

マイクロブログ上の話題抽出とユーザの態度の分類に基づく 流言検出支援システム

藤川 智英[†] 鍛冶 伸裕[‡] 吉永 直樹[‡] 喜連川 優[‡]

[†] 東京大学 大学院 情報理工学系研究科 電子情報学専攻 〒153-8505 東京都目黒区駒場 4-6-1

[‡] 東京大学 生産技術研究所 〒153-8505 東京都目黒区駒場 4-6-1

E-mail: [†] {fujikawa,kaji,ynaga,kitsureg}@tkl.iis.u-tokyo.ac.jp

あらまし 現在ツイッターなどのマイクロブログ上では様々なニュースやそれに対する反応が数多くツイートされているが、その中には真偽不明の情報（流言）も多い。本研究では、流言の検出を支援するため、話題のニュースやそれに対する反応の抽出を行うシステムを提案する。まず、各単語に対してその出現回数の確率の低さから、出現回数の急上昇（バースト）の度合いを計算し、話題となっている単語群の抽出・クラスタリングを行う。そして、その話題に関連するツイートの抽出し、ユーザの態度を「信疑の有無」「根拠の有無」で分類を行い、各ユーザに対して流言かどうかを判断する手がかりを与える事を考える。

キーワード 自然言語処理, テキストマイニング, 話題抽出, 情報信頼性

1. はじめに

ツイッターなどのマイクロブログの登場は情報の共有を手軽にし、現在ネット上では様々な情報が非常に速く広がるようになってきている。しかし、それらの中には根拠のない風説（流言¹）も含まれており、特に災害などの非常時には爆発的に流言が広がることが多い。例えば東日本大震災では流言が少なくとも 80 件広がった [1]。こうした流言は人間の安全や名誉に関わるものも多いため、迅速に発見し、サイトなどを通じてその存在を明らかにすることが重要と考えられる。

そこで、本研究では、話題の抽出を行い、さらにその話題に対してユーザがとっている態度を分類することで、流言の可能性のある情報を提示する流言検出システムを提案する。

話題抽出では流言であるかどうかにかかわらずユーザがその内容について多くツイートをこなっている話題を、話題タイトルと関連ツイートのペアで抽出する。具体的にはまず、話題の内容を記述する文に出現する頻度が急上昇（バースト）することを手がかりに、既存のバースト検出手法で検出した話題語をクラスタリングすることで話題語クラスタを得る。次に話題語を手がかりに各話題クラスタに関連するツイートを抽出する。最後に各クラスタに含まれる話題語を多く含む表現をツイートから話題タイトルとして抽出する。このように、話題語を手がかりに話題抽出を行うことで、既存の Tweetbuzz のような URL を手がかりにした話題抽出手法に比べて頑健かつ高いカバレッジで話題の認識が可能となる。

次に、関連ツイートに含まれるユーザの態度の分類では、ユーザが話題の内容を疑っているかどうか、また、その根拠を述べているかどうかという 2 つの観点

でツイートの分類を行い、その結果をユーザに提示する。

2. システムの概要

全体のシステム概要は **Figure 1** で表される。

まず、Streaming API でツイッターのツイートを収集し、後述する話題抽出アルゴリズムを適用して、話題を抽出する。

次に、解析した話題からクエリを作成し、Filtering API を用いて関連するツイートを収集する。それに対して「疑っているかどうか」と「根拠の有無」で分類を行う。

疑っている人がどの程度いるかを表示するとともに、根拠を伴って意見を述べているツイートを提示し、流言かどうかを判断する手がかりを与える。

3. 話題の抽出

3.1. バーストスコアの計算

話題の内容に関する単語を発見するためには、まず頻度が急上昇（バースト）している単語を見つける必要がある。一定期間（例：30 分）収集したツイートにおいて、ある単語がバーストしているかどうかを判断するため、そのような事象が発生する確率がどれほど低いかを計算することを考える。

