

# 潜在状態ネットワークに基づくソーシャルゲームユーザの行動抽出

Analysis of user behaviors using hidden Markov model

河津 裕貴<sup>\*1</sup>  
Hirotaka Kawazu

鳥海 不二夫<sup>\*1</sup>  
Fujio Toriumi

高野 雅典<sup>\*2</sup>  
Masanori Takano

和田 計也<sup>\*2</sup>  
Kazuya Wada

福田 一郎<sup>\*2</sup>  
Ichiro Fukuda

<sup>\*1</sup> 東京大学  
The University of Tokyo

<sup>\*2</sup> 株式会社サイバーエージェント  
CyberAgent, Inc.

Social network game, the online games provided on the Social Networking Sites, used by a wide range of people because of the popularity of mobile phones. Therefore, it is considered that there is a variety of play styles. In this study we propose a method to discover play styles of social network game users. Firstly, we extract abstract states of user's actions with hidden Markov model. Then, we build a state network in which nodes are abstract states and edges are transition probabilities. We also apply network clustering to extract nodes communities which are considered to be play styles.

## 1. はじめに

### 1.1 背景

ソーシャルゲームとは主にソーシャルネットワーキングサービス (SNS) 上で提供されるオンラインゲームである。ソーシャルゲームは多くの人々が所有するモバイル端末で利用できるため、ユーザ層は幅広い。また従来のパッケージゲームと違い、ゲームの内容は常に追加や更新が可能であり、それに起因し明確なゴールが設定されていないことが多いため [9]、ユーザの楽しみ方は多岐にわたると考えられる。

ユーザの楽しみ方を把握することはサービスの品質向上やより面白いゲームの開発に役立つと考えられる。またゲームの楽しみ方に関する知見はゲーミフィケーションに応用することができる。ゲーミフィケーションの基本的な着想は、ゲームの持つ人間のモチベーションを刺激し高める仕組みを、様々な課題解決に応用するというものであり [10]、様々なサービス [3] や教育 [4] への応用が模索されている。ユーザがゲームに感じる面白さの適切な分析ができれば、より効率的なゲーミフィケーション施策へ貢献できると期待される。

本研究では、実際のゲームログの分析に基づき、ソーシャルゲームにおけるユーザの楽しみ方を検出する手法を検討する。まず隠れマルコフモデルを用いて、ユーザ群の行動系列から抽象的な状態を抽出する。そしてこれらの遷移確率を基に状態遷移ネットワークを構築し、ユーザの滞在しやすい状態群を抽出することで、ユーザの楽しみ方を検出する。

### 1.2 関連研究

Bartle らは MUD (Multi-User Dungeon) と呼ばれるテキストベースのオンラインゲームについて、ユーザのゲームの面白さに関するディスカッションを観察することで、ゲームプレイヤーは「Achiever」、「Explorer」、「Socializer」、「Killer」の4つに大別できると報告している [1]。また Yee はユーザにゲームの面白さに関するアンケートを実施し、アンケート結果を基にユーザのプレイスタイルを検出した [6]。

以上のように、ゲームの楽しみ方に関する従来の研究はユーザへの意見調査によるものであった。しかし、意見調査の手法

では、調査対象のユーザが調査に答えるユーザのみに限られてしまい、データの偏りが懸念される。そこで我々は実際のゲームログの分析に基づいたユーザの楽しみ方を検出する手法の構築を目指す。

## 2. データ

本稿では表 1 のデータを対象に分析した。本研究を行うにあたって、対象サービスを運営する株式会社サイバーエージェント<sup>\*1</sup> から学術的な目的でデータの提供を受けた。

本稿ではユーザの行動系列としてアクセスログ系列を使用した。アクセスログとはユーザがゲーム上のページを閲覧する度に出力されるログであり、ゲーム上のユーザの行動のほとんどはページ遷移を伴うこと、及び比較的サービス固有のデータにならないことから選択した。提供を受けたアクセスログは匿名化されたユーザ ID、閲覧ページの URL 及びタイムスタンプから構成される。ユーザ毎に、期間中のアクセスログを時系列順に一つに繋げた物を行動系列と定義する。

アクセスログはページを開いた場合のみ出力するため、現状ではユーザがゲームを閉じたことを直接判別できるデータはない。そこでユーザがゲームから離れている状態を表現するためにダミー変数を導入した。ユーザが一定時間アクセスしなかった場合、その間隔に応じたダミー変数を挿入した。ダミー変数の種類とそれぞれが表す間隔を表 2 に示す。

