

新入学生アンケートのテキストマイニングによる 大学の価値分析（第一報）

林 俊克（就実大学経営学部）

University Value Analysis by Text Mining of New Student Questionnaire (First Report)

Toshikatsu Hayashi

新入学生アンケートをテキストマイニングすることで、大学の備えるべき価値を可視化・定量化することを試みた。共起分析・対応分析・価値ポートフォリオ分析の結果、「就職が良い」「資格が取得できる」「友達と良好な関係が築ける」が理想の大学として必要であり、「資格がとれるカリキュラムの充実」「良い就職の斡旋・確保」「通学のしやすさ」「良い教員の採用」「割安感のある授業料」に関して未だ不満であることがわかった。

The value of the university is attempted to be visualized and quantified by Text Mining of the questionnaire of new students.

As a result of co-occurrence analysis, correspondence analysis, and value portfolio analysis, “Advantageous for employment”, “Qualifications can be obtained”, and “Good relationships with friends” were necessary as an ideal university, and Examinee were not satisfied enough with “Enhancement of qualified curriculum”, “Recruitment and securing of good employment”, “Ease of attending school”, “Recruitment of good teachers”, and “Discounted tuition”.

キーワード：大学の価値、テキストマイニング、共起分析、対応分析、価値ポートフォリオ分析

【はじめに】

私立大学の経営環境は、近年の少子化や大卒の経済的メリットの低下により大変厳しい状況にある。文部科学省高等教育局私学部参事官付が2018年に公表した『私学行政の現状と課題等について』によると、地方中小私大の収支状況は約半数が赤字傾向にあり、教学的な指摘事項として、「設置学部等の学生確保に向けた対応策を立案し着実に実施」「定員管理の適正化、定員超過の改善」が記されている。

文部科学省の指摘を待つまでもなく、大学の経営環境を改善するためには、組織の再編や経費の削減などによる経営の効率化が必要だけでなく、入学希望者を増やすことが不可欠である。

受験生の獲得に向けて、新しい時代の要請に応えた学部・学科の新設や特色ある教育活動の展開等、様々な努力によって成果を上げている大学も少なくない一方で、状況が改善しない或いは悪化している大学も散見されるが、いったい何がその成否を分けているのだろうか？多くの受験生を獲得するために、大学はどのような施策を打つべきなのだろうか？この問いに答える事は多くの大学にとっての福音となることは間違いない。そこで本稿では、感性工学や魅力工学をはじめ、多くの学問分野で汎用されるようになったテキストマイニングの手法を用いて、大学の備えるべき価値の可視化・定量化を試みた。

感性工学や魅力工学の分野において、価値の定量化・可視化手法は多数提案され、数多くの書籍や研究論文が発表されているが、ここでは、定型自由文形式のアンケートを、形態素解析によってデータ化し、周知の解析手法である共起分析、対応分析を行うとともに、価値分析の為に開発した価値ポートフォリオ分析「Dohkan,K., Hayashi,T., Masuda,M. and Fukuchi,Y. (1999)」「真柳麻誉美、林俊克、平野広隆 (2001)」「林 (2001, 2002)」を行った結果について報告する。

ちなみに、テキストマイニングとは、自由文のようなテキスト型データを分析して役立つ知識・情報を取り出そうとする技術の総称であり、テキスト型データを計量的に扱えるようにデータ変換する「形態素解析」技術の進歩によって実用可能となり、1997年頃から多分野の学術論文の中で頻繁に使われるようになった。当初は、形態素解析エンジンが非公開であったり非常に高価であったため、限られた一部の研究機関でのみ実施が可能であったが、奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座（松本研究室）が、広く自然言語処理研究に資するための無償のソフトウェア「ChaSen」を開発したことで、テキスト型データの計量的解析が可能となった。その意味で、テキストマイニングは本質的にはデータマイニングと呼ばれる旧来からある統計手法による解析をテキスト型データまでに拡張した技術と言えるものであり、多変量解析を含む古典的な統計手法、クラスター分析、共起分析、リンク分析、決定木、記憶ベース推論、ニューラルネットワーク、遺伝的アルゴリズム等々様々な解析手法の適用によって、多量のテキストのなかから必要な情報を探し出したり、多量のテキストを要約し論旨を明確にしたり、元テキストの著者でさえも気づかなかった新しい情報を発見したりすることを可能にしている。

