

トピックモデルによる大学生の運動実践に関する研究(1) 模擬授業に対する省察分析

堀井大輔*, 金田啓稔**

Evaluation of Movement Practice on College Students using Topic Model (1) Analysis of Self-Reflections about Trial Teaching in Physical Education

Daisuke HORII*, Hiratoshi KANEDA**

キーワード：トピックモデル，運動実践，模擬授業，自己省察

Abstract

The aim of the present study was to identify patterns in the self-reflections of college students about movement practice using topic model. We analyzed college students' texts which were written as self-reflections of trial teaching in physical education. As a result, we found 6 topics from the text, it was found that six characteristics can potentially classify the terms of self-reflections. Topic 1 comprised words like “good” and “looking.” Similarly, Topic 2 was composed of “small game” and “final.” Topic 3 consisted of “originality” and “content.” Topic 4 consisted of “difficult” and “sports.” Topic 5 consisted of “rain” and “number of people,” and Topic 6 consisted of “ball” and “first.” The method we used in this analysis would be a useful way to review an analysis of self-reflections about trial teaching in physical education.

1. はじめに

大学生による模擬授業における運動実践の自己省察について、テキストマイニング手法を用いて分析・検討を行った報告^[1]では次のようにまとめられている。1) 単語の出現頻度に基づく手法で分析した結果、「生徒」「できる」「良い」「もっと」「指導案」「やる」「感じる」「今後」「考える」「時間」「楽しい」などの40個のカテゴリが生成され、授業担当者は授業のマネジメントや学習者の主体性に目を向けることができていた。2) 生成された各カテゴリ間の対応関係から、授業担当者は学習者の特徴が把握できており、彼らの反応を感じ取った省察が行われていた。3) 省察の変化について、当初は「反省点」「だめ」という授業全体を否定するような表現がみられたが、後半には「やる」「実際」「意識」「展開」「時間」という視点の広がりや具体的な内容への深まりが示された。

これらの結果は、教師を目指す学生の身につけるべき能力が高まっていることを示すものであり、学生の自己省察という質的なデータにテキストマイニングの手法を用いることで、授業を通じた学生の内的な成長を客観的に確認することが可能であった。しかし、ここで用いられたテキ

* 大阪電気通信大学医療福祉工学部健康スポーツ科学科

** 大阪電気通信大学工学部人間科学研究センター

ストマイニングの手法は、自己省察の自由記述データに対して単語の出現頻度や単語間の共起頻度をデータ化したものを分析・検討したものであり、少数の単語からなる周辺的な話題は比較的よく分離できるが、多くの単語が複雑に共起する中心的な話題は大きなかたまりとなって内容を把握するのが難しいという問題を抱えていた^[2]。

その解決策のひとつは、文書中の大量のデータの中に隠れている共通のパターンを発見するための統計解析モデルであるトピックモデルを用いることである。この手法では、文書の中には潜在的にいくつかのトピックが存在し、表出している単語の組み合わせはそれらトピックの基で作られることを考慮するため、単語分布が類似した文書を集めることによって、あるトピックに関して書かれた文書群を見つけることができる。さらに、共起しやすい単語をモデル化することができ、潜在的な共起性を推定することができるため、データ表層上に現れていないが、共起する可能性がある単語も推定することができる^[3]^[4]。このようにトピックモデルを用いれば、文書集合全体の中にどのような単語の集合から構成されるトピックが含まれるのかという概要把握が可能になるのに加え、各文書がどのようなトピックから構成されているかということについても推定が可能となる。そのうえトピックモデルでは、潜在的な単語間の関係も考慮されるため、従来の手法では難しかった、文脈によって同じ単語が違う意味を表すケースなどにも対応が可能であり、高い分類精度が期待できる^[5]。

したがって本稿では、大学3年生後期に履修される保健体育科教育法の模擬授業における運動実践後に、自己省察として提出された自由記述形式の回答について、トピックモデルを用いて分析することを試み、今後の運動実践・授業研究に有用な資料を得ることを目的とする。

2. 方法

2.1 分析対象

A大学で中学校・高等学校保健体育科の教員免許の取得を目指し、2017年度～2018年度に保健体育科教育法4の授業を履修した大学生が対象となる。本授業の受講生は2017年度が22名、2018年度が17名であり、全員が翌年度に教育実習を控えている。彼らが自己省察として提出した自由記述データのうち、有効データとして2017年度の延べ282名分と2018年度の延べ346名分の合計627名分のテキストデータを分析対象とした。

2.2 分析対象の運動実践

本稿における大学生の運動実践については、以下に示すとおりA大学で開講されている授業において大学生が教師役を演じる模擬授業のうちの身体活動を取り上げることとした。

A大学で開講されている授業のうち保健体育科関連の教職科目「保健体育科教育法」は、中学校および高等学校保健体育科の教員免許取得を目指す学生が履修しなければならない科目である。

2年生前期で配当される「保健体育科教育法1」では、「保健体育とは何か」や「授業を通して何を教えるか」などの保健体育科教育についての基本的内容について理解を深め、実際の授業における教授技術・指導過程の一般原則を身につけ、保健体育科教育の理念に基づいた授業展開を実践できる能力を身につけることを目標としている。2年生後期に配当される「保健体育科教育法2」では、保健体育科教育法1で学習した「目標」や「内容」を指導計画・年間計画・単元計画を踏まえて、授業計画を作成できる「方法作り」についての理解を深めることを目的とす

