



主催 DS⁴ 運営委員会

「データ科学におけるPBL学習の効果と限界」

日 程： 2022/03/11(金) 14:00-15:00

開催方法： Zoomによるウェビナー形式

プログラム

- 実践データ科学演習@神戸大学
講演者： 斎藤 政彦
- データ科学PBL@大阪大学
講演者： 狩野 裕
- 討論

1

第7回 データ科学教員のための教育FD会議

多くの大学でデータ科学のPBLが行われているようになった。しかし、新しい試みであるが故、適切な実施方法を模索しているのが現状であろう。そこで、今回の教育FD会議では、データサイエンティスト育成事業「DS⁴」で現在行われているPBL形式の授業を二件紹介することで、楽しく有効なPBL授業の構築に向けての話題提供としたい。

2

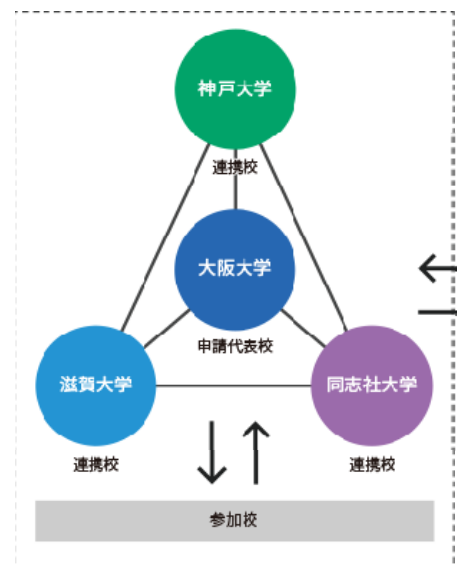
データ科学PBL@大阪大学

講演者： 狩野 裕

3

「DS⁴」カリキュラム(一部)

- 独り立ちデータサイエンティスト人材育成プログラム(DS⁴)
- 修了要件単位数
 - 12単位(6コマ)
 - 修了証を交付
- 選択必修科目A群(6単位以上)
 - DSインターンシップ
 - 実証型研究法
 - データ科学PBL@大阪大
 - データ科学各論
 - 数理特論III(意思決定とデータ科学)
 - 実践データ科学演習A,B@神戸大
- 選択必修科目B群(4単位以上)
 - …
- 選択科目
 - …



4

Problem-Based Learning (PBL)

- cf. Project-Based Learning
- Active Learning(AL)の一つ
 - 学修者が能動的に学習していく教育方法
- 自発的学習と小グループ学習が特徴
- 与えられた問題について、解決方法を能動的に模索しゴールに向かうことによる学習方法
 - 問題自体を発見させる場合もある. 社会問題も対象
 - 問題解決に既存知識を活用させるトレーニング. 必要な新知識は自習
 - 学術的理解をふまえて問題解決へ
- 試行錯誤のプロセスに意味がある
 - ゴールが無くてもよい?
→最後に、考えられるベストなプロセスとゴール(の一つ)を提示する方がよい
- 解決のプロセスに未習事項が含まれる
 - 問題解決が優先されるため、中途半端な理解で結果だけを使うことが横行する
→欠点
 - 目的的な学習の欠点

5

系統(的)学習

- 系統(的)学習(systematic study)
 - 伝統的な学習方法でPBLと対比される
 - 既習事項の上に未習事項を学習する
 - 体系的な知識が効率よく身に付く
 - 具体的な問題を解決できるとは限らない→欠点
 - どの知識をどのように活かすか、が身に付いていないことがある
 - 問題解決に未習事項が必要になると、その責をカリキュラムに帰属させることがある
- 参考
 - <http://www.cscd.osaka-u.ac.jp/user/rosaldo/061127pbl.html>

