文脈を考慮した観点に基づく意見ツイートクラスタリング

鷹栖 弘明　　内海 彰
電気通信大学 大学院情報理工学研究科 総合情報学専攻
takasu@utm.inf.uec.ac.jp, utsumi@inf.uec.ac.jp

1 はじめに

Web 上には，様々な製品やサービスに対する評判や時事問題に対する意見が存在している．近年では，マイクロブログサービスの普及により，個人が簡単に意見を述べられるようになったため，膨大な量の意見が存在している．このような膨大な意見をセンチメントや観点などに基づいて分類することは，意見の比較や整理が容易にできることに繋がる．本研究では，マイクロブログサービスの中でも Twitter を取り扱う．

Twitter を対象とした意見や評判の自動分類の研究として，橋本ら [1] は詳細な評判傾向の抽出を目的として，意見文中に用いられる感情表現の違いから意見文をその感情ごとに分類している．また，Jiang ら [2] は，ユーザの関連ツイートを考慮して意見文を肯定/否定/中立に分類している．

しかし，時事問題などに対する意見を述べたツイートは，肯定/否定や感情とは別に意見の観点で分類することもできる．例えば,「原発」というトピックに関する意見には「安全性」や「エネルギー」,「健康」といった様々な観点の意見が混在している．意見をこのような観点に基づいて分類することで，観点ごとに意見を把握・比較することができ，新たな観点の意見を発見する手がかりにもなる．

Twitter を含め，Web 上の意見を観点ごとに分類する研究は，ほとんど行われていない．ブログ記事を対象とした横本ら [3] の研究では，トピックを表す話題語を含む Wikipedia 記事集合を取得し，それらの記事タイトルの中で分類対象のブログ記事に多く出現したものを分類に用いる観点として設定しているが，Wikipedia の記事タイトルにない観点は設定できず，適切に分類することができない可能性がある．また，Twitter を対象として，意見が述べられたツイート（以下，意見ツイート）を観点ごとにクラスタリングする研究が鷹栖ら [4] によって行われている．この研究では，意見ツイートだけでは類似度を計算するのに十分な情報がないことから，意見に関連するツイート集合（以下，関連ツイート集合）を考慮してクラスタリングを行っている．しかし，鷹栖らの手法では観点に基づいたクラスタリングを目指しているものの，観点の性質を活かしきれていないという問題点がある．また，クラスタリングにおいて，意見ツイートとその関連ツイート集合をまとめて 1 つの意見ツイートとして意見ツイート間の類似度を計算しているが，関連ツイート集合の情報に引きずられて誤ったクラスタリングがされてしまう問題点もある．

そこで本研究では，予め観点を設定することなく，Twitter 上の意見ツイートを観点ごとに分類することを目的とする．観点ごとに意見を分類するために意見ツイート間の類似度を計算して文書クラスタリングを行うが，意見ツイートとその関連ツイート集合に含まれる情報を別々に扱い，意見ツイート間の類似度を計算する．また，一般的な類似度計算には BoW(Bag of Words) が用いられるが，共通語を多く含む意見どうしが同じ観点を示すとは限らない．そこで本研究では，意見ツイート間の類似度を適切に計算するために，意見ツイートに関連するユーザのツイートや，名詞と動詞の係り受け関係といった文脈情報を考慮して意見ツイートを観点ごとにクラスタリングする手法を提案する．

2 提案手法

提案手法では，ある特定のトピック（時事問題）に関する意見ツイートの集合について，1 つの意見ツイートに単一の観点が付与されると仮定し，排他的なクラスタリングを行う．

また，本研究では，意見を観点ごとに適切に分類するにあたり，名詞と動詞の係り受け関係が有用であると仮定する．例えば,「**燃料から作れる**」「**燃料を消費する**」という 2 つの文では,「**燃料**」という共通の名詞が存在するが，それぞれ「**作れる**」と「**消費する**」という動詞に係ることから，それぞれ「発電技術」「発電コスト」といった異なる観点を示す文であると推測される．そこで，文節の係り受け関係から名詞と動詞のペア（名詞・動詞ペア）を抽出し，名詞どうしの類似度に加え，動詞どうしの類似度を計算することで観点の差異を考慮できると考えられる．

