

# 口コミデータを用いた観光スタイルと観光行動の関係分析

プロフィットエンジニアリング研究

5220F013-0 後藤孝輔  
指導教員 大野高裕

## A Relationship Analysis between Sightseeing Style and Behavior Using Word-of-Mouth of Tourists

Kousuke Goto

### 1. 研究背景と目的

現在、消費者の購買行動はモノ消費からコト消費に変化し、その代表例である観光業界の市場規模は拡大している[1]。そこで観光地や旅行代理店では、観光マーケティングに取り組み、さらなる観光客獲得のために観光ニーズに寄り添った観光プランの提案が必要とされている。そのためには、どのような観光客が、どのような観光行動をとっているかといった構造を把握する必要がある。ここで、どのような観光客が存在するか検討するためには、年齢や性別といったデモグラフィック属性の把握だけでは不十分である。より詳細に観光客を理解するために性格や価値観といった心理的属性を把握する必要がある。また、観光行動の構造に関して野守ら[2]が観光テーマ（観光動機や目的）と選択観光地の関係を検証している。そこでは、観光客が持つ観光テーマによって選択される観光地が異なることを示している。このことから、観光行動は目的である「観光テーマ」と、それに対する行動である「観光地選択」の二つで構成されると考えられる。以上をまとめると、把握すべき構造は、どのような心理的属性を持つ観光客が、どのような観光テーマを持ち、どの観光地を選択しているかという構造である。

上記構造を把握するにあたって、選択観光地は外形的に明らかである。一方で、心理的属性と観光テーマは本人の内面にあり、本人が明らかにしない限り把握が困難である。しかし、現在では観光客が口コミ・SNS上に投稿を行うため、この投稿からそうした内面的情報を得られる可能性がある。これに関して、観光テーマは野守ら[2]が口コミデータを用いてこれを把握する研究を行っている。一方で、心理的属性を把握する研究は見られない。というのも、口コミ・SNS投稿に観光客の性格や価値観が直接書かれることが少ないことや、分析のために一人の複数投稿が必要であることが理由として挙げられる。しかし、Engelら[3]の研究よれば心理的属性は観光スタイルという観光行動タイプに集約できると考えられ、観光スタイルであれば口コミ・SNSデータから得られる可能性がある。

上記背景を基に本研究の目的を述べる。観光地や旅行代理店が観光マーケティングを行うためには、心理的属性、観光テーマ、選択観光地の因果関係を把握する必要がある。しかし、口コミ・SNSデータから直接心理的属性を把握することは困難である。そこで本研究は心理的属性に起因する観光スタイルの把握を試みる。すなわち本研究では口コミデータから観光スタイル、観光テーマの抽出を行った上で、これらと観光地選択の3者の因果関係を分析する。

これによって、観光客の詳細理解と、効果的な観光プランの策定に貢献することを本研究の目的とする。なお本研究は、観光サイトとして著名であり口コミが豊富に集まる株式会社リクルートが運営する「じゃらん」[4]の口コミデータを扱う。

### 2. 従来研究と本研究の特徴

#### 2.1. アンケートを用いた従来研究

観光マーケティングに関してアンケートを用いた従来研究について述べる。石井ら[5]は、アンケート調査結果から有効な知見を導き出すための手法を提案している。具体的には、アンケート結果の分析に従来の基礎統計分析に加えて、テキストマイニング手法を適用している。これによって、調査項目間の関係性を可視化し、従来では得られなかった観光ニーズの把握に貢献している。また原田[6]は、内面的要因である「心理的属性」、「観光動機」を含めた観光ニーズを把握するための包括的仮説モデルを構築して、アンケート調査と統計的分析によって実証している。このように、アンケートを用いた観光ニーズ・観光客理解に関する研究は多数行われている。その中で、心理的属性を用いてより詳細にそれらを捉えようとする取り組みは一般的になってきている。

#### 2.2. 口コミ・SNSを用いた従来研究

観光マーケティングに関して口コミ・SNSを用いた研究について述べる。竹岡[7]は口コミデータからテキストマイニングによって抽出された消費者の体験と、施設の来場者数、延床面積などの外形的データを総合的に分析し、観光施設の特徴を可視化している。また、野守ら[2]は口コミデータから観光客が話題にする観光テーマを抽出し、そのテーマを用いて観光地の特徴を定量的に評価している。このように、観光地の特徴を定量的に評価、可視化する研究は多数行われている。一方で、投稿者自身の理解や、心理的属性を把握するといった試みは見られない。

