

質問回答サイトにおける QARank を用いたユーザ貢献度の推定法

QARank: An Analysis Method of User Contributions in a QA Community

佐藤弘樹¹ (さとう ひろき・Hiroki SATO) ・ 島田諭² (しまだ さとし・Satoshi SHIMADA)
伏見卓恭³ (ふしみ たかやす・Takayasu FUSHIMI) ・ 福原知宏⁴ (ふくはら ともひろ・Tomohiro FUKUHARA)
斎藤和巳⁵ (さいとう かずみ・Kazumi SAITO) ・ 佐藤哲司⁶ (さとう てつじ・Tetsuji SATOH)

筑波大学大学院図書館情報メディア研究科 {¹博士前期課程, ²博士後期課程, ⁶教授}

静岡県立大学大学院経営情報学研究科 {³博士前期課程, ⁵教授}

⁴東京大学人工物工学研究センター 特任助教

[Abstract]

Knowledge sharing communities are quite rapidly spread on the World Wide Web (WWW). Yahoo! Chiebukuro (YC) is one of the largest question-answer(QA) sites in Japan. This paper proposes an evaluation method of degree of user contributions in QA sites. This method is called "QARank". QARank values of question messages in YC are estimated. We applied QARank to the network of the first answer(FA) users who replied quickly to a question, and the best answer(BA) users who wrote the most useful answer in each question. Then, QARank was examined and compared with random networks built on the FA and the BA network. As a result, QARank figured out the properties of users depending on categories in YC.

[キーワード]

QARank, ユーザ貢献度, 質問回答サイト, ネットワーク分析

1. はじめに

現在, Web 上には多数の質問回答サイトが存在し, 日本語で投稿する質問回答サイトに限っても Yahoo! 知恵袋^{*1} や OKWave^{*2}, 教えて!goo^{*3}, MSN 相談箱^{*4}など多くのサイトが活況を呈している. 質問回答サイトとは, あるユーザが投稿した質問に他のユーザが回答を行うことで, 様々な知識を多くの人たちで共有を目的とした知識共有コミュニティの一形態である. 質問回答サイトでは, 各ユーザが興味を持っていることに対する疑問や自分が持っている知識を生かした投稿活動によってコミュニティが形成されている. これらコミュニティは各々が異なる性質を持ち, その性質の形成に様々な要因が関係していると考えられる. その要因を特定することは円滑なコミュニティ運営にとって重要である反面, これまでは運営者の主観的な観察に頼らざるを得なかった.

筆者らはこれまでに, 質問回答サイトに参加するユーザの性質, 及びユーザの行動によって得られるコミュニティの性質を調査してきた [1, 2]. 本論文は, これまでに得られた知見に基づいて, 質問回答サイトにおけるユーザの貢献度を推定する手法: QARank を提案する. QARank はコミュニティにおけるユーザの貢献度の評価値計算法であり, この評価値によってユーザの貢献度を定量化する.

本論文は, 関連する先行研究の概観, 質問回答サイトとそのネットワーク構造, 提案する QARank 手法の詳細, 提案手法による分析結果, およびコミュニティの性質に関する考察によって構成される.

*¹ <http://chiebukuro.yahoo.co.jp/>

*² <http://okwave.jp/>

*³ <http://oshiete.goo.ne.jp/>

*⁴ <http://questionbox.jp.msn.com/>

2. 関連研究

オンラインコミュニティの分析に関する研究が数多く発表されている。コミュニティの成功要因を分析した研究に Bourhis ら[3] がある。コミュニティの中で率先して発言するなどリーダーシップを発揮するユーザの存在が、そのコミュニティの成功と関連があることを明らかにしている。また、津川ら[4] は、ユーザのソーシャルネットワークが成すトポロジ構造からユーザのリーダーシップ度を推定し、オープンソースソフトウェアの開発者および利用者コミュニティの成功度合いを推定する指標を提案している。野村ら[5] は、掲示板のスレッドにおいて初期に投稿するユーザの行動から、将来の盛り上がり予測する手法を提案している。この手法は、あるスレッドにおいて一定時間内に多数の発言が見られた場合に、その直前に投稿したユーザに高いスコアを与える。スコアの高いユーザの行動から、将来の発言の増加を予測する研究である。松村ら[6] は、掲示板における議論を主導し、他のユーザに影響を与えるユーザを *Influencers* として定義し、リンクと共起語を介する影響力の伝播をモデル化した IDM 手法を提案している。提案法をブログに適用したところ、良好な結果が得られたと報告している[7]。