る。また、様々な教育方法・学習形態・授業実践に伴う課題や問題点を学び、自己における授業展開だけでなく授業分析や評価ができ、授業改善する能力を身につけることを目標としている。

3年生前期に配当される「保健体育科教育法3」は、おもに保健分野の模擬授業を行い、保健体育の授業の目標・学習内容・学習づくり、教員に求められる職務・資質など、教育内容や指導法に関する全般的事項について理解を深めることを目標としている。3年生後期に実施される「保健体育科教育法4」は、これまで学習した内容を踏まえ、保健体育の授業に関する教材研究や授業研究を通して、体育・スポーツの特性を明確にしながら指導案を作成し、模擬授業をすることで教育的実践力の資質を向上させ、教育現場の問題や課題に即した実践的な指導法を身につけることを目的にしている。この授業は本稿で対象としている授業であり、保健体育科教員に求められる実務的知識・技術について総合的な理解を深め、実践的な教科の指導法の習得を目指す教育実習前の重要な科目として位置づけられている。

2.3 模擬授業としての運動実践の展開

保健体育科教育法4の授業展開は、初回の授業時に受講生たちで話し合いを行い、3回目から実施される模擬授業の担当者（教師役）として、自分自身が専門とする種目以外の運動種目を決定して振り分ける。2回目の授業時には、担当する運動種目の教材研究や指導案の作成などの準備、先輩の授業をビデオで鑑賞する機会を設ける。その後は模擬授業を中心に合計12回の授業が実施される。最終回には、学習指導案や実技の指導方法などを包括的に確認するまとめの討議が行われる。

とくに3回目からの授業の進め方は、以下の1)～4)の通りであり、準備・導入から展開・まとめに至るまで教師役の学生に一任される。受講生は最低1回以上の教師役を経験し、その他の受講生が生徒役を演じることになる。

1) 事前準備：授業を担当する学生は、学習指導要領や実技の専門書、国立教育政策研究所の例を参照して、単元計画と授業指導案を作成する。その指導案にもとづき、事前に他の教員と打ち合わせを行い、実技の実習場所を確保する。屋内外の実習施設の大きさなどの環境条件、用具の種類・個数などの過不足について、十分に情報収集して確認を行う。

2) 模擬授業の実践：1回の模擬授業は50分とし、実習場所の設営や用具類の準備などの時間を含めるかどうかは教師役の学生に一任される。記録用にデジタルビデオカメラで授業風景を撮影し、模擬授業終了後の授業検討会の際に適宜視聴する。

3) 授業検討会：模擬授業終了後、数名のグループに分かれてディスカッションし、話し合った内容を報告し合ったり、担当教員からコメントをもらったりする形式で行う。

4) 自己省察：授業検討会終了後、他の受講生や担当教員からの評価・助言・批判的指摘などを誠実に受け止め、自分自身の授業の実態を確認し、自覚したことをまとめてmoodle (Modular Object-Oriented Dynamic Learning Environment)^[6]へ入力する。入力については、とくに制約を設けず、学習課題や環境、教師としての行動、学習者の反応などを自由に記述する。

なお、模擬授業における運動実践種目は、順不同で「体づくり運動」「器械運動」「バスケットボール」「バレーボール」「ソフトボール」「テニス」「サッカー」「バドミントン」「卓球」「柔道」「ダンス」であり、天候などによっても適宜場所や実施日の入れ替えが行われる。

2.4 トピックモデルを用いた分析方法

トピックモデルは、文書中に複数の潜在的なトピックがあることを仮定し、出現した単語の確率に基づいて生成するモデルである。本稿においては、言語処理における文書を「自己省察」、単語を「名詞・形容詞・動詞」に置き換えた頻度行列に対してトピックモデルを用いることで、自己省察ごとの各トピックに属する確率（自己省察ごとのトピック分布）と、単語ごとの各トピックにおける出現確率（トピックごとの単語分布）を推定した。自己省察ごとのトピック分布を生成することにより潜在変数としてのトピックパターンが明らかになる。また、トピックごとの単語分布を生成することにより各トピックが持つ特徴が明らかになる。

手順としては、毎回の授業終了後に受講生がmoodleへ入力する自己省察（自由記述）に対し、テキストデータとしての前処理後にトピックモデルを用いて分析・検討する。

受講生の自己省察は、定型化されていないデータであるので、データ解析・マイニングの方法で分類し、テキストを構成する要素を定量化することが必要となる。テキストデータを定量化する最も簡単な方法は、テキストに現れる文字の頻度を集計することである。しかし、文字を単位としたn-gram（検索対象を単語単位ではなく文字単位で分解して行う索引文字列の抽出手法）データには言語学的に解釈できないパターンが数多く含まれる。そこで、言語学的解釈を可能とするため、文法理論に基づいて形態素解析、あるいは構文解析を行い、その単位に基づいてテキストデータを定量化することが必要となる。形態素解析とは、文を単語毎に分割し、単語の品詞属性などを付けることである。

