

# ソーシャルネットワークユーザ間情報伝播量測定

## Measuring the Quantity of Information Propagation of Users in Social Network

宮城 涼<sup>1\*</sup> 高木 友博<sup>1</sup>

Ryo Miyagi<sup>1</sup>, Tomohiro Takagi<sup>1</sup>

<sup>1</sup> 明治大学大学院理工学研究科基礎理工学専攻

<sup>1</sup> Computer Science Course, Graduate School of Science and Technology, Meiji University

**Abstract:** These days information propagation method that transmits information to many people is proposed. Many previous method regard the network structure as important. However, these methods decide the quantity of information propagation without sufficiently considering a conversation topic of users. Therefore, We propose quantification method of information propagation considering communication between users in Social Network. We verified experimental results in twitter data.

### 1. はじめに

近年ソーシャルネットワークが台頭しており,その重要性が高まっている.国家レベルではアラブの春や東日本大震災でソーシャルネットワークの重要性が広く認知されるようになり,個人レベルではソーシャルネットワークを通じて今までは注目されていなかった人や物に注目が集まるようになった.企業レベルではソーシャルネットワークを積極的に活用しようという流れがあり,その活用方法が模索されている.企業のソーシャルネットワークを使用したマーケティング活動の一つにバイラルマーケティングがある.バイラルマーケティングとはソーシャルな影響を利用したマーケティングであり,口コミなどが人々の行動,意見,購買につながるという考え方に基づいている.このバイラルマーケティングに関連する研究は数多く行われており,ネットワーク構造を重視し,全体最適解を求める研究が特に活発に行われている.しかし,これらのネットワーク構造を重視した研究ではユーザの嗜好をあまり考慮していない.そのため,現実的な場面でネットワーク構造を重視した技術でバイラルマーケティングを試みようとしたとき,3つの課題が想定される.1つ目は,ターゲットユーザに向けてどういった情報を伝えたいのか分からない.2つ目は,ターゲットユーザにどのように情報が伝わっているのか分からない.3つ目は,ターゲットユーザが他ユーザにどのような情報を伝播しているのか分からない.これらの課題に対してユーザ間という局所に注目し,どのような種類

の情報がどのくらい伝播しているかを定量化した情報伝播量を測定し対処することが考えられる.情報伝播を扱った研究は存在するが,情報伝播量の測定値を十分検証した研究はなされていない.従って本稿では情報伝播量を測定し,測定値が実際の情報伝播を反映しているか検証する.具体的には, Twitter ユーザの各発言の嗜好を LDA によってトピックで表現し,それを時系列入力データとして移動エントロピー及びトピック移動エントロピーで情報伝播量を測定し検証する.

又,本稿で提案するトピック移動エントロピーを用いることによって,従来の移動エントロピーを用いた手法では測定できなかったトピックごとの情報伝播量を測定できることを示すと同時に,実際の情報伝播量をより反映した測定ができることを示す.

### 2. 関連研究

この章ではバイラルマーケティングに関連する研究について,ネットワーク構造を重視した研究とネットワーク内の局所に注目した研究に分類し説明する.

#### 2.1. ネットワーク構造を重視の研究

バイラルマーケティングの主な目的は,口コミによってより多くのユーザに情報を伝播することにある.よって,より多くのユーザに情報伝播を行うことができるユーザを検知し,伝播させたい情報をそのユーザに流すことは理にかなっている.ソーシャルネットワーク内で多数のユーザからの被リンクを得ているインフルエンサーはこの条件を満たしている

\*連絡先: 明治大学理工学研究科基礎理工学専攻  
〒214-0034 川崎市 多摩区 東三田 1-1-1,  
E-mail: miyagi.wsl@gmail.com

といえる.このような考えに基づき PageRank を拡張することによってインフルエンサーを検知する研究 [1][2]が行われている.その中でネットワーク構造だけでなく,ユーザの嗜好をトピック分布で表現し PageRank を用いてトピック類似度を考慮した研究 [3]がある.

