

# Twitter 投稿情報を活用した記念日お祝い支援システム An anniversary celebration support system using posted information to Twitter

越野 亮<sup>\*1</sup>

Makoto Koshino

<sup>\*1</sup> 石川工業高等専門学校  
Ishikawa National College of Technology

In this study, we have developed an anniversary celebration support system, which helps to think about anniversary gift. The system extracts preference information such as where a person wants to go and what a person wants from a person has been posted to Twitter in the past by morphological analysis and TF-IDF method, which informs by the tag cloud.

## 1. はじめに

Twitter はサービスの月間アクティブユーザー数が 2 億を超え、急速に普及している。Twitter ユーザーは現在の心境や今やっていることなどツイート(投稿)しているが、特に「欲しい物」や「食べたいもの」などの嗜好情報を含んだものとしてツイートされることが多い。本研究では、このような Twitter に投稿された嗜好情報を利用した推薦システムとして、友人や家族の記念日にプレゼントをする際に参考となるシステム「記念日お祝い支援システム」Loving it!(図1)を開発した。

本研究のように Twitter からユーザーの嗜好情報を抽出する研究は広く行われており特に新しくはないが、本研究の特徴としてはキーワード抽出の TF-IDF を使ってキーワードをタグクラウドで表示したところである。一見関係ないキーワードばかり表示されるが、その中にいくつかはプレゼントを選ぶ参考になるキーワードが表示されることがある。これを「欲望語」と名付けたいいくつかの用語検索と結びつけて表示するところがポイントである。

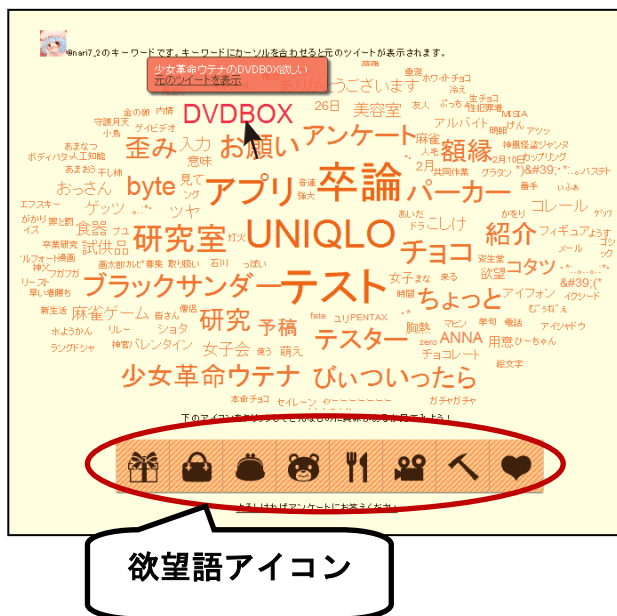


図1 記念日お祝い支援システム

## 2. 嗜好情報の抽出

### 2.1 全体の流れ

嗜好情報の抽出方法について大まかな流れを図 2 に示す。まず、形態素解析によりツイートから名詞を抜き出す。次に TF-IDF 法により、特徴的な名詞に重みをつけることで、個人のキーワードを抽出する。そのキーワードと共につぶやかれている、「欲しい」「行きたい」といった、個人の欲望を表す語句(以下、「欲望語」)を元に、個人が特に興味を持っているものを嗜好情報として抽出する。

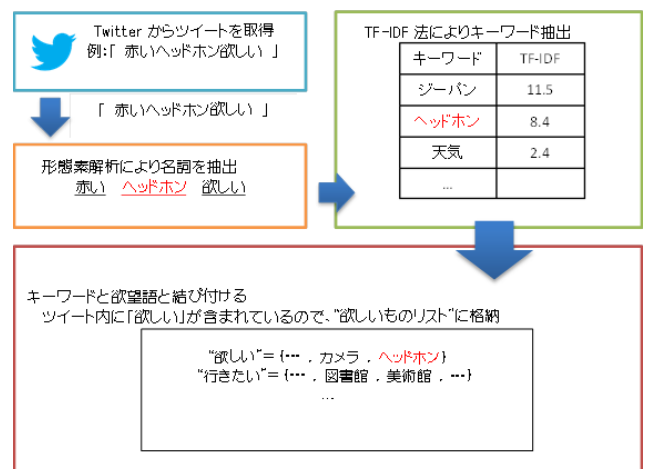


図2 全体の流れ

前処理として、取得したツイートから、他の人との返信等に使われる「@付きの TwitterID」、「#マーク付きのハッシュタグ」、「http://(もしくは https://)から始まる URL」を除外する。これは、結果に TwitterID などの欲望の対象になり得ない単語が抽出されるのを防ぐためである。

