

# 一般常識を用いた非明示的な談話関係解析

岸本 裕大<sup>†</sup> 村脇 有吾<sup>†‡</sup> 黒橋 禎夫<sup>†‡</sup>

<sup>†</sup> 京都大学大学院情報学研究科 <sup>‡</sup> 科学技術振興機構 CREST

<sup>†‡</sup> {kishimoto, murawaki, kuro}@nlp.ist.i.kyoto-u.ac.jp

## 1 はじめに

我々は文章を読解する際、文単位や節単位で読解するのではなく、周辺の文や節との意味的なつながりを理解しながら読み進めている。このようなつながりを談話関係と呼び、談話関係を解析するタスクを談話関係解析と呼ぶ。談話関係は“because”や“however”といった談話関係を直接示す表現（以降、談話標識と呼ぶ）が含まれる場合（以降、明示的と呼ぶ）と談話標識が含まれない場合（以降、非明示的と呼ぶ）に分類される。明示的な場合は談話標識が大きな手がかりとなるため解析精度は非常に高いが、非明示的な場合は解析精度が低く、非常に挑戦的なタスクとなっている [1]。本研究では非明示的な談話関係解析を扱う。

人間が談話関係を理解する際、一般常識を用いている。次の例について考える。

Arg1: *Not counting the extraordinary charge it would have had a net loss of \$3.1 million, or seven cents a share*

Arg2: **A year earlier, it had profit of \$7.5 million, or 18 cents a share**

(Comparison, Contrast)

この例では、Arg1の“loss”とArg2の“profit”が反義関係であるという点がArg1とArg2が対比関係であることの大きな手がかりとなる。このような手がかりは大量の学習セットさえあれば、システムが自ら学習することが期待できる。しかし談話関係解析で使われている学習セットは規模が小さいため、このような関係を直接学習することは困難である。そこで、反義関係などの一般常識を外部知識として直接モデルに与える方法を提案する。

本研究では、MAGE-GRU [2] と呼ばれる外部知識を直接的に活用することができるモデルを用いた談話関係解析器を提案する。MAGE-GRUとはGated Recurrent Unit (GRU) [3]を拡張したモデルであり、通常の単語列を拡張した有向非循環グラフを入力としている。前述の例の場合、“loss”と“profit”のような反義関

係をエッジで結んだ有向非循環グラフをMAGE-GRUに入力することで、直接的に外部知識を活用することが可能となる。本研究では、ConceptNet [4]と共参照解析結果を一般常識として用いて実験を行なった。実験の結果、ConceptNetを用いたことで非明示的な談話関係解析の精度が向上した。

## 2 関連研究

非明示的な談話関係解析タスクでは、Feedforward Neural Network [5] や CNN [6]、Bi-LSTM [7] など様々なニューラルネットワークが用いられている。Rutherfordら [5] は1層のpooling層と1層の隠れ層からなるFeedforward Neural Networkモデルを構築し、LSTMやTree LSTMで構築したモデルよりも解析精度が高くなることを示した。Qinら [6] はCNNと敵対的学習を組み合わせたモデルを提案した。このモデルは、非明示的な談話関係を解析するCNNの出力が明示的な談話関係を解析するCNNの出力に近づくように敵対的学習を行うことで、非明示的な談話関係解析の精度を向上させている。Liuら [7] は、Bi-LSTMと2段階のアテンション機構を組み合わせたモデルを構築した。

近年、Brown Clusters や VerbNet、モダリティなどの外部知識を用いることで非明示的な談話関係解析の精度が向上することが報告されている [8][9]。これらの研究では、ニューラルネットワークの出力層の直前に素性を追加する、という形で外部知識を与えている。談話関係解析以外のタスクにおいて、Dhingraらは外部知識をより直接的にニューラルネットワークに与えることができるMAGE-GRUというモデルを提案している [2]。彼らは、共参照解析結果を外部知識としてMAGE-GRUに与えることで様々なタスクで精度が向上することを示した。

