

Twitter を用いた電車遅延の自動通知

築瀬 拓弥^{†1} 増田 英孝^{†1} 山田 剛一^{†1} 荒牧 英治^{†2} 中川 裕志^{†3}

本研究では電車の運行状況をリアルタイムに一般の Twitter のユーザのツイートから取得することを目的とする。対象として常磐線をキーワードとして含むパブリックタイムライン上のツイートを半年分収集した。このデータを用いて単位時間あたりのバース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を試作した。また、遅延や運転見合わせ時のユーザのツイートの特徴の分析を行った。

Automatic Notification for Train Delay from Twitter

TAKUYA YANASE^{†1} HIDETAKA MASUDA^{†1} KOICHI YAMADA^{†1}
EIJI ARAMAKI^{†2} HIROSHI NAKAGAWA^{†3}

Our goal is to extract train services from ordinary twitter users' tweets in real-time. We have collected tweets which includes '常磐線' from public time line over six months. We distinguish abnormal state from normal state by using a burst per minutes in the collected tweets, and we implemented a prototype system which can be notified abnormal state. Also, we analyzed the characteristic of users' tweets when trains are delayed or postponed.

1. はじめに

現代の通勤通学における一般的な交通手段として電車が挙げられる。しかし電車の問題点は、突発的な事故や点検による遅延や運転見合わせ等のダイヤの乱れである。駅に到着してから電車遅延や運転見合わせを知り困惑する人も少なくない。

このような電車の運行状況を知るために、実際に駅に行かなくても各鉄道会社のホームページで電車の運行状況を確認することができる。しかし各鉄道会社のホームページの情報は自分自身で確認をしなければ知ることはできない。また電車の運行状況を通知してくれるサービスもある。例えば、YAHOO! JAPAN 路線情報は遅延や運転見合わせの際にメールでダイヤの乱れを通知するというサービスを提供している。このサービスを利用することでユーザは自分で運行情報を調べなくてもダイヤに乱れが発生した際、自動的に情報を取得することができる。しかし、このサービスの問題点は実際の遅延や運転見合わせが発生してからユーザに通知がくるまでにタイムラグが生じてしまうことである。

電車が遅延した場合の人の流れをシミュレーションにより解析する試み[1]や、列車の遅延状況をシミュレーション

により解析する試み[2]が行われているが、リアルタイムな解析は行うことができない。

そこで本研究では突発的な電車の遅延や運転見合わせ等の運行状況をよりリアルタイム性のある情報として取得して通知することを目的とする。運行状況をリアルタイムに取得し通知することにより、電車を利用する人が遅延や運転見合わせ等によるダイヤの乱れを素早く察知し、別の路線を利用する、バスやタクシー等の代替交通手段を利用する等のアクションを取ることができる。

ソーシャルネットワーキングサービス(SNS)の一つである Twitter を利用して注目話題[3]やイベント[4][5]に関するツイートの抽出や解析を行う試みや、ツイートから地域の特徴を調査する試みが行われている[6][7]。我々も Twitter を用いて、よりリアルタイム性のある電車の運行状況を一般の Twitter のユーザのツイートから予測し分析を行う。

2. ツイートの分析

本節では、電車運行の定常時と異常時の差を分析するために開発したツールについて述べる。

2.1 分析ツールの目的

電車の異常時を自動的に検出するためには、定常時と異常時の差を知る必要がある。時系列データから異常を検出する手法も研究されている[8]。本研究では時系列データとその時に発言されたツイートの内容を利用する。

手始めに、路線名を含むツイートを取得し、定常時と異常時のツイート数の変化およびその内容の特徴を調査する。このために、単位時間あたりのツイート数のグラフ化機能、

^{†1} 東京電機大学 未来科学部
Tokyo Denki University School of Science and Technology for Future Life
^{†2} 東京大学 知の構造化センター
The University of Tokyo Center for knowledge Structuring
^{†3} 東京大学 情報基盤センター
The University of Tokyo Information Technology Center

