Vol.2018-NL-237 No.6
2018/9/26

Wikipedia から獲得した外部知識を用いた賛否分類

塙 一晃[1,a)]　佐々木 彬[2]　岡崎 直観[3]　乾 健太郎[1,4]

概要：本研究では賛否分類においてトピックに関する外部知識を利用するための手法を提案する．Wikipedia 記事に紐づいた 7 トピックに関する 6,701 件のツイートからなるデータセットを作成し，分析することで賛否分類における外部知識の必要性が明らかとなった．また，本研究では Wikipedia 記事から獲得した知識を賛否分類で利用するために，関連する知識を参照しながら文をエンコードすることができるモデルを提案する．Wikipedia から獲得した知識を使用する提案手法は外部知識を使用しないものよりも高い精度で賛否の予測ができることが実験結果より明らかとなった．

Hanawa Kazuaki[1,a)]　Sasaki Akira[2]　Okazaki Naoaki[3]　Inui Kentaro[1,4]

1. はじめに

賛否分類とはある文とトピックが与えられた時に，その文の書き手がトピックに対して賛成（肯定的）か反対（否定的）か，またはどちらでもないかを予測するタスクである．[1]．このタスクはオンラインディベート分析 [2], [3], [4] や意見クラスタリング [5], [6], [7]，選挙結果予測 [8], [9]，フェイクニュースの検出 [10] などの様々な応用分野における中核となるタスクである．

賛否分類ではトピックとそれに関連する “もの” や “こと” についての外部知識を必要とする事例が少なくない [11], [12], [13]．例えば，環太平洋パートナーシップ協定 *(TPP)* というトピックに対する意見を述べている “自由貿易を取り入れるべきだ” という一文を考える．この文には *TPP* というトピックそのものを表す語が含まれないため，計算機にとって賛否を予測することは簡単でない．しかし，人間は TPP と自由貿易の関係 (*TPP* が自由貿易を促進する) を知ってさえいればこの文が TPP に対して肯定的だということを予測することができる．3.1 節で述べるように，このようなトピックそのものを含まずにトピックと関連のある何かに対して意見を述べている事例が数多く存在している．

しかし，外部知識を賛否分類に取り込むことは非自明で困難な問題である．Sasaki ら [11] は文中のトピックと関連のある表現に対してアノテーションをするという手法に取り組んだ．Boltuzic and Snajder [12] はドメイン特有の表現間の関係を表す *microstructures* という枠組みを提案した．これらの研究では人手のアノテーションをテキストに対して行うことで賛否分類における精度向上が見られることを示している．Bar-Haim ら [13] は 整合的 (類義語など) または対照的 (対義語など) の関係にある名詞句ペアを外部知識として賛否分類に利用した．しかし，彼らの実験では通常の尺度における精度の向上は見られなかった．

本研究では賛否分類を二つの部分問題に切り分ける (図 1)：

i. トピックについての Wikipedia 記事をの読むことで，トピックと関連する “もの” や “こと” についての知識を得る (知識獲得)；

ii. 得られた外部知識を参照しながら与えられた文のトピックに対する賛否を予測する．

本研究の貢献は以下の四つである．

(1) トピックが Wikipedia 記事と紐づいているような新たな賛否分類のデータセットを作成した[*1]．このデータセットは 7 つのトピック，6,701 件のツイートから成り，これを分析することで三分の一以上のツイートにおいて外部知識が必要であることがわかった (3.1 節)．

(2) トピックに関連する知識を獲得するために Wikipedia

[1] 東北大学
[2] 株式会社リクルートテクノロジーズ
[3] 東京工業大学
[4] 理研 AIP
[a)] hanawa@ecei.tohoku.ac.jp

[*1] http://www.cl.ecei.tohoku.ac.jp/index.php?OpenResources

ⓒ 2018 Information Processing Society of Japan

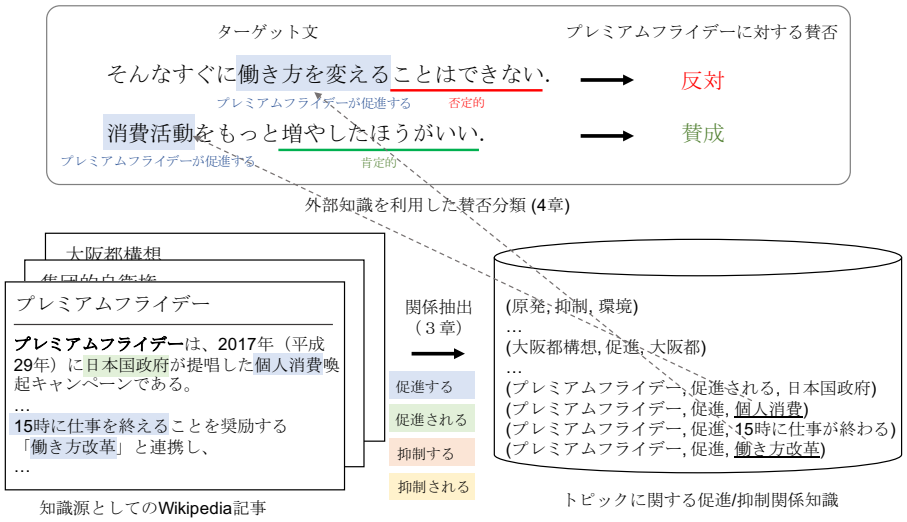


図 1 Wikipedia 記事から獲得された知識を利用した賛否分類.

記事に対してアノテーションを行なった[*1](3 節).

(3) 関連する知識を参照しながら文をエンコードする手法の提案. (4 節).

(4) 実験によって獲得された知識を利用することで賛否分類の F 値が約 3 ポイント向上することがわかった (5 節).

