深層学習を用いた経済テキスト可視化の検証

伊藤友貴 [1*]　坂地泰紀 [1]　和泉潔 [1]
Tomoki Ito[1]　Hiroki Sakaji[1] Kiyoshi Izumi[1]

[1] 東京大学工学系研究科システム創成学専攻
[1] Graduate School of Engineering, The University of Tokyo

Abstract: 経済文書のような専門的な文書は非専門家にとって読みにくい場合が多い．そのため，非専門家を対象に経済文書上のセンチメントを単語単位で可視化するようなサポートシステムを構築することには一定の需要があると思われる．経済文書上のセンチメントを可視化する手段の一つとして近年提案された「ニューラルネットワークモデルの解釈」に関する手法，LRP (Layer-wise Representation Propagation) を用いるという手段がある．しかし現状 LRP が日本語の経済文書の可視化に有用かどうかは調査されておらず，その性質についての詳細な分析もあまりされていない．また，LRP の Attention RNN への適用方法は未だ提案されていない．本報告では LRP の Attention RNN への適用方法を提案し，また，LRP が日本語金融テキストの可視化に有用かどうかを検証する．さらに，実データを用いた検証の中で LRP を用いた日本語文書可視化の性質について分析する．

1　はじめに

経済文書のような専門的な文書は非専門家にとって読みにくい場合が多い．この原因の一つとして，経済文書のような専門的な文書を読み解くためにはドメイン特化の専門的な知識を要する必要がある場合が多々あることが挙げられる．例えば,「動きが悪い」という表現が景気動向に文書に出てきた場合，この表現の意味は「客足が悪い」,「消費が悪い」に近く，一般的な意味とは少し違う意味で使われる．このような問題を解決する策の一つとして，文書内の単語についてセンチメント値を与え，図 1 のように単語単位での可視化するという手がある．このような可視化によって非専門家であっても文書内のセンチメントを単語単位で簡単に把握することができる．これは非専門家が専門文書を読み解く上で大きな助けになると期待できる．本研究ではこのような経済文書の可視化を大規模なコメントとそのポジネガタグからなるデータセットを用いて行う手法について考える．このような問題設定の下で

競合 店 が 1 店舗 少ない なる 現 段階 で は 上向き だ ある

図 1: 非専門家のための経済文書可視化例: ネガティブ単語を青色，ポジティブ単語を赤色に着色

文書の可視化を行う手法としてロジスティック回帰モデルなどのような線形回帰モデルの重みベクトルを利用する手法，[13]．Attention メカニズム [4, 11, 20] を利用する手法が考えられる．上記に加え，深層学習モデルの解釈を行う手法 [1, 2, 5, 6, 10, 14, 16, 17, 18] も有用であると考えられる．その中でも 特に Layer-wise Representation Propagation(LRP)[2] は LSTM をセルに持つ RNN モデルへも適用できる [1] state-of-the-art な深層学習モデルの解釈を行う手法の一つであり，文書の可視化にも有用だと考えられる．しかし，LRP には現状，以下のような課題がある．

- LRP が日本語の金融文書に適用できるかの検証がほとんどされていない．
- LRP を文書可視化に適用した場合の言語的な観点からの分析がほとんど行われていない．
- RNN Attention モデルは多くのタスクにおいて RNN よりも高い予測性能を出すことが知られているにも関わらず，LRP の RNN Attention モデルへの適用方法が提案されていない．

また，LRP に限定する話ではないが，今後深層学習を利用するケースが増えることが予想されることを踏まえると，深層学習モデルを解釈する手法について様々な観点から分析しておくこと，その汎用性を高めることには一定の研究意義があると考えられる．

そこで，本研究では以下について取り組む．

- LRP の RNN Attention モデルへの適用法を提案する．
- LRP が日本語金融文書の可視化に適用可能かどうかを検証する．
- LRP による文書可視化の性質を調査する．

*Email: m2015titoh@socsim.org

本研究の貢献は以下のようにまとめられる．

- LRP の RNN Attention モデルへの適用法を提案した (第 2 節).
- オリジナルセンチメントと文脈センチメントという文書の可視化を評価するための新しい指標を提案すると共にそれらを評価するためのデータセットを作成した (第 3 節).
- LRP を用いた文書可視化が日本語金融文書に適用可能か検証し，その性質を分析した (第 4 節).

