

震災前後における Twitter 上のユーザ行動の変化

Change in User Behavior on Twitter before and after the Disaster

海鋒健太 *¹ 鳥海不二夫 *¹ 大橋弘忠 *¹
 Kenta Kaihoko Fujio Toriumi Hirotada Ohashi

*¹ 東京大学工学系研究科

School of Engineering, The University of Tokyo

Such large disasters as earthquakes and hurricanes are very unpredictable. During a disaster, we must collect information to save lives. Social media attract attention for sharing information, especially Twitter, which is a hugely popular social medium that is now being used during disasters. In this paper, we classify twitter users from thier behaviors(replies, retweets, and tweets) to clarify how people change their behaviors before and after the disaster.

1. 序論

スマートデバイスなどの普及によりインターネット環境が身近になるに伴い、コミュニティ型の会員サービスである SNS などのソーシャルメディアが多くの人に利用され生活に密着したものとなった。ソーシャルメディアの中でも、Facebook*¹ 及びマイクロブログサービス Twitter*² の利用者は世界的にも多く、アクティブユーザが Facebook では 12.8 億人、Twitter では 2 億 4100 万人と莫大なユーザを抱えている。

日本においてもその発展は著しく、株式会社 ICT 総研 [1] によると 2014 年 8 月の調査で日本のインターネット利用者への SNS 普及率は 61% と急速に普及している。このようなソーシャルメディアの普及によりコミュニケーションの利便性は高まっており、その利用方法も個人によってさまざまに変化している。個人間のメッセージ送信といったやり取りだけではなく、社内コミュニケーションの円滑化や重要情報の共有、周知徹底のためにソーシャルメディアを導入する企業も見られる。さらに、ソーシャルメディアは東日本大震災などの緊急時には情報ソースともなりえ、情報拡散や情報取得の有用なツールになるという面ももちあわせている [2]。

一方で、ユーザの利用方法には多様性があり、どのような利用方法をしているユーザが震災情報をどのように扱っていたかを知ることは、今後の震災時のソーシャルメディア利用に役立つと考えられる。

本研究の目的は、ソーシャルメディアの特徴であるユーザ間の関係性の中からユーザ行動を抽出しクラスタリングを行った上で、ユーザがどのように Twitter を利用していたかを明らかにしていく。さらに、その結果を用い東日本大震災時にユーザの行動がどのように変わったのかを明らかにする。

2. GMM を用いたユーザ分類

本研究では、ユーザのコミュニケーション行動に基づいてクラスタリングを行い、ユーザがコミュニケーション構造の観点から Twitter をどのように利用していたのか、またその利用がどのように変化したかを分析する。

ユーザの分類には混合ガウスモデルによるクラスタリング手法 [3] を用いた。

そのために、まず各ユーザの特徴抽出を行う。特に、ここではリツイート行動から作られるリツイートネットワーク及び、リプライ行動から作られるリプライネットワークの双方を用いて、各ユーザのコミュニケーション行動に基づいた特徴を作成する。

本研究で利用するデータは 2011 年 3 月 6 日から 2011 年 3 月 23 日までのメンションに関するツイートのログデータである。ログデータには TwitterID、スクリーンネーム、ツイート内容、タイムスタンプなどが含まれる。

特徴抽出に用いる指標は以下の 21 種類である。

1. Reply 入次数 (被リプライ数):Reply From
2. Reply 出次数 (リプライ数):Reply To
3. Reply クラスタ係数:Reply Clst
4. Reply 相互率:Reply Int
5. Reply 出リンクエントロピ (被リプライ数):Reply FrEN
6. Reply 入リンクエントロピ (リプライ数):Reply ToEN
7. 出隣接 Reply 出次数 (リプライ数):Reply NX-To
8. 入隣接 Reply 入次数 (被リプライ数):Reply NX-Fr
9. 出隣接 Reply 入次数 (リプライ数):Reply NX-FrTo
10. 入隣接 Reply 出次数 (被リプライ数):Reply NX-ToFr
11. Retweet 入次数 (リツイート数):Retweet From
12. Retweet 出次数 (被リツイート数):Retweet to
13. Retweet クラスタ係数:Retweet Clst
14. Retweet 相互率:Retweet Int
15. Retweet 出リンクエントロピ (リツイート数):Retweet FrEN
16. Retweet 入リンクエントロピ (被リツイート数):Retweet ToEN
17. 出隣接 Retweet 出次数 (リツイート数):Retweet NX-To
18. 入隣接 Retweet 入次数 (被リツイート数):Retweet NX-Fr
19. 出隣接 Reply 入次数 (被リプライ数):Retweet NX-FrTo
20. 入隣接 Reply 出次数 (リプライ数):Retweet NX-ToFr
21. 総ツイート数:Tweet