2. 賛否分類データセットの構築

2.1 Twitter における意見分析

我々の究極的な目標はデータジャーナリズム [14] や自動記事作成 [15] である. 特に, 世間で話題となっている事柄に関するソーシャルリスニングや SNS 上の意見分析を行うことの需要は高い. そのため, 本研究では Twitter で活発に議論されているトピックを研究対象とする. また, トピックに関する知識を Wikipedia 記事から獲得することを想定しているため, 対象のトピックに関する Wikipedia 記事が存在することも仮定する. この節では, 活発に議論されており Wikipedia 記事と紐づけることができるようなトピックを含む賛否分類データセットを構築する. このデータセットは Wikipedia が賛否分類のための知識源としてどれだけ有用かを評価するためのデータとしても有用である.

2.2 トピックの選出とツイートの収集

2015 年 4 月から 2017 年 6 月までででクローリングした約 260 億件のツイート対象のデータとして使用した. まず, 週ごとのハッシュタグの TF*IDF を計算した. ここで TF はある週の中であるハッシュタグが出現した回数であり, DF はあるハッシュタグが出現した週の数である. この手順によって週ごとの集中的に利用されているハッシュタグのリストが得られる. このハッシュタグのリストから, 活発に議論されており, かつ対応する Wikipedia 記事が存在するようなトピックを人手で選出した. 選出されたトピックは “TPP”, “プレミアムフライデー”, “共謀罪”, “原発”, “大阪都構想”, “安保法案”, “集団的自衛権” の 7 つである. それぞれのトピックについて対応するハッシュタグを含むツイートの中からランダムに 2,000 ツイートを抽出した. また, 収集されたツイートから全てのハッシュタグを削除した. このようにして 7 つのトピックに対して合計 14,000 件のツイートが得られた.

トピック	賛成	反対	どちらでもない
TPP	53	802	230
プレミアムフライデー	153	744	218
共謀罪	86	592	308
原発	47	783	202
大阪都構想	239	259	380
安保法案	168	352	262
集団的自衛権	160	468	195
合計	906	4,000	1,795

表 1 賛否分類データセット中の賛否ラベルの分布.

2.3 ツイートに対する賛否のラベリング

収集したツイートから賛否分類のためのデータセットを作成するために, 各ツイートが対応するトピックに対して賛成, 反対, どちらでもない, のいずれであるかをクラウドワーカに訊ねた. 1 ツイートに対して 5 人のワーカによるラベリングを行なった後, 4 人以上が同じラベルを付与したものだけを残し, それ以外は除外した. 表 1 はデータセット中のトピックごとの賛否ラベルの数を示している.

2.4 トピックに関する知識の必要性

このデータセット上での賛否分類におけるトピックに関する知識を考慮することの影響はどれだけあるのだろうか. 影響の大きさを推測するため賛成または反対のラベル

必要な知識	%	例文
なし (トピックそのものがツイート中に出てくる)	56.3	原発 は絶対に必要！ (トピック: 原発)
促進/抑制関係 (Wikipedia に記述あり)	26.3	関税が機能するべきだと思うんだけど (トピック: *TPP*)
促進/抑制関係 (Wikipedia に記述なし)	13.9	遺伝子組み換え食品とか今後大丈夫なのかな (トピック: *TPP*)
その他の関係	2.5	いやはや治安維持法を復活させたいのか (topic: 安保法案)

表 2 促進/抑制関係に着目した，賛否分類に必要な知識.

が付与されたツイートのうち 491 件 (10%) のツイートをランダムに抽出し，文中で直接意見を述べられている対象がトピックどのような関係にあるのか人手で分類した．

ここで，我々はトピックと文中で直接意見を述べられている対象との間の関係として促進/抑制関係 [16], [17], [18] に着目した.[*2]．“A が B を促進する” とは A が活性化した時に B も活性化するような関係であり，“A が B を抑制する” とは A が活性化した時に B が不活性化するような関係である．

表 2 に分析を示す．1 行目は 56.3%のツイートはトピックそのものを表す表現 (“原発” など) を含んでいることを示している．このような場合はトピックに対する感情分析を行えば十分である．表 2 中の例では，この意見は “*X* は絶対に必要” という肯定的な意見を述べるときの表現パターンでを含んでおり，*X* = “原発” であることから賛成の意見であることが予測できる．

しかし，40.2% (= 26.3% + 13.9%) のツイートにおいて促進/抑制関係の知識が必要であることがわかった．2 行目,3 行目の例ではトピック *TPP* と例文の間には “*TPP* が関税を抑制する” や “*TPP* が遺伝子組み換え食品を促進する” という関係が裏に存在している．更に分析を進めると Wikipedia で述べられているような促進/抑制関係を必要としているツイートは全体の 26.3%であることがわかった．このことから，Wikipedia 記事から促進/抑制関係を抽出してくることは賛否分類の精度向上に寄与することが十分に期待できると考えられる．

3. Wikipedia からの促進/抑制関係知識の獲得

3.1 Wikipedia 記事への促進/抑制関係知識のアノテーション

節の分析に基づき，我々は促進/抑制関係は賛否分類における重要な知識であるという仮定を置き，トピックについての Wikipedia 記事を読むことで促進/抑制関係の知識を得ることを目指した．Hanawa ら [18] に倣い，Wikipedia 記事の文中のどの区間の表現が記事のタイトルと促進/抑制関係にあるのかを予測する系列ラベリング問題として定式化した．言い換えれば，関係の片方の項を記事のタイトルに固定し，もう片方の項が文中のどの区間で，以下の 4 つのうちどの関係であるかを予測するような手法である．

- PRO: “[*title*] が B を促進する”
- SUP: “[*title*] が B を抑制する”
- PROBY: “A が [*title*] を促進する”
- SUPBY: “A が [*title*] を抑制する”

