

# BERT-CRF を用いた日本語機能表現意味ラベル付与

田所 佑一 平 博順  
大阪工業大学大学院

{m1m22a21,hirotoshi.taira}@oit.ac.jp

## 概要

複数の形態素から構成され、一つの機能語として働く表現である機能表現は、モダリティ解析などの応用タスクでも利用されており、機能表現解析の精度向上は、重要な課題である。本研究では、解析モデルとして、BERT-CRFを用いることで、日本語機能表現意味ラベル付与の精度向上を試みた。評価実験の結果、BERT-CRFは既存モデルと比較して、高い精度を達成し、機能表現意味ラベル付与においても有効であることを確認した。

## 1 はじめに

複数の形態素から構成され、一つの機能語として働く表現は、「機能表現」と呼ばれ、文の意味に大きく関わっている。機能表現を正確に抽出し、意味ラベルの付与を行うことは、事実性解析やモダリティ解析などの応用タスクでも重要である。佐野ら [1] は機能表現を考慮したモダリティを用いて、文末述語を用いた2文間の接続関係推定を行っている。亀甲ら [2] は、将棋の解説文生成において、機能表現を考慮した、モダリティ情報のアノテーションを行っている。また、友利ら [3] は、モダリティ表現認識と事象の事実性解析のマルチタスク学習を行い、モダリティ情報の解析を行っている。

一方、様々な自然言語処理のタスクで、BERT [4] などの、大規模汎用言語モデルを利用した手法が提案されており、系列ラベリングタスクにおいても、優れた性能を達成している。田川ら [5] は、医療分野における日本語固有表現認識において、BERT-CRF [6] を用いた系列ラベリングタスクを行い、優れた性能が得られることを報告している。

本研究では、系列ラベリングタスクにおいて、大規模汎用言語モデルを用いた手法が有効であることに注目し、系列ラベリングタスクの1つと見なすことができる、機能表現意味ラベル付与においても、BERT-CRFが有効であるかを検証する。これまで、

日本語機能表現に対して系列ラベリングを行っている研究としては、CRFを用いた手法 [7] などがある。このCRFをベースラインモデルとし、事前学習済みのBERTモデルに対して、ファインチューニングを行うことで、機能表現意味ラベル付与を行った。CRF、BERT、BERT-CRFのそれぞれに対して実験を行い、性能の比較を行った。

## 2 関連研究

本節では、日本語の機能表現に関する研究について述べる。吉松ら [8] は、日本語機能表現の網羅的な辞書である「つつじ」を作成している。また、土屋ら [9] は、機能表現の検出をチャンク同定問題みなし、機械学習を用いることで高い精度が得られることを示している。機能表現に対するラベリングの研究としては、今村ら [10] が、機能表現辞書「つつじ」と最尤パス探索を用いて、述部に対する機能表現にラベリングを行っている。上岡ら [7] は、機能表現意味ラベルを付与したコーパスを作成し、そのコーパスに対して機能表現解析を行っている。文末述語における機能表現検出については、佐野ら [1] が、文末述語解析システム「Panzer」を開発している。

また、系列ラベリングに関する研究としては、友利ら [3] が、モダリティ表現認識のマルチタスク学習において、CRFのみを用いた手法に対して、エンコーダにBERT、出力層にCRF層を用いた手法が、有効であることを示している。田川ら [5] は、固有表現抽出タスクにおいて、CRFとBERT-CRFの比較を行い、日本語においてもBERT-CRFが有効であることを確認している。BERT-CRFは、BERTの出力層にCRFを繋げたものであり、ラベルの制約を学習することができるため、単純なBERTと比較して高い性能が期待できる。

## 3 BERT-CRF

### 3.1 CRF

CRF (Conditional Random Field) [11] は、系列ラベリングタスクを解くことができる識別モデルである。入力文の系列を  $X = (x_1, x_2, \dots, x_n)$ 、出力ラベルの系列を  $Y = (y_1, y_2, \dots, y_n)$  とした際に、 $X$  から  $Y$  の系列を予測するモデルであり、条件付き確率  $P(Y | X)$  は、以下の式で表すことができる。

$$P(Y | X) = \frac{1}{Z_X} \exp\left(\sum_{i=1}^N \sum_{k=1}^K \lambda_k f_k(y_{i-1}, y_i, X)\right) \quad (1)$$