#### 2.3. 本研究の特徴

上記の従来研究を見ると、アンケートを用いた観光客の心理的属性の把握や観光ニーズの抽出といった試みは行われている。しかし、口コミ・SNS投稿データを用いた同様の研究は見られない。これに対して本研究は、口コミデータの分析によって、心理的属性に起因する観光スタイルを網羅的に捉え、これと観光行動との関係を明らかにする。これによって、口コミデータを用いた観光客の詳細理解と観光ニーズの抽出に貢献する。

### 3. 本研究の提案

#### 3.1. 研究対象エリアと観光地の決定

研究対象とする場所の選択に関して述べる。本研究対象地に対する分析実行条件として、①ロコミの数が一定数必要、②多くの種類・ジャンルの観光資源を持つこと、以上の2つが考えられる。これに対して、北海道は魅力が高く、季節を問わず観光客を集めており、ロコミが豊富に集まる。また、自然、建造物、アクティビティなど豊富な種類の観光資源を有している。よって、研究対象エリアは北海道とする。次に、具体的な選択先となる観光地に関して述べる。まず「じゃらん」サイト内で使用されている観光地ジャンルを6つのグループにまとめ、それに北海道の観光地を当てはめた。その結果が表1である。ここで、表1の「食事・買物場所の選択」、「宿泊先の選択」は観光地先での意思決定である。そのため、訪問する観光地の選択とは選択要因が異なると考えられる。また、従来研究でもこれらを同様に扱っていない。よって、「買物・グルメ」、「温泉・宿泊施設」グループは研究対象から外す。そして、残ったグループでロコミ数が上位2位の観光地を研究対象とする。ここで、アクティビティの「旭山動物園」、「円山動物園」は非常に似通ったものであるため、片方の「旭山動物園」のみを研究対象とする。以上によって「函館山」、「青い池」、「小樽運河」、「大通公園」、「旭山動物園」、「北海道神宮」、「博物館網走監獄」を選択する。

#### 3.2. 観光スタイルという概念とその抽出方法

一般生活において、Engelら[3]は心理的属性がライフスタイルという行動様式、傾向に集約されることを立証している。このことから観光では、心理的属性に起因する観光スタイルという観光行動タイプが現れると考えられる。観光スタイルという言葉の定義は、「観光地においてどのように過ごすか」とする。では具体的にどのような観光スタイルが存在するか様々な文献を基に考察すると、①「能動的⇔受動的」、②「個人行動⇔団体行動」、③「計画的な行動⇔その場その場で行動」、④「人と違う行動⇔人と同じ行動」、⑤「楽しさを追及⇔自己成長など行動に意味を追求」、⑥「その他（観光の中で気に入る点）」、以上の6つの軸で観光スタイルが構成されると考えられる。ここで、観光スタイルの違いによって気になる点や、心を揺さぶられる点異なるため、ロコミ投稿時に着目・言及する点が異なることが考えられる。よって、その違いを探求することでロコミから観光スタイルを抽出することが可能であることが考えられる。

#### 3.3. 本研究の提案

本研究ではロコミデータから観光スタイル、観光テーマを抽出して、これらと選択観光地の因果関係を検証することを目的としているが、以下にそのアプローチ方法の全体像を述べる。まず、研究背景で述べたようにロコミデータから直接心理的属性を得ることは困難であるので、本研究では心理的属性に起因する観光スタイルを抽出する。この抽出には、トピックモデルの一つであるLDA手法を用いる。その際に、3.2項で示した6つの軸を基にロコミの着目・言及点を探求することで観光スタイルの抽出を試みる。次に、野守ら[2]の研究より観光行動は観光テーマと観光

表1 北海道の観光地を整理した表

観光地ジャンル	観光地
自然景観	「函館山」、「藻岩山」、「洞爺湖」、「知床五湖」、「青い池」、「モエレ沼公園」、「摩周湖」
人工景観	「小樽運河」、「大通公園」、「札幌テレビ塔」、「五稜郭タワー」
アクティビティ	「旭山動物園」、「円山動物園」
歴史・文化・芸術	「北海道神宮」、「五稜郭公園」、「博物館網走監獄」、「北海道庁旧本庁舎」
買物・グルメ	「六花亭本店」、「かま栄」、「すすきの」、「金森赤レンガ倉庫」、「クランベリー本店」
温泉・宿泊施設	「登別温泉」、「洞爺湖温泉」、「定山溪温泉」

