

QA サイトにおける質問応答グラフの成長パターン分析

Analysis on Evolution of Question-Answer Pattern in a QA Site

甲谷 優[▼] 川島 晴美[▼]
藤村 考[▼] 奥 雅博[▼]

Yutaka KABUTOYA Harumi KAWASHIMA
Ko FUJIMURA Masahiro OKU

教えて! goo, Yahoo Answers をはじめとする QA サイトの利用が増加してきている。Adamic らは Yahoo Answers の各カテゴリをコミュニケーションタイプ「知識交換」「相談」「議論」に分類し、カテゴリ毎に良回答(良回答者)を予測する手法を提案している。本研究では、教えて! goo に対してコミュニケーションタイプ分類が可能か検討した結果を報告する。まず、Adamic らが利用した素性では教えて! goo の分類に適さないことを示す。次に、新しい素性として Adamic らの QA ネットワーク分析に時系列の概念を追加した、成長パターン分析の手法を提案する。エゴネットワークを時系列分析した結果から、成長パターンがコミュニケーションタイプ分類の素性として利用できる可能性を示した。

Recently, the question-answer (QA) site, such as Oshiete! goo or Yahoo Answers becomes more and more popular. Adamic et al. analyzed question-categories in Yahoo Answers, and they classified the categories into three clusters. The clusters are different from each other in communication types. First, we classify categories in Oshiete! goo into three groups according to the same way as Adamic et al. did, and compare Oshiete! goo with Yahoo Answers. Furthermore, we study the dynamic structure of social interaction pattern which Adamic et al. did not addressed.

1. はじめに

近年、QA サイトと呼ばれる Web サービスの利用が拡大してきている。QA サイトとは、あるユーザの自然文で表わされた質問に対して別のユーザが回答することによる、人どうしの知識の共有を目的としたサービスである。代表的な QA サイトに、教えて! goo¹ や Yahoo! 知恵袋がある。たとえば教えて! goo の場合、現在約 360 万件の質問が投稿されており、現在日本では 2 番目の規模の QA サイトである。

Adamic ら[1]は英語圏で最大規模の QA サイトである Yahoo Answers において、カテゴリ毎に多角的な分析を行っている。その結果、QA サイトで行われているコミュニケ

ーションには「知識交換」「相談」「議論」の 3 つのタイプが存在することを示した。

日本語の QA サイトである教えて! goo においても、カテゴリごとに、3 つのタイプのコミュニケーションのうちいずれかがとられていると考えられる。教えて! goo に対する何らかのアプリケーションを考えた場合、コミュニケーションタイプごとの適切な分類はきわめて重要な課題である。たとえば、回答分の内容の解析に基づくベストな回答を予測するシステムは、「知識交換」のように回答が比較的客観的な場においては有効であると考えられるが、一方で「相談」「議論」のように主観的な回答の多い場ではあまり有効でないと考えられる。

そこで本研究では、教えて! goo の各カテゴリをコミュニケーションタイプごとに適切に分類するために、さまざまな分類のための素性の妥当性を検証する。まず正解セットとして、教えて! goo の各カテゴリを筆者らの主観でもって前述の 3 つのコミュニケーションタイプごとに分類した。

Adamic ら [1] は、Yahoo Answers の各カテゴリに対して 3 つの特徴量を提案し、それらを素性としてコミュニケーションタイプ別に分類できることを示した。それらの特徴量について、教えて! goo における各カテゴリをコミュニケーションタイプごとに分類するための素性として有効かどうかを検証した。

QA ネットワークとは、各ユーザをノード、各回答を回答者から質問者へのエッジで表すことにより QA サイトにおける各ユーザ間のインタラクションをグラフにモデル化したものである。Adamic ら [1] は QA ネットワークに対し多角的な分析を行い、結果コミュニケーションのタイプが異なるカテゴリ間では分析結果が異なることを示した。

しかしコミュニケーションは時間とともに変化する。QA ネットワークも同様に、新規ユーザが増えれば疎になり、逆に新規ユーザが増えず質問・回答の投稿数が増えると密になる性質がある。したがって、ユーザ数、ないし質問・回答の投稿数が十分でなければ、同様の分析では異なる結果が出ると考えられる。そこで本研究では、コミュニケーションタイプを分類するための素性として、QA ネットワークの時間ともなう成長パターンに着目した。

2. 関連研究

QA サイト以外の Web フォーラムにおいて、コミュニケーションの性質を分析するために、ソーシャルネットワーク分析を行った研究がある。Usenet を対象として、分析する手法 [2]、可視化する手法がいくつか提案されている [3,4]。また、Zhongbao ら [5] は BBS を分析し、オンラインのコミュニケーションとユーザの興味に関連性について述べている。

