

## 取り上げる論文について

タイトル： ECサイトにおけるソーシャルログイン会員のサイト内行動分析  
—ゴルフ用品 ECサイトを対象とした会員の特徴と購買行動分析—  
著者： 大竹 恒平、橋本 鴻、生田目 崇  
掲載： 日本ソーシャルデータサイエンス論文誌 第1巻 第1号 2017年3月

※おことわり： 発表時間と僕の理解力の関係で省略した箇所が多々あります。それに関しては原著論文をご参照下さい。

### 概要

仮説： ECサイトにおいて、ソーシャル会員が行うサイト内行動は、一般登録会員が行うそれとは異なっているのではないかと。

結果： 購買に至るセッション時の滞在時間において、ソーシャル会員と一般会員のサイト内行動には差異が見られた。

### 1. はじめに

・近年、ビジネス領域におけるデータ活用が進んでいる。とりわけ ECサイトにおいては、ビッグデータを活用したマーケティング活動が求められている。学術領域においても、マーケティング領域に関する研究が盛んにおこなわれており、中でも消費者の購買行動に着目した研究は産学問わず高い関心を集めている。他方で、近年インターネット上における情報発信・共有ツールとして、ソーシャルメディアが広く利用されている。

→ ソーシャルメディアを利用した広告・マーケティング活動が注目されている。

Ex)

- ・商品の紹介、活用方法、レビューなど、企業の製品・サービスの広告メディアとしての利用。
- ・商品開発・販売戦略の策定や、自社の顧客特性の解明に必要な、ライフログに基づく詳細な行動データの収集場所としての活用。
- ・ソーシャルログインサービス。

#### ソーシャルメディア

ユーザーが発信した情報によってコンテンツを作り出しているメディア。

Ex) 2ch、Youtube、Twitter、facebook 等

#### ソーシャルログイン

アカウント連携機能の一種。ソーシャルメディア上のIDを用いて、企業のECサイトやポータルサイトにログインする機能のこと。「ソーシャル会員」とは、ソーシャルログインへの登録を行った会員のことを指している。

・ソーシャルログインは、前述したように、アカウント連携サービスであり、SNSを定期的に利用している会員が登録していると考えられる。ソーシャルメディアの利用目的として「情報探索」が上位に挙げられることから、ソーシャル会員は、インターネット情報探索能力の高い会員であると考えられる。

→ ソーシャル会員は、ウェブ上での比較・検討、流入経路、閲覧回数といった、ECサイト上で購買に至るまでのサイト内行動に関して、一般会員と異なる特徴を持つことが想定される。

## 2. 研究目的

- ・ EC サイトにおいてソーシャル会員と一般登録会員それぞれが行うサイト内行動の特徴とその差異について評価することを目的とする。

GDO

株式会社ゴルフダイジェストオンライン社

## 3. データ概要

- ・ ゴルフ商品を扱う EC サイト（GDO が運営）における以下の 3 種類のデータを用いる。
  - （1）ID 付き POS データ ・ ・ ・ 商品名、商品金額、購買日などの商品に関するデータ
  - （2）会員登録データ ・ ・ ・ 職業、性別、年齢、ゴルフスコアなどの会員の属性に関するデータ
  - （3）アクセスログデータ ・ ・ ・ GDO の運営するサイトへのアクセス回数、利用デバイス、ページ滞在時間や回遊数などに関するデータ
- ・ 2014 年 7 月 1 日～2015 年 6 月 30 日の 1 年分のデータを対象とする。

## 4. 分析の流れ

**STEP 1** : 全サイト会員の中から、ソーシャル会員を抽出する。

**STEP 2** : 抽出したソーシャル会員の ID 付き POS データに記録された購買商品を用い、該当期間に 1 回以上購買があった会員を分類する。

※ 先行研究の結果を参考に、事前にソーシャル会員を「クラブ、アパレル、用品・小物」の 3 カテゴリに分類し、それぞれのカテゴリに対して判別モデルを作成する。これにより、特性が異なる商品カテゴリごとに、ソーシャル会員の探索行動に対する特徴づけを行うことができる。