訓練データ量の観点から、期間中 5 日間以上ログインしたユーザからランダムサンプリングした 10000 人のデータを使用した。

また、データ処理の簡便さの観点から、ページ名を 1 階層に集約した。例えば、[/card/A] という URL のページと [/card/B] という URL のページは同じ [card] というページとして扱った。

表 1: 使用データ

サービス名	ガールフレンド (仮)
種類	カードバトル系ソーシャルゲーム
期間	2013/3/20 - 2013/5/19

連絡先: 河津裕貴, 東京大学工学系研究科システム創成学専攻, 〒113-8656 東京都文京区本郷 7-3-1 工学部 8 号館 526, TEL: 03-5841-6991, E-mail: kawazu@crimson.q.t.u-tokyo.ac.jp

\*1 <https://www.cyberagent.co.jp>

表 2: ダミー変数

ダミー変数	表す間隔
DUM-L	48 時間以上
DUM-M	8 時間以上
DUM-S	2 時間以上

### 3. 手法

#### 3.1 隠れマルコフモデル

ユーザの行動系列はその時々楽しみ方に基づき形成されると考えられる。そこで、まず隠れマルコフモデル [8] を用いて、ユーザの行動系列から抽象的な状態の系列を抽出する。

隠れマルコフモデルとは系列データを扱う状態空間モデルの一種である。直接観測されない離散的な潜在変数がマルコフ過程を構成しており、観測される変数は各時刻の潜在変数に対応する出力確率分布に従って出力される出力変数であると仮定して、データの生成モデルを構築する。時刻  $n$  における潜在変数を  $z_n$  とすると、一次マルコフ連鎖を構成している仮定から、

$$p(z_n | z_1, z_2, \dots, z_{n-1}) = p(z_n | z_{n-1})$$

となる。出力変数を  $x_n$  とすると、時刻  $N$  までにおける潜在変数と出力変数の同時確率  $Q$  は以下で表される。

$$\begin{aligned} Q(X, Z) &= p(x_1, x_2, \dots, x_N, z_1, z_2, \dots, z_N) \\ &= p(z_1) \left[ \prod_{n=1}^N p(z_n | z_{n-1}) \right] \prod_{n=1}^N p(x_n | z_n) \end{aligned}$$

上式から、同時確率を記述するパラメータには、以下の式で表される初期確率  $\pi$ 、各時刻における潜在変数の遷移確率  $A$ 、出力変数の条件つき確率  $\Sigma$  があることが分かる ( $i, j = 1, 2, \dots, L$ ,  $k = 1, 2, \dots, K$ )。

$$\begin{aligned} \pi_i &= p(z_i) \\ A_{ij} &= p(z_j | z_i) \\ \Sigma_{ik} &= p(x_k | z_i) \end{aligned} \quad (1)$$

隠れマルコフモデルでは Baum-Welch アルゴリズム [2] を用いて、同時確率  $Q$  を最大化するパラメータ集合  $\theta^*$  を算出する。

$$\begin{aligned} \theta &= \pi, A, \Sigma \\ \theta^* &= \arg \max Q(X, Z | \theta) \end{aligned}$$

本稿では 1000 ユーザの行動系列、つまり 1000 個の観測データを基に 1 つの隠れマルコフモデルを構築した。出力変数  $x_k$  はページ  $k$  を表す離散変数となる。得られる潜在変数  $z_i$  は、以下の式で表される出力確率分布  $\Sigma_i$  によって特徴づけられる。

$$\Sigma_i = (p(x_1 | z_i), p(x_2 | z_i), \dots, p(x_K | z_i))$$

各々にページの出力確率分布を持つ潜在変数は、ユーザの様々な行動に対応する抽象的な状態と考えられる。本稿ではこれを潜在状態と呼ぶことにする。また潜在状態の数は 100 個とした。

#### 3.2 Walk Trap

隠れマルコフモデルにより抽出された潜在状態はマルコフ過程を構成しており、ユーザの滞在しやすい状態群は長いタイ

ムスケールにおけるユーザの楽しみ方であると考えられる。まず潜在状態をノードとし潜在状態間の遷移確率をエッジとした状態遷移ネットワークを構築する。このネットワーク上でエッジが密になっている個所はその間で遷移が起こりやすいことを示すため、ユーザが滞在しやすい状態群だと考えられる。そこで滞在しやすい状態群を抽出するためにコミュニティ分割を行う。

