

コミュニティ構造を用いた Web サービスにおけるユーザ推薦手法の検討

User Recommendation System using community structure in SNS

関 喜史^{*1}
Yoshifumi Seki

福田 一郎^{*2}
Ichiro Fukuda

松尾 豊^{*1}
Yutaka Matsuo

^{*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.

This paper presents a user recommendation system for social networking services (SNSs). In a SNS, a user belongs to communities which she is interested in and communicates with other users, who may belong to other communities. We assume that communities affects the generation of friendship on SNS and propose a user recommendation system based on a community structure. We evaluate our system on a dataset of a Japanese large web community service called Ameba.

1. はじめに

近年ソーシャルメディアに対する関心が非常に高まっている。ソーシャルネットワーキングサービス(SNS)がその代表であり、現在世界最大の SNS である Facebook^{*1} は 2011 年 9 月に全世界でアクティブユーザ数が 8 億人に達したと発表した^{*2}。日本国内でも 2012 年 3 月の時点でアクティブユーザ数が 1000 万人に達したと発表している^{*3}。

こうした SNS の広がりに伴い、SNS 機能をベースにした様々なサービスも生まれている。中でもゲームと組み合わせたソーシャルゲームが注目を集めており、Facebook 上で提供されている Zynga 社の CityVille は月間アクティブユーザが 5000 万人ともいわれている^{*4}。国内では GREE^{*5}、Mobage^{*6} といったソーシャルゲームを主体とした SNS が展開されており、GREE は約 2800 万人^{*7}、Mobage は 3500 万人^{*8} が登録している。またビジネスのつながりに特化した LinkedIn^{*9}、動画を介したコミュニケーションを図る Youtube^{*10}、音楽を中心とした Last.fm^{*11}、1 ユーザあたりの友人関係数を少数に制限する Path^{*12} など、目的に応じて多種多様な SNS が運営されている。

SNS においてユーザは他のユーザと友人関係を形成し情報をやり取りする。さらに特定のトピックについて関心を持つユーザ同士がグループ機能を用いて情報の共有や発信を行なっている。このようにユーザ間の友人関係が増えるほどサービスは活発化し拡大していく。このため多くのサービスでは友人関係を

拡大させるための機能としてユーザ推薦機能を提供している。

SNS を分析する研究や、社会ネットワークにおけるリンク予測の研究は数多くあるが、ユーザ推薦機能に関する研究はまだ少ない。ユーザの発言内容を元に同じような興味を持つユーザを推薦する方法や[Pennacchiotti 11]、ユーザ間の交流情報を元に推薦を行うものなどが提案されている[Blackstorm 11]。しかし多くの SNS ではすでに友人関係があるユーザとの共通の友人関係の数をみる、いわゆる Common Neighbors を元にした推薦が一般的になっている。

ここまで紹介したように SNS も多様化しており、趣味をベースとした友人関係を形成することに特化した SNS や、実際の人間関係を反映した友人関係を形成することを目的とした SNS もある。さらに最近では Facebook のように SNS 上に様々なサービスが展開されはじめている。こういった状況ではユーザごとに構築したい友人関係の属性が異なる。たとえば同じ趣味のユーザを探したいというユーザや、過去の同級生を探したいというユーザ、将来のビジネスにつながりそうなユーザを探したいというユーザなど様々なケースが考えられる。

本研究の目的は SNS における新たなユーザ推薦手法を構築することである。過去さまざまな研究により SNS のネットワークがもつ特徴は明らかになっている。その中でもグループ機能に関する特徴を用いて、ユーザの友人関係形成に対する志向性を反映した推薦を構築することを目指す。

以下2章で関連研究を紹介し、3章で検証に用いたデータを紹介し全体的な分析を行う。4章で本研究のユーザ推薦手法について述べ、5章で手法について既存手法との比較実験を行う。最後に6章で本論文のまとめを述べる。