ユーザのふるまいとコミュニケーションの性質の関連性について言及した研究も多い。たとえば、Wenger [6] はコミュニティが形成され継続される状況下でのユーザの役割の重要性について言及している。Preece ら [7] は異なる Web 上のフォーラムにて、メッセージを読むだけで投稿しないようなユーザのふるまいについて分析を行っている。Donath [8] は Web フォーラムのユーザの仮想的な ID をマイニングし、さらに嘘を検出する手法を提案している。また、Welser ら [9] は Web 上のコミュニティをモデル化したグラフに対して、エゴネットワーク分析を行った。エゴネットワーク分析とは、あるユーザに着目し、そのユーザにあ

[▼] 正会員 日本電信電話株式会社 NTTサイバーソリューション研究所 {[kabutoya.yutaka](mailto:kabutoya.yutaka@lab.ntt.co.jp), [kawashima.harumi](mailto:kawashima.harumi@lab.ntt.co.jp), [fujimura.ko](mailto:fujimura.ko@lab.ntt.co.jp), [oku.masahiro](mailto:oku.masahiro@lab.ntt.co.jp)}@lab.ntt.co.jp

¹ <http://oshiete.goo.ne.jp/>

たるノードを中心に 1 ホップで結ばれるノードと、それらの間のエッジを可視化し, “answer person” と “discussion person” を比較する手法である. QA ネットワークの成長パターン分析にこのエゴネットワーク分析に基づく手法を利用している.

Web フォーラムのスレッドや投稿に着目した研究も存在する. Sack [10] は議論スレッドにおける会話パターンをみつけるために可視化を行っている. Arguello ら [11] および Joyce ら [12] は新規ユーザの投稿メッセージと, 新規ユーザが継続して Web 上のフォーラムを利用するきっかけとなる別ユーザからの返信メッセージを分析している.

グラフの時系列分析を行う研究も数多く存在する. たとえば, Google のランキングアルゴリズムである PageRank には, 被リンク数は時間に対して単調非減少であるため「古い」ページの評価が高くなる傾向がある. この問題に対処すべく, Cho ら [13] は被リンク数ではなくその増加率でページを評価する手法を提案している. また, 今藤ら [14] は HITS アルゴリズムによって抽出された Web コミュニティの成長パターンについて分析を行っている.

3. 教えて! goo のデータセット



図 1 教えて! goo のスレッドの例
Fig.1 Sampled thread of Oshiete! goo

教えて! goo ではユーザどうしのインタラクションはすべて質問とそれに対する回答という形式をとる. あるユーザが質問を投稿すると, 別のユーザがその質問に対して回答をつける.

また, 教えて! goo では, 他の多くの QA サイトと同様, カテゴリが存在し, 質問者は質問したいトピックに応じてカテゴリを選択しその中で質問を投稿する. 教えて! goo のカテゴリは葉の深さが一定でない木構造になっており, 葉は合計で 329 種類存在する. それらはコンピュータ関係のものからスポーツ, 美容, 政治と多岐にわたる.

図 1 に, 教えて! goo のスレッドの例を示す. 質問は, オレンジ色の枠の中に表示される. また, 各回答は紫色の枠の中に表示され, 色彩として明確に質問と区別される. さらに, 図 1 から明らかに, 質問の枠と回答の枠の間の距離は

回答の枠間のそれよりも長く, この点でも質問と回答は区別される. したがって, ユーザは質問・回答という形式を意識することになり, 他の BBS のようなシステムとは異なりコミュニケーションの方向性が回答者から質問者へと一定になる(すなわち, 回答者間のコミュニケーションの存在の可能性を考えなくて良い).

「マナー・冠婚葬祭」カテゴリにおいて, 以下のような質問があった.

私は, 小 4 と中 1 の娘をもつ母です. 次女の運動会と甥の結婚式が重なってしまいました. 先輩に相談したところ, 一生に一度しかない結婚식을優先するのが当然で常識だと言われました. 今後のつきあひも考えるなら結婚式を優先するのがベストなのでしょうか?

このような質問には合計で 20 件程度の回答がつき, さまざまなユーザの多種多様な意見が寄せられた. この「マナー・冠婚葬祭」カテゴリは平均で 3 件以上の回答がついている. QA サイトでは質問・回答という形のコミュニケーションのフォーマットをとっているものの, 単純な知識交換ではなくこの質問のように相談がなされているものも存在する.