得られた自己省察（自由記述）を以下の手順に従って分析する。

1) テキストデータについて、MeCab Ver0.996^[7]を用いた形態素解析を行う。分析には表層語ではなく基本形を用い、助詞・助動詞・記号を除く名詞・形容詞・動詞を中心にまとめた。

2) 上記の結果をデータベースとして保存し、統計分析ソフトR（64ビット版）3.6.1^[8]を用いたトピックモデル分析^[9]を行う。トピックモデル分析パッケージには、トピック分布にディリクレ分布を仮定し、ギブスサンプリングによるベイズ推定を行う、潜在的ディリクレ配分法（Latent Dirichlet Allocation, 以下LDA）^[10]を採用する。LDAは、文書内の各単語に潜在変数（トピック）を、文書ごとにトピックの出現確率分布をそれぞれ仮定し、文書における単語の生成過程をモデル化した代表的なトピックモデルである。なお、単語文章行列を作成する場合、入力データとなる文章それぞれの長さ（文章量）が大きく異なっていると、続く解析に影響を与えることがあるため、長さの影響を調整する正規化を行う必要がある。本稿のように文章からそれぞれの特徴を見つけ出そうという課題の場合、すべての文章に出現する単語と、一部の文章にしか出現しない単語を区別する必要があるためにTF（term frequency）-IDF（inverse document frequency）を用いることとした。この手法は、自然言語処理やテキストマイニングではよく知られた重み付けである^[11]。

図1にグラフィカルモデルを示す^[4]。

(1) 文書ごとに、事前分布のディリクレ分布（ α ）に従い、トピックの出現確率分布（ θ ）を生成する。

(2) トピックごとに、事前分布のディリクレ分布（ β ）に従い、単語出現確率分布（ ϕ ）を生成する。

(3) 文書内の単語ごとに、(1)で生成したトピックの出現確率分布（ θ ）に従い、トピック（ Z ）を生成する。

(4) 文書内の単語ごとに、(2) で生成したトピックの単語出現確率 (ϕ) に従い、単語 (W) を生成する。上記 (3) と (4) を、全文書・全単語に関して行い、文書における単語の生成過程をモデル化している。

