

# 一般ユーザ目線での Twitter からの人物関係の可視化

西村 章宏<sup>†</sup>      土方 嘉徳<sup>†</sup>      三輪 祥太郎<sup>‡</sup>      西田 正吾<sup>†</sup>

<sup>†</sup>大阪大学大学院 基礎工学研究科    <sup>‡</sup>三菱電機株式会社 先端技術総合研究所

**概要** マイクロブログサービスの 1 つである Twitter では、その時々で話題になる政治家や芸能人など有名人に関する一般ユーザの発言を豊富に得ることができる。さらに近年では、SNS から得られる評判情報をマーケティングやその他サービスに応用しようという試みが活発に行われている。そこで本研究では、Twitter から得られる評判情報のうち、一般ユーザの有名人に関する発言とその発言を行ったユーザのプロフィール等に注目する。これらの各情報源から得られるデータに対し、抽出の妨げとなるノイズへの前処理を経て、一般ユーザの観点が反映された特徴量であるトピックの抽出を行う。そして得られたトピックの分布を元に人物の類似関係を獲得し、それを基に各人物を平面上に配置することで、人物関係の可視化を行う。この可視化結果に対して、使用した情報源毎、および可視化手法毎に、妥当性と発見性に着目した特徴の分析を行う。

**キーワード** Twitter, 評判情報抽出, トピックモデル, 可視化

## 1 はじめに

Web からあらゆる情報を容易に得ることができるようになった現代、ユーザは通学や通勤途中、家事の最中、友人との会話の最中など、日常生活の様々なシーンにおいて情報をパソコンやスマートフォンを使って獲得している。この様なシーンで獲得したい情報は、仕事や勉強の問題を解決するためのものだけでなく、ふと気になった些細な内容であることも多い。この些細な気になる事柄の 1 つとして、有名人に対する興味や関心を取り上げる。これらは、例えばテレビや Web ページを見ている時や、友人と会話をしている時に、ふと目にした・耳にした有名人に対するものである。具体的な例を挙げると、

- 最近ニュース番組でよく耳にする名前であるが、その人がどういった人物なのかあまり知らない。
- 友人間でよく話題になる芸能人だけど、その人についてよく知らず、会話の内容についていけなかった。
- 自分が支持している政治家に対して、世間一般の印象がどの様なものか気になった。

といった場面が想像できる。この様な際にユーザの多くは、wikipedia やタレント名鑑等のオンライン百科事典を利用し、年齢や所属といった客観的事実からその人物を把握していると思われる。しかし、その有名人が世間一般ではどの様に捉えられているかを表す印象や感想といった客観的ではない情報 (以下、一般ユーザの観点到に基づく人物情報) は、上記の様な情報源からは得ることができない。上記の様にふと気になった有名人については、客観的な事実に関する情報だけでなく、世間一般の

印象や評判についても知りたいところである。この一般ユーザの観点到に基づく人物情報が多く存在するサイトとしては、ソーシャルネットワークサイト (SNS) や掲示板が考えられる。これらのサイトを巡回し、他の一般ユーザが投稿した多数の文章に目を通すことで、一般ユーザが有名人に対して抱いている印象や感想を知ることができる。ただし、そのためには膨大な文章を読む必要があり、自力で一般ユーザの観点到に基づく人物情報を得ようと思うと多大な時間と労力が必要になる。そこで本研究では、SNS における一般ユーザの有名人に関する発言から自動で一般ユーザの観点到に基づく人物情報を抽出し、これを可視化して提示することを目指す。

今回我々は、マイクロブログの 1 つで、一般ユーザの発言が豊富に存在する Twitter を対象に一般ユーザの発言を収集する。Twitter では投稿の単位となる 1 つの発言は tweet と呼ばれ、日本国内において日平均で約 8 千万の tweet が投稿されている [12]。tweet はそのユーザの日常に関する内容であることが一般的だが、他にもその時々で話題になっているニュースや有名人に関する内容の tweet も多く見られる。このため、有名人に関する一般ユーザの発言を得るためのデータソースとしても Twitter は適している。