又,ネットワークを重視した研究に影響最大化 (Influence Maximization)がある. バイラルマーケティングに「影響力が強いと予測される少数のノード群をターゲットとし,社会ネットワーク上でのノードからノードへの情報伝搬を引き起こし,そして,それら情報伝搬の結果として,より多くのノード群にその情報を伝える」[4]という戦略が存在する.影響最大化はこの戦略に従い,ターゲットとするノード(ユーザ)の選択を目的としている. 情報拡散モデルによって,他ノードに多くの情報伝播をできると判定したノードをターゲットノードとして選択する. 情報拡散モデルの代表的なモデルとして,IC(independent cascade)モデルや LT(linear threshold)モデルなどがあり,これらを使用した研究[5]が行われている.又これらのモデルを拡張した様々なモデルが提案されており,その中で情報伝播確率を学習,予測するモデル [6][7]が提案されている.

## 2.2. ネットワーク内の局所注目の研究

ソーシャルネットワーク内の局所に注目した研究は数多く存在するがその中から,特定のソーシャルネットワークに関する研究,部分的なネットワークを使用した研究,ネットワークを使用しない研究を取り上げる.

様々なソーシャルネットワークを利用した研究が存在するが,特定のソーシャルネットワークのみに特化した研究がある.特に Twitter に関する研究において,Web Ecology Project[8]では 10 日間で 12 人の有名な Twitter ユーザのフォロワーについて調査している. Cha ら[9]はフォロワー数,リツイート数,メンション数を用いた影響力測定の比較をしている.

部分的なネットワークを使用した研究も行われている. Tang ら[10]はトピックベースの情報伝播の分析を行っている. Topical Affinity Propagation(TAP)という手法を用いて,ネットワークと各ユーザを表したトピック分布でメンバー間のトピックごとの情報伝播量を測定している.この研究はトピックモデルを用いてトピックごとの情報伝播量を測定しているという点で本研究と関連している.

ネットワークを使用しない研究も存在する. greg ら[11]は時系列なユーザの発言を LDA でトピック分布として表現し,それと移動エントロピーによってユーザ間情報伝播量を測定している. 特に,本稿は彼

らの研究にならってシステムを提案している.彼らは条件付き相互情報量推定を用いて,移動エントロピーの推定値を使用している.また,ユーザ間情報伝播量のみ計算を行っている.それに対し,本稿ではトピック移動エントロピーを提案し,トピックごとのユーザ間情報伝播量を測定するという点で異なる.

## 3. 潜在的ディリクレ配分法

Blei ら[12]によって導入された潜在的ディリクレ配分法(LDA:Latent Dirichlet Allocation)はデータに隠れた潜在的意味(トピック)を推定する確率的トピックモデルの 1 つである. Hofmann[13]による確率的潜在意味解析モデル(pLSI:probabilistic Latent Semantic Indexing)におけるトピックと単語の多項分布それぞれにディリクレ事前分布を導入して拡張し pLSI の問題を解決した.一文書に複数トピックが存在することを表現でき,文書の生成過程をトピックに基づいて確率的に表現するモデルである.具体的には, LDA は指定したトピック数だけ各単語を割り振り,文書のトピック分布を出力する.

## 4. 移動エントロピー

この章では移動エントロピーについて説明した後,ソーシャルネットワーク情報伝播量測定でどのように使用するかを説明する.又,移動エントロピーを拡張したトピック移動エントロピーを説明する.

### 4.1. 移動エントロピー

移動エントロピー (Transfer Entropy:TE) は Schreiber[14]によって導入された. 時間  $t$  における確率変数  $X, Y$  の要素をそれぞれ  $x_t, y_t$  とする.  $X$  に対する  $Y$  の影響を表す移動エントロピー  $TE_{Y \rightarrow X}$  は式(1)のように定義される.

$$TE_{Y \rightarrow X} = \sum p(x_{t+1}, x_t^{(k)}, y_t^{(l)}) \log \frac{p(x_{t+1} | x_t^{(k)}, y_t^{(l)})}{p(x_{t+1} | x_t^{(k)})} \quad (1)$$

$k, l$  は時間  $t$  における状態から何個分遡った状態までを含めるかを表している.

