

ウェブ閲覧履歴を用いた農産物購入者の志向度の推定 Estimating consumer inclination for agricultural products from Web browsing history

竜田 力[†] 中西 惇[†] 梶原 祐輔[‡] 島川 博光[‡]
Riki Tatsuta Sunao Nakanishi Yusuke Kajiwara Hiromitsu Shimakawa

1. はじめに

近年、農産物を販売するECサイトが増加している。インターネットを活用した農産物販売は3つの良い影響を持つ。農家は売上を増加できる。また、農家は消費者との交流でやりがいが生まれる。高品質農産物提供の実績によりブランド力が獲得できるなどの効果がある[1]。農産物を販売するECサイトは、家に届けるなどの便利さがある。農産物を販売するECサイトは、今後子育て世帯や高齢者世帯において利用拡大が見込まれることが調査で示されている[2]。一方で、農家がECサイトでの農産物の販売の成功には、多くの消費者に自サイトを訪問してもらう集客作業が欠かせない。そのため、農産物を販売するサイトには消費者の購入意欲を増加させるような情報が提示されている。一般的な通販サイトに含まれるレビュー・価格・内容量などが提示されている。ECサイトは上記の項目に加え、農産物の栽培方法や農家のこだわり、品種の説明、おいしさをアピールするような情報も掲載されている。これらの情報は消費者の購入意欲を増加させる[3]。

購入意欲を向上させる情報は、消費者の嗜好によって異なる。例えば、食材、メニュー、食感、味、味付け、風味への嗜好がある[4]。また農産物の購入には、これらの嗜好とは別に、安全・安心や価格など、ユーザ固有の考え方やこだわりである志向が存在する[5]。たとえば、高価でもおいしさと安全性に拘る消費者もいれば、見た目よりも価格を気にする消費者もいる。個々の消費者の嗜好と志向の組み合わせは、その消費者の購買活動を規定する特定の志向(傾向)を形成する。栽培方針にあった消費者に的を絞るマーケティングが農家には必要である。しかし、高度な集約農業に従事する我が国の農家は、多種類の作物の栽培に忙殺され、マーケティングを考える余裕は無い。ECサイトでの集客支援の方法として、自サイトが検索時に推薦されやすくするSEO(Search Engine Optimization)があるが、これではユーザの嗜好と志向を考慮した上での的を絞った集客はできない。一方で、消費者は、自らの嗜好と志向に合致する農家から納得したうえで、得心して農産物を購入したい。たとえば、安全志向の消費者ならば、農家の評判を考慮し、価格との均衡を取りながら、できるだけ安全な農産物を購入するために、ECサイト上の記述を丁寧に読むであろう。現状のECサイトでの取引では、農家もつ栽培上の方針と、消費者がもつ味の嗜好と購入に対する志向との整合性が考慮されていない。消費者は、検索エンジンを駆使して、膨大な数の農産物販売のECサイトから、自らにあったものを手作業で探しているのが実情である。しかし、検索エンジンは期待はずれの農産物を扱うサイトも多く紹介し、消費者は選択に戸惑っている。

本研究では、消費者と農産物ECサイトを適切にマッチングするために、消費者の嗜好と志向を、ECサイトに書かれたWebコンテンツとそれに対するユーザの閲覧履歴から推定する手法を提案する。この手法は、LDA(Latent Dirichlet Allocation)を用いて、ECサイト上のWebページ内のトピック混合比とそのWebページの閲覧時間の積をトピック注目度とする。本手法は、このトピック注目度を説明変数とする志向度モデルを、多重ロジスティック回帰モデルを基に構築する。実験の結果、構築された志向度モデルは、消費者の安全志向度を正しく算出できることがわかった。

2. ユーザの志向にあった情報・商品の提示

消費者は農産物の味や外観、香りなどに関する嗜好を持っている。この嗜好は、年齢、世代間、性別、地域、職業などによって異なると言われている。個々の消費者の嗜好と志向の組み合わせは、その消費者の購買活動を規定する特定の志向(傾向)を形成する。