ここで、 $f_k$  は素性関数、 $\lambda_k$  は素性関数  $f_k$  に対する重みである。また、 $Z_X$  は、全系列の確率和を 1 にするための正則化項であり、

$$Z_X = \sum_Y \exp\left(\sum_{i=1}^N \sum_{k=1}^K \lambda_k f_k(y_{i-1}, y_i, X)\right) \quad (2)$$

で表される。入力文の系列  $X$  に対する最適なラベル  $y^*$  は、以下の式で求めることができる。

$$y^* = \underset{Y}{\operatorname{argmax}} P(Y | X) \quad (3)$$

### 3.2 BERT

BERT [4] は、双方向エンコーダにより構成される事前学習済みモデルである。事前学習済みの BERT に対して、ファインチューニングを行い、系列ラベリングタスクを解けるように調整を行う。入力文の系列  $X = (x_1, x_2, \dots, x_n)$  を BERT に入力し、分散表現に変換する。得られた系列の分散表現を Softmax を用いることで各系列のラベルの予測スコアに変換する。予測スコアに対して、ビタビアルゴリズムを用いることで、最終的なラベルの系列  $Y = (y_1, y_2, \dots, y_n)$  を決定する。

### 3.3 BERT-CRF

BERT-CRF [6] は、BERT の出力層に CRF 層を組み合わせたモデルである。入力文の系列を  $X = (x_1, x_2, \dots, x_n)$ 、出力ラベルの系列を  $Y = (y_1, y_2, \dots, y_n)$  とすると、系列のスコアは以下の式で求めることができる。BERT の出力スコア  $P$  は、CRF 層に与えられる。ここで、 $A$  は遷移スコアを表す行列であり、 $A_{i,j}$  は行列  $A$  の  $(i, j)$  要素を表す。また、 $P_{i,y_i}$  は系列  $P$  の  $i$  番目の要素が、クラス  $y_i$  で

あるスコアを表す。

$$s(X, Y) = \sum_{i=0}^n A_{y_i, y_{i+1}} + \sum_{i=1}^n P_{i, y_i} \quad (4)$$

正しい出力ラベルの系列が得られるように、対数確率が最大となるように学習を行う。

$$\log(p(Y | X)) = s(X, Y) - \log\left(\sum_{\tilde{Y} \in Y_X} e^{s(X, \tilde{Y})}\right) \quad (5)$$

ここで  $Y_X$  は、ありうる全ての系列であり、ビタビアルゴリズムを用いて最終的なラベルの系列を求める。

## 4 評価実験

本節では、BERT-CRF を用いた日本語機能表現解析に対する評価実験を行った結果について述べる。

### 4.1 実験設定

評価実験には、「機能表現タグ付与コーパス」[12] (ver. 2.1a) の開発セットを用いた。このコーパスは、1,627 文からなり、文内のすべての機能表現に、IOB2 フォーマットで意味ラベルが付与されている。意味ラベルのクラス数としては、67 種類であり、特殊なラベルとして、意味ラベルチャンクの一つ前の系列に現れる C ラベルを含んでいる。また、C ラベルを除いた、意味ラベルが付与された機能表現の総数は、9,843 個である。

評価実験では、機能表現タグ付与コーパスを、訓練データ、検証データ、テストデータとしてそれぞれ 8:1:1 に分割して使用した。各手法で、実験条件を揃えるため、同一の分割されたデータセットを利用し、シャッフルなどは行っていない。また、CRF、BERT、BERT-CRF のそれぞれの実験で、訓練データとテストデータを揃えるために、CRF の実験の際には、検証データは利用していない。出力ラベルの形式としては IOB2 形式を使用している。

CRF の学習には、特徴量として、前後 2 トークンまでの表層形態素と品詞、品詞細分類を利用した。また、CRF の実装には CRFSuite [13] を用いた。BERT と BERT-CRF の実験には、事前学習済みモデルとして、東北大学の BERT-base を使用した。バッチサイズを 32、学習率を  $1 \times 10^{-5}$  として学習を行い、検証データを用いて評価し、最も精度が高かったエポックのモデルをテストで使用した。実験は、10 分割交差検定で評価を行った。