本稿では、有向重み付きネットワークのコミュニティ分割手法として walk trap 法 [7] を用いる。walk trap 法とはネットワーク上で行うランダムウォークに基づきノード間の距離を定義し、距離に基づきクラスタリングすることでノードを分類する手法である。ノード  $i$  からランダムウォークを開始し、 $t$  ステップ歩いた際にノード  $j$  にいる確率は、数式 1 の遷移確率行列  $A_{ij}^t$  を用いて  $A_{ij}^t$  で表される。ノード  $i$  とノード  $j$  が同じコミュニティに属する場合、同じコミュニティのノード  $k$  に対して  $P_{ik}^t$  と  $P_{jk}^t$  が比較的近くなる傾向があることを利用し、ノード  $i, j$  間の距離  $r_{ij}$  を

$$r_{ij} = \sqrt{\sum_{k=1}^n \frac{(P_{ik}^t - P_{jk}^t)^2}{d(k)}}$$

と定義する。但し  $d(k)$  はノード  $k$  の入次数を表す。距離  $r$  を基にノードの階層型クラスタリングを行い、得られるクラスタをコミュニティとする。本稿ではランダムウォークのステップ数は 3 とし、コミュニティ数は 5 とした。

### 4. 結果

#### 4.1 潜在状態のクラスタリング

潜在状態の概要を把握するため、潜在状態の分類を行う。出力確率分布を特徴量とし、k-means 法 [5] を用いて 10 個のクラスタにクラスタリングした。表 3 にクラスタの特徴として、各クラスタの中心点において出力確率の高いページ 5 つとその確率を載せた。ただし中心点の出力確率分布とは、クラスタに属する潜在状態の出力確率分布の平均をとったものとする。また各ページの詳細な説明や、ソーシャルゲームの用語の説明は付録 A に示す。

表 3 より、クラスタ C0 は [battle] の出力確率が高く、主にソーシャルゲームの基本要素である他ユーザとのバトルを遊ぶ状態群と考えられる。クラスタ C2 は [raid]、クラスタ C9 は [marathon] が高い確率で出力されている。主にソーシャルゲームで頻繁に開催されるレイドイベントやマラソンイベントを遊ぶ状態群と考えられる。クラスタ C3 はアイテムを受け取るページが最も高くなっているが、これは受け取り待ちアイテムが多数あると目的のアイテムを探す際に何回かページ遷移をするため、その行動系列を抽出した潜在状態だと考えられる。クラスタ C4 はガチャのページや手に入れたアイテムを受け取るページがあるため、新しいカードを手に入れて楽しむ状態群と考えられる。クラスタ C6, C8 はカードを一覧するページやカードを育てるページ、カードを受け取るページが見られ、カードを強くして楽しむ状態群と考えられる。C7 は [DUM-S], [DUM-M], [DUM-L] などが出力されており、ユーザがゲームから離れる状態を表す状態群と考えられる。一方クラスタ C1 は最も高い確率のページでも 0.12 ほどしかない。これは様々な種類の潜在状態が集まってしまっているクラスタであると考えられる。また多くのクラスタで [mypage] が上位に出力されている。これは、ユーザは様々な場面で自分のステータス

