

FrameNet を利用した談話関係の認識

李 凌寒* 加藤 恒昭†

東京大学 大学院総合文化研究科

ryou0634@g.ecc.u-tokyo.ac.jp*; kato@boz.c.u-tokyo.ac.jp†

1 はじめに

談話関係認識とは、テキスト中の文や節といった談話単位の修辞関係や談話関係を同定するタスクであり、包括的な意味理解を必要とする自動要約や対話システムの高度化のために重要である。例えば、例 (1) では2つの談話単位 *Arg1* と *Arg2* の間に、〈原因〉の談話関係が認められる。

- (1) *Arg1*: I can't go hiking tomorrow.
Arg2: It will be raining.

談話単位間にはしばしば「談話マーカ」(*because* などの接続詞) と呼ばれる明示的な (explicit) 手掛かりがあり、その場合の処理は比較的容易であるが、上記の例のように談話マーカが存在しない非明示的な (implicit) な談話関係認識は非常に困難である。現状の手法では談話単位内の単語といった表層的な情報のみを用いたモデルが主であるが、非明示的な談話関係を正確に捉えるためには、世界知識や常識を用いた推論が必要になると考えられる。

そこで本研究では、言語学知識データベースの FrameNet[1] の有用性を調査する。FrameNet は、フレーム意味論 [7] の「単語が使用されるとき話者はとある背景知識 (フレーム) を前提にしてその語を使用している」という考えに基づき、1200 余りのフレームの知識をまとめており、特にフレーム間関係が談話関係認識に役立つことが期待される。本研究では FrameNet のシンボリックな知識をニューラルネットワークに入力するために、知識グラフ埋め込み (Knowledge Graph Embedding) の手法 [2] を用いて、フレームの分散表現を獲得する。分散表現は FrameNet に記述されていない知識を補うことも期待できる。

実験として、談話関係認識で一般的に用いられる the Penn Discourse Treebank (PDTB) を用いた非明示的な談話関係を分類するタスクを行う。PDTB では談話関係が階層的に分類されているが、今回は最上位の分類を行った。分類器として多層パーセプトロン (MLP) を用い、ベースラインとして単語埋め込み (Word Embedding) を入力とするエンコーダーを用いたモデルを採用する。フレーム情報を用いる際は、

入力文で認識されたフレームのペアについて重み付けをし、それらの分散表現の重み付き和を MLP に入力する。

結果として、フレーム情報のみのモデルとランダムベースラインを比較したところ、フレームのモデルが有意に高い F 値を達成した。このことから、フレームは談話関係認識に有用な情報を持っていると考えられる。しかし、フレームと単語の情報を合わせたモデルと、単語のみのモデルを比較したところ、フレームの情報による F 値の有意な向上は見られなかった。原因として、(1) フレームから得られる情報がすでに単語埋め込みに含まれていること (2) 単純なフレーム間関係だけでは談話関係認識の手がかりにならないことが考えられる。

2 背景

2.1 談話関係認識

談話関係認識には、以前、人手で作成した特徴量を元に機械学習を用いる手法が用いられていた [10]。しかし最近では、煩雑な特徴量設計を必要としない単語埋め込みを用いた畳み込みニューラルネットワーク (CNN) [11] やリカレント・ニューラルネットワーク (RNN) [9] を用いたモデルが高い精度を上げている。

しかし、既存の手法では単語からの表層的な情報しか考慮することができず、その背景にある知識を用いることができない。そこで、次に紹介する FrameNet の情報がその欠点を補えるのではないかと考える。

2.2 FrameNet

FrameNet[1] は、フレーム意味論 [7] に基づき 1,221 のフレームの定義に加え、フレームの意味役割 (frame elements; FEs)、フレームを喚起する単語 (lexical units; LUs) をまとめている。また、*Inherits from/Is Inherited of* や *Precedes/Is preceded by* など全 13 種のフレーム間関係 (frame-to-frame relation) が定義されており、これらが特に談話関係特定に役立つことが

