

アドミッション・ポリシーの分類と作成支援の試み

齋藤 朗宏 (北九州市立大学)

アドミッション・ポリシーに関するこれまでの研究では、頻出単語のクラスタリングなどを通してポリシーの分類などを行ってきたが、方法によっては分類に使う単語の恣意性などの問題があった。そこで本研究では、第一にアドミッション・ポリシーの各段落の文章に対してクラスター分析を行った。結果、クラスターごとに特徴的な単語の共起パターンなどを発見することができた。第二に、各ポリシーに対して BERT モデルを用いてポリシーの文章同士の類似度を算出した。これを元にするこ

とで、何らかの文章を入力することで、それに類似するポリシーを抽出することができた。この手法を用いることで、実際のポリシーを作成する際の支援を行うことができると考えられる。

キーワード：アドミッション・ポリシー (AP) , クラスタ分析, NLP, BERT

1 はじめに

1.1 アドミッション・ポリシー

入学者受け入れの方針 (アドミッション・ポリシー、以下 AP) の制定が中央教育審議会答申で提言 (文部科学省 中央教育審議会, 1999) されてから 20 年以上が経過した。その間、AP の研究も大きく進展しており、鳴野ほか(2004)による各大学へのアンケート調査をはじめとして、山村ほか(2014)による AP の効果検討など様々である。AP は後に制定が提言された教育課程編成・実施の方針であるカリキュラム・ポリシー (以下 CP) , 学位授与の方針であるディプロマ・ポリシー (以下 DP) と共に文部科学省 中央教育審議会大学分科会大学教育部会(2016)において具体的なガイドラインが示されるなど、実体を伴うものとして制定の重要性が年々増している。

1.2 研究目的

以上の背景を踏まえると、AP の制定は他大学の事例なども参考にした上で慎重に行うのが望ましいと考えられる。よって、そのための内容分析を通じた AP の作成支援は有用であると考えられる。たとえば齋藤(2020)では、AP を含む 3 つのポリシーについて、たとえば「建学の精神」や「選抜方法」など、それぞれ 6 つずつの代表的な内容を設定し、その内容に言及しているかという観点から、3 つのポリシー間の関係をベイジアンネットワークを用いて分析している。結果、DP から CP への影響は大きいですが、CP から AP の影響は小さいことなどを確認している。ただ、この分析では、ある内容に言及しているかという観点について、いくつかの代表的な単語の有無に絞って決定しており、その点が分類の正確さ等の観点から問題であった。こういった内容分析のためには、より細かく、文章単位

での内容分析が必要と考える。そこで本研究では、第一に全学 AP に関して文章単位での単語抽出や分類など基本的な分析を行う (分析 1) 。

その上で、文章の分析を実際の AP 作成支援に繋げる方法を検討する。齋藤(2020)で確認した通り、全体の 3/4 程度の大学において全学 AP が制定されている。そのため、この全学 AP の作成を支援することが実務上の支援に繋がると期待できる。そこで第二に、ある文章から、それに類似する全学 AP を抽出することを試みる (分析 2) 。このような AP を作成したいというイメージとなる文章を入力すれば、それに類似する実際の AP を簡単に見つけることができるようになるため、それらの AP を参考にして説明の過不足等を調整し、より良い AP を作成することができるようにすると期待できる。

2 方法

2.1 使用するデータ

データは齋藤(2020)で作成した 2019 年 5 月現在の 3 つのポリシーのデータベースから、AP のみを取り出したものを用いた。データ収集の条件等は以下の通りであった。

分析対象は、2 つ以上の学部を持つ四年制総合大学における、全学の 3 つの方針とする。

総合大学の定義は困難だが、便宜上ここでは、2 つ以上の学部を持ち、少なくとも 1 つ以上の文科系課程を持っていることを条件とした。単科大学を条件から外したのは、単科大学においては、学部の方針と全学の方針との区別が困難なためである。

複数学部であっても、経済学部と経営学部の 2 学部のみという場合のように、類似した教育を行っていることが想像される学部のみで構成されている場合に

は、同様の理由で分析対象から除外した。文科系を少なくとも 1 学部含むことを条件としているのも、この点を考慮している。

データベースは国立 50 大学、公立 35 大学、私立 319 大学の 404 大学により構成され、このうち、全学 AP が含まれ、本研究の対象となったのは国立 47 大学、公立 24 大学、私立 230 大学の合計 301 大学であった。

