消費者の買物に関する話題抽出
ーTwitterにおけるテキストデータを用いた分析からのアプローチー

三坂 昇司
公益財団法人流通経済研究所　研究員

1 はじめに

近年TwitterやFacebook、Instagramなど、SNS（Social Networking Service）の普及により、自身の日常における思ったことをつぶやいたり、日記として書き込んだり、思い出を写真とともにアップロードしたりできるようになった。また、スマートフォンを利用することで、自宅だけではなく、街中、商業施設内、電車の中などあらゆる場所で、いつでも知人や友人とコミュニケーションが取れるようになった。消費者だけでなく、一部の小売業でもSNSアカウントを開設し、消費者とのコミュニケーションや宣伝・販売促進に取り組んでいる。

小売業側が消費者に向けてコミュニケーションを図ったり、宣伝・販売促進に関する情報発信に取り組む一方で、消費者側が小売業に関する情報を発信し、SNS上で話題にすることはあるのであろうか。寺本（2016）が、小売業は情報共有の対象としてSNSの話題になっていないことを問題意識として挙げている通り、消費者が小売業に関する内容を話題にすることは少ないように思われる。寺本（2016）が提唱するように、小売業が「他人に言いたくなってしまう存在」になることにより、競争環境が激化する中で、新規顧客のトライアルを促進したり、ロイヤルティを育成するなどの新しい施策を検討できる可能性があるだろう。

本稿では、Twitterにおけるテキストデータを用いて、消費者が小売業について、どのようなことを話題にしているのか確認する。特に小売業に関する話題の中でも、消費者の行動側面に注目し、「買物」というキーワードを含むツイートについて、分析することとする。現時点で小売業が話題に上ることは、ボリュームとしては少ないと考えられるものの、話題のタイプを把握することは、今後「他人に言いたくなってしまう存在」になるために意味があると考えられる。

2 小売業におけるTwitter活用の概観

本稿の目的である消費者のTwitterデータを確認する前に、小売業がどのようにTwitterを活用しているか概観する。水野他（2013）は、Twitterの11の企業アカウントを対象にツイートの内容を分析し、「感謝」、「挨拶」、「謝罪」、「宣伝」の４タイプに分類できることを示しており、本稿では、この４つのタイプを参考に整理することとした。

Twitterを活用している主な小売業については**図表１**にまとめた。対象とした小売業は売上高ランキング上位30社[1)]のうち、Twitter社より公式アカウント認定がされて

図表1

主な小売業におけるTwitter利用状況

業態	アカウント名（※1）	アカウントID	登録年月	ツイート数（※2）	フォロー数（※2）	フォロワー数（※2）	ツイート内容（※3）					
							コミュニケーション				宣伝	
							返信	感謝	挨拶	謝罪	イベント等	商品紹介
総合スーパー 食品スーパー	イトーヨーカドー	Lets_go_Yokado	2011年9月	5,422	55	83,774	-	-	-	-	○	○
	イオンリテール	AEON_JAPAN	2015年10月	343	1	87,230	-	-	-	-	○	○
コンビニエンスストア	セブン・イレブン・ジャパン・認証済みアカウント	711SEJ	2010年4月	1,838,240	8,274	2,538,465	○	○	-	○	○	○
	ローソン	akiko_lawson	2010年2月	604,741	201,045	1,494,805	○	-	-	-	○	○
	ファミリーマート	famima_now	2009年11月	2,202,323	52,820	808,022	○	○	-	-	○	○
専門店	ヤマダ電機	yamada_official	2010年1月	2,120	190	69,071	-	-	-	-	○	-
	ビックカメラ	biccameraE	2012年8月	5,041	5,490	17,203	○	○	-	-	○	○
	ヨドバシカメラ【公式】	Yodobashi_X	2016年9月	2,808	0	116,611	-	-	-	-	○	○
	ユニクロ	UNIQLO_JP	2009年11月	6,347	3	514,500	-	-	-	-	○	○
ディスカウントストア	驚安の殿堂 ドン・キホーテ	donki_donki	2010年4月	72,087	16,273	36,627	○	○	-	○	○	○

※1　各企業とも、公式アカウントのみ抽出
※2　ツイート数、フォロー数、フォロワー数は2017年7月26日時点の数値
※3　「○」は2017年7月12日から7月25日（2週間）のツイートの中に含まれていた内容

