

Twitter ネットワーク上のユーザコミュニティ抽出と話題分析

Community extraction and topic analysis on Twitter

新保 直樹^{*1} 織田 瑞夫^{*1} 城 沙友梨^{*1} 米山 照彦^{*1} 水野 誠^{*2}
 Naoki Shinbo Mizuo Oda Sayuri Jo Teruhiko Yoneyama Makoto Mizuno

^{*1} 株式会社 構造計画研究所
 Kozo Keikaku Engineering

^{*2} 明治大学
 Meiji University

ソーシャルメディア上に日々大量に投稿される自社製品のクチコミデータを「傾聴」することは企業にとって重要な課題になっている。本研究では、Twitter 上でのスマートフォンに関する発言や RT についてデータを収集し、他のユーザに対して強い影響力を持つ可能性があるユーザを中心としたコミュニティを抽出する。そして、そこでの話題の差異を分析する。

1. はじめに

ソーシャルメディアの普及によって、企業と顧客のコミュニケーションのあり方が大きく変わりつつある。[1-2008]はその最も基本的な段階として、顧客の声をソーシャルメディアから「傾聴」することを挙げた。最近では、それはソーシャル・リスニングといわれている。

ソーシャルメディアでの顧客の発言は、通常のマーケティングリサーチとは違い、顧客の自然な状態での発言である。また、企業に直接に寄せられる苦情ほど切羽詰まっていないが、逆に潜在的に重要な情報を含んでいる可能性がある。したがって、ソーシャル・リスニングは、マーケティングリサーチの重要なツールになりつつある[萩原 2011]。

ソーシャル・リスニングの目的が顧客や社会の平均的な意見を把握することだとしたら、対象を特に絞り込む必要はない。しかし、変化の芽をなるべく早い段階で見つけることが目的であれば、影響力の大きい人々、すなわちインフルエンサーの発言に重点を置くことが一つの有望な方法になる。

過去の研究では、SNS 上で友人数(ネットワークの次数)が非常に多い少数のユーザ(ハブ)の影響力が大きいことが示されている[Goldenberg 2009]。Twitter のように、一方的にフォロー対象を選べるメディアの場合、価値ある情報を流していると思われるユーザほどフォローされる可能性が高い。だとすれば、フォロワーの多いハブは影響力が大きくなる傾向があるはずである。

一方、ソーシャルメディア上に少数のインフルエンサーが存在するという考え方への異論も存在する。Watts によれば、大規模なクチコミは山火事のようなもので、どこで出火したかは偶然の結果にすぎない。山火事がどこから起きたかを調べても意味がないように、クチコミがどこから起きたかを調べても意味がないという[Thompson 2008]。Watts の主張を裏づける根拠の1つが、ある個人の影響力は持続しないという研究である[Bakshy 2011]。ただし、そもそも影響力をどう測るのか、それが時間とともに変化することは問題なのか、といった点で議論の余地を残している。

本研究はこうした問題を解決することを目的とした研究の第一ステップである。Twitter を対象に、スマートフォンに関するツイートと RT を分析してインフルエンサー候補を探す。次に、こうして集められたインフルエンサー候補を中心とするコミュニティを抽出し、そこでどのようなツイートがなされているかを分析する。

そして、インフルエンサーを中心としたコミュニティの実態を把握する。

2. データセット

2011年11月1日から2012年3月31日の間に Twitter API を用いて「iPhone」または「Android」を含む日本語のツイートを収集した。収集したツイートの件数は iPhone が 8,237,176 件、Android が 2,369,315 件である。また、ユーザ数は iPhone が 1,258,519 人、Android が 157,554 人である。月別の収集ツイート数を図 1 に示す。2月のデータが極端に少ないが、これはデータ収集の失敗による欠損が主な原因である。

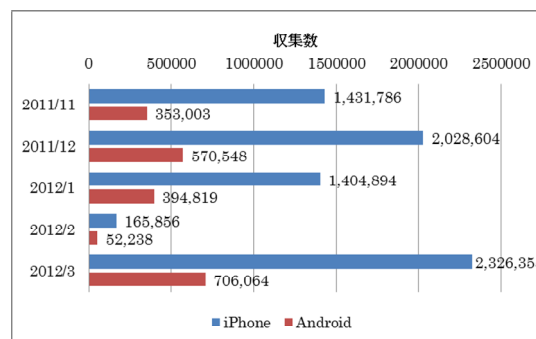


図 1: 収集した月別ツイート数

3. 分析方法

3.1 インフルエンサー候補の抽出

iPhone または Android を含むツイートの中で、多くのユーザが興味関心をもつ話題についての中心人物となるユーザを抽出する。この中心となるユーザをインフルエンサー候補、それ以外のユーザをインフルエンシーと定義する。ここでインフルエンサー候補の選定は以下の指標を用いておこなう。

