

〈自由論文〉

## ソーシャルメディアのコンテンツから分析した コンビニエンスストアの特徴と分類

### Characteristics and Categorizing of Domestic Major Convenience Store Brands by Analyzing Contents in Social Media

小林 稔  
Minoru Kobayashi

#### 【Abstract】

The purpose of this paper is to show some characteristics and categorizing about domestic major convenience store brands by analyzing contents in social media which are a kind of big data in the Internet. This paper acquired contents in social media by using some free analysis tools mainly. This paper also carried out various text mining methods for analyzing these contents in social media by using some free analysis tools. It is another purpose of this paper to inspect whether we can execute various text mining methods to big data in the Internet by using free analysis tools. Specifically, at first this paper executed the frequency analysis to the morphemes included in contents in social media and made word clouds from the result of the frequency analysis. Then, this paper made a term-document matrix and applied correspondence analysis to the term-document matrix. Furthermore, this paper analyzed N-gram from that data and made N-gram graphs. From those results, this paper shows some characteristics and categorizing of each convenience store brands.

#### 【キーワード】

ソーシャルメディア, テキストマイニング, ワードクラウド, コレスポネンシ分析,  
コンビニエンスストア, インターネット, ビッグデータ

## 1. はじめに

今日では日常生活に密着した存在となっているコンビニエンスストアは、地域社会のインフラストラクチャとして様々な商品やサービスを提供しマルチな機能を担っている。本研究では、国内の主要な8ブランドのコンビニエンスストアについて、各コンビニエンスストアのブランドをキー

ワードとしてTwitter<sup>1)</sup>のツイートを取得し、そのテキストデータを分析することによって各ブランドの特徴を抽出した。また、その結果から各ブランドの分類を試みた。

Twitterは、ユーザ数の多いソーシャルメディア<sup>2)</sup>の一つであり、少ない単語数の制限の中でユーザが感じたことを気軽に共有することに特徴があり、その他のソーシャルメディアと比較して不特定多数のユーザへ向けたツイートが多数を占

める。したがって、各コンビニエンスストアのブランドに関連した Twitter のツイートから各コンビニエンスストアのブランドの特徴の抽出と分類が可能ではないかというのが本研究の出発点になっている。

また、本研究では、Twitter のツイートというインターネット上のビッグデータを分析して問題意識の解明を図るという研究の方法論自体についても研究対象としている。この点に関して本研究では、特に近年になって急速に整備されつつあるオープンソース<sup>3)</sup>のソフトウェアをはじめとしたフリーな研究ツールを利用して、目的とする研究対象の分析がどこまで可能であるかという検証を行っている。本研究に利用した主な研究ツールは、インターネット上のビッグデータの取得や主要な分析に関しては、全てがオープンソースのフリーなものであり、今後同様な研究を進める場合における具体的な分析ツールの選択や利用法について参考になるものと考えている。

## 2. 国内コンビニエンスストアの現状

日本フランチャイズチェーン協会の調査によれば、同協会の正会員である(株)スリーエフ、(株)セコマ、(株)セブンイレブン・ジャパン、(株)ファミリーマート、(株)ポプラ、ミニストップ(株)、山崎製パン(株)デイリーヤマザキ事業統括本部、(株)ローソンの国内8社の国

内店舗数は2017年6月末で5万5026店、2017年6月の売上高は8897億円、2016年の年間では10兆5722億円と10兆円を上回る規模にまで達している。2016年は、店舗数2.8%増、売上高3.6%増、来客数2.7%増と着実に成長を続けている。しかし、その内訳をみると新規の出店を除いた既存店ベースでは、売上高は0.5%増、来客数は0.5%減と一服感が出ており、業界の成長は新規出店に支えられている構図がうかがえる。全国のコンビニエンスストアは、すでに飽和状態に達しつつあり、新規出店に支えられた成長路線には限界もあろう。

このような状況の中で、各コンビニエンスストア間では競争が激化しており、新しい商品やサービスの開発、各種のキャンペーンの実施などにより他社のコンビニエンスストアとの差別化と顧客の取り込みに躍りになっている。また、コンビニエンスストア業界にも一部で合従連衡の動きが出始めている。2004年にサークルKとサンクスが統合してサークルKサンクスとなり、さらに、サークルKサンクスを配下に持つ大手スーパーグループであるユニーとファミリーマートが2016年9月に合併して、ファミリーマートとサークルKサンクスが統合した。2017年7月時点で、サークルKサンクスの店舗は順次ファミリーマートへ移行しつつある。参考までにコンビニエンスストア各社の出店数と売上高を表1に示す。

2020年に東京オリンピックを控え、わが国の

表1 コンビニエンスストア運営会社の国内店舗数と国内売上高

	店舗数(店)		売上高(億円)	
	2016年度末	2015年度末	2016年度	2015年度
セブンイレブン※	19,638	18,572	45,156	42,910
ローソン	12,575	11,880	21,580	20,496
ファミリーマート※	13,383	11,656	30,094(注1)	20,056
ミニストップ※	2,255	2,221	3,405	3,363
サークルKサンクス※	3,674	6,350	(注2)	9,746
デイリーヤマザキ	1,571	1,561	619	639
セイコーマート※	1,183	非上場のため不明	非上場のため不明	1,827

