

グラフデータベースを活用した事例推薦システムの開発 —実世界エビデンスの活用に向けて

黒宮 寛之^{1,a)} 中西 太郎^{1,b)} 緒方 広明^{2,c)}

概要: 新型コロナウイルスの流行によりオンライン教育が広がりを見せる中で、効果的な学習ツールの使い方や効率的な成績評価の方法を機関内で共有する試みが多く多くの大学で行われている。このような事例（実世界エビデンス）は現場の教員にとって非常に有用であるが、多忙な教員にとって数ある事例の中から自分に関連する事例を探し出すのは困難である。関連する事例を収集するうえで、教員は自身の運営する授業の特性を踏まえながら、適切なキーワードを選定し、適切な条件で絞り込みを行う必要がある。本論文ではこの問題を解決するため、教員のアクセスログを活用し、収集した事例をユーザーに自動的に推薦するエビデンスポータルシステム REAL (Real-world Evidence Analysis Library) を提案する。本システムはグラフデータベースに事例を蓄積することによって、蓄積された事例とユーザーを関連付け、ユーザーにとって最適なエビデンス（事例）をシステム側から提案することができる。本システムの実現によって教員が過去の事例を探す手間が削減され、有用な知見がより多く授業で活用される効果が期待できる。本論文では提案システムの設計案とその利用例を示す。

1. はじめに

新型コロナウイルスの流行によりオンライン教育が広がりを見せる中で、効果的な学習ツールの使い方や効率的な成績評価の方法を機関内で共有する試みが多く多くの大学で行われている。例えば東京大学では情報基盤センター／大学総合教育研究センターが中心となって「オンライン授業・Web会議ポータルサイト」(<https://utelecon.adm.u-tokyo.ac.jp/good-practice/>)を運営している。ここでは「グッドプラクティスの共有」という形でオンライン期間中に行われた授業の工夫や苦勞した点などを公開している。ほかにも慶応義塾大学 (https://www.sfc.keio.ac.jp/campuslife/online2020_spring/g-practices/)、東北大学 (<https://olg.cds.tohoku.ac.jp/forstaff>) などでも同様の事例の共有が行われている。

このような事例は「実世界エビデンス」[1]と呼ばれ、統制された実験条件などがいないため因果関係の同定には不十分なものの、現場の教員にとっては非常に有用であることが指摘されている。しかしインターネット上に多くの事例

が集まる一方で、多忙な教員にとって数ある事例の中から自分に関連する事例を探し出すのは困難である。関連する事例を収集するうえで、教員は自身の運営する授業の特性を踏まえながら、適切なキーワードを選定し、適切な条件で絞り込みを行う必要がある。

本論文ではこの問題を解決するため、教員のアクセスログを活用し、収集した事例をユーザーに自動的に推薦するエビデンスポータルシステム REAL (Real-world Evidence Analysis Library) を提案する。他のエビデンスポータルシステムと比較して、本システムの特徴的な点は以下の2点である。

- (1) ユーザーの閲覧履歴をもとに、ユーザーにとって最適なエビデンス（事例）をシステムが自動的に推薦することができる。
- (2) 上記の目的を達成するため、効果的な教授方法に関する事例をユーザーやコースと関連付けながらグラフデータベース上に蓄積していく。

本システムの実現によって、教員が事例を探す手間が削減され、過去の有用な知見がより多く授業で活用される効果が期待できる。本論文では提案システムの設計案とその利用例、またシステムの稼働・評価実験に際しての問題点を概観する。

