

平成30年6月28日現在

機関番号：82111

研究種目：基盤研究(C) (一般)

研究期間：2014～2017

課題番号：26450370

研究課題名(和文) インターネット通販の「お客様の声」から探る青果物の消費者ニーズ

研究課題名(英文) Mining the customer preferences from online agricultural product reviews

研究代表者

竹崎 あかね (Takezaki, Akane)

国立研究開発法人農業・食品産業技術総合研究機構・農業技術革新工学研究センター・上級研究員

研究者番号：40550520

交付決定額(研究期間全体)：(直接経費) 1,900,000円

研究成果の概要(和文)：野菜商品レビューから消費者の野菜に関する価値観等の情報を効率的に収集するために、記述内容(概念)を表すキーワードを対象テキスト特有の表現を考慮して抽出する工程、およびその工程で利用するユーザー辞書を提案した。提案工程、辞書を利用して抽出した野菜品目特有のキーワード群から、消費者が重視する品質保証要素(ブランド、健康機能性など)に品目間差があることが示唆された。

研究成果の概要(英文)：In this study, we present the procedure of concept extraction based on natural language processing (NLP) from online vegetable-product reviews. The unique characteristics of vegetable product reviews made difficult to process accurately using general NLP. To improving performance of NLP on vegetable-product reviews, we proposed the customer dictionary and the concept extraction methods such as 1) morpheme analysis by reference to an additional custom dictionary, 2) acquisition of negative meaning, 3) synonymous substitution, 4) identification of target nouns that adjectives have a relationship with according to the analysis purpose.

研究分野：農業語彙体系の構築

キーワード：自然言語処理 野菜商品レビュー 形態素解析 インターネット通販 品種

1. 研究開始当初の背景

(1) 農作物の高付加価値化による農業所得向上が喫緊の課題である中、消費者ニーズを反映した農作物生産が今まで以上に重要となっている。一方、近年急速に蓄積するインターネット上の掲示板やブログのクチコミ等のテキストを解析し、商品やサービスの改善案、あるいは改善活動に進む気づきを得ることが試みられている。農業分野では現時点でこのようなテキスト活用事例は少ないものの、今後消費者の価値観を探る手段の一つになる可能性がある。

(2) 定性データであるテキストから効率的に記述内容(概念)を抽出するために、形態素解析・構文解析などの自然言語処理が行われる。自然言語処理は新聞記事等一般分野のテキストでは実用化レベルの高い精度を達成しており公開ツールで簡単に試行できる。しかし、異なる分野のテキストでは分野特有の言語表現により公開ツールの解析精度が劣るとされている。

(3) 農業分野のテキストについては付属辞書を参照した形態素解析(以下簡易自然言語処理)を適用した報告はあるものの分野特有の表現が解析精度に及ばず影響について定量的な評価はなされていない。また、農業分野のテキストに適した概念抽出工程も示されていない。

2. 研究の目的

(1) 今まで農作物の消費者ニーズ調査で十分活用されてこなかった野菜商品レビューを対象とし、簡易自然言語処理からレビュー内容を把握する際の問題点を明らかにし高精度の概念抽出工程を提案する。

(2) 野菜商品レビューの解析精度向上に資する辞書拡張を効率化するため、野菜商品レビューに出現する語の特徴を明らかにし、既存ソフトウェアによる網羅的収集や研究者語彙の利用可能性を検討する。

3. 研究の方法

(1) 簡易自然言語処理により野菜商品レビューを品詞タグ付けした語に分割した。簡易自然言語処理で抽出した単語もしくはその連続した複合語からなるキーワードの集合体から、テキストに含まれる概念を把握する際の問題点を明らかにした。なお全ての研究に共通して、データセットには楽天市場の3年間(2010年1月1日~2012年12月31日)の「野菜・キノコ」ジャンルの商品レビューを利用した。

(2) 野菜商品レビュー特有の語(以下ドメイン概念語)の特徴を明らかにし、ドメイン概念語の効率的収集を検討した。

形態素解析の付属辞書に登録された語を

一般分野の語とし、未登録語の中に野菜商品レビュー特有の語が含まれると仮定した。用語抽出モジュールTermExtractの重要度スコアに基づきドメイン概念語を選定、分類することでその特徴を明らかにした。

また、ドメイン概念語の効率的な収集のため、TermExtractによる選定の網羅性を確認するとともに体系的に整理された研究者語彙が利用可能かを検討した。研究者語彙を農業・食品産業技術総合研究機構;農研機構ホームページ

