2N-06

# 技術文書共有サービスにおける，PageRank に基づくユーザの順位付け

中村 大貴 †　　矢吹 太朗
千葉工業大学 社会システム科学部 プロジェクトマネジメント学科 ‡

## 1　序論

ソーシャルメディアの普及により情報の発信・収集は容易となった．技術文書共有サービスもその一例である．

しかし情報発信が容易になった結果，情報が膨大になり適切な情報の取捨選択が困難となる情報オーバーロードが生じている [1].

技術文書共有サービスでは情報オーバーロードの問題に加えて，専門的な文書の内容が良いかどうかを判断するのが難しいという問題がある．他のユーザからの評価によって判断することもできるが，評価するユーザが良い評価をしているのかの判断も困難である．

上記のような問題点がユーザの情報に対する取捨選択を困難にしてしまっている．

## 2　目的

本研究では Qiita（`https://qiita.com/`）を対象として，PageRank に基づくユーザの順位付けによる良い情報を発信するユーザを発見する手法を提案する．

Qiita とはプログラミングに関する知識を記録・共有するためのサービスである．Qiita には，ユーザを評価するための Contribution という指標がある．運営もとによれば Contribution は，「投稿した記事がいいねされる，編集リクエストが受け入れられた記事がその後いいねされる，コメントに対していいねされる，で数値が上がっていく.」というものである [2].

Qiita における，ユーザと「いいね」の関係は，ユーザをノード，「いいね」をエッジとするグラフで表現できる．グラフのノードの重要度の指標の一つに PageRank がある [3]．PageRank は本来，ウェブページのランキングのための指標ではあるが，一般的なグラフでも計算できる．そこで，Qiita のグラフで PageRank を計算し，Contribution と比較することを試みる．

## 3　手法

PageRank に基づくユーザの順位付けを行い，PageRank と Contribution の関係を分析するために以下の手法で研究を行う．

1. API とスクレイピングを用いてユーザ ID とユーザ別のいいねした記事 ID，Contribution のデータを取得する（Contribution を取得できないユーザは本研究の対象外である）．
2. 記事 ID を著者のユーザ ID で置き換える．
3. ユーザ ID をノード，いいねをエッジとし有向グラフを作成し，PageRank を計算する．
4. PageRank と Contribution に対し回帰分析を行う．

## 4　結果

取得したユーザ ID（406056 件）といいねされた記事 ID（67380 件）から有向グラフ（ノード数 : 141581，エッジ数 : 6789904）を作成した．

PageRank を $y$，Contribution を $x$，両者の関係を $y = \beta_1 + \beta_2 x$ とし，線形回帰分析によって $\beta_1, \beta_2$ を求めた結果を表 1 に掲載する．

表 1　パラメータ推定値

| | Estimate | Standard Error | t-Statistic | P-Value |
|---|---|---|---|---|
| 1 | $1.06999 \times 10^{-6}$ | $4.44637 \times 10^{-8}$ | 24.0644 | $1.07771 \times 10^{-127}$ |
| X | $1.04144 \times 10^{-7}$ | $9.14426 \times 10^{-11}$ | 1138.9 | 0. |

表 1 から，「$\beta_1 = 0$」や「$\beta_2 = 0$」という帰無仮説は棄却されることがわかる．

PageRank と Contribution の散布図と，回帰直線 $y = 1.07e-06 + 1.041e-07x$ を図 1 に示す．決定係数は 0.903 となった．

PageRank のヒストグラムを図 2 に，Contribution の

PageRank-based user ranking in technical document sharing services.

† Hiroki NAKAMURA (`s1642093hd@s.chibakoudai.jp`)
‡ Department of Project Management, Faculty of Social Systems Science, Chiba Institute of Technology.



Copyright ©2020 Information Processing Society of Japan.
All Rights Reserved.

