

Deep Learning を用いた中途退学者推論に関する一検討

Research on Inference of Dropout Students using Deep Learning

高橋 良貴^{*1}, 小松川 浩^{*2}
Yoshitaka Takahashi^{*1}, Hiroshi Komatsugawa^{*2}^{*1}千歳科学技術大学 総合光科学部^{*1} Chitose Institute of Science and Technology^{*2}千歳科学技術大学 理工学部^{*2} Faculty of Science and Technology, Chitose Institute of Science and Technology
mail: taakhashi213@kklab.spub.chitose.ac.jp

あらまし：本稿では限られたデータの中で Deep Learning の教育分野での有用性について検討した。具体的には Deep Learning を用いた中途退学者を推論するプログラムを作成し、A 大学の学生の学習履歴などのデータを用いて中途退学者の推論を行った。推論をする際の学習回数とデータを学習と推論に用いる割合を変更し推論結果の変化を調べた。推論の結果は推論精度が約 30%となった。

キーワード：Deep Learning, データ活用, 中途退学者

1. はじめに

近年, Deep Learning が注目されており画像認識だけではなく経済や医療の分野など様々な分野での活躍が期待されている。一方で教育の分野において中途退学者の増加が問題となっており, これまでは学生の SNS(Social Networking Service)の投稿を構文解析したり LMS(Learning Management System)の学習履歴を分析したりしてデータを活用していた。Deep Learning ではビックデータなどの多量のデータを用いて学習を行うが, 個別の教育機関単位ではデータはそれほど多くない。本研究ではこうした限られたデータの中での Deep Learning の教育分野における有用性の検証を目的とし A 大学を事例に, 中途退学者の推論の可能性について調べた。

2. 推論

本研究における Deep Learning を用いた中途退学者の推論をする流れを図 1 に示す。本研究では Deep Learning を用いた中途退学者を推論するプログラムを作成し, A 大学のデータを用いて推論を行った。推論では学習の設定を変えて推論の結果の変化を調査した。

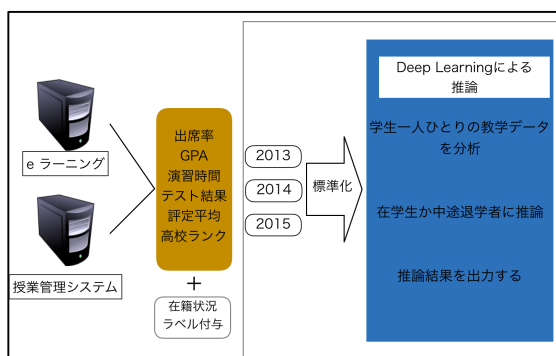


図 1. 推論プログラムの流れ

2.1 推論プログラム

中途退学者の推論には Java 言語の Deep Learning 用フレームワークである DL4J を用いた。DL4J から提供されている CSVExampleEvaluationMetaData.java を改変し中途退学者を推論するプログラムを作成した。推論プログラムが中途退学者を推論した人数と実際に退学した人数で割った数を推論精度とする。Deep Learning が推論した中途退学者の数が実際に退学した学生の数に近いと推論精度が高くなる。推論精度は推論に用いるデータの量や学習の設定や層の構成によって決まる。

2.2 活用データ

本研究で用いたデータは本研究チームの先行研究において主成分分析と SOM (Self-Organizing Map) によって中途退学者をクラスタリングした際に用いた既にある程度中途退学者の傾向がわかるデータを採用した(1)。データは A 大学における, 2013 から 2015 年度の入学前教育を受けた学生群の計 539 名の学習履歴などである。A 大学における, 2013 から 2015 年度の入学前教育を受けた学生の在学状況を表 1 に示す。A 大学における 2013 から 2015 年度の入学前教育を受けた学生の退学率は 7.2% である。

