

機械学習を用いたコンピテンシー評価の分析

Analysis of competency assessment using machine learning

紅葉 亜練^{*1}, 桶田 昂史^{*1}, 山川 広人^{*1}, 小松川 浩^{*1}

Aren MOMIJI^{*1}, Koji OKETA^{*1}, Hiroto YAMAKAWA^{*1}, Hiroshi KOMATSUGAWA^{*1}

^{*1} 公立千歳科学技術大学 理工学部

^{*1} Faculty of Science and Technology, Chitose Institute of Science and Technology

Email: b2171880@photon.chitose.ac.jp

あらまし：大学では専門的知識の習得を重要視しているが、現代社会では行動特性コンピテンシーの育成を重要視している。そこで科研は、数学を事例に学生のコンピテンシーが授業態度に現れると予想し、コンピテンシーを測るためのアンケートを実施し、教員に共通の指標が不明瞭なままコンピテンシーの評価をつけてもらっていた。本研究は、教員が何か教員の中にある指標をもって特徴量を捉えていると予想し、機械学習によって再現性の有無を分析した。

キーワード：コンピテンシー、評価、機械学習

1. 背景

従来、大学では学問領域を学ぶための基礎学力や知識修得、専門性の理解等の認知能力を意識した教育体系が構築されてきた。一方、現代社会ではこれらとは異なる問題解決力や、継続的な学習力、主体性、協調性などのテストで測ることのできない行動特性であるコンピテンシーの育成を重要視している。そこで、本研究では、数学を事例に学生の授業態度にもコンピテンシーが現れると予想した。そして、複数の大学の数学教員と連携して、実際に数学の授業の開始前と開始後にコンピテンシーの変化を問うためのアンケートを実施した。学期のはじめに実施したアンケートの内容を図1に示し、図2は学期終了後に実施した図1のアンケート内容に加えた設問内容を示す。その上で、その学生のアンケートの回答に対して、各所属機関の数学教員がコンピテンシーの有無をループリック評価を行うこととした。具体的には数字で1から3までの値で紐付けられている。本研究では、こうした教員の評価は、学生の文章から何らかの能力に関わる特徴量を推定して判断していると考え、機械学習で分析をして特徴量把握できるかを検証することを目的とした。

Q3. この授業で最も興味をもって取り組んだことを具体的に書いてください。ただし、回答は箇条書きで書いてください。

図1 授業の開始後に追加する設問の内容

2. 目的

本研究の目的は、教員による学生のコンピテンシーの評価を総合的に判断する手法について機械学習を用いて分析し、教員による学生のコンピテンシーの評価を再現可能であるか検証を行うことである。具体的には、文章中の単語を Word2Vec を用いて次元圧縮し、ニューラルネットワークの入力ベクトルを作成し、教員による評価を機教師信号として学習を行い結果を見る。実際に高い精度で再現可能ならば、学生の文章に特徴量が存在すると言える。

3. 分析の方法

本研究では、学生の振り返りなどの文章データを形態素解析によって単語に分け、Word2Vec を用いて単語ベクトルに次元圧縮する。単語ベクトルから文章ベクトルを作成し、それに特徴量に含まれると予想される単語の長さ、単語の数、キーワードの数などを足し入力ベクトルを作成する。また、教員による学生のコンピテンシーの評価を教師信号として学習させ、テストデータを用いて精度を見る。テストデータの評価を1から3までの値に高い精度で分類ができていれば教員による評価の再現が可能となり、入力ベクトルの中に特徴量があることがわかる。

4. 本研究で使用する技術

本研究では機械学習を機能させる技術の一つであるニューラルネットワークを用いる。ニューラルネットワークは適確な推論を行うため、最適な重みとバイアスを決定する。決定した目標と実際の出力の誤差を損失関数とし、勾配法を用いて重みとバイアスを更新し、損失関数を最小値に近づけることで精

Q1. 選択
今のあなたにとって数学は大切なものですか？いずれの1つをチェックしてください。
5: とても大切 4: やや大切 3: どちらでもない 2: あまり大切でない 1: 全く大切でない