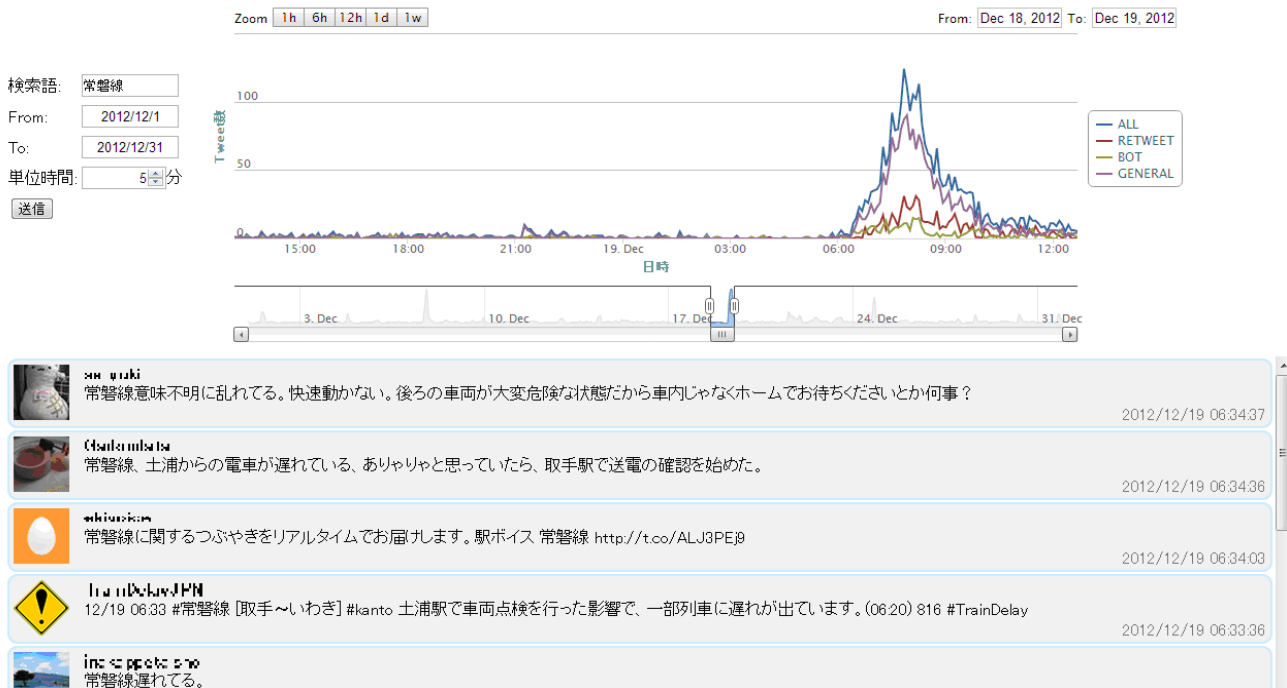


図 1 ツイート分析ツール

指定した期間のツイートの表示機能を持つツールを実装した。電車運行状況に関連するツイートの蓄積には Twitter API[9]と Java 用 Twitter ライブラリ Twitter4J[10]を用い、リレーショナルデータベースである MySQL で管理を行った。蓄積したツイートはパブリックタイムラインから「常磐線」を検索語として 2012 年 7 月 24 日から 2013 年 1 月 15 日に取得した 144,049 件のツイートである。