連絡先: 海鋒健太, 東京大学, 東京都文京区本郷 7-3-1,
 kaihoko@crimson.q.t.u-tokyo.ac.jp

*¹ <https://www.facebook.com/>

*² <https://www.twitter.com/>

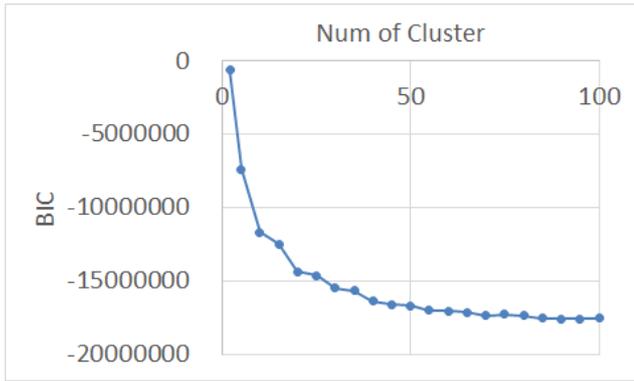


図 1: 解析対象データの各クラスタ数における BIC

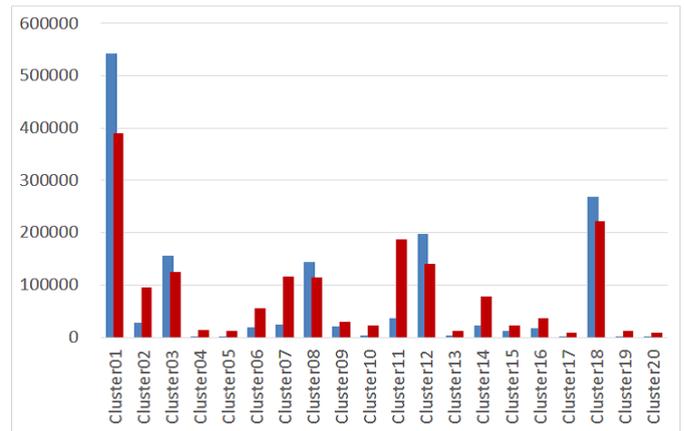


図 2: 2011 年 3 月 6,12 日のクラスタ分布

3. 震災による Twitter 利用方法の変化

3.1 混合ガウス分布によるクラスタリング

本研究では 2011 年 3 月 6-24 日のユーザ群から 1% を抽出し学習用データとして利用し、クラスタリングを行った。これは、震災前後ではユーザの利用の仕方が大きく異なると想定されることから、双方のデータを含めたクラスタリングを行わなければ、偏ったデータとなる可能性が高いと判断されたためである。

適切なクラス多数を決定するため、クラスタ数を変化させたときのベイズ情報量基準 [4] の変化を求めた。その結果を Fig.1 に示す。クラスタ数が 20 以下の時は BIC が改善されているが、20 を超えた時点で BIC の値の変動が小さくなる。そこで、ここではクラスタ数を 20 としてクラスタリングを行った結果を用いて分析を行う。

3.2 得られたクラスタの特徴分析

混合ガウス分布から得られたクラスタについて、各クラスタの代表点の特徴を Table 1 に示す。

得られたクラスタについて見ると、たとえばクラスタ 01 は Reply, Retweet とともにほとんど使われず、Tweet も平均 2.4 回程度で、あまり積極的にツイッターを使っていないユーザ層であると考えられる。逆にクラスタ 05 に所属するユーザは Reply, Retweet とともに積極的に使っており、ヘビーユーザであることがうかがえる。

3.3 震災直後の利用方法の変化

震災前後でユーザの利用方法がどのように変化したかを確認するため、3 月 6, 12 日の各クラスタに所属するユーザ数の分布を Fig.2 に示す。

まず震災前 (3 月 6 日) についてみると、クラスタ 01 に最も多くのユーザが所属していることが分かる。すでに述べた通りクラスタ 01 に所属するユーザは、利用率が低いユーザである。

