

# ニューラルネットワークを利用した 日本語意味役割ラベル付与システムの構築

岡村 拓哉

岡山大学工学部

pxyl1pi0@s.okayama-u.ac.jp

竹内 孔一

岡山大学大学院自然科学研究科

koichi@cl.cs.okayama-u.ac.jp

石原 靖弘

岡山大学工学部

ishiharay@okayama-u.ac.jp

## 1 はじめに

本論文では「昨日/時間 太郎が/動作主 詰め将棋の本を/対象 買って来た」のように述語(「買う」)に対して係り元がどのような意味的な係り関係かを識別する意味役割付与システムに着目し、ニューラルネットワークを利用した手法について記述する。英語ではFrameNetの体系による意味役割付与システム [1] が提案されて以降、SemEval-2007<sup>1</sup>やCoNLL2009<sup>2</sup>などshared taskも開催され様々な手法が提案されている [2]。

日本語では格フレームを基にした京都大学テキストコーパス、NAISTテキストコーパスが提案され深層学習を利用した手法も提案されている [3, 5, 9]。一方で、日本語を対象に係り元に対して意味的關係を記述している言語資源データが存在する [6](GDAコーパス<sup>3</sup>、日本語フレームネット [4]、EDR<sup>4</sup>、述語項構造シソーラス<sup>5</sup>)。

そこで本研究では述語項構造シソーラスに着目し、現代日本語書き言葉均衡コーパス(BCCWJ)<sup>6</sup>に意味役割が付与されているデータBCCWJ-PT<sup>7</sup>を基に機械学習による意味役割付与手法の比較を行う。先行研究 [8]では述語の係り元(ここでは項と呼ぶ)の末尾表現(例えば、「を」や「ならば」など)を利用することが意味役割付与に有効であることを示した。本論文では従来の機械学習に比べてニューラルネットワークを利用したモデル化によりどの程度意味役割付与の精度が向上するか実験的に明らかにする。

実験の結果、SVMに比べて、ニューラルネットワークによる意味役割付与の精度が優れており、さらに、3層ニューラルネットワーク(3LNN)、畳込みニューラルネットワーク(CNN)、GRUを利用した場合、CNNが最も精度が高かったことを報告する。

## 2 意味役割付与タスク

BCCWJ-PTは述語項構造シソーラスの体系に基づいて、BCCWJ内の1文に対して1述語を対象に項を同定して、項全体に対して意味役割を付与している。

<sup>1</sup><http://nlp.cs.swarthmore.edu/semeval/>

<sup>2</sup><http://ufal.mff.cuni.cz/conll2009-st/trial-data.html>

<sup>3</sup><http://www.gsk.or.jp/catalog/gsk2009-b/>

<sup>4</sup><http://www2.nict.go.jp/ipp/EDR/ENG/indexTop.html>

<sup>5</sup><http://pth.cl.cs.okayama-u.ac.jp/>

<sup>6</sup><http://pj.ninjal.ac.jp/corpus.center/bccwj/doc.html>

<sup>7</sup><http://pth.cl.cs.okayama-u.ac.jp/>

図1の例では述語「買う」に対する項と意味役割が付

[対象 グランツーリスモ 4 を] [補語相当(を) 新品 で]  
[買う] と思っている のです が

図1: BCCWJ-PTの意味役割の例

与されている。意味役割の種類はデータ中で64種類定義されている。また語の単位はBCCWJにおける長単位で付与されており、品詞、活用形と項の範囲が付与されている<sup>8</sup>。

本研究では述語に対する意味役割ラベルの付与に焦点を絞ったタスクを設定する。つまり、係り受けなど項の同定は正しい情報を利用し、項の意味役割ラベルの推定を行う。さらに、タスクを簡素化し、文全体を入力とするのではなく述語と項の組合せのみを与え、意味役割を推定するタスクを設定する。よって幅広い

意味役割ラベル	入力
対象	買う グランツーリスモ 4 を
補語相当(を)	買う 新品 で

図2: 意味役割付与タスク

文脈情報なしで、述語と項の表現を入力として意味役割を予測する。

## 3 意味役割付与モデル

本研究では、SVMおよびニューラルネットワークを利用した意味役割付与モデルを構築する。特徴量として図2に示した形態素を基に、nwjc2vec[7]のfasttextを利用したskip-gram版(200次元)を利用する。それに合わせて上記の形態素の単位をUniDic MeCabで切り直した単位を利用する。下記にそれぞれのモデルについて説明する。

### SVMモデル

SVM(scikit-learnを利用)では入力部分の形態素のBOWを基本特徴量とする。拡張として各項の内容語の主辞<sup>9</sup>と述語のskip-gramを特徴量として追加す

<sup>8</sup>国語研よりdownload可能。 <https://bccwj-data.ninjal.ac.jp/mdl/>

<sup>9</sup>例えば図2の各項の主辞は「グランツーリスモ4」と「新品」である。

る。さらに、先行研究の知見 [8] から項の末尾 2 形態素を特徴量とする<sup>10</sup>(表 1)。64 種類の意味役割を識別

表 1: SVM モデルの特徴量

特徴量	説明
BOW	項と動詞の表層と基本形
BOW+skip	項の主辞と動詞の基本形の skip-gram を追加
BOW+skip+two	末尾 2 形態素のベクトルを追加