[2]では単語が二項分布に従って出現すると仮定している。つまり、まず、以下のように単語の平均出現率 p を計算する。

$$p = \frac{\text{その単語が出現したツイートの数}}{\text{全ツイート数}}$$

¹ 流言という用語は従来、真疑不確かな情報一般に対して用いられるが、本稿では文脈上誤解のない場

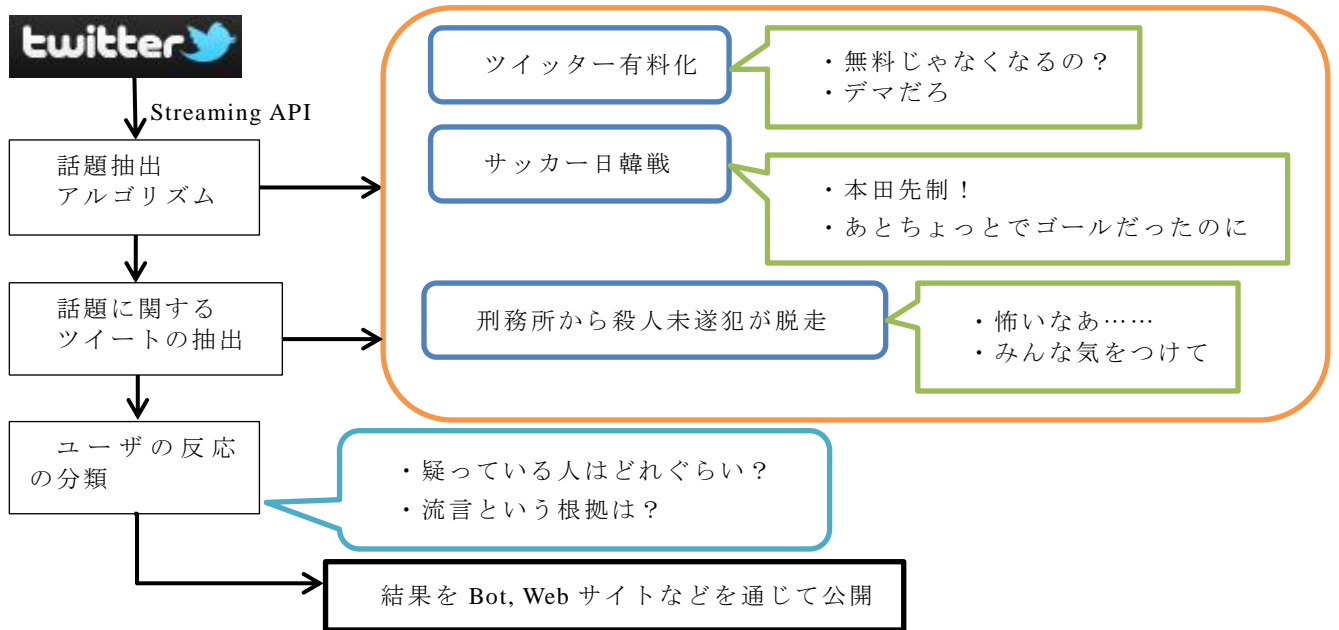


Figure 1 システムの概要

各ツイートに確率 p で単語が独立に出現すると仮定すると、 n 個のツイートを収集した時、そのうち k 個のツイートに単語が出現する確率は、以下の式になる。

$$\text{prob}(n, k) = p^k (1 - p)^{n-k} \binom{n}{k}$$

二項分布では単語の出現率 k/n が平均出現率と一致する場合最も確率が高くなり、そこから離れるに従って急激に確率が減少する。

本研究では、単語の出現確率が平均出現率 p を上回り、かつ、出現回数が観測された出現回数 k 以上になる確率が 0.05 以下のとき、その単語がバーストしている可能性があるものとみなす。このとき、そのバーストのスコア（バーストスコア）を、以下の式で定義する。

$$\text{score} = -\log \sum_{m=k}^n \text{prob}(n, m)$$

3.1.1. PageRank を応用したバーストスコアの調整

前節では確率が 0.05 以下の事象を「バーストの可能性はある」とみなしたが、例えば 1 万種類の単語があれば、そのうち 0.05 、 500 個はたとえバーストしてなくても、この 0.05 の領域に入ってしまう。そのため、実際に話題になっているものと、偶然バーストしているものを区別する必要がある。

そこで何か話題になる場合、複数の単語が同時にバーストすることが多いことに着目する。例えばサッカーの日韓戦では「サッカー」「日本」「韓国」「本田」（選手名）などの単語が同時にバーストする。また、「スマトラ沖で地震」というニュースでは「スマトラ」「地震」「津波」「マグニチュード」「インドネシア」が

同時にバーストする。

そこで、多くの単語と共起している単語のバーストスコアが高くなるように調整を行う。このために、本研究では PageRank [3] をテキストに対して応用した TextRank [4] を用いる。

PageRank では、多くの重要なページからリンクされているページほど重要なページであるというアイデアに従い Web ページの重要性を判断する。同様に、TextRank ではリンクを共起度に置き換え、共起度が高いものほど重要であるとみなす。本研究では以下のような改良を加えて使用した。

