

電子教材の閲覧データとコンテンツ 情報を用いた学習者のスコア予測

2023/2/13

兵庫県立大学 社会情報科学部 社会情報科学科

指導教員 川嶋 宏彰教授

JB19S029 小岸沙也加

目的

学生の講義資料の閲覧行動から教材の理解度を推定する

⇒ 小テスト・定期テストの前の段階で学生にアプローチできる

入力：閲覧ログデータ、コンテンツ情報（講義資料）

出力：小テストごとに予測した点数

評価：小テストごとに求めたRMSEの平均

⇒ コンテンツ情報を使用することでどこまで精度があがるのか

使用データ

九州大学講義（2020年 情報系科目）

閲覧データ（講義回数：7回、対象学生：100名、200,818ログ）

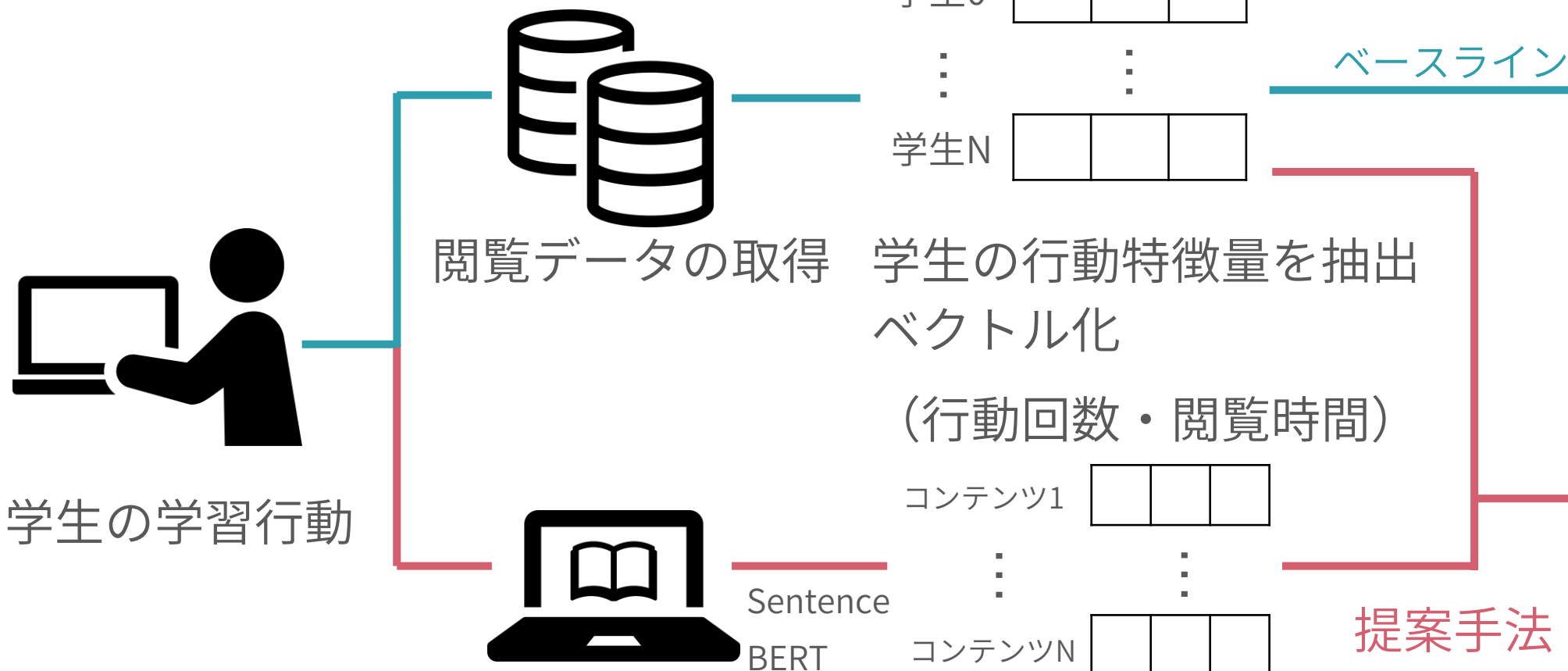
コンテンツのテキスト情報

2週目に2回
行われている

小テストのデータ（小テスト回数：8回、5問択一式）

（学生番号・問題文・学生の選択・正解かどうか・提出時間）

全体像



閲覧データのみ
を使用した場合

学生	予測点数
0	5
1	4
2	3

LightGBMで
スコアを予測
RSMEで評価

学生	予測点数
0	5
1	3
2	5

閲覧データとコンテンツ
を使用した場合

ベースライン

閲覧データから各ページにおける各操作の操作回数、閲覧時間を求める

⇒ 行動特徴ベクトル $\mathbf{u}_c^{(i)}$

行動特徴ベクトル $\mathbf{u}_c^{(i)}$ のみを使用

学生の行動を

「講義時間外」も含めて

「講義時間内＋前後1時間」に絞って

次元数はページ数 \times (12 (行動) + 1 (ページ閲覧時間))

User id	Open 1	Close 1	...	Next 15	Prev 15
0	3	0		4	3
1	2	1		3	2
2	0	0		2	1

↑ 行動特徴ベクトル $\mathbf{u}_c^{(i)}$

i : 学生

C : コンテンツ

学生がよく閲覧したページ情報を使用 (提案手法)