**STEP 3** : それぞれのソーシャル会員のアクセスログデータと会員登録データを用いて、判別モデルに利用する説明変数の作成を行う。

(STEP 4 : 決定木分析による、ソーシャル会員と一般会員の判別を目的としたモデルの作成を行う)

**STEP 5** : ソーシャル会員、一般会員の両者において、決定木分析による購買生起の判別を目的としたモデルの作成を行う。

## 5. データセットの作成及び判別モデルに利用する説明変数の作成

**STEP 1** : 2015 年 6 月 30 日時点で Facebook アカウントを用いてソーシャルログイン登録を行っている 55,333 人のソーシャル会員を抽出。

**STEP 2** : 上記のソーシャル会員に関して、購買商品を抽出したところ、7,771 人のソーシャル会員が 35,742 件の購買を行っていた。これらの商品をキーとして、商品カテゴリごとに会員を分類した。その結果、「クラブ」では 2,675 人（総購買件数：5,984 件）、「アパレル」では 4,171 人（総購買件数：14,624 件）、「用品・小物」では 5,113 人（総購買件数：15,134 件）のソーシャル会員が購買を行っていることが分かった。

**STEP 3** : アクセスログデータから、「PV 数」「滞在時間」といった説明変数をセクション毎に集計した。会員登録データから、「年齢」「サイト登録年数」といった説明変数を作成し、会員ごとに集計を行った。以下の分析では、アクセスログデータから作成した説明変数を行動変数、会員登録データから作成した説明変数を会員属性変数と呼ぶ。

**セッション数** : ある会員の EC サイトへの訪問から離脱までを 1 として集計した値。

**PV（ページビュー）数** : 1 セッション数に訪問したページ数を集計した値。

**CV（コンバージョン）数** : 実際に商品の購買に至った回数を集計した値。

## 補足

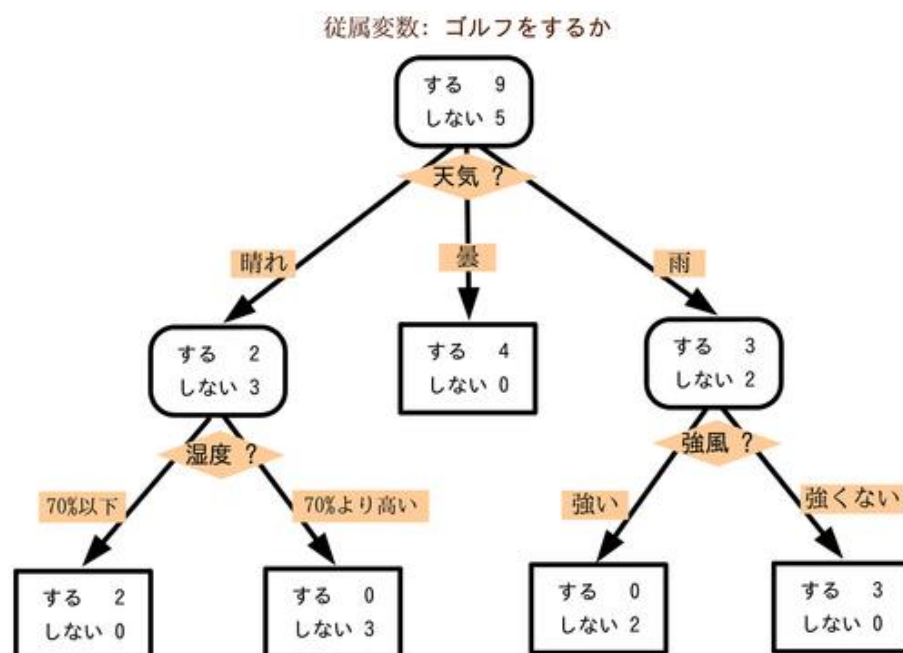
## 決定木分析 について

今年度に照井ゼミで使用する教科書、『Rによるデータマイニング』によると、決定木分析とは、「教師ありデータを用いて説明変数がある基準に従って分割し、樹木状の『決定木』を用いて分類していく手法」だそうです。

