

# テキストマイニングのアンケート解析への応用の試み

小棹 理子<sup>a</sup> 石田 英弥<sup>b</sup>

<sup>a</sup>湘北短期大学 <sup>b</sup>早稲田大学大学院博士前期課程

## 【抄録】

テキストデータを対象とするテキストマイニング手法を、本学のオープンキャンパス来場者（高校生）に対するアンケート調査の回答部分のうち、感想・コメントなど、フリーコメント部分の分析に適用することを試みた。参加生徒は、いずれの学科においても「楽しく」「良い雰囲気」、と答え、ポジティブ評価であるが、学科により多少印象が異なることがわかった。また、6月時点の感想「また来たい」、が7月では見られなくなったことが興味深い。

## 【キーワード】

テキストマイニング    オープンキャンパス    単語頻度解析    特徴語抽出

## 1. はじめに

最近消費者やユーザーの動向を探る方法として、従来のアンケートのような選択肢を用意した質問票を用いないスタイルが増えている<sup>[1]</sup>。質問票で得られた結果とは異なった行動をとる消費者が少なくないためである。そこで、ブログやコミュニティーサイトなど、いわゆる口コミサイトに現れる消費者の意識や考え方をテキストデータから分析する手法が登場している。そのような、「文章の中に埋もれている共通性や相関関係を発掘し、有用な知見を獲得する分析手法」は、テキストマイニング<sup>[2]</sup>と呼ばれ、アンケート調査などにおいてフリーコメント部分を客観的に、かつ効率よく分析したい、というニーズに応える手法として

1990年代から利用されはじめた。

筆者らは、データマイニングを用いて、データベースのアクセスログの解析を行い、アソシエーション分析によりエコ技術に関する一般動向を探った<sup>[3]</sup>。また、同手法を湘北短期大学のWebサイトのアクセスログに適用することにより、サイト閲覧者の興味の違いが読み取れることを示した<sup>[4]</sup>。そこで今回は、データマイニングの中でもテキストデータを対象とするテキストマイニング手法をアンケートのフリーコメント部分の分析に適用することを試みた。アンケートデータとして、①全国高校のキャリア教育担当者から得られたフリーコメント、ならびに②オープンキャンパス来場者の感想・コメント、を用いた。①を用いた目的は、大学側では明確になっていない、本学に対する印象や高校における問題点を探るためであり、②は、学科ごとにニーズや印象をさぐり今後の広報活動の一助とするためである。