6

データ科学におけるPBL・実習演習で強調されていること

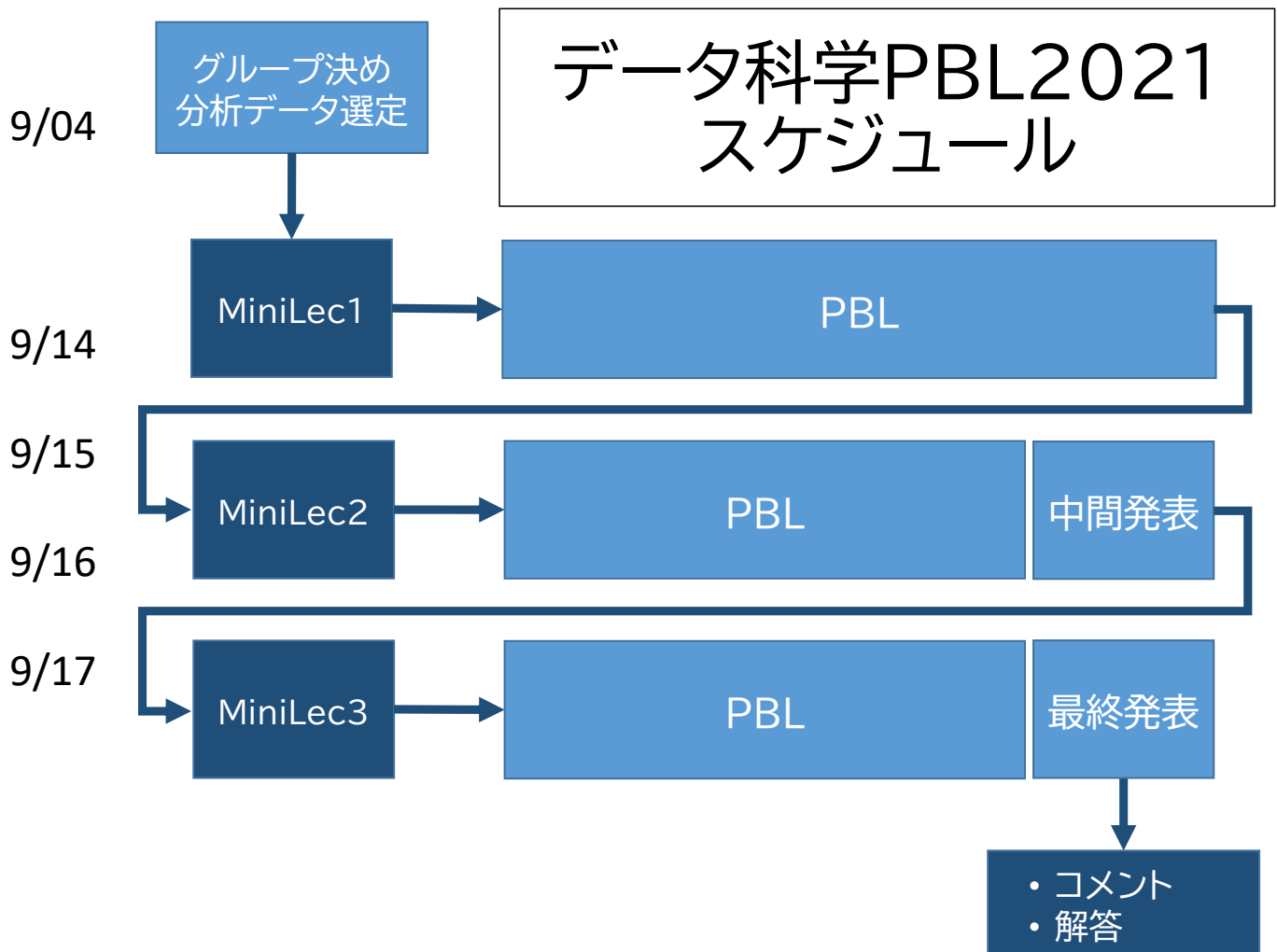
- **リアルデータ**
 - 企業様から提供された(ビッグ)データ
 - 受講生に好評
 - 他
- **課題発見・課題解決**
 - 業務上の課題
 - 社会的課題(e.g., SDGs)
 - 要人・賓客の前でのプレゼン
- **社会と大学との関係強化**
 - 社会の課題解決へ資する
 - 社会による大学への貢献も
 - 象牙の塔に籠るな
 - 社会のニーズに合う教育
 - イノベーションを産む人材育成

7

DS⁴のデータ科学PBL

- 三泊四日合宿形式
 - 大学教室での集中実習(コロナ禍による)
- 履修者:約25名, 4名×6グループ, 学生所属は様々
- PBLと系統的学習の併用により, データ科学を正しく理解し, データ科学による問題解決の方法を身に付ける
 - 与えられたデータから問題を発見し(課題を立て)解決へと導く
 - 与えられたデータと問題を解決へと導く
 - 関連したデータ科学の基礎事項をレクチャー等で補う
 - 最後に, 解決方法の一つを示し, 受講者へ反省と近未来の学修の方向性を示す
- 基本的に, 与えられたデータ以外の情報はWeb等から得て頂くが, 教員は関連する情報を持っているかもしれない. 教員に尋ねられたい

