

楽天市場における

コーヒー関連商品レビューの統計・分析

s17161430 高田 真成

s17161609 山根 拓真

目次

1. はじめに
 - 1-1 研究動機
 - 1-2 楽天市場について
2. 研究方法
 - 2-1 目的と分析手法
3. 研究結果
 - 3-1 形態素解析を用いた分析
 - 3-2 テキストマイニングツールを用いた分析
4. おわりに
 - 4-1 まとめ
 - 4-2 今後の展望

1.はじめに

1-1 研究動機

昨今、感染症による影響でオンラインショッピングは以前よりはるかに広く浸透している。商品が届くまでの時間、価格、商品数など日々進歩し続け、いまや我々にとって身近で、生活に欠かせないものとなっている。そんなオンラインショッピングにおいて、購買意欲に寄与する一因がレビューである。私たちはレビューに対する理解をより一層深めたいと考え、テーマを選定した。

その中でも楽天市場について研究したいと考えた際、はじめはこのテーマではなく、俗にいうサクラレビューという、ある目的をもって作られたレビューを見つけてその傾向や使用単語を分析しようと考えていた。しかし、すでにネット上にサクラレビューかどうかを判断できるサイトが存在したこと、レビューの有意性の調査よりも、レビューの役割について掘り下げていきたいという考えなどがあり、このテーマは破棄された。そこで、研究がしやすい程度のレビュー数に絞り込めるコーヒーという単語でレビューを 11759 個に絞り込み、あらゆる視点からレビューデータを分析し、レビューデータに対し深く理解を得て、その中から特徴的なものを探していこうと研究を進めた。その結果、様々な成果を得たためまとめていこうと思う。

1-2 楽天市場について

楽天市場は出店数 41,568 店（2015 年 3 月末）国内最大の電子商店街である¹。楽天トラベル、楽天デリバリー、ラクマ、東北楽天ゴールデンイーグルス（野球）、ヴィッセル神戸（サッカー）など多種多様な業界に進出している楽天株式会社の祖となる通販サイトであり、サイト設立は 1997 年 5 月 1 日となっている。

¹ 「楽天」ウィキペディア

<https://ja.wikipedia.org/wiki/%E6%A5%BD%E5%A4%A9%E6%A5%BD%E5%A4%A9%E5%B8%82%E5%A0%B4>

2.研究方法

使用するレビューデータは次の項目から構成される。ユーザー名、年齢、性別、商品ページの URL、商品名、出品者による商品詳細ページの URL、価格、購入目的（サイト上の選択肢から選択）、誰のために購入したか（サイト上の選択肢から選択）、過去に同じ商品を購入した回数、評価（5段階）、レビュータイトル、レビュー本文、レビュー投稿日時、以上 15 項目である。今回のデータは 2010 年 1 月 1 日～31 日までのデータとなっている。レビュー一つ書くのにかなりの項目をユーザーが入力していることにも驚いたが、この中からまず初めに「購入目的」に注目した。

購入目的は、「実用品、普段使い」「プレゼント」「ビジネス」「おもたせ、ギフト」「趣味」「イベント」「空白（回答無し）」以上の 8 項目に分かれている。コーヒーという単語で分析対象を選定しているため、圧倒的に「実用品、普段使い」とするユーザーが多数であった。そのレビュー数は 9653 個になった。次いで「空白（回答無し）」が 903 個、「プレゼント」が 518 個、「おもたせ、ギフト」292 個、「趣味」206 個、「ビジネス」120 個、「イベント」67 個となった。空白はおそらく指定することが煩わしいと感じたユーザーによるものであろうが、実用品以外のどの項目よりもデータ数が多く、これほどの数のレビューデータを無視するのは少し問題があるという判断に至り、あまり傾向などを参考にはしていないが念のために他の項目と同じ分析を行った。

2-1 目的と分析手法

以下の分析の目的は、用途と頻出する単語を比較することで、その用途における特徴的な単語を見つけることである。この分析において、実用品においては値段やコストに関する単語が、ギフト・ビジネス・プレゼント等の人に贈与する場合には味を重視し、それに関連する単語が頻出するという仮定を立てた。

分析手順は、mecab という形態素解析ができるツールを使い単語に分解し、その単語が使われているレビューの個数を求め、そこから用途別の個数も分類する。それらをグラフ化して比較を行う。形態素解析とは、自然言語処理の分野で行われる文学的ではない自然言語のテキストデータを品詞ごとに分解することである。

