

深層学習を利用した PropBank 形式の 日本語意味役割付与モデル

タロック カラム¹ 竹内 孔一² バトラー アラスティア³

長崎 郁⁴ パルデシ プラシャント⁵

¹ 岡山大学大学院 ² 岡山大学学術研究院

³ 弘前大学 ⁴ 名古屋大学 ⁵ 国立国語研究所

poq93z0h@s.okayama-u.ac.jp

takeuc-k@okayama-u.ac.jp

概要

本論文では日本語に対して述語項構造シソーラスの概念フレームを基に PropBank 形式の意味役割を付与したタグ付きコーパス NPCMJ-PT に対して、複数の深層学習モデルを適用することで、意味役割ラベルとその範囲 (スパン) を予測する実験を行ったので報告する。深層学習で利用するモデルとして BERT の他にスパンを選択するためのモデルを導入する。また NPCMJ-PT のデータからは意味役割が付与された事例約 53,000 事例 (1 事例 1 述語の意味役割) を利用する。評価実験において BERT の利用層の異なりや特徴量を変えることで最も高い精度を出すモデルについて検討する。

1 はじめに

意味役割ラベル付与というタスクは文内の述語と係り関係にある句に対して意味的な関係を付与するタスクである。例えば図 1 では「壊す」の述語と係り関係にある「彼は」「木の扉を」「手で」3つの句に対して Arg0, Arg1, Arg-MNS のラベルを付与するタスクである。この Arg0 などのラベルはここでは Arg0 は動作主 (壊す人), Arg1 は壊される対象, ArgM-MNS は壊す際に使う手段を表しており、これは述語の概念フレームによって異なる。図 1 には概念フレームは示していないがここでは壊すは [破壊] という概念 (FID=277) である。これらは述語項構造シソーラス (PT)¹⁾ で公開している。

一方、国立国語研究所では Web 上で公開可能な

1) <https://pth.cl.cs.okayama-u.ac.jp/>. 「壊す」の他の概念フレームとして [身体部分の症状 (195)] 「異を壊す」や [消滅 (137)] 「計画を壊す」がある。それぞれ概念フレームによって意味役割の意味が異なる。

彼	は	木	の	扉	を	手	で	壊	し	た
Arg0		Arg1				ArgM-MNS	V	O		

図 1 「彼は木の扉を手で壊した」の意味役割ラベル

日本語テキストに対して構文木を付与したデータ NPCMJ を作成して公開している²⁾。NPCMJ に対して述語項構造シソーラスの体系を基に PropBank 形式の意味役割を付与したデータが NPCMJ-PT であり現在も付与を続けている。

そこで、本論文では NPCMJ-PT に付与されている意味役割データ (意味役割は 32 種類) に対して、深層学習モデルを適用することで係り関係にある句とその意味役割ラベルを推定する。

2 関連研究

意味役割付与の研究は主に英語の PropBank を利用したデータで研究されており、CoNLL2005 や CoNLL2012 のデータが利用されている [1, 2, 3]。上述の通り、意味役割は係り関係にある句を取り出す必要があるため、複数の単語列を文から取り出す必要がある。文献 [1] にあるように大きく分けて 3 種類の方法がある。1) IOB2 タグで意味役割のラベルと区間をモデル化する手法 (例えば [4])、2) スパンベースで解く方法 ([2, 3])、3) 各フレーズの主辞単語を同定してから解く方法 (例えば [5]) が提案されている。本論文では 2) のスパンを利用した手法 [3] を参考に日本語意味役割付与モデルについて検討する。

3 スパンに基づく意味役割付与モデル

2) <https://npcmj.ninjal.ac.jp/index.html>.

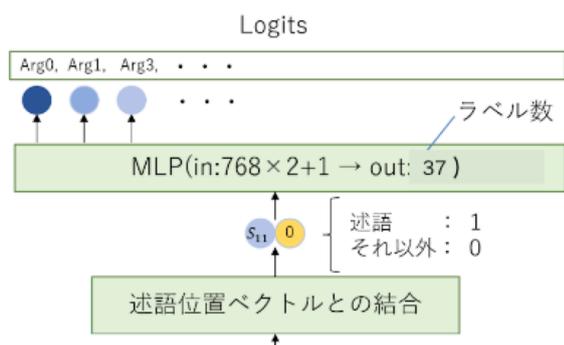


図2 スパンベクトル「彼は木の扉を手で壊した」の意味役割ラベル

3.1 スパン選択モデル

本論文ではスパン選択モデル [3] を基にした深層学習による意味役割付与モデルを構築する。まず入力文である日本語はすべてトークン化された状態で入力するとし、またどの述語に対する項 (述語と係り関係にある句) を取り出すか指定することとする。

スパンは文中の対象の述語に対する1つの項を表し、その最初と最後の形態素に対応したトークン番号を i, j とし、意味役割のラベルを l としたとき $\langle i, j, l \rangle$ で表す。各トークンに対する分散表現ベクトルを h_i, h_j とすると、最大のトークン列について全ての組 $[h_i, h_j]$ を作成する。このベクトル組を連結して1つのベクトルにしたものを本論文ではスパンベクトルと呼ぶ。

これらのスパンベクトルに対して意味役割が付与されているものには Arg0 や Arg1 など付与し、その他の部分について O や述語 V や N(Null) タグ、さらに意味役割の句よりも短いものは F(Fragment) などラベルを設定して、全スパンの組み合わせについてラベルを用意する³⁾。誤差学習の部分では、全スパンの組に対して正解ラベルとの誤差を計算して、クラス分類に基づく誤差でニューラルネットワークの重みを学習する。

