

## テキストマイニングによる学習者の特性と理解度の分析

### Analysis of Relationship between Learner's Characteristics and Level of Understanding using Text-mining

東 るみ子

Rumiko AZUMA

日本大学商学部

College of Commerce, Nihon University

Email: azuma.rumiko@nihon-u.ac.jp

**あらまし**：近年初等教育向けに、AI やビッグデータを活用することで、学習者のつまづきを分析し、学習者のレベルに合わせた教育支援ができるシステムが開発されている。一方大学教育では、統一された指導要領が存在しない為、システムの導入が難しい状況にある。本稿では、大学で利用されているリアクションペーパーなどの学生の自由記述の文章から、授業に対する姿勢、理解度など学習者特性を定量的に分析することで、成績に関連する要因を探る。

**キーワード**：テキストマイニング，学習分析，リアクションペーパー

#### 1. はじめに

現在多くの大学では、授業改善を目的とした授業評価アンケートが実施されている。しかし授業評価アンケートでは、選択肢による回答が中心となっているため、各項目に対する多くの情報が捨象されてしまうことがデメリットとなる。そこで大学の講義では、学生の学びを把握するために、担当教員が積極的に自由記述を中心とするリアクションペーパーなどを利用するケースが増えてきている。

学習者のつまづきや理解度を把握する手法として、近年機械学習を使った分析も進んでおり、多くの学習データを用いることで、学習者の理解度を予測することが可能となっている。しかしながら、大学の一講義において機械学習を利用するにはハードルが高く大量のデータを集めることも容易ではない。

そこで本稿では、大学で利用されているリアクションペーパー等の自由記述を定量的に分析し、記述内容のどのような要素が学生の理解度と深い関連があるのかを分析する。本稿では、学生の記述内容をより細かく分類し、成績との関連を分析することで、学生の記述から理解度を把握することを目的とする。

#### 2. LMS を使った学習の振り返りシート

リアクションペーパーに関する先行研究では、その機能を検討した研究や、使用することによる効果を明らかにした研究が多く存在する<sup>(1)</sup>。一方、リアクションペーパーから得られる情報と学生の理解度の関連を分析している研究は少ない。本稿では、カリフォルニア大学バークレー校で実施された Minute Paper<sup>(2)</sup>（「今日学んだことで一番大切なことは何か。今日の講義における疑問点は何か。」を1分間で書かせるもの）をベースにし、学生の主観的評価項目を加えた学習の振り返りシートを用いて分析を行った<sup>(3)</sup>。シートの内容は次の通りである。

設問1：本日の講義の理解度（5段階評価）

設問2：予備知識（予習）（5段階評価）

設問3：講義に関する「振り返り」（その日に学んだ内容、理解した内容、分からなかったこと、質問など）（自由記述）

#### 3. 分析方法

首都圏の私大文系学部で1年生向けに開講されている科目「統計入門（必修）」の中で、学習の振り返りシートを使用した。シートの配布と回収には、LMS（Learning Management System）を利用し、学生には毎回講義終了前に入力を促した。学生がLMSにアクセスする際、学生自身のIDでログインが必要となるため、各学生の識別には、回答に付随するログインIDを使用した。シートの設問1～3の内容に加え、受講生の定期試験の得点、小テストの得点も合わせて分析を行った。

#### 4. 分析結果

##### 4.1 理解度と予備知識度合いに関する集計

講義13回分（2回は中間試験と解説のため除く）の理解度と予備知識の集計を行ったところ、理解度は5段階評価で平均3.79、標準偏差0.58、予備知識は平均2.48、標準偏差0.72であった。このことから、学生は授業内容をある程度理解していると主観的に感じており、予備知識としては単語程度を聞いたことがあるレベルであることが分かった。

##### 4.2 講義に関する「振り返り」記述の分析

自由記述の文章を定量的に分析するために、テキストマイニングのソフトウェア KHCoder<sup>(4)</sup>を用いた。その際、頻度が多いがそれ自体意味を持たない一般的な語は分析からはずした（例：ある、いる、なる）。また、複合語としてまとめられる語（例：標準偏差、相関分析）は1つにまとめた。分析の対象となる文書数は783（学生66名分）、語の数は1592、出現回数平均は9.47であった。

本稿では、振り返りの内容により学生の理解度が

異なるのではないかと考え、振り返り内容を表1のように分類し、各文書にタグ付けを行った。「軽い振り返り」「深い振り返り」に関しては、コメントの文字数なども考慮した。また、「軽い振り返り」「深い振り返り」「授業内容」「感想」「未分類」タグは、各文書にいずれか1つ付与されており、複数のタグが付与されている文書に関しては、人間が直接目視で確認することで分類の修正を行った。

