

# 企業事例セッション

---

---

---

---

---

---

---

---

---

---

## 経営情報学会 2019 年春季全国研究発表大会

経営情報学会 2019 年春季全国研究発表大会では、2018 年秋季大会に続いて企業事例セッションを開催しました。本セッションは産業界の方に（1）経営情報学に関わる企業での実践例や（2）経営情報学に関わる技術・サービスをご紹介いただくことで、他企業や学識者との情報交換・交流の促進を意図するものです。企業同士となると日常ではなかなか踏み込んだ情報交換の機会が少ないということが考えられ、学識側もまた企業のリアルに触れられる機会はその多くはありません。そこで、学会が交流の場を設定することは実務・学術の両者にとってよい機会になり得ます。また産学交流という視点でも共同研究（事業）への発展の可能性もあることから、本セッションにより大会がより一層魅力的な場となることを目指しています。

本誌をご覧の読者の方にも参考になる部分が多いと思いますので、今後の事業・研究にご活用いただければ幸いです。なお、本セッションは次回大会でも開催予定です。企業の皆様はぜひ発表をご検討ください。

### 【経営情報学会 2019 年春季全国研究大会優秀報告賞】

○齊藤武史 氏（株式会社マクロミル）

「発生／収集における時間概念が相違するデータの直接比較～購入ログとアクセスログの関係性について～」

# 発生／収集における時間概念が相違するデータの直接比較 ～購買ログとアクセスログによるレコメンドエンジン～

齊藤武史（さいとう たけし）  
株式会社マクロミル

## 1. はじめに

本稿は2019年6月に開催された経営情報学会で発表した内容を、おおよそぞったものである。

具体的には、弊社に蓄積されている同一人と定義されたモニター、約28,000人のインターネットアクセスログと購買ログを合わせて分析したもので、レコメンドエンジンとして、ファクトリゼーションマシン（以下、『FM』とする。）とWord2Vecのどちらの手法が優れているかを比較考察した。

インターネット広告費が地上波TV広告費と肩を並べる水準に成長し、インターネット上の行動と実購買行動の関係が重要性を増していることから、このような分析を行った。

## 2. 分析データ

購買ログについては、弊社のQPRのデータを使用した。QPR（Quick Purchase Report）<sup>TM</sup>は携帯型バーコードスキャナー／スマートフォンアプリを用いて、全国（沖縄除く）3万人のモニターが毎日の商品購入履歴を記録する購買データ。モニターに対する意識調査と掛け合わせ、消費者の日々の購買動向を捉える商品購買調査サービスを提供している。QPRでは、各バーコード（janコード）がどのような商品カテゴリに属するものであるかの情報も有している。

インターネットアクセスログについては、Web広告やWebサイトに敷設されたタグを介して月間100億レコードを上回る水準のログを取得しており、クッキーというWebブラウザの同一性を追跡できる技術を基に、各モニターのログであることを特定している。このログから得られるURLを、一定のロジックに基づきカテゴリ分類し、DDS（Digital Data Safari）としてサービス提供している。

## 2.1 具体的な分析対象データ

期間：2019年1月

モニター数：約28千ID

レコード数：購買約3百万、アクセス約5.7百万

## 2.2 両データの本質的な違い

購買ログが、実在店舗への物理的移動や購入費用が必要なものであることに対し、インターネットアクセスログは、一般的に特段の物理的移動は不要で、金銭的負担も限りなく低いものである。

またログの発生が即時に捕捉できているかについて違いが存在する。すなわち購買ログは購買後にモニターが購入商品のjanコードをスキャンすることでログとして収集できるのに対し、インターネットアクセスログは敷設タグへの接触が即時にログとして反映されるものである。

## 3. 分析方法

手法の一つとして、FMを選択した。これはレコメンドエンジンで一般的な手法を進化させたもので有ることに加え、timestampの厳密な前後関係の要素なしに学習が可能なることから、両ログデータ間に存在する性質の違いによる影響を排除できるものと考えたからである。

もう一つの手法として、Word2Vecを選択した。これは、文章を単語単位に分割させたものを学習させることで、単語をベクトルで表現する方法論のことである。単語の前後関係である文脈を手掛かりに学習している。文章をモニターと考え、単語を購買ログ・インターネットアクセスログと考えることで、各ログをベクトルで表現できる。また、モニターごとのログのベクトル平均値を当該モニターのベクトルと定義し、ベクトルの類似度を手がかりに当該モニターに好まれるであろうアイテムを予測することができる。

このようにログの前後関係を加味した手法と、そ

うでない手法を試行することで、精度比較において興味深い結果が期待できると考えた。

#### 4. 学習データの作成とモデルの構築

Word2Vecのモデル構築において、購買ログが同じtimestampである程度まとまって存在するため、インターネットアクセスログとの前後関係を上手く学習できないことが想定される。これを回避するため、購買ログだけで学習したモデル（以下、『モデルA』とする。）と、購買ログを暦日単位でまとめたログ（以下、『暦日購買』とする。）と、インターネットアクセスログと組み合わせたログで学習した（以下、『モデルB』とする。）二つのモデルを構築した。このモデルAとモデルBは、暦日購買によって架橋され、一体的に運用できるものである。詳しくは次節で解説する。