3.研究結果

3-1 形態素解析を用いた分析

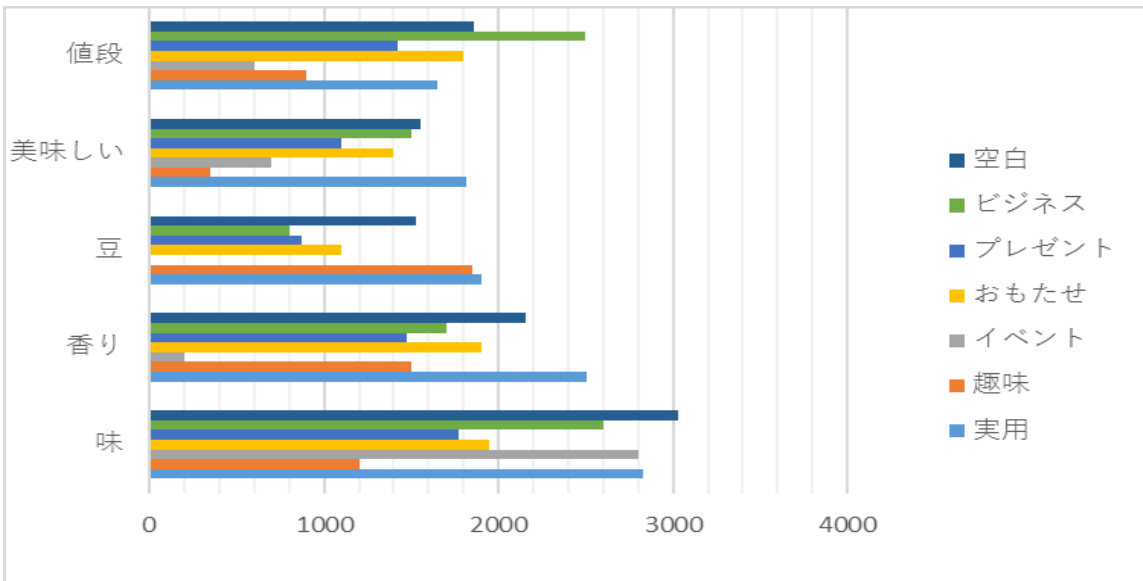
結果は以下である。

		82%	2%	1%	2%	4%	1%	7%
全体		実用	趣味	イベント	おもたせ	プレゼント	ビジネス	空白
11621	し	9284	218	96	315	524	122	763
8738	コーヒー	7200	158	40	168	341	77	561
6307	い	5181	119	29	114	244	65	406
4816	の	3899	77	32	107	180	37	339
4745	購入	3834	74	30	112	189	40	333
2786	味	2320	24	28	39	71	26	212
2680	思い	2114	56	17	73	120	27	200
2533	この	2100	27	14	43	85	28	186
2395	香り	2051	30	2	38	59	17	151
1945	とても	1538	28	17	65	116	15	116
1794	豆	1559	37	#N/A	22	35	8	107
1794	飲ん	1554	15	#N/A	16	39	7	117
1741	美味しい	1491	7	7	28	44	15	109
1660	値段	1355	18	6	36	57	25	130

意味のない単語である「し」や「い」などは当然ではあるが考慮しない。表を見れば明らかのように、実用に関連した単語の出現が圧倒的に多く、絶対出現数で比較することは意味をなさない。そこで、表の上部のパーセンテージに注目する。これは全体のデータ数、11759個に対する各項目の占める割合である。この割合で各数値を割った数値が以下である。この処理により各数値を同じ条件で比較することが出来るようにした。

	実用	趣味	イベント	おもたせ	プレゼント	ビジネス	空白
し	11321.95	10900	9600	15750	13100	12200	10900
コーヒー	8780.488	7900	4000	8400	8525	7700	8014.286
い	6318.293	5950	2900	5700	6100	6500	5800
の	4754.878	3850	3200	5350	4500	3700	4842.857
購入	4675.61	3700	3000	5600	4725	4000	4757.143
味	2829.268	1200	2800	1950	1775	2600	3028.571
思い	2578.049	2800	1700	3650	3000	2700	2857.143
この	2560.976	1350	1400	2150	2125	2800	2657.143
香り	2501.22	1500	200	1900	1475	1700	2157.143
とても	1875.61	1400	1700	3250	2900	1500	1657.143
豆	1901.22	1850	#N/A	1100	875	800	1528.571
飲ん	1895.122	750	#N/A	800	975	700	1671.429
美味しい	1818.293	350	700	1400	1100	1500	1557.143
値段	1652.439	900	600	1800	1425	2500	1857.143