※店舗数は2017年7月。なお、ローソンは2017年2月の、デイリーヤマザキは2016年2月の店舗数である。

(注1) ファミリーマートとサークルKサンクスが統合したため、サークルKサンクスの売上高を含む。

(注2) 2016年9月に親会社であるユニーグループ・ホールディングスがファミリーマートに吸収合併されたため、サークルKサンクスの売上高はファミリーマートの売上高に含まれている。

(出所) 有価証券報告書、会社概要、会社案内よりデータを抜粋。

経済は一部で拡大へ向かう動きもみられるが、デフレからの脱却は遅々として進んでいない。企業の内部留保が拡大する中で、個人所得は思ったように増えておらず、国内の個人消費はなかなか拡大に転じていない。百貨店やスーパーは前年比で売上高を減少させている。その意味では、着実に売上高を拡大させているコンビニエンスストアは、今後の流通業のあり方を模索する上で重要な位置付けにあるといえよう。

本研究では、このようなコンビニエンスストアの流通業としての位置付けを分析するための第一歩として、コンビニエンスストアのブランド別の特徴と商品やサービス、店舗運営に対する顧客の意識を Twitter のツイートをテキストマイニングすることで明らかにしていく。コンビニエンスストアは、顧客からみれば各ブランドによって認識されていることを考慮して、企業別ではなくブランド別に分析することとした。具体的には、セブンイレブン、ローソン、ファミリーマート、ミニストップ、サークル K、サンクス、デイリーマート、セイコーマートの 8 ブランドについて研究を進めていく。

### 3. 研究の方法

本研究では、国内の主要なコンビニエンスストア 8 ブランドをキーワードとして Twitter からツイートを取得するとともに、取得したツイートをテキストマイニングした上で適切な分析を行うことによって消費者が抱いている各ブランドの商品やサービス、店舗運営などに対する意識を明らかにする。つまり、Twitter でツイートされている各コンビニエンスストアのブランドに関連した商品、サービス、店舗運営に対する期待、不満などに関わる単語の頻度などから消費者が持っている各ブランドへの意識を明示するとともに、その結果から各ブランドの特徴を抽出し、またその特徴によって各ブランドの分類を試みる。

まず、各ブランドをキーワードにした Twitter のツイートの取得を行っていく。Twitter はユーザ登録とともに必要な手続きを行えば、一定の条

件の下で Twitter からツイートを取得するための API (Application Programming Interface)<sup>4)</sup> の提供を行っている。この API を利用して必要なキーワードや条件を指定して Twitter からツイートを選択的に取得することができる。まずは、研究に必要なツイートを取得するためのユーザアカウントの登録を行い、次に API の利用に関わる手続きをする。手続きは無料であり、手続きが完了すると API を利用するための consumerKey, consumerSecret, accessToken, accessSecret が発行される。

実際に API を使ってツイートを取得するために、本研究では“Twitter”という数理統計パッケージ R と連携したツールを利用した。“Twitter”はフリーで利用できるツールであり R の標準ライブラリーとして公開されているのでインターネット経由でインストールが可能である。この“Twitter”は、キーワード、日時などを指定して R からコマンドを入力することで必要なツイートを取得することが可能である。

取得したツイートは各コンビニエンスストアのブランドごとに整理して、データクレンジング<sup>5)</sup>を行いツイートに含まれる URL などの記号や分析に不必要な文字を除去する。なお、データクレンジングはツイート取得と同時に記号や分析に不必要な文字を除去するとともに、後で説明する形態素解析を実施する際に不必要な品詞を削除する。つまり 2 段階に分けて実施していく。1 段階目のデータクレンジングを実施した後のデータは、形態素解析によって品詞ごとに分類し、2 段階目となるデータクレンジングを品詞ごとに実施するとともに、実際に品詞ごとに分類された単語の内容を検証して、コンビニエンスストアのブランドの特徴を抽出するために必要な単語とそうではない単語に振り分けていく。絵文字や顔文字などもこのプロセスで除去する。

なお、形態素解析には、RMeCab というツールを利用した。RMeCab は京都大学情報学研究所と日本電信電話株式会社コミュニケーション科学基礎研究所の共同研究ユニットプロジェクトによって開発された形態素解析エンジンである。フ







各コンビニエンスストアのブランドをキーワードとしてAPI経由で取得したツイートにデータクレンジングを2段階で施したテキストデータを作成する。

このテキストデータをRMeCabを使って形態素解析を行い品詞別に形態素を分類し、同時に出現頻度を整理する。次に各ブランドの品詞別の形態素の頻度表を作成して、形態素の意味などを詳細にチェックする。その際に、各コンビニエンスストアのブランドの特徴や消費者の意識などとの関係のない意味不明の形態素を適切に選択して削除していく。

さらに、不必要な形態素を削除した各ブランドの形態素の頻度表を一つにまとめて、本研究で対象とした8つのコンビニエンスストアのブランドを列名、それぞれの形態素を行名とし、各ブランドにおける各形態素の頻度を要素とした行列を作成する。この行列を、検索語・文書行列 (term-document matrix) という。ここでは、各ブランド名をキーワードとして取得したツイートの集合が文書に相当する。つまり、行列の列名が各ブランド名となる。このように、ある term (形態素) が、各ブランドをキーワードとして取得したツイートの集合の中に現れる頻度を要素として作成した検索語・文書行列が表5-1である。この検索語・文書行列を使ってコレスポネンス分析を行う。