---

<連絡先>

小棹 理子    ozao@shohoku.ac.jp

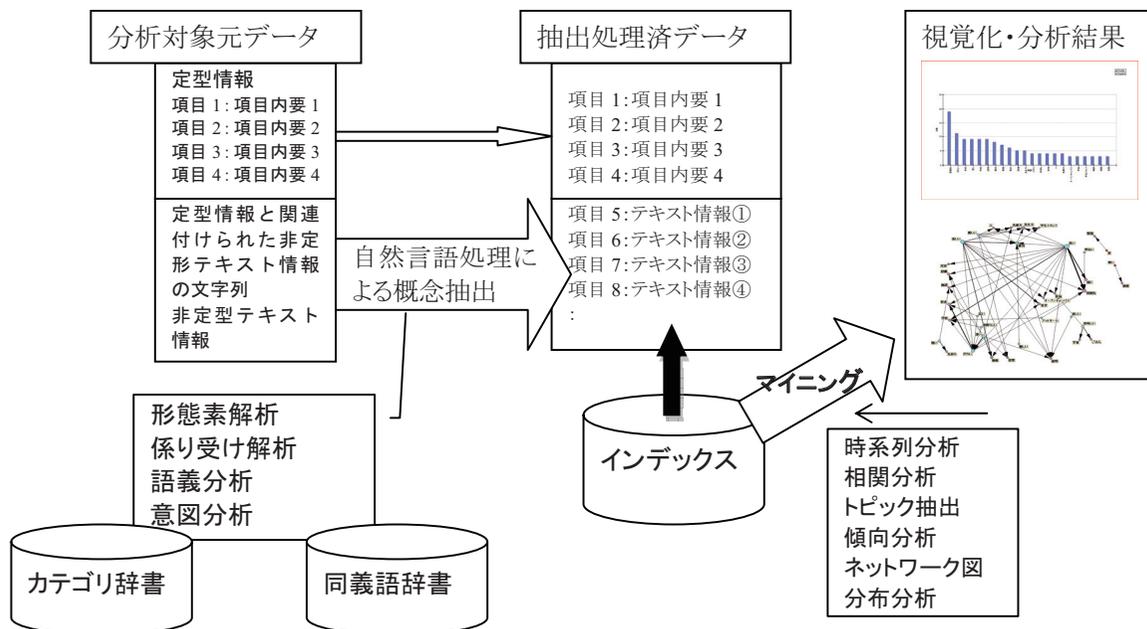


図1 テキストマイニングの処理の流れ

## 2. テキストマイニングの手順

全体的な処理の流れを図1に示す(「テキストマイニングを使う技術/作る技術」<sup>[5]</sup>を参考に作成)。自然言語処理における「自然言語」とは、通常われわれが使う言語である。情報処理の分野において「プログラミング言語」と区別するために用いられている。自然言語をコンピュータで処理するための基本技術として形態素解析と構文解析がある。形態素解析は、文字列を文法的に意味のある単位の構成要素に分割し、各要素の文法的素性(品詞など)を決定する。構文解析は、要素間の文法的・意味的關係を示し、係受け構造を見出だす。このような概念抽出を行うことによって新たな定型情報が得られることになる。この結果、多次元解析が可能になり、データマイニングの一種であるバスケット分析(相関ルールの発見)が可能になる。

なお、解析ソフトとして(株)数理システムのTextMiningStudio3.1を用いた。同ソフトでは、Chasenを形態素解析に組み込んでいるが、ネッ

トワークグラフの描画に定評がある。本ソフトを用いて以下を行った。

- ① 基本情報分析
- ② 単語頻度解析
- ③ 係り受け頻度解析
- ④ 注目語分析
- ⑤ 特徴分析
- ⑥ 評判抽出
- ⑦ ことばネットワーク

以上に加え、類義語辞書登録、原文参照を活用した。

## 3. 分析結果

### 3.1 全国高校へのアンケートのフリーコメント分析

キャリア教育に関する全国高校のキャリア教育担当者から得られたアンケート回答のフリーコメント部分を分析対象とした。アンケート<sup>[6]</sup>を5211校に送付し、753通の回答が得られたが、うちフ

リーコメントの記述は252件であった。

フリーコメント部分の単語頻度解析結果を表1に示す。圧倒的に大学が多い。このアンケートのフリーコメント部分に至る設問では、高校における問題点を問うものであったので、フリーコメントでは大学の問題を指摘するコメントが増えたものと思われる。図2に、「大学」で注目語解析した結果を示す。大学へ進学する際に無目的である、大学教育が卒業後に結び付かない、選ばなければ大学入学は可能、など、図3のネットワーク図に示すようにさまざまな問題が明らかとなった。

表1で3番目と12番目のAO入試と推薦入試に関して原文参照すると、これらの入試制度があるために高校での授業が成り立たなくなるなどのネガティブなコメントが多く見られた。

定型属性が多く、内容を絞り切れなかった点が反省点である。

表 1 単語頻度解析（全国高校向けアンケート）

単語	品詞	頻度
大学	名詞	65
生徒	名詞	55
AO入試	名詞	34
多い	形容詞	31
進学	名詞	23
基礎学力	名詞	18
必要	名詞	17
学力	名詞	16
入学	名詞	15
安易	名詞	14
高校	名詞	14
推薦入試	名詞	14
入試	名詞	14
問題	名詞	14
指定校	名詞	13
早い	形容詞	12
合格	名詞	11
少ない	形容詞	11
増える	動詞	11
本校	名詞	11

