

深層距離学習を用いた日本語動詞の意味フレーム推定

矢野 千紘¹ 山田 康輔¹ 笹野 遼平^{1,2} 武田 浩一¹
¹ 名古屋大学 ² 理化学研究所
 {yano.chihiro.j3,yamada.kosuke.v1}@s.mail.nagoya-u.ac.jp
 {sasano,takedasu}@i.nagoya-u.ac.jp

概要

本研究では、深層距離学習でファインチューニングした BERT を用いた日本語動詞の意味フレーム推定に取り組む。英語では人手で構築された大規模なフレームリソースである FrameNet が存在する一方で、日本語では教師データとして利用可能なリソースが日本語フレームネットなど比較的小規模なものに限られている。そのため本稿では、少量だが評価言語で構築された日本語フレームネットを教師信号として BERT をファインチューニングする手法と、言語間での知識の転移を期待して、他言語だが規模の大きい FrameNet を教師信号として多言語 BERT をファインチューニングする手法を実施し、各手法による意味フレーム推定実験の結果を報告する。

1 はじめに

人は言葉を理解するにあたり、経験から得た語の背景知識を利用しているとされる [1]。意味フレーム知識とはそれらの背景知識をまとめた知識リソースのことであり、英語については、大規模なフレーム知識リソースが人手で整備されている [2]。また、大規模なテキストコーパスから意味フレーム知識を自動構築することを目的とし、動詞を喚起する意味フレームごとにまとめる **意味フレーム推定** の研究も行われており、近年は、BERT [3] などの事前学習済み言語モデルから得られる文脈化単語埋め込みを利用した手法が多く提案されている [4, 5, 6]。中でも、山田らの一連の研究 [7, 8] では、動詞ごとにその用例をクラスタリングした後に動詞横断的なクラスタリングを行うという 2 段階クラスタリングを提案し、さらに、人手で整備されたフレーム知識を教師データとして深層距離学習を用いて事前学習済み言語モデルをファインチューニングすることで、高い精度で意味フレーム推定が可能であることが報告されている。

しかし、英語以外の言語において人手で整備されたフレーム知識リソースの規模は限定的であり、自動構築を行う研究もほとんど行われていない。そこで本研究では、日本語動詞の意味フレーム推定を目的とし、山田らの手法 [8] を日本語動詞に適用する。日本語への適用に際し、日本語のフレーム知識リソースは小規模であることから十分なデータを学習できず、英語と同様の手法を用いても高い性能は得られない可能性がある。したがって、日本語のフレーム知識リソースを教師信号としてファインチューニングした BERT を用いる手法に加え、英語のフレーム知識リソースを教師信号としてファインチューニングした多言語 BERT を用いる手法でも意味フレーム推定を行う。

2 日本語動詞の意味フレーム推定

本節では、山田らが提案した英語動詞の意味フレーム推定手法 [8] を紹介した後、日本語への適用方法について検討する。

2.1 英語動詞の意味フレーム推定手法

山田らはフレーム推定対象動詞の埋め込みをクラスタリングすることでフレーム推定を行った。その際、より高精度なフレーム推定を実現するため、以下の 3 つの工夫を取り入れている。

まず、動詞の表層が持つ情報による影響が過度に強くなることを防ぐため、対象動詞の BERT による埋め込み v_{word} をそのまま用いるのではなく、その動詞をマスクトークンに置き換えたときの埋め込み v_{mask} との加重平均による埋め込み v_{w+m} を対象動詞の埋め込みとして用いている。この埋め込みは α を重みとした場合、式 (1) で定義される。

$$v_{w+m} = (1 - \alpha) \cdot v_{word} + \alpha \cdot v_{mask} \quad (1)$$

また、同じ動詞の用例が過度に多くのクラスタへ分散してしまうことを防ぐため、一度にすべての埋

め込みをクラスタリングする1段階クラスタリング手法に加えて、動詞ごとに用例のクラスタリングを行った後に動詞横断的にクラスタリングを行う2段階クラスタリング手法を検証している。