本研究では、一般ユーザの観点到に基づく人物情報の 1 つとして、一般ユーザ目線での有名人間の類似度の抽出に焦点を置く。すなわち、一般ユーザーの主観的な観点到から似ている有名人を見つけ、その関係を俯瞰できる様にすることを目指す。我々は、A. 有名人に関して言及している tweet, B. 有名人に関して言及している tweet を行ったユーザのプロフィール という 2 点の情報に着目する。また比較対象として、一般ユーザの観点到が反映されない C. 有名人本人の tweet を用いる。

類似度の元となる特徴量の算出には Latent Dirichlet Allocation(LDA)[1] を利用し、有名人毎のトピック分布を求める。そして、推定したトピック分布を元に主成分分析、多次元尺度構成法、自己組織化写像のいずれかをを用いて平面上に可視化して提示するシステムを提案する。

## 2 関連研究

Twitter 上の tweet を利用して情報抽出や意見要約を行う研究には様々なものがある。中でも、商品やテレビ番組などのアイテムに対する評判情報を分析し、マーケティングやその他のアプリケーションに応用しようという研究は多く見られる [7, 9]。一方、アイテムだけでなく人物に対する評判情報を扱った研究も近年増えてきている。Meng らは、あるエンティティ(アイテムや人物などの総称)に関する Twitter 上での意見要約を行った [4]。彼らは、あらゆるエンティティに関しての tweet 群から、ハッシュタグに注目してクラスタリングにより  $N$  個のトピックを見つけ、それぞれのトピック毎に極性も考慮した要約を行った。また、Twitter 以外の SNS の情報を利用した研究には次の様なものもある。Park らは、実名公開型の SNS である Cyworld における政治家への一般ユーザのコメントから、セマンティックネットワーク解析および感情分析を用いてその政治家への集団感情を分析した [5]。Guy らは、社内 SNS における 9 つのソーシャルメディアから他者への評価を元にユーザ間の類似度を求めている [3]。

本研究は、一般ユーザの対象人物に関する発言だけでなく、その発言を行ったユーザの特徴(すなわちプロフィール文)に注目した分析も行っている。また、情報源の種類に着目した先行研究 [6] の内容に加え、複数の可視化手法を用いて結果を提示し、各手法毎の特徴に関して分析を行っている。

## 3 提案システム

我々は Twitter 上の情報から、一般ユーザの観点に基づく人物情報を抽出し、その結果を可視化して提示するシステムを提案する。このシステムは大きく 3 つのモジュールから構成されている。

1. 前処理: tweet 中に含まれるノイズに対する処理および形態素解析を行う。
2. 特徴量算出: 人物毎のトピックに基づく特徴量を求める。
3. 可視化: 人物間の類似関係を平面上に可視化する。

### 3.1 情報源選択

人物情報の情報源として何を用いるかによって、人物に対して注目する観点が変わってくると思われる。我々は一般ユーザの観点に基づく情報源として次の A と B を、比較対象として C を用いる。

#### A. 有名人に関して言及している tweet

対象有名人  $i$  の名前を含む tweet 群に対して、前処理を行って得られた単語素性集合を文書  $d_i$ 、文書  $d_i$  の集合を  $D$  とする。この入力データは、一般ユーザの対象有名人に対する率直な発言であることが多い。そのため、一般ユーザが抱いている対象有名人に対する印象・感想という観点で有名人間の比較を行うことができる。

#### B. 有名人に関して言及したユーザのプロフィール

対象有名人  $i$  の名前を含む tweet(上記情報源 A)を行ったユーザの集合を作成する。この集合中の各ユーザのプロフィール文<sup>1</sup>に対して、前処理を行って得られた単語素性集合を文書  $d_i$ 、文書  $d_i$  の集合を  $D$  とする。この入力データは、その tweet を行ったユーザの社会的情報や趣味・興味に注目している。そのため、対象有名人に関心を持っているユーザ層の違いという観点で有名人間の比較を行うことができる。

