

ユーザの行動、選択を支援するリアルタイム情報の提供

眞保宏史[†] 岡誠[†] 森博彦[†]

[†]東京都市大学大学院工学研究科システム情報工学専攻

1. 研究背景

近年、インターネットの普及に伴い、インターネット上の情報を収集し活用をする機会が増えてきている。そのため様々な場面でキーワードを入力するweb検索が利用されている。さらにSNSの一部であるTwitter等のリアルタイムで情報を配信出来るサービスの発展により、リアルタイムで変動する情報を取得するためにこれらのサービスを使う機会も増えてきた。Twitter上での情報の投稿をツイートと称し、ツイートをを用いて様々な場面でユーザに対するリアルタイムな情報を提供する研究が行われている。しかし、突発的に発生し、時間の経過により変動する事象をユーザが逐一認知し検索に繋げることは難しい。

2. 関連研究

人の行動、移動パターンからツイートに付与されたジオタグを使い、ある時間にツイートが集中する場所とツイートされたユーザの移動範囲を特定し、それらのツイートの特徴語からイベントの発見と影響範囲を推定する藤坂ら[1]の研究がある。この研究において、実際に人の流動とツイートからその場所で起きているイベントを特定できるといった結果が出ている。しかしこの研究では、イベントの名称は分かっても、具体的にそのイベントの中で発生しているリアルタイムな出来事を把握することはできない。さらにジオタグが付与してあるツイートが少数であるという問題もある。

築瀬ら[2]はリアルタイムな電車の運行状況のある路線に関する単位時間あたりのツイート数を元に、定常時と異常時を判別し、通知を行うシステムの研究を行っている。この研究において、リアルタイムで変化する事象に対するツイートの有効性が示されている。

しかし運行に支障が発生しているという事実は認識できても、具体的な運行状況に繋がる情報を得ることは困難である。

3. 研究目的

本研究ではユーザの行動からユーザの目的を推測し、推測した目的からユーザの行動、選択に必要なとされるリアルタイムな情報を提供することを目的とする。

3. 提案方法

ユーザが起こしている行動を、ユーザの情報端末から把握する。ユーザは行動を起こす際、アプリやWebページ等を閲覧し、目的を調べることが多いと考え、これらの情報を解析することでユーザの目的を推測し、目的、手段の分かる語句として抽出する。さらにユーザの現状を把握するためにユーザの存在する現在地、現在時刻を抽出することにする。抽出した情報に基づいて、ユーザがリアルタイム情報を意図しているか判断し、情報提供の有無を決定する。ユーザに情報提供を行う場合、抽出した目的の分かる語句を検索ワードとして用いて、ユーザの目的に関するリアルタイムな情報を取得する。そして取得した情報を解析することで、ユーザに適したリアルタイムな情報を提供するためのキーワードを生成する。

生成したキーワードと検索ワードを用いて、さらに情報を取得する。最後に取得したデータであるリアルタイムの情報が投稿された時刻と現在時刻がどのくらい離れているのかをふまえた上で提供することとした。

今回はリアルタイム情報を収集するためのデータベースとしてTwitter上のツイート情報を用いることとした。

5. 予備調査

ユーザの目的、手段に関係する情報の含まれるツイートを取得するためにはどのような語句が有効であるかを調査した。今回は鉄道の運行情報を例とし調査を行った。ユーザが乗車する路線名を路線検索サイトから検索ワードとして抽出し、TwitterAPIを用いて検索ワード毎に最新100件のツイートを収

Provides real-time information to support user actions, the selection

Hiroshi Shimpo[†], Makoto Oka[†] and Hirohiko Mori[†]

[†]Systems and Information Engineering, Tokyo city University 158-0087 Tokyo Jap

集した。収集したツイートのうちユーザの乗車路線の運行情報が容易にわかるツイートを正解ツイートとした。同様に日時、路線を変更し5事例を調査した。そして正解ツイートの中に共通して登場する語がどのような語であるかを調べた。

正解ツイートの中に多く含まれる語として遅延、遅れ、運転、見合わせ、再開、止、動があることが分かった。調査した5事例の正解ツイートの中にそれぞれどのくらい含まれているのかの平均をまとめたものを図1に示す。