¹ 京都大学大学院情報学研究科
Yoshida-Honcho, Sakyo-ku, Kyoto, Kyoto 606-8315, Japan

² 京都大学学術情報メディアセンター
Academic Center for Computing and Media Studies

a) khiroyuki1993@gmail.com

b) nakanishi.taro.26r@st.kyoto-u.ac.jp

c) ogata.hiroaki.3e@kyoto-u.ac.jp

2. 関連研究

Learning Analytics の分野では 2016 年ごろから欧州を中心として効果的なラーニングアナリティクスツールの活用方法を事例として蓄積するためのプラットフォーム開発が続けられてきた [2], [3]. LACE Evidence Hub と呼ばれるこのデータベースは、ラーニングアナリティクスに関する事例を A) 学習効果を改善するための事例, B) 教授活動を支援するための事例, C) ラーニングアナリティクスを広めていくための事例, D) ラーニングアナリティクスの倫理的な配慮に関する事例, の 4 つのカテゴリに分け、それぞれにその事例が生まれた地域や結果の positive/negative というラベルを付加することで、どのようなツールがどのような文脈で役立つのかについての知見を蓄積していくことを目的としている。このデータベースは英国の Open University など実際に活用され [4], 学習データに基づいて介入を行い、評価するというラーニングアナリティクスの基本原則 [5] を補強するものとして作用している。

しかしこのようなシステムは事例を蓄積することが主な目的となっており、蓄積された事例を分析して教員にシステム側から適切な事例を推薦するというところまではまだ踏み込めていない。したがって、本研究では事例の推薦を可能にするエビデンスポータルシステム REAL の提案を行う。以下にシステムの構成を示す。

3. システム構成

3.1 データベース

本システムではユーザーに事例の推薦を行うため、事例を蓄積するデータベースとしてグラフデータベースの Neo4j [7] を使用する。グラフデータベースとはグラフデータモデルを公開する CRUD メソッドを備えたオンラインデータベース管理システムである [6]。ここではグラフモデルとしてプロパティグラフと呼ばれる「ノード」「リレーション」「プロパティ」の 3 要素に構成されるグラフを扱う。「ノード」とはグラフのエンティティとなる要素のことで「ラベル」によって種類分けがなされる。「リレーション」はノード間の関連性を表したもので方向とタイプを有する。「プロパティ」はノードとエッジの属性情報を表したもので、主に key-value 形式で保持される。本システムで使用するノードの種類を表 1、リレーションの種類を表 2 に示した。

図 1 に本システムによって構成されるグラフの例を示す。図中のオレンジのノードがユーザー、水色のノードがコース、赤のノードが事例を表している。この図から User 2 が Course A を担当しており、このコースから Case 1, Case 2 の 2 つの事例が登録されていることが分かる。このうち Case 1 は User 1 から、Case 2 は User 1, User 3

表 1 本システムにおけるノードとそのプロパティ

ラベル	プロパティ
User	名前, ユーザー ID, 教員歴, メールアドレス, 所属 (部局, 大学名)
Course	コース名, 開講部局, 対象学年, 学生数, 教科, 形態 (オンライン or 対面)
Example	介入 (使用したツール名, 教授方略, 評価方法), 学生の反応 (量的/質的), 適用コスト (苦勞したところなど)

表 2 本システムにおけるリレーションの定義

ラベル	対象ノード	説明
OWN	User-Course	コースを担当している
HAS	Course-Example	事例が属するコースである
VIEWED	User-Example	ユーザーが事例を閲覧した

から閲覧されている。本システムではこの場合、User 3 に Case 1 を推薦することによって適切な事例の推薦ができる考える。なぜなら Case 1, 2 は同一のユーザーから閲覧されているため関連性があり、User 3 は Case 2 は閲覧しているものの Case 1 をまだ閲覧していないからである。