以上の考えのもと，本研究で提案するクラスタリング手法の手順を以下に示す．

1. あるユーザ x_i がつぶやいた意見ツイート o_i の周りに存在するユーザ x_i のツイート集合 A_i から，意見に関連するツイート集合 R_i を抽出する．
2. 各意見ツイート o_i および，各関連ツイート集合 R_i への係り受け解析で得られた文節の係り受け関係から，それぞれ名詞・動詞ペアを抽出する．
3. 各意見ツイート o_i とその関連ツイート集合 R_i をそれぞれに含まれる名詞・動詞ペア集合 $\boldsymbol{P}(o_i), \boldsymbol{P}(R_i)$ で表現し，意見ツイート間の類似度を名詞・動詞ペア集合間の類似度として計算する．
4. 手順 3. で計算した意見ツイート間の類似度を用いて，Ward 法による階層型クラスタリングを行う．

2.1 関連ツイートの抽出

Twitter では，文字数の制限や投稿のしやすさから，あるトピックに対するツイートが連続して投稿されることがある．つまり，複数回の投稿により 1 つの意見が構成される場合があるため，意見に関連する複数の（投稿）ツイートを抽出し，クラスタリングに利用する．

関連ツイートの抽出は，鷹栖ら [4] と同様に行う．形態素単位の意味空間と文字 bigram 単位の意味空間，ツ

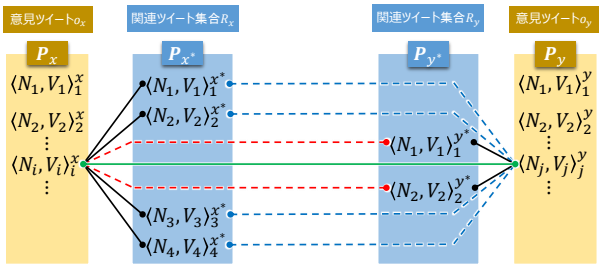

図 1: 名詞・動詞ペア間の関係図

イートの投稿時間から意見ツイート o_i とその周りのツイート $a_{ij} \in A_i$ との類似度を計算し，類似度が閾値を超えた a_{ij} を関連ツイートとして抽出する．

2.2 名詞・動詞ペアの抽出

ツイートに対して係り受け解析を行い，得られた文節の係り受け関係から名詞 N とそれが係る動詞 V のペア（名詞・動詞ペア）$\langle N, V\rangle$ を抽出する．例えば，「**自然の脅威に備える**」という文に対しては，$\langle$ **脅威**, **備える** $\rangle$ というペアが抽出される．また，抽出したペアの名詞 N に係る文節（修飾節）に自立語 W が含まれる場合は，原則として N と W を 1 つの名詞（複合名詞）として抽出する．先述の例文では，$\langle$ **自然**, **脅威** $\rangle$ を複合名詞とみなし，最終的に $\langle\langle$ **自然**, **脅威** $\rangle$, **備える** $\rangle$ という名詞・動詞ペアが抽出される．

なお，ツイートは文字数が少なく，文法が不完全という特徴があるため，ツイートによっては係り受け解析が正しくできず，名詞・動詞ペアを抽出できない可能性がある．そのため，1 ツイート中で名詞・動詞ペアを抽出できなかった場合は，動詞との係り受け関係を考慮せずに名詞のみを抽出し，2.5 節で述べる名詞・動詞ペア間の類似度計算に用いる．このような場合，動詞の情報を利用することができないため，動詞どうしの類似度は 0 とする．

2.3 意見ツイートどうしの類似度計算

意見ツイート o_x, o_y から抽出した名詞・動詞ペアの集合をそれぞれ $\boldsymbol{P}_x = \{\langle N_i, V_i\rangle_i^x\}, \boldsymbol{P}_y = \{\langle N_j, V_j\rangle_j^y\}$ とし，o_x, o_y の関連ツイート集合 R_x, R_y から抽出した名詞・動詞ペアの集合をそれぞれ $\boldsymbol{P}_{x^*} = \{\langle N_k, V_k\rangle_k^{x^*}\}, \boldsymbol{P}_{y^*} = \{\langle N_l, V_l\rangle_l^{y^*}\}$ とする．