コレスポネンス分析は、フランスの Benzecri のグループ<sup>7)</sup>によって方法論が示され普及した分析方法であり、対応分析とも呼ばれている。コレスポネンス分析と同様の分析手法は日本やその他の国でも同時並行的に研究されていたものの、言語の壁によって各国共通の方法論として確立するには至らなかったため、当初はコレスポネンス分析と同様な分析手法が異なる方法論として示されていた。

例えば、1940年代に林知己夫によって示された数量化Ⅲ類は、コレスポネンス分析とは異なる方法論として扱われていたこともあるが、それぞれの分析結果は変換によって一致させることが可能であり、現在では数量化Ⅲ類と数理的には同

等の分析であることが示されている。英語によるコレスポネンス分析に関連する文献は、1970年以前は、Benzecriのグループの文献以外にほとんど存在せず<sup>8)</sup>、このことがフランスにおいて普及が進んでいった理由とされている。1980年代以降、米国を中心にコレスポネンス分析の普及が進み、今日ではSPSS<sup>9)</sup>やSAS<sup>10)</sup>、Rなどの統計パッケージでも分析手法の一つとして実装され利用できるようになっている。

一般にコレスポネンス分析では、分析対象とするクロス集計表の行の項目と列の項目の相関が最大になるように数量化する。分析の考え方は、主成分分析や因子分析と同様であり、コレスポネンス分析を制約付き主成分分析とする見方もある<sup>11)</sup>。つまり今回の場合は、作成した検索語・文書行列の形態素とブランドの相関が最大になるように数量化していく。本論文の分析には、数理統計パッケージRで利用可能なコレスポネンス分析のパッケージ mass と ca のそれぞれを利用した。mass, ca それぞれのパッケージで分析を試みてその結果が一致することを確認している。

## 5.2 コレスポネンス分析の結果

本論文でコレスポネンス分析を行った結果を表5-2～5-3、図5-1～5-7に示す。

まず図5-1を見ると各コンビニエンスストアのブランドの相対的な位置関係を確認することができる。コレスポネンス分析の結果である図5-1で近い位置にプロットされている各ブランドは、商品、サービス、キャンペーン、顧客対応、店舗運営などに関して同じような特徴を持っていると考えることができる。

そこで、図5-1にプロットされている各ブランドの位置関係から各ブランドをグループ1～5の5つのグループに分類してみた。また分類した5つのグループの共通性を確認するため、図5-1で各グループの近くにプロットされている主な形態素を抜き出して5つのグループに分けてプロットしたのが図5-3～図5-7である。5つの各グループの近くにプロットされている形態素の内容を検証すると以下のような各グループの特徴を抽出す

表5-1 形態素別の各コンビニをキーワードとして取得したツイートに出現する頻度（単語文書行列）

形態素	セブンイレブン	ローソン	ファミリーマート	ミニストップ	サークルK	サンクス	デイリーヤマザキ	セイコーマート
アイス	42	22	24	59	18	1,031	0	31
アイテム	22	3	42	6	196	12	11	13
アカウント	2	2	1	83	17	3	0	2
あげる	25	67	7	3	10	6	0	2
あなた	7	6	34	3	5	8	0	0
いただく	4	38	3	1	29	18	0	3
いただける	3	3	11	0	55	6	0	7
イベント	5	24	7	2	20	14	0	2
イラスト	9	4	43	2	243	14	0	5
インド	1	0	0	0	0	25	0	0
うまい	10	22	3	22	34	10	2	42
おいしい	17	26	13	29	37	5	6	21
オーク	42	0	0	0	0	0	0	0
おそい	2	29	4	34	4	1	0	0
おにぎり	30	20	7	4	25	8	1	29
オムニ	43	0	0	0	0	0	0	0
オムニセブン	41	0	0	0	0	0	0	0
オムニバス	38	0	0	0	0	0	0	0
おもしろい	1	21	1	1	1	2	0	0
オリジナル	4	42	0	67	2	4	2	4
お客様	1	7	3	98	1	0	0	1
カード	28	24	18	23	36	89	1	20
カフェ	2	5	24	4	3	2	2	26
カロリー	24	5	34	0	0	0	0	1
キャラ	2	89	5	1	3	2	0	1
キャンペーン	289	167	42	531	15	28	3	12
きれい	608	0	0	0	1	0	0	2
くださる	19	23	10	6	8	4	0	1
クッション	1	1	0	52	1	0	0	0
グッズ	17	20	178	12	13	6	5	0
クリーム	2	5	31	9	2	5	1	3
クロノトリガー	42	0	0	0	0	0	0	0
ケーキ	1	13	22	4	5	0	1	0
ケロロ	42	0	0	0	0	0	0	0
コイン	1	1	24	0	0	0	0	0
コード	27	22	2	11	27	12	0	0
コーヒー	22	17	21	168	10	11	2	5
コール	5	0	84	0	1	0	0	0
こだわる	1	0	0	29	2	0	0	11
こちら	32	45	95	2	256	19	3	4
ゴディバ	2	29	0	0	0	0	0	0
コラボ	242	135	22	80	6	3	27	4
コンビニ	289	127	248	76	191	128	16	270
コンプリートシングル	24	0	0	0	0	0	0	0
サイズ	16	5	2	80	6	2	0	3
サラダ	23	14	15	3	6	1	0	10
サンド	1	44	1	1	6	0	0	10
サンドイッチ	4	86	1	0	2	0	0	2
シート	5	6	2	43	1	0	0	0
シール	6	35	1	0	0	0	4	0