2.2 分析 1 : 文章の分類

第一に、KH Coder 3.Beta.02c(樋口, 2004)を用いて文章のクラスター分析を行った。分析の対象は名詞とし、分析単位を段落とした。これは、文章単位にした場合、AP とは直接的に関係のない文章が含まれる部分など、解釈不能なクラスターが多く出現してしまうこと、逆に大学単位にした場合には、その大学の AP の中には多様な内容が含まれるため、内容の有無などからクラスター分けをするのが難しいことを考慮したためである。

クラスター分けには、名詞のうち 30 段落以上で出現した 157 語を用い、解釈のしやすさなども踏まえて 10 のクラスターに分類を行った。単語間の距離には Jaccard 距離を、クラスターの併合には Ward 法を用いた。

2.3 分析 2 : 類似文章の抽出

第二に、BERT モデルを用いて類似した文章同士の類似度を求めた。BERT は、深層学習のモデルである Transformer を元にして Google によって発表された自然言語処理 (NLP) モデルである。たとえば①「彼はとても腕がいい。」という文章に出てくる「腕」は、②「筋肉痛で腕が痛い。」という文章に出てくる「腕」よりも、③「高い技術を持った職人に仕事を依頼した。」という文章に出てくる「技術」の方が意味としては近い。しかし、単純に単語のみを分析対象とした場合には、①の「腕」と②の「腕」とが同じ単語として扱われ、①の「腕」と③の「技術」は異なる単語として扱われることになる。一方 BERT の場合には、①の「腕」と③「技術」との類似度は、①と②の「腕」同士の類似度よりも高い値と判定される。このように、BERT は文脈に基づいた単語の使われ方も含めて検討可能なモデルであり、BERT を用いることで、より高い精度で類似した文章を抽出できると期待できる。Google Colaboratory 上で Python を用いて分析を行った。Python のバージョンは 3.7.13 であり、Transformer のバージョンは 2.9.0 であった。

また、BERT では自身で準備した文章だけではなく、事前学習のデータを元に分析を行うことができる。本研究では、東北大学で公開している Pretrained Japanese BERT models¹を用いた。

こちらは、文章から参考となる別の文章を見つけることを目的としているため、より細かく、文単位での検討を行った。また、文章には「アドミッションポリシー」のみで 1 文として扱われてしまっているケースのように、1 単語だけの実質的に見出しとになってしまっているものもあった。「アドミッションポリシー」が 11 字であることから、12 字を下回る文章については、分析の対象から取り除いた。全体で 2723 文あったが、12 文字を下回る文章を取り除いた結果分析対象は 2419 文となった。これらと、このような AP を作成したいというイメージとなる文章を加えて、それらの間のコサイン類似度を求めている。

NLP では、単語や文章などの特徴をベクトルで表現している。コサイン類似度は、そのベクトル間の類似度を測定する指標である。ここでは、2 つの文章の特徴を示すベクトル同士がなす角を利用し、コサインを求めている。結果、なす角が小さく類似している場合にはコサイン類似度は 1 に近づき、逆になす角が大きく類似していない場合には -1 に近い値が得られる。

3 結果と考察

3.1 分析 1 : クラスター分類結果

各クラスターに分類された文書数は、それぞれ表 1 の通りであった。分類不可とされたものは、特定のクラスターに分類できなかったものであり、内容を確認すると、通し番号だけの段落などが含まれていた。また、特に分類された文書数の少なかったクラスター 9 を確認すると、こちらは「求める学生像は次のとおりです。」といった内容のみが入っている段落であり、次に少なかったクラスター 1 も、「一般入試 (前期日程)」のような入試区分名のみであり、クラスター 2 も同様に、「入学者選抜評価の基本方針」のような文章のみであることが多かった。これらも分析上意味はないので今後の検討の対象からは外すものとする。

¹ <https://github.com/cl-tohoku/bert-japanese>

表 1 各クラスターに含まれる文書（段落）数

	段落数
分類不可	65
クラスター1	114
クラスター2	167
クラスター3	184
クラスター4	125
クラスター5	413
クラスター6	146
クラスター7	149
クラスター8	367
クラスター9	37
クラスター10	513

