

ゼクシィnet におけるユーザの行動パターンの分析

User Behavior Analysis on Zexy.net

飯塚 修平*1

Shuheii Iitsuka

濱野 将司*1

Masashi Hamano

川上 和也*2

Kazuya Kawakami

松尾 豊*1

Yutaka Matsuo

萩原 静蔵*3

Seigen Hagiwara

川上 登福*4

Takayoshi Kawakami

浜田 貴之*4

Takayuki Hamada

*1 東京大学大学院工学系研究科技術経営戦略学専攻

Department of Technology Management for Innovation, The University of Tokyo

*2 東京大学工学部システム創成学科知能社会システムコース

Program for Social Innovation, The Department of Engineering, The University of Tokyo

*3 株式会社リクルートマーケティングパートナーズ ブライダル事業本部

Recruit Marketing Partners Co.,Ltd

*4 株式会社 経営共創基盤

Industrial Growth Platform, Inc.

It has become increasingly important for e-commerce platforms to effectively use their huge amount of data. Although various methods have been suggested, many of them have regarded conversion as only one single action, *purchase*. In this paper, we regard it as a two-stage action, comparison and purchase. This idea enables us to propose a method to visualize the competitive relationships of merchandise depending on users' preferences.

1. はじめに

近年のウェブの著しい発展は、企業のみならず一般の人々の情報収集の形態を大きく変化させた。生活にまつわるさまざまな情報がウェブから収集できるようになり、普段は馴染みがない結婚・挙式にまつわる情報の収集にもウェブが大いに活用されるようになってきた。現在では、挙式するカップルのほとんどが結婚式場の比較・検討にウェブを活用している。生活にまつわる情報検索の軸は紙媒体やテレビなどの旧来のメディアから段々とウェブに移ってきており、今後ますますウェブの活用は進んでいくだろう。

一方、ウェブサイト上にはユーザの商品閲覧履歴や予約履歴などのログデータが大量に蓄積されているにも関わらず、それらをウェブサイトの改良やマーケティング手法の改善に十分に生かされている企業は少ない。うまく活用することが出来れば、ユーザの特性から適切な商品を推薦して、コンバージョンを増やすことが出来るはずである。コンバージョンとは商品購入や資料請求など、ウェブサイト運営者がユーザにしてもらうことを望むの行動のことである。

例としてコンビニの販売分析を考えてみよう。ある店舗の商品ごとの販売数のデータには、商品の配置や広告などの店舗上の施策の効果が加味される。こうした影響を排除して、商品そのもののコンバージョンを計測しようとする、そもそも手に取った人の中で、その商品を買った人の割合を計測する必要がある。しかし、実店舗の場合「手に取った」という事象を計測することは事実上不可能である。ところが、ウェブの場合はこの事象を「商品ページの閲覧」というユーザ行動で定義することによって、そもそも閲覧されなかったのか、閲覧されたがコ

ンバージョンが発生しなかったのかを容易に区別することができる。

特に、閲覧されたがコンバージョンが発生しなかったという事象は、明確に「選ばれなかった」ことを意味する。この考え方をさらに推し進め、複数の商品を手に取った上である商品を購入した場合は、購入されなかった残りの商品は購入された商品に「負けた」という考え方を導入することができる。このようにして商品間の勝敗関係を明確に定義することができるため、この勝敗関係を商品閲覧したすべてのユーザについて合算してネットワークとして可視化することによって、商品の勝敗関係のダイナミクスを捉えることができる。

しかも、ユーザごとにこの勝敗ネットワークを構築すれば、ユーザによって商品の勝敗関係が変化する様子を観察することが出来る。全体として見れば人気が高くない商品であったとしても、その中では圧倒的な人気を見せるユーザカテゴリを発見できるかもしれない。これは、様々なウェブサービスのレコメンデーションに応用可能な手法であると考えられる。