質問回答サイトを対象とした研究も幅広く取り組まれている。三浦ら[8] は、Yahoo! 知恵袋の利用者にアンケートによる意識調査を行い、利用者の行動パターンや属性について報告している。甲谷ら[9] は、質問者と回答者のネットワークの成長パターンを分析している。また、甲谷ら[10] は、ネットワークの成長パターンに基づく記事推薦の方法も提案している。Adamic ら[11] は、Yahoo! Answers を対象として、カテゴリを様々な観点から分析している。その結果として、Yahoo! Answers の質問者の知識レベルやカテゴリ内で行われるコミュニケーションのタイプを明らかにしている。Zhang ら[12] は、Java Forum 参加者の投稿関係に着目し Expertise Network を構築、PageRank や Hits アルゴリズムを適用して分析を行っている。その結果、これらのアルゴリズムが Expertise Network の評価に有効であること、参加者の専門レベルや、コミュニティの特性を明らかにできることを示している。

本研究は、コミュニティの参加者の役割とその貢献度に着目し、参加者間のネットワークを形成したことに新規性がある。統計的分析手法やネットワーク分析手法を適用することで大域的な貢献度や役割を評価する。

3. ユーザ貢献度の推定手法: QARank

筆者らは、質問回答サイトに参加するユーザの性質、及びユーザの行動によって得られるコミュニティの性質について調査してきた[1, 2]。Yahoo! 知恵袋を対象とした調査・分析によって、質問に対する第一回答(FA: First Answer)やベストアンサー(BA: Best Answer)となる回答者の分布が、「PC」や「恋愛相談」などのカテゴリに依存して異なっていることを明らかにした。また、FA や BA の投稿がカテゴリを性質付ける重要なユーザの役割の一つであることも明らかとなった。しかしながら、これまでの分析では、このようなユーザの中でより重要なユーザの存在までは考慮していなかった。例えば、同一カテゴリにおいて、100 件の回答が付いた質問で BA となったユーザ A と、5 件の回答が付いた質問で BA となったユーザ B がいたとする。この時、より多くの競合がいる中で BA を獲得したユーザ A の方が、そのカテゴリにおける貢献度が大きいとするのが妥当と考えられるが、これまでの研究では同一視していた。

本論文では、質問回答サイトにおける各ユーザの貢献度を推定する手法として QARank を提案する。以下に、提案手法の概念について述べる。初めに、質問回答サイトにおける質問・回答関係から、ユーザの貢献度を推定するための QA ネットワークを構築する手法を述べる。次に、構築したネットワークにおける評価値計算手法、すなわち本論文で提案する QARank の算出法を詳説する。

3.1. QA ネットワーク

質問回答サイトにおいて、ユーザの集合を $U = \{u_n: n=1, \dots, N\}$ とし、質問の集合を $Q = \{q_m: m=1, \dots, M\}$ とする。ユーザ u_n が質問 q_m に回答したことを (u_n, q_m) で表せば、これらを要素とする回答の集合 A を定義できる。ここで、ユーザ集合 U と質問集合 Q の各要素をノードとし、回答集合 A の各要素 (u_n, q_m) をノード u_n からノード q_m への有向リンクとみなせば、 $G=(U \cup Q, A \cup BA)$ で規定する 2 部グラフを構築できる。このような質問と回答者の関係を表した QA ネットワークを図 1(a) に示す。