表 3 クラスター4 の頻出単語

単語	全体	共起
人	590 (0.259)	118 (0.944)
目標	134 (0.059)	17 (0.136)
文化	108 (0.047)	15 (0.120)
人間	151 (0.066)	14 (0.112)
他者	95 (0.042)	14 (0.112)
次	111 (0.049)	12 (0.096)
自分	77 (0.034)	12 (0.096)
尊重	42 (0.018)	12 (0.096)
努力	33 (0.014)	11 (0.088)
関心	113 (0.050)	10 (0.080)

3.2 分析 1：各クラスターの特徴

分類不可であったものと、分析上意味がないと判断されたクラスター1, 2, 9を除いた7つのクラスターについて、それぞれに含まれる単語の特徴を確認した。それぞれのクラスターについて、共起確率（ここではその段落中に出現する確率）が最も高い10単語を抽出した結果が表2から8であった。ここで、全体は全体の中での出現段落数であり、括弧内は全体での出現率である。共起はクラスター内の出現段落数と出現率である。

表 2 クラスター3 の頻出単語

単語	全体	共起
社会	472 (0.207)	156 (0.848)
人	590 (0.259)	87 (0.473)
貢献	200 (0.088)	82 (0.446)
地域	159 (0.070)	65 (0.353)
国際	138 (0.061)	53 (0.288)
意欲	396 (0.174)	43 (0.234)
発展	74 (0.032)	28 (0.152)
人間	151 (0.066)	25 (0.136)
活躍	64 (0.028)	23 (0.125)
分野	122 (0.054)	20 (0.109)

表 4 クラスター5 の頻出単語

単語	全体	共起
人	590 (0.259)	126 (0.305)
意欲	396 (0.174)	123 (0.298)
課題	130 (0.057)	52 (0.126)
積極	113 (0.050)	50 (0.121)
主体	114 (0.050)	49 (0.119)
態度	113 (0.050)	48 (0.116)
活動	124 (0.054)	47 (0.114)
解決	110 (0.048)	44 (0.107)
コミュニケーション	87 (0.038)	39 (0.094)
関心	113 (0.050)	39 (0.094)

表 5 クラスター6 の頻出単語

単語	全体	共起
評価	197 (0.086)	114 (0.781)
面接	124 (0.054)	76 (0.521)
選抜	264 (0.116)	75 (0.514)
総合	137 (0.060)	73 (0.500)
試験	264 (0.116)	67 (0.459)
能力	364 (0.160)	67 (0.459)
入試	268 (0.118)	60 (0.411)
入学	469 (0.206)	59 (0.404)
学力	293 (0.129)	51 (0.349)
書類	67 (0.029)	49 (0.336)

表6 クラスター7の頻出単語

単語	全体	共起
表現	172 (0.075)	112 (0.752)
思考	188 (0.082)	110 (0.738)
判断	168 (0.074)	105 (0.705)
知識	369 (0.162)	63 (0.423)
技能	198 (0.087)	56 (0.376)
基礎	345 (0.151)	40 (0.268)
身	255 (0.112)	31 (0.208)
学力	293 (0.129)	30 (0.201)
能力	364 (0.160)	30 (0.201)
態度	113 (0.050)	28 (0.188)

表7 クラスター8の頻出単語

単語	全体	共起
基礎	345 (0.151)	189 (0.515)
知識	369 (0.162)	166 (0.452)
学力	293 (0.129)	134 (0.365)
学校	245 (0.107)	122 (0.332)
人	590 (0.259)	109 (0.297)
入学	469 (0.206)	95 (0.259)
技能	198 (0.087)	89 (0.243)
教育	484 (0.212)	86 (0.234)
身	255 (0.112)	83 (0.226)
試験	264 (0.116)	77 (0.210)

表8 クラスター10の頻出単語

単語	全体	共起
教育	484 (0.212)	315 (0.614)
本学	390 (0.171)	217 (0.423)
社会	472 (0.207)	161 (0.314)
理念	169 (0.074)	150 (0.292)
人材	194 (0.085)	142 (0.277)
精神	171 (0.075)	125 (0.244)
能力	364 (0.160)	123 (0.240)
大学	228 (0.100)	122 (0.238)
入学	469 (0.206)	122 (0.238)
学生	288 (0.126)	121 (0.236)