### 4.2. 移動エントロピーを使用した情報伝播量測定

確率変数  $X$  のエントロピー  $H(X)$  を式(2)のように定義する.エントロピー  $H(X)$  は  $X$  のあいまいさの量

を表す.

$$H(X) = - \int dx p(x) \log p(x) \quad (2)$$

このとき,相互情報量(Mutual Information)と条件付きエントロピーはそれぞれ式(3),式(4)で定義される.

$$I(X; Y) = H(X) + H(Y) - H(X, Y) \quad (3)$$

$$H(X|Y) = H(X, Y) - H(Y) \quad (4)$$

X の次の時刻の状態をX' とするとき,移動エントロピーは条件付き相互情報量 (Conditional Mutual Information)で式(5)のように表される.

$$TE_{Y \rightarrow X} = I(X': Y|X) = H(X'|X) - H(X'|Y, X) \quad (5)$$

2 行目の第 1 項は X を考慮したときのX'のあいまいさであり, 第 2 項は X と Y を考慮したときのX'のあいまいさである. よって,式(5)の 2 行目は Y を考慮したことによるX'のあいまいさの減少度を表しており,Y を考慮したことによるX'の予測性の向上度ともいえる.これを Y から X への情報伝播量とみなす.

又,実際の計算は $k=1, x_t^{(1)} = x_t, y_t^{(1)} = y_t$  として式(6)で計算する.具体的な計算方法については後述する.

$$TE_{Y \rightarrow X} = \sum_{x_{t+1} \in X'} \sum_{x_t \in X} \sum_{y_t \in Y} p(x_{t+1}, x_t, y_t) \log \frac{p(x_{t+1}, x_t, y_t) p(x_t)}{p(x_{t+1}, x_t) p(x_{t+1}, y_t)} \quad (6)$$

#### 4.3. トピック移動エントロピー

本稿では,移動エントロピーを拡張したトピック移動エントロピー(Topic Transfer Entropy:TTE)を提案する.これによって特定トピックの情報伝播量を測定することができる.具体的には,Y からX' に伝わった情報が同一であったときのみ計算を行う. トピック移動エントロピーを式(7)で定義する.具体的な計算方法については後述する.

$$TTE_{Y \rightarrow X}(\text{topic}) = \sum_{x_t \in X} \{ p(x_t, Y = \text{topic}, X' = \text{topic}) \log \frac{p(x_t, Y = \text{topic}, X' = \text{topic}) p(x_t)}{p(X' = \text{topic}, x_t) p(X' = \text{topic}, Y = \text{topic})} \} \quad (7)$$

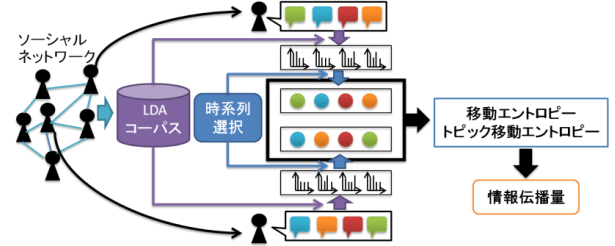


図 1 情報伝播量測定システム概要

## 5. 情報伝播量測定システム

この章では,LDA 及び移動エントロピー,トピック移動エントロピーを使用した情報伝播量測定システムについて述べる.ソーシャルネットワーク内のユーザの発言を LDA によってトピックとして表現する.その後,ユーザ間の発言を時系列に列挙し,その中からデータを選択する.選択したデータを入力として移動エントロピー,トピック移動エントロピーで情報伝播量を測定する.図1に情報伝播量測定システム概要を示しシステムの処理過程を説明する.

### ① LDA コーパス : 各発言のトピック分布

ユーザの発言をトピック分布で表現するまでの処理過程を説明する.形態素解析器 Kuromoji[15]を用いて発言文書から名詞を抽出し,発言を Bag-of-words で表現する.この Bag-of-words を LDA で解析することで各発言をトピック分布で表現する.

### ② 各ユーザ間の時系列トピック選択

各ユーザ間の時系列トピック選択までの処理過程を説明する. まず 2 人のユーザを選択し,ユーザ間の発言を時系列に列挙する.ユーザの発言は①のトピック分布で一番高い所属確率のトピック番号で表される.この時,一番所属確率が高いトピックが複数ある場合,移動エントロピーでは一番数が小さいトピック番号を選択し,トピック移動エントロピーでは全トピックを選択する.ところで,移動エントロピー,トピック移動エントロピーはどの発言がどの発言に対して影響を与えたかを定義しなければ計算できない.ここで,情報の受け手,情報の送り手について,任意の発言トピックを $x_i, y_j$ ,発言数を  $m, n$  としたとき,時間の古い方から並べた時系列トピックは $X = \{x_0, x_1, \dots, x_m\}, Y = \{y_0, y_1, \dots, y_n\}$ と定義される.この 2 つの時系列に並んだ発言トピックを同一時系列上に時間の古い方から列挙し,以下のアルゴリズムに従いトピックを選択する.