いる小売業とした[2)]。なお、ツイートを確認した期間は2017年７月12日から７月25日の２週間である。

「感謝」、「挨拶」、「謝罪」、「宣伝」の４つについて、水野他（2013）の分類を参考に、「感謝」は「ありがとうございます」といった感謝の言葉が含まれる内容、「挨拶」は「みなさま、おはようございます」のような挨拶の言葉が含まれる内容、「謝罪」は「申し訳ございません」のような謝罪の言葉が含まれる内容、「宣伝」については「イベント等の紹介」と「商品紹介」に分け、各アカウントのツイートにどのような内容が含まれるか確認した。また、４タイプに加え、他のアカウントとコミュニケーションを行っているかどうかも確認するため、返信の有無についても確認した。

確認した小売業30社中、公式アカウントが確認できたのは10社であった。その他の小売業は店舗別にアカウントを開設していたり、Twitterではなく他のSNSを活用している小売業もあった。

Twitterの登録年月について、2010年前後にアカウント開設を行っている小売業が多いことが確認できる。調査時点でのツイート数は、コンビニエンスストア３社が他の小売業よりも多い。これは、登録年月による違いだけでなく、小売業が実施しているキャンペーンに応募したアカウントへの返信なども含まれ、ツイート数が多くなっていることによるものである。

ツイートの内容について、「イベント等の紹介」は10社中すべて、「商品紹介」は10社中９社がツイートしており、「宣伝」として活用する小売業が多いことが分かる。店舗でのイベントやセールの情報を発信したり、こだわりのある商品や注目の商品を紹介することが、Twitterの一般的な活用方法であることを改めて確認できる。一方で、他のアカウントへの「返信」や「感謝」などのコミュニケーションに関する内容については、各企業によって傾向が異なるようである。

3　分析方法

一部の小売業においては、積極的にTwitterを活用する動きが見られる一方で、消費者はTwitter上でどのようなことを話題としているのであろうか。ここからは「買物」という単語を用いて、一般的にどのようなことが話題とされているのかを確認する。

[1] 本分析に用いるデータ

本分析では、分析に用いるデータとして、NTTデータ社のサービスである「なずきのおと」を利用し、2016年10月25日から10月31日までの１週間のTwitterデータ74,940件（テキストデータ）を取得した。抽出に用いたキーワードは「買物」で、この「買物」という単語が本文に含まれるデータを分析対象とした[3)]。なお、Twitterデータは、いわゆるビッグデータであり、利用者の膨大な量のツイートが日々蓄積されている。分析を効率的に行うため、本分析ではランダムに５％抽出を行った。

[2] データクリーニングの条件

上述の通り、「買物」をキーワードにTwitterデータを取得した。データクリーニングの主な条件は、次の通り設定した。

本分析においては、「bot」や「RT」などリツイートのデータは削除した。ただし、「bot」についてはアカウント名に含まれている場合も削除の対象とした。また、ツイートへのリプライについては、引用部分を示す「Re」以降の文章を削除した。その他にも、他のWebサイトの引用にあたるURL（「http://」や「https://」から始まる文字列）は削除し、「Re」や「RT」を含まない同一文章の投稿は対象外とした。また、キーワードのみをツイートしているアカウントについても削除し、公序良俗に反するような不適切なツイートについては、目視作業も含めて削除を行った。

これらの手順にてクリーニングを行った結果、分析対象となるデータの件数は、45,501件（取得データの60.7%）となった。

[3] 分析手順

本分析では、分析の手順として２段階の分析を設定した。

分析１は、潜在意味解析とクラスター分析を用い、話題の抽出を行うことを目的とした。潜在意味解析（LSA：latent semantic analysis）とは、言葉の同義性や多義性に対処するために発展した統計的技法であり、複数の単語の背後に共通して潜在する意味構造を抽出するために用いられる（豊田編著2008）。潜在意味解析を実施するにあたり、ターム・文書行列（term document matrix）を作成する必要がある。本分析では形態素解析の段階で、名詞、形容詞、動詞の３つの品詞に限定し、助詞や副詞、感嘆符など、分析のノイズとなる単語は除外した。共起頻度が２以上のターム・文書行列を作成後、特異値分解によって単語の用いられる次元を集約した。その後、k-means法によって単語をクラスター分析によって分類し、得られたクラスター１つずつを話題として定義した。なお、本分析においては、解釈しやすいよう、クラスター数を10に設定した。