してその確率の比が大きなもの（表の太字部分）に着目することでクラスターの特徴が確認できる。それぞれのクラスターの特徴的な単語とそれを元にしたクラスターの解釈は以下の通りである。

- クラスター3：社会，地域，国際，貢献。「社会貢献への意欲」
- クラスター4：人，目標，文化。「目標設定と努力への意思」
- クラスター5：課題，積極，主体。「主体性」
- クラスター6：評価，面接，選抜，総合，試験。「入試制度」
- クラスター7：表現，思考，判断。「思考力と表現力」
- クラスター8：基礎，知識，学力。「基礎学力」
- クラスター10：教育，本学，理念，精神。「教育理念の理解，共感」

該当段落数が多かったクラスター10と、他部分と比較してやや抽象度が高く、結果がややわかりにくかったクラスター4, 5 に関して単語の共起ネットワークを作成し、単語同士の共起パターンを確認することで、クラスター分類通りの内容になっているかより詳細に検討を行った。その結果はそれぞれ図1～3の通りであった。



図1 クラスター10の共起ネットワーク

クラスター10は教育理念への理解，共感と解釈している。実際，図左下の部分にそういった特徴は現れているが，一方で右下では人間性や教養，専門知識を

共起確率の高い単語の中で，全体での出現率と比較

身につけることが謳われ、右上では選抜方法に関する言及があるなど、必ずしも特定の内容だけで段落が構成されているわけではないことが確認できる。これは他のクラスターにおいても同様である。



図 2 クラスタ4の共起ネットワーク

クラスター4は目標設定と努力への意思と解釈していた。実際、先ほど挙げた単語以外にも、自ら、積極なども共起しており、解釈が概ね間違いないことが確認できた。



図 3 クラスタ5の共起ネットワーク

クラスター5は主体性となっており、クラスター4と類似している。ただ、共起単語を見ると、他者や尊重、豊かなども挙げられており、主体性を含む人間性全般の涵養を示していると理解した方が良さそうである。

3.3 分析2：文章分類の結果

APの例として、「大学で学ぶ意欲がある人を求めています。」という文章(文例1)を作成し、それらと残りの文章とのコサイン類似度を求め、類似度が高い順に並べた結果は以下の通りである。

1. まちと暮らしに関心をもてる人を求めています。(コサイン類似度：0.913)
2. 自らが積極的に学び、社会で活躍できる力を身につけたいと望む学生を求めます。(0.912)
3. その専門知を生かして狭い専門領域を超えて統合し、共生社会の創造に貢献したいという意欲を持った人を求めています。(0.912)
4. 伝えることの大切さが分かる人を求めています。(0.912)
5. 経験から学び取る力やコミュニケーション能力を有する学生を求めます。(0.907)
6. 意欲をもっている人を本学は求めています。(0.905)
7. 「生きていること」に関心をもとうという人を求めています。(0.893)
8. 活動の中で中核的役割を担おうとする人材を求めます。(0.891)
9. 貢献しようとする人を本学は求めています。(0.890)
10. 生涯にわたって学び成長し続ける持続性・社会性・能動性をもつ社会人の受け入れを目的としています。(0.889)

文例1では「意欲」という言葉を使っているが、抽出されたAPの中では、必ずしも意欲という単語を使っている訳ではなく、「関心」や「望む」、「～しようとする」という言葉で表現されているものが似た意味として抽出されることになった。これらは、単語を元にした分析で見つけることは難しく、BERTを用いた成果であると考えられる。

同様に、「高等学校において習得した基礎的な学力を一般選抜を通して確認します。」という文章(文例2)と類似度が高い文章を抽出すると以下の通りであった。

1. 高等学校教育段階においてめざす基礎学力を確認します。(0.927)