ユーザの志向を推定する研究としては[6]がある。これらはユーザに対して自由記述式のアンケートやプロフィールの入力などを行ってもらい、取得した情報をテキストマイニングで解析し、農産物に対する志向を明らかにしている。しかし、ユーザの志向は時勢や流行によって異なる。例えば福島原発事故や食品偽装の問題でユーザの安全、安心への志向が高まった。このことから、ユーザの考え方は日々変化するものと考え、その上でユーザの志向を逐次更新する必要がある。ユーザに何度も自由記述式のアンケートやプロフィールを記載してもらうことは、ユーザの負担となる。したがって、日々変化するユーザの志向をリアルタイムで抽出する手法が必要となる。

3. 志向にあわせた農産物の推薦

3.1 消費者の志向の抽出

ユーザの志向をリアルタイムで抽出するために、ユーザの閲覧履歴を用いる。また、ユーザの志向を抽出するアプローチとして、今回はECサイト上のコンテンツであるWebページを用いる。ECサイトには、多種多様なコンテンツが提示されている。その中には、ユーザの志向に適合しており、ユーザを刺激するようなコンテンツがある。したがって、ユーザの志向は商品を選択する時にあらわれると考えられる。また志向ごとにユーザが閲覧するWebページが異なることが考えられる。例えば、安全を重視するユーザは、安全に配慮した無農薬のトマトや国産のトマトを買う傾向にあると考えられる。この場合、ユーザは農産物の種類やどこでこのトマトが作られたのかを調査すると予想される。一方、安全よりも経済性を重視するユーザは価格が安いので、外国産のトマトを好んで買う傾向にあると考えられる。この場合、ユーザはECサイト内の農産物を示すページを、価格が低い順に閲覧

