社団法人　電子情報通信学会
ヒューマンコミュニケーショングループ
第二種時限研究会

BERT を用いたコミックにおける固有表現の抽出に関する検討

朴 炳宣††　　松下 光範†,a

†関西大学 総合情報学部　††関西大学大学院 総合情報学研究科

a) k281401@kansai-u.ac.jp b) mat@res.kutc.kansai-u.ac.jp

概要　コミックにおける自然言語処理において，Wikipedia やレビュー，ブログといったインターネットからの資源は，コンテンツの推薦やキャラクターの特徴抽出など積極的に活用されている．しかし，これらの資源を活用するためには，登場人物の名前や道具名といったコミック特有の固有表現の抽出が必要となる．本稿では，近年多様なタスクにおいて高性能を示した汎用 Encoder である BERT (Bidirectional Encoder Representations from Transformers) を用いたコミックにおける固有表現の抽出と，標準語彙とのクラスタリングについて検討する．

キーワード　自然言語処理, 固有表現抽出, 深層学習, 転移学習, BERT

1　はじめに

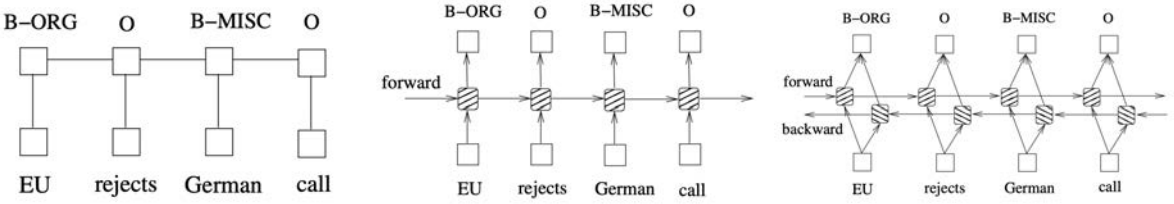
コミックの内容情報の獲得のために，コミックに描かれた内容に対する人による解釈が含まれた情報源として，Wikipedia[1]やレビューがよく用いられている [1, 2, 3]．これらの情報は，ユーザが読んだことのない未知の作品に対する情報収集に積極的に用いられるため，コミックの特徴抽出や推薦への活用が期待される．しかし，コミックに関する記事やレビューには，作品名や登場人物，世界観に関する用語といった作品内で定義がなされた未知語に基づく表現（以下，固有表現と記す）が多く含まれている．固有表現はユーザが持つ作品に対する認知度によって内容推測に役立つ情報の量が大きく変化する [4]．また，コミック作品が増えるたびに固有表現の数もまた増加するため，形態素解析や特徴抽出などの固定された語彙に基づく自然言語処理を用いた処理が困難となる．そこで，著者らはコミックに対するレビュー文中に存在する NER (Named Entity Recognition, 固有表現認識) のために CRF (Conditional Random Fields) [5] を用いた機械学習行うことで固有表現認識モデルを作成した．さらに，作成したモデルを用いて固有表現の数を調整する実験によって，ユーザがより容易に内容を推測できる情報を提示できることを確認できた．しかし，認識モデルの評価時の精度は 52.6% に留まっており，性能の向上が求められる．本稿では，NLP (Natural Language Processing, 自然言語処理) において多く用いられる LSTM (Long Short-Term Memory) [6] モデルと，幅広いタスクに対して大きな貢献を果たしている汎用エンコーダーである BERT (Bidirectional Encoder Representations from Transformers) [7] を用いたコミックのレビューに対する NER モデルを検討する．

Copyright is held by the author(s).
The article has been published without reviewing.

