

# 極性辞書を利用した句構造による 注意型ニューラル評価極性分類

宮崎 亮輔<sup>1,a)</sup> 小町 守<sup>1,b)</sup>

**概要:** 評価文は、評価極性（ポジティブ・ネガティブ）を反転させるような表現を含むことがあるが、このような現象を単語のみから捉えることは難しく、この問題を解決するためには統語的構造を考慮する必要がある。構文木を利用した方法はこれまでも考えられているが、単語間の係り受け関係で定義されるスパースな素性を設計する必要がある。素性のスパース性を解消するために再帰的ニューラルネットワーク (Recursive Neural Network) による手法も提案されているが、文単位の評価極性分類で既存の手法と同程度以上の性能を得るためには、各句毎に極性アノテーションされたコーパスが必要である。各句に人手でアノテーションを行うことはコストが高く、現在は英語などでの限られた言語でしかコーパスが存在しない。また極性分類に影響する句や単語は文内でも一部分だけであり、再帰的ニューラルネットワークにおいて木構造上で深い位置にある重要な句や単語の情報に誤差を伝播するのは難しい。そこで、本研究では極性辞書と注意機構を利用した再帰的ニューラルネットワークによるモデルを提案する。提案手法では、極性辞書による部分アノテーションの活用により、文単位の極性アノテーションしか得られない言語においても、利用可能な教師データの拡充を図る。また、各句や単語を重み付きで考慮する注意機構を再帰的ニューラルネットワークの一種である Tree-LSTM へ適用することで、終端ノードに近い場所にある句の情報も効果的に分類に用いることを狙った。実験の結果により、提案手法が日本語の評価極性分類において最高精度を示すことがわかった。

## 1. はじめに

ユーザが投稿するテキストは有用な情報を含むことが多い。例えば、ユーザによるレビュー文やその評価は、他のユーザが製品を比較したり購入の決断をしたりする際に役に立つ。自動で評価極性分類を行うことは、そのようなユーザが投稿するテキストを解析するためにも重要な技術の一つである。

文書分類の標準的なアプローチは単語の共起や bag-of-words などの素性を利用しているが、これらの単純な素性は極性分類のタスクには効果的ではない [21]。例えば、“詐欺を撲滅する。”という文は“詐欺”と“撲滅”というネガティブな評価極性を示す語を含んでいるが、表現全体はポジティブな評価極性と考えられる。このような極性が変わる問題に対処するためには、単語や句間の関係を考慮する必要がある。

Nakagawa ら [18] は依存構造木を利用することでこの問題に取り組んだ。依存構造木の各部分木に隠れ変数として

トレーニングデータには含まれない評価極性を推定する。文全体の極性は Tree-CRF [31] によって各部分木の隠れ変数を考慮して分類される。これによって、このモデルは評価極性分類特有の極性の反転などの特徴を捉えることができる。しかしながら、彼らの手法は素性テンプレートを利用して組み合わせ素性を展開しているため、素性スパースネスの問題を抱えている。

近年、密な素性表現を用いたディープニューラルネットワークによる手法が多く注目を集めている [9, 10, 28–30, 35]。特に、再帰的ニューラルネットワークと呼ばれる木構造を用いたアプローチが極性分類のタスクで高い分類精度を示している [10, 28–30]。Tree-CRF が疎なバイナリ素性を利用している一方で、再帰的ニューラルネットワークは密な実数値の素性表現を学習することで素性スパースネスの問題を解決している。しかし評価極性分類タスクにおいて、再帰的ニューラルネットワークはしばしば句（構文木の各部分木）単位にアノテーションされたコーパスを必要とする。この句単位のアノテーションは、再帰的ニューラルネットワークにおいて“悪くない”などの極性の反転や、“とても良い”などの極性の強調を学習するため重要な情報である。句単位の極性の情報は英語以外

<sup>1</sup> 首都大学東京 システムデザイン研究科

<sup>a)</sup> miyazaki-ryosuke@ed.tmu.ac.jp

<sup>b)</sup> komachi@tmu.ac.jp

の言語で利用可能な資源が未だないため、再帰的ニューラルネットワークのモデルを他の言語に適用することを妨げている。

そこで、本論文では評価極性辞書と句構造に対する注意（アテンション）機構を導入し、部分木の情報を再帰的ニューラルネットワークに直接伝える手法を提案する。アテンション機構はNLP分野でも多くの注目を集めている [1, 3, 4, 12, 13, 24, 33, 34]。例えばニューラル機械翻訳では、アテンション機構を導入し、直接原言語の各単語に対応する隠れ層を参照することで、長い文に対して入力 of 原言語の文全体を1つの隠れベクトルとして表現する既存のエンコーダ・デコーダモデルの難しさを解消した [1]。ニューラル機械翻訳においては、エンコーダ・デコーダのような回帰的ニューラルネットワーク (Recurrent Neural Network) のみならず、再帰的ニューラルネットワークに対してもアテンション機構の研究がなされている [3]。本論文では、この再帰的ニューラルネットワークにアテンション機構を適用したモデルを評価極性分類に利用する。さらに、評価極性辞書を用いることで木構造上終端ノードに近い位置にある句や単語レベルの分散表現の学習に利用可能な教師データの拡充を行う。

我々の提案する方法では、評価極性辞書とアテンション機構の導入により、文単位にアノテーションされたコーパスのみから学習することができる。

本論文の主要な貢献は以下に示すとおりである。

- アテンション機構を適用することで、文単位のアノテーションのみのコーパスからでも再帰的ニューラルネットワークによる評価極性分類を学習でき、Tree-CRFのような素性エンジニアリングが不要であることを示した。
- 再帰的ニューラルネットワークによる評価極性分類で評価極性辞書を効果的に利用できることを示した。
- 日本語の評価極性分類タスクにおいて最高精度を達成した。