Q1. 記述
そう思う理由は何ですか？箇条書きでお答えください。

Q2. 選択
数学は現実場面でどれくらい役に立つと思いますか？いずれか1つをチェックしてください。
5: とても役に立つ 4: やや大切 3: どちらでもない 2: あまり役に立たない 1: 全く役に立たない

Q2. 「とても役に立つ」あるいは「やや大切」を選んだ人は、数学が役に立つ例を具体的に箇条書きで挙げてください(1つ以上)。思いつかない場合は「思いつかない」と書いてください。

図1 授業の開始前に行う設問の内容の例

度の高い推論を可能にする。今回はニューラルネットワークを用いるため、TensorFlow と Keras というこの二つのライブラリを用いて分析を行う。

4.1 TensorFlow

TensorFlow は、python だけでなく C 言語や Java など様々な環境で動作する、オープンソースの機械学習ライブラリ⁽¹⁾である。データフローグラフを使用し、計算、共有状態、およびその状態を変更ができる。

4.2 Keras

Keras は Python で書かれた、TensorFlow または CNTK, Theano 上で実行可能な高水準のニューラルネットワークライブラリ⁽²⁾で、迅速な実験を可能にすることに重点を置いて開発された。数学的理論の部分を短いソースコードで実装することができる。

4.3 Word2Vec

Word2Vec とは、文章から単語をベクトル化する技術である⁽³⁾。単語をベクトル化することで、単語と単語、文章と文章の類似度を測ることができる。

本研究では、教員からの評価が高い文章と類似した文章を推論することに用いる。

5. 分析の結果

今回の分析では、コンピテンシーを測るために大学生、高専生を対象とした合計 11 機関から収集したアンケート結果を用いる。本研究にて用いたデータ数を表 1 に示し、分析結果を表 2 に示す。訓練データとテストデータのレートは 0.7 とする。また、入力ベクトルの作成には、Word2Vec で 100 次元に圧縮した単語ベクトルを用いる。

表 1 表 2 に用いたデータ数

評価	データ数
1	356
2	564
3	682

今回用いたデータは、教員に 3 の評価をつけられたデータに偏っており、1 の評価をつけられたデータは少なかった。

表 2 均一なデータ数ではない場合の検証結果

回数	損失	精度
1 回目	0.229	0.511
2 回目	0.225	0.516
3 回目	0.235	0.491
4 回目	0.226	0.547
5 回目	0.264	0.428
平均	0.236	0.497

表 1 から表 2 は偏ったデータ数で分析を行っている。そのため用いるデータの中で教員に 2 の評価を

つけられたデータ数、3 の評価をつけられたデータ数をそれぞれ教員に 1 の評価をつけられたデータ数と同じになるようランダムにデータを省き、再度分析を行った。その結果を表 4 に示す。

表 3 表 4 に用いたデータ数

評価	データ数
1	356
2	356
3	356

表 4 均一なデータ数である場合の検証結果

回数	損失	精度
1 回目	0.253	0.486
2 回目	0.256	0.489
3 回目	0.242	0.520
4 回目	0.241	0.536
5 回目	0.263	0.470
平均	0.251	0.500

6. 考察

表 1, 表 3 の精度に違いがあまり見られなかった。また、表 1 と表 3 からテストデータの評価 1, 2, 3 の 3 分割中の精度が 0.49 という結果が得られた。精度としては高いとは言えないが、教員の中にある指標をもって感覚的に特徴量を捉えて評価している可能性が示唆される結果となった。また、表 2, 表 4 を比較し精度に違いがあまり見られない点、そして精度が上がらなかった要因としてデータ数が少なく機械学習で特徴量を捉えきれなかったことが原因であると考えられる。具体的な解決策として、他大学の協力の下データの総数を増やし再度分析を行う、他に特徴量に含まれると思われる要素を足し、入力ベクトルを作成するなどが挙げられる。

参考文献

- (1) TensorFlow ホワイトペーパー
<https://www.tensorflow.org/about/bib?hl=ja> (2021 年 2 月 6 日アクセス)
- (2) Keras: Python の深層学習ライブラリ <https://keras.io/ja/> (2021 年 2 月 6 日アクセス)
- (3) Word2vec embeddings
<https://radimrehurek.com/gensim/models/word2vec.html> (2021 年 2 月 9 日アクセス)