ここで A と B は文中のある区間を示す変数であり [*title*] は記事のタイトルを示している．

Hanawa ら [18] は 1,494 件の Wikipedia 記事の概要文について促進/抑制関係がアノテーションされたデータを公開しているが，我々は今回これに加えて作成した賛否分類データセット中の 7 つのトピックに対応する Wikipedia 記事の全文に対して新たにアノテーションした．促進/抑制関係のアノテーションはクラウドソーシングを用いて，PRO, SUP, PROBY, SUPBY を記事中の適切な区間に対してラベリングする作業として定式化した．Yahoo!クラウドソーシング[*3]を用いて 1 記事あたり 10 人のワーカによってアノテーションを行なった．10 人中 2 以上が全く同じ区間に同じラベルを付与したもののみを採用した．表 3 は文数とアノテーションされた区間の数を関係ごとに示している．

3.2 Wikipedia 記事からの促進/抑制関係の自動獲得

PRO, SUP, PROBY, SUPBY の 4 つのラベルがアノテーションされたデータは Wikipedia から関係知識を抽出するための教師データとして利用することができる．我々はこのタスクを IOB2 記法を用いた系列ラベリング問題として定式化した．より詳細には，このタスクでは与えられた単語の系列に対してラベル (B-PRO, I-PRO, B-SUP, I-SUP など) の系列を予測する．

我々はこの系列ラベリング問題を BiLSTM-CRF[19] を用いてモデル化した．単語埋め込みと隠れ層の次元はいずれも 300 次元であり，単語埋め込みは Wikipedia 記事で事前に学習されたもの[*4]を用いて初期化した．モデルを学習する際には Hanawa ら [18] によって公開されているデータ

[*2] 促進/抑制関係は整合的/対照的 [13] と概ね同じものである．Boltuzic and Snajder [12] は促進/抑制を含むより詳細な 8 つの関係を使用した．ある主張の賛否を予測することを考えるのであれば，Boltuzic and Snajder が使用している 8 つの関係は促進/抑制に集約することができる．例えば，A と B が同一であるという関係を A が B を促進するという関係に落とし込んだとしても賛否を予測する上では問題ない．

[*3] http://crowdsourcing.yahoo.co.jp/

[*4] https://github.com/overlast/word-vector-web-api

	TPP	プレミアムフライデー	共謀罪	原発	大阪都構想	安保法案	集団的自衛権
文数	333	6	179	257	169	121	39
Pro	257	17	122	190	165	120	42
Sup	67	2	163	74	115	46	25
ProBy	131	7	77	108	64	45	21
SupBy	145	3	86	96	51	30	6

表 3 各記事の文数と関係ごとのアノテーション数

と 3.1 節で作成したデータの両方を合わせて利用した．

予測された IOB2 ラベルはその予測確率が閾値 α を超えた場合にのみ採用し，いずれのラベルも α を超えなかった場合には O ラベルを採用した．このようにして各事例が Wikipedia 記事のタイトル $\mathfrak{t}$，関係タイプ $\mathfrak{r}$，Wikipedia 記事中の表現 $\mathfrak{m}$ から成る三つ組であるような知識ベース (KB)$\mathcal{D}$ を構築した．

$$(\mathfrak{t}, \mathfrak{r}, \mathfrak{m}) \in \mathcal{D}. \tag{1}$$

KB $\mathcal{D}$ には全ての Wikipedia 記事から獲得された関係事例が含まれている．

4. 賛否分類モデル

この節では 3 節で獲得された促進/抑制関係を利用する賛否分類のためのモデルを提案する．トピック z と N 語からなる文 $s = w_1, w_2, \cdots, w_N$ が与えられた時，モデルは 3 つの賛否ラベルの確率分布 $y \in \mathbb{R}^3$ を計算する．ここで，y の各次元はそれぞれ s が賛成，反対，どちらでもない である確率を表している．

4.1 ベースラインモデル (KB 不使用)

ベースラインモデルでは文 s から BiLSTM と最大値プーリングを用いてベクトル h を計算する．より詳細には，まず単語 w_t に対する埋め込み $x_t \in \mathbb{R}^{d_w}$ と ELMo [20] $ELMo_t \in \mathbb{R}^{d_e}$ を得る．ここで d_w, d_e はそれぞれ単語埋め込みと ELMo の次元数を示している．これらを結合することで w_t の表現 $x'_t = x_t \oplus ELMo_t$ を得る．ここで，$\oplus$ ベクトルの結合操作を表す演算子である．2 層の BiLSTM を用いて $x'_1, \cdots, x'_N$ の順方向でエンコードした $\overrightarrow{h}_1, \cdots, \overrightarrow{h}_N$ と逆方向でエンコードした $\overleftarrow{h}_1, \cdots, \overleftarrow{h}_N$ を得る．

$$\overrightarrow{h}_t, \overrightarrow{c}_t = \overrightarrow{\text{LSTM}}(x'_t, \overrightarrow{h}_{t-1}, \overrightarrow{c}_{t-1}), \tag{2}$$
$$\overleftarrow{h}_t, \overleftarrow{c}_t = \overleftarrow{\text{LSTM}}(x'_t, \overleftarrow{h}_{t+1}, \overleftarrow{c}_{t+1}). \tag{3}$$

ここで，$\overrightarrow{h}_t, \overrightarrow{c}_t \in \mathbb{R}^{d_h}$ $(t = 1, \cdots, N)$ は順方向 LSTM($\overrightarrow{\text{LSTM}}$) の隠れ状態とメモリセルであり，$\overleftarrow{h}_t, \overleftarrow{c}_t \in \mathbb{R}^{d_h}$ $(t = 1, \cdots, N)$ は逆方向 LSTM($\overleftarrow{\text{LSTM}}$) の隠れ状態とメモリセルである．また，$d_h$ はベクトル $\overrightarrow{h}_t, \overrightarrow{c}_t, \overleftarrow{h}_t, \overleftarrow{c}_t$ の次元数である．

$[\overrightarrow{h}_N; \overleftarrow{h}_1]$ など $\overrightarrow{h}_t$ と $\overleftarrow{h}_t$ から h を構成する方法はいくつか考えられるが開発セット上で最も高い精度を出した最大値プーリングを利用することにした．よって，$h \in \mathbb{R}^{2d_h}$ を最大値プーリングを $[\overrightarrow{h}_t; \overleftarrow{h}_t]$ $(t = 1, \cdots, N)$ に対して適用することで得られる文ベクトルとする．