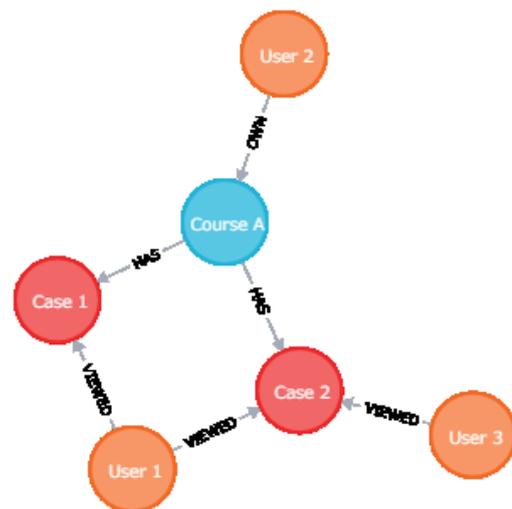


図 1 グラフデータベースの例

3.2 システムの全体構成

図 2 にエビデンスポータルシステム REAL の概観を示す。本システムでは LTI (Learning Tools Interoperability) 認証技術を利用し、ユーザーが学習管理システム (LMS) のユーザー情報を引き継いでログインできるようにする。これによって、本システムでは LMS からユーザー名やコース名を引き継ぐことができ、事例の登録や閲覧をユーザーと関連付けることができる。事例とユーザーを関連付けることができれば、本システムはグラフデータベースに適切なクエリを発行することによって、ユーザーに関連する事

例を推薦することができる。アプリケーション側からデータベースの操作には Cypher[8] と呼ばれる Neo4j 専用のクエリ言語を使用する。



図 2 システムの全体構成

4. 利用例

ここでエビデンスポータルシステム REAL の利用例を「事例登録」「事例閲覧」「事例推薦」の3つに分けて紹介する。これはエビデンスに基づく教育の定義 [9] である, 1) エビデンスの作成, 2) エビデンスの使用の二つのレイヤーに対応しており, ユーザー (教員) が自身の実践の中で本システムを活用する一つの理想的なワークフローを表したものである。各項目にはシステムから発行されるクエリ言語の一例を付記した。

4.1 事例登録

本システム的前提として, 教員は自身の授業運営を行うにあたって LMS を利用するものとする。教員は LMS を通してリソースの配布や課題の採点などを行うが, 自身の教授活動の中で他の教員にとって役立つような「事例」を発見した場合には, LMS のコースから「REAL (エビデンスポータルシステム)」を起動する。システムを起動すると「事例の作成」「事例の閲覧」の2つの画面が現れるが, 「事例の作成」を選ぶことで図3のような事例登録画面が現れる。ここでユーザーはこの事例登録フォームから自身の事例をデータベースに登録することができる。実際にシステムからデータベースに対して発行されるクエリは以下のとおりである (図4)。

図 3 事例登録画面

```
MATCH (C:Course {id: context_id (LMS から取得)})
CREATE (Ex1:Example {action: '授業のディスカッションに Zoom のブレイクアウトルームを使用', reaction: '授業後のアンケートの結果, 32名の学生がブレイクアウトルームの利用に肯定的な評価をしていた', cost: '授業中3名の学生に接続トラブルあり'})
CREATE (C)-[:HAS]->(Ex1)
```

図 4 事例登録クエリ

4.2 事例閲覧

また, 教員は自身の実践の内容をデータベースに登録するだけでなく, データベースに登録された他の事例を閲覧することができる。実際の事例閲覧画面の例を図5に示した。ユーザーが事例を閲覧した場合, そのユーザーと事例を関連付けるため以下のようなクエリがデータベースに対して発行される (図6)。



図 5 事例閲覧画面

```
MATCH (u:User {id: user_id (LMS から取得)}),
(e:Example {id: example_id (事例 ID)})
CREATE (u)-[:VIEWED]->(e)
```

図 6 事例閲覧クエリ

4.3 事例推薦

本システムでは事例閲覧画面を選択した場合, 検索フォームの下にグラフネットワークに基づいた関連する事例を推薦する。具体的には自分が閲覧した事例を閲覧した人が閲覧した事例のうち, 自分がまだ読んでいないものを関連度が高い順にリスト表示する (図7)。また (図8) に, システムからデータベースに対して発行するクエリを示した。このようなクエリは従来のリレーショナルデータベースでは複雑なクエリを記述する必要があったが, グラフデータベースを使うことで効率的に記述することができる。

タイトル	概要	リンク
Allen (2010)	Ecler	http://www.example.com
Alan (2020)	Jellybean	http://www.example.com
Jonathan (2019)	Lollipop	http://www.example.com

図 7 事例推薦画面 (例)

```
MATCH (me: User {id: user_id})-[:Viewed]->
(e:Example)-[:VIEWED]-(other_user:User)-[:VIEWED]->
(recommend_case:Example) WHERE NOT (me)-[:VIEWED]->
(recommend_case)
RETURN recommend_case.title, count(recommend_case)
as count
ORDER BY count DESC
```

