

GCN による議論要素の特定

Graph Convolutional Networks for Argument Component Identification

森尾 学†

† 東京農工大学大学院 工学府

概要

本研究では、文書の中から議論的な要素を抽出する「Argument Component Identification」と呼ばれるタスクに対して、依存構造解析結果のグラフ畳み込み機構を追加することによる有用性を評価する。

1 背景:どこに問題意識を感じているか

1.1 Argument Component Identification

文書の中から、特定の属性を持っている部分（範囲）を抽出するタスクは、自然言語処理の重要な役割である。近年では Argument Mining が注目されているが、Argument Mining における議論的な要素の抽出タスクは Argument Component Identification (ACI) と呼ばれる。ACI とは議論的な文章から「主張」や「前提」になっていそうな部分をトークンレベルで抽出するタスクである [9]。ACI によって抽出された部分は Argument Component (AC) と呼ばれる。

例えば、Essay データセット [9]（学生が書いたエッセイのデータセット）には「As a result, it is clear that people can freely express their patriotic feelings during international sports events.」という文章が含まれる。この中から AC を抽出すると、「it is clear that people can freely express their patriotic feelings during international sports events」となる。

1.2 BIO エンコーディング

AC の抽出は、系列ラベリングの問題に落とし込める。一般的には BIO エンコーディングといって、AC の最初のトークンを「B」、AC 中のトークンを「I」、AC の外のトークンを「O」で表し、系列として扱う。先程のエッセイの冒頭を BIO で表現すると次のようになる：

Output	O	O	O	O	B	I	I	...
	↑	↑	↑	↑	↑	↑	↑	
Input	as	a	result	,	it	is	clear	

input が入力トークン列であり、output が BIO 系列である。すなわち、ACI を行うことは、BIO 系列の推定にほかならない。BIO 系列の推定のために従来とられて来た手法は主に以下の通り：

- ・ 特徴ベースの CRF（条件付き確率場）
- ・ 双方向 LSTM
- ・ 双方向 LSTM+CRF
- ・ 文字レベル CNN+双方向 LSTM+CRF

これらのモデルは語彙的な特徴をうまく学習することができることから、多くの場面で用いられている。しかし、これまで「依存構造」が陽に入力として用いられることは多くなかった。依存構造とは単語間の係受けの関係であるが、我々が文章を見たときに議論的であるかどうかというのは、単語列だけでなく単語間の構造によっても判断しているはずである。そこで本研究では、依存構造解析器によって得られる係受け構造を明示的にニューラルネットで扱うことで、ACI のパフォーマンスにどのような影響を与えるかを調べる。

2 関連研究がどんな貢献をしたか

Graph Convolutional Network (GCN) [3, 2] はグラフを扱う多層ニューラルネットの技術である。グラフの全てのノードに対して、GCN は隣接ノードの特徴を含めてエンコードすることができる。自然言語処理に置いては各単語をノードとみなして GCN を適用する研究がある。Marcheggiani ら [8] は係り受け解析結果を用いて、Semantic Role Labelling を行った。彼らの手法は、各ノード（単語）の特徴ベクトルを双方向 LSTM (BLSTM) によって求め、その特徴ベクトルに対して GCN を適用するというものである。[4] は固有表現抽出に対して初めて GCN を適用した。また、[6] では、ニューラル機械翻訳の BLSTM に対して GCN を適用することにより、意味的なバイアスを注入することを試みた。特に、彼らの Semantic Role Label パーサ [7] で推定した依存構造を GCN に組み込んだことにより、BLEU スコアを向上させている。

3 提案手法

Marcheggiani ら [8] の手法にならい、BLSTM と GCN を組み合わせることを提案する。しかし、ACI タスクでは通常、CRF を損失の計算に用いるので、CRF を組み込んだモデルも評価する。GCN の中間ベクトルは次のようにして求める（スペースの都合で詳細は省略）：

$$h_v = \sigma \left(\frac{1}{N(v)} \sum_{u \in N(v)} V_{dir(u,v)} h_u + b_{L(u,v)} \right) \quad (1)$$

ここで、 $N(v)$ は単語 v の隣接ノード集合、 $dir(u,v)$ は u と v が順方向か、逆方向か、もしくは self-loop の 3 つを含む。 $L(u,v)$ は、係り受けの種類ラベルを表し、有向である。

係り受け解析は CoreNLP [5] を用いて行った。39 種類の係り受けラベルが存在するので、 $L(u,v)$ は 39 次元

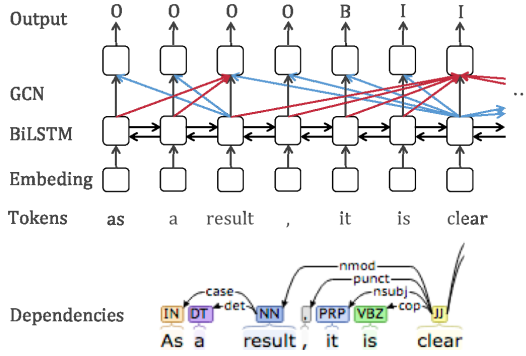


図 1: 提案アーキテクチャ (BLSTM + GCN)。下部は CoreNLP による係り受け解析の結果。

	essays	news
BLSTM	86.35	58.60
BLSTM+GCN	87.05	58.59
BLSTM+CRF	86.86	58.97
BLSTM+CRF+GCN	87.88	59.44