今，関連ツイート集合 R_x は意見ツイート o_x に関連するものであることから，$\langle N_k, V_k\rangle_k^{x^*}$ は o_x が示す観点を特徴づける材料であると仮定すると，$\langle N_j, V_j\rangle_j^y$ と $\langle N_k, V_k\rangle_k^{x^*}$ の類似度が高いとき，$\langle N_k, V_k\rangle_k^{x^*}$ は意見ツイート o_y が示す観点を特徴づける材料でもあると言える．つまり，2 つの意見ツイートが示す観点の材料が同じであれば，その意見ツイートどうしは観点が似ていると考えることができる．

以上のことを示す意見ツイートおよび関連ツイート集合の名詞・動詞ペア間の関係を図 1 に示す．なお，図 1 では例として $|\boldsymbol{P}_{x^*}| = 4, |\boldsymbol{P}_{y^*}| = 2$ としている．

図 1 の点線・実線は名詞・動詞ペア間の類似度を示している．赤と青の点線は，片方の意見ツイートに含まれる $\langle N, V\rangle$ と他方の意見ツイートの関連ツイート集合に含まれる $\langle N, V\rangle$ との類似度を示しており，この類似度が高くなるほど 2 つの意見ツイートが示す観点の材料が同じだと言える．なお，黒の実線は，関連ツイート集合に含まれる名詞・動詞ペアが意見ツイートの示す観点を特徴づけるものであることを指す．意見ツイート o_x, o_y から抽出した任意の名詞・動詞ペア $\langle N_i, V_i\rangle_i^x, \langle N_j, V_j\rangle_j^y$ 間の最終的な類似度は，赤の点線における最大類似度と青の点線における最大類似度，緑の実線で示される類似度の平均とする．

以上のことから，意見ツイートどうしの類似度 $\mathrm{sim}_o(o_x, o_y)$ を式 (1) のように定義する．

$$\mathrm{sim}_o(o_x, o_y) = \frac{\mathrm{nvSim}_x + \mathrm{nvSim}_y}{|\boldsymbol{P}_x| + |\boldsymbol{P}_y|} \tag{1}$$

$$\mathrm{nvSim}_x = \sum_{i=1}^{|\boldsymbol{P}_x|} \max_j \left[M\left(\langle N_i, V_i\rangle_i^x, \langle N_j, V_j\rangle_j^y\right)\right]$$

$$\mathrm{nvSim}_y = \sum_{j=1}^{|\boldsymbol{P}_y|} \max_i \left[M\left(\langle N_i, V_i\rangle_i^x, \langle N_j, V_j\rangle_j^y\right)\right]$$

nvSim_x は意見ツイート o_x の各名詞・動詞ペア $\langle N_i, V_i\rangle_i^x$ に対する意見ツイート o_y の $\boldsymbol{P}_y$ との最大類似度の和である．nvSim_y は逆に，意見ツイート o_y の各名詞・動詞ペア $\langle N_j, V_j\rangle_j^y$ に対する意見ツイート o_x の $\boldsymbol{P}_x$ との最大類似度の和である．なお，意見ツイート間の名詞・動詞ペアどうしの類似度は式 (2) を満たす関数 M により計算される．

$$M\left(\langle N_i, V_i\rangle_i^x, \langle N_j, V_j\rangle_j^y\right) = \frac{m_0 + m_1 + m_2}{1 + f(\boldsymbol{P}_{x^*}) + f(\boldsymbol{P}_{y^*})} \tag{2}$$

$$m_0 = \mathrm{sim}_{nv}\left(\langle N_i, V_i\rangle_i^x, \langle N_j, V_j\rangle_j^y\right)$$