また, 「国内アーティスト」カテゴリにおいて以下のような質問があった.

衝撃を受けた歌謡曲の曲名と, それを歌っている歌手名が知りたいな!!

このような質問に対し, 最終的に 70 件以上もの回答がつけられた. この質問でとられているコミュニケーションの形式は議論であると考えられる. 教えて! goo ではあくまで質問・回答というコミュニケーションの形式をとっているため, 質問に回答するという形で議論に参加することになる. このように教えて! goo においても議論・対話というコミュニケーションは存在する.

今回, データとして 1999 年から 2007 年までの教えて! goo のデータを使って実験を行った. 2,859,606 件の質問に対し, 合計で 9,364,359 件の回答がつけられた. すなわち, 1 質問あたり平均で 3.275 件の回答が得られていることになる. この数字は, Adamic ら [1] の用いた Yahoo Answers のデータセット (質問数: 1,178,983, 回答数: 8,452,337) における 1 質問あたりの平均回答数 7.170 よりずっと少ない. ただし, 1 件でも回答が得られている質問の割合は 0.994 で, ほぼすべての質問が何らかの回答を得ている.

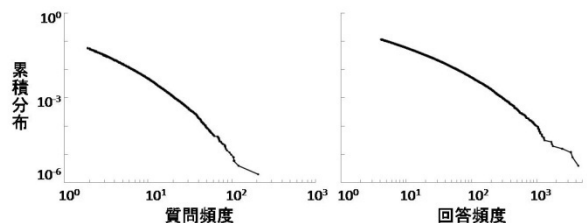


図 2 質問・回答の投稿数とユーザの分布
Fig.2 Distribution of asking frequency (number of users one has posted questions) and replying frequency (number of users one has answered)

一意な質問者・回答者の数はそれぞれ 507,370, 253,072 であり, 質問者の方が回答者よりもずっと多いことがわかる.

図 2 に質問・回答の投稿数とユーザの分布を示す。横軸は各ユーザの質問および回答の投稿数、縦軸はその投稿数以上質問・回答を行ったユーザの割合を表す(たとえば、質問頻度-累積分布グラフ上の点 (134, 4.2×10^{-5}) は、「134 回以上質問したユーザの割合は 4.2×10^{-5} である」ことを表す)。両対数グラフにおいてほぼ直線になるので、ユーザの質問・回答の投稿数に対する分布はべき分布、すなわちスケールフリー性 [15] を示している。また、質問も回答も投稿しているユーザの数は 179,701 で、回答するユーザの多くは質問の投稿も行っていることがわかる。

4. カテゴリの特徴量分析

Adamic ら [1] は以下に示す 3 つの手法を用いて Yahoo Answers の各カテゴリの特徴量を抽出し、その次元 3 の特徴ベクトルを *k-means* 法を用いてクラスタリングした。

- (a) thread length
1 質問につく平均回答数。
- (b) post length
1 回答あたりの平均語数。
- (c) asker/replier overlap
カテゴリごとに質問者ベクトル・回答者ベクトルを作る。それぞれの各成分は各ユーザのそのカテゴリでの質問の頻度および回答の頻度を表す。このとき両ベクトルのコサイン類似度を asker/replier overlap とする。その結果計 3 つのクラスが得られ、それぞれは異なるコミュニケーションタイプが多くみられるカテゴリの集合であることがわかった。1 つ目は、thread length と asker/replier overlap が小で、post length が大になるクラスで、「知識の交換」を目的とする質問が多く含まれるカテゴリであることがわかった。2 つ目は、thread length と asker/replier overlap, post length とともに中くらいの値のクラスで、「相談」を目的とする質問が多く含まれるカテゴリであることがわかった。3 つ目は、thread length と asker/replier overlap が大で、post length が小になるクラスで、「議論」を目的とする質問が多く含まれるカテゴリであることがわかった。

教えて!goo にも同様の素性でコミュニケーションタイプごとにカテゴリを分類できるか検証する。まずは正解セット作成のため、コミュニケーションタイプを以下のように定義する。

1. 知識交換
解が一意に定まる質問と、それに対する回答をさす。Yahoo Answers ではたとえば“Programming”カテゴリがこれにあてはまる。
2. 相談
質問の解が一意に定まらず、かつ質問者にはその質問に対して明確な回答ができないか、あるいは回答者の回答によってその質問に対する質問者自身の回答が変わりうる状況でのやりとりをさす。Yahoo Answers ではたとえば“Marriage”カテゴリがこれにあてはまる。
3. 議論
質問の解が一意に定まらず、かつ質問者が自身なりのその質問に対する回答をもっている状況でのやりとりをさす。Yahoo Answers ではたとえば“Wrestling”カテゴリがこれにあてはまる。