- 条件 1. 被 RT 数の絶対数が多いこと
- 条件 2. 被 RT 日数が多いこと
- 条件 3. 複数期間において条件 1 と条件 2 を満たすこと

ある個人の影響力が高ければフォロワーの数は増えるし、被 RT 数も増えることから、条件 1 は必要条件と考える。条件 2 は影響力の高いユーザはツイートが頻繁に RT されること、条件 3

はインフルエンサーが恒常的にフォロワーに影響を与える存在であるとの仮定を満たす要件である。具体的には以下の手順でインフルエンサー候補を抽出する。

1. データセットの期間を前期・後期の 2 期間(2011/11/1 ~ 2011/12/31, 2012/1/1 ~ 2012/3/31)に分割する
2. 前期・後期それぞれに対して被 RT の絶対数の上位 1000 かつ被 RT の日数が上位 1000 のユーザを抽出する
3. 前期・後期ともにインフルエンサー候補となるユーザを最終的なインフルエンサー候補とする

この手法を適用した結果、iPhone は 270 人、Android は 355 人のインフルエンサー候補が抽出された。このインフルエンサー候補を中心としてコミュニティの抽出をおこなう。

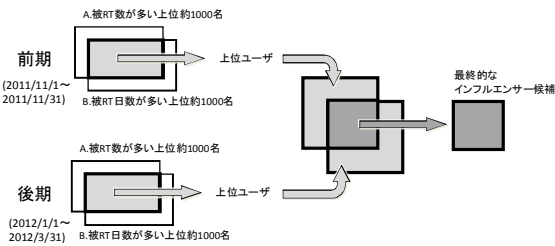


図 2:インフルエンサー候補の抽出方法

3.2 ネットワークの構築とコミュニティ抽出

インフルエンサー候補とインフルエンシーをノード、両者の間の非公式 RT をエッジとしたネットワークを構築する。ただし、偶発的なリツイートを除外するため、インフルエンサー候補とインフルエンシー間において非公式 RT が 2 回以上かつインフルエンシーがインフルエンサー候補をフォローしていることを条件とした。ユーザが非公式 RT としてツイートする方法は様々であるが、「RT」もしくは「QT」と「@」を用いたリツイートが一般的に用いられる。その中でも半角スペースやダブルクォーテーションの有無などの各種のパターンがあるが、今回は表 1 に示すパターンに基づいて非公式 RT の判定をおこなっている。

表 1.非公式 RT の判定基準

RT,QTによるリツイート(引用符なし)	iPhone		Android	
	iPhone	Android	iPhone	Android
RT@foo	4570	1585	3214	287
RT@_foo	6	2	0	0
RT_@foo	1230293	417732	6	0
RT_@_foo	254	41	0	0
QT@foo	79	43	2076	212
QT_@foo	0	0	1	0
QT_@_foo	17086	8713	5	0
QT_@_@_foo	30	3	0	0
QT"@foo			114	18
QT"@_foo			0	0
QT"@_@foo			0	0
QT"@_@_foo			103	14
QT"@_@_@foo			0	0
QT"@_@_@_foo			0	0

本研究では構築した RT ネットワークの最大コンポーネントを分析対象とする。最大コンポーネントが全体に対して占める割合は約 91%である。また、iPhone についてのみ言及したユーザが 3818 人、Android について言及のみのユーザが 578 人、iPhone と Android の両方に最低一回以上言及しているユーザが 13,770 人である。ツイート数は iPhone が 527,198 件、Android が 301,903 件となった。図 3 に抽出した RT ネットワーク

を示す。次に、構築したネットワークからコミュニティの抽出をおこなう。コミュニティ抽出にはモジュラリティに基づくクラスタリング手法である CNM アルゴリズム[Clauset 2004]を用いてユーザを 51 のコミュニティに分類し、各コミュニティを C00~C50 と名付けた。各コミュニティのサイズを図 4 に示す。