最後に文ベクトル h から 3 つのラベルに対する確率分布を計算する．

$$y = \text{softmax}(W \cdot h + b). \tag{4}$$

ここで，$W \in \mathbb{R}^{3 \times 2d_h}$ と $b \in \mathbb{R}^3$ はそれぞれ重み行列とバイアスベクトルである

4.2 完全一致に基づく KB の参照

ベースラインモデルに外部知識を取り入れる素朴な方法として，KB $\mathcal{D}$ 中にあるトピック z といずれかの関係にあるような文 s 中の適当な区間に対して自動でアノテーションする方法が考えられる (図 2)．より具体的には，文 s 中のある区間と $\mathcal{D}$ 中のある $\mathfrak{m}$ の（最長の）完全一致をみることで $\mathfrak{m}$ に対応する $\mathfrak{r}$ をアノテーションする．式 5 によって $\mathcal{D}$ からタイトル $\mathfrak{t}$ がトピック z であるような三つ組の集合を取り出す．

$$D_z = \{(\mathfrak{m}, \mathfrak{r}) \mid (\mathfrak{t}, \mathfrak{r}, \mathfrak{m}) \in \mathcal{D} \wedge \mathfrak{t} = z\} \tag{5}$$

p_t を単語 w_t が D_z 中のどの関係を持つ事例と関連づいたか（つまり z とどの関係にあるか）を示す変数として定義する．p_t の値は Pro, Sup, ProBy, SupBy, None のいずれかである．ここで，None は単語 w_t が D_z 中のどの事例とも関連づかないことを示すための特殊な関係である．例として “自由貿易を受け入れるべき” という文の TPP というトピックに対する賛否を予測することを考える．つまり， $z = TPP$, $s =$ (自由, 貿易, を, 受け入れる, べき) である．また知識ベースとして $D_z = \{$(自由貿易, Pro)$\}$ を考える．この時，p_1, p_2, p_3 の値は None であり，p_4, p_5 の値は Pro となる．

w_t の z との関係の埋め込みと x'_t を結合することで新たな w_t の表現 x''_t を得る．

$$x''_t = x'_t \oplus \text{emb}(p_t). \tag{6}$$

ここで，$\text{emb}(p) : p \longmapsto \mathbb{R}^{d_r}$ は与えられた関係 p に対する埋め込みベクトルを返すような関数であり，d_r は関係埋め込みの次元数である．このようにして得られた x''_t を 式 2 と式 3 中で x'_t の代わりに用いることで，ベールラインモデルに関係知識の情報を取り入れる．

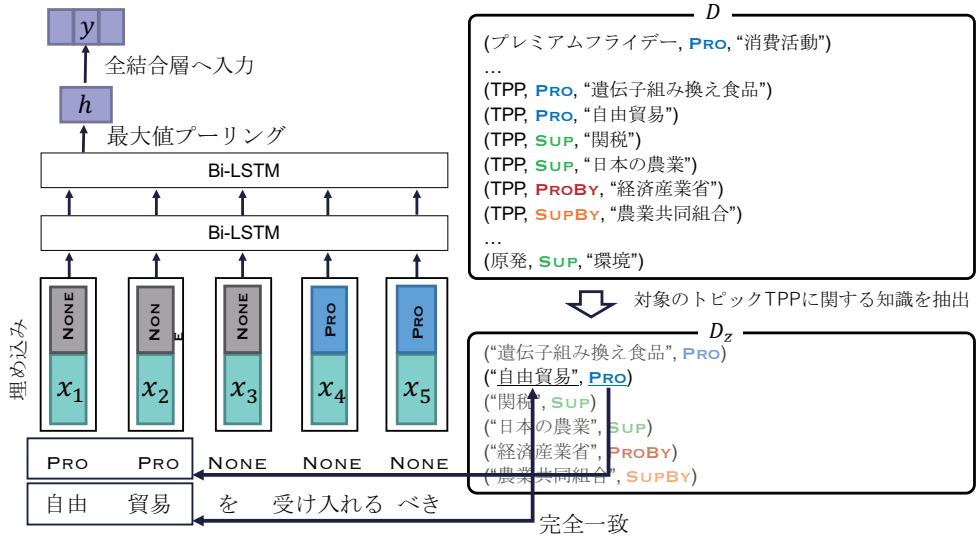


図 2　完全一致による促進/抑制関係の賛否分類での利用.

4.3　アテンションに基づく KB の参照

KB $\mathcal{D}$ 中の表現と全く同じ表現が文 s 中で使用されるとは限らない．そこで，我々はより柔軟な KB の参照をを行うことができるニューラルネットワークモデルを提案する．図 3 に本モデルの概要を示す．本提案モデルでは入力文の各時刻 t において，パラメータが共有された LSTM によってそれぞれエンコードされた w_t と $\mathfrak{m}_i$ の隠れ状態を比較することで KB 中の i 番目の事例に対するアテンションスコアを計算する．

まず，式 5 によって与えられたトピック z に対する知識源 D_z を得る．ここで，$\mathfrak{m}_i$ と $\mathfrak{r}_i$ はそれぞれ D_z 中の i 番目の事例 $(\mathfrak{m}_i, \mathfrak{r}_i)$ を表す $(i \in \{1, ..., |D_z|\})$．$D_z$ 中の事例はキーが $\mathfrak{m}$ で値が $\mathfrak{r}$ であるような key-value 機構として保存されている．キー $(\mathfrak{m})$ を単語系列として 1 層の BiLSTM を用いてエンコードする．例えば (自由貿易, PRO) という事例に対して (自由, 貿易) を BiLSTM に入力することでこの事例に対するキーのエンコードを得る．ここで，$\overrightarrow{v}_i \in \mathbb{R}^{d_h}$ と $\overleftarrow{v}_i \in \mathbb{R}^{d_h}$ は D_z 中の $\mathfrak{m}_i$ をそれぞれ順方向，逆方向でエンコードした時の最終時刻の隠れ状態とする．