この定義に基づき、教えて!goo における計 329 件のカテ

ゴリそれぞれを 3 つに分類した。具体的には、各カテゴリの「このカテゴリで人気の Q&A」である 10 個の質問を、筆者らの主観により「知識交換」か「相談」か「議論」か判別する。そしてもっとも数が多かったコミュニケーションタイプを各カテゴリのものとした。その結果が表 1 である。

表 1 各コミュニケーションタイプのカテゴリ例
Table 1 Categories for each communication type

タイプ	カテゴリ数	カテゴリの代表例
知識交換	177	お茶・ドリンク, 書道, ゲーム機ディスプレイ, 登山・キャンプ
相談	91	恋愛相談, 就職, 公認会計士, 生命保険, 融資, 食べ歩き(全国)
議論	61	DIY(日曜大工), オンラインゲーム, バイク, 俳優・女優, サッカー

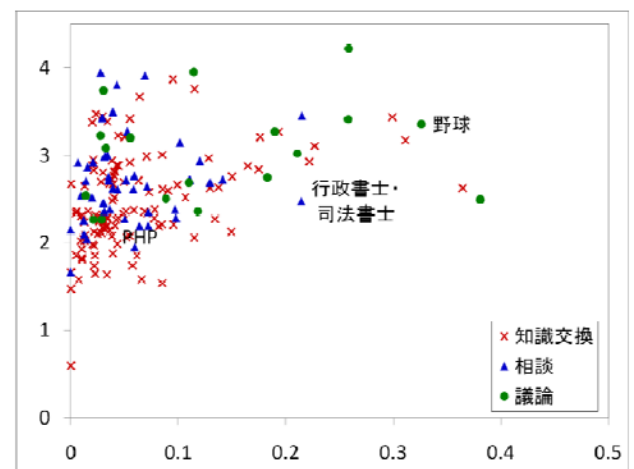


図 3 各カテゴリの post length と thread length
Fig.3 Correlation between post length and thread length

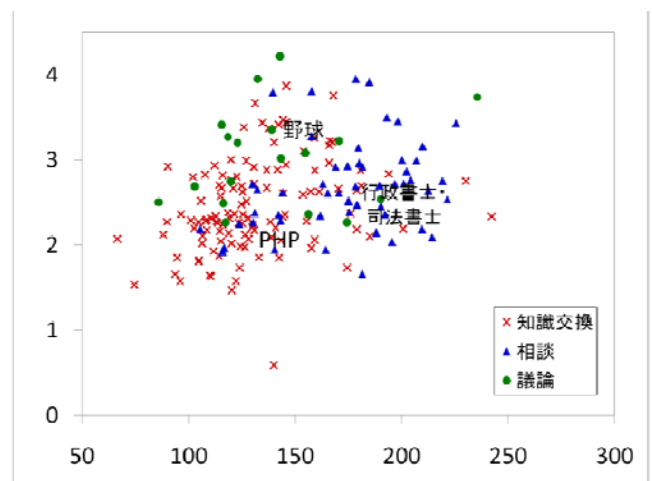


図 4 各カテゴリの asker/replier overlap と thread length
Fig.4 Correlation between asker/replier overlap and thread length

図 3 に横軸に post length, 縦軸に thread length をとった際の教えて!goo の各カテゴリの分布を表すグラフを、図 4 に横軸に asker/replier overlap, 縦軸に thread

length をとった際の教えて！goo の各カテゴリの分布を表すグラフを示す。ただし、post length 算出の際日本語の自然文を扱うにあたって、MeCab [16] のわかち書きの機能を用いた。

教えて！goo では、カテゴリの 3 つの特徴量 (post length, thread length, asker/replier overlap) の値について、3 つのコミュニケーションタイプごとにはっきりした差がなかった (特に、「知識交換」と「議論」との差が明確に出なかった)。したがって、教えて！goo におけるカテゴリを適切にコミュニケーションタイプごとに分類するには Adamic らのものとは別の素性・手法が必要であると考えられる。

また、教えて！goo の分析結果では、Yahoo Answers の分析結果とは異なり「知識交換」カテゴリの方が「相談」よりも post length が小になる傾向にあることがわかった。原因としては以下に示す 2 つが考えられる。