2. 関連研究

関連研究として、SNS の分析、リンク予測、SNS におけるリンク予測の研究について以下で述べる

2.1 SNS の分析

SNS は学術的にも注目を集めており、様々な研究が行われている。SNS を分析した例としては、Massachusetts Institute of Technology と Stanford University それぞれの SNS データを分析した研究がある[Adamic 03]。Mislove らは Flickr, Youtube, LiveJournal, Orkut それぞれのネットワークを比較分析している[Mislove 07]。また Benevenuto らは Orkut, MySpace, Hi5, LinkedIn といった大規模な SNS 上でのユーザのクリックデータを分析することで、各サービスにおけるユーザ行動の違いを示した[Benevenuto 09]。Viswanath らは Facebook においてユー

連絡先: 関 喜史, 東京大学大学院工学系研究科技術経営戦略学専攻, seki@biz-model.t.u-tokyo.ac.jp

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

*2 f8 での発表

*3 Mobile Hack Tokyo での発表

*4 <http://www.appdata.com/> より

*5 <http://gree.jp/>

*6 <http://www.mbga.jp/>

*7 グリー株式会社 2012 年6月期第二四半期決算説明会資料より概算

*8 Tech-In-Asia に株式会社ディー・エヌ・エーが提供した数値より <http://www.techinasia.com/japan-social-dena-gree-mixi/>

*9 <http://www.linkedin.com/>

*10 <http://www.youtube.com/>

*11 <http://www.lastfm.jp/>

*12 <https://path.com/>

ザの持つ友人関係数とそのユーザのサービス上におけるアクティビティの関係性を分析した[Viswanath 09]. 国内の SNS においては mixi を対象に湯田らがネットワーク全体の特性について論じている他[湯田 06], 松尾らはコミュニティと呼ばれるグループ機能がネットワークに及ぼす影響について論じている[松尾 07]. また So-net SNS を用いた研究として鳥海らは小規模な SNS の比較分析を行い SNS ごとにユーザの行動に差があることを示した他[鳥海 10], 山本らはコミュニケーション構造の時系列変化の分析を行なっている[山本 11].

2.2 リンク予測問題

SNS におけるユーザ推薦は社会ネットワークにおけるリンク予測問題として捉えることができる. 鹿島はリンク予測問題を「ネットワーク構造の既知の部分が与えられたとき, これを手がかりに未知の部分を予測する」として定義している[鹿島 07]. 代表的な研究として Liben-Nowell らが行ったものがある[Liben-Nowell 04]. Liben-Nowell らはすべてのノード間をリンク指標の大きい順に並べて, 閾値以上のノード間にエッジが存在すると予測出来るとして論文の著者ネットワークを元に様々なリンク指標の比較実験を行っており, 結果として Adamic-Adar[Adamic 03]が最も良い結果を示したと報告されている. また最近ではネットワークにおけるコミュニティ構造を用いたリンク予測や, 教師ありモデルなども研究が進められている.

2.3 SNS におけるユーザ推薦手法

このように SNS 自体の研究や, リンク予測に関する研究は活発に行われているがユーザ推薦手法に関する研究は驚くほど少ない. その理由としては有効性を検証するためのハードルが非常に高いことが挙げられる. Backstorm らは教師ありランダムウォークモデルを用いたリンク予測手法を提案し, 論文の著者データに加えて Facebook のデータを用いて検証を行い, ユーザ推薦手法としての可能性について言及している[11 Backstorm 11]. また Pennacchiotti らはトピックモデルを用いた推薦システムを提案しており, twitter のフォロー情報を用いて検証を行なっている[Pennacchiotti 11].

しかしこれらの手法は SNS 固有の特徴を生かしているとはいえない. 先に述べたとおり SNS に対する研究は盛んであり, SNS の持つ特徴は明らかになりつつある. 本研究では SNS のネットワークが持つ特徴を生かした推薦手法を提案しその有効性を検証する.

3. 本研究で用いたデータ

本研究では株式会社サイバーエージェントが運営するコミュニケーションサービスサイト「Ameba」^{*13}のデータを用いる.

