

ダイエット口コミデータセットにおけるダイエット食品および ダイエット飲料に関する語彙解析

大塚 敬義

目白大学短期大学部 ビジネス社会学科

t.ootsuka@mejiro.ac.jp

概要

ダイエットカフェ株式会社から国立情報学研究所 (NII) を通じて研究者に提供されているデータセットである「ダイエット口コミデータセット」の貸与を受けた。合計約 19 万 9 千件の口コミから成る当該データセットのうち、ダイエットドリンク約 4 万 3 千件、ダイエット食品約 8 千件について、出現する単語の共起ネットワーク図を作成したので報告する。

1 背景

ダイエットを目的とした医療行為（内服薬・注射等）、食品（青汁・ゼリー等）、着用品（加圧シャツ等）、器具（振動マシン等）、痩身エステ等が巷間にあふれている。そうした製品やサービス（以下「ダイエット商品」）の中には、はたして本当に減量効果を有するかが疑わしいものも少なくない。あるいは最悪の場合には健康被害をもたらすものが市場に流通し、人々により摂取されている可能性も否定できない。我が国において「食生活の欧米化」が昭和期から叫ばれるのと軌を一にして、肥満者 (BMI ≥ 25 kg/m²) の割合が増加し、令和 4 年国民健康・栄養調査結果（厚生労働省）では、20 歳以上の男性では 31.7%、女性では 21.0% となっている[1]。

ダイエット商品について、利用者がそれらに対する評価すなわち「口コミ」は健康雑誌等の紙媒体に掲載され、同時に減量試行者が集うインターネット上の掲示板やそれとは別の SNS に書き込まれてきた。ただし紙媒体で掲載された口コミは、集計・解析対象とする際に文字情報を人手で電子化する膨大な手間を要するがためにほぼ放置されてきた。他方でインターネット上に存在する口コミは紙媒体の口コミに比べて電子化しやすいが、無数ともいえる掲示板や SNS の書き込み群から不要情報（一例：アスキーアート等による無意味な荒し目的の書き込み等）を人手で除去して、本来のまともな口コミを収集する労苦がひとつの障害になっていた。

減量を主目的とするダイエット商品は、老若男女を問わず一定の関心を集めている。

ダイエットが主目的ではないが、血中の悪玉コレステロールレベルを下げる効果を謳ったある健康食品で、最悪の場合は死亡を伴う可能性のある健康被害が 2024 年 3 月に報道された。「健康食品」とは、健康に良いことを謳った食品の総称である。そのなかには、国が定めた安全性と効果に関する基準等に従って機能性が表示されている「保健機能食品」と、それ以外のいわゆる「健康食品」がある。さらに保健機能食品は、「特定保健用食品」、「栄養機能食品」及び「機能性表示食品」に分けられる。今回健康被害が報告された「紅麹サプリ（小林製薬）」は、「機能性表示食品」に該当する（通称「紅麹事件」）。国が定めた基準に従った「機能性表示食品」で健康被害が発生したことは、健康食品の信頼性が問われる重大な事件である。

よく個人のブログ等にダイエット商品の使用体験記が掲載されているが、それはあくまで一個人での感想や経験に過ぎない。一個人に焦点を当てるのではなく、ある特定のダイエット商品に限定して焦点を当て、複数の人物による商品評価の蓄積を検証することで初めて当該商品の有効性等の検証につながる。各個人のブログや各種 SNS の書き込みを横断して特定商品の商品名で検索をかけて口コミを収集するのでなく、ダイエット商品の口コミ書き込みに特化しかつ一定の信憑性があるウェブサイトを検索先を絞って口コミを収集したほうが良いであろうことは以前からある程度想定されていた。

2 目的

著者は前述の健康食品「紅麹サプリ事件」報道に接し、こういった服用者が高齢者に比較的偏りがちな商品でなく、生産年齢人口に当たる世代が主たる服用者であるダイエット商品の中にも、特に効果の高い商品や問題のある商品が潜在的に存在する可能性に着目した。