$$m_1 = \begin{cases} \max_l \left[\mathrm{sim}_{nv}\left(\langle N_i, V_i\rangle_i^x, \langle N_l, V_l\rangle_l^{y^*}\right)\right] & (\boldsymbol{P}_{y^*} \neq \varnothing) \\ 0 & (otherwise) \end{cases}$$

$$m_2 = \begin{cases} \max_k \left[\mathrm{sim}_{nv}\left(\langle N_j, V_j\rangle_j^y, \langle N_k, V_k\rangle_k^{x^*}\right)\right] & (\boldsymbol{P}_{x^*} \neq \varnothing) \\ 0 & (otherwise) \end{cases}$$

$$f(\boldsymbol{P}) = \begin{cases} 0 & (\boldsymbol{P} = \varnothing) \\ 1 & (otherwise) \end{cases}$$

m_0 は緑の実線が示す類似度を，m_1, m_2 はそれぞれ赤の点線における最大類似度と青の点線における最大類似度を表している．関数 M は m_0, m_1, m_2 の平均を返すが，関連ツイート集合が空の場合は，m_1, m_2 を平均から除くようにしている．

以降の節では，名詞・動詞ペアどうしの類似度 sim_{nv} の計算方法について述べる．また，sim_{nv} は名詞どうし，動詞どうしの類似度を用いて計算するため，まず 2.4 節で単語どうしの類似度計算方法について述べたのち，2.5 節で名詞・動詞ペアどうしの類似度計算方法について述べる．

2.4 単語どうしの類似度計算

単語 w_i, w_j 間の日本語 WordNet を用いた類似度 jwn_w[6] と潜在意味インデキシング（以下，LSI）[5] により構築した意味空間を用いた類似度 lsi_w から，w_i と w_j の類似度 sim_w を式 (3) のように定義する．

$$\mathrm{sim}_w(w_i, w_j) = \mu \times \mathrm{jwn}_w + (1 - \mu) \times \mathrm{lsi}_w \tag{3}$$

μ $(0 \leq \mu \leq 1)$ は，どちらの類似度の影響を強くするかを示すパラメータであり，その値が大きいほど日本語 WordNet を用いた類似度を重視することになる．ただ

し，w_i, w_j のどちらかが日本語 WordNet に存在しない場合は，$\mu = 0$ とする．

2.5 名詞・動詞ペアどうしの類似度

名詞・動詞ペア $\langle N_i, V_i \rangle_i, \langle N_j, V_j \rangle_j$ 間の類似度 sim_{nv} を，名詞 N_i, N_j 間の類似度 sim_n と動詞 V_i, V_j 間の類似度 sim_v を用いて式 (4) のように定義する．

$$\begin{aligned}&\mathrm{sim}_{nv}\left(\langle N_i, V_i \rangle_i, \langle N_j, V_j \rangle_j\right)\\&\quad = \mathrm{sim}_n + \left((1-\lambda) + \lambda(\mathrm{sim}_n)^2\right) \times \mathrm{sim}_v \qquad (4)\end{aligned}$$

式 (4) は，sim_n と係数付きの sim_v の和を取る形になっている．sim_{nv} の計算式を式 (4) とした理由は，sim_n が小さければ sim_v の大小に関わらず 2 つの名詞・動詞ペア $\langle N, V \rangle$ が異なる内容を表す可能性が高く，sim_{nv} を小さくする必要があると考えたからである．そのため，sim_n が大きくなるほど sim_v が sim_{nv} に与える影響が大きくなるように sim_v の係数が設定されている．パラメータ λ はその影響度合いを示しており，本研究では $\lambda = 2/3$ とした．

