

トピックモデルを用いた自由記述アンケートの分析と活用

— オンライン授業に関する学生アンケートの回答結果から —

西山 慶太（専修大学）

1. 背景

Institutional Research (IR) は高等教育機関において情報の調査および分析を実施する機能または部門であり、大学内外に対して必要な情報を提供することは IR の主要な役割である[1]。2020年現在、COVID-19の影響により多くの高等教育機関で前例のない規模でのオンライン授業が展開されている。この対応は急を要したため、従来の対面授業からオンライン授業に変更することによって、どのような利点や問題が生じるかについて詳細に検討することは難しかったと考えられる。よって、ある種の見切り発車を行わざるを得なかったわけであるが、その後の約半年の間に実施されたオンライン授業について、学生がどのような利点および問題点を感じているか把握する必要があると考えられる。実際、調査実施規模の大小はあるが、多くの大学からオンライン授業に関する報告が集まりつつある[2]。筆者が所属する専修大学においても、オンライン授業の実施状況を把握するために「オンライン授業に関する学生アンケート調査（以下、本アンケート調査）」を実施した。なお、本稿では本アンケート調査の自由記述設問に関する分析事例を報告するが、本アンケート調査に関する概要等については本研究集会において別途報告が予定されているため、そちらを参照していただきたい[3]。

近年、web環境が発展したことで学生調査は従来の調査紙を用いた調査ではなく、webを用いたコンピュータを使用した調査方法にシフトしている。また、自由記述に関して言えば調査方法の変化により、回答が直接的にテキストデータとして取得できる状況になった。実際に本アンケート調査で使用した Google Forms は無償で利用できる調査ツールであり幅広く利用されているが、このツールにおいても選択式と自由記述式の2種類の方法を用いて調査を行うことができる。また、これまで調査紙を用いた調査での自由記述回答は、調査が大規模になるとその入力に多大な労力を要するため、その分析には大きなハードルがあった。加えて、近年の自然言語処理技術や分析環境の発展に伴い、その統計的分析も容易になったと言える。

高等教育分野においても自由記述の情報を定量的に扱うことで、従来の方法では得られなかった知見が集まりつつある。松河らは授業評価アンケートの結果を、トピックモデルを用いて分析し、その有用性について報告している。これによればトピックモデルを用いることで従来の方法に比べて、分析者の認知的負荷や分析時間が軽減され、自由記述の分量にも左右されにくいとされている[4]。

このように自由記述情報を分析することで、高等教育機関においてもより有益な情報を得られる可能性がある。本稿では本アンケート調査において得られた自由記述回答に対しトピックモデルを用いた分析を行うこととした。

2. 方法

本稿で対象とするのは、本アンケート調査によって得られた自由記述設問に関するものである。使用するデータは 2020 年 6 月 29 日から 7 月 7 日までの間に Google Form を用いて得られた回答の内、自由記述での回答を求めた五つの設問の結果である。本アンケート調査では合計 7,307 名からの回答が得られ、その内いずれかの自由記述設問に回答した学生は 4,008 名であった。また、各設問の内容、回答者数、回答率、文字数などの回答状況の概要は表 1 の通りである。この内、Q6A から Q9A は選択式設問と関連付けており、選択式設問回答後に任意で記述を求めている。また、Q10 は独立した設問であり話題を特定しない幅広い回答を期待した設問であり、他の設問とは少し性質が異なる。

表 1 自由記述の設問内容と回答状況

設問	アンケート 回答者数	自由記述 回答者数	自由記述 回答率	合計文字数	自由記述あたりの 平均文字数	トピック数
Q6A オンライン授業のメリットとして上記以外のことがあったら、自由に記入してください	7,307	881	12.1%	29,133	33.1	15
Q7A オンライン授業のデメリットとして上記以外のことがあったら、自由に記入してください	7,307	2,296	31.4%	129,260	56.3	17
Q8A 後期の授業や大学生活について上記以外の考えがあったら、自由に記入してください	7,307	1,186	16.2%	70,480	59.4	8
Q9A 新型コロナウイルス感染症の流行と、それともなうオンライン授業の実施に関連して、上記以外に感じることがあったら、自由に記入してください	7,307	1,120	15.3%	68,602	61.3	11
Q10 オンライン授業全般について、感想、意見、提案などありましたら、自由に記入してください	7,307	2,441	33.4%	307,971	126.2	22

