Vol.2023-CLE-41 No.18
Vol.2023-CE-172 No.18
2023/12/3

少量の学習ログからの成績予測モデル構築に関する予備的検討

尾﨑 優也[1,a)]　出口 大輔[1]　村瀬 洋[1]　川西 康友[2]　久徳 遙矢[3]

概要：本発表では，学習ログからの成績予測モデル構築手法に関する予備的な検討結果について報告を行う．現在，デジタル教材配信システムによって提供される教材を使用して講義を行い，受講生の学習活動をログとして取得する取り組みが大学等の高等教育機関で実施され始めている．このような学習ログを用いて講義内容に対する学生の理解度を推定するといった様々なデータ利用法が模索されており，その一例として At-risk 検出への応用が提案されている．At-risk 検出では，最終的に良い成績が得られない学生を At-risk 学生，良い成績を収める学生を No-risk 学生と定義し，それらを分類するタスクである．この At-risk 検出は学期初めの数週間経過後に実施し, 早期に対策を講じる目的で行われる．しかし, 期間の早い段階で At-risk 検出をする場合，利用できる学習ログが少ないため高い精度が得られないという課題がある．そこで本発表では，学期初めの数週間という少量の学習ログから精度良く At-risk 検出を行うための手法について検討した結果について報告する．

1. はじめに

近年，様々な教育機関で Information and Communication Technology（ICT）技術を活用した新しい教育システムの導入が進んでいる．この ICT を活用した教育システムの一つに，九州大学や京都大学で利用されている「BookRoll」と呼ばれるデジタル教材配信システムがある．このデジタル教材配信システムでは，**図 1** に示すように，「教材を開く (Open)」,「次のページに移動 (NEXT)」,「前のページに移動 (PREV)」，といった学習行動およびその行動がいつ行われたかなど，学習者が学習中に行った行動がログとして記録されている．現在，これらの情報を活用して学習行動を分析し，学生の学習支援への活用が試みられている．

このような学習ログを利用した分析の研究には，ページ滞在時間を特徴とするもの [1,2]，項目ごとに得点化して特徴とするもの [3,4]，ログの特定イベントの発生数を特徴とするもの [5,6] などがある．これらの分析では，学習ログをもとに何らかの特徴表現を抽出しているが，各ログから算出した統計量を特徴量とするため詳細な学習行動は表現できていない．

例えば,「次のページに移動」を短時間に繰り返し実行する学生と，1 ページずつゆっくりと読みながら学習した学生がいた場合, 両者は同じ理解度で受講しているとは考えにくい．しかし, 特定イベントの発生数をカウントしたヒストグラムを特徴とした場合,「次のページに移動」を実行した回数が同じならば，同じ特徴量が得られてしまう．この問題を解決するために提案された新しい特徴表現手法として，学習行動の系列を文章として扱い，自然言語処理で用いられる fastText [7] モデルを通して分散表現を得る E2Vec [8] がある．この fastText はどのような長さの文章でも一定次元数のベクトルに変換できる．そのため, E2Vec は学習ログを一定次元数の分散表現に変換できる．

宮崎ら [8] は, 提案した E2Vec を用いて At-risk 検出の評価を行っているが, その評価では学期終了までのログを使用している．しかし, At-risk 検出は, 将来良い成績が得られない学生を早期に検出し, それを防ぐ目的で行われる．そのため, 学期初めの数週間の早い段階での At-risk 検出を考慮する必要がある．

そこで, 本研究では，E2Vec で変換された分散表現を使用し，学習ログが十分得られていない学期初めの数週間の学習ログから At-risk 検出を行う手法について検討を行う．具体的には，学期初めの数週間の学習ログから得られる分散表現から学期終了時点での分散表現を算出する特徴変換ネットワークについて有効性の検証を行う．

[1] 名古屋大学
Nagoya University
[2] 理化学研究所 ガーディアンロボットプロジェクト
RIKEN GRP
[3] 愛知工科大学
Aichi University of Technology
[a)] ozakiy@vislab.is.i.nagoya-u.ac.jp