1. 教えて！goo には Yahoo Answers とは異なり、回答者が回答に対して参考 URL をつけるインターフェイスが用意されている。たとえば、ソースファイルの注釈以外のどこかに全角文字が無い検索するにはどうすべきかという質問に対して、

GetStringTypeEx() などいかが？様々な文字や全角半角を見分ける事が出来ますよ。

という一言に加えて msdn ライブラリへのリンクが貼られた回答が存在する。

2. 宮原 [17] は日本人のコミュニケーションの特徴を「高コンテキスト」であると捉えている。「高コンテキスト」であるとは、場面・文脈・状況に強く依存するという意味であり、したがって英語圏の QA サイトである Yahoo Answers と比較して、日本語の教えて！goo では、知識交換において回答者が質問者にある程度の知識を持っていることを前提にやりとりを行う傾向があるために、メッセージが短くなるのではないかと考えられる。逆に相談においては質問者と回答者の間でコンテキストの共有ができていないことや、議論においては回答が質問者のコンテキストに依存しないことが理由でメッセージが短くならないと考えられる。

5. QA ネットワークの成長パターン分析

別のコミュニケーションタイプの分類手法として、Adamic ら [1] の行った QA ネットワークを用いた手法を検討する。ただし、コミュニケーションは時間とともに変化するという性質を考慮し、QA ネットワーク分析に時系列の概念を導入し、より詳細な分析を行った。

5.1 QA ネットワークとは

各カテゴリにおいて、1 度でも使用したことのあるユーザー (質問者・回答者両方を含む) をノード、スレッド内における各回答をその回答者から質問者へのエッジで表わすことによりグラフにモデル化する。このとき、そのグラフをそのカテゴリの QA ネットワークと呼ぶ。以下、知識交換カテゴリ・相談カテゴリ・議論カテゴリのサンプルとしてそれぞれ「PHP」、「行政書士・司法書士」、「野球」の 3 つを選択し、QA ネットワークの成長パターンの分析を行った。

5.2 エゴネットワークの時系列分析

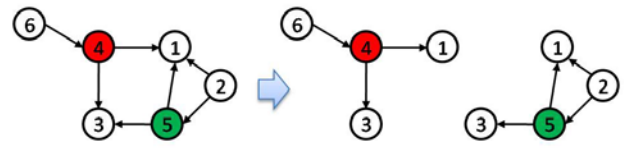


図 5 エゴネットワークの抽出例

Fig.5 Extraction of ego networks

エゴネットワークとは、あるユーザーと、そのユーザーと直接やりとりしたユーザー、そしてそれらのユーザーの間のエッジからなる QA ネットワークの部分グラフをさす。たとえば図 5 において、左図のグラフのうちユーザー 4、ユーザー 5 に対するエゴネットワークは右図のそれぞれのようになる。

Welser ら [9] によれば、あるユーザーに対して、「エゴネットワーク」においてそのユーザーでない周辺ユーザー間のノードにエッジが 1 つもない場合 “answer person” であり、周辺ユーザー間にエッジが 1 つでもある場合 “discussion person” である。たとえば、図 5 において、ユーザー 4 はそのエゴネットワークの周辺ユーザーであるユーザー 1, 3, 6 のいずれのノード間にもエッジが存在しないため、“answer person” である。一方、ユーザー 5 はそのエゴネットワークの周辺ユーザーであるユーザー 1, 2 のノード間にエッジが存在するため “discussion person” である。

本研究では、エゴネットワークの周辺ユーザー数の平均と、“answer person” および “discussion person” の割合の時間変化について、以下の手順にて分析を行った。

1. カテゴリ c に対して、1999 年 8 月から過去のある時刻 t までに投稿された全質問・回答を集約して QA ネットワーク $g(c, t)$ を作成
2. $g(c, t)$ に含まれる全ノード、つまりカテゴリ c に時刻 t の時点で 1 度でも質問・回答を投稿したことのある全ユーザーに対してエゴネットワークを抽出
3. エゴネットワークの周辺ユーザー数の平均を計算
4. “discussion person” の数の、周辺ユーザーが 2 人以上のユーザー数に対する比 (“discussion person” の定義上、周辺ユーザーの数が 1 以下の場合 “discussion person” になりえないため) を計算