自由記述による調査は、文章として直接的に回答者からの情報を得ることができるが、その量が膨大になると読解・解釈が困難となる問題点がある。本調査では表 1 のように大量のテキストデータが得られたため、自然言語処理の技術であるトピックモデルを用いて自由記述の分析を行うこととした。トピックモデルは自動的に文書の分類ができる自然言語処理技術の一つであり、複数の単語の共起性によって創発される潜在的意味を統計的に分類し、文書の内容を解釈しやすくすることができる。また、本研究では分析環境として R 言語[5]および RStudio[6]を使用した。以下、簡単に分析の手順を示す。

- ① Google Forms から出力された csv 形式の回答データを RStudio に読み込む。
- ② 回答者数や回答文字数など要約統計量を計算する[7]。
- ③ 設問ごとに単語分かち書きし、不要な単語を取り除き、設問ごとのワードセットを作成する[8][9]。
- ④ 基のテキストデータとワードセットを使用しコーパスを作成、Document Term Matrix (DTM) 形式にデータ整形し保存する[10][11]。
- ⑤ DTM 形式で保存したデータを使用してトピックモデルを作成[12]。

なお、本稿において、推定方法は潜在ディリクレ配分法 (LDA)、サンプリングは Gibbs

サンプリングを用いた。トピックモデル作成する際の重要なパラメータであるトピック数 k については `ldatuning` パッケージ[13]を用いて検討するとともに、`tidymodels` パッケージ[14]を用いて確率モデルの精度を表す指標である `preplexity` を求め、最適であると思われるトピック数 k を表1の通りとした。

3. 結果

トピックモデルによって得られた各設問に対する結果を集計・加工したものが表2～表6である。「解釈列」は分析者が `keyword` および自由記述の原文を参考に意味づけを行ったものである。また「件数」列は一つの自由記述ごとに属する確率が最も高いトピックを選定し、その件数の総和を求めた値である。つまり、この総和が大きいトピックほど多くの学生が記述したトピックということになる。「`keyword`」列は、そのトピックを代表する単語上位10位までを選択している。以下、設問ごとの結果について簡単に説明する。

まず、表2の「オンライン授業のメリット」の分析結果として通学時間の削減や時間の有効活用といったトピックが見られた。また、オンライン授業の利点についての具体例や感想を述べているトピックが見られる。続いて、表3の「オンライン授業のデメリット」に関しては、授業の質に関することや課題の多さを指摘する意見、オンライン授業の環境やパソコンを使うことで起こる疲労などのトピックが見られる。これらのオンライン授業のメリット・デメリットに関する結果は、他大学の調査でも確認されており、大学でのオンライン授業に広く当てはまる事項であると考えられる[15]。また、表1の回答状況からもわかるようにデメリットに関する設問への回答者数や記述文字数の方がメリットに関する設問よりも大きい。

表2 Q6A「オンライン授業のメリット」の集計結果

解釈	件数	keyword
通学時間短縮	79	時間,通学,自分,睡眠,省ける,勉強,確保,余裕,学習,より
パソコンへの習熟	78	パソコン,特に,すぐ,調べる,向上,オンライン,ストレス,使用,多く,家
オンライン授業の利点	76	授業,オンライン,時間,先生,特に,学生,受講,就職,体調,ほとんど
オンライン授業の利点 (対面との比較)	75	授業,より,対面,意見,勉強,気,良い,質問,効率,集中
時間の有効活用	74	時間,有効,自由,活用,自分,通学,使える,移動,使う,節約
授業に集中できる	72	授業,時間,集中,受ける,オンライン,対面,内容,理解,出席,用事
周囲への気遣い不要	66	気,授業,いい,講義,必要,周り,人,他,目,トイレ
体調不良時でも受講できる	64	体調,自分,受講,受ける,授業,講義,良い,ペース,生活,不良
通学定期代の節約	62	電車,通学,済む,定期,満員,浮く,交通,食費,移動,声
通学コスト減少	60	交通,かかる,減る,就活,学校,通学,大学,いい,機関,行く
対面よりも理解が進む	59	授業,課題,理解,テスト,内容,勉強,確認,自分,提出,学習
講義資料等の利便性	59	資料,音声,管理,思う,学習,授業,場合,より,講義,教授
コロナウイルス感染リスクの抑制	56	感染,コロナ,リスク,ウイルス,無い,人,必要,心配,少ない,大学
Google Classroomへの評価	48	課題,Google,提出,管理,楽,必要,クラス,ルーム,期限,classroom
空きコマの有効活用	48	時間,授業,空き,コマ,自分,家,移動,作業,家事,学校