【方法】

2016年度より、担当する講義「データサイエンス入門」のオリエンテーションの際に実施している新入学生アンケートの設問のうち、定型自由文形式の以下の設問部分（図1）の結果4年分（2016年～2019年）を分析に供した。アンケートの際、データの学術利用の可否を聞き、利用可とする学生346名分のデータについて解析を行った。アンケートは、学生の回答入力の利便性、結果をリアルタイムに表示可能であること、結果をエクセル等の汎用性の高い形式でエクスポート可能であることから、Google Formを利用した。設問では、「一般的な大学」「理想の大学」「就実大学経営学部」について短文でイメージを回答することとし、回答された短い定型自由文をテキストマイニングすることで、大学の備えるべき価値の可視化・定量化を試みた。

形態素解析による定型自由文テキストのデータ化には、KH Coderを使用した。KH Coderは、樋口耕一氏が開発したフリーのテキストマイニングソフトウェアであり、詳細な仕様・使用マニュアルも提供されている。「樋口 (2018)」

The figure consists of four separate survey form panels arranged in a 2x2 grid. Each panel has a dark header bar with a question and three short-answer input fields below. The questions are: 1. '一般的な大学' (General University), 2. '理想の大学' (Ideal University), 3. '就実大学経営学部' (Keio University Faculty of Business Administration), and 4. '就実大学経営学部新入生' (Keio University Faculty of Business Administration New Students). Each panel also includes a title, a brief description of the survey, and a note about the use of the data.

図1. アンケート画面 (一般的な大学、理想の大学、就実大学経営学部について短文で回答)

KH Coderでは、テキスト型データを計量的に扱えるようにテータ変換する機能として、ChaSenの形態素解析結果をほぼそのまま利用している。また統計解析機能としてKH Coderは、統計解析とグラフィックスのために、オープンソースの統計解析言語・環境である「R」を利用しており、汎用的なテキストマイニング手法として強力な、対応分析 (数量化Ⅲ類)・クラスター分析・多次元尺度構成法 (MDS)・自己組織化マップ・共起ネットワーク・機械学習 (ナイーブベイズ) などを実装している。共起分析では、解析結果を関連が特に強い語同士を線で結んだネットワーク図としてアウトプットすることができるため、定量化は勿論のこと可視化にも優れている。対応分析でも、解析結果を2次元散布図としてアウトプットできるため、結果の考察が非常に容易である。

KH Coderは、このような形態素解析機能や高度な分析機能・グラフィック機能を日本語だけでなく英語・中国語・韓国語・イタリア語・オランダ語・カタロニア語・スペイン語・ドイツ語・フランス語・ポルトガル語・ロシア語データについても提供しており、驚くべきフリーソフトウェアであると言える。

価値ポートフォリオ分析は、KH Coderの形態素解析結果をエクスポートした後、エクセルを用いてデータを編集加工して行った。価値ポートフォリオ分析とは、顧客価値の定量化・可視化を目

的に林らが提唱する分析手法の一つで、顧客にとっての価値とは何なのか？を明らかにする為の手法である。①顧客の理想を実現するほど価値が高い。②顧客の現在の不満を解消するほど価値が高い。③多くの顧客に認識されるほど価値が高い。との観点に立ち、簡単な定型自由文形式のアンケートで商品やサービスの「一般的現状」「一般的理想」「特定の商品やサービス」のイメージを定義のような単意の簡潔な言葉で聞くことから「定義法」とも呼んでいる。「一般的現状」と「一般的理想」を2次元散布図に展開し、「一般的現状」と「一般的理想」の情報を表すポイントが散布図の45度のラインを境にして現状側か理想側かで「既に満たされている価値」なのか「未だ満たされていない価値」なのかを分析できるようにした図を「価値ポートフォリオ」と呼ぶことから、この手法による価値分析を価値ポートフォリオ分析と呼んでいる。また、この価値ポートフォリオ分析を拡張すると、「一般的現状」と「一般的理想」の情報を価値ベクトルと考え、価値ベクトルの角度と長さから顧客にとっての価値を定量化することもできる。更に、「特定の商品やサービス」がその価値や理想にどの程度コミットできているかも調べることができる。今回のアンケートで、「一般的な大学」「理想の大学」に加えて「就実大学経営学部」について調べたのはこの分析を行う為である。