[†]立命館大学情報理工学部

[‡]立命館大学大学院理工学研究科

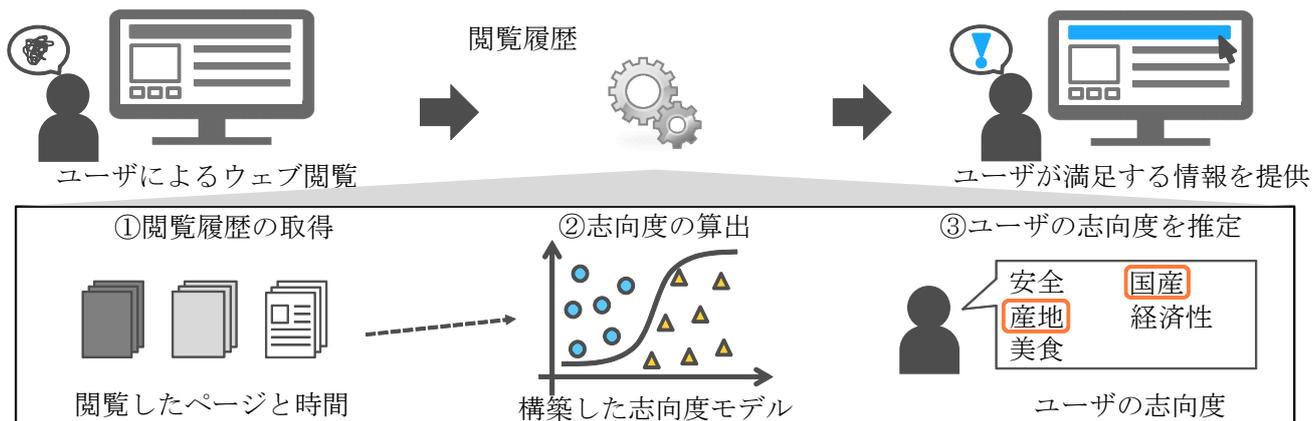


図1: システム概要図

していくと予想される。またユーザーの志向を刺激するコンテンツが多く記載されている Web ページは閲覧時間が長くなると考えられる。そこで本手法では Web ページに記載しているトピックと Web ページを閲覧した時間からユーザーの志向を推定する。本研究では、食に関する志向の調査結果 [5] や農産物購入時の選択基準に関する研究 [8], [9] から、安全志向・国産志向・産地志向・経済性志向・美食志向に着目する。図1に示すように、本研究は閲覧履歴から推定されたユーザーの志向に基づいて、農産物を示した Web ページを推薦する手法を提案する。まず、農産物を販売する EC サイトにおいて、ユーザーの閲覧履歴を取得し、閲覧したページと時間を抽出する。次にあらかじめ構築したモデルに、取得した情報を入力し、そのユーザーがどれくらい各志向を持っているかの度合い（以下、志向度）を算出する。最後に算出された志向度にあわせ、ユーザーの志向にあった農産物の情報を推薦する。

例えば、「有機トマト」「フルーツトマト」「ミディトマト」などはトマトという同じカテゴリに属する。本論文では、このようなカテゴリを商品カテゴリと定義する。消費者は、多様なサイトを訪問して農産物を探す。本手法では、特定の商品に関するページを Item ページ、複数の商品が一覧となって表示されるページを Search ページ、Google 検索や Wikipedia などの EC サイト以外を other ページと定義する。Item ページに含まれる単語や商品のタイトルを用い、商品カテゴリごとに Item ページを分類する。各々の消費者の閲覧履歴から、各商品カテゴリ毎に Item ページを抽出する。本手法は、以下に手順を用い、各々の消費者の志向を推定するモデルを構築する。

1. Bag-of-words を用い、Item ページから単語の集合を抽出する。LDA を用い、Item ページ毎のトピック混合比を取得する。本手法は、各々の Item ページにおける閲覧履歴とトピック混合比を用い、トピック注目度を算出する。
2. 消費者の志向度を算出するために、重要な変数を選択する。
3. 各々の消費者のために重要な変数を用い、志向度モデルを構築する。

3.2 閲覧時間を用いたトピック混合比の重み付け

文書を形態素解析し、単語に切り分ける。切り分けられた単語の中から志向と関連のあると考えられる名詞・形容詞・動詞だけを抽出し、bag-of-words を適用する。bag-of-words を適用し、取得した単語群に LDA (Latent Dirichlet Allocation [10]) を適用し、Item ページごとにあらわれる各トピックをあらわす単語群とトピック混合比を取得する。Item ページには志向と関連する単語が多数含まれているため、LDA を用いることで、志向ごとに分類されたトピックが取得できる。

ユーザーの志向の有無を判別するために、ユーザーの興味の対象があらわれる閲覧時間を用いる。ユーザーの興味の対象と閲覧時間の関連を調査した既存研究では、閲覧時間とページに含まれる文書に対する興味の度合いが関連しているのが明らかにされている [11]。LDA は文書が複数のトピックから成ることを仮定したモデルである。そのため、文書に対する興味の度合いは、その文書に含まれるトピックに対する度合いと言える。したがって、ユーザーの閲覧時間とそのページのトピックに対する興味の度合いも関連している。農産物を販売する EC サイトの Item ページには、複数の志向に関する単語が含まれている可能性がある。しかし、Item ページごとに閲覧時間で重み付けすると、ユーザーが Item ページに含まれるどの志向に着目したのかを考慮できない。

したがって、1つのページに含まれる志向に関する単語を分類する必要がある。そこで本手法では、Item ページごとに取得するトピック混合比を用いることで、単語を細分化し、ユーザーのそれぞれの志向に対する度合いの取得を可能にする。式1にトピック混合比のベクトル T_{cp} を示す。

$$T_{cp} = [T_{cp1} \quad \cdots \quad T_{cpk} \quad \cdots \quad T_{cpK}]^T \quad (1)$$

$$(1 \leq c \leq C), (1 \leq k \leq K), (1 \leq p \leq P_c)$$

k, c, p はそれぞれ特定のトピック、特定の商品カテゴリ、特定のページを示す。 K, C は、それぞれトピック数、商品カテゴリ数である。 P_c は c 番目の商品カテゴリに属するページ数を示す。 T_{cpk} は c 番目の商品カテゴリに属する p 番目のページのトピック混合比の第 k 成分である。式2

