

解釈可能なニューラルネットワークによるレビュー可視化

伊藤 友貴¹ 坪内 孝太² 山下 達雄² 坂地 泰紀¹ 和泉 潔¹

¹ 東京大学大学院 工学系研究科 ² ヤフー株式会社

{m2015tito}@socsim.org

1 はじめに

SNS, 掲示板, 商品レビュー, クチコミといったセンチメントを含む web 上のテキスト情報はユーザーあるいは国民の生の声であり, これらの情報は商品やサービスの改善案, 政策提案などを考える上で重要であると考えられる. しかし, このような web 上のテキストデータは膨大であり, その全てを読み込んで把握することは現実的ではない. そのため, これらのテキストからセンチメントを抽出し, わかりやすく可視化するような技術はビジネス上重要であると考えられる. このとき, 文書あるいはコメント全体のセンチメントの抽出だけでは不十分であり, 「何がポジティブなのかネガティブなのか (Content-based Sentiment)」と「どのようにポジティブなのかネガティブなのか (Sense-based Sentiment)」を単語単位でのセンチメントを抽出し, わかりやすく可視化することが求められる.

以上を踏まえ, 本研究では文書内の Sense-based Sentiment 及び Content-based Sentiment を抽出できる, 解釈可能な Network Model, Text Visualization Neural Network Model (TVNN) を用いた Sense-based Sentiment 及び Content-based Sentiment のユーザーフレンドリーな可視化手法を提案する.

2 TVNN

2.1 Architecture

TVNN は Original word-level sentiment layer (OWSL), Relation matrix between terms, contextual word-level sentiment layer (CWSL) (Figure 1) からなる, コメント $\mathbf{Q} = \{w_t^Q\}_{t=1}^n$ 入力すると, そのポジネガ予測 $y^Q \in \{0(\text{negative}), 1(\text{positive})\}$ を出力する.

本論文ではコーパスに出現する語彙数 v の語彙集合を $\{w_i\}_{i=1}^v$, 単語 w_i の語彙 ID を $I(w_i)$, $\mathbf{w}_i^{em} \in \mathbb{R}^e$ を単語 w_i の次元 e の skip-gram [5] 及び用意されたコー

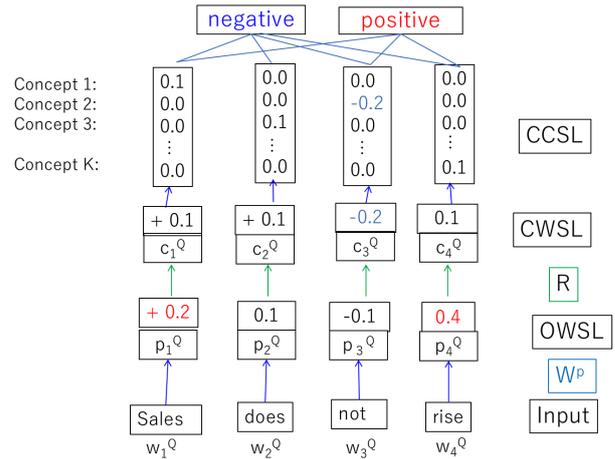


図 1: TVNN Architecture

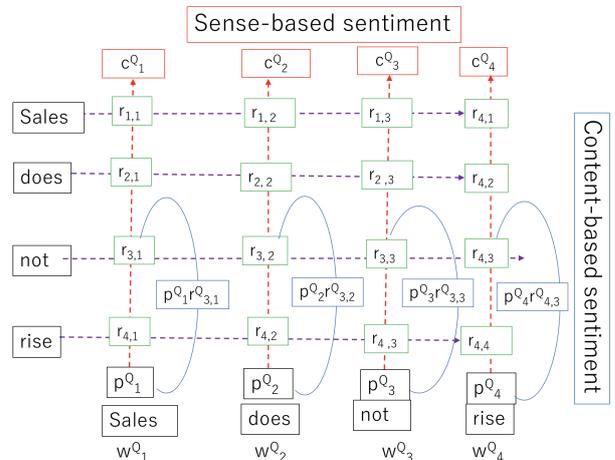


図 2: Relation matrix between terms

パスから計算された分散表現とし, さらに $\mathbf{W}^{em} \in \mathbb{R}^{v \times e} := [\mathbf{w}_1^{emT}, \dots, \mathbf{w}_v^{emT}]^T$ とする.

OWSL: この層ではコメント $\mathbf{Q} = \{w_t^Q\}_{t=1}^n$ の各単語をその単語が文脈に左右されずに持つセンチメント値, オリジナルセンチメント値に変換する.

$$p_t^Q := w_{I(w_t^Q)}^p \quad (1)$$

ここで, $\mathbf{W}^p \in \mathbb{R}^v$ は各単語のオリジナルセンチメン

ト値を表す. w_i^p は \mathbf{W}^p の i 番目の要素を表し, w_i^p の値が w_i のオリジナルセンチメント値に対応する.