【結果と考察】

(1) 共起分析からの知見

KH Coderによる、「一般的な大学」「理想の大学」「就実大学経営学部」についてのイメージの共起分析結果を(図2)に示す。KH Coderの分析条件は(図3)の通りである。

一般的な大学は「現状」、理想の大学は「理想」、就実大学経営学部は「就実」とラベルし、ラベルと頻出語との共起について、共起の度合い(Degree)、共起関係の強さ(Coefficient)、語の出現頻度(Frequency)をそれぞれ色、線の太さ、バブルの大きさに示している。

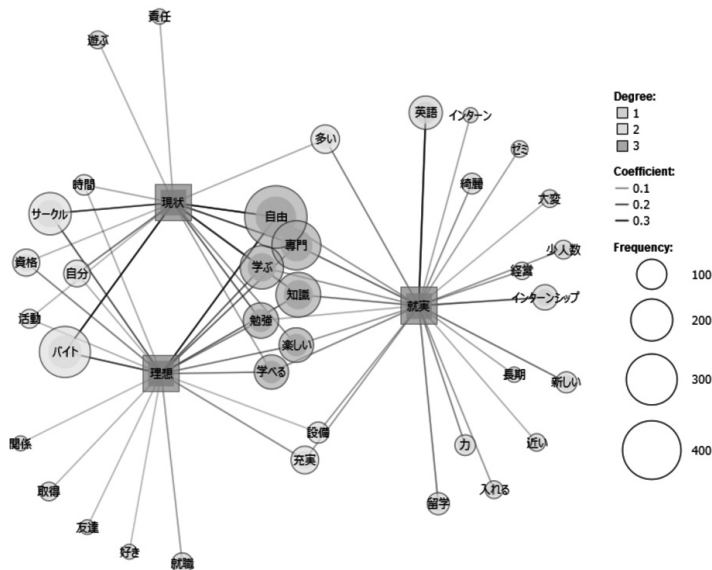


図2. 大学について思っているイメージ(共起分析)

