5N-3

語彙連鎖構造を用いたTwitterユーザ推薦手法の提案

大村　涼†　　佐藤　真†　　赤石　美奈†

†法政大学情報科学部

1. はじめに

Twitterをよりよく利用するための1つの方法としてユーザは自分と趣味の合う他のユーザの発見が挙げられる．それを支援するためにTwitterにはおすすめユーザを表示する機能があるが，必ずしも自分と趣味の合うユーザを見つけるのには適しているとは言えない．また，興味のあるキーワードを呟いているユーザを検索することもできるが膨大な結果の中から自分と趣味の合うユーザを見つけるのは困難である．

そのためTwitterに関する研究は多数存在している．Golder[1]らは，Twitterにおけるユーザ同士の結びつき方について，どのような条件でフォローが為される傾向にあるかを述べている．北川[2]らは，コミュニケーションに着目したTwitterフォローユーザ推薦手法を提案し，その有効性を検証している．しかし，どちらもツイートの内容を推薦に用いておらず，被推薦ユーザにとって関係性の薄いユーザが推薦されることがある．

そこで本研究では，ツイートやプロフィールの内容の語彙連鎖構造を用いて，より趣味の合うユーザを推薦する手法を提案する．

2. 提案手法

本提案手法では被推薦ユーザとフォロー候補となるユーザのプロフィール及びツイートの語彙連鎖構造に着目し，ユーザの興味のある語を抽出し，それらの語の共起関係から自分とより趣味の合う他のユーザを推薦する．主に興味語の抽出及びユーザ推薦の二つの工程から成る．

2.1. 興味語の抽出

ユーザの興味を示す興味語の抽出を以下のような手順で行う．

（1）被推薦ユーザのプロフィール及び最新200件のツイートを取得する．

（2）被推薦ユーザのプロフィール及び最新200件のツイート中に出現した語から以下の語を除外した語を興味語とする．

・時間を表す語（「今日」「最近」など）
・代名詞（「自分」「私」「その」など）
・記号（「♪」「？」「ω」「▽」など）
・Twitter特有の語（「RT」「QT」「http」など）
・重要度の低い語．なお重要度は以下のように定義する吸引力で表す．

まず語 t の t'に対する出現依存度 $td(t,t')$を以下の式で計算する．

$$td(t,t') = \frac{sentences(t \cap t')}{sentences(t)} \quad (1)$$

ここで，$sentences(t)$ は，文書中における語 t を含む文の数であり，$sentences(t \cap t')$ は，t と t'を同時に含む文の数である．

次に語 t が文書中の他の語を引き付ける力を吸引力と呼び，他の語から語 t に対する出現依存度の総和として，以下のように定義する．T は文書に含まれるすべての語の集合を表す．

$$attr(t) = \sum_{t'' \in T} td(t'', t) \quad (2)$$

2.2. ユーザ推薦

興味の合うユーザを絞り込み，推薦を行う．手順は以下の通りである．

(1)被推薦ユーザの興味語の中で最も頻度の高い語について呟いた最新 50 人のユーザを取得し推薦候補とする．

(2)推薦候補の50人のユーザのプロフィール及び最新200ツイートを取得する．

(3)被推薦ユーザの興味後と各推薦ユーザの興味語の共通する語の数が多いユーザ上位 20 名を推薦ユーザとする．ただし共通する語の数が同じだった場合，被推薦ユーザの興味語の中で最も頻度の高い語について吸引力の大きいユーザから推薦する．

3. 評価実験と結果

本研究の評価方法としてアンケートを実施する．被験者としてTwitterを普段から使用しているユーザ8名にアンケートを採った．

被験者には本提案手法を用いたおすすめユーザ20名と公式のおすすめユーザ20名に関するアンケートを行った．以下にアンケート内容とその結果を記載する．

Twitter User Recommendation using Lexical Chain Structures
†Ryo Oomura, Hosei University
‡Makoto Sato, Mina Akaishi, Hosei University

Copyright ©2013 Information Processing Society of Japan.
All Rights Reserved.