2 LRP

LRP は提案されたニューラルネットワークモデルにおいて入力値が出力値に対して与える影響を計算する手法である [2] ．出力層から入力層にかけて Chain Rule に近い形で影響度を伝搬させることで求める手法である．LSTM[7], GRU[3] を含む RNN について計算するために linear connections と multiplicative connections という二種類の影響計算方法が提案されている [1].

2.1 Lnear connections

ノード z_j が $z_j = \sum_i z_i \cdot w_{ij} + b_j$ (z_i は z_j につながるノード) と 順伝搬時に計算できるとする．ここで，w_{ij} は重みベクトル，b_j はバイアスベクトルである．また，ノード z_j の 出力層に与える影響を R_j とする．このときノード z_i の影響度 R_i を以下のように計算する．

$$R_{i\leftarrow j} = \frac{z_j \cdot w_{ij} + \frac{\epsilon sign(z_j)) + \delta b_j}{N}}{z_j + \epsilon sign(z_j)}, R_i = \sum_j R_{i\leftarrow j}$$

N は z_j につながる下位層のノードの数，ϵ は十分に小さい値である．先行研究 [1] と同様に $\epsilon = 0.001$, $sign(z_j) := (1_{z_j \geq 0} - 1_{z_j < 0})$, $\delta = 0$ とした．

2.2 Multiplicative connections

上位層のノード z_j が下位層のノード z_g, z_s によって $z_j = z_g \cdot z_s$ と計算される場合について考える．これは LSTM[7], GRU[3] などに見られる積の演算である．ここで，LSTM, GRU において sigmoid 関数によって $[0, 1]$ の値になる方を z_g，ならない方を z_s とする．このとき，$R_g = 0$, $R_s = R_j$ と計算する．

2.3 LRP for Attention RNN (提案手法)

LRP を LSTM cell を持つ Attention RNN に適用することを考える．先行研究 [1] 通り，線形結合については linear connections，LSTM における Gate 部分の結合については multiplicative connections を適用する．

さらに，Attention 部分についても multiplicative connections を適用する．この Attention 部分への 適用が本研究の提案部分となる．

3 実データによる可視化検証

本節では LRP が日本語経済文書の可視化に適用できるかどうかについて実データを用いて検証する．まず，LRP による日本語経済文書の可視化が妥当かどうかについて第 3.1 節にて定義するオリジナルセンチメント値，文脈センチメント値を正しく割り当てられるかどうかという観点から検証する．その後，LRP による可視化の結果をいくつか紹介する．

3.1 単語センチメント

本研究では以下のようにオリジナルセンチメント値，文脈センチメント値を定義する．

- オリジナルセンチメント値: 単語本来のセンチメント値.
- 文脈センチメント値: 文脈における反転情報考慮後のセンチメント値.

例として,「売上が伸びない」という文における「伸びる」という単語について考える．オリジナルセンチメント値をこの単語に与える場合はプラスの値がつくことが正しい．一方，文脈センチメント値をこの単語に与える場合には「伸びる」が「ない」によって反転を受けていることを考えてマイナスの値がつくことが正しい．

3.2 テキストコーパス

本実験においては内閣府から提供されている日経景気ウォッチャーデータ [1] を実テキストデータとして用いた．期間としては 2002 年 1 月から 2017 年 4 月までのテキストデータを用いた．これらのデータはコメントとそのセンチメントタグからなるデータセットであり，センチメントタグの種類は {1 (悪い), 2 (やや悪い), 3 (変わらない), 4 (やや良い), 5 (良い) } の 5 つのタグがついている．本研究では「悪い」,「やや悪い」

[1] http://www5.cao.go.jp/keizai3/watcherindex.html

のものをネガティブタグ,「良い」,「やや良い」のものをポジティグタグとして扱い，センチメント予測モデル構築に利用した．また，このコーパスの形態素解析には MeCab[9] を用いた．