表 1: ACI タスクの F 値のマクロ平均。8 回以上訓練させ、その平均値をとった。各モデルにおいて validation ロスが最小となるモデルを選んだ。

となる。図 1 に、提案アーキテクチャを示す。この図において、各単語の特徴を BLSTM によって得る。その後、GCN によって係り受け解析の結果に基づいたグラフ畳み込みの処理を行う。

4 評価実験と考察

今回、ACI タスクのためのデータセットとして、エッセイ (essay) [9] とニュースコメント (news)[1] の 2 つを用いる。さらに、GCN の効果を検証するために次の 2 つの実験結果を得た。

実験結果 1: GCN による BIO タグ付けの識別精度向上

各モデルの識別精度の評価結果を表 1 に示す。この結果より、BLSTM と CRF を用いた従来モデルに対して、GCN を組み込むことが最も効果が高いことが分かった。よって、陽に係り受け解析結果をニューラルネットに注入することが有効であると言える。しかしながら、この表からも分かるように、データセットやドメイン、アノテーションの質によって結果が左右されると言える。

実験結果 2: GCN による BIO タグ付けの識別精度向上

もし、構文情報が ACI に有用であるならば、GCN を用いることで、単語情報が少なくなったとしても、効果的に ACI を行えるはずである。そこで、意図的に訓練データの量を落として実験することにした。例えば、訓練データの量を 1K (トークン数が 1000 個) に制限したのならば、essay データでは 15 個程度のパラグラフしか訓練に含まれない。すなわち、テストデータにおいて未知な語彙が登場しやすい。今回は、 $k = 1K, 5K, 12K, 20K, 60K, 80K$ を用意し、各 k における $\Delta(k) = \{BLGCN_{norm}(k) - BL_{norm}(k)\}$ を図 2 に示した。

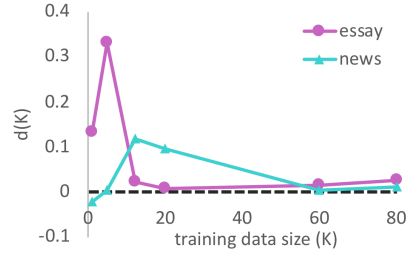


図 2: 横軸は訓練データの量。縦軸は BLSTM+CRF と BLSTM+CRF+GCN のマクロ F 値の正規化された差。

なお、 $\sigma_{norm}(k) = \frac{\sigma_{norm}(k)}{BL_{norm}(k)}$ である。すなわち、図 2 は 1K を起点として訓練データを増やしたときにパフォーマンスの増減を正規化したグラフである。この図より、訓練データが少ない (特に 5K, 12K) 場合に、GCN が効果的であることが分かる。これは、訓練データが少ない場合でも構文情報を学習することによる効果であると推測する。しかしながら、1K では精度が下がっているため、構文情報を使うといえども極端に少ない訓練データでは学習しにくいことも明らかになった。

今後は、様々な条件で実験を繰り返し、GCN を用いる場合の性質をより詳しくしていくつもりである。

参考文献

- [1] Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. The argument reasoning comprehension task: Identification and reconstruction of implicit warrants. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, p. (to appear), New Orleans, LA, USA, June 2018. Association for Computational Linguistics.
- [2] Steven Kearnes, Kevin McCloskey, Marc Berndl, Vijay Pande, and Patrick Riley. Molecular graph convolutions: moving beyond fingerprints. *Journal of Computer-Aided Molecular Design*, Vol. 30, No. 8, pp. 595–608, Aug 2016.
- [3] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2017.
- [4] Maria Kunilovskaya and Andrey Kutuzov. Universal dependencies-based syntactic features in detecting human translation varieties. In *Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories*, pp. 27–36, 2017.
- [5] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pp. 55–60, 2014.
- [6] Diego Marcheggiani, Joost Bastings, and Ivan Titov. Exploiting semantics in neural machine translation with graph convolutional networks. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pp. 486–492. Association for Computational Linguistics, 2018.
- [7] Diego Marcheggiani, Anton Frolov, and Ivan Titov. A simple and accurate syntax-agnostic neural model for dependency-based semantic role labeling. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pp. 411–420. Association for Computational Linguistics, 2017.
- [8] Diego Marcheggiani and Ivan Titov. Encoding sentences with graph convolutional networks for semantic role labeling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 1506–1515. Association for Computational Linguistics, 2017.
- [9] Christian Stab and Iryna Gurevych. Parsing argumentation structures in persuasive essays. *Computational Linguistics*, Vol. 43, No. 3, pp. 619–659, 2017.