

Twitter の流行語発言者の抽出に基づく フォロワー推薦システムの開発

白木原 渉† 倉門 浩二†† 大石 哲也††† 越村 三幸††† 藤田 博††† 長谷川 隆三†††
†九州大学工学部電気情報工学科
††九州大学システム情報科学府
†††九州大学システム情報科学研究所

1 はじめに

ここ十数年の技術の向上により、インターネットは、情報検索のツールとして大きく発展してきた。欲しい情報があるとき、多くの人がインターネットの情報検索エンジンに頼っている。しかし、情報検索エンジンで検索される情報は、誰かが更新するまでは、もとの情報のままであり、最新の情報、特に流行している事柄を検索するのは難しい。

昨今では SNS(Social Network Service) が普及している。SNS とは、友人・知人間のコミュニケーションを促進する手段や場、あるいは新たな人間関係を構築する場を提供するサービスである。SNS は主に情報交換の場として機能しており、情報検索エンジンで情報を検索するよりも速く正確に、目的の情報を得られることが多い。しかし SNS では基本的に、築かれた人間関係(ネットワーク)の中でしか情報交換できず、欲しい情報を持っている人間とのネットワークを作らなければ意味をなさなかった。

近年、SNS の一種である Twitter が急激に普及し始めた。今までの SNS と違うのは、mixi を例にとると、自分からネットワークを作らなくても、すべての人の書き込んだ情報が見れることである。しかも、その書き込みはリアルタイムで更新されるため、常に最新の情報が手に入る。さらに、Twitter では世の中で流行している事柄(流行語)について、多くの人が発言する傾向がある。流行に敏感な人を知り、その人の書き込みを注目することで、最新の流行している事柄についての情報を簡単に手に入れることができる。本研究では、そのような人を推薦することで、さらに情報を得やすくするシステムを考案する。

2 節では Twitter, 3 節ではシステムの概要, 4 節では実験について述べる。

2 Twitter

Twitter(ツイッター)とは、個々のユーザーが 140 文字以内の「つぶやき」を投稿する、いわば「ミニブログ」である。他のユーザーを登録(フォロー)すると、その人(フォロワー)が書き込んだつぶやきを読むことが

できる。また、自分がフォローされることもある。フォローした人のつぶやきは、書き込まれた順番に「タイムライン」と呼ばれる、自分の専用画面に表示される。さらに、誰が誰をフォローしているかが一覧でき、相手にメッセージを送ってコミュニケーションをとることができる。このような特徴から、Twitter を「情報交換のためのツール」として利用しているユーザーが多い。また、多くのユーザーが世の中で流行している事柄についてつぶやきを投稿している。したがって、このような「流行に敏感な人」を知ることができれば、その人をフォローすることで、最新の流行についての情報を知ることができる。

他の特徴として、Twitter のほぼ全ての機能(つぶやきの投稿、ユーザーやつぶやきに対する検索など)に対応する API(Application Program Interface)がある。これを利用した多くの関連サービスが公開されており、これも Twitter 普及の一因となっている。その Twitter の関連サービスの一つに Buzztter がある。Buzztter は、Twitter で多くの人がつぶやいた語を流行語とみなし、それを抽出してユーザーに提供するサービスである。この Buzztter から流行語を得て、それをクエリとして Twitter で検索することによって、流行語の発言者を知ることができる。

3 システムの概要

我々は、Twitter 上での流行語を抽出するサービス(Buzztter)より流行語を得て、それをクエリとして Twitter の API を用いて検索し、大勢の Twitter ユーザーの中から、より多くの種類の流行語を発言し、かつより早く発言している人を数人抽出するシステムを考案する。そのような人たちを「流行に敏感な人」とする。以上のことを繰り返し、データベースに「流行に敏感な人」のユーザー名やプロフィールなどの情報を蓄積していく。その際、その人が、どの分野の流行に敏感なのか(例えば、音楽や映画など)をその人が発言した流行語から判断し、カテゴリ分けを行う。

