

# 電子教材の閲覧データと コンテンツ内容を用いた点数予測

川嶋研究室 小岸沙也加

# 背景

講義ではオンライン上で講義資料が閲覧できる機能が使われる

詳細な閲覧データを取得することができる



**学生の行動から理解度が推定できれば  
はやめのアプローチが学生にできるのでは？**

# 問題設定

## 電子教材の閲覧データと コンテンツ内容を用いた点数予測

入力：講義回ごと to 取得した閲覧データ、コンテンツ内容

出力：小テストの予測した点数（小テストごと）

評価：RMSE

RQ

⇒ コンテンツ内容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 サイバーセキュリティ基礎論）

閲覧データ（講義回数：7回、100名、200,818ログ）

コンテンツ画像・内容（npzファイル）

小テストのデータ（7回）

（学生番号・問題文・学生の選択・正解かどうか・提出時間）

Log id	Contents id	Contents name	...	Operation date	Operation name	Page no	...	User id
331	1	イントロダクション		2020/05/12 12:55:15	OPEN	1		0
332	1	イントロダクション		2020/05/12 12:56:00	NEXT	1		0
333	1	イントロダクション		2020/05/12 12:56:15	NEXT	2		0
334	1	イントロダクション		2020/05/12 12:56:20	PREV	3		0
335	1	イントロダクション		2020/05/12 12:57:00	ADD_MERKER	2		0
336	1	イントロダクション		2020/05/12 12:57:04	NEXT	2		0

...

## 小テストデータ

Quest ionid	State	Time	Question summary	Response summary	User id
333	Right	2020/05/12 14:55:15	ICT について～	コンピュータセキュリティに ついて～	30
336	Wrong	2020/05/12 14:55:15	パスワードの扱いについて～	異なっていれば～	30
339	Right	2020/05/12 14:55:15	セキュリティについて～	設計段階から～	30
343	Right	2020/05/12 14:55:15	日本国政府の～	サイバーセキュリティ戦略 ～	30
349	Right	2020/05/12 14:55:15	サプライチェーンに関する正し い説明を～	取引の連鎖には～	30
332	Right	2020/05/12 14:57:04	ICT について～	メール送信のタイミング～	26