2.2 ツイート数のグラフ化

図 1 に示すように、本ツールは上部左の入力フォーム、上部右のグラフ表示部、下部のツイート内容表示部で構成されている。

図 1 上部右に示すように、本ツールは上部にユーザが指定した期間を X 軸、図 1 上部左の入力フォームで指定した単位時間あたりのツイート数を Y 軸とした折れ線グラフを表示する。グラフは全体のツイート数を計測するための「ALL」、Bot アカウントのツイートの速度や数を計測するための「BOT」、「ALL」のツイートの中でリツイートがどの程度の割合を占めるのか計測するための「RETWEET」、その他一般のユーザのツイート数を計測するための Bot によるツイートでもリツイートでもないツイート「GENERAL」の 4 種類のデータについて描画を行う。グラフを 4 種類に分類した理由は、常磐線に関するツイートの中で、Bot のツイートとリツイートがどのくらいの割合を占めるのか明確にするためである。グラフをこの 4 種類に分類し、表示することでユーザは数値として、また視覚的に Bot のツイートやリツイートが全体のどの程度の割合を占めるのかを知ることができる。

また本研究で Bot として扱う Bot アカウントは、パブリ

ックタイムラインから「常磐線」を検索語として 2012 年 7 月 24 日から 2013 年 1 月 15 日に取得した 144,049 件のツイートからツイート数の多い 100 アカウントを抽出し、目視にて各アカウントのツイート内容を確認して Bot であると識別した 98 アカウントである。

2.3 ツイートの表示

図 1 下部に示すように本ツールでは下部にツイートの内容を表示する。グラフのデータをクリックすることで、クリックした時間のツイートを一覧で表示する。詳細を知りたい時間のツイートを表示する操作を、グラフのデータをクリックするというアクションにすることにより、直感的に操作を行うことができる。

2.4 単位時間あたりのツイート数の推移

定常時と遅延や運転見合わせ等の異常時にツイート数に変化が生じるのかを調査するために、本ツールを使用し、単位時間を 5 分として電車運行状況に関するツイート数の推移を調査した。

定常時の「ALL」のツイートは概ね、0 件～5 件で推移し、「これから常磐線で帰宅」や「常磐線に乗り遅れた」等の日常的な内容のツイートや、「常磐線からスカイツリーが綺麗に見える」や「常磐線で酒盛りするサラリーマンが多い」等の路線の特徴を示すような内容のツイートが多く見られた。

遅延時や運転見合わせ等の異常時にはツイート数は増加し、ツイート数に大きな変化が見られることが確認できる。

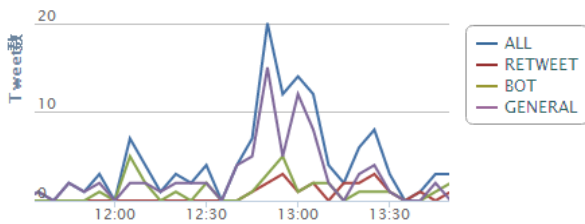


図 2 線路立ち入りによる遅延時のツイート数の増加

図 2 に示すように、2012 年 12 月 9 日に発生した線路への人立ち入りの影響による遅延では、12 時 40 分時点で 4 件だった「ALL」のツイートは 5 分後に 7 件に増加し、その直後に 20 件まで増加した。

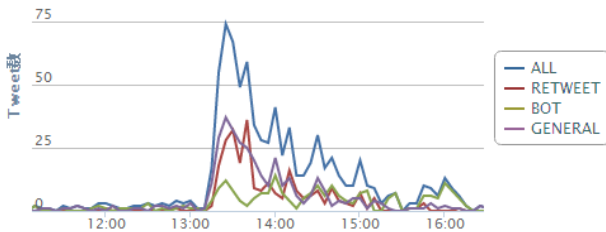


図 3. 人身事故による運転見合わせ時のツイート数の増加

また図 3 に示すように、2012 年 12 月 1 日に発生した人身事故による運転見合わせ時には 13 時 10 分時点で 1 件だった「ALL」のツイートは、人身事故発生直後の 13 時 15 分には 17 件まで増加し、その後 10 分後のピーク時には 74 件まで増加した。

