

タイミングを考慮した将来イベントリマインダー

古賀 健士郎†

庄司 功樹‡

小林 亜樹†

† 工学院大学工学部情報通信工学科

‡ 工学院大学大学院工学研究科

1 はじめに

情報ツールとして普及した Twitter は、多数の人をフォローすることで、それらの人々のツイートが自らのタイムライン上に流れる構造となっている。一方、ツイートには将来情報も含まれ、これを自らの予定に反映させるような使い方もされている。しかし、多数のツイートが流れるため見落としがちな場合もあり、また、予定を忘れてしまうことも考えられる。そこで本稿では、タイムライン上を流れる将来イベントを自動的にリマインドするシステムについて提案する。

2 提案システム

ツイートに日付に関する表現があると、その日に「イベント」又は「予定」があると推測できる。このような、ツイートに日付に関する表現を含み、かつ、その日付が未来であると判断されるものを「将来情報ツイート」(Future Information Tweet: 以下 FIT と記す) と呼び、FIT に記述されている未来の日付を未来日時 t_f とする。すなわち、FIT $\mathcal{F}$ は、投稿日時 t_c と含んでいる未来日時 t_f を持つから、 $\mathcal{F}(t_c, t_f)$ と表記する。提案システムでは、これをリマインドするために適当な時期に再度タイムライン上に現れるよう、通知する。このタイミングは、 t_c から t_f までの期間 $d = t_f - t_c$ に依存する。一方、通知日は t_f にも依存し、かつ、複数回を許容するため、通知日集合 R は、これを計算する関数 r を用いて

$$R(\mathcal{F}) = r(t_c, t_f) \quad (1)$$

のようになる。

システムの動作は、日付表現の抽出により FIT $\mathcal{F}$ を検出し、 $R(\mathcal{F})$ を計算し、その各通知日に $\mathcal{F}$ を通知する。具体的には

1. API の `home_timeline` メソッド*を利用してシステム利用者のタイムラインを取得する。
2. ツイートから日付表記や日付を推定できる表現を抽出し、FIT を検出すると共に t_c と t_f を推定する。
3. t_c から t_f までの期間の長さを基に $R(\mathcal{F})$ を求める。

Considering the timing for Future Event Reminder

†Kenshiro Koga ‡Kouki Syouji

†Department of information and Communications Engineering, Faculty of Engineering, Kogakuin University

‡Kogakuin University

*アプリケーションを認証したユーザのタイムライン (ユーザのツイート・リツイート、ユーザがフォローしている人のツイート・リツイートが流れているもの) を取得する

4. $R(\mathcal{F})$ の時刻に TwitterAPI の `update` メソッド†を用いてシステム利用者としてシステムが FIT の内容をツイートする。

以上の動作により、システム利用者のタイムライン上に再び FIT を出現させ、再認識してもらう。

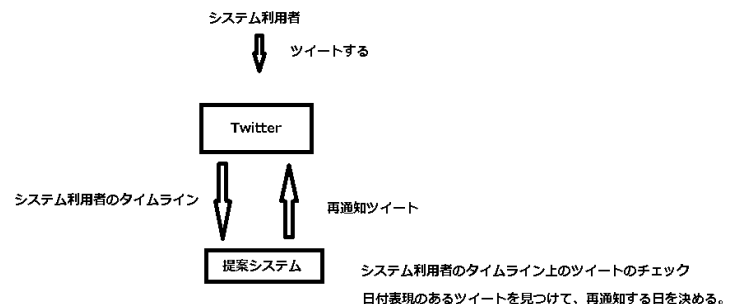


図 1: システムの概要

3 日付表現の抽出

3.1 日付表現

Twitter 上で一般ユーザ同士がやり取りを行って予定を決めたり、企業が告知などに利用する際に、FIT を多く含む傾向がある。特に企業の FIT では「〇月〇日」や「〇/〇」といった絶対日付表現をしていることが多い。他にも、「来週〇曜日」や「〇日後」といった相対日付表現がある。

絶対日付表現においても、年までは記述していない例が多い。この場合、FIT の投稿日時から一年以内の日付を表していると考えられるので、年を補完し、相対日付表現の場合には FIT の持つ t_c を基準に推測することで、 t_f を推定した。このため、 $d \leq 300$ とする。また、月だけの表現や日だけの表現では十分に t_f を推定できないので、対象外とし、複数の日付表現がある場合には先頭の日付表現のみを抽出対象とした。

表 1: 日付表現の例

絶対表現	平成 24 年 1 月 1 日, 一二月一五日, 12/15
相対表現	来週, 来週月曜, 来月の 15 日, 10 日後

3.2 抽出手法

日付表現は固有表現の一形態であり、固有表現抽出に関する研究は数多く、「NEXt」[1] として知られる固

† ツイートを投稿するメソッド。アプリケーションを認証したユーザとしてツイートすることが出来る

有表現抽出ツールでは、固有表現と判断された箇所にはタグ付けされる。これを実際の FIT について試用してみたところ、「〇月〇日」「来週」といった表現には日時表現としてのタグ付けがされたが、「〇／〇」「木曜日」などにはタグ付けされなかった。このため、現在公開されている版では、本研究の目的には不適切である。

そこで、形態素解析によって形態素に分割した後、名詞を連結して複合名詞を抽出する手法 [2] を用いた。また、連続する名詞の連結だけでなく、連体詞「の」によって説明される表現[‡]も抽出可能とするために、これらも連結し一つの語句として扱う。その後、正規表現を用いて抽出した。

