

# Web サイト閲覧行動のモデリングと評価

## Modeling online behaviors and their reason on a recipe-site

西尾 義英<sup>\*1</sup>  
NISHIO, Yoshihide

藤井 絵美子<sup>\*1</sup>  
FUJII, Emiko

安松 健<sup>\*1</sup>  
YASUMATSU, Ken

<sup>\*1</sup> シナジーマーケティング株式会社  
Synergy Marketing, Inc.

For improving the user experience on a web service, the designer should understand about users and their goal first, and then build a model of user behaviors. In this paper, we discuss how to infer the recipe-site user attributes and their goal at which time using the service. We apply probabilistic latent semantic analysis (pLSA) on online behaviors to get their semantic classes. The result shows that classes of recipe attributes can explain the viewer's attributes, classes of when recipes are viewed can distinguish the viewer's mode in the day or week, and classes of screen transition can distinguish what content the user wants.

### 1. はじめに

マーケティング活動において求められる顧客理解の粒度は、市場全体の傾向という統計的なものからセグメント毎の差違、さらには消費者一人一人へと進み、突き詰めるとその時間毎の変化にまで向かいつつある。この理由は消費者のニーズが多様化したことだけでなく、彼らが求めるものが、モノの所有・消費だけでなくコトとしてのサービス体験といった高次なものに変わってきたことが挙げられる。また、この顧客理解はかつて定性的かつ属人的なものであったが、大量消費や電子的な商取引が主流の現代ではこれを再現可能で、自動的なものとするのが求められている。

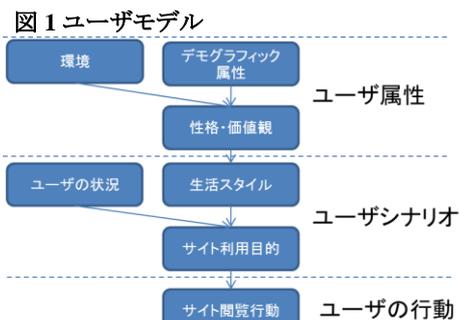
そのためのアプローチとして有効と考えられるのが、様々なデータを元に消費者と商品のカテゴリを同時に生成し、消費者の行動を確率モデルで表現するカテゴリマイニング[石垣 2011]という手法である。

本稿では、レシピサイト「ボブとアンジー」を題材にし、カテゴリマイニングによるユーザモデルの構築方法について述べる。

### 2. モデリングの概要

#### 2.1 モデリングの目的と想定するモデルの概念的構造

レシピサイトにおける顧客理解の動機はサイトにおけるユーザの体験を改善し、これによってサイトを継続的に利用してもらうことにある。このためにはユーザがサイトを利用する目的を知ることが最も重要と考えられる。利用目的はユーザの家族構成、料理スキルや価値観といった要素だけでなく、料理にまつわる生活スタイルやその日の気分といった要素によって時々刻々で変わるものと考えられる。この考えを概念的なモデルとして表現したものを図 1 に示す。



連絡先: 西尾 義英, nishio.yoshihide@synergy101.jp

### 2.2 モデル構築に必要なデータと収集方法

#### (1) ユーザ属性

「ボブとアンジー」は非会員制サイトであるため、ユーザについて知るにはアンケート調査を行うことにする。

#### (2) ユーザシナリオ

ユーザがどのような状況でどのような目的を持ってどのようにサイトを利用するか、という知識をアクセスログのような限定的なデータから得ることは難しい。アンケート調査するにしても質問者が想定する範囲あるいはユーザが言語化できる範囲でしか把握できない可能性がある。そこで本研究ではアンケート回答者からインタビュー協力者を募集し、インタビューを行うことでこれを補うこととした。

#### (3) ユーザの行動

一般に Web サイトのアクセスログを見ることでどのページをいつ表示しているかという情報が得られる。さらに非会員制のサイトにおいてユーザ毎の行動を識別するには Cookie を使う方法がある。アンケート調査をオンラインで実施し回答者の ID(Cookie ID)を Cookie に書き込んでおく、回答者のブラウザは Cookie ID をサイト訪問の度にサイトへ送信するようになるので、サイト側でこの ID をアクセスログにも書き込むようにしておく。こうしてユーザ属性と行動を関連付けることができる。

なおこの仕組みの限界として、サイトへの Cookie 送信をユーザが拒否できる点、ページを実際読んだか、ページのどこに注目したかということまでは分からない点、また人ではなくブラウザを識別しているため、異なる端末を使い分けてサイトを利用している場合に同一ユーザの行動として追跡ができない点、には注意が必要である。

