

教育データ × 教育データ

島田敬士

九州大学 大学院システム情報科学研究所 教授

九州大学 ラーニングアナリティクスセンター センター長

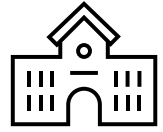


KYUSHU UNIVERSITY

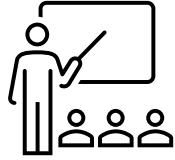
① 成績予測・At-risk学習者予測

② 閲覧行動と学習コンテンツ分析

研究背景：成績予測・At-risk学習者予測

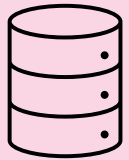


学校

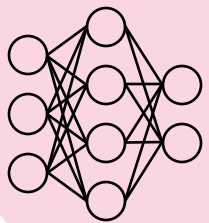


授業

- 学習困難者の早期特定，早期介入
- 学習活動と成績の関係を分析／モデル化



教育データ



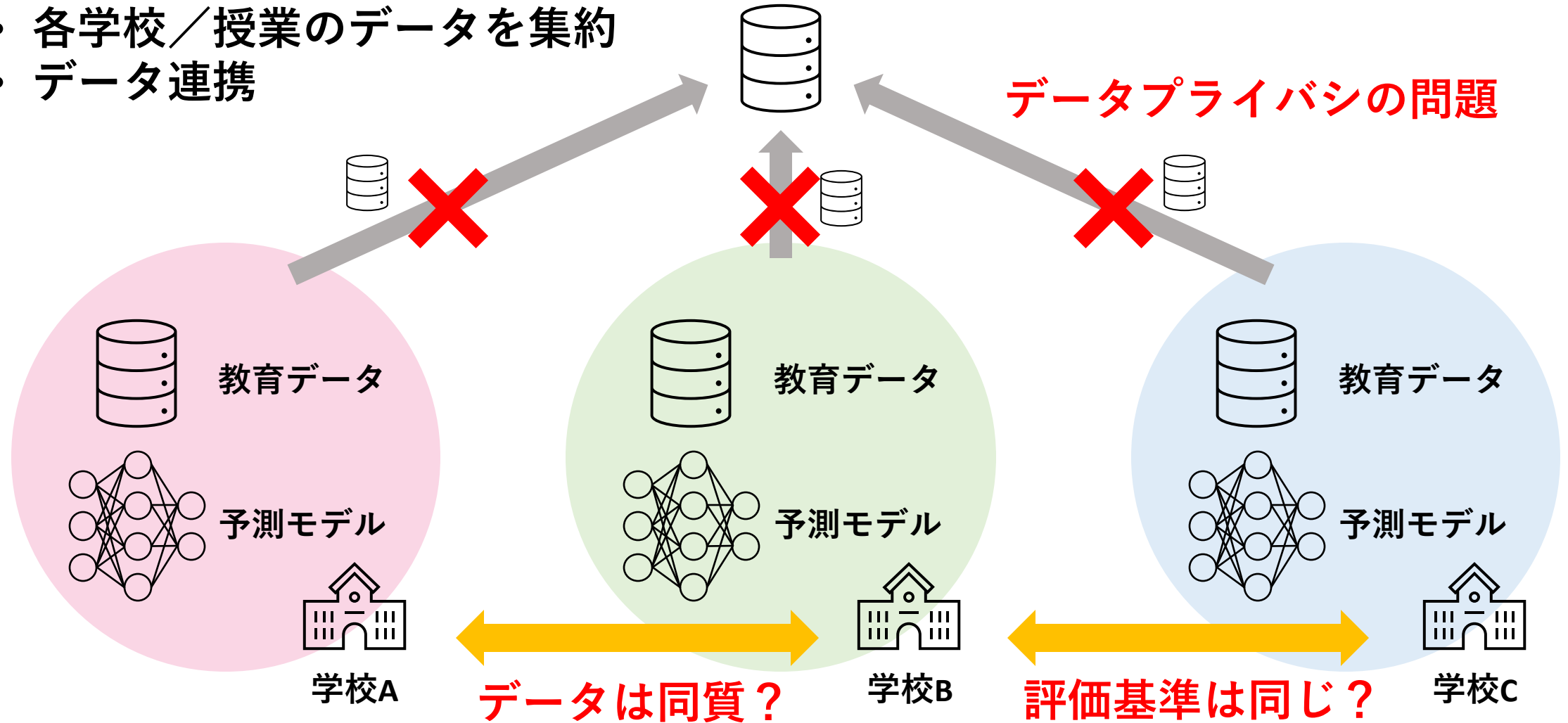
予測モデル

課題

- 各学校・授業で収集されるデータは小規模
 - モデルが低性能
- 学校／授業単位でモデルを構築
 - モデルの準備が非効率

データ連携と課題

- 各学校／授業のデータを集約
- データ連携



課題1：データの横断利用

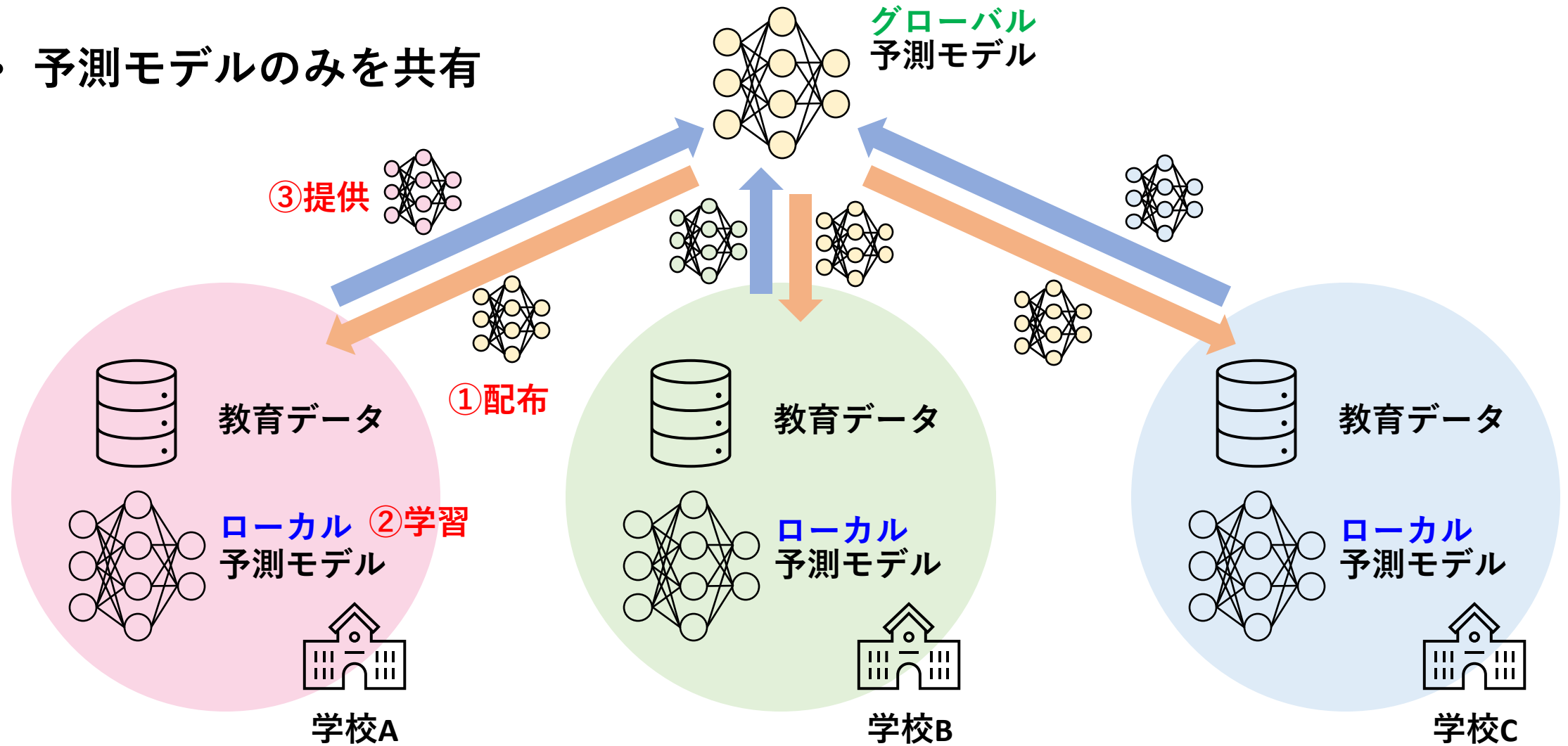
- 組織間，コース間，授業間でデータプライバシーを保証
- 限られたリソースでモデル構築の効率化