## 2. 再帰的ニューラルネットワークによる評価極性分類

統語構造を考慮した文の分散表現を得るために、これまでに様々な再帰的ニューラルネットワークのモデルが研究されている [22, 27–30, 36]。再帰的ニューラルネットワークでは、構成関数と呼ばれる関数によってある句を構成する2つの子からその句の分散表現を構成し、これを再帰的に適用して文の分散表現を得る。Socher ら [28] は最初に単語や句間の相互関係を考慮しないシンプルな再帰的ニューラルネットワークを提案し、後に行列やテンソルによる単語や句間の相互作用を考慮するよう拡張した構成関数を提案し [27, 29]、英語の評価極性分類タスクの精度の向上に貢献した。Qian ら [22] は句カテゴリの情報を構成関数に組

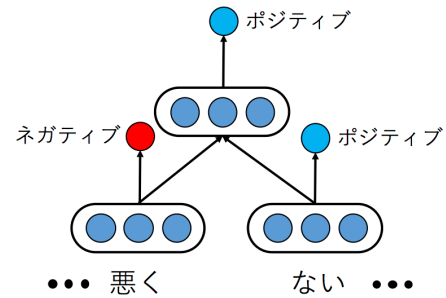


図 1: RNN による句の構成

み入れる方法を提案した。これらの再帰的ニューラルネットワークには回帰的ニューラルネットワークと同様に、長い履歴を考慮できなくなる問題が存在する。Tai ら [30] はこの長い履歴を考慮できなくなる問題を解消するために、LSTM (Long Short-Term Memory) [7] と呼ばれるユニットを再帰的ニューラルネットワークに適用した。

この章では、Socher ら [28] が最初に提案したシンプルな再帰的ニューラルネットワークと、Tai ら [30] が提案した LSTM を再帰的ニューラルネットワークに適用したモデルの2つを説明し、それらを用いてソフトマックス分類器によって評価極性分類を行う方法を述べる。

### 2.1 再帰的ニューラルネットワークによる文の分散表現のモデル化

再帰的ニューラルネットワーク (RNN) [28, 29] は句の表現をその子の表現から構成する。図 1 に示すように、この構成プロセスはより長い句の分散表現を得るために再帰的に繰り返すことができる。文の分散表現は句構造木のような構文木に沿って RNN を適用することで得られる。構成関数を複雑にしないため、二分木化した構文木がしばしば用いられる。構文木内では根ノード、非終端ノード、終端ノードはそれぞれ文、句、単語を表す。

$j$  番目のノードの表現  $h_j$  は構成関数  $g$  によって計算される。

$$h_j = f(g(h_j^l, h_j^r)), \quad (1)$$

$$g(h_j^l, h_j^r) = W \begin{bmatrix} h_j^l \\ h_j^r \end{bmatrix} + b, \quad (2)$$

ここで、行列  $W \in R^{d \times 2d}$  とバイアスベクトル  $b \in R^d$  は学習するパラメータであり、 $h_j^l, h_j^r \in R^d$  は  $d$  次元のベクトルで表されたノード  $h_j$  の左右の子の表現である。活性化関数  $f$  には  $\tanh$  を用いられることが多い。

### 2.2 Tree-LSTM による再帰的ニューラルネットワークの改善

一方、回帰的ニューラルネットワークを学習する過程で、長い系列を入力するとパラメータの勾配が爆発もしくは消失してしまう問題がある [2, 6]。この勾配の消失を防ぐため

に, LSTM と呼ばれる方法が回帰的ニューラルネットワークのために提案された [7].

RNN についても勾配消失問題が同様に生じるため、Tai ら [30,36] によって Tree-LSTM (Tree-Structured Long Short-Term Memory Network) が導入され、RNN における勾配消失の問題を改善した。本論文では、Tai ら [30] によって提案された二分木の Tree-LSTM (Binary Tree-LSTM) を利用する。Tree-LSTM では以下のように子ベクトルを構成する。

$$i_j = \sigma \left( U^{(i)} \begin{bmatrix} h_j^l \\ h_j^r \end{bmatrix} + b^{(i)} \right), \quad (3)$$

$$f_{jl} = \sigma(U^{(fl)}h_j^r + b^{(fl)}), \quad (4)$$

$$f_{jr} = \sigma \left( U^{(fr)} h_j^l + b^{(fr)} \right), \quad (5)$$

$$o_j = \sigma \left( U^{(o)} \begin{bmatrix} h_j^l \\ h_j^r \end{bmatrix} + b^{(o)} \right), \quad (6)$$

$$u_j = \tanh \left( U^{(u)} \begin{bmatrix} h_j^l \\ h_j^r \end{bmatrix} + b^{(u)} \right), \quad (7)$$

$$c_j = i_j \odot u_j + f_{jl} \odot c_{jl} + f_{jr} \odot c_{jr}, \quad (8)$$

$$h_j = o_j \odot \tanh(c_j), \quad (9)$$

行列  $U \in R^{d \times 2d}$  ( $U^{fl}$ ,  $U^{fr}$  においては,  $U \in R^{d \times d}$ ) とバイアスペクトル  $b \in R^d$  は学習するパラメータである.  $\sigma$  はシグモイド関数を,  $\odot$  はアダマール積を表す. 記憶状態  $c$  は必要な情報を必要になるタイミングまで保持しておくために, 入力ゲート  $i$ , 忘却ゲート  $f$ , 出力ゲート  $o$  と呼ばれるゲートによって制御される. 入力ゲートは  $u_j$  のどの次元をどれくらい入力するべきなのか, 忘却ゲートは  $c_{jl}$  と  $c_{jr}$  のどの次元をどれくらい忘れるべきなのか, 出力ゲートは  $\tanh(c_j)$  のどの次元をどれくらい出力すべきなのかをそれぞれ選択する役割を担う.

