

Twitter を用いたインフォーマルコミュニケーション支援に利用可能な情報の抽出方法の検討

笹嶋 亮太[†] 大竹 恒平[†] 櫻井 彰人[†]

慶應義塾大学 理工学研究科 開放環境科学専攻[†]

1 研究背景

近年、人と人とのコミュニケーションを支援するツールである SNS・マイクロブログが人気を博しており、インターネットを介したコミュニケーションが盛んに行われている。中でも手軽に情報発信・収集が可能な、マイクロブログの 1 つである Twitter の利用者は増加の一途を辿っている。これらを背景に、Twitter に投稿された情報を用いた研究がおこなわれている [1][2]。それらは Twitter 上でのコミュニケーション支援を目的としている。

一方で Twitter の情報を用いて、現実世界で行われる対面コミュニケーション (face to face (以下 FTF)) のコミュニケーション支援を目的とした研究は、十分に行われていない。Twitter 上には、ユーザの趣味・趣向をはじめとする、ユーザ属性を表す様々な情報が蓄積されている。これらの情報を適切に用いることで、インターネット上のコミュニケーションに限らず、FTF のコミュニケーションの活性化に役立てることができると考える。

2 研究対象および研究目的

本研究ではマイクロブログの 1 つである Twitter に着目し、ユーザの過去のツイートから、趣味・趣向などの、コミュニケーションに利用可能な情報を取得し、FTF のコミュニケーション活性化を目指す。なお、本研究においては、インフォーマルコミュニケーション (以下 IC) を想定する。IC の例としては、休憩スペースでの雑談や、組織内の知人同士での情報交換などが挙げられる。また、本研究では、Twitter の主要な利用層であり、IC の機会が多い大学生に着目する。

3 現状分析

現状分析として、SNS・マイクロブログの利用状況及び IC の現状を明らかにするため、大学生 70 人を対象にアンケート調査及び半構造化インタビュー調査を行った。

この調査の結果、約 7 割の回答者が Twitter を利用していた。投稿内容としては、友人とのコミュニケーション、興味・近況に関する内容

が上位にあげられた。

一方、IC の現状に関するアンケート調査の結果、IC を苦手だと感じている人は、話題に困るなどのネガティブな感情を抱く傾向が確認された。また、初対面の相手とのコミュニケーション時のきっかけとして、共通の友人、所属よりも、共通の趣味、興味が重要視される傾向が確認された。

現状分析の結果を踏まえ、本研究では Twitter の情報のうち、趣味・興味に関する情報に注目し、インフォーマルコミュニケーションに利用可能なトピックの抽出手法について検討を行う。

4 インフォーマルコミュニケーションにおける会話の内容に関する分析

IC で実際に会話される内容 (トピック) を明らかにするため、1 対 1 での IC を想定した実験を行った。被験者は、3 グループで計 11 人 (男性 8 人女性 3 人) の学生であり、初対面の組み合わせを 22 通り行った。実験室には初対面の 2 人とカメラ、レコーダーを設置した。実験の様子は動画、音声で記録した。また、各組み合わせ終了後、被験者に対してアンケート調査を行った。調査項目の概要を以下に示す。

- ✓ 話した内容に満足しているか (5 段階)
 - ✓ 会話が盛り上がったと感じるか (5 段階)
- これらについて分析を行った結果、以下のことが明らかになった。
- 自分が話した内容への満足度の高評価と盛り上がりの高評価には、正の相関がある (相関係数 0.93)
 - 共通の興味、現在の状況に関する内容が話された場合、双方の盛り上がりに関する評価が高い
 - 話題の不一致もしくは第一印象が悪い場合に盛り上がりに関する評価が低くなる
- 共通の興味での具体的なトピックとしては「音楽」「スポーツ」「ゲーム」に関する内容が良く話されている傾向が確認された。