表3 Q7A「オンライン授業のデメリット」の集計結果

解釈	件数	keyword
自宅で集中できない	189	集中,家,気,授業,自分,家族,生活,自宅,学習,部屋
教授法に関する意見	187	授業,音声,資料,理解,説明,内容,動画,ファイル,講義,先生
グループワーク等が困難	185	授業,人,グループ,友人,ゼミ,ワーク,難しい,対面,友達,コミュニケーション
授業の質に差がある	182	授業,質,オンライン,先生,教員,差,対面,内容,教授,学生
パソコン特有の疲労	180	目,画面,長時間,疲れる,パソコン,疲れ,見る,低下,視力,時間
課題提出時の不安や改善点	180	提出,課題,不安,先生,期限,Google,クラス,ルーム,方法,多い
印刷に関わる経費等	162	印刷,資料,かかる,コピー,インク,レジュメ,お金,プリント,家,プリンター
質問に対する教員の返信	154	質問,先生,授業,返信,人,教員,場合,答え,教科書,自分
学費,特に施設費の減額要望	151	大学,学費,施設,利用,図書館,授業,学校,参考,学生,設備
授業に対する不満	150	授業,内容,課題,先生,理解,自分,オンライン,中,いい,資料
課題が多く時間が撮られる	150	課題,時間,授業,多く,量,毎回,勉強,復習,毎日,非常
通信環境の問題	150	授業,通信,環境,回線,問題,ネット,生徒,オンライン,先生,場合
課題が多く授業に支障	149	課題,講義,時間,教授,多い,授業,提出,教科,中,内容
成績評価への不安	132	課題,評価,レポート,成績,授業,不安,テスト,提出,単位,どの
PCや通信機器	122	パソコン,Wi,通信,人,授業,環境,Fi,受講,声,使用
meetやzoomなどのリアルタイム配信	108	meet,授業,zoom,場合,先生,人,Google,資料,同じ,受講
PCや通信機器(授業時)	94	時間,授業,かなり,PC,パソコン,フォン,スマート,テスト,感じ,制限

次に、表4「後期授業や大学生活への考え方」の結果であるが、この設問では後期授業や大学生活への要望が記述されていると考えられる。そのトピックとしては、コロナウイルスへの対応に関することや、施設費等の学費減免に関する意見、また大学施設の開放要望や課題の量を減らすことなどへの要望があった。そして表5「コロナ禍とオンライン授業に関して感じていること」ではここまでの設問でも現れた学費等の減額要望や、後期授業への対応などの意見が見られたが、授業や友人関係、課外活動に関する不安といったトピックも見られた。これらからは現状に対する不満だけでなく、今後の動向が不透明なゆえに不安を感じている学生が多いことが示唆される。

表4 Q8A「後期授業や大学生活への考え方」の集計結果

解釈	件数	keyword
コロナウイルスへの対応	242	授業,オンライン,感染,コロナ,後期,大学,対面,人,ウイルス,通学
対面授業再開時の対応	209	授業,対面,オンライン,ゼミ,いい,人数,希望,人,講義,科目
オンラインと対面の並行実施への懸念	208	授業,オンライン,対面,時間,通学,行う,学校,良い,方式,受ける
施設費等の学費減額	141	施設,大学,学費,授業,利用,オンライン,欲しい,学校,この,給付
図書館等の大学施設開放	140	図書館,欲しい,利用,卒論,使える,施設,卒業,緩和,自由,開放
課題の量	120	授業,課題,オンライン,時間,量,欲しい,より,後期,自分,大学
課外活動の再開	107	活動,サークル,再開,部,課外,対面,大学,良い,面,ゼミナール
キャンパスライフ	81	生活,大学,不安,いい,友達,後期,キャンパス,オンライン,同じ,機