分析２では、「買物」に関するツイートの中で、買物における感情やカテゴリーに関連する単語と関係が強い単語が、どのようなものか確認することを目的とした。設定したキーワードと関係が強い単語を確認することで、感情を喚起するための単語やカテゴリーにおける話題になるためのヒントを検討することができると考えられる。分析手順としては、特異値分解によって得られた単語ベクトルに対し、コサイン尺度を用いて単語間の類似度計算を行った。本分析で用いる施策に関連する単語は、感情に関する単語として「楽しい」、「テンション」、「感動」、「残念」をキーワードとして設定した。また、買物における購買頻度が高く、計画性が高いと考えられるカテゴリーの単語として「野菜」、「魚」、「肉」、「酒」をキーワードとして設定した。

なお、分析に用いたソフトウェアはRで、形態素解析にはMeCabを用いた[4]。

4 分析結果

[1] 分析1：潜在意味解析による話題抽出

分析１で得られた10のクラスターを**図表２**に示した。

クラスター10には、本分析では記述しきれない様々な固有名詞など、全体における登場頻度が低い単語が多く存在している。このクラスター10を「その他」の話題に用いられる単語として扱い、また、含まれる単語が１となったクラスター９も除き、本分析では、比較的登場頻度が高い単語が含まれるクラスター１からクラスター８までを詳細に確認することとした。

図表３に各クラスターに含まれる主な単語と例文（ツイートの一部）を記した。なお、図表３における「話題」は分析結果より命名した。

クラスター１は、「今日」、「買う」、「私」、「服」などの単語が含まれるクラスターであった。これらよりクラスター１は「衣服の買物」の話題であると考えられる。クラスター２は、「帰る」、「一緒」、「店」、「寒い」などの単語が含まれるクラスターであった。これらよりクラスター２は「帰宅時の出来事と買物」の話題であると考えられる。クラスター３は、「人」、「時間」、「できる」、「仕事」などの単語が含まれるクラスターであった。これらよりクラスター３は、「仕事と買物」の話題であると考えられる。クラスター４は、「あと」、「食べる」、「自分」、「可愛い」などの単語が含まれるクラスターであった。これらよりクラスター４は「直近の予定としての買物」の話題であると考えられる。クラスター５は、「行く」、「いく」、「明日」、「スーパ

図表2

単語のクラスター分析結果と所属する単語の割合

クラスターNo.	含まれる単語の数	単語の構成割合	
			クラスター9,10除く
クラスター1	18	0.2%	15.9%
クラスター2	17	0.2%	15.0%
クラスター3	17	0.2%	15.0%
クラスター4	9	0.1%	8.0%
クラスター5	10	0.1%	8.8%
クラスター6	17	0.2%	15.0%
クラスター7	10	0.1%	8.8%
クラスター8	15	0.2%	13.3%
クラスター9	1	0.0%	－
クラスター10（頻度が低い語句の所属）	7,676	98.5%	－
合計	7,790	100.0%	－
合計（クラスター9,10除く）	113	1.5%	100.0%

※合計（クラスター9，10除く）の構成割合は合計7,790に対する割合

図表3

クラスター分析結果のまとめ

クラスターNo.	話題	含まれる主な単語	各クラスターにおけるツイート文のイメージ
クラスター1	衣服の買物	今日　買う　私　服　カード　家　etc.	今日ストレス解消に服買ってしまったー
クラスター2	帰宅時の出来事と買物	帰る　一緒　店　寒い　作る　高い　車　etc.	帰りに一緒に買物に行った
クラスター3	仕事と買物	人　時間　できる　仕事　ポイント　考える　etc.	仕事帰りに買物できる時間はなさそう
クラスター4	直近の予定としての買物	あと　食べる　自分　可愛い　今　etc.	何も食べるものがないからあとで買物に行ってこよう
クラスター5	明日の予定としての買物	行く　いく　明日　スーパー　ご飯　とき　来る　etc.	明日の朝ご飯ないからスーパー行こう
クラスター6	嫌々な買物	円　前　雨　降る　購入　寝る　気づく　etc.	雨が降る前に買物しないと・・・
クラスター7	期待以上の買物	笑　すぎる　楽しい　好き　万　昨日　みたい　etc.	昨日いっぱい買物したー、ほんと楽しすぎた（笑）
クラスター8	（拡散した話題）	問題　理系　文系　答え　分かれる　まとめ　おつり　etc.	文系と理系で買物おつりの計算は分かれるーまとめサイトー
クラスター9	-	する	-
クラスター10	（その他）	（固有名詞や人名などその他多数の単語）	-