これだけでは全く意味が分からないので、Wikipedia に載っていた例を取り上げて説明します。下の二つの図表をご覧ください。上の表はゴルフクラブの経営者が、特定の日の天候と来場者の有無をまとめたものです。そして下の図が、上の表から作成される決定木分析の結果です。

### ゴルフクラブ来場状況

独立変数				従属変数
天気	気温(℃)	湿度(%)	風が強い	ゴルフをするか
晴れ	29	85	強くない	しない
晴れ	27	90	強い	しない
曇	28	78	強くない	する
雨	21	96	強くない	する
雨	20	80	強くない	する
雨	18	70	強い	しない
曇	18	65	強い	する
晴れ	22	95	強くない	しない
晴れ	21	70	強くない	する
雨	24	80	強くない	する
晴れ	24	70	強い	する
曇	22	90	強い	する
曇	27	75	強くない	する
雨	22	80	強い	しない



「ゴルフをするか」という従属変数は、「天気」、「湿度」、「強風」という3つの説明変数で、上図のように説明できます。晴れていてじめじめした日、風の強い雨の日は来場者がおらず、それ以外の日には来場者がいるという調査結果が一目瞭然で分かります。上の表より下の図の方が、変数間の関係が見やすいのが分かります。ちなみに、第一分岐である「天気」が最も説明するのに重要な変数とのことです。

### 参考・引用

Rによるデータマイニング入門 オーム社 山本義郎、藤田友和、久保田貴文 共著

Wikipedia <https://ja.wikipedia.org/wiki/%E6%B1%BA%E5%AE%9A%E6%9C%A8>

## 6. 決定木分析を用いたソーシャル会員の判別モデルの作成

※アクセスログデータと会員登録データを利用して、ソーシャル会員と一般会員の判別を目的とした分析を行っていますが、大して面白くない結果だったので、失礼ながら今回は省略させていただきます。

## 7. ソーシャル会員及び一般会員の購買生起の判別モデル

### (1) 購買生起の判別モデルの作成

表 7 購買生起の判別モデルに用いた説明変数の一覧

	説明変数名		データ形式
基準変数	セッションにおける購買の有無 (購買 1,非購買 0)		2 値
行動変数	外部リファラーサイト	google	2 値
		yahoo	2 値
		bing	2 値
		facebook	2 値
		rakuten	2 値
		Ameblo	2 値
		Kakaku.com	2 値
		amazon	2 値
	内部リファラーサイト	www (トップページ)	2 値
		shop (ショップページ)	2 値
		reserve (ゴルフ場予約ページ)	2 値
		lesson (レッスンページ)	2 値
		news (ニュースページ)	2 値
		usr (ユーザページ)	2 値
		myp(マイページログイン)	2 値
		style (ゴルフスタイルページ)	2 値
		sns (コミュニティページ)	2 値
		faq (ヘルプページ)	2 値
	morning:閲覧時間帯(朝:5:00~10:59)		2 値
	daytime:閲覧時間帯(昼:11:00~16:59)		2 値
	night:閲覧時間帯(夜:17:00~22:59)		2 値
	pv_count(ページビュー数)		連続値
	stay_time(滞在時間)		連続値