### 3. 調査

具体的なデータ収集方法と収集結果について説明する。

#### 3.1 アンケート

価値観フレームワーク Societas[馬場 2013]の規定するアンケートの他、食に関する意識調査事例、ならびに共同研究者の株式会社エルネットが過去に実施した主婦を対象とした料理に関する調査事例を参考にして設問を設計した。年齢、性別、職業、家族構成といったデモグラフィック属性、価値観(性格・ポジネガ・人間関係・時間・お金、食や料理)、料理の習慣や料理スキルなどを

問うものとなった。募集はレシピサイトのユーザを対象とするメールマガジンやサイトバナーを通じて行われ、1300 件の回答が集まった。ただし非ユーザからの懸賞目的の応募が多く、採用した回答は 400 件程度に留まった。

予備調査としてアンケート回答を元に価値観類型の分布を確認した。我々の調査に基づく日本人の平均と比較したときに、ボブとアンジーのユーザ特徴は「#5-2 社交的な堅実ホームメーカータイプ」(余裕のある専業主婦タイプ)に強く表れることを確認した。

### 3.2 インタビュー

ボブとアンジーの特徴である価値観類型#5-2 と、#5-2 に比較的近い主婦的な価値観類型#5-1 が強い回答者をインタビュー対象とし、デモグラフィック条件を「30～40 代前半の、小学生の子供のいる(専業)主婦」に揃えて 3 名を選定した。訪問インタビューを実施した結果、主婦の生活パターンから一日の中でのレシピサイトのアクセス数ピークに次のような背景があることを把握できた。

#### (1) 10 時頃～昼

家族を送り出した後、掃除・洗濯などの家事の合間に、気分転換がてらレシピを見てメニューのヒントを得る

#### (2) 16 時頃を中心とする夕方

慣れない料理は作りながら、または、なじみの料理も、忘れがちな調味料の配合を確認するため等に見る

#### (3) 21 時以降

子供を寝かせた後、落ち着いて明日以降のメニューを考えたり、食材をまとめ発注したりするためにレシピをダウンロードする。

また、料理スキルの高いユーザからレシピの詳細情報は不要で、料理写真のみを見ているという意見があった。その理由は写真を見れば作り方は分かる、盛りつけの参考とする、新しい料理のヒントとしているというものが挙げられ、実際のログからも料理の検索結果一覧のみを閲覧していたということが分かった。

### 3.3 アクセスログ

「ボブとアンジー」のアクセスログに Cookie ID を記録し始めたのは 2013 年 9 月以降で、この期間のログを分析対象とした。

利用可能な情報は Cookie ID、アクセス日時、リファラ(リンク元 URL)、URL、検索サイトの検索ワード(リファラから取れる場合のみ)等で、リファラと URL からはページ内容と無関係なパラメータ(クエリ文字列)を除去した。

レシピページの URL からはレシピの ID を確認することができる。レシピには写真、材料、栄養価、作り方という情報が含まれる他、検索用に、料理の種類や調理方法、関係する季節のイベント、美容や健康についての効用といったカテゴリ情報が付与されている。今回はデータ処理のしやすさから、このカテゴリ情報のみをレシピの属性変数として使用する。

### 4. 分析

データが集まったところで、概念レベルのユーザモデルを構成する変数の候補とそれらの関係を確認する。

#### 4.1 ユーザ属性と行動の関連付け

分析のしやすさからレシピの閲覧に絞って、閲覧されたレシピの特徴とユーザ属性の関係を調べることで、料理の習慣やスキル、料理に対する価値観の違いが現れると考えた。

具体的には、まず Cookie ID と閲覧したレシピの検索カテゴリ情報に対して pLSA(確率的潜在クラス分析)を行う。情報量基準(AIC)によりクラス数 8 の結果を採用し、Cookie ID 毎のクラス所属確率に対するユーザ属性有無(1/0)の相関分析を行った。Cookie とレシピ ID に対する pLSA では共起行列の欠損率が高く、解釈が難しかったため採用しなかった。分析結果を定性的に解釈した結果を表 1 に示す。