## 3 提案手法

本研究で提案するシステムの概略を図1に示す。本システムはRutherfordらのモデル [5]を基にしている。提案手法は2つのテキストスパン（以降 Arg1、Arg2

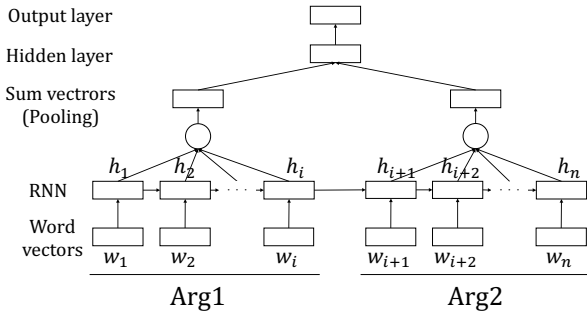


図 1: 提案システムの概略図

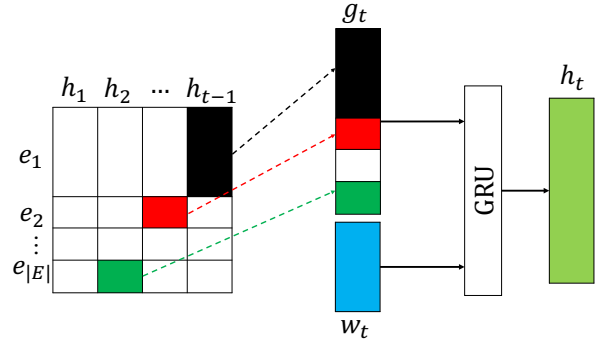


図 3: MAGE-GRU の処理

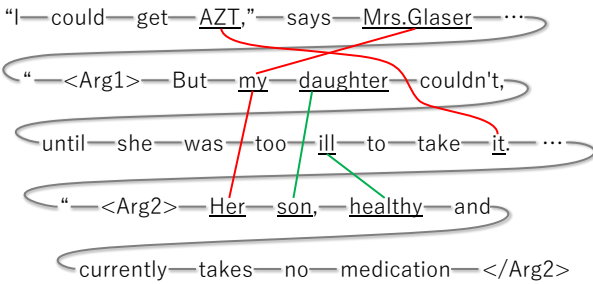


図 2: MAGE-GRU に入力する単語列。(通常の単語列に灰色のエッジを、共参照を示す単語同士に赤色のエッジを、反義関係を示す単語同士に緑色のエッジを張っている。)

と呼ぶ) に含まれる単語ベクトル列  $w_1, \dots, w_n$  を入力とし、RNN を用いて隠れ状態ベクトル  $h_1, \dots, h_n$  を生成する。その後 pooling 層で Arg1、Arg2 それぞれの特徴ベクトルを生成し、それらのベクトルを連結した後に softmax でどの談話関係ラベルに該当するかを出力する。

### 3.1 MAGE-GRU

本研究における Rutherford らのモデルからの改良点は、MAGE-GRU [2] を使用する点である。MAGE-GRU は Gated Recurrent Unit (GRU) [3] の拡張であり、より直接的な情報として外部知識を活用することができるニューラルネットワークモデルである。

MAGE-GRU は、単語列を拡張した有向非循環グラフを入力する。具体例を図 2 に示す。グラフの構築方法としては、まず単語列を特別なエッジとみなして繋げる。その上で、因果関係や反義関係など強い関係を持つ単語ペアにエッジを張る。このような単語ペアを (単語 A, 関係ラベル, 単語 B) という形で表現する。この際、単語 A は単語 B より前に出現していることとする。例えば、図 2 中では反義関係を持つ単語ペア (daughter, 反義, son) と (ill, 反義, healthy) に緑色のエッジを張っている。なお、原因・結果のように一定の方向を持つ関係ラベルに対しては、逆向きの関係ラベル

を作ることで方向性を維持している。また、複数の単語からなるフレーズを扱う場合は、フレーズの最後の単語にエッジを張る。エッジを結ぶ際、計算量を削減するために各単語は各関係ラベルごとに最大 1 つの入力エッジを持つように制限している。そのため、単語 B と連結できる単語 A が 2 つ以上ある場合は単語列の上で距離が最も近いもの 1 つのみとエッジを張る。