二番目に多くのユーザが所属するクラスタ 18 も、平均ツイート数が 4.0 と増加し、クラスタ 01 と比べると Reply が使われているユーザ群である。

次に、震災直後の 3 月 12 日についてユーザ分布を見る。震災直後はクラスタ 01 の所属ユーザが減少し、クラスタ 02, クラスタ 07, クラスタ 12 といったクラスタに所属するユーザが増加したことが分かる。

クラスタ 02 はクラスタ 01 と比較すると Reply の利用が多く、友人とのコミュニケーションが多いクラスタと言える。また、Retweet についてはリツイートが平均で 2.2 であることから多少情報の拡散は行っているが、被 Retweet は 0 であるこ

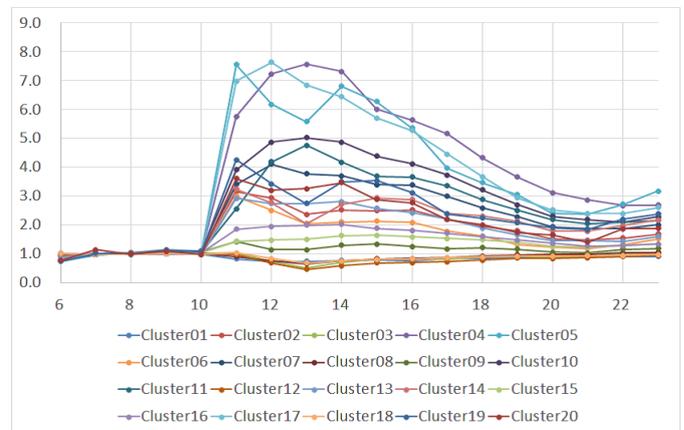


図 3: ユーザ数の変化

とから自分自身は情報源とはなっていない。クラスタ 12 も、クラスタ 02 と同様 Reply 機能を使っているが、Retweet 機能はほとんど使っていないユーザである。

クラスタ 07 はクラスタ 02 と似ているが、自分から Reply を発信するのではなく、他のユーザから Reply を受けることの方が多かったユーザである。

以上のように、震災前後でユーザの分布が変化したことが明らかとなった。

3.4 各クラスタのユーザ数変化

一日ごとの各クラスタに所属するユーザ数の変化を Fig.3 に示す。ここでは、震災前 (6-10 日) の各クラスタに所属するユーザ数の平均を 1.0 としたときの、その後の変化を示している。

このグラフから、震災前と比べ、クラスタ 04, 05, 17 がユーザ数を特に大きく増加させ、クラスタ 01, 03, 08, 18 がユーザ数を減少させていたことが分かる。

3.4.1 大幅に増加したクラスタ

クラスタ 04 は、Reply, Retweet とともにある程度利用しているユーザである。特徴としては、Retweet 元となるユーザの被 Retweet 数がそれほど大きくない (NX-FrTo=179.86) であることから、他のクラスタに比べると有名人の Tweet を Retweet した割合が低いと推測される。

クラスタ 05 は最も積極的に Twitter を利用していたユーザ

群であると言える。積極的に Reply, Retweet ともに利用しており、Tweet 数も最大となるクラスタである。ここから、積極的に Twitter を使うユーザが震災後に大幅に増加したことが分かる。

次に、クラスタ 17 は被 Retweet 数が極めて大きいクラスタである。有名人クラスタであると考えられ、多くの情報を他のユーザに拡散した起点となったユーザと考えることが出来る。これは、本クラスタに所属するユーザの Tweet を Retweet したユーザがさらに Retweet された数 (NX-To) も最大であることから、強い拡散力を持っていたといえる。

3.4.2 減少したクラスタ

ユーザ数が減少したクラスタであるクラスタ 01 は、先に述べたとおりあまり Twitter を利用していなかったユーザである。クラスタ 03 は Reply は多いが Retweet は少ないユーザであり、クラスタ 18 もやはり Retweet を使っていないユーザである。

一方、クラスタ 08 も Retweet はほとんど使っていないが、Reply, 被 Reply はともに 1.0 という特徴を持つ。相互 Reply 率も 1.0 であることから、一人とだけやりとりを行ったユーザであると言える。

以上より、まったく Retweet を使っていないようなユーザは震災直後から減少しており、Retweet 機能を利用して情報拡散を行うユーザが増加したことが示された。

