

ユーザのコミュニティ分割を用いた Web インタフェース設計のための操作ログ解析 The Operation Log Analysis for Web Interface Design using Dividing User Community.

滝見 優太†
Yuta TAKIMI

鈴木 育男†
Ikuo SUZUKI

山本 雅人†
Masahito YAMAMOTO

古川 正志†
Masashi FURUKAWA

1. 序論

近年、様々なサービスが Web を介して行われるようになってきている。サービスは性能そのものよりもむしろ使いやすさが求められており誰であっても見やすい、操作しやすい Web インタフェースが望まれている。

本研究では WEB インタフェース設計のアプローチとして、既存のシステムのユーザの操作ログから、人によって操作時間に極端な差がある、操作の手順を迷いやすい、などの WEB インタフェースとしての改善点として発見することを目的とする。

ここでユーザのクラスタリングの必要性について記述する。本研究で扱うシステムは一つのシステムで様々なサービスを提供しており、ユーザによって主に使う操作と殆ど使わない操作が分かれている。前述のとおり、最終的には操作フローから個々のユーザ同士を比較する手順があるがこの際に全く異なる操作を主に行うユーザ同士を比較したところで有用な結果が得られるとは到底考えられない。そのため、詳細な比較をする前段階としてまずシステムを利用するユーザを主に使う操作の頻度から分類することで後の作業の効率化を図る。

この目的のための手順として「1.利用するユーザのクラスタリング」、「2. クラスタリングされたユーザごとに操作のフローを作成、解析」、「3. 操作フローの比較から、問題を抱えている操作を推測」のステップに分けて進める。本研究では「1.利用するユーザのクラスタリング」を担当する。

2. 実験で用いる手法

2.1 ネットワーク

今回の実験ではユーザを操作頻度の類似度からユーザ類似度ネットワークを作成し、このネットワークのコミュニティ分割を行うことでクラスタリングを図る。ネットワークの作成方法についてはシステムの使用と併せて後述する。ネットワーク化する利点として一対一の類似度というミクロな視点と、全体として似たものがコミュニティとして現れるというマクロな視点の双方を表現できる点を利点として挙げる。

2.2 コミュニティ分割

ネットワークにおけるコミュニティとはリンクで密に繋がったノードの集合を指す。同じコミュニティに存在するノード同士はリンクを張っている、もしくは少ない次数で

繋がっていることが多い。現実のネットワークでもコミュニティは頻繁に形成され、似た特徴を有するノード同士が集まっているなどの現象が見られる。

コミュニティ分割には Modularity[1]を指標として用いた手法の1つである CNM アルゴリズム[2]を採用した。

このアルゴリズムではまず個々のノードを所属ノード数 1 の仮コミュニティと仮定し、各段階で最もモジュラリティ増分 ΔQ が大きいペアを結合する。結合したペアは新しい仮コミュニティとなるので、各仮コミュニティの結合した場合の増分 ΔQ を再計算する。このようにコミュニティの結合を進めていき、いずれの結合についても ΔQ が負となる(どのペアを結合してもモジュラリティが増加しなくなる)状態になるまで繰り返し、この最終時点での状態をコミュニティ分割結果とする。

3. 実システムデータを用いた実験

3.1 システム使用について

実験ではある会社(以下 A 社と記述)の提供する既存のシステムを用いる。

このシステムは画面に表示されたボタンをクリックすることで画面が切り替わっていく。ボタンを押した際に情報がレコードとしてログに記録されていく。具体的なログについては表 1 のように記録される。

一つのレコードには操作一回分の情報が記録されており、誰が操作を行ったか(ユーザ ID)、行われた操作(操作コード)、行われた時間(操作時間)が記録してある。

今回の実験で用いたログデータは 10 日間分となっており今回の実験では操作回数(総レコード数)の多いユーザ上位 500 名を対象とした。なお、操作コードはこれらのユーザの内誰かが一度でも期間内に操作したものを調査に用いておりコード総数は 405 となった。

3.2 ユーザ間類似度ネットワーク

この情報からユーザ間の類似度ネットワークを生成する。ノードは各ユーザに該当する。対象となるユーザにはレコードを元に操作頻度ベクトルを設定する。操作頻度ベクトルは各ユーザが行った操作の割合である。あるユーザ a の持つ操作頻度ベクトルのある要素 i はユーザ a が操作 i を行った割合と表現できる。

操作頻度から定義した各ユーザ間の類似度をリンクとする。類似度は操作頻度ベクトルの各要素の二乗誤差の総和から算出した値とする。

作成した類似度ネットワークは重み付き無向ネットワークとなる。

†北海道大学 大学院情報科学研究科 複合情報学専攻

Table 1 Record sample

No.	User ID	Code	Time
1	aaaabbbccc	Abc00	2011.04/10 11:00:00
2	aaaabbbccc	Ddc02	2011.04/10 11:00:15
3	eeeeeeeggg	Abc01	2011.04/10 11:30:00
4	eeeeeeeggg	Ddc02	2011.04/10 11:30:10
5	aaaabbbccc	Abc03	2011.04/10 12:00:00