Vol.2023-CLE-41 No.18
Vol.2023-CE-172 No.18
2023/12/3

user	contents	operation	···	eventtime
user1	content1	Open		2023-10-19 8:45:10
user2	content1	Next		2023-10-19 8:45:12
user1	content1	Next		2023-10-19 8:45:28
user3	content2	Prev		2023-10-19 8:45:36
user2	content1	Next		2023-10-19 8:45:40

図 1 学習ログの例

2. 関連研究 (E2Vec)

従来，学習ログから得られる統計量が学習活動分析の特徴表現として広く用いられてきた [1–6]．しかしながら，このような統計量に基づく特徴表現では，学習行動の時間的な情報が失われてしまい，詳細な学習活動を捉えることは難しい．このような問題の解決を目的として，宮崎らは自然言語処理のモデルである fastText を活用した E2Vec [8] を提案している．E2Vec は，以下の手順により学習ログから分散表現を算出する．

(1) 学習ログ中のイベントを対応する記号に置き換えた文章に変換
(2) fastText による分散表現の生成
(3) 分散表現の集約による特徴表現の生成

まず, 学習ログは図 1 で示した形式となっていることから，それぞれのイベントを対応する記号（文字）に置きかえる．そして，イベント間隔を一定の時間長で量子化し，それらに対応する記号として s，m，l の文字を挿入する．これにより，学習ログを文章 S に変換する．次に，得られた文章 S を fastText に入力することで分散表現を生成する．

E2Vec では, すべてのログを一度に文章 S に変換するのではなく, 教材の代わり目や, イベントがしばらく続かなかった場合にログを分割して文章 S を作成している．そのため，学生一人あたり複数の文章 S が得られ，対応する分散表現も複数得られる．そこで，E2Vec では得られた複数の分散表現を集約して一定次元にしている．

宮崎ら [8] の検討では，集約の方法として総和型と連結型が提案されている．本発表では最も精度が得られている総和型の集約により分散表現を得る方法を採用する．

3. 提案手法

本研究では，E2Vec により得られる分散表現を用いて早期（学期初めの数週間の段階）に At-risk 検出を行うことを目的としている．通常 At-risk 検出を行うネットワークは, 学期終了時点までの全学習ログを使用して学習することが望ましい．しかし, 学期終了時点までの全ログから生成された分散表現の分布は, 学期初めの数週間のログから生成された分散表現の分布とは大きく異なることが予想される．そのため, 学期初めの数週間の学習ログから生成された分散表現を入力とし，学期終了時点までの学習ログから生成されるであろう分散表現を推測し, データの分布を近づける特徴変換ネットワークを提案する．

3.1 特徴変換ネットワークの概要

対象の学生が At-risk であったかどうかの判断を考えた場合，学期終了時点までのすべての学習ログを利用したほうが高い精度が得られるはずである．このことから，学期終了時点までの学習ログから作成した分散表現を入力とする At-risk 検出器の利用が望ましい．しかし，この場合, At-risk 検出のためには学期終了時点まで待つ必要があり，学期の途中での介入といった用途には利用できない．

そこで，学期初めの数週間の学習ログから生成した分散表現を入力とし，学期終了時点までの学習ログから生成された分散表現を予測するネットワークを構築し，その出力として得られる分散表現を前述の At-risk 検出器に入力する手法を提案する．

以降, 特徴変換ネットワーク, At-risk 検出ネットワーク, At-risk 検出の手順について, 順に詳しく説明する．

3.2 特徴変換ネットワーク C_Θ の構築

図 2 は，本発表で提案する特徴変換ネットワークの訓練方法を示している．以下では，具体的な訓練手順について詳しく説明する．

学習者 i の学期終了時点までの全学習ログを $\mathcal{L}_i^{\mathrm{a}}$ とする．ここで，学期初めからある時点までの学習ログのみを抽出する関数を clip() とすると，学期初めの数週間（e 週間分）のログ $\mathcal{L}_i^{\mathrm{e}}$ は式 (1) のように表すことができる．

$$\mathcal{L}_i^{\mathrm{e}} = \mathrm{clip}(\mathcal{L}_i^{\mathrm{a}}) \tag{1}$$