時刻 t は 2002 年 4 月 1 日から 2007 年 8 月 1 日まで 1 月毎に、計 65 の時点を選択した。

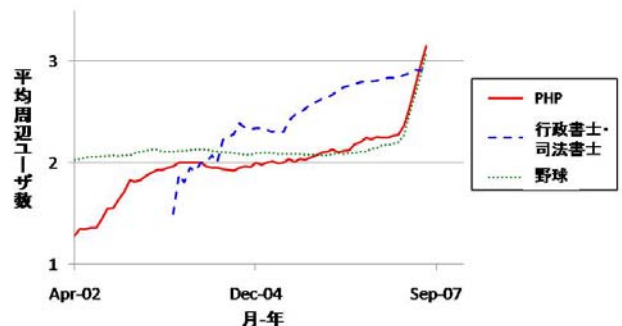


図 6 1 エゴネットワークあたりの周辺ユーザー数
Fig.6 Number of neighbors for ego network

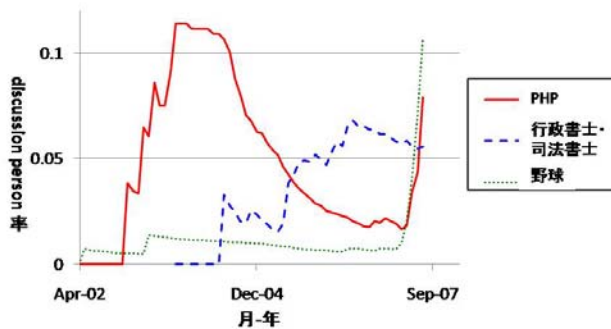


図7 “discussion person” の割合
Fig.7 Fraction of “discussion persons”

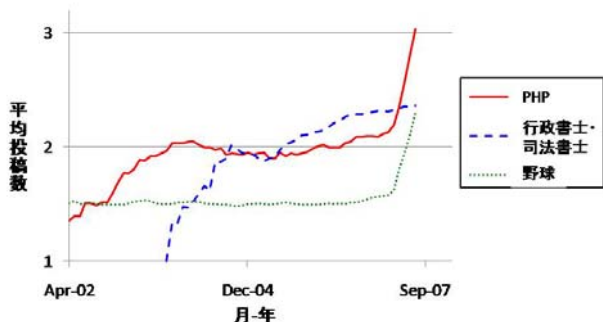


図8 1 ユーザあたりの平均投稿数
Fig.8 Average of posting frequency (number of users one has posted)

図6に3つのカテゴリのQAネットワークにおける各時点ごとのエゴネットワークの周辺ユーザ数の平均の時間変化の様相を、図7に“discussion person”の割合の時間変化の様相を示す。また、エゴネットワークの周辺ユーザ数および“discussion person”の割合はQAネットワークの枝密度と強い相関があると考えられる。また、QAネットワークの枝密度は1ユーザあたりの質問・回答の平均投稿数と強い相関があるのは自明である。ここで、図8に1ユーザあたりの平均投稿数を示す。

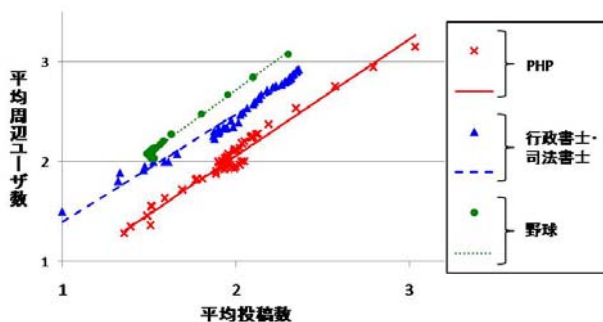


図9 1 ユーザあたりの平均投稿数と周辺ユーザ数
Fig.9 posting frequency vs. neighbors number for a user

図6, 8から、3カテゴリともエゴネットワークの周辺ユーザ数の平均は1ユーザあたりの平均投稿数と強い正の相関があると考えられる。実際に両者の正準相関係数を求めると、「PHP」、「行政書士・司法書士」、「野球」においてそ

れぞれ 0.979, 0.987, 0.994 であった。すなわち両者はほぼ正比例する。今、両者の関係をプロットすると図9のようになる。また、最小二乗法により観測点を近似する直線を引くと、「PHP」、「行政書士・司法書士」、「野球」のそれぞれで近似直線の傾きは 1.168, 1.078, 1.266 と、ほぼ同じような値になった。したがって、エゴネットワークの周辺ユーザ数は1ユーザあたりの平均投稿数のみによって決定されると考えられる。

