

楽天市場及びダイエットカフェのレビューデータに対するテキストマイニング - 数量化Ⅰ類の適用可能性について -

王 帥, 三道 弘明 (関西学院大学)

背景・目的

- 口コミの投稿はサイトによって評価の違いがあるかを検証する
- 市場に流行しているダイエットサプリメントはどんな評価されているかを明らかにする
- 数量化Ⅰ類の適用可能性について検討する

データについて



みんなのレビュー：2018年～2019年
 ダイエットカフェ：2008年～2020年

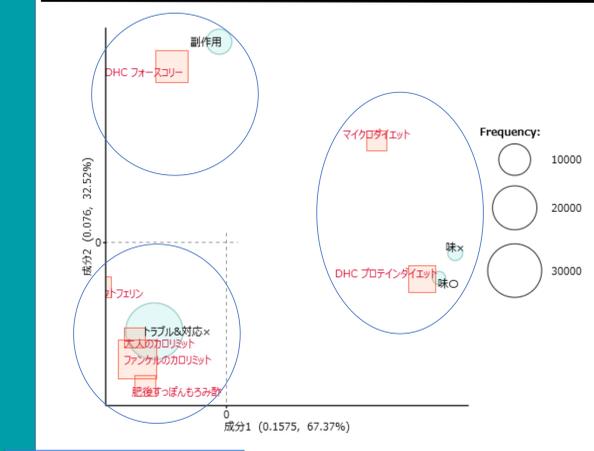
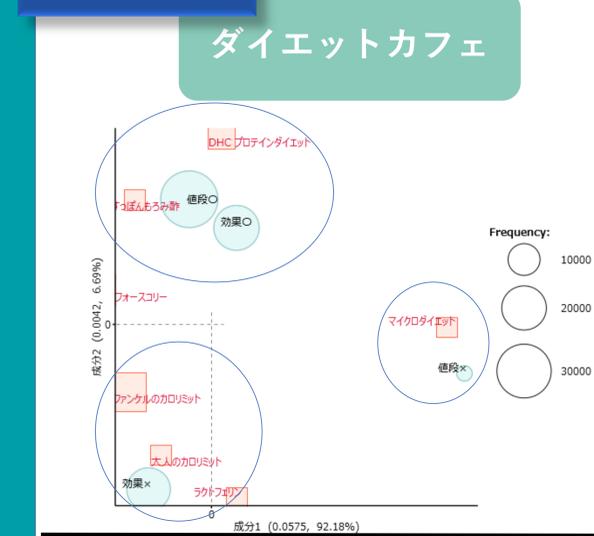
商品名	ダイエットカフェ	みんなのレビュー	合計
ファンケルのカロリミット	9,386	1,917	11,303
DHCフォースコリー	5,996	360	6,356
DHCプロテインダイエット	3,739	180	3,919
ラクトフェリン	2,454	2,266	4,720
大人のカロリミット	2,387	1,141	3,528
マイクロダイエット	2,260	206	2,466
肥後すっぽんもろみ酢	2,212	140	2,352
合計	28,434	6,210	34,644

コーディングルール

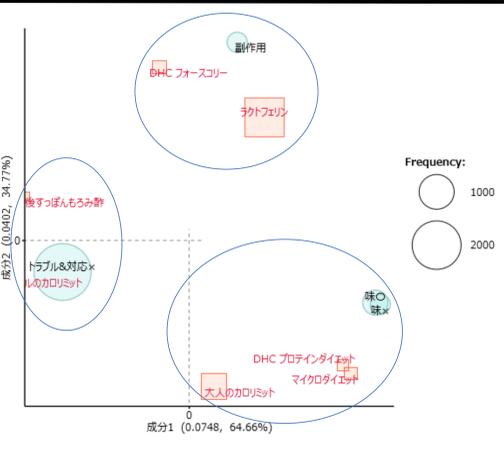
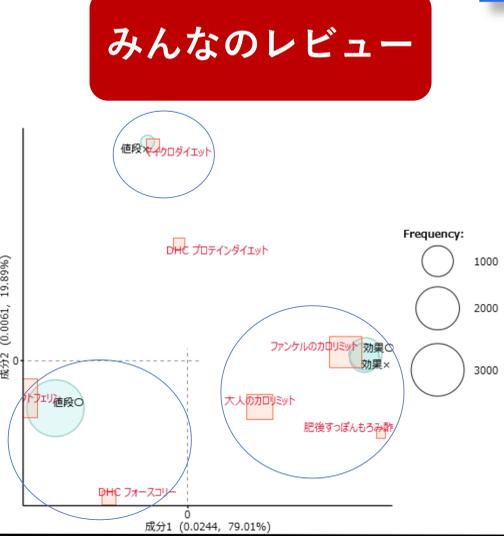
- コーディング：共通する概念を一つのキーワードに集約
- コーディング・ルール：論理演算子を用いて記述する