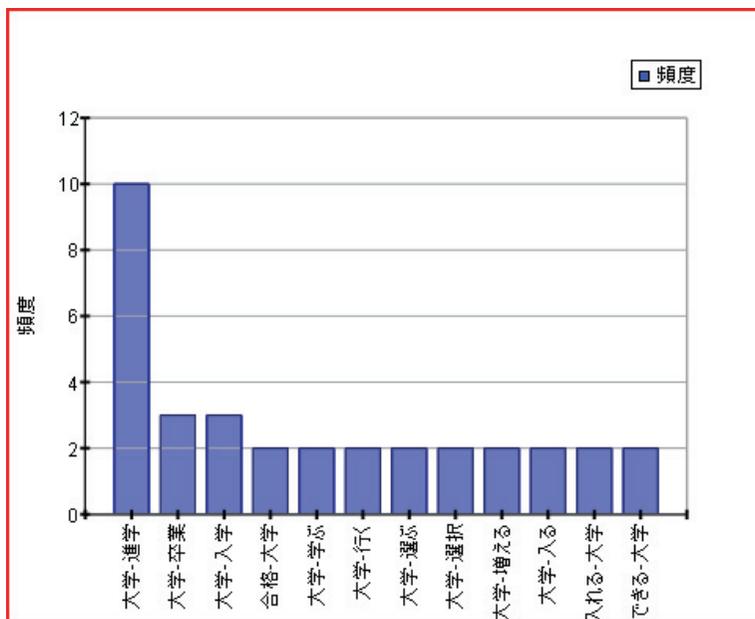


図 2 注目語（「大学」）解析

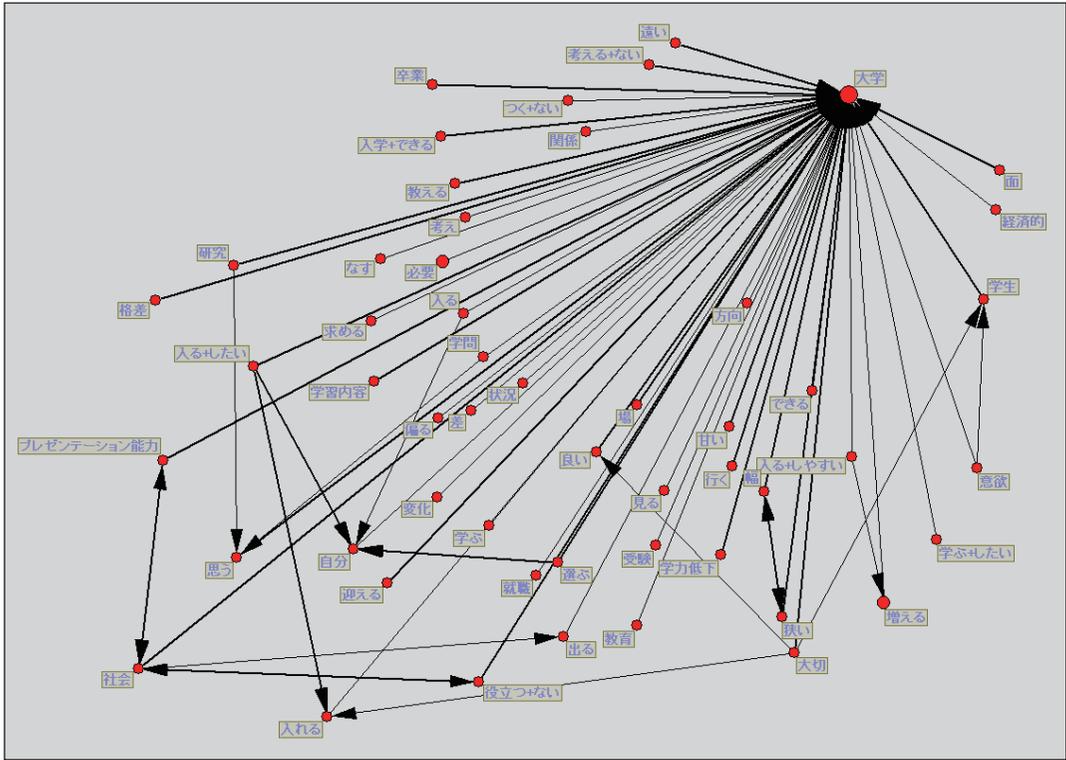


図3 注目語「大学」の注目語情報（共起抽出）

3.2 オープンキャンパスアンケートのフリーコメント分析

オープンキャンパス（OC）に来場した高校生のアンケートのフリーコメント部分の解析を行っ

た。アンケート回答は、(A) 2008年6月7日と (B) 2008年7月28日（模擬授業を実施）に得られたものである。学科別内訳を示す。(B) では、来場者の志望状況も示した。

表2 アンケート回答を得たオープンキャンパス実施日（2008年）

実施日	6月7日(土)		7月28日(月)				
	来場者	来場者	第一志望	出願したい	出願の方向	迷っている	検討中
情報メディア	23	32	12	5	4	0	11
生活プロデュース	38	52	28	3	2	1	16
総合ビジネス	56	58	37	3	6	3	9
保育	99	88	42	11	6	10	18
不明	8	12	4	3	1	0	4
合計	224	242	123	25	19	14	58

### 3.2.1 単語頻度解析

単語頻度解析はテキストマイニングの基本である。表2に、(A)、(B)の解析結果を示す。品詞ごとの出現回数からみると、比較的動詞が多い。このような場合は、行動自体がデータ中で重要なことがある。6月7日では生活プロデュース学科と保育学科の「わかりやすい」、保育学科の「来たい」が目をはく。形容詞が多いのは感性的な文章であることの表れであることが多いが、保育学科の「凄い」は突出している。