### 2.2 形態素解析

形態素解析を行い、名詞と判断された単語を抽出する(図 3)。本システムでは、オープンソースの形態素解析エンジン「MeCab」の Python バインディングを用いて、形態素解析を行なう。

連絡先: 越野 亮, 石川工業高等専門学校,  
〒929-0392 石川県津幡町北中条, 076-288-8132,  
koshino@ishikawa-nct.ac.jp

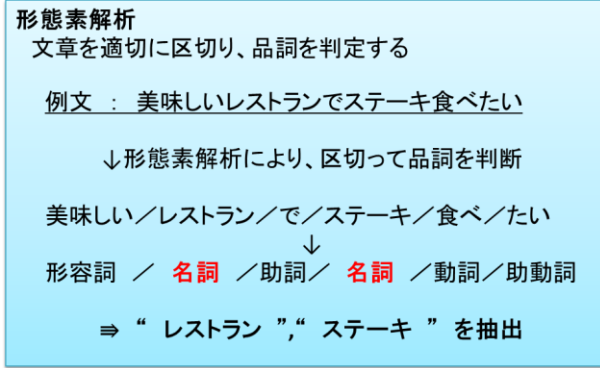


図3 形態素解析

MeCab では、IPA辞書(mecab-ipadic)がデフォルトである。この辞書では、人名などの固有名詞の形態素解析が難しいため、名詞の解析に特化したユーザ辞書として、Wikipedia データベース(<http://dumps.wikimedia.org/jawiki>)と、はてなダイアリーキーワードふりがなリスト([http://d.hatena.ne.jp/images/keyword/keywordlist\\_furigana.csv](http://d.hatena.ne.jp/images/keyword/keywordlist_furigana.csv))から csv ファイルをダウンロードし、MeCab の辞書形式に整形することでユーザ辞書を作った。現在の Wikipedia データベースは約 110 万語、はてなダイアリーキーワードふりがなリストは約 27 万語であった。

Wikipedia のデータベースより生成したユーザ辞書には、「こんにちは」や「ありがとう」などの挨拶や、「熱い」、「可愛い」、「大きい」などの形容詞が含まれている。このような、ユーザの欲望と結びつかない語については、「除外ワードリスト」を作成し、解析の対象から除外した。

### 2.3 TF-IDF 法

ツイートから出現した名詞がどれだけ特徴的であるかを、TF-IDF 法(図 4)により求める。本システムにおいて、出現頻度 TF は最新の個人のツイート 200 件における単語の「出現頻度」を用いる。最新の 200 件を取得するのは、最近の興味や趣味を取得するためでもあるが、Twitter API(ver.1.0) の仕様で、一度で取得できる最大の件数が 200 件であるためである。

IDF の計算には、パブリックタイムラインから取得したツイート 19,644 行を利用した。「全文書数」を 19644 行とし、「その単語が含まれる文書数」は、その単語が含まれるツイートの行数とした。パブリックタイムラインのツイートを形態素解析し、単語とその IDF 値をカンマ区切りでまとめたファイルを作った。実際のアプリケーションではこのファイルを読み込んで、TF と乗じることで TF-IDF 値を導出した。

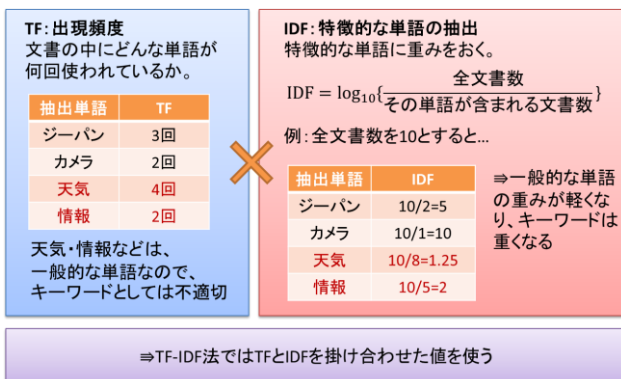


図4 TF-IDF 法

### 2.4 欲望語との結びつけ

今までの手順によりユーザ特有のキーワードを求めることができるが、キーワードだけでは、ユーザがその単語についてどう思っているかという情報を得ることができない。そこで、ツイート内に、「欲しい」や「行きたい」といった欲望語が含まれているかどうかを調べ、欲望語とともにつぶやかれている単語を、ユーザの欲望としてリストに格納した。本アプリで取り扱った欲望語は「欲しい」、「行きたい」、「買いたい」、「可愛い」、「食いたい」、「見たい」、「作りたい」、「大好き」の8つである。

