

潜在状態を用いたコミュニティサービスの分析

Analysis of Community Service Focusing on Latent States of Users

垣内弘太^{*1} 鳥海不二夫^{*1} 高野雅典^{*2} 和田計也^{*2} 福田一郎^{*2}
 Kota Kakiuchi Fujio Toriumi Masanori Takano Kazuya Wada Ichiro Fukuda

^{*1} 東京大学 The University of Tokyo ^{*2} 株式会社サイバーエージェント CyberAgent, Inc.

A lot of people communicate with others in Web services such as Social Networking Services because of the popularity of mobile phones and PCs. In this study we analyze communities on Web service focusing on latent states of users. Firstly, we extract latent states of users from action data with hidden Markov model. Then we classify users according to latent states of them in order to realize how they enjoy the service. Finally, we analyze communities on Web service. We find the correlation between how users enjoy the service and their friendships. Because of the correlation, users whose ways of enjoying the service are similar to each other are likely to be contained in the same community. The result of this study is useful for improving Web services focusing on how users enjoy them to make users communicate with others more often and develop communities.

1. はじめに

1.1 背景

近年、パソコンやスマートフォンといった情報通信端末が広く普及しており、それに伴いソーシャルネットワーキングサービス (SNS) などの Web サービス上におけるコミュニケーションが活発になっている。ビジネスの視点から見ると、ユーザが他ユーザと活発に交流することで、利用時間の増加やコアユーザになることが期待できるため、Web 上の人間関係のコミュニティを分析することは、サービスの発展に関わる重大なタスクである。また、Web 上での人間関係の分析は、実際の人間の交流や行動の理解にもつながるため、学術的にも非常に関心が高い分野である。

Web サービスにおける人間関係の分析を行う際、Web 特有のデータとして、ユーザのログデータを用いることができる。しかし、ログデータでは、ユーザの潜在的な目的や考えが考慮されていない。よって、我々は、河津らの手法 [1] を用いてログデータからユーザの潜在状態を抽出し、それらを用いて Web 上の人間関係の分析を行う。

本研究では、ユーザの潜在状態に着目して Web 上のコミュニティを分析する。まず、ユーザの行動ログから潜在状態を抽出する。そして、潜在状態から各ユーザのサービスの楽しみ方を定義し、コミュニティに属するユーザを楽しみ方で分類する。最後に、友人関係とサービスの楽しみ方の近さの関係を明らかにした上で、各コミュニティをユーザの楽しみ方に着目して分析し、サービスの楽しみ方がコミュニティの組成に影響を与えていることを明らかにする。コミュニティごとのサービスの楽しみ方の特徴を抽出することで、ユーザが楽しんでいる機能の改善によるコミュニティの活発化へのアプローチや、ユーザを、その楽しみ方から定着しやすいコミュニティに誘導し、コアユーザを増やすといった施策への活用が期待できる。

1.2 関連研究

Web サービスにおける人間関係の研究は、これまで数多く行われている。しかし、それらの多くは、例えば、Twitter^{*1} のフォロー、フォロワーのつながりによるネットワークの構造分析 [6][7] や、mixi^{*2} における友人関係ネットワークの分析 [8] など、人間関係ネットワークの形成原理や構造に焦点を当てたものである。

ユーザに注目した研究として、F.Benevenuto ら [9] らは、4 つの SNS について、クリックストリームを用いて、ユーザの Web 上での行動履歴から友人間の交流の仕方を分析している。クリックストリームとは、ユーザの Web ページの遷移の軌跡を示すログデータである。F.Benevenuto らの研究では、友人のページを訪れた履歴から直接的に交流の分析をしているが、我々は全ログデータを用いてユーザのサービスの楽しみ方を抽出することで、ユーザの性質に着目した友人関係の分析を行う。

