

お土産の類似商品抽出のための Twitter の利用可能性の検証

熊倉 梨華¹ 平林 (宮部) 真衣² 吉野 孝¹
¹和歌山大学システム工学部 ²東京大学大学院医学系研究科

1 はじめに

旅行に行った際に多くの方が、友人や職場へのお土産を購入する¹。食品関連のお土産を選ぶ際、自分の食べたことのないものはおいしさの想像がしづらく、選択しづらい可能性がある。そのため、レビューサイトや SNS の口コミを参考にする人は多い²。しかし、SNS やレビューサイトには多数の情報が投稿されている場合があり、自分の求める情報に容易にたどり着けるとは限らないため、自分の食べたことのないお土産のおいしさの容易な想像を支援する必要がある。

我々は、食べたことのないお土産に自分が既に食べたことのある食品と類似する点がある場合、類似する食品とその類似点を提示することで、未知のお土産のおいしさの想像を支援できるのではないかと考えた。そこで本研究では、類似商品情報の提供によるお土産のおいしさ想像支援の実現を目指し、類似商品情報の抽出可能性について検討する。

類似商品情報を抽出する媒体として、我々は Twitter³に着目した。Twitter は人々が率直な意見や感想を投稿しているメディアであり、「ある商品が別の商品と似ている」といった意見が投稿される可能性がある。そこで本稿では、類似を意図した表現（以降、「類似意図表現」と表記する）を含むツイートの分析により、類似商品抽出における Twitter の利用可能性を検証する。

2 関連研究

お土産の類似性に関する研究として、宇衛らはお土産と類似商品の実食実験を行い、お土産の類似商品を抽出するための要素について分析した [1]。原材料、シズルワード、ツイートにおいて、それぞれの類似度の判定分析を行い、お土産の類似商品を抽出するための特徴を明らかにしている。本研究は Twitter に焦点を当て、Twitter を用いた類似商品の抽出可能性を検証する。

類似する食品に関する研究として、加藤らは Twitter、Web 検索、CookPad⁴ 上のシズルワード [3] と食品との関係の分析を行った [2]。メディアごとに食品とシズルワードの関係を比較分析し、それぞれの特徴を明らかにした。本研究は Twitter において、あるお土産と類似商品名が共起した際の特徴を分析する。

Testing the feasibility of using Twitter to extract similar products for souvenirs

Kumakura Rika¹ Mai Miyabe Hirabayashi² Takashi Yoshino¹

¹ Faculty of Systems Engineering, Wakayama University

² Graduate School of Medicine, The University of Tokyo

¹ お土産に関するアンケート調査:

https://www2.fgn.jp/mpac/_data/8/?d=201309_06

² 口コミサイト・インフルエンサーマーケティングに関する

アンケート結果: https://www.caa.go.jp/policies/policy/consumer_policy/policy_coordination/internet_committee/pdf/internet_committee_180927_0003.pdf

³ <https://twitter.com>

⁴ <http://cookpad.com/>

表 1: お土産名を含むツイートの取得結果

お土産名	ツイート数
わかさいも	81,773 件
カスドース	43,964 件
甘金丹	11,175 件
ナッツヴェセル	2,619 件
田むらの梅	2330 件
菜種の里	974 件

表 2: 選定した類似意図表現

類似意図表現
似て、似た、っぽい、同じ、みたい、類似、酷似、ような、ほぼ、一緒、パクリ、ジェネリック、そっくり、同等

3 Twitter 上の類似意図表現の分析

本分析では、Twitter で用いられている類似意図表現について分析する。

3.1 分析対象データ

今回は、宇衛ら [1] の実験で用いられた 6 つのお土産（わかさいも・カスドース・甘金丹・ナッツヴェセル・菜種の里・田むらの梅）を調査対象とした。ツイートの取得期間は 2006 年 3 月 21 日～2022 年 11 月 20 日で、分析対象のお土産名を含むツイートを取得した。ツイートの取得結果を表 1 に示す。

3.2 分析の手順

本研究では、以下の手順により Twitter 上で用いられる類似意図表現を調査する。

- 今回は、類語辞典⁵における「類似」の類語として挙げられている表現 5 語、および連想類語辞典⁶に挙げられている表現のうち、類似した商品の言及に用いられる可能性があるとして著者らが判断した表現 9 語、合計 14 語を調査対象の類似意図表現として用いることとした。選定した類似意図表現を表 2 に示す。
- 分析対象ツイートから、選定した類似意図表現を含むツイートを抽出する。
- 手順 2 で抽出したツイートから、類似商品名が含まれるツイートを抽出する。