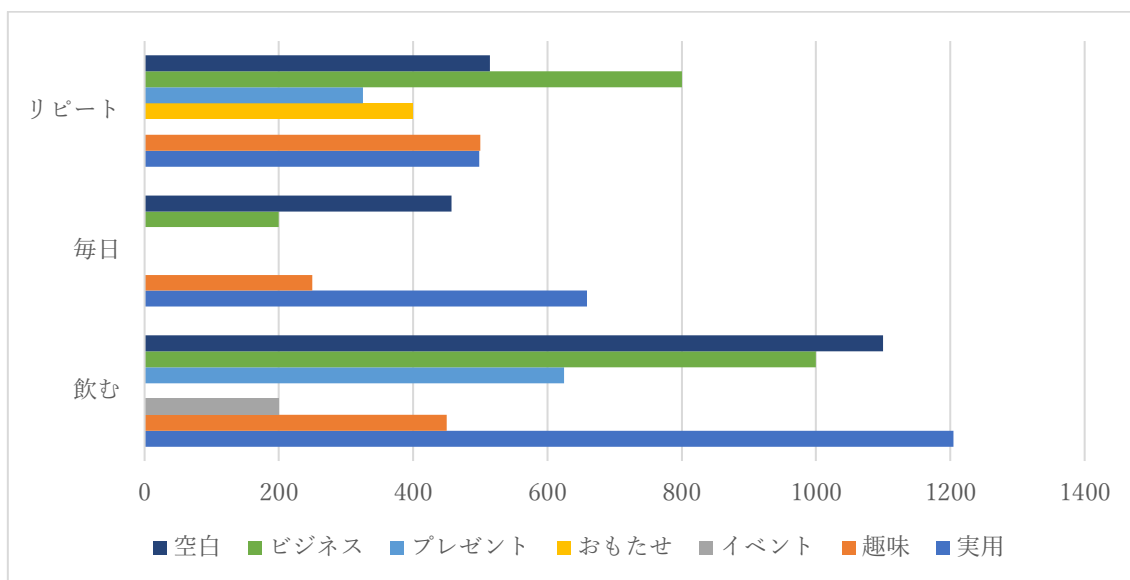
こうすることでデータとして使えるようになったため、これを項目ごとの上位の単語をグラフに表したものが以下である。



今回の分析においてノイズになる動詞「し」や「思い」を除いた、上位5つの単語をグラフ化した。こうしてグラフに表すと、「豆」という単語においては、実用品、普段使い、趣味の目的のレビューに多く見られ、おそらく各人のこだわりがレビューに詰まっているのだろうと思われる。反対に、味に関する単語においては、趣味用途でのレビューが他の用途と比較して少なく、豆を挽くまでの過程を注視し、味に関する考えを書かないレビューが多いとも推測できる。「値段」という単語においては、かなり突出してビジネス用途のレビューにおいて頻出していることがわかる。このことから、一個人が消費するよりも多く消費されるビジネスの現場では、値段や味へのこだわりは強いのだと考えられる。

これらの結果から、分析前に立てた仮定との差異が読み取れた。実用品においては値段を、ビジネスやプレゼントにおいて味・質を重視するものと仮定していたが、値段という単語がビジネス用途のレビューにおいて他用途よりも出現回数が多いこと、味という単語においてはおもたせ・プレゼント用途の数が実用などよりも少ないという結果が出た。

また、上位以外の単語でコーヒーの消費において特徴的でありそうな単語「レポート」「毎日」「飲む」の3語を選定し比較したところ、以下の結果も得られた。



ビジネス用途での「レポート」の出現が多く、イベント用途では出現していない。「毎日」という単語はイベント、おもたせ用途での出現がなく、コーヒーの消費という面においては妥当と思える結果が出た。また、「飲む」という単語がおもたせ用途では出現していないが、この用途ではレビュー本人が飲むという行動を取るわけではないため、このような結果が得られたと考えられる。

以上の分析手法は、既存商品の改善もしくは新規市場参入の際の計画に利用できると考えている。各用途でどんな単語・要素が出現しているかを比較することで、購入者がどのような要素に注目しているかを窺い知ることが出来るためである。今回の分析の結果から言えば、ビジネス用途の購入者をターゲットに含める場合、レポートし易い値段設定を行う必要があるようだ、といった具合である。

ここまでの特徴的な単語の抽出は私見によるものであったが、テキストマイニングツールを併用した分析についても記す。テキストマイニングとは、ビッグデータ解析においてマストな、ありふれた自然言語のテキストデータから有益な情報を取り出す手法である。今回はネット上で一般公開されているテキストマイニングツール²を使用した分析を行った。利用するデータはショップ「加藤珈琲店」と「澤井珈琲」各店の評価 5・4 のデータである。加藤珈琲店側は 140561 文字、澤井珈琲店側は 131452 文字のデータを使用する。このサイトは、単語出現数の計上、品詞の分類を行うのはこれまでと同様であるが、分析の際、TF-IDF 法による単語のスコア付けを行ってくれるのが特徴である。

TF-IDF は、文書に含まれる単語の重要度を評価する手法の 1 つであり、主に情報検索やトピック分析などの分野で用いられている。TF-IDF は、TF (英: Term Frequency、単語の出現頻度) と IDF (英: Inverse Document Frequency、逆文書頻度) の二つの指標に基づ