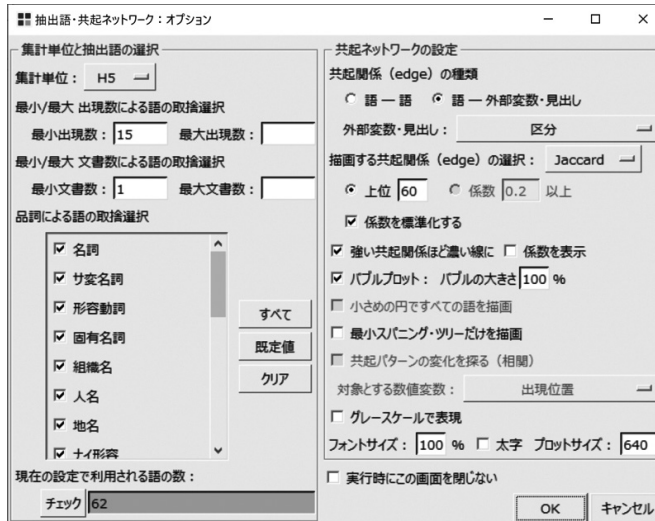


図3. KH Coderの分析条件（共起分析）

図2からわかるとおり、「一般的な大学」「理想の大学」「就実大学経営学部」に共通するイメージとしては、

- ・自由がある
- ・専門知識を学べる
- ・楽しい

といったイメージが持たれていることがわかる。

また、「一般的な大学」にも「理想の大学」にも挙げられるイメージとして、

- ・サークル活動をする
- ・バイトをする
- ・資格をとる
- ・時間がある

が挙がっており、これらは現状の大学でも満たすことができる価値であると考えられる。

一方で、「理想の大学」にのみ挙げられるイメージとしては、

- ・就職が良い
- ・資格が取得できる
- ・友達と良好な関係が築ける

があり、昨今の大学選択における就職至上主義の一端を垣間見ることができる。

「就実大学経営学部」と「理想の大学」の両方に共通するイメージは、

- ・設備が充実している

であり、受験生に就実大学の魅力をアピールする際は、いの一歩に「設備の充実」を取り上げるべきであると考えられる。

(2) 対応分析からの知見

一般的な大学、理想の大学、就実大学経営学部についてのイメージの対応分析結果を(図4)に示す。KH Coderの分析条件は以下(図5)の通りである。一般的な大学は「現状」、理想の大学は「理想」、就実大学経営学部は「就実」とラベルし、語の出現頻度(Frequency)をバブルの大ききで示している。

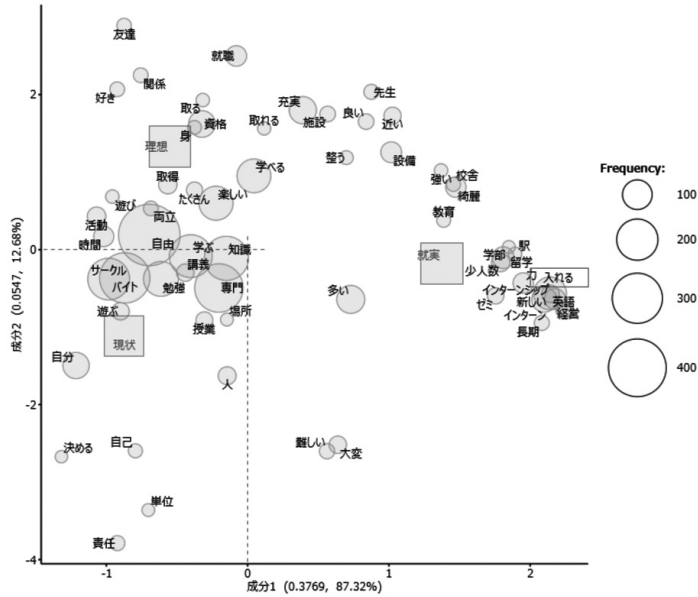


図4. 大学について思っているイメージ (対応分析)

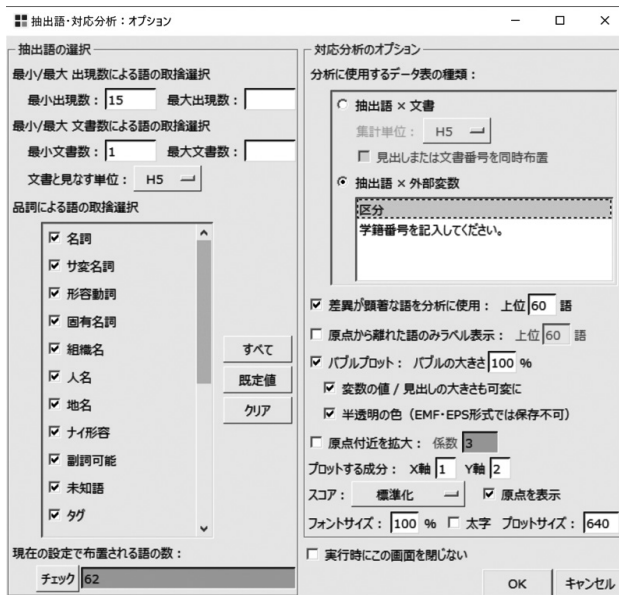


図5. KH Coderの分析条件 (対応分析)

図4からわかるとおり、原点付近には、

- ・ 専門知識を学べる
- ・ 自由がある
- ・ 勉強とサークル・バイト・遊びの両立
- ・ 楽しい

といったイメージが見られることから、それらが「一般的な大学」「理想の大学」「就実大学経営学部」に共通するイメージであると考えられる。またこれは先の共起分析結果と同様の傾向を示している。