1.  $x_0$  を取得し,それ以降を探索
  2.  $y_j$  の次に  $x_i$  がくる組み合わせをすべて取得
- このアルゴリズムの具体例を図2に示す.

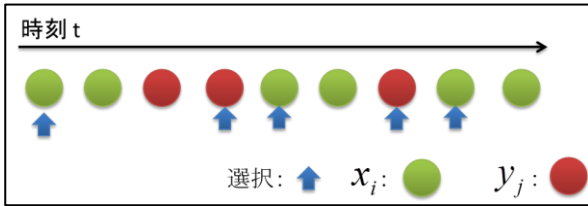


図 2 時系列トピック選択

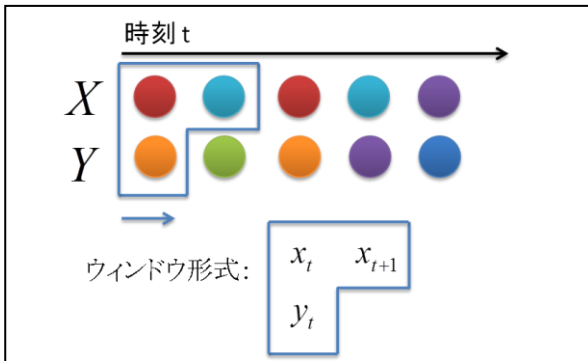


図 3 情報伝播量測定

$$p(x_{t+1}, x_t, y_t) = p(\text{blue circle, red circle, orange circle}) = \frac{2}{4}$$

$$p(x_t) = p(\text{red circle}) = \frac{2}{5}$$

$p(x_{t+1}, x_t, y_t)$ は4ウィンドウ中の2つが一致しているため、 $\frac{2}{4}$ .  
 $p(x_t)$ は5つ中2つが赤であるため、 $\frac{2}{5}$ .

図 4 確率計算方法

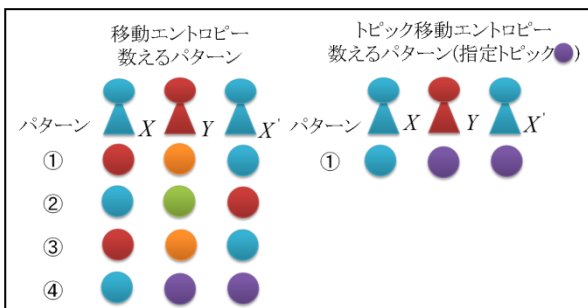


図 5 移動エントロピーとトピック移動エントロピーの差異

### ③ 情報伝播量測定

移動エントロピーを用いてユーザ間の情報伝播量を測定する処理を説明する.図 3 のように,ウィンドウをスライドさせて要素のパターン一致数から式(6)の各確率を計算する.例えば,図 3 のような X,Y の時系列トピックが並んでいる時, $p(x_{t+1}, x_t, y_t)$ ,  $p(x_t)$  は図 4 の値になる.

トピック移動エントロピーも図 3 のように,ウィン

ドウをスライドさせて計算する.ただしトピック番号 topic を指定したとき, $x_{t+1} = y_t = \text{topic}$ のみ計算する.

移動エントロピーは Y のどのような発言も X に少なからずは影響を与えており, X' にその影響が反映されているという考えに基づいた情報伝播量を測定しているのに対し,トピック移動エントロピーは Y が X に影響を与えた時, X' は Y と同調的な発言をするという考えに基づいた情報伝播量を測定している.この差異を具体的に示すために,図 3 でウィンドウを左から右へずらした時の移動エントロピーが数えるパターンとトピック移動エントロピーが数えるパターンを図 5 に示す.ここで, X, Y, X' は X が Y に影響されて X' になったことを表している.移動エントロピーは Y の発言のトピックとそれに影響を受けた X' の発言のトピックの種類に関係なく伝播しているとみなし X, Y, X' の全発言パターンを数えることで情報伝播量を計算するのに対し,トピック移動エントロピーは Y と X' の発言トピックが同じとき,つまり X' が Y に影響を受けて同じような発言をするときのみ情報が伝播したとみなし情報伝播量を計算する.