本研究では、結婚情報サイト「ゼクシィnet」*1を題材に取り上げ、ウェブサイト上でのユーザの商品閲覧履歴と式場見学予約履歴から結婚式場間の競合関係を勝敗ネットワークとして可視化する手法を提案する。さらに、嗜好に沿って分類されたユーザカテゴリによって勝敗ネットワークの構造が変化することを示す。

2. 従来研究

ウェブマーケティングにおけるログデータの活用において、商品間の関係を知る手法の一つとして、協調フィルタリングを用いたレコメンデーションがある。協調フィルタリングとは、

連絡先: 飯塚修平, 東京大学工学系研究科技術経営戦略学専攻, 東京都文京区弥生 2-11-16, 03-5841-7672, iitsuka@weblab.t.u-tokyo.ac.jp

*1 ゼクシィnet <http://zexy.net>

ユーザの商品に対する評価をもとにユーザ - 商品の「格付け行列」を構築し、既に得た格付けから未知の格付けを推定して推定値が高い商品を推薦する手法である。商品自体の特徴や性質を把握する必要がないため、多様なコンテンツに応用することができるのが利点である [山下 09]。しかし、ウェブサイトの商品閲覧データのように暗黙的に収集されたデータは膨大になるため、格付け行列が巨大化し非常に大きな計算リソースを必要とする [Zhao 10]。これに対して、特異値分解を組み合わせることで格付け行列のランクを落として商品の推薦を行う手法が提案されている [井上 01]。

また、商品間の関係を可視化する試みとして、同一の消費者に購入された商品同士をつなげることで商品ネットワークを構築し、商品ネットワークが成長していく様子をモデル化したものがある [伊藤 07]。しかしながら従来の研究の多くは、商品のコンバージョンを購入という行為のみで捉えていた。本研究のように、コンバージョンを「検討」と「購入」という2段階に分け、商品を「そもそも検討すらされなかった商品」、「検討されたが購入には至らなかった商品」、「検討されて購入に至った商品」という3種類に区別して商品間の関係を分析するという視点は新しい。

3. 提案手法

本研究では、まず式場の共起行列の特異値分解によって各式場の特徴ベクトルを算出する。そして、あるユーザの嗜好をそのユーザが閲覧した式場の特異ベクトルの足し合わせによって推定する。今回はブライダルフェア*2予約（以後BF予約とする）を目的のコンバージョンとし、各ユーザの式場閲覧履歴の中でBF予約があった式場を勝者、予約のなかった式場を敗者と定義する。これらの式場間の勝敗関係を表した勝敗ネットワークを生成し、ユーザの嗜好によってそのネットワークの構造が変わることを示す。

3.1 ユーザの分類

ユーザをカテゴリ分類する方法はさまざまなものが考えられるが、ここでは、ユーザの閲覧履歴をベクトルとして捉え、ユーザの嗜好に応じて分類する方法を用いる。今回は同一ユーザによって閲覧されることを「共起」と定義し、その共起回数を要素に持つ式場 - 式場の共起行列を構築する。式場数が n の場合、これは $n \times n$ の対称行列 M になる。この行列を LSI (Latent Semantic Indexing) の手法を用いて分解する。

$$M \sim M_k = V_k \Sigma_k V_k^T \quad (1)$$

ここで得られる $n \times k$ 行列 V_k の各行、 k 次元ベクトル \mathbf{v} は特異ベクトルの $k+1$ 次元以降を削減したものであり、各式場の特徴を表している。ここでは \mathbf{v} を各式場の特徴ベクトルと名付ける。

ユーザが閲覧している式場の特徴が、そのユーザ自身の嗜好を表していると仮定すると、ユーザ u の嗜好 \mathbf{t}_u は、閲覧した各式場 $p \in P_u$ の特徴ベクトル \mathbf{v}_p の足しあわせを正規化したものによって表現することができる。すなわち、

$$\mathbf{t}_u = \frac{\sum_{p \in P_u} \mathbf{v}_p}{|\sum_{p \in P_u} \mathbf{v}_p|} \quad (2)$$