3.3 分析結果

類似意図表現を含むツイートの割合を表 3 に示す。ツイートでは口語的な言葉遣いになることが多いため、「類似」や「酷似」などよりも、日常的に使われやすいと考えられる「っぽい」や「みたい」などが含まれる割合が高かった。

わかさいもに関するツイート中に出現した類似意図表現と抽出された類似商品名を表 4 に示す。「みたい」を含むツイートは 1,999 件と多いものの、「行ってみたい」「食べてみたい」という使い方が大部分を占め、類似商品名を含むツイートの割合は 2.1 % と他の類似意図表現に比べて低い。すなわち、類似以外を意図する表現としても用いられる類似意図表現では、十分な類似商品名を抽出できない可能性がある。一方で、「類似」や「酷似」は類似の言及のみに使用される単語である

⁵ <https://thesaurus.weblio.jp/>

⁶ <https://renso-ruigo.com/>

表 3: 類似意図表現を含むツイートの割合

	似て	似た	っぽい	同じ	みたい	類似	酷似	ような	ほぼ	一緒	パクリ	ジェネリック	そっくり	同等	総数
わかさいも	0.41% (334)	0.13% (106)	1.24% (1016)	0.49% (403)	2.45% (1999)	0.10% (78)	0.01% (4)	0.69% (561)	0.16% (128)	0.46% (377)	0.04% (33)	0.05% (39)	0.05% (40)	0% (1)	6.26% (5119)
カスタード	0.13% (57)	0.06% (28)	0.21% (93)	0.57% (250)	25.06% (11017)	0.01% (2)	0% (1)	2.95% (1296)	0.07% (29)	0.28% (124)	0.01% (4)	0% (1)	0.02% (7)	0.01% (3)	29.37% (12912)
甘金丹	1.25% (140)	0.09% (104)	0.44% (49)	0.80% (89)	2.84% (317)	0.22% (24)	0.02% (2)	0.99% (111)	0.22% (24)	0.57% (64)	0.46% (51)	1.40% (156)	0.69% (77)	0.01% (1)	10.82% (1209)
ナッツヴェセル	0.12% (3)	0.04% (1)	0.12% (3)	0.15% (4)	0.34% (9)	0% (0)	0.00% (0)	0.04% (1)	0.08% (2)	0.31% (8)	0% (0)	0% (0)	0% (0)	0% (0)	1.18% (31)
田むらの梅	0.26% (6)	0.17% (4)	0.04% (1)	0.56% (13)	0.86% (20)	0% (2)	0% (0)	0.69% (16)	0.09% (2)	0.30% (7)	0% (0)	0% (0)	0.04% (1)	0% (0)	3.09% (72)
菜種の里	0.62% (6)	0% (0)	0.10% (1)	0.21% (2)	1.95% (19)	0% (0)	0% (0)	0.62% (6)	0% (0)	0.92% (9)	0% (0)	0% (0)	0% (0)	0% (0)	4.41% (43)
平均値	0.38% (91.0)	0.17% (40.5)	0.81% (193.8)	0.53% (126.8)	9.37% (2230.2)	0.07% (17.7)	0% (1.2)	1.39% (331.8)	0.13% (30.8)	0.41% (98.2)	0.06% (14.7)	0.13% (32.7)	0.09% (20.8)	0% (0.0)	2.26% (3231)

表 4: 「わかさいも」に関するツイート中で出現した類似意図表現と抽出された類似商品名

類似意図表現	類似商品名を含むツイートの割合	類似商品名(ツイート数)
みたい (1,999)	2.1% (42)	いもすけどん (5), ままどおる (5), ひよこ (5), ハーバー (4), 黄金芋 (4), モンモオ (4), 黄金芋 (4), なんと金時 (3), なごやん (3), ざびえる (2), 長寿いも (1), ぼくぼく (1), かんざし (1)
っぽい (1,016)	0.6% (6)	通りもん (1), モンモオ (1), ひよこ (1), いも風味 (1), 甘藷日記 (1), なんと金時 (1)
同じ (403)	4.0% (16)	いも風味 (5), いもすけどん (4), 手焼いも (2), 黄金芋 (1), なんと金時 (1), 能登のいも (1), ひよこ (1)
似て (334)	10.5% (35)	いもすけどん (5), ハーバー (5), ひよこ (3), いも風味 (3), 通りもん (3), ままどおる (3), モンモオ (3), 黄金芋 (3), なごやん (2), 能登のいも (2), 手焼いも (1), 長寿いも (1), 奥能登 (1)
類似 (78)	12.9% (10)	いも風味 (8), いもすけどん (2)
酷似 (4)	25.0% (1)	いもすけどん (1)