ー」などの単語が含まれるクラスターであった。これらよりクラスター５は「明日の予定としての買物」の話題であると考えられる。クラスター６は、「円」、「前」、「雨」、「降る」などの単語が含まれるクラスターであった。これらよりクラスター６は「嫌々な買物」の話題であると考えられる。クラスター７は、「笑」、「すぎる」、「楽しい」、「好き」などの単語が含まれるクラスターであった。これらよりクラスター７は「期待以上の買物」の話題であると考えられる。クラスター８は、「問題」、「理系」、「文系」、「答え」などの単語が含まれるクラスターであった。これは分析期間中にTwitterで拡散した雑学系の話題であると考えられる。

図表２と合わせて含まれる単語の構成割合を確認すると、クラスター１からクラスター８の中で最も構成割合が高かったのはクラスター２で、最も構成割合が低かったのはクラスター５であった。

[2] 分析２：買物における感情やカテゴリーに関連するキーワードと関係が強い単語

分析２で得られた感情、カテゴリーに関するキーワードと関係が強かった単語を**図表４**にまとめた。

「楽しい」と特に関係が強かった単語は、「揃える」、「憂鬱」、「日帰り」、「浮かれる」であった。このうち小売業と関係がありそうな単語は、「揃える」と「憂鬱」であると考えられる。「揃える」は「品揃え」に関する内容であり、「品」と「揃える」が分かれて認識されたものである。品揃えによって、楽しさを高め、Twitterによる発信を行う行動につなげられる可能性が示唆される。一方、「憂鬱」は、「楽しい」と正反対の単語であり、反動があることを表していると考えられる。買物は憂鬱な気分を晴らし、楽しくすることでツイートに至る可能性があるだろう。

「テンション」と特に関係が強かった単語は、「やる気」、「安売り」、「出会う」、「チョコレート」、「カラフル」であった。このうち小売業に関係がありそうな単語は、「安売り」、「チョコレート」、「カラフル」である。「安売り」は、消費者のテンションを高める機能があると考えられる。「チョコレート」、「カラフル」については、テンションと関係しているとも考えられるが、分析期間がハロウィンの直前であることから、ツイートされたことも考えられる。この点についてはさらなる分析が必要であろう。

「感動」と特に関係が強かった単語は、「キャンディ」、「デザイン」、「三井」であった。このうち小売業に関係がありそうな単語は、「キャンディ」、「三井」である。「キャンディ」カテゴリーでは感動体験を創造できる可能性がある。「三井」については「アウトレット」とも関係しており、アウトレットモールのことを意味していると考えられる。訪問時の感動体験がツイートされたのであろう。

ここまではポジティブな感情に関する単語との関係を確認したが、ネガティブな感情として「残念」についても確認を行った。特に関係が強かった単語は、「急」、「接客」、「スタッフ」、「おとなしい」、「天井」、「柚子」であった。このうち小売業に関係がありそうな単語は、「接客」、「スタッフ」、「天井」、「柚子」である。「接客」、「スタッフ」については、小売業のスタッフの対応によっては「残念」な感情を生起させ、ツイートによって拡散してしまう危険性があると考えられる。「接客」、「スタッフ」の教育を強化する小売業が増えているが、方向性としては間違っていないと考えられるだろう。一方、「天井」、「柚子」が関係の強い単語として挙がってきたことも興味深い。これらについては、飲食店だ

図表4

感情、カテゴリーに関するキーワードと関係が強い単語

キーワード		関係が強い単語		
感情	楽しい	揃える 浮かれる	憂鬱	日帰り
	テンション	やる気 チョコレート	安売り カラフル	出会う
	感動	キャンディ	デザイン	三井
	残念	急 おとなしい	接客 天井	スタッフ 柚子
カテゴリー	野菜	カジュアル 玉ねぎ	餅 キノコ	高騰 すき焼き
	魚	醤油 冷凍	塩 冷蔵	蟹
	肉	すき焼き グラタン	焼肉 コンソメ	ワイン ピーマン
	酒	ナッツ	餃子	

けでなく、食品スーパーでも惣菜などで提供されることもある。盛付けなどにより見た目が悪いと、「残念」という感情を生起している可能性があるだろう。

買物における購買頻度が高く、計画性が高いと考えられるカテゴリーの単語として、「野菜」、「魚」、「肉」、「酒」についても確認を行った。

まず「野菜」と関係が強かった単語については、「カジュアル」、「餅」、「高騰」、「玉ねぎ」、「キノコ」、「すき焼き」が確認された。野菜は食品スーパーでは入口近くに陳列されているため、一般的には消費者に季節感や鮮度感に影響を与える。加えて、色彩も鮮やかであるため、カジュアルな印象がツイートされている可能性がある。