図2.3に、人数で規格化した頻度分布を示す。6月と7月を比較するため、6月の頻度トップ20語を用いている。「良い」「楽しい」は二強であり、大きな変化はない。

図2を参照すると、保育学科では「凄い」の頻度が飛びぬけて高く、逆に情報メディアでは皆無であることがわかる。「凄い」には、「すごい」「すごい」「すごく」などが含まれている。原文を参照すると、「すごい」は「勉強・参考になった」、「わかりやすい」「きれい」「良い」「整っている」などにかかっている場合と、「設備」「ピアノ」「造形」「環境」などが「すごい」、の場合があるが、いずれにしても見学者の強い感情の表れであることが示唆される。また、生活プロデュース学科では、「説明」が「わかりやすい」ことが特徴的であったことが明らかである。情報メディア学科では「設備」が突出しているが、「くわしい(=委しい、詳しい、詳細、などを含む)」と「親切」がまったく表れないことがわかる。総合ビジネス学科は「先生方(先生)」に対する良い印象が強い。

図3の7月を参照すると、「来たい(来る+したい)」が減ったことがわかる。「来る」は、「来てわかった」「はじめて来た」「友人と来た」「来てみてわかった」、とオープンキャンパスそのものへの参加を意識しているのに対し、「来る+したい」は、「また来たい」「ぜひ来ます」「来てよかった」など

のコメントが含まれている。6月では、「来たい」はとくに保育に顕著であったが、7月になって減ったのは、すでに本学への出願を決定している生徒が多かったのであろうか。また、6月、7月を通して、総合ビジネスでは「先生方」への評価が良い。

上述のとおり、本学に対して「良いイメージ」で語られていることは明らかであるが、評判抽出した結果をことばネットワークで示す(図4)。ことばと属性の共起分析よりイメージ分析が可能である。矢印で相関ルール(組み合わせ)をあらわすが、太さがサポート(支持度)を示す。「良い」→「雰囲気」「感じ」「学校」であり、感覚的で具体的に何か、とは分析しづらい。「楽しい」は、「楽しく」→「勉強になった」、「親切で」、といった感謝の気持ちも表れている。

表3 単語頻度解析 (学科別)

(A) 2008年6月7日						
単語	品詞	情報 メディア	生活 プロデュース	総合 ビジネス	保育	文字列 欠損
楽しい	形容詞	8	6	21	35	3
良い	形容詞	7	10	16	28	1
説明	名詞	5	14	13	17	1
凄い	形容詞	0	6	8	29	0
わかる+しやすい	動詞	3	9	9	19	1
来る	動詞	6	4	10	16	0
優しい	形容詞	1	4	5	17	0
学校	名詞	3	4	6	12	1
オープンキャンパス	名詞	4	4	6	10	0
わかる	動詞	2	5	4	12	0
きれい	名詞	1	2	4	14	0
湘北	名詞	1	5	4	10	0
雰囲気	名詞	1	3	7	8	0
来る+したい	動詞	1	5	2	11	0
思う	動詞	3	2	5	7	0
美味しい	形容詞	3	5	5	4	0
委しい	形容詞	0	2	4	8	0
校舎	名詞	1	2	3	7	0
親切	名詞	0	1	3	9	0
先生方	名詞	1	1	7	4	0
(B) 2008年7月28日						
単語	品詞	情報 メディア	生活 プロデュース	総合 ビジネス	保育	文字列 欠損
良い	形容詞	17	28	37	69	3
楽しい	形容詞	14	17	42	67	5
説明	名詞	8	22	25	31	2
生徒	名詞	6	13	28	33	2
凄い	形容詞	3	12	15	48	1
わかる+しやすい	動詞	8	13	21	31	2
来る	動詞	6	10	18	33	2
学校	名詞	6	6	14	22	2
先生	名詞	3	5	19	19	0
わかる	動詞	5	10	12	18	0
オープンキャンパス	名詞	6	6	12	18	1
湘北	名詞	5	8	11	17	1
優しい	形容詞	3	11	7	21	0
きれい	名詞	3	8	9	21	0
雰囲気	名詞	2	8	11	14	1
校舎	名詞	3	7	8	12	0
勉強	名詞	2	6	9	10	1
模擬授業	名詞	2	5	7	12	0
参加	名詞	6	5	3	11	0
聞く+できる	動詞	1	4	6	13	1