※いずれかのツイートで出現頻度20以上の形態素306からランダムに50選択した。

表 5-2 コレスポネンシ分析の結果  
- rscore の内ランダムに 30 項目を抜粋

形態素	Dimension1	Dimension2
アイス	▲ 1.458	▲ 0.467
アイテム	0.343	▲ 0.546
アカウント	0.185	1.565
あげる	0.707	▲ 0.340
あなた	0.197	▲ 0.479
いただく	0.134	▲ 0.351
いただける	0.241	▲ 0.558
イベント	0.128	▲ 0.365
イラスト	0.259	▲ 0.581
インド	▲ 1.723	▲ 0.632
うまい	0.388	0.028
おいしい	0.483	0.048
オーク	1.592	▲ 1.310
おそい	0.418	0.925
おにぎり	0.619	▲ 0.461
オムニ	1.592	▲ 1.310
オムニセブン	1.592	▲ 1.310
オムニバス	1.592	▲ 1.310
おもしろい	0.523	▲ 0.065
オリジナル	0.368	1.130
お客様	0.230	1.960
カード	▲ 0.291	▲ 0.305
カフェ	0.427	▲ 0.197
カロリー	0.815	▲ 0.805
キャラ	0.662	▲ 0.075
キャンペーン	0.597	0.705
きれい	1.587	▲ 1.305
くださる	0.659	▲ 0.349

表 5-3 コレスポネンシ分析の結果 - cscore

コンビニエンスストア	Dimension1	Dimension2
seven	1.370069353	▲ 1.058333543
Loson	0.633674214	▲ 0.012032019
Famima	0.243691420	▲ 0.472835490
ministop	0.149204998	1.825343959
circleK	0.268600420	▲ 0.480768499
thanks	▲ 1.596922006	▲ 0.488707582
dailyY	0.595003132	▲ 0.119741463
seikoM	0.532350211	▲ 0.074740883

図 5-1 コレスポネンシ分析の結果

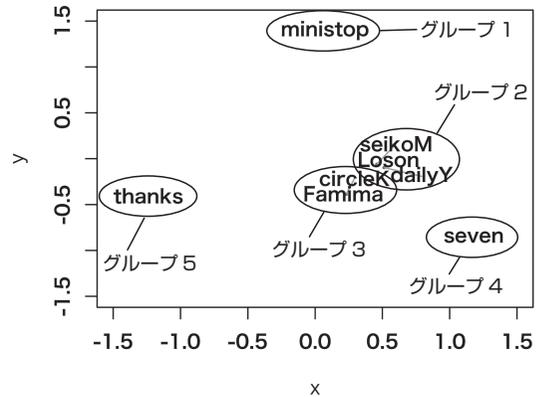


図 5-2 コレスポネンシ分析の結果

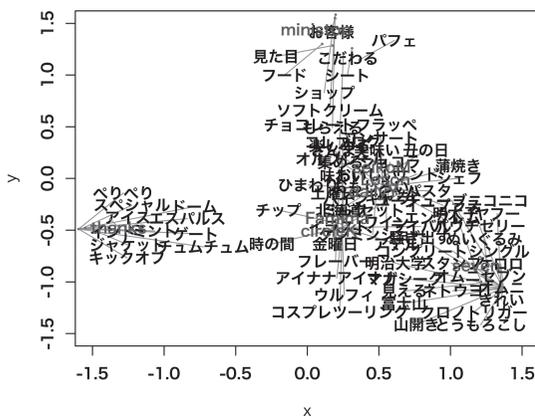
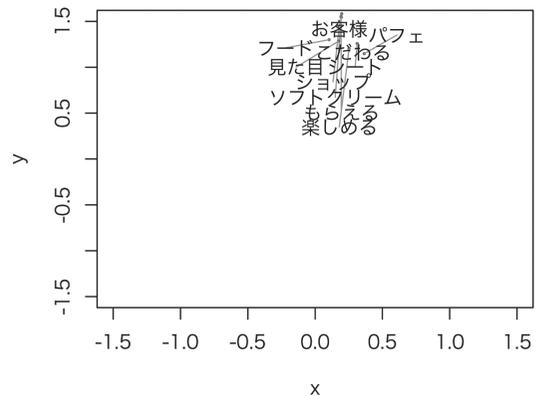


図 5-3 コレスポネンシ分析の結果



ることができる。

まず図 5-1 から、ミニストップは他のブランドとは異なる位置に単独でプロットされておりグループ 1 に分類する。グループ 1 の近くには図

5-3 から“ソフトクリーム”，“フード”，“パフェ”などデザートや軽食といった食品や“楽しめる”，“もらえる”など顧客への対応に関わる形態素がプロットされている。つまり、ミニストップは食