学習ログから分散表現を生成する関数を E2Vec, 学期初めの数週間の学習ログを用いて E2Vec により生成した分散表現を $\mathbf{v}_i^{\mathrm{e}} \in \mathbb{R}^{100}$, 学期終了時点までのすべての学習ログを使用して生成した分散表現を $\mathbf{v}_i^{\mathrm{a}} \in \mathbb{R}^{100}$ とする (100 は定数) と，$\mathbf{v}^{\mathrm{a}}$ と $\mathbf{v}^{\mathrm{e}}$ は,

$$\begin{aligned} \mathbf{v}_i^{\mathrm{a}} &= \mathrm{E2Vec}(\mathcal{L}_i^{\mathrm{a}}) \\ \mathbf{v}_i^{\mathrm{e}} &= \mathrm{E2Vec}(\mathcal{L}_i^{\mathrm{e}}) \end{aligned} \tag{2}$$

により得られる．

ここで $\mathbf{v}^{\mathrm{e}}$ を $\mathbf{v}^{\mathrm{a}}$ に変換するネットワーク C_Θ を考える．訓練データに含まれる学習者の集合を $\mathcal{U}$, 学習者の数を N とすると，C_Θ の訓練は損失最小化として式 (3) のように定式化される．

$$\min_\Theta \frac{1}{N} \sum_{i \in \mathcal{U}} \left| C_\Theta(\mathbf{v}_i^{\mathrm{e}}) - \mathbf{v}_i^{\mathrm{a}} \right|^2 \tag{3}$$

これにより得られる C_Θ を用いることにより，学期初めの数週間の学習ログ $\mathcal{L}_i^{\mathrm{e}}$ に基づいて生成された分散表現 $\mathbf{v}_i^{\mathrm{e}}$ か

Vol.2023-CLE-41 No.18
Vol.2023-CE-172 No.18
2023/12/3

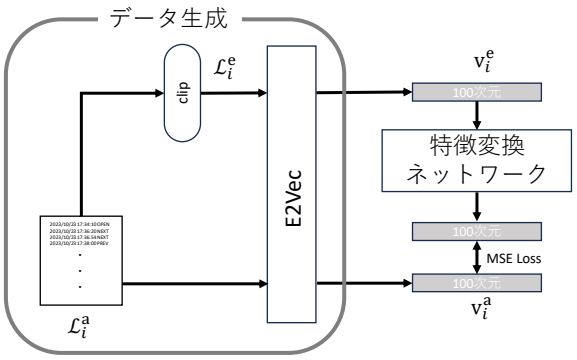

図 2 特徴変換ネットワークの訓練

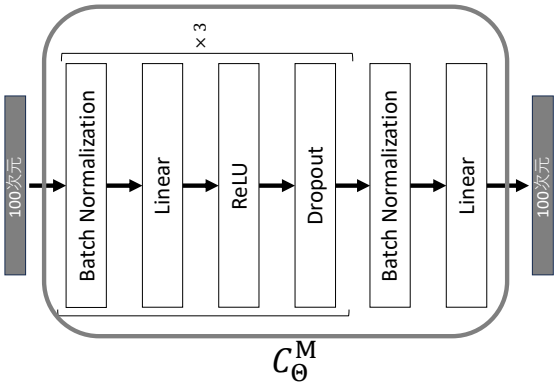

図 3 変換ネットワークの構成 (MLP)

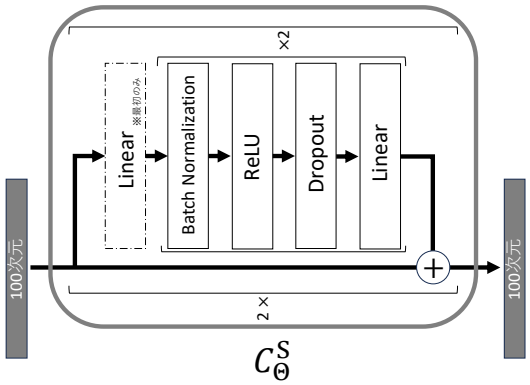

図 4 変換ネットワークの構成 (SkipLayer)

