

# 研究報告書

## 「時空間粒度の異なる教育ビッグデータの非同期ストリーム処理基盤の構築」

研究タイプ: 通常型

研究期間: 2015年10月～2019年3月

研究者: 島田 敬士

### 1. 研究のねらい

都市、人流、医療、広告、購買など様々なビッグデータの利活用時代の到来により、個人やスポットのような「点」に注目していた従来の解析から、群衆や面のように「空間」的に実世界を捉えて、社会全体の最適化や個人の快適性を向上させるための解析に軸足が移りつつある。従来のビッグデータ解析においては収集したデータを様々な時空間粒度で分析して、その結果を将来や次世代にフィードバックするリレー形式の活用が多く見受けられる。一方、データの収集からフィードバックまでの時間をできるだけ短縮するためのストリーム処理が実現できれば、ビッグデータの提供元である現場や現場のユーザが直接的な恩恵を受けることができる社会を実現できる。

本研究では、時空間粒度の異なるビッグデータのストリーム処理基盤の構築を目的とし、特に教育分野への応用を考える。近年、教育現場へのICT導入が積極的に行われており、その結果として教育ビッグデータが形成されるようになりつつある。従来のICT教育は主に個人学習の解析を対象にしていたが、教育ビッグデータを活用することで他の学習者との比較やグループ・教室・科目・大学のような異なる空間粒度での解析や、リアルタイム・週単位・学期単位・年単位などの異なる時間粒度での解析が可能になる。教育ビッグデータには、電子教材や情報端末、デジタル学習環境などから学生や教師の膨大な活動ログが非同期に収集されているため、本研究で開発する処理基盤技術の有効性検証に利用できる。

本研究で具体的に扱う教育ビッグデータは、学習管理システム(e-Learningシステム)とデジタル教科書システム(e-Bookシステム)から収集される大学生の学習活動ならびに講義を行う教師の教育活動データである。各学生がシステムを利用するタイミングは各々で異なるため、データは非同期に収集される。また、同一講義、同種の講義間、学期間、大学間などで同種の情報であっても異なる時間や空間の粒度でデータが収集される。そのような教育ビッグデータを利用して、学習困難に陥っている学生を発見したり、教室全体の学習理解度を把握したり、学習到達度を予測したり、学生に理解を助長する資料を提供したり、講義進行の調整案を教師に提示したりできるフィードバックシステムを実現し、教育学習環境の最適化を目指す。

### 2. 研究成果

#### (1) 概要

研究者が所属する九州大学で学習管理システム(Moodle)、デジタル教科書システム(BookRoll)を中心としたデジタル学習環境による教育ビッグデータ収集基盤を構築し、PC必携化が行われている大学教育において時空間粒度の異なるビッグデータのストリーム処理基盤の構築を行った。研究成果は空間粒度、時間粒度の観点で分類できる。空間粒度としては、個人レベルの粒度では、デジタル学習環境で収集されるイベントストリーム(学生がシス

テムを操作した際に得られる記録)から個人の学習パターンをモデル化することで異常行動を検知する技術, 個人の学習理解度に応じて復習教材を自動生成する技術の開発を行った。また, 教室規模(100~200 名程度が受講する講義)の粒度では, デジタル教科書の閲覧パターン分析技術, 予習用要約教材の自動生成技術, 講義中のリアルタイム閲覧状況分析技術, 着席位置と学習活動の関係解析技術の開発を行った。さらに, 同種の講義を対象とした科目レベルの粒度では, コース間の類似・非類似性を把握するコース横断分析技術, 成績予測技術の開発を行った。これらの成果は, 時間粒度の観点でも整理可能であり, 以下の研究成果詳細では教育ビッグデータの分析結果をフィードバックする際の時間粒度の違いで成果を分類し, 取り組み内容を報告する。

研究テーマ A 「現場の教育・学習を支援するリアルタイム学習分析技術の開発」

研究テーマ B 「各講義の内容理解を助長する教材分析技術の開発」

研究テーマ C 「長期間データを利用した学習活動分析技術の開発」

## (2) 詳細

研究テーマ A 「現場の教育・学習を支援するリアルタイム学習分析技術の開発」

本テーマでは, 現場の教育・学習をリアルタイムに支援するための教育ビッグデータ処理基盤を開発した。本成果を利用することで, 講義を行う教師は現場の学生の学習状況に応じて柔軟な講義進行を行えるようになる。

予習状況分析技術では, デジタル教科書のアクセス履歴からページ単位で講義受講者の予習状況を集計し, 未予習者の多いページを特定することが可能である[1]。また, 予習理解度を確認するための小テストを実施し, 正答率の低い設問に対応する教材内の該当ページを教師に提示することが可能である。設問と教材を接続するために, 設問文内の単語と教材の各ページに含まれる単語間の類似度を計算することで, 関連ページを特定する技術を開発した。教師は提示された教材のページの説明時間を長く確保したり, 説明を補足したりするなど柔軟な対応を行えるようになった。