また、「理想の大学」に特徴的なイメージとしては、

- ・ 就職が良い
- ・ 資格が取得できる
- ・ 友達と良好な関係が築ける

があり、こちらも先の共起分析結果と同様の傾向を示している。

「就実大学経営学部」と「理想の大学」の両方に共通に持たれやすいイメージとしては、

- ・ 設備が充実している
- ・ 良い先生が居る
- ・ 家から近い

等が挙げられており、設備の充実だけでなく、教員の充実度や通学のしやすさも大きなアピールポイントになると推測できる。

(3) 価値ポートフォリオ分析からの知見

価値ポートフォリオ分析の結果を(図6)に示す。有効回答364名中、何%の学生がその語をイメージとして挙げたかを、それぞれ一般的な大学は「現状(%)」、理想の大学は「理想(%)」とし、「現状(%)」をx軸、「理想(%)」をy軸として散布図を描いている。この時、分析対象語は277語となったが、図を見やすくするため、頻出上位50語を図示した。

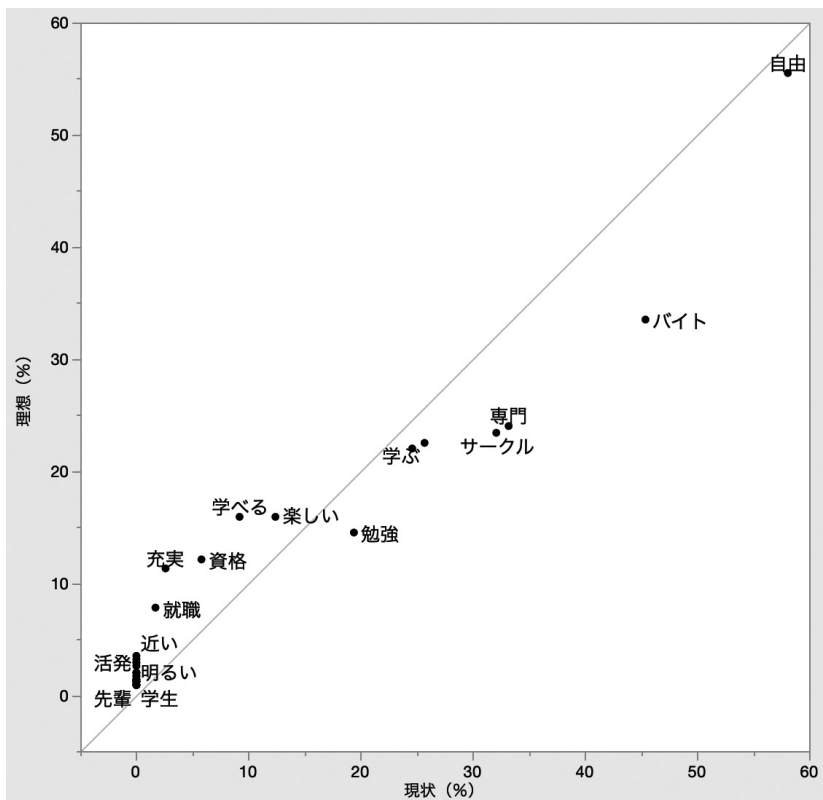


図6. 価値ポートフォリオ分析結果

図6において、「現状%」と「理想%」の値が一致する45度のラインを境にして、現状側か理想側かで「既に満たされている価値」なのか「未だ満たされていない価値」なのかを分析することができる。すなわち、価値ポートフォリオにおける右下は「既に満たされている」度合いが高く、逆に左上は「まだ満たされていない」と考えることができる。その意味では、殆どのイメージは45度ラインに近い位置にあることから、現在の一般的な大学は、ほぼ丁度よい満足度にあることが推察できる。

ここで、語の出現頻度の低い20%以下の部分に注目すると、45度ラインよりもやや左上にあるイメージが散見されることから、20%以下の部分を拡大した（図7）を作成した。

