.: Data mining : A prediction for performance improvement of engineering students using classification.World of Computer Science and Information Technology Journal(WCSIT),vol.2,no.2,pp.51-56,2102.
[4] Y.Qiu,Y.Qi,H.Lu,Z.Pardos,and N.Hefferman.: Does time matter? modeling the effect of time with bayesian knowledge tracing. Proceeding of the 4th International Conference on Educational Data Mining(EDM 2011),2011,pp.139-148.
[5] J.Tane,C.Schmiz,andG.Stumme.:Semantic resource management for the web: an e-learning application. Proceeding of the 13th International World Wide Web conference on Alternate track papers posters.ACM,2004,pp.1-10.
[6] 奥村 学:自然言語処理シリーズ 8 トピックモデルによる統計的潜在意味解析. コロナ社.(2015).
[7] 石田 基広:R によるテキストマイニング入門. 森北出版.(2012).
[8] 北 研二，津田 和彦，獅子堀 正幹:情報検索アルゴリズム. 共立出版.(2003).

.

表 15 データセット 2 の成績推定精度 (F 値) の分散

月	PCA-SVM	PCA-決定木	PCA-ランダムフォレスト	LDA-SVM	LDA-決定木	LDA-ランダムフォレスト
質問 1	0.017	0.003	0.001	0.008	0.020	0.008
質問 2	0.007	0.005	0.004	0.007	0.006	0.007
質問 3	0.035	0.005	0.012	0.008	0.028	0.007

[9] 東京工芸大学: 適合率 (precision) と再現率 (recall) ”http://www.cs.t-kougei.ac.jp/SSys/Pre_Rec.htm”2017/2/1
[10] データ分析相談所:クロスバリデーションは一つだけではない ”http://univprof.com/archives/16-11-24-8652578.html”2017/2/4