2. 分析データ概要

2.1 使用データ

本研究では、表 1 のサービスにおけるユーザの行動ログを分析対象とした。用いたデータの中では、サービス内のユーザの行動は全部で 142 種類の activity で表現される。行動ログは、匿名化されたユーザ ID、タイムスタンプ、activity、activity を行った場所を表す場所 ID の 4 要素から構成される。ユーザごとに、期間中の行動ログを時系列順の一つにつなげたものを、各ユーザの行動系列と定義する。

2.2 サービス内の友人関係

本研究では、アメーバビグ内の友人関係のコミュニティを分析の対象とする。しかしながら、利用データには明示的な友人関係データが存在しない。そこで、まず行動ログから、友人関係のコミュニティを抽出する。その方法としては、全行動ログの内、activity がアメーバビグ内での発話を表す [chat] である行動ログのタイムスタンプ、場所 ID を用いて、同じ場所で短

連絡先: 垣内弘太, 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:kakiuchi@crimson.q.t.u-
 tokyo.ac.jp

*1 <http://twitter.com/>

*2 <https://mixi.jp/>

表 1: 使用データ

サービス名	アマーバビグ
運営	株式会社サイバーエージェント
種類	コミュニティサービス
期間	2016/5/10~2016/5/31

表 2: 各クラスタの楽しみ方の特徴

クラスタ	特徴
0, 1, 9, 13	会話
2	釣りゲーム
4, 5, 6	音楽動画
7	音楽動画, 釣りゲーム
8, 12	ギフト
10	友人の家を歩き来して交流
11	コミュニケーションの比重が低く多様
14	釣りゲーム, カジノ
15	広場を遷移して交流

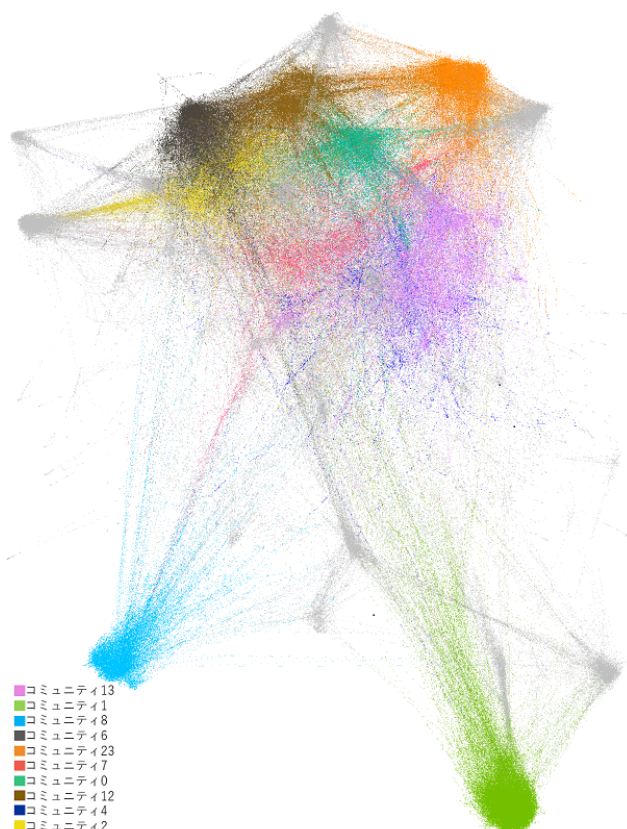


図 1: 友人関係ネットワーク

い期間に [chat] を繰り返しているユーザー同士を発見することで、アマーバビグ内の友人関係を抽出する。そして、抽出した友人関係からネットワークを構築し、その友人関係ネットワークをクラスタリングして、コミュニティに分割する。クラスタリング手法としては、louvain 法 [2] を用いた。含まれるユーザー数 top10 のコミュニティが色付けされた友人関係ネットワークを図 1 に示す。ノード数は 27,836、リンク数は 160,722 である。本研究では、ある程度の人数が所属するコミュニティを分析対象とするために、ユーザー数 top10 のコミュニティについて分析を行う。