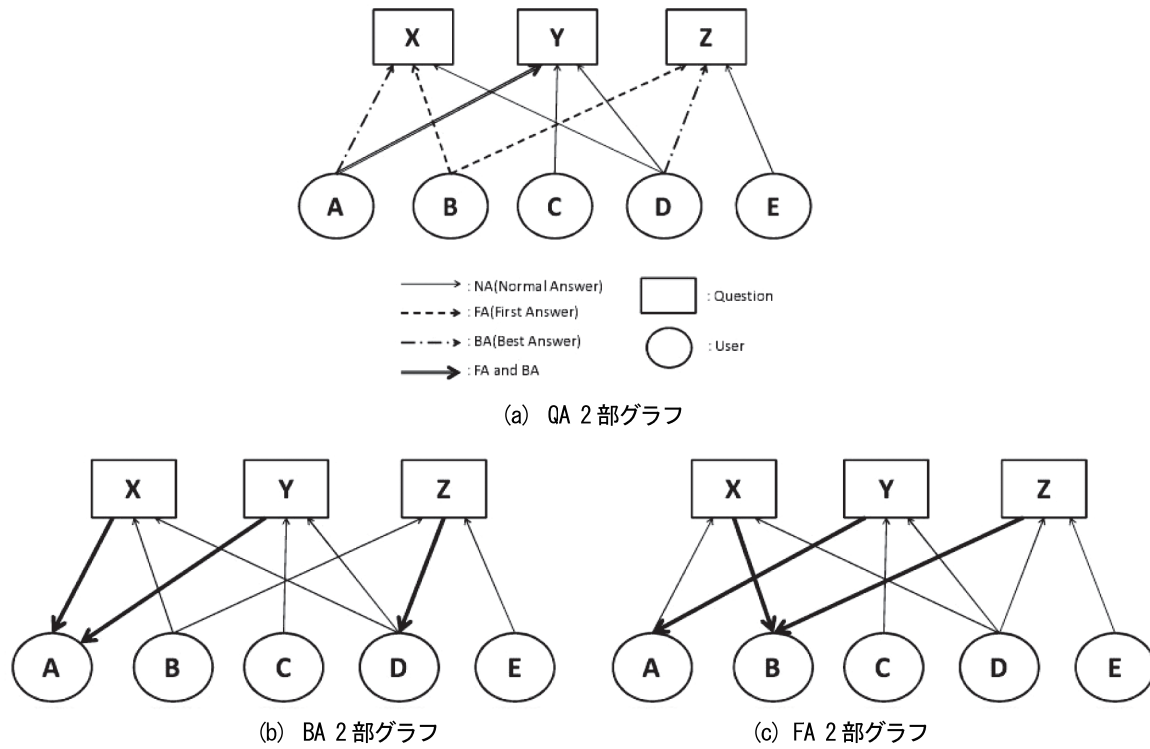


図1 QA ネットワーク

質問回答サイトでは、各質問 q_m に対して、BA (Best Answer) およびFA (First Answer) に相当するユーザがそれぞれ指定される。いま、 q_m に対するBA ユーザを $BA(q_m) \in U$ とすれば、 $(q_m, BA(q_m))$ を要素とするBA 集合を定義できる。ここで、BA 集合の各要素 $(q_m, BA(q_m))$ をノード q_m からノード $BA(q_m)$ への有向リンクとみなせば、 $GBA=(U \cup Q, A \cup BA)$ で規定するBA2 部グラフを構築できる (図 1(b))。同様に、 q_m に対するFA ユーザを $FA(q_m) \in U$ とすれば、 $(q_m, FA(q_m))$ を要素とするFA 集合を定義でき、 $GFA=(U \cup Q, A \cup FA)$ で規定するFA2 部グラフを構築できる (図 1(c))。以下では、BA2 部グラフについてのみ述べるが、FA2 部グラフに対しても同様の議論を展開できる。

3.2. QARank

各質問 q_m に対して、ある種の評価値 (重要度など) $v(q_m)$ が与えられているとする。これら評価値は、BA を行ったユーザに加えられるものとする。つまり、ユーザ u_n が BA となった質問部分集合を $BA(u_n) \subset Q$ とすれば、質問評価値 $v(q_1), \dots, v(q_M)$ に対し、ユーザ評価値 $v(u_1), \dots, v(u_N)$ のそれぞれを次式で定義する。

$$v(u_n) = \sum_{q_m \in BA(u_n)} v(q_m) \quad (1)$$

逆に、各ユーザ u_n に対して、ある種の評価値 (貢献度など) $v(u_n)$ が与えられているとする。これら評価値は、回答した質問に重み付けして加えるものとする。つまり、質問 q_m に回答したユーザ部分集合を $QA(q_m) \subset U$ とすれば、ユーザ評価値 $v(u_1), \dots, v(u_N)$ に対し、質問評価値 $v(q_1), \dots, v(q_M)$ のそれぞれを次式で定義する。

$$v(q_m) = \sum_{u_n \in QA(q_m)} w(u_n)v(u_n) \quad (2)$$

ここで、 $w(u_n)$ は上述した重み付けを表す。典型的な $w(u_n)$ の設定法として、本論文では、ユーザ u の回答数の

逆数を用いる。すなわち、回答数が少ないユーザの重みを高くする。ユーザ u_n が回答した質問部分集合を $QA(u_n) \subset Q$ とすれば、 $w(u_n)$ は以下で与えられる。

$$w(u_n) = \frac{1}{|QA(u_n)|} \quad (3)$$

ここで、 $|QA(u_n)|$ は集合 $QA(u_n)$ の要素数である。この設定で、以下の条件を仮定する。

$$\sum_{n=1}^N v(u_n) = 1, v(u_n) > 0 \quad (4)$$

このとき、以下が成立する。

$$\sum_{m=1}^M v(q_m) = \sum_{m=1}^M \sum_{u_n \in QA(q_m)} \frac{1}{|QA(u_n)|} v(u_n) = \sum_{n=1}^N |QA(u_n)| \frac{1}{|QA(u_n)|} v(u_n) = 1 \quad (5)$$

一方、上述した計算式に従い、質問評価値 $v(q_1), \dots, v(q_M)$ からユーザ評価値 $v(u_1), \dots, v(u_N)$ を計算すれば、以下が成立する。

$$\sum_{n=1}^N v(u_n) = \sum_{n=1}^N \sum_{q_m \in BA(u_n)} v(q_m) = \sum_{m=1}^M v(q_m) = 1 \quad (6)$$