Ameba ではブログサービスである「アメーバブログ」, マイクロブログサービスである「アメーバなう」, コミュニティサービスである「Ameba グルっぽ」, コミュニティサービス「アメーバピグ」など多数のサービスが運営されている. 本研究ではこのうち「アメーバブログ」における「メンバー」という SNS 機能におけるユーザ間の友人関係情報, そして「グルっぽ」におけるユーザの参加情報を利用した.

アメーバブログはブログサービスであるが, 「メンバー」という SNS 機能を有しており, ユーザはブログ記事, 写真, 動画などをメンバーとして承認したユーザに対してのみ公開できるようになる. メンバーは承認した側とされた側で区別されているが機能上は変わらないため, 今回はすべて方向なしのグラフとして扱っている.

グルっぽはコミュニティサービスであり, 同じグルっぽに所属するユーザ同志が掲示板を介してコミュニケーションをとることができる. グルっぽへの参加は誰でも参加できるものから, 管理人が承認するもの, また設定した問題に答え, その回答結果によって参加の可否が判断される「入試」という機能もある.

4. Ameba データの分析と推薦手法の提案

4.1 Ameba データの分析

Ameba のデータについて松尾らの mixi におけるデータと比較し分析を行う. データにおけるメンバーサービスの利用ユーザ数は 14,306,578 人, メンバー数の平均は 4.11 人である. このうちメンバーが 3 人以下のユーザは全体の 83.2%となっている. mixi において平均友人数は 10.48 人, 3 人以下のユーザは全体の 44.3%であったことを踏まえると, mixi のものに比べて更に疎なネットワーク構造となっていることが分かる.

グルっぽ数は 265,415 個, 参加人数の平均は 74.5 人である. グルっぽに一つでも所属しているユーザは 3,421,085 人, 一人あたりのグルっぽ参加数は 5.95 個であった. mixi では全ユーザのうち 66.4%がコミュニティに所属しており, 一人あたりの参加数が 24.97 と一人あたりのグルっぽ参加数は非常に少なくなっている.

参加人数が多いグルっぽ上位 10 件を表 1 に示す. なお公式のグルっぽ以外はグルっぽの概要を示している.

表1 参加人数の多いグルっぽ

グルっぽ名・概要	参加人数
【公式】ピグライフ グルっぽ	103,775
「ダイエットしたい人のグルっぽ」	101,236
[公式]アージュカグルっぽ	93,690
「メンバーが欲しい人のグルっぽ」	88,723
[公式]モグ グルっぽ	69,014
「ブログにコメントを付け合うグルっぽ」	58,391
[公式]前田敦子グルっぽ	58,185
「音楽が好きな人のグルっぽ」	55,723
アメーバピグ[公式]グルっぽ	54,459
「ブログにつけるタグについてのグルっぽ」	53,227

表2 結合性の高いグルっぽ

グルっぽの概要	参加人数	結合性
「知り合い同士のグルっぽ」	70	0.99
「ブログパーツのユーザグルっぽ」	94	0.56
「中学校のグルっぽ」	51	0.65
「専門学校のグルっぽ」	33	0.98
「アメーバピグの釣り友達のグルっぽ」	154	0.2
「中学校のグルっぽ」	61	0.47
「地域の遊び友達のグルっぽ」	61	0.46
「アイドルファンのグルっぽ」	32	0.85
「中学校のグルっぽ」	60	0.44
「アイドルファンのグルっぽ」	121	0.22

松尾らの研究において, グルっぽ機能とユーザ間の友人関係の関わりを表す指標として結合性(Community Connection)が提案されている[松尾 07]. これは同一グループに所属するユーザ間のネットワーク密度であり, これが低いほどグループ内の交流が活発であることを示すとされている. この指標をグルっぽに対しても適用した. 表 2 に結合性にグルっぽの参加ユーザ数を