ここで、式 4,5 に示されている通り、左の子の記憶状態  $c_{jl}$  のための忘却ゲート  $f_{jl}$  は、入力 of 隠れ層ベクトルとしてその逆の右の子からの出力  $h_j^r$  のみを受け取り、その逆もまた同様であることに注意されたい<sup>\*1</sup> [30].

## 2.3 ソフトマックス分類器

構文構造をニューラルネットワークに組み込むために、RNN を用いて句や文の分散表現を計算し、任意の句や文の評価極性ラベル  $\hat{y}$  を予測したい。句構造木を利用する場合は任意の単語・句は対応するノードのベクトル表現で表され、文はルートノードのベクトル表現に対応し、これらのベクトル表現を入力として評価極性を推定する。予測したい対象のノード  $j$  の評価極性ラベル  $\hat{y}_j$  の予測にはソフトマックス分類器を用いる。

ノード  $j$  のベクトル表現  $h_j$  を入力として受け取り, 分

\*1 Zhu ら [36] の手法では忘却ゲートへの入力も、式 3,6 のように左右両方の隠れ層ベクトルを結合して受け取る。

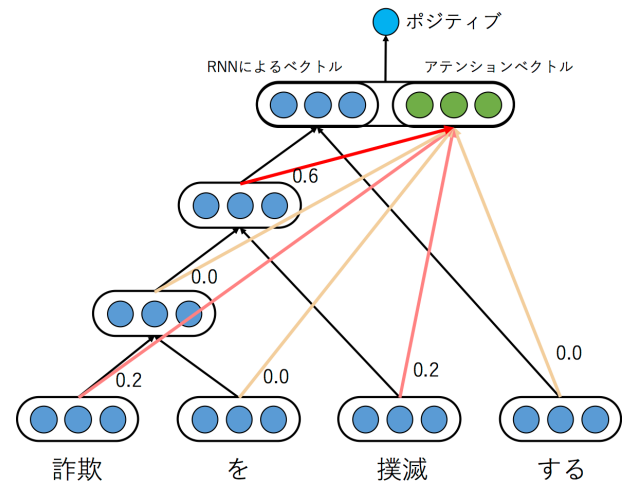


図 2: 注意型ソフトマックス分類器が張るアテンション

類器はそのラベル  $\hat{y}_j$  を以下の式によって予測する.

$$\hat{y}_j = \arg \max_y \hat{p}_\theta(y \mid h_j), \quad (10)$$

$$\hat{p}_\theta(y \mid h_j) = \text{softmax} \left( W^{(s)} h_j + b^{(s)} \right), \quad (11)$$

ここで、 $W^{(s)} \in R^{d^l \times d}$  と  $b^{(s)} \in R^d$  はパラメータ行列とバイアスであり、 $d^l$  はラベルの数を表している（ポジティブ・ネガティブの分類なら 2 値）。ソフトマックス分類器はラベルの確率分布  $y \in R^{d^l}$  を生成し、その中で最も高い次元に対応するラベルを予測ラベルとして選択する。

### 3. 極性辞書を利用した句構造注意型ニューラル評価極性分類

この節では、2 節で説明した RNN による評価極性分類器を拡張し、提案手法となるアテンション機構を導入した注意型ソフトマックス分類器、追加で利用する評価極性辞書の利用方法、最後に学習の目的関数について説明する。

### 3.1 注意型ソフトマックス分類器

句単位の極性アノテーションが得られない場合、比較的長い句の場合、特に文の場合には、学習時に木構造の葉ノード付近まで誤差を伝えるのが難しい。後の実験で示すように、評価極性分類は単語の情報だけでも比較的高い分類精度が得られるタスクであり、終端ノードや前終端ノードのような深い層の情報も有効に活用したい。この問題を解消するために、注意（アテンション）機構を導入する。

ソフトマックス分類器は入力として新たにアテンションベクトル  $a_j$  を入力として受け取る.

$$\hat{p}_{\theta}(y \mid h_j) = \text{softmax} \left( W^{(s')} \begin{bmatrix} a_j \\ h_j \end{bmatrix} + b^{(a)} \right), \quad (12)$$

$$a_j = \sum_i a_{ji} \odot h_i, \quad (13)$$

$$a_{ji} = \frac{\exp \left( W^{(a2)} \tanh \left( W^{(a1)} \begin{bmatrix} h_i \\ h_j \end{bmatrix} \right) \right)}{\sum_k \left( \exp \left( W^{(a2)} \tanh \left( W^{(a1)} \begin{bmatrix} h_k \\ h_j \end{bmatrix} \right) \right) \right)}, \quad (14)$$

ここで、 $W^{(s')} \in R^{d^l \times 2d}$ ,  $W^{(a1)} \in R^{d^a \times 2d}$ ,  $W^{(a2)} \in R^{1 \times d^a}$  はパラメータ行列である。式 14 では表記の簡略化のために、 $W^{(a1)}$  と  $W^{(a2)}$  に対応するバイアス項を記述していない。

アテンションベクトル  $a_j$  は分類しようとしているターゲットノード以下の全ノードに対して、それぞれどれくらい強く考慮してターゲットノードの分類に利用するかを表現している。スカラー値  $a_{ji}$  は各ノードの重みとして利用され、ターゲットノード以下の全ノードの重み付き和でアテンションベクトル  $a_j$  が構成される。

図 2 に示すように、注意型ソフトマックス分類器の左半分は 2 節に説明したような再帰的ニューラルネットワークによって文の分散表現を計算するが、右半分は句構造に対応する隠れベクトルにアテンションを張ることにより、どの単語あるいは句に注目して分類を行うか、という情報を分類で明示的に考慮することができる。