よって、上述した計算式による評価値の更新は、マルコフ連鎖での確率更新式と見なすことができる。すなわち、PageRank[13] でのランク値更新と同様に、定常状態の確率ベクトルを得ることができる。以下では、この設定に基づく評価値計算法を QARank と称する。

QARank の実際の適用では、ランキング結果の安定性を考慮し、PageRank と同様に、一様なランダムジャンプを導入する。詳細には、式(1) と式(2) を次のように拡張して用いる。

$$v(u_n) = \frac{\epsilon}{N} + (1 - \epsilon) \sum_{q \in BA(u_n)} v(q) \quad (7)$$

$$v(q_m) = \frac{\epsilon}{M} + (1 - \epsilon) \sum_{u \in QA(q_m)} w(u) v(u) \quad (8)$$

ここで、 ϵ は一様ジャンプ確率を表し、標準的な設定値として $\epsilon=0.15$ とする。

4. QARank を用いたユーザ貢献度の推定

4.1. 評価に使用したデータセット

国立情報学研究所を通じて提供されている Yahoo! 知恵袋データ^{*5} を用いてユーザの QARank を算出し、ユーザの貢献度推定における QARank の有用性を評価する。Yahoo!知恵袋では、約 300 のカテゴリが設けられ、すべての投稿はいずれか 1 つのカテゴリに属する。また、質問と回答は同じカテゴリに属し、投稿後にカテゴリを変更することはできない。分析対象データのカテゴリ、ユーザ数、および投稿数を表 1 に示す。これらの数は、提供されているデータの全体にあたる 2004 年 4 月から 2005 年 10 月までの期間における投稿の統計である。投稿数において上位となる「恋愛相談、人間関係の悩み」、「Yahoo! オークション」、「パソコン、周辺機器」、「政治、

^{*5} <http://research.nii.ac.jp/tdc/chiebukuro.html>

表 1 分析対象データのカテゴリ、ユーザ数、および投稿数

カテゴリ名	異なり回答者数	質問記事数	回答記事数
恋愛相談, 人間関係の悩み	62,717	210,052	1,416,280
Yahoo!オークション	37,727	190,363	1,242,153
パソコン, 周辺機器	27,420	171,786	474,510
政治, 社会問題	25,766	78,749	403,169

社会問題」(以下,「恋愛」,「オークション」,「PC」,「政治」と記す)について,カテゴリ別および4 カテゴリ全体での分析を試みる.表1において,異なり回答者数,質問記事数,回答記事数は,図1(a)に示すQA ネットワークの User 数, Question 数, リンク数と対応している.

本論文では,BA 2 部グラフと FA 2 部グラフに対する, QARank の特性を明らかにするため,2 種類のランダム 2 部グラフを考える.質問 q_m に回答したユーザ集合を $QA(q_m) \subset U$ で定義する.明らかに, $BA(q_m) \subseteq QA(q_m)$ である.第一のランダムグラフでは, $QA(q_m)$ からランダムに要素を 1 つ選択して BA と見なして構築したグラフである.以下では,このグラフを GR グラフと呼ぶ.

第二のランダムグラフでは,どのユーザがどの質問に回答したかもランダムに入れ替えて構築する.ただし,各ユーザが回答した質問数とともに,各質問に回答したユーザ数は,元の BA 2 部グラフと同じままとするグラフである.このグラフの詳細な構築法は以下となる.

1. 順番に回答 (u_n, q_m) を一つ選定する.
2. 選定された回答に対して,別の回答 (u_n, q_m) をランダムに選択する.
3. 2 つの回答ペアに対して,回答した質問を入れ替え, (u_n, q_m) と (u_n, q_m) を新たな回答ペアとする.

明らかに,この処理で各ユーザと各質問の回答数(回数)は不変である.また,このような入れ替えにより,元の BA の意味は殆どなくなるので,GR グラフと同様に,新たにできた $QA(q_m)$ からランダムに要素を 1 つ選択して BA と見なす.以下では,このグラフを GR2 グラフと呼ぶ.

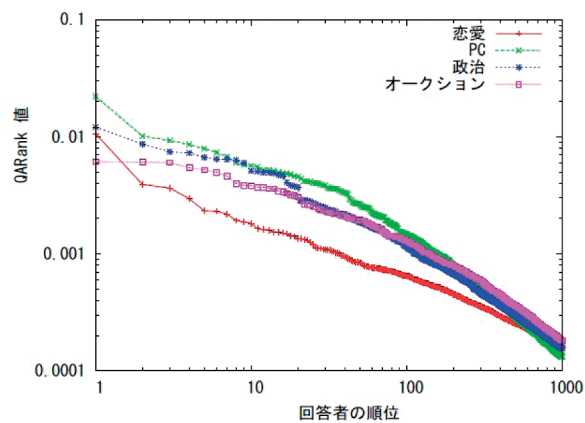
上記 2 種類のランダムグラフを表 1 に示した 4 カテゴリに対してそれぞれ生成する.このとき,ランダムグラフのノード数,リンク数,密度は対応するカテゴリの BA/FA 2 部グラフと等しい.