そこで解析対象データとして実用に耐えうるダイエット商品のロコミデータを探し求めていたところ、「ダイエットカフェ株式会社」が国立情報学研究所(NII)を通じて研究者に提供しているデータセットである「ダイエットロコミデータセット」[2] (以下「当該データセット」と呼ぶ: 図1に概念図を提示)の存在を認知することができた。ロコミとあることから、画像データに比べて取り扱いしやすい文字データであることが分かった。また社会的信用を有する研究機関であるNIIを経由して入手できる点から、当該データセットは一定の信頼性を有して解析対象として耐えうると判断できた。

当該データセットが含むロコミは原則としてすべて日本語の文字情報である。単純な数値(定量的データ)として扱えず、日本語の文字列部分である「ロコミ本文」を自然言語処理(テキストマイニング等)により一定程度解析できる。それゆえロコミに対してテキストマイニング等を施せば何らかの学術的知見や各ダイエット商品の特徴を獲得できるのではないかと推定した。

情報工学(言語工学, 自然言語処理)の知見を利用し、ダイエット商品の一部を構成する数々のダイエット食品の中から有用なもの, 効果に乏しいもの, 有害なものである傾向のある特定のダイエット食品を検出することにより, 人々を効果が疑わしい経済的支出や健康被害から救済し, 人々の健康増進に寄与したい。

実社会における利点として, いわゆるメタボリックシンドロームの減少につなげ, ひいては国全体の医療費の削減につながる。

著者が過去に従事していた医療情報学の研究[7][8]からつながる新たな地平を切り開く。

その一端として, 当該データセットの解析着手に当たり, 語彙どうしの関係性を共起ネットワーク図で把握することを試みるとしたい。

3 方法

当該データセットの本体はカンマ区切り形式(.csv)のデータ合計2件で構成される。

前者1件は, 図2のように話し言葉で記述されたロコミ約19万9千レコードを収録した「ロコミデータ」(review.csv)である。

後者1件は, 図3のようにダイエット商品の一覧を記した1万267レコードを収録した, 「ダイエット商品データ」(product.csv)である。本研究はカ

テゴリ6つおよびそれらに含まれない「ダイエットサプリ」のうち, 「ダイエットドリンク」「ダイエット食品」の合計2つはいずれも「直接体内に入る」ものであり, 最優先の解析対象データとする。

CiNiiにおいて「ダイエット」「ロコミ」の2語をAND検索した結果, 2024年9月9日現在で3件の先行研究論文がヒットした。うち文献2件[4][5]は発表年がともに2021年であり, かつ研究の著者が2名の同一人物で, このうち文献[4]は当該データセットに収録された「ダイエット器具」のデータのみを, 文献[5]は「ダイエット器具」に加え「ダイエットサプリ」のデータを取り扱ったのみにとどまっている。論文検索でヒットした残り1件の文献[6]は, 当該データセットの提供が開始された2019年よりも以前の研究であるため, 論文刊行以後に発生したロコミをカバーしていない。

本研究では2016年以降の比較的新しいロコミも解析対象とし, かつ文献[4][5]が取り扱わなかった「ダイエットドリンク」「ダイエット食品」のカテゴリ2つも解析対象として取り扱う点に新規性, 独創性がある。

当該データセット[2]に収録され, 固有の商品IDが振られた10266件の「ダイエット商品データ」は, 文字列に当たる部分をなす「カテゴリ名」「成分名」を含むほか, 具体名を記述した「商品名」に「商品ID」が付与されている。データ収集期間は2008年8月3日から2023年10月6日までである[3]。また「ロコミデータ」のうち, 特に文字列に当たる部分をなすフィールド「ロコミ本文」に重点的に注目する。「ロコミデータ」は「商品ID」のフィールドを含んでいるため, 「商品ID」をキーとすることにより, 「ダイエット商品データ」と照合して具体的な商品名とロコミとを紐付けることが可能である。なお同一の商品に対して付されるロコミは1件のみとは限らず, ロコミ複数件が付された商品もある。データ収集期間は前述の「ダイエット商品データ」と同一であり, ロコミ数は198,870件である[3]。