閲覧状況分析技術では, 講義に参加している学生がデジタル教科書システムにアクセスして教科書を開いたり, ページを移動したりする際に記録される学習活動ログを瞬時に集計し, 1分毎の閲覧分布を教師のPC上に可視化するシステム(図 1)を開発した[1]。デジタル教科書システムへのアクセス履歴とアクセス者の履修情報を統合分析することで, 各授業に参加している学生を特定しつつ, 教室内の学習活動を分単位で統合処理する技術を開発した。教師へのフィードバックは各ページの閲覧者数をヒートマップ風に表示することで行われ, 表示は1分ごとに更新される。ヒートマップ上では教師が説明しているページが赤枠でハイライトされる。教師が説明しているページ付近に多くの学生が閲覧している状況が確認できれば, 授業進行は概ね順調とみなせるが, 閲覧分布に偏りが見受けられれば, 学生が講義進行についていけない可能性や, ページの先々を見ている可能性が考えられる。教師はそのような状況を瞬時に把握し, 授業進行のスピードを遅くして説明を補足したり, 学生がキャッチアップするための時間を設定したりするなど, 現場の実状況に応じた柔軟な講義進行が可能になる。

学習活動変化検出技術では、講義に参加している各学生のシステムへのアクセスパターンをモデル化し、モデルから逸脱するパターンを検出することを可能にした[2]。先行研究では教室全体のアクセスパターンをモデル化する技術が提案されていたが、実際のアクセスパターンを分析すると個人による特徴が強いことが分かったため、アクセスパターンの個人化を行った。また、現場でのリアルタイム変化検出に向けて、個人のアクセスパターンをポアソン過程により逐次的にモデル化する方法を開発した。本技術により異常パターンとして検出された学生は、他の学生よりも事後テストの成績が低いことが確認されたことから、開発した技術の有用性が確認できた。

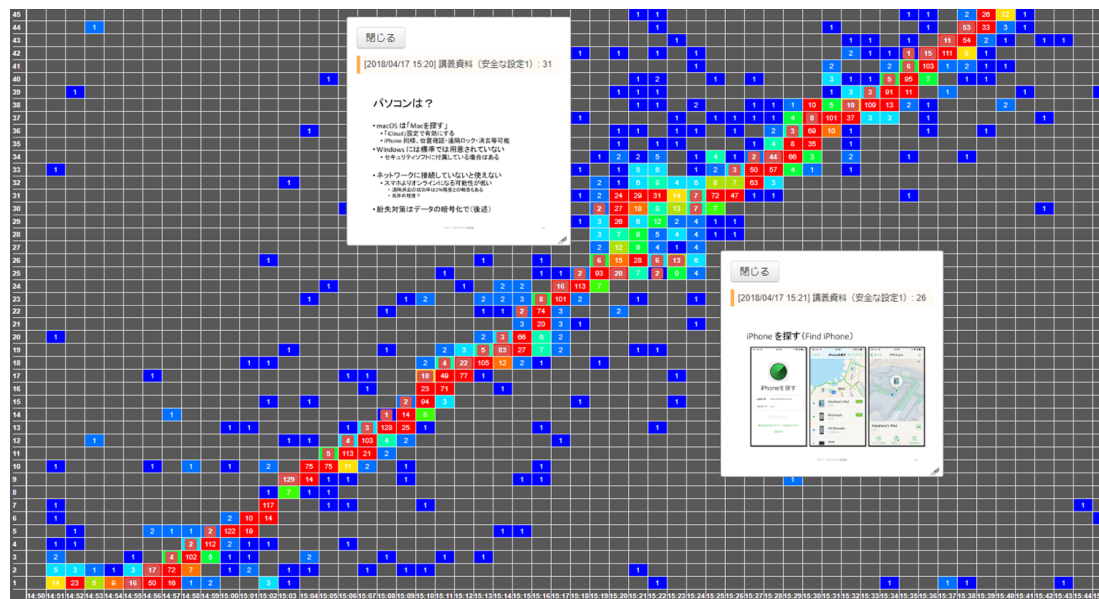


図 1 教材閲覧ヒートマップ。横軸は時刻，縦軸は教材のページ番号に対応。同時刻に同じページを閲覧している学生が多いほど赤系の色で表示される。ページ番号をクリックすると該当ページの内容がポップアップ表示される。

### 研究テーマB「各講義の内容理解を助長する教材分析技術の開発」

本テーマでは、デジタル教科書システムに登録されている教材を分析対象とし、学生の予習復習を支援するための教材生成技術を開発した。本成果を利用することで、予習復習の効率・効果の向上ならびに教材改善を行うことができる。