4.2. カテゴリ間の比較

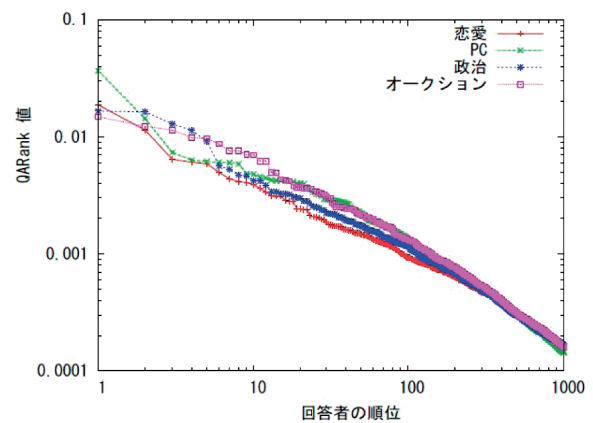
BA 2 部グラフと FA 2 部グラフにおける QARank のカテゴリ間の比較を行なう.BA 2 部グラフの結果を図 2(a), FA 2 部グラフの結果を図 2(b) に示す.各図において,縦軸は回答者の QARank 値,横軸は回答者の順位となり, QARank 値の高い順に 1,000 人を抽出しプロットしている.

図 2(a) の BA 2 部グラフにおいて,恋愛カテゴリは 1 位の回答者の QARank 値が高く下位との差がある.しかし,恋愛カテゴリを 1,000 位までの全体で見た場合, QARank のカーブは緩やかに減少しており QARank 値の減少幅は 4 カテゴリの中で 1 番小さい.一方,PC カテゴリでは,他のカテゴリよりも上位 100 位では QARank 値が高くなっているが,100 位を境目として急な減少を見せ,結果的に 1,000 位では 4 カテゴリで最も QARank 値が低くなっている.また政治カテゴリとオークションカテゴリでは,上位の QARank 値に大きな差はなく,それ以降もまた緩やかな減少となっている.

図 2(b) の FA 2 部グラフにおいて,上位 10 位にカテゴリ間の差が見られる.PC カテゴリでは,トップの QARank 値が高く上位 10 位の中でも上位と下位で差が大きくなっている.この傾向は,恋愛カテゴリでも PC カテゴリほど顕著ではないが観察される.この 2 つのカテゴリに対して,政治カテゴリとオークションカテゴリでは,上位 10 位の中で QARank 値はほぼ横ばいとなっている.特に,オークションカテゴリでは 10 位以降においても大きな差は見られない.しかし,FA 2 部グラフでは BA 2 部グラフほどカテゴリ間に差は見られない.1000 位までの全体で見たとき,どのカテゴリにおいても上位 10 位の値が高く,それ以降では緩やかに一定の傾向を保って減少しているという点で共通している.