⋮

# 全体像

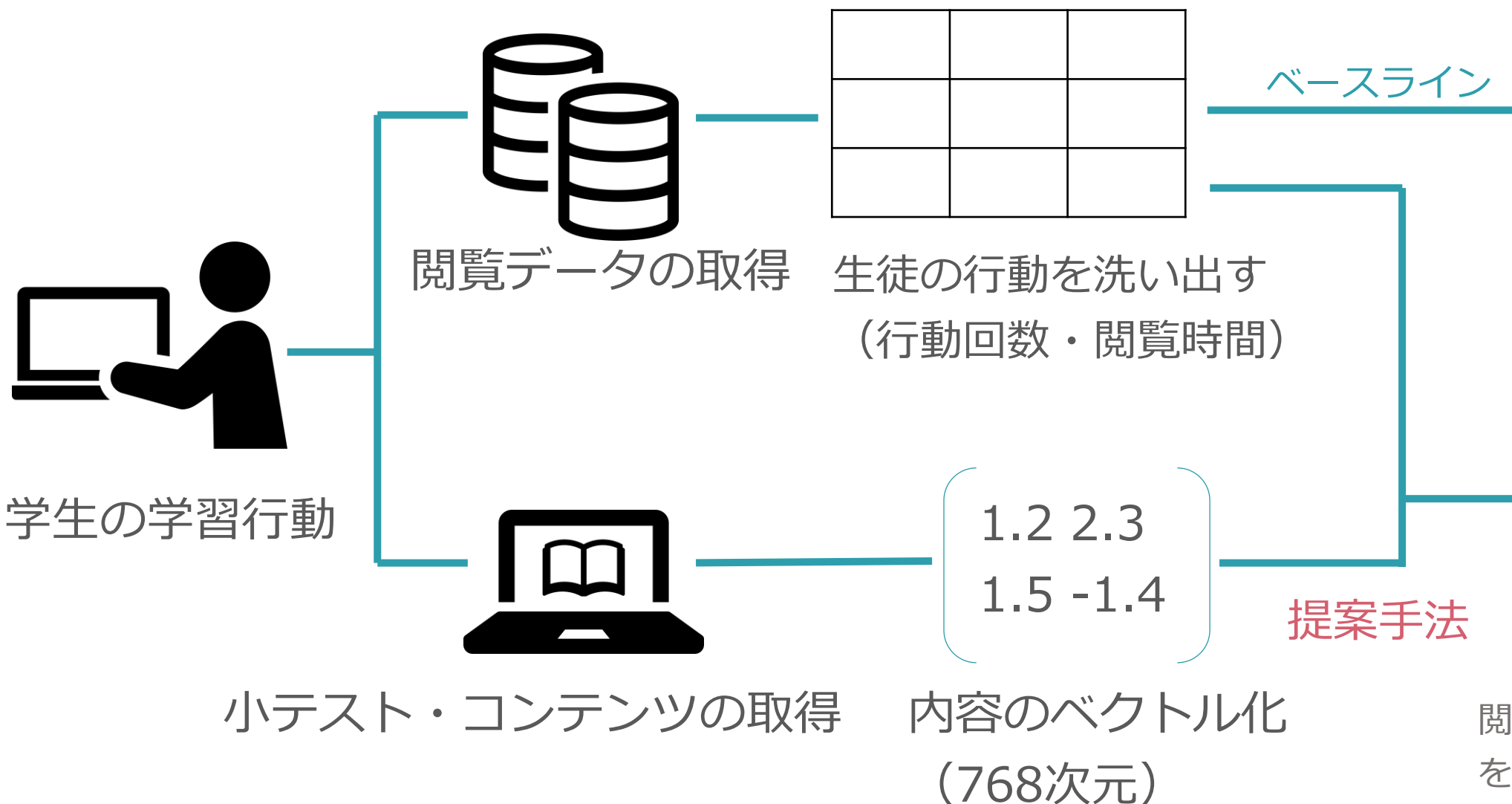
閲覧データのみを使用した場合

学生	予測点数
0	5
1	4
2	3

LightGBMで  
点数を予測  
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



# アプローチ

## ベースライン

閲覧データのみを使用

行動を講義時間外も含めて

講義時間内 + 前後1時間絞って

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

↑ 行動特徴量

## 提案手法

閲覧データとコンテンツ内容を使用

### 手法 1:

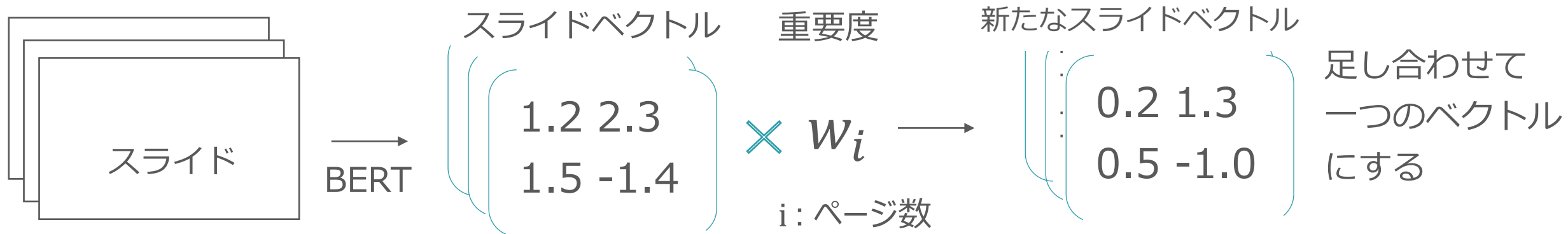
コンテンツ内容を行動特徴量に入れる

### 手法 2:

小テストに関するスライドにおける  
行動だけとりだす



# 手法 1 : スライドベクトルを行動特徴量に入れる



閲覧時間が長いスライドほど重要度を高くする

重要度は連続値

閲覧時間が長すぎる部分は省く (セッションタイムアウト等)

## 手法2：小テストとの関係を重視する

小テストベクトルとスライドベクトルから求めたコサイン類似度を使用

コサイン類似度0.4以上の和を重みとする

↓ 行動特徴量

contents page	q1	q2	q3	q4	q5	$w_i$
1	0.3	0.4	0.3	0.6	0.7	1.7
2	0.6	0.3	0.2	0.1	0.5	1.1
3	0.5	0.5	0.1	0.7	0.8	2.5
4	0.1	0.2	0.1	0.1	0.3	0.0