#### C. 有名人本人の tweet

対象有名人  $i$  が発信した tweet に対して、前処理を行って得られた単語素性集合を文書  $d_i$ 、文書  $d_i$  の集合を  $D$  とする。これは対象有名人が自ら発信した tweet であるため、対象有名人がどう見られたいかを反映していると思われる。そのため、一般ユーザの観点は反映されていないといえる。

### 3.2 前処理

収集した tweet には、一般ユーザの有名人に関する発言以外に、外部のサイトの宣伝や blog の記事名が含まれた文章(以降、ノイズテキスト)が存在する。これらが多く含まれてしまうと一般ユーザが発信する印象・感想に関する情報の抽出が妨げられる。上記の様なノイズテキストは、URL を伴いかつ複数の tweet により参照されることが多い。そこで、URL を伴う tweet に注目し、2 つ以上の tweet に共通して出現するテキスト部分を発見することで除去を行う。処理内容の詳細に関しては、先行研究 [6] を参照。

また、この処理とは別に URL とハッシュタグの除去も行う。加えて、URL を含んでいなくても tweet が長文で完全一致するものが複数ある場合、これらはスパムであることが多いため、tweet の文字数が  $M$  文字以上<sup>2</sup>でテキストが完全一致するものは除外する。

このノイズ処理を情報源 A,C に対して行う。その後、各テキストに対して MeCab[10] を利用して形態素解析を行い、テキストを単語単位に分解する。なお本実験では、固有名詞や専門用語を扱える様にするため、事前に

<sup>1</sup>ユーザが自由に記入できる 160 文字以内の自己紹介文。

<sup>2</sup>我々は tweet の文字数上限の 3 割 (42 文字) 以上に設定した

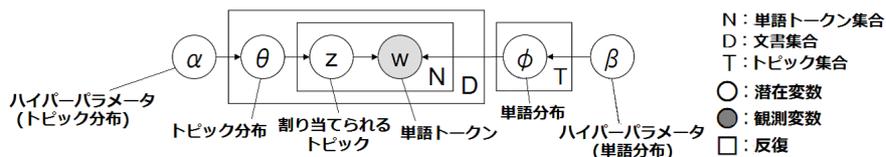


図 1 LDA のグラフィカルモデル

MeCab で利用する辞書に対して Wikipedia に存在する項目名の登録と, TermExtract[11] を用いて Tweet 中によく出現する専門用語やフレーズを抽出して登録する処理を行った. 形態素解析後は, 有名人毎に単語の品詞が名詞・形容詞・動詞と判定されたものを bag-of-words 表現で表し, これを文書  $d_i$  とする. この  $d_i$  の集合を文書集合  $D$  として以降用いる.

### 3.3 特徴量算出

前節で作成した文書集合  $D$  を入力とし, 各人物 (文書  $d_i$  と対応している. 以降, 添字の  $i$  を省略して  $d$  と表す) のトピックに基づく特徴量を求める. 文書毎の特徴量を求める手法としては, 単語の出現頻度と稀少度を考慮した TF-IDF を算出する方法が一般的によく用いられている. ただしこの方法では, 文書中の完全一致する単語しか考慮されておらず, 類似した意味の語の影響を反映させることができない. そこで, 文書の背後に存在するトピックを考慮したソフトクラスタリングを行う手法であるトピックモデルを用いる. 我々はトピックモデルとして, 近年注目されている LDA を使用し, 文書毎のトピック分布を推定する.

#### 3.3.1 LDA

LDA は, Blei らにより提案された確率的トピックモデルである [1]. トピックモデルとは, ある文書 ( $N_d$  個の単語からなるトークン列)  $\mathbf{w}^d = (w_1, w_2, \dots, w_{N_d}) \in D$  が単一または複数のトピックに属する単語から構成されているという仮定をおき, その文書を構成するトピックの比率 (トピック分布)  $\Theta^d = (\theta_1, \theta_2, \dots, \theta_{|T|})$  と, それらのトピック毎の単語生成確率 (単語分布)  $\Phi^t = (\phi_1, \phi_2, \dots, \phi_{|W|})$  に基づき確率的に文書を生成するモデルである. ただし, 文書集合を  $D$ , 各文書のトークン数を  $N_d$ , 全トークン数 ( $N_d$  の和) を  $N$ , 全文書中に出現する全語彙の集合を  $W$ , トピックの集合を  $T$  とする. この LDA の生成過程を表したグラフィカルモデルは, 図 1 のようになる.

