

# 予備実験:GCN による議論要素の特定

## *Graph Convolutional Networks for Argument Component Identification*

江川 綾†

† 東京農工大学 工学部

### 1 背景:どこに問題意識を感じているか

Argument Component Identification (ACI) は Argument Mining における最も基礎的なサブタスクである。簡単に言えば, ACI 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 [?]. 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 [?] (学生が書いたエッセイに対して, Argument Mining のアノテーションを行ったデータセット) 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。

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 の中のトークンを"I", AC の外のトークンを"O"で表す。図 1 に, 先程のエッセイの例を BIO で可視化したものを示す。

この BIO エンコーディングを行うために, 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

入力には, 単語の BoW や Glove 埋め込み表現, 品詞埋め込み表現などが用いられてきた。しかし「依存構造」が用いられることは多くなかった。また, ニューラルネットにおいて依存構造解析結果を入力としたアーキテクチャが開発されて来なかった。依存構造とは単語間の係り受けの関係である。本研究では, 依存構造解析器によって得られる係り受け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

あああ

### 3 提案手法

ううう

### 4 実験結果から何がわかったか

えええ