各話題語に対して TextRank を計算するにあたって、まず各単語の共起の強さ（共起度）を表す行列が必要となる。本研究の場合では w_i, w_j の共起度 $c_{i,j}$ を次のように計算する。

$$c_{i,j} = \frac{|w_i \cap w_j|}{|w_i \cup w_j|} \cdot \text{idf}_i$$

ここでは、共起度として Jaccard 係数に w_i, w_j の idf (逆文書頻度) をかけたものを用いている。idf をかけるのは、珍しい単語ほど共起しにくく、そのぶん重要であると考えられるためである。なお、 $c_{ii} = 0$ である。また、計算量削減とノイズの除去のため、 $c_{i,j}$ が 0.10 以下の場合 0 と再設定した。

次に、推移行列 A は C を各列において正規化したものである。I 番目の列の要素の合計を s_i とおくと以下のようなになる。

$$A = (a_{i,j}) = \begin{cases} \frac{c_{i,j}}{s_i} & (s_i \geq 1 \text{ のとき}) \\ c_{i,j} & (s_i < 1 \text{ のとき}) \end{cases}$$

PageRank を用いたバーストスコアの再計算は以下の

式の繰り返しで行う。

$$R = dAR + (1 - d)R_0$$

ここで、 R_0 は i 番目の要素が単語 w_i の最初のバーストスコアを正規化したものであるような列ベクトルである。パラメタ d は一般的な PageRank の例にのっとり、0.85 と設定する。

このようにして得られた上位 N 個の単語を話題語として選択する（実験では $N=150$ とした）。

3.1.2. 話題語のクラスタリング

前節で得られた話題語をクラスタリングし、ひとつの話題に対応する話題クラスタを得る。クラスタリングは階層的クラスタリングを用いる。ただし、類似度 r の計算は以下のように行う。クラスタ $C1, C2$ が存在した時、そこに含まれる単語を $C1 = \{w_{11}, w_{12}, \dots, w_{1n}\}, C2 = \{w_{21}, w_{22}, \dots, w_{2m}\}$ とすると、

$$r = \sum_i \sum_j W_i W_j c_{ij}$$

ここで、 W_i, W_j は w_{1i}, w_{2j} のそのクラスタ内での正規化された重みであり、ランクの高い単語との共起ほど重要とみなすようになっている。

3.2. 関連するツイートの収集

次に、クラスタと関連のあるツイートを抽出する。各ツイートに出現する単語に対して、前節で得られたスコアを求め、その合計をそのツイートのスコアとし、それが一定の閾値以上ならば関連するツイートであるとみなす。

例えば「少女時代、KARA が紅白に内定」というニュースで、各単語の重み付けが次のようになっていたとする。

- 少女 : 0.15
- 時代 : 0.25
- KARA : 0.30
- 紅白 : 0.30

このとき、あるツイートに「少女」「時代」が含まれていればスコアが $0.15+0.25=0.40$ 、「KARA」「紅白」が出現していればスコアは $0.30+0.30=0.60$ というように計算する。

これにより、URL などと関係なく、「中心的な単語が多く含まれているかどうか」のみで話題と関連があるかどうかを判定するため、テレビ番組などが発信源の話題に対しても適用することができる。

3.3. フレーズと文の取得

クラスタリングを行ったあとは、そのクラスタの要約のため、[5]を用いる。この手法では、キーとなる単語を中心として、隣接する単語のグラフを作り、最も重みが大きくなるグラフを取り出し、それを要約とみなす。

例えば以下のようなツイートが存在する場合を考

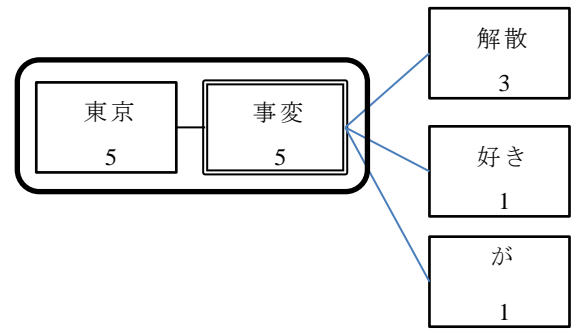


Figure 2 フレーズと文の抽出

える。

- 東京事変解散、公式 HP で発表 <http://t.co/AAA>
- えっ、東京事変解散したの
- そんな、東京事変が……
- 東京事変解散とかショックすぎる
- 東京事変好きだったのに……

ここで、クラスタ内のスコアが上位の単語（ここでは「事変」とする）を中心にグラフを作成し、出現数が1のものを除いた、これにより Figure 2 のようなグラフを得る。

元の論文では各グラフから最も重みの大きい経路を選択しているが、本研究では次のような改良を加えている。まず、出発点の単語の出現回数の 3/4 以上出現している単語の連なりを「フレーズ」（ひとかたまりの言葉）とみなす。この例では「東京」-「事変」がこれに当たるため、「東京事変」というフレーズが取り出すことができる。