期待できる。例えば、*precedes/is preceded by* の関係は Temporal の談話関係を認識するのに強力な手がかりになると予想される。

2.3 知識グラフの埋め込み

FrameNet のシンボリックな知識をニューラルネットワークに適した表現に変換するための、フレームの分散表現を得る手法について述べる。FrameNet のデータは、フレームを節点、フレーム間関係を辺とする知識グラフとみなすことができる。知識グラフを連続ベクトル空間に埋め込む手法は、知識グラフ補完の文脈で研究されてきたが、FrameNet のデータに応用した研究もあり [3]、そこでは TransE モデル [2] が欠損データの予測に関して最も高い精度を取っている。そこで、本研究でも以下に説明する TransE モデルを採用する。

フレームの分散表現を得る際には、FrameNet に定義された全フレームを F 、7つのフレーム間関係（逆関係と See_also を除いた）を R とし、関係の事例を表す三つ組 $\{(h, r, t) | h, t \in R, r \in F\}$ の訓練データ集合 S から、 F と R の D_f 次元ベクトル表現を学習する。TransE モデルでは関係は要素間の変換として表現され、ある関係 (h, r, t) が成立する場合、対応するベクトルは $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ を満たし、そうでない場合 $\mathbf{h} + \mathbf{r}$ は \mathbf{t} から離れている。学習では以下の目的関数を最小化する。

$$\mathcal{L} = \sum_{(h,r,t) \in S} \sum_{(h',r',t') \in S'_{(h,r,t)}} [\gamma + d(\mathbf{h} + \mathbf{r}, \mathbf{t}) - d(\mathbf{h}' + \mathbf{r}', \mathbf{t}')]_+ \quad (1)$$

$[x]_+$ は $\max(0, x)$ を表す。 $\gamma > 0$ はマージンを表すハイパーパラメータであり、 d は距離関数として L_1 または L_2 ノルムが用いられる。集合 S' は、ある関係について成り立ちうる組み合わせ全ての集合であり、必ずしも成立しない負例として用いられる。

$$S'_{(h,r,t)} = \{(h', r, t) | h' \in E\} \cup \{(h, r, t') | t' \in E\} \quad (2)$$

学習の際は、目的関数 \mathcal{L} を単にノルムを大きくすることで小さくするのを防ぐため、要素ベクトルの L_2 ノルムは 1 になるよう正規化される。本研究では先行研究 [3] にない、 $\gamma = 1$ 、距離関数 d は L_1 ノルム、学習されるフレームの分散表現の次元は $D_f = 50$ とした。

3 手法

本研究では、単語の情報のみを考慮したモデルと、それをフレーム情報と合わせたモデル（図 1）とを比

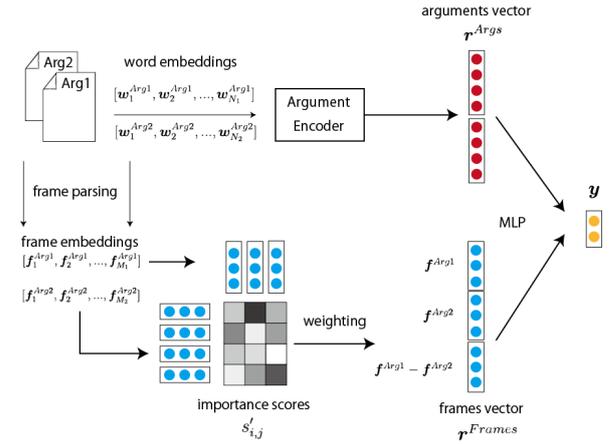


図 1: 単語とフレーム情報を考慮したモデル

較する。最終的な分類には、単語またはフレームからの情報 \mathbf{r}^{Args} , \mathbf{r}^{Frames} を受け取り、談話関係の二値分類結果を出力する隠れ層一層を持つ MLP を用いる。以下に単語またはフレームの情報をエンコードするモデルの詳細を記述する。