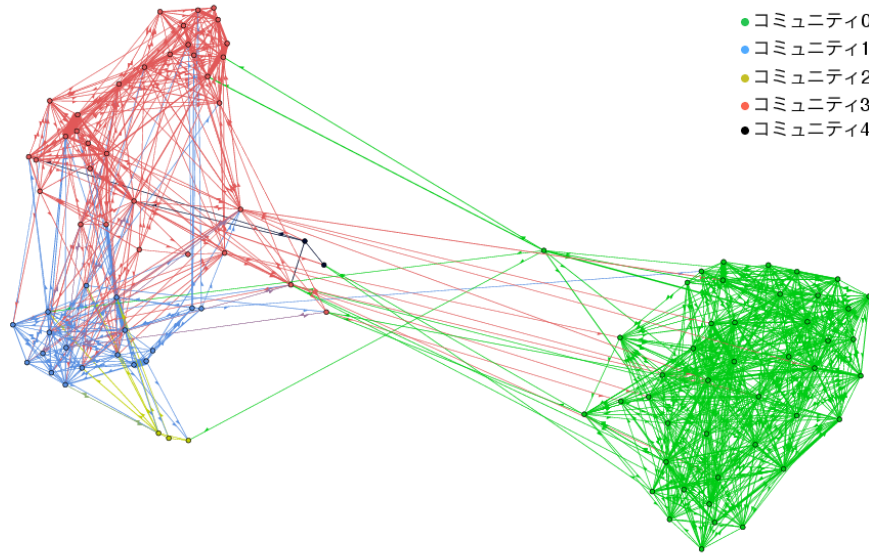


図 1: 状態遷移ネットワーク

表 3: 各クラスタの中心点において出力確率の高い 5 ページ

id	数					
C0	6	battle	mypage	marathon	grade	treasure
		0.7032	0.1036	0.0949	0.0206	0.0122
C1	28	mypage	job	quest	card	club
		0.1169	0.0956	0.0752	0.0747	0.0702
C2	17	raid	mypage	""	giftbox	top
		0.9122	0.0678	0.0042	0.0032	0.0013
C3	6	giftbox	raid	mypage	upgrade	marathon
		0.9722	0.0088	0.0066	0.0063	0.0009
C4	8	cupid	giftbox	mypage	upgrade	landing
		0.8728	0.0517	0.0307	0.0160	0.0089
C5	18	mypage	job	""	quest	giftbox
		0.7484	0.0508	0.0328	0.0203	0.0183
C6	3	evolution	mypage	card	upgrade	giftbox
		0.9737	0.0080	0.0041	0.0039	0.0034
C7	7	DUM-S	DUM-M	DUM-L	mypage	raid
		0.4411	0.3752	0.0517	0.0412	0.0223
C8	5	upgrade	mypage	evolution	card	cupid
		0.8909	0.0600	0.0234	0.0077	0.0057
C9	2	marathon	battle	mypage	giftbox	upgrade
		0.9425	0.0158	0.0155	0.0059	0.0049

を確認したり、他ページへ遷移する際のハブとして [mypage] を使っているからだと考えられる。

#### 4.2 コミュニティ分割

隠れマルコフモデルにより得られた状態遷移ネットワークを walk trap 法によりコミュニティ分割する。潜在状態が属するコミュニティに基づき色分けしたネットワークを図 1 に示す。

各コミュニティがどのような特徴をもつ状態群なのか調査するため、各コミュニティのノード数を前節で得られたクラスタ別に集計し、表 4 に示す。同表より、クラスタ C2 のノードは全てコミュニティ 0 に属しており、コミュニティ 0 はレイドイベント時に滞在する状態群だと考えられる。同様にクラスタ C9 のノードは全てコミュニティ 3 に属しており、コミュニティ

3 はマラソンイベント時に滞在する状態群だと考えられる。またコミュニティ 2 に属するノードは 3 つとなっており、各潜在状態が出力するページを確認したところ表 5 のようになっていた。各ノードは [clubcup] を出力しており、クラブカップイベント時に滞在する状態群だと考えられる。一方コミュニティ 1、コミュニティ 4 にはクラスタ C2、C9 のノードは属しておらず、イベントの開催されていない期間やイベントを遊ばないユーザが滞在しやすい状態群だと考えられる。

以上のように、得られたコミュニティは各イベント期間もしくはイベント期間外にユーザが滞在しやすい状態群と考えられる。対象サービスにおいて各種イベントは期間限定でかつ順々に開催されるため、隠れマルコフモデル学習時に時間軸の影響を強く受けすぎたためであると考えられる。このコミュニティはユーザの長いタイムスケールの行動に対応しているものの、楽しみ方ではなく遊んでいる時期を表してしまっているため、楽しみ方を抽出できるよう改善が必要であると考えられる。

#### 5. 結論

本稿では、ソーシャルゲームにおいてユーザがどのような楽しみ方をしているのか、ユーザの行動ログに基づいて定量的に検出する手法の検討を行った。まず隠れマルコフモデルを用いて、ユーザの行動を潜在状態として抽象的に抽出した。どのような潜在状態があるか傾向を調査し、ソーシャルゲームのイベントや基本的要素を遊んでいる状態が抽出されていることを確

表 4: コミュニティ及びクラスタの所属ノード数マトリクス

	C0	C1	C2	C3	C4	C5	C6	C7	C8	C9
コミュニティ 0	1	8	17	3	2	7	1	3	2	0
コミュニティ 1	2	7	0	0	2	5	0	2	0	0
コミュニティ 2	0	2	0	0	0	1	0	0	0	0
コミュニティ 3	3	11	0	2	4	5	2	2	2	2
コミュニティ 4	0	0	0	1	0	0	0	0	1	0