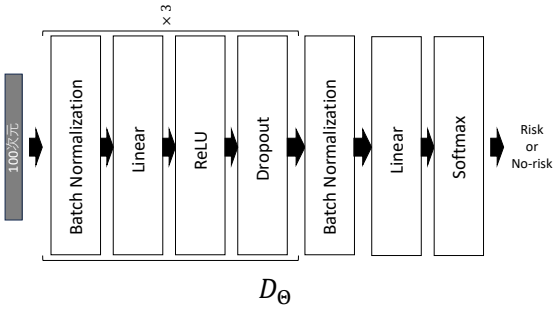

図 5 At-risk 検出ネットワークの構成

ら, 学期終了時点までの全学習ログ $\mathcal{L}_i^{\mathrm{a}}$ から生成された分散表現 $\mathbf{v}_i^{\mathrm{a}}$ を推測することが可能となる.

3.3 At-risk 検出ネットワークの構築

At-risk 検出ネットワークは，特徴変換ネットワーク C_Θ の構築とは別に行う．学習者 i が At-risk 学生であるかどうかを表すラベルを $g_i \in \{0, 1\}$ とすると，At-risk 検出ネットワーク D_Θ の訓練は式 (4) の損失最小化として定式化される．

$$\min_\Theta \frac{1}{N} \sum_{i \in \mathcal{U}} \sum g_i \log D_\Theta(\mathbf{v}_i^{\mathrm{a}}) \tag{4}$$

3.4 At-risk 検出の手順

At-risk 検出を行う際は，特徴変換ネットワーク C_Θ と At-risk 検出ネットワーク D_Θ を組み合わせて用いる．具体的な At-risk 検出の手順は以下の通りである．

(1) 学期初めの数週間の学習ログ $\mathcal{L}^{\mathrm{e}}$ を入力する
(2) 学習ログの分散表現を $\mathbf{v} = \mathrm{E2Vec}(\mathcal{L}^{\mathrm{e}})$ として得る
(3) 特徴変換ネットワークを用いて学期終了時点の分散表現を $\mathbf{v}' = C_\Theta(\mathbf{v})$ により予測する
(4) $\widehat{g} = D_\Theta(\mathbf{v}')$ により At-risk 検出結果を得る

以上をまとめると，全体としての At-risk 検出処理は次式により表される．

$$\widehat{g} = D_\Theta(C_\Theta(\mathrm{E2Vec}(\mathcal{L}^{\mathrm{e}}))) \tag{5}$$

4. 実験方法

分散表現を変換する変換ネットワークの有効性を確認するため, 実際の学習ログを用いて実験を行った．以下では，まず実験に用いた変換ネットワークの構成，At-risk 検出ネットワークの構成についてそれぞれ述べた後，具体的な実験内容について述べる．

4.1 変換ネットワークの構成

本実験では 2 種類の変換ネットワークの評価を行う．実験に用いたネットワーク構造の 1 つ目は，MLP を基本構造とした**図 3** に示すネットワーク C_Θ^{M} である．もう一つのネットワーク構造は ResNet [9] で採用されている残差接続を導入したものであり，**図 4** に示すような構成のネットワーク C_Θ^{S} である．なお，活性化関数, 線形層, バッチ正則化の挿入順序については He ら [10] の報告を参考にし，最初の入力層のみ線形層を追加した構成とした．

4.2 At-risk 検出ネットワークの構成

At-risk 検出ネットワーク D_Θ の構成を**図 5** に示す．図 5 に示すように，At-risk 検出ネットワークは，バッチ正則化，線形結合，ReLU，Dropout の順で 3 回繰り返し，バッチ正則化と線形層を通して最終出力を得る構成となっている．

表 1　各実験条件の比較

	訓練	テスト
比較 1	$\mathbf{v}_i^{\mathrm{e}}$	$\mathbf{v}_i^{\mathrm{e}}$
比較 2	$\mathbf{v}_i^{\mathrm{a}}$	$\mathbf{v}_i^{\mathrm{e}}$
提案	$\mathbf{v}_i^{\mathrm{a}}$	$C_\Theta(\mathbf{v}_i^{\mathrm{e}})$

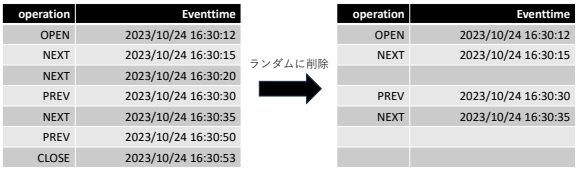

図 6　学習ログのデータ拡張例

表 2　各データの At-risk と risk の学習者数

	At-risk	No-risk	合計
A-2019	22	30	52
A-2020	10	50	60
A-2021	24	30	54
A-2022	24	28	52

4.3　実験手順

本発表で提案する変換ネットワークの有効性を確認するため, 以下の 3 種類の手法の比較実験を行った.