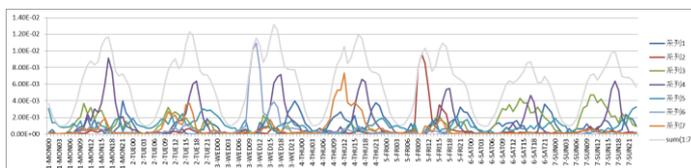
表 1 レシピカテゴリクラス

| # | クラス名        | レシピの特徴                           | ユーザの特徴                    |
|---|-------------|----------------------------------|---------------------------|
| 1 | ケーキ・焼き菓子    | オーブンを使ったお菓子                      | 調理頻度と料理スキル低め/食材はお金を掛けても良い |
| 2 | スイーツ        | オーブンを使わないお菓子                     | 健全な食事がしたい/レシピは人に教わる       |
| 3 | 汁物・煮物       | スープや味噌汁、煮物といったカテゴリ               | 食は団らん重視/食材は安く抑えたい         |
| 4 | 西洋焼き料理      | おもてなしやパーティ向けのオーブンを使うような西洋料理      | 料理は自己表現/健康に配慮したメニュー       |
| 5 | 副菜          | サラダなどヘルシーなおかず                    | 料理は昔からしている/ゴミの出にくい食材を選ぶ   |
| 6 | 主菜          | 晩ご飯のおかず                          | 団らん食卓/料理は省力で/調理スキル低め      |
| 7 | 中華・エスニック・麺類 | 中華料理や東南アジアのスパイスを使った料理、麺類にはパスタも含む | 料理には達成感がある/調味料にこだわり       |
| 8 | ごはんもの       | 炊き込みご飯やちらし寿司など                   | 料理スキル低め/食材へのこだわり薄い        |

#### 4.2 生活パターンと閲覧傾向の関係

主婦の生活パターンに関するインタビュー調査から、サイトを閲覧する曜日や時間帯の傾向は、生活時間に由来するサイト利用目的のクラスによって生じるのではと考えた。閲覧日時を曜日と時間で離散化した変数(SUN00～SAT23)を生成し、閲覧されたレシピ ID との pLSA を実施した。情報量基準(AIC)に基づきクラス数 7 の結果を採用する。時間帯特徴を見るため、時間帯とクラスの同時確率をプロットしたものを図 2 に示す。

図 2 月曜から日曜の時間帯毎のクラス確率分布



大まかに夕方、夜間、休日(土日)というパターンが見え、これは概ね想定通りであった。また水曜、木曜、金曜の午前 10 時にピークを持つクラスが抽出された。これは週次のコンテンツ更新に関連したものと推測でき、息抜きにレシピを見ているという仮説にマッチしていると考えられる。クラスをレシピに割り振り、レシピの検索カテゴリの特徴を見てみるとわずかだが以下のような傾向も見られた。

- 夕方⇒主菜や副菜で比較的手間の掛からないもの
- 夜間⇒デザート中心でパーティやイベント向き

- 休日⇒特徴が薄い、副菜が減り、麺類がわずかに増える
- 水、木、金午前⇒和食で汁物や副菜が多くダイエット向き

### 4.3 画面遷移に見るサイトの利用目的

前二つの分析はレシピのページを対象にしているが、インタビュアーからレシピそのものは見ないというユーザがいることが分かっており、レシピ閲覧以外の行動についても分析が必要である。今回は、リファラと遷移先の URL のパターンを分類することを試みた。サイトへの流入経路とサイト内の主な画面遷移にパターンが見られるならば、それはある瞬間におけるサイトの利用目的に関係していると考えたためである。同様にクラス数  $k$  を変えながら pLSA を実施し、AIC によってクラス数 6 の結果を採用した。クラスの解釈を表 2 に示す。

表 2 画面遷移クラス

| # | クラス名                                | 主な遷移元            | 主な遷移先              |
|---|-------------------------------------|------------------|--------------------|
| 1 | 検索サイトから流入                           | Google など        | さまざま               |
| 2 | 健康レシピを探す                            | 健康ポータルサイトや献立検索   | ダイエット・ヘルシーカテゴリのレシピ |
| 3 | トップページのリンクをクリック                     | トップページ           | さまざま               |
| 4 | コラムを閲覧<br>※コラム: あるテーマに沿ってレシピをまとめたもの | コラム一覧            | コラム                |
| 5 | お気に入り経由                             | リファラ無し           | トップページなど           |
| 6 | レシピ検索結果を巡回                          | レシピ検索<br>※サイト内検索 | レシピ検索結果            |

なお、分析にあたり、共起行列の欠損率を下げるためリファラに対していくつかの丸め処理を行った。

- 外部サイトからの流入はドメイン名のみを残す
- レシピページ(~/recipes/99999)など多数のコンテンツが階層的に現れる場合、コンテンツの種類が分かる階層(~/recipes/)までを残す

### 4.4 レシピのレコメンド実験

これまでの分析を検証するために、サイトに変更を加えることでユーザの行動を変えられるかどうかなを実験する。ユーザ属性の推定には多くの行動データが必要と考えられるため、時間帯のクラスと画面遷移のクラスに基づいて、レシピページに表示するお勧めレシピの表示を制御するという方針を立てた。

#### (1) 実験方法

サイトに大きな変更を加えることは困難だったため、あるレシピに対するお勧めレシピは固定で最大 3 つまでという制約の下で実験を行う。まず全 8055 件のレシピに対し時間帯クラスおよび画面遷移クラスを MAP で割り当て、それぞれ 7 および 6 のセグメントを作成する。次にこのセグメント毎にセグメントが一致するレシピの中からお勧めレシピを 3 つ選んでおく。ここでレシピ全体をランダムに 2:2:1 の比率で 3 グループに分け、一つめのグループに時間帯クラスに基づくお勧めレシピを設定、二つ目のグループは画面遷移クラスに対して同様にお勧めレシピを設定、三つ目のグループは比較対象とし、全てのレシピからお勧めレシピを 3 つ選んで設定する。