図 3:抽出した RT ネットワーク

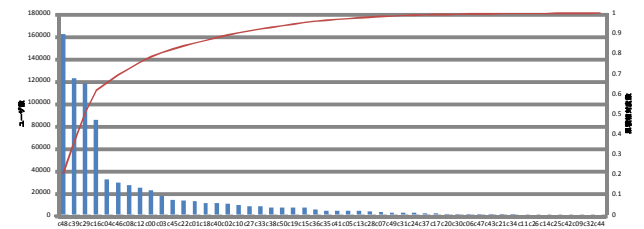


図 4:クラスタリング結果

3.3 トピックの抽出

本研究ではツイートからの話題の抽出に潜在的ディリクレ配分法を用いる[Blei 2003]。潜在的ディリクレ配分法は1つの文章に複数のトピックが存在すると想定した文書生成過程の確率的なモデルである。ここでトピック数 K は予め与える必要がある。本研究では K=40 としてツイートのトピック分類をおこない、トピック内の特徴語から各トピックにラベルを付与した(表 2)。また、抽出したトピックのうち、類似したトピックに同一のカテゴリ名を付与することで、Twitter 上のスマートフォンに関する話題を「ニュース」、「アプリ情報(一般)」、「アプリ情報(ゲーム)」、「端末情報」、「キャンペーン」、「市場動向」、「通信キャリア」、「製品情報」の 8 つのカテゴリに分類した。各カテゴリの概要について表 3 に示す。

表 2.抽出したトピックとラベル

トピック	ラベル	トピック	ラベル
topic0	電子書籍アプリ情報	topic20	タブレット端末普及
topic1	スマホ全般のニュース	topic21	スマホ全般のニュース
topic2	内容共有, 写真加工, 画像の話題	topic22	スマホ本体機能の話題
topic3	端末のリリース情報, 新製品の噂	topic23	代理店存在の話題
topic4	通信の話題(テキスト通信不具合など)	topic24	音楽アプリ情報
topic5	無料情報や製品情報	topic25	端末修理やサポート
topic6	スマホ全般のニュース	topic26	アプリ勉強会
topic7	電子書籍宣伝	topic27	ハンズフリー, 常電の話題
topic8	アプリ情報	topic28	通信速度の話題
topic9	android日本導入の話題	topic29	アプリ開発の話題
topic10	アプリリリース情報	topic30	アプリ宣伝
topic11	ゲームアプリリリース情報	topic31	キャンペーンの話題
topic12	スマホ全般のニュース	topic32	アプリセキュリティの話題
topic13	スマホ本体関連	topic33	スマホ市場動向全般の話題
topic14	集約のニュース(版権問題等)	topic34	アプリアップデート情報
topic15	Androidアプリ開発情報	topic35	iphone周辺グッズのプレゼント情報
topic16	NTTドコモ動向	topic36	キャリア間比較(通信速度比較)
topic17	名刺交換アプリ情報	topic37	スマホ性能評価
topic18	ゲームアプリリリース情報	topic38	iphoneゲームアプリリリース情報
topic19	ニュースの話題(topic)	topic39	スマホ全般のニュース

表 3.抽出したトピックのカテゴリ分類

カテゴリ	ツイート概要
ニュース	スマホに関連する様々なニュース
アプリ(ゲーム)	ゲームアプリの話題
アプリ(一般)	ゲーム以外のアプリの話題
端末本体	端末本体の操作方法や通信障害情報
キャンペーン	端末の割引情報や特典情報
市場動向	端末のシェア、メーカーの参入・撤退情報など
通信キャリア	通信キャリア各社についての話題
製品情報	新製品のスペックや価格など

4. 考察

4.1 コミュニティの特徴分析

各コミュニティの iPhone/Android のツイート数の構成比を図 5 に示す。iPhone、Android それぞれの話題を中心としたコミュニティと、iPhone と Android 両方についてバランスよくツイートをおこなっているコミュニティが存在することがわかる。

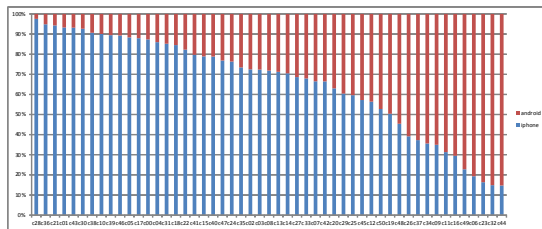


図 5:iPhone/android のツイート数の構成比

次に各コミュニティの話題分析をおこなう。コミュニティ数が多いことから分析対象を所属ユーザ数の上位 3 つのコミュニティ (c29, c48, c39) とした。選択した 3 つのコミュニティにおける特徴的なトピックを表 4 に、コミュニティ内のユーザの iPhone ツイートの比率と人数の関係を図 6 に示す。