で T_{cp} をそれぞれのページの閲覧時間で重み付けし、商品カテゴリごとに総和したベクトル I_c を示す。

$$I_c = [I_{c1} \ \cdots \ I_{ck} \ \cdots \ I_{cK}]^T = \sum_{p=1}^{P_c} (\mathbf{T}_{cp}) d_{cp} \quad (2)$$

I_{ck} は c 番目の商品カテゴリの第 k トピックに関する、重み付けされたベクトル成分である。 d_{cp} は c 番目の商品カテゴリに属する p 番目のページの閲覧時間を示す。 I_{ck} はユーザが同じ商品カテゴリに属するページを多く閲覧するほど、値は大きくなる。また、ユーザによっては閲覧時間や閲覧回数が異なるため、 I_{ck} を正規化する必要がある。式3に I_{ck} を正規化し、取得できる $S(I_{ck})$ を示す。

$$S(I_{ck}) = \frac{I_{ck}}{|I_c|} \quad (3)$$

商品カテゴリごとに抽出する $S(I_{ck})$ をトピック注目度とする。

3.3 志向度推定において重要な変数を選択

本節では、節3.4にて作成するモデルに入力する説明変数を選択する。それにより、志向の有無によってトピック注目度の差があらわれるトピックを抽出する。式4にユーザごと取得したトピック注目度の集合 $G\{n_{ck}\}$ を示す。

$$G\{n_{ck}\} := \{S(I_{ck})_1, \dots, S(I_{ck})_u, \dots, S(I_{ck})_U\} \quad (4)$$

$$(1 \leq u \leq U)$$

u は u 番目のユーザである。 U はユーザ数である。式5.6で集合 $G\{n_{ck}\}$ の内、 n 番目の志向を持たない群・持つ群でユーザを分類し2つの集合である、 $X\{n_{ck}\}$ と $Y\{n_{ck}\}$ を示す。

$$X\{n_{ck}\} := \{G\{n_{ck}\} | P(\{G_{ck}\})\} \quad (5)$$

$$Y\{n_{ck}\} := \{G\{n_{ck}\} | Q(\{G_{ck}\})\} \quad (6)$$

P は志向を持たないという条件である。 Q は志向を持つという条件である。 n は n 番目の志向を示す。 N は志向の数である。 n_{ck} は n 番目の志向に対する c 番目の商品カテゴリのトピック k を示す。

ユーザの志向はあらかじめアンケートやインタビュー法などにより取得する。このことはアンケートやインタビューに答えるユーザに既存研究と同様の負担を強いてしまう。しかし、本研究では未知のユーザに対して適用可能な一般的なモデルを構築することを目指している。本研究では未知のユーザに対しての負担はない。志向を持たない群・持つ群とそれらの群に属するユーザのトピック注目度の差を算出する。この群におけるトピック注目度の差が大きいトピックを見ることにより、志向を持たない群・持つ群で着目しているトピックの違いを把握できる。志向を持たない群・持つ群とでトピック注目度の差を算出するために、ノイズに左右されにくい中央値を用いる。式7,8にて志向を持たない群・持つ群の要素であるトピック注目度の中央値を算出する。

$$MedianX(n_{ck}) = \begin{cases} x_{(a+1)/2}, & \text{if } a \text{ is odd.} \\ \frac{1}{2}(x_{a/2} + x_{a/2+1}), & \text{if } a \text{ is even.} \end{cases} \quad (7)$$

$$MedianY(n_{ck}) = \begin{cases} y_{(a+1)/2}, & \text{if } a \text{ is odd.} \\ \frac{1}{2}(y_{a/2} + y_{a/2+1}), & \text{if } a \text{ is even.} \end{cases} \quad (8)$$