表1 実験結果

	CRF			BERT			B-CRF		
	Pre.	Rec.	F1	Pre.	Rec.	F1	Pre.	Rec.	F1
マイクロ平均	0.762	0.739	0.750	0.758	0.782	0.770	0.801	0.825	<b>0.813</b>
マクロ平均	0.603	0.552	0.555	0.583	0.592	0.566	0.663	0.668	<b>0.648</b>

表2 クラスごとの結果

ラベル	頻度	CRF	BERT	B-CRF	ラベル	頻度	CRF	BERT	B-CRF	ラベル	頻度	CRF	BERT	B-CRF
願望	68	0.785	0.773	<b>0.789</b>	理由	308	0.901	0.881	<b>0.907</b>	名詞化	154	0.738	0.797	<b>0.832</b>
順接確定	229	0.606	0.642	<b>0.680</b>	無意志	30	0.545	0.609	<b>0.635</b>	同時性	1	0.000	0.000	0.000
順接仮定	196	0.711	0.734	<b>0.765</b>	添加	41	0.543	0.391	<b>0.690</b>	可能	23	0.380	0.660	<b>0.690</b>
逆接確定	189	0.676	0.707	<b>0.738</b>	比較	15	0.089	0.183	<b>0.238</b>	受身	108	0.825	0.839	<b>0.882</b>
逆接仮定	48	0.514	<b>0.560</b>	0.555	比況	32	0.437	0.577	<b>0.639</b>	受益	16	0.463	0.354	<b>0.487</b>
起点	2	0.000	0.000	0.000	様態	11	0.071	0.000	<b>0.321</b>	反復	3	0.000	0.000	0.000
話題	78	0.502	0.570	<b>0.623</b>	方向	28	0.534	0.493	<b>0.636</b>	勧誘	11	0.222	0.000	<b>0.233</b>
試行	45	0.861	0.811	<b>0.865</b>	推量-高確実性	110	0.797	<b>0.842</b>	0.840	勧め	153	0.705	0.712	<b>0.761</b>
許可	26	0.258	0.404	<b>0.418</b>	推量-不確実	65	0.560	0.648	<b>0.656</b>	判断	1049	0.825	0.821	<b>0.873</b>
自発	2	0.000	0.000	0.000	授与	20	0.905	0.753	<b>0.923</b>	内容	199	0.744	0.767	<b>0.808</b>
自然発生	111	0.521	<b>0.685</b>	0.680	手段	12	0.167	0.000	<b>0.521</b>	依頼	275	0.830	0.845	<b>0.928</b>
習慣	72	0.344	0.308	<b>0.368</b>	態度	239	0.856	0.806	<b>0.859</b>	例示	46	0.514	0.551	<b>0.583</b>
継起	9	0.250	0.167	<b>0.528</b>	感嘆	5	0.000	0.000	0.000	使役	17	0.604	0.749	<b>0.795</b>
継続	97	0.446	0.529	<b>0.536</b>	意志	34	0.493	0.443	<b>0.564</b>	伝聞	58	0.713	0.703	<b>0.768</b>
結果状態	209	0.526	0.654	<b>0.681</b>	当為	15	<b>0.417</b>	0.259	<b>0.417</b>	付帯-続行	24	<b>0.552</b>	0.415	0.495
経歴	1	0.000	0.000	0.000	強調	6	0.000	0.000	0.000	付帯-並行	15	0.667	0.667	<b>0.724</b>
終点	5	0.333	0.000	<b>0.889</b>	対比	1	0.000	0.000	0.000	事前	4	0.000	0.000	<b>0.250</b>
程度	29	0.512	<b>0.713</b>	0.613	容易	25	0.664	0.749	<b>0.877</b>	並立	70	0.528	<b>0.640</b>	0.630
着継続	14	0.442	0.188	<b>0.570</b>	完了	377	0.892	0.917	<b>0.934</b>	不許可	6	0.467	0.467	<b>0.667</b>
目的	30	0.441	0.608	<b>0.769</b>	場合	2	0.000	0.000	0.000	不必要	7	0.200	0.000	<b>0.238</b>
発継続	3	0.000	0.000	<b>0.667</b>	困難	8	0.567	0.000	<b>0.960</b>	不可避	13	<b>0.867</b>	0.811	<b>0.867</b>
疑問	428	0.805	0.823	<b>0.869</b>	否定	440	0.845	0.897	<b>0.921</b>	不可能	4	0.000	0.000	0.000