表 5: 抽出された類似要素の分類

類似要素(大分類)	類似要素(細分類)	該当表現(ツイート数)
抽象的類似部分	系統	ルーツ (4), カテゴリ (2), ジャンル (2)
	視覚的要素	見た目 (43), 外側 (2), 中身 (3), 形状 (2), 色 (1), 断面 (1), 包装 (1)
	主観的要素	面白い (3), 好き嫌い分かれる (1)
	嗅覚的要素	食感 (3)
具体的類似部分	味覚的要素	味 (44), テイスト (1), 風味 (1)
	シズルワード (味覚系)	甘い (41), 激甘 (5)
	シズルワード (食感系)	シャリシャリ (6), フワフワ (4), フワッフワ (1), パサパサ (3), パッサパサ (3), もさもさ (3), しっとり (2), ホクホク (1)
	シズルワード (情報系)	懐かしい (3)
	素材名	砂糖 (7), カスタード (3), 白餡 (3), 皮 (4), しょうゆ (1), 芋 (3)
	由来の地名	九州 (7), 南蛮 (6), 長崎 (2)
	その他	口の中の水分奪われる (2)

ため、ツイート数自体はそれぞれ 78 件、4 件と少ないものの、類似商品名はそれぞれ 12.9%, 25% のツイートから抽出できた。

4 言及される類似要素の分析

類似商品に言及するユーザは、「“味”が似てる」や「“見た目”は同じ」など、類似要素について言及する可能性がある。そこで、類似商品名が含まれるツイートに類似要素が含まれるかどうかを調査する。

4.1 分析の手順

以下の手順により、言及される類似要素を抽出する。

- 3章の分析において類似商品名が含まれたツイートを分析対象とする。
- 分析対象ツイートに含まれる類似要素を抽出する。

4.2 抽出結果

表 5 に、今回抽出された類似要素を示す。また、わかさいもの類似商品名を含むツイートから抽出した類似要素を表 6 に示す。類似要素としては、大別して二つの情報があることが分かった。1つ目は“抽象的類似部分”で、例えば「見た目」「味」「形状」「ルーツ」など、類似商品のどの部分が似ているのかを抽象的に言及するものである。2つ目は“具体的類似部分”であり、シズルワードなどが含まれる。わかさいもの場合、「パサパサ」「もさもさ」「懐かしい」などが該当する。

5 おわりに

本研究では、類似商品の抽出のための、お土産名を含むツイートにおける類似意図表現および言及される類似要素に関する分析を行った。分析の結果、以下の点を明らかにした。

- 今回対象とした 14 語の類似意図表現は、表現によって取得できるツイートの割合に差があるものの、複数の類似意図表現を用いることでツイート

表 6: 「わかさいも」の類似商品名を含むツイートから抽出された類似要素

お土産名	類似商品名	類似要素の該当表現
わかさいも	いもすけどん	形状, 味, 芋, 見た目
	ひよこ	味, 皮, 白あん, 中身, 懐かしい, もさもさ, 口の中の水分奪われる, パッサパサ, パサパサ
	ハーバー	見た目, 味, 形状, 芋, パッサパサ
	いも風味	ルーツ, 味, 包装, 見た目, しょうゆ風味
	通りもん	餡, 中身, パサパサ
	なごやん	味, 中身
	黄金芋	芋, ホクホク
	ままどおる	白餡, パッサパサ, 断面, 食感, 味, パサパサ, 懐かしい
	モンモオ	見た目, 食感

が取得できる可能性がある。ただし、類似意図表現を含むツイートにおいて必ずしも類似商品名が言及されているとは限らず、また類似商品名と共にしやすい表現が存在する可能性がある。

- 言及される類似要素は、“抽象的類似部分”と“具体的類似部分”の2種類に大別できる。

今後は、Twitter から十分なデータが抽出できなかったお土産について、他の媒体を対象とした分析を行う。

謝辞

本研究の一部は、JSPS 科研費 18K11555 および 21K12138 の助成による。

参考文献

- 宇衛 穂乃実, 平林 (宮部) 真衣, 吉野 孝: お土産のおいしさ想起支援のためのお土産と類似商品の類似性分析, 情報処理学会関西支部支部大会, G-24, pp. 1-7(2021).
- 加藤 大介, 宮部 真衣, 荒牧 英治, 灘本 明代: インターネット上のメディア毎の「おいしさ」表現比較分析, データ工学と情報マネジメントに関するフォーラム, F5-5, pp. 1-8(2015).
- B・M・FT ことばラボ: Sizzle Word シズルワードの現在 2018 改訂「おいしいを感じる言葉」調査報告, 株式会社 B・M・FT 出版部 (2018).