「魚」と特に関係が強かった単語は、「醤油」、「塩」、「蟹」、「冷凍」、「冷蔵」であった。「蟹」は食品スーパーの店頭では鮮魚コーナーの中心ではないものの、ツイートという観点では鮮魚コーナーの中心商材となりえると解釈することもできるかもしれない。

「肉」と特に関係が強かった単語は、「すき焼き」、「焼肉」、「ワイン」、「グラタン」、「コンソメ」、「ピーマン」であった。「すき焼き」は「野菜」にも関連がある単語として確認されていることから、消費者の情報発信という観点からも、２つのカテゴリーをつなぐメニューであると考えられる。「ワイン」、「グラタン」、「コンソメ」についても同様に情報発信につながる商材、メニューであると考えられる。

「酒」と特に関係が強かった単語は、「ナッツ」、「餃子」であった。これらからは「おつまみ」が連想できることから、ともにツイートされやすい単語であると考えられる。

5 まとめ

[1] 本分析のまとめ

本稿では、寺本（2016）が提唱した、小売業が「他人に言いたくなってしまう存在」になるための基礎分析として、消費者がどのようなことを話題にしているのか、Twitterにおけるテキストデータを用いて確認した。

まず分析１では、ツイートされている話題を潜在意味解析とクラスター分析によって確

認した。登場頻度が低いその他の単語が非常に多かったものの、一部の特徴的な話題を確認することができた。商品、あるいはカテゴリーに関する話題は、クラスター１の衣服の買物に関する話題のみ確認することができた。クラスター２からクラスター６は、日常の買物に関する話題であったが、買物の中身を伝えるような話題ではなかった。クラスター７に見られるように「楽しい」や「好き」+「すぎる」といったような期待を超えるような経験が話題として発信されることも、本分析において確認できた。

次に分析２では、買物における感情やカテゴリーに関するキーワードと関係が強かった単語を確認した。感情に関するキーワードと関連する単語では、その感情を高める施策の方向性について、確認することができた。ネガティブな感情である「残念」と関連する単語からも施策の方向性を確認することができた。一方、カテゴリー名称と関連する単語からは、カテゴリーと関連するメニューや商材を確認することができた。近年は、ID-POSデータを活用したクロスMD企画の検討が行われているが、購買実績のデータだけでなく、消費者による情報発信という観点から、クロスMD企画を検討できる可能性が示唆された。

[2] 本分析における課題

本分析における課題、ならびにテキストデータの分析における課題も記述しておく。

近年、テキストデータの分析については、様々な手法が提案されている。本分析においては、Twitterデータを分析するための最適な手法について検討を行っておらず、一般的に用いられている潜在意味解析を用いた。分析に用いる手法の詳細を再検討することでより精度の高い話題を抽出できる可能性がある。

また、形態素解析においては、一般に提供されている分析ソフトウェアであるMeCabを用いた。MeCabは辞書と呼ばれる基本的な単語の分類機能を利用して、形態素解析を行っている。本分析においては、辞書整備にかかる莫大な労力やコストを節約し、独自に辞書を整備していない。従って、日々生成される新語や仲間同士のコミュニケーションに用いられる特殊な単語、専門用語などはカバーできていない可能性がある。これは、本分析における限界であると言える。

加えて、Twitterデータのクリーニングについても、さらに精度を高められる余地がある。Twitterから取得したデータには、多くの分析に値しないデータが存在する。サンプリングしたデータを用いたとはいえ、データクリーニングの段階で分析に値しないデータをすべて除外することは不可能であり、本分析においても完全にデータクリーニングできているとは言い切れない。このことについても、今後の分析における課題として残されている。

〈注〉

1） 売上高ランキングは『DIAMOND Chain Store』2016年9月15日号を参照した。
2） 各社の状況は2017年7月26日時点の数値を記載した。
3） 「買物」の他に、「買い物」「買いもの」をキーワードとして設定している。
4） R上でMeCabを操作できるよう、分析パッケージとしてRMeCabを用いた。

〈参考文献〉

寺本高（2016）「『ネタ』になる売場とは?ー売場情報のSNS投稿と反応の関係ー」『流通情報』第519号、pp56-67

豊田秀樹編著（2008）『データマイニング入門ーRで学ぶ最新データ解析ー』東京図書

水野誠、高階勇人、新保直樹 (2013)「Twitter を用いた顧客とのコミュニケーション：対話と拡散（＜特集＞声・質的データを経営・マーケティングに活かす）」『オペレーションズ・リサーチ: 経営の科学』、第58巻第8号、pp427-435