## 4.2 評価手法

実験の評価には、適合率、再現率、F値を使用し、ラベルのクラスとスパンが共に一致した場合に、正解とした。Cクラスは、機能表現意味ラベルではないため、評価の際は除外した。

## 4.3 実験結果

CRF, BERT, BERT-CRFのそれぞれについて、10交差検定を用いた場合の結果を、表1に示す。実験の結果、BERT-CRFが、適合率、再現率、F1のそれぞれに対して、最も精度が高い結果となった。CRFと比較して、BERT, BERT-CRFではマイクロ平均とマクロ平均の両方で、F値の向上がみられた。CRFと比較し、BERTでは適合率が少し減少したものの、再現率が値が向上した。また、BERT-CRFでは適合率と再現率が、ともに向上しており、特に再現率

は、CRFと比較して大きく向上した。今回使用したデータセットは、クラスごとの要素数の偏りが大きい。そのため、マクロ平均でも評価したところ、BERT-CRFは、適合率、再現率、F1の全てにおいて大きく改善しており、本タスクにおいて最も精度の高いモデルであることがわかった。

10交差検定を実施した結果における、各クラスでの精度の平均値を表2に示す。各クラスの結果についても、ほとんどのクラスで、BERT-CRFを用いた場合に、精度が高かった。特に、クラスの要素数が多い場合には、CRFと比較してBERT, BERT-CRFでF値が改善する傾向にあった。一方、要素数が比較的少ない場合には、CRFと比較してBERTのF値が低くなる傾向があったが、BERT-CRFのF値は、高くなる傾向があった。

## 4.4 考察

CRFと比較して、BERTを使用した実験では、要素数が多いクラスで、F値が改善する傾向があった。逆に、要素数が少ないクラスでは、F値が悪化する傾向が見られた。BERTでは、チャンクに対するラベルをそれぞれ判定しているため、ラベルの遷移を考慮した推論を行うことが難しく、学習データ量が少ないクラスではF値が悪化したと考えられる。また、適合率が減少したものの、再現率は改善しており、CRFで検出できなかった意味ラベルが、検出できるようになったと考えられる。

BERT-CRFの特徴としては、多くのクラスでF値が、改善していたことがあげられる。BERTの傾向と同じく、要素数が多いクラスでは、F値が改善しており、また改善している項目も多かった。CRFと比較した際に、BERTでは、要素数が少ない場合に、F値が悪化する傾向がみられた。一方、BERT-CRFでは、「付帯-続行」を除き、全てのクラスでF値が改善していた。BERT-CRFではBERTの埋め込みを用いることで、CRFでは検出できなかった意味ラベルが検出できるようになった。また、CRF層を用いることで、ラベルの遷移を学習し、学習データが少ない場合にBERTで悪化していたクラスについても、CRFと同程度かそれ以上の精度を達成することができたと考えられる。

各手法に共通して、正しく抽出できなかったラベルの例として、同じ助動詞でも、他の形態素と連結することで、異なるラベルに変化するラベルなどがあった。その例を表3に示す。文末の「です」は「判断」としてラベルが付与されていることが多いが、「事も無いです」では、一つの機能表現として「不必要」のラベルが付与されている。そのため、各手法で「I-不必要」とラベリングするところを、「B-判断」としてラベリングしている。「判断」の学習データ量が多いのに対して、「不必要」のデータ量が少ないため、学習にバイアスが発生し、正しく抽出できていないと考えられる。

## 5 おわりに

本稿では、機能表現タグ付与コーパスに対してBERT-CRFを用いた機能表現意味ラベル抽出を行い、有効性を検証した。実験の結果、BERT、BERT-CRFを用いた手法は、CRFを用いた手法と比較して、有効であることが分かった。特に、