なお「商品データ」および「ロコミデータ」は著者へ提供時点ですでに匿名化済みであり, プライバシーや人権において倫理的な配慮がなされている。「ロコミデータ」の中から[親カテゴリ名]が[ダイエット食品]または[ダイエットドリンク]に該当するロコミレコードを抽出し, ツール「KH Coder」[9][10]を用いて形態素解析処理を行い, 統計ツールRにより描画を行い, 語彙どうしの共起ネットワー

ク図を獲得し、ロコミデータの全体概要を把握する。

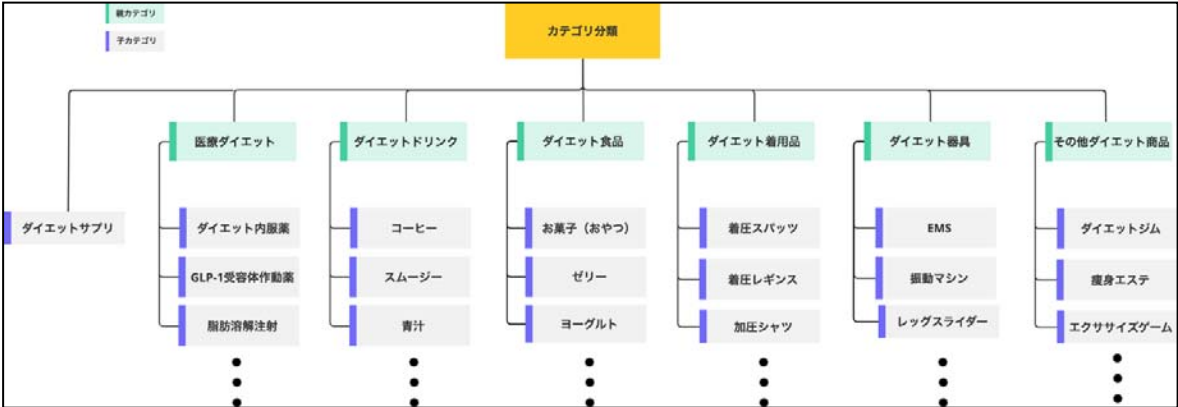


図 1 「ダイエットロコミデータセット」 [2]におけるカテゴリ分類の概念図（出典 [3]）

ロコミID	投稿者名	投稿日時	商品ID	ロコミ本文
1	椿	2008/8/11 11:34	150	<p>リエータのドリンクタイプはオススメです！</p> <p>数ある置き換えドリンクの中ではお手頃価格で、味の種類が豊富で、甘過ぎず飲みやすいです。水350mlに溶かすので、他の置き換えに比べると量が多いところもいいです。</p> <p>ただし、置き換えは運動も兼用しないとリバウンド率が高いです。元の食事に戻したら・・・ということもありますので、あくまでサポート感覚で使うのがいいかと思います。</p>

図 2 「ダイエットロコミデータセット」 [2]におけるロコミの一部（出典 当該データセットが含む「review.csv」）

商品ID	商品名	親カテゴリ名	子カテゴリ名	認証制度	成分名	商品登録日時	ロコミ数	評価数	評価平均点
412	清浄茶房	ダイエットドリンク	その他のお茶	特定保健用食品（トクホ）	グロビン由来バリン-バリン-チロシン-プロリン	2008/11/10 11:33	4	4	4.5
414	生葉純瘦茶	ダイエットドリンク	その他のお茶			2008/11/11 6:59	2	2	2.5
434	ジキニデルバック	ダイエットドリンク	その他ドリンク			2008/11/26 9:39	0	0	
439	五葉茶	ダイエットドリンク	その他のお茶			2008/11/25 9:49	43	37	3.76
440	スリムアップスリム	ダイエットドリンク	シェイク			2008/11/26 9:51	687	677	3.49
441	カロリー専科生搾（イキイキ）ぞうすい	ダイエット食品	主食系食品（ご飯、パン、麺類等）			2008/11/27 9:54	56	53	2.89
442	マンナンヒカリ	ダイエット食品	主食系食品（ご飯、パン、麺類等）			2008/11/27 9:56	73	71	3.59
443	まんぷくダイエット	ダイエット食品	その他食品			2008/11/20 9:58	8	8	3.38