いったデザート系の食品に共通性を見出すことができる。

図5-1においてグループ3から少し離れた位置にセブンイレブンが単独でプロットされており、ここではグループ4とする。図5-6から、グループ4の近くには、「明太子」パウチゼリー，“とうもろこし”などの食品に関わる形態素，“富士山”，“山開き”などの地域性に関わる形態素がプロットされている。“パウチゼリー”は、セブンイレブンが7月中にパウチゼリーを扱ったキャンペーンを実施しており，“明太子”は、セブンイレブンで明太子を使った食品の販売が多く、人気食品の特徴が出ているためであろう。“とうもろこし”は新商品である「1本分の焼きとうもろこし」が人気となったこと，“富士山”については、山開きに合わせて富士山が見える店舗に関するツイートを企業側が投稿していたことの影響であろう。

一方、図5-1では、サンクスが他のグループとは離れた位置にあり、ここではグループ5とする。図5-7より、この近くには“アイス”，“チュムチュム”，“エスパルス”，“ミント”，“ぺりぺり”といった形態素がプロットされている。サンクスは、アイスなどの食品や各種のキャンペーンで他のブランドとは異なる特徴を持っているためと考えられる。具体的にみると“エスパルス”については、サッカーチームであるエスパルスのチケットの一般先行販売を実施していたこと，“ぺりぺり”は、「ホームランバー…“ぺりぺり”と紙を剥がして食べる白いアイス…」など昔を懐古するアイス販売していたことの影響であろう。

以上のように、コレスポネンス分析の結果から、各コンビニエンスストアのブランドをその特徴によって5つのグループに分類することができた。しかし、この分類には、期間限定でのキャンペーン、商品、サービスなど時節に関係する項目の影響があることを考慮しなくてはならない。つまり、分析する時期や期間によっては、異なる分類となる可能性がある。今回の結果は、ツイートを取得した期間の分析による分類であることを確認しておく。

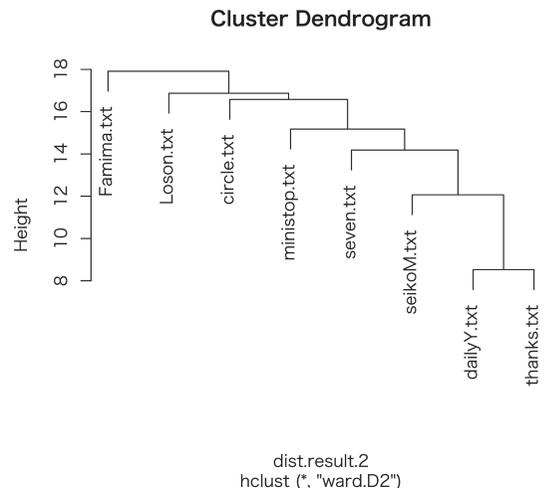
## 6. 階層型クラスター分析と考察

ここでは、階層型クラスター分析によって各コンビニエンスストアのブランドの類似性を検証する。また、先に行ったコレスポネンス分析の結果と比較検討を行うことにする。

まず各ブランドの非類似度を距離として表し距離の近いものをまとめてクラスターを作っていくこととする。階層型クラスター分析では、分析対象が未分類の状態から少数のクラスターを形成していき、結果として全ての分析対象を含む大きなクラスターを形成する。したがって、分析対象間の非類似度を測定するための距離、またどのようにクラスター間の距離を計算していくのかが重要である。ここでは、分析対象間の非類似度の測定にユークリッド距離 (Euclidean distance)<sup>12)</sup> を用い、クラスター間の距離はウォード法 (Ward's method)<sup>13)</sup> を用いた。

階層型クラスター分析の結果を図6-1に示す。階層型クラスター分析では、任意の階層に線引きをすることでクラスターをいくつかのグループに分類することが可能である。ここでは、図6-1の階層を任意に線引きして前章と同様に各ブランドをグループ6～グループ10の5つのグループに分類してみる。

図6-1 階層型クラスター分析の結果



まず図6-1を見ると、7つのクラスターが階層を形成していることが分かる。ここで図6-1の最上位の階層にあるクラスターから3番目の階層にあるクラスターまでを1つのグループとしてグループ6に分類する。次に4番目にあるクラスターをグループ7、5番目の階層にあるクラスターをグループ8、6番目の階層にあるクラスターをグループ9、そして最後に7番目の階層にあるクラスターをグループ10とする。

つまり、最上位のクラスターにあるファミリーマート、2番目の階層のクラスターにあるローソン、3番目の階層のクラスターにあるサークルKをグループ6とする。ミニストップが単独でグループ7、セブンイレブンがやはり単独でグループ8に、同様にセイコーマートがグループ9に、そしてデイリーマートとサンクスをグループ10に分類する。

グループ6は、ファミリーマート、ローソン、サークルKが同じグループとなり類似性が高いことが分かる。図5-1のコレスポンデンス分析による結果では、この3つのブランドはそれぞれグループ2およびグループ3に分類されているが、グループ2とグループ3は極めて近い位置にありこの3つのブランドの類似度が高いことについてはここでの分類の結果と矛盾はない。