一方、1ユーザあたりの平均投稿数と“discussion person”の割合の両者の正準相関係数は「行政書士・司法書士」および「野球」ではそれぞれ 0.917, 0.946 と強い正の相関があったが、「PHP」では 0.312 となった。実際に、図7と図8における「PHP」に関するデータを比較すると、2004年5月頃から2007年5月頃まで、1ユーザあたりの平均投稿数は変化しないが、“discussion person”の割合は減少していることがわかる。一方、「行政書士・司法書士」「野球」の両カテゴリにおいてこのような現象はみられない。本現象がコミュニケーションタイプと何らかの因果関係があるか、さらなる分析を行う必要がある。

また、“discussion person”の割合はひじょうに時間変化が大きい。たとえば、2004年までは「PHP」においても“discussion person”の割合が多いが、2005年4月からは「行政書士・司法書士」の割合が多くなる。また、2007年4月ほどから「野球」の“discussion person”の割合が急増する。このことは「野球」カテゴリにおけるユーザ数・投稿数の急増に起因すると考えられる(2007年3月までのユーザ数、投稿数は1,827人、2,890件だが、2007年8月には3,084人、7,100件になる)。このように、ユーザ数・投稿数が十分でない状態では、“discussion person”の割合が異なるふるまいをすることがある。このことは、単純な1時点における“discussion person”の割合ではコミュニケーションタイプ分類の素性としては不十分で、時系列分析の必要性を示唆している。

Adamicら[1]は、Yahoo Answersの各カテゴリに対しエゴネットワーク分析を行うことにより、「議論」、「相談」、「知識交換」の順でエゴネットワークの周辺ユーザの数が多く、また「知識交換」に対して「相談」「議論」カテゴリの方が“answer person”と“discussion person”の割合が高いことを、エゴネットワークの可視化により示した。しかしながら、本分析によりエゴネットワークの周辺ユーザ数は1ユーザあたりの平均投稿数により決定され、コミュニケーションタイプとは独立であることが判明した。また、“discussion person”の割合に関しても、時間変化が大きいため、時系列を踏まえたより深い分析が必要ながわかった。

6. まとめと今後の課題

本研究では、1999年から2007年の教えて!gooのデータを用いて、各カテゴリを「知識交換」「相談」「議論」という3種類のコミュニケーションタイプに分類した上で、Yahoo Answersの同様の分析結果と比較した。その結果、Adamicらの用いた素性では教えて!gooに対しては3つのコミュニケーションタイプを適切に分類できないことがわかった。また、post length (1回答あたりの平均語数)に関しては、Yahoo Answersでは「相談」「議論」よりも「知識交換」の方がpost lengthが長い、教えて!gooでは「相談」「議論」よりも「知識交換」の方が逆にpost lengthが

短いことがわかった。

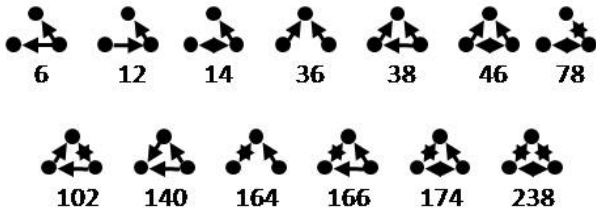


図 10 QA ネットワーク分析のためのモチーフ
Fig.10 Network motifs for analysis on QA network

また、教えて！goo におけるコミュニケーションの時間変化に着目し、QA ネットワークの成長パターンについて調査した。特に、本研究ではエゴネットワークの時系列を分析した。その結果、コミュニケーションタイプの分類の素性として QA ネットワークを用いるには時系列の概念の導入が必要であることを示した。

コミュニケーションタイプ分類のための適切な素性を検討するために、今後以下を課題として研究を続けていく：

- 教えて！goo における各カテゴリについて、3 つの尺度 (post length, thread length, asker/replier overlap) および別の尺度を用いて特徴量を抽出し、クラスタリングを行った結果とコミュニケーションタイプの関係性について分析を行う。
- QA ネットワークの成長分析について、他のさまざまな手法を用いて多角的に分析を行い、コミュニケーションタイプ分類の素性として適切なものを発見する。特に、モチーフの時系列分析について検討している。図 10 に示すモチーフが、過去の時点の QA ネットワークにてどのくらいの頻度部分グラフとして出現しているのかについて分析する。また、同じ 3 ユーザが別のタイプのモチーフに変化することも考えられる(たとえば図 10 における 36 から 38 へ変化する等)。
- QA ネットワークの成長パターンの分析結果が何らかのカテゴリないしユーザの性質抽出に利用できないか検討を行う。

[文献]

[1] L.A. Adamic, J. Zhang, E. Bakshy, and M.S. Ackerman. Knowledge sharing and Yahoo Answers: Everyone knows something. Proceedings of the 17th international conference on World Wide Web, 2008.