ここで、 \mathbf{t}_u の第 a 成分 ($1 \leq a \leq k$) を t_{ua} と表すと、 $|t_{ua}|$ がユーザ u の特徴 a への関心の大きさを表している。このと

*2 結婚式場の下見をすることができるイベントのこと。模擬挙式や模擬披露宴を見学したり、料理の試食をしたりすることができる。

き、各ユーザ u が最大の関心を払う特徴 s は

$$s = \arg \max_{1 \leq a \leq k} |t_{ua}| \quad (3)$$

と表される。この場合はユーザ u をユーザカテゴリ U_s に分類する。このようにして、全ユーザ $u \in U$ を k 種類のユーザカテゴリ $\{U_1, \dots, U_k\} \subset U$ に分類する。

3.2 式場勝敗ネットワークの構築

ゼクシィnet では、あらゆるユーザが自分の嗜好にあった結婚式場を探すために、ウェブサイト内の複数の式場を閲覧している。その中から、実際に式場に足を運んでブライダルフェアに参加したいと思う式場を選択し、ゼクシィnet 上でBF予約を完了する。この際に予約される式場はひとつとは限らず、複数の式場が同時に予約されることもある。

このとき、下記のルールに基づいて式場間の勝敗関係を定義することができる。

- 同一ユーザによって閲覧された式場のなかで、BF予約が行われた式場は「勝者」、行われなかった式場は「敗者」とラベル付けされる。
- 閲覧された順番には関係なく、すべての勝者式場とすべての敗者式場の間に勝敗関係が結ばれる。
- 勝者同士、敗者同士には勝敗関係は結ばれない。

このルールに基づいて、商品の勝敗ネットワークを構築する。

まずは、あるユーザカテゴリに属するユーザ $u \in U_a$ に閲覧された式場の集合 P_u を、BF予約された勝者式場集合 P_w とBF予約されなかった敗者式場集合 P_l のふたつに分ける。各式場をノードと見立てて、敗者式場集合に属するすべての式場 $\forall p_l \in P_l$ から勝者式場集合に属するすべての式場 $\forall p_w \in P_w$ にむけて勝敗関係を表す有向エッジを引く (図1参照)。これをユーザカテゴリに属するすべてのユーザについて繰り返すことにより、あるユーザカテゴリ U_a における式場勝敗ネットワーク G_a を構築することができる。ある式場間で複数回勝敗関係が結ばれた場合は、負け越した式場から勝ち越した式場にエッジが結ばれ、その重みは勝ち越し数である。たとえば、式場 A が式場 B に対して 8 勝 2 敗の成績を取った場合は、式場 B から式場 A に重み 6 の有向エッジが結ばれる。

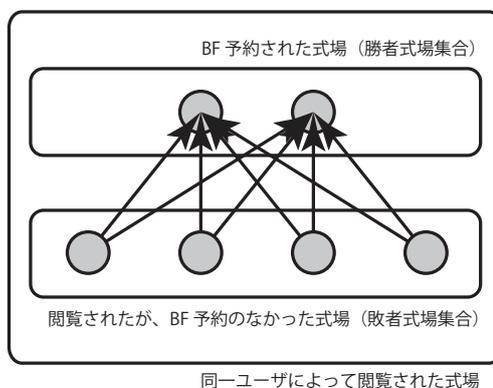


図 1: 式場勝敗ネットワークの構築方法

エッジの重みとしては、勝ち越した数の他にも勝率やその統計的偏りなど、さまざまな指標を用いることができるが、ここでは簡単に勝ち越し数を用いている。

4. 解析

4.1 データセット

式場共起行列の構築に用いた式場は、首都圏版に登録されている 418 式場を対象とした。ある期間に観測された、BF 予約が含まれている 3,187 のセッションを解析の対象とした。なお、今回は簡易的にセッション=ユーザとして解析を進める。また、ユーザカテゴリ数は $k=7$ とした。