図 2 には分析のイメージ図を示す。

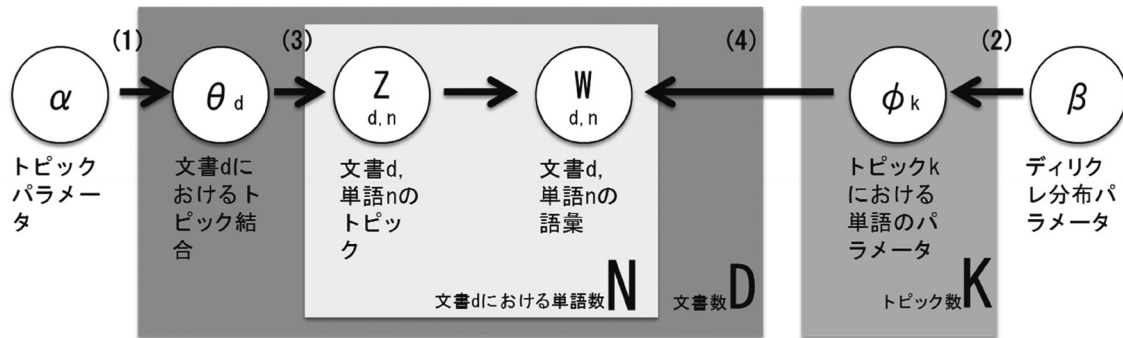


図 1 LDA のグラフィカルモデル

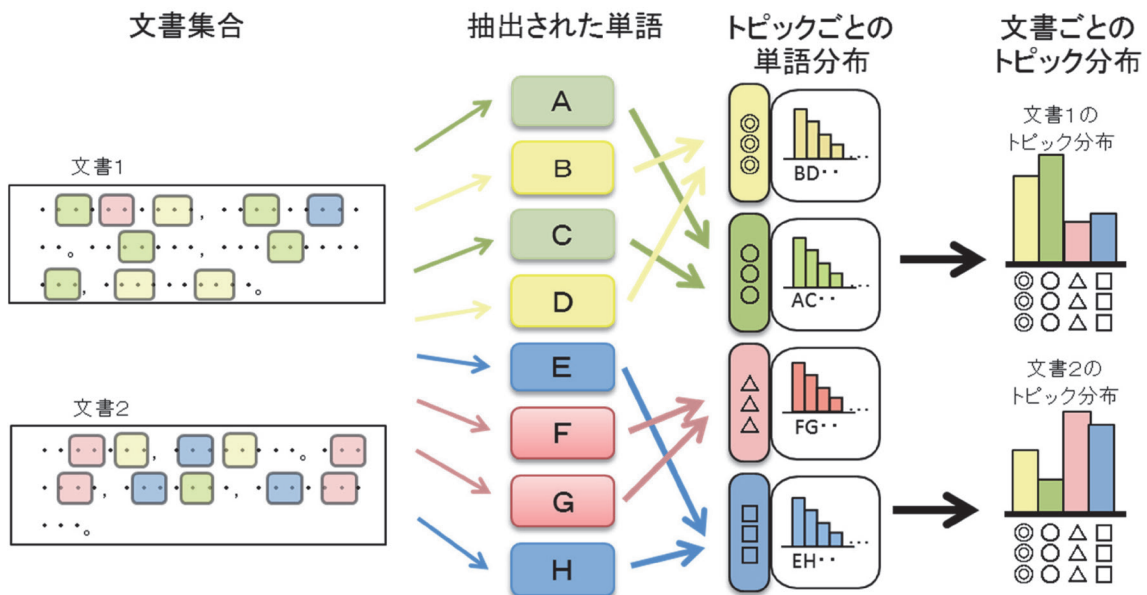


図 2 文書集合から文書ごとのトピック分布作成の流れ

3. 結果と考察

3.1 トピック数の決定

言語処理における文書を「自己省察」、単語を「名詞・形容詞・動詞」に置き換えた頻度行列に対し、トピックモデルLDAを用いて、それぞれのトピック数、文章-トピックの割合、トピックの意味解釈、トピックに出現する確率の高い単語に着目して検討した。結果、単語ごとの各トピックにおける出現確率（トピックごとの単語分布）（図 3）と、自己省察ごとの各トピックに属する確率（自己省察ごとのトピック分布）を推定した。

トピック数の決定は、各トピックにおける単語の分布（図 3）と各トピックの γ 値、Perplexityの値を参考にした。一般に、文書内に特定のトピックに分類される単語が多ければ多

いほど、文書-トピック分類の重みである γ 値は大きくなる^[12]。本稿では6つのトピックが妥当と判断し、その後の分析を行った。

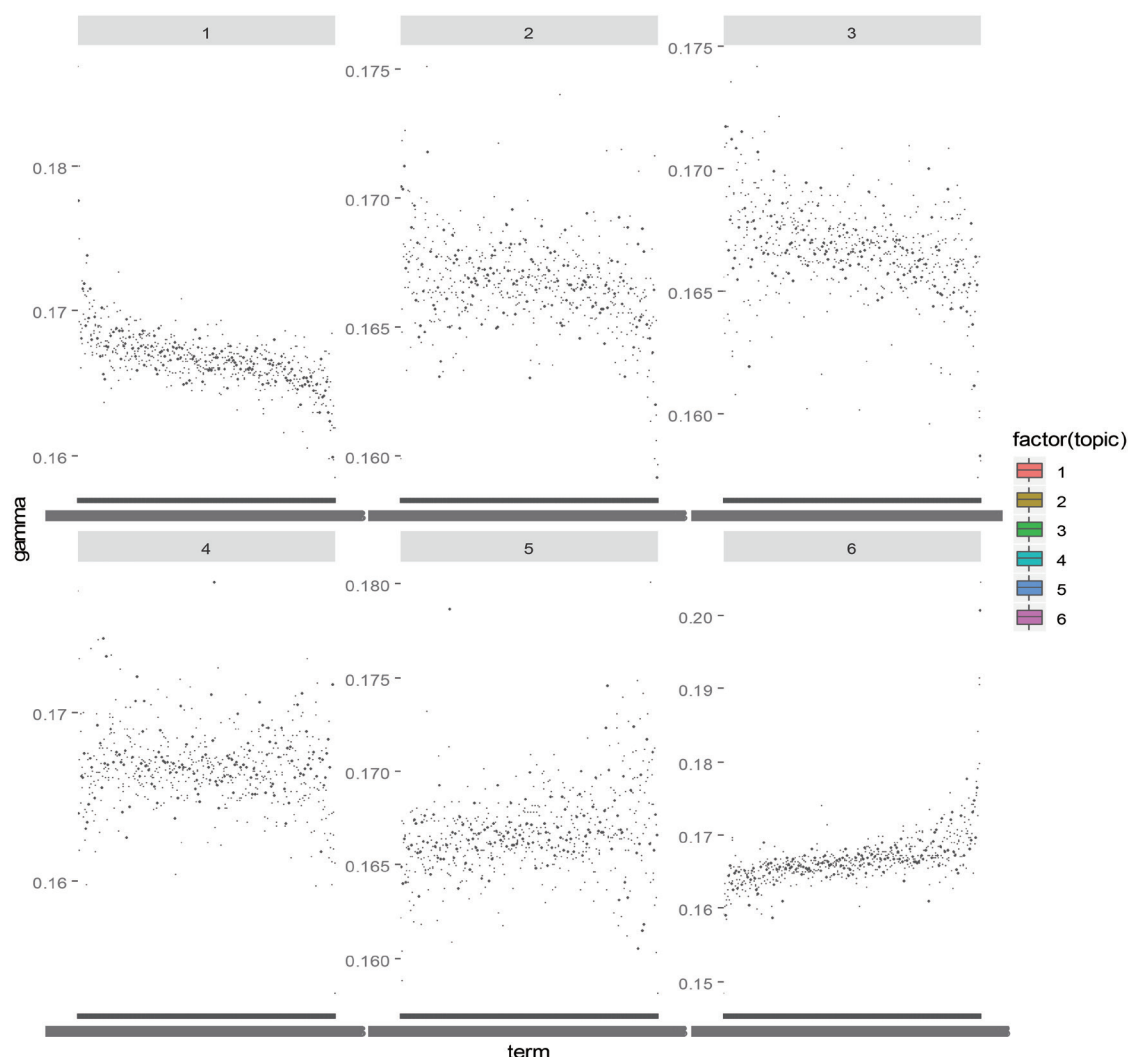


図3 各トピック（1～6）における単語の分布（ γ 値）

3.2 抽出されたトピックとラベル

LDA によるトピック抽出の結果、同じ種目に関する省察が、同じトピックの上位に入っているとは限らないことから、文章の意味内容によって解析されたトピック分類が行われていることが確認された。各トピックにおける出現確率の高い上位20の単語は表1の通りとなった。