表 5: コミュニティ 2 に属する各潜在状態において出力確率の高い 5 ページ

id					
0	club	mypage	clubcup	job	DUM-S
	0.2074	0.1862	0.1824	0.0713	0.0653
1	mypage	clubcup	club	DUM-S	giftbox
	0.7341	0.1363	0.0939	0.0133	0.0048
2	clubcup	mypage	giftbox	club	item
	0.08891	0.03768	0.0332	0.0162	0.0084

認した。次に状態遷移ネットワークを構築し、コミュニティ分割を行うことで長いタイムスケールにおけるユーザの楽しみ方の抽出を試みた。抽出されたコミュニティは期間限定コンテンツであるイベントの影響を強く受けており、各種イベント期間にユーザが滞在する状態群であることを確認した。

今後の展開としては、期間限定のコンテンツに影響を受けないように、各種イベント期間毎にデータ区分を分け実験を行うことが挙げられる。また、ユーザに焦点をあてて、各ユーザのネットワーク上の移動系列を利用して楽しみ方を分類することが考えられる。

## 参考文献

- [1] Bartle and Richard. Hearts, clubs, diamonds, spades: Players who suit muds. *Journal of MUD research*, Vol. 1.1, p. 19, 1996.
- [2] Baum L. E., Petrie T., Soules G., and Weiss N. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, pp. 164–171, 1970.
- [3] Zichermann G. and Cunningham C. *Gamification by design: Implementing game mechanics in web and mobile apps*. O’Reilly Media, Inc., 2011.
- [4] Huizenga J., Admiraal W., Akkerman S., and Dam G. T. Mobile game based learning in secondary education: engagement, motivation and learning in a mobile city game. *Journal of Computer Assisted Learning*, Vol. 25(4), pp. 332–344, 2009.
- [5] MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1967.
- [6] Yee Nick. Motivations for play in online games. *CyberPsychology behavior*, Vol. 9.6, pp. 772–775, 2006.
- [7] Pons, Pascal, and Matthieu Latapy. Computing communities in large networks using random walks. *Computer and Information Sciences-ISCIS*, pp. 284–293, 2005.
- [8] Rabiner and Lawrence R. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, Vol. 77.2, pp. 257–286, 1989.
- [9] まつもとあつし. ソーシャルゲームのすごい仕組み. アスキー・メディアワークス, 2012.

- [10] 神馬豪, 石田宏実, 木下裕司. ゲームフィクション. 大和出版, 2012.

## A URL の説明

本稿の分析対象サービスにおける、主な URL が示すページの説明を表 6 に記す。

表 6: URL の示すページの説明

url	ページ
mypage	ハブページ。主に自分のステータスや様々なページへのリンクが表示される。
top	ゲームのトップページ。主にゲームを開始した際に開かれる。
””	[top] のエイリアス。
battle	バトルページ。主な目的は、他ユーザとカードバトルを行うこと。
quest	クエストページ。主な目的は、「クエスト」と呼ばれるミッションをクリアすること。
club	ギルドページ。主な目的は、ギルドのステータスを確認したり、同じギルドのユーザと交流すること。
raid	レイドイベントページ。レイドイベントとは、出現する敵を倒すことなどで得られるイベントポイントを獲得し、他ユーザとイベントポイントを競うコンテンツのこと。イベントは期間限定で開催される。
marathon	マラソンページ。マラソンイベントとは、特別に用意されたミッションをクリアすることなどで得られるイベントポイントを獲得し、他ユーザとイベントポイントを競うコンテンツのこと。イベントは期間限定で開催される。
clubcup	ギルド対抗イベントページ。主な目的は、期間限定で開催させるギルド対抗イベントに参加し、ギルドとしてたくさんのイベントポイントを獲得すること。イベントは期間限定で開催される。
giftbox	ギフトボックスページ。主な目的は、イベント等で獲得が確定したアイテムを実際に受け取る手続きをすること。
upgrade	カード強化ページ。主な目的は、自分の持っているカードを「強化」し、より強くすること。
evolution	カード進化ページ。主な目的は、自分もっているカードを「進化」し、より強くすること。
card	カード一覧ページ。主な目的は、自分の持っているカードを確認したり、カードのステータスを確認すること。
cupid	ガチャページ。主な目的は、ゲーム内ポイントを用いて新しいカードを手に入れること。