表 5 Q9A 「コロナ禍とオンライン授業に関して感じていること」 集計結果

解釈	件数	keyword
学費, 特に施設費の減額要望	186	施設,学費,授業,利用,大学,オンライン,設備,一部,使用,返金
後期のコロナウイルス対応	158	授業,オンライン,大学,対面,感染,後期,コロナ,人,この,時間
授業や課題への不安	135	課題,授業,提出,時間,先生,レポート,不安,オンライン,多く,全て
友人関係等の不安	112	不安,大学,人,友達,友人,相談,生活,コミュニケーション,就活,面
一律給付金の支給	111	大学,給付,学生,他,専修大学,一律,支援,状況,Wi,奨学
課外活動への不安	102	活動,サークル,情報,特に,不安,就職,相談,卒業,人,論文
印刷費用等の負担	93	印刷,コピー,授業,家,学校,資料,プリンター,負担,学費,使用
オンライン授業の質	92	授業,オンライン,講義,環境,音声,通信,質,課題,先生,多く
情報連絡への不満	89	情報,メール,特に,ポータル,連絡,欲しい,確認,更新,大学,重要
経済的問題	88	授業,バイト,実家,家賃,収入,欲しい,オンライン,今,アパート,経済
図書館やパソコン	67	図書館,パソコン,どう,本,多い,目,生活,学習,難しい,学校

最後に表6「オンライン授業に関する意見全般」の結果であるが、ここまでの設問の結果とほぼ同様のトピックが見られる。その中でも件数として多かったのは課題の量および学費等の減額要望、授業に関連すること全般であり、ここまでの設問の結果を踏まえて考えれば学生の置かれている状況や要望が具体的に示されたと言えるだろう。

表 6 Q10 「オンライン授業に関する意見全般」の集計結果

解釈	件数	keyword
課題の量が多い	257	課題,量,多く,毎日,多い,アルバイト,出す,一つ,毎週,睡眠
学費, 特に施設費の減額要望	236	施設,学費,利用,大学,一部,同じ,納得,減額,疑問,通常
課題提出方法等の統一	207	提出,課題,meet,期限,Google,統一,クラス,方法,もう少し,classroom
友人関係等への不安	206	とても,いい,不安,オンライン,自分,友達,勉強,友人,しっかり,直接
後期のコロナウイルス対応	204	オンライン,後期,コロナ,感染,ウイルス,学校,希望,通学,登校,数
成績評価, レポート評価への不安	196	課題,レポート,評価,テスト,成績,不安,方法,試験,単位,出席
授業コンテンツへの意見	188	授業,資料,音声,理解,内容,説明,科目,形式,ファイル,配布
後期対面授業再開への意見	186	授業,対面,オンライン,後期,受ける,良い,行う,人数,ゼミ,方式
通学時間等のメリット他	176	時間,より,学習,通学,復習,多い,予習,課題,良い,特に
パソコンに関する要望	175	欲しい,授業,パソコン,教員,いただける,ありがたい,集中,オンライン,使用,PC
印刷費や図書館利用	168	ほしい,分,負担,特に,図書館,印刷,非常,大きい,コピー,制限
授業の質の差への不満	166	授業,質,差,意味,理解,ほとんど,不満,改善,ただ,程度
オンライン授業のメリット・デメリットの対比	162	授業,感じ,オンライン,メリット,対面,とても,個人,デメリット,良い,グループ
通信環境への配慮	158	授業,オンライン,環境,かなり,通信,場合,行う,受講,考慮,悪い
教員への不満	155	先生,授業,質問,生徒,大変,問題,中,答え,改善,初めて
講義の配信方法	150	講義,教授,理解,内容,ほしい,目,配信,中,リアルタイム,より
要望文体	142	この,状況,思い,お願い,どう,大変,今,検討,よろしく,可能
大学生活やアルバイトなど	140	大学,生活,今,生徒,自分,バイト,学校,必要,活動,実家
授業や履修の情報	129	授業,オンライン,情報,考える,履修,キャンパス,実施,より,どの,学部
長文意見の傾向	127	大学,学生,側,他,対応,考え,意見,全て,面,この
現状への思いや意見, 考え方など	121	思い,人,少し,考え,中,良い,そう,形,自分,アンケート
給付金要望	92	学生,給付,支援,/,...授業,一律,-,オンライン,) _