するために one-vs-rest 法を利用する。

### 3 層ニューラルネットワークモデル (3LNN)

中間層は 256 次元で活性化関数は ReLU を利用した。また学習には Adam を利用した。Adam のパラメータは  $\alpha = 0.001$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$  とした<sup>11</sup>。入力の特徴量として上記の SVM モデルと全く同じ表 1 の 3 種類の特徴量を利用したモデルを構築する。

### 畳み込みニューラルネットワークモデル (CNN)

入力として図 2 に示すように述語のあとに項を連結した形態素列を仮定し、これらの skip-gram を適用した入力に対して skip-gram と同次元 (200 次元) のフィルタを適用する。連続形態素長として、3, 4, 5 形態素のフィルタを用意し、各形態素長に対して 128 種類のフィルタを生成した。ネットワークは skip-gram で embedding した入力に対して畳み込み層、プーリング層、最終出力層を適用する。異なるモデルとして、最終出力層の手前に、BOW, skip-gram, 最終 2 形態素のベクトルを加えたモデルも作成した。

表 2: CNN モデル

モデル	説明
conv	畳み込み層+プーリング層のみ
conv+BOW+skip+two	BOW と項の主辞と動詞の基本形の skip-gram と末尾 2 形態素ベクトルを追加

### GRU モデル (GRU)

Gated Recurrent Unit を利用した意味役割付与モデルを構築する。図 2 の入力形態素列を時系列と見なして GRU を適用し、最終状態に対して出力層を結合して意味役割を識別させる。時系列の長さは 60 形態素を最大とする。時系列の順番の違いによる精度の異なりを見るために、表 3 に示す複数のモデルを設定した。モデル GRU3 と GRU4 では活用形の違いも考慮するように活用形も入力として利用する。上記全てのニューラルネットワークは Tensorflow を利用して構築した。

<sup>10</sup>例えば図 2 の各項では、それぞれ「4 を」と「新品 で」となる。

<sup>11</sup>以降のニューラルネットワークのモデルは全て Adam を利用する。

表 3: GRU モデルにおける時系列処理の異なり

モデル名	時系列の扱い
GRU1	グランツーリスモ 4 を 買う
GRU2	買う グランツーリスモ 4 を
GRU3	買った グランツーリスモ 4 を
GRU4	買った 買う グランツーリスモ 4 を

## 4 意味役割付与実験

### 4.1 実験設定

BCCWJ-PT から図 2 の形式に変換したデータを利用して前述のモデルを学習させる。この時、項の形態素が 60 以上のものはデータから排除した。SVM モデルの評価では 5 分割交差検定で精度を求める。一方、ニューラルネットワークを利用したモデルでは下記のように学習、開発、テストデータに分割する。学習データで

表 4: ニューラルネットワークで利用するデータの内訳

データの種類	データ数 (割合)
学習データ	6,753 (6.5 割)
開発データ	520 (0.5 割)
テストデータ	3,117 (3 割)

ニューラルネットワークを学習させ、開発データの精度を参考に最適な学習回数 (epoch) を決定して、その求めた重みでテストデータを解析する。学習回数を増やすと過学習で開発データに対して精度が悪くなる傾向が見られたので下記実験結果では全て 20 回で打ち切ることとした。

評価は全てのモデルは項に対して 1 つの意味役割だけ出力するのでテストデータに対する意味役割ラベルの一致率で精度を求める。表 5 に全データでの意味役

表 5: 意味役割上位 5 件

ラベル	事例数
対象	3100
動作主	1200
様態	549
修飾	443
副詞	439

割ラベルの内訳を示す。表 5 から [対象] と [動作主] が多く<sup>12</sup>、続いて、[様態] や [副詞] など述語を修飾する係り元が多いことが分かる。意味役割ラベルの [修飾] はいわゆる外の関係であり述語が連体修飾する場合も少なくないことがわかる。

<sup>12</sup>ここで意味役割ラベルは [] で示すことにする。また意味役割の具体的な事例は <http://pth.cl.cs.okayama-u.ac.jp/> を参照。

## 4.2 実験結果

SVMと3LNNの意味役割付与結果について表6に示す。表6の特徴量の違いによる精度を比較すると、

表 6: SVMと3LNNの意味役割付与精度

モデル	特徴量		
	BOW	BOW+skip	BOW+skip+two
SVM	0.508	0.562	0.598
3LNN	0.538	0.610	0.650

単純なBOWに比べて主辞と述語のskip-gramの特徴量を加えるとSVMでは6%程度、3LNNでは7%程度精度が向上した。さらに末尾の2形態素のベクトルを加えるとSVMで3.6%、3LNNで4%も向上した。このことから、意味役割ラベル付与において、これらの特徴量は有効に働くことが分かる。