3.1 単語情報の利用

談話単位内の単語は、訓練済みの Google-News Vectors¹ で初期化された単語埋め込み層を通じて、 $D_w (= 300)$ 次元の分散表現の並び $[\mathbf{w}_1^{Arg1}, \mathbf{w}_2^{Arg1}, \dots, \mathbf{w}_{N_1}^{Arg1}]$, $[\mathbf{w}_1^{Arg2}, \mathbf{w}_2^{Arg2}, \dots, \mathbf{w}_{N_2}^{Arg2}]$ へと変換される。これを以下に記述するエンコーダーによって、 $\mathbf{r}^{Args} \in \mathbb{R}^{D_a}$ へと変換する。

単語エンコーダーとして、シンプルかつ比較的高い精度を達成している Zhang ら [11] のモデルを採用した。このエンコーダーは 3つのプーリング操作、*average*, *min*, *max* を用いて、入力となるベクトル列から固定長の特徴量を抽出する。

$$\mathbf{c}_{avg} = \frac{1}{N} \sum_{i=1}^N \mathbf{w}_i \quad (3)$$

$$\mathbf{c}_{min} = \min(\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_N) \quad (4)$$

$$\mathbf{c}_{max} = \max(\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_N) \quad (5)$$

min, *max* は入力となる行列の行からそれぞれ最小/最大値を抽出する。これらのベクトルは縦に連結されたのち、非線形の *tanh* を通し、正規化され出力される。

$$\mathbf{a} = [\mathbf{c}_{Arg1}^{avg}; \mathbf{c}_{min}^{Arg1}; \mathbf{c}_{max}^{Arg1}; \mathbf{c}_{Arg2}^{avg}; \mathbf{c}_{min}^{Arg2}; \mathbf{c}_{max}^{Arg2}] \quad (6)$$

$$\mathbf{r}^{Args} = \frac{\tanh(\mathbf{a})}{|\tanh(\mathbf{a})|} \quad (7)$$

¹<https://code.google.com/archive/p/word2vec/>

3.2 フレーム情報の利用

フレームの分散表現をニューラル分類器に入力するために、固定長表現に変換する手法を説明する。この手法は、談話関係認識で単語ペアについて重み付けをする Chen らの手法 [4] を参考にした。まず、 $Arg1$, $Arg2$ をフレーム意味論に基づいたパーサー SEMAFOR[5] で解析し、認識されたフレームを前述の D_f 次元の TransE ベクトルに変換し、 $[f_1^{Arg1}, f_2^{Arg1}, \dots, f_{M_1}^{Arg1}], [f_1^{Arg2}, f_2^{Arg2}, \dots, f_{M_2}^{Arg2}]$ を得る。このうち、談話関係認識に関連する情報のみを得るために $Arg1$, $Arg2$ 間のフレーム対に以下のように重みを計算する。

$$s_{i,j} = z^T \tanh(W_a \begin{pmatrix} f_i^{Arg1} \\ f_j^{Arg2} \\ f_i^{Arg1} - f_j^{Arg2} \end{pmatrix} + b_a) + b_z \quad (8)$$

$$s'_{i,j} = \frac{\exp(s_{i,j})}{\sum_{i,j} \exp(s_{i,j})} \quad (9)$$

ここで、TransE モデルの、ベクトルの差分が関係を表すという性質から $f_i^{Arg1} - f_j^{Arg2}$ の項をフレーム間関係の情報として加える。最終的なフレームの表現 r^{Frames} は以下の通りとなる。

$$r^{Frames} = \sum_{i,j} s'_{i,j} \begin{pmatrix} f_i^{Arg1} \\ f_j^{Arg2} \\ f_i^{Arg1} - f_j^{Arg2} \end{pmatrix} \quad (10)$$

4 実験

4.1 設定

実験データには、先行研究にならい、PDTB のうちセクション 2-20 を訓練セット、0-1 を評価セット、21-22 をテストセットとして用いた。タスクは PDTB の一番上の階層の 4 つの関係についての正負の二値分類とした。表 1 に示されている通り、訓練データ内の正負例の分布が関係によって偏っているため、訓練の際はエポック毎に、正負例の数を揃えるようにダウンサンプリングを行った。