リファラー：EC サイトにアクセスする直前に閲覧していたページ。

外部リファラー：GDO 外のサイトからの流入を示すフラグ。

内部リファラー：ゴルフ場予約やゴルフレッスン、ニュース等の GDO のコンテンツからの流入を示すフラグ。

## (2) 購買生起の判別モデルの評価

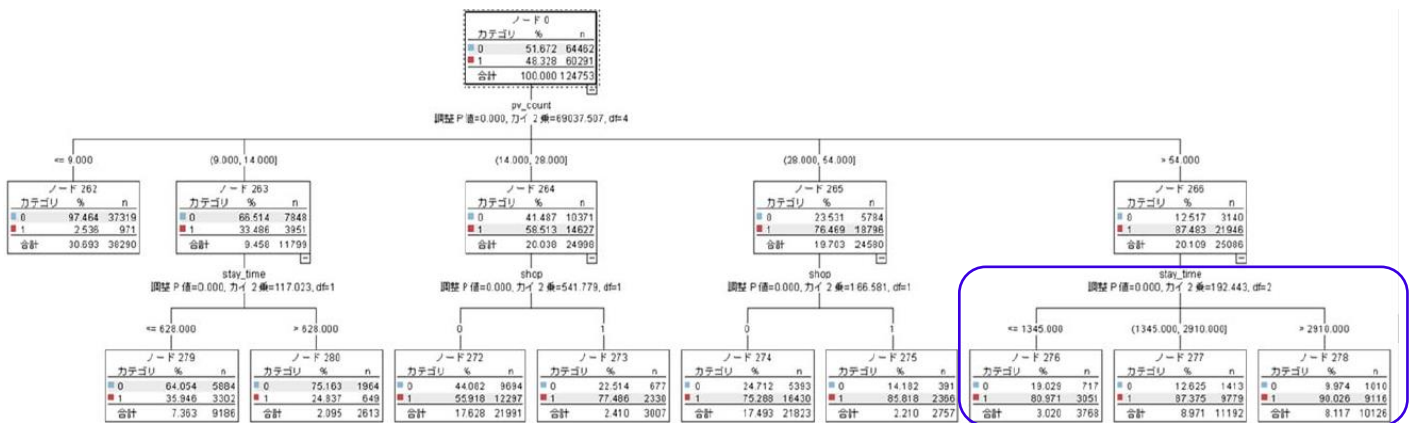


図 2 全データを用いた一般会員の購買判別モデルの結果

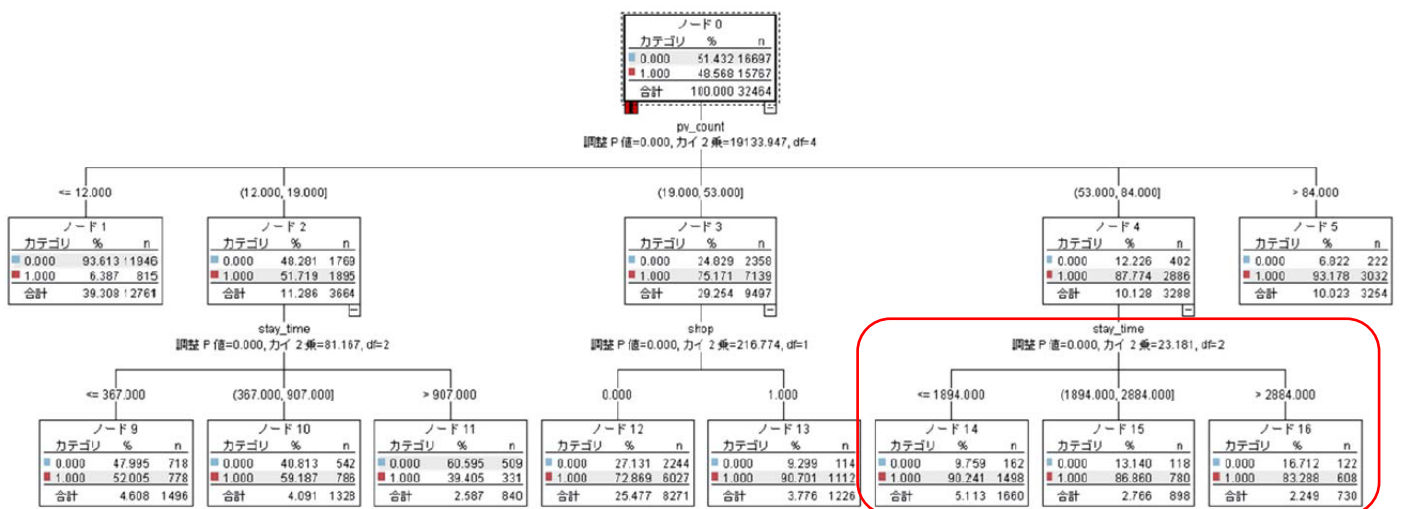


図 1 全データを用いたソーシャル会員の購買判別モデルの結果

## (3) 購買生起の判別モデルの考察

- ・ ソーシャル会員、一般会員ともに、第一分岐においては「PV 数 (pv\_count)」が採用された。PV 数が多くなるほど、購買に至るセッション数が多くなる。ソーシャル会員では、PV 数が 12 回以上、一般会員では 14 回以上になると購買に至るセッション数の割合が 5 割を上回る。
- ・ ソーシャル会員、一般会員ともに、第二分岐においては「滞在時間(stay\_time)」及び「内部リファラー」の「ショッピングページ(shop)」が採用された。一般会員の滞在時間に着目すると (図 2)、ノード 276、277、278 (青い線で囲んだ部分) では、滞在時間が長くなるほど、購買に至るセッションの割合が高くなっていることが分かる。一方で、ソーシャル会員の滞在時間に着目すると (図 1)、ノード 14、15、16 (赤い線で囲んだ部分) では、滞在時間が長くなるほど、購買に至るセッションの割合が低くなっていることが分かる。

→ ソーシャル会員においては、購買に至るセッション数の割合及び実数共に、滞在時間が短いセッション程、購買に至りやすい。

∴ サイト内での購買を考えると、複数の商品間の比較・検討や、商品の詳細情報の閲覧などの購買行動により滞在時間は増加する。一方で、滞在時間が短く購買に至る消費者は、外部サイトで何らかの探索行動を経て購買に至ることが推察される。