(<<http://www.naro.affrc.go.jp/> >

2015年6月23日~8月7日調査)に出現した語彙とし、ドメイン概念語;消費者語彙が含まれるか否かを確認した。

(3) 提案した概念抽出工程(図1)、ユーザー辞書を用いて野菜商品レビューを解析し、野菜品目ごとの商品レビューの特徴を明らかにした。野菜品目毎にキーワード(名詞、動詞、形容詞と品詞判定された語)の出現レビュー数を集計し、対象品目とそれ以外の品目、対象キーワードとそれ以外のキーワードの出現レビュー数についてクロス集計表を作成した。クロス集計表のフィッシャー正確確率検定を行い確率値が0.01より低い上位100キーワードを対象品目に特徴的なキーワードとし、ブランド(品種名;例“ペニハルカ”、品種名総称;例“安納芋”、地域等で設定されたブランド名;例“鳴門金時”を含む)、安全性、健康機能性に分類した(表2)。またレビューの解析結果を実際の市場実態と比較するため、生産者と卸売業者間で利用されるベジフルコードの商品名を調査した。

4. 研究成果

(1) 簡易自然言語処理では、野菜商品レビューで高出現した“安納芋”などの重要語が抽出できないこと、“青梗菜”、“チンゲン菜”、“チンゲンサイ”のような同義語のために“チンゲンサイ”の出現数が少ないと判定されること、高出現した重要語の否定概念が欠落し反対語が抽出されることを確認した(例;“無農薬”の76.5%が“農薬”で抽出)。また、形容詞“安い”のような語は単独でも概念が把握できるのに対して“高い”のように形容詞の対象となる語(例;“値段”、“糖度”)と一緒に抽出しないと正確な概念が把握できないと判断した。<研究の方法(1)>

(2)(1)の結果に基づき、簡易自然言語処理済みテキストから抽出すべき概念を提案し、野菜商品レビューの概念抽出工程として提示した。1)解析対象に合致した参照辞書を構築して形態素解析を行う(図1工程2)。2)構文解析後、動詞“する”は、その直前に出現する名詞と一語に集約し、具体的動作を示す動詞に変換する(工程4)。3)否定概念を付与するため、語の変換、集約処

理をする（工程 5）。4）同義語を正規化する（工程 6）。5）解析対象に合わせて係り受け関係の語を抽出する（工程 7）。＜研究の方法（1）＞

また、上記工程で利用する、形態素解析用、同義語正規化用のユーザー辞書を構築した。

（3）野菜商品レビュー特有の語；ドメイン概念語の 60%は品種やブランドの名称等を含む商品の名称であった（表 1）。品種やブランドの名称には トウモロコシ品種の“味来（みらい）”、トマトブランドの“優糖星（ゆうとうせい）”のように、ひらがな表記で一般分野の語と同表記異義語になるもの、ジャガイモ品種の“とうや”、トウモロコシ品種の“ゴールドラッシュ”のように一般分野の語と同表記異義語になるものがあった。以上から、野菜商品レビューの解析精度向上には高出現する商品の名称をユーザー辞書へ登録すること（図 1 工程 2）、品種・ブランドの名称で確認された同表記異義語を区別して辞書登録する必要があることを明らかにした。＜研究の方法（2）＞

（4）TermExtract を利用することでドメイン概念語の 56%が選定できた。“農林 1 号”のような数字を含む複合名詞、“インカのめざめ；インカ（名詞）・の（助詞）・めざめ（動詞）”のような名詞以外の品詞を含む複合語、接頭辞が付く語（例；“無農薬”・“低農薬”）は TermExtract では選定できなかった。

研究者語彙には、ドメイン概念語；消費者語彙の 66%の概念が含まれた（表 2）。出現率は健康機能性、安全性カテゴリで高く、商品の名称、特にブランドの名称で有意に低かった。以上から TermExtract によるドメイン概念語の選定は効率的であるものの、複合語の構成ルールが特異的な品種の名称などは選定できないため網羅的な選定はできないと判断した。また、研究者語彙からは消費者語彙に含まれる健康機能性・安全性カテゴリの語を収集できると結論づけた。＜研究の方法（2）＞