## 6. 実験

### 6.1. 検証データ

情報伝播システムの検証として Twitter データを使用した.検証には「東京ディズニーリゾート PR 【公式】」のツイート情報に影響を受けているユーザが多いと見込まれる為, Twitter API を使用し「東京ディズニーリゾート PR 【公式】」をフォローしているユーザのうち,2,972 ユーザとユーザが発言した 421,336 ツイートを取得した.取得したツイートが作成された期間は 2009 年 5 月 8 日から 2014 年 3 月 3 日である.

特定のユーザ名から始まるツイートをリプライといい,そのユーザのタイムラインにのみ表示される.本稿では,情報の送り手のリプライは全て考慮せず,情報の受け手のリプライは情報の送り手に対するもののみ用いた.ツイート内の URL とハッシュタグは削除せず,形態素解析を行う.ここで,文書頻度の高い名詞はストップワードとして除去する.除去するストップワードは表 1 に示した通りである.

### 6.2. LDA の結果

検証データを入力として用いたときの LDA の結果を提示する.LDA のハイパーパラメータ  $\alpha, \beta$  の値は 1.0,0.3 とし,トピック数は 50,75,100,125,150 を定性的に評価し 75 とした.推論にはギブスサンプリング

を用い,繰り返し回数は1500回とした.トピック数75のLDAの結果は表2に示す.表2では各トピックに所属する所属確率が高い5つの単語を列挙している.

表1 ストップワード

http	一	二	三	ゼロ-zero	皆	四	九	八	五	月
------	---	---	---	---------	---	---	---	---	---	---

表2 「東京ディズニーリゾートPR【公式】」のLDA結果

解釈	topic番号	単語 トピック1~5				
カップル	0	男	馬鹿	女	嘘	LINE
ラブライブ	1	ライブ	キャス	七	chiaki	チアキ
?	2	方	訳	ファン	風	英語
?	3	放送	本日	女性	生	開始
天気	4	雪	雨	朝	天気	今日
自然災害 警報	5	県	市	駅	トウキョウ	運転
バイト	6	バイト	今日	パン	マン	伯父
?	7	絶対	駄目	love	筆	renai
?	8	七	六	恋	日	桜
原発	9	零	日本	ニュース	問題	新聞
ライブ	10	曲	映画	発売	決定	イベント
テレビ	11	姉	テレビ	山	疲れ	フジ
デザート	12	ケーキ	音	菓子	味	アイス
顔文字	13	スバ	月	ノ	日	今度
スバ	14	一緒	御腹	風呂	一杯	六
?	15	所	今	名前	家族	人
お詫び・心配	16	御免	(´・`・´)	先輩	元氣	草
デート・遊び	17	今日	久々	練習	カラオケ	帰宅
六本木	18	ブログ	更新	応援	ロッポンギ	紹介
気持ち	19	気	最近	全部	所為	体調
軽い返事	20	笑い	本真	蓋	917	ため
金	21	円	万	金	以上	そう
プレゼント	22	フォロー	当選	プレゼント	ツイート	方法
動物	23	七	母	夜	猫	六
家族	24	子	目	いけメン	娘	ママ
ジョジョ	25	前	普通	普通	適当	能力
歌詞	26	中	頭	電車	番	涙
冠婚葬祭	27	画像	卒業	結婚	式	花
セル	28	出演	為	監督	バーゲン	安売り
恋愛	29	事	為	勝手	喧嘩	魔
ハリーポッター	30	六	mymelody	sayamero	hide	七
携帯	31	顔	メール	携帯	絵	先
家	32	内	奴	家	婆	旦那
ソーシャルゲーム	33	ゲーム	参加	七	スマホ	リアル
曜日	34	日	人	誕生	診断	友達
?	35	六	live	chigatv	stream	momoclo
友人	36	友達	感じ	女子	男子	最近
DVD	37	七	話	六	箱	(´・`・´)
テスト勉強	38	物	勉強	一人	テスト	(´・`・´)
?	39	七	六	sakura	(´・`・´)	showma
Twitter拡散呼びかけ	40	人	フォロー	全員	夢	おた
嵐	41	嵐	ドラマ	ショウ	ジュン	役
季節旅	42	度	最高	何	春	夏
ジェジュン	43	今	個の学習書	ジェジュン	時代	ファン
運動・体	44	上	体	ライン	ロ	下
1年回想	45	年	今年	;	数	来年
Twitter	46	bot	Twitter	定期	ツイート	サイト
?	47	笑い	aya	mentaiko	☺	爆笑
感情	48	暇	意味	間	びっくり	本気
?	49	祭り	チーム	ゆき	ユウ	メンバー
占い	50	今日	仕事	運	健康	恋愛
テラスハウス	51	格好	ハウス	テラス	舞台	ケン
子育て	52	子供	爆笑	部	部屋	親
?	53	手	https	旅行	aruaru	動画
Twitter:ジャニーズ	54	方	声	物	最後	希望
Facebook	55	写真	無理	泣き	撮影	理由
出勤	56	時	分	頃	出勤	半
女性恋愛	57	本当	幸せ	気持ち	感謝	互い
ディズニー	58	ディズニー	トウキョウ	クリスマス	disney	ランド
心理	59	事	自分	人	言葉	心
休日	60	明日	休み	朝	学校	機
学校	61	金	先生	学校	高校	大学
?	62	六	七	班	今	場所
正月	63	予定	御節	オオサカ	年	料理
マラソン	64	回	七	結果	何	最終
?	65	笑い	706	aaii	4337	miuhiro
予定	66	楽しみ	疲れ	次	明日	(´・`・´)
くつろぎ時間	67	時間	後	週間	人生	睡眠
出前寿司	68	イン	弁当	予約	寿司	個人
バレンタイン	69	チョコ	♥	バレンタイン	無し	七
疲労	70	今日	昨日	疲れ	・	連
ご飯	71	毎日	御飯	箇月	ラーメン	カレー
風邪	72	感	脳	風邪	会社	病院
プロスポーツ	73	日本	選手	世界	試合	マオ
アニメ	74	パス	歌	鳥	キス	マイ

