

機械学習モデリングを用いたコンピテンシー評価の分析

紅葉 亜練^{*1}, 小松川 浩^{*1}

^{*1} 公立千歳科学技術大学大学院

Analysis of competency assessment using machine learning modeling

Aren Momiji^{*1}, Hiroshi Komatsugawa^{*1}

^{*1} Chitose Institute of Science and Technology

現代社会では、コンピテンシーが重要視され、この能力の養成には、学習者が自らの学びを振り返り、教員がその評価を返す教育上のやり取りが重要となる。しかし、教員の学生に対する評価は暗黙知によって行われ、数値的評価や可視化を伴うシステム化の障壁となっている。そこで学生の自由記述形式の振り返りデータから、教員の暗黙知に基づく評価と関連する特徴量の抽出し、機械学習的アプローチによるモデル化を試みた。

キーワード: コンピテンシー, 機械学習, ルーブリック, アンケート, 数学, 文章

1. 背景

従来大学では、基礎学力や専門性などの知識といったテストなどで数値化して測ることのできる認知能力を意識した人材育成を行っている。

一方現代社会では、問題解決力や主体性といった数値化できない非認知能力の養成が重要視される。こういった能力は主に行動特性に左右され、コンピテンシーと呼ばれる。この能力の養成には、学習者が自らの学びを振り返り、教員がその評価を返す教育上のやり取りが基本となる。しかし、教員の学生に対する評価は暗黙知によって行われ、数値的評価や可視化を伴うシステム化などのモデル化が難しい。

2. 目的

本研究の目的は、自然言語処理を伴う機械学習的アプローチによるモデル化を通じて、学生の自由記述形式の振り返りデータから、教員の暗黙知に基づく評価をどこまで再現できるか検証し、教員の判断要素を検討することである。具体的には、数学の振り返りに関する既に行っていたアンケートの結果を活用し、学習者の自由記述回答に教員の協力のもと3段階でスコアを付与する。教員が学習者のコンピテンシーを評価す

る基準であると考えられる特徴の候補を挙げ、各特徴を用いて XGBoost と BERT による機械学習モデリングを行い、2つのモデルの精度の比較を行う。その精度比較から、特徴が教員の判断とどれほど関係があるかを調べる。

3. 分析手法

本研究では、数学を事例として複数の大学の数学教員と連携し、数学の授業開始前と開始後にコンピテンシーの変化を問うアンケートを実施した。

実施したアンケートの内容を図1に示す。学期開始前には Q1, Q2 のみを実施し、学期終了時には Q1~Q4 を回答してもらった。Q1, Q2 の回答は学習者の自己評価のルーブリックと自由記述式の回答で構成される。

アンケートには、学習者の自由記述回答に教員の協力のもと3段階でスコアを付与した。スコアは教員が評価した学習者のコンピテンシーを表す。アンケート結果から得られたルーブリックや記述された文章を入力として、教員による3段階のスコアを教師信号として学生の振り返りに機械学習的なアプローチを適用することで、教員の暗黙知に基づく評価をモデル化できるか検証する。また、図2のような別アンケートへの

適用可能性の検証も行う。

Q1. 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つを選択してください。また、そう思う理由を文章で具体的に教えてください。

1: 全く大切でない 2: あまり大切でない 3: どちらでもない 4: やや大切 5: とても大切

Q2. 数学は現実場面でどれくらい役に立つと思いますか？また、数学が役に立つ例を文章で具体的に教えてください。

1: 全く役に立たない 2: あまり役に立たない 3: どちらでもない 4: やや役に立つ 5: とても役に立つ

Q3. この授業で最も興味をもって取り組んだことを文章で具体的に教えてください。

Q4: この授業で学んだことを、今後どのような学びにつなげていきたいですか。文章で具体的に教えてください。

図1 実施したアンケートの内容

Q1(責任感): この授業で、最後まで意識して、最後まで行うように心がけたことを書きましょう。

Q2(主体性): この授業でもっとも興味をもって取り組んだことを具体的に書きましょう。

Q3(協調性): この授業で自らが他者に積極的に働きかけた・貢献できたことがあれば書きましょう。その結果、自らの学習成果の向上に繋がっているかも考えてください。

Q4(倫理観): この授業で将来の自分に役立つと思ったことを書きましょう。それに関連して、将来社会で貢献したい・できると感じたことがあれば書いてみてください。また、その影響で自らの学習態度に変化があったかも考えてください。

図2 本学で実施した別アンケートの内容

特徴として考えられる要素として単語数、語彙数、キーワード数を挙げる。キーワードは、学習者の文章から2種類のキーワードを定義し、数学に関する単語として理数ワードを、コンピテンシーに関する単語として能力ワードとして定義する。

本研究では、教員による3段階のスコアを教師信号とし、XGBoostには、ルーブリック、単語数、語彙数、

理数ワード数、能力ワード数を特徴として入力し、特徴量を明示的に指定することで、教員が専門キーワードを含んでいるか、単語数は多いか、などで捉えているかを判断する。BERTには、自由記述から得られた文章を入力し、教員が学習者の文章を総合的に捉えているかを見る。その後、XGBoostにBERTの推論結果を特徴として入力しモデリングを行い、shapを用いて入力した各特徴の影響度を見て、RandomForestで推論の際にどの要素が重要視されているかを見る。

4. 使用技術

4.1 XGBoost⁽¹⁾

XGBoostとは、複数の決定木を用いて予測モデルを作成する機械学習手法であり、特徴を明示的に入力することができる。数値データ以外を入力することができず、2本目以降の木は、目的変数とこれまでの木による予測値の差に対して学習が行われる。予測値は各木で属する葉のウェイトの和をとって計算される。これは、教員に置き換えると学習者の文章に明示的に指定した特徴量で捉えているかを判断することに相当する。