### 3.2 評価極性辞書による訓練事例の拡充

3.1 節でも述べたように、文単位でアノテーションされたコーパスでは学習のための情報が不足する。そこで、追加の事前知識として評価極性辞書の利用を検討する。具体的には、本研究では Nakagawa ら [18] と同じように、単語や短い句の評価極性辞書を利用する。これは、単語や句に関する極性情報付きの辞書は、どの言語でも比較的容易に入手・（半）自動構築可能であると考えられるためである。

学習コーパス内で評価極性辞書のエントリとマッチする単語・句が出現した際には、その単語や句に対応する評価極性を教師ラベルとして与え、訓練事例に追加して学習に利用する。Nakagawa ら [18] との違いは、我々のモデルでは評価極性辞書による極性を正解ラベルとして遠距離教師あり学習 (distantly supervised learning) [17] のように利用するのに対して、Nakagawa らは CRF の隠れ変数の初期値として利用している点である。

### 3.3 学習

ニューラルネットワークの誤差関数は、真のラベルの分布（すなわち one-hot ベクトル）とソフトマックス関数によって予測した分布との交差エントロピーの総和を利用する。

$$J(\theta) = - \sum_{k=1}^m t_k \log \hat{y}_k + \frac{\lambda}{2} \|\theta\|_2^2, \quad (15)$$

ここで、 $m$  は学習データ中に存在する教師ラベルの数を表し \*2  $\lambda$  は荷重減衰係数である。

## 4. 日本語評価極性分類実験

提案したモデルを評価するために、日本語の評価極性分類の実験を行った。

### 4.1 実験データ

RNN の入力単語ベクトルの初期値として、word2vec の Skip-gram [14–16] による事前学習を行ったベクトル表現を利用した。日本語 Wikipedia (2014.11) を KyTea [19]\*3 によって単語分割した後に、200 次元の単語分散表現を学習した。句構造解析には Ckylark [20]\*4 を利用し、Ckylark への入力には KyTea による単語分割の出力を用いた。実験には、NTCIR の日本語意見分析コーパス (NTCIR-J) を利用した [25, 26]。NTCIR-J はポジティブな文 997 文とネガティブな文 2,400 文を含んでおり、Nakagawa ら [18] と同様に NTCIR-6 と NTCIR-7 を組み合わせたものである。評価極性辞書も Nakagawa ら [18] と同様の、小林ら [11] が作成したものと、東山ら [5] が作成したものをを用いる \*5。全体として 5,447 のポジティブな表現と 8,117 のネガティブな表現を含んだ辞書が得られた。

実験は 10 分割交差検定にて行い 1 つをテストセット、1 つを開発セット、残りを学習セットとして利用した。

### 4.2 比較手法

我々の手法を他の 6 つの手法と比較する。単語分散表現を入力として利用する方法では全て word2vec [14–16] による事前学習を行ったものを利用する。提案手法と、LogRes, RNN, Tree-LSTM はそれぞれ Chainer [32]\*6 を用いて実装した。

#### 4.2.1 MFS

常に最頻出のラベルを選択する最も基本となるベースライン手法。この実験データの場合は常にネガティブを選択する。

#### 4.2.2 LogRes

線形分類器であるロジスティック回帰を利用する手法。入力には文ベクトルとして、単語の分散表現を平均したベクトルを扱う。

\*2 もしデータセットが文単位アノテーションのコーパスであれば、 $m$  は事例数となる。我々の手法の場合、評価極性辞書にマッチする表現の数だけ、事例数が増加する。

\*3 利用したバージョンは 0.4.7。

\*4 利用したのは 2016 年 7 月時点のバージョン。

\*5 <http://www.cl.ecei.tohoku.ac.jp/index.php?Open%20Resources/Japanese%20Sentiment%20Polarity%20Dictionary>

\*6 利用したバージョンは 1.9.0

表 1: ハイパーパラメータの設定

| パラメータ    | 値          |
|----------|------------|
| 隠れ層次元    | 200        |
| 最適化手法    | AdaDelta*7 |
| 荷重減衰係数   | 0.0001     |
| 勾配クリッピング | 5          |

表 2: 各手法の正解率

| 手法                      | 正解率          |
|-------------------------|--------------|
| MFS                     | 0.704        |
| LogRes                  | 0.771        |
| SdA [35]                | 0.817        |
| Tree-CRF [18]           | 0.826        |
| RNN                     | 0.517        |
| Tree-LSTM               | 0.709        |
| Tree-LSTM w/ attn       | 0.810        |
| Tree-LSTM w/ dict       | 0.829        |
| Tree-LSTM w/ attn, dict | <b>0.844</b> |

#### 4.2.3 RNN

2.1 節で説明した、最もシンプルな RNN のモデルとソフトマックス分類器によって分類する手法。

#### 4.2.4 Tree-LSTM

2.2 節で説明した、RNN に LSTM を適用し、ソフトマックス分類器によって分類する手法。フレーズ単位でアノテートされた Stanford Sentiment Treebank [29] では最高精度を示している [30]。

#### 4.2.5 SdA

その他のニューラルネットワークを利用した手法として、多層雑音除去自己符号化器 (Stacked Denoising Autoencoder) による手法 [35]。NTCIR-6 のデータでは Nakagawa らの手法による評価極性分類精度を上回っていたが、本実験ではデータが違うので数字は単純に比較できない。