3. ユーザー別サービスの楽しみ方の抽出

3.1 抽出手法

河津ら [1] は、ログデータを用いて、Web サービスにおけるユーザーの楽しみ方を検出する手法を提案している。河津らの研究では、ユーザーの行動系列はその時々での楽しみ方、すなわちユーザーの潜在状態に基づき形成されるという前提のもとに、ユーザーの行動系列に隠れマルコフモデル [3] を適用し、ユーザーの潜在状態という抽象的な事象の系列を定量的に抽出している。

本研究では、この手法をアマーバビグにおけるユーザーの行動系列に適用する。アマーバビグにおける一般的な楽しみ方を抽出するために、全ユーザーからランダムサンプリングした 30,000 ユーザーの行動系列を用いて、1つの隠れマルコフモデルを構築した。隠れマルコフモデルの各潜在変数は、全 activity の出力確率分布によって特徴づけられる。それぞれの activity の出力確率分布を持つ潜在変数は、ユーザーの様々な行動に対応する抽象的な潜在状態だと考えられる。この潜在状態の数は、アマーバビグの楽しみ方の多様性を考慮して 100 個とした。

このようにして抽出した潜在状態を用いて、各ユーザーのサービスの楽しみ方を定量的に定義する。まず、各ユーザーの行動系列を潜在状態系列に変換する。そして、潜在状態系列におけるそれぞれの潜在状態の出現回数をカウントして、長さ 100 のベクトルとする。そのベクトルを標準化したものを、潜在状態ベクトルと呼び、各ユーザーのサービスの楽しみ方を表す特徴量であるとする。

3.2 ユーザーの楽しみ方による分類

ユーザーの楽しみ方を把握するために、ユーザーの潜在状態ベクトルを X-Means[5] を用いてクラスタリングし、ユーザーの分類を行う。X-Means は、K-Means[4] の拡張手法であり、K-Means による分割と、BIC (ベイズ情報量規準) による分割停止を繰り返してクラスタリングを行う。クラスタ数を指定する必要がある K-Means と違い、最適なクラスタ数を推定し、分類することができる。分類した結果、クラスタ数は 16 個となった。各クラスタに含まれるユーザー群の、サービスの楽しみ方の特徴を表 2 に示す。

4. サービスの楽しみ方に着目したコミュニティ分析

4.1 サービスの楽しみ方と友人関係

友人関係とサービスの楽しみ方の関係について考える。現実世界においては、一般的に興味嗜好が似ていたり、生活パターンが近いような人たちが友人関係になりやすいということが言える。これを Web サービスに当てはめると、サービス内で、同じような楽しみ方をしているユーザー同士は友人関係になりやすいと想定される。

よって、友人間、非友人間それぞれにおけるサービスの楽しみ方の近さを定量的に表し、両者を比較する。2 ユーザー間の楽しみ方の近さは、2 ユーザーの潜在状態ベクトルのユークリッド距離で定義する。対象ユーザーの内、友人関係にあるユーザーのペアは全部で 105,339 ペアあり、それらすべての潜在状態ベクトル

ル間のユークリッド距離を求め、ヒストグラムで表す。また、対象ユーザの中で、友人関係にないユーザのペアをランダムに 105,339 ペア選び、同様に全ペアについて距離を求めヒストグラムで表す。それらを、友人間の距離のヒストグラムを黄色で、非友人間の距離のヒストグラムを青色で図 2 に示す。横軸はペアの潜在状態ベクトル間の距離、縦軸はペア数を示す。