4. 考察

これらの分析結果は後期授業の実施方法を検討する大学内の会議において報告を行った。本稿では割愛したが、報告時にはトピック選択確率の高い自由記述の原文を上記の各

表に加える形で提供した。意思決定層が学生からの自由記述の原文を直接見ることによって、学生の置かれている状況が直接的に理解でき、各設問におけるトピックの選択件数を量的に示すことでどのような意見が多いかをすぐに理解することができるという点が、施策を考える上で有効であったと思われる。加えて、トピックモデルによって膨大な情報が整理されたことで、分析者と意思決定者の双方の認知的・時間的負荷が大幅に軽減された。そして専修大学では本アンケート調査の結果を踏まえて十分な議論を行い、「学長伝書鳩[16][17]」という学長発のコラムにおいて、大学としての考え方や今後の方針を示した。その詳しい内容については参考文献に挙げたホームページで確認していただきたいが、かなりの部分が本アンケート調査の結果と対応した施策であると言えるだろう。したがって、本稿で示したような自由記述回答の分析および報告は、機関の意思決定に対して非常に強力な影響を与えることが示唆され、今後も幅広い応用が期待される。

【参考文献】

- [1] 中央教育審議会大学分科会, “「教学マネジメント指針」用語解説,” 2020. [Online]. Available: https://www.mext.go.jp/content/20200206-mxt_daigakuc03-000004749_005.pdf. [Accessed: 01-Sep-2020].
- [2] 国立情報学研究所, “4月からの大学等遠隔授業に関する取組状況共有サイバーシンポジウム,” 2020. [Online]. Available: <https://www.nii.ac.jp/event/other/decs/>. [Accessed: 01-Sep-2020].
- [3] 金井雅之, “オンライン授業に関する学生アンケートの概要,” in *MJIR2020 第9回 大学情報・機関調査研究集会*, 2020.
- [4] 松河秀哉, 大山牧子, 根岸千悠, 新居佳子, 岩崎千晶, and 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*, vol. 41, no. 3, pp. 233–244, 2018, doi: 10.15077/jjet.41018.
- [5] R Core Team, “R: A Language and Environment for Statistical Computing.” Vienna, Austria, 2019.
- [6] RStudio Team, “RStudio: Integrated Development Environment for R.” Boston, MA, 2020.
- [7] H. Wickham *et al.*, “Welcome to the {tidyverse},” *J. Open Source Softw.*, vol. 4, no. 43, p. 1686, 2019, doi: 10.21105/joss.01686.
- [8] 工藤拓, 山本薫, and 松本裕治, “Conditional Random Fields を用いた日本語形態素解析,” *情報処理学会研究報告. NL, 自然言語処理研究会報告*, vol. 161, pp. 89–96, 2004.
- [9] M. Ishida, “RMeCab: interface to MeCab.” 2018.
- [10] I. Feinerer and K. Hornik, “tm: Text Mining Package.” 2019.
- [11] J. Silge and D. Robinson, “tidytext: Text Mining and Analysis Using Tidy Data Principles in R,” *JOSS*, vol. 1, no. 3, 2016, doi: 10.21105/joss.00037.
- [12] B. Grün and K. Hornik, “{topicmodels}: An {R} Package for Fitting Topic Models,” *J. Stat. Softw.*, vol. 40, no. 13, pp. 1–30, 2011, doi: 10.18637/jss.v040.i13.
- [13] M. Nikita, “ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters.” 2020.
- [14] M. Kuhn and H. Wickham, “tidymodels: Easily Install and Load the ‘Tidymodels’ Packages.” 2020.
- [15] 植原啓介, “慶應SFCにおける遠隔授業とアンケート調査結果,” **【第10回】4月からの大学等遠隔授業に関する取組状況共有サイバーシンポジウム**, 2020. [Online]. Available: https://www.nii.ac.jp/event/upload/20200605-5_Uehara.pdf. [Accessed: 02-Sep-2020].
- [16] “「学長伝書鳩」No.6,” 専修大学ホームページ. [Online]. Available: <https://www.senshu-u.ac.jp/news/nid00011437.html>. [Accessed: 02-Sep-2020].
- [17] “「学長伝書鳩」No.7,” 専修大学ホームページ. [Online]. Available: <https://www.senshu-u.ac.jp/news/nid00011602.html>. [Accessed: 10-Sep-2020].