Relation matrix between terms: この層ではコメント内の各単語を分散表現に $\{e_t^Q\}_{t=1}^n$ using \mathbf{W}^{em} により変換し, その後 bidirectional long short-term memory, namely, LSTM [6] によ文脈表現 $\{h_t^Q\}_{t=1}^n$ に変換する.

$$h_t^Q = \text{LSTM}(e_t^Q). \quad (2)$$

$\{h_t^Q\}_{t=1}^n$ により以下のように単語間の関係を表現する.

$$g_{t,t'}^Q = \tanh(h_t^Q \cdot h_{t'}^Q), \quad (3)$$

$$g_t^Q = [g_{t,1}^Q, \dots, g_{t,n}^Q]^T, \quad (4)$$

$$R^Q = \text{Softmax}([g_1^Q, \dots, g_n^Q])^T \in \mathbb{R}^{n \times n} \quad (5)$$

$r_{t,t'}^Q$ を R^Q の (t, t') 成分としたとき, $r_{t,t'}^Q$ は単語 w_t^Q に単語 $w_{t'}^Q$ が与える影響を表す.

CWSL: この層では $\{p_t^Q\}_{t=1}^n$ を単語単位での文脈考慮後のセンチメント値 $\{c_t^Q\}_{t=1}^n$ に変換する.

$$c_t^Q := p_t^Q \sum_{t'=1}^n r_{t,t'}^Q$$

出力: 最後にセンチメントタグ

$y^Q \in \{0(\text{negative}), 1(\text{positive})\}$ を出力する.

$$v_t^Q = c_t^Q b_t^Q. \quad (6)$$

$$a^Q = \text{Softmax}(\mathbf{W}^O \sum_{t=1}^n v_t^Q), y^Q = \text{argmax } a^Q.$$

ここで, $b_t^Q := \max(\text{Softmax}(\mathbf{W}_c e_t^Q - t_c), 0)$, $v_t^Q \in \mathbb{R}^K$, $b_t^Q \in \mathbb{R}^K$, $t_c > 0$ はハイパーパラメーター, $\mathbf{W}_c \in \mathbb{R}^{K \times e}$ は spherical k-means method [3] によって K 個に単語をクラスタリングした際のセントロイドのベクトル $\{w_i^{em}\}_{i=1}^v$ を表す.

2.2 学習

TVNN の各パラメーターはコメントとそのポジネガセンチメントタグ $\{(\mathbf{Q}_i, d^{\mathbf{Q}_i})\}_{i=1}^N$ からなる大きさ N のデータセットを用いて学習できる. ここで, $d^{\mathbf{Q}_i}$ はコメント \mathbf{Q}_i のポジネガセンチメントタグであり, \mathbf{Q}_i がポジティブのとき $d^{\mathbf{Q}_i} = 1 d^{\mathbf{Q}_i}$, ネガティブのとき $d^{\mathbf{Q}_i} = 0 d^{\mathbf{Q}_i}$ である. このとき, a^Q と d^Q の間の cross entropy をコスト関数とする backpropagation による学習を基本としつつも Algorithm 1 の process 4–6 のように \mathbf{W}^O において制約をかけながら学習をする.

このような制約をかけることにより, OWSL が各単語のオリジナルセンチメント値を表しつつ (Proposition 2.1 より) も backpropagation による学習を十分な回数の更新後には阻害しない (Proposition 2.2.)

Algorithm 1 Learning

```

1: for  $i \leftarrow 1$  to  $v$  do  $w_{p,i} \leftarrow 0$ 
2: while learning has not been finished do
3:   Update  $\mathbf{W}^O$ ,  $\mathbf{W}_p$  and the LSTM cells using the
   backpropagation method.;
4:   for  $k \leftarrow 1$  to  $K$  do
5:     if  $w_{1,k}^o > 0$  then  $w_{1,k}^o \leftarrow 0$ ;
6:     if  $w_{2,k}^o < 0$  then  $w_{2,k}^o \leftarrow 0$ ;
7:   end while

```

Proposition 2.1 *If we use the Update, then, for every $p_t^Q \in \{\{p_t^Q\}_{t=1}^n\}$,*

$$\frac{\partial L^Q}{\partial w_{I(w_t^Q)}^p} = \frac{\partial L^Q}{\partial p_t^Q} \begin{cases} < 0 & (d^Q = 1) \\ > 0 & (d^Q = 0) \end{cases}, \quad (7)$$

where Ω^{tr} means passages in a training dataset.