トピック1は、「行う」「生徒」「する」「良い」「出来る」「思う」「なる」「行く」「感じる」「多い」「使う」「ある」「分かる」「考える」「ゲーム」「楽しい」「うまい」「ルール」「見る」「グループ」。

トピック2は、「思う」「行う」「する」「良い」「ある」「出来る」「なる」「感じる」「人」「行く」「教具」「分かる」「ゲーム」「多い」「先生」「ミニゲーム」「ルール」「ない」「見る」「最後」。

トピック3は、「思う」「行う」「良い」「生徒」「なる」「考える」「出来る」「する」「行く」「教具」「楽しい」「人」「分かる」「声」「使う」「オリジナリティ」「ない」「ある」「内容」「多い」。

トピック4は、「思う」「行う」「良い」「出来る」「ある」「行く」「人」「する」「生徒」「分かる」「ない」「難しい」「ルール」「スポーツ」「なる」「コート」「少ない」「自分」「取り入れる」「最後」。

表1 各トピック内における単語出現確率 (β) の高い上位 20 単語

1	2	3	4	5	6
"行う"	"思う"	"思う"	"思う"	"思う"	"する"
"生徒"	"行う"	"行う"	"行う"	"行う"	"良い"
"する"	"する"	"良い"	"良い"	"ある"	"行う"
"良い"	"良い"	"生徒"	"出来る"	"良い"	"ある"
"出来る"	"ある"	"なる"	"ある"	"する"	"ボール"
"思う"	"出来る"	"考える"	"行く"	"生徒"	"ゲーム"
"なる"	"なる"	"出来る"	"人"	"考える"	"なる"
"行く"	"感じる"	"する"	"する"	"多い"	"出来る"
"感じる"	"人"	"行く"	"生徒"	"ない"	"思う"
"多い"	"行く"	"教具"	"分かる"	"使う"	"考える"
"使う"	"教具"	"楽しい"	"ない"	"雨"	"人数"
"ある"	"分かる"	"人"	"難しい"	"体"	"楽しい"
"分かる"	"ゲーム"	"分かる"	"ルール"	"声"	"入る"
"考える"	"多い"	"声"	"スポーツ"	"入れる"	"最初"
"ゲーム"	"先生"	"使う"	"なる"	"人数"	"声"
"楽しい"	"ミニゲーム"	オリジナリティ	"コート"	"内容"	"持つ"
"うまい"	"ルール"	"ない"	"少ない"	"教具"	"自分"
"ルール"	"ない"	"ある"	"自分"	"ルール"	"生徒"
"見る"	"見る"	"内容"	"取り入れる"	"先生"	"少ない"
"グループ"	"最後"	"多い"	"最後"	"やる"	"走る"

トピック 5 は、「思う」「行う」「ある」「良い」「する」「生徒」「考える」「多い」「ない」「使う」「雨」「体」「声」「入れる」「人数」「内容」「教具」「ルール」「先生」「やる」。

トピック 6 は、「する」「良い」「行う」「ある」「ボール」「ゲーム」「なる」「出来る」「思う」「考える」「人数」「楽しい」「入る」「最初」「声」「持つ」「自分」「生徒」「少ない」「走る」。

それぞれのトピックに共通性の高い単語を降順に並べると表 2 のようになり、「行う」はトピック 1 に 20.4%，トピック 2 に 10.1%，トピック 3 に 12.6%，トピック 4 に 6.6%，トピック 5 に 8.8%，トピック 6 に 3.1% の確率で出現することが確認された。同様に、「思う」「する」「良い」「ある」も各トピックの上位 20 単語に出現しており、「出来る」「生徒」「なる」「考える」「行く」「分かる」「多い」「ない」などの単語もほとんどのトピックで上位に出現していた。つまり、この分析対象となった文章自体が、模擬授業の自己省察という観点で作成されていることから、「生徒」の立場から「出来る」か「出来ない」か、「分かる」か「分からない」か、「多い」か「多くない」かなどをどうすれば「良い」と「思う」か「考える」という振り返りが行われたと考えられる。

それぞれのトピックに特徴的な単語としては、トピック 1 の「うまい」「見る」「グループ」などであり、うまい生徒を見て学習するような場面を設定したグループ活動に関するトピックと考えられる。トピック 2 は「ミニゲーム」「最後」などであり、授業の最後に行うゲームやミニゲームをどのようなルールで実施するかに関するトピックと考えられる。トピック 3 は「オリジナリティ」「内容」などであり、教具を使いながらオリジナルな内容を検討しているトピックと考えられる。トピック 4 は「難しい」「スポーツ」「コート」などであり、スポーツ実施の際のルールやコートの難しさに関するトピックと考えられる。トピック 5 は「雨」「人数」などであり、天候に左右される実践種目の問題についてのトピックと考えられる。トピック 6 は「ボール」「最初」「走る」などであり、球技系の種目に関するトピックと考えられる。