入力文の w_t に対する $\overrightarrow{h}_t$ と $\overleftarrow{h}_t$ も $\overrightarrow{v}_i$ と $\overleftarrow{v}_i$ をエンコードする際に用いたものとパラメータを共有している LSTM を用いて計算する．その後，アテンションスコア $\overrightarrow{a}_{t,0}, \overrightarrow{a}_{t,1}, \cdots, \overrightarrow{a}_{t,|D_z|}$ を計算する．ここで，$\overrightarrow{a}_{t,0}$ は単語 w_t が D_z 中のどの事例とも適合しなかったことを表すスコアであり，$\overrightarrow{a}_{t,i}$ $(i \in \{1, ..., |D_z|\})$ は w_t が D_z 中の i 番目の事例とどれだけ適合したかを表すスコアである．

$\overrightarrow{a}_{t,i}$ を順方向の隠れ状態の $\overrightarrow{h}_t$ と $\overrightarrow{v}_i$ から計算されるアテンションスコアとする．

$$\overrightarrow{a}_{t,i} = \frac{\exp(\mathrm{sim}(\overrightarrow{h}_t, \overrightarrow{v}_i))}{\sum_{i'=0}^{|D_z|} \exp(\mathrm{sim}(\overrightarrow{h}_t, \overrightarrow{v}_{i'}))} \tag{7}$$

ここで，関数 sim は以下のような関数である．

$$\mathrm{sim}(\overrightarrow{h}_t, \overrightarrow{v}_i) = \begin{cases} \overrightarrow{h}_t \cdot \overrightarrow{v}_i & (\text{if } 0 < i) \\ \kappa & (\text{if } i = 0) \end{cases}, \tag{8}$$

すなわち，$i < 0$ の時は二つのベクトルの内積を計算し，$i = 0$ の時は定数 κ(ハイパーパラメータ) を返す．これによって w_t が D_z のどの事例とも適合しない場合には相対的に見て κ の値が大きくなるため $\overrightarrow{a}_{t,0}$ の値が大きくなる．

r_i は $i \in \{1, ..., |D_z|\}$ の時に i 番目の事例 $(\mathfrak{m}_i, \mathfrak{r}_i) \in D_z$ の $\mathfrak{r}_i$ を表し，それ以外の時 $(i = 0)$ に NONE を表すものとする．w_t に対する関係埋め込みを以下のように計算する．

$$\overrightarrow{q}_t = \sum_{i=0}^{|D_z|} \overrightarrow{a}_{t,i} \cdot \mathrm{emb}(r_i). \tag{9}$$

逆方向の場合の関係埋め込み $\overleftarrow{q}_t$ も順方向の場合と同様の方法で計算する．最後にベースラインモデルの 2 層目の LSTM への入力 $[\overrightarrow{h}_t; \overleftarrow{h}_t]$ の代わりに $[\overrightarrow{h}_t; \overleftarrow{h}_t; \overrightarrow{q}_t; \overleftarrow{q}_t]$ を用いることで関係埋め込みをベースラインモデルに取り入れる．

5.　実験

5.1　設定

外部知識 (促進/抑制関係) がどれだけ寄与したかを評価するためにトピックごとの 7 分割交差検証を行なった．5 トピックを訓練セット，1 トピックを開発セット，1 トピックをテストセットとして用いた．Wikipedia からの関係知識獲得モデルを訓練する際には対象のトピックと開発セットに使われているトピックについてのデータは訓練データから取り除かれる．この設定ではモデルは常に未知のトピックについての予測をすることになるため難しい設定である．

各ハイパーパラメータはそれぞれ，$d_w = d_e = d_h = 300$,

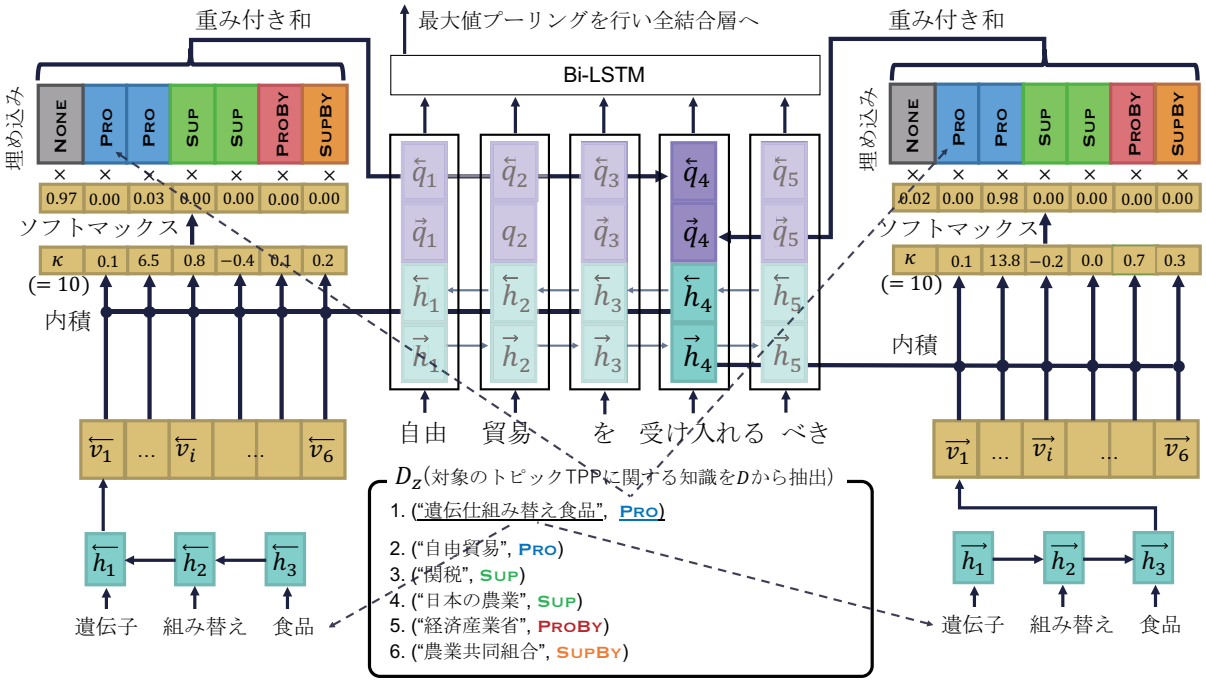


図 3　アテンションによる促進/抑制関係の賛否分類での利用.