⋮



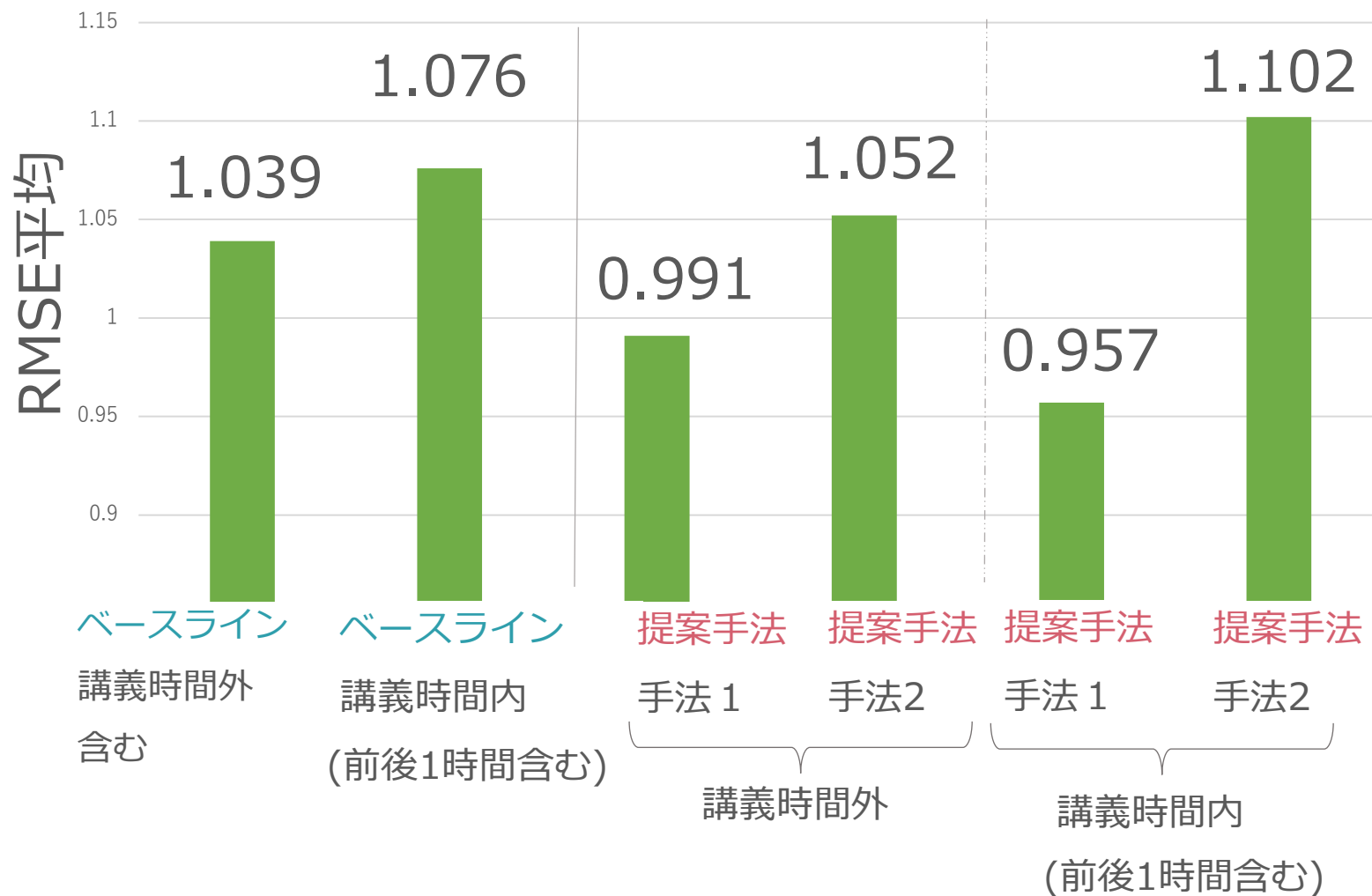
User id	Open 1	Close 1	...	Next 15	Prev 15
0	3	0		4	3
1	2	1		3	2
2	0	0		2	1
3	1	1		5	1
	×	×		×	×
重要度	$w_1$	$w_1$		$w_{15}$	$w_{15}$

# 結果

閲覧時間がより重要  
講義時間内外を絞るのは  
関係なさそう

コンテンツ内容を含める  
ことは精度向上に繋がる  
のではないかな

週ごとに求めたRMSEの平均



# 今後の計画

手法2は各学生に対し同じ値を重みとして使用するためかベースラインと結果があまり変わらなかった

行動の重要度及びスライドの重要度によって重みを変える  
⇒ 学生・行動によって重みが変わるので結果が期待できる

## 計画

10月～11月 上の手法を進める・論文執筆

12月～2月 論文執筆、修正、提出

## 結果(P11の補足)

小テスト	ベースライン	ベースライン		提案手法	提案手法		提案手法	提案手法
1	1.078	0.921		0.989	1.049		0.779	0.965
2-1	0.740	0.868		1.172	0.730		1.180	0.910
2-2	1.346	1.430		0.808	1.349		0.714	1.361
3	1.176	1.169		0.494	1.176		0.404	1.157
4	0.867	0.962		1.082	0.909		1.143	0.942
5	0.987	0.932		0.883	0.929		0.858	1.047
6	0.790	0.855		0.969	0.790		1.268	0.920
7	1.324	1.469		1.530	1.484		1.324	1.511

# 講義について

講義は7週間

1週目と2週目は二つのコンテンツを用いている

講義開始から80分授業、残りの10分で小テスト

2週目のみ小テストが2回行われている

小テストは一回につき5点満点、4択で一問一答形式