[2] S. Whittaker, L. Terveen, W. Hill, and L. Cherny. The dynamics of mass interaction. Proceedings of the 1998 ACM conference on Computer supported cooperative work, pp. 257-264, 1998.

[3] D. Fisher, M. Smith, and H.T. Welsler. You Are Who You Talk To: Detecting Roles in Usenet Newsgroups. Proceedings of the 39th Hawaii International Conference on System Sciences, 2006.

[4] T.C. Turner, M.A. Smith, D. Fisher, and H.T. Welsler. Picturing Usenet: Mapping Computer-Mediated Collective Action. Journal of Computer-Mediated Communication, Vol. 10, No. 4, 2005.

[5] K. Zhongbao and Z. Changshui. Reply networks on a bulletin board system. Physical Review E, Vol. 67, No. 3, p. 36117, 2003.

[6] E. Wenger. Communities of Practice: Learning, Meaning, and

Identity. Cambridge University Press, 1999.

[7] J. Preece, B. Nonnecke, and D. Andrews. The top five reasons for lurking: improving community experiences for everyone. Computers in Human Behavior, Vol. 20, No. 2, pp. 201-223, 2004.

[8] J.S. Donath. Identity and Deception in the Virtual Community. Communities in Cyberspace, pp. 29-59, 1999.

[9] H. Welsler, E. Gleave, D. Fisher, and M. Smith. Visualizing the Signatures of Social Roles in Online Discussion Groups. The Journal of Social Structure, Vol. 8, No. 2, 2007.

[10] W. Sack. Conversation map: a content-based Usenet newsgroup browser. Proceedings of the 5th international conference on Intelligent user interfaces, pp. 233-240, 2000.

[11] J. Arguello, B.S. Butler, L. Joyce, R. Kraut, K.S. Ling, and X. Wang. Talk to me: foundations for successful individual/group interactions in online communities. Proceedings of the SIGCHI conference on Human Factors in computing systems, pp. 959-968, 2006.

[12] E. Joyce and R.E. Kraut. Predicting Continued Participation in Newsgroups. Journal of Computer-Mediated Communication, Vol. 11, No. 3, pp. 723-747, 2006.

[13] J. Cho, S. Roy, and R.E. Adams. Page quality: in search of an unbiased web ranking. Proceedings of the 2005 ACM SIGMOD international conference on Management of data, pp. 551-562, 2005.

[14] 今藤紀子, 喜連川優. ウェブコミュニティ出現におけるリンク構造成長パターン分析. DBSJ Letters, Vol. 4, No. 1, 2005.

[15] A.L. Barabasi and R. Albert. Emergence of Scaling in Random Networks. Science, Vol. 286, No. 5439, p. 509, 1999.

[16] MeCab: Yet Another Part-of-Speech and Morphological Analyzer. <http://mecab.sourceforge.net/>.

[17] 宮原哲. 入門コミュニケーション論. 松柏社, 東京, 1992.

甲谷 優 Yutaka KABUTOYA

NTT サイバーソリューション研究所 所属。2008 年京都大学大学院情報学研究科博士前期課程修了。同年、日本電信電話株式会社入社。Web マイニングの研究開発に従事。情報処理学会, 日本データベース学会, 言語処理学会, 各会員。

川島 晴美 Harumi KAWASHIMA

NTT サイバーソリューション研究所 主任研究員。1990 年山梨大学大学院工学研究科博士前期課程修了。同年、日本電信電話株式会社入社。現在インターネットからの話題情報抽出技術の研究開発に従事。電子情報通信学会会員。

藤村 考 Ko FUJIMURA

NTT サイバーソリューション研究所 主幹研究員。電気通信大学大学院情報システム学研究科客員教授。1989 年北海道大学大学院工学研究科博士課程修了。同年、日本電信電話株式会社入社。トランザクション処理記述言語, 汎用電子チケットシステム, 電子決済システム, blog マイニングの研究開発に従事。博士(工学)。情報処理学会, 電子情報通信学会, 日本社会情報学会, 各会員。

奥 雅博 Masahiro OKU

NTT サイバーソリューション研究所 主幹研究員。1984 年大阪府立大学大学院工学研究科博士前期課程修了。同年、日本電信電話公社(現 NTT)入社。機械翻訳, 日本文推敲支援技術等の自然言語処理, 検索をはじめとするポータルサービスの研究開発に従事。博士(工学)。情報処理学会, 電子情報通信学会, 言語処理学会, 各会員。