我々が入力として与える必要があるのは, 文書集合  $D$ , トピック数  $|T|$ , 反復回数,  $\Theta$  と  $\Phi$  の事前分布であるディリクレ分布のハイパーパラメータ  $\alpha$  と  $\beta$  である. ここで,  $\alpha$  と  $\beta$  は一般にベクトルであるが, Griffiths に従いすべての要素を同じ値に設定する [2]. なお, パラメータ  $\Theta$ ,  $\Phi$  の推定には, Griffiths らの Gibbs Sampling に

よる手法 [2] を用いた.

### 3.4 可視化

前節で求めた人物毎の特徴量を用いて, 人物間の類似関係を平面上に可視化する.

#### 3.4.1 主成分分析

主成分分析 (PCA) により, トピックに基づく人物間の関係を最も表現できる 2 軸を求める. 3.3 で求めたトピックを元の軸とすると, PCA は複数のトピックの影響を反映した人物間の違いを表すのに効果的な新たな軸を発見する. また, PCA は各軸に対して意味的な解釈を行うことが可能であり, 何らかの観点に基づく要約が行われることが期待できる. これはユーザへ提示する際に, ユーザの提示内容に対する理解を促進する効果があると予想される.

#### 3.4.2 多次元尺度構成法

多次元尺度構成法 (MDS) は, 2 者間の距離尺度を維持した配置を求める方法であり, PCA とは異なり配置した空間における軸には意味的な解釈が存在しない. MDS は入力に要素間の距離行列が必要であり, まず 3.3 で求めたトピック分布から距離行列を算出する. 今回用いるのは, 古典的多次元尺度構成法 (CMDS) と呼ばれる最も基本的な MDS 手法である. 次に, 距離行列を算出する方法について説明する.

#### JS 情報量.

確率分布の類似度を測る手法としては, 比較対象の一方の確率分布の値に 0 が存在しないことを制約条件に持つ Kullback-Leibler divergence (KL 情報量) や, 2 つの平均的な確率分布までの KL 情報量の平均を求める Jensen-Shannon divergence (JS 情報量) などがある. 本研究では, 比較の対称性を満たす JS 情報量を用いて距離を求める.

#### 3.4.3 自己組織化写像

自己組織化写像 (SOM) は, ニューラルネットワークによる教師なし学習の一種であり, 類似した特徴を持つ入力同士を近くに配置する様な出力を行う. 入力には 3.3 で求めたトピック分布を  $T$  次元のベクトルとして用いる. ニューラルネットワークである SOM は, 細かな手法のパラエティや多数のパラメータが存在し, これらの設定により結果が変化する性質がある.

春名風花	宮迫博之	松本人志	スギちゃん	山本太郎	ローラ	剛力彩芽	西川貴教	GACKT	きゃりーぱみゅぱみゅ
鬼龍院翔	喜矢武豊	宇治原史規	伊集院光	篠田麻里子	平野綾	中川翔子	安倍晋三	東国原英夫	太田順也 (ZUN)
橋下徹	堀江貴文	孫正義	香川真司	ダルビッシュ有	東浩紀	乙武洋匡	茂木健一郎	田原総一朗	前山田健一 (ヒャダイン)