表 1 A 大学の在学状況

2013 年度	2014 年度	2015 年度
在学生 204 名	在学生 163 名	在学生 133 名
退学者 25 名	退学者 9 名	退学者 4 名

データは e ラーニングの学習履歴(演習時間) や, 授業の出席率などのデータ全 7 項目を標準化し

たものを用いる。入力データから学生の特徴を学習し、在學生か中途退學者のどちらかが推論結果として出力される。

2.3 試行実験

本研究では学習回数と学習の割合による推論精度の変化に注目した。推論精度が高くなる学習回数と推論の割合の設定を検証し、推論精度が高かった設定で繰り返し推論を行いその時の変化を調査した。学習回数の調査では学習回数を50回から2000回まで変化させた時の推論精度の変化を調べた。

学習の割合の調査では、入力データの学習に用いるデータと推論に用いるデータに分ける割合を変動させて推論結果の変化を調べた。例えば、入力データから8割のデータを学習に用いて2割のデータを推論に用いるときは、539名の8割である431名を学習に用いて残りの108名を推論に用いる。推論に用いる108名の中でも在學生と中途退學者の比率が8:2となっており、108名の2割にあたる22名からさらに2割の8名の中途退學者を推論に用いる。学習と推論に用いる在學生と中途退學者の割合を変えて推論精度の変化を調べた。

学習回数と学習の割合で推論精度が高かった組み合わせの設定で繰り返し推論した時の推論結果を調べた。

3. 推論結果

学習回数の変化を図2に示す。青い線が推論の精度を示しており、オレンジ色の線がA大学の退學率である7.2%を示している。

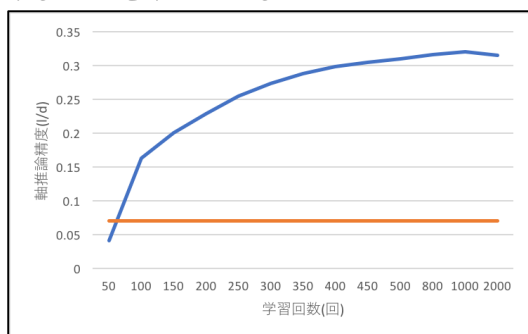


図2. 学習回数による推論精度の変化

図2の通り学習回数を50回から2000回までの推論結果の変化を調査したところ、学習回数が400回を超えると推論精度が約30%に収束した。Deep

Learning では早期停止と呼ばれる現象があり、学習に用いるデータと学習回数の割合によっては学習の精度が収束せずに低下することがある(3)。今回の結果から推論精度が低下せずに早期停止の状態にならなかったことからデータ量と学習回数の設定が妥当であったと言える。

学習比率は学習に8割のデータを用いて推論に2割のデータを用いたときであることがわかった。

学習回数を400回に設定し学習と推論に用いるデータを8:2に設定した推論を1000回繰り返して行った。さらにその推論を3回繰り返したときの標準偏差を表2に示す。

表2 繰り返し推論を行ったときの推論精度

1回目	2回目	3回目
31.57%	31.72%	31.45%

表3 繰り返し推論を行ったときの標準偏差

1回目と2回目	2回目と3回目	1回目と3回目
0.0010	0.0011	0.0001

繰り返し推論を行った結果、今回は約31%の推論精度であることがわかった。

4. 今後の課題

今回は限られたデータの中で推論を行ったため推論精度が約30%であった。推論精度を向上するためにデータ量の増加と推論プログラムの深層化があげられる。データ量に関しては毎年データが蓄積されるため毎年増えることが期待できる。学習回数の調査においては早期停止が見られなかったことから、データ量と学習回数を増やすことで推論精度が向上する可能性がある。本研究では本研究チームの先行研究が用いた7つの項目で推論を行ったが、今後は項目を変えることで中途退學者の推論精度の向上に期待できる。これまでの主成分分析やSOMでは扱わなかったデータを用いた推論の結果を示す可能性がある。また、項目数の増加にともない、学習の設定と層の変更が必要となるため、あらためて最適な設定を調査し中途退學者の推論精度の向上が期待できる。

参考文献

- (1) 高橋駿嗣:”自己組織化マップを用いた退學者特性発見システムに関する研究”, 千歳科学技術大学, 修士論文, 2016
- (2) SKYMAIND社:”Early Stopping”, <https://deeplearning4j.org/earlystopping> (2017年2月12日アクセス)