このように遅延や運転見合わせ等の異常が発生した場合にはグラフにピークが確認できる。

しかし、遅延や運転見合わせ発生時以外でも、面白いネタツイートのリツイートや探しものや落し物に関する拡散ツイートによってグラフにピークが生じることがある。

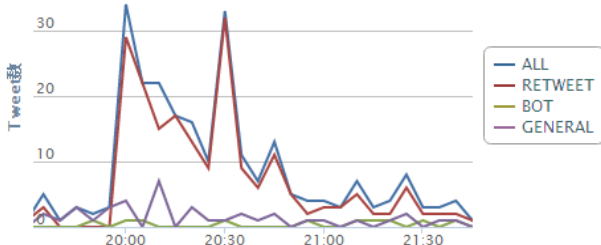


図 4. ネタツイートによるツイート数の増加

例えば、図 4 に示す 2012 年 11 月 2 日 20 時に発生したグラフの山では 5 分あたり 34 件「ALL」のツイートがされたが、34 件中 29 件は電車が止まる風力についての【簡単に分かる風力定義表】風力 1：常磐線が橋を渡るご

とに徐行 風力 2：武蔵野線が双六モードになる 風力 3：武蔵野線が死ぬ 風力 4：京葉線も死ぬ 風力 5 東西線・常磐線あたりまで死ぬ 風力 6：総武線が死ぬ 風力 7：京成が死ぬ（千葉という世界の終わり）」

といった内容のネタツイートのリツイートであった。

同じようにピークが確認できる場合でも、運行状況の異常に関するグラフのピーク時とネタツイートで発生するグラフのピーク時はリツイートを取り除くことで判別できることがわかった。

2.5 異常時ツイートの特徴

遅延や運転見合わせ等の異常時のツイート内容の特徴語を調査するために、2012 年 7 月 24 日から 2012 年 9 月 23 日の「常磐線」を含む 36,203 件のツイートに対して形態素解析を行い、調査を行った。形態素を品詞ごとに出現頻度順にランク付けし、遅延や運転見合わせに関係のある表現を人手によって抽出した。

表 1. 形態素解析結果

名詞		動詞	
語	出現回数	語	出現回数
運転	4,261	止まる	2,116
影響	2,668	見合わせる	1,303
遅れ	2,521	死ぬ	1,050
再開	1,567	遅れる	1,024
遅延	1,442	停る	941
点検	843	混む	747
見合わせ	715	座れる	651

その結果、表 1 に示すように遅延に関する名詞では「遅れ」や「遅延」が多く出現し、動詞では「遅れる」が多く出現した。運転見合わせに関する名詞では「見合わせ」が多く出現し、動詞では「見合わせる」が多く出現した。また遅延と運転見合わせに共通した語として「止まる」、「停る」、「死ぬ」という動詞が多く出現した。

3. 電車遅延の自動通知システム

本節では、前節で得られた知見を元に、電車の遅延や運転見合わせを路線情報に関するツイートから検出し通知するシステムについて述べる

3.1 自動通知システムの目的

ツイートから取得した遅延や運転見合わせ等の運行状況をユーザに素早く通知をするため、自動通知システムを試作した。本自動通知システムではユーザの持つ携帯端末に依存しないよう、JavaMail ライブラリ[11]を用いてメールでの通知として実装した。将来的にはスマートフォンのプッシュ通知機能を用いたアプリケーションとして実装を行う予定である。

3.2 遅延・運転見合わせの判定条件

自動通知システムで遅延と運転見合わせの判定を行うために、2012年12月の遅延と運転見合わせが発生した際のツイート数の調査を行った。

遅延と運転見合わせが発生した際の直後の Bot のツイートでもリツイートでもないツイート数は表2のようになった。Bot のツイートでもリツイートでもないツイート数について調べた理由は、一般のユーザによるツイートを元に遅延と運転見合わせを検出したいため、またリツイートを含めてしまうとネタツイートのリツイートによってツイート数がバーストした際にシステムで誤検出をしてしまうためである。