テキストマイニングのアンケート解析への応用の試み

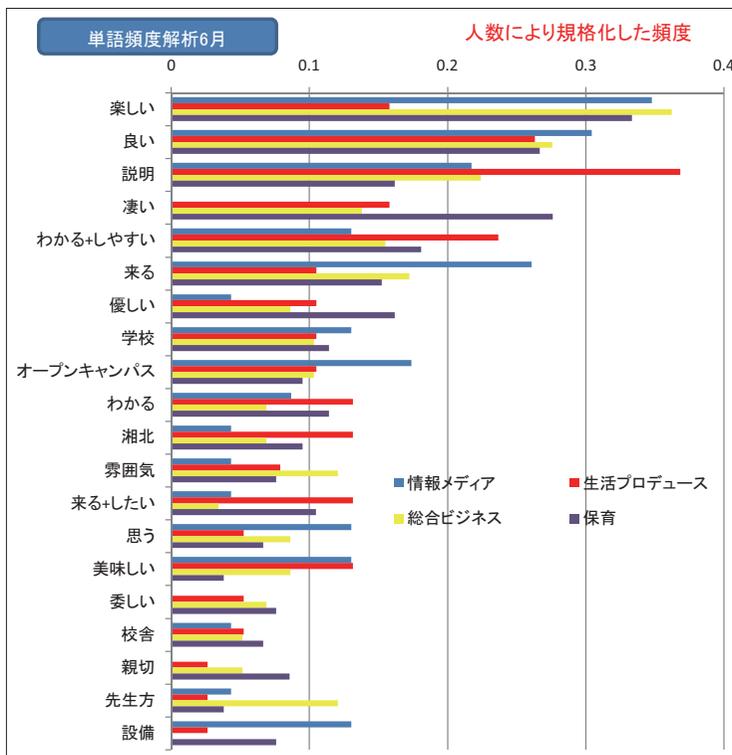


図 4 単語頻度解析（学科別、人数で規格化6月）

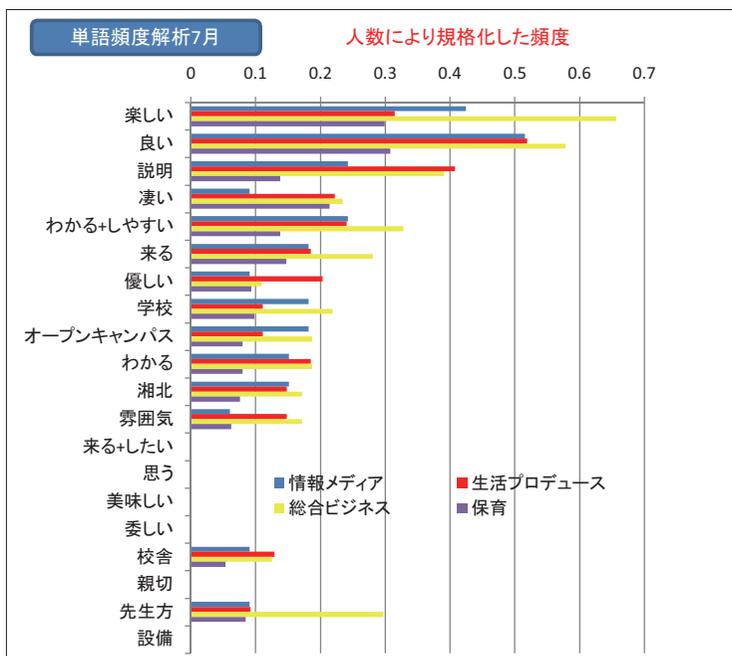


図 5 単語頻度解析（学科別、人数で規格化7月）



表 5 特徴分析 (学科ごと)

学科-E	指標値	学科-L	指標値	学科-C	指標値	学科-P	指標値
ペーパークラフト	12.062	作る+できる	14.222	相手	16.621	先生	20.636
楽しい	11.644	簡単	13.749	考える	14.448	楽しい	19.605
やる+したい	6.725	作る+したい	10.666	内容	8.404	体	10.761
フォトショップ	6.031	コースター	7.111	勉強	8.404	授業	10.37
使う+したい	6.031	良い	6.729	心	7.555	面白い	7.645
家	5.368	作る	5.986	気持ち	6.044	受ける+したい	7.327
使う	5.368	可愛い	5.333	心理学	6.044	凄い	6.936
作る	4.704	不器用	5.333	人間関係	6.044	動く	5.38
凄い	3.439	ていねい	4.771	分かる+しやすい	6.044	明るい	5.38
体験+できてほしい	2.684	優しい	4.208	説明	5.382	過ごす+できる	4.035
知る	2.684	うまい+できる	3.555	自分	4.908	気分	4.035
分かる+ない	2.684	コースター+できる	3.555	わかる	4.533	子ども	4.035
やさしい	2.352	スタッフ	3.555	大切	4.533	笑顔	4.035
やる	2.352	雑	3.555	分かる	4.533	模擬授業	3.292
楽しい+できる	2.352	笑	3.555	変わる	4.533	やる+したい	3.008
機会	2.352	面白い	3.263	難しい	4.059	楽しむ+できる	2.69
好き	2.352	教える	2.993	学ぶ+したい	3.871	元氣	2.69
ていねい	2.02	好き	2.993	授業	3.385	講師	2.69
わかる+しやすい	2.02	家	2.43	コミュニケーション	3.022	自然	2.69
学ぶ+したい	2.02	おしえる	1.215	ロールプレイング	3.022	実感	2.69
模擬授業	2.02	かんたん	1.215	学ぶ	3.022	受ける+できる	2.69
		ユニーク	1.215	学ぶ+できる	3.022	小さい	2.69
		学生さん	1.215	気	3.022	入学	2.69
		簡単そう	1.215	言う	3.022	遊ぶ+できる	2.69
		驚く	1.215	言葉1つ	3.022		
		参加+したい	1.215	考える+できる	3.022		
		手	1.215	行動	3.022		
		多い	1.215	短い	3.022		
		体験+できてほしい	1.215	話	3.022		
		体験+できる	1.215				
		分かる+ない	1.215				
		話す	1.215				