²AI テキストマイニング by ユーザーローカル <https://textmining.userlocal.jp/>

いて計算される³。

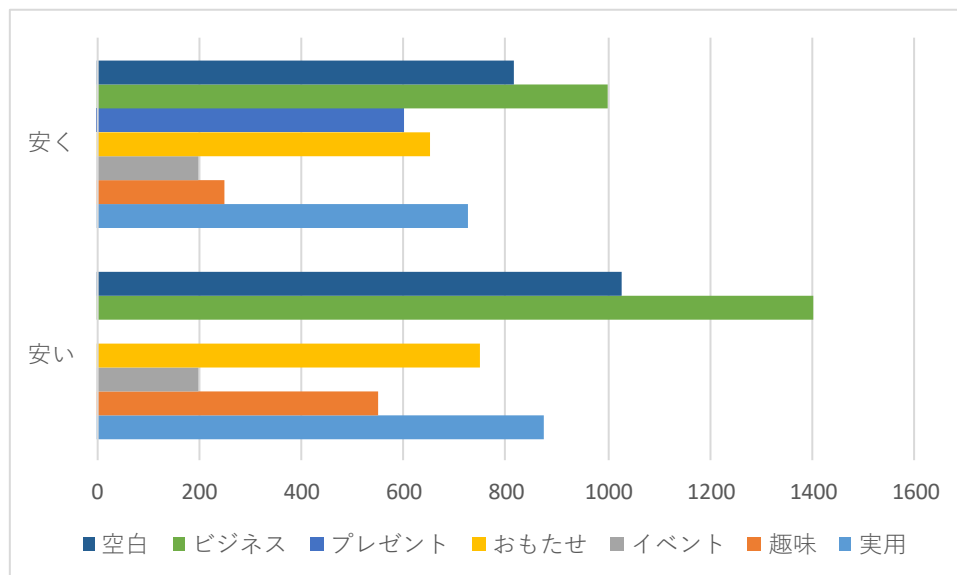
加藤珈琲店と澤井珈琲の単語がスコア付けされたものが以下である。

形容詞	スコア	出現頻度	形容詞	スコア	出現頻度
おいしい	960.34	395	おいしい	996.57	405
美味しい	854.62	575	美味しい	881.31	586
安い	236.01	173	安い	761.08	362
飲みやすい	129.53	42	飲みやすい	272.81	72
良い	93.69	278	お安い	240.24	59
お安い	69.46	24	良い	94.33	279
濃い	57.47	46	いい	40.61	238
香ばしい	32.82	13	濃い	39.80	37
よい	26.31	116	よい	30.38	125
いい	25.11	186	うれしい	29.51	42

澤井珈琲

加藤珈琲店

「おいしい(美味しい)」に次いでスコアが高い単語は「安い」となる。ここで使用しているデータは評価が5及び4のデータであるため、安いことが評価されているレビューが多いと推測できる。ここで、用途別の単語分類データの「安い」をグラフ化する。



出現割合が高いのはビジネス用途、空白(記入なし)に次ぐのは実用・普段使用用途となった。以上の結果から、「安い」という単語がレビューにおいて特徴的であるという結果を得られたショップは、ビジネス用途での消費を検討している既存及び将来的な顧客へのアプローチを強めるべきだ、といった予想が立てる事が可能ではないだろうか。テキストマイニングツールを利用する際には、こうした数値や割合との関係についても注視するべきだろう。

³ 「tf-idf」ウィキペディア <https://ja.wikipedia.org/wiki/Tf-idf>

3-2 テキストマイニングツールを用いた分析

先ほど評価5・4のレビューで行った分析を、すべてのレビューに対して評価別に行った。この分析はテキストマイニングツールで評価された特徴的な単語類は、本当にその店舗や商品の特征と合致するのかを検証したものである。今回の解析に使用した加藤珈琲店のレビューデータの文字数が約19万文字に対し、澤井珈琲が約13万文字と少し差があるが、このサイトの分析では同等の数に均されるため問題ないことを先に説明しておく。

・「加藤珈琲店」と「澤井珈琲」評価1~3における頻出単語と偏り

katou321にだけ出現	katou321によく出る	両方によく出る	sawai321によく出る	sawai321にだけ出現
深い セット ひどい 冬 求める ラブソング 利用 加藤 ありがたい かわいい もったいない それなり ハズレ ホット リピート 不満 売る	思う 安い 値段 袋 今回 商品 レビュー 珈琲 お安い 使う スーパー 物 きつい いただく くれる 得 量 煎る おく 変わる 思える 持つ 出す 膨らむ	良い コーヒー 美味しい 飲む 味 購入 香り おいしい 豆 いい 挽く 買う よい 入れる 期待 注文 好み 好き 無い 酸味 残念 感じる ミルク 強い 届く できる コク ドリップ 粉 入る	薄い 悪い 飲める ブレンド すぎる つく 毎日 すごい 早い 遅い 使用 3つ かかる 種類 大きい 忙しい 甘い もらう 分かる 待つ 比べる	ごめんなさい すみません 有難う 細かい うるさい っばい 朝 缶 こぼれる 知れる ちやう なくなる 付く 濁る 無くなる 貰う 間に合う うれしい おきにくい かぐわしい こぼれやすい すばらしい めんどくさい わかりにくい 小さい 待ち遠しい 淡い 温い 痛い 美味しい