[1] https://www.wikipedia.org

2　関連研究

NLP に対する機械学習によるアプローチの中で，近年では NN を基盤とする深層学習によるアプローチが活発に提案されている．既存では，文章データを時系列データとして捉え，CRF[5] (図 1 の (a) 参照) などのモデルに基づき系列ラベリングを行う方法が主流であったが、入力と出力の系列の長さが可変長でない点やラベルとその組み合わせによって推論するコストが膨大である点が課題となっていた．一方 NN を基盤とした RNN (Recurrent Neural Network)[8] は，ある時系列 `t` における各入力 `x` とその出力 `y` を計算する隠れ層 `h` に対して `t-1` の入力を与える構造を持つ．CRF は与えられた系列全体に対して出力を求める反面，RNN では各時点 `t` に対して出力を求めるため，出力を可変長することが可能となる．さらに，初期の RNN では長期依存が必要なタスクにおいて勾配損失が起きていた問題を補う LSTM [6] モデルが提案された．LSTM は，隠れ層が出力ゲート，忘却ゲート，入力ゲートの 3 つの取捨選択機構を持っており (図 1 の (b) 参照)，各ゲートの出力によって隠れ層に与えられた状態を長期に保持するかどうかについて決定する．これによって長期に維持されることが望ましい入力が失われることなく学習を行うことが可能となる．さらに，現在までの入力から未来の出力を予測する LSTM に，未来の入力から過去の出力を予測する逆方向のレイヤーを加え、その出力を同一の出力層に統合する Bi-LSTM (Bidirectional LSTM) [9] (図 1 の (c) 参照) が登場し，NER などのタスクにおいて大きな成果を成し遂げている [10, 11]．特に，Lample らは Bi-LSTM と CRF を組み合わせた Bi-LSTM-CRF モデル (図 2 参照) を提案し，Bi-LSTM によって得られた各ノード間でも文脈と，CRF による全体のラベルに依存性を応用することで，高精度を実現している．



(a) CRF (b) LSTM (c) Bi-LSTM

図 1 CRF, LSTM, Bidirectional LSTM の構造 [10]

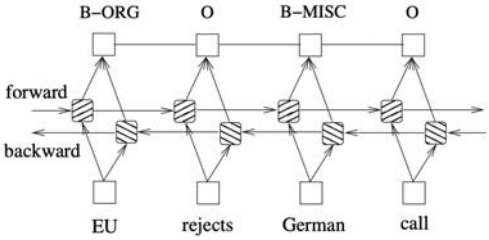


図 2 Huang らの Bidirectional LSTM CRF の構造 [10]

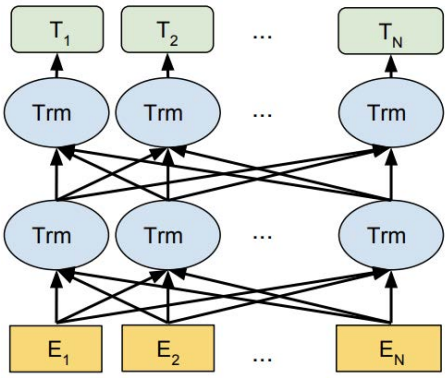


図 3 BERT の構造 [7]

また，Devlin らは，隠れ層の各内部状態について注意 (Attention) を表現したレイヤーによって構成された Transformer [12] を双方向で組み合わせ大規模データによって事前学習された BERT [7] を提案した (図 3)．BERT に対して事前学習されたモデルに新しいタスクのためのレイヤーを追加し，一部の既存レイヤーを再学習する Fine-Tuning を行うことで，NLP における多くのタスクに対して既存手法より高い精度を達成されている．

次章では，コミックにおける NER モデルを実現するために，LSTM や Bi-LSTM，さらにそれらのモデルにおける Embedding レイヤーを BERT に置き換えたモデルを検討する．

3 実験

本実験では，コミックにける NER タスクについて，既存の CRF モデルに基づく手法の他に，LSTM 及び Bi-LSTM-CRF に基づくモデルを作成し，精度や汎用性について検討を行う．また，入力データに対して BERT によって得られる特徴量を空間ベクトルとして用いることで Word-Embedding (単語埋め込み) を行い，LSTM や Bi-LSTM-CRF の入力として活用する．これによって，小規模のデータのみの学習であっても，大規模のコーパスによって学習されたモデルによる特徴量を活用するこで，精度の向上の期待できる．BERT は多言語モデルも公開されているが，1,800 万文の日本語のコーパスのみによって学習されたモデルが京都大学の黒橋研究室[2]によって公開されている．本実験に用いるコーパスは全て日本語で作成されているため，黒橋らのモデルを用いることが望ましいと考えられる．実験に用いる各モデルのレイヤー構造を図 4 に示す．