#### 4.2.6 Tree-CRF

このデータセットにおいて最高精度を示していた手法。依存構造木をもとに Tree-CRF によって学習する [18]。

#### 4.2.7 Tree-LSTM w/ attn, dict

我々の提案する手法。3.1 節と 3.2 節で述べたように、アテンション機構や評価極性辞書を Tree-LSTM に適用する。

### 4.3 ハイパーパラメータ

10 分割交差検定の開発セットの正解率の平均値でハイパーパラメータをチューニングした。得られたパラメータは表 1 に示す。

### 4.4 結果

実験の結果を表 2 に示す。単純な RNN では MFS にも大差で負け、全く学習できていない。また、句単位の極性

アノテーションが得られる他のデータセットでは最高精度を示している Tree-LSTM でも MFS と同程度で、LogRes の方が高い精度になっている。

我々の提案したアテンション機構を利用した Tree-LSTM では、単語レベルのレイヤーの情報を減衰させることなく用いることができる SdA と同程度の性能を示している。Tree-LSTM と SdA のどちらも単語分散表現のみを入力としている。一方、評価極性辞書も利用した Tree-CRF と比べると、アテンション機構と評価極性辞書の両方を利用した Tree-LSTM による提案手法が高い精度を示しており、このデータセットにおいて最高精度を得ることができた。

## 5. 議論

結果が示す通り、単純な RNN では評価極性分類に有効な文の分散表現を学習することが困難であることがわかる。LSTM を適用することで RNN の勾配消失を解決すれば分類精度は改善されるが、MFS ベースラインをわずかに上回る程度である。これは、LogRes や SdA のように、構文木を用いないため勾配消失問題がない手法が高い精度を示すのと比較すると対照的である。

しかしながら、RNN は他の任意の句レベルの情報を受け取ることができればより正確に表現を獲得できることがわかった。例えば、我々のモデルでは、アテンションの情報と評価極性辞書の情報を句レベルの情報として Tree-LSTM に与えた。アテンションの情報のみでは Tree-CRF による分類手法よりも低い精度であるが、評価極性辞書を組み合わせることで、Tree-CRF よりも 1.8 ポイント高い精度を達成できた。アテンションよりも評価極性辞書による貢献の方が大きい。RNN における各句に対する教師ラベルの役割は大きいものと考えられる。Tree-CRF も我々の提案手法と同じ評価極性辞書を用いているが、フレーズに対する極性の初期値として与えて学習につれて更新するより、遠距離教師あり学習のように正解ラベルとして学習する方が、最終的な分類精度向上につながるがわかった。Nakagawa らは 10 分割交差検定において開発セットを用いていないのに対して、我々はハイパーパラメータのチューニングに開発セットを利用しており、学習セットの量が 10 ポイント少ない量になっているにもかかわらず、彼らの手法より高い精度を達成することができている。

図 3 に正解した例を、図 4 に不正解した例をテストデータ中からサンプルして示す。ノード内の数字はルートノードでの分類時にどれだけアテンションされたかの重みを表し、重い順にトップ 5 を赤いノードで示している。三重線で囲まれたノードはラベルが付与されたノードで、ルートノードの他に評価極性辞書とマッチした単語・句も該当する。ラベルの付与されたノードでは、T が正解、Y が予測の結果をそれぞれ示し、0 がネガティブ、1 がポジティブを示している。