情報メディア学科では、模擬授業で「ペーパークラフト」を作成したことが「楽しかった」ことがわかり、順当に抽出が行われていることがわかる。生活プロデュース学科でも「簡単に」「作れた・作りたい」「コースター」、ときちんと模擬授業の内容とそれに対する感想が抽出されていることが

わかる。総合ビジネス学科では「相手」を「考える」ことを学んだことがわかる。保育も「先生」と「体」を使った「楽しい」「授業」が展開されたことがわかる。

以上よりテキストマイニングで適切に特徴抽出がなされていることがわかる。

※ 補完類似度：フィルタをかけて抽出した指定の下限頻度以上の単語  $w$  の集合  $W = \{w_i\}$  が、取りうる属性  $R = \{r_j\}$  を持つとき、単語  $w_i$  について補完類似度は次式で表される：

$$\frac{ad - bc}{\sqrt{(a+c)(b+d)}} \quad (\text{ただし全単語数 } N = a+b+c+d)$$

	属性 $r_j$	属性 $r_j$ 以外
単語 $w_i$	a	b
単語 $w_i$ 以外	c	d

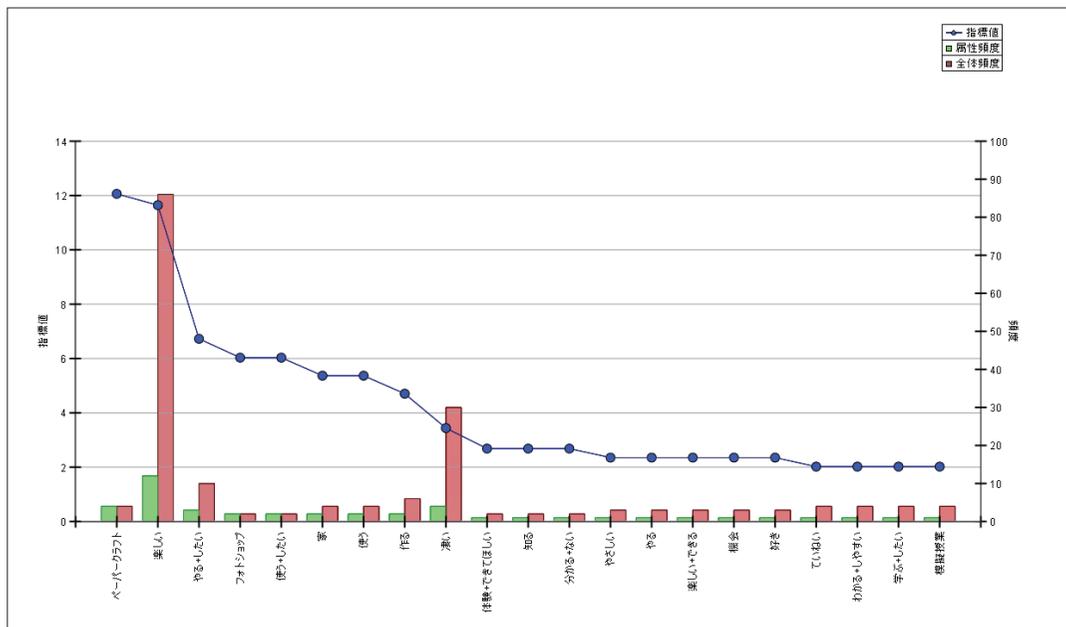


図7 特徴語抽出 (情報メディア学科)