さらに、類似した意味フレームの事例が近くに位置し、異なる意味フレームの事例が遠くに位置する埋め込み表現となるように、人手で整備されたフレームリソースを教師データとし、深層距離学習を用いてBERTのファインチューニングを行っている。深層距離学習の手法としては、Contrastive 損失 [9], Triplet 損失 [10], ArcFace 損失 [11], AdaCos 損失 [12] に基づく手法を比較している。

2.2 日本語への適用

山田らの手法では英語動詞の意味フレーム推定に対し、同言語である英語リソースを用いた学習を行っているが、日本語のフレーム知識リソースは英語のリソースに比べ規模の面で大きく劣っており、山田らの学習方法が最適であるとは限らない。例えば日本語フレームネット [13] は FrameNet [2] と比較して3.2%程度の用例しか含まない。そのため、事前学習済み言語モデルおよび教師信号として何を用いるべきか検討が必要となる。

事前学習済み言語モデルとしては、大きく分けて日本語単言語モデルと多言語モデルの2つが考えられる。同程度の規模であるなら単言語モデルの方が対象言語の言語モデルとして高性能である可能性が高いが、フレームの概念は言語間で共通する部分があるため、多言語モデルも有用な可能性がある。

教師信号としては日本語フレームネットを用いることで、日本語のフレームに関する直観をそのまま取り入れることが可能となる一方、事前学習済み言語モデルとして多言語モデルを採用した場合は、FrameNetを教師信号とすることで、大規模なフレームリソースが活用できると考えられる。

そこで本研究では以下の4設定を提案する。

S_{JFN} 単言語モデルまたは多言語モデルに対し、日本語フレームネット (JFN) を教師信号とした深層距離学習を実施

S_{FN} 多言語モデルに対し、FrameNet (FN) を教師信号とした深層距離学習を実施

$S_{FN \rightarrow JFN}$ 多言語モデルに対し、FN, JFN を順に教師信号とした深層距離学習を実施

S_{FN+JFN} 多言語モデルに対し、FN, JFN をまとめて教師信号とした深層距離学習を実施

3 実験

提案する学習手法の有効性を検証するため、日本語動詞のフレーム推定実験を行う。

3.1 実験設定

損失 深層距離学習で用いる損失には、山田らによる英語動詞のフレーム推定実験 [8] において高い性能を示した Triplet 損失と AdaCos 損失を用いる。Triplet 損失は距離ベースの損失であり、事例の3つ組に対して、アンカーと負例の距離を、アンカーと正例の距離より一定の-margin以上遠くなるように学習する。AdaCos 損失は事例とクラスの距離を近づける学習を行う ArcFace 損失に対し、性能がハイパーパラメーターに依存しないよう-marginを省きスケールを動的に決定するようにした損失である。

データセット FrameNet [2], 日本語フレームネット [13] から、動詞がフレームを喚起する用例を抽出し、データセットを作成した。このうち、FrameNetから抽出された用例数は82,610であった。評価に利用する日本語フレームネットについては、多義動詞の割合が一定となるように動詞単位で、学習データ、開発データ、評価データの3つのサブセットに分割し、3分割交差検証を行った。各サブセットの用例数の平均は786であった。

事前学習済みモデル 単言語モデルとして東北大BERT¹⁾を、多言語モデルとして multilingual BERT(mBERT) [3]²⁾を用いた。

比較手法 東北大BERTに対しては、深層距離学習を行わない設定 (w/oFT) と S_{JFN} の2つの設定を比較する。mBERTに対してはさらに S_{FN} , $S_{FN \rightarrow JFN}$, S_{FN+JFN} を加えた5つの設定を比較する。予備実験の結果に基づき、JFNのみを学習に用いる場合は10 epoch, FNを用いる場合は3 epoch 学習させる。