8



9

データ科学PBL2021 スケジュール

- 9月04日(土) グループ分け,
データ事前説明と選択
- 9月14日(火)
 - 受付(10:00~10:30)
 - Opening, MiniLec1
 - **SQLの使い方**
 - PBL(11:00~12:00)
 - 昼食
 - PBL(13:00~18:00)
 - 休憩適宜
- 9月15日(水)
 - MiniLec2(10:00~10:30)
 - **カテゴリカルデータ解析:対数線形モデル**
 - PBL(10:30~12:00)
 - 昼食
 - PBL(13:00~16:00)
 - 休憩適宜
 - 中間報告会(16:00~18:00)

- 9月16日(木)
 - MiniLec3(10:30~11:00)
 - **SEMIによる企業評価分析**
 - PBL(11:00~12:00)
 - 昼食
 - PBL・中間発表
(13:00~18:00)
 - 休憩適宜
- 9月17日(金)
 - 受付(10:00~10:30)
 - PBL(10:30~12:00)
 - 昼食(12:00~13:00)
 - 最終報告会(13:00~16:00)
 - 休憩適宜
 - 講評・解説・意見交換
(16:00~18:00)

写真撮影

10

2021年度データ科学PBLで 解析するデータリスト

- (A1) ブランドデータ
 - (A2) 人口流動データ
 - (B1) 失業率が1%増えると2000人が亡くなる!?
 - (B2) コロナ禍のデータ(履修生が収集・分析, 資料なし)
 - (C1) 素敵なつながり: 1994 and 2017
 - (C2) 水瓶座は宝くじに強い!?
 - (C3) イメージトレーニング効果の検証
 - (C4) 三段論法は正しいか
- } 2問選択

11

データの種別

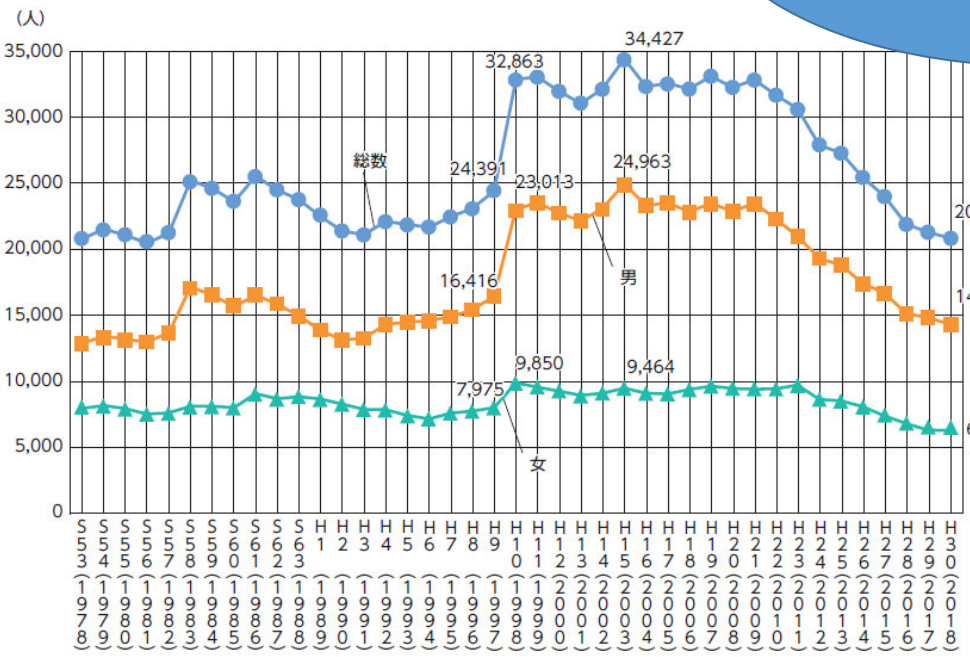
- A) 企業様の有するビッグデータ
 - 誓約書
- B) 社会的課題
 - データ: 受講生がWeb等から取得
- C) DSの考え方(Project-Based Learning)
 - データの観方
 - 実験データ
 - 基礎統計
 - 他

12

(B1) 失業率が1% 増えると...