次に、出現回数が全体の 1/20 以上で、次の条件を満たす経路を、そのクラスタを表す代表的な「文」とみなすことを考える。

1. 「は」「が」などの係助詞、格助詞が含まれる
2. 「存在」「解散」などのサ変接続名詞が含まれている

ここでは「東京事変解散」が「代表的な文」としてとり出される。

また、URL が同時に出現している場合は、その内の一つを代表として抽出する。

3.4. 実験

3.4.1. 関連ツイート抽出

関連ツイート抽出においては、閾値を小さく設定すると少しでも関係ある単語が含まれていれば「関連ツイート」とみなされるため、取りこぼすことはなくなる一方関係ないツイートを「関連ツイート」としてとり出してしまふことが多くなる。逆に高く設定すると関係ないツイートを「関連ツイート」とみなすことは少なくなるが、取りこぼしも多くなる。そのため抽出

精度は Figure 3 のようなグラフを描く。

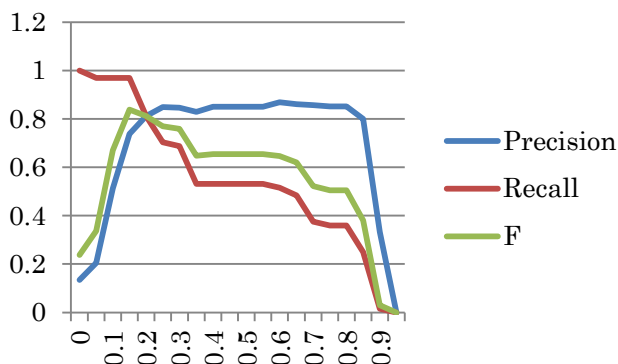


Figure 3 閾値と抽出精度の関係

適切な閾値を調べるため、以下のような実験を行った。11個のニュースに対して、関連するツイートかどうかの判定精度を調べた。なお、すべてのツイートを調べると莫大な数になり、かつ、そのうち関連するツイートは多くても全体の0.5%ほどであるため、クラスター内の単語が最低でもひとつ含まれているもののみを対象とした。

各ニュースに、閾値を0.05刻みで変化させ、Precision, Recall, F-値を計算した。例えば「ツイッターが国の状況により、検閲を行う意向を表明」というニュースでは Figure 3 のようになった。

各ニュースに対して、F 値が最大となるしきい値、更にもその F 値から 0.1 引いたものを「許容値」と仮定して、それを満たす最小のしきい値と最大のしきい値を調べた。その結果をグラフにしたものが Figure 4 である。ここでは、F 値が「許容値」以上となる閾値の範囲を誤差棒で表した。

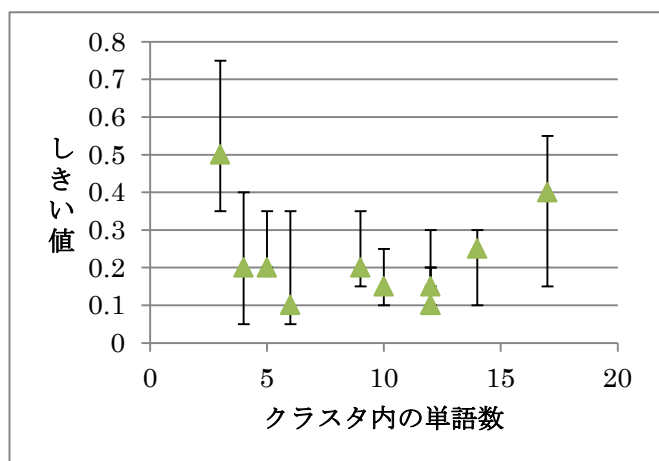


Figure 4 F 値を最大にする閾値の分布

当初クラスタ内の単語数が多いほど、スコアが分散するためしきい値は低くなると予想したが、実際には特に関連はなかった。図を見ると、どの場合でも 0.1

から 0.3 の付近に位置していることがわかる。

各閾値の中央値は Table 1 のとおりである。なお、最大 F 値の平均は 0.827 である。

これらの結果から、本研究では閾値を 0.2 と設定した。

Table 1 各閾値の中央値

F 値が最大になる閾値	「許容値」を満たす最小閾値	「許容値」を満たす最大閾値
0.2	0.1	0.35

3.4.2. 実際の結果

以上の手順を行い得られる例を Table 2 に示す。なお、これは 2012 年 2 月 14 日~16 時に得られたものである。

Table 2 結果の一例

代表的な文	福島第一原発:2号機の温度、79.1度 に上昇 - 毎日 jp(毎日新聞)
話題語	号機、上昇、温度、冷温、停止、注水、容器、圧力、規定、保安、上限、読売、原子、炉
関連ツイート	2号機、誤差が±20℃の温度計で計測してたらしいのに 79.1℃…コンマ 1℃ってどういうアレ?(° ° ;)\(--;)
	…え? 原発 2号機やばくね…?
	本日 12日現在、2号機の温度、79.1度 に上昇。80度を超えると冷温停止状態とは完全に言えなくなる。