行動の中でもページ閲覧時間に注目

「閲覧時間の長いページのコンテンツ情報を多く含むベクトル」

⇒ 学生 i のコンテンツ c に対する「閲覧コンテンツベクトル $\mathbf{v}_c^{(i)}$ 」

行動特徴ベクトル $\mathbf{u}_c^{(i)}$ と閲覧コンテンツベクトル $\mathbf{v}_c^{(i)}$ を使用

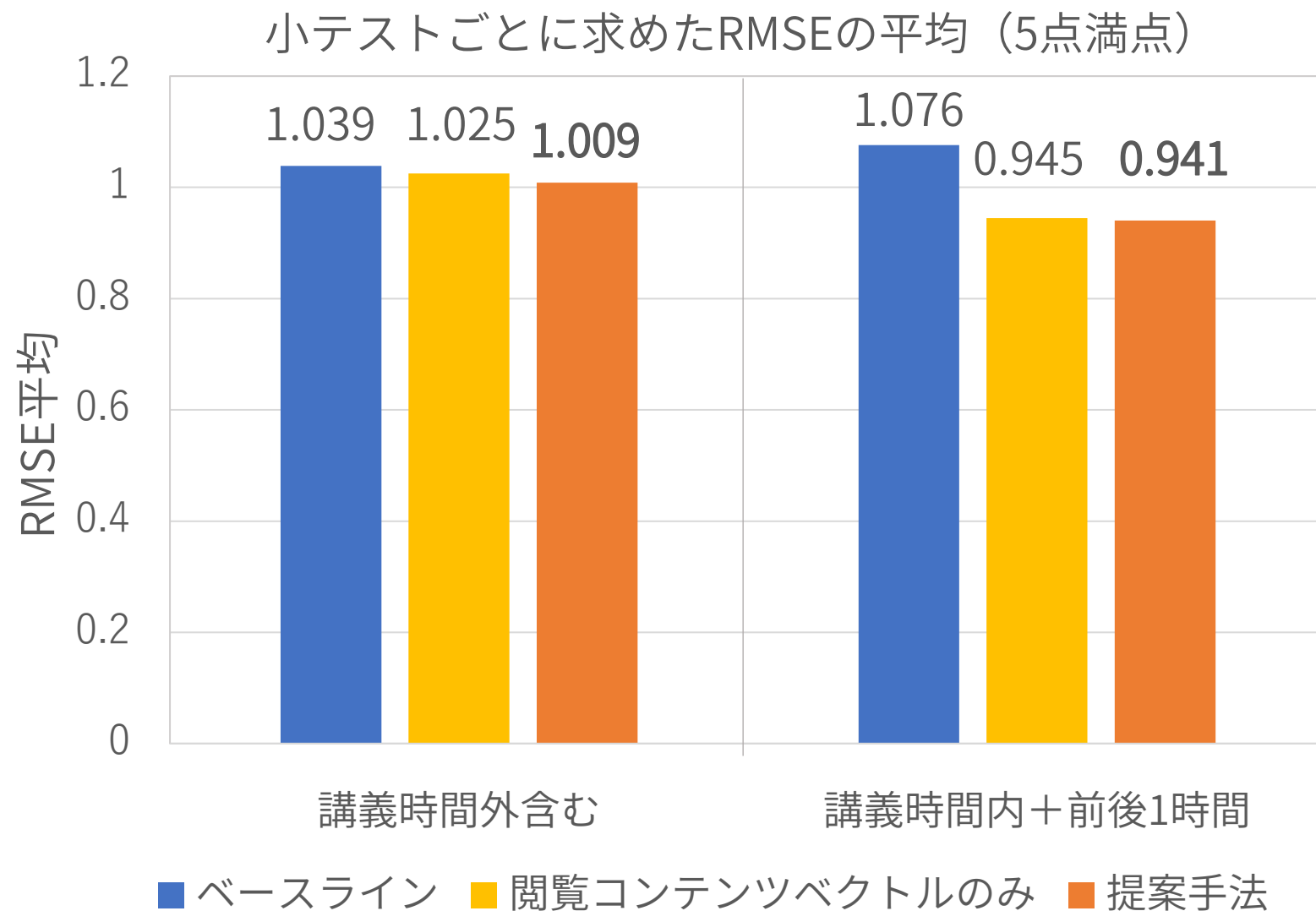
$$\begin{array}{ccccccc} \text{ページベクトル} & & \text{閲覧時間} & & \text{閲覧コンテンツベクトル} & & \begin{array}{l} i : \text{学生} \\ c : \text{コンテンツ} \\ p : \text{ページ} \end{array} \\ \mathbf{v}_{(c,p)} & \times & t_{(c,p)}^{(i)} & \longrightarrow & \sum_p t_{(c,p)}^{(i)} \mathbf{v}_{(c,p)} & = & \mathbf{v}_c^{(i)} \end{array}$$

1回の閲覧時間が5分より長く開いていたページは放置されたものとして省く

結果

5-fold 交差検証でRSME
を計算

講義時間内＋講義前後の
行動がより予測に関わる
(コンテンツ含める場合)



まとめ

コンテンツ内容を含めることは精度向上に繋がる
学生が長く閲覧したコンテンツ情報を用いることはより重要

今後の課題

ベクトルの次元数を減らす

ページベクトルに画像の情報や配置の情報を含める

問題1問ごとに正解 / 不正解を予測する