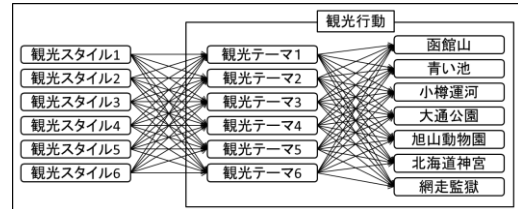


図1 本研究のモデル図

地選択で構成されると考えられることから、本研究ではロコミ投稿データから観光テーマをLDA手法を用いて抽出して、観光地選択は外形的データとしてサイト上からデータ取得を行う。そして本研究では、観光スタイルが観光テーマの醸成に影響を与え、醸成された観光テーマが観光地選択に影響を与えると考える。これをモデル図化したものが図1であり、本研究はこの因果関係の連鎖を検証する。なお、この検証にはベイジアンネットワークを用いる。

### 4. 研究方法

本研究は、①データの取得と加工、②観光テーマ、観光スタイルの取得、③モデル図の検証という手順を踏む。以下にその詳細を述べる。

#### 4.1. データの取得と加工

データ取得・加工は次の3つの手順で行った。まず、「じゃらん」のロコミページよりwebスクレイピングにより研究対象とする各観光地に対するロコミをすべて取得した。その際に、取得されたロコミのうち文字数が100文字より少ないものを除外した。その理由は、文字数が過度に少ない場合、次に行う観光テーマ、観光スタイルの抽出ができないからである。これは事前いくつかの投稿を参照して、100文字程度がそれらを抽出できる限界だと判断したことに因る。そしてスクリーニングに残った直近400件にロコミを分析データとした。上の手順により、2021年10月26日以前の対象観光地のロコミで、7観光地×400件の計2800件のロコミを分析データとした。

#### 4.2. LDA手法による観光テーマと観光スタイルの抽出

観光テーマ、観光スタイルの抽出にLDA手法を用いる。LDA手法とは、トピックモデルの一つであり、テキストデータを任意の分類数に定量的に分類することが可能である。ここで、本研究でこのLDA手法を用いる理由を述べる。理由は3つある。1つ目は、教師無しデータのクラスタリングが可能なことである。本研究で扱うのは正解情報（本研究では観光テーマ、観光スタイル）が付与されていないテキストデータである。これに対してLDA手法は定量的な分類を行うことができる。2つ目は、分類数を指標に基づいて考察できることである。LDA手法ではPerplexity(モデルの予測精度に関する指標)と

Coherence(人による解釈のしやすさ)を用いて、分類数を考察することが可能である。3つ目は、分類ごとに影響力の高い単語を見ることによって、分類結果を解釈することが可能なことである。以上の理由から LDA 手法を用いることを決定した。

次に具体的な抽出方法を述べる。各要因の抽出には図 2 のアルゴリズムを用いた。概説すると、まず LDA 手法をトピック数を 2~50 で設定して行い、その中で Perplexity が低く Coherence が高いトピック数の分類を目視する。その時に、その分類が解釈できれば終了となる。解釈できなかった場合は、stopwords(分類で考慮しない単語)に「各要因に関係ないと考えられる単語」、「複数トピックに出現する単語」、「極端に出現頻度が高い単語」を追加して最初に戻る。その時、観光スタイルの抽出時は、1.4 項で述べた 6 つの軸を基に stopwords の検討を行う。これによって、最終的に観光テーマ、観光スタイルを浮き上がらせる。

分析には Python の自然言語処理ライブラリである gensim の LdaModel を用いた。

### 4.3. ペイジアンネットワークによるモデル図の関係分析

観光スタイルと観光テーマの関係、観光テーマと選択する観光地の関係を、事象間の確率的な依存関係から視覚的にモデルを構築して確率推論を行うペイジアンネットワークにより検討する。モデル構造は、観光スタイルが観光テーマに影響を与え、観光テーマが選択する観光地に影響を与えるという階層構造とした。このモデル構築により、観光地ごとに持たれやすい観光テーマ、観光テーマごとに持たれやすい観光スタイルを検討し、マーケティング施策検討に役立つ知見を獲得する。