・「加藤珈琲店」と「澤井珈琲」評価4~5における頻出単語と偏り

katou54にだけ出現	katou54によく出る	両方によく出る	sawai54によく出る	sawai54にだけ出現
加藤 ラブソング まあまあ うわー ただいま ヨロシク	安い 値段 セット 価格 送料 好み お安い 買える スーパー 有難う 早い お買い得 無料 酸味 少ない おめでとう うるさい すばらしい かわいい	ありがとう 美味しい コーヒー 飲む おいしい 思う 購入 香り 良い 味 買う いい 挽く 豆 いただく 珈琲 満足 できる よい 届く しまう 入れる 入る 今回 袋 飲める 注文 好き 商品 得	福袋 おまけ ドリップ かかる 付く 行く 甘い 薄い 忙しい 優しい 有り難い 遅い	食べる クッキー おから ごちそうさま 手軽い

先に説明しておくが、楽天市場のレビューデータはそもそも評価が5のレビューが圧倒的に多く、極端に低い評価のレビューはわずかしかない。しかし、そのことを認識しておけばこの図に示した解析結果は興味深いものが出ていると思う。

まず、純粋に単語の数が違っている。高い評価のレビューが多いことは説明したが、それはつまりデータの文字数も多いということになる。つまり、あらゆる単語が存在しているはずである。しかし、「両方によく出る」の欄は4~5の方が多く単語が出ているが、そ

それぞれの店舗ごとにおける特徴的な単語になるにつれて、頻出単語はどんどん少なくなっている。しかし、反対に評価 1~3 の方はまんべんなく一定の文字数がカウントされている。ここからわかることとしておそらく、評価 4~5 のレビューはレビュー投稿者が皆似たような単語で投稿しているのであると思われる。そして反対に評価 1~3 のレビュー投稿者は、様々な単語を使用して投稿したのだと思われる。こうした結果だけで、低い評価を書いているユーザーの感情が伝わってくるようである。評価 3 のデータも入っているのでポジティブな単語も見受けられるが、ここで評価 1 のレビューの分析結果を見る。

・「加藤珈琲店と澤井珈琲」評価 1 における頻出単語

sawai1にだけ出現	sawai1によく出る	両方によく出る	katou1によく出る	katou1にだけ出現
入れる 届く つく 濁る っばい わかりにくい 温い 痛い 細い 細かい 美味しい お腹 イマイチ 以上 品物 脂 説明 返品 たつ ちゃう できる ふく むらす ちらう わかる 任せる 含む 届ける 待つ 扱う	豆 挽く 飲める	コーヒー 美味しい 思う 買う よい 良い 購入 飲む 悪い 味 不味い 無い 薄い 粉 今回 焙煎 少ない 高い しれる 書く 全部 期待 くれる すぎる 合う 感じる 訳	安い 香り 商品 入る 袋 しまう 言う いい 残念 リピ	おいしい セット 値段 きつい ひどい もったいない 濃い 物 使う 思える 捨てる 開ける ミル 以前 普段 最低 おもい お安い くさい ものすごい 使いにくい 厳しい 強い 怖い 新しい 酷い 黒い お勧め しない せい

評価 1 のレビューのみを比較すれば、大幅にポジティブな単語が減り多様な罵詈雑言が見られ、より直接的に感情が伝わると思う。

次に、加藤珈琲店と澤井珈琲それぞれのレビューにのみ見られる単語を見ていくと、まず加藤珈琲店のレビューには「値段」という単語が多く見受けられる。先に説明していたように評価の高いレビューが多いので、値段が高い、ではなく安いのであると考えた。そこで、レビューデータの価格の項目を参照し同じコーヒー豆（ブルーマウンテン）で値段を比較してみたところ、澤井珈琲は 500g2000 円。それに対して、加藤珈琲店の値段は 500g5000 円と、約 3000 円もの差で澤井珈琲のほうが安くなっていた。頻出単語として出ていても必ずしもその店舗の特徴になるとは限らないようだ。結果としては、予想に反し澤井珈琲のほうが安くなっていた。

次に、澤井珈琲にのみ「福袋」という単語が見られているのはなぜか見ていく。説明しておくのと加藤珈琲店にも福袋は存在しているが、なぜ澤井珈琲に多く出ているのか。これも比較してみると、まず加藤珈琲店は、「冬のラブソング」という商品名のオリジナルブレンド豆、これは解析結果でもよく出ているとされている「ラブソング」という単語の正体である。ほかにもパプアニューギニア、コスタリカ、グアテマラの 4 種類各 500g の計 2 kg で値段は 3148 円となっていた。そして澤井珈琲の福袋はオリジナルブレンドのコーヒー豆が 3 種類 500g、グアテマラが 300g の合計 1,7 kg で 2999 円だった。内容量が違うので加藤珈琲店の値段を澤井珈琲に合わせると 2676 円となり、両者の差は 323 円とわずか