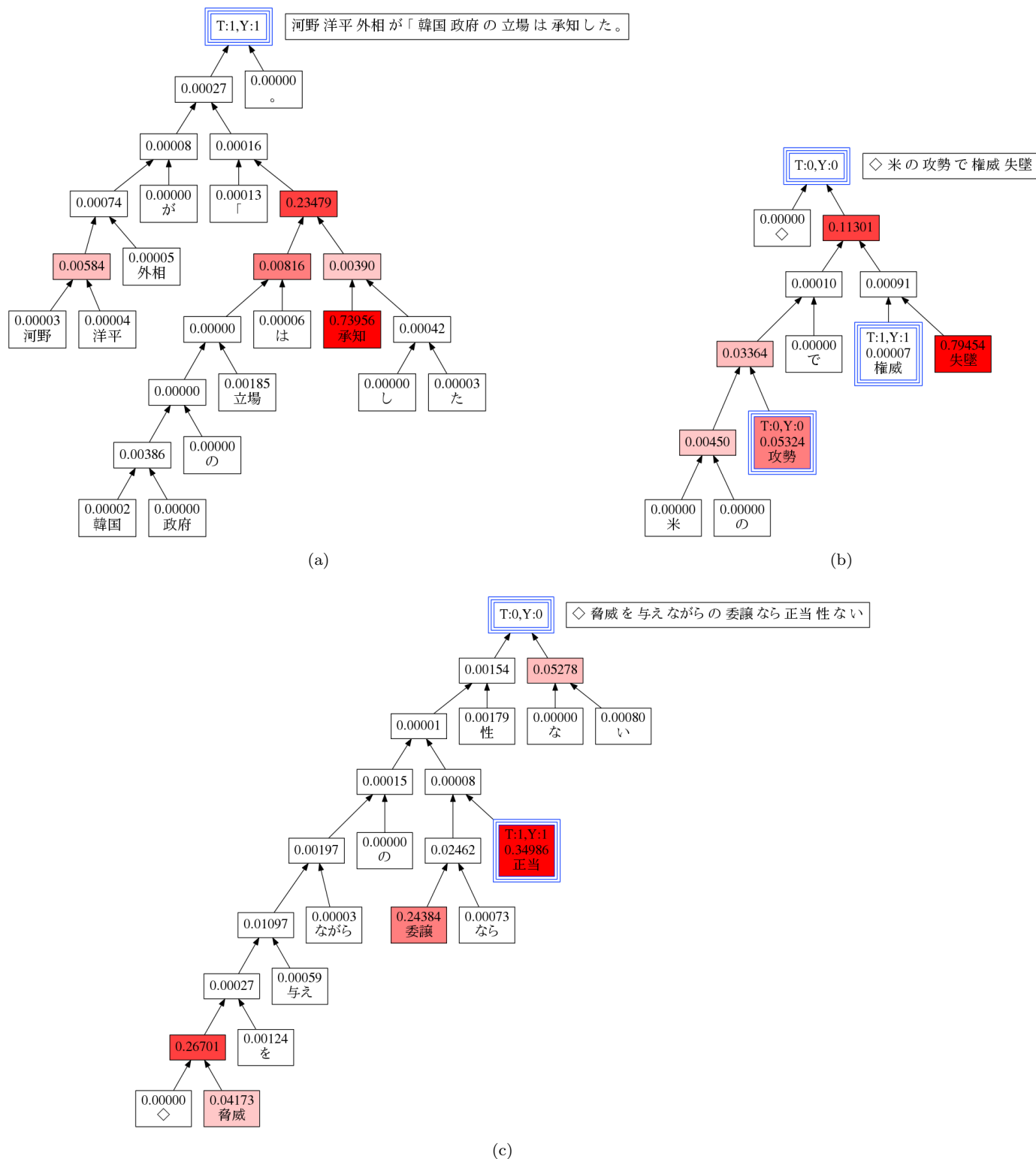


図 3: 正しく分類できた事例

図 3a を見ると“承知”という単語や“韓国政府の立場は承知した”という句にうまくアテンションを張れていることがわかる。図 3b も同様に“失墜”という単語や“米の攻勢で権威失墜”という句にアテンションを張れている。図 3c では、“正当”という単語がポジティブだと正しく分類しているが、文末に出てくる否定表現の“ない”や前半の

“脅威”と合わせてアテンションを張り、文の表現では正しくネガティブだと推定できている。

一方で図 4a では、“心配”という単語がネガティブだと正しく分類していて、かつ否定表現まで含めた“心配するな”にもアテンションを張っているが、アテンションの 9 割強が“心配”で占めており最終的な文の表現はネガティ

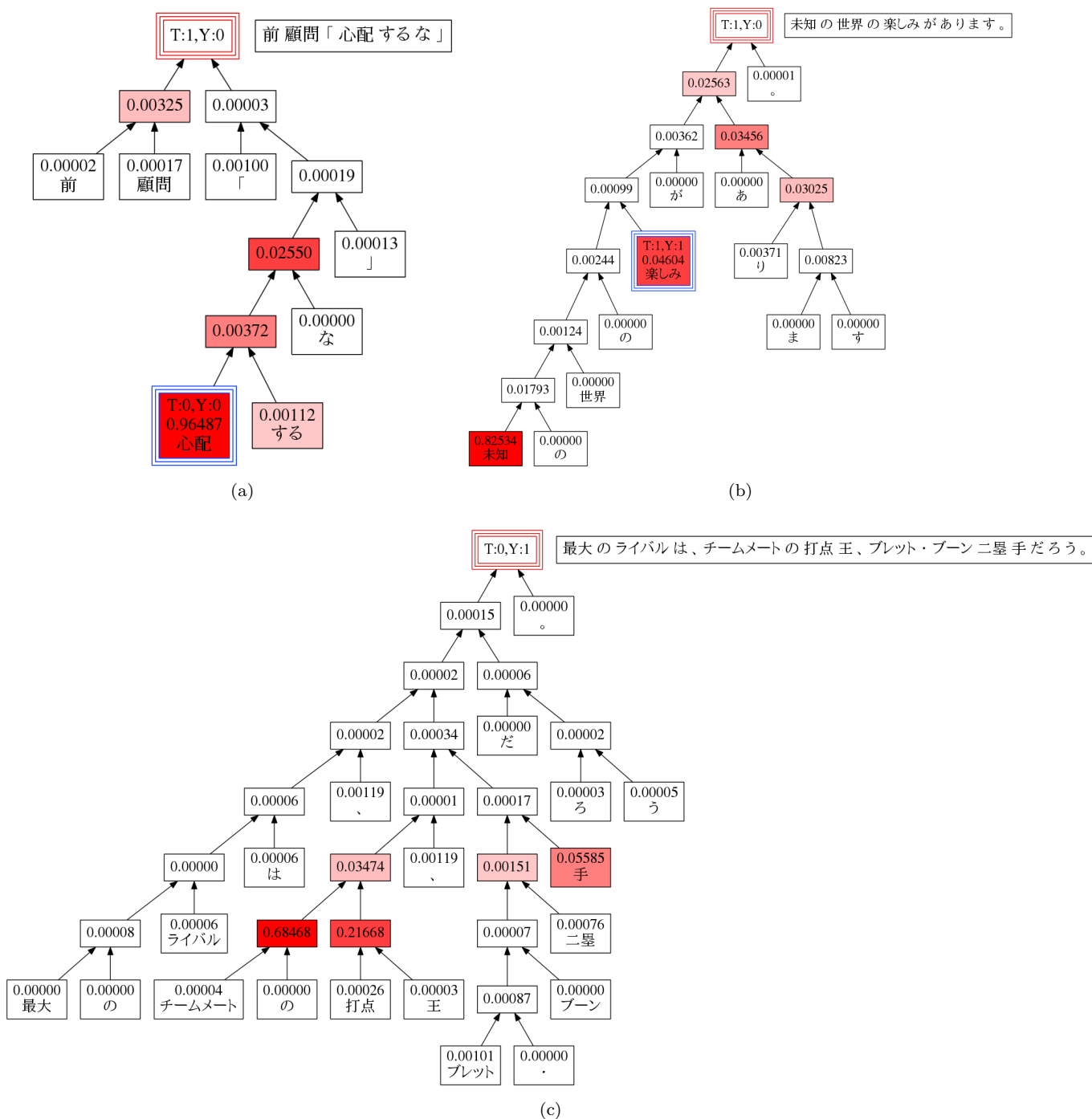


図 4: 正しく分類できなかった事例

ブと分類し間違っしてしまっている。これは文が短すぎるためにアテンションを張る対象が少なくなり、一つの単語や句に偏ってアテンションを張ってしまっているのではないかと考えられる。同じく図 4b でも、“楽しみ”という単語をポジティブだと正しく分類していて、かつ“あります”という句に対してもアテンションを張っているが、“未知”という単語へのアテンションが強すぎて文の表現をネガティブだと間違えて分類している。図 4c では“チームメート”や“打点王”などにアテンションが張られ、チームメート