また、話題を bot² @trend_words_jp³とサイト⁴を通じて公開を行なっている。

4. ユーザの反応の分類

4.1. 反応を分類することの必要性

Web 情報をユーザがどう信じるかに関する研究では、正しい情報というのは必ずしも客観的に定義できないため、最終的な判断を行うのはユーザであり、正しさをシステムによって完全に客観的に決定することはできないと指摘している。なぜなら、情報の信憑性は「情報の受け手によって認知される特性」であるため、その判断基準は人によって異なるからである [6]。

そこで、情報信頼性に関する研究では、「信憑性」を直接判断するのではなく、「信憑性」を判断するための材料を提示することを目指す研究も多い。情報分析システム WISDOM [7]では、「電気自動車は環境に良い」「裁判員制度」など知りたい意見を入力すると、それに対する賛成意見、反対意見がどの程度存在するか、

² 自動的に発言を行うツイッターアカウント

³ https://twitter.com/#!/trend_words_jp

⁴ <http://www.tkl.iis.u-tokyo.ac.jp/~fujikawa/trend/>

どのような人がどのような意見を述べているかなどを一覧表示することで判断の手助けを行うことを目指している。

流言の可能性のある情報の検出においても、ユーザーに判断材料を与えるためには、話題に対して人々がとっている反応を分析することが重要であると考えられる。

こうした研究としてはすでに疑っているかどうかで分類を行う [8]が存在するが、本研究ではさらに「根拠の有無」でも分類を行う。

4.2. 分類基準

まず、「信疑の分類」の分類基準について説明する。一口に疑いと言っても完全に疑っているものから半信半疑、「本当ならひどい話です!」のように一応疑っているがかなり信用よりのものまで幅広い。しかし、ここでは、わずかでも疑っているものは疑いに分類する。なぜなら、細かく分類するとモデルが複雑になり、ラベル付の際にも判断に個人差が出てくるためである。また、根拠のはっきりした情報なら「本当なら」のように留保をつけることもないと思われるため、真偽不明の情報特有のツイートであると考えられるからである。

例えば有名人が死んだという流言が流れたとき、「○○(有名人)が死んだって本当?」「デマだろ」のような発言を「疑い」、「○○(有名人)が交通事故でなくなったそうです。ショック……」のように、まったく疑っておらず、普通のニュースと同様に受け取っている場合は「信用」と分類する。

次に、各発言におけるユーザーの態度(信疑)に対する根拠の有無の判断基準について述べる。根拠としては、具体的な URL やツイートの引用など、その根拠の内容が明白なものに限らず、発言の文面からユーザーが確信を持って判断する材料を持つと考えられる場合に根拠ありとする。例えば「○○が死んだって話、ジョークサイトの嘘記事ですよ。」という発言では、単にそれが誤りであると指摘するだけでなく、なぜ誤りと言えるかの根拠も示しているため、「根拠あり」と分類する。

4.3. 分類手法

本研究では、テキスト分類問題で広く使われる分類器である SVM を、教師付きデータから学習し、各分類タスクに適用する [9]。流言に対するツイートを観察した結果、おおよそ次のような特徴が見られた。

- ユーザーが情報を信じている場合は「ひどい」「感動した」などの感情的表現が含まれることが多い
- 流言であると指摘しているツイートは「流言」「ガセ」など直接的な言葉を使っているツイートが多い
- 根拠を示しているツイートは、何らかの事実を提示するため、「○○は××のはず」「××じゃない」など特定の言い回しのパターンが好んで用いられる

これらの特徴から考えて、出現する単語や文体を素性とすれば分類が可能と考えられる。実際に用いた素性の詳細は次のとおりである。

- 単語の 1、2、3-gram
- 品詞(第一階層と第二階層)の n-gram
 - 例: “東京都が株主”
 - 第一階層: 名詞-助詞-名詞
 - 第二階層: 固有名詞-格助詞-名詞
- 文章長(140で割り、1以下となるように正規化)

ここで、品詞(POS)の n-gram を使用しているのは、品詞の連なりは文体を表すためである [9]。根拠を提示している文では、次のように話題によって提示される根拠も異なり、当然のことながら含まれる単語も異なる。