3.5 クラスタ間移動の分析

最後に、震災前後で所属するクラスタがどのように変化したのかを分析した。

特に、積極的なユーザ群であるクラスタ 05 と、情報源として活用されたクラスタ 17 について、どのクラスタに所属していたユーザがこれらのクラスタに変化したかを確認した。Fig.4 に、どのクラスタに所属していたユーザのうちどの程度の割合がクラスタ 05, 07 に変化したかを示す。

これより、震災後積極的に Twitter を利用するユーザ (クラスタ 05) には、震災前にはクラスタ 05 や 19 に所属していたユーザがなりやすいことが明らかとなった。また、情報源ユーザ (クラスタ 17) には、震災前にクラスタ 04, 05, 09, 10, 17 に所属したユーザがなりやすいことが分かった。

これより、積極的なユーザはもともと積極的なユーザが変化しやすいことが明らかとなった。また、情報源ユーザについては、やはりもともとある程度 Retweet されていたユーザが変化しやすいことが明らかとなった。一方で、震災前には全く Retweet されなかったクラスタ 02 の 1% 程度のユーザがクラスタ 17 に変化しているなど、大きな変化がもたらされる場合も存在する。

一方で、震災前にある程度 Retweet されていたクラスタ 20 が情報源であるクラスタ 17 に変化した割合は低く、他のクラスタ 04, 05, 09, 10 との違いを明らかにすることが必要と考えられる。

4. 終わりに

本研究ではユーザのソーシャルメディアの利用方法を明らかにするため、Twitter 上でのユーザの行動から、ユーザのクラスタリングを行った。その上で更に震災前後のユーザ行動の変化とその特徴を抽出した。

クラスタリングを利用して震災前後のクラスタへの所属の分布変化を分析した。その結果、積極的に Twitter を利用するユーザや情報源となるユーザが大幅に増加したことが示された。

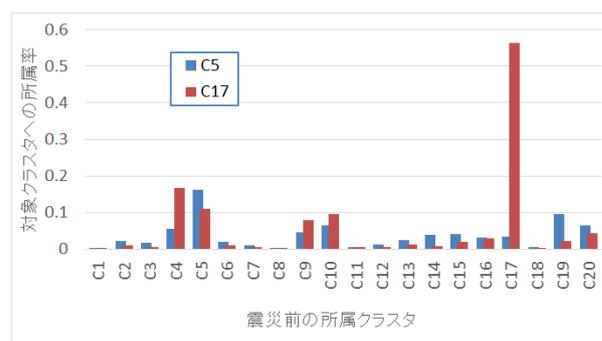


図 4: 各クラスタに所属していたユーザが震災後クラスタ 5, 17 に変化した割合

また、大幅に増加したクラスタの性質を分析した結果、情報源となるユーザはもともと被 Retweet 数がある程度あるユーザが多かったが、被 Retweet が多くても情報源とならないユーザが多いクラスタも存在することが明らかとなった。

今後は、どのようなユーザがどのように利用方法を変化させるのかをより詳細に分析し、情報源ユーザのような重要なユーザがどのように生まれるのか、事前に重要ユーザを把握しておく手法を確立することが必要である。それによって、ソーシャルメディア内でのユーザ行動を把握するツールとしてユーザの問題行動を検出できるシステム構築に寄与できることが望まれる。

5. 謝辞

本研究の実施にあたり、ツイートデータの収集に協力していただいたクックパッド株式会社の兼山元太氏に感謝する。また、本研究の一部は学術振興会「先進的人文社会科学振興事業」によって行われた。

参考文献

- [1] 2014 年度 SNS 利用動向に関する調査：レポート | ICT 総研 市場調査・マーケティングカンパニー 2014 年度 SNS 利用動向に関する調査：レポート | ICT 総研 市場調査・マーケティングカンパニー
- [2] 篠田 孝祐, 榊 剛史, 鳥海 不二夫, 風間 一洋, 栗原 聡, 野田 五十樹, 松尾 豊: 東日本大震災時における Twitter の活用状況とコミュニケーション構造の分析 知能と情報 Vol.25 No.1 pp.598-608(2013)
- [3] パターン認識と機械学習 下 C.M. ビショップ著
- [4] Schwarz, G.: Estimating the Dimension of a Model, *Annals of Statistics*, Vol. 6, No. 2, 461-464 (1978)