式9で $MedianX(n_{ck})$ と $MedianY(n_{ck})$ の差である $diff(n_{ck})$ を算出する。

$$diff(n_{ck}) = MedianX(n_{ck}) - MedianY(n_{ck}) \quad (9)$$

式10で n 番目の志向に関する中央値の差の集合である $Diff_n$ を示す。

$$Diff_n := \{diff(n_{11}), \dots, diff(n_{CK})\} \quad (10)$$

着目しているトピックの違いの中でも、中央値の差が大きいものは志向を持たない群・持つ群を特徴付ける重要な値だと考えられる。本手法では、この中央値の差を昇順にし、上位 R 個を $High_n$ 、下位 R 個を Low_n とする。 $High_n$ に対応するトピックは、 n 番目の志向を持つ群が注目し、それを持たない群は興味を示さないと云える。 Low_n に対応するトピックは、 n 番目の志向を持たない群が注目し、それを持つ群は興味を示さないと云える。

3.4 志向度モデルの構築

節3.3にて取得した $High_n$ 、 Low_n に属するトピック注目度を説明変数、あらかじめ取得しておいた各ユーザの志向を目的変数として、ユーザの志向度を推定するモデル（以下、志向度モデル）を作成する。志向度モデル作成には、多重ロジスティック回帰分析を適用する。ステップワイズ法を用い、モデルへの当てはまり度が低い説明変数を除去し、当てはまり度が高いモデルを作成する。

作成した志向度モデルに、トピック注目度の入力することで、志向が未知のユーザに対しても、志向の有無を判別できる。あらかじめECサイト上のコンテキストに対してLDAを適用することで、事前に各Webページのトピックを把握することができる。そのため、志向度を推定するためにリアルタイムで取得する必要がある項目は、ユーザの閲覧時間のみである。

4. 実験

本手法の有用性を検証するのに実験した。本実験の目的を以下に示す。

- ユーザの志向の有無によって、閲覧したトピックに差の有効性の検証
- 志向度モデルの妥当性の検証

4.1 実験方法

被験者17名に対し、ECサイトを使って農産物を購入する実験した。実験開始から終了までの被験者の閲覧履歴を取得した。食に関する志向があらわれやすくなるように、価格を意識させない想定した次のようなシチュエーションに沿って農産物を購入してもらった。親戚10人くらいで集まり、おばあちゃんの還暦祝いをしたい。お母さんからは「お金がいくらでもかかってもいいので料理を作ってあげなさい。お金は私が出すから。」と言われた。おばあちゃんの家で料理をしようと思ったが、家は田舎にあるため、近くにスーパーがない。だからネット通販で

表 1: アンケート項目

志向	アンケート内容
美食志向	Q1. 【味の良さ】を気にしましたか Q2. 【新鮮さ】を気にしましたか
安全志向	Q3. 【安全性の高さ】を気にしましたか
国産志向	Q4. 【国産】を気にしましたか
産地志向	Q5. 【産地】を気にしましたか
経済性志向	Q6. 【価格】を気にしましたか

表 2: 志向に関するアンケート結果

	User	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Q1.	○	○	○	○	○	○	○	○	X	○	○	○	○	△	○	△	○	X
Q2.	○	○	X	○	○	○	○	△	X	X	△	X	○	○	○	○	○	△
Q3.	△	X	○	○	○	○	△	△	○	X	X	X	X	○	○	○	○	○
Q4.	○	○	○	○	△	○	○	○	○	X	○	○	○	○	○	○	○	○
Q5.	○	○	X	○	○	○	○	△	○	X	X	○	X	X	○	○	○	○
Q6.	○	X	○	X	○	○	○	X	△	○	○	X	○	○	X	X	○	○

(注: ○…はい, △…わからない, X…いいえ)