パラメータ 埋め込み v_{w+m} の重み α , クラスタ数, marginは開発セットを用いて決定する。このうち、 α は0から1まで0.1刻み, Triplet 損失において必要となるmarginは, $\{0.1, 0.2, 0.5, 1.0\}$ から探索する。バッチサイズは32, 学習率は $1e-5$ とし, 最適化アルゴリズムには AdamW [14] を使用する。

クラスタリング手法 クラスタリングは、1段階クラスタリングでは群平均法を、2段階クラスタリングでは1段階目にX-means [15], 2段階目に群平

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

2) <https://huggingface.co/bert-base-multilingual-uncased>

表 1 日本語動詞を対象としたフレーム推定実験の結果. スコアはいずれも 100 倍したものを記載している.

事前学習済みモデル	設定	学習手法	1 段階クラスタリング			2 段階クラスタリング		
			BcP	BcR	BcF	BcP	BcR	BcF
東北大 BERT	w/oFT	-	61.3	61.4	<u>61.2</u>	47.3	72.2	57.1
	S_{JFN}	Triplet	65.1 \pm 1.4	67.4 \pm 1.5	65.9 \pm 1.2	65.5 \pm 0.6	64.5 \pm 1.9	64.6 \pm 1.3
		AdaCos	63.5 \pm 0.3	66.4 \pm 0.2	64.6 \pm 0.2	47.9 \pm 0.4	74.0 \pm 0.4	58.1 \pm 0.3
mBERT	w/oFT	-	61.5	57.9	<u>59.3</u>	46.8	71.5	56.6
	S_{JFN}	Triplet	62.6 \pm 0.4	58.3 \pm 0.8	60.1 \pm 0.6	60.3 \pm 5.5	62.7 \pm 6.0	60.8 \pm 0.6
		AdaCos	58.8 \pm 1.1	60.8 \pm 2.0	59.3 \pm 0.3	55.0 \pm 1.3	66.4 \pm 1.6	60.1 \pm 0.5
	S_{FN}	Triplet	61.6 \pm 0.8	58.1 \pm 0.3	<u>59.4</u> \pm 0.3	57.3 \pm 0.2	60.1 \pm 0.8	58.4 \pm 0.5
		AdaCos	61.0 \pm 0.5	57.6 \pm 1.0	59.0 \pm 0.8	53.4 \pm 2.7	66.6 \pm 2.0	59.1 \pm 1.0
	$S_{FN \rightarrow JFN}$	Triplet	61.7 \pm 0.6	57.9 \pm 1.3	59.4 \pm 1.0	59.0 \pm 2.9	62.3 \pm 2.0	60.4 \pm 1.1
		AdaCos	59.4 \pm 0.1	61.9 \pm 1.0	60.1 \pm 0.5	54.6 \pm 1.3	67.2 \pm 1.6	60.1 \pm 1.6
	S_{FN+JFN}	Triplet	58.2 \pm 1.4	57.7 \pm 0.8	57.5 \pm 1.1	57.8 \pm 1.2	63.1 \pm 0.6	60.3 \pm 0.4
		AdaCos	61.7 \pm 1.1	61.2 \pm 1.4	61.2 \pm 1.2	58.2 \pm 0.8	69.0 \pm 2.3	62.8 \pm 0.5

均法を用いて実施する.

評価方法 評価指標には, クラスタリングの代表的な評価指標の 1 つである B-cubed Precision (BcP), Recall (BcR), およびその調和平均である F 値 (BcF) [16] を使用する. 全ての設定において, 異なるランダムシードで 3 回実験を行い, 評価値の平均と標準偏差を算出する.