値段	効果	味	副作用	トラブル
<ul style="list-style-type: none"> • 値段が安い 高い • コスパが良い 悪い 	<ul style="list-style-type: none"> • 効果がある ない • 痩せた 痩せない 	<ul style="list-style-type: none"> • 味がいい 悪い • 飲みやすい にくい 	<ul style="list-style-type: none"> • 便秘 下痢 • 頭痛い 体だるい 	<ul style="list-style-type: none"> • 勧誘電話 解約 • 返事来ない 遅い

商品ごとにサイト間の比較



みんなのレビュー



クロス集計表

みんなのレビュー

	*値段○	*値段×	*味○	*味×	*効果○	*効果×	*トラブル&対応	*副作用	ケース数
DHCフォースコリー	196 (26.34%)	5 (0.67%)	9 (1.21%)	10 (1.34%)	45 (6.05%)	44 (5.91%)	107 (14.38%)	27 (3.63%)	744
DHCプロテインダイエット	126 (25.40%)	8 (1.61%)	41 (8.27%)	25 (5.04%)	37 (7.46%)	35 (7.06%)	86 (17.34%)	4 (0.81%)	496
ファンケルのカロリミット	1100 (25.30%)	53 (1.22%)	37 (0.85%)	478 (11.00%)	376 (8.65%)	841 (19.35%)	30 (0.69%)	4347	
マイクロダイエット	114 (18.18%)	19 (3.03%)	51 (8.13%)	38 (6.06%)	32 (5.10%)	26 (4.15%)	96 (15.31%)	2 (0.32%)	627
ラクトフェリン	1261 (19.79%)	55 (0.86%)	249 (3.91%)	151 (2.37%)	255 (4.00%)	200 (3.14%)	1093 (17.15%)	232 (3.64%)	6372
大人のカロリミット	668 (24.47%)	29 (1.06%)	111 (4.07%)	68 (2.49%)	235 (8.61%)	176 (6.45%)	498 (18.24%)	17 (0.62%)	2730
肥後すっぽんもろみ酢	60 (19.67%)	1 (0.33%)	4 (1.31%)	1 (0.33%)	35 (11.48%)	30 (9.84%)	51 (16.72%)	5 (1.64%)	305
合計	3525 (22.57%)	170 (1.09%)	502 (3.21%)	313 (2.00%)	1117 (7.15%)	887 (5.68%)	2772 (17.75%)	317 (2.03%)	15621
カイ2乗値	69.147**	29.782**	196.914**	142.256**	214.690**	165.827**	18.216**	172.390**	

ダイエットカフェ

	*値段○	*値段×	*味○	*味×	*効果○	*効果×	*トラブル&対応	*副作用	ケース数
DHCフォースコリー	6438 (25.76%)	136 (0.54%)	72 (0.29%)	67 (0.27%)	4013 (16.06%)	3805 (15.22%)	6319 (25.28%)	2892 (11.57%)	24992
DHCプロテインダイエット	4081 (22.78%)	229 (1.28%)	766 (4.28%)	1216 (6.79%)	2427 (13.55%)	1575 (8.79%)	3958 (22.09%)	626 (3.49%)	17914
ファンケルのカロリミット	11799 (32.30%)	412 (1.13%)	252 (0.69%)	118 (0.32%)	7593 (20.79%)	7433 (20.35%)	11691 (32.01%)	984 (2.69%)	36527
マイクロダイエット	2181 (22.29%)	896 (9.16%)	282 (2.88%)	478 (4.88%)	1655 (16.91%)	763 (7.80%)	2034 (20.78%)	807 (8.25%)	9786
ラクトフェリン	2746 (26.34%)	273 (2.62%)	23 (0.22%)	16 (0.15%)	1597 (15.32%)	1916 (18.38%)	3022 (28.98%)	429 (4.11%)	10427
大人のカロリミット	3183 (32.34%)	142 (1.44%)	75 (0.76%)	43 (0.44%)	1918 (19.49%)	2102 (21.36%)	3131 (31.81%)	344 (3.49%)	9843
肥後すっぽんもろみ酢	3042 (29.35%)	62 (0.60%)	45 (0.43%)	48 (0.46%)	1556 (15.01%)	1544 (14.90%)	3706 (35.76%)	165 (1.59%)	10363
合計	33470 (27.93%)	2150 (1.79%)	1515 (1.26%)	1986 (1.66%)	20759 (17.32%)	19138 (15.97%)	33861 (28.25%)	6247 (5.21%)	119852
カイ2乗値	914.776**	3482.463**	1962.489**	4539.597**	613.539**	1974.015**	1319.402**	3163.107**	

数量化Ⅰ類

ダイエットカフェ：7商品の全口コミデータ (28,434レコード) 2008～2020
 みんなのレビュー：2018年～2019年の7商品の口コミデータ (6,210レコード)
 各データをランダムに100件を選び、分析した
 なお、ヒット率の高い
値段○、効果○、効果×、トラブルの4コードに基づき