4.2 BERT⁽²⁾

BERTとは、2018年にGoogleの論文で発表された自然言語処理モデルである。学習できるのは文章データのみであり、従来の自然言語処理モデルのWord2Vecとは異なり、文を双方向に学習する。そのため、BERTは文脈を理解することができる。これは、教員に置き換えると学習者の文章を総合的に捉えて判断することに相当する。

4.3 shap⁽³⁾

shapとは、SHapley Additive exPlanationsの略であり、協力ゲーム理論のシャープレイ値を機械学習に応用したものである。ある特徴を抜いたとき、他の特徴との組み合わせを考慮した上でその特徴がどれほど推論精度に影響を与えるかを見ることができる。

4.4 RandomForest

RandomForestとは、複数の決定木を用いて予測モ

デルを生成する機械学習手法であり、XGBoost とは異なり、前の木の結果を考慮しないため入力した特徴がそのまま結果に影響を与える。そのため特徴を抜いたときの影響が分かりやすく、ジニ不純度から、ある特徴で分割することでどれくらいジニ不純度を下げられるかを求めることができる。

5. 結果

XGBoost と BERT のモデルの精度を表 1 に示す。精度は 3 段階のスコアから全てのスコア、1 のスコアを除いたもの、2 のスコアを除いたもの、3 のスコアを除いたものを示している。

表 1 モデルの推論精度の一覧

	1, 2, 3	2, 3	1, 3	1, 2
XGBoost	61%	63%	86%	79%
BERT	64%	70%	86%	78%

表 1 より、BERT モデルの精度が 3%高くなっている。そのため、ここから得られる数値的な指標を特徴量と定義し、BERT の推論結果を加える。また、実際にどのような特徴が有効であるか見るため、shap を用いて各特徴の影響度を見たものを図 3 に示し、RandomForest を用いて各特徴の強さを見たものを図 4 に示す。

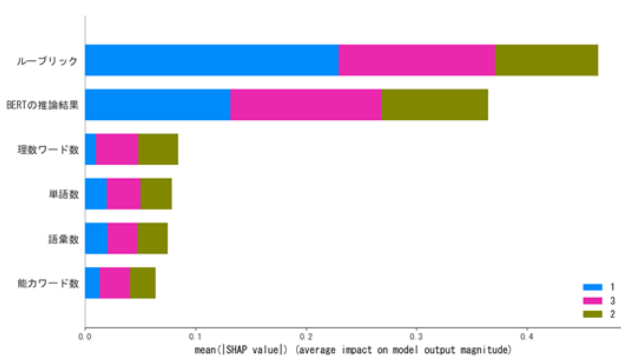


図 3 shap の各特徴の影響度

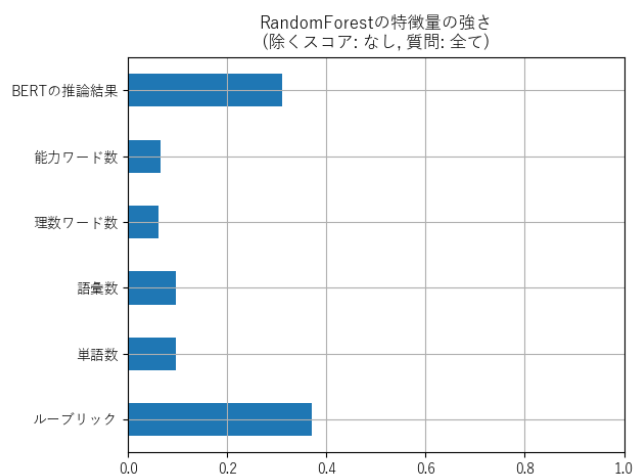


図 4 randomforest の各特徴の強さ

図 3, 図 4 より、BERT の推論結果とルーブリックを除いた時精度が大きく下がったことがわかる。よって BERT の推論結果とルーブリックを入力したものを本研究の機械学習モデリングとし、XGBoost モデルの推論精度の比較を表 2 に示す。本学の授業の振り返りのアンケートで同様に検証した結果を表 3 に示す。

表 2 XGBoost モデルの推論の精度の比較

	1, 2, 3	2, 3	1, 3	1, 2
特徴分析前	61%	63%	86%	79%
特徴分析後	72%	75%	93%	83%

表 3 授業の振り返りアンケートにおける精度

	1, 2, 3	2, 3	1, 3	1, 2
XGBoost	59%	64%	81%	75%

6. 考察

表 1 より、BERT モデルの精度が 3%高くなっていることから教員は文脈を考慮して判断していると考えられる。図 3, 図 4 より、教員は、明示的に評価しているはずなので、BERT の推論結果、ルーブリック、特徴と予想した要素を入力して評価すると、BERT の推論結果とルーブリックを組み込むのが最も効果があり、他はあまり効果がないといえる。表 2 より、XGBoost モデルの精度は 72%であったため、教員の暗黙知を 7 割程度の精度で定量的に把握できることが示された。表 3 より、XGBoost モデルの精度は 59%であったため、問いかけ方に関わらず、分類できる可能性が示唆された。

参 考 文 献

- (1) Tianqi Chen, Carlos Guestrin. “XGBoost: A Scalable Tree Boosting System”. cornell university. [1603.02754.pdf \(arxiv.org\)](#). (accessed 2022-02-07)
- (2) Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. cornell university. [1810.04805.pdf \(arxiv.org\)](#). (accessed 2022-02-07)
- (3) Scott Lundberg, Su-In Lee. “A Unified Approach to Interpreting Model Predictions”. [1705.07874.pdf \(arxiv.org\)](#). (accessed 2022-6-16)