次にモデルの違いによる精度の比較を行う。SVMは5分割交差検定なので、テストデータの量は全体の2割程度であるのに対して、3LNNでは3割がテストデータである。つまりSVMは残りの8割を全て学習データとして利用できる一方で、3LNNでは6.5割が利用できる学習データである。それにも関わらず表6は3LNNの精度が大きくSVMを上回ったことを示している。全ての特徴量を入れた場合には、特徴量が全く同じであるにもかかわらず3LNNがSVMに比べて5%以上高い精度を示した。先行研究[8]と比較した場合、データサイズが異なるため単純に比較できないが、先行研究の結果よりも精度は大きく上回っている。このことから、3LNNはSVMやCRFに比べて意味役割付与で有効なモデルであることがわかる。

次にCNNモデル結果を表7に示す。学習データお

表 7: CNNの意味役割付与精度

モデル	精度
conv	0.641
conv+BOW+skip+two	0.660

よびテストデータは3LNNと全く同じである。表6と比較するとCNNにおけるconvモデルでは3LNNのBOW+skip+twoモデルの精度には及ばないものの、高い精度を示している。convモデルでは入力形態素列は全てskip-gramに変換された後、畳込み層によるフィルタが適用されているので、skip-gramも利用している。しかしながら、最終出力層にはプーリングした結果のみを伝えているため抽象化した情報のみを利用して識別している。それでいて部分的にskip-gramを利用した3LNN(BOW+skip)モデルを大きく上回る精度を得ていることから、畳込み層により、形態素連続列の中から意味役割の識別に有効な特徴量が獲得できていることが窺える。さらに、CNNのconv+BOW+skip+twoモデルでは3LNNモデルの精度を1%上回った結果が示されている。このことから畳込みにより得られている特徴量はBOW+skip+two

で得られる情報と異なっており両方利用することでさらに精度が向上することがわかる。

次に、GRUであるがGRUでは他のモデルと異なり、BOWではなく時系列として入力を扱うため、入力順序の異なりによる意味役割付与の精度への影響を調べた。GRUの意味役割付与精度の結果を表8に

表 8: GRUの意味役割付与精度

モデル名	精度
GRU1	0.599
GRU2	0.633
GRU3	0.619
GRU4	0.631

示す。大きな特徴としては、述語を時系列の最後に置くGRU1が最も精度が低く、述語を先頭に持ってきたGRU2と比較すると3%以上の精度の差が開くことが示されている。これは既に先行研究[8]で指摘しているように、項の末尾表現が意味役割の識別に強く関与するため時系列の最後に識別に有効な特徴量が来ると意味役割付与精度が向上することがわかる。

一方で、3LNNの(BOW+skip+two)モデルやCNNと比較すると識別精度は高くない。特に入力はskip-gramなのでCNN(conv)モデルと入力は同じである。よって、この実験結果から、形態素の連続列からあるラベルを1つ決めるタスクでは単純なりカレントニューラルネットワークよりも畳込みが有効に働くことが推察される。

## 5 おわりに

本論文では日本語の意味役割ラベルを付与する手法として、従来の機械学習手法とニューラルネットワークによる手法を比較し、ニューラルネットワークが高い精度を示すことを実験的に明らかにした。さらにニューラルネットワークの中でも畳込み層によるフィルタを適用したモデルが最も高い精度を示すことを明らかにした。今後、ネットワーク構成を洗練してより高い精度を示す意味役割ラベル付与手法を検討したい。

## 参考文献

- [1] Daniel Gildea and Daniel Jurafsky. Automatic Labeling of Semantic Roles. *Computational Linguistics*, Vol. 28, No. 3, pp. 1–45, 2002.
- [2] Martha Palmer, Daniel Gildea, and Nianwen Xue. *Semantic Role Labeling*. Morgan & Claypool Publishers, 2010.
- [3] Tomohide Shibata and Daisuke Kawahara and Sadao Kurohashi. Neural network-based model for Japanese predicate argument structure analysis. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 1235–1244, 2016.

- [4] 小原京子. 日本語フレームネットの構文アノテーション. 言語処理学会第 20 回年次大会発表論文集, pp. 161–164, 2014.
- [5] 大内啓樹, 進藤裕之, 松本裕治. 深層リカレントニューラルネットワークを用いた日本語述語項構造解析. 情報処理学会, 自然言語処理研究会, Vol2016-NL-229, pp. 1–8, 2016.
- [6] 竹内孔一. 語彙概念と語彙概念構造. 松本裕治, 奥村学 (編), コーパスと自然言語処理, pp. 94–113. 朝倉書店, 2017.
- [7] 浅原正幸, 岡照晃. nwjc2vec: 『国語研日本語ウェブコーパス』に基づく単語の分散表現データ. 言語処理学会第 23 回年次大会, pp. 94–97, 2017.
- [8] 石原靖弘, 竹内孔一. 係り元の末尾表現に着目した Hierarchical Tag Context Tree を利用した日本語意味役割割付与システムの構築. 情報処理学会論文誌, Vol. 57, pp. 1611–1626, 2016.
- [9] 松林優一郎, 乾健太郎. ニューラルネットワークによる日本語述語項構造解析の素性の汎化. 言語処理学会第 23 回年次大会, pp. 394–397, 2017.