予習教材生成技術では、学生に電子教材の要約版を自動生成する技術を開発した[3]。具体的には、デジタル教科書システムに登録された電子教材に対して、画像処理とテキスト処理を適用し、視覚的な観点と単語の重要度の観点を統合し、各ページの重要度を算出する。さらに、各ページの閲覧に要する推定時間を用いて、短時間でできるだけ閲覧効果の高いページの組み合わせを自動的に抜粋し、要約版資料として再構成する技術を開発した。要約版資料を予習に利用した学生群からは、予習達成率の飛躍的な向上ならびに事前テストでの成績向上の効果が確認できた。

復習教材生成技術では、学生の講義内容の理解度に応じて復習を支援するための要約版資料を生成する技術を開発した[4]。具体的には、学習管理システムで実施される事後テストの結果を分析し、誤答した設問に対応する教材のページを自動発見するアルゴリズムを開発

した。設問文に含まれる単語を解析し、その単語に対して関連度の高い教材のページをランキング形式で出力する。また、各学習者の予習状況も同時に分析し、未予習箇所についても自動推薦することが可能である。本技術を利用して生成された復習教材を利用することで、事後テストでの理解度の向上が確認できた。

#### 研究テーマC「長期間データを利用した学習活動分析技術の開発」

本テーマでは、週単位や学期単位で収集された学習ログを分析対象とし、学生がどのような学習活動を行っているかを定量的に分析するための学習活動分析技術を開発した。本成果は、講義を実施する教師や、授業・コースを管理する科目マネージャなどが授業改善の方策を検討したり、困難に陥っている学生に対する対応を検討したりする際に活用できる。

教材の閲覧パターン分析技術では、同一教材を閲覧した学生群の閲覧パターンを要約する技術を開発した[5]。各学生の各ページの閲覧時間を行列表現し、行列分解手法を適用することで、閲覧パターンを抽出した。本技術を1週間ごとの予習期間中の学習ログに適用したところ、代表的な閲覧パターンが複数発見され、その中には効率の良い閲覧方法で予習を実施し、結果として事前テストでよい成績を修めている学生群を発見することができた。

着席位置と学習活動の関係分析技術では、学期中に学生が教室内で着席していたエリア情報と講義内外での学習活動の関係を分析する技術を開発した[6]。学生の着席エリアを収集するために、学習管理システム上にクリッカープラグインを実装し、毎回の講義開始時に自分が着席しているエリアを学生に投稿してもらう方式を採用した。また、学習活動については、出席状況、小テストの結果、課題の提出状況、予習復習状況、授業中の閲覧状況などを数値化した。14週間に渡りデータを収集し、開発した分析技術を適用したところ、教室前方に着席傾向の強い学生の方が教室後方に着席傾向の強い学生よりも学習活動の活発さが高いことが確認された。また、着席エリアを頻繁に変更する学生は成績が低く、欠席傾向が高いことも明らかになった。

成績予測技術では、学期中に収集された学習活動ログと最終成績の関係をモデル化し、週ごとの学習活動から最終成績を予測するための成績予測モデルを開発した[7]。出席状況、小テストの結果、課題の提出状況、授業外での教材閲覧状況を週ごとに数値化し、8週(クォーター)あるいは15週(セメスター)後の成績を予測するためのリカレントニューラルネットワークモデルを生成した。ネットワークの学習には、昨年度以前のデータを利用し、実際の成績予測には現行年度のデータを利用した。全体期間の半分程度の期間情報をネットワークに入力できれば、8割以上の精度で成績を予測できることが確認できた。

### 3. 今後の展開

本研究を通して、様々な時空間粒度での教育ビッグデータの分析技術を開発してきたが、今後は本技術の横展開を考えていく必要がある。同様のデジタル学習環境を導入している大学への展開は比較的容易であると考えるが、学習ログの形式が異なるシステムを利用している場合は、ログフォーマットの変換あるいは分析技術の改修が必要になる。そのため、学習ログの標準化についての研究についても今後進めていく。

また、他大学とのデータ連携や学習分析結果の連携を行うためには、データの匿名化やプライバシーポリシーの策定が急務である。プライバシーポリシーについては、学内では実験倫理委員会



を中心とした関連組織での議論を進めており、ボトムアップな研究を推進するうえでのデータの取り扱いや管理方針を定める予定である。またトップダウン的には大学の枠を超えた国家レベルでの教育データの取り扱いについて今後議論が加速していくことを期待している。その際に本研究の取り組みがひとつの先行事例になるよう取り組みを続けていく。

#### 4. 自己評価

1)教育ビッグデータを収集するためのデジタル学習環境の構築, 2)学習管理システムならびにデジタル教科書システムを利用して収集される大規模な学習活動ログを様々な時空間粒度で分析する技術, 3)現場の講義環境や学生, 教師に対して分析結果をフィードバックし, 学習改善ならびに教育改善に資する成果が得られた, という点において, 本研究が ICT を活用した教育に対して一定の成果をあげることができたと考える。