課題2：異質データへの対応

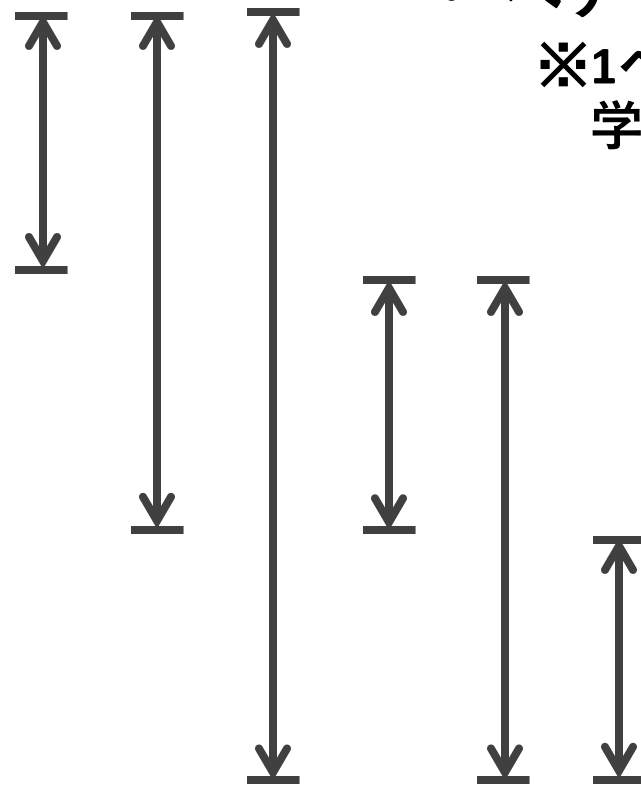
- 特に，データ間の評価基準の差異を吸収
- At-risk検知の観点では，検出対象者の基準が曖昧

課題1の解決策：連合学習

- 予測モデルのみを共有



課題2の解決策：相対特徴の導入



- 絶対指標（評価）に左右されにくい
- ペアワイズでデータ量を増強
※1ペアで2種の相対特徴を生成
学習者A→B, 学習者B→A

例：30人クラス
30サンプル → 870ペアサンプル



活動の差

20点(85-65)

成績の差

評価結果：連合学習の有効性

- At-risk学生の検出精度で検証
- リスク順に対象者をソートして，Top n Precisionを評価

評価指標	予測データ： 人数	評価データ①		評価データ②	
		連合	一括	連合	一括
Top n precision	n = 5	1.0	1.0	1.0	1.0
	n = 10	0.70	0.90	1.0	1.0
	n = 15	0.60	0.80	0.93	0.93
	n = At-risk数	0.63	0.58	0.76	0.76
AUC		0.65	0.68	0.82	0.84