- 株主は東京都です。都知事じゃありません。
- 現在ではザイルという国は存在しません。

そのため、共通する要素である文体を捉えることが重要と考えられる。

Table 3 信疑の分類結果

	Baseline	tf			tf-idf		
		単語 n-gram のみ	+固定長品詞 n-gram	+可変長品詞 n-gram	単語 n-gram のみ	+固定長品詞 n-gram	+可変長品詞 n-gram
Accuracy	0.63	0.749	0.785	0.758	0.779	0.779	0.789
Precision	0.63	0.83	0.872	0.839	0.848	0.852	0.841
Recall	1	0.763	0.777	0.767	0.796	0.791	0.825
F-measure	0.773	0.795	0.822	0.801	0.822	0.82	0.833

Table 4 根拠の有無の分類結果

	Baseline	tf			tf-idf		
		単語 n-gramのみ	+固定長品 詞 n-gram	+可変長品 詞 n-gram	単語 n-gramのみ	+固定長品 詞 n-gram	+可変長品 詞 n-gram
Accuracy	0.438	0.704	0.749	0.785	0.692	0.737	0.722
Precision	0.438	0.758	0.804	0.825	0.795	0.845	0.811
Recall	1	0.476	0.566	0.648	0.4	0.49	0.476
F-measure	0.609	0.585	0.664	0.726	0.532	0.62	0.6

ここで、より長い品詞の連なりを使用すればそれだけ文体をとらえやすくなるが、その場合素性の数が爆発的に増えてしまう。そこで、長さを3までに固定した固定長品詞 n-gram を使用する場合と、長さは固定しないが、[10]で提案されているアルゴリズムを使用して、出現するドキュメントの割合が閾値(minsup)以上の可変長品詞 n-gram だけを使用する場合で実験を行った。

また、素性の重み付けとして tf (単語の出現回数) と idf (逆文書頻度、珍しい単語ほど大きくなる関数) を用いる。

なお、非公式 RT の場合、「RT」以前の部分に意見が書かれているため、最後の RT 以前の部分のみを分類に使用する。例えば、「そうなんだ RT @YY: デマだろ RT @XX: ツイッターが情報統制」というツイートがあった場合、「RT @XX: ツイッターが情報統制」より前の部分を使用する。ここで、最初の部分(「そうなんだ」)だけを用いないのは、この例のように、デマだという意見を信じている場合でも、「そうなんだ」だけを取り出して「信用している」と誤って判断しないためである。

4.4. 実験

4.5. データ

Togetter の「デマ」とタグがついたページから合計 1986 ツイートを収集して使用した。対象とした特定の流言に関するまとめページは 35 で、ラベル付けを行ったのは 1986 ツイート、内 761 が「疑い」で、410 が「根拠あり」だった。

4.6. 実験結果

実験ではラベルづけしたデータのうち、「東京電力株主 5 位は石原氏」というトピックに関する流言をテストデータ、それ以外を訓練データとして用いた。

また、SVM 分類器の実装には liblinear⁵を用いた。

SVM を用いて分類器の学習を行う場合、ソフトマージンパラメタ C の値が分類精度に大きく影響することが知られている。そこで、C を 2^{-s} ($s = -4, -2, 0, \dots, 14, 16$) の範囲で動かし、陽性(疑い、根拠あり)に対する F-measure が最も高くなる C を選んだ。また、可変長品詞 n-gram における閾値 minsup に関しても、0.05 から 0.30 まで 0.05 刻みで動かし、陽性に対する F-measure が最も高くなる minsup を選んだ。

比較のため、全ての発言を疑いまたは根拠ありと分類した場合を Baseline とする。

4.6.1. 分類精度

Table 3 と Table 4 に分類実験の結果を示す。文体を表す品詞の n-gram を使用した場合、信疑分類の精度に大きな変化は見られないが、根拠の有無の分類精度は大きく向上している。これは前述したように、根拠を提示している文は共通の単語はあまりないが文体は共通しているためと考えられる。

tfidf に関しては、信疑の分類では精度を向上させたが、根拠の有無の分類では精度がかえって下がってしまった。これは、前述したように同じカテゴリであっても使われている単語に多様性があり、出現割合の少ない単語だからといって重要なわけではないためと考えられる。

4.6.2. 重み付けの大きい素性

Table 5 と Table 6 に、素性の上位 10 件を示す。これらの表から次のことがわかる。まず、信疑分類の「疑い」に関する素性では、「デマ」「本当」「？」など直感的にも関係があると思われる素性が並んでいる。また、「信用」では感情的な表現につきやすい「！」がトップに来ている。一方、「新宿」など特定の話題にしか出現しない素性に対して大きな重み付けが与えられているという問題がある。