デジタル学習環境を安定運用することは, 研究成果(論文)として直接的には表れないが, 研究を実施するうえでは非常に重要な課題である。研究補助者にシステム運用を支援してもらえたことで, 研究に集中することができた。また, プロトタイプシステムを実装し, それを実運用システムに実装する際には, 専門の業者にソフトウェア開発を依頼することができたため, 開発効率が飛躍的に向上した。さらに, 国内外で研究成果発表を数多く行うために研究費を活用できた。

大学教育におけるPC必携化が進んでおり, また近い将来初等・中等教育に対して教育の情報化が実施されることを鑑みると, 本研究成果を社会展開する環境が国家レベルで整いつつあると考えられる。教育の情報化が進めば, 自ずとデータ利活用の動きが活発化すると推測されるため, 本研究成果は先行的な研究事例として注目される可能性が極めて高い。また, 開発した技術については社会展開を見据えて, 汎用性を重視した実装を進めてきており, 今後様々な教育学習システムに本研究成果が導入されることを期待する。またそうなるように研究者自身も活動の幅を広げていく。

#### 5. 主な研究成果リスト

##### (1)論文(原著論文)発表

1. Atsushi Shimada, Shin'ichi Konomi, Hiroaki Ogata, Real-time learning analytics system for improvement of on-site lectures, Interactive Technology and Smart Education, Vol.15, No.4, pp.314-331, 2018.
2. Atsushi Shimada, Yuta Taniguchi, Fumiya Okubo, Shinichi Konomi, Hiroaki Ogata, Online Change Detection for Monitoring Individual Student Behavior via Clickstream Data on e-Book System, 8th International Conference on Learning Analytics & Knowledge (LAK'18), pp.446-450, 2018.
3. Atsushi Shimada, Fumiya Okubo, Chengjiu Yin, Hiroaki Ogata, Automatic Summarization of Lecture Slides for Enhanced Student Preview -Technical Report and User Study-, IEEE Transactions on Learning Technologies, Vol.11, Issue2, pp.165-178, 2017.
4. Atsushi Shimada, Fumiya Okubo, Chengjiu Yin, Hiroaki Ogata, Automatic Generation of Personalized Review Materials Based on Across-Learning-System Analysis, Cross-LAK2016, 2016.

- |   |
|---|
| 5. Atsushi Shimada, Fumiya Okubo, Hiroaki Ogata, Browsing–Pattern Mining from e–Book Logs with Non–negative Matrix Factorization, the 9th International Conference on Educational Data Mining, pp.636–637, 2016.  |
| 6. Atsushi Shimada, Fumiya Okubo, Yuta Taniguchi, Hiroaki Ogata, Rin–ichiro Taniguchi, Shin’ichi Konomi, Relation Analysis between Learning Activities on Digital Learning System and Seating Area in Classrooms, 11th International Conference on Educational Data Mining, 2018. |
| 7. Fumiya Okubo, Takayoshi Yamashita, Atsushi Shimada, Hiroaki Ogata, A Neural Network Approach for Students’ Performance Prediction, The 7th International Conference on Learning Analytics & Knowledge Understanding, 2017.   |

(2)特許出願

研究期間累積件数:3件

1.

発明者: 島田敬士, 緒方広明

発明の名称: METHOD FOR ARRANGING SUMMARY SLIDE SET

出願人: 九州大学

出願日: 2015年9月

出願番号: 62/212608

2.

発明者: 島田敬士, 緒方広明

発明の名称: スライド要約装置、学習支援システム、スライド選択方法及びプログラム

出願人: 九州大学

出願日: 2016年4月

出願番号: 特願 2016–083102

3.

発明者: 島田敬士, 緒方広明, 清田麻寛, 毛利考佑

発明の名称: 情報表示装置、学習管理システム、情報表示方法及び情報表示プログラム

出願人: 九州大学

出願日: 2017年3月

出願番号: 特願 2017–042856

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

受賞

第1回 IMS Japan 賞 最優秀賞

第13回日本 e–Learning 大賞 教育ビッグデータ特別部門賞

IPJSJ/IEEE–CS Young Computer Researcher Award(授賞式は2019年6月予定)

報道発表



2017年1月25日掲載、日本経済新聞(31面)、「学習支援にビッグデータ」  
2017年1月26日掲載、西日本新聞夕刊(1面)、「九大デジタル教育革命」

招待講演

日本教育工学会 第34回全国大会

「実時間学習分析による教育改革の可能性」

一般社団法人データサイエンティスト協会 5th シンポジウム

「ラーニングアナリティクスを用いた教育・学習支援の可能性」