4.2 解析結果

式場共起行列の構築と式場特徴ベクトルの算出を行い、ユーザの分類を行った結果、表 1 のように各ユーザが分類された。ユーザを分類後、各ユーザカテゴリについて式場勝敗ネット

表 1: 各ユーザカテゴリに含まれるユーザ数

カテゴリ	ユーザ数
1	1001
2	708
3	161
4	680
5	180
6	125
7	332

ワークを構築した。図 2 はユーザカテゴリ 1 について構築した式場勝敗ネットワークである。ノードの大きさはその会場の入次数、すなわちその会場の強さを表している。エッジの太さはそのエッジの重み、すなわち勝ち越し数を表している。

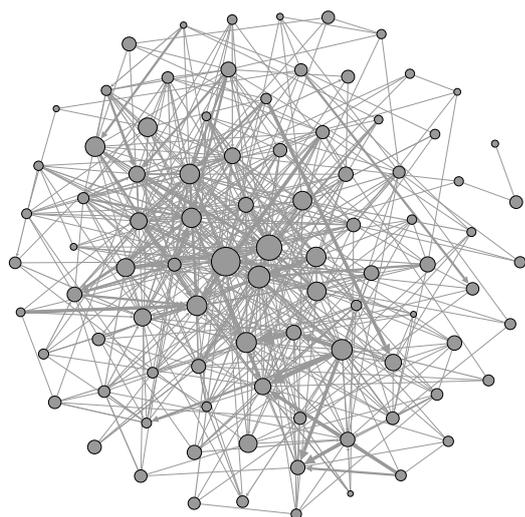


図 2: ユーザカテゴリ 1 の式場勝敗ネットワーク

各ユーザカテゴリごとに構築されたネットワークを比較してみると、式場間の競合関係が変化していることがわかる。図 3 はユーザカテゴリ 1 の式場勝敗ネットワークの、図 4 はユーザカテゴリ 3 の式場勝敗ネットワークの一部分である。両方の図に見られる都内の和風式場 A と同じく都内の和風式場 B の両者に着目すると、ユーザカテゴリ 1 では式場 A が優位に立っていたにも関わらず、ユーザカテゴリ 2 では式場 B が優位に立っており、ユーザカテゴリ間で勝敗関係が逆転している

ことがわかる。また、前者には人気の高い東京都内のホテル D がこれらの勝敗関係に大きく関わっているにも関わらず、後者には現れていない。代わりに都内の和風式場 E が重要なプレイヤーとして関わっている。ネットワーク全体を見ても、前者は式場 A などの和風の由緒ある式場以外にもホテル、ゲストハウスや大聖堂など多様なスタイルの人気式場が含まれているが、後者はその多くを和風の式場が占めている。このことから、ユーザカテゴリの違いが勝敗ネットワークに現れる式場の違いとして現れていることが分かる。