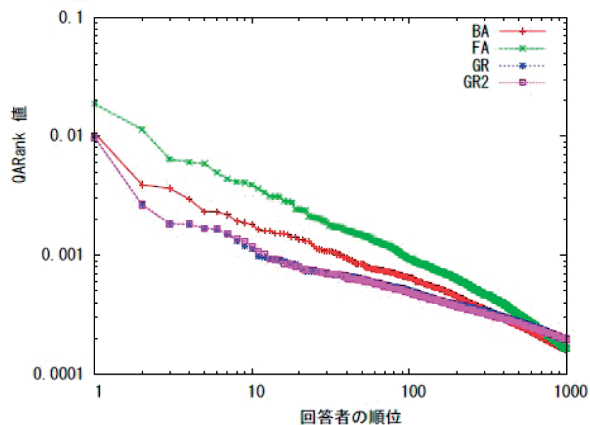


(a) BA 2部グラフ

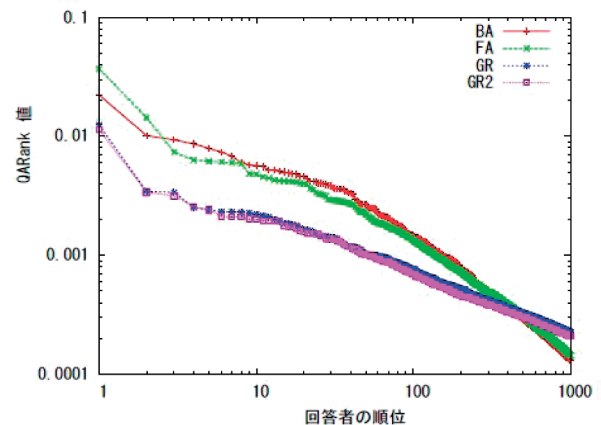


(b) FA 2部グラフ

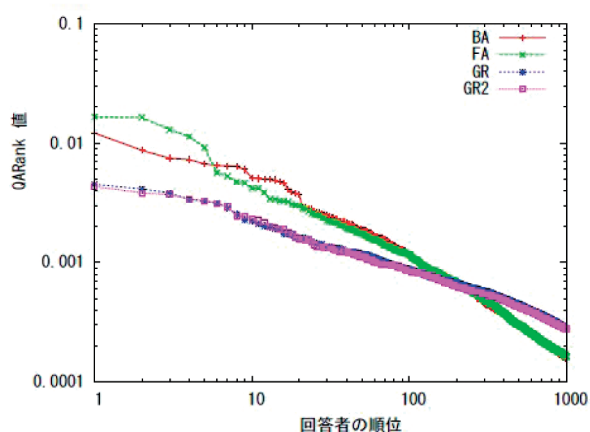
図2 QARank 値のカテゴリ間比較



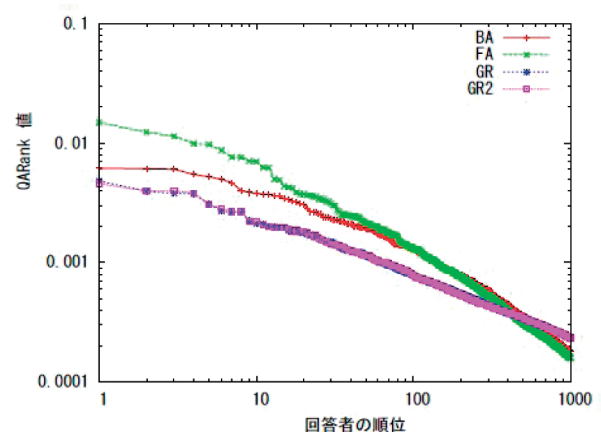
(a) 恋愛カテゴリ



(b) PC カテゴリ



(c) 政治カテゴリ



(d) オークションカテゴリ

図3 評価対象カテゴリの各グラフ (BA, FA, GR, GR2) のQARank 値比較

4.3. ランダムグラフとの比較

BA2 部グラフと FA2 部グラフに対する QARank の特性を見るため、2 種類のランダム 2 部グラフである GR グラフと GR2 グラフとの比較を行った。各カテゴリの BA2 部グラフと FA2 部グラフ、そしてランダムグラフとの比較結果を図 3(a) から図 3(d) に示す。各図において、QARank 値の高い順に 1,000 人を抽出しプロットしている。

図 3(a) の恋愛カテゴリでは、ランダムグラフと BA, FA を比較したとき、BA は比較的ランダムグラフに近い QARank 値となっている。それに対して、FA はランダムグラフよりも明らかに高い値となっており、さらに 4 カテゴリの中で最も BA と FA が離れている。図 3(d) のオークションカテゴリでは、恋愛カテゴリと似た傾向が見られる。オークションカテゴリでも BA とランダムグラフの QARank 値が近く、FA とは離れている。しかし、恋愛カテゴリとオークションカテゴリではグラフのカーブの形状が異なっている。

図 3(b) の PC カテゴリと図 3(c) の政治カテゴリでは上記 2 カテゴリとは傾向が異なる。PC カテゴリと政治カテゴリでは BA とランダムグラフとでは QARank 値が離れている。そして、BA と FA の QARank 値が近くなっている。特に、PC カテゴリでは、BA と FA の QARank 値がほぼ一致しており、4 カテゴリの中で最も BA と FA が近いと言える。また全てのカテゴリに共通して、FA の QARank 値が各図の中で最も高く、ランダムグラフと離れている。

5. 考察

5.1. First Answer の特性

評価を行った 4 カテゴリにおいて、FA2 部グラフの QARank 値にカテゴリ間の差異はほとんど見られなかった。これは、質問に早く回答を行うユーザは、カテゴリに依存せずに存在していることを示している。ランダムグラフとの比較実験の結果から、いずれのカテゴリにおいても、FA とランダムグラフを比較すると、QARank 値が離れていた。ランダムグラフと QARank 値が離れることは、FA を獲得するユーザにランダム性がないことを示している。すなわち、ある特定のユーザは、質問に対して早い回答を行っており、このようなユーザは、カテゴリに依存せず一定の比率で存在していると考えられる。