名詞 N_i, N_j 間の類似度 sim_n は，N_i, N_j がそれぞれ単一の名詞である場合と修飾語を含む複合名詞である場合とで計算方法が異なる．

■ **N_i と N_j が単一名詞の場合：**
2.4 節で定義した式 (3) で計算する．

■ **N_i と N_j の片方のみが複合名詞の場合：**
N_i が複合名詞（$N_i = \langle N_{i,1}, N_{i,2} \rangle$）のとき，式 (5) のように N_i に含まれる修飾語 $N_{i,1}$ と N_j 間，被修飾語 $N_{i,2}$ と N_j 間の類似度を式 (3) で計算し，パラメータ ω $(0 \leq \omega \leq 1)$ を用いて和を取る．

$$\begin{aligned}\mathrm{sim}_n(N_i, N_j) &= \omega \times \mathrm{sim}_w(N_{i,1}, N_j)\\&\quad + (1-\omega) \times \mathrm{sim}_w(N_{i,2}, N_j) \qquad (5)\end{aligned}$$

■ **N_i と N_j が複合名詞の場合：**
複合名詞 $N_i = \langle N_{i,1}, N_{i,2} \rangle$ と $N_j = \langle N_{j,1}, N_{j,2} \rangle$ に対して，式 (6) のように，両複合名詞に含まれる修飾語どうし，被修飾語どうしの類似度を式 (3) で計算し，式 (5) と同じパラメータ ω を用いて和を取る．

$$\begin{aligned}\mathrm{sim}_n(N_i, N_j) &= \omega \times \mathrm{sim}_w(N_{i,1}, N_{j,1})\\&\quad + (1-\omega) \times \mathrm{sim}_w(N_{i,2}, N_{j,2}) \qquad (6)\end{aligned}$$

なお，式 (5), (6) に共通するパラメータ ω は，修飾語に基づく類似度が全体の類似度に与える影響の度合いを示しており，その値が大きいほど修飾語に基づく類似度の影響が強くなる．

動詞 V_i, V_j 間の類似度 sim_v は，式 (3) で計算する．

3 評価実験

評価実験には，鷹栖ら [4] の研究で用いた「原発」「衆議院選挙」「尖閣諸島」の 3 つのトピックに関する意見ツイートを利用した．また，各トピックにおいて 2 名ずつ個別に観点ごとに意見ツイートを分類してもらい，各トピックごとに 2 種類のクラスタリングの正解データを用意した．本研究の評価実験では，人手による分類と同じ観点の数で提案手法によるクラスタリングを行った．

提案手法により生成されたクラスタ群と人手により生成された正解クラスタ群（正解データ）がどの程度近いかの指標として F 値を用いた．鷹栖らと同様に，提案手法により生成されたクラスタ群 $\boldsymbol{S}$ と人手により生成された正解クラスタ群 $\boldsymbol{L}$ に含まれるクラスタ間で計算した重み付き平均 F 値が最大となるクラスタの組み合わせを決め，そのときの重みを除いた F 値を用いた．

3.1 比較手法

比較手法には，鷹栖ら [4] の手法と LSI 法を用意した．

鷹栖らの手法では，LSI により構築した形態素単位の意味空間を利用して，各意見ツイートおよび各関連ツイートの特徴ベクトルを生成する．最終的には意見ツイートとその関連ツイート集合の特徴ベクトルの重心ベクトルを意見ツイートの特徴ベクトルとして意見ツイート間のコサイン類似度を計算し，Ward 法による階層型クラスタリングを行う．意味空間はすべての意見ツイート O と関連ツイート集合全体 $\boldsymbol{R}$ に含まれる単語の出現頻度を要素とする単語・文書行列に対して次元圧縮を施して構築する．なお，関連ツイート $r_{ij} \in R_i$ については，行列の要素として，r_{ij} に含まれる単語の出現頻度に r_{ij} と意見ツイート o_i の類似度を掛けた値を用いる．また，鷹栖らと同様に各意見ツイートの関連ツイート集合は，人手で関連ツイートと判定されたものを用いた．

LSI 法は，関連ツイート集合の情報を利用せず，LSI により構築した意味空間から生成した意見ツイートのみの特徴ベクトルを用いて意見ツイート間のコサイン類似度を計算し，Ward 法による階層型クラスタリングを行う方法である．意味空間は各意見ツイートに含まれる単語の出現頻度を要素とする単語・文書行列に対して次元圧縮を施して構築した．