*13 <http://www.ameba.jp/>

乗じた数値を指標としたときの上位 10 件のグルッぽを示す。松尾らの mixi の分析ではリアルにつながりのあるコミュニティやオフ会を頻繁に行なっているようなコミュニティが上位にみられた。グルッぽにおいては現実世界でつながりのある知り合いや学校関連のグルッぽに加え、アメーバピグでのつながりのあるグルッぽやアイドルのファンが集まるグルッぽなどがみられる。またこれ以外に結合性の高いコミュニティとして、アニメキャラになりきってブログを書くユーザの集まりや、オフ会を定例的に開催しているグルッぽなど多種多様なつながりを見ることができる。リアルな人間関係を反映したネットワークを形成しやすくなっている mixi に対して、Ameba では様々な関係が構築されていることがここからもわかる。

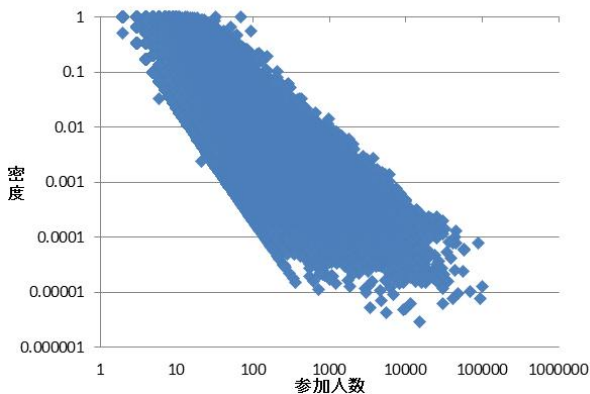


図 1 グルッぽの参加人数と密度の分布

5. 提案手法

Facebook や mixi は主にリアルな人間関係を反映したネットワークであり、mobage や GREE ではソーシャルゲームでのつながりが反映されたネットワークを形成している。それに対して Ameba のサービスにおけるユーザ間のつながりの有り様はこのようなグルッぽの結合性に現れるように、ウェブ上のつながりとリアルなつながりが様々に混在している。そしてもちろんユーザごとにどういったつながりを形成したいかという意思は異なる。今回提案する手法はグルッぽの結合性をベースにそういったユーザごとのつながりに対する意思を反映して推薦することを目指す。

あるユーザ u のメンバーの集合を F_u 、あるグルッぽ g の参加ユーザの集合を G_g 、ユーザ u が参加しているグルッぽの集合を UG_u 、グルッぽ g の結合性を $cc(g)$ とする。ここで結合性にグルッぽ内のメンバー数を乗じることでグルッぽ g がユーザ u の友人形成に与える影響を $ConnectionValue(CV)$ として以下のように定義しユーザ間のスコアを定めスコアの高い順に推薦する。

$$CV(u, g) = cc(g) \cdot |F_u \cap G_g|$$

$$score(u_i, u_j) = \sum_{g \in UG_{u_j}} CV(u_i, g)$$

グルッぽにおいて結合性とユーザ数を乗じることでそのグルッぽの友人形成における影響度を見ることができた。それを踏まえ、そのグルッぽに所属しているメンバーの数を乗じることでそのグルッぽがユーザの友人形成に与える影響を表すことができると仮定する。CV はユーザのメンバーがそのグルッぽに所属していないと 0 となるため、友人推薦の対象となるユーザはユ

ーザのメンバーが所属しているグルッぽの所属ユーザとなる。つまり実際に推薦を行う場合には、対象ユーザのメンバーが所属しているグルッぽのメンバーの中で、対象ユーザのメンバーでないユーザについてスコアの計算を行い、スコア順にランク付けを行うこととする。

SNS の友人推薦において提案手法の既存手法に対する利点は 2 つある。第一に各グループの結合性を事前に計算しておくことで **common neighbors** とほぼ同等の計算量で計算できる点である。数千万人のユーザデータすべてを定期的に計算する必要があるため、計算コストは低ければ低いほどよい。

第二に **common neighbors** モデルでは推薦できないユーザも推薦できる可能性がある点である。推薦システムの評価には **coverage** という指標がある。これは全データ中推薦システムが対象とできる推薦対象アイテムの割合を表すものである。提案手法では友人が参加しているグループに所属しているメンバーを推薦対象とできるため、**common neighbors** では推薦できないユーザも推薦できる可能性がある。

