

SNSでの情報探索行動に関する予備的検討：情報要求ツイートの分類

Preliminary Analysis of Information Seeking Behavior on SNS: Classifying Information Needs Tweets

齋藤 ひとみ[†], 林 穂波[†]

Hitomi Saito, Honami Hayashi

[†] 愛知教育大学

Aichi University of Education

hsaito@aecc.aichi-edu.ac.jp

概要

本研究では、Twitter上で情報を求めるつぶやき(情報要求ツイート)を収集し、どのような質問がされているのかを分析した。Twitter APIを用いて質問系のハッシュタグを含むツイートを収集し、機械学習の手法を使って情報要求ツイートを自動抽出した。抽出された918件のツイートについて、タイプやトピックの分類、リプライの有無を調査した。分析の結果、Twitterで多く投稿される情報要求のタイプや回答率の高い情報要求のタイプなどが明らかになった。

キーワード: Twitter, Information Seeking Behavior, Information Needs Tweets

1. はじめに

近年、Web上の情報が日々増大しており、人々はサーチエンジンやQ&Aサイト、SNSなどの様々なサービスを使い分けながら情報探索を行っている。中でもリアルタイム性が特徴とされるTwitterは、ツイート検索、ユーザ間の質問行動、流れてくる情報を受動的に得る等の方法により、リアルタイムな情報探索に利用されている。本研究では、それらの中でも能動的で対話的な情報探索行動である質問行動に焦点を当て、人々はTwitterを利用してどんな情報を要求しているのか、その傾向について検討する。そこで、Twitter上で情報を要求する「情報要求ツイート」について分類・分析する。

2. 背景と目的

Web上の質問行動を分析した研究として、栗山・神門(2009)はQ&AサイトであるYahoo!知恵袋に投稿された質問のタイプを分類した。その結果、質問タイプを、サーチエンジンや本等で調べることで適切で客観的な回答が得られる「情報検索型」と、適切な回答の基準がなく個人的な意見を尋ねる「社会調査型」の2つに大別できるとし、これらに「非質問型」を加え

た3つの型と、その下位13項目に分類した。Q&Aサイトの質問カテゴリと質問タイプとの関係を分析した結果、Q&Aサイトの質問の特徴として、科学的・技術的な質問は情報検索型に広く分布し、人間関係や個人の生活等に関する質問は社会調査型が多く、その中でも特に助言や意見を求めるものが多いことを明らかにした。

またSNS上の質問行動について分析した研究として、Morris, Teevan, and Panovich(2010)は、624人に対してSNSで質問する頻度や質問する・答える理由を尋ねた。さらに実際の質問やその回答も提出してもらい、質問のタイプとトピックを分類することで質問行動について調査した。その結果、質問のタイプを情報検索系1つ、社会調査系7つの合計8つに、トピックを9つに分類している。またPaul, Hong, and Chi(2011)は、Morris et al.(2010)を使用してTwitter上の質問約120万件に対し分類を行った。各ツイートに対する回答の分析もしており、回答率や回答の関連性についても調査している。

先行研究から、海外のTwitterユーザを対象とした質問行動についての研究は多く見られた。しかし、海外と日本とでは言語や文化が異なるため、質問行動の傾向も変わってくると考えられる。そこで、本研究では日本語の情報要求ツイートを収集し、分類・分析することによって、日本のTwitterユーザがTwitterを利用してどんな情報を求めているのか明らかにする。さらに収集したツイートに対する回答の調査も行い、どんな情報要求ツイートに回答が付きやすいのかについても検討する。

3. 方法

日本語の情報要求ツイートは多様で、抽出が難しい。そこで、収集をより効率的に行うために、教師ありの機械学習による自動分類を用いた。抽出の流れを図2に示す。まず質問系のハッシュタグを含むツイートを