表3 ラベリングエラー例

形態素	Gold	CRF	BERT	BERT-CRF
これ	O	O	O	O
が	O	O	O	O
初めて	O	O	O	O
の	O	O	O	O
ドラクエ	O	O	O	O
で	B-逆接仮定	O	O	O
も	I-逆接仮定	O	O	O
問題	O	O	O	O
ない	O	O	O	O
です	B-判断	B-判断	B-判断	B-判断
。	O	O	O	O
心配	O	O	O	O
する	O	O	O	O
事	B-不必要	B-不必要	B-名詞化	B-不必要
も	I-不必要	I-不必要	O	I-不必要
ない	I-不必要	I-不必	O	I-不必要
です	I-不必要	B-判断	B-判断	B-判断
。	O	O	O	O

BERT-CRFでは、再現率が大きく改善するとともに、学習データ量が少ないクラスについても精度が改善し、より多くの意味ラベルを抽出できることが分かった。しかし、極端にデータ量が少ないクラスでは、精度が低くなる場合もあるため、意味的に近いクラスのデータを援用するなどの方法で、そのようなクラスの意味ラベル付与の精度を向上させることも今後、検討する予定である。

## 参考文献

- [1] 佐野正裕, 佐藤理史, 宮田玲. 文末述語における機能表現検出と文間接続関係推定への応用. 言語処理学会第26回年次大会発表論文集, pp. 1483–1486, 2020.
- [2] 亀甲博貴, 松吉俊, 牛久敦, 笹田鉄郎, 村脇有吾, 鶴岡慶雅, 森信介. 将棋解説文への固有表現・モダリティ情報アノテーション. 自然言語処理, Vol. 28, No. 3, pp. 847–873, 2021.
- [3] 友利涼, 村脇有吾, 松吉俊, 亀甲博貴, 森信介ほか. モダリティ表現認識・事象の事実性解析の同時学習. 研究報告自然言語処理 (NL), Vol. 2019, No. 13, pp. 1–8, 2019.
- [4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 4171–4186, Minneapo-



- lis, Minnesota, June 2019. Association for Computational Linguistics.
- [5] 田川裕輝, 西塾徹, 谷口友紀, 大熊智子. 生成された読影所見の自動評価に向けた固有表現認識とモダリティ推定. 言語処理学会 第 26 回年次大会, 2020.
  - [6] Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. Portuguese named entity recognition using bert-crf. **arXiv preprint arXiv:1909.10649**, 2019.
  - [7] Yudai Kamioka, Kazuya Narita, Junta Mizuno, Miwa Kanno, and Kentaro Inui. Semantic annotation of Japanese functional expressions and its impact on factuality analysis. In **Proceedings of the 9th Linguistic Annotation Workshop**, pp. 52–61, 2015.
  - [8] 松吉俊, 佐藤理史, 宇津呂武仁. 日本語機能表現辞書の編纂. 自然言語処理, Vol. 14, No. 5, pp. 123–146, 2007.
  - [9] 土屋雅稔, 注連隆夫, 高木俊宏, 内元清貴, 松吉俊, 宇津呂武仁, 佐藤理史, 中川聖一. 機械学習を用いた日本語機能表現のチャンキング. 自然言語処理, Vol. 14, No. 1, pp. 111–138, 2007.
  - [10] 今村賢治, 泉朋子, 菊井玄一郎, 佐藤理史. 述部機能表現の意味ラベルタガー. 言語処理学会第 17 回年次大会発表論文集, pp. 308–311, 2011.
  - [11] John Lafferty, Andrew McCallum, and Fernando Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.
  - [12] 機能表現タグ付与コーパス, (2022-12 閲覧). [https://www.cl.ecei.tohoku.ac.jp/Open\\_Resources-Japanese\\_FE\\_Corpus.html](https://www.cl.ecei.tohoku.ac.jp/Open_Resources-Japanese_FE_Corpus.html).
  - [13] Naoaki Okazaki. CRFsuite: a fast implementation of Conditional Random Fields (CRFs), 2007. <http://www.chokkan.org/software/crfsuite/>.