

自由記述からの講義キーワード検出と 主成分分析を用いた学習者の理解度評価

Student Understanding Evaluation from Descriptive Responses based on Principal Component analysis

森田 博人^{*1}, 健山 智子^{*2}, 松本 慎平^{*3}

Hiroto MORITA^{*1}, Tomoko TATEYAMA^{*2}, Shimpei MATSUMOTO^{*3}

^{*1}広島市立大学大学院 情報科学研究科 知能工学専攻

^{*1} Graduate School of Information Science, Graduate School of Intelligent Engineering,
Hiroshima City University Graduate School

^{*2}滋賀大学 データサイエンス教育研究センター

^{*2}The Center for Data Science Education and Research, Shiga University

^{*3}広島工業大学 情報学部 情報コミュニケーション学科

^{*3}Dept. of Information and Communication, Hiroshima Institute of Technology

Email: mq67017@e.hiroshima-cu.ac.jp^{*1}

あらまし: 本研究では、講義における学習者の理解度評価手法の確立を目的として、講義で取得したアンケートから講義の目的キーワードを抽出し、それを用いた情報解析による履修前後の比較について議論する。キーワード単語の頻度は学生の講義理解度に相関があると仮説を立て、その検証を学生の自由記述アンケートから解析した。この解析は学習者に対する理解度評価として可能であるか、について議論する。
キーワード: 講義キーワード検出, 理解度評価, 習熟度解析, 主成分分析

1. はじめに

情報通信技術の進展により、教育現場における学習者の学習ログデータ収集は容易かつ膨大になった。収集されたデータに対して解析・解釈をすることで、学習者の理解度評価や講義改善などの効果を客観的に明示することが期待される。特に、学習者が講義の振り返り機会であるリフレクションデータは、学習者が学んだ内容に関するデータであり、その中でも記述データを用いた学習者個々の理解度評価手法は検討されている。

テキストマイニングによる文章分類による理解度評価⁽¹⁾や学習効果の可視化⁽²⁾など提案されているが、学習者個人に焦点を当てた学習到達度の評価には至っていない。その理由として、収集された膨大な情報のうち、その集約情報間における関係性が明らかになっていないためである。本研究では、この課題解決を目指し、履修前後のテキストマイニングより講義のキーワードと学習者の文章特徴から主成分解析による可視化による理解度評価手法を提案する。

2. 対象講義の概要と講義キーワードの定義

本研究の対象講義概要と講義キーワードの定義を説明し、研究の流れを下記と図1で示す。

- 履修前後の講義終わりに、受講者らに講義内容に沿って、1000文字程度で「10年後、どのような情報システムが社会に必要とされているか」について自由記述する
- 各文章に対して MeCab を用いて名詞を抽出
- 2.のデータからテキスト解析を用いた、学習者の履修前後文章から理解度変化を把握

〈2.1〉対象講義の概要

対象講義は広島工業大学情報学部の 2019 年度 3 回生前期の「組織活動と情報システム」で、講義の概要は以下の通りである。

- 情報技術進展による生活や社会の変化
- 企業、自治体の組織内外での情報のあり方
- 今後想定される技術、備えるべきリスク
- AI 社会との関わり

上記の内容をもとに講義展開が行われた⁽³⁾。

〈2.2〉講義キーワードの定義

学生文章から講義の理解度を評価するため、文章内のキーワードがどれだけ講義内のキーワードと一致しているか確認を行う。そこで、本研究では講義キーワードを定義し、これを用いた。講義キーワードの定義と、本講義での講義キーワードは以下である。

- ◆ 講義キーワードの定義
講義を通して教員が受講者らに、学習、理解して欲しいキーワード群
- ◆ 本講義での講義キーワード

人, もの, 金, 企業, IoT, 分析, 可視化, 情報化, セキュリティ, ビッグデータ, データサイエンス, など (16 キーワード)

また、本研究では受講者らの自由記述において講義キーワードを含めるなどの指示は行わない。これにより外部からの意図的な記述ではなく、学習者自身からの理解に特化した文章取得につながる。