表 1: クラスタの特徴

Reply	From	To	Clst	Int	FrEN	ToEN	NX-To	NX-Fr	NX-FrTo	NX-ToFr
C01	0.026	0.073	0.000	0.000	0.000	0.000	0.000	0.000	0.026	0.084
C02	1.949	1.326	0.000	0.245	0.402	0.491	3.525	6.764	10.671	4.629
C03	4.089	5.543	0.179	0.557	1.428	1.077	28.177	20.020	26.583	29.404
C04	1.884	1.154	0.000	0.233	0.354	0.566	3.019	6.339	8.961	5.365
C05	10.534	11.573	0.045	0.523	2.202	2.090	80.446	71.322	90.238	88.551
C06	2.873	4.887	0.146	0.390	1.312	0.775	22.021	12.524	18.591	86.142
C07	0.515	1.341	0.000	0.485	0.203	0.000	3.611	1.537	1.735	8.105
C08	1.000	1.000	0.000	1.000	0.000	0.000	3.736	3.255	3.736	3.255
C09	3.818	3.192	0.032	0.327	0.887	1.014	11.703	15.459	21.430	13.139
C10	1.399	1.347	0.000	0.291	0.405	0.404	3.606	4.347	5.661	6.506
C11	0.016	0.031	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.031
C12	2.687	1.846	0.000	0.504	0.555	0.879	5.935	9.181	12.900	6.197
C13	2.332	2.516	0.117	0.472	0.699	0.559	9.445	8.348	11.969	17.497
C14	4.674	5.736	0.067	0.709	1.489	1.299	29.719	23.193	28.683	27.942
C15	1.657	1.851	0.030	0.326	0.645	0.596	5.694	4.995	5.657	5.622
C16	1.091	1.110	0.000	0.263	0.294	0.267	2.576	3.118	4.315	4.221
C17	7.296	2.517	0.011	0.161	0.752	1.522	7.911	26.887	45.545	13.786
C18	0.480	1.109	0.000	0.094	0.263	0.000	1.513	1.109	1.703	4.016
C19	8.693	10.337	0.060	0.576	2.065	1.889	67.663	55.876	72.232	72.900
C20	4.341	5.339	0.088	0.414	1.289	1.088	25.972	22.659	30.130	40.167

Retweet	From	To	Clst	Int	FrEN	ToEN	NX-To	NX-Fr	NX-FrTo	NX-ToFr	Tweet
C01	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	2.435
C02	2.218	0.000	0.000	0.000	0.000	0.491	0.000	4.897	567.86	0.000	11.321
C03	0.000	0.000	0.000	0.000	0.000	1.077	0.000	0.000	0.000	0.000	18.908
C04	2.389	9.028	0.026	0.046	2.070	0.566	29.267	6.766	179.86	27.757	15.138
C05	4.469	8.323	0.002	0.048	1.940	2.090	12.044	14.110	832.32	31.448	58.901
C06	2.279	0.000	0.000	0.000	0.000	0.775	0.000	5.110	398.96	0.000	23.751
C07	2.054	0.000	0.000	0.000	0.000	0.000	0.000	4.751	629.11	0.000	8.062
C08	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	3.313
C09	0.000	2.453	0.000	0.068	0.672	1.014	1.627	0.000	0.000	6.983	18.019
C10	2.995	3.196	0.000	0.000	1.073	0.404	3.018	8.344	488.14	6.600	15.520
C11	1.794	0.000	0.000	0.000	0.000	0.000	0.000	4.068	496.75	0.000	3.757
C12	0.000	0.000	0.000	0.000	0.000	0.879	0.000	0.000	0.000	0.000	7.631
C13	1.671	1.144	0.000	0.441	0.094	0.559	0.651	3.936	96.393	3.279	15.909
C14	2.440	0.000	0.000	0.000	0.000	1.299	0.000	5.259	698.00	0.000	25.984
C15	0.710	1.000	0.000	0.000	0.000	0.596	0.000	1.299	12.603	1.247	12.483
C16	0.893	1.000	0.000	0.000	0.000	0.267	0.887	1.819	19.321	1.793	9.738
C17	2.171	83.897	0.001	0.013	4.277	1.522	250.79	6.014	120.15	267.83	23.160
C18	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	4.058
C19	3.745	1.348	0.000	0.079	0.231	1.889	1.379	10.139	863.64	6.428	52.007
C20	2.299	3.691	0.188	0.070	1.132	1.088	9.712	6.143	100.68	14.934	33.546