お勧めレシピの選出方針は、類似するレシピをお勧めするという方針をとった。具体的には、レシピ名同士の最長共通部分列(LCS)距離とレシピ検索カテゴリのビット列同士のハミング距離

の和をレシピ間の距離と定義し、距離が近いレシピを上位 3 件抽出するという方法である。全てのレシピから選ぶ方が平均的にはレコメンドの成功率は高いと考えられるが、レシピが閲覧される条件(時間帯または画面遷移)がクラスとマッチする場合は対応するセグメントからのレコメンド成功率が上回るかも知れないという仮定を置き、これを実験によって確かめる。

計測期間は 2014 年 2 月 17 日から 2014 年 3 月 2 日の 2 週間である。

#### (2) 実験結果

事前に置いた仮定に対しては否定的な結果を得た。つまりクラスに基づくセグメント内で類似度を基準にお勧めレシピを選ぶという方法では、そのクラスとマッチする閲覧状況であっても、全てのレシピからお勧めレシピを選ぶ方法を上回ることはなかった。

ただ、閲覧状況によってレコメンドの成功率に差があることは確認できた(表 3,4)。これは閲覧状況が異なると閲覧中のレシピ以外に興味を示すかどうか、または、レコメンドの方針がユーザにマッチしているかどうかは異なるということを示唆している。類似レシピのレコメンドが効きやすいのは、時間帯クラスで言うと「木曜午前」および「夜間」、画面遷移クラスで言うと「トップページのリンクをクリック」、「お気に入り経由」、「健康レシピを探す」、「コラムを閲覧」、であることが分かった。

表 3 時間帯クラス別  
レコメンド成功率

| クラス名 | 成功率   |
|------|-------|
| 木曜午前 | 0.109 |
| 夜間   | 0.102 |
| 休日   | 0.083 |
| その他  | 0.083 |
| 夕方   | 0.069 |
| 金曜午前 | 0.058 |
| 水曜午前 | 0.054 |

表 4 画面遷移クラス別  
レコメンド成功率

| クラス名            | 成功率   |
|-----------------|-------|
| トップページのリンクをクリック | 0.291 |
| お気に入り経由         | 0.177 |
| 健康レシピを探す        | 0.144 |
| コラムを閲覧          | 0.099 |
| レシピ検索結果を巡回      | 0.055 |
| 検索サイトから流入       | 0.051 |

## 5. まとめ

### 5.1 分析の結論

4 節で行った分析結果からは、行動データを用いることで、ユーザの属性やある瞬間のサイトの閲覧目的を推測できそうという感触を得た。ただし、モデルの有用性という面で予測に対する情報量や予測精度を定量的に評価するという課題が残る。

### 5.2 モデル構築の試み

アクセスログやユーザ属性からサイトの利用目的を予測するモデルを構築すべく、2013 年 12 月に、ボブとアンジーのユーザに対してサイトの利用目的を問う追加アンケートを行った。利用目的は複数選択式でインタビュー調査を元にして選択肢を設計した。アクセスログと紐付け可能な回答者が 20 人程度であったため、モデルの評価は保留している。

### 5.3 今後の課題

本研究を通じてデータを収集・統合する仕組みやそれを分析する流れを作ることができたが、データはまだ不足している。アクセスログは季節性を考慮すると最低 1 年分は必要であろうし、アクセスログと紐付けできるアンケートデータも確率モデルを構築するには不十分である。またアンケート調査方法の見直しも必要かも知れない。

今後はデータを増やしながらかめてモデルを構築し,サイトの改善を通じた検証を行っていきたい.

### 謝辞

本研究は株式会社エルネットとの共同研究です.同社には各種データのご提供にはじまり,ユーザインタビューの設計・実施・分析,ならびに一部データ分析を担当くださいました.またユーザモデルの検討や実験においても多大な協力をいただいております.ここに感謝の意を表します.

なお,本論文にて同社による分析結果を使用しましたが,編集は筆者によるものであり,文責は全て筆者にあります.

### 参考文献

- [石垣 2011] 石垣 司,竹中 毅,本村 陽一: 日常購買行動に関する大規模データの融合による顧客行動予測システム 実サービス支援のためのカテゴリマイニング技術, 人工知能学会論文誌 26 巻, 2011.
- [馬場 2013] 馬場 彩子,ベルタン マチュー,谷田 泰郎: 社会知としての消費者価値観構造モデルと類型「Societas」の構築, 人工知能学会全国大会(第 27 回),2013.