表 1 対象有名人一覧

トピック解釈 上位単語	ジャニ系	アイドル	V系	アニメ・ゲーム	政治・経済	学問	スポーツ	学生	BOT
	嵐	AKB	金爆	アニメ	日本	物理	サッカー	好き	紹介
	NEWS	推す	ゴールデンボンバー	好き	政治	伊織	野球	フォロー	名言
	組	前田敦子	V系	ゲーム	反対	ぼるぼる	香川真司	音楽	面白い
	二宮	篠田麻里子	ギルド	ボカロ	脱原発	村上春樹	選手	気軽	よろしくお願ひします
	相葉	SKE48	シド	漫画	TPP	文学	MLB	野球	ニュース
	潤	NMB48	己龍	声優	日本人	ゲンロン	SAMURAI	高校	2ch

表 2 主なトピック (情報源 B: 有名人に関して言及したユーザのプロフィール)

## 4 実験

3種類の情報源に対する可視化結果の特徴の分析と、3種類の可視化手法に対する可視化結果の特徴の分析をそれぞれ行う。分析する特徴としては、提示結果がある程度事実と則しており一般的に納得できるものであるか(妥当性)、人物間の意外な関係や集団を見つけられるか(発見性)という2点に着目する。なお、本実験では可視化結果のスクリーンショットを示し、その中で発見されたグループ(密に集まった有名人集合)に注目するが、そのグループはスクリーンショット上で四角で囲んで、また(a)~(e)の記号を振って示すことにする。各グループを本文中で参照するときには、(a)の様に表記する。

### 4.1 データセット・パラメータ

#### データセット

Twitter公式APIを利用して、表1の有名人に対してそれぞれ情報源A~Cのデータを収集した。これらの有名人は、有名人ツイッターランキング<sup>3</sup>を参考に政治・芸能・経済・スポーツ・文化人など様々な分野から、十分な量のデータが収集可能なTwitterにおいて人気の人物を選んでいる。収集するデータの詳細は次の様になっている。情報源Aは、対象有名人の名称・呼称<sup>4</sup>をクエリとして検索で得られた2013年9月2日から8日までの1週間分のtweetである。情報源Bは、情報源Aに含まれるtweetを行った各ユーザの2013年9月時点でのプロフィールである。情報源Cは、対象有名人の2013年9月時点から最大で3200まで遡ったtweetである。

#### パラメータ

LDAのトピック数 $|T|$ は、[8]の手法により形成されたクリーク数である17と設定した。LDAのハイパーパラメータは $\alpha = 50/|T|$ 、 $\beta = 0.1$ 、反復回数は1000とした。SOMに関しては、 $30 \times 30$ の正方形格子、逐次型学習、反復回数は1万回、近傍関数はガウス関数、ベクトル間の比較はコサイン類似度、参照ベクトルの初期値

は乱数で設定した。

### 4.2 情報源毎の結果

先行研究[6]では、情報源毎の可視化結果における特徴を調べた(可視化手法はMDSに固定)。ここでは、各情報源で得られたトピックの特徴と、妥当性・発見性の最終的な評価結果のみを記す。

情報源Aは、対象人物に対する発言内容のトピックに注目しており、このため時事ニュースや人物に対する現在の世論の影響を強く受ける傾向がある。トピックの種類としては、好意的・批判的といったポジティブ・ネガティブを表すものや、アイドル・俳優・お笑い芸人・政治・ゆるキャラなど話題に関連するグループ、時事ニュースがトピックとして見られた。情報源Bは、対象人物に言及したユーザのプロフィールであり、これを用いれば同じユーザ層に関心を持たれている人物同士が近くに配置されることが期待される。トピックの種類としては、主にユーザが興味のある対象が得られている他、学生やBOTの様なユーザ層そのものも得られた。情報源Cは対象人物本人のtweetに注目しており、一般ユーザの観点は反映されていない。トピックの種類としては、政治・モデル撮影といった有名人本人の職業に関するトピックの他、ライブイベントの告知やブログ更新の告知といったトピックが見られたが、それ以外のトピックは解釈できないものが多い結果となった。