比較 1　At-risk 検出ネットワークの訓練とテストのいずれも学期開始後 3 週間分の学習ログを使用する.

比較 2　At-risk 検出ネットワークの訓練には学期終了時点までのすべての学習ログを使用し，テスト時は学期開始後 3 週間分の学習ログを使用する.

提案　At-risk 検出ネットワークの訓練時には学期終了時点までのすべての学習ログを使用し，テスト時は学期開始後 3 週間分の学習ログを使用する．ただし，特徴変換ネットワークを用いて学期終了時点の分散表現に変換したものを At-risk 検出に用いる.

上記 3 種類の実験条件の比較を**表 1** に示す. 学習ログは 7 週間または 8 週間にわたって行われる講義のログを使用しており, 3 種類の方法すべてにおいて，テスト時の At-risk 検出に用いる学期初めの数週間の学習ログ $\mathcal{L}_i^{\mathrm{e}}$ は，学期開始後 3 週目の講義終了時点までの学習ログとした.

また，使用したデータの At-risk の人数と No-risk の人数を**表 2** に示す．表に示す 4 年間分のデータ”A-2019”, ”A-2020”, ”A-2021”, ”A-2022”のうち 3 年分を訓練に用い，1 年分をテストデータとして用いる 4 分割交差検証により評価した.

今回の学習ではモデルの汎化性能を高めるために, **図 6** に示すような学習行動ログをランダムドロップアウトさせるデータ拡張を施したものを訓練に用いた.

表 3　At-risk 検出の正解率 (MLP を使用)

	A-2019	A-2020	A-2021	A-2022	平均
比較 1	0.750	0.633	0.667	0.692	0.686
比較 2	0.673	0.300	0.630	0.538	0.535
提案	**0.827**	**0.817**	**0.685**	**0.712**	**0.760**

表 4　At-risk 検出の正解率 (SkipLayer 使用)

	A-2019	A-2020	A-2021	A-2022	平均
比較 1	0.780	0.667	0.648	**0.712**	0.702
比較 2	0.654	0.283	0.648	0.538	0.531
提案	**0.865**	**0.750**	**0.704**	0.654	**0.743**

5.　実験結果および考察

5.1　特徴変換ネットワークの有効性

特徴変換ネットワークに C_Θ^{M}（MLP）を使用した際の At-risk 検出の正解率を**表 3** に，C_Θ^{S}（SkipLayer）を使用した際の At-risk 検出の正解率を**表 4** に示す．表の各列は，テストデータとして使用した年度を表しており，赤字は各年度で最も高い精度を示した手法を表している．これらの結果より，テスト時に特徴変換ネットワークを使用する提案手法が，変換ネットワークを用いない比較 1 ならびに比較 2 よりも正解率が向上することを確認した．このことから，学期初めの数週間の学習ログから At-risk 検出を行うというタスクにおいて，特徴変換ネットワークを用いた学期最後まで受講した際の分散表現の予測が，At-risk 検出の精度向上に寄与することを確認した.

比較 1 は，学期終了時点までの学習ログではなく，学期初めの数週間の学習ログを At-risk 検出ネットワークの学習に用いる方法である．それに対して，提案手法は学期終了時点までの学習ログすべてを At-risk 検出ネットワークの学習に用いている．この比較より，3 週間たった時点で At-risk 検出を行う場合, 3 週間のデータで学習したモデルよりも, 8 週間のすべてのデータで学習したモデルの方がよい精度を出すことができるということが言える.