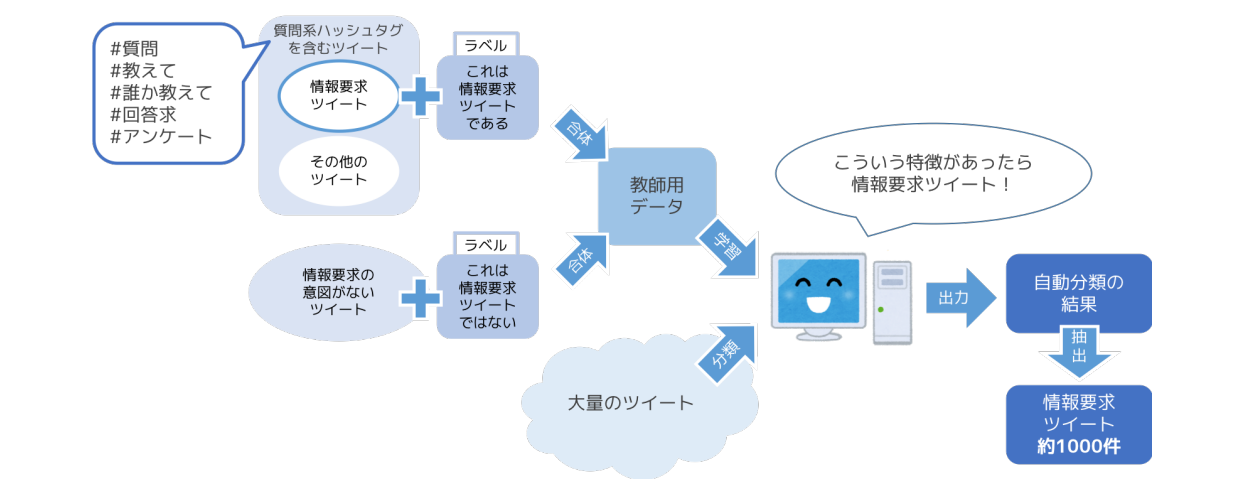


図 1 情報要求 Tweets の抽出方法

表 1 タイプ別結果

型	タイプ	件数	割合	回答率
情報探索型	事実	68	6.9%	13.5%
	真偽	275	27.9%	17.6%
	定義・記述	28	2.8%	13.0%
	方法・手段	20	2.0%	31.3%
	原因・理由	41	4.2%	15.6%
	効果・結果	17	1.7%	7.1%
	助言	22	2.2%	33.3%
社会調査型	意見	96	9.7%	25.3%
	嗜好	23	2.3%	58.8%
	推薦	17	1.7%	21.4%
	経験	14	1.4%	27.3%
	相手の状況・性質	94	9.5%	41.0%
	同意・反語	164	16.6%	16.8%
	勧誘・募集	102	10.3%	34.4%
	許可	5	0.5%	60.0%

表 2 トピック別結果

トピック	件数	割合	回答率
コンピューターテクノロジー	14	1.5%	18.2%
教養・学問・サイエンス	27	2.9%	17.4%
エンターテインメント・趣味	481	52.4%	24.2%
恋愛・人間関係の悩み	23	2.5%	6.3%
美容・健康・ファッション	38	4.1%	18.5%
地域・場所・旅行	21	2.3%	21.4%
料理・グルメ	19	2.1%	35.7%
買い物	15	1.6%	15.4%
その他生活	10	1.1%	12.5%
マナー・規則	7	0.8%	0%
倫理・哲学	0	0.0%	
ニュース・政治	10	1.1%	0%
現在・最近のイベント・流行	119	13.0%	19.8%
Twitter	23	2.5%	22.2%
雑談・クイズ	111	12.1%	25.3%

収集し、その中で実際に質問の意図があるもの 336 件を教師用の正解データとして使用した。次に明らかに情報要求の意図がないツイート 664 件を教師用の不正解データとして使用した。これら 1000 件を教師用データを使った機械学習を行い、約 5 万件のツイートを自動分類した。そして、自動分類で情報要求ツイートと判断された 2527 件のツイートから、実際に質問の意図があるかどうかを手作業で確認し、抽出した合計 918 件についてタイプ・トピックを分類した。

分類に用いたタイプ・トピックは、栗山・神門 (2009) の分類を基盤に、Morris et al. (2010) の分類も参考にし、実際に分類を行う過程で適宜新たな分類を追加しながら作成した。分類の結果、タイプについては表 1 に示すように情報探索型が 6 項目、社会調査型が 9 項目となった。またトピックについては表 2 に示す 15 項目となった。