最終的な妥当性・発見性の評価結果を表3に示す。有名人の塊が意味のある分け方になっているかどうかを表す妥当性に関しては、情報源Aと情報源Bの両方で優れた結果を得られていると判断できた。興味深い新たな塊を見つけることができたかどうかを表す発見性に関しては、情報源Bが最も優れており、情報源Aは情報源Bに比べるとやや劣る結果となった。一方、情報源Cは情報源Aと比較して妥当性に関して劣る結果となった。また、情報源Bにおいては集団がより密になる傾向があるため、グループの見つけやすさという点でも優れていた。

<sup>3</sup>www.talenttwit.com

<sup>4</sup>Wikipediaに登録された名称に加え、Wikipediaの本人記事中に記載された愛称・略称も用いる

情報源	A	B	C
主なトピック	話題内容, 感情極性	ユーザの趣味・興味	有名人の仕事内容, 広報
妥当性	○	○	△
発見性	△	○	△

表 3 情報源毎の結果まとめ

可視化手法	PCA	MDS	SOM
妥当性	○	◎	△
発見性	◎	◎	△

表 4 可視化手法毎の結果まとめ

### 4.3 可視化手法毎の結果

情報源を B に固定した場合の可視化手法毎の特徴を確認した。なお、このマップは情報源に B を用いているので、その人物に関心を持っているユーザ層に基づいて類似度を求めていることになる(表 2 に具体的なトピックの一部を示す)。PCA による可視化結果を図 2 に、MDS による可視化結果を図 3 に、SOM による可視化結果を図 4 に示す。ただし、PCA の可視化結果は離れた位置に大きく 3 つの集団が現れる結果となっており、紙面の都合上その中間部分を省略(白抜きで表示)している。

PCA による可視化結果の主な特徴として、縦横の軸に意味が存在しているため集団の解釈がしやすいという利点がある。今回の結果では、縦軸がカジュアル・フォーマルを表していることが分かる一方、横軸の解釈は困難であった。妥当性に関しては、やや外れ値が見られるが、ユーザが持つ関心の対象を政治家(図 2-(a))と芸能人(図 2-(c))に大きく分けることができている。発見性に関しては、知識系芸能人(図 2-(b))の他、最下部にはネットをよく利用するユーザや V 系バンドに関心があるユーザといった特定のユーザ層に人気がある有名人(図 2-(d,e))が見られる。

MDS による可視化結果の主な特徴としては、基本的に PCA と似た様な人物配置となっている。4.2 節の情報源 B で説明した様に、図 3 は人物が集中している部分はより密集している傾向が見られ、PCA よりも集団が明確に表れている。特に、図 2 の PCA の結果では、芸能人の塊がややばらけているが、MDS ではこれらが密になっている。妥当性で言うと、外れ値が少ない点で、PCA よりも優れていると言える。発見性に関しても、4.2 節の情報源 B で述べたとおり、多くの興味深い塊が見つかっており、これは PCA と同等と言える。

SOM による可視化結果の主な特徴として、格子状の座標へ飛び石の様に人物が配置される(図 4 に格子は表示していない)。妥当性に関しては、政治家が割り当てられたノード(図 4-(a))は妥当なものであるといえる。中央部やや右下の密集した部分(図 4-(b))に注目すると、芸能人が多く集まっているが、田原総一郎と孫正義といった評論家・実業家やネット上の有名人も含まれているため妥当性には疑問が残る。加えて、(c)の集団に含まれていた方がよい芸能人として、PCA・MDS と比較してきゅりーばみゅばみゅ、MDS と比較して篠田麻里子が(c)の集団から孤立している点も妥当性に疑問を残

す要因となっている。発見性に関しては、PCA や MDS と比べて V 系バンドの人物が多く現れている(図 4-(c))が、知識系芸能人やネット上の有名人といったグループは見られなかった。

以上の結果をまとめたものを表 4 に示す。これまで述べてきたとおり、有名人の塊が意味のある分け方になっているかどうかを表す妥当性に関しては、MDS が最も優れた結果を得られた。興味深い新たな塊を見つけることができたかどうかを表す発見性に関しては、PCA と MDS のどちらも同等の優れた結果を得られていると判断する。PCA の妥当性に関しては、やや外れ値が多いため、MDS より劣る結果となった。また、SOM は PCA と比べて、妥当性と発見性の両方で劣る結果となった。