表3 移動エンタロピーの情報伝播量上位10 ユーザ

Xの名前	総ツイート数	TE
千葉県 の担当	19	0.30103
しょーくん	173	0.30103
ちょこ	25	0.30103
中川美里	23	0.30103
(怜美@櫻宮.° —° )	8	0.30103
41.0℃プロジェクト	11	0.30103
yuka(*´`*)♡	50	0.30103
ごみ人間ごみ	12	0.30103
ちわ	29943	0.291807
はじまりの国さいごの話	159	0.286273

表4 トピック移動エンタロピーの情報伝播量上位10 ユーザ

Xの名前	総ツイート数	TTE(58)
doughnut@甘党	58	0.15904
FK531(for D)	15	0.15904
ロマンスタ	14	0.15904
はな88	38	0.150515
じゅんじゅん@D垢	40	0.150515
ドナもち	86	0.150515
ゆうGoofFTW@D垢	78	0.150515
kendisney	19	0.150515
ペヌ吉	95	0.150515
安藤裕介	658	0.11928

表5 トピック移動エンタロピーの情報伝播量下位3 ユーザ

Xの名前	総ツイート数	TTE(58)
ぼおと	76	-0.029035
HIKACHU	71	-0.029349
Minami	22	-0.058697

このアカウントは、千葉県のアカウントさんを、リスト @Chiba\_eyes/lists に、カテゴリー分けしながら、フォローして居マッスル！ by ほつとあいず @hot\_eyes ♪(^o^)♪ ツ(^o^)シ

図6 「千葉県 の担当」のツイート



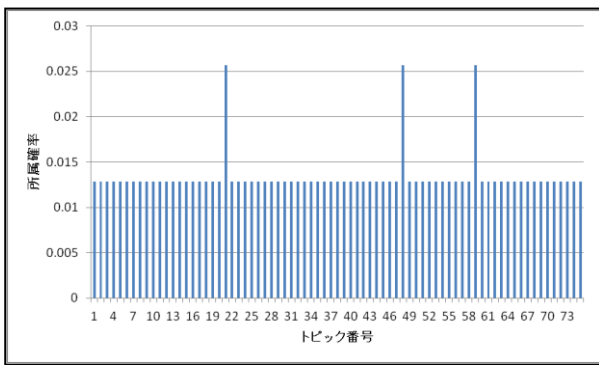


図 7 複数のトピックが一番高い所属確率である発言トピック分布

### 6.3. LDA の考察

表 2 中の「解釈」という項目はトピックの中身を観察し、主観で付けたタグである。「解釈」が「?」になっているトピックは、トピック中の単語から意味を解釈できない、または複数の意味が含まれており一意に解釈できないことを示している。表 2 のようにほとんどのトピックの意味は解釈することができる。特にトピック 58 は東京ディズニーリゾートと判断できる。