表2. 2012年12月に発生した異常時のツイート数

異常発生時間	ツイート数(件)
2012/12/01 13:20	8
2012/12/03 08:10	8
2012/12/04 18:30	6
2012/12/05 20:20	7
2012/12/06 16:35	8
2012/12/07 17:25	18
2012/12/08 00:05	8
2012/12/08 12:05	12
2012/12/09 12:55	15
2012/12/10 19:45	6
2012/12/11 17:45	6
2012/12/12 21:00	6
2012/12/13 15:45	9
2012/12/14 07:40	13
2012/12/14 17:50	10
2012/12/14 23:20	7
2012/12/15 05:00	6
2012/12/15 09:45	8
2012/12/19 06:30	10
2012/12/22 12:05	7
2012/12/24 17:45	18
2012/12/25 23:10	7
2012/12/27 17:45	5
2012/12/27 22:40	11
2012/12/30 06:15	6
2012/12/31 09:00	10

2.5 節で述べた遅延や運転見合わせ時のツイートの特徴と表2のツイート数から、遅延や運転見合わせの判定条件として「5分あたりの Bot のツイートでもリツイートでもないツイート数が6件以上」、遅延の場合には「5分あたりのツイートの内容に『遅延』または『遅れ』が含まれてい

る」、運転見合わせの場合には「5分あたりのツイートの内容に『見合わせ』が含まれる」を設定した。図5に通知条件とアルゴリズムを示す。

```

if(ツイート数 >= Th) {
    if(ツイート内容.contains("遅延")
        || ツイート内容.contains("遅れ")) {
        遅延通知;
    }
    else if(ツイート内容.contains("見合わせ")) {
        運転見合わせ通知;
    }
}
Thは閾値(実験ではTh=6とした)
    
```

図5. 通知条件とアルゴリズム

ツイート数の条件を「5分あたりの Bot のツイートでもリツイートでもないツイート数が6件以上」とした理由は、条件となるツイート数を5件以下とした場合に、定常時にも条件が満たされてしまう可能性が高いため、また表2の結果から8割以上の遅延と運転見合わせを検出するためである。

3.3 自動通知システムの試作

本自動通知システムを用いることで、収集したツイートデータから遅延や運転見合わせを検出し、通知を行うことを試みた。本システムの長所は、Twitter ユーザのリアルタイム性のあるツイートをを用いることで素早く遅延と運転見合わせを検出し通知することができることである。

しかし問題点も明らかになった。ツイートや電車利用者の少ない早朝や深夜などには利用する Twitter ユーザの母数が少ないために通知の条件を満たすことができず、遅延や運転見合わせが発生しているにもかかわらず、通知ができない点である。また、ツイート数が条件に達し、「学校に『遅れ』そう」といったツイートが出現してしまった場合には、遅延や運転見合わせが発生していなくても通知を行ってしまう点が挙げられる。このような改善点については今後研究を進め、より精度の高い通知条件を調査していく必要がある。

3.4 他サービスとの比較

本自動通知システムと既存のサービスの長所と短所を調査するために本自動通知システムによる通知時間と Yahoo! 運行情報による通知時間、また Twitter の Bot アカウントによる遅延や運転見合わせ発生時の最初のツイートとの時間の比較を行った。比較したデータは2012年12月のものである。