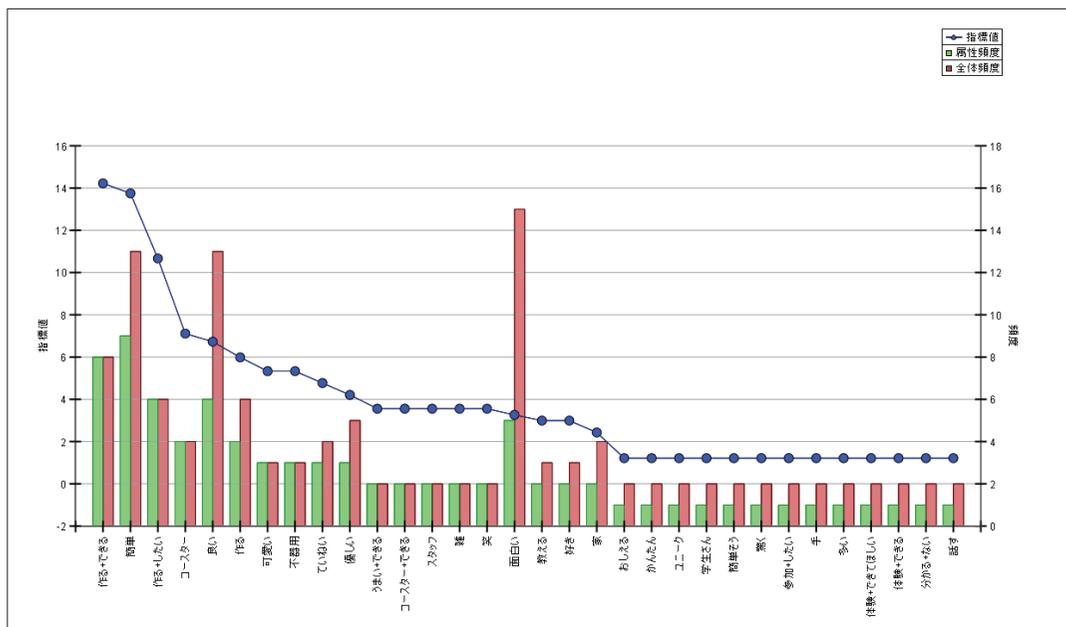


図8 特徴語抽出 (生活プロデュース学科)

テキストマイニングのアンケート解析への応用の試み

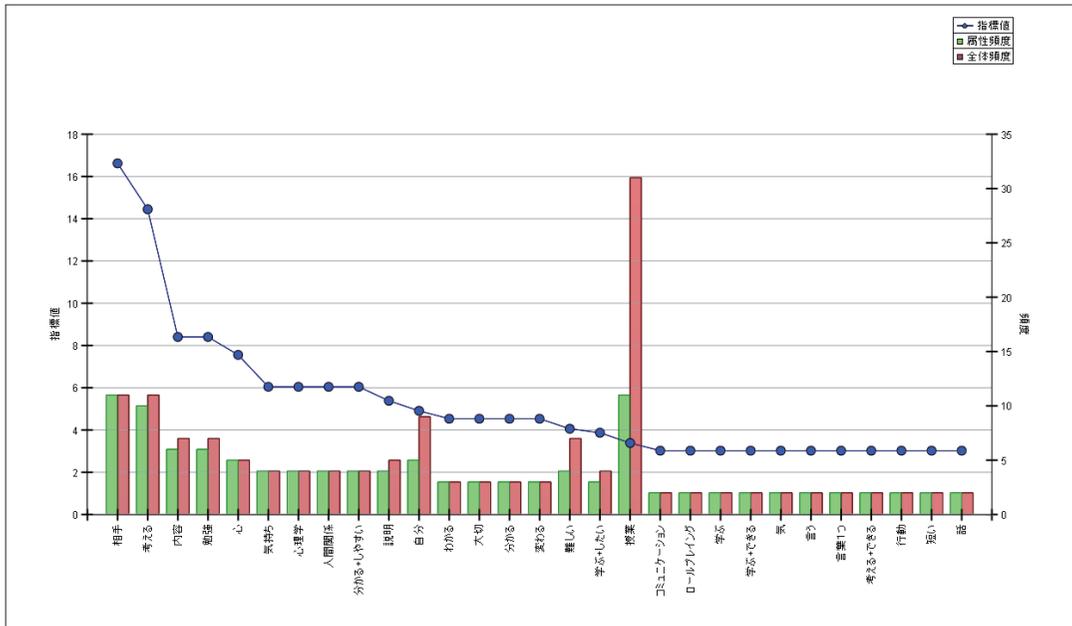


図9 特徴語抽出 (総合ビジネス学科)