表 1: PDTB のデータラベルの分布

	Train	Dev	Test
Comparison	1928/10704	196/987	151/895
Contingency	3237/9395	283/900	272/774
Expansion	6988/5644	668/515	573/473
Temporal	757/11875	63/1120	85/961
Total	12632	1183	1046

モデルの訓練について、最適化するパラメーターは単語埋め込み E_w 、フレームに重み付けをする際の W_a, b_a, z, b_z 、また分類器の W_{out}, b_{out} であり、フレームの分散表現は含まれない。訓練は全て訓練データにおけるモデルの出力と正解ラベルとの交差エントロピー誤差を最小化することで行った。最適化アルゴリズムとして Adam [8] を用い、過学習を防ぐために評価セットでの F 値の最高値が 10 epoch 変わらなかった時点で訓練を打ち切り、評価セットで最高値を記録したモデルでテストセットの評価を行った。

4.2 結果

予備実験として、テストセットのラベル分布に従ってランダムに予測するベースラインと、フレームのみの情報を用いるモデルの F 値を比較した結果を表 2 に示す。ランダムベースラインに比べ、フレームのみのモデルが 5 ポイントから 16 ポイントと有意に高い F 値を達成している。従って、フレームは談話関係認識に有用な情報がある程度含んでいると考えられる。

表 2: ランダム/フレームのみモデルの比較 (F 値)

Relation	ランダム	フレームのみ
Comparison	14.43	28.78
Contingency	26.00	42.87
Expansion	54.78	63.10
Temporal	8.12	13.14

次に、単語情報のみのモデルと、それにフレーム情報を加えたモデルの F 値を比較した結果を表 3 に示す。フレームありのモデルは、ベースラインモデルに比べて最大 2 ポイント低いまたはわずかに 1 ポイント高い F 値を記録している。フレーム情報はパフォーマンスの向上に有意に寄与せず、むしろ過学習を引き起こしていることが多いと考えられる。

表 3: フレームあり/なしモデルの比較 (F 値)

関係	単語のみ	単語・フレーム
Comparison	39.61	37.30
Contingency	53.14	53.47
Expansion	68.71	68.54
Temporal	33.19	32.51

5 分析

フレームの情報は、単語情報を考慮したモデルの改善に寄与しなかった。これには 2 つの主な原因が考えられる：(1) FrameNet のデータ量では人間の複雑

な知識を完全に表現することは難しく、単語埋め込みにフレームの情報とほぼ等価なものが既に含まれている; (2) 談話関係認識は単純なフレーム間関係に注目するだけでは解くことができない。以下にそれぞれの分析を述べる。

(1) フレームが提供する情報として、単語の一般化が挙げられる。例えば、Sleep のフレームは LU として、*asleep, drowse, nap, unconscious* などをもち、これらを睡眠に関連する単語として一般化する。しかし、大規模コーパスによって学習された単語埋め込みにも、同様の情報は含まれており、実際に *sleep* の単語ベクトルと他の単語のコサイン類似度をとると、上位 50 の単語のうちに Sleep の LU が大部分確認できる。さらに、フレーム間関係も同様に大規模コーパスから学習されている可能性がある。

また、FrameNet のデータ量が人間の世界知識を補うには小さすぎるということも考えられる。FrameNet は 1,221 フレームを要素としてもち、逆関係を除いた 7 つの関係、要素間の 1,935 の関係例を持つが、他の知識データセット、例えば WordNet [6] の持つ 117,000 要素、20 関係、500,000 の関係例に比べると、大幅に小さい。

(2) 談話関係認識は、そもそもフレーム間関係に注目するだけで精度が向上するものではないという可能性がある。実験前の直感的な仮説として、*precedes* のフレーム間関係は Temporal の談話関係認識に有用である、というものがあつた。しかし、フレームありモデルが実際にフレーム対に割り当てた重み s' を調べると、モデルは Temporal の関係認識の際に *precedes* の関係に高い重みを割り当てているわけではないことが分かつた。