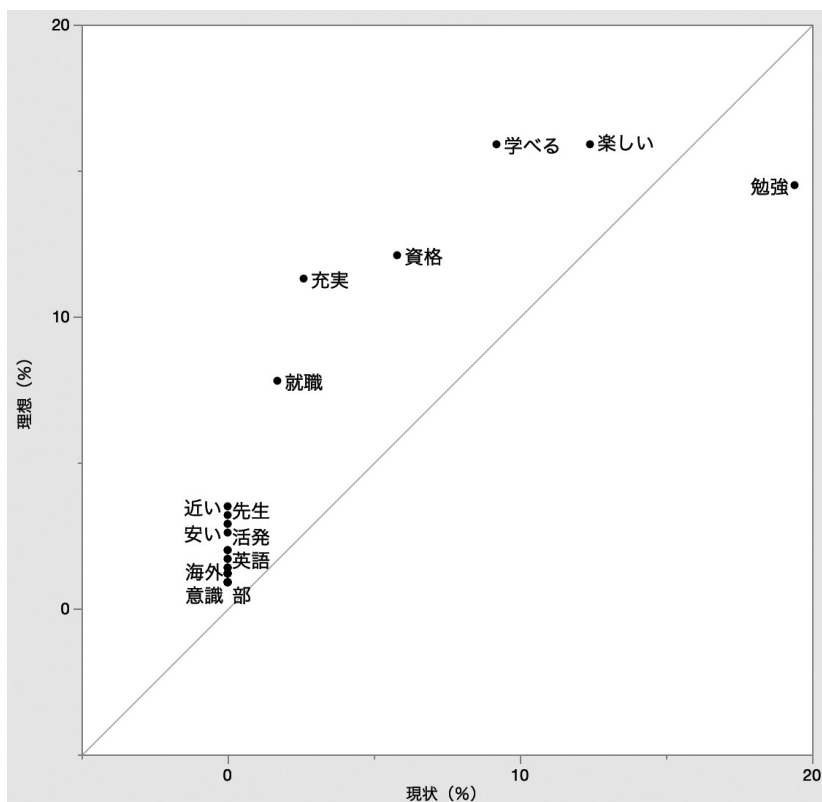


図7. 価値ポートフォリオ分析結果 (20%以下部分)

図7を見ると、「資格」「充実」「就職」「近い」「先生」「安い」「英語」「海外」といったイメージが45度ラインの左上側、すなわち「まだ満たされていない」と考えられる領域に位置していることがわかる。このことから、今後大学が多くの受験生を獲得するために為すべきこととしては、「資格がとれるカリキュラムの充実」「良い就職の斡旋・確保」「通学のしやすさ」「良い教員の採用」「割安感のある授業料」といった価値を高めることが必要と考えられる。

更に、便宜的に、どの程度の価値があるか（ちょうどした満足な状態を100としたとき、相対的にどの程度の価値になるか）を表すCS指数と、就実大学経営学部のイメージと理想の大学のイメージとのギャップを算出した。ギャップは単純に引き算（就実% - 理想%）、CS指数はエクセルの関数を使って、

$$\text{CS指数} = \text{ATAN}(\text{理想}\% / \text{現状}\%) / \text{ATAN}(1) * \text{SQRT}(\text{現状}\%^2 + \text{理想}\%^2) / \text{SQRT}(2)$$

によって求めた。この数式の意味するところは、散布図のプロットポイントを価値ベクトルと考えたとき、「傾き」をアークタンジェントで求め、「長さ」を三平方の定理から求め、その両者のいずれにも比例関係を持つように「積」を総合指標とし、100%の人が現状でも理想でもイメー