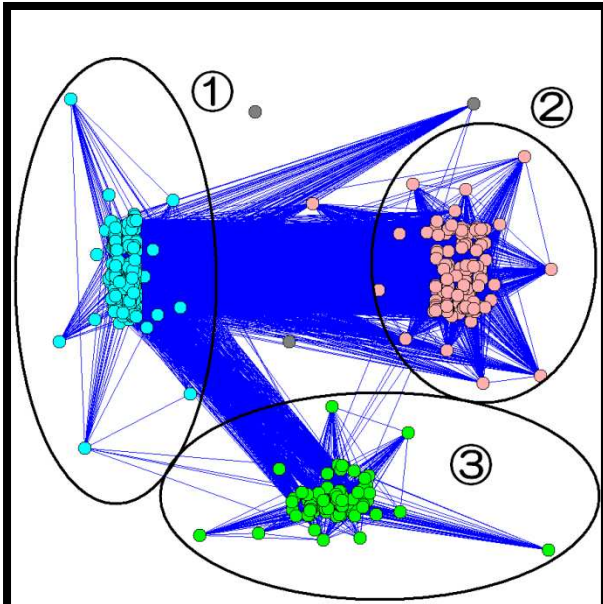


Figure1. Result of dividing community

3.3 実験結果

作成したネットワークに対し CNM アルゴリズムを適用しコミュニティ分割を行った。結果を図1に示す

円が一つのノードを示し、それらを繋ぐ線がリンクを表している。また、わかりやすいように分割されたコミュニティごとに円で囲んで図示し、説明の便宜上コミュニティ番号を丸文字で付与した。

この結果から大きく三つのコミュニティに分かれたことがわかる。またコミュニティ2とコミュニティ3同士はリンクが少なく各コミュニティに属するノード同士は類似度が低いことがわかる。またコミュニティ1は他の二つのコミュニティどちらともリンクが多いことからこれらの中間の性質を持つノードが集まっていることが予測される

3.4 考察

分割された各コミュニティ内の操作頻度平均グラフを図2に示す。横軸が操作コード、縦軸が操作頻度を表し、コミュニティごとに異なる色で分かれている。

今回使用した操作コード数は405と膨大であるため主に使用された操作コードのみを選定し、各コミュニティでの操作頻度平均の和の大きい順に50コードを選択してグラフに表した。

このグラフから各コミュニティの特徴がわかる。コミュニティ2、及び3の所属ユーザはある特定の操作を集中的に行っている。特にコミュニティ3では特定の操作が操作

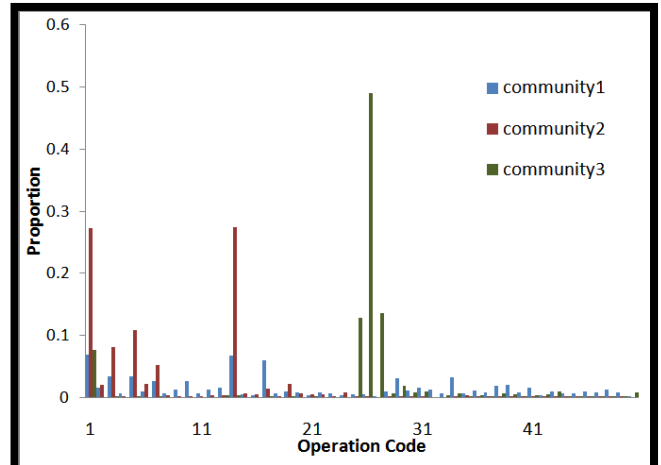


Figure2. Operation code chart in community average

の半分を占めており、扱う操作にかなり偏りがあることが見える。

コミュニティ1は前述の予想通りある程度幅広く操作を行っている。操作コード1から15あたりに見えるような操作頻度の多い部分はコミュニティ2と似ているが、操作コード30以降のコミュニティ3の中でも少なくともはあるが行われている操作が似ており、結果としてどちらのコミュニティとも類似した部分を持つようになっている。

4. 結論

システム利用ユーザの操作頻度からユーザ同士の操作頻度類似度ネットワークを生成、更にネットワークをコミュニティ分割を行った。この結果、分割したコミュニティはそれぞれが異なる特徴を示しており、目的とした主に行っている操作が似ているユーザの集合を抽出することに成功した。

本論文で扱った内容は序論で説明したとおり全体の大きな流れの中の1ステップであり、次のステップでは今回分類された同じ操作を行う傾向にあるユーザ同士を操作フローやある操作にかかる時間といった情報の比較から更に詳細な考察を重ね、最終的な目的であるシステムのWebインタフェースが持つ問題点の発見ができるものと期待する。

参考文献

- [1] M.E.J.Newman, M.Girvan“Finding and evaluating community structure in networks”,Physical.Review.E, 69:026113(2004)
- [2] A..clauset, M.E.J.Newman, C.Moore.“Finding local community structure in very large networks”,Physical.Review.E, 70:066111(2004)