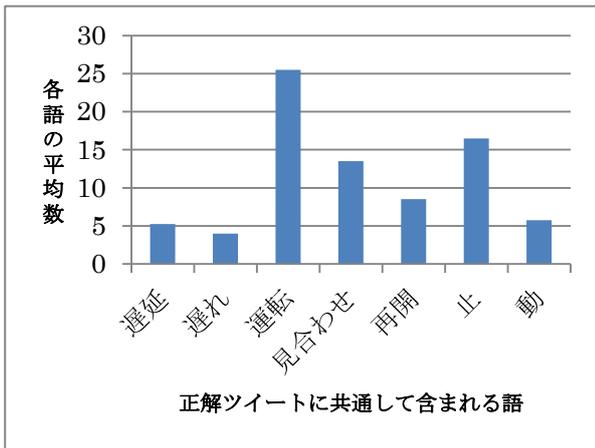


図1 正解ツイートに含まれる各語の数の平均

図1より運転という語が一番多く含まれていることが分かった。他の語は一時的な状態を表す語が多いことが分かった。さらに事例を9件に増やし、正解ツイートを電車が不通になっている状態、運行は再開直後の遅延が発生している状態、ほぼ平常通りに戻りつつある状態の3つの状態に分け、それぞれの状態でどのくらい語が含まれているのかの平均を調べた。

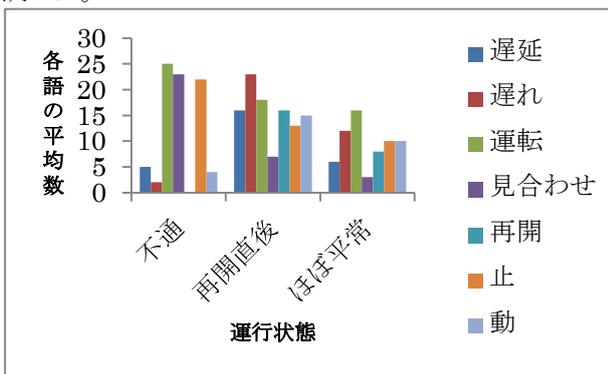


図2 正解ツイートに含まれる各語の数の平均

図2の結果より、リアルタイムの事象変化に伴い、正解ツイートの中に多く含まれる語の数も変化すると考えられる。

6 システム概要

提案システムは、大きく分けてユーザ情報抽出部、情報提供判断部、キーワード生成部、出力部の4つの要素からなる。

6.1 ユーザの情報抽出部

ユーザの状況を把握するためにユーザの情報端末を用いた行動をつかみとり、その行動を起こした日時、現在地、閲覧したサイトの使用したソース情報アプリ等のタイトル名から目的地、手段を抽出する。

6.2 情報の提供判断部

ユーザが行動を起こしている日時、現在地の情報をもとにリアルタイムの情報を提供するかどうかを判断する。初めにユーザの現在地と目的地または経由地との位置関係を把握しておく。今回は5.1にて抽出したユーザの現在地と目的地の座標を

geocoding API を用いて特定する。特定した座標よりユーザの現在地から目的地までの距離を算出する。算出した距離をもとに徒歩で移動することを想定した時間を算出する。算出された時間が2時間以上の場合、および明らかに検索した日時とユーザが行動すると思われる日時がかけ離れている場合には情報の提供は行わないこととした。

6.3 キーワード生成部

6.2 でリアルタイム情報を提供する状態にふさわしいと判断された場合のみ行う。6.1にて抽出した目的地、手段を検索ワードとし、TwitterAPI を用いて最新100件のツイートを収集する。予備調査の結果から1回で収集したツイートの中に含まれる出現頻度が高い語をキーワードとすることによって、効率よくリアルタイムな情報を収集することが出来る考えた。

6.4 出力部

6.1にて抽出した検索ワードと6.3にて生成したキーワードを用いて検索し、検索結果のツイートを時系列順に並べ出力する。なお現在時刻とツイートの時刻との差に10分以上の差があった場合には、出力はしないものとする。

7.まとめと今後の課題

今回はユーザの行動に関係し、その判断材料となるリアルタイムな情報を収集し、ユーザに適したタイミングで、提供する手法を提案した。今後は収集した情報から、各ユーザにとって最善となる方策を提示する方法を提案することでさらにユーザの負担を減らすことが出来ると考えている。

参考文献

- [1] 藤坂達也,李龍,角谷和俊” 実空間マイクロブログ分析による地域イベントの影響範囲推定”, 情報処理学会研究報告, 2010-7-4
- [2] 築瀬拓弥,増田英孝,山田剛一,荒牧英治,中川裕志” 時間的距離に着目した Twitter を用いた電車遅延の自動通知” 情報処理学会研究報告,2013-02-21