ジするポイント（現状100%、理想100%）の価値ベクトルで基準化し、%表示に変換したものである。算出結果を（表1）に示す。

表1. 価値ポートフォリオ分析結果

ワード	現状(%)	理想(%)	就実(%)	CS指数	理想とのギャップ
近い	0.0	3.5	5.5	108.3	2.0
綺麗	0.0	3.2	9.8	91.0	6.6
先生	0.0	2.9	3.8	75.2	0.9
良い	0.0	2.9	4.0	75.2	1.2
安い	0.0	2.6	0.6	60.9	-2.0
自由	58.1	55.5	18.2	55.2	-37.3
活発	0.0	2.0	0.0	36.9	-2.0
距離	0.0	2.0	0.0	36.9	-2.0
有利	0.0	2.0	0.6	36.9	-1.4
バイト	45.4	33.5	8.4	32.3	-25.1
豊富	0.0	1.7	0.6	27.1	-1.2
留学	0.0	1.7	13.0	27.1	11.3
専門	33.2	24.0	23.4	23.1	-0.6
サークル	32.1	23.4	3.8	22.5	-19.7
学ぶ	25.7	22.5	14.2	22.2	-8.4
知識	24.6	22.0	19.9	21.6	-2.0
家	0.0	1.4	0.0	18.8	-1.4
環境	0.0	1.4	0.0	18.8	-1.4
強い	0.0	1.4	3.8	18.8	2.3
校舎	0.0	1.4	4.3	18.8	2.9
学べる	9.2	15.9	13.0	17.3	-2.9
楽しい	12.4	15.9	10.4	16.5	-5.5
充実	2.6	11.3	10.4	14.0	-0.9
勉強	19.4	14.5	6.9	13.9	-7.5
資格	5.8	12.1	5.2	13.6	-6.9
英語	0.0	1.2	34.7	12.0	33.5
管理	0.0	1.2	0.0	12.0	-1.2
教育	0.0	1.2	4.0	12.0	2.9
広い	0.0	1.2	0.6	12.0	-0.6
趣味	0.0	1.2	0.0	12.0	-1.2
少人数	0.0	1.2	9.0	12.0	7.8
制度	0.0	1.2	0.0	12.0	-1.2
大学	0.0	1.2	0.9	12.0	-0.3
仲	0.0	1.2	0.0	12.0	-1.2
就職	1.7	7.8	3.8	9.7	-4.0
子	0.0	0.9	0.0	6.8	-0.9
意識	0.0	0.9	0.0	6.8	-0.9
海外	0.0	0.9	0.9	6.8	0.0
外	0.0	0.9	0.9	6.8	0.0
学生	0.0	0.9	0.0	6.8	-0.9
後輩	0.0	0.9	0.0	6.8	-0.9
合う	0.0	0.9	0.0	6.8	-0.9
清潔	0.0	0.9	1.4	6.8	0.6
生徒	0.0	0.9	0.9	6.8	0.0
積極	0.0	0.9	0.0	6.8	-0.9
先輩	0.0	0.9	1.2	6.8	0.3
多く	0.0	0.9	0.0	6.8	-0.9
部	0.0	0.9	0.0	6.8	-0.9
分かる	0.0	0.9	0.0	6.8	-0.9
明るい	0.0	0.9	1.2	6.8	0.3

表1より、大学において最も大きな価値は「近い」ということであることが考えられた。次いで「綺麗」「先生」「安い」「自由」「留学」「専門」「サークル」といった価値が続いた。先の価値

ポートフォリオから導出された「資格」「就職」といったイメージが上位に上がらなかったのは、CS指数の算出方法が価値ベクトルの長さに比例するため、それらをイメージする学生が少なかったことに起因しているが、高校を卒業したばかりの新入生へのアンケートである事から、まだ「資格」「就職」といった社会に出る際に意識されるような問題に思いが至っていないものと考えられる。

また、就実大学経営学部は、CS指数上位の価値については比較的高く評価されており、「近い」「綺麗」では理想を上回っていた。また、上位の価値ではないものの、「留学」「英語」でも大きく理想を上回っていることがわかった。このことから、就実大学経営学部をPRする際には、「通学のしやすさ」や「施設の綺麗さ」を前面に押し出し、加えて「英語に力を入れている」「留学もできる」ことをアピールすると良いと考えられる。

【結論】

私立大学の経営環境改善に資するべく、多くの受験生を獲得するために大学はどのような価値を備えるべきかを、テキストマイニングを用いて可視化・定量化したところ、アンケートの共起分析・対応分析・価値ポートフォリオ分析より、