FMについては、モニターID、アイテム、回数のデータを作成した。回数は5を超えるものは5に置

き換えて学習データとした。

なおインターネットアクセスログについて、両手法ともドメインTopページと広告配信ドメインのページのログは除いて使用した。除かれたページは、モニターの興味関心を示唆するものでないと考えられることから、このように処置した。

#### 5. 両分析手法の結果比較

比較結果は以下の手順で導出した（図1）。レコメンドされたアイテムのカテゴリと、学習データに使用した実ログのカテゴリが合致するか否かを正解指標とした。

表1 比較指標

	FM	Word2Vec
合致数	536	557
平均合致率	約7.1%	約7.2%

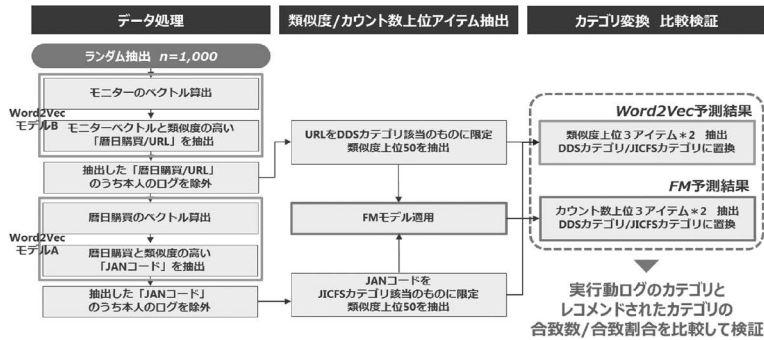


図1 比較結果の導出方法

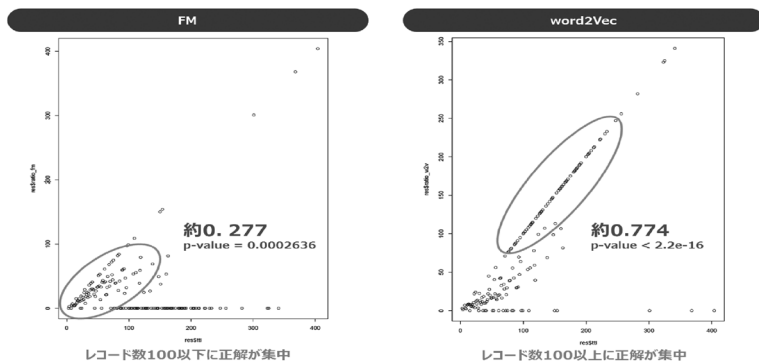


図2

比較指標として定義したカテゴリの合致数、平均合致率において大きな差はなかった（表1）。

## 6. 結果に対する考察と仮説、その検証

FMは5を超えるレコード数を無視して学習している一方で、Word2Vecは前後関係を重要な要素として学習する仕組みである。これらのことからモニターレコード数の多寡と、各手法の正解割合に相関が有るのではないかと、この仮説に至った。

上記仮説を検証するため、レコード数毎の両手法の正解割合をレコード数で重みづけした数値を作成し、その数値と各モニターレコード数との相関係数を算出した（図2）。

するとFMはレコード数100以下のモニターに正解が集中して相関係数が低いのにに対し、Word2Vecはレコード数100を超えるモニターに正解が集中して相関係数が高いことが確認できた。

## 7. まとめ

### 7.1 モデルの評価

モニタレコード数が少なくても一定の精度を

揮できるうえ、一定の属性情報を予測に付加的に取り込めるFMが精度面で優れていると考えた。

一方で、計算速度が速くアイテム同士の類似度を直接に算出できるなど、実用面ではWord2Vecの方が優れていると考えた。

蓄積されているデータ量や分析結果の利用が想定される場面などに応じて、適切に使い分けがされるのが肝要であると結論付けた。

### 7.2 今後の課題

計算処理能力と使用したRのパッケージ仕様の都合上、FMの予測はWord2Vecの結果を前提としたものになっている。そのため純粋に両分析手法の比較ができていない。この点今後の課題としたい。