根拠の有無の分類の「根拠あり」に関しては、URL が上位に来ている他、固有名詞を含む品詞 n-gram が上位に来ている。これは、根拠を示す際、何らかの特定の事物について述べることが多いからと考えられる。一方で、信疑の分類と同様に「コンゴ」のように特定のトピックとしか関連しない単語に大きな重みが与え

⁵ <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>

られてしまっているものも多い。また、「根拠なし」の方はなぜそれが大きな重みが与えられているかわかりにくい物が多い。

Table 5 信疑の分類についての重みの大きい素性の一覧

	疑い		信用	
1	デマ	7.08	!	-2.68
2	ソース	3.96	中国	-2.61
3	本当	3	統制	-2.59
4	という	2.74	…	-2.57
5	確認	2.64	ね	-2.35
6	ガセ	2.5	新宿	-2.18
7	嘘	2.35	情報統制	-2.15
8	記事	2.06	事	-2.13
9	.	2.01	歴史	-2.1
10	ほんとに	1.99	日本	-2.04

Table 6 根拠の有無の分類についての重みの大きい素性の一覧

	根拠あり		根拠なし	
1	URL	3.32	名詞_	-2.01
2	固有名詞_格助詞_自立動詞	2.37	一般記号	-1.95
3	代名詞_一般名詞_	1.93	RT	-1.67
4	山	1.85	空白記号	-1.66
5	固有名詞	1.80	サ変接続名詞_空白記号	-1.63
6	富士山	1.77	自立動詞_句点記号	-1.55
7	コンゴ	1.74	助詞_	-1.47
8	括弧閉記号_空白記号	1.70	係助詞_自立動詞_接続助詞_非自立動詞	-1.47
9	固有名詞_格助詞	1.60	ていない	-1.45
10	ソース無い	1.54	形容動詞語幹名詞_助動詞	-1.44

4.6.3. マージンとテストデータの分布

Figure 5 Figure 6 に、SVM を用いて分類を行ったとき各テストサンプルの分離超平面からの距離の分布を示す。信疑タグ付きサンプル、根拠タグ付きサンプルともに、陽性のほうがピークが右よりになっており、スコアの大きさが実際のラベルと連動していることがわかる。

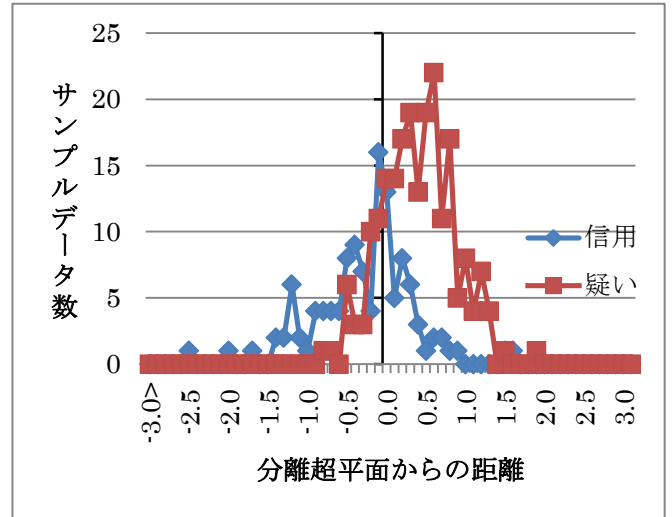


Figure 5 マージンとテストデータの分布(信疑タグ付きデータ)

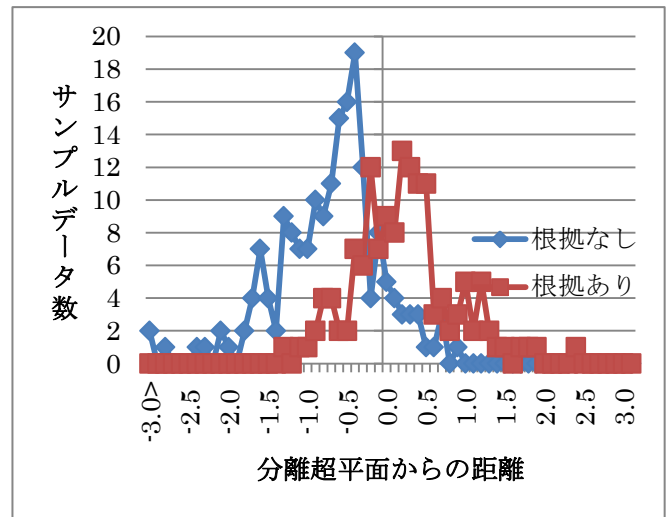


Figure 6 マージンとテストデータの分布(根拠の有無タグ付きデータ)