回答の調査は、各情報要求ツイートに対するリプラ

イ、またはアンケート投票が 1 件以上あれば回答有とし、回答の有無を確認したツイートのうち回答有であった割合を回答率として算出した。

4. 結果

分類結果と回答率について、タイプごとの結果を表 1、トピックごとの結果を表 2 に示す。なお、1 つのツイートが複数のタイプに属する場合がある。分類の信頼性を確認するため、全タイプ・トピック含まれるよう抽出した情報要求ツイート 100 件を協力者 1 人に分類してもらい、筆者との一致率を調べた。その結果、タイプについては一致率 90%、トピックについては一致率 71%、Kappa 係数は共に 0.6 以上となり、高い一致度であった。

タイプの分類では、2 つの型の件数に大きな差は見られなかったが、情報探索型においてはその過半数が「真偽」で突出している。それに対し社会調査型では「同意・反語」が 3 割ほどを占めたものの、「勧誘・募

集」,「意見」,「相手の状況・性質」の3つも多く,1つのタイプだけに大きく偏ることはなかった。トピックの分類では,「エンターテインメント・趣味」が過半数を占めており,次いで「現在・最近のイベント・流行」,「雑談・クイズ」が多い。また,海外で見られた「倫理・哲学」は,今回は1件も当てはまるものがなかった。

回答率は,全体では22%であった。タイプごとでは「許可」,「嗜好」,「相手の状況・性質」の順で高く,トピックごとでは「料理・グルメ」,「雑談・クイズ」,「エンターテインメント・趣味」の順で高かった。しかし,「許可」と「料理・グルメ」については,分類されたツイート数が少ないため割合が高くなってしまっていると考えられる。

5. 考察

分類を行った結果,タイプ分類において「真偽」の多さが目立った。「真偽」は物事の真偽や可否を尋ねるものであるが,中でも自分の予想・記憶や話題になっている物事の真偽について確認を取るようなものが多かった。このことから,人々はTwitter上では,未知の物事の名称や仕組みを尋ねるより,自分の知識や記憶の曖昧な部分について確認を取る傾向があるのではないかと考える。また,栗山らやPaulらの分類結果と比較した結果,いずれにおいてもこの傾向は見られなかった。しかしPaulらの分類においては,そもそも情報検索型にあたる質問が1つにまとめられており,「真偽」にあたるような質問が抽出されたかどうかは定かでない。よってこの傾向は,少なくともQ&Aサイトでは見られない, Twitter特有の傾向であるといえる。また,タイプ「同意・反語」,「勧誘・募集」,「意見」,「相手の状況・性質」や,トピック「エンターテインメント・趣味」が多く見られたことについては,仲の良い友達や趣味の繋がりなど,自分で構築した社会ネットワークに向けて質問できるSNSだからこそ見られる傾向であるといえる。また,回答率についてはタイプ「嗜好」やトピック「雑談・クイズ」が高かったが,これらは知識の有無に関わらず誰でも気軽に答えられるものが多いため,高くなったと考えられる。

6. おわりに

本研究では,日本のTwitterユーザが情報要求ツイートの投稿を通してどんな情報を得ようとしているのか,その傾向と,Q&Aサイトや海外のSNSについての研究では見られなかった特徴を明らかにした。今後は,文面だけでなく画像・動画やアンケート機能等

を考慮した分類方法や,複数のタイプに当てはまる場合の扱い方,リプライ以外の回答の調査方法等が課題となるだろう。

文献

- 栗山 和子・神門 典子 (2009). Q&A サイトにおける質問と回答の分析. 『情報処理学会 研究報告データベースシステム』, **148** (19), 1-8.
- Morris, M., Teevan, J., & Panovich, K. (2010). What do people ask their social networks, and why?: a survey study of status message q&a behavior. *SIGCHI2010*, 1739-1748.
- Paul, S., Hong, L., & Chi, E. (2011). Is Twitter a Good Place for Asking Questions? A Characterization Study. *AAAI2011*.