- ・就職が良い
- ・資格が取得できる
- ・友達と良好な関係が築ける

が「理想の大学」として必要であり、

「就実大学経営学部」と「理想の大学」の両方に共通に持たれやすいイメージとしては、

- ・設備が充実している
- ・良い先生が居る
- ・家から近い

があることがわかった。

また、大学が備えるべき未だ十分には満たされていない価値としては、

- ・資格がとれるカリキュラムの充実
- ・良い就職の斡旋・確保
- ・通学のしやすさ
- ・良い教員の採用
- ・割安感のある授業料

が分析されたことから、こういった価値を高める努力を怠らないことが重要と考えられ、

- ・通学のしやすさ
- ・施設の綺麗さ
- ・英語に力を入れている
- ・留学できる

については、理想を上回る評価が得られるものと考えられることから、就実大学経営学部の強み

となることが示唆された。

【おわりに】

本研究は、就実大学経営学部に入学者を対象に行ったアンケート結果に基づいたものである。つまり、「自己の大学選択を正当化したい」という認知バイアスの存在を否定できない状況での価値分析となっている。このため、一般的な大学の備えるべき価値が分析できたとは言うべきでない。しかし、特定の大学の「学生確保に向けた対応策を立案し着実に実施する」ための実践可能な大学の価値の可視化・定量化方法としては、一定の可能性を提供できたものと考えられる。今後も同様のアンケートを継続し、時系列的な分析方法を導入することで、大学に求められる価値の時代による変化の把握や、実際に行った対応策の効果検証も行っていきたい。

【引用文献】

- Dohkan,K., Hayashi,T., Masuda,M. and Fukuchi,Y. (1999) An Application of Kansei Engineering to a Cosmetic Product— Application of Dohkan Method—, Kansei Engineering II —Human sensibility ergonomics—, Edited by Soon Yo Lee, p.13-22, Ingankyngyungsa
- 真柳麻誉美、林俊克、平野広隆（2001）「非定型自由記述法と定型自由記述法の比較」『日本行動計量学会大会発表論文抄録集』 29, 106-107
- 文部科学省高等教育局私学部参事官付（2018）『私学行政の現状と課題等について』
- 林俊克（2001）「感性工学と化粧品開発」『FRAGRANCE JOURNAL』 2001/APRIL 4
- 林俊克（2002）『Excelで学ぶテキストマイニング入門』 オーム社

【参考文献】

- 「特集テキストマイニング」『日本人工知能学会誌』 2001／ Vol.16 No. 2
- M.J.A.ベリー&G.リノフ（1999）『データマイニング手法』 海文堂
- R.バゼル. B.ゲイル（1988）『新PIMSの戦略原則 —業績に結びつく戦略要素の解明』 ダイアモンド社
- トム・ピーターズ（1995）『トム・ピーターズの経営創造』 TBSブリタニカ
- 関根聡（1999）「テキストからの情報抽出 —文書から特定の情報を抜き出す—」『情報処理学会誌』 Vol.40. No. 4
- 根上生也（1993）『離散構造（情報数学講座）』 共立出版
- 佐藤郁哉（2008）『質的データ分析法 —原理・方法・実践』 新曜社
- 森典彦（1991）『デザインの工学 —ソフトシステムの設計計画』 朝倉書店
- 神田範明（1994）「新製品開発のための商品企画7つ道具」『品質管理』 Vol.45, No. 7 （財）日本科学技術連盟
- 神田範明（1995）『商品企画7つ道具』 日本科学技術連盟出版社

- 杉山和雄、井上勝雄（1996）『EXCELによる調査分析入門』海文堂
- 石田基広（2017）『Rによるテキストマイニング入門』森北出版
- 石田基広・金明哲 編著（2012）『コーパスとテキストマイニング』共立出版
- 大澤幸生（1999）「KeyGraph：単語共起グラフの分割統合によるキーワード抽出」『電子情報通信学会論文誌』J82-D1. No. 2
- 朝野熙彦 偏（2001）『魅力工学の実践 —ヒット商品を生み出すアプローチ—』海文堂
- 那須川哲哉、諸橋正幸、長野徹（1999）「テキストマイニング —膨大な文書データの自動分析による知識発見—」『情報処理学会誌』Vol.40. No. 4
- 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』
- 樋口耕一（2012）「社会調査における計量テキスト分析の手順と実際 —アンケートの自由回答を中心に—（第10章）」「今日から始めるテキストマイニング —計量テキスト分析の環境『KH Coder』—」
- 樋口耕一（2014）『社会調査のための計量テキスト分析 —内容分析の継承と発展を目指して』ナカニシヤ出版
- 樋口耕一（2018）KH Coder 3 リファレンス・マニュアル
- 樋口正美（2000）『経営品質を高めるCS（顧客満足）調査法』品質月刊委員会
- 末吉美喜（2019）『テキストマイニング入門：ExcelとKH Coderでわかるデータ分析』オーム社