- 各コードに該当すれば1
 該当しなければ0

例えば：
 1瓶飲み切りましたがなんの変化も感じませんでしたのでやめました。それでも携帯に電話がかかります。めんどろなので1度も出てないです。着拒否したら違う番号でもかかってきました。もう何か月も前ですが未だに月1くらいでかかってくる。すごく迷惑です。
 評価ポイント：★
 値段○：0
 効果○：0
 効果×：1
 トラブル：1

数量化Ⅰ類の結果を使って評価ポイントを予測
 $Y(\text{評価ポイント}) = 2.12331$

サイト	口コミ本文	評価	値段○	効果○	効果×	トラブル&対応	副作用	予測値
ダイエツト毎日飲み続けて、早		1	1	1	0	1	1	2.650554
ダイエツト1500円のキャンペーン		2	1	0	0	1	2	2.12331
ダイエツト500円で試せるという		2	1	1	0	1	2	2.421149
ダイエツト食べ過ぎた時や、飲		4	1	1	1	1	4	2.650554
ダイエツト2ヶ月飲み続けました		2	1	0	0	1	2	2.12331
ダイエツト試せと言われ、ま		3	0	1	1	0	3	3.099799
ダイエツト全然効果なし 運動も		1	1	0	1	0	1	2.352715
ダイエツト増えもせず、減りも		3	0	1	0	0	3	3.868534
ダイエツト定期購入を利用して		4	0	0	0	0	4	2.570695
ダイエツトオルファリスト、マ		3	0	0	1	0	3	2.9001
ダイエツト夕食のみ置き換えで		2	0	0	0	0	2	2.570695
ダイエツト味とかは問題ない		3	1	0	0	1	3	2.12331
ダイエツト1瓶飲み切りました		1	1	0	0	1	1	2.12331

みんなのレビュー

100件のデータを用いて、数量化Ⅰ類に適用した結果

分析結果	アイテム	カテゴリ	例数	カテゴリ	範囲	偏相関係数	
*値段○	0	62	-0.03263	0.085878	0.036785		
	1	38	0.053244				
*効果○	0	85	0.031705	0.211367	0.090197		
	1	15	-0.17966				
*効果×	0	82	0.042063	0.233682	0.102817		
	1	18	-0.19162				
*トラブル	0	68	0.095914	0.299732	0.120333		
	1	32	-0.20382				
定数項				4.52			
重相関係数R			0.232354				
決定係数R2			0.053989				
自由度修正済み決定係数			0.014157				
Y評価の標準誤差			0.828676				
要因効果の検定							
アイテム	残差平方和	自由度	平均平方	F値	P値	F(0.05)	F(0.01)
*値段○	65.32534	1	65.32534	95.12872	5.66E-16	3.941222	6.90941
*効果○	65.77203	1	65.77203	95.77921	4.72E-16	3.941222	6.90941
*効果×	65.93396	1	65.93396	96.015	4.45E-16	3.941222	6.90941
*トラブル	66.19546	1	66.19546	96.39681	4.04E-16	3.941222	6.90941
全体	65.23695	95	0.686705				

ダイエットカフェ

分析結果	アイテム	カテゴリ	例数	カテゴリ	範囲	偏相関係数	
*値段○	0	31	-0.26698	0.386932	0.049192		
	1	69	0.119949				
*効果○	0	53	-0.13998	0.297839	0.12951		
	1	47	0.157855				
*効果×	0	50	-0.1147	0.229405	0.100246		
	1	50	0.114702				
*トラブル	0	29	0.592365	0.834317	0.103633		
	1	71	-0.24195				
定数項				2.5			
重相関係数R			0.265195				
決定係数R2			0.070329				
自由度修正済み決定係数			0.031185				
Y評価の標準誤差			1.123564				
要因効果の検定							
アイテム	残差平方和	自由度	平均平方	F値	P値	F(0.05)	F(0.01)
*値段○	120.2185	1	120.2185	95.23044	5.42E-16	3.941222	6.90941
*効果○	121.9735	1	121.9735	96.02061	3.82E-16	3.941222	6.90941
*効果×	121.145	1	121.145	95.96438	4.5E-16	3.941222	6.90941
*トラブル	121.2296	1	121.2296	96.03136	4.43E-16	3.941222	6.90941
全体	119.9276	95	1.262396				

数量化Ⅰ類の結果を使ってすべてのデータの評価ポイントを予測し、実際の評価ポイントとの相関を計算した結果

評価	予測値	相関
評価	1	1
予測値	0.107044	1

評価	予測値	相関
評価	1	1
予測値	0.107917	1

考察

- コーディング・ルールの適用によってテキストマイニングの結果はより簡潔、明確になった
- 商品の間、価格や味、効果などそれぞれブロックごとに明確に分かれた
- 口コミはサイトによって(楽天、ダイエットカフェ)商品の評価が異なる部分がある
- 数量化Ⅰ類に関しては
- 分析結果は芳しくないが、コーディングルールの見直しにより精度が向上する可能性
- 価格を含むようなデータが入手できれば、更なる知見を得られる可能性
- 数量化Ⅱ類(判別分析)の適用