5.2. Best Answer の特性

BA 2 部グラフの QARank 値をカテゴリ間で比較する。PC カテゴリは QARank 値の減少幅が大きくグラフのカーブ形状が急峻となっているのに対し、恋愛カテゴリは QARank 値の減少幅がわずかでカーブ形状が緩やかになっている。このことは、PC カテゴリでは BA を獲得するユーザが限られていること、すなわち、質問に対する専門性が要求される PC カテゴリでは、BA の獲得が特定のユーザに偏っていることを表していると考えられる。一方、恋愛カテゴリでは、PC カテゴリに比べ BA が回答者全体に均等に獲得されていると見ることができる。

これらのことは、ランダムグラフとの比較実験の結果からも示唆される。恋愛カテゴリでは、BA2 部グラフとランダムグラフの QARank 値の変化に大きな差が見られないことから、BA のランダム性が高いといえる。また PC カテゴリでは、BA 2 部グラフとランダムグラフの QARank 値が離れているのは、BA にはランダム性が少ない、すなわち BA 獲得にふさわしい内容を書いた回答者が選ばれる傾向が強いと考えられる。

5.3. Best Answer の貢献度に着目したカテゴリのタイプ

Best Answer の特性で述べたように、QARank 値の上位と下位の減少幅は BA ユーザ選定の特定性、つまり BA 獲得ユーザが限られた人かどうかということを示している。また、BA グラフとランダム (GR) グラフの QARank 値の差は、BA ユーザ選定のランダム性、つまり BA 獲得ユーザは質問にふさわしい回答を書いた人が選ばれているかを示している。今回、FA グラフに比べ BA グラフについてカテゴリ間でランク値に異なる特徴が見られた。この点に着目して分類した回答者のタイプを図 4 に示す。



図4 回答者の貢献度 (BA) から見たカテゴリのタイプ

A. 問題解決タイプ:

BA ユーザ選定のランダム性が低く、かつ、特定性が高いのが問題解決タイプのカテゴリである。問題解決タイプのカテゴリでは、解の存在する質問である場合が多く、回答には質問に対する専門性が必要となってくる。そのため、回答するユーザが限定されることで特定性が高くなる。また BA の獲得も、より専門性の高いユーザが選ばれやすくなることでランダム性が低くなる。分析した 4 カテゴリの中では PC カテゴリがこのタイプに該当すると考えられる。

B. 議論タイプ:

BA ユーザ選定のランダム性が低く、かつ、特定性も低いのが議論タイプのカテゴリである。議論タイプのカテゴリでは、誰でも回答しようと思えば投稿できるが、BA の獲得には専門性が必要となる。例えば、スポーツチームの勝敗予想についての質問があった場合、各ユーザは自分の好きなチームについて自由に回答できるので投稿は容易であり、BA ユーザの特定性が低くなる。しかし、BA を獲得するには、他のユーザを納得させなければならず、そのスポーツチームの戦力や現状など根拠に基づいた回答である必要があり、ランダム性は低くなる。分析した 4 カテゴリの中で政治カテゴリは、このタイプの傾向が強いといえる。

C. 相談タイプ:

BA ユーザ選定のランダム性が高く、かつ、特定性が低いのが相談タイプのカテゴリである。このタイプのカテゴリは、投稿の敷居が低く誰でも回答を行え、また BA の選定が専門性に左右されず誰でも獲得できる可能性がある。例えば、人間関係の悩みの相談が質問として投稿された場合、各ユーザは自分の考えや意見・体験を回答できるので投稿の敷居は低い。また、このタイプのカテゴリでは、質問に対する専門性だけでなく、質問者が回答にどれだけ共感できるかも BA を獲得する要件となる場合が多いと思われる。そのため、いずれの回答であっても BA の候補となり、BA ユーザ選定のランダム性は高くなるといえる。分析した 4 カテゴリの中では、恋愛カテゴリがこのタイプに該当すると考えられる。また、オークションカテゴリもこのタイプの傾向が見られる。

D. 井戸端会議タイプ:

BA ユーザ選定のランダム性が高く、かつ、特定性も高いのが井戸端会議タイプである。井戸端会議タイプでは、投稿の敷居が高いため特定少数のユーザによって質問と回答が行われており、この特定少数のコアメンバーの中で持ち回りのように BA が選ばれている。例えば、一部の人の間で人気のあるアニメに関する質問が投稿された場合、このアニメを知っている人は限定されるために特定性が高くなる。また、この回答できる人たちの間では専門性に大きな差異はなく、回答を行った人誰でもが BA 獲得の可能性があるのでランダム性も高くなる。今回調査した範囲には、このタイプに属すると思われるカテゴリは見当たらなかった。