失業率が1%増えると
(あってはなりません)
国全体で一年間、2000人
が亡くなると言われている

第1-1図 自殺者数の推移 (自殺統計)



資料：警察庁「自殺統計」より厚生労働省自殺対策推進室作成



(C4) 三段論法(推移律)は正しいか？

- “X→Y, Y→Z” ⇒ “X→Z”？
 - X: 運転者の性別
 - Y: 車の走行距離
 - Z: 車の修復歴(事故歴)

性別 × 走行距離

	短い	長い	
女	178	74	オッズ比 5.79
男	74	178	

走行距離 × 修復歴

	なし	あり	
短い	178	74	オッズ比 5.79
長い	74	178	

性別 × 修復歴

	なし	あり	
女	126	126	オッズ比 1.00
男	126	126	

性別	走行距離	修復歴		計
		なし	あり	
女	短い	115	63	178
	長い	11	63	74
	計	126	126	252
男	短い	63	11	74
	長い	63	115	178
	計	126	126	252

(人工データ)

(C4) 三段論法(推移律)は正しいか？

--本課題によって習得できる事柄--

- (統計的)因果分析の重要性の理解
 - 学術であれ実務であれ
- 効果の分解の概念の習得とその役割の理解
 - 直接効果, 間接効果, 総合効果
 - 相関係数, 偏回帰係数...
 - 交絡変数の影響
- モデルの局所評価と全体評価の重要性
 - 三段論法が不適であることの見つけ方
- 条件付き独立の意味とデータ分析での使い方
- 対数線形モデルによる分析

19

データ分析に一意の正解はあるか

- 最後に, 解決方法の一つを示し, 受講者へ反省と近未来の学修の方向性を示す(DS⁴のPBL)
- 近年強調されていること
 - 不確実性の高い時代、世の中は正解のない問題だらけ
 - 人によって答えが違ふ
 - 100点の試験でも、500点、1000点という突き抜けた結果を出す人がいる
 - 平尾丈『起業家の思考法「別解力」で圧倒的成果を生む問題発見・解決・実践の技法』
- 目的に対してデータ科学の「定石」がある
 - 異なるアプローチがあり得る
 - 結果として, 定石を超えるアプローチがあり得る
 - 結論はともかく, 定石を試さないデータ分析はあり得ない

20

DS4PBLの受講生感想(2019)

• データ・課題について

- 企業の実データに触れたこと
- データの変数や量が多いため自由度の高い分析が実施できた
- 課題に多様性があり工夫されていた
- 統計を専門としない人にとっても取り組みやすい内容だった

• データ分析について

- 基本的手法の重要性を再認識した
- 試行錯誤しながら課題に取り組めたこと
- 分析ストラテジーが多様であること
- データの可視化などいつもは気づかない部分に気づけた
- 統計を学ぶ重要性を再認識した
- 自分の研究データ以外ほとんど扱ったことがないので貴重な経験だった

• 教員の指導について

- 講師による分析例が大変わかりやすく参考になった
- データ分析手順についてたくさんのアドバイスがあったこと
- 同じデータを用いた分析例がとても参考になった

• グループワークについて

- チームで一緒に同じデータを分析するのはよい取り組みだと思う
- 他分野の方のデータに関する考え方を知ることができてよかった
- 他研究科の人たちが自分の知らない手法を数多く知っていたこと
- 他学、専攻の違う人と作業、交流できた
- 様々な分析手法に触れ勉強になった

21

改善を要求された点

• 2019

- 事前に必要とする最低限の能力やできるべきことを明記してもらえると良いと感じた。
- **自己紹介がちゃんとあった方がいい**
 - Ice-breaking の重要性 (狩野 記)
- データ選択はもう少し吟味させるべき

• 2020

- **履修条件以外に、一定のプログラミングレベルを設けてほしい**
- 泊まりでしたかった
- もう少し時間があればよかった
- 指導が少なかった

• 2021

- 途中のフィードバックがもう少し頻繁にあってくれる
- 各人のメンバーについて理解度の差異が大きい

22

まとめ: データ科学における PBL・実習演習の在り方

- ゴールは, DSによってその課題を解決することではなく, 近未来DSとして活躍できる素養を身に付けること
 - 残差より予測誤差?
 - そのためには
 - 山(課題)は登りきることが大事?
 - 成功体験は大事だが, 実際業務に近い課題の解決は相当に困難
 - その課題解決のための深いスキルが必要. 汎用性は?
 - 「実課題とは何か」や「課題解決は難しい(給料が出るだけある)」などを知り, 大学での勉強・研究の糧にする
- 教材作成がすべて
 - そのPBL・実習演習でDSの何が身に付くのか
 - この目的のための教材作成・適切な分析データセット収集が極めて重要
 - DSの定石の理解と実践が重要