ではあるが、ここでは先ほどの値段の比較結果と違い、加藤珈琲店のほうが安くなっている。さらに商品ページを読み進めると澤井珈琲の福袋は「ショップ・オブ・ザ・イヤー⁴」を受賞しており、値段のわずかな差よりはこちらの功績がより購入される理由であるという結論に至った。

次により一つ一つの単語を詳しく分析した。図は評価2のレビューで使われた単語を双方の偏り具合とともに、品詞ごとに分類した図である。

・「加藤珈琲店」と「澤井珈琲」評価2の単語分類

名詞			動詞			形容詞		
sawai2	単語	katou2	sawai2	単語	katou2	sawai2	単語	katou2
45	香り	55	65	飲む	35	50	良い	50
61	味	39	33	思う	67	54	美味しい	46
57	コーヒー	43	93	感じる	7	0	おいしい	100
29	購入	71	100	いける	0	28	安い	72
60	残念	40	100	すぎる	0	0	お安い	100
64	期待	36	100	知れる	0	0	無い	100
61	値段	39	65	異う	35	100	うるさい	0
21	レビュー	79	78	できる	22	100	かぐわしい	0
0	今回	100	0	書く	100	100	こぼれやすい	0
66	好み	34	100	あける	0	100	よい	0
100	フタ	0	100	いらっしゃる	0	100	嬉しい	0
100	牛乳	0	100	おとす	0	100	忙しい	0
51	袋	49	100	こぼれる	0	100	新しい	0
0	焙煎	100	100	つく	0	0	欲しい	100
0	風味	100	100	ねだる	0	100	淡い	0
57	コク	43	100	もれる	0	0	深い	100
30	豆	70	100	入れる	0	66	いい	34
64	開封	36	100	割る	0	66	少ない	34
0	ドリップ	100	100	合う	0	0	うい	100
0	バッグ	100	100	挽く	0	0	おかしい	100
0	前回	100	100	比べる	0	0	ほしい	100
84	粉	16	100	置く	0	0	まずい	100
0	ミルク	100	100	落とす	0	0	不味い	100
0	リピート	100	100	開める	0	0	可愛い	100

⁴ 5周年記念 箱が壊れるまでカフェ福袋 ムチャしました大赤字セールで送料無料で、このボリュームで驚きの2999円!!しかも、澤井珈琲自慢のコーヒーがさらにもう1袋プレゼント

~!!<https://web.archive.org/web/20100317195008/http://item.rakuten.co.jp/sawaicoffee-tea/445822/>

・「加藤珈琲店」と「澤井珈琲」評価4の単語分類

名詞			動詞			形容詞		
sawai5	単語	katou5	sawai5	単語	katou5	sawai5	単語	katou5
50	コーヒー	50	50	飲む	50	50	美味しい	50
42	購入	58	48	思う	52	48	おいしい	52
51	香り	49	41	買う	59	51	良い	49
47	味	53	59	挽く	41	34	安い	66
36	値段	64	57	いただく	43	43	いい	57
38	豆	62	48	できる	52	49	よい	51
41	珈琲	59	52	届く	48	53	多い	47
44	満足	56	61	入る	39	46	嬉しい	54
43	今回	57	53	しまう	47	34	飲みやすい	66
32	セット	68	51	入れる	49	45	高い	55
45	好き	55	55	飲める	45	52	濃い	48
54	袋	46	55	気に入る	45	36	すごい	64
39	注文	61	48	頂く	52	37	うれしい	53
42	商品	58	54	楽しむ	46	48	お安い	62
56	得	44	51	おる	49	42	無い	58
57	ブレンド	43	59	開ける	41	31	早い	69
39	毎日	61	56	違う	44	30	深い	70
60	ミル	40	44	言う	56	53	欲しい	47
34	価格	66	41	使う	59	27	少ない	73
43	リピート	57	56	喜ぶ	44	38	ありがたい	62
39	リピ	61	54	楽しめる	46	42	強い	58
42	大変	58	49	くれる	51	68	甘い	32
44	量	56	48	なくなる	52	42	楽しい	58
76	福袋	24	36	助かる	64	29	大きい	71
54	楽しみ	46	48	出来る	52	63	香ばしい	37
31	送料	69	58	頼む	42	48	苦い	52
38	コーヒー豆	62	57	つく	43	56	うまい	44
43	家族	57	36	買える	64	59	優しい	41
41	利用	59	50	感じる	50	59	薄い	41
52	プレゼント	48	51	もらう	49	52	濃い	48
33	好み	67	42	探す	58	7	かわいい	93
35	スーパー	65	54	試す	46	56	旨い	44
33	お買い得	67	84	食べる	16	56	美味しい	44
51	毎朝	49	54	持つ	46	48	使いやすい	52
53	以前	47	60	送る	40	67	忙しい	33
18	酸味	82	48	無くなる	52	52	素晴らしい	48
1	加藤	99	30	煎る	70	44	細かい	56
28	無料	72	57	選ぶ	43	71	有り難い	29
54	対応	46	75	付く	25	56	物足りない	44
80	おまけ	20	63	行く	37	29	小さい	71
73	ドリップ	27	54	いく	46	100	手軽い	0
43	期待	57	52	比べる	48	51	悪い	49
47	朝	53	61	あげる	39	22	程よい	78
38	コーヒーメーカー	62	63	ひく	37	56	まずい	44
46	箱	54	39	考える	61	56	新しい	44
45	お嬢い	55	55	わかる	45	71	不味い	29
47	我が家	53	50	くださる	50	12	たまらない	88
44	最高	56	52	いれる	48	50	しやすい	50
40	大好き	60	49	合う	51	50	すばらしい	50
36	ショップ	64	41	すぎる	59	38	こい	62