（5）野菜商品レビューでは“美味しい”等いずれの品目でも高出現するキーワードを認めた。そこで、レビューの記述内容を高出現キーワードではなく、品目に特徴的なキーワード（分布の偏りが大きいキーワード）の集合体で表現し野菜品目間で比較することとした。品目に特徴的なキーワードを観点毎に整理したところ、サツマイモ、トウモロコシ、ジャガイモ、トマトは、カテゴリ_ブランドに分類されるキーワードが多かった（表 3）。生産者と卸売業者間で利用されるベジフルコードにおいてもトマト・ジャガイモ・ネギ・トウモロコシ・サツマイモではカテゴリ_ブランドに分類される商品名が多く、ネギ以外は野菜商品レビューと同様の傾向を示した。以上から、

トマト・ジャガイモ・トウモロコシ・サツマイモについては、ネット通販でも実際の市場と同様に品種・品種総称・地域等のブランド名で販売されブランド志向の消費者が多いこと、ネギについては今後ネット通販でブランド名での販売が増える可能性があることを推測した。また、野菜商品レビューでは、嶽きみ（トウモロコシ）、シンシア（ジャガイモ）、トマトベリー（トマト）のようなベジフルコードの商品名にないキーワードが出現しており、希少ブランドの販売が確認できた。3 カテゴリに分類した品目に特徴的なキーワードの数により、レビューの記述内容を 3 分類できた。この記述内容の分類結果から、消費者の品質知覚に影響を及ぼす品質保証要素は品目によって異なることが示唆された。＜研究の方法（3）＞

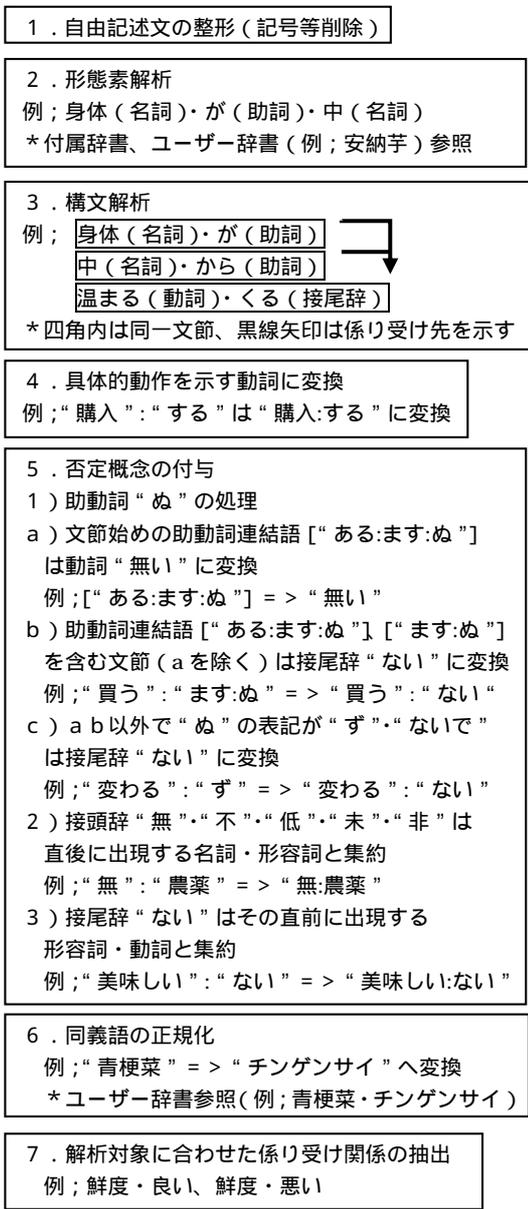


図 1 野菜商品レビューの概念抽出工程

表1 TermExtract に基づき選定したドメイン概念語の分類

カテゴリ	概念数(割合)	ドメイン概念語例
商品の名称 ^{注)}	278 (60.3)	
・品種	94	安納芋、ピュアホワイト
・品種群	57	ミニトマト、サニーレタス
・品物	65	ゴボウ茶、豆苗、新玉葱
・作物	28	小松菜、冬瓜、ツルムラサキ
・作物分類	5	葉物、根菜、緑黄色野菜
・ブランド	29	鳴門金時、深谷ねぎ、嶽きみ
健康機能性	32 (6.9)	食物繊維、栄養価、冷え性予防
安全性	29 (6.3)	無農薬、放射能汚染有機野菜
料理・食事	28 (6.1)	冷蔵庫、離乳食、アルミホイル
商品特性	30 (6.5)	食感、肉厚、大きめ、辛味
注文・配送	12 (2.6)	クール便、メール便
地名	9 (2.0)	安納、淡路島、西日本
植物部位	7 (1.5)	葉先、外葉、外皮、大根葉
農法	6 (1.3)	水耕、特別栽培、自然農法
その他	30 (6.5)	送料無料、生産者、説明書
合計	461 (100)	