MAGE-GRU は前述の有向非循環グラフを入力とし、各時刻  $t$  毎に隠れ状態ベクトル  $h_t$  を出力する。MAGE-GRU の処理を図 3 に示す。通常の GRU は、 $h_{t-1}$  と単語ベクトル  $w_t$  を入力とするが、MAGE-GRU では、 $h_{t-1}$  に代わって別のベクトル  $g_t$  と  $w_t$  を入力としている。 $g_t$  は過去の隠れ状態ベクトル  $h_1, \dots, h_{t-1}$  の一部を選択、結合して生成する。

$$g_t = [g_t^{e_1}; g_t^{e_2}; \dots; g_t^{e_{|E|}}]$$

$$g_t^{e_i} = \begin{cases} h_{t'}^{e_i} & \text{if } (x_{t'}, e_i, x_t) \text{ exists} \\ \mathbf{0} & \text{otherwise} \end{cases}$$

$h_{t'}$  は  $[h_{t'}^{e_1}; h_{t'}^{e_2}; \dots; h_{t'}^{e_{|E|}}]$  からなる。また  $e_i$  ( $1 < i < |E|$ ) は関係ラベルを指す。したがって、 $h_{t'}^{e_i}$  とは関係ラベル  $e_i$  に対応する部分ベクトル  $h_{t'}$  を指す。なお、 $e_1$  は通常の単語列を指している。例えば、 $w_t$  が図 2 の “healthy” を指す場合、 $g_t$  は灰色のエッジで連結した “son” の部分ベクトル  $h_{\text{son}}^{\text{sequence}}$  と緑色のエッジで連結した “ill” の部分ベクトル  $h_{\text{ill}}^{\text{antonym}}$  からなる。なお、それ以外の部分は全てゼロベクトルである。

### 3.2 使用する一般常識

本研究では、一般常識を表す外部知識として ConceptNet [4] と共参照解析結果を用いる。

ConceptNet とは人手による知識構築や Wikipedia などの Web ページからの自動抽出、ゲーミフィケーションを活用して、言語的な定義や一般常識を集約している知識源である。この知識源に含まれる関係の中には談話関係解析タスクに有用な知識が一部含まれている。例えば、“Causes” というラベルが付与されている関係

Sense	Train	Develop	Test
Comparison.Concession	179	19	21
Comparison.Contrast	1,672	185	206
Contingency.Cause	3,332	370	411
Contingency.Pragmatic	57	6	6
Expansion.Alternative	146	16	18
Expansion.Conjunction	2,787	309	344
Expansion.Instantiation	1,131	125	139
Expansion.List	314	34	38
Expansion.Restatement	2,519	279	310
Temporal.Asynchronous	527	58	65
Temporal.Synchrony	143	15	17
Total	12,807	1,416	1,575

表 1: テストセットの分布

は PDTB が定義する “Contingency.Cause” というラベルと強く関係している。このように談話関係と強く関係している知識が多く含まれているため、ConceptNet を用いることで談話関係解析の精度向上が期待される。

また、共参照解析結果も外部知識として活用する。共参照解析結果は談話関係解析を行う際に重要な文脈情報を示す直接的な手がかりであるため、談話関係解析の精度向上に寄与することが期待される。

## 4 実験

### 4.1 実験設定

#### 4.1.1 Penn Discourse TreeBank

本研究では、Penn Discourse TreeBank (PDTB) [10] を使用して性能評価を行う。PDTB は Wall Street Journal に談話関係をアノテーションした英語のコーパスである。アノテーションは 2 つのテキストスパン (arg1、arg2) と談話関係ラベルから構成されている。

PDTB の談話関係ラベルは 3 階層の構造から成り立っており、上位ラベルを使った 4 値分類 [7]、中位ラベルを使った 11 値分類 [5]、中位ラベルを一部改変した 15 値分類 [1] など複数の実験設定がある。本研究では、中位ラベルを使った 11 値分類で評価する。