分析の際には、分析データをモデルの学習用に 1800 件、検証用に 1000 件というように分けて、モデル学習と検証を行った。そして分析には、株式会社 NTT データ処理システムが提供する「BayoLinkS」を用いた。

## 5. 研究結果

### 5.1. 観光テーマと観光スタイルの抽出結果

分析で得られた観光テーマ分類結果を表 2、分類結果を観光地ごとに整理したグラフを図 3 に示す。表 2 のように、本研究では 8 つの観光テーマを抽出することができた。また、図 3 より観光地ごとに抽出された観光テーマの割合が異なっている。よって、各観光地で持たれやすい観光テーマが異なると言える。

次に分析で得られた観光スタイル分類結果を表 3、分類結果を観光地ごとに整理したグラフを図 4 に示す。表 3 のように、本研究では 9 つの観光スタイルを抽出することができた。また、図 4 から観光テーマごとに抽出された観光スタイルの割合が異なっていることがわかる。よって、観光テーマごとに持たれやすい観光スタイルが異なると言える。

### 5.2. ペイジアンネットワークによる関係分析結果

上記の抽出結果を基にモデルを構築して学習を行った。その後、選択観光地を証拠状態として設定した時の観光テーマの出現確率、各観光テーマを証拠状態として設定した時の観光スタイルの出現確率を計算した。その結果を示したのが表 4 と表 5 である。色がついている部分は、事前確

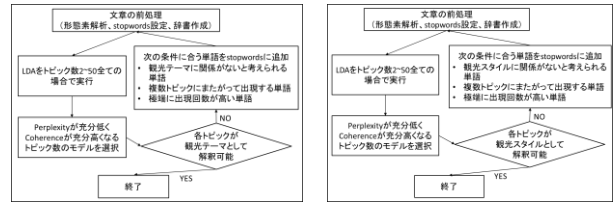


図 2 観光テーマ(左)、観光スタイル(右)抽出アルゴリズム

表 2 各分類の関連単語群とそこから命名した観光テーマ

LDA結果	関連する単語	命名した観光テーマ
1	博物館、展示、見学、散歩、写真	展示物を鑑賞
2	参拝、歴史、パワースポット	寺社仏閣を参拝
4	写真、見学、夜景、景色	風景・写真を撮る
5	お茶、公園	公園で過ごす
6	写真、夜景、お土産、ツアー	風景・写真を撮る
7	夜景、歴史、風呂、お祭り、撮影、写真	夜景を楽しむ
8	散歩、お土産、散策、お茶	散策を楽しむ
9	展示、動物、動物園	動物鑑賞
10	歴史、出店、イベント、見学、映像	歴史を知る

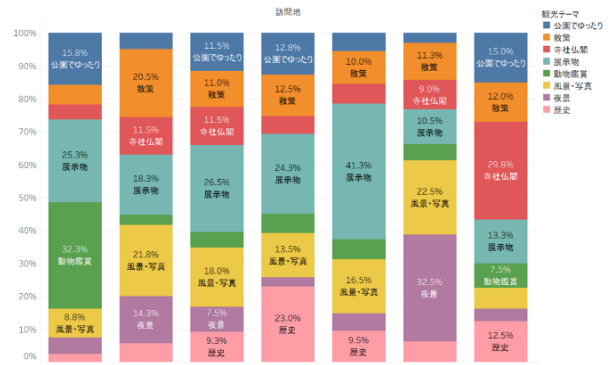


図 3 観光地ごとに抽出された観光テーマ割合の図

表 3 各分類の関連単語群とそこから命名した観光スタイル

LDA結果	関連する単語	命名した観光スタイル
1	温かい、暖房、静か、家族、極寒	過ごしやすい気候を気にする
2	コース、有料、渋滞、寒さ	移動の容易さを気にする
3	対策、天気、防寒、深い、予定	予め予定を立てる
4	混雑、余裕、一人、静か、予約	混雑を気にする
5	環境、トイレ、衝撃、完備、静か	施設・サービスを気にする
6	ツアー、順路、ガイド、案内	定められたものに従う
7	道路、老人、トイレ、サービス、整備	施設・サービスを気にする
8	案内、余裕、アクセス	アクセスを気にする
9	子供、便利、混雑	家族を気にする
10	徒歩、案内、静か、あちこち	静かにゆっくり回る
11	家族、息子、友人、天気	家族を気にする
12	ガイド、ツアー、子どもたち、予定	定められたものに従う