※評価用コースを除く14コース分のデータで学習

※両手法で相対特徴を利用．評価データ①，②のNo-risk/At-risk者数は，それぞれ28/24名，68/25名

連合学習を導入しても一括型学習と同等の性能を実現可能

評価結果：相対特徴の有効性

- At-risk学生の検出精度で検証
- リスク順に対象者をソートして，Top n Precisionを評価

評価指標	予測データ： 人数	評価データ①		評価データ②	
		相対	絶対	相対	絶対
Top n precision	n = 5	1.0	0.60	1.0	1.0
	n = 10	0.70	0.70	1.0	0.90
	n = 15	0.60	0.67	0.93	0.67
	n = At-risk数	0.63	0.46	0.76	0.73
AUC		0.65	0.52	0.82	0.71

※評価用コースを除く14コース分のデータで学習

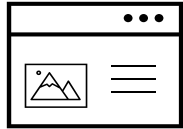
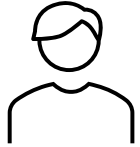
※両手法で連合学習を利用．評価データ①，②のNo-risk/At-risk者数は，それぞれ28/24名，68/25名

相対特徴の導入で性能向上

① 成績予測・At-risk学習者予測

② 閲覧行動と学習コンテンツ分析

研究背景：電子教材の閲覧行動分析



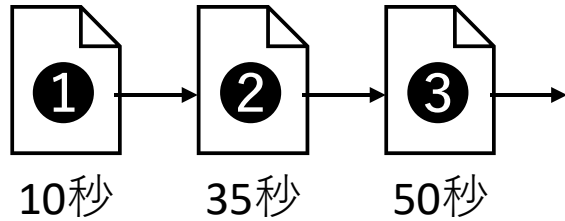
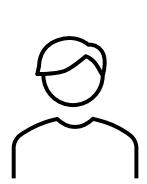
Student ID	Contents ID	page	operation	Time stamp	...
X-2022_Cn	X-2022_Ua	1	NEXT	2022-04-17 hh:mm:ss	...
X-2022_Cn	X-2022_Ua	2	NEXT	2022-04-17 hh:mm:ss	...

- 電子教材の操作履歴を保存
- 学習者の閲覧行動として分析
 - ページ単位, 全体の閲覧時間
 - マーカー, メモ
 - 注目ページ

課題

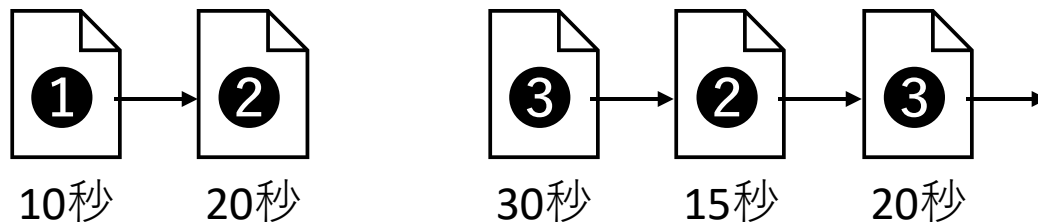
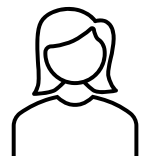
- 閲覧行動の上手な捉え方は？
 - 閲覧時間集計では「どのように」閲覧したか分からない
- 学習内容まで把握できる？
 - ページ番号のみでは「何を」学習したか分からない

閲覧行動と学習コンテンツの横断分析の課題



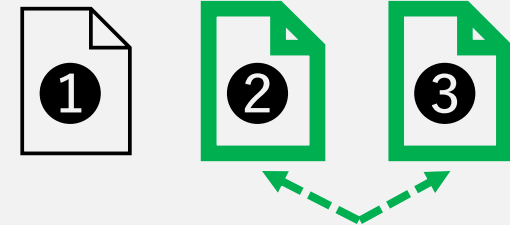
- ✓ 閲覧時間を集計すると二者間を区別不可能
 - 閲覧順序の欠如
 - 学習タイミングが無視