6. 評価実験

6.1 実験方法

ユーザ推薦の精度評価として 5 分割交差検定を行う。推薦を行いたいユーザのメンバーを 5 分割し、そのうちひとつをテスト集合として取り出し、残りのメンバーを用いて推薦を行う。テスト集合のユーザ数 n 人と、推薦結果スコア上位 n 人を比較しどれだけ当てることができたかを評価する。これをすべての分割に対して行い、結果の平均を精度として評価する。

提案手法において適切な推薦を行うためには一定以上の結合性を持ったグルッぽに対象ユーザが所属している必要がある。そのため今回評価実験ではグルッぽの結合性に所属ユーザ数を乗じた値の上位 50 件のグルッぽに所属する 1758 人のユーザを対象として行う。

6.2 実験結果と考察

まずユーザ全体に対する精度とメンバー数 50 人以上のユーザに対する精度の比較を表 3 に示す。精度はテスト集合が n 件あったとき、スコア上位 n 件に何%テスト集合のユーザが含まれているかを表している。メンバーを 5 分割していることから、メンバー数が 50 人以上であればテスト集合が 10 件以上となるため、50 人以上についても別途精度を求めるとした。なおメンバー数 50 人以上のユーザは 1758 人中 911 人である。

表 3 推薦精度比較実験の結果

	提案手法
ユーザ全体	0.231
メンバー50人以上のユーザ	0.3

図 2 に結合性と精度の分布図を示す。提案手法では結合性が高くなるにつれ精度が高くなっていくことがわかる。このことからグルッぽの結合性がメンバー形成に関係していることがいえる。

結合性と参加ユーザ数を乗じた値の上位 10 件に対してグルッぽごとの精度を比較したものを表 4 に示す。特に参加ユーザ数が 30 人程度で結合性が高いグルッぽにおいては非常に高い精度を示しており、こういった小規模で結合性が高いグループ構造が SNS における友人関係形成に対して大きな影響を持っていると推測できる。反対に参加ユーザ数が 50 人以上でありながら結合性が 0.2 程度と参加ユーザ数に対して結合性が比

較的高いグループでは、大幅な優位性は見られなかった。今回結合性に対してそのグループに存在するメンバーの人数を乗じるという単純な手法を用いたが、グループの規模に基づく重みづけの方法については、今後より詳細に検討していく必要がある。

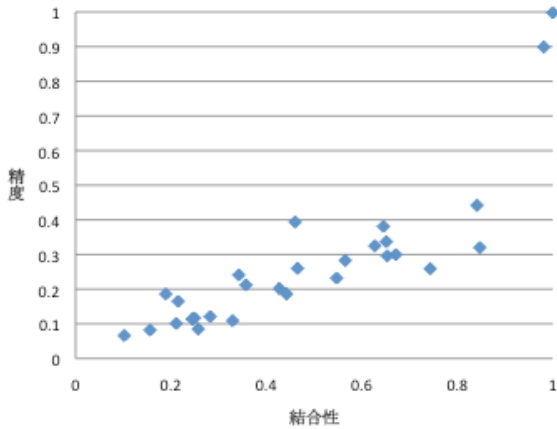


図 2 結合性と精度の分布

表 4 グループごとの精度の比較

参加人数	結合性	提案手法
70	0.999	0.998
94	0.565	0.283
51	0.645	0.381
33	0.981	0.899
154	0.189	0.186
61	0.465	0.26
61	0.46	0.394
32	0.847	0.32
60	0.442	0.186
121	0.215	0.165

7. まとめ

SNS におけるユーザ間には同級生、同じ趣味、職場といった様々な属性があり、それを特定することは非常に意義があることである。我々はグループを用いることでそうしたユーザ間の関係を表すことができると考えている。本研究では SNS におけるグループ機能が友人形成に与える影響に着目し、友人形成の指向性をユーザ推薦手法に取り込むことを目指した。