ヒストグラムを図 3 に示す．図 2 と図 3 を見比べるとどちらも低い値ほど圧倒的に数が多く，傾きが右肩下がりの直線のような分布となっている．

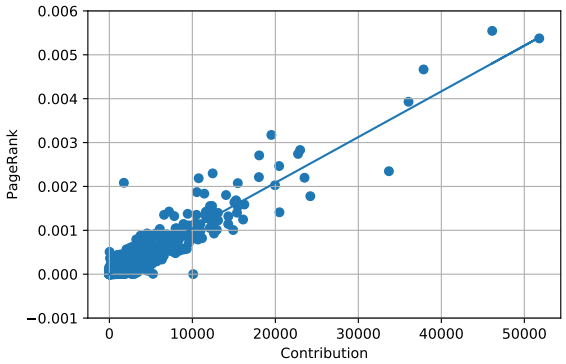


図 1　PageRank と Contribution の散布図と回帰直線

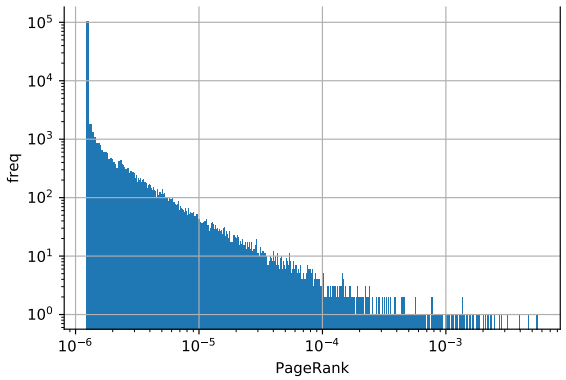


図 2　PageRank のヒストグラム（両対数）

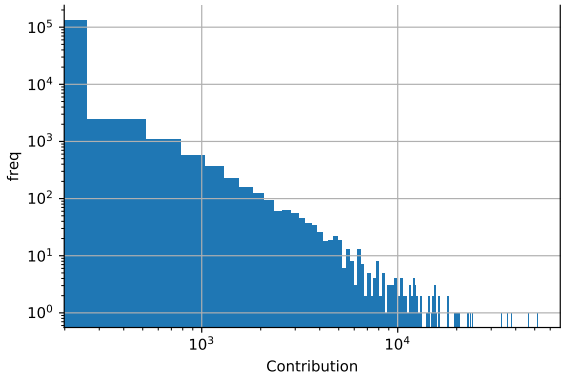


図 3　Contribution のヒストグラム（両対数）

## 5　考察

決定係数が 0.903 であったことから Contribution が PageRank をよく説明できていることが判明した．図 1 を見ても右肩上がりの直線にデータが集まっており，強い正の相関があると言える．

次に，PageRank と Contribution との共通点と相違点について考察する．

共通点についてはどちらも多くのリンクを集めることで順位が上がるということで共通している．これは決定係数の高さにも表れている．

また，図 2 と図 3 からスケールフリー性が見られる．このことから，ユーザはわずかないいね数しか持たない記事よりも多くのいいね数を獲得している著者の記事を優先的にリンクする傾向にあることがわかる．

相違点は PageRank は記事の被リンク以外で数値が上がらないのに対し，Contribution は記事の被リンク以外でも数値を上げる手段があるという点である．この違いが最も顕著な例が図 1 の座標が (10101,0.000001) に位置するユーザである．このユーザは投稿記事数が 0 であるため PageRank は低い．しかし，記事の被リンク以外の方法で Contribution を集めておりデータ取得時点での Contribution ランキングでは 55 位となっている．PageRank は被リンクのみを評価するのに対し，Contribution では被リンク以外での活動も評価する指標であると言える．

## 6　結論

本研究では Qiita のユーザを PageRank に基づいて順位付けした．その結果，PageRank を Contribution で十分に説明可能であり，強い正の相関が確認された．

PageRank と Contribution は，多くのリンクを集めることで順位が上がるという点では共通している．

しかし，Contribution にはそれ以外の活動も評価される要素がある．そのため，両者を比較することで，興味深いユーザを発見することが可能になると期待できる．

## 参考文献

[1] 笹原和俊. フェイクニュースを科学する. 株式会社科学同人, 2017.

[2] Qiita. Contribution とは - qiita:support. https://help.qiita.com/ja/articles/qiita-contribution（2020.01.07 閲覧）.

[3] Amy N.Langville and Carl D.Meyer. Google PageRank の数理. 共立出版, 2009. 岩野和生，黒川利明，黒川洋 訳.



Copyright ©2020 Information Processing Society of Japan.
All Rights Reserved.