3.1 データ

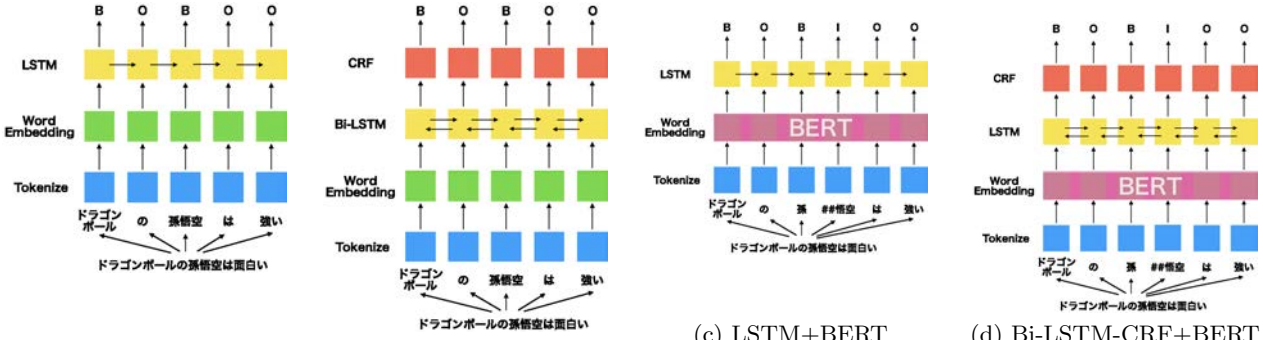
本実験のデータは，著者らの CRF に基づく NER モデルの作成時に用いられたコミックのレビュー 5,170 文を用いる．これらのコーパスは，コミックのレビューを掲載している Web サイトの人気作品 20 作品の Wikipedia の記事から各作品の固有表現を収集し，レビュー文に登場する箇所に固有表現タグを付与したものである．固有表現タグは各トークンに対して,「始まり (Beginning)」,「中間 (Inside)」,「外部 (Outside)」の 3 種類の状態を表現した IOB フォーマットを用いた．学習には全体の 70% (3,618 文) を，評価には 30% (1,552 文) を用いた．

3.2 学習

上記のモデル設計とデータを用いて学習を行った．各モデルを学習するにあたり，コーパスの量が少ない点，単方向 LSTM に比べ Bi-LSTM-CRF は双方向 LSTM と CRF を組み合わせた複雑な構造を持つため学習ステップがより多く必要となる点を考慮し，LSTM ベースのモデルはエポック数を 10 に，Bi-LSTM-CRF ベースのモデルは 50 に設定した．各モデルの学習過程図 5 に示す．また，学習終了後にテストデータを用いて評価を行った際の各モデルの成果を表 1 に示す

学習の結果，実験モデルすべて Loss が 0.5 以下，精度も 97% 以上であることが確認された．また，LSTM

[2] http://nlp.ist.i.kyoto-u.ac.jp/



(a) LSTM (b) Bi-LSTM-CRF (c) LSTM+BERT (d) Bi-LSTM-CRF+BERT

図 4 実験に用いられる各モデルの構造

表 1 各モデルの精度

	Accuracy
CRF[4]	52.6%
LSTM	97.87%
LSTM+BERT	97.63%
Bi-LSTM-CRF	**98.29%**
Bi-LSTM-CRF+BERT	97.02%

表 2 各モデルによるテスト結果 (1)

	Sentence (Color blocks means Named Entity)
Raw	ドラゴンボール はとても面白いです
LSTM	ドラゴンボール はとても面白いです
LSTM+BERT	ドラゴンボール はとても面白いです
Bi-LSTM-CRF	ドラゴンボール はとても面白いです
Bi-LSTM-CRF+BERT	ドラゴンボール はとても面白いです

表 3 各モデルによるテスト結果 (2)

	Sentence (Color blocks means Named Entity)
Raw	エドワード・エルリック はかっこいい
LSTM	エドワード・ エルリック はかっこいい
LSTM+BERT	エドワード・エルリック はかっこいい
Bi-LSTM-CRF	エドワード・ エルリック はかっこいい
Bi-LSTM-CRF+BERT	エドワード・エルリック はかっこいい