このような背景からグループ機能における結合性という使用をベースにしたユーザ推薦手法を提案し、アメーバブログのメンバーとグループのユーザデータを用いて評価実験を行った。結果として結合性が高いグループのユーザほどより良い推薦ができていたことから、グループの結合性が友人関係形成に大きな影響を与えていることを示すことができた。

しかしグループの持つユーザへの影響を定める際の結合性への重みづけにはより詳細な検討が必要になる。これはサービス全体におけるグループの特性が関係してくることが予想されるため、参加ユーザ数などの分布をみながら決定する方法を検討していきたい。また推薦手法として考えたとき、結合性の高いグループは全体のごく一部であるため、そういったグループに所

属していないユーザに対しても推薦ができるような手法も検討していきたいと考える。

今後は既存手法と比較しながら上記のような検討を行い手法のブラッシュアップに努めていきたい。

近年 SNS の普及はさらに拡大しており、様々な SNS や SNS に紐づくサービスが登場している。この流れの中で SNS におけるユーザの行動はますます複雑化しており、学術的にも非常に興味深い課題となっている。本研究がユーザ推薦にとどまらず、SNS における友人関係の形成プロセスを解き明かすための一助となれば幸いである。そしてこの知見が学術的な観点だけでなくよりよいサービス提供に生かされることを期待したい。

謝辞

本研究は株式会社サイバーエージェントが開催した「サマーインターンシップ・テクノロジーコース・ラボインターン」の下で行われました。

「Ameba」のデータ提供、及び Ameba Technology Laboratory の社員の皆様の多くのご助言に感謝いたします。

参考文献

- [Adamic 03] Lada A. Adamic, and Eytan Adar :Friends and neighbors on the Web, SocialNetworks, Vol. 25, No. 2, pp. 211-230 (2003).
- [Backstorm 11] Lars Backstorm, and Jure Leskovec . : Supervised Random Walks: Predicting and Recommending Links in Social Networks. , WSDM'11 (2011) .
- [Benevenuto 09] Fabricio Benevenuto, Tiago Rodrigues, Meeyoung Cha, and Vrigilio Almeida: Characterizing User Behavior in Online Social Networks, In IMC'09 (2009) .
- [Liben-Nowell 04] David Liben-Nowell, and Jon Kleinberg : The Link-Prediction Problem for Social Networks. , CIKM'04 (2004).
- [Mislove 07] Alan Mislove, Massimiliano Marcon, and Krishna P. Gummadi :Measurement and Analysis of Online Social Networks, In IMC'07 (2007).
- [Pennacchiotti 11] Marco Pennacchiotti, and Siva Gurumurthy. : Investigating Topic Models for Social Media User Recommendation. , WWW2011 Poster (2011).
- [Viswanath 09] Bimal Viswanath, Alan Mislove, Meeyoung Cha, and Krishna P. Gummadi. : On the Evolution of User Interaction in Facebook. , WOSN'09 (2009).
- [湯田 06] 湯田 徳夫, 藤原 義久 :ソーシャル・ネットワーキング・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, Vol. 47, No. 3, pp. 865 - 974 (2006).
- [松尾 07] 松尾 豊, 安田 雪 :SNS における関係形成原理 - mixi のデータ分析 -, 人工知能学会論文誌, Vol. 22, No. 5, pp. 531-541 (2007).
- [鹿島 07] 鹿島 久嗣 : ネットワーク構造予測, 人工知能学会誌, Vol. 22, No. 3, pp. 344-351 (2007).
- [鳥海 10] 鳥海 不二夫, 山本 仁志, 諏訪 博彦, 岡田 勇, 和泉 潔, 橋本 康弘 : 大量 SNS サイトの比較分析, 人工知能学会論文誌, Vol. 25, No. 1 pp. 78-89 (2010).
- [山本 11] 山本 仁志, 諏訪 博彦, 岡田 勇, 鳥海 不二夫, 和泉 潔, 橋本 康弘 : コミュニケーション構造の推移による大量 SNS サイトの分類, 日本社会情報学会誌, Vol. 23, No. 1, pp. 33-43 (2011).