に打点王がいることを考慮してポジティブを出力しているが、“ライバル”などのキーワードからそれが相手のチームであることを理解してネガティブを出力しなければならない。

## 6. 関連研究

この節では日本語評価極性分類、再帰的ニューラルネットワーク、アテンションモデルに関する関連研究をそれぞれ述べる。

## 6.1 日本語評価極性分類

Ikeda ら [8] は単語の周囲の窓内で極性が反転するか否かを分類する手法を提案し、日本語評価極性分類に適用した。Ikeda らの手法では構文情報は利用されていない。また、単語や句、文の分散表現も考慮されていない。

Nakagawa ら [18] は依存構造木にもとづいた評価極性分類手法を提案した。このモデルは Tree-CRF によって、依存構造木の部分木に対応する評価極性も考慮しながら分類する。それぞれの部分木に対応する隠れ変数が評価極性を表し、隠れ変数どうしの相互作用も考慮する。彼らの手法では、句の隠れ変数の初期値を決定するために評価極性辞書利用されている。また、極性反転を捉えるための極性反転辞書も利用して評価極性分類を行っているが、我々の提案する手法では単純な評価極性辞書のみを利用している。彼らの手法で用いられる素性は基本的にバイナリ素性であり、単語や句の分散表現は考慮されていない。

Zhang ら [35] は多層雑音除去自己符号化器と単語および文の分散表現を利用して評価極性分類を行った。入力文内の単語の分散表現の平均ベクトルを文の分散表現として、多層雑音除去自己符号化器へと入力した。統語的情報を一切扱っていないが、多層に事前学習を行いかつ雑音除去による汎化能力によって高い精度を示した。一方、彼らの手法では統語情報の他に句の分散表現も考慮せず、評価極性辞書も用いていない。彼らの手法で評価極性辞書を活用する方法は自明ではないが、本研究は遠距離教師あり学習と類似した方法で自然に評価極性辞書を利用している。

## 6.2 再帰的ニューラルネットワーク

2 節で述べたように、評価極性分類において様々な再帰的ニューラルネットワークのバリエーションが考案されてきた [22, 27–30, 36]。これらモデルは、文の分散表現を得るために構文木に沿ってボトムアップに計算を行い、文の表現は根ノードに対応する表現によって表される。この方法では、それぞれの句に対して構成的に計算を行い、その構成関数を学習することを目的としている。

構成関数に焦点を置いた研究は RNN の構成能力を向上させている [22, 27, 29]。Socher ら [27] は単語をベクトルだけでなく行列としてもパラメータ化し、ベクトルと行列の両方を用いた構成関数によって極性の変化を捉えるような構成関数を提案した。しかし、出現する語彙全てにベクトルと行列の両方を対応させるとパラメータの増大によって学習が困難になってしまう。そこで、Socher ら [29] は双線形形式の行列をベクトルの次元毎に用意し、ベクトルの相互作用を考慮する構成関数を提案しパラメータの増大を抑えた。Qian ら [22] は句のカテゴリラベルを構成関数に取り入れ、親や子のカテゴリによって異なる構成を行うような方法を提案した。

Tai ら [30] は 2.2 節で述べた通り、LSTM ユニットの

RNN へ適用させた。LSTM の導入によって増えたパラメータは、上で述べた手法とは違いすべての構成ステップで勾配が計算され更新が行われるため、問題なく学習を行うことができる。我々はこの Tree-LSTM のモデルを RNN として利用した。

## 6.3 注意機構

人間の注意（アテンション）能力 [23] に関するインスピレーションを心理学の分野から取り入れ、コンピュータサイエンスの分野でもアテンション機構が使われるようになった。アテンションの主な着眼点は、入力のどの部分により焦点を当てるかという機能である。

NLP の分野では機械翻訳 [1, 3, 13]、文書要約 [4, 24, 33]、単語分散表現の学習 [12]、画像キャプション生成 [34] などの様々なタスクにアテンション機構が導入されてきた。

これらのアプローチは、入力の各単語や領域が単語の生成にどれだけ寄与するかをアテンション機構によって考慮している。これらのアプローチのほとんどが単語系列の生成のタスクに取り組んでいる<sup>\*8</sup>のに対して、我々のモデルでは評価極性分類にアテンション機構を導入している。江里口ら [3] は再帰的ニューラルネットワークおよび回帰的ニューラルネットワークの両方を用いたニューラル機械翻訳において、木構造に対するアテンション機構を提案しているが、我々のモデルでは、単語レベルの回帰的ニューラルネットワークは除外している。

## 7. おわりに

本論文では、アテンション機構と評価極性辞書の利用によって、文単位でしかアノテーションされていないコーパスにおいて再帰的ニューラルネットワークを学習する際の勾配消失問題の解決方法を提案した。我々の方法では、句のベクトル表現はアテンション情報として文の極性を分類する際に利用される。また、評価極性辞書に含まれる単語や句は教師ラベルとして学習に利用される。実験の結果、アテンションと評価極性辞書を同時に利用した手法が既存の手法を上回り、日本語評価極性分類タスクにおいて最高精度を示した。

今後の課題として、Qian ら [22] が取り組んだ句カテゴリの利用を我々のモデルにも取り組むことや、句単位のアノテーションの有無による結果の差異を Stanford Sentiment Treebank [29] などのデータセットを用いて比較実験することを考えている。

## 参考文献

- [1] Dzmitry Bahdanau, KyungHyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to