23

discussion

24

データ科学におけるPBL・実習演習で 強調されていること

- **リアルデータ**
 - 企業様から提供された(ビッグ)データ
 - 受講生に好評
- **課題発見・課題解決**
 - 社会的課題(e.g., SDGs)
 - 要人・賓客の前でのプレゼン
- **社会と大学との関係強化**
 - 社会の課題解決へ資する
 - 社会の大学への貢献
 - 象牙の塔に籠るな
- → 社会のニーズに合う教育
イノベーションを産む人材

26

データ科学におけるPBL・実習演習で 強調されていること

- **リアルデータ**
 - 実データ解析の経験は必要
 - データクリーニング, 守秘義務...
 - DSを理解する目的に適しているか?
 - 将来DSとして羽ばたく準備になっているか?
- **課題発見・課題解決**
 - 真の課題に取り組んでいるか?
- **社会と大学との連携強化**
 - 企業様・団体様の様々な声
 - 大学では就職後OJTできないことを学んできてほしい
 - DSとして採用したが, 思うように活躍してくれない

27

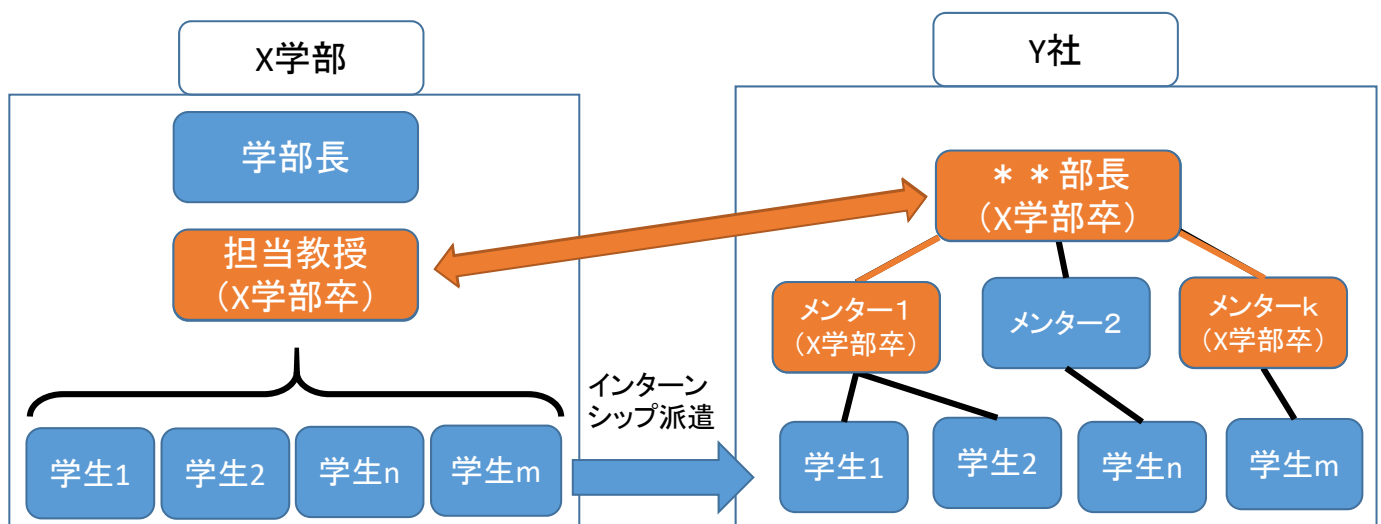
あるべき姿？

- 企業さんと正式に契約し、相当しっかりと実施しているPBLもある
 - 現場の検分, メンターとのコミュニケーション, 部長クラスの前でのプレゼン, etc.
- 企業さんのメンターはIS学生や受講生を叱れるか
 - 部下の指導のように
 - (企業側は)なぜ, そこまでやる必要があるのか

28

ある工学系学部の インターンシップの例

- 大学での成果発表は企業メンターにご足労頂き, その前で行う. 不十分な内容に対しては, 容赦なく強い指導を(メンターの)前で行う
- 不真面目な学生については部長(メンター)から大学にクレームが入る



29

ご清聴ありがとうございました