表 4 各モデルによるテスト結果 (3)

	Sentence (Color blocks means Named Entity)
Raw	モンキー・D・ルフィ の技は ゴムゴムの銃 である
LSTM	モンキー・D・ルフィの技はゴムゴムの銃である
LSTM+BERT	モンキー・D・ルフィの技は ゴムゴムの銃 である
Bi-LSTM-CRF	モンキー・D・ルフィの技はゴムゴムの銃である
Bi-LSTM-CRF+BERT	モンキー・ D ・ルフィの技はゴムゴムの銃である

や Bi-LSTM モデルよりも，各モデルに BERT を加えた際，Loss の減少がより早いことが確認された．しかし，テストデータによるモデルの精度は，BERT を加えていない状態のモデルが高いことも確認された．

3.3 追加実験

モデルの汎用性を確認するために，学習や評価時に用いられたコーパスに存在しない文章を任意に作成し，各モデルによって推論を行った．入力文章に対する各モデルの出力結果を表 2 から 5 に示す．各表では，入力文の固有表現の箇所と各モデルによって固有表現と認識された箇所に色を塗っている．実験に用いられた 4 文のうち先頭 3 文はコーパスに含まれた語彙を含めた文であり，最後の 1 文はコーパスの対象外のコミック作品の固有表現が含まれている．

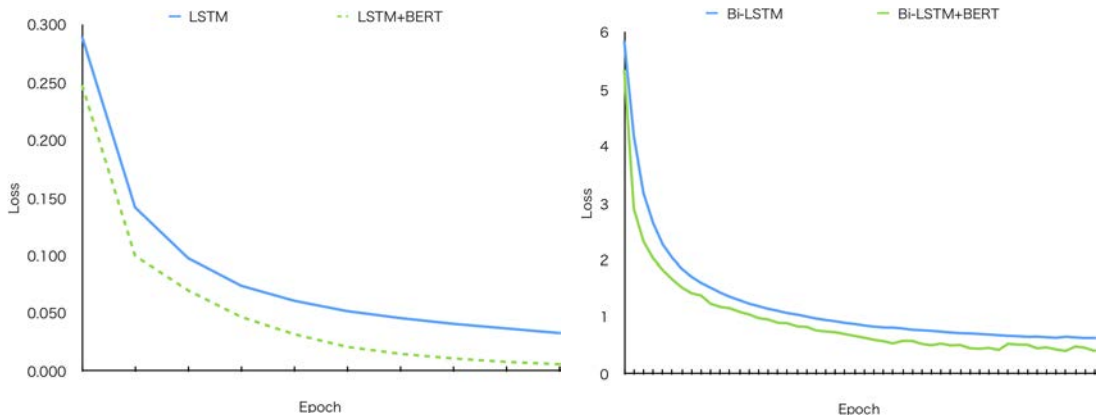
実験の結果，入力文に対して最も近い出力が得られたモデルは，LSTM+BERT であった．LSTM+BERT モデルは，他のモデルに比べほぼ全て固有表現を認識しており，その中でも学習時の語彙に含まれない固有表現を認識することができた．

4 議論

実験によって，既存手法よりも本稿で提案されたモデルは高精度であり，モデルにとって未知の固有表現を認識するケースも存在することが確認された．しかし，著者らは LSTM や Bi-LSTM に BERT により Word-Embedding を組み合わせることにより精度の向上を期待していたが，実際は BERT を組み込んでいないモデルの精度がより高いことも確認された．NLP タスクにおける最先端の成績を示す SOTA (State of the Art)[3] とこれまでのスコア上では，単方向の LSTM よりも Bi-LSTM-CRF や BERT がより優れた成績を残していることが確認できる．つまり，今回の実験結果は通説とは外れた状態であると言える．

これらの結果の原因として，下記の 2 点が予想されるまず 1 つ目の理由は，コーパスが不足している点である．今回モデルの学習に用いられたのは 3,618 文であり，NN を基盤とした学習時に用いられるコーパスの量として非常に小規模である．すでに大規模のコーパスによって学習された BERT による Word-Embedding によってこの問題が解決することを試みたものの，結果とし