時間
集計



中断

読み返し



内容が関連

- ✓ ページ番号では中身が分からない
 - 同内容の学習時間
 - その学習の進め方

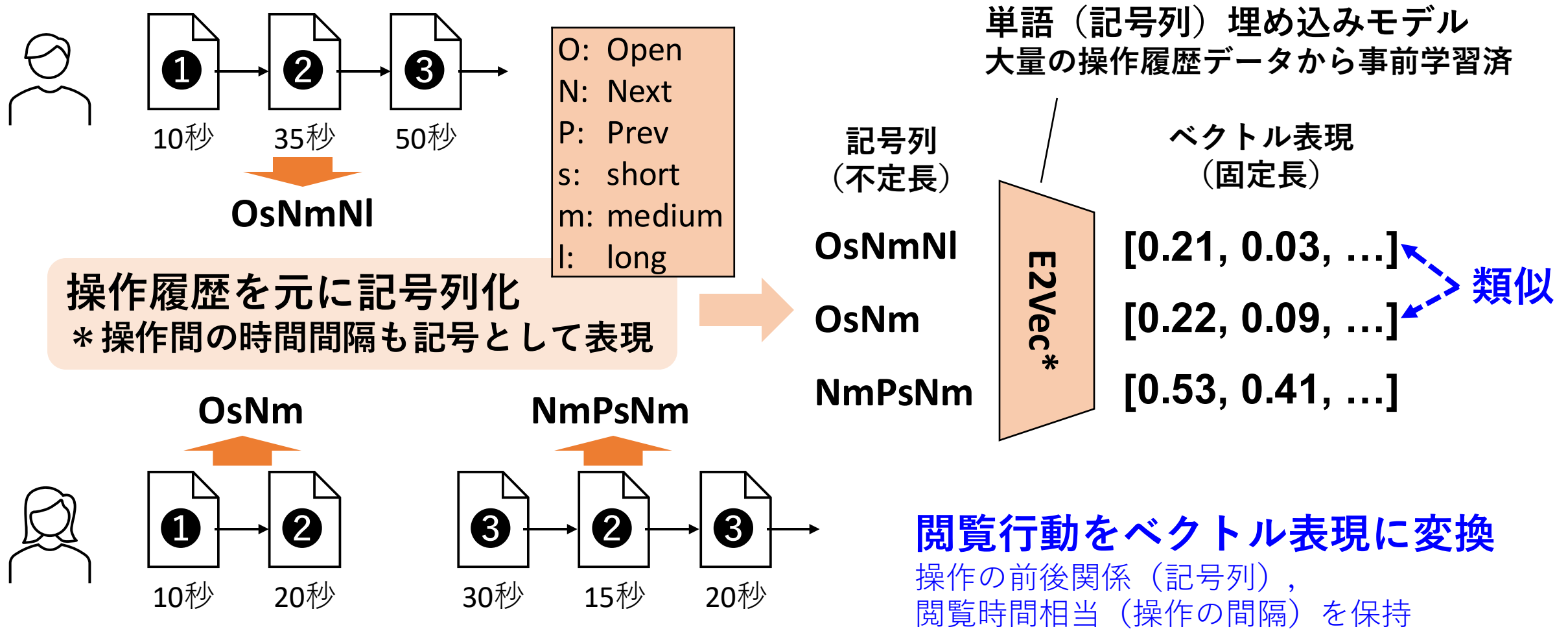
課題1：閲覧行動の特徴表現

- 閲覧順序，閲覧時間，閲覧タイミングを考慮
- 類似した閲覧行動には類似した特徴を付与

課題2：学習コンテンツの特徴表現

- 学習内容（ページの中身）を特徴化
- 類似した内容のページには類似した特徴を付与

課題1の解決策：閲覧行動の分散表現



課題2の解決策：学習コンテンツの分散表現


標本化レベルと画素数の関係 25

細かく標本化するほど、元の画像に近いデジタル画像が生成される



粗い ← 標本化 → 細かい
少ない ← 画素数 → 多い

第1回：デジタル信号処理の概要



内容説明
図表、構図
にも対応

標本化レベルと画素数についての関係を示しています。細かく標本化するほど、元の画像に近いデジタル画像が生成されることを説明しています。
左から右へ向かって標本化が粗い状態から細かい状態へ変化しており、それに伴って画素数も少ない状態から多い状態へ変化していることが視覚的に示されています。

Text Embedding

Open AI: text-embedding-3-largeを利用

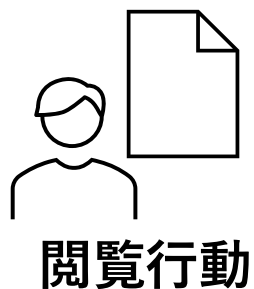


Embed.