6. まとめと今後の課題

本論文では、知識共有コミュニティにおけるユーザ各々の貢献度を推定する手法として QARank を提案した。QARank はコミュニティにおけるユーザの貢献度の評価計算法であり、QARank によって算出された貢献度は、ユーザを特徴付ける要因となると考えられる。質問とその質問に回答した回答者の関係から、第一回答者およびベストアンサーのネットワークを構築し、構成したネットワークに PageRank アルゴリズムを拡張・適用して QARank 値を計算する。実際に運用されていた Yahoo! 知恵袋データで投稿数が多い4カテゴリに提案法を適用して、参加者の貢献度を分析した。カテゴリ間での QARank 値の比較や、ランダムグラフとの比較を行った結果、BA2 部グラフはカテゴリ間に大きな差異が見られたのに対して、FA2 部グラフではカテゴリ間で大きな差異は見られなかった。またランダムグラフとの比較結果では、恋愛カテゴリの BA とランダムグラフの QARank 値が同程度の値を示し、PC カテゴリの BA とランダムグラフの QARank 値は離れた値を示した。また、FA とランダムグラフの比較では、いずれのカテゴリにおいても QARank 値が離れていることを確認した。以上の結果は、BA 選定がカテゴリによって異なること、また、FA を獲得するユーザの性質の差異を表していると考えられることから、これらの指標を用いたカテゴリのクラス分けを議論した。

今後の課題は、よりユーザ行動に着目した分析を行うことで、カテゴリ間の関係を明確にすることである。また、ブログなどの他の知識共有コミュニティに対しても提案手法が適用できるかを評価することも課題である。

謝辞

本研究の一部は科研費 (21500091) の助成を受けて行ったものである。また、実装・評価に際し、大学共同利用機関法人国立情報学研究所から提供を受けた Yahoo! 知恵袋のデータを利用している。ここに記して謝意を示す。

[参考文献]

- [1] 佐藤弘樹, 島田諭, 福原知宏, 佐藤哲司: コミュニティの活性度評価に関する一検討, 情報社会学会 知識共有コミュニティワークショップ論文集, pp. 29–38, 2008.
- [2] 佐藤弘樹, 島田諭, 福原知宏, 斉藤和巳, 佐藤哲司: 質問回答サイトにおける投稿種別に注目したコミュニティ分析手法, Web とデータベースに関するフォーラム (WebDB Forum) 2009, 2A-3, 2009.
- [3] A. Bourhis, L. Dube, and R. Jacob: The success of virtual communities of practice: The leadership factor. *Electronic Journal of Knowledge Management*, Vol. 3, pp. 23–34, 2005.
- [4] 津川翔, 杉山浩平, 大崎博之, 今瀬真: ソーシャルネットワークのトポロジ構造を用いた開発型オンラインコミュニティの成功度推定手法, 電子情報通信学会 技術研究報告 (IN2008-61), pp. 105–110, 2008.
- [5] 野村拓也, 小野稔, 下山洋一, 高木友博: 目利き人を利用した電子掲示板の盛り上がり予測, 電子情報通信学会 Webインテリジェンスとインタラクション研究会(WI2), WI2-2008-58, pp. 31–35, 2008.
- [6] 松村真宏, 大澤幸生, 石塚満: テキストによるコミュニケーションにおける影響の普及モデル, 人工知能学会論文誌, Vol. 17, No. 3, pp. 259–267, 2002.
- [7] N. Matsumura, H. Yamamoto, and D. Tomozawa: Finding influencers and consumer insights in the blogosphere, *International Conference on Weblogs and Social Media (ICWSM2008)*, pp. 76–83, 2008.
- [8] 三浦麻子, 川浦康至, 地福節子, 大瀧直子, 岡本真: 知識共有コミュニティを創り出す人たち, 第20回人工知能学会全国大会, 3D3-1, 2006.
- [9] 甲谷優, 川島晴美, 藤村孝: QAサイトにおける質問応答グラフの成長パターン分析, 情報処理学会 研究報告, データベース・システム研究会報告, No. 88, pp. 247–252, 2008.
- [10] 甲谷優, 川島晴美, 藤村孝: QAコミュニティの成長パターンに基づく回答者への質問推薦, 日本データベース学会論文誌 (DBSJ Journal), Vol. 8, No. 1, pp. 89–94, 2009.
- [11] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman: Knowledge sharing and yahoo answers: Everyone knows something, *17th International Conference on World Wide Web (WWW2008)*, pp. 665–674, 2008.
- [12] J. Zhang, M. S. Ackerman, and L. Adamic: Expertise networks in online communities: Structure and algorithms, *16th International Conference on World Wide Web (WWW2007)*, pp. 221–230, 2007.
- [13] L. Page, S. Brin, R. Motwani, and T. Winograd: The pagerank citation ranking: Bringing order to the web, *Technical report, Stanford Digital Library Technologies Project*, 1998.