### 6.4. 移動エントロピー、トピック移動エ

#### ントロピーの結果

表 3 は移動エントロピーを用いた情報伝播量上位 10 ユーザの結果であり、表 4 はトピック番号 58 のトピック移動エントロピーを用いた情報伝播量上位 10 ユーザの結果である。トピック 58 はディズニーに関するトピックであり、ユーザ「東京ディズニーリゾート PR【公式】」は 402 ツイート中 299 ツイートがトピック 58 のツイートだった為、トピック 58 のトピック移動エントロピーの情報伝播量のみ結果を示す。

### 6.5. 移動エントロピー、トピック移動エン

#### トロピーの考察

移動エントロピーとトピック移動エントロピーを用いた情報伝播量上位 10 ユーザの結果に関する考察を行い、測定した情報伝播量が実際の情報伝播を反映できているかという精度について述べる。評価方法は高い情報伝播量のユーザ間で情報伝播が行われているか主観で評価した。

#### ① 移動エントロピー

移動エントロピーを用いた情報伝播量上位 10 ユーザ中 9 ユーザがディズニーに関するツイートを行っていないかった。また、「東京ディズニーリゾート PR【公式】」から影響を受けて発言したというような事は観察できなかった。唯一ディズニーに関してツイートをしていた「yuka(\*~\*)♡」も「東京ディズニーリゾート PR【公式】」から影響を受けて発言したというような事は観察できなかった。このような結果になった原因として、スパムユーザとトピック選択の問題がある。以下で両問題を具体的に説明し、対処方法を述べる。

上位 10 ユーザの特徴は同じようなツイートを繰り返しており、総ツイート数が少ない傾向があるスパムのようなユーザだった。具体的には「千葉県 の担当」は図 6 のようなツイートを繰り返しており、すべてトピック 5 に分類される発言である。このようなユーザはスパムユーザと判定して除去するという対策が挙げられる。

移動エントロピーを用いるとき、事前処理の時系列トピック選択で一番高いトピックが複数ある場合、一番数が小さいトピック番号のみを選択している。実際に発言トピック分布を観察すると図 7 のような複数のトピックが一番高い所属確率である発言がほとんどであった。図 7 の発言はトピック 20, 48, 58 が同じ所属確率であるにもかかわらず、現在の手法はトピック 20 と判定してしまう。このようなことから、適切な情報伝播量を測定できていないと想定できる。この問題に対して、greg ら[11]のように発言トピック分布全体を入力として条件付き相互情報量推定を用いて情報伝播量測定を行うことで対処することが挙げられる。

#### ② トピック移動エントロピー

トピック移動エントロピーを用いた情報伝播量上位 10 ユーザ中 9 ユーザがディズニーに関するツイートをしていた。特に「FK531(for D)」、「じゅんじゅん@D 垢」、「ドナもち」、「ゆう GoofFTW@D 垢」、「kendisney」は積極的にディズニーに関するツイートをしていた。このことから、トピック移動エントロピーはトピックを考慮することにより、移動エントロピーよりも正確な情報伝播量が測定できているといえる。ただし、トピック移動エントロピーも 2 つの問題が存在する。

1 つ目はディズニーに関するツイートをしていない「ペヌ吉」の情報伝播量が高くなったという問題である。原因は「ペヌ吉」の発言の多くがトピック 58 を含んでいた為である。それらの発言は図 7 のように一番高い所属確率が複数存在し、トピック 58 を含む。このようなあいまいな発言が多く存在し、ト

ピック 58 の発言として複数回加算する為、情報伝播量が高くなってしまった。このようなあいまいな発言を判定し除去することで上記のような問題に対処することができると考えられる。