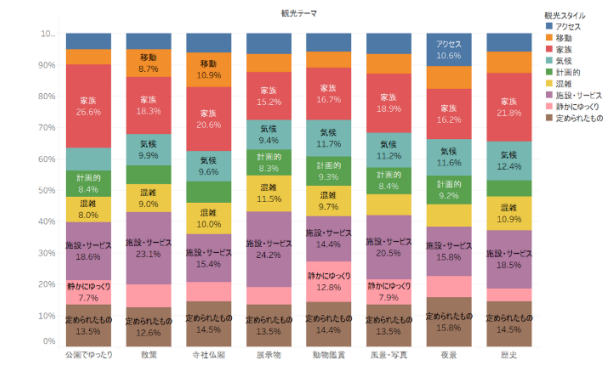


図 4 観光テーマごとに抽出された観光スタイル割合の図



表 4 観光地が証拠状態時の観光テーマ出現確率

観光地	展示物	寺社仏閣	風景・写真	公園	夜景	散策	動物鑑賞	歴史
事前確率	0.406	0.349	0.367	0.35	0.336	0.347	0.343	0.34
函館山	0.4195	0.3837	0.4040	0.3832	0.3825	0.3834	0.3767	0.3749
青い池	0.4264	0.3823	0.3994	0.3852	0.3685	0.3800	0.3745	0.3740
小樽運河	0.4226	0.3834	0.4018	0.3826	0.3722	0.3853	0.3729	0.3723
大通公園	0.4288	0.3802	0.3980	0.3869	0.3656	0.3806	0.3739	0.3809
旭山動物園	0.4247	0.3787	0.3935	0.3871	0.3667	0.3766	0.3910	0.3695
網走監獄	0.4374	0.3806	0.3966	0.3828	0.3674	0.3797	0.3745	0.3739
北海道神宮	0.4185	0.3936	0.3931	0.3885	0.3674	0.3809	0.3760	0.3766

表 5 観光テーマが証拠状態時の観光スタイル出現確率

観光地	家族	公園	歴史	散策・アクセス	動物鑑賞	青い池	定められたもの
事前確率	0.107	0.072	0.081	0.098	0.193	0.061	0.183
函館山	0.1051	0.0718	0.0822	0.1038	0.2031	0.0656	0.1689
青い池	0.1081	0.0811	0.0814	0.1011	0.1806	0.0612	0.181
寺社仏閣	0.1122	0.07	0.0861	0.0973	0.1874	0.0677	0.1798
風景・写真	0.1036	0.0713	0.0844	0.102	0.1833	0.0616	0.1876
公園	0.1073	0.0739	0.0872	0.0984	0.1807	0.0675	0.1771
散策	0.1097	0.077	0.0796	0.1033	0.1924	0.06	0.1761
動物鑑賞	0.1114	0.0747	0.0827	0.103	0.1771	0.0648	0.1736
歴史	0.1078	0.0758	0.0808	0.105	0.1822	0.0625	0.1846

率からの増加幅が大きいもので、赤色、朱色、緑色の順に大きい。この表の解釈例として一例を挙げると、表 4 から函館山を選択した人は観光テーマ「夜景」、「風景・写真」、「散策」を持ちやすく、表 5 から観光テーマ「夜景」を持つ人は観光スタイル「アクセス」、「計画的」、「定められたものに従う」を持ちやすいと言える。

## 6. 考察

### 6.1. 観光地と観光テーマの関係に関して

観光テーマの抽出は野守ら[3]が既に行っている部分であるため、その結果と本研究で得られた結果を比較する。まず旭山動物園に対する結果に関して、野守らの研究では「動物鑑賞」のみが観光テーマとして抽出されていた。しかし、本研究では旭山動物園に対して「公園で過ごす」テーマが得られた。この結果から、動物園は動物を見るだけでなく、公園として利用されてゆっくり時間を過ごすことも利用動機となることが考えられる。また、大通公園に関して見ると本研究で新たに「歴史を知る」テーマが得られた。これは大通公園沿いには歴史的建造物がいくつか存在し、それを楽しむ目的が抽出できたと考えられる。このように、本研究では従来研究では捉えられなかった観光地に対する観光テーマを得られたと考えられる。