通常の談話関係解析タスクでは、Section 2–21 を学習セット、Section 22 を開発セット、Section 23 をテストセットとして活用している [5]。しかし Shi らは、中位ラベルの分類タスクを行う場合、通常のテストセットだと規模が小さいため正しく評価できないことから Section 0–24 を 10 分割した交差検定で評価している [9]。我々も予備実験において Shi らの主張を確認したため 10 分割交差検定で性能評価を行う。

表 1 に談話関係ラベルの分布を示す。なお、本実験では Shi らの実験設定にできる限り近づけたが、談話

Description	Values
入力	単語列
単語埋め込み	100,000 語 (300 次元)
optimizer	AdaGrad
pooling 層	summation
隠れ層	1 層 (600 次元)
外部知識	共参照解析結果 and/or ConceptNet (10 次元/関係ラベル)
ミニバッチサイズ	1
early stopping	実施

表 2: 学習パラメータの設定

関係ラベルの分布が若干異なっている。

#### 4.1.2 モデル設定

入力となる有向非循環グラフを作成する際、Arg1 と Arg2 を連結し 1 つの単語列として扱った。その際、それぞれの始点と終点を示す特別なマーカー <Arg1>、</Arg1>、<Arg2>、</Arg2> を単語列に挿入した。

Word embedding としては、CoNLL2016 の Shared Task で使用された word2vec のモデルを使用する。単語埋め込みは <Arg1> や </Arg2> などの特別なものについてのみ学習を行った。

外部知識としては 3.2 節で述べた通り ConceptNet と共参照解析結果を使用した。なお外部知識を使わない場合、MAGE-GRU は通常の GRU と同じ挙動をする。

ConceptNet とマッチする単語ペアにエッジを貼った結果、1 例あたり 1.6 個のエッジが張られた。なお、ノイズを削減するために NLTK corpus<sup>1</sup> の Stopwords リストに載っている単語にはエッジを張っていない。この状態ではカバー率が非常に低いため、以下の手続きをとってエッジ数を増やした。まず単語列に含まれる全ての単語に対して、コサイン類似度が 0.6 以上の単語を最大 10 個までリストアップした。その上で、リストアップされた全単語内において ConceptNet とマッチする単語ペアに対してエッジを張った。これにより Conceptnet によって張られたエッジ数が 1 例あたり 7.0 個と、約 4.4 倍に増加した。なお、本実験では 42 種類の関係ラベルに “AtLocation” や “Causes” といった方向を持つ関係ラベルの逆方向を示すラベル 35 種類を加えた、77 種類の関係ラベルを使用した。

共参照解析結果は、Stanford CoreNLP<sup>2</sup>(ver.3.7.0) の出力結果を使用する。CoreNLP は共参照解析を行うモデルを 3 つ用意しており、今回は最も精度の良い neural model を使用する。表 2 にその他の実験設定を記載する。

<sup>1</sup><http://www.nltk.org/>

<sup>2</sup><https://stanfordnlp.github.io/CoreNLP/>

	MicroAve.
提案手法 (GRU)	0.387
+共参照解析結果	0.389
+ConceptNet (拡張なし)	0.390
+共参照解析結果+ConceptNet (拡張なし)	0.390
+共参照解析結果+ConceptNet (拡張あり)	0.393
Shi ら (LSTM) [9]	0.344
+モダリティ	0.377
Rutherford ら (Feedforward、再実装) [5]	0.366

表 3: 実験結果。(“拡張なし”は ConceptNet とマッチする際にコサイン類似度を使用しなかった場合、“拡張あり”はコサイン類似度を使用した場合を指す。)

## 4.2 実験結果

実験結果を表 3 に示す。交差検定を用いて評価した場合の指標として、外部知識を用いている Shi らのモデル [9] と提案手法のベースである Rutherford らのモデル [5] の 2 つの精度を記載する。Shi らは LSTM のモデルに、Brown Cluster やモダリティなどの表層的な素性を追加した解析器を構築しており、表 3 では彼らの論文で報告されたモデルの中で最もスコアが良かった LSTM+モダリティのスコアを記載している。また、Rutherford らのモデルは交差検定で性能評価を行っていませんでしたので我々が再実装した結果を記載している。