こちらの結果は、先ほどの低評価レビューとは違いデータの偏りがわかりにくくなって

いる。細かく見ていくと、澤井珈琲側の名詞には先ほども特徴として挙げた「福袋」がやはり挙がっている。他に「おまけ」「ドリップ」といった単語が挙がっている。澤井珈琲側では「おまけ」が名詞の中で最も特徴的であるが、加藤珈琲店側にも単語自体はある程度出現している。このことから澤井珈琲側の「おまけ」は、より高い評価を受けていることがわかる。また、澤井珈琲の商品のおまけについて調べると、ドリップコーヒーがおまけとして少しつく商品があった。このことから「ドリップ」という単語も多い理由がわかった。

次に、加藤珈琲店側に偏っている単語を見ていくと、やはりこちらでも先に特徴として出た「安い」「お買い得」といった単語がある。その他に「酸味」「コーヒーメーカー」「煎る」「かわいい」といった単語が挙がっている。

この中でも「かわいい」に関してはかなりの差が出た。しかし、理由がこの単語分類だけではわからなかったため「かわいい」を含むレビューをいくつか見ると、パッケージ、ドリップパック、中にはコーヒーの入っていた麻袋に対するレビューもあり、以上の点から何か一つの商品が秀でているわけではないと判断した。

次に、「酸味」という単語に関して、先ほど頻出単語としても挙げたオリジナルブレンド「冬のラブソング」の商品ページを見ると、コーヒーの味が香り、コク、甘み、苦み、酸味の5つの観点からそれぞれ説明されていた。そして酸味の項目は5段階中の0、つまり酸味がないということだった。「酸味」という単語が特徴として出ただけで酸味があるというわけではなく、ないことが特徴となっていたのは予想外だった。

また、「コーヒーメーカー」という単語に関しては先ほどの評価2のレビューにおける澤井珈琲側の特徴のある単語から、澤井珈琲のコーヒー以外の関連商品に多く低評価がついているという考察が生かせる。しかし、評価4においては評価2の場合より偏りが少ない。この理由はおそらく、数あるコーヒー関連商品の中で「コーヒーメーカー」に限った結果だからではないだろうか。以上の考察から、「コーヒーメーカー」に関しての高評価レビューは加藤珈琲店側の商品に多く寄せられていることがわかった。

次に、「煎る」という動詞に関しては、コーヒー豆の焙煎についてだろうと考え、現在の両店のコーヒー豆の商品紹介ページを見た。両店とも焙煎にはこだわっていると書かれていたが、少し違いがあった。まず加藤珈琲店の紹介には、少量ずつ時間をかけて焙煎を行うとあった。次に澤井珈琲の紹介には、最新の焙煎機を導入してより深く均等に焙煎ができるとあった。紹介の文字数は澤井珈琲のほうが多く、あまり身近ではない焙煎機の良い部分をわかりやすく伝えようとしていると感じた。一方の加藤珈琲店の紹介文は、豆のこだわりを3つにまとめていて、その1つとして焙煎を説明してあるのだが、3行でまとめてあり、澤井珈琲側の紹介と比べると魅力が伝わってこないと感じた。繰り返しになるが、今回のレビューは2010年1月のデータであり、当時のwebページの状態はわからないのでおそらく当時は加藤珈琲店の方に軍配が上がったとは考えられるが、現在はむしろ逆転していることもありうるのではないかと、という結果になった。