注) 商品の名称カテゴリは品種・品種群、品物、作物・作物分類、ブランドの名称カテゴリを含む。

表2 . 消費者と研究者語彙の概念比較

カテゴリ名	概念数		P 値 ^{注2)}	出現率 B/A*100
	消費者 ^{注1)} (A)	研究者 (B)		
商品の名称 ^{注3)}	278	146	<0.01	52.5
・品種	94	51	0.27	54.3
・品種群	47	37	0.44	78.7
・品物	65	29	0.08	44.6
・作物・作物分類	33	22	1.00	66.7
・ブランド	29	7	0.03	24.1
健康機能性	32	27	0.38	84.4
安全性	29	29	0.12	100
料理・食事	28	21	0.74	75.0
商品特性	30	25	0.43	83.3
注文・配送	12	7	0.99	58.3
地名	9	8	0.70	88.9
植物部位	7	5		71.4
農法	6	6		100
その他	30	28	0.20	93.3
合計	461	302		65.5

注1) 消費者；ドメイン概念語が示す概念の数

研究者：農研機構 HP 出現語が示す概念の数

注2) 概念数 A と B をカテゴリ毎にカイ 2 乗検定した場合の p 値を示す。は期待値が 5 未満のため未検定。

注3) 商品の名称は品種・品種群、品物、作物・作物分類、ブランドの名称を含む。

表3 品目特徴的キーワード¹⁾の品目間差異

品目名	カテゴリ		
	ブランド	安全性	健康機能性
サツマイモ	安納芋 ベニハルカ ベニアズマ 鳴門金時、安納紅 ベニコガネ、紫芋	安心する 国産 安全だ	
ショウガ	黄金生姜 金時生姜	国産	温まる、冷え性、冷え性予防、風邪、冷え性改善
タマネギ		国産 安心する 無農薬	血液さらさら、ダイエット、温まる
トウモロコシ	ピュアホワイト 味来、ゴールド ラッシュ、嶽きみ、 甘々娘、恵味 サンライズコーン サニーショコラ	安心する 安全だ	
ニンジン		無農薬 安心する 有機 国産	療法、病気 酵素、ダイエット、癌
ジャガイモ	キタアカリ インカのめざめ メークイン レッドムーン インカ、男爵芋 とうや ホッカイコガネ シンシア レッドアンデス アカリ	安心する 安全だ 安全だ	
トマト	塩トマト トマトベリー アイコ、アメーラ 優糖星、桃太郎 赤糖房 徳谷トマト	国産 安心する 無農薬 安全だ	リコピン
ゴボウ			アンチエイジング、便秘、若返る、美容、便通、浮腫む、利尿、便秘解消
ニンニク		国産	体調、体力 風邪、疲労回復、肩こり

出現レビュー数 10 以上、フィッシャーの正確確率検定の P 値が低い 100 キーワードを品目特徴的キーワードとする。

5. 主な発表論文等

〔雑誌論文〕(計2件)

竹崎あかね、大浦裕二、河野恵伸、木浦卓治、林武司、自然言語処理を利用した農産物テキストからの概念抽出-野菜商品レビューを対象事例として-、農業情報研究、査読有、25、2016、47-58

DOI: <https://doi.org/10.3173/air.25.47>

竹崎あかね、大浦裕二、河野恵伸、木浦卓治、林武司、野菜商品レビューに出現する消費者語彙の特徴、農業情報研究、査読有、25、2016、105-115

DOI: <https://doi.org/10.3173/air.25.105>

〔学会発表〕(計2件)

竹崎あかね、大量の農産物テキストの効率的解析手法提案 ~野菜商品レビューを対象として~、平成29年度革新工学センター研究報告会、2018

竹崎あかね、大浦裕二、河野恵伸、木浦卓治、林武司、インターネット通販における野菜購入者の特徴抽出法、農業情報学会2014年度年次大会、2014

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

名称：
発明者：
権利者：
種類：
番号：
出願年：
国内外の別：

取得状況(計0件)

名称：
発明者：
権利者：
種類：
番号：
取得年：
国内外の別：

〔その他〕

ホームページ等

6. 研究組織

(1)研究代表者

竹崎 あかね (TAKEZAKI, Akane)
農業技術革新工学研究センター
高度作業支援研究領域・上級研究員
研究者番号：40550520