### 6.2. 観光テーマと観光スタイル抽出結果に関して

表 5 より観光テーマによって持たれる観光スタイルが異なることが言える。なぜそのような結果が現れたのかについて、観光スタイルが「家族を気にする」の時に観光テーマ「公園でゆっくり過ごす」、「歴史を知る」の割合が大きいことを一例として考察する。家族を気にする人は家族と一緒に時間を楽しむことや、子供に貴重な経験・体験をしてほしいという思いを持っていると考えられる。よって、公園でゆっくり家族の時間を作りたいという思いや、歴史を学べる場所で子供に色々な経験をさせたいという思いを持っているため、このような結果が得られたと考えられる。このように、他の関係性でも同様な考察を行うことができ、観光スタイルと観光テーマの因果関係を考察できる。

### 6.3. 研究結果より効果的だと考えられる観光地推薦例

本研究の結果から、効果的だと考えられる観光地の推薦例を述べる。ここでは、観光スタイル「家族を気にする」人が観光テーマ「公園でゆっくり過ごす」、「歴史を知る」を持ちやすく、それぞれの観光テーマで選択されやすい観光地を持つという関係を一例として述べる。まず、ロコミ

の分析やアンケートからある観光客の情報として観光スタイルが「家族を気にする」、選択観光地が「青い池」という情報を得られたとする。この時、研究結果からこの観光客は「公園でゆっくり過ごす」ことを観光テーマとして「青い池」を訪れたことが予測できる。そして、この観光スタイルを持つ人は「歴史を知る」というテーマも持つ可能性が高いことが研究結果から考えられる。よって、この観光客には、「歴史をテーマとして子供さんと博物館網走監獄にてその歴史を学ぶ」といった観光を提案できると考えられる。このように、本研究の結果とロコミやアンケートから得られる情報を基に効果的な観光地推薦案を考えることができる。

## 7. 結論と今後の課題

本研究は観光客の詳細理解と、効果的なマーケティング施策の策定に貢献することを目的として、LDA 手法によって観光スタイル、観光テーマの抽出を行い、それらと選択観光地の関係を、ベイジアンネットワークを用いて明らかにした。その結果、LDA 手法による要因抽出の有効性を示すことができ、関係分析の結果から観光スタイルという観光客属性を考慮した効果的な観光地推薦例を提案することができた。

今後の課題は、LDA 手法による観光テーマ、観光スタイルの抽出精度の向上と、観光スタイルと心理的属性の関係分析が考えられる。現状の方法でも上記要因の抽出が行えてはいるが、実際の文章と抽出結果を見た時に正しいと言えない結果も少々見られた。よって、この改善を行う必要がある。そして、ロコミから観光客の心理的属性を考察するために、今回得られたような観光スタイルと心理的属性の関係を分析することが求められる。

## 参考文献

- [1]北村倫夫,「自治体や DMO が展開すべき世界標準の観光マーケティング戦略, 国際文化研修 2018 秋, vol.101, pp.54-57 (2018)
- [2]野守耕爾, 神津友武,「ロコミデータに PLSA を適用した観光客目線による観光地分析」, 人工知能学会全国大会論文集, 29, pp.1-4 (2015)
- [3]Engel, Kollat, Blackwell, “Customer Behavior (3rd ed.)”, New York: John Wiley & Sons, P174-175, P294-296. (1978)
- [4]株式会社リクルート, 「じゃらん」, <https://www.jalan.net/>, (最終閲覧日 2021/12/27)
- [5]石井康夫, 大久保あかね, 鈴木大介, 「観光マーケティングにおける新たな分析手法の提案-伊豆半島の観光魅力度に関するテキストマイニング分析を事例として-」, 知識と情報(日本知能情報フェジ学会誌)Vol.31, No.4, pp.745-753(2019)
- [6]原田倫妙, 「観光消費行動への影響要因に関する研究(日本人の台湾観光と台湾人の日本観光を例として)」, 鹿児島国際大学大学院経済学研究科地域経済政策専攻博士学位論文(2018)
- [7]竹岡志郎「機械学習を活用したテキストマイニング-外形的データを併用することによる特徴分析-」, 経営学論集第 89 集 自由論題, pp.F17-1-F17-7(2018)