グループ7とグループ8は、ミニストップとセブンイレブンでありそれぞれ単独でグループを形成している。図5-1のコレスポンデンス分析の結果ではミニストップが単独でグループ1に、セブンイレブンは単独でグループ4に分類されておりここでの結果と同じである。

次に、グループ9は、セイコーマートが単独でグループを形成し、デイリーマート、サンクスがグループ10を形成している。サンクスが最下層のクラスターに位置して他のグループとの類似性が低いことは、コレスポンデンス分析の結果と同じ傾向である。また、セイコーマートとデイリーマートがそれぞれ隣接するクラスターにあり類似性が比較的高くなっていることも、コレスポンデンス分析の結果で両ブランドが同じグループ2に分類されたことと同様の傾向である。しかし、グ

ループ9のセイコーマート、グループ10のデイリーマートがグループ6のファミリーマート、ローソン、サークルKとは異なるグループとなり、またクラスターも隣接しておらず類似性が低くなっている。このことは図5-1のコレスポンデンス分析の結果とは異なっている。共に最下層のクラスターにあるデイリーマートとサンクスの類似性が高いことも図5-1のコレスポンデンス分析とは異なる結果である。

それぞれの分析において異なる結果となったのは、コレスポンデンス分析は、分析結果に大きな影響を与える要因となる特徴的なデータに着目して分析するのに対して階層型クラスター分析では、全てのデータ間の距離を測定してクラスターを形成していることから、このように一部で異なる結果が出ているものと考えられる。つまり、セイコーマートとデイリーマートは、図5-1のコレスポンデンス分析の結果から考えると、消費者から注目されている商品やサービスなどに関しては、ファミリーマート、ローソン、サークルKと類似性がある。しかし、取得したツイートから抽出した全てのデータ間の距離という判断基準で分類すると必ずしも類似性は高くないという解釈ができる。

ここでは、階層型クラスター分析によって、各ブランドを5つに分類することを試みた。また、その結果をコレスポンデンス分析の結果と比較を行った。それぞれの分析で共通した傾向は、以下の①～④となる。①ファミリーマート、ローソン、サークルKは消費者からは類似性が高いブランドとして意識されていること。②ミニストップ、セブンイレブンは、それぞれ他のブランドとは異なる特徴のあるブランドとして意識されていること。③セイコーマートとデイリーマートは、比較的類似性が高いブランドと意識されていること。④サンクスは、他のブランドとは類似性が低く特徴のあるブランドとして意識されていること。

このように、Twitterのツイートを取得しテキストマイニングならびに統計的な処理を行うことによって、消費者の各コンビニエンスストアのブランドに対する意識を分析することができた。従

来から消費者の意識調査を行う場合は、アンケート調査やインタビュー調査などが行われてきた。しかし、今回のようにソーシャルメディアのコンテンツを取得することによって、アンケート調査などと同様に、多くの消費者の意識に関わる膨大なテキストデータを抽出することができることを示した。ソーシャルメディアのコンテンツには、消費者のタイムリーな意識が反映されており、そのようなソーシャルメディアのコンテンツを取得して分析することによって、アンケート調査とは異なる迅速かつタイムリーに消費者の意識の分析が可能になる。

今回の分析においては、期間限定のキャンペーンや季節に関連した消費者の意識が結果に反映されており、特定の期間の特徴が結果に影響を与えることを確認することができた。この点は今回の分析によって得られた成果の一つであり、また今後の同様の分析を行う場合には十分に留意すべきである。

## 7. N-gram による共起関係の分析

文書内に存在する任意の文字や形態素、品詞が2言語、3言語、 $N$ 言語として隣接して出現する言語単位の共起<sup>14)</sup>関係を、2グラム、3グラム、 $N$ -gram と呼ぶ。ここで  $N$  は任意の正の整数である。 $N$ -gram は、文書の特徴を示すものと解釈さ

れている。言語単位で  $N$ -gram を考えるとき、言語の総数を  $X$  とすると、連続する  $N$  単位の組み合わせの数は  $X^N$  となる<sup>15)</sup> から、 $X$  が大きい場合に  $N$ -gram の出現頻度を調べる作業は  $N$  の増加とともに膨大なものになってしまう。

ここでは、各コンビニエンスストアのブランドをキーワードとして取得したツイートから抽出した形態素について、 $N = 2$ 、一般的にはバイグラムといわれるケースについて分析を試みる。本論文では、 $N$ -gram の出現頻度を調べるため RMeCab に用意されている Ngram 関数を用いて行った。その結果を各コンビニエンスストアのブランドごとに図 7-1 ~ 図 7-8 に示す。

図 7-1 からセブンイレブンのバイグラムを見ると“キャンペーン”、“実施”、“山開き”、“富士山”、“きれい”、“見える”、“スタンプ”、“抽選”、“当たる”、“景品”などが共起関係にあることが分かる。「セブンイレブンの静岡県の店舗では富士山が良く見える」ことがバイグラムでも確認できる。

図 7-2 からローソンでは、“沖縄”、“県”、“産”、“黒糖”、“使用”、“ローソンチキ”、“キャラ”、“選挙”が共起関係にある。ローソンのフライドチキが「新エルチキ」として2017年6月27日から販売され、「新エルチキ」キャラクターへリツイートすることで選挙が行われたこと、また新商品の「黒糖レーズン蒸しパン」に沖縄県産の黒