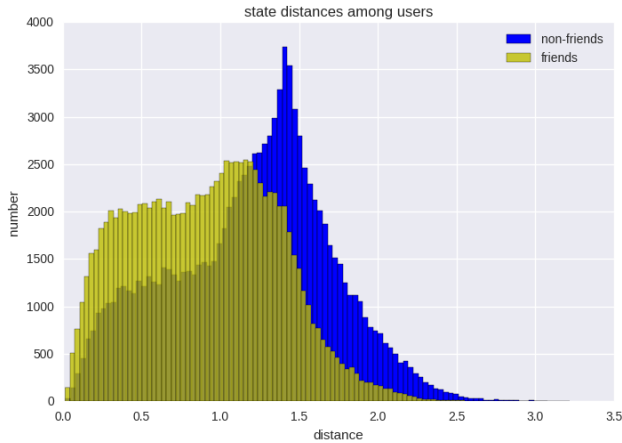


図 2: 友人間, 非友人間の状態距離のヒストグラム

図 2 における 2 つの分布を比較すると、明らかに友人間の方が非友人間より潜在状態ベクトルの距離が短い傾向にあることがわかる。これより、サービスの楽しみ方が近いユーザは友人関係になりやすいことがわかる。

4.2 各コミュニティにおけるユーザの傾向

友人関係にあるユーザはサービスの楽しみ方も似ていることが多いことから、各コミュニティには似た楽しみ方をするユーザが集まりやすく、全体として他とは違う特徴を持つと考えられる。よって、まず、各コミュニティにおける所属ユーザの、3.2 節で分類したクラスター別の割合を求めることで、コミュニティの組成を明らかにする。そして、外れ値検出の手法である smirnov-grubbs 検定を用いて、各コミュニティにおける特徴的なクラスターを取り出す。

例として、まず図 3 に、コミュニティ 13 の組成を示す。ここで、図 3 において、横軸は各クラスターを、縦軸はコミュニティ 13 に所属するユーザのクラスター別の割合を、全対象ユーザにおけるクラスター別の割合、すなわち平均的なコミュニティの組成で割った値を示す。赤色の線は縦軸の値が 1 となるところに引かれており、平均的な組成を示す基準を表す。検定の結果、コミュニティ 13 で特徴的なクラスターはクラスター 2 であり、図 3 を見ると、クラスター 2 の割合が非常に高いことがわかる。ここで、表 2 より、クラスター 2 は釣りゲームをより楽しむユーザ群であることから、コミュニティ 13 は釣りゲームをより楽しむユーザが集まりやすいコミュニティであるといえる。

また、図 4 に、コミュニティ 23 の組成を示す。検定の結果、コミュニティ 23 で特徴的なクラスターはクラスター 4 とクラスター 14 であった。まず、クラスター 4 については、表 2 より、音楽動画をより楽しむユーザ群である。また、クラスター 4 に加えて、クラスター 5, 6, 7 も音楽動画をより楽しむユーザ群である。ここで、図 4 を見ると、クラスター 4 だけでなく、クラスター 5, 6, 7 の割合も平均より高い。クラスター 14 については、表 2 より、釣りゲームとカジノゲーム両方をより楽しむユーザ群である。よって、コミュニティ 23 は、音楽動画をより楽しむ