2 つ目は表 5 のように積極的にディズニーに関するツイートをしているユーザの情報伝播量が負の値になり、「東京ディズニーリゾート PR【公式】」へ影響を与えていると測定されてしまった問題である。しかし、「東京ディズニーリゾート PR【公式】」が Twitter 上の 1 つの発言に影響を受けるとは考えにくい。この問題に対しては不明瞭な点が多く今後の課題として挙げられる。

## 7. むすび

本稿では、確率的トピックモデルである LDA を用いてユーザの発言の嗜好をトピックで表現し、ユーザ間の情報伝播量を測定した。移動エントロピーを用いた場合、スパムやトピック選択の問題が存在した為、測定した情報伝播量は実際の情報伝播を十分に反映することができなかった。本稿で提案したトピック移動エントロピーを用いた場合特定のトピックのみ計算することで、測定した情報伝播量は実際の情報伝播をより反映し、ユーザ間でどのような種類の情報がどのくらい伝播しているかを定量化できることを示した。

考察では様々な問題点を挙げたが、それ以外の課題としては、計算時間や大規模ネットワークへの対処である。実際にバイラルマーケティングを行う際、この 2 点は非常に重要であるが、本稿ではこれらについて考慮していない。

今後の展望としては、可視化、影響最大化、推薦エンジン、リンク予測、スパム判定、ユーザの影響力への応用が挙げられる。

## 参考文献

- [1] Kwak, H., Lee, C., Park, H., and Moon, S.: What is Twitter, a social network or a news media?, Proceedings of the 19th international conference on World wide web, pp. 591-600, (2010)
- [2] Ding, Z. Y., Jia, Y., Zhou, B., Han, Y., He, L., and Zhang, J. F.: Measuring the spreadability of users in microblogs, Journal of Zhejiang University SCIENCE C, pp. 701-710, Vol. 14, No. 9, (2013)
- [3] Weng, J., Lim, E. P., Jiang, J., and He, Q.: Twitterrank: finding topic-sensitive influential twitterers, Proceedings of the third ACM international conference on Web search and data mining, pp. 261-270, (2010)
- [4] 木村昌弘, 斉藤和巳, 中野良平: 社会ネットワーク上の情報伝搬における強影響力ノード抽出の効率化, 電子情報通信学会論文誌, pp. 1004-1015, (2008)
- [5] Kempe, D., Kleinberg, J., and Tardos, E.: Maximizing the Spread of Influence through a Social Network, Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 137-146, (2003)
- [6] Goyal, A., Bonchi, F., and Lakshmanan, L.V.S.: Learning Influence Probabilities in Social Networks, Proceedings of the third ACM international conference on Web search and data mining, pp. 241-250, (2010)
- [7] Saito, K., Nakano, R., and Kimura, M.: Prediction of information diffusion probabilities for independent cascade model, Knowledge-Based Intelligent Information and Engineering Systems, pp. 67-75, (2008)
- [8] Leavitt, A., Burchard, E., Fisher, D., and Gilbert, S.: The influentials: New approaches for analyzing influence on twitter, Web Ecology Project, Vol. 4, No. 2, pp. 1-18, (2009)
- [9] Cha, M., Haddadi, H., Benevenuto, F., and Gummadi, P. K.: Measuring User Influence in Twitter: The Million Follower Fallacy, 4th Int'l AAAI Conference on Weblogs and Social Media, pp. 10-17, (2010)
- [10] Tang, J., Sun, J., Wang, C., & Yang, Z.: Social influence analysis in large-scale networks, Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 807-816, (2009)
- [11] Ver Steeg, G., & Galstyan, A.: Information-theoretic measures of influence based on content dynamics, Proceedings of the sixth ACM international conference on Web search and data mining, pp. 3-12, (2013)
- [12] Blei, D. M., Ng, A. Y., & Jordan, M. I.: Latent dirichlet allocation. the Journal of machine Learning research, Vol. 3, pp. 993-1022, (2003)
- [13] Hofmann, T.: Probabilistic latent semantic indexing, Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, pp. 50-57, (1999)
- [14] Schreiber, T.: Measuring information transfer, Physical review letters, Vol. 85, No. 2, pp. 461, (2000)
- [15] Atilika. kuromoji - japanese morphological analyzer. (オンライン) (引用日: 2014 年 3 月 6 日) <http://www.atilika.org/>