表3に示すように、本自動通知システムの通知及び Bot によるツイートは Yahoo! 運行情報の通知に比べ、およそ 10

表 3. 通知の時間比較

日付	種別	自動通知システム	Yahoo!運行情報	Bot ツイート	自動通知システムの勝敗
2012/12/01	運転見合わせ	+00:01	+00:04	13:19	×
2012/12/03	遅延	08:10	+00:16	+00:13	○
2012/12/04	遅延	—	+00:03	09:30	×
2012/12/04	運転見合わせ	+00:03	+00:16	18:27	×
2012/12/05	運転見合わせ	20:20	+00:03	+00:03	○
2012/12/06	運転見合わせ	16:35	—	16:35	○
2012/12/07	運転見合わせ	+00:02	—	17:23	×
2012/12/08	遅延	00:05	—	+00:02	○
2012/12/08	運転見合わせ	12:05	+00:15	+00:05	○
2012/12/08	遅延	—	18:03	—	×
2012/12/09	運転見合わせ	+00:03	—	12:52	×
2012/12/10	遅延	+00:01	—	19:44	×
2012/12/11	運転見合わせ	17:45	+00:20	+00:01	○
2012/12/12	運転見合わせ	+00:05	—	20:55	×
2012/12/13	運転見合わせ	+00:04	—	15:41	×
2012/12/14	運転見合わせ	07:40	+00:22	+00:01	○
2012/12/14	運転見合わせ	+00:05	—	17:45	×
2012/12/14	運転見合わせ	—	20:17	—	×
2012/12/14	遅延	23:20	—	—	○
2012/12/15	運転見合わせ	+00:25	04:35	+00:20	×
2012/12/15	遅延	+00:05	—	09:40	×
2012/12/19	遅延	06:30	+00:39	06:30	○
2012/12/22	運転見合わせ	12:05	—	12:05	○
2012/12/22	遅延	—	21:53	+00:13	×
2012/12/24	運転見合わせ	17:45	+00:12	+00:02	○
2012/12/25	運転見合わせ	23:10	+00:14	23:10	○
2012/12/27	遅延	+00:02	—	17:43	×
2012/12/27	運転見合わせ	+00:04	—	22:36	×
2012/12/30	運転見合わせ	+00:03	06:12	06:12	×
2012/12/31	遅延	09:00	—	+00:05	○
2012/12/31	遅延	—	18:57	+00:08	×

分から 30 分程度早く遅延や運転見合わせを検出することができた。表 3 の符号なし部分は最速のシステムまたはサービスの通知時刻を表し、+は最速のものからの遅れを分:秒で示し、-は通知できなかったことを示す。しかし Yahoo! 運行情報の通知には運転再開見込みや具体的な遅延時間等、詳細な情報が含まれており 1 つの通知情報としては最も多くの情報が含まれていた。

また本自動通知システムと Bot によるツイートを比較した場合、Bot によるツイートがやや早い傾向があった。ただし、本調査においては単一の Bot ではなく、目視で確認し登録した 98 件の Bot アカウントによるツイートの中から

一番早いツイートを遅延や運転見合わせ発生時の最初のツイートとした。そのため、遅延や運転見合わせ発生時の Bot アカウントの最初のツイートは毎回同じアカウントによるものではない。よって、一般の Twitter ユーザが同様の情報を得るためには同じく 100 件近い Bot アカウントをフォローしなければならない。ユーザが 100 件近い Bot アカウントをフォローしなければならない点を考慮すると、一つの情報ソースとしての本自動通知システムは Bot によるツイートにやや速度は劣るが、容易に遅延や運転見合わせ等の通知情報を取得することができる。

各サービスを本自動通知システムと比較した場合の

2012年12月の異常時における平均通知時間は表4のようになった。表4では、本システムを基準として、遅かった場合を+分:秒で示し、早かった場合を-分:秒で示している。Yahoo!運行情報の平均通知時間は、遅延と運転見合わせに共通して本自動通知システムより遅くなり、Botのツイートの時間は本自動通知システムの通知時間よりやや早かった。最終的な本自動通知システムと他サービスとの通知速度の勝敗数は表5のような結果となった。

また、通知の時間比較結果の表3から本自動通知システムとYahoo!運行情報、Botに共通して、遅延や運転見合わせの検出漏れが発生することがわかった。このことから、複数の電車運行情報通知サービスを組み合わせることでより漏れのない遅延や運転見合わせの情報を取得できることがわかった。