表 4.上位 3 つのコミュニティのトピック

トピック	ラベル	カテゴリ	c29		c48		c39	
			iPhone	Android	iPhone	Android	iPhone	Android
top06	スマホ全般のニュース	ニュース	○	○				
top12	スマホ全般のニュース	ニュース						
top39	スマホ全般のニュース	ニュース						
top23	中身良対応の話題	通信キャリア		○				
top29	アプリの話題	通信キャリア						
top38	アプリの話題	通信キャリア						
top02	内蔵カメラ、写真加工、画像の話題	端末本体	○		○		○	
top13	スマホ本体の話題	端末本体					○	
top19	ケースの話題 (iPhone)	端末本体						○
top27	スマホ本体機能の話題	端末本体						○
top25	端末修理や整備	端末本体						○
top27	iPhone の修理、交換の話題	端末本体						○
top37	スマホ特許情報	端末本体						
top03	端末のリリース情報、新製品の噂	製品情報			○	○	○	
top05	端末価格や製品情報	製品情報						
top14	海外のニュース (総論的)	市場動向	○		○			
top16	NTTドコモの話題	市場動向	○		○			
top33	スマホ市場の動向の話題	市場動向			○	○		
top07	競争環境	キャンペーン						
top20	タブレット端末の話題	キャンペーン						○
top31	プレゼントの話題	キャンペーン						○
top35	iPhone 周辺グッズのプレゼント情報	キャンペーン						○
top11	ゲームアプリのリリース情報	アプリ(ゲーム)						○
top17	ぬいぐるみアプリの話題	アプリ(ゲーム)						○
top18	ゲームアプリのリリース情報	アプリ(ゲーム)						○
top38	iPhone ゲームアプリのリリース情報	アプリ(ゲーム)						○
top15	Android アプリのリリース情報	アプリ(ゲーム)			○		○	○
top09	携帯アプリの話題	アプリ(一般)						
top01	スマホ全般のニュース	アプリ(一般)			○			
top08	アプリの話題	アプリ(一般)						○
top10	アプリのリリース情報	アプリ(一般)			○			○
top24	携帯アプリの話題	アプリ(一般)			○			○
top34	アプリのアップデート情報	アプリ(一般)			○			○

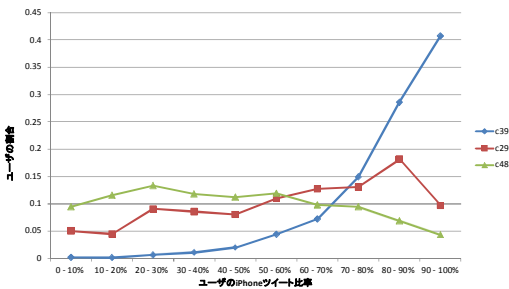


図 6:iPhone ツイートの比率とユーザ数の関係

最も特徴的なコミュニティは C39 の iPhone に関する話題を中心としたコミュニティである。ツイート内容の特徴として、本体の機能、バッテリー持続時間など既に発売済みの端末についての話題が多い。

C29, C48 は iPhone、Android 両方のキーワードに対して活発にツイートしているコミュニティで、新しい製品や、スマートフォン市場の動向についてのツイートが活発なグループである。一方で既存の端末の操作や機能など、端末利用としての話題への関心は C39 と比較して薄い傾向にある。このように本研究の手法を用いることで、指定したキーワードについて Twitter 上でのコミュニティ単位での話題を分析することが可能となる。

5. まとめと今後の課題

本研究では RT を指標としてスマートフォンについて Twitter 上で影響力の高いユーザを中心とした RT のネットワークによるコミュニティの抽出を行い、各コミュニティの話題についての分析をおこなった。

今後の課題としては、各コミュニティの中心となっているインフルエンサー候補について詳細に調べることでコミュニティの特徴についてより深い理解を得ることがまずは挙げられる。またインフルエンサー候補に本当に影響力があったかどうか、またその影響力がどこまで持続するかについては別の検証が必要で、それも今後に予定される研究課題である。

参考文献

[リー2008] シャーリーン・リー, ジョシュ・バーノフ (2008), グランズ ウェル ソーシャルテクノロジーによる企業戦略, 伊東美奈子 訳, 翔泳社。
 [萩原 2011] 萩原雅之 (2011), 次世代マーケティングリサーチ, ソフトバンククリエイティブ。
 [Goldenberg 2009] Goldenberg, J., Han, S., Lehmann, D. R. & Hong, J.W. (2009), The Role of Hubs in the Adoption Process, Journal of Marketing, 73 (March), 1-13.
 [Bakshy 2011] Bakshy, E., Hofman, J. M., Mason, W. A. & Watts, D. J., Everyone's an Influencer: Quantifying Influence on Twitter, WSDM'11.
 [Thompson 2008] Thompson, C. (2008), Is the Tipping Point Toast? Fast Company, Feb.8. <http://www.fastcompany.com/641124/tipping-point-toast>
 [Clauset 2004] Clauset, A., Newman, M. E. J. & Moore, C.. (2004), Finding Community Structure in Very Large Networks, Physical Review E, 70, 066111.
 [Blei 2003] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003), Latent Dirichlet Allocation, Journal of Machine Learning Research, 3:993-1022.