### 3. 本システムの説明

Web システムの画面遷移を図 5～図 8 に示す。ページにアクセスするとまずトップページ(index.html)が表示される(図 5)。Twitter ID を入力し、送信ボタンをクリックすると、解析された結果画面(reslut.html)が表示され(図 6)、単語をクリックすると Amazon の検索結果が表示される。表示されたタグクラウドの単語にカーソルを合わせると、単語の元となったツイートが吹き出しに表示される(図 7)。タグクラウドの下にあるイラストアイコンをクリックすると、欲望語に応じて単語の色が変わる(図 8)。



図5 トップページ



図6 結果画面

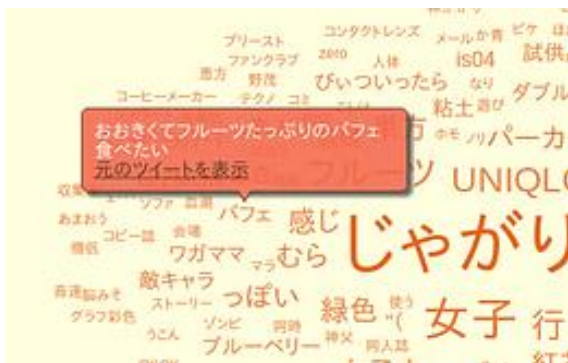


図7 「パフェ」にカーソルを合わせた時の表示



図8 「欲しい」のアイコンをクリック

## 4. 評価

### 4.1 評価方法

普段 Twitter を利用しているユーザ 43 名に本システムを利用してもらい、アンケートを実施した。このシステムでは、友人の Twitter ID を入力するのが本来の使い方だが、本評価では自分の Twitter ID を入力して結果を見てどれだけ自分の意思と合っているかを確認して頂いた。アンケートは、Google ドキュメントのフォーム(図9)を用いて行なった。



図9 実施したアンケート

### 4.2 アンケート結果

本システム使用後に実施したアンケートの結果を以下のとおりである。

質問1. 友人の誕生日や記念日に、何をプレゼントしようか、どうやってお祝いしようかと考えたことがあるか

ある	ない
88%	12%

この質問については、88%の被験者が、「考えたことがある」を選択した。この結果から、本システムの必要性が伺える。

質問2. タグクラウドに自分らしいキーワードが表示されたか

はい	どちらかといえ ばはい	どちらかといえ ばいいえ	いいえ
60%	38%	2%	0%

「はい」、「どちらかといえはい」を答えた被験者が 98%を占めた。この結果から、特徴的な単語が、キーワードとして抽出されているということが確認できた。

質問3. イラストアイコンを押したときに、自分の欲しいと思ったものや、行きたい場所などの色が変わったか

変わった	変わらなかった
50%	50%

「変わった」と「変わっていない」がちょうど半分であった。色が変わらないということは、本研究で考えた8つの欲望語をツイートしていなかったということである。イラストアイコンの欲望語について吟味する必要がある。

質問4. 友人がその結果を見たときに色が変わった単語から、自分の好きなものや興味のあるものが連想できそうであるか また、元のツイートを見て、どうだったか

連想できる	元のツイートを見れば連想できる	連想できない
19%	62%	19%

「元のツイートを見れば連想できる」と答えた人が 62%と一番多かった。欲望語により抜き出された単語について、不適切なものが抜き出されていたことが考えられる。欲望語に問題があるのではないかと考えられる。

質問5. このシステムは友人の記念日をお祝いするのに役立つそうか

はい	どちらかといえ ばはい	どちらかといえ ばいいえ	いいえ
23%	47%	23%	7%

70%の被験者が、「役立つそうだ」「どちらかといえれば役立つそうだ」と答えたのに対し、30%のユーザが「役立つそうではない」「どちらかといえれば役立つそうにない」と答えた。今までのアンケートを見る限り、欲望語を改善することで実用性が上がるのではないかと考えられる。

#### 4.3 考察

本システム「Loving it!」にはイラストアイコンをクリックすることで欲望に関連する単語の色が変わる機能が付いている。この欲望語は「友人の記念日を祝う際に、参考になるだろうと思われる欲望語」を選定したものであり、一般的に必要な欲望語について考察する。そこで、4.1のアンケートと同様に、Googleドキュメントのフォームで以下のようなアンケートを実施し、13名のユーザが回答した。

質問1. どのアイコンを押した時に、単語の色が変わったか(複数選択可)

質問2. アイコンの中でこれは要らない(必要ない)と思ったアイコンはどれか(複数選択可)

質問3. アイコンの中でこれは要るな(便利だ)と思ったアイコンはどれか(複数選択可)

結果を表1にまとめる。

表1 欲望語アイコンに関するアンケート結果

	質問1	質問2	質問3
欲しい	6	0	9
行きたい	3	1	4
買いたい	2	2	4
可愛い	4	0	6
食べたい	4	0	8
見たい	0	5	2
作りたい	0	9	0
大好き	1	0	6