## 5 おわりに

本稿では、Twitter から得られる情報を元に、一般ユーザの観点に着目した人物間の類似関係の可視化を行うことを目指した。我々はこの目的を達成するために、多くのノイズが含まれたテキストに対して前処理を行い、各情報源から得られるトピックに基づいた人物毎の特徴量を算出し、これを元に 2 次元平面上での配置を決定した。そして、可視化結果に対して使用した情報源毎、および可視化手法毎に妥当性と発見性に着目して特徴の分析を行った。分析の結果、我々が用いた一般ユーザの観点に基づく情報源は事実に基づく妥当性を保ちつつ、知識人として捉えられている芸能人やネット上で人気の有名人といった容易に思いつかない発見も可視化結果に反映させられることを示した。また、情報源の中では、有名人の名前を含んだ tweet の著者のプロフィール文が最も妥当性が高く、なおかつ多くの発見が行えることが分かった。可視化手法の中では、2 者間の距離尺度を維持した配置を求める方法である MDS が最も妥当性を保ち、多くの発見を行えることが分かった。

今後は、人物同士だけでなく、人物とアイテムといった異なるクラス間の一般ユーザの観点に基づく関連性の抽出を行う研究や、それらの関連性を元に推薦を行うシステムの作成を検討していく予定である。