図 7-1 セブンイレブンの共起関係 (バイグラム)

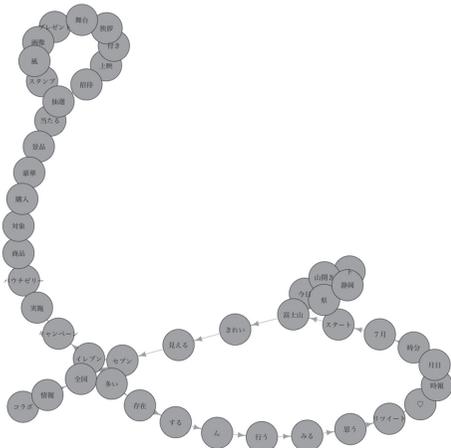
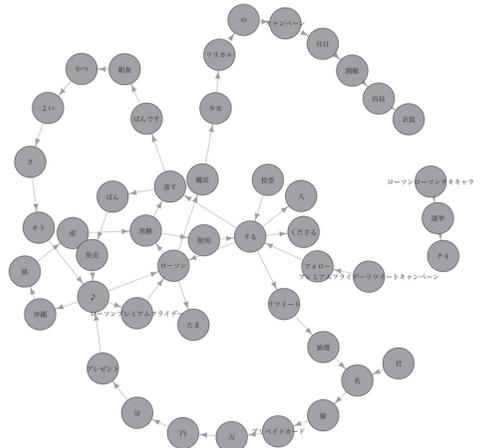


図 7-2 ローソンの共起関係 (バイグラム)





糖が使用されるなど沖縄に関連した商品が販売されたことなどがバイグラムに表れている。

図7-3からファミリーマートのバイグラムを見ると“ファミマ”、“チョコレートフラッペ”、“チョコチップ”、“増量”、“復活”、“サークル”、“サンクス”、“マルチ”、“コピー”、“機”が共起関係にある。ファミリーマートでは、昨年チョコレートフラッペ200万食が3週間で完売したことから、今年は500万食を用意しチョコチップを増量して復活させたことがバイグラムに表れている。

図7-4からミニストップのバイグラムを見ると“オリジナルイラストグッズ”、“もらえる”、“ボトル”、“カバー”、“ローソン”、“ミニストップ”、“タイアップ”、“キャンペーン”、“スタート”が共起関係にある。ミニストップでは、7月3日から7月23日まで「カードキャプターさくら」を扱ったキャンペーンが行われていたことがバイグラムに表れている。

図7-5からサークルKのバイグラムを見ると“刀剣”、“乱舞”、“くじ”、“ファミマ”などが共起関係にある。これは、サークルK、サンクス、ファミリーマートなどでアニメ「刀剣乱舞」のくじなどの関連商品が扱われていたことがバイグラムに表れているためであろう。

図7-6からサンクスのバイグラムを見ると“チョコ”、“ミント”、“アイス”、“風味”が共起関係にあることが分かる。これはサンクスでチョコミントのアイス入りの飲み物が人気となっていたことが要因であろう。

図7-7からデイリーヤマザキのバイグラムを見ると“魔法使い”、“黒猫”、“ウイズ”が共起関係にある。これは、デイリーヤマザキ、セイコーマートなどで「魔法使いと黒猫のウイズ」、「魔法書、フル・エーテル」に関連したキャンペーンを実施していたことがバイグラムに表れているためと思われる。

図7-8からセイコーマートのバイグラムを見ると“札幌”、“市”、“北海道”、“ケンミンショー”、“豆”、“パン”、“カツ”、“丼”などが共起関係にある。これは、北海道を地場としているセイコーマートの特徴である札幌や北海道などの地名がツ

イートされていること、テレビ番組でセイコーマートなどの北海道の小売店舗では豆パンが人気であることが放送されたこと、さらにセイコーマートのホットシェフという店頭で調理する商品の中でカツ丼が一番人気であることがバイグラムに表れているためと考えられる。

このように、バイグラムでは分析するテキストデータ中の形態素の共起関係を概観することができる。また、その共起関係からどのような形態素の連続、つまり文章が多数出現しているのかを確認することができる。つまり、各コンビニエンスストアのブランドをキーワードとして取得したツイートから抽出した形態素の集合からバイグラムを作成することによって、どのような形態素の連続が多くなっているのかを確認することができる。またその要因に関連する資料から考察することによって、各コンビニエンスストアのブランドが扱っている商品やサービス、キャンペーンの中で消費者から意識され、注目されているのは何であるのかを探ることができるのである。さらに、将来に向けて消費者から注目され売上を拡大することができる商品やサービス、キャンペーンについて検討することも可能であろう。

## 8. おわりに

本論文では、国内の主要なコンビニエンスストア8ブランドに関わるTwitterのツイートをAPIを利用して取得するとともに、取得したツイートからテキストデータを抽出しテキストマイニングならびに統計的な分析を行うことによって、各コンビニエンスストアのブランドに対して消費者がどのようなことに注目し、またどのような意識を持っているのかを検証した。Twitterのツイートの取得には、数理統計パッケージRにTwitterのAPIを利用するためのパッケージであるTwitterをインストールして使用した。ツイートの取得は、国内の主要なコンビニエンスストアのブランドをキーワードとして2017年7月初旬に行った。本論文では、数理統計パッケージRをはじめとして、オープンソースなどのフリー