$d_r = 100$, $\kappa = 10$, $\alpha = 0.85$ と設定した．単語埋め込みは Wikipedia 記事を用いて訓練されたもの[*6]を用いて初期化した．ELMo は Wikipedia 記事を用いて訓練された 3 層の BiLM から計算した．モデル中のパラメータ (LSTM，全結合層，単語埋め込み，関係埋め込み) は Adam を用いて最適化した．Mohammad ら [1] に倣って賛成と反対ラベルの F 値の平均である $F_{avg} = (F_{favor} + F_{against})/2$ を評価尺度として用いた．また，別の初期値によって訓練された 10 個のモデルによる多数決アンサンブルを行った場合の結果についても実験を行なった．予測されたラベルの数が最大のものが複数個存在した場合には訓練データ中でより多くのラベルが存在した順で優先して採用した (どの場合でも反対，どちらでもない，賛成の順番で優先される)．

5.2　結果

表 4 に賛否分類の結果を示す．まず，アンサンブルを行なった場合は単体のモデルの場合と比較して常に性能が高いことがわかる．全体の F 値は F_{avg} のトピック間のマイクロ平均を計算している．ベースラインモデルは全体の F 値が 0.478 であった．ベースラインの素朴な拡張 (表中のトピック) はトピックと全く同じ表現を PRO，PROBY 関係にあると考えた時の設定である．例えばトピック z が TPP の場合，$D_z = \{(TPP, \text{PRO}), (TPP, \text{PROBY})\}$ である．使用する知識がトピックの設定の場合，精度の向上は見られなかった: 0.475 (完全一致)，0.476 (アテンション)．

また，どのような知識を利用するかについて自動 と 人手の 2 つの設定で実験を行なった．自動 の設定では 3.2 節で述べた手法によって自動で獲得した知識を用いている．言い換えれば，この設定は計算機が Wikipedia 記事を読み，そこからトピックに関する知識を獲得し，賛否を予測するという一連の手順を行うことに対応している．人手 の設定ではクラウドワーカによって Wikipedia 記事にアノテーションされた知識をそのまま使用している．この設定では本提案手法で Wikipedia 記事からの知識獲得が完璧に行えた場合の賛否分類の性能の上限を調べることができる．

最も高い性能を出したのはアテンションに基づく KB の参照を行うモデルでゴールドの促進/抑制関係知識が与えられた時であった (0.507)．自動獲得した知識を用いた場合には，F 値は 0.491 まで下がるが知識を使わない場合と比べて高い性能を出すことができている．これらの結果はトピックに関する知識を利用することは賛否分類の精度向上に寄与することを示している．

また，この表からアテンションに基づく参照を行うモデルの方が基本的に完全一致で参照を行うモデルよりも性能が高いことがわかる．図 3.2, 3.2 に示すアテンションスコアの可視化から，アテンションに基づく手法は全く同じではないフレーズに対しても正しく参照できていることがわかる．

表 5 に促進/抑制関係の自動獲得の精度を示す．ここでは人手のアノテーションに対する文字単位の Precision(p) と Recall(r) を示している．開発セット上での探索によって比較的高い閾値 ($\alpha = 0.85$) が選ばれたため，少数の信頼度の高い関係知識が予測される傾向にある．そのため，Precision は比較的高く，反対に Recall は低い結果となった．

モデル	知識	アンサンブル	TPP	プレミアムフライデー	共謀罪	原発	大阪都構想	安保法案	集団的自衛権	全体
Majority baseline			0.375	**0.400**	0.431	0.425	0.228	0.310	0.363	0.374
ベースライン			0.490	0.258	0.489	0.430	0.332	0.533	0.470	0.466
		あり	0.498	0.267	0.490	0.439	0.348	0.543	0.488	0.478
完全一致	トピック		0.488	0.257	0.502	0.431	0.331	0.528	0.475	0.470
		あり	0.499	0.260	0.502	0.448	0.352	0.549	0.494	0.475
	自動		0.484	0.255	0.488	0.447	0.359	0.544	0.477	0.475
		あり	0.489	0.258	0.514	0.448	0.360	0.561	0.482	0.486
	人手		0.501	0.264	0.525	0.444	0.368	0.548	0.477	0.483
		あり	0.516	0.262	0.529	0.452	**0.369**	0.566	0.489	0.490
アテンション	トピック		0.491	0.250	0.488	0.425	0.347	0.539	0.475	0.468
		あり	0.489	0.251	0.495	0.436	0.352	0.546	0.479	0.476
	自動		0.491	0.252	0.508	0.430	0.352	0.564	0.481	0.473
		あり	0.509	0.263	0.527	0.442	0.360	0.570	0.499	0.491
	人手		0.500	0.249	0.515	0.448	0.361	0.566	0.486	0.490
		あり	**0.523**	0.258	**0.539**	**0.465**	0.362	**0.582**	**0.500**	**0.507**

表 4　トピックごとの各設定での賛否分類の精度 (F 値)