3.3 各トピックを代表する省察

複数のトピックが存在する文章において、それぞれのトピックの占める割合が高い文章を一部抜粋すると、トピック 1 では「グループ内で発表を行うのも良いが、グループごとにうまい人を見てみんなで評価すると良いと思いました。課題をレベルごとに分けて全員に課題クリアのチャンスを与えてあげると良いと思いました。（ $\gamma=18.7\%$ ）」であった。同様に、トピック 2 では「ミニゲーム 1 ではルールがあまり理解できず、だらけている印象があった。ミニゲーム 2 では範囲を指定したりして飛びすぎないように行うことで経験者との差がなくなると思う。（ $\gamma=17.4\%$ ）」。トピック 3 では「授業の構成にオリジナリティや工夫が多く含まれていて良かったと感じた。アイスブレイクなど他の人が普段行わないようなアップを設けていたり、授業でもみんなが笑いながら笑顔でやれていたのは素晴らしいことだと思った。（ $\gamma=17.1\%$ ）」。トピック 4 では「e スポーツを取り入れる事はすごく難しかったと思いますが、まだまだ反省点を残しながらも授業としてなんとか成り立たせていた点では良かったと思いました！しかし、やはり運動量の不足や生徒の集中力の欠如の問題、ニュースポーツとの繋がりの少なさが反省点として挙げられると思いました。（ $\gamma=17.4\%$ ）」。トピック 5 では「体育館では雨の日の場合さまざまな学年が合同で体育館を使う場合があるが、今回先生役の声が通っていないと思った。また、先生の立ち位置もすごく重要であり、生徒は他方ばかり気にしていて先生の話がしっかり通っていなかったのでもそこを改善するべきだと思った。（ $\gamma=17.9\%$ ）」。トピック 6 では「ゴロ捕球の時の人数配置が少なく慌ただしかった。ボールの置き場所が危なかった。カバーをつけるか、つけないか統一

すると良かった。チームで考える時間が良かった。先生が入ってくれたけど入った位置がボールのよく回る位置だったので他のところに入った方が良かった。ボール回しの時に女子にはワンバウンドとかにすると良かった。ゲームの勝敗の有無があったら良かった。ゲームをゴロスタートで行ったら良かった。(γ=19.1%)」。

このような自己省察のトピックパターンの特徴として、図4のように気づきの内容を示す単語のトピック構成率に差異が存在することや、表2に示す単語の割合推定によって単語から構成される文章が複数のトピックを含んでいることが確認された。また、各トピック内にどのような重要な単語が多いかについて、事後確率としての数値を算出し、パターン分類が困難であった数多くの組み合わせを客観的に区分できたことにより、それらの特徴や関連性の把握が容易になったと考えられる。

省察とは、ある授業について良かった点と悪かった点を列挙することや、生徒や授業者をほめたり非難したりするような評価をする営みではなく、そこで起きていた事実の意味を生成するプロセスである。そこには自分の意思で能動的に行う思考が求められ、自分自身が直面している問題を問題として認識し、「何を」「何のために」「どのように」していくべきかを、自分自身の経験について思考することによって学んでいく過程だと考えられる。授業中や授業後の討論で終わるのではなく、その後もさらに調査研究を行って、事実関係を探究していく作業を必要とする場合も多いことから、省察は即興的に行われるものから、数時間、数日、時には数年にわたって続く、長期的で探究的な営みも含んでいるのである^[13]。

これらの結果が省察の観点の把握だけでなく、種目別の教材・教具の設定、授業の展開の計画立案、授業全体のマネジメントの検討に際して有用な資料になるだけでなく、学生間の協同的な学習や複眼的な振り返りと学生の成長に寄与することを期待したい。