質問項目 1:そのユーザを知っていましたか？
①知っていた
②芸能人や著名人等の有名人だった
③会社などの公式アカウントまたは bot だった
④知らなかった

表 1. 既知の推薦候補数 (/20 人)

被験者		A	B	C	D	E	F	G	H
①	提	0	0	0	0	0	0	0	0
	公	0	4	5	4	0	7	6	11
②	提	0	0	0	0	0	0	1	1
	公	2	6	5	10	2	1	1	0
③	提	0	0	1	0	0	0	0	0
	公	5	1	1	5	0	5	4	4
④	提	20	20	19	20	20	20	19	19
	公	13	9	9	1	18	7	9	5

質問項目 2:そのユーザに興味を持ちましたか？もしくはフォローしようと思いましたか？

表 2. 興味をもった推薦候補数(/20 人)

被験者	A	B	C	D	E	F	G	H
提案	14	11	5	9	12	3	4	2
公式	5	6	8	4	11	8	4	3

また，興味を持ったユーザの中で未知だったユーザ数は表 3 のようになった．

表 3. 興味を持った未知の推薦候補者の割合(%)

被	A	B	C	D	E	F	G	H
提	100	100	100	100	100	100	100	50
公	0	17	13	0	82	0	25	0

質問項目 2 の結果から本提案手法は公式のおすすめユーザよりもやや低い推薦精度となったが，質問項目 1 からほぼ 100%の未知かつ bot でもないユーザから推薦できた．そのため本提案手法は未知のユーザかつ同一の興味を持つユーザの発見に有効である．

また，本提案手法における語彙連鎖構造の視覚化がユーザの興味を推測するのに有用か調べるために興味語について 2.2 で定義した共起依存度と吸引力を可視化した WordColony[3]（図 2）を提示し，以下のようなアンケートを採った．

質問項目 3:WordColony の画像からそのユーザの興味が推測できましたか？

表 4. 20 名中,興味の推測ができた人数

A	B	C	D	E	F	G	H	計
19	14	10	13	16	9	19	10	110

これらの結果から半数以上のユーザに対し興味の推測ができた．そのため，プロフィール及びツイートの語彙連鎖構造の視覚化は興味の推測に有効である．

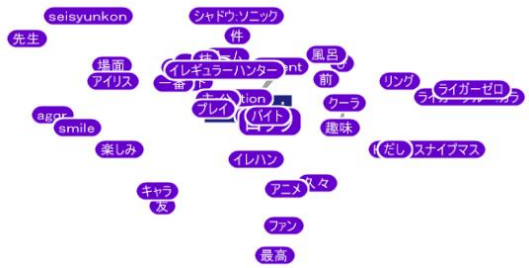

図 1. 語彙連鎖グラフ WordColony

4. おわりに

Twitter において未知かつ興味の合うユーザの発見を支援するために，語彙連鎖構造視覚化ツール WordColony を用いて興味語の抽出し，興味語の共起関係からユーザを推薦する手法を提案した．また評価実験の結果，Twitter 公式のおすすめユーザよりも未知のユーザから被推薦ユーザと興味の合うユーザの発見について良好な結果を得ることが出来た．また，語彙連鎖構造の視覚化がユーザの興味の推測に有効であることも確認した．

今後の課題として，興味語となり得ないノードの除去することで推薦精度を高める必要がある．また，本稿では提案手法を全て自動化するまでは至っていないため，本提案手法により抽出されたユーザの興味語をもとにした全て自動で行うユーザ推薦システムの構築を目指したい．

参考文献

[1] Scott A. Golder, S. Yardi：Structural Predictors of Tie Formation in Twitter: Transitivity and Mutuality, 2010 IEEE Second International Conference on Social Computing, socialcom, pp.88-95, 2010.

[2] 北村 太一,小川 祐樹, 諏訪 博彦：コミュニケーションに着目した Twitter フォローユーザ推薦, 2012 年度人工知能学会全国大会 ,2012

[3] 赤石美奈,佐藤健,堀浩一：語の吸引力に基づく主題遷移解析と視覚化,電子情報通信学会技術研究報告,言語理解とコミュニケーション,2006

Copyright ©2013 Information Processing Society of Japan.
All Rights Reserved.