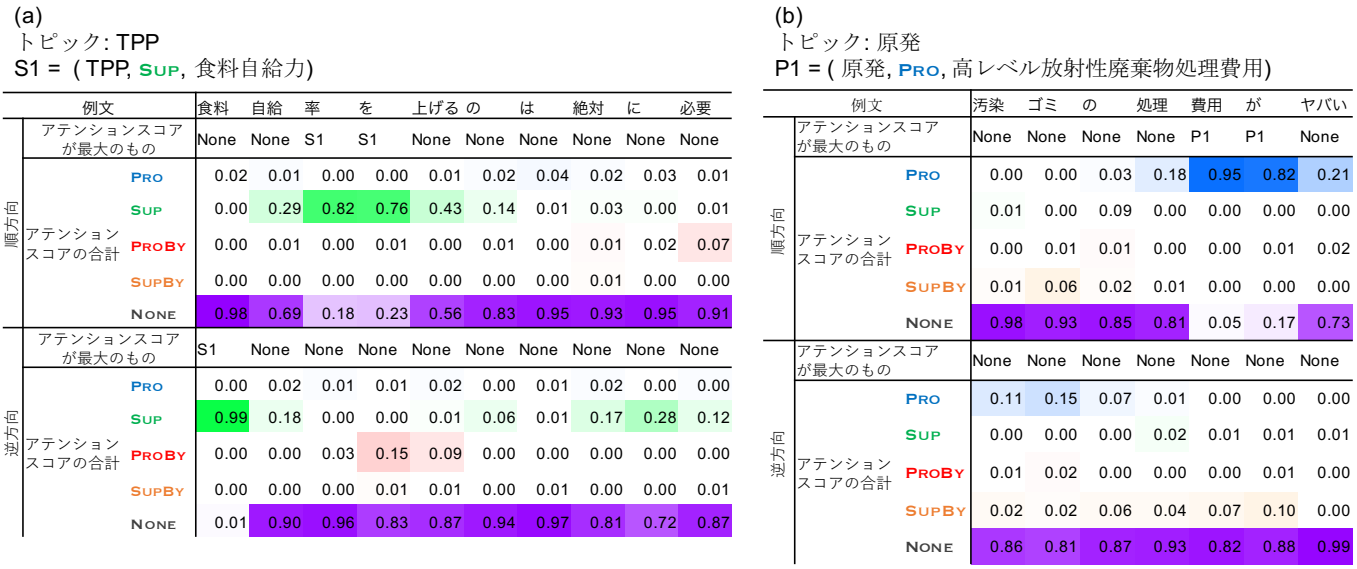


図 4　アテンションスコアの可視化

6. 関連研究

近年，多くの賛否分類に関する多くの研究が行われてきた． [1], [2]．しかし，特に未知のトピックに対する賛否分類については未だ多くの問題が残っている．既知のトピックについての賛否分類は比較的高い性能を出すことが可能になってきたが，未知の (訓練データ中に出現しない) トピックになると性能は大きく落ちる．[21], [22], [23].

未知のトピックに関する賛否分類を行なっている先行研究は大きく 2 つに分類することができる．一つ目は何かしらの手がかり (SNS でのやユーザのプロフィルール [24] やハッシュタグ [25]) をもとに未知のトピックに関する擬似的な訓練データを作成する手法である．しかし，これらの手法にはいくつかの欠点が存在する．これらは大量のラベル付けされていないデータが必要になるため，あまり活発に議論されていないトピックに対して応用するのは困難である．また，これらの手法ではあるトピックやある SNS に特有の手がかりを利用しているため，あらゆるトピックやあらゆる SNS に拡張するのは困難である．

もう一つは外部知識を用いる方法である．先行研究では様々な種類のトピックに関する関係知識 (言い換え [10]，対比 [26]，アスペクト [27], 含意 [28], and 因果 [11] など) を考慮している．Boltuzic and Snajder [12] は *microstructures* という枠組みで意見を分析するために促進/抑制関係を含む 8 つの関係を定義した．粒度の細かい関係を使用するのはもっともらしい方法であるが，これらの関係を識別するのは困難である．本研究では粒度の細かい関係を促進/抑制関係に落とし込むことで，トピックとそれに関連する物事の関係の極性だけを考慮し賛否分類に利用している．

Bar-Haim ら [13] は整合的/対照的という関係を用いるという，本研究と近い手法をとっている．しかし彼らはこの外部知識を獲得する際に Web の検索クエリや Wikipedia のセクション名から “A vs B” や “A and B” などの言語

	TPP		プレミアムフライデー		共謀罪		原発		大阪都構想		安保法案		集団的自衛権	
	p	r	p	r	p	r	p	r	p	r	p	r	p	r
Pro	0.41	0.12	0.48	0.37	0.26	0.27	0.29	0.30	0.38	0.15	0.48	0.18	0.43	0.10
Sup	0.67	0.09	1.00	1.00	0.58	0.06	0.34	0.18	1.00	0.09	1.00	0.15	0.59	0.29
ProBy	0.18	0.17	0.38	0.29	0.14	0.18	0.23	0.10	0.19	0.13	0.16	0.13	0.50	0.06
SupBy	0.32	0.08	0.0	0.0	0.29	0.15	0.19	0.05	0.25	0.18	0.30	0.08	0.85	0.10

表 5 促進/抑制関係の自動獲得の文字単位での Precision (p) と Recall (r).

パターンを用いて獲得している．この手法は競合関係にあるもの同士 (*TPP* と *NAFTA* など) の獲得には向いているが，因果関係 (*TPP* と 関税 など) にあるものの獲得には向いていない．また，彼らは予測の信頼度が高い一部のテストデータだけを使用した設定において，性能の向上を示したものの，テストセット全体を使用した設定では性能の向上は見られなかった．

7. おわりに

本研究では賛否分類においてトピックに関する外部知識を利用する新たな手法を提案した．7 トピックに対する 6,701 件のツイートから成る賛否分類のためのデータセットを作成した．データセットを分析することで 40.2%のツイートにおいて賛否の予測のためにはトピックに関する促進/抑制関係の知識が必要であることがわかった．また，Wikipedia 記事に対して促進/抑制関係のアノテーションを行なった．さらに，入力文に応じて得られた知識の参照を行いながら賛否分類を行うニューラルネットワークモデルを提案した．実験結果から促進/抑制関係知識を利用することによって賛否の予測精度が改善することがわかった (自動獲得した知識の場合と人手アノテーションした知識の場合でそれぞれ F 値で 1.4 ポイントと 2.9 ポイント向上した)．

今後はより多くの知識を集めるために知識源として Wikipedia 記事だけでなく，ニュース記事や SNS に拡張することを検討している．また，今回は知識獲得と賛否分類を別の 2 つのタスクとして切り分けたが，これらを end-to-end で行うようなモデルの設計も検討している．

謝辞

本研究は JST CREST(課題番号: JPMJCR1301) の支援を受けて行った．

参考文献

[1] Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X. and Cherry, C.: Semeval-2016 task 6: Detecting stance in tweets, *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval)*, pp. 31–41 (2016).