Proposition 2.2 *Let the value of \mathbf{W}^O before and after performing Update in Algorithm 1 in the t th update iteration be $\mathbf{W}_t^{O,a}$ and $\mathbf{W}_t^{O,b}$, respectively. Then, $\frac{\|\mathbf{W}_t^{O,a} - \mathbf{W}_t^{O,b}\|_2}{\|\mathbf{W}_t^{O,b}\|_2} \xrightarrow[t \rightarrow \infty]{} 0$ is established.*

2.3 TVNN による可視化

TVNN を用いて $CS(w_i) := \sum_{t=1}^n c_t X(w_t^Q, w_i)$ を計算することで各単語 w_i の Sense-based Sentiment を抽出でき, また $CS'(w_i) := \sum_{t'=1}^n \sum_{t=1}^n p_{t,t'}^Q r_{t,t'}^Q X(w_t^Q, w_i)$ を計算することで各単語 w_i の Content-based Sentiment 抽出できる. ここで, $X(w', w) = \begin{cases} 1 & (w' = w) \\ 0 & (w' \neq w) \end{cases}$ である. これらのスコアを用いて, 本研究では図 3 のように円の内側にクラスタ単位での Content-based Sentiment を, そして円の外側に各クラスタに影響を与えた単語の Sense-based Sentiment の値を word cloud 形式で可視化することを試みる. このような可視化を用いることで, ユーザーは生テキストを読む場合よりも効率よくレビューの情報を収集できる可能性がある.

3 評価実験

3.1 事前評価

可視化結果を評価する前に、TVNN において、OWSL 及び CWSL が対応する値をそれぞれ表すことができているかどうかを実データを用いて検証した。

3.1.1 評価基準

(1) OWSL の妥当性: OWSL の妥当性を訓練データを用いて学習後の W^p を用いて構築した極性辞書が人手で構築した辞書に合致するかどうかで評価した。Logistic fixed weight model (LFW)[7], Sentiment-oriented NN (SONN) [4], 及び PMI [7] による予測結果と比較して妥当性を検証した。

(2) CWSL の妥当性: CWSL の妥当性を訓練データのコメントとそのセンチメントタグを用いて学習後の TVNN が正しく評価データのコメントのセンチメントタグを予測できるかどうかで評価した。これは W^p の値が正しければ Term Relation Property の妥当性が y^Q の妥当性に直結するという仮説に基づく。Logistic 回帰 (LR) 及び Bidirectional RNN with LSTM cells (RNN) による予測結果と比較して妥当性を検証した。

3.1.2 データセット

コメントとそのセンチメントタグのデータとして景気ウォッチャー調査の現状に関するコメントのデータセット (Economic Review (Now)), 景気ウォッチャー調査の先行きに関するコメントのデータセット (Economic Review (Future)), 2014 年 9 月の Yahoo Finance 掲示板のコメントのデータセット (Yahoo Microblog), ヤフーショッピングレビュー (Shopping Review), Amazon Product Review の Book に関するレビューのデータセット (Book), Amazon Product Review の家電に関するレビューのデータセット (Electronic), Amazon Product Review の Movie・TV に関するレビューのデータセット (Movie), Sentiment 140¹ を用いた。

OWSL の妥当性の評価の際には、それぞれのコーパスに対応するポジネガ単語リスト [2, 1] をを用いた。

¹<https://www.kaggle.com/kazanova/sentiment140>

3.1.3 評価結果

表 1 は original sentiment assignment Property の評価結果である。TVNN は他の手法より高い性能でオリジナルセンチメントの予測を行うことができた。これより十分に高い Original Sentiment Assignment Property を TVNN は持つと考えられる。表 2 は Term Relation Property の評価結果である。TVNN は RNN より低いもののそれ以外の手法より高く Sentiment Tag の予測を行うことができた。これより十分に高い Term Relation Property を TVNN は持つと考えられる。

Model	Economic	Economic	Yahoo	Shopping
	Review (Now)	Review (Future)	Microblog	Review
LFW	0.728	0.724	0.786	0.704
SONN	0.719	0.748	0.733	0.767
PMI	0.754	0.757	0.796	0.785
TVNN	0.825	0.815	0.810	0.795

Model	Book	Movie	Electronic	Sentiment 140
	LFW	0.640	0.600	0.681
SONN	0.643	0.636	0.690	0.705
PMI	0.682	0.664	0.692	0.733
TVNN	0.704	0.685	0.731	0.735