さらに, 比較 2 についても考える．比較 2 は学期終了時点までの学習ログを用いる場合である．この結果を見ると, 学期終了時点までのすべての学習ログを使用して学習した At-risk 検出ネットワークは，学期初めの数週間の学習ログを入力すると At-risk 検出がうまくいかないことが分かる．この原因は，学期初めの数週間の学習ログを E2Vec に入力して得られる分散表現と学期終了時点までのすべての学習ログから生成した分散表現が大きく異なるためだと考えられる．具体的には，学期初めの数週間しか学習をしなかった学習者，つまり，学期中間，終盤は一切学習していない学習者を表す分散表現が得られているためだと考えられる.

6. むすび

本発表では，学期初めの数週間の学習ログを用いる At-risk 検出の精度向上を目的として，特徴変換ネットワークによる分散表現の変換手法を提案した．

At-risk 検出ネットワークをすべてのログデータ $\mathcal{L}_i^{\mathrm{a}}$ を使用して訓練し, テスト時には, 3 週間のログから得られた分散表現を変換ネットワーク C_Θ で変換して学期終わり分散表現を推定することで, 早期の At-risk 検出精度が向上することを確認した．この結果より, 早期に At-risk 検出を実施する場合, 特徴変換ネットワークの導入が有効である可能性が示唆された．

本発表では, clip によりログを抽出する期間を学期開始 3 週間としたが, 今後は, clip を適用する時刻を変化させた場合に At-risk 検出の精度にどのような変化が生まれるかを確認する必要がある．1 週目までで予測するモデル, 2 週目までで予測するモデル, 3 週目までで予測するモデル... のように, 複数の変換モデルをそれぞれ訓練する必要があるのか, それとも, 1 週目まで, 2 週目まで, 3 週目まで... のログをすべて混合し, 1 つの変換モデルとして構築することが可能なのかを確認する必要がある．

また, 本発表では, 同じ授業のみでの適用に限られていたが, 他授業のデータで At-risk 検出を行う場合にも有効であるかどうかの確認をする必要がある．

謝辞 本研究の一部は，JST，CREST，JPMJCR22D1 の支援を受けたものである．

参考文献

[1] Shimada, A., Mouri, K., Taniguchi, Y., Hiroaki, O., Taniguchi, R. and Konomi, S.: Optimizing assignment of students to courses based on learning activity analytics, Proceedings of the 12th International Conference on Educational Data Mining, pp. 178-187, 2019.

[2] Okubo, F., Ogata, H. and Shimada, A.: Browsing-pattern mining from ebook logs with non-negative matrix factorization, Proceedings of the 9th Internation Conference on Educational Data Mining, pp. 636–637, 2016.

[3] Okubo, F., Yamashita, T., Shimada, A., Taniguchi, Y. and Konomi, S.: On the prediction of students' quiz score by recurrent neural network, CEUR Workshop Proceedings, Vol. 2163, 2018.

[4] Okubo, F., Yamashita, T., Shimada, A. and Konomi, S.: Students' performance prediction using data of multiple courses by recurrent neural network, Proceedings of the 25th International Conference on Computers in Education, pp. 439–444, 2017.

[5] 椎野 徹也, 峰松 翼, 島田 敬士, 谷口 倫一郎：デジタル教材の学習ログと成績の関連分析, 研究報告教育学習支援情報システム (CLE), Vol. 2020, No. 10, pp. 1–4 , 2020.

[6] Yin, C., Ren, Z., Polyzou, A. and Wang, Y.: Learning Behavioral Pattern Analysis Based on Digital Textbook Reading Logs, Proceedings of HCII 2019, pp.471–480, 2019.

[7] Bojanowski, P., Grave, E., Joulin, A. and Mikolov T.: Enriching word vectors with subword information, Transactions of the Association for Computational Linguistics, vol.5, pp. 135–146.

[8] 宮崎 佑馬, 峰松 翼, 谷口 雄太, 大久保 文哉, 島田 敬士：教育データの分散表現生成手法の提案と At-risk 学生検知への応用, 第 40 回教育学習支援情報システム研究発表会 (CLE40), 2023.

[9] He, K. M., Zhang, X., Ren, S. and Sun, J.: Deep residual learning for image recognition, Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, 2016.

[10] He, K., Zhang, X., Ren, S. and Sun, J.: Identity mappings in deep residual networks, Proceedings of the 14th European Conference on Computer Vision. pp. 630–645, 2016.