3.2 実験結果

表 1 に実験結果を示す. まず, 事前学習済み言語モデルをそのまま用いた設定 (w/oFT) の性能を比較すると, mBERT を使った場合よりも東北大 BERT を用いた場合, 2 段階クラスタリングよりも 1 段階クラスタリングを用いた場合の方がそれぞれ高い BcF となった. 東北大 BERT と 1 段階クラスタリングを組み合わせた場合の BcF は 61.2 であった.

教師信号として日本語フレームネットを使った設定 (S_{JFN}) に注目すると, ほとんどの設定において w/oFT より高い BcF を達成しており, 深層距離学習の有効性が確認できる. 東北大 BERT を使ったモデルと mBERT を使ったモデルを比較すると, AdaCos と 2 段階クラスタリングの組み合わせ以外では, 東北大 BERT の方が大幅に高い BcF を達成しており, 単言語に閉じた実験の場合, 言語モデルも単言語モデルの方が高い性能となる傾向があると言える. この設定でもっとも高い BcF となったのは東北大 BERT と 1 段階クラスタリングを組み合わせた場合で, 全実験でもっとも高い 65.9 であった.

続いて, 教師信号として FrameNet を使った設定

(S_{FN}) に注目すると, 深層距離学習を実施しなかった mBERT を使う設定 (w/oFT) と比べ, 2 段階クラスタリングに基づく手法において僅かに性能向上が見られたものの, 各モデルで最高性能となる設定の BcF はそれぞれ 59.4, 59.3 と大きな差異は確認できなかった. 教師信号として FrameNet, 日本語フレームネットの両方を使った設定 ($S_{FN \rightarrow JFN}$, S_{FN+JFN}) を日本語フレームネットのみ使った設定 (S_{JFN}) と比較すると, もっとも高い BcF である 62.8 を達成したのは S_{FN+JFN} と 2 段階クラスタリングを組み合わせたときであったが, 性能が低下しているケースもあり, FrameNet を使うことの有効性は本実験では確認できなかった.

3.3 考察

図 1 に, 日本語フレームネットを教師信号とし (S_{JFN}), 東北大 BERT に対し深層距離学習によるファインチューニングを行った場合の, 学習前と学習後のフレーム推定対象動詞の埋め込みの変化の一例を示す. 埋め込みは t-SNE [17] を用いて 2 次元に圧縮して可視化し, 事例数の多い上位 10 フレームを喚起する動詞の埋め込みに対応する色を付与している. 学習前の段階でもある程度, 喚起するフレームごとにまとまって分布しているものの, 深層距離学習を行うことでその傾向が強くなるのが確認できる. たとえば, **Statement** フレームや **Change_position_on_a_scale** フレームは学習前は, 他のフレームを喚起する用例と近接して分布しているのに対し, 学習後は該当のフレームを喚起する用例だ