3.2 実験結果

関連ツイート集合については，鷹栖ら [4] における関連ツイート抽出の評価実験において，抽出性能が最も高くなったときの関連ツイート集合をそのまま用いた．

提案手法については，パラメータ μ, ω をそれぞれ 0 〜1 の 0.1 刻みで変化させてクラスタリングを行った．また，提案手法・比較手法ともに LSI により構築した意味空間の次元数を 5 から全意見ツイート数（ただし，鷹栖らの手法では意見ツイートと関連ツイートの総数）までの 5 刻みで変化させた．

なお，各トピックにおいて F 値は正解クラスタ群ごとに計算するが，先述したように提案手法や比較手法ではパラメータや意味空間の次元数を変化させてクラスタリングを行っていることから，交差検定を用いて F 値を計算した．各トピックごとに，1 つの正解データで学習したパラメータを用いて，もう 1 つの正解データに対してクラスタリングを行い，得られた 2 つの F 値のマクロ平均をそのトピックの F 値とした．

以上の実験結果を図 2 の棒グラフで示す．すべてのトピックにおいて提案手法で最も高い F 値が得られたことから，本研究の提案手法は観点に基づくクラスタリング手法として有用であると言える．

4 考察

4.1 関連ツイートと名詞・動詞ペアの有用性

提案手法において，関連ツイートや名詞・動詞ペアの利用が精度にどのような影響を与えるかを調べるために，以下の 3 条件におけるクラスタリング性能を求め，評価実験で得られた提案手法の性能と比較した．

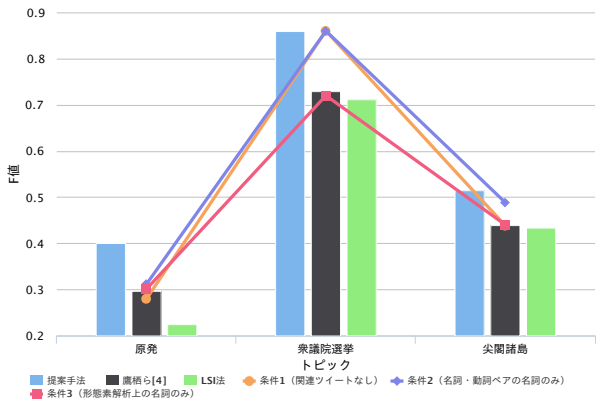

図 2: クラスタリングの評価実験結果

条件 1 関連ツイートの情報を利用せずに類似度を計算する．(式 (2) において m_0 のみを計算する．)

条件 2 名詞・動詞ペアの名詞のみを利用して類似度を計算する．(式 (4) において $\mathrm{sim}_v = 0$ とする．)

条件 3 条件 2 に加えて，名詞・動詞ペアの（複合）名詞を，動詞との係り受け関係を考慮せずに形態素解析上で名詞と解析されたすべての語をもとにして抽出する．

特に，条件 2 における性能を求めることで，動詞どうしの類似度が精度に与える影響を調べることができる．また，条件 3 で抽出された（複合）名詞には文中で動詞の役割を担う語（「開発する」の「開発」などのサ変名詞 [1]）が含まれることから，条件 2 と性能を比較することで，動詞との係り受け関係を考慮した名詞の抽出が精度に与える影響を調べることができる．

以上の各条件におけるクラスタリング性能を図 2 の折れ線グラフで示す．

図中の提案手法（青い棒グラフ）と条件 1 を比較すると，トピック「衆議院選挙」では同じ精度となったが，他のトピックでは提案手法で最も良い精度が得られたことから，関連ツイートの情報を利用することは有用であると言える．提案手法と条件 2 の比較においても，トピック「衆議院選挙」では同じ精度となったが，他のトピックでは提案手法で最も良い精度が得られたことから，動詞どうしの類似度を考慮することで，より観点に基づいたクラスタリングができることが示された．また，条件 2 と条件 3 を比較すると，すべてのトピックにおいて条件 2 の方が精度が良いことから，動詞との係り受け関係を用いて抽出した名詞をクラスタリングに利用することは有用であると言える．