図 3 「ダイエットロコミデータセット」 [2]におけるロコミ群の一部（出典 当該データセットが含む「ダイエット商品データ」（product.csv））

4 結果

まずロコミ全 198,870 件に対して第 1 キーに「商品 ID」を当て昇順にソートし、第 2 キーに「商品 ID」を当て昇順に、ソートした。その後に「親カテゴリ名」でロコミを分類した。[ダイエットドリンク]

についてのロコミは構成比が全体の 21.7%（43,084 件）を占めた一方で、[ダイエット食品]については 4.0%（8,042 件）であった。形態素解析後に「集計単位」を「文」でなく「段落」として共起ネットワークの図を描画した結果を、図 4 および図 5 に示す。

5 考察

〔ダイエット食品〕ロコミにおける解析結果（図4）では、「置き換える」（出現頻度 1,998 回数・第 14 位）が見られる。従来の食習慣を変えた、あるいは変えようとしている減量挑戦者たちの行動変容が示唆されている。「美味しい」（出現頻度 1,988 回・第 15 位）は、原文が「美味しくない」として出現している可能性を否定できない。よってダイエット食品一般は美味しいものであるという短絡的解釈を避けねばならない。Subgraph02 に属する「飽きる」（出現頻度 821 回・第 31 位）が登場しており、ダイエット食品ばかりの食生活は飽きやすいという現象を一定の確率で裏付けているといえる。また「バナナ」（出現頻度 460 回・第 75 位）は、朝、昼、夜の時間帯 3 つのうち「朝」と結びついていることから、減量試行時にバナナを摂食するならば朝の時間帯が良いのではないかという推測が可能である。さらに「体」と「脂肪」が共起しやすいとして検出されたことから、「体脂肪」という語句を単一の形態素として強制的に抽出すべきであったと考える。

〔ダイエットドリンク〕ロコミにおける解析結果（図5）では、「下痢」と「腹痛」のペアで、「野菜」と「不足」のペアで単語どうしの共起が見られることから、特定商品とは限定されないがダイエットドリンクの飲用時に何らかの問題が起きていることがわかる。また身体の部位で見ると上位 79 位以内に「顔」がランクインせず、「お腹」（出現頻度 6,886 回・第 20 位）がランクインしていることからやはり人々にとって痩せたい部位は腹部であることが示されている。また「解約」（出現頻度 4,081 回・第 39 位）が上位 79 位以内にランクインしており、一定程度の割合で利用者から効果なしと判断された商品群が存在すると解釈できる。

なお明確な体調不良を示す語句である「下痢」を著者が目にし、共起ネットワーク図には表出しないが危険な副作用を示す語句がロコミの原データ中に存在することを目視で知覚した。

〔商品 ID〕が 3494 番のダイエットドリンク（カテゴリ名は酵素ドリンク）は、「不正出血」の語句を含むロコミが 2 件検出された。うち 1 件は 2016 年 9 月に性別を女性と回答しており、2012 年 9 月の他方 1 件は性別が未回答だった。ただし「卵巣機能が低下」（原文ママ）の記述があることから、臓器的特徴からみて女性であると推定できた。投稿文章

中には、下記の深刻な内容が記載されており、商品自体に何か問題があるのではと懸念される事例があった（原文ママ）：

やり始めてからの生理がおかしくなったこと。通常、28 日周期で 5 日で終わるのですが、すぐの生理は少量で 10 日続き。生理と言うより不正出血？ その次は 22 日周期で始まって、それも量は多くはなかったのですが、なんと 3 週間も続いたので婦人科を受診した。