3.3 モデル構築

各可視化手法の検証にあたって Logistic Regression model (LR), RNN model with LSTM cells (RNN)[7], Bidirectional RNN model with LSTM cells (BiRNN)[15], RNN Attention Model with LSTM cells (AttRNN)[11] の 4 つのセンチメントタグ予測モデルを構築した．このとき，3.2 節で紹介したデータのうちポジティブコメント及びネガティブコメント 20000 件ずつ (計 40000 件) を訓練データとして, ポジティブコメント及びネガティブコメント 2000 件ずつ (計 4000 件) をハイパーパラメータの探索及び学習の早期終了を行うための検証データとして用いた．

学習後にポジティブコメント及びネガティブコメント 4000 件ずつ (計 8000 件) からなるテストデータについてそのポジネガ予測力を検証した．LR, RNN, BIRNN, AttRNN それぞれの Macro F_1 score の結果はそれぞれ 0.878, 0.920, 0.921, 0.923 であった．

訓練データ，検証データ，テストデータの間に被りはない．その他の実験設定は以下の通りである．RNN, BiRNN, AttRNN における各隠れ層の次元数は 200 とし，埋め込みベクトルには 3.2 節で紹介したテキストコーパスをもとに skip-gram model (window size = 5, negative sampling を使用)[12] によって学習したものを使用した．また，RNN モデル，AttRNN モデルは共に埋め込み層 1 層，LSTM cell を含む 逆方向 RNN の層 1 層，線形結合層 1 層からなるモデルであり，BiRNN モデルは埋め込み層 1 層，LSTM cell を含む 両方向 RNN の層 1 層，線形結合層 1 層からなるモデルであった．AttRNN における Attention には dot 関数による global attention [11] を用いた．また，各 RNN モデルは Dropout 法 [19] (dropout rate = 0.5)，Adam Optimizer[8] を用い， 最大 epoch 数 = 50 という条件のもとでの学習を行った．

3.4 評価用データセット

可視化手法の評価のために評価用データセットを構築した．まず，テストデータからポジティブコメント 500 件，ネガティブコメント 500 件を抽出した．その後，人手で各コメント内の重要単語についてオリジナルセンチメント値が正か負かのタグ (オリジナルセンチメントタグ) と文脈センチメント値が正か負かどうかのタグ（文脈センチメントタグ）を付与した．オリジナルセンチメントタグについてはポジティブタグ 1,794 件，ネガティブタグ 1,062 件が付与され，文脈センチメントタグについてはポジティブタグ 1,340 件，ネガティブタグ 1516 件が付与された．

3.5 評価基準

次の 2 指標について macro F_1 値をもとに評価した．

オリジナルセンチメント: 各可視化手法によって各単語に付与するオリジナルセンチメント値の正負が人手でつけたオリジナルセンチメントタグの正負に一致する度合い．

文脈センチメント: 各可視化手法によって各単語に付与する文脈センチメント値の正負が人手でつけた文脈センチメントタグの正負に一致する度合い．

3.6 比較手法

オリジナルセンチメント，文脈センチメントの評価を以下の手法を用いて行い，結果を比較し，各手法の性質を調査した．

LR: LR Model の重みベクトルを用いて各単語へセンチメント値を付与する手法.

LRP with BiRNN: BiRNN に LRP を適用することで各単語へセンチメント値を付与する手法.

LRP with RNN: RNN に LRP を適用することで各単語へセンチメント値を付与する手法.

LRP with Attention RNN: AttRNN に LRP を適用することで各単語へセンチメント値を付与する手法．Attention 部分については 2.3 節で提案した方式に沿って LRP を適用する．

Gradient with RNN: RNN に Gradient 法 [5, 6] を適用することで各単語へセンチメント値を付与する手法.

Gradient with BiRNN: BiRNN に Gradient 法 を適用することで各単語へセンチメント値を付与する手法.

Gradient with Attention RNN: AttRNN に Gradient 法 を適用することで各単語へセンチメント値を付与する手法.

Attention RNN: Attention 層の計算に使われる dot 関数の演算結果をもとに各単語へセンチメント値を付与する手法.