原因として、訓練データ内のフレーム間関係の分布が考えられる。実際に *precedes* の関係が Temporal の訓練データ内でどのように分布しているかを調べると、4.7% (4/85) の正例が *precedes* の関係を持つフレームのペアを含む一方で、4.1% (48/951) の負例も同様に *precedes* のフレーム対を持つことが分かつた。*precedes* の関係が、正/負例にわたってほぼ同じ比率で分布しているため、Temporal の認識に *precedes* の関係を利用することが難しくなっていると思われる。

このような直感に反する言語学的素性の分布は他の談話関係においても報告されており (例: Pitler ら [10] の研究における、単語極性ペアの素性と Comparison の談話関係)、単純な言語学的素性に注目するだけでは、談話関係認識の精度向上は難しく、より深い意味や文脈を捉えるような情報が必要であることが示唆される。

6 おわりに

本研究では、FrameNet の談話関係認識への有用性を調査した。結果としてフレーム情報を加えることによる既存モデルの有意な精度向上は見られなかつた。今後の方向性として、人手で整備されたデータだけでなく、タグ付けされていないテキストデータなどから、より大規模な知識を獲得する手法の研究などが挙げられる。

謝辞 本研究は JSPS 科研費#17H01831 および #15K12873 の助成を受けた。ご支援に深く感謝する。

参考文献

- [1] C. F. Baker. FrameNet: A Knowledge Base for Natural Language Processing. In *Proc. of Frame Semantics in NLP A Workshop in Honor of Chuck Fillmore*, pp. 1–5, Baltimore, Maryland USA, 2014.
- [2] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating Embeddings for Modeling Multi-relational Data. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in NIPS 26*, pp. 2787–2795. Curran Associates, Inc., 2013.
- [3] T. Botschen, H. Mousselly-Serigh, and I. Gurevych. Prediction of Frame-to-Frame Relations in the FrameNet Hierarchy with Frame Embeddings. In *Proc. of the 2nd Workshop on Representation Learning for NLP*, pp. 146–156, Vancouver, Canada, 2017.
- [4] J. Chen, Q. Zhang, P. Liu, X. Qiu, and X. Huang. Implicit Discourse Relation Detection via a Deep Architecture with Gated Relevance Network. In *Proc. of the 54th ACL*, pp. 1726–1735, Berlin, Germany, 2016.
- [5] D. Das, D. Chen, A. F. T. Martins, N. Schneider, and N. A. Smith. Frame-Semantic Parsing. *Computational Linguistics*, Vol. 40, No. 1, pp. 9–56, 2014.
- [6] C. Fellbaum. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA, 1998.
- [7] Charles J. Fillmore. Frame Semantics and the Nature of Language. *Annals of the New York Academy of Sciences*, Vol. 280, No. 1, pp. 20–32, 1976.
- [8] D. P. Kingma and J. L. Ba. Adam: A Method for Stochastic Optimization. *The International Conference on Learning Representations*, pp. 1–15, 2015.
- [9] Y. Liu and S. Li. Recognizing Implicit Discourse Relations via Repeated Reading: Neural Networks with Multi-Level Attention. In *Proc. of the 2016 EMNLP*, pp. 1224–1233, Austin, Texas, 2016.
- [10] E. Pitler, A. Louis, and A. Nenkova. Automatic sense prediction for implicit discourse relations in text. In *Proc. of the ACL and the 4th IJCNLP of the AFNLP*, pp. 683–691, Suntec, Singapore, 2009. ACL and AFNLP.
- [11] B. Zhang, J. Su, D. Xiong, Y. Lu, H. Duan, and J. Yao. Shallow Convolutional Neural Network for Implicit Discourse Relation Recognition. In *Proc. of the 2015 EMNLP*, pp. 2230–2235, Lisbon, Portugal, 2015.