4 通知タイミング

FIT の投稿日時と推定されたイベント日時から、イベントまでの期間がわかるが、それをいつ通知するかは、重要な案件である。大学生 8 名、大学院生 1 名に「1~300 日後にイベントがあるとしたら、イベントの何日前に通知してほしいか」アンケートを取り、アンケート結果をグラフにしたものを下に示す。

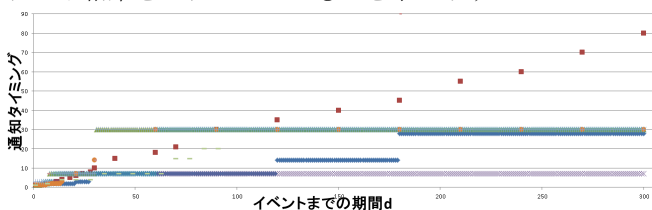


図 2: イベントまでの期間と通知タイミング

このグラフでは、通知タイミングの変化が横ばいになっている部分を青色で表し、各人のほぼ線形に変化している部分を各色で表している。図 2 から、ほぼ線形に増加し続ける場合と、ある値から横ばいになっている場合があるのがわかる。9 人中 7 人が頭打ちになる形をしており、内 5 人が 30 付近の値で横ばいになっていた。このことから、線形の部分と横ばいの部分で通知のタイミングを分けることにした。

各人の線形に変化している部分で回帰分析を行い、その結果の平均を取ることににより、 t_f の何日前に通知するかというタイミング $n(d)$ を求めた。

$$\begin{cases} n(d) = [0.28d + 0.99] \\ \text{ただし, } d \geq 2 \text{ のとき} \end{cases} \quad (2)$$

ここでの $d \geq 2$ という条件は、 $d = 1$ とは、すでに前日なので必要ないと判断し、 $d \geq 2$ と条件を付けた。

イベントまでの期間 d が長いと、通知後からイベントまでの期間 $n(d)$ も長くなる。その間にイベントを忘れてしまっは意味がないので、通知した時を初めて FIT がツイートされた時と同様に考え、 $n(d)$ より、もう一度通知するタイミングを決める事を繰り返す。し

かし、再度の通知では中一日での通知は頻度が高く好ましくないので、 i 回目 ($i \geq 2$) の通知タイミングは、

$$\begin{cases} n^i(d) \\ \text{ただし, } i \geq 2 \text{ かつ } n^{i-1}(d) - n^i(d) > 1 \text{ のとき} \end{cases} \quad (3)$$

と定めた。

横ばいの部分は、一番多かった 30 の値を採用し、式 (2) (3) と通知の頻度を考慮することにより、

$$\min\{|n^i(d) - 30|\} \geq 2 \quad (4)$$

のとき、30 日前の通知タイミングでの通知を行う。これをまでの通知タイミングを図示すると、図 3 のようになる。

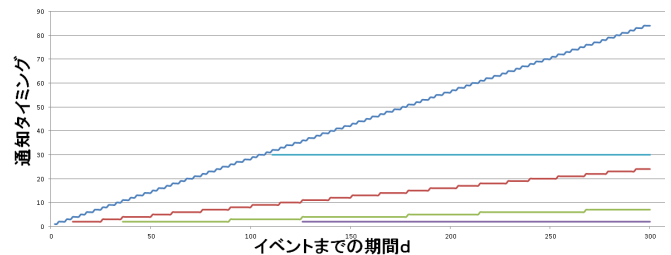


図 3: イベントまでの期間と通知タイミング

以上により、通知日 $r(t_c, t_f)$ は

$$r(t_c, t_f) = \begin{cases} t_f - n(d) & (d \geq 2) \\ t_f - n^i(d) & (i \geq 2, n^{i-1}(d) - n^i(d) > 1) \\ t_f - 30 & (\min\{|n^i(d) - 30|\} \geq 2) \end{cases} \quad (5)$$

となる。

5 動作例

FIT $\mathcal{F}$ として、次のツイートを抽出すると。

【2012 年の主な天体ショー一覧】 3 月 6 日 火星最接近 5 月 21 日 金環日食 6 月 4 日 部分月食 6 月 6 日 金星の太陽面通過 8 月 12 日 ペルセウス座流星群の極大 9 月 30 日 中秋の名月 12 月 14 日 ふたご座流星群の極大

t_c : 2012 年 1 月 9 日 15:56, t_f : 3 月 6 日

$t_f - t_c = 57$

$R(\mathcal{F})$: 3 月 1 日, 2 月 19 日

のように動作する。

6 おわりに

ユーザの忘れ防止、思い出しをさせることが出来るリマインダーシステムを提案した。システムのもたらす効果に関する評価や、現在抽出対象になっていない日付を推測できる表現や曖昧な表現などに対応することなどが今後の課題である。

参考文献

- [1] 渡辺一郎, 榊井文人, 福本淳一: "固有表現抽出ツール NExT の精緻化とユーザビリティの向上", 第 10 回言語処理学会年次大会発表論文集, pp.413-415, 2004.
- [2] 中島泰, 黒木さやか, 櫻井宏樹, 山名早人: "ブログにおける話題語の出現理由の抽出と話題に関する詳細記事推薦", IECE W12-2009-28 45, pp1-6, 2009.

[‡] 「来週」「の」「木曜日」: 名詞のみの連結。「来週の木曜日」: 名詞と連体詞「の」の連結