[2] Thomas, M., Pang, B. and Lee, L.: Get out the vote: Determining support or opposition from Congressional floor-debate transcripts, *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 327–335 (2006).

[3] Murakami, A. and Raymond, R.: Support or oppose?: classifying positions in online debates from reply activities and opinion expressions, *Proceedings of the 23rd International Conference on Computational Linguistics (COLING)*, pp. 869–875 (2010).

[4] Somasundaran, S. and Wiebe, J.: Recognizing stances in ideological on-line debates, *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pp. 116–124 (2010).

[5] Abu-Jbara, A., Dasigi, P., Diab, M. and Radev, D.: Subgroup Detection in Ideological Discussions, *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 399–409 (2012).

[6] Qiu, M., Yang, L. and Jiang, J.: Modeling interaction features for debate side clustering, *Proceedings of the 22nd ACM international conference on Conference on information and knowledge management (CIKM)*, pp. 873–878 (2013).

[7] Hasan, K. S. and Ng, V.: Extra-Linguistic Constraints on Stance Recognition in Ideological Debates, *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 816–821 (2013).

[8] Kim, S.-M. and Hovy, E.: Crystal: Analyzing predictive opinions on the web, *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 1056–1064 (2007).

[9] Bermingham, A. and Smeaton, A.: On using Twitter to monitor political sentiment and predict election results, *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP)*, pp. 2–10 (2011).

[10] Ferreira, W. and Vlachos, A.: Emergent: a novel data-set for stance classification, *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pp. 1163–1168 (2016).

[11] Sasaki, A., Mizuno, J., Okazaki, N. and Inui, K.: Stance Classification by Recognizing Related Events about Targets, *Proceedings of the 2017 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pp. 582–587 (2016).

[12] Boltuzic, F. and Šnajder, J.: Toward Stance Classification Based on Claim Microstructures, *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA)*, pp. 74–80 (2017).

[13] Bar-Haim, R., Bhattacharya, I., Dinuzzo, F., Saha, A. and Slonim, N.: Stance Classification of Context-Dependent Claims, *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pp. 251–261 (2017).

[14] Gray, J., Chambers, L. and Bounegru, L.: *The Data*

ⓒ 2018 Information Processing Society of Japan

Journalism Handbook, O'Reilly Media (2012).
[15] Graefe, A.: Guide to Automated Journalism, Technical report, The Tow Center for Digital Journalism (2016).
[16] Hashimoto, C., Torisawa, K., De Saeger, S., Oh, J.-H. and Kazama, J.: Excitatory or inhibitory: a new semantic orientation extracts contradiction and causality from the web, *Proceedings of the 2012 Conference on Empirical Methods on Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 619–630 (2012).
[17] Fluck, J., Madan, S., Ellendorff, T. R., Mevissen, T., Clematide, S., van der Lek, A. and Rinaldi, F.: Track 4 Overview: Extraction of Causal Network Information in Biological Expression Language (BEL), *Proceedings of the Fifth BioCreative Challenge Evaluation Workshop*, pp. 333–346 (2015).
[18] Hanawa, K., Sasaki, A., Okazaki, N. and Inui, K.: A Crowdsourcing Approach for Annotating Causal Relation Instances in Wikipedia, *Proceedings of the 31st Pacific Asia Conference on Language, Information and Computation (PACLIC)* (2017).
[19] Huang, Z., Xu, W. and Yu, K.: Bidirectional LSTM-CRF models for sequence tagging, *arXiv preprint arXiv:1508.01991* (2015).
[20] Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K. and Zettlemoyer, L.: Deep Contextualized Word Representations, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pp. 2227–2237 (2018).
[21] Anand, P., Walker, M., Abbott, R., Tree, J. E. F., Bowmani, R. and Minor, M.: Cats Rule and Dogs Drool!: Classifying Stance in Online Debate, *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA)*, pp. 1–9 (2011).
[22] Zarrella, G. and Marsh, A.: MITRE at SemEval-2016 Task 6: Transfer Learning for Stance Detection, *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval)*, pp. 458–463 (2016).
[23] Du, J., Xu, R., He, Y. and Gui, L.: Stance classification with target-specific neural attention networks, *Proceedings of the 26th International Joint Conferences on Artificial Intelligence (IJCAI)*, pp. 3988–3994 (2017).
[24] Ebrahimi, J., Dou, D. and Lowd, D.: Weakly supervised tweet stance classification by relational bootstrapping, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1012–1017 (2016).
[25] Wei, W., Zhang, X., Liu, X., Chen, W. and Wang, T.: pkudblab at semeval-2016 task 6: A specific convolutional neural network system for effective stance detection, *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval)*, pp. 384–388 (2016).
[26] Jindal, N. and Liu, B.: Identifying comparative sentences in text documents, *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR)*, pp. 244–251 (2006).
[27] Somasundaran, S. and Wiebe, J.: Recognizing stances in online debates, *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP (ACL-IJCNLP)*, pp. 226–234 (2009).
[28] Cabrio, E. and Villata, S.: A Natural Language Bipolar Argumentation Approach to Support Users in Online Debate Interactions, *Argument and Computation*, Vol. 4, No. 3, pp. 209–230 (2013).

ⓒ 2018 Information Processing Society of Japan