また〔商品 ID〕3532 番の投稿文章にも、何らかの問題を想起される記載があった（原文ママ）：

プラセンタ(胎盤)の影響か？、不正出血止まらず、お腹がぱんぱんになり、胸が張って辛いのでやめます。出血が、18 日止まらないので、明日婦人科にいきます。

このような、危険と思われるダイエット食品やダイエットドリンクが市場に一定数流通している可能性がある。体調不良を示唆する語句で機械的に検索をかけ、症状がダイエット飲食製品に起因するのか、あるいは属人的（製品のせいではなく利用者本人の体質に依るもの）か等を自然言語処理の知見で解明できればより良いと考える。

謝辞

本研究では、ダイエットカフェ株式会社（代表取締役 福田尚広様）から国立情報学研究所（NII）を通じて研究者に提供されている「ダイエットロコミデータセット」の貸与を受けました。データセット解析の機会提供にご尽力くださった関係諸氏に深謝いたします。また日頃の論文読解輪講を通じて著者に自然言語処理の勉学の機会をくださっている中央大学 理工学部 難波英嗣先生に深謝いたします。

参考文献

- [1] 厚生労働省：令和 4 年 国民健康・栄養調査結果の概要，
[<https://www.mhlw.go.jp/content/10900000/001296359.pdf> (cited 1-Nov-2024)] (2024).
- [2] ダイエットカフェ株式会社, 国立情報学研究所: 情報学研究データリポジトリ ダイエットロコミデータセット，
[<https://www.nii.ac.jp/dsc/idr/diet-review/> (cited 1-Aug-2024)] (2023).
- [3] ダイエットカフェ株式会社, 国立情報学研究所: 情報学研究データリポジトリ ダイエッ

- トロコミデータセット データスペック説明表, [data_spec_20231026.xlsx] (2023).
- [4] 木暮 美菜, 諸上 茂光: ロコミが行動変容促進型商品の評価に与える影響, マーケティングレビュー, 第(2)巻, 第1号, 日本マーケティング学会, pp. 22-29 (2021).
- [5] 木暮 美菜, 諸上 茂光: 努力が影響する製品に対するロコミ閲覧者の製品評価 : ダイエット器具とダイエットサプリのロコミを比較して, 日本経営システム学会全国大会講演論文集, 第(67)巻, pp. 196-199 (2021).
- [6] 大谷 鉄平: ダイエット系「ロコミ」にみられる語彙の特性 : 新聞折り込みチラシの場合との差異に注目して, 学習院大学大学院日本語日本文学, 第11号, 学習院大学大学院人文科学研究科日本語日本文学専攻, pp. 58-79 (2015).
- [7] 大塚 敬義: 教育用電子カルテのデータベース上に収録された模擬患者診療記録のレコード群に対する自然言語処理ツールを用いた単語の種類や出現頻度に関する考察, 安田女子大学紀要, 第42号, 安田女子大学, pp. 313-316 (2014).
- [8] 大塚 敬義: 医学辞書 2014 for ATOK における医療専門用語登録状況に関する調査, 新島学園短期大学紀要, 第36号, 新島学園短期大学, pp. 27-31 (2016).
- [9] KH Coder: 計量テキスト分析・テキストマイニングのためのソフトウェア,
[<http://khcoder.net/> (cited 1-Oct-2024)].
- [10] 樋口 耕一, 中村 康則, 周 景龍: 『動かして学ぶ! はじめてのテキストマイニング』, ナカニシヤ出版, (2023).