なソフトウェア資源を使ってインターネット上のビッグデータを取得し、さらに取得したデータを分析して新たな知見を得るという一連のプロセスを実験的に検証している。この一連の試みも本論文の一つの目的である。

本論文で行ったコレスポネンス分析および階層型クラスター分析の結果から国内の主要なコンビニエンスストア8ブランドをその商品・サービス構成、顧客対応、キャンペーン・イベント、店舗運営、消費者の意識などに関連して複数のグループに分類することができた。また、分類したグループごとの特徴や消費者の意識を明らかにすることができた。本論文の結果として分類した各コンビニエンスストアのブランドの特徴を詳細に考察することによって、各ブランドの強みを確認することが可能である。同様に各ブランドの弱点を見出すことも可能である。さらに、本論文の結果から消費者の注目する商品やサービス、キャンペーンなどを明確にすることができた。したがって、本論文の結果を詳細に検討することによって、各コンビニエンスストアのブランドが消費者から注目され、意識されるためには何をなすべきかが明確になってくる。また弱点を克服する方向性を明らかにすることが可能となる。

一方、本論文の一連の分析で使用した主な研究ツールは、オープンソースを中心としたフリーなものである。このことは、経済的なコストを必要としないフリーな研究環境であってもインターネット上のビッグデータを取得、分析して有用な知見を得ることができることを実証している。インターネット上のビッグデータに関わる研究を推進するにあたって、必ずしも高価な研究環境を整備することが必須ではないことを明らかにした。

ただし、本研究では、テキストマイニングの形態素分析にあたって、不必要な形態素を削除するプロセスに一定の課題が浮き彫りになった。つまり、形態素分析にあたっては、品詞や記号などによって機械的に不必要な形態素を削除するだけでは不十分であり、形態素の意味を理解するとともに、また時節に応じた形態素の意味も検討して分析に不必要な形態素を取捨選択することが必要で

あり、そのための効率的な研究手法の開発が不可欠であることが判明した。今後は、この点に関して十分な配慮と対応をとって研究を進めていかなくてはならない。

本論文では、各コンビニエンスストアのブランドに対する消費者の意識を明らかにするとともに、各ブランドを複数のグループに分類することができた。今後は、各ブランドの売上高や利益などの企業業績と今回の分析結果を詳細に考察することによって、どのような商品、サービス、キャンペーン、店舗運営が企業業績とどのように相関があるのかを考究していくことが課題である。消費者の注目する商品、サービス、キャンペーン、店舗運営と企業業績との相関が明確になれば、今後のさらなる成長を実現するための最適な経営戦略の立案が可能となるであろう。

#### 【注】

- 1) ツイートと呼ぶ140字以内の短文の投稿をユーザーが共有する情報サービスの一つである。
- 2) Twitterを運営するTwitter社は、「社会的な要素を備えたコミュニケーションネットワーク」でありソーシャル・ネットワーキング・サービス (Social Networking Service) ではないとしている。
- 3) コンピュータプログラムのソースコードの公開や自由な再頒布、派生物の自由な利用、原作者のソースコードとの区別など、Open Source Initiativeにより策定された定義に合致しているフリーなソフトウェアのことである。
- 4) コンピュータプログラムがデータなどを取得するためのインターフェイスの仕様のことである。
- 5) ここでいうデータクレンジングとは、テキストデータから分析に不要なデータを除去することである。
- 6) コンピュータプログラムのソースコードをコンパイル、リンク、インストールするための手順やルールを記述したmakefileと呼ばれるテキストファイルを使って、コンピュータプログラムを利用可能にするための一連の自動化された作業のことである。
- 7) Benzecri, J. P., "L'Analyse des Correspondances", 1973.
- 8) 1980年代以降は、例えば、Michael J. Greenacre, "Theory and Applications of Correspondence Analysis", 1984, Sten Erik Clausen, "Applied Correspondence Analysis: An Introduction", 1998. など、コレスポネンス分析に関する英語の文献が発行されている。
- 9) IBM社の統計パッケージ。2009年にIBMがSPSS社を買収し、現在はIBM社が開発、販売を行っている。
- 10) SAS Instituteが開発、販売する統計パッケージである。
- 11) 高根芳雄 (1995) 『制約つき主成分分析法』朝倉書店。
- 12) ユークリッド距離 (Euclidean distance) とは、2地点間の

通常距離のことでありピタゴラスの定理によって計算が可能である。1次元から $N$ 次元まで計算が可能である。ただし、 $N$ は任意の自然数である。

- 13) ウォード法 (Ward's method) とは、偏差平方和が最小となるようにクラスタリングする方法論のことである。

- 14) 任意の文章において、ある形態素が出現した時に別の形態素が同時に出現することである。

- 15)  $X$  を  $N$  回かけ算を行うべき乗のことである。

(2017年10月10日 受稿)  
(2017年10月20日 受理)