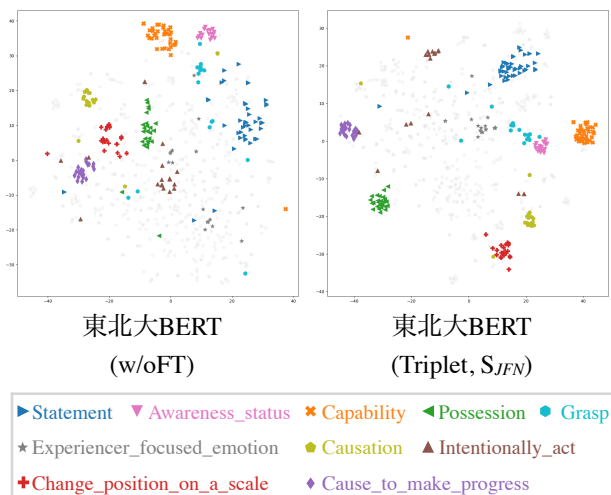


図1 深層距離学習前後の東北大BERTの分布の変化例

けでまとまって分布していることが確認できる。

次に、図2にFrameNetと日本語フレームネットをまとめて教師信号とし(S_{FN+JFN})、mBERTに対し深層距離学習を行った場合の、学習前と学習後のフレーム推定対象動詞の埋め込みの変化の一例を、実際の用例とともに示す。同じフレームに属する用例ペア(1)と(2)、(5)と(6)は、学習前は空間上の離れた場所に位置しているのに対し、学習後は近くに位置することが確認できる。このうち、用例ペア(1)と(2)で喚起語となる動詞は共に「出来る」であるが、学習前は用例(1)のように未然形・連用形で出現した用例群と、用例(2)のように終止形・連体形の一部として出現した用例群とに分かれて分布していた。また、用例ペア(5)と(6)で喚起語となる動詞は共に「持つ」であるが、これらの埋め込みは学習前において、平仮名の「もつ」を含む用例群と漢字の「持つ」を含む用例群に分かれて分布していた。

一方、用例ペア(7)と(8)の埋め込みは接近せず、学習後も異なるクラスターに属した。これらの用例は学習前後一貫して、それぞれ同じ動詞「言う」を喚起語とするものの、平仮名で記載された「いう」を含む用例群と漢字を用いて記載された「言う」を含む用例群に分かれて分布した。また、平仮名の「いう」を含む用例群の周囲には同じく平仮名の喚起語を持つ用例が分布しており、文字種の特徴が埋め込みに強く反映されていた。このような表記揺れは日本語特有の問題であり、FrameNetを教師信号としてmBERTに対して深層距離学習を適用した場合に意味フレーム推定の性能向上が確認できなかった一因である可能性が考えられる。

フレーム	用例
Capability	(1) 私は泣くことが でき なかった。
	(2) よく できる 子どもたちは、このことをよく知っており、学業にいそむ。
Change_position_on_a_scale	(3) [...] きめ細やかに対応した米の安定的供給の必要性が 高ま っている。
	(4) [...] 国連が国際の平和と安全を維持する事例も 増加 した。
Possession	(5) コウモリも素晴らしい生物時計を もっ ていて、日が暮れたのを [...]
	(6) 私は A+B+C からなる発明について特許権を 持 っています。
Statement	(7) [...] 結婚を遅らせる晩婚の人が増えていると い われている。
	(8) 話している途中でユナが 言 った。

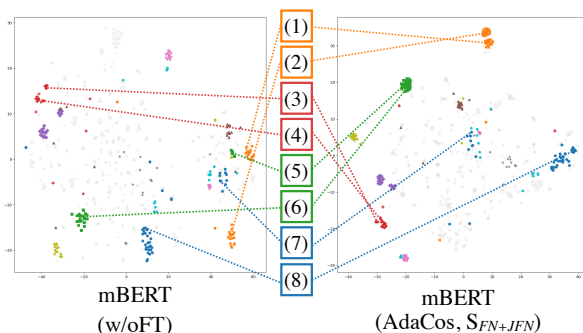


図2 深層距離学習前後のmBERTの分布の変化

4 おわりに

本稿では山田らの提案した意味フレーム推定手法を日本語動詞の意味フレーム推定に適用し、性能を検証した。日本語フレームネットに含まれる用例数は、山田らが英語動詞の意味フレーム推定に利用したFrameNetの用例数の3%と大幅に少ないにも関わらず、教師信号として日本語フレームネットを使う設定では、言語モデルとして日本語単言語モデルである東北大BERTを用いた場合、多言語モデルであるmBERTを用いた場合、いずれにおいても深層距離学習を行うことで意味フレーム推定性能が向上することが確認できた。一方、多言語モデルであるmBERTを対象に、FrameNetを教師信号として用いる設定では、深層距離学習による性能向上は確認できなかった。

今後の課題としては、日本語フレームネットのフレームを付与したデータを増やし、より大規模な評価セットを用いた実験を行う予定である。また、FrameNet形式のフレームリソースは、ブラジル語やポルトガル語、中国語をはじめ、多くの言語を対象に開発が進んでいる[18]ことから、英語と日本語以外のリソースを用いた実験に取り組みたい。