5. 関連研究

ツイッター上の流言の検出については、先行研究が存在する [8]。しかし、この研究では、流言に関するツイートかどうかの判定を行う分類器を話題ごとに作成する必要がある。また、疑っているユーザの割合を知ることができるが、それらのユーザがどのような理由で疑っているかまでは知ることができないため、個々のユーザが流言かどうかを判断するには不十分である。

話題解析の研究に最初に取り組んだものとしては、確率オートマトンを用いた [11]が存在するが、この手法は計算的なコストが高く、また、パラメータの設定が必要である。 [2]では、確率分布を仮定し、確率の低い事象が発生するほど話題になっているとみなす手法が提案されており、本研究もこれに従った。

なお、話題抽出の研究では「何が話題になっている

か」を客観的に定義することが難しいためか、評価実験は行なっていないものも多い [2] [12] [13]。

6. まとめ

本研究では TextRank などを用いてバーストしている単語のクラスタを抽出し、話題を解析、さらに、そのスコアと各ツイートに含まれる単語から関連するツイートを収集する手法、そして、ユーザの反応を分類して流言かどうかの判断を手助けするシステムについて提案を行った。

将来的な課題としては、まず、共通の単語が多く含まれる似たような話題に対してはクラスタリングが上手くいかないことがあげられる。例えば「松本復興大臣が知事を恫喝」と「梶川ゆきこ氏が知事を批判」という2つのニュースがあった場合、「知事」という共通の単語が含まれる近い事象の話題のため、2つを同じクラスタとみなしてしまう。

また、似たような単語（「昼食」「昼飯」など）は、同時に使われることがないため、共起度のみを尺度とすると別々のクラスタとみなされてしまう。一方で、時間的な同時性を用いると、同じ時間に始まる番組など、時間的な同時性は高いが別々の話題を同じクラスタとみなしてしまう問題が生じる。このため、意味が似ている単語を自動的に同じクラスタとみなすなどの対応策を取る必要が考えられる。

今後は以上のような問題点の解決、また、実際のシステムの構築を目指したい。

参 考 文 献

- [1] (2011) 震災後のデマ 80 件を分類整理して見えてきたパニック時の社会心理. [Online]. <http://news.livedoor.com/article/detail/5477882/>
- [2] Jeffrey Xu Yu, Philip S. Yu, Hongjun Lu Gabriel Pui Cheong Fung, "Parameter Free Bursty Events Detection in Text Streams," in *VLDB '05 Proceedings of the 31st international conference on Very large data bases*, Trondheim, Norway, 2005.
- [3] Lawrence and Brin, Sergey and Motwani, Rajeev and Winograd, Terry Page, "The PageRank Citation Ranking: Bringing Order to the Web.," Stanford InfoLab., 1999.
- [4] P. Tarau R. Mihalcea, "TextRank: Bringing Order into Texts," in *Proceedings of EMNLP-04*, Barcelonaand Spain, 2004.
- [5] Mark-Anthony Hutton, and Jugal Kalita Beau Sharifi, "Summarizing Microblogs Automatically," in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL*, Los Angeles, California, 2010, pp. 685-688.
- [6] 山本祐輔, "ウェブ情報の信憑性分析に関する研究," Kyoto University, PhD thesis 2011.
- [7] 情報信頼性判断支援システム. [Online]. <http://ici.wisdom-nict.jp/>
- [8] Emily Rosengren, Dragomir R. Radev, Qiaozhu Mei Vahed Qazvinian, "Rumor has it: Identifying Misinformation in Microblogs," in *Conference on Empirical Methods in Natural Language Processing*, Edinburgh, Scotland, UK, July 27-31, 2011, pp. 1589-1599.
- [9] P. Raghavan, and H. Schutze C.D. Manning, *Introduction to Information Retrieval.*: Cambridge University Press, 2008.
- [10] A. Mukherjee and B. Liu, "Improving gender classification of blog authors," in *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, MIT, Massachusetts, USA, 9-11, Oct. 2010, pp. 207-217.
- [11] Jon Kleinberg, "Bursty and hierarchical structure in streams," *Data Mining and Knowledge Discovery*, vol. 7, no. 4, pp. 373-397, October 2003.
- [12] Jinfeng WEI Hong LI, "Netnews Bursty Hot Topic Detection Based ," in *International Conference on E-Business and E-Government*, 2010, pp. 1437-1440.
- [13] Nick Koudas Nilesh Bansal, "Searching the Blogosphere," in *Proceedings of the 10th International Workshop on Web and Databases*, Beijing, China, 2007.
- [14] Member, IEEE, and Shumeet Baluja, Member, IEEE Yushi Jing, "VisualRank: Applying PageRank to Large-Scale Image Search," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 11, no. 30, pp. 1877 - 1890, Nov. 2008.