表 1: OWSL の妥当性評価の結果 (macro F_1 score)

tion Property の評価結果である。TVNN は RNN より低いもののそれ以外の手法より高く Sentiment Tag の予測を行うことができた。これより十分に高い Term Relation Property を TVNN は持つと考えられる。

Model	Economic	Economic	Yahoo	Shopping
	review (now)	review (future)	microblog	review
LR	0.876	0.879	0.741	0.956
RNN	0.922	0.932	0.749	0.971
TVNN	0.915	0.936	0.766	0.969

Model	Book	Electronic	Movie	Sentiment 140
	LR	0.915	0.856	0.871
RNN	0.960	0.936	0.925	0.837
TVNN	0.954	0.926	0.911	0.829

表 2: CWSL の妥当性評価の結果 (macro F_1 score)

4 ユーザー評価

4.1 評価基準

TVNN による可視化の有用性を一つあたりに特定の店舗のレビュー 100 件が含まれるようなレビュー集 200 件 (Long Set), 及びレビュー 30 件が含まれるようなレビュー集 180 件 (Short Set) を用いて評価した。クラウドソーシングによって、一つのレビューにつき 3 人ずつアノテーターを割り当て、図 3 のような質問を「出力された画像を用いて回答する場合」と「テキ

ストを読んで回答する場合」それぞれについて実施し、その結果をもとに以下の形で「画像の正しさ」と「画像の使いやすさ」を検証した。

(1) 画像の正しさ: テキストを読んだ場合の回答結果(多数決)と画像を見て回答した場合の回答結果(多数決)が一致するかどうかで「画像の正しさ」を検証した。レビューの評価(1-5)の平均点が3より大きければポジティブ, 3より小さければネガティブとした場合の結果(ベースライン)と比較した。

(2) 画像の使いやすさ: テキストを読んだ場合(ベースライン)の回答時間と画像から判断した場合の回答時間を比較した。

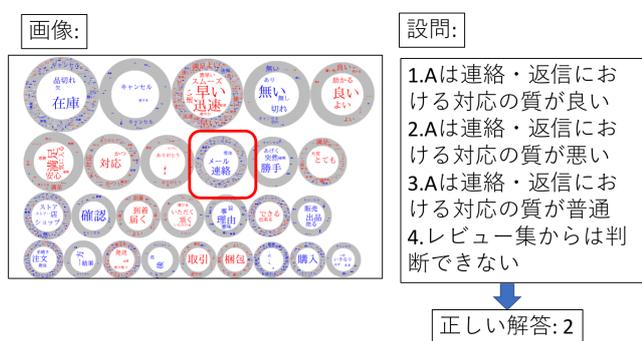


図 3: TVNNSample

図 4: クラウドソーシングにおける質問例

4.2 結果

表 3 及び表 4 が結果である。この結果より、本手法を用いることでそれぞれのベースラインよりも正確かつ短期間の間にレビューの概要を掴むことが可能であることが検証できた。

表 3: (1) 画像の正しさ

	Short Reviews	Long Reviews	Total
ベースライン	0.510	0.540	0.547
TVNN	0.660	0.598	0.695

表 4: (2) 画像の使いやすさ

	Short Reviews	Long Reviews
ベースライン(テキスト)	116 sec	156 sec
TVNN(画像)	92 sec	134sec

5 おわりに

本研究では何がポジティブなのかネガティブなのか(Content-based Sentiment)と「どのようにポジティブなのかネガティブなのか(Sense-based Sentiment)」を単語単位でのセンチメントを抽出し可能な解釈可能なNN, TVNNを用いたレビューの可視化手法を提案し、その性能を実データを用いて評価した。今後はよりわかりやすい可視化を目指し、可視化手法の改善を行う予定である。

参考文献

- [1] M. Hu and B. Liu. Mining and summarizing customer reviews. In *KDD-2004*, 2004.
- [2] Tomoki Ito, Kota Tsubouchi, Hiroki Sakaji, Tatsuo Yamashita, and Kiyoshi Izumi. Csn: Contextual sentiment neural network. In *IEEE ICDM 2019*, 2019.
- [3] M. Kober K. Hornik, I. Feinerer and C. Buchta. Spherical k-means clustering. *Journal of Statistical Software*, Vol. 50, No. 10, pp. 1–22, 2012.
- [4] Q. Li. Learning stock market sentiment lexicon and sentiment-oriented word vector from stocktwits. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pp. 301–310, 2017.
- [5] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *NIPS*, pp. 3111–3119, 2013.
- [6] M. Schuster and K. Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, Vol. 45, No. 11, pp. 2673–2681, 1997.
- [7] D. T. Vo and Y. Zhang. Don’t count, predict! an automatic approach to learning sentiment lexicons for short text. In *ACL 2016*, pp. 219–224, 2016.