(4) 商品カテゴリ（クラブ、アパレル、用品・小物）毎に作成した判別モデルの結果と考察

- ・一般会員の3商品カテゴリにおいては、滞在時間が分岐基準に採用された場合は、滞在時間が長くなるほど、購買に至るセッションの割合が高くなる傾向にあった。
- ・ソーシャル会員の「クラブ」、「用品・小物」の2つのカテゴリにおいては、滞在時間が短くなるほど、購買に至るセッションの割合が高くなる傾向にあった。ただし、「アパレル」に関しては、滞在時間と購買に至るセッションの割合に大きな差は発見されなかった。

∴「アパレル」は好みが反映されやすい製品であり、品揃えも多様である。生地や販売メーカーなどの詳細情報以外にも、色見や会員自身の有するアイテムとのコーディネートなど、個人の感性により検討する視点が異なる。そのため、滞在時間は人によって異なり、そういった個人差をモデルに反映することは困難であったと考えられる。

- ・「用品・小物」に関しては、基本的に安価で価格差が小さい。比較項目も他の商品カテゴリに比べると少ない。従って、一般会員においても、滞在時間は短い。
- ・「クラブ」は高価格帯に分類され、価格差が大きい。加えて、商品詳細情報が細かく記載されているため、検討項目も多くなる。ただし、商品によって規格は定められており、事前にメーカーの Web サイトや他の EC サイトを通じて探索行動を行っていた場合には、容易に検索から購入まで進めることができると思われる。

## 8. 結びに

### まとめ

- ・一般会員の場合は滞在時間が長くなるほど購買に至るセッションの割合が高くなるのに対し、ソーシャル会員の場合は滞在時間が長くなるほど、購買に至るセッションの割合が低くなっている。これらの結果から、ソーシャル会員は外部サイトでの比較・検討を十分に行った上で当該 EC サイトに訪問し、購買のみ行うといった探索行動の一端が推察された。