<sup>\*8</sup> 単語分散表現の学習 [12] は系列の生成ではなく点推定を行っている。

- align and translate. In *Proceedings of the 3rd International Conference on Learning Representations*, 2015.
- [2] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, Vol. 5, No. 2, pp. 157–166, 1994.
- [3] Akiko Eriguchi, Kazuma Hashimoto, and Yoshimasa Tsuruoka. Tree-to-sequence attentional neural machine translation. In *Proceedings of the 54th Annual Meeting of the Association for Natural Language Processing*, 2016.
- [4] Karl Moritz Hermann, Tomáš Kočiský, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. Teaching machines to read and comprehend. In *Advances in Neural Information Processing Systems 28*, pp. 1693–1701, 2015.
- [5] Masahiko Higashiyama, Kentaro Inui, and Yuji Matsumoto. Learning sentiment of nouns from selectional preferences of verbs and adjectives. In *Proceedings of the 14th Annual Meeting of the Association for Natural Language Processing*, pp. 584–587, 2008.
- [6] Sepp Hochreiter. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 6, No. 2, pp. 107–116, 1998.
- [7] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [8] Daisuke Ikeda, Hiroya Takamura, Lev-Arie Ratinov, and Manabu Okumura. Learning to shift the polarity of words for sentiment classification. In *Proceedings of the 3rd International Joint Conference on Natural Language Processing*, pp. 296–303, 2008.
- [9] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. A convolutional neural network for modelling sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Natural Language Processing*, 2014.
- [10] Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1746–1751, 2014.
- [11] Nozomi Kobayashi, Kentaro Inui, Yuji Matsumoto, and Kenji Tateishi. Collecting evaluative expressions for opinion extraction. *Journal of Natural Language Processing*, Vol. 12, No. 3, pp. 203–222, 2005.
- [12] Wang Ling, Lin Chu-Cheng, Yulia Tsvetkov, Silvio Amir, Ramón Fernández Astudillo, Chris Dyer, Alan W Black, and Isabel Trancoso. Not all contexts are created equal: Better word representations with variable attention. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1367–1372, 2015.
- [13] Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1412–1421, 2015.
- [14] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In *Proceedings of the 1st International Conference on Learning Representations*, 2013.
- [15] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26*, pp. 3111–3119, 2013.
- [16] Tomas Mikolov, Wen tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In *Proceedings of Human Language Technologies: The 2013 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 746–751, 2013.
- [17] Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. Distant supervision for relation extraction without labeled data. In *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*, pp. 1003–1011, 2009.
- [18] Tetsuji Nakagawa, Kentaro Inui, and Sadao Kurohashi. Dependency tree-based sentiment classification using CRFs with hidden variables. In *Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 786–794, 2010.
- [19] Graham Neubig, Yosuke Nakata, and Shinsuke Mori. Pointwise prediction for robust, adaptable Japanese morphological analysis. In *Proceedings of The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 529–533, 2011.
- [20] Yusuke Oda, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura. Ckylark: A more robust PCFG-LA parser. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pp. 41–45, 2015.
- [21] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*, pp. 79–86, 2002.
- [22] Qiao Qian, Bo Tian, Minlie Huang, Yang Liu, Xuan Zhu, and Xiaoyan Zhu. Learning tag embeddings and tag-specific composition functions in recursive neural network. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pp. 1365–1374, 2015.
- [23] Ronald A. Rensink. The dynamic representation of scenes. *Visual cognition*, Vol. 7, No. 1-3, pp. 17–42, 2000.
- [24] Alexander M. Rush, Sumit Chopra, and Jason Weston. A neural attention model for sentence summarization. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 379–389, 2015.
- [25] Yohei Seki, David Kirk Evans, and Lun-Wei Ku. Overview of opinion analysis pilot task at NTCIR-6. In *Proceedings of the 6th NTCIR Workshop*, pp. 265–278, 2007.
- [26] Yohei Seki, David Kirk Evans, and Lun-Wei Ku. Overview of multilingual opinion analysis task at NTCIR-7. In *Proceedings of the 7th NTCIR Workshop*, pp. 265–278, 2008.
- [27] Richard Socher, Brody Huvaland Christopher D. Manning, and Andrew Y. Ng. Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 1201–1211, 2012.
- [28] Richard Socher, Jeffrey Pennington, Eric H. Huang,

- Andrew Y. Ng, and Christopher D. Manning. Semi-supervised recursive autoencoders for predicting sentiment distributions. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 151–161, 2011.
- [29] Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1631–1642, 2013.
- [30] Kai Sheng Tai, Richard Socher, and Christopher D. Manning. Improved semantic representations from tree-structured long short-term memory networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pp. 1556–1566, 2015.
- [31] Jie Tang, Mingcai Hong, Juanzi Li, and Bangyong Liang. Tree-structured conditional random fields for semantic annotation. In *Proceedings of International Semantic Web Conference*, pp. 640–653, 2006.
- [32] Seiya Tokui, Kenta Oono, Shohei Hido, and Justin Clayton. Chainer: a next-generation open source framework for deep learning. In *Proceedings of Workshop on Machine Learning Systems in The 29th Annual Conference on Neural Information Processing Systems*, 2015.
- [33] Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. Pointer networks. In *Advances in Neural Information Processing Systems 28*, pp. 2692–2700, 2015.
- [34] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *Proceedings of the 32nd International Conference on Machine Learning*, pp. 2048–2057, 2015.
- [35] Peinan Zhang and Mamoru Komachi. Japanese sentiment classification with stacked denoising auto-encoder using distributed word representation. In *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation*, pp. 150–159, 2015.
- [36] Xiaodan Zhu and Parinaz Sobhani. Long short-term memory over recursive structures. In *Proceedings of the 32nd International Conference on Machine Learning*, pp. 1604–1612, 2015.