予め自宅から注文して、おばあちゃんの家に材料を届けてもらい、そのまま料理するという形を取りたい。出来合いのものは嫌いなので、手料理が食べたい。あとお米がなくなりそう。だからオススメのものをついでに買ってほしい。シチュエーション内で想定した料理は、サラダ、松茸ご飯、煮物、肉料理である。購入させた料理に必要な材料は、トマト、きゅうり、松茸、お米、大根、里芋、鶏肉、肉料理に必要な肉である。購入させた料理に必要な材料の量はあらかじめ被験者にレシピと一緒に提示した。農産物を購入する EC サイトは指定し、情報を収集するための検索は許可した。今回は、ユーザの持つそれぞれの志向をアンケートで取得する。アンケートは志向の調査 [6] を参考に、それぞれの志向 (美食志向・安全志向・国産志向・産地志向・経済性志向) に対して、ユーザが気にする EC サイト上の情報を関連付けし、作成した。美食志向に関しては、その農産物の美味しさを気にすると考えられる。例えば、糖度が他より高く甘い、こだわりの栽培方法で味が他よりも良いなどの情報や朝採りで鮮度が高く美味しいといった情報が EC サイト上で提示されている。美食志向はこれらの味の良さや鮮度を気にする。価格や量などの情報が EC サイト上で提示されている。経済性志向に関しては、値段が安く、量の多さを気にすると考えられる。したがって価格や量を気にすると考えられるが、今回は実験方法で、購入してもらう量は指定していたため、経済性を志向するユーザは、価格を気にしたかどうかだけを問う。安全志向・国産志向・産地志向に関してのアンケート項目は自明である。農産物購入後、被験者に対し、商品を購入する時に表 1 のアンケート項目を気にしたかどうかをはい・いいえ・わからないの三段階評価でアンケートを行った。

4.2 評価

4.2.1 閲覧した Item ページの抽出

トピック注目度を取得するのに用いたデータは被験者 17 名が閲覧した Item ページに含まれる文書である。Item ページ数は 503 ページである。ただし、節 4.1 にて肉料理に必要な肉を被験者に選んでもらった結果、牛肉と豚肉に 2 通りに分かれたため、今回はそれらを同じ商品カテゴリとして分類した。Item ページには、ユーザの検索意図とは関係のない情報が表示される場合があった。例えば、推薦される商品の情報やサイト特有のランキング商品などが挙げられる。今回はこれらの文書はユーザの意図とは関係がないため、除外した。また、Item ページに含まれるレビューも今回においては対象外のデータとして除外した。ページを商品カテゴリごとに分類した。

4.2.2 Item ページからトピックの抽出

分類した Item ページ群に対し、節 3.1 を適用した。ただし、頻出度が高い単語の中でも、志向と関係のない商品カテゴリに関する単語や記号はトピックを分けた後、そのトピックがどのような志向を表しているかを考察できない。したがってこれらの単語を除外し、bag-of-words を作成した。作成した bag-of-words にギブスサンプリング法 (collapsed Gibbs sampling) [12] を用いた LDA を適用した。LDA のハイパーパラメータ α を 0.1, β を 0.1, 試行回数は 30000 とした。トピック数は $2 \cdot 4 \cdot 6 \cdot 8 \cdot 16$ の 5 パターンを試し、トピックを分類するのに単語が志向ごとに分かれていると考えられるトピック数 4 を選択した。抽出された 4 つのトピックに対し、各トピックに含まれる上位 5 つの単語のラベル付けをした。

4.2.3 ユーザの志向の取得

ユーザごとの志向の有無に関しては、表 1 のアンケート項目に対して、はいと答えた場合、その志向を持つとした。わからないという回答に対しては、その志向を持っていたとしても、答え方が曖昧なので、そのユーザにとって本当に必要な情報は他の志向である可能性や無くてもよい情報だと考えられる。したがってわからない・いいえと回答した場合はその思考を持っていないとした。美食志向に関しては、美味しさ・新鮮さの項目に対してどちらもはいと答えた場合のみ美食志向を持っているとした。わからない・いいえの扱いに関しては上記と同様である。