・「欲しい」のアイコンを押した時に単語の色が変わったと答えた人が一番多く、6人いた。同様に、「欲しい」のアイコンは、記念日をお祝いするために必要だと答えた人も一番多く、9人いた。

・次点で「可愛い」、「食べたい」、も多く、同時に必要であるという回答も多かった。

・「作りたい」、「見たい」の項目については、不要であるという回答が多かった。

・「大好き」の項目については、記念日をお祝いするために参考になる項目であるという意見が6人いるのに対し、単語の色が変わった被験者は1人と、少なかった。

・「買いたい」と思ったものは、「欲しい」とツイートするので、まとめれば良いのではないかという意見がいくつかあった。

以上の結果から記念日をお祝いするのに必要だと思われる欲望語について再考する。従来通りの欲望語として、「欲しい」、「食べたい」、「大好き」「可愛い」、「行きたい」を残し、「作りたい」、「見たい」は削除する。「買いたい」については、アイコンを表示させずに、「欲しい」と同じカウントにする。

また、ツイートする際の表記について、欲望語がひらがなで書かれている事があるため、それぞれのひらがな表記もカウントする。例えば「欲しい」は「ください」や「くれ」、「大好き」は単に「好き」という表現を用いることがあるため、それもカウントする。「可愛い」も「きかわたん」という流行的に使われている口語表現があり、これもカウントすべきである。

#### 5. 課題

本システムの本評価結果から、タグクラウドに表示される単語を見るだけではユーザの嗜好情報を得るのが難しいことがわか

った。ユーザのキーワードだけではなく、似ているユーザの嗜好情報もあわせて紹介することで、よりユーザの嗜好を知ることができると考えられる。

Twitter では、友人や気になる人物のアカウントをフォローすることで自分のタイムラインに表示させることができる。自身が興味のある人物や、嗜好の似ている人物をフォローする傾向にあるため、フォローが関心を示す物は自分も関心のある物である可能性が高い。そこで、フォローの嗜好情報を同時に紹介することを提案する。フォローのIDを取得して同様に解析を行ない、出てきた単語を紹介することで、ユーザはより多くの情報から記念日をお祝いすることができると考えられる。

#### 6. 関連研究

本研究の位置づけを明確化するために、Twitter を使った関連研究を紹介する。

[渡邊 12]では、Twitter から語の関連性を分析し、その情報を基にユーザの嗜好を表す興味語を抽出する手法を提案している。ユーザ評価実験の結果、語の関連性を考慮することでTF-IDF法に比べ興味語抽出性能が向上しており、未投稿であるユーザの興味語を抽出することも可能となっている。

[齋藤 11]では、ソーシャルブックマークサービスで使用されているタグをもとに構築した類語辞書を用いてユーザの興味・嗜好・属性をキーワードベースで抽出し、類似するユーザの推薦を行う研究である。Twitter ではソーシャルネットワークサービスの名の通り、ユーザの興味に近い人を見つけたり、人とのつながりを支援する研究が多い。

[矢野 12]では、個人のライフログを生成するために、Twitter に投稿された行動を表す単語(一般的な動詞とサ変名詞、「なう」)を使っている。本研究では行動ではなく、嗜好情報を抽出するために、より単純な欲望語というキーワードと共にツイートしている情報を抽出するようにしている。

#### 7. まとめ

記念日をお祝いしたくても、何をしたらよいかかわからないといった人が存在し、本研究にて開発した、記念日のお祝いを支援するシステムが有用であることは、アンケートの結果より明らかである。しかし、機能の実用性については、まだ検討すべき点がいくつかある。Web 上で行った評価により、一般的な Twitter ユーザから多くの改良案が寄せられた。この意見をもとに改良を重ねることでより精度の高い解析結果が得られ、ユーザの満足度向上につながると考えられる。

#### 参考文献

- [齋藤 11] 齋藤 準樹, 湯川 高志: ソーシャルブックマークを基にした Twitter ユーザの興味語抽出・推薦手法の提案と評価, 情報処理学会研究報告[デジタル・ドキュメント], Vol. 2011, No. 2, pp. 1-8 (2011)
- [渡邊 12] 渡邊 恵太, 加藤 昇平: Twitter における語の関連性に着目したユーザ興味語抽出手法の提案, 第 26 回人工知能学会全国大会, pp.1B1-R-3-7, 2012.
- [矢野 12] 矢野 裕司, 横井 健, 橋山 智訓: 行動辞書を利用した Twitter からの行動抽出, 第 11 階情報科学技術フォーラム (FIT2012), RO-003, 第 4 分冊, pp.51-56, 2012.