4 検証結果

4.1 LRP の有用性

オリジナルセンチメント値，文脈センチメントスコアの結果は表 1 の通りである．これらの結果より RNN, BiRNN，AttRNN を用いた場合のどの場合においても LRP が Gradient 法 よりも正しくオリジナルセンチメントスコア，文脈センチメントスコア共に正しく付与できていることが確認でき，その有用性を確認できた．

4.2 Attention RNN への LRP 適用

今回提案した AttRNN への LRP への適用手法，LRP with Attention RNN が AttRNN を用いた他の可視化場合に比べ．オリジナルセンチメント，文脈センチメントの両面にて高性能で可視化でき，その妥当性を検証できた．また，本実験では LRP with Attention RNN はオリジナルセンチメントよりも文脈センチメントについてそのセンチメント値を正しく与えていた．

表 1: オリジナルセンチメント値，文脈センチメント値の付与結果 (macro F_1 スコア)

Method	オリジナル	文脈
LR	**0.910**	0.793
Grad with RNN	0.590	0.632
LRP with RNN	0.834	**0.816**
Grad with BiRNN	0.708	0.738
LRP with BiRNN	0.867	0.805
Attention RNN	0.633	0.713
Grad with Attention RNN	0.281	0.356
LRP with Attention RNN	0.680	**0.815**

4.3 各可視化手法の性質分析・考察

各可視化手法の性質を分析するためにセンチメントについて反転がある場合とない場合それぞれの場合における文脈センチメントの付与結果を見てみた（表 2)．本結果より，各手法を以下の四グループに大別できた．

- 反転がある場合にもできない場合にもそこそこ対応できるもの: LRP with RNN Attention.
- 反転がある場合は高性能はだが反転がない場合には低性能のもの: Grad with BiRNN, LRP with RNN Attention, Attention RNN.

表 2: 反転がある場合・ない場合における文脈センチメント付与性能検証結果 (macro F_1 スコア)

Method	反転あり	反転なし
LR	0.166	**0.938**
Grad with RNN	0.473	0.640
LRP with RNN	0.340	0.905
Grad with BiRNN	0.422	0.778
LRP with BiRNN	0.268	0.919
Attention RNN	0.537	0.715
Grad with Attention RNN	0.474	0.316
LRP with Attention RNN	**0.667**	0.809

- 反転がない場合は高性能はだが反転には対応ができないもの: LR, LRP with RNN, LRP with BiRNN.
- 反転があるなしに関わらず低性能のもの: Grad with RNN, Grad with BiRNN.

ただ，このグループ分けが他のデータセットでも同様かどうかは現段階では不明であり，これは検証すべき事項である．

各手法の特徴を可視化例から説明する．図 2 はセンチメント反転が起こっている場合の可視化例である．この例において「伸びる」は「ない」によってセンチメントが反転しているので「青く」色がつくのが正しい．この反転に LR, LRP with RNN は対応できていないが，Grad with RNN, Attention RNN, LRP with Attention RNN は対応できている．図 3 はセンチメント反転が起こっていない場合の可視化例である．この例において「少ない」に青く色がつく (ネガティブ) のが正しい．LR, LRP with RNN, LRP with Attention RNN は正しく色がついているが，Grad with RNN, Attention RNN は正しく色をつけることができていない．Attention RNN では全体的に赤く色がついており，コメント全体の極性がポジティブであることに引きづられて 失敗しているように見える．

5 結論

本研究では LRP が日本語金融文書の可視化に適用可能であることを実データを用いて検証し，LRP による文書可視化の性質を調査した．さらに深層学習の解釈の可視化手法 LRP の RNN Attention モデルへの適用法を提案した．実データを用いて今回提案した LRP の RNN Attention モデルへの適用法が他の RNN

LR	天候 不順 も 重なる 売上 は 伸びる ない た
LRP With RNN	天候 不順 も 重なる 売上 は 伸びる ない た
Grad With RNN	天候 不順 も 重なる 売上 は 伸びる ない た
Attention RNN	天候 不順 も 重なる 売上 は 伸びる ない た
LRP with Attention RNN	天候 不順 も 重なる 売上 は 伸びる ない た