4.3 結果

節 3.2 を適用し、トピックを抽出した。また、単語群からラベル付けした。

4.3.1 トピック注目度の抽出

節 3.2 を適用し、トピック注目度を抽出した。図 2 に抽出されたトピック注目度の中央値を示す。

表 2 に志向に関するアンケート結果を示す。

4.3.2 志向度を用いたユーザの志向の有無を判別

作成した志向度モデルを用いて、ユーザにおける各志向の有無を志向度から判別できるかどうかを分析した。表 3 に多重ロジスティック回帰分析の結果を示す。

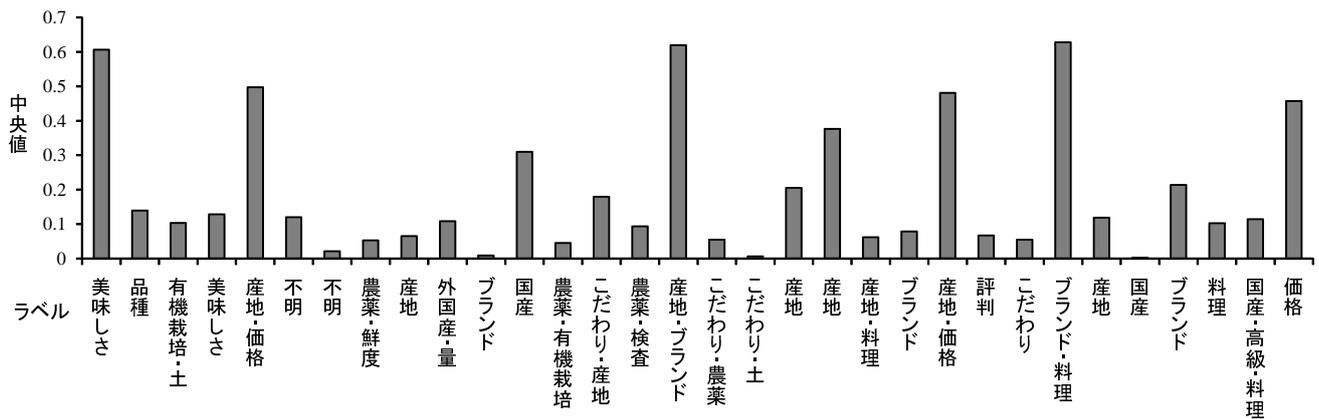


図2: トピック注目度の中央値

表3: 多重ロジスティック回帰分析の結果

志向	ラベル	Estimate	Std.Error	z value	Pr(> z)
国産	農薬, こだわり	44.99	20068.94	0.002	0.998
	ブランド	-494.47	168178.07	-0.003	0.998
	国産, 高級, 料理	1074.23	358333.88	0.003	0.998
美食	美食	-898.71	305721.58	-0.003	0.998
	AIC	12.773			
経済性	料理	30.8814	21.4463	1.44	0.1499
	AIC	18.025			
	農薬, こだわり	-7001	741189	-0.009	0.992
安全	美食	-23338	2467722	-0.009	0.992
	産地, 価格	5922	626767	0.009	0.992
	AIC	8			
産地	国産	2.802	1.984	1.412	0.1579
	価格	6.75	3.32	2.033	0.042
	AIC	20.537			
産地	産地	-3.823	2.396	-1.596	0.111
	美食	8.476	6.372	1.33	0.183
	AIC	20.306			

5. 考察

多重ロジスティック回帰分析の結果から、安全志向以外は推定に有効な変数を抽出できなかった。本節では、有意確率の向上に向けての考察を述べる。本研究では、ユーザの志向の有無と閲覧しているページの内容・時間が関連していると仮定した。時間を用いた理由は、閲覧している時間が長いほどそのページに興味があり、閲覧している時間が短いほどそのページは興味がないと考えたからである。しかし、ページに興味の有無とは別に、ページの走査の仕方によっても閲覧している時間は変わる。今回はページの走査の仕方を考慮せず重み付けしたため、有意確率が低くなった原因の1つとして考えられる。この原因を検証するのに、ページの走査の仕方によって閲覧時