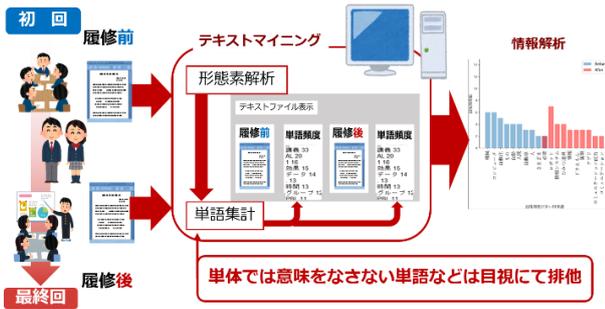


図1：本研究の流れ

3. テキスト解析と主成分分析による解釈

テキスト解析では、履修前後の全文章に対して、形態素解析から名詞抽出し、それらの単語はスコア化として、TF-IDF (式(1)) による数値化をした。

$$P_{t_i} = TF - IDF_{i,j} = TF_{i,j}IDF_{i,j}$$

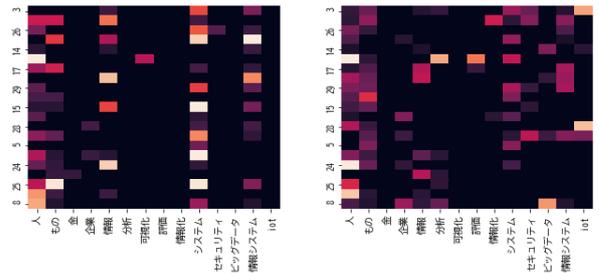
$$= \frac{n_{i,j}}{\sum_k n_{k,j}} \times \left(\log \frac{|D|}{\{d: \ni t_i\}} + 1 \right) \quad (1)$$

$n_{i,j}$: i 番目の文章における j 番目の単語頻度

D : 総文書数

$\{d: \ni t_i\}$: 任意単語 t_i を含む文章数。

本研究では、対象講義で履修前後両方を記述した受講者 21 人の計 42 篇文章から、テキスト解析を行った。式(1)をもとに各受講者の履修前後 2 文章内の講義キーワードの P_{t_i} 値の変化を図2で示す。縦軸は、受講者の期末試験点数の結果をもとに昇順で学生番号を表示し、横軸は文章全体で出現した講義キーワードあり、色を P_{t_i} 値をもとに設定した。図2から、履修前全体で {システム, 人, 情報システム} が使用されていた。履修後の講義キーワードでは、専門単語の出現と受講者個々では {分析, 評価} などの関連性が想定される 2 単語が出現した。この結果をもとに、先述の 42 篇文章の各講義キーワードの P_{t_i} 値をもとに主成分分析 (以降: PCA) を行った。しかし、PCA の第 2 主成分までの累積寄与率は 80% に到達しておらず、結果として文章の大まかな特徴抽出のみに留まったと考えられる。そこで、各講義キーワードに加えて受講者変数を加えた Biplot を図3に示す。図3から、個々の受講者、および全体での講義キーワードに対する文章傾向の変化を捉え、履修前に比べ履修後が専門的な講義キーワードの変位から講義に対する理解度の向上が示唆された。また、講義の担当教員が期末試験結果、対面を通じて講義をよく理解している学生 3 人と、そうではないと思われる学生 3 人に注目する。各学生の履修前後の文章を同様に式(1)よりスコア化し、履修後上位 10 単語を学生間で比較すると、講義キーワードとそれに関連する単語が出現し、講義キーワードの種類、数値の増加傾向が確認された。以上より、テキスト解析から講義理解度の評価がの可能性が示唆された。



(a)履修前での TF-IDF (b)履修後での TF-IDF

図2：履修前後の講義キーワード TF-IDF

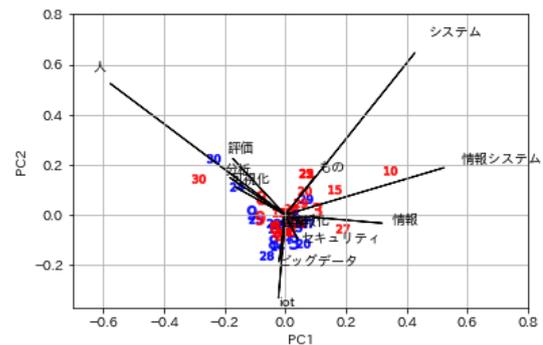


図3：学生 6 人に注目した履修前後 Biplot

4. おわりに

本研究では、受講者の理解度評価手法として、履修前後の自由記述に対してテキストマイニングを用いた講義キーワード検出、TF-IDF の数値化をし、主成分分析から、受講者と講義キーワード間の関係性を Biplot での可視化と文章解釈をした。結果、履修前後の文章内での講義キーワードの変化、教員の受講者評価上位下位と文章特徴傾向が一致すると示唆された。以上の結果から、履修前後の自由記述における文章キーワードと講義キーワードとの関係性から理解度確認の効果的な手法であると示した。

今後、他科目でも同様の解析から本研究、手法の妥当性を検証する。

謝辞

本研究は、広島工業大学 HIT 教育機構次世代 ICT 教育情報可視化研究開発センター助成の下、研究を遂行した。

参考文献

- (1) 東 るみ子：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，第 42 回教育システム情報学会全国大会 (II-08)，北九州国際会議場，2017
- (2) 二瓶，他：アクティブラーニングにおける学習効果の可視化と教育改善への取り組み，平成 26 年度 ICT 活用による教育改善研究発表会，B - 11，ICT 活用教育方法研究第 17 巻第 1 号，2014
- (3) 健山，他：効果的な PBL 型アクティブ・ラーニング実施のための教育 ICT 活用とデータ解析，第 44 回教育システム情報学会全国大会，P1-34，静岡大学，2019