表 3 より、知識を加えていない GRU のモデルが Shi らのモデルを上回る精度を示した。しかし、Shi らは外部知識を加えることで精度が大幅に向上したが、我々の手法では精度は向上したが限定的であった。次の例を用いて検討する。

例 1:

Arg1: *Saturday morning, a resident was given 15 minutes to scurry into a sagging building and reclaim what she could of her life’s possessions*

Arg2: **Saturday night she dined in an emergency shelter on salmon steaks prepared by chefs from one of the city’s four-star restaurants**

(Temporal.Asynchronous)

例 2:

Arg1: *It fell \$1.875 a share Friday, to \$6.25, a new low, in over-the-counter trading*

Arg2: **Its high for the past year was \$16.375 a share**

(Comparison.Contrast)

例 1 の場合、GRU では誤答していたが、ConceptNet を加えると正しいラベルを答えられた。これは、Arg1

の“morning”と Arg2 の“night”に反義関係のエッジが張られていたからではないかと推測する。しかし例 2 では“low”と“high”に反義関係のエッジが張られていたが、誤ったラベルを付与している。実験結果を分析すると、ノイズが非常に多く、その結果外部知識を上手く活用できなかったのではないかと考えられる。そのため、ConceptNet の整理や別の知識源を利用する必要があると考えられる。

## 5 おわりに

本研究では、外部知識を効率的に組み込むことができる MAGE-GRU を用いることで、非明示的な談話関係解析の精度向上を目指した。実験の結果、MAGE-GRU と ConceptNet を用いることで解析精度が限定的であるが向上した。今後は、MAGE-GRU の機構を LSTM や bi-GRU に移植したニューラルネットワークモデルの構築や、ConceptNet の整理、Freebase<sup>3</sup>など他の外部知識を活用することで、非明示的な談話関係解析の精度向上を目指す。

## 参考文献

- [1] Nianwen Xue, Tou Hwee Ng, Sameer Pradhan, Attapol Rutherford, Bonnie Webber, Chuan Wang, and Hongmin Wang. CoNLL 2016 shared task on multilingual shallow discourse parsing. In *Proceedings of CoNLL2016 shared task*, pp. 1–19, 2016.
- [2] Bhuwan Dhingra, Zhilin Yang, William W. Cohen, and Ruslan Salakhutdinov. Linguistic knowledge as memory for recurrent neural networks. *arXiv*, 2017.
- [3] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of EMNLP2014*, pp. 1724–1734, 2014.
- [4] Robert Speer and Catherine Havasi. Representing general relational knowledge in conceptnet 5. In *Proceedings of LREC2012*, pp. 3679–3686, 2012.
- [5] Attapol Rutherford, Vera Demberg, and Nianwen Xue. A systematic study of neural discourse models for implicit discourse relation. In *Proceedings of EACL2017*, pp. 281–291, 2017.
- [6] Lianhui Qin, Zhisong Zhang, Hai Zhao, Zhiting Hu, and Eric Xing. Adversarial connective-exploiting networks for implicit discourse relation classification. In *Proceedings of ACL2017*, pp. 1006–1017, 2017.
- [7] Yang Liu and Sujian Li. Recognizing implicit discourse relations via repeated reading: Neural networks with multi-level attention. In *Proceedings of EMNLP2016*, pp. 1224–1233, 2016.
- [8] Attapol Rutherford and Nianwen Xue. Discovering implicit discourse relations through brown cluster pair representation and coreference patterns. In *Proceedings of EACL2014*, pp. 645–654, 2014.
- [9] Wei Shi and Vera Demberg. Do we need cross validation for discourse relation classification? In *Proceedings of EACL2017*, pp. 150–156, 2017.
- [10] Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. The Penn discourse treebank 2.0. In *Proceedings of LREC2008*, pp. 2961–2968, 2008.

<sup>3</sup><https://developers.google.com/freebase/>