これらの結果を踏まえ、本研究では、趣味・興味に関する情報の内、「音楽」「スポーツ」に注目し、それらの情報を Twitter 上から取得する手法の検討を行う。

5 ツイートの分析

はじめに、大学生 100 人の Twitter アカウントをランダムに抽出し、ツイートを取得した。収集したツイートに対して、形態素解析を行い、各品詞の出現頻度を求めた。単語に注目した理由としては、Twitter は 140 文字以内の文章が投稿されるため、分析の情報が少なく、単語単位で文章分類を行った方が、より適切に分類を行うことができると考えた為である。これらの品詞の内、本研究では、名詞に注目した。

次に、名詞の出現頻度上位 900 件に対して、筆者が内容を元に分類を行った。その結果、「生活」「興味・趣味」「地域」の 3 つのトピックに分類できることが明らかになった。それぞれの頻度を表 1 に示す。

表 1 から、約 2 割のツイートが興味・趣味に関する内容であることが分かる。なお、その内容に注目すると、属人的で、多岐にわたることが明らかになった。

表 1 名詞の上位 900 件に対するトピック分け

トピック名 (出現頻度)	中分類	小分類	出現頻度
生活 (311)	学業	勉強	89
		バイト	9
	社会活動	衣食住	152
		状態 (結婚など)	23
	生活活動	イベント	21
		人間関係	17
興味・趣味 (193)	コンテンツ	スポーツ	74
		アニメ、漫画	18
		音楽	27
		映画	16
		ゲーム	20
		旅行	5
		その他	33
		地名 (1ヶ所のみ)	259
地域 (396)	行動	組織	104
		インフラ	33

6 トピックの抽出手法の検討

本研究では、個人の Twitter アカウントから IC に利用可能な「興味・趣味」に関する情報を抽出するため、Wikipedia の見出し記事を用いた分類方法の検討を行った。

はじめに、Wikipedia の記事の内、「スポーツ」「音楽」の 2 つのトピックに対し、辞書を作成した。辞書の作成方法としては、各トピックに関連する記事を用いて、tfidf 値を算出し、スコ

アの高い単語と記事のタイトルを辞書に登録した。作成した辞書を用いて、被験者 10 名計 5000 件のツイートに対して分類を行い、その精度を確認した。結果を表 2 に示す。

表 2 トピック別分類の結果

	Accuracy	Precision	Recall
音楽	95.6%	26.0%	40.7%
スポーツ	87.2%	41.0%	39.1%

表 2 から、本研究において作成した辞書の有効性は確認されなかった。正しく分類されていないツイートに注目すると、愛称・略語が含まれるものや独立した 1 ツイートだけでは意味を理解することができないものが多くみられた。そのため、本研究ではカテゴリ内の記事の本文に対して tfidf 値を計算し、上位単語を登録することで略語や特徴語を取得できる様に修正を行ったものの、反対に一般的な単語が登録されてしまうことがあった。その結果 Precision が非常に低い結果になったと考えられる。

これらの結果から、本研究において作成した辞書を用いて、トピックに関連するツイート抽出を試みた場合、現状では十分な精度を担保できていない。しかしながらコミュニケーションに必要な「興味・趣味」情報の分類に対して、可能性の一端を示したと考える。

7 結論および今後の課題

本研究では、マイクロブログの 1 つである Twitter に注目し、IC に利用可能な情報の抽出を行った。初対面同士の 1 対 1 での IC を想定した実験を行った結果、「共通の興味・趣味」が重要視される傾向が確認された。また、大学生を対象にツイートの分析を行った結果、ツイートを「生活」「興味・趣味」「地域」の 3 つのトピックに分類できることが明らかになった。

これらの結果をうけ、本研究では、Wikipedia の見出し記事を用いて辞書を作成し、興味・趣味に関するツイートの分類を試みた。今後は、LDA 等、トピックの自動生成を行う手法を試みる。もしくは、人手により単語を辞書に登録することで精度の向上を図る。また、分類結果を実際の IC 時に話題として提供することで、有効性を検証する予定である。

参考文献

- [1] 古賀裕之, 谷口忠大: "潜在トピックに着目した Twitter 上のユーザ推薦システムの構築", ヒューマンインターフェースシンポジウム (2010)
- [2] 岩木祐輔, アダム・ヤトフト, 田中克己: "マイクロブログにおける有用な記事の発見支援", DEIM Forum (2009)