

(続紙 1)

京都大学	博士 (情報学)	氏名	清丸 寛一
論文題目	Studies on Fundamental Problems in Event-Level Language Analysis (イベントレベルの言語解析における基礎的課題に関する研究)		
(論文内容の要旨)			
<p>自然言語処理において、テキストの構造的解析は依然として重要である。本論文は、5W1Hに対応する情報単位であるイベントに着目し、イベントレベルの言語解析における基礎的課題に取り組んでいる。イベントレベルの言語解析は、イベントが持つ属性を予測するイベント分類、イベント間の関係を認識するイベント間関係認識、ある条件を満たすイベントを予測するイベント予測の3つに大別される。本論文は、それぞれの基礎的課題である意志性分類、談話関係認識、後続イベント予測に取り組んでおり、それぞれに1章を割り当て、序論、関連研究、結論の章を合わせた全6章から構成されている。</p> <p>第1章は序論である。まず、テキストの構造的解析が、end-to-end 学習が主流となった現在でも依然として重要な役割を果たすことを説明している。次に、イベントレベルの構造的解析の有用性を論じている。最後に、イベントレベルの言語解析を上述の3分類に類型化し、いずれも自然言語処理の応用システムにとって重要であることを論じている。</p> <p>第2章は、イベントレベルの言語解析の関連研究をまとめている。まず、イベントの代表的な表現方法として、統語的構造表現、意味的構造表現、スパン表現の3つがあることを説明し、それぞれの利点と欠点を論じている。次に、イベント分類、イベント間関係認識、イベント予測の文脈で取り込まれてきた諸課題を整理している。最後に、イベントレベルの言語解析の応用を紹介している。</p> <p>第3章では、意志性分類の手法を提案している。意志性はイベントが人や組織等の意志的な行為であるか否かを表すイベントの基本的な属性であり、その潜在的応用は広い。しかし、これまで意志性に着目した研究はほとんどなく、容易に利用可能な意志性分類器もそれを学習するための訓練データも存在しなかった。この問題を解決するため、本研究は意志性分類のための弱教師あり学習手法を提案している。まず、「わざと」や「うっかり」といった意志性を表す副詞を頼りに意志性のラベル付きイベントを収集する。分類器の学習時には、手がかり語の副詞を含まないイベントにも汎化するよう、手がかり語の副詞に頼らない予測を促す正則化を導入する。さらに、意志性の必要条件である主語の有生性に着目し、その同時学習を提案している。実験により、提案手法によって、手作業でラベル付き訓練データを整備することなく、高精度な意志性分類器を学習できることを示している。</p> <p>第4章では、談話関係解析の手法を提案している。談話関係解析は、イベント間の意味的関係を認識するタスクである。談話関係解析は、近年、汎用言語モデルが作り出すイベントのベクトル表現の上で分類器を学習することで大きな精度向上を果たした。これは、適切な目的関数を設計すれば、大規模テキストから談話関係解析に資するイベントのベクトル表現が得られる可能性を示唆している。本研究では、談話関係を捉える上ではイベントの些末な詳細を理解することよりも文章を貫く論理の流れが分かる程度に汎化された意味を理解することが有効だと考え、対照学習に基づく非生成的目的関数を通じてイベントのベクトル表現を学習することを提案している。実験では、英語と日本語の標準的データセットにおいて、提案手法によって学習したイベ</p>			

ントのベクトル表現を利用することで談話関係解析の精度が向上することを示している。

第5章では、後続イベント予測の手法を提案している。後続イベント予測は、あるイベントを入力として、そのイベントの次に起こりうるイベントを予測するタスクである。後続イベント予測は、近年、sequence-to-sequence (seq2seq) のタスクとして定式化されることが主流になりつつある。従来研究は、機械翻訳のために提案された seq2seq モデルを適用してきたが、機械翻訳と比べて後続イベント予測は入力に対して妥当な出力が多様であり、通常の seq2seq モデルではこのことを陽に考慮できない。本研究は、後続イベント予測を確率的生成モデルである conditional variational autoencoder (CVAE) によって学習することで、入出力の一对多関係を捉えている。加えて、モデル自身が出力した後続イベントから入力イベントを再構成する機構を併せて学習することで、モデルの多様な出力を促すことを提案している。また、一つの入力に対して複数の妥当な出力が紐付いた評価データを新たに構築し、そのもとでモデルの出力の妥当性と多様性を評価する尺度を提案している。実験を通じて、後続イベント予測を CVAE で学習することにより妥当かつ多様な後続イベントが生成できることを示している。

第6章は結論である。本論文を総括し、今後の課題および展望について述べている。

(続紙 2)

(論文審査の結果の要旨)

本論文は、5W1Hに対応する情報単位であるイベントの粒度の言語解析における基礎的課題に取り組んだものである。得られた主要な成果は以下の通りである。

1. 意志性はイベントの基本的な属性であり、意志性分類には多くの潜在的応用がある。これまで意志性に着目した研究はほとんどなく、容易に利用可能な意志性分類器もそれを学習するための訓練データも存在しなかった。本論文では、意志性分類器を学習可能な弱教師あり学習手法を提案し、日英の両方でその有効性を確認した。

2. 談話関係解析は、イベント間の意味的关系を認識するタスクである。談話関係解析は汎用言語モデルが作り出すイベントのベクトル表現の上で分類器を学習することで大きな精度向上を果たした。このことを受け、本論文では談話関係解析に有効なイベントのベクトル表現を学習するための対照学習に基づく自己教師あり学習フレームワークを提案した。実験では、英語と日本語の標準的データセットのもとで、提案手法が談話関係解析の精度向上に有効であることを確認した。

3. 後続イベント予測は、あるイベントを入力として、そのイベントの次に起こりうるイベントを予測するタスクである。先行研究が機械翻訳のために提案された sequence-to-sequence (seq2seq) モデルを素朴に適用するのに対して、本論文は、後続イベント予測における入出力間の一対多関係を陽に考慮し、conditional variational autoencoder (CVAE) の枠組みで、後続イベント予測を確率的生成の問題として学習することを提案した。加えて、一つの入力に対して複数の妥当な出力が紐付いた評価データを新たに構築し、そのもとでモデルの出力の妥当性と多様性を評価することを提案した。実験を通じて、CVAE は通常の seq2seq モデルと比べて後続イベントの妥当性と多様性をより良く捉えられることを示した。

よって、本論文は博士(情報学)の学位論文として価値あるものと認める。また、令和4年2月24日、論文内容とそれに関連した事項について試問を行った結果、合格と認めた。なお、本論文のインターネットでの全文公開についても支障がないことを確認した。

要旨公開可能日： 年 月 日以降