ユーザーが、与えられたカテゴリから自分の探したい情報に合うカテゴリを選択すると、システムがそのカテゴリに含まれる「流行に敏感な人」の情報を返す。

ユーザーは、その中から好きなだけ選び、その人をフォローすることで、目的の情報を得る。

システムの動きを具体的に以下に示す。

1. Buzztter から、HTML パーサを用いて、流行語 t を得る。
2. Twitter の API を用いて、流行語 t を含むつぶやきを検索し、そのタイムライン (ユーザー名とつぶやいた時刻) を得る。
3. 流行語 t を含むつぶやきが急増する時間帯をタイムラインから探し、その時間帯の中で最も早くつぶやいた数名を、(流行語 t に対して) 流行に敏感な人 p_1, p_2, p_3, \dots とする。
4. 1~3 を繰り返す。また、流行に敏感な人の総人数 m と、流行語の総数 n を得る。
5. ここで、 m 行 n 列の行列を用意する。行列の要素は bool 値で、人 p_i が流行語 t_j に敏感だったならば要素 $[i][j]$ を 1、さもなければ 0 とする。
6. 行列の i 行の要素の総和を、人 p_i の流行に対する敏感さ r_i とする。
7. r_i の値が大きかった数名を、「流行に敏感な人」とする。

4 実験

4.1 実験方法

3 節で説明したシステムで得られた「流行に敏感な人」が、その後もより多くの種類の流行語を発言し、かつより早く発言しているかどうかを確認することで評価実験とする。具体的には、システムで得られた「流行に敏感な人」のその後の発言に対して形態素解析を行い、その発言から名詞を取り出す。その各名詞の中に、Buzztter から得られた流行語と合致する語があるかを確認し、あった場合は、その流行語をクエリとして Twitter で検索し、その流行語が多くのユーザーに発言された時間帯のどのタイミングで、「流行に敏感な人」のもとの発言がなされたかを調査する。

4.2 予備実験

実際に、流行に敏感な人が得られるかどうかを確認するため、簡単な予備実験を行った。

12 月 9 日、Buzztter より流行語「google chrome」を得た。API で検索したところ、多くの発言は 12 月 8~9 日にされていたが、最も早い発言は 12 月 2 日にされていた。以下の文章は、そのときの実際の発言 [4] である。

Wed, 02 Dec 2009 11:21:17 +0900

google が Mac 版 Chrome のベータ版をそろそろリリースするそうです。Win 版より機能制限があるらしいけど、そこまですて Mac 版を出す意味があるのか。ブラウザなら Safari で十分だし、あえて Chrome に乗り換えるにはなにかメリットがなきゃね。

さらに、その人の別の発言を検索した。

新 Opera がサーバー機能を有したらしい。自分の PC にあるデータをそのまま公開できるらしい。アクセス権を個別に設定できるので知り合い同士のデータ共有が可能に。ファイル容量の制限がないらしいので何でもデータで逝けちゃうんですね。1GB のファイルを共有して実用に耐えうるかは疑問。

以上のように、ブラウザ「Opera」の新しいバージョンについての情報が得られた。この発言が誰よりも早いものかどうかは調査していないが、このシステムを利用することで、流行についての情報が手に入ることがわかった。

5 おわりに

システムの大まかな流れは考案したが、流行に敏感な人たちをどのようにしてカテゴリ分けするか、流行語 t を含むつぶやきが急増する時間帯の検出のしかたを詳細まで考案していない。

予備実験の結果から有効性が期待できるので、以上 2 点の問題を解決したうえで、これから実際にシステムを構築していき、流行の検索が簡単にできるようなインターフェースを作りたいと考えている。

謝辞 本研究は科研費 (21500102) の助成を受けたものである。

参考文献

- [1] 神田 敏晶, “Twitter 革命” ソフトバンク, 2009
- [2] コグレ マサト, いしたに まさき, “ツイッター 140 文字が世界を変える” 毎日コミュニケーションズ, 2009
- [3] 大力 慶祐, 大向 一輝, 武田 英明, “ソーシャルブックマークにおけるイノベータに注目した情報推薦手法の提案” The 22nd Annual Conference of the Japanese Society for Artificial Intelligence, 2008, 2D3-3
- [4] <http://twitter.com/F15sBase>