2. 大学入試センター試験の結果で、志望する学部学科入学後の学習内容に必要な基礎学力および応用力、理解力を評価します。(0.926)
3. 学力検査では、志望する学部学科へ入学した後の学習内容に必要な基礎学力および応用力、理解力を評価します。(0.924)
4. 専門分野を学ぶために必要な基礎学力を測る3科目または2科目の学科試験によって能力を判定します。(0.920)
5. 修学に必要な最低限の学力を確認するため、高校での学業成績をも評価の対象とします。(0.917)
6. 書類審査、学力試験、小論文、面接によって総合的に能力を判定します。(0.909)
7. 一般入試(第3年次編入学・学士入学入試)では、個別学力試験を重視し、「実践的な知力」を修得するために必要不可欠となる教養と基礎学力を有し、短期大学・高等専門学校・大学で得た知識、理解力、論理的な思考力、判断力等を確認します。(0.909)
8. 本学では、本学が実施する個別学力検査により、総合的な基礎学力とともに、思考力・判断力・表現力が身につけているかを問うています。(0.907)
9. 大学入試センター試験の得点により、本学の求める日本語能力(国語)と、得意な1教科(地理歴史・公民、数学、外国語のいずれか)について、本学での学修における基礎学力を有する学生を求めます。(0.904)
10. 本学において専門知識を学ぶための「知識・技能」が身につけているか、日本留学生試験の成績を利用して評価し、面接において「主体性を持って学ぶ態度、思考力・判断力・表現力」、基礎知識などの評価を加え、総合的な選抜を行います。(0.904)

文例1では、大雑把に意欲を指し示す文章を元に類似する文章を抽出したため、抽出された文章もまた、意欲に関連したものが広く抽出されることとなった。一方で文例2では、高等学校で学んだ基礎的学力、一般選抜といったより具体性が高い文章になっているため、抽出される文章もまた、高等学校での学習内容を試験で問うということに関係したものが多くなっていることが確認できた。

さらに、どの程度の文章を入力すればより明確な文例が得られるか確認するため、単に「学力検査。」とした場合(文例3)、また、文例2をより具体的な科

目に言及した「数学的思考力を一般選抜を通して確認します。」(文例4)も作成し、それぞれ類似文書の抽出を行った。結果、文例3では「海外帰国生徒入学試験」といった単に入試形態を示す単語のみのものをはじめとして、「社会人、外国人留学生及び帰国生徒を対象とした入試。」など、文例としても、また、入力した文章との関係性からも十分なものは抽出できなかった。一方文例4では、「この試験により基礎知識、論理的な理解力および洞察力などを確認します。」といった、かなり関係性の深い文章が抽出できている、本方法では、作成したい文章を象徴する単語を複数入れた文例を作成することでより作成したいAPに関連する文章を抽出できると考えられる。

4 まとめと今後の課題

従来のAP分類の試みにおいては、単語のクラスタリングを通してその単語出現の有無などから検討してきたが、本研究では、AP中の各段落をクラスタ分けすることで、より直接的にAPの分類を行うことができた。ただ、段落を用いたクラスタリングでは、段落内で複数の話題を取り上げているケースが珍しくないことも図1などから確認できた。

BERTを用いた文章分類からは、同じようなことを異なる表現で説明している文章を抽出することができた。ここから、当初期待していたとおり、AP作成時に、内容的な不足の確認や、表現の洗練などの参考にすることが可能になったと考えられる。

一方で、文例1では、類似度が高い文章の中は、意味で見たときにあまり近いとは言いがたいものも散見された。また、文例4でも、数学的思考力という文例であったにもかかわらず、科目名ということで類似していると判断された国語に関する説明なども抽出されている。こういったことは、ある程度人力で分類した上で、教師あり学習を用いることにより精度が上がる可能性が考えられる。今後は、そのような観点からのモデルの精緻化にも取り組みたい。

参考文献

- 齋藤朗宏(2020). 「3つのポリシーの連携に関する分析」『大学入試研究ジャーナル』30, 74-79.
- 嶋野英彦ほか(2004). 「アドミッション・ポリシーと入学受入方策」大学入試センター研究開発部共同研究報告書.
- 樋口耕一(2004). 「テキスト型データの計量的分析: 2つのアプローチの峻別と統合」『理論と方法』19(1), 101-115.
- 文部科学省 中央教育審議会(1999). 『初等中等教育と高等教育との接続の改善について(答申)』(平成20年12月24日).

文部科学省 中央教育審議会大学分科会大学教育部会(2016).

「『卒業認定・学位授与の方針』(ディプロマ・ポリシー),
『教育課程編成・実施の方針』(カリキュラム・ポリシー)
及び『入学者受入れの方針』(アドミッション・ポリシー)
の策定及び運用に関するガイドライン」(平成28年3月31
日)

山村滋ほか(2014). 「アドミッション・ポリシーの効果に関する研究」大学入試センター研究開発部共同研究報告書.