表1 分類の種類

| タグ名    | 分類内容  | 文書数 |
|--------|---|-----|
| 軽い振り返り | 自分の理解に関する単純な記述。単に「～を理解した」「～が分かった」「～を知った」「分からなかった」などの記述。                 | 455 |
| 深い振り返り | 自分の理解度に関する記述で、理解した理由やどのようなところが理解できたのか、なぜ分からなかったのかなど、内容の記述が詳しく書かれているケース。 | 187 |
| 授業内容   | 単にその日学んだ内容のみが書かれている記述。  | 24  |
| 感想     | 「～頑張りたい」「もっと復習したい」「計算が面倒くさい」など感想のみの記述。                                  | 115 |
| 未分類    | 上記4つの分類に当てはまらない内容。  | 2   |
| 要望     | 授業や説明に対する要望の記述。   | 23  |
| 疑問     | 具体的にどの部分が、と指定した上で「なぜ」そうなるのかがわからない」とか「～に疑問を感じた」などの記述。                    | 88  |
| 質問     | 具体的に書いた内容の後に「～でしようか?」「～知りたい」などの記述。                                      | 11  |

### 4.3 テストの得点との関連分析

はじめに、学生の主観的な理解度および予備知識と中間・期末試験・小テストの得点の相関を調べた(表2)。理解度と小テストの得点では、中程度の正の相関がみられた。また、理解度は中間試験とも弱い正の相関がみられた。

表2 主観的評価と各テストの得点の相関係数

|      | 期末試験   | 中間試験    | 小テスト      |
|------|--------|---------|-----------|
| 理解度  | 0.167  | 0.270 * | 0.477 *** |
| 予備知識 | -0.177 | 0.091   | 0.122     |

\* $p<.05$ , \*\* $p<.01$ , \*\*\* $p<.001$

次に、学生の自由記述に対する各分類のタグ数と各テストの得点の相関を調べた。その結果を表3に示す。「深い振り返り」は、期末試験と小テストにおいて弱い正の相関を示した。また「疑問」タグに関しても、同様に弱い正の相関がみられた。

さらに、期末試験の得点を予測するのに最適な項目を分析するために、期末試験の得点を目的変数、理解度、予備知識および各分類タグの集計数を説明変数とし、ステップワイズ法による重回帰分析を行

った。その結果、「軽い振り返り」( $\beta=0.422, p<.001$ )と「深い振り返り」( $\beta=0.570, p<.001$ )を説明変数としたとき、AICが最も小さく最適なモデルとなった。一方、決定係数は0.211( $p<.001$ )と低かった。

以上の結果から、学生の主観的な理解度と同様に、自由記述に関しても、「深い振り返り」「疑問」の記述内容が、試験やテストの得点と関連があることが分かった。すなわち、「深い振り返り」を行う学生ほど、講義内容をより理解しようとしており、その結果得点も高くなる傾向にあった。このことから、学生自身が行う評価だけではなく、自由記述の内容も定量的に分析をすることで、学生の真の理解度をより精度が高いかたちで測ることが出来るのではないかと考える。さらに興味深い結果として、「感想」のみ多く述べている学生ほど、中間試験や小テストの得点が低くなる傾向にあった。これは、「難しかった」や「簡単だった」などの単純な授業感想が、授業内容に深くコミットしていないことを表しており、その結果として試験や小テストの得点との間に負の相関が表れたのではないかと考える。

表3 自由記述の分類と各テストの得点の相関係数

|      | 期末試験    | 中間試験     | 小テスト       |
|------|---------|----------|------------|
| 軽い理解 | 0.076   | 0.223    | -0.096     |
| 深い理解 | 0.314 * | 0.090    | 0.432 ***  |
| 授業内容 | -0.062  | -0.191   | 0.004      |
| 感想   | -0.367  | -0.319 * | -0.466 *** |
| 要望   | 0.030   | 0.150    | 0.116      |
| 疑問   | 0.272 * | 0.091    | 0.431 ***  |
| 質問   | 0.210   | 0.111    | 0.192      |

\* $p<.05$ , \*\* $p<.01$ , \*\*\* $p<.001$

### 5. おわりに

本稿では、中間試験までの7回分のデータと中間試験の得点の関係、および全13回分のデータと期末試験、小テストの得点の関係を分析した。今後の課題として、期間ごとの集計データだけではなく、毎回の講義ごとのデータを使って分析を行う必要がある。また、自由記述の分類に関しても、本稿で使用したタグとは別の観点から分類を行うことで、学生の理解度と関連の深い要素を見つけることができな

#### 参考文献

- (1) 須田昂宏: “リアクションペーパーの記述内容をデータとしてどう活用するか—研究動向の検討を中心に—”, 教育論叢, 58, pp.19-34 (2015)
- (2) B.G. Davis, R. Wilson, L. Wood: 授業をどうする!—カリフォルニア大学バークレー校の授業改善のためのアイデア集”, 東海大学出版会 (1995)
- (3) 東るみ子: “振り返りシートを用いた学習者理解度の分析”, 教育システム情報学会第41回全国大会, pp.263-264 (2016)
- (4) 樋口耕一: “社会調査のための計量テキスト分析”, ナカニシヤ出版, (2014)