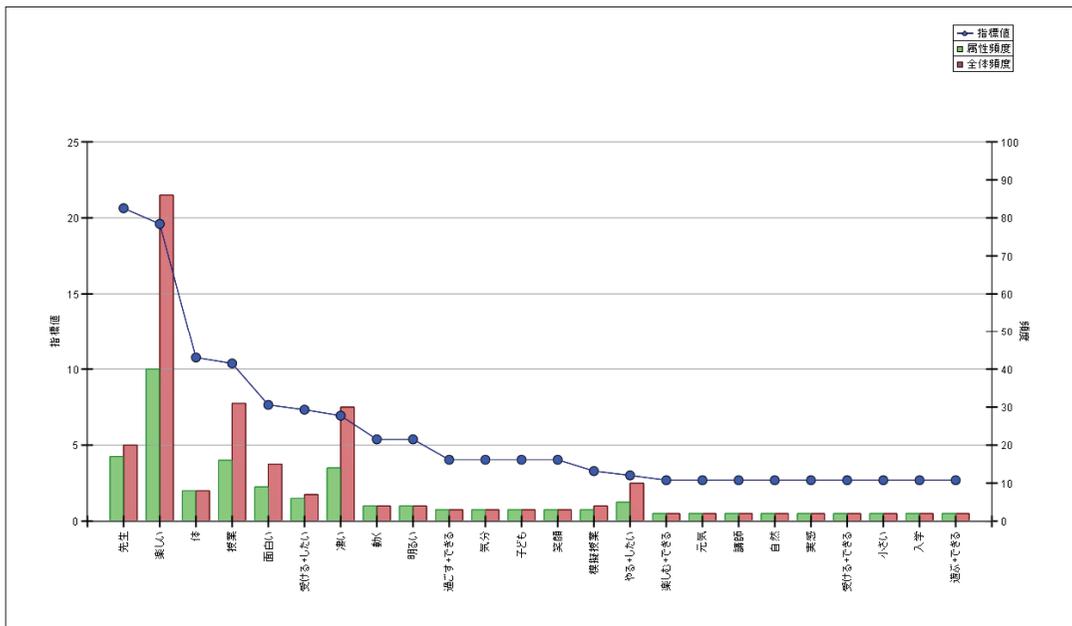


図10 特徴語抽出 (保育学科)

#### 4. おわりに

テキストマイニングを導入し、本学で蓄積されているキャンパス見学会時に取得しているアンケートデータ等のマイニングを行った。同研究の目的のひとつであるテキストマイニングの有効性の確認を行うことができた。

また、キャンパス見学に来校する生徒の動向の分析を試みた。参加生徒は、いずれの学科においても「楽しく」「良い雰囲気」と答え、ポジティブ評価であるが、学科により多少印象が異なることがわかった。アンケートの定型感想の選択肢である「とても楽しかった」「とても勉強になった」「学生スタッフがとても感じがよかった」「校舎がきれいだった」「先生方がとても親切だった」「今日来て本当によかった」「説明が分かりやすくて面白かった」などがフリーコメントで繰り返されているが、学科間で差異が出るのは興味深い。

6月での感想「また来たい」、が7月では見られなくなったことが興味深い。アンケート項目として満足度調査を追加するのも可能であろう。

テキストマイニングを推移分析に応用することも可能である。時間軸に対し変化を追うのも興味深いと思われる。

#### 謝辞

平成20年度湘北短期大学助成金を得て実施された研究結果の一部を報告したものです。ここに謝意を表します。

#### 【参考文献】

- 1 萩原雅之、「ネット調査新手法」日経産業新聞(2009年2月3日)4面.
- 2 マイケル・J.A. ベリー & ゴードン・S. リノフ(著)、江原 淳、上野 勉、河野 順一、佐藤 栄作、朝稲 努(訳)、「データマイニング手法—営業、マーケティング、CRMのための顧客分析」海文堂(2005)
- 3 R. Ozao, T. Sawaguchi, H. Ishida, M. Iji, T. Furuyama, Y. Shinohara, K. Yamada, K. Halada, "Eco-MCPS: A newly developed web-based database for eco-materials in Japan", Mater. Trans, 48 (12) pp. 3043-3049 (2007) .
- 4 小椋理子・石田英弥、「湘北短期大学ウェブサイトのアクセスログから見られる閲覧者の各学科への興味の違い」、湘北紀要 29、pp1-6 (2008)
- 5 那須川哲哉、「テキストマイニングを使う技術／作る技術」(東京電機大学出版局(2006)
- 6 小椋理子・伊藤善隆・田村新吾・岩崎敏之・藤澤みどり・高橋可奈子・原満・住谷勉・佐藤明宏・小林久美子・石田英弥、「高大連携による教育交流ネットワークの構築—コミュニケーション教育研究会の活動とコミュニケーションリテラシー—」、湘北紀要 30 (印刷中)

## Text Mining for Public Relations

OZAO Riko ISHIDA Hideya

### **[abstract]**

Text data mining was applied to free comments obtained as text data from open campus participants. Almost all high school students answered they had good impressions, but the most impressed points differed from a department to another. Furthermore, the sentiment of the participants seemed to change from June to July.

### **[key words]**

Text data mining, text clustering, concept/entity extraction, sentiment analysis, public relations