図 8 事例推薦クエリ

5. まとめと今後の研究課題

本稿では、ユーザーに適切なエビデンスを推薦するためのエビデンスポータルシステム REAL の概要とそのシステム構成について紹介した。本システムではグラフデータベースに事例を蓄積していくことによって、ユーザーと事例を関連付け、適切な事例を教員に推薦することを可能にした。教育におけるエビデンスデータベースの中では、事例の自動推薦にまで踏み込んだアプリケーションはなく、本システムの開発によって、教員のエビデンスに基づく実践を支援することができると考えている。

ただし、ユーザーに適切な事例を推薦するための方法はグラフデータベースの利用だけが唯一の方法というわけではない。例えば、事例の内容をテキストマイニングによって抽出し、特徴量化することによって、コサイン類似度などを用いたコンテンツベースの推薦を行うことも可能である。したがって本研究では今後のシステムの評価実験として 1) グラフデータベースに基づいた推薦を行う場合 (本提案)、2) 自然言語処理を用いたコンテンツベースの推論を行う場合、3) グラフデータベースとコンテンツベースの推論をうまく組み合わせた場合、でどの手法がもっともユーザーにとって有用な事例を推薦することができるのかを評価する予定である。

また、そもそも「本システムにどのように大量の事例を登録していくのか」というのも未解決の問題として残っている。現在では各大学の事例をスクレイピング技術などを使って登録していくことを考えているが、著作権の問題などもあり、個別の対応が必要であると考えられる。さらに、公開された情報だけでは事例が少ないので、ラーニングアナリティクスの基盤システム [10] で蓄積したデータを利用して、いかに効率的に事例を登録するか、という点についても併せて研究を行っていきたい。

謝辞 本研究は JSPS 科研費基盤研究 (S)16H06304 および JST ACT-X グラント番号 JPMJAX20AA の支援を受けた。

参考文献

[1] Sherman, R. E., Anderson, S. A., Dal Pan, G. J., Gray, G. W., Gross, T., Hunter, N. L., LaVange, L., Marinac-Dabic, D., Marks, P. W., Robb, M. A., & Others. (2016). Real-world evidence—what is it and what can it tell

us. *The New England Journal of Medicine*, 375(23), 2293–2297.

[2] Ferguson, R., & Clow, D. (2016). Learning Analytics Community Exchange: Evidence Hub. 6th International Learning Analytics and Knowledge (LAK) Conference. 6th International Learning Analytics and Knowledge (LAK) Conference, Edinburgh, Scotland. <http://oro.open.ac.uk/45314/>

[3] Ferguson, R., & Clow, D. (2017). Where is the evidence?: a call to action for learning analytics. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 56–65.

[4] Rienties, B., Borooa, A., Cross, S., Kubiak, C., Mayles, K., & Murphy, S. (2016). Analytics4Action Evaluation Framework: A Review of Evidence-Based Learning Analytics Interventions at the Open University UK. *Journal of Interactive Media in Education*, 2016(1). <https://eric.ed.gov/?id=EJ1089327>

[5] Clow, D. (2012). The learning analytics cycle: closing the loop effectively. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*, 5.

[6] Robinson, I., Webber, J., & Eifrem, E. (2015). Graph databases: new opportunities for connected data. "O'Reilly Media, Inc."

[7] Webber, J. (2012, October). A programmatic introduction to neo4j. In *Proceedings of the 3rd annual conference on Systems, programming, and applications: software for humanity* (pp. 217-218).

[8] Francis, N., Green, A., Guagliardo, P., Libkin, L., Lindaaker, T., Marsault, V., ... & Taylor, A. (2018, May). Cypher: An evolving query language for property graphs. In *Proceedings of the 2018 International Conference on Management of Data* (pp. 1433-1445).

[9] Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies*, 47(2), 108–121.

[10] Flanagan, B., & Ogata, H. (2018). Learning analytics platform in higher education in Japan. *Knowledge Management & E-Learning: An International Journal*, 10(4), 469–484.