4.おわりに

4-1 まとめ

まず、今回の研究を進める際に感じたことは、研究を二人で行う事のメリットの多さであった。約1年半の間、このテーマで研究を進めてきたが研究発表、進捗の報告などをする際に使用するパワーポイントの作成、またこの論文の執筆の際、一人だと絶対に気づけない視点からの考えや、単純なミスへの気づきが早くかなり快適に研究を進められた。

また、今回の研究自体も進めている際は面白いと感じることがいくつもあった。一つは低評価レビューにおける特徴的な単語の、直接的な表現の多さである。今回は「コーヒー」と言う単語での研究だったためレビュアーの年齢は平均約37歳と高いのだが、その年代の発する「まずい」という単語の持つ意味は決して軽くはないだろう。

分析に際して気づいたことを挙げる。Mecab を利用した形態素解析と用途別分類の分析においては、実用・普段使い用途が全体の8割を占める形となり、他用途では出現していない単語が実用・普段使い用途及び記入なしにおいてのみ頻出しているといったことが見受けられた。最終的に同じ数値の重みで比較してはいるが、絶対数に大きな違いがあるというのは留意せねばならない点である。レビュー記入期間を1か月間ではなく1年など長期に渡るものに設定すれば改善されると考えられるが、その場合膨大なデータ量が更に増加することが懸念される。

次に、テキストマイニングツールを利用したのショップの特徴比較に関して述べる。高評価レビューにおける単語の類似性についてだ。この結果は予想していたものではあったが、「加藤珈琲店」「澤井珈琲」の比較の際、あらゆる単語がどれも大抵が6:4、4:6の範囲に収まっていてあまり多くの特徴を見出すことはできず、また分析した結果もわずかな差でしかないために有用であるかはあまり自信がもてないというのが正直なところである。また、結果と考察の項にも記述したが、加藤珈琲店と澤井珈琲の比較において、「安い」という単語が特徴的と評価された前者であるが、実際の価格の比較においてはそのような特徴は無いと見られた。このことから、テキストマイニングのみで事象を評価・分析することの限界が窺い知れた。例えば、「安い豆と比べて美味しかった」というレビューが見られたが、この場合の「安い」の意味は安い豆より「高い」ということになる。文脈も考慮して分析するのは容易では無いため、1つ目の分析のように他の数値等に関連付けてマイニングの結果を利用すべきだと考えられる。

4-2 今後の展望

まず、使用するデータが2010年のものと古いため、最新のデータセットを使用すべきであったと感じる。2010年かつ期間が1月中と限定されたデータを使用する事には福袋という季節性の商品が多いこと、現在と比較してECサイト利用率が低かったことから、所謂、サクラレビューと呼ばれる作為的なレビューが少ないと予想されるメリットがあった。しかし、今回の分析はレビューの有意性を検証するものではなくレビューを受ける側がデータをどのように活用できるかの検討、レビューからショップの特徴を見出せるかの検証であったため、比較的新しいデータを使用すべきであったと感じる。そして、今回は「コーヒー」という商品群で選別・分析を行ったが、他の商品群でも同様のことが言えるのかを検証すべきであったとも感じる。幅広い年齢層の購入者がいると予想されるスポーツ用品や、サイズや品質への言及が多いであろうファッション関連まで、楽天市場には多様な商品が存在し、それに対するレビューの数も膨大である。消費する嗜好品という、予想の立てやすい商品群を選ぶことは分析を進める上で便利であったが、他の商品群では違った結果が見られたのではないだろうか。

また、今回の研究・分析を通じて将来的に実現されそうな機能構造が考えられ、それは、レビューから商品やショップの特徴を分析し、それらを活用して顧客へのレコメンデーションを行うといった仕組みである。勿論、それを利用するための作為的なレビューへの対処といった問題も発生すると考えられるため、レビューの有意性に関する研究も更に発展していくのが理想ではないだろうか。

レビュー・口コミとは、CtoCの情報共有という側面もあるが、CtoBのデータ提供という面もある。データマイニング技術が発展している今、アンケートでの「はい/いいえ」という問答以上に、自由回答・レビュー・口コミが様々な分野に活用されるデータとなっていく筈である。これらを適切に扱える人材が、これからの時代に必要になっていくと、改めて考えさせられる機会となった。

参考資料

「楽天」 ウィキペディア

<https://ja.wikipedia.org/wiki/%E6%A5%BD%E5%A4%A9%E6%A5%BD%E5%A4%A9%E5%B8%82%E5%A0%B4>

AI テキストマイニング by ユーザーローカル <https://textmining.userlocal.jp/>

「tf-idf」 ウィキペディア <https://ja.wikipedia.org/wiki/Tf-idf>

5 周年記念 箱が壊れるまでカフェ福袋 ムチャしました大赤字セールで送料無料で、このボリュームで驚きの 2999 円!!しかも、澤井珈琲自慢のコーヒーがさらにもう 1 袋プレゼント～!!

<https://web.archive.org/web/20100317195008/http://item.rakuten.co.jp/sawaicoffee-tea/445822/>