図 2: 反転が起こっている場合の可視化例

LR	競合 店 が 1 店舗 少ない なる 現 段階 で は 上向き だ ある
LRP With RNN	競合 店 が 1 店舗 少ない なる 現 段階 で は 上向き だ ある
Grad With RNN	競合 店 が 1 店舗 少ない なる 現 段階 で は 上向き だ ある
Attention RNN	競合 店 が 1 店舗 少ない なる 現 段階 で は 上向き だ ある
LRP with Attention RNN	競合 店 が 1 店舗 少ない なる 現 段階 で は 上向き だ ある

図 3: 反転が起こっていない場合の可視化例

Attention を解釈する手法に比べ，今回利用した日本語金融文書を可視化する上では有用であることを示した．

今後の課題として他のテキストデータによる解析も行い，LRP による可視化の性質について一般化すること，センチメントの反転があるなしに関わらず正しく文脈センチメントを割り振れるような可視化手法の構築，及びオリジナルセンチメント・文脈センチメントどちらにも柔軟に対応可能な可視化手法の構築が考えられる．

謝辞

本研究の一部は JSPS 科研費 JP17J04768 の助成を受けたものである．

参考文献

[1] L. Arras, G. Montavon, K. R. Muller, and W. Samek.: Explaining Recurrent Neural Network Predictions in Sentiment Analysis. *EMNLP Workshop* (2017)

[2] S. Bach, A. Binder, G. Montavon,F. Klauschen, K. R. Muller and W. Samek.: On pixel-wise explanations for nonlinear classifier decisions by layer-wise relevance propagation. *PLOS ONE*, Vol. 10, No. 7, 1–46 (2015)

[3] J. Chung, C. Gulcehre, K. Cho, Y. Bengio.: Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. In *NIPS Workshop* (2014)

[4] Y. Dong, H. Su, J. Zhu and B. Zhang.: Improving Interpretability of Deep Neural Networks with Semantic Information. In *CVPR* (2017)

[5] S. Karen, Ve. Andrea and A. Zisserman.: Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv:1312.6034* (2013)

[6] Y. Hechtlinger.: Interpretation of prediction models using the input gradient. *arXiv:1611.07634* (2016)

[7] S. Hochreiter and Jurgen Schmidhuber.: Long short-term memory. *Neural computation*, Vol. 9, No. 8 , 1735–1780 (1997)

[8] D. P. Kingma, J. L. Ba.: ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION. *arXiv:1412.6980* (2014)

[9] T. Kudo, K. Yamamoto, Y. Matsumoto.: Applying Conditional Random Fields to Japanese Morphological Analysis. In *EMNLP*. 230–237.

[10] J. Li, D. W. Monroe, D. Jurafsky.: Understanding Neural Networks through Representation Erasure. *arXiv:1612.08220* (2016)

[11] M. Luong, H. Pham and C. D. Manning.: In *EMNLP*. Effective Approaches to Attention-based Neural Machine Translation 1412–1421 (2015)

[12] T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean.: Distributed Representations of Words and Phrases and their Compositionality. In *NIPS*. 3111–3119 (2013)

[13] K. Ravi, V. Ravi.: 2015. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*. **89**(C), 14–46 (2015)

[14] M. T. Ribeiro, S. Singh, C. Guestrin.: 2016. "Why Should I Trust You?" Explaining the Predictions of Any Classifier. In *KDD*

[15] M. Schuster and K. K. Paliwal.: Bidirectional Recurrent Neural Networks. *IEEE Transactions on Signal Processing*, Vol. 45, No. 11, 2673–2681 (1997)

[16] A. Shrikumar, P. Greenside and A. Kundaje.: Learning Important Features Through Propagating Activation Differences. In *ICML* (2017)

[17] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. A. Riedmiller.: Striving for simplicity: The all convolutional net. In *ICLR Workshop* (2015)

[18] M. Sundararajan, A. Taly, Q. Yan.: Axiomatic Attribution for Deep Networks. In *ICML* (2017)

[19] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *JMLR*, Vol. 15, No. 1, 1929–1958 (2014)

[20] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, R. Zemel, and Y. Bengio.: Show, attend and tell: Neural image caption generation with visual attention. In *ICML*. 77–81 (2015)