[0.24, 0.19, ... , 0.04]

内容をベクトル表現に変換

ここまでの整理



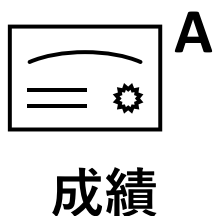
Student ID	Contents ID	pages	unit	Unit vector
X-2022_Ua	X-2022_C1	[2]	Nm	[0.21, 0.03, ...]
X-2022_Ua	X-2022_C1	[3,4]	NsNm	[0.25, 0.12, ...]
...

- ページの読み進め方を数値として表現
- 行動が類似している学習者の探索が容易



Contents ID	Page num	content vector
X-2022_C1	1	[0.01, 0.43, ...]
X-2022_C1	2	[0.22, 0.52, ...]
...

- 教材のコンテンツ（中身）をページごとに数値として表現
- 内容が類似しているコンテンツの探索が容易



Student ID	grade	At-risk
X-2022_Ua	A	0
X-2022_Ub	D	1
...

横断分析！！

成績高群と低群で、読み進め方が異なる学習コンテンツは??

分析例：プログラミング科目

大

成績高群と低群の閲覧行動の差異

小

ページ群A

ページ群B

ページ群C

仕事の分割
この仕事には

プログラム設計法
プログラム開発に必要な活動

プログラム設計法: Contract, Header

Contract の定式化のために使うことができるもの

- number や symbol のような不可分データのクラス名
- posn のようなデータ定義で導入した名前

```
;; distance_to_0: posn -> number  
(define (distance_to_0 a_posn) ... )
```

本日の内容

構造体の定義(一般形)

様々なデータ

- 単純データ
 - 数値、ブール値、シンボル、文字列、...
- 合成データ: 複数要素で1つのデータを構成
 - 構造体 (structure): 固定数の要素で1つのデータ
 - 2次元平面座標(x,y座標)、名簿の一人分の項目数、...
 - リスト(後日説明): 不定数の要素で1つのデータ
 - 窓口の行列の人数、...

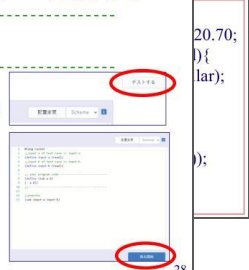
例題1. ベクトルの内積

演習1. ドルから円への変換

TechFULでの課題取組

- 問題文をよく読む
- 「your program code」で囲まれた範囲内にコードを記入する

```
;; your program code  
このエリアに自分のコードを書く  
このエリアに自分のコードを書く
```
- TechFULのエディタ横の「テストする」ボタンでサンプルケースに対して正しい振る舞いをするかどうか確認する
- サンプルケースの実行に成功したら、TechFULのエディタ下の「採点開始」ボタンを押して、テストケースに対する結果を得る



プログラム設計法

データ形式

演習・例題

授業内容の理解に重要なページ群

- 高群はじっくり読む
- 低群はページ操作が速い

閲覧行動データから取得可能

高群・低群で閲覧行動にあまり差が無い

① 成績予測・At-risk学習者予測

- プライバシーを保護しつつ，予測精度を向上
- 連合学習，相対特徴を導入

② 閲覧行動と学習コンテンツ分析

- 教材の読み進め方，教材の中身まで詳しく分析
- 『かゆいところに手が届く』分析結果を提供

ご清聴ありがとうございました



KYUSHU UNIVERSITY