表2 全文書におけるトピック別の出現単語確率 (β) の高い上位単語

	1	2	3	4	5	6	M	SD
行う	0.204	0.101	0.126	0.066	0.088	0.031	0.103	0.059
思う	0.030	0.118	0.132	0.153	0.154	0.016	0.100	0.062
する	0.050	0.096	0.024	0.020	0.026	0.119	0.056	0.042
良い	0.047	0.042	0.041	0.055	0.028	0.098	0.052	0.024
出来る	0.038	0.028	0.026	0.048	0.002	0.018	0.027	0.016
生徒	0.066	0.002	0.040	0.020	0.024	0.008	0.026	0.023
ある	0.017	0.029	0.009	0.035	0.038	0.029	0.026	0.011
なる	0.029	0.026	0.039	0.011	0.003	0.024	0.022	0.013
考える	0.014	0.007	0.030	0.000	0.024	0.015	0.015	0.011
行く	0.020	0.017	0.020	0.024	0.002	0.003	0.014	0.009
分かる	0.015	0.016	0.014	0.019	0.007	0.001	0.012	0.007
多い	0.018	0.013	0.009	0.004	0.021	0.004	0.012	0.007
ない	0.005	0.010	0.009	0.016	0.020	0.003	0.011	0.007
使う	0.017	0.005	0.013	0.002	0.019	0.006	0.011	0.007
人	0.003	0.018	0.014	0.023	0.000	0.001	0.010	0.010
ゲーム	0.013	0.015	0.001	0.002	0.002	0.026	0.010	0.010
楽しい	0.012	0.003	0.015	0.006	0.004	0.014	0.009	0.005
教具	0.001	0.016	0.015	0.003	0.010	0.004	0.008	0.006
声	0.003	0.005	0.014	0.002	0.016	0.010	0.008	0.006
ルール	0.010	0.010	0.006	0.011	0.009	0.001	0.008	0.004
感じる	0.018	0.019	0.001	0.000	0.005	0.004	0.008	0.008
ボール	0.000	0.002	0.005	0.003	0.005	0.027	0.007	0.010
先生	0.006	0.013	0.007	0.004	0.009	0.002	0.007	0.004
人数	0.005	0.004	0.003	0.004	0.010	0.015	0.007	0.005
少ない	0.008	0.007	0.002	0.009	0.004	0.008	0.006	0.003
自分	0.002	0.007	0.003	0.009	0.006	0.009	0.006	0.003
内容	0.001	0.006	0.009	0.005	0.010	0.003	0.006	0.003
難しい	0.000	0.006	0.002	0.016	0.005	0.001	0.005	0.006
最後	0.003	0.009	0.006	0.008	0.002	0.002	0.005	0.003
見る	0.010	0.009	0.001	0.007	0.001	0.001	0.005	0.004
うまい	0.011	0.000	0.002	0.006	0.007	0.002	0.005	0.004
コート	0.004	0.002	0.007	0.010	0.002	0.001	0.004	0.003
入れる	0.001	0.001	0.002	0.005	0.010	0.007	0.004	0.004
持つ	0.003	0.007	0.003	0.003	0.001	0.009	0.004	0.003
体	0.004	0.001	0.000	0.001	0.016	0.003	0.004	0.006
オリジナリティ	0.001	0.002	0.009	0.003	0.005	0.003	0.004	0.003
雨	0.001	0.000	0.001	0.001	0.016	0.004	0.004	0.006
ミニゲーム	0.000	0.011	0.003	0.004	0.004	0.000	0.004	0.004
スポーツ	0.004	0.001	0.002	0.011	0.000	0.002	0.003	0.004
やる	0.001	0.001	0.000	0.007	0.009	0.003	0.003	0.004
グループ	0.009	0.001	0.004	0.003	0.002	0.002	0.003	0.003
入る	0.001	0.001	0.001	0.001	0.002	0.013	0.003	0.005
最初	0.003	0.001	0.000	0.001	0.001	0.011	0.003	0.004
取り入れる	0.002	0.001	0.001	0.008	0.000	0.003	0.003	0.003
走る	0.001	0.000	0.000	0.000	0.002	0.008	0.002	0.003