[3] http://nlpprogress.com/english/named_entity_recognition.html



(a) LSTM, LSTM+BERT　(b) Bi-LSTM-CRF, Bi-LSTM-CRF+BERT

図 5 各モデルの学習グラフ (LSTM ベースの 場合 10 epoch まで，Bi-LSTM-CRF ベースの場合 50 epoch まで学習)

表 5 各モデルによるテスト結果 (4)

	Sentence (Color blocks means Named Entity)
Raw	怪盗ルパン 3 世は言いました
LSTM	怪盗ルパン 3 世は言いました
LSTM+BERT	怪盗ルパン 3 世は言いました
Bi-LSTM-CRF	怪盗ルパン 3 世は言いました
Bi-LSTM-CRF+BERT	怪盗ルパン 3 世は言いました

て目的は達成されなかったと言える．今後，コミックのレビューのみではなく，Wikipedia やニュース記事などを活用することで，コーパスを拡大していくことが必要であると考えられる．2 つ目の理由は，Fine-Tuning をしなかった点である．今回モデルの学習時には，BERT のパラメータの更新は行われていない．Fine-Tuning は事前学習されたモデルの一部のレイヤー (通常では出力層に近いレイヤー) のパラメータを学習時に更新することで，より新しいタスクに適した出力が得られるとされる．今後は，BERT の既存のパラメータのままではなく，Fine-Tuning を行うことで精度の向上を試みる．

5　おわりに

本稿ではコミックの NLP タスクの中でも NER に対して LSTM や Bi-LSTM-CRF，BERT といった最先端の深層学習技術の応用を試みた．実験の結果，既存手法を大きく上回る精度 (98.29%) と汎用性を持ち合わせたモデルが作成することができた．しかし，本来精度や汎用性の向上に貢献すると期待していた BERT は，一部の結果において本来の意図とは異なる働きをしていることが確認された．今後は，深層学習を応用するにあたり十分なコーパスの確保と，Fine-Tuning などの小規模のデータセットに適した学習手法の適用について検討していきたい．

参考文献

[1] 池田　郁，土方嘉徳，岩井秀成，西田正吾： ネタバレを見ずにレビュー文を閲覧できるシステムの開発，第 19 回インタラクティブシステムとソフトウェアに関するワークショップ論文集，pp. 36–41 (2011).

[2] 前田恭佑，土方嘉徳，中村聡史： ストーリー文書内のネタバレの記述に関する調査とレビュー文書でのネタバレ検出の試み，第 8 回 Web とデータベースに関するフォーラム論文集，pp. 32–39 (2015).

[3] 山下　諒，朴　炳宣，松下光範： コミックの内容情報に基づいた探索的な情報アクセスの支援，人工知能学会論文誌，Vol. 32, No. 1, pp. 1–11 (2017).

[4] 朴　炳宣，松下光範： レビューを用いたコミックの特徴抽出における固有表現の影響に関する調査，第 31 回人工知能学会論文集，pp. 4F1–2in2 (2017).

[5] Lafferty, J., McCallum, A. and Pereira, F. C.: Conditional random fields: Probabilistic Models for Segmenting and Labeling Sequence Data, *Proc. 18th International Conference on Machine Learning*, pp. 282–289 (2001).

[6] Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, *Journal of Neural Computation*, Vol. 9, No. 8, pp. 1735–1780 (1997).

[7] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (2018).

[8] Elman, J. L.: Finding Structure in Time, *Cognitive Science*, Vol. 14, No. 2, pp. 179–211 (1990).

[9] Graves, A. and Schmidhuber, J.: Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures, *Neural networks*, Vol. 18, No. 5-6, pp. 602–610 (2005).

[10] Huang, Z., Xu, W. and Yu, K.: Bidirectional LSTM-CRF Models for Sequence Tagging (2015).

[11] Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K. and Dyer, C.: Neural Architectures for Named Entity Recognition, *Proc. Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 260–270 (2016).

[12] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N. and Łukasz Kaiser: Attention Is All You Need (2017).