表4: タブ型・クリック型

	タブ型	クリック型
閲覧時間の平均 (秒)	51.945	42.354
閲覧回数の平均 (回数)	1.935	1.457

間に変化するかを調査した。調査方法は、Item ページを開く時にタブを使って開くか、クリックして開くかの走査の仕方によって閲覧時間・回数の差を見た。表4にItem ページを開く時に、タブを使って開くか、クリックして同じタブにそのページを開くかの回数をユーザごとに示す。

Item ページの閲覧時間・回数の平均を比較した場合、タブ型の方が閲覧時間・閲覧回数が多いと言える。したがって、ユーザの走査の仕方の違いによって閲覧時間に差があらわれるため、閲覧時間を用いてトピックに重み付けを行う場合、個人の走査の仕方を考慮する必要がある。個人の走査の仕方を考慮すれば、より当てはまり度が良いモデルの作成を期待できる。

6. おわりに

本論文では、ECサイトにおける農産物購入者の志向に基づいた情報を提示するために、志向度を推定した。Item ページに含まれる文書から、bag-of-words を作成し、LDA を用いてトピック混合比を抽出した。トピック混合比と閲覧時間を重み付けし、トピック注目度を算出する。ユーザの志向と重要なトピック注目度を用いて、多重ロジスティック回帰分析を適用し、志向度モデルを構築した。構築した志向度モデルにトピック注目度を入力することで、ユーザの志向を推定できる。統計的有意水準を5%未満と設定すれば、安全志向に関する志向度に関しては、価格に関するトピックが有意確率0.042と有意な変数として抽出された。今後は、ページの遷移などを考慮し、より当てはまり度がよい志向度モデルの作成を試みる。

参考文献

- [1] 日本政策金融公庫: 農業の6次産業化に関する調査結果 <https://www.jfc.go.jp/n/findings/>

- pdf/topics_111202_1.pdf(accessed 2015-03-08).
- [2] 一般社団法人 JC 総研『野菜・果物の消費行動に関する調査結果』, http://www.jc-so-ken.or.jp/pdf/agri/investigate/141203_01.pdf(accessed 2015-02-02).
- [3] 農林水産情報交流ネットワーク事業: ”野菜の生産流通情報に関する意識・意向調査結果”, <http://www.maff.go.jp/j/finding/mind/pdf/20040820cyosa.pdf>(accessed 2015-02-02).
- [4] 「日本人の味覚と嗜好」 <http://www.maff.go.jp/j/keikaku/syokubunka/culture/mikaku.html>(accessed 2015/01/29)
- [5] 日本政策金融公庫: 消費者の「食の志向調査」, https://www.jfc.go.jp/n/findings/pdf/topics_140911a.pdf(accessed 2015-02-02).
- [6] 磯島昭代. ”テキストマイニングによる農産物に対する消費者ニーズの把握.” フードシステム研究 16.4 (2010): 38-42.
- [7] 山口和子, and 高橋史人. ”食品の嗜好に関する研究(第2報).” Science of Cookery 15.2 (1982): 104-113.
- [8] 磯島昭代. ”テキストマイニングを用いた米に関する消費者アンケートの解析.” 農業情報研究 15.1 (2006): 49-60.
- [9] 磯島昭代. ”米購入時の選択基準と消費者像.” 農業経営研究 36.3 (1998): 22-32.
- [10] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. ”Latent dirichlet allocation.” the Journal of machine Learning research 3 (2003): 993-1022.
- [11] Morita, Masahiro, and Yoichi Shinoda. ”Information filtering based on user behavior analysis and best match text retrieval.” Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval. Springer-Verlag New York, Inc., 1994.
- [12] Griffiths, Thomas L., and Mark Steyvers. ”Finding scientific topics.” Proceedings of the National academy of Sciences of the United States of America 101.Suppl 1 (2004): 5228-5235.