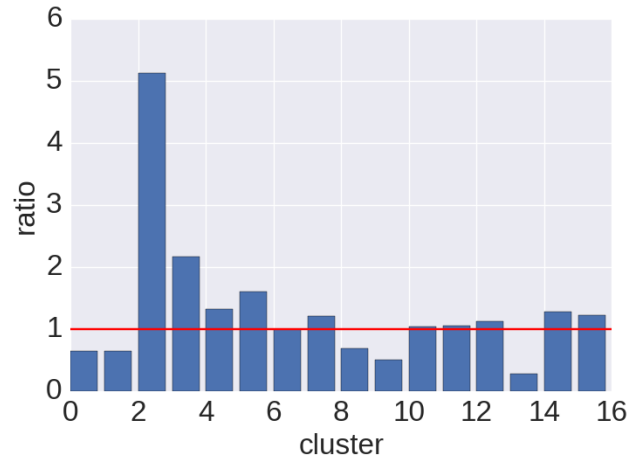


図 3: コミュニティ 13 における各クラスターの割合

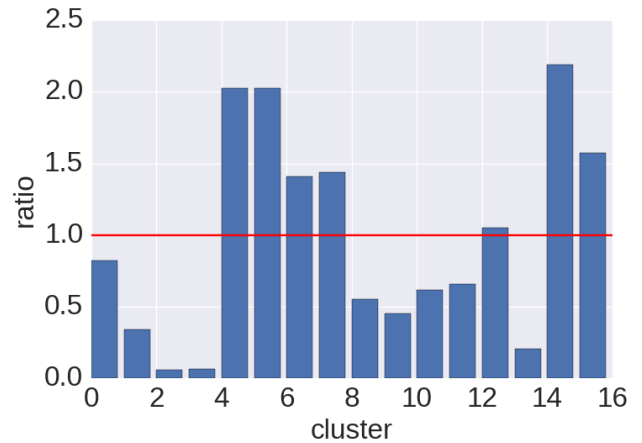


図 4: コミュニティ 23 における各クラスターの割合

ユーザ、釣りゲームとカジノゲームをより楽しむユーザが集まりやすいコミュニティであるといえる。

このような組成の偏りが、例として挙げた上記の 2 つのコミュニティだけでなく、対象のすべてのコミュニティで見られた。よって、サービスの楽しみ方の似たユーザが集まることで、コミュニティの組成に偏りが生まれ、それぞれのコミュニティにサービスの楽しみ方の特徴が現れることがわかる。

5. 結論

本稿では、Web サービスにおける友人関係のコミュニティを、ユーザのサービスの楽しみ方という視点から分析を行った。まず、河津ら [1] の手法を用いて、隠れマルコフモデルを Web サービスのログデータに適用し、ユーザのサービスの楽しみ方を定量的に抽出した。次に、分析対象のコミュニティに属するユーザのサービスの楽しみ方を把握するために、楽しみ方に基づいた、クラスターリングによるユーザの分類を行った。最後に、友人関係とユーザのサービスの楽しみ方について分析を行い、サービスの楽しみ方が近いユーザは友人関係になりやすい傾向があることを示した。そして、各コミュニティをユーザの楽しみ方に着目して分析し、コミュニティの組成に偏りが

あり、それぞれのコミュニティにサービスの楽しみ方の特徴が表れることを明らかにした。このようなコミュニティごとの楽しみ方の特徴に着目してアプローチすることで、Web サービスにおけるコミュニケーションを活性化し、サービスの発展につなげることが期待される。

今後の展開としては、ユーザの潜在状態を用いた、コミュニティの形成過程に着目した分析が挙げられる。この分析により、ユーザのサービスの楽しみ方がコミュニティの発展に与える影響に関する知見が得られると考えられる。

参考文献

- [1] 河津裕貴, 鳥海不二夫, 高野雅典, 和田計也, 福田一郎, "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", 第30回人工知能学会全国大会, 4D4-2, 2016.
- [2] Blondel, V. D., Guillaume, J. L., Lambiotte R., and Lefebvre E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 10, p. P10008, 2008
- [3] 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.
- [4] MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1967.
- [5] Pelleg, Dan, and Andrew W. Moore. "X-means: Extending K-means with Efficient Estimation of the Number of Clusters." *ICML*. Vol. 1. 2000.
- [6] Teutle, Abraham Ronel Martnez. "Twitter: Network properties analysis." *Electronics, Communications and Computer (CONIELECOMP)*, 2010 20th International Conference on. IEEE, 2010.
- [7] Kwak, Haewoon, et al. "What is Twitter, a social network or a news media?." *Proceedings of the 19th international conference on World wide web*. ACM, 2010.
- [8] 湯田聡夫, 小野直亮, and 藤原義久. "ソーシャル・ネットワークワーキング・サービスにおける人的ネットワークの構造." *情報処理学会論文誌* 47.3 (2006): 865-874.
- [9] Fabrcio Benevenuto, Tiago Rodrigues, Meeyoung Cha, Virglio Almeida, Characterizing user behavior in online social networks, *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, November 04-06, 2009, Chicago, Illinois, USA