トピック「衆議院選挙」では，提案手法と条件 1,2 で同じ精度となり，比較手法においても他のトピックに比べて高い精度が得られた．このことから，「衆議院選挙」では，クラスタリングするには意見ツイートに含まれる情報（名詞・動詞ペア）だけで十分であり，意見ツイートやその関連ツイート集合に含まれる名詞が直接意見の観点を示すことが多いという特徴があると考えられる．

トピック「衆議院選挙」では，ツイートに含まれる名詞が直接意見の観点を示すことが多いという特徴があるものの，提案手法と条件 3 を比較すると，提案手法の方が精度が良いことから，動詞との係り受け関係を用いて抽出した名詞の情報，つまり間接的に動詞の情報がクラスタリングに有用であると言える．

[1] 名詞に動詞「する」が付属することで動詞化するもの

4.2 エラー分析

正しくクラスタリングができなかった主な原因として，意見とは関係のない関連ツイート集合が抽出されたことが挙げられる．関連ツイート集合については，鷹栖ら [4] における関連ツイート抽出の評価実験において，抽出性能が最も高くなったときの関連ツイート集合をそのまま用いているため，意見とは関係のない関連ツイートが少なからず含まれていた可能性がある．つまり，意見とは関係のない名詞・動詞ペアが類似度計算に利用されてしまったために，異なる観点の意見どうしで類似度が高くなってしまったと考えられる．

また，1 つの意見に異なる観点を示す名詞・動詞ペア $\langle N, V \rangle$ が含まれていたために，誤ったクラスタに意見が属してしまったことも原因の 1 つだと考えられる．例えば，ある意見に「安全性」という観点を示す $\langle N, V \rangle$ と「政治」という観点を示す $\langle N, V \rangle$ が含まれているとき，正解データではその意見が「安全性」という観点を示すクラスタに属していたとしても，排他的なクラスタリングを行うと「政治」という観点を示すクラスタに属してしまうことがある．そのため，非排他的なクラスタリングに適するような類似度の計算方法を考案する必要がある．

5 おわりに

本研究では，Twitter 上に存在する意見ツイートに対して，ユーザの意見ツイートに関連するツイートと文節の係り受け関係から抽出した名詞・動詞ペアという文脈情報を用いて，観点に基づいて意見ツイートをクラスタリングする手法を提案した．評価実験の結果より，従来のクラスタリング手法に比べて高い精度が得られたことから，提案手法の有用性を確認することができた．

今後の課題としては，関連ツイート集合の抽出方法の改善や，関連ツイート集合を用いた意見ツイートどうしの類似度計算方法のさらなる改善が挙げられる．また，本研究ではクラスタリングしか行っておらず，得られたクラスタがどのような観点を示しているのか分かりづらいことから，クラスタへのラベリング手法の考案も今後の課題である．

参考文献

[1] 橋本 和幸，中川 博之，田原 康之，大須賀 昭彦：センチメント分析とトピック抽出によるマイクロブログからの評判傾向抽出, 電子情報通信学会論文誌-D, J94-D(11), 1762-1772 (2011).

[2] L.Jiang, M.Yu, M.Zhou, X.Liu and T.Zhao: Target-dependent twitter sentiment classification, *Proc. of ACL2011*, 151-160 (2011).

[3] 横本 大輔，林 東權，牧田 健作，宇津呂 武仁，河田 容英，福原 知宏，神門 典子，吉岡 真治，中川 裕志，清田 陽司：特定トピックに関するブログ記事集合の観点分類における Wikipedia の利用, 第 3 回データ工学と情報マネジメントに関するフォーラム論文集, A4-3 (2011).

[4] 鷹栖 弘明，小林 聡，内海 彰：Twitter における観点に基づいた意見文クラスタリング, 言語処理学会 第 19 回年次大会論文集, A4-3 (2013).

[5] S.Deerwester et al.: Indexing by latent semantic analysis, *Journal of the American Society for Information Science*, 41, 6, 391-407 (1990).

[6] P.Resnik: Using information content to evaluate semantic similarity in a taxonomy, *Proc. IJCAI1995*, 1, 448-453 (1995).