附録

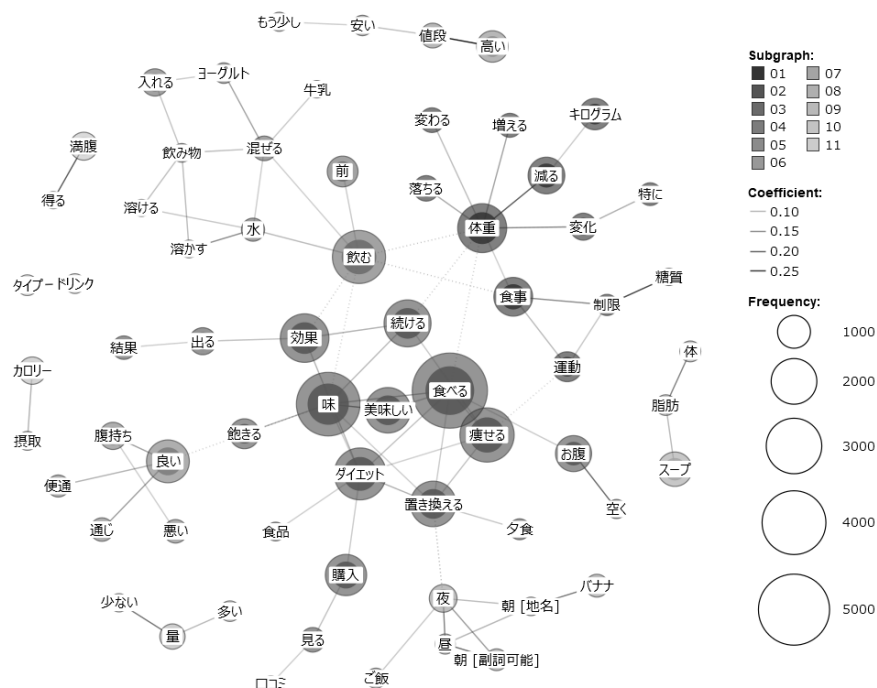


図4 [ダイエット食品] ロコミにおける共起ネットワーク図（上位80語・最小出現数200）

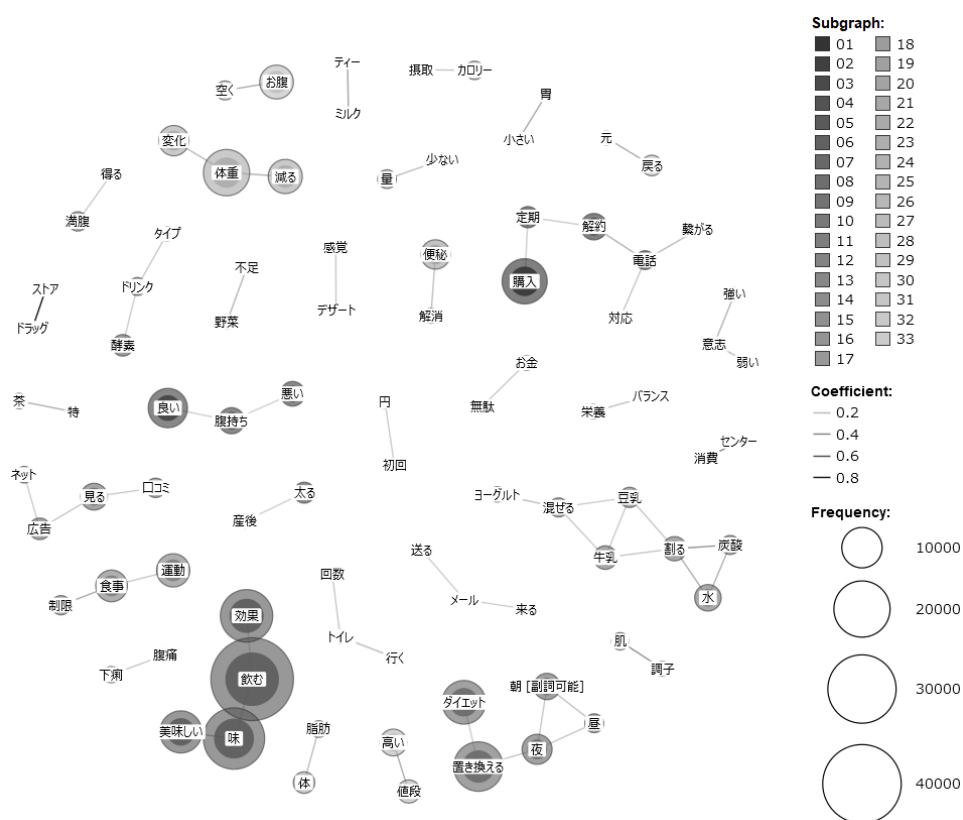


図5 [ダイエットドリンク] ロコミにおける共起ネットワーク図（上位60語・最小出現数300）