## 謝辞

この研究は、三菱電機株式会社先端技術総合研究所との共同研究である。

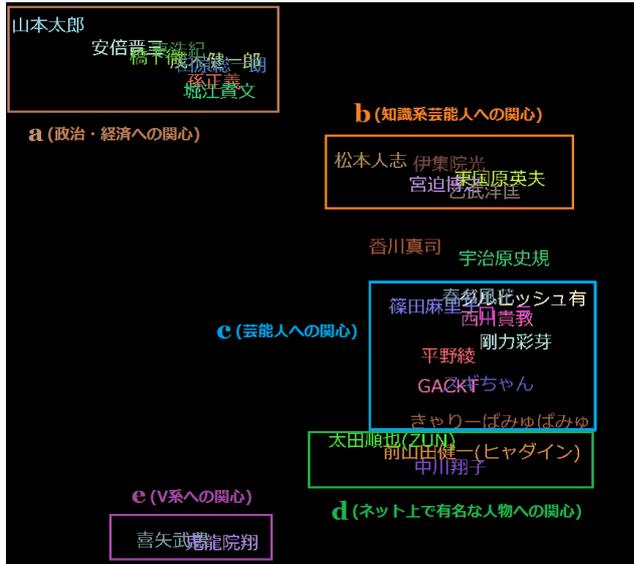


図3 情報源 B(関心を持つユーザ層) - MDS

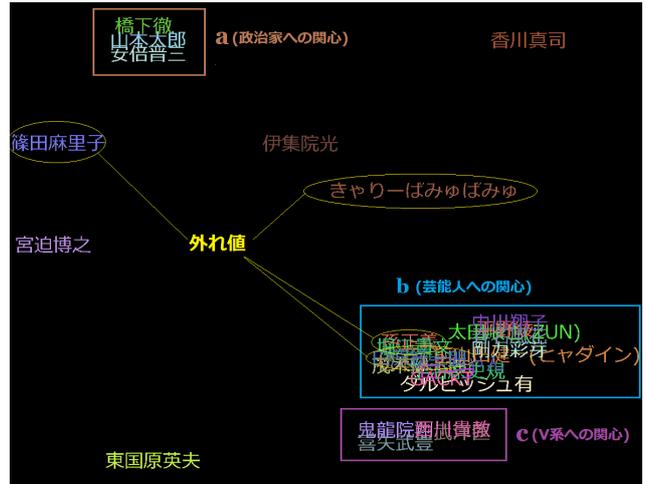


図4 情報源 B(関心を持つユーザ層) - SOM

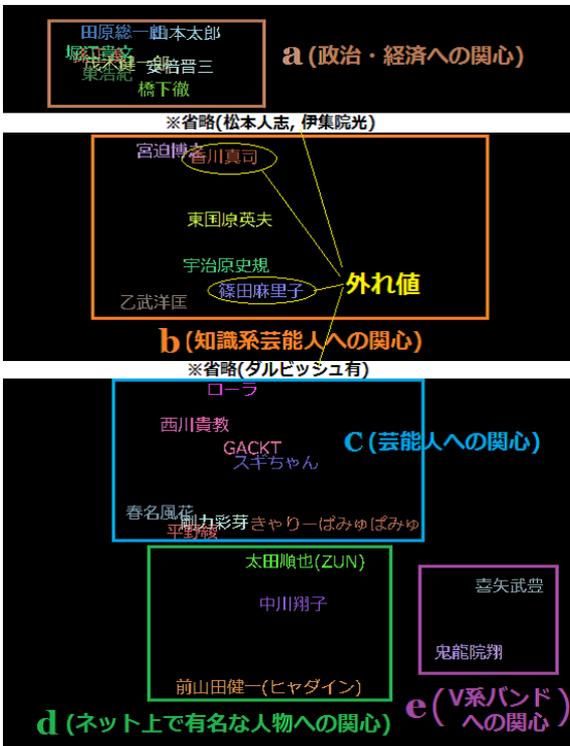


図2 情報源 B(関心を持つユーザ層) - PCA

参考文献

- [1] Blei, D.M., Ng, A.Y. and Jordan, M.I.: Latent Dirichlet Allocation, The Journal of Machine Learning Research, Vol.3, pp.993-1022 (2003)
- [2] Griffiths, T.L. and Steyvers, M.: Finding Scientific Topics, Proc. of National Academy of Sciences, Vol.101, pp.5228-5235 (2004)
- [3] Guy, I., Jacovi, M., Perer, A., Ronen, I. and Uziel, E.: Same Places, Same Things, Same People? Mining User Similarity on Social Media, Proc. of ACM CSCW, pp.41-50 (2010)
- [4] Meng, X., Wei, F., Liu, X. and Zhou, M.: Entity-centric topic-oriented opinion summarization in twitter, Proc. of ACM SIGKDD, pp.379-387 (2012)
- [5] Park, S.J., Lim, Y.S., Sams, S., Sang, M.N. and Park, H.W.: Networked Politics on Cyworld: The Text and Sentiment of Korean Political Profiles, The Journal of Social Science Computer Review, Vol.29, No.3, pp.288-299 (2011)
- [6] 西村章宏, 土方嘉徳, 三輪祥太郎, 西田正吾: 一般ユーザの観点に基づく Twitter からの人物関係の可視化, 情報処理学会研究報告, DBS 研究会 (2014)
- [7] 池田和史, 服部元, 松本一則, 小野智弘, 東野輝夫: マーケット分析のための Twitter 投稿者プロフィール推定手法, 情報処理学会論文誌 コンシューマ・デバイス&システム, Vol.2, No.1, pp.82-93 (2012)
- [8] 芹澤翠, 小林一郎: 潜在的ディリクレ配分法に基づくトピック類似度を考慮したトピック追跡, DEIM (2011)
- [9] 山本祐輔, 浅井洋樹, 上田高徳, 秋岡明香, 山名早人: テレビ番組に対する意見をもつ Twitter ユーザのリアルタイム検出, DEIM (2013)
- [10] MeCab: Japanese morphological analyzer, 入手先 (https://code.google.com/p/mecab/), (参照 2013)
- [11] TermExtract: 専門用語 (キーワード) 自動抽出用 Perl モジュール, 入手先 (http://gensen.dl.itc.u-tokyo.ac.jp/termextract.html), (参照 2013)
- [12] ビッグローブ: 5月のTwitter利用動向を発表 (https://www.biglobe.co.jp/pressroom/release/2014/06/140609)