表4. 自動通知システムと他サービスの平均通知時間の差

種別	Yahoo! 運行情報	Bot ツイート
遅延	+00:27.5	+00:01.7
見合わせ	+00:07.4	-00:01.2

表5. 自動通知システムの通知状況

異常発生 回数	最速通知 回数	最速でない 通知回数	未検出
31	13	13	5

4. おわりに

本研究ではTwitterを用いた電車の遅延や運転見合わせ時のツイートの分析を行い、その分析で得た結果を用いて遅延と運転見合わせの判定、及び自動通知方法を提案した。

結果として、一般ユーザのツイート量が遅延や運転見合わせをきっかけにパーストし、対象路線では定常時の0~5件前後から6件以上に増加すること、遅延や運転見合わせ時のツイートの内容の特徴として「遅延」や「遅れ」、「見合わせ」等の語が多く出現することがわかり、電車の遅延や運転見合わせを自動通知システムで検出することができた。

今後の研究課題として以下の2点が挙げられる。

1 つ目は遅延や運転見合わせ感知の正確性の向上である。遅延や運転見合わせ時のツイートを形態素解析したところ、「遅延」や「遅れ」、「見合わせ」と言った語の他に、「点検」や「事故」などの遅延や運転見合わせの原因に関する語が現れた。原因を特定し分類することで、遅延や運転見合わせをより正確に検出し、また原因によって遅延や運転見合わせと判断する条件を変化させることでより正確な検出と通知が出来る。

2 つ目は遅延や運転見合わせ時の他路線推薦である。本

研究では路線を常磐線に特定し、研究を進めた。今後、常磐線以外の路線に関して同様の研究を進め、各路線の運行状況をTwitterから取得し、その分析結果を組み合わせることで、1つの路線が遅延や運転見合わせになった際に、平常通り運行している他路線を推薦することが出来る。

謝辞 本研究の遂行にあたり、終始ご指導とご助言を頂きました東京大学情報基盤センター吉田稔助教、佐藤一誠助教ならびに東京大学知の構造化センター宮部真衣特任研究員に心より感謝いたします。

参考文献

- 1) 大山大輔, 原和弘, 古関隆章: 判断の因果関係を正しく表現する列車運転異常時の乗客流推定法, 平成19年電気学会全国大会, (2007)
- 2) 中村幸史, 田口東: 通勤電車運行スケジュールにおける遅延計算モデルの構築, 日本オペレーションズ・リサーチ学会春季研究発表会アブストラクト集 2004, pp.142-143 (2004)
- 3) 木原大志, 白木原渉, 越村三幸, 藤田博, 長谷川隆三: Twitterの時系列解析による注目話題の抽出, 情報処理学会第74回全国大会講演論文集, pp.625-627 (2012)
- 4) H Becker, M Naaman, L Gravano: Beyond Trending Topics Real-World Event Identification on Twitter, Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, pp.438-441 (2011)
- 5) Jianshu Weng, Yuxia Yao, Erwin Leonardi, Francis Lee: Event detection in twitter, Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, pp401-408 (2011)
- 6) 東口大樹, 仲野雅幸, 佐野博之, 白松俊, 大園忠親, 新谷虎松: ユーザ間のTweet頻度偏り補正に基づくローカルバースト検出手法, 情報処理学会第73回全国大会講演論文集, pp.369-371 (2011)
- 7) Mor Naaman, Amy Xian Zhang, Samuel Brody, Gilad Lotan: On the Study of Diurnal Urban Routines on Twitter, Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media, pp.258-265 (2012)
- 8) 櫻井保志: 時系列データのためのストリームマイニング技術, 情報処理 Vol.47, No.7, pp.755-761 (2006)
- 9) Twitter Developers, <https://dev.twitter.com/>, 2012-01-28 アクセス
- 10) Twitter4J, <http://twitter4j.org/ja/index.html>, 2012-01-28 アクセス
- 11) JavaMail, <http://www.oracle.com/technetwork/java/javamail/index.html>, 2012-01-28 アクセス