謝辞

本研究は、JST 創発的研究支援事業、JPMJFR216N の支援を受けたものである。

参考文献

- [1] Charles J Fillmore. **Frame Semantics**. Hanshin Publishing Company, 1982.
- [2] Josef Ruppenhofer, Michael Ellsworth, Myriam Schwarzer-Petruck, Christopher R Johnson, and Jan Scheffczyk. **FrameNet II: Extended theory and practice**. International Computer Science Institute, 2016.
- [3] 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 (NAACL 2019)**, pp. 4171–4186, 2019.
- [4] Nikolay Arefyev, Boris Sheludko, Adis Davletov, Dmitry Kharchev, Alex Nevidomsky, and Alexander Panchenko. Neural GRANNy at SemEval-2019 task 2: A combined approach for better modeling of semantic relationships in semantic frame induction. In **Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval 2019)**, pp. 31–38, 2019.
- [5] Saba Anwar, Dmitry Ustalov, Nikolay Arefyev, Simone Paolo Ponzetto, Chris Biemann, and Alexander Panchenko. HHMM at SemEval-2019 task 2: Unsupervised frame induction using contextualized word embeddings. In **Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval 2019)**, pp. 125–129, 2019.
- [6] Eugénio Ribeiro, Vânia Mendonça, Ricardo Ribeiro, David Martins de Matos, Alberto Sardinha, Ana Lúcia Santos, and Luísa Coheur. L2F/INESC-ID at SemEval-2019 task 2: Unsupervised lexical semantic frame induction using contextualized word representations. In **Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval 2019)**, pp. 130–136, 2019.
- [7] Kosuke Yamada, Ryohei Sasano, and Koichi Takeda. Semantic frame induction using masked word embeddings and two-step clustering. In **Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021)**, pp. 811–816, 2021.
- [8] 山田康輔, 笹野遼平, 武田浩一. 深層距離学習を用いた動詞の意味フレーム推定. 言語処理学会第 29 回年次大会, 2023.
- [9] Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In **Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006)**, Vol. 2, pp. 1735–1742, 2006.
- [10] Kilian Q Weinberger and Lawrence K Saul. Distance metric learning for large margin nearest neighbor classification. **Journal of machine learning research**, Vol. 10, No. 2, 2009.
- [11] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019)**, pp. 4690–4699, 2019.
- [12] Xiao Zhang, Rui Zhao, Yu Qiao, Xiaogang Wang, and Hongsheng Li. AdaCos: Adaptively scaling cosine logits for effectively learning deep face representations. In **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019)**, pp. 10823–10832, 2019.
- [13] Kyoko Ohara, Seiko Yamaguchi Fujii, Toshio Ohori, Ryoko Suzuki, Hiroaki Saito, and Shun Ishizaki. The Japanese framenet project: An introduction. 2004.
- [14] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In **Proceedings of the 5th International Conference on Learning Representations (ICLR 2017)**, 2017.
- [15] Dan Pelleg and Andrew Moore. X-means: Extending k-means with efficient estimation of the number of clusters. In **Proceedings of the 17th International Conference on Machine Learning (ICML 2000)**, pp. 727–734, 2000.
- [16] Amit Bagga and Breck Baldwin. Entity-based cross-document coreferencing using the vector space model. In **Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (ACL-COLING 1998)**, pp. 79–85, 1998.
- [17] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. **Journal of Machine Learning Research**, Vol. 9, pp. 2579–2605, 2008.
- [18] Collin F. Baker, Michael Ellsworth, Miriam R. L. Petruck, and Swabha Swayamdipta. Frame semantics across languages: Towards a multilingual FrameNet. In **Proceedings of the 27th International Conference on Computational Linguistics: Tutorial Abstracts**, pp. 9–12, Santa Fe, New Mexico, USA, August 2018. Association for Computational Linguistics.