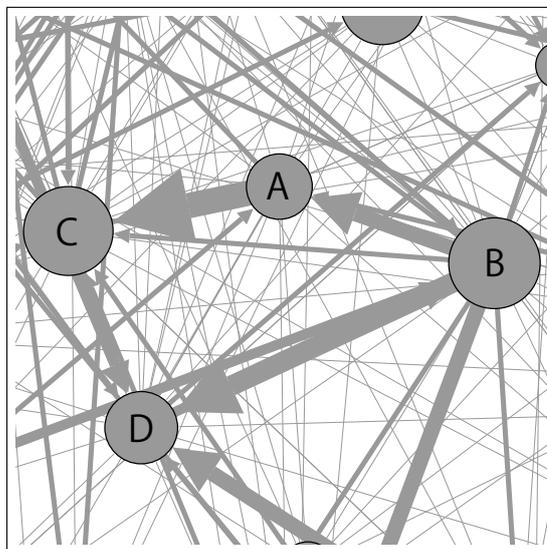


図 3: ユーザカテゴリ 1 に見られる競合関係

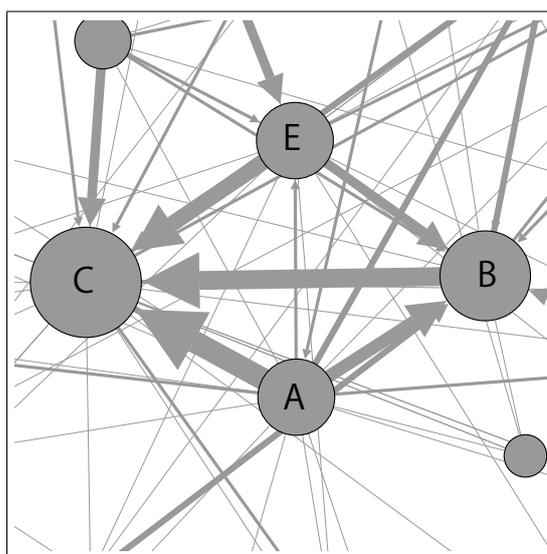


図 4: ユーザカテゴリ 3 に見られる競合関係

5. まとめ

本研究では、結婚情報サイト「ゼクシィnet」を題材に取り上げ、ログデータから得られた商品の共起行列の特異値分解によるユーザの嗜好の推定、およびコンバージョンを元にした商

品の勝敗ネットワークの構築手法を提案した。本研究の新規性は、以下の3点を組み合わせたとこにある。

1. コンバージョンに加えて、閲覧の有無という指標を加える事で、「勝敗」という概念を定義した。
2. 上記の勝敗関係をネットワークとして可視化した。
3. 閲覧という行為に基づいてユーザを分類した上で、ユーザカテゴリごとのネットワークの変化に着目した。

ユーザはウェブサイト上で最後に閲覧した商品でコンバージョンするとは限らず、多数の商品を見て検討した後に、結局最初に見た商品でコンバージョンするという事も往々にしてある。ユーザの遷移のみに着目した遷移ネットワークではこのことを見落とす危険性がある。そのため、今回はあえて時間軸に着目せず閲覧とコンバージョンのみに着目してネットワークを構築した。こうすることで、よく閲覧はされているがコンバージョンに繋がっていない要改善商品の発見にもつなげることができる。

また、このネットワークをユーザカテゴリ毎に構築することによって、全体として見ると人気がなくとも特定のカテゴリ、ユーザ層においては多くのコンバージョンを集めることができる商品の発見にもつなげることができる。

本研究の手法は、協調フィルタリングと同様に商品に関する知識を必要としない。そのため、Amazon.comの閲覧と購入の関係の分析やFacebook.comの閲覧と友達申請の関係の分析など、2つのアクションがあるサイトの分析において一般的に応用可能なものである。今後ユーザの行動履歴から得られるビッグデータの活用が重要視されるなか、今回提案したような嗜好推定、商品ネットワーク構築の試みはますます増えていくであろう。

参考文献

- [Zhao 10] Z.-D. Z. Z.-D. and M.-S. S. M.-S.: User-Based Collaborative-Filtering Recommendation Algorithms on Hadoop, *Knowledge Discovery and Data Mining 2010 WKDD 10 Third International Conference on*, pp. 478–481 (2010)
- [伊藤 07] 論志, 崇: 商品ネットワークの成長モデル: 市場の秩序形成の探究に向けて (セッション 3), 情報処理学会研究報告. MPS, 数理モデル化と問題解決研究報告, Vol. 2007, No.43, pp. 41–44 (2007)
- [井上 01] 光平, 喜一: データ行列の特異値分解に基づく協調フィルタリング, 電子情報通信学会論文誌. D-I, 情報・システム, I-情報処理, Vol.84, No.1, pp. 116–119 (2001)
- [山下 09] 晃弘, 秀憲, 恵二, 東: ユーザ間・アイテム間協調フィルタリングの適応的な融合手法: MovieLens での格付けデータに基づく検証 (社会システムと知能), 情報処理学会研究報告. ICS, [知能と複雑系], Vol. 2009, No.16, pp. 105–110 (2009)