## 8. 最後に ～弊社の取り組みについて～

弊社は、インターネットを介したアンケート調査サービスを中核に、あらゆるニーズに応える、豊かなリサーチサービスを展開している（図3）。

また直近では、以下の商品に注力している。

### OTSリサーチ / OTSパネル

アンケートを通じて取得したテレビ視聴実態デー

<p><b>インターネットリサーチ</b> <span>定量</span></p> <p>インターネットを利用した定量調査です。調査対象者に対しメールやアプリを通じて調査を依頼し、PCやスマートフォンから回答を求めます。従来の市場調査よりも速く、かつ低コストで定量調査を実施することができます。</p>	<p><b>動画視聴調査</b> <span>定量</span></p> <p>調査対象者が、動画を視聴したうえでアンケートに回答する調査手法です。テレビCMやソーシャルをはじめとした動画コンテンツの調査を、スマートフォンやタブレット等、マルチデバイスに配信が可能です。</p>	<p><b>MROC</b> <span>定性</span></p> <p>「特定の商品・ユーザ」や「同じ生活背景の人」など、共通の要素を持った人がオンライン上のコミュニティで特定のテーマについて自由に議論します。これらを観察することでインタビューよりも深い生活者インサイトを探索することができます。</p>	<p><b>家計調査サービス</b> <span>データベース</span></p> <p>家計簿アプリをインストールしてもらい、レシートスキャン・手入力等によって日々のすべての支出データを把握することが可能です。支出データを元にした追跡調査も可能です。</p>
<p><b>リサーチシステム提供</b> <span>定量</span></p> <p>お客様がご持ちの顧客リストや会員リストを調査対象者とし、マクロレベルのリサーチシステムを使ってネットリサーチが実施できます。顧客満足度調査や会員アンケートなどに最適です。</p>	<p><b>オータメイト調査</b> <span>定量</span></p> <p>お客様ご自身の都合に合わせてオータメイトリサーチを実施します。複雑な条件分岐やWeb上にリアル（朝・雑談）を再現する調査、写真調査、テレビ視聴調査など、あらゆる調査が可能です。</p>	<p><b>動画収集調査</b> <span>定性</span></p> <p>行動観察調査が低コストで手軽に実現できる調査手法です。生活者がWeb上にアップロードした自宅で撮影した動画を専用画面からいつでも閲覧することができます。手軽に全員の生活者の行動を観察することが可能です。</p>	<p><b>ライフスタイル分析</b> <span>データベース</span></p> <p>ブランド・メディア・価値観など200項目におよぶ回答データをデータベース化しているため、生活者のライフスタイルを多角的に把握することが可能です。</p>
<p><b>HUT（ホームユーステスト）</b> <span>定量</span></p> <p>製品サンプルを調査対象者に送付し、実際に使用した感想をリサーチします。調査はネットリサーチで実施しますので、アンケート終了と同時に回答結果を納品します。</p>	<p><b>インタビュー調査</b> <span>定性</span></p> <p>調査対象者に特定の会場に集まってもらい、座談会形式、あるいは1対1でインタビューを行う調査手法です。生活者の態度・行動の背景にある気持ちや要因の関係を把握することができます。</p>	<p><b>デジタルマーケティングリサーチ</b> <span>定量・定性</span></p> <p>Cookie取得済みのパネルの、特定のオンライン広告の接触履歴や行動履歴を継続しており、それらログベースでの追跡調査が可能です。</p>	<p><b>セルフリサーチ</b> <span>定量・定性</span></p> <p>Web上で会員登録を行うだけで、調査会社を経由せずリサーチを実施することができる。セルフリサーチのWebサービスです。</p>
<p><b>CLT（会場調査）</b> <span>定量</span></p> <p>調査対象者に特定の会場に集まってもらい、実際に試作品や製品を「見る」「触れる」「飲食」することで、評価を測定する手法です。実際の店舗や室内に近い環境を再現しての調査が可能です。</p>	<p><b>オンラインインタビュー調査</b> <span>定性</span></p> <p>オンラインで行うインタビュー調査です。調査対象者が自宅から参加できるため、育児や介護で外出の難しい方、地方在住者などへインタビューが可能です。</p>	<p><b>消費者購買履歴データ</b> <span>データベース</span></p> <p>パネルにバーコードスキャナを貸出し、日々の購買データを収集。どこでどのように商品を購入しているかが把握できます。また、購入者へのアンケートにより、購入理由等の把握が可能です。</p>	<p>マーケティングリサーチにはさまざまな種類があり、調査目的に応じて最適な手法を選択することが重要です。マクロレベルはあらゆるニーズにお応えできる、豊富なリサーチサービスを取り揃えています。</p>

図3

タとTV番組/CM出稿データを掛け合わせ、接触状況を特定する調査手法。アドホックOTSリサーチ：60万人（東名阪）への調査が可能。OTSパネル：1万人（東名阪）は毎週テレビ視聴実態データを取得している。

#### Digital Data Safari (DDS)

パーミッション取得済みのCookie保有モニターから、自然なWEB行動のアクセスログを収集。ターゲットユーザが、日頃どのような内容のWEBサイトにアクセスしているのか把握することで、事実ベースで興味・関心を測ることが可能。

#### millBoard

自社モニターをベースに、「リアルタイム」で「デモグラフィック属性別」に「WEB/APP/TV」のリーチを捉えるマクロミル独自開発のダッシュボード。アンケート調査と掛け合わせることで日次/週次でリーチから態度変容までを捉える事が出来る。

#### 参考文献

- [1] 池田裕一，リクルート式自然言語処理技術の適応事例紹介，<https://www.slideshare.net/recruitcojp/ss-56150629>
- [2] 電通，2018年（平成30年）日本の広告費。

#### 略歴

---

##### 齊藤武史（さいとう たけし）

2017年 株式会社マクロミル入社  
入社以来、アクセスログデータの商品化プロジェクトに、アナリストとして従事している。