次に、テストデータに対する意味役割付与において、各スパンベクトルの意味役割ラベルを推定した値 (Softmax や Logits を利用) を利用して値が大きいものから順に Greedy で意味役割のスパンとラベルを決定する。この処理を Decode と呼ぶ。この際、スパンが重ならないように選択する。また、テストデータに対しては学習中に利用した F や O タグは無視して、Arg のついたラベルのみでスパンを構成す

3) 意味役割 Arg は全部で 32 種類、その他 O, V, N と F-A, F-P を含めて学習時は 37 種のタグで学習する。

る。図2に最終層の部分の処理について記述する。スパンベクトルには指定する述語を含むかどうかを知らせる1次元のベクトルを連結している。

3.2 スパンベクトルの作成

スパンベクトル作成のために様々な深層学習モデルを利用する。まず文書は BERT を利用してトークン化を行い、各トークンに対する分散表現ベクトルとする。BERT は東北大が作成した Hugging Face の BERT (Small)⁴⁾ を利用する。この基本モデルを基に、複数のモデルを作成する。1) Base: BERT の最終出力層のベクトルを利用する。また Decode 時に softmax の値を利用する。2) Logits: Decode 時に logits の値を利用する。3) ExO: Decode 時に O タグを予測しない。4) DelO: 学習時に O ラベルを N に統合。5) AddF: 学習時に F ラベルを F-A0, F-A1 などより詳細に付与する。6) Enc4: Base モデルの BERT のエンコーダの学習層を最終4層に変更。これら6種類のモデルを作成して実験を行う。

3.3 実験

上記で定義したモデルに対して意味役割付与データ (NPCMJ-PT) で学習およびテストを実施する。これによりどのモデルが効果的であるか実験的に明らかにする。

3.4 実験設定

NPCMJ-PT から抽出した学習データは1述語に対して関係する意味役割が付与されている形式である。そのため同じ文であっても対象とする述語が異なれば別の1件としてデータが作成されている。1述語のデータの件数は5483件で、これらから意味役割の部分だけ取り出すと全部で53,867件である。これを学習、開発、テストデータとして8:1:1に分割して利用する。

評価としてスパンと意味役割タグが一致した部分のみ正解とする。また評価値としては適合率 (Precision), 再現率 (Recall), 調和平均 (F) を利用する。

3.5 実験結果と考察

上記のモデルについてテストデータに対して意味役割付与を行った結果を表1に実験結果を記述

4) <https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>.

表1 NPCMJ-PT に対する意味役割付与実験

モデル	Precision	Recall	F
Base	0.724	0.470	0.570
Logits	0.738	0.442	0.553
ExO	0.505	0.538	0.521
DelO	0.72	0.465	0.565
AddF	0.73	0.469	0.571
Enc4	0.702	0.558	0.622

する。

表1から、まずBaseモデルと比較してsoftmaxではなくlogitsを利用した場合はF値が下がったためsoftmaxを利用した方が良いことが分かる。また副次的に出力しているOタグやNタグ、Fタグであるが、基本的にはこれらは最終の出力としては利用されないタグではあるが、タグそのものは正解の意味役割ラベルから与えられているため、ある種の異なるタスクをモデルに解かせている状態と考えられる。ExOやDelOなどタグを削除する方向では精度が下がったため、逆にFタグを細かく設定して学習させたところ精度の向上が見られた。また、BaseにたいしてF値を上回ることができた。

また、一方でBERTから取り出すベクトルを変更したEnc4は大きく精度が向上し、提案したモデルの中で最も高い値をしめした。BERTの最終出力層だけでなく最終の4層部分を利用した方がよいことがわかる。

また表1の結果は英語のでCoNLLを利用した結果よりF値が低いことが分かる。しかしながら学習データが意味役割の項で約20万以上⁵⁾と大きいことから、学習データ量による差があると考えられる。

3.6 おわりに

日本語のPropBank形式の意味役割付与データであるNPCMJ-PTに対して、スパンに基づく意味役割付与モデルを構築して評価実験を実施した。その結果F値として約0.62の精度を得ることができた。意味役割の種類数は32である。英語の結果に比べてF値が劣る理由としてデータ量が挙げられるが、NPCMJ-PTも約5万例と少なくない意味役割データ量であるため、より外部知識や深層学習モデルを工夫することにより精度の向上を行う予定である。

謝辞

本研究の遂行にあたってJSPS 科研費JP22k00530の助成を受けた。

参考文献

- [1] Zhisong Zhang, Emma Strubell, and Eduard Hovy. Comparing span extraction methods for semantic role labeling. In **Proceedings of the 5th Workshop on Structured Prediction for NLP**, pp. 67–77, 2021.
- [2] Luheng He, Kenton Lee, Omer Levy, and Luke Zettlemoyer. Jointly Predicting Predicates and Arguments in Neural Semantic Role Labeling. In **Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics**, pp. 364–369, 2017.
- [3] Hiroki Ouchi, Hiroyuki Shindo, and Yuji Matsumoto. A Span Selection Model for Semantic Role Labeling. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing**, pp. 1630–1642, 2018.
- [4] Luheng He, Kenton Lee, Mike Lewis, and Luke Zettlemoyer. Deep Semantic Role Labeling: What Works and What’s Next. In **Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics**, pp. 473–483, 2017.
- [5] Zhisong Zhang, Xiang Kong, Zhengzhong Liu, Xuezhe Ma, and Eduard Hovy. A two-step approach for implicit event argument detection. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pp. 7479–7485, 2020.

5) <https://www.cs.upc.edu/~srlconll/st05/slides/intro.pdf>