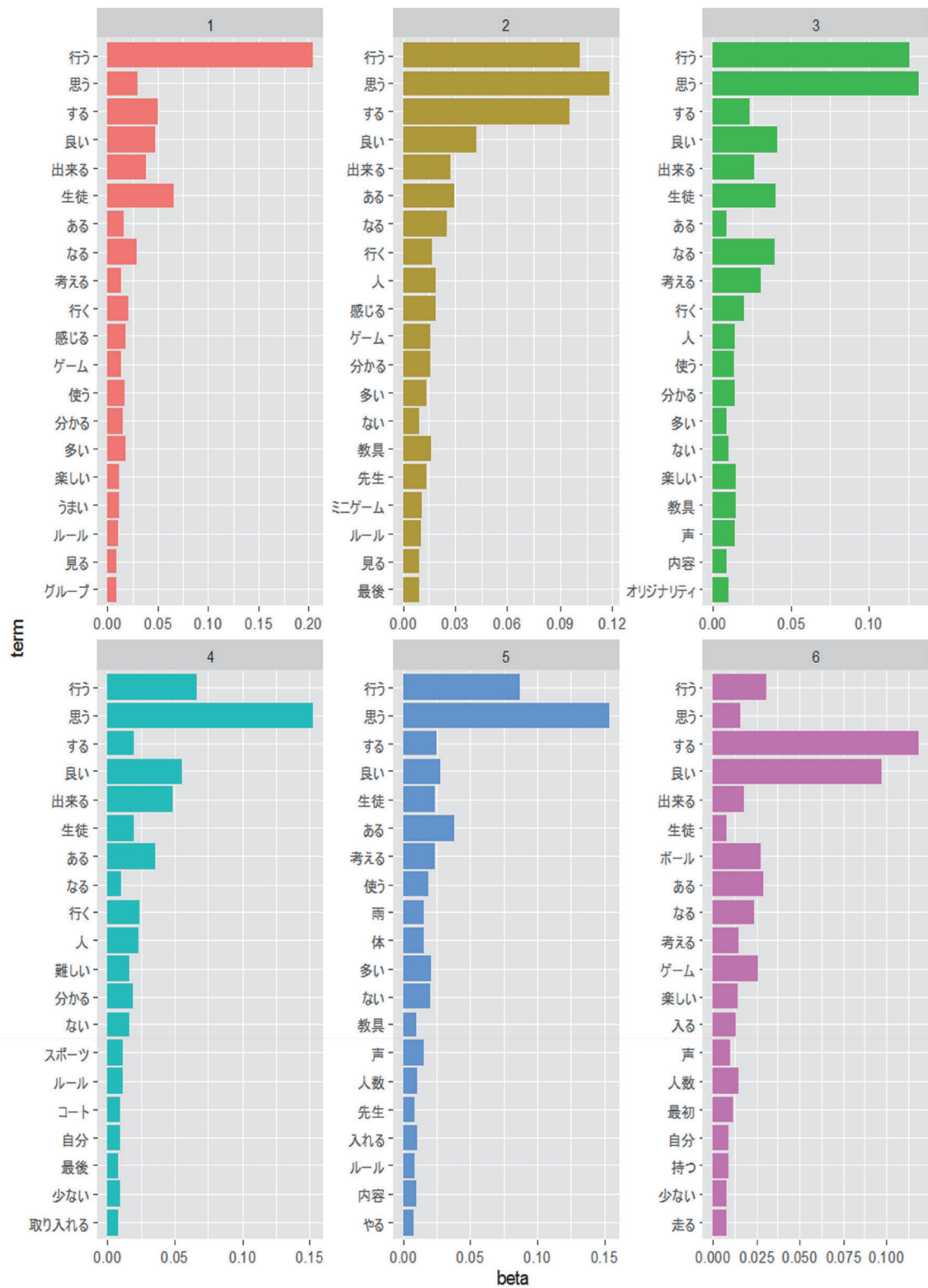


図4 トピック別の全文書における出現確率の高い上位20単語

4. まとめ

本稿では、保健体育科教員免許の取得希望学生が受講する教科教育法における模擬授業の運動実践内容を報告するとともに、学生の自己省察（自由記述形式）についてトピックモデルを用いて分析、検討を行った。トピックモデルには、自己省察に含まれる出現単語の単純な頻度のみで推定するのではなく、ディリクレ分布を前提とした生成モデルで意味割合の低い単語の表出を抑制できること、さらに個々人の自己省察に異なる複数トピックを考慮できることが利点としてあげられる。

学生の自己省察という質的なデータにトピックモデルLDAの手法を用いることで、6つのトピックを抽出することができ、言語としての意味解析の分析結果を示すことが可能となった。

一方で、今後の課題として、トピック数の決定に用いた指標の信頼区間を明確にする工夫や、トピックに応じたラベリング作業とその効率化、トピックラベルの妥当性検証は必要だと考えられる。さらに、トピック間の経年変化を分析することで学生の内的変容をみることも可能であると考えられる。

5. 引用・参考文献

- [1] 堀井大輔, 奥田援史, “体育指導における模擬授業の効果—テキストマイニングによる自己省察の分析—”, 滋賀大学教育学部附属教育実践センター紀要, 第26巻, (2018).
- [2] Matsukawa, H., Arai, Y., Iwasaki, C., Kinjo, Y. and Hotta, H., “Semantic Classification of Text Messages Using the Concept of Community in Social Network Analysis”, IADIS International Conference e-society 2015, Madeira, Portugal, (2015).
- [3] 岩田具治, “トピックモデル（機械学習プロフェッショナルシリーズ）”, 講談社, 東京, (2015).
- [4] 奥村学, 佐藤一誠, “トピックモデルによる統計的潜在意味解析”, コロナ社（自然言語処理シリーズ）, 東京, (2015).
- [5] 松河秀哉, 大山牧子, 根岸千悠, 新居佳子, 岩千晶, & 堀田博史, “トピックモデルを用いた授業評価アンケートの自由記述の分析”, 日本教育工学会論文誌, 41 (3), (2017).
- [6] 喜多敏博, 穂屋下茂, 大西淑雅, 奥村晴彦, 上木佐季子, 木原寛, 長谷川理, 不破泰, “Moodleの開発体制と日本の大学における管理運用事例”, 教育システム情報学会誌, Vol.32, No.1, (2015).
- [7] 工藤拓, “MeCab: Yet Another Par-of-Speech and Morphological Analyzer”, <https://taku910.github.io/mecab/>, (2013).
- [8] “The R Project for Statistical Computing”, <https://www.r-project.org/>, (2019) .
- [9] Bettina Grün, Kurt Hornik, David M Blei, John D Lafferty, Xuan-Hieu Phan, Makoto Matsumoto, Takuji Nishimura, Shawn Cokus, “Package ‘topicmodels’ : Topic Models”, <https://cran.r-project.org/web/packages/topicmodels/topicmodels.pdf>, (2019).
- [10] Jonathan Chang, “Package ‘lda’ : Collapsed Gibbs Sampling Methods for Topic Models”, <https://cran.r-project.org/web/packages/lda/lda.pdf>, (2015).
- [11] 石田基広, “Rによるテキストマイニング入門”, 森北出版, 東京, (2018).
- [12] 大橋真也, 長尾高弘, “Rによるテキストマイニング”, オライリー・ジャパン